

Article

Comparative Genome Analysis of 19 *Trueperella pyogenes* Strains Originating from Different Animal Species Reveal a Genetically Diverse Open Pan-Genome

Zoozeal Thakur ^{1,†} , Rajesh Kumar Vaid ^{1,*,†} , Taruna Anand ¹ and Bhupendra Nath Tripathi ^{1,2}

¹ Bacteriology Laboratory, National Centre for Veterinary Type Cultures, ICAR-National Research Centre on Equines, Hisar 125001, India

² Division of Animal Science, Krishi Bhavan, New Delhi 110001, India

* Correspondence: rajesh.vaid@icar.gov.in

† These authors contributed equally to this work.

Abstract: *Trueperella pyogenes* is a Gram-positive opportunistic pathogen that causes severe cases of mastitis, metritis, and pneumonia in a wide range of animals, resulting in significant economic losses. Although little is known about the virulence factors involved in the disease pathogenesis, a comprehensive comparative genome analysis of *T. pyogenes* genomes has not been performed till date. Hence, present investigation was carried out to characterize and compare 19 *T. pyogenes* genomes originating in different geographical origins including the draftgenome of the first Indian origin strain *T. pyogenes* Bu5. Additionally, candidate virulence determinants that could be crucial for their pathogenesis were also detected and analyzed by using various bioinformatics tools. The pan-genome calculations revealed an open pan-genome of *T. pyogenes*. In addition, an inventory of virulence related genes, 190 genomic islands, 31 prophage sequences, and 40 antibiotic resistance genes that could play a significant role in organism's pathogenicity were detected. The core-genome based phylogeny of *T. pyogenes* demonstrates a polyphyletic, host-associated group with a high degree of genomic diversity. The identified core-genome can be further used for screening of drug and vaccine targets. The investigation has provided unique insights into pan-genome, virulome, mobiliome, and resistome of *T. pyogenes* genomes and laid the foundation for future investigations.

Keywords: *Trueperella pyogenes*; comparative genome analysis; pan-genome; pyolysin; virulence factors; phylogeny; antimicrobial resistance



Citation: Thakur, Z.; Vaid, R.K.; Anand, T.; Tripathi, B.N. Comparative Genome Analysis of 19 *Trueperella pyogenes* Strains Originating from Different Animal Species Reveal a Genetically Diverse Open Pan-Genome. *Antibiotics* **2023**, *12*, 24. <https://doi.org/10.3390/antibiotics12010024>

Academic Editors: Renata Urban-Chmiel, Kinga Wieczorek, Marta Dec, Agnieszka Marek and Dagmara Stępień-Pyśniak

Received: 16 November 2022
Revised: 14 December 2022
Accepted: 16 December 2022
Published: 24 December 2022



Copyright: © 2022 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

Trueperella pyogenes, formerly *Arcanobacterium pyogenes*, is a commensal member of skin biota and mucosa of the upper respiratory, gastrointestinal, and urogenital tract of a wide range of animal species including cattle, sheep, and pigs [1]. Infection has also been reported in wild animals [2,3]. This ubiquitously found opportunist bacterium is the etiological agent of abortions and chronic suppurative infections such as mastitis, pyometra, pneumonia, and abscesses leading to significant losses in the livestock industry, especially in intensive systems of animal husbandry [1,4]. The bacterium is mostly observed in multispecies infection with anaerobe *Fusobacterium necrophorum*, which highlights the pathogenic synergy between the two species [5]. Although its importance in animal husbandry is well established, it is being also recognized as playing a role in zoonotic infections in humans [6,7].

The factors that play a pertinent role in the establishment and development of disease includes virulence determinants encoded by pathogens, underlying chronic medical conditions of host, animal husbandry practices, and adverse environmental conditions, among others [1,8]. A few virulence determinants associated with the pathogenic potential have been recognized till date such as pyolysin (*pl*), fimbriae (*fimA*, *fimC*, *fimE*, and *fimJ*),

collagen-binding protein A (*cbpA*), and neuraminidases (*nanH* and *nanP*) [9,10]. The pyolysin (*plo*) gene belonging to the family of cholesterol-dependent cytolysins is the most recognized and characterized virulence factor of *T. pyogenes*. The haemolytic exotoxin protein encoded by the gene is involved in transmembrane pore formation owing to its cytolytic activity. Notably, *plo* is reported to be present in all of the investigated wild type *T. pyogenes* strains until now [11,12]. Other virulence determinants which target host tissue and lead to the persistence of *T. pyogenes* include neuraminidase (*nanH* and *nanP*), extracellular binding proteins (*cbpA*), and fimbrial subunits (*fimA*, *fimC*, *fimE*, and *fimJ*) [11,13]. Notably, the knowledge regarding underlying virulence factors involved in the infection development is sparse [1,14]. In order to decode the genetic basis of pathogenesis as well as the unravelling of mechanisms that act as a switch for conversion of a commensal to a pathogen, and comprehensive analysis of *T. pyogenes* strains at genomic level is the first required step.

The first complete genome sequence of *T. pyogenes* species was reported in 2014 of *T. pyogenes* strain 6375 [15]. Till date, 11 complete genome sequences of *T. pyogenes* isolated from various host species and from different countries have been published. In this comparative genomics study of *T. pyogenes* strains from all over the world, we have also included the first Indian draft genome sequence of *T. pyogenes* species (Bu5 strain), which was isolated from wound infections of water buffalo (*Bubalus bubalis*). Notably, infections due to *T. pyogenes* in water buffalo have been reported from Mediterranean and Middle East countries [16–18]. In India, cases of *T. pyogenes* infection have been reported from small ruminants [19]; however, reports from buffaloes in India have been elusive.

Trueperella pyogenes strains associated with different host species and clinical manifestations have been characterized biochemically, genotypically as well as with different genotype combinations for candidate virulence markers such as *plo*, *fimA*, *fimC*, *fimE*, *fimJ*, *cbpA*, *nanH*, and *nanP* [9,10,20,21]. Pyolysin encoding gene *plo* is detected in all of the pathogenic strains of *T. pyogenes*; however, other candidate virulence determinants such as *nanH*, *nanP*, *cbpA*, and fimbrial genes are not always detected [11,21–23]. The high level of variation in biochemical testing and haemolytic reaction on Sheep Blood Agar, resulting in eight different biotypes of *T. pyogenes* isolated from metritis in cattle, also indicates a need for comparative studies between different strains of *T. pyogenes* [22]. However, a comprehensive comparative analysis describing pan-genome and diversity of virulome, mobilome, as well as resistome of *T. pyogenes* strains has not been reported till date. Although few studies which focussed on fimbrial genes have characterized *T. pyogenes* strain genomes, these studies have utilized a limited number of genes and genomes [9,13,24]. In order to fill the gap, a comparative genome investigation was required to obtain insights of commonalities, dissimilarities, and evolutionary history of the prevalent *T. pyogenes* strains. In fact, identification of new virulence factors significant for pathogenesis could aid in the designing of new strategies for combating of the disease [25,26].

The current study has elucidated pan-genome architecture, which includes core-genome and strain-specific genes in the investigated *T. pyogenes* genomes. In addition, the investigation has also detected diversity in fimbrial genes, candidate virulence genes, genomic islands, prophage sequences, and antibiotic resistance genes via usage of number of computational tools. The identified core-genome can be employed for metabolic profiling as well as for screening of drug or vaccine targets for therapeutic intervention. On the other hand, dissimilar genomic features can be employed as markers for strain characterization, tracking of evolutionary changes, diversity assessment, and determination of transmission history for well-timed detection and epidemiological surveillance. To the best of our knowledge, this is the first comprehensive report of *T. pyogenes* genome analysis encompassing strains from across the world.

2. Materials and Methods

2.1. Isolation, Characterization, and Genome Sequencing of *T. pyogenes* Bu5

The *T. pyogenes* (Bu5) strain was isolated from pyogenic wound infection of a water buffalo (*Bubalus bubalis*) from an organized herd at Hisar, (Haryana), India. The isolate was identified by phenotypic colony characteristics, cell-morphology, biochemical tests, and 16S rRNA sequencing. The isolate was grown on Sheep Blood Agar (SBA) at 37 °C in 5% CO₂ atmosphere. Genomic DNA of *T. pyogenes* Bu5 was extracted using the DNeasy kit (Qiagen), as per the manufacturer's instructions.

The whole genome sequencing of the Bu5 strain was achieved by shotgun sequencing strategy using 454 pyrosequencing and assembled de novo using Newbler version 2.60. The strain *T. pyogenes* Bu5 (Accession No. VTCCBAA267) is available at the Indian Council of Agricultural Research, National Centre for Veterinary Type Cultures (NCVTC), National Research Centre on Equines (NRCE), Hisar, Haryana, India.

2.2. Sequence Information and Quality

The NCBI and the Pathosystems Resource Integration Center (PATRIC) database search for *T. pyogenes* genomes resulted in 24 and 29 entries, respectively, in April, 2021 [27,28]. Nineteen *T. pyogenes* genomes were included in the investigation after exclusion of genomes of low sequence length (GCA_015264715.1, GCA_015264725.1, GCA_015264745.1, GCF_001070855.1, GCA_900299135.1), high number of contigs (GCA_001068695.1), suppressed assembly (GCA_003346995.1), phage genome (GCA_017347915.1), and plasmid sequences (AY255627, U83788) (Table S1). The sequence information of 19 *T. pyogenes* genomes that were isolated from various parts of the world and consisted of 11 complete and 8 good quality draft genomes including an *in-house* draft genome *T. pyogenes* strain Bu5 were retrieved from NCBI database in fasta format (Table S1). The investigated genomes have been reported to be isolates of different livestock hosts such as cattle, goats, water buffalo, and pigs (Table S2a,b). *Trueperella pyogenes* strain TP6375 was utilized as the reference genome in comparative circular plot visualization, pan-genomic analysis as well as in synteny plots. The strain TP6375 was chosen as reference genome in the current investigation due to the following reasons, which include the (i) first complete genome sequence of *T. pyogenes* species, (ii) experimentally utilized for investigation of virulence factors such as fimbrial genes and surface anchored proteins and (iii) also being used as reference genome in the NCBI genome database as well as in previous studies [13,29,30]. The sequence quality information of investigated genomes such as sequencing depth, sequencing platform, assembly method, and other parameters provided by PATRIC is also listed in Table S2a,b.

2.3. Genome Characteristics

The investigated genome sequences were annotated by using the Prokka pipeline available at Usegalaxy server.

(<https://usegalaxy.eu/> accessed on 14 April 2021) [31,32]. Basic genomic features, such as the number of CDS, tRNA, rRNA, tmRNA, and repeat regions, were determined.

2.4. Pan-Genome, Core-Genome, and Strain-Specific Gene Calculations

Pan-genome describes the complete genetic repertoire of the genomes under investigation, whereas core-genome refers to the set of orthologous genes shared by all the genomes under analysis [33]. On the other hand, strain specific genes or singletons are the unique genes possessed by a genome under study and have no other homologs in the rest of the genomes under analysis [34,35]. In order to determine the pan-genome, core-genome, and strain-specific genes of 19 *T. pyogenes* genomes under investigation, EDGAR 3.0 web interface (<https://edgar3.computational.bio.uni-giessen.de> accessed on 15 July 2021) was employed with default settings [36]. EDGAR is a widely used platform for comparative genome analysis that employs the gene orthology criterion which uses BLAST Score Ratio Values (SRVs) for detection for pan- and core-genome with visual representations. Sta-

tistical extrapolation of detected pan- and core-genome is performed by employment of non-linear least-squares curve fitting of the detected core- and pan-genome sizes as a function of the number of investigated genomes in EDGAR [36]. In case of core-genome extrapolation, an exponential decay function as described by Tettelin et al. is used, where c is the amplitude of the function, n is the number of genomes, Ω is the extrapolated size of the core-genome for $n \rightarrow \infty$, and τ is the decay constant representing the speed at which f converges to Ω [37]:

$$f(n) = c \cdot \exp\left(\frac{-n}{\tau}\right) + \Omega$$

However, in pan-genome extrapolations, a Heaps' power law function is employed, where n is the number of investigated genomes, c is a proportionality constant and γ the growth exponent that depicts at which speed the pan-genome is expanding:

$$f(n) = c \cdot n^\gamma$$

A customized project for pan-genome calculations was created by EDGAR on request.

2.5. Functional Annotation

Functional annotation of the representative core-genome subset and complete set of singletons was carried by eggNOG-mapper v2 available at (<http://eggno-mapper.embl.de> accessed on 2 December 2022) with default parameters (minimum hit value: 0.001, minimum hit bit score: 60, percent identity: 40, minimum query coverage: 20%, and minimum subject coverage: 20%). The tool utilizes pre-calculated orthologous Groups (OGs) and phylogenies of the EggNOG database (<http://eggno5.embl.de> accessed on 2 December 2022) to assign functional annotation to an input dataset of nucleotide sequences. The tool provided COG category assignment, gene ontology (GO), E.C number, KO identifier, and PFAM domain information [38,39].

2.6. Average Nucleotide Identity (ANI) Determination

The Average Nucleotide Identity (ANI) is a robust measure of nucleotide-level genomic similarity between two investigated genomes. The estimation of ANI among the investigated *T. pyogenes* genomes as well as the generation of an all versus all comparison matrix was carried at a customized setup at the EDGAR web interface. A clustering tree was also built by utilizing the ANI genome-based distance matrix of the investigated strains [40].

2.7. Phylogenetic Tree Construction

Phylogenetic tree was built by utilizing the core-genome of the investigated *T. pyogenes* strains. Each core gene set was aligned by MUSCLE and subsequently the alignments were joined together to form one huge alignment [41]. Finally, the phylogenetic tree was generated from the alignment by using FastTree software for inference of phylogeny. The FastTree method computes local support values calculated by the Shimodaira–Hasegawa (SH) test as metrics of phylogenetic tree reliability [42].

2.8. Synteny Plot Analysis

Synteny plots were created by EDGAR for depiction of gene order and detection of large-scale genomic rearrangements if present in the ten investigated complete *T. pyogenes* genomes with reference to *T. pyogenes* 6375 genome [36]. Draft genomes were not included in the synteny plot analysis as their inclusion gives rise to erroneous findings.

2.9. Detection of Candidate Virulence Factors (CVFs)

Candidate virulence factors were detected in the investigated 19 *T. pyogenes* genomes by using literature mining, a VFAnalyzer tool, BLASTN searches, and manual curation [43,44] as summarized in Figure 1. Firstly, nucleotide sequence information of known virulence genes,

namely pyolysin (*plo*), collagen-binding protein A (*cbpA*), neuraminidases (*nanH* and *nanP*), and fimbriae (*fimA*, *fimC*, *fimE*, and *fimJ*) were located and extracted from the genomes of *T. pyogenes* by literature mining and manual curation of the sequence information files available at the NCBI genome database (Table S7a–c). Next, homologues of these virulent genes were searched for in the rest of the investigated *T. pyogenes* genomes using BLASTN searches. Subsequently, BLASTN search results were then manually checked and, afterwards, corresponding protein sequences were mapped for hit regions in the GenBank or RefSeq assembly sequence records and extracted in fasta format. Multiple sequence alignment of protein sequences of the investigated virulence genes (*plo*, *cbpA*, *nanH*, *nanP*, *fimA*, *fimC*, *fimE*, and *fimJ*) was carried by NCBI COBALT for further inspection [45].

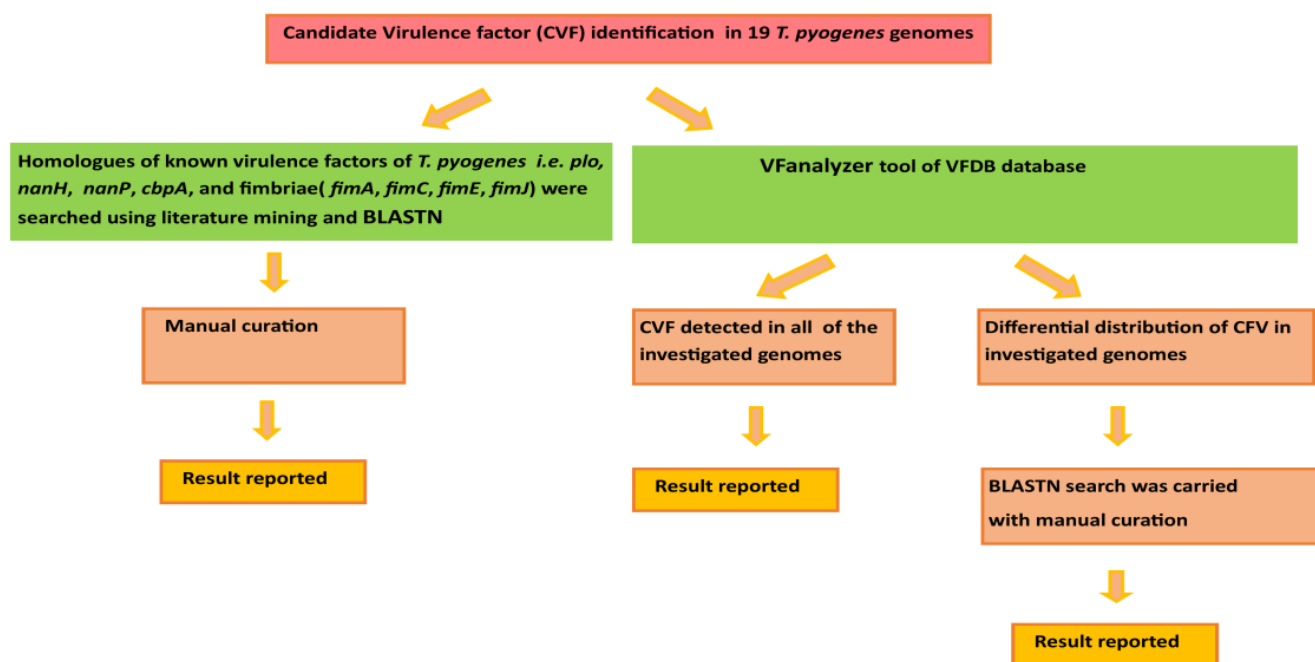


Figure 1. Overview of methodology implemented in identification of candidate virulence factors (CVF's) in nineteen investigated *T. pyogenes* genomes. The first approach involved the detection of homologues of known virulence factors i.e., pyolysin (*plo*), collagen-binding protein A (*cbpA*), neuraminidases (*nanH* and *nanP*), and fimbriae (*fimA*, *fimC*, *fimE*, and *fimJ*) in the investigated genomes via using literature mining, BLASTN searches, and manual curation. The second approach involved utilization of the VFAnalyzer tool of the VFDB database to detect homologues belonging to various classes of virulence factors.

Secondly, the VFAnalyzer tool of the VFDB database was employed for the detection of CVFs related to adherence, iron uptake, regulation, toxins, amino-acid and purine metabolism, anti-apoptosis factor, lipid and fatty acid metabolism, phagosome arresting, protease, and stress adaptation among others. The candidate virulent genes that showed differential presence in the investigated genomes by the VFAnalyzer tool were searched again using BLASTN searches. In order to maintain stringency, BLASTN results were manually curated and hits that showed query coverage $\geq 30\%$ and percent identity $\geq 60\%$ only were considered as homologs.

2.10. Detection of Genomic Islands (GIs)

GIs are the cluster of genes of horizontal origin that are attributed to be the source of genetic diversity and contribute to virulence and evolution [46]. Islandviewer 4 was used for detection of GIs in the studied *T. pyogenes* genomes. Islandviewer 4 uses four different GI detection methodologies i.e., Islander, SIGI-HMM, Islandpath-DIMOB, and Islandpick [47]. However, GIs in draft genomes were predicted by using SIGI-HMM as

well as Islandpath-DIMOB tools by taking *T. pyogenes* 6375 as a reference genome, which was required for contig reordering.

2.11. Detection of Prophages

PHASTER (Phage Search Tool Enhanced Release) available at (www.phaster.ca accessed on 17 July 2021) was used to identify and annotate candidate prophage sequences within the *T. pyogenes* genomes. The tool classified identified prophage sequences into the categories of intact, incomplete, and questionable [48].

2.12. Searching of Antibiotic Resistance Genes (ARGs)

Searching of ARGs was achieved by employing the CARD database available at (<https://card.mcmaster.ca> accessed on 22 July 2021). The CARD database archives' sequences and mutations reported to confer AMR. The resistance gene finder (RGI) tool available at CARD employs detection models on the basis of SNP as well as homology models and subsequently classifies the identified hits in the categories of perfect, strict, and complete category [49].

3. Results

3.1. Isolation, Characterization, and Sequencing of *T. pyogenes* Bu5

The pus sample obtained aseptically from an adult water buffalo was streaked on 5% Sheep Blood Agar (SBA) and incubated at 37 °C in the CO₂ incubator for 24–48 h. After 48 h of incubation, pure, minute haemolytic colonies grew on 5% SBA. *Trueperella pyogenes* strain Bu5 (pus) isolate was Gram-positive cocco-bacilli, non-motile as well as catalase and oxidase negative (Table S2a,b). A total of 217,058 reads of 424 bp were generated using the GS FLX Titanium system (454 Life Sciences Corporation, Branford, CT, USA), giving ~41 × coverage. The data generated 15 large contigs with an average contig size of 148,335 bp and a largest contig size of 741,606 bp. The total size of the genome was 2,225,039 bp, with an N50 of 506,009 bp and a Q40 of 99.95%. This is the first whole genome sequence of *T. pyogenes* isolated from water buffalo (*Bubalus bubalis*) in the world.

3.2. Data Availability

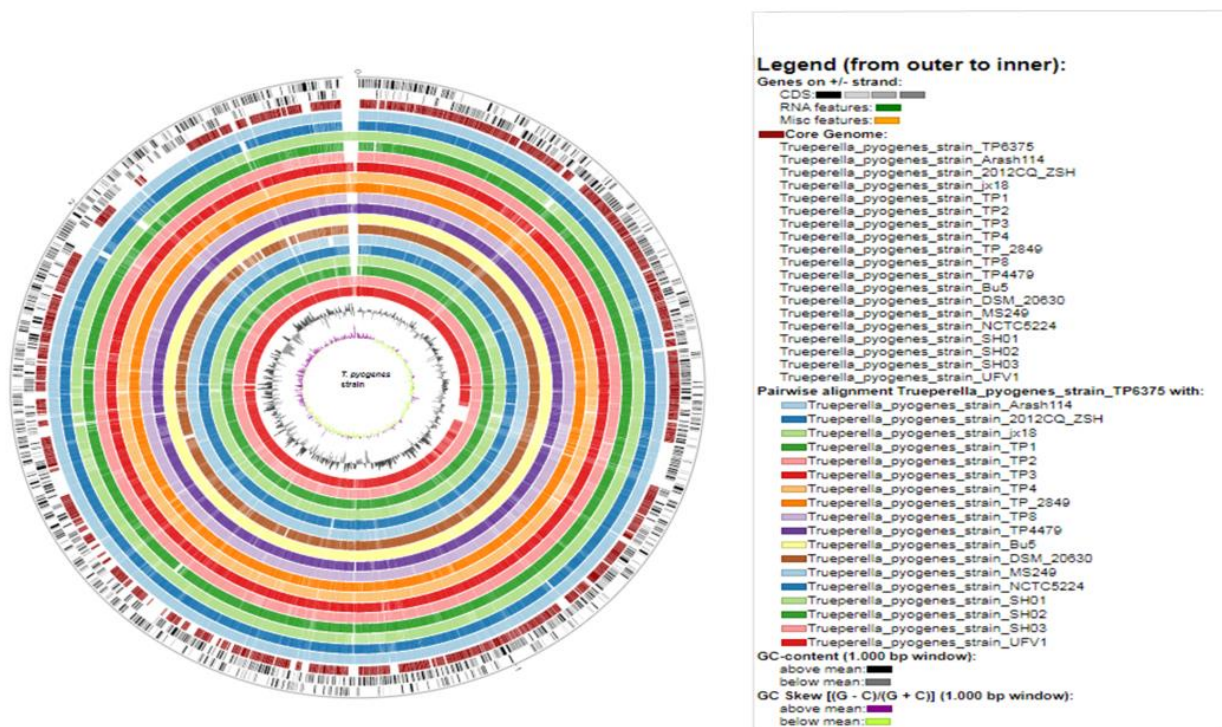
This Whole Genome Shotgun project was deposited at DDBJ/ENA/GenBank under the accession PESV00000000, BioprojectPRJNA416992.

3.3. Comparative Genome Statistics

Basic genomic features of *T. pyogenes* genomes were determined and compared by using the Prokka pipeline as listed in Table 1. The average genome size of the investigated genomes is 2,327,522.5 bp, ranging from 2,187,257 (*T. pyogenes* DSM 20630) to 2,427,168 (*T. pyogenes* TP4). The average GC% is 59.54% ranging from 59.33% of *T. pyogenes* strain jx18 to 59.8% of the *T. pyogenes* strain MS249. Genomic features such as CDS, rRNA, tRNA, tmRNA, and repeat regions were determined by the Prokka pipeline. The average CDS were observed to be 2079.21 with the highest (2180) in *T. pyogenes* jx18 and the lowest (1948) in *T. pyogenes* Bu5. The tRNA were detected to be in the range of 45–51 with 45 possessed by *T. pyogenes* TP8 and 51 harbored by *T. pyogenes* UFV1. All of the genomes were detected to harbor one transfer-messenger RNA (tmRNA). Repeat regions were observed in the range of 1–10, with two being observed in *T. pyogenes* strains (Bu5, UFV1, and SH01) and the highest number (10) possessed by the *T. pyogenes* strain MS249. On the other hand, the rest of the strains harbored only one repeat region in their genome. The circular plot visualization of other investigated 18 *T. pyogenes* genomes with reference to *T. pyogenes* str. TP6375 depicts varied GC content and GC skew along with core-region similarity (Figure 2).

Table 1. Basic genomic features of the investigated *T. pyogenes* genomes such as isolation country, bases, number of CDS, rRNA, tRNA, tmRNA, and repeat regions.

S.No	Genome and Strain	Host	Country	Bases	GC	CDS	rRNA	tRNA	tmRNA	RR
1	<i>T. pyogenes</i> 2012CQ-ZSH	Goat, <i>Capra hircus</i>	China	2295822	59.67	2045	6	46	1	-
2	<i>T. pyogenes</i> Arash114	Water Buffalo, <i>Bubalus bubalis</i>	Iran	2338282	59.49	2109	6	46	1	1
3	<i>T. pyogenes</i> jx18	Pig, <i>Sus scrofa</i>	China	2415007	59.33	2180	9	46	1	1
4	<i>T. pyogenes</i> TP1	Cow, <i>Bos taurus</i>	China	2332403	59.76	2126	9	46	1	1
5	<i>T. pyogenes</i> TP2	Cow, <i>Bos taurus</i>	China	2245225	59.68	1993	9	46	1	1
6	<i>T. pyogenes</i> TP3	Pig, <i>Sus scrofa</i>	China	2384650	59.35	2112	9	46	1	1
7	<i>T. pyogenes</i> TP4	Pig, <i>Sus scrofa</i>	China	2427168	59.43	2169	9	47	1	1
8	<i>T. pyogenes</i> TP8	Musk deer, <i>Moschus berezovskii</i>	China	2272494	59.58	2069	3	45	1	1
9	<i>T. pyogenes</i> TP6375	Cow, <i>Bos taurus</i>	USA	2338390	59.5	2100	6	46	1	1
10	<i>T. pyogenes</i> TP4479	Pig, <i>Sus scrofa</i>	China	2382253	59.35	2114	9	46	1	1
11	<i>T. pyogenes</i> TP-2849	Pig, <i>Sus scrofa</i>	China	2384672	59.35	2113	9	46	1	1
12	<i>T. pyogenes</i> Bu5	Water Buffalo, <i>Bubalus bubalis</i>	India	2218921	59.66	1948	3	46	1	2
13	<i>T. pyogenes</i> MS249	Cow, <i>Bos taurus</i>	UK	2216617	59.8	1984	3	46	1	10
14	<i>T. pyogenes</i> UFV1	Cow, <i>Bos taurus</i>	Brazil	2407507	59.75	2149	2	51	1	2
15	<i>T. pyogenes</i> NCTC5224	Pig, <i>Sus scrofa</i>	-	2310711	59.57	2073	9	48	1	1
16	<i>T. pyogenes</i> SH02	Pig, <i>Sus scrofa</i>	China	2380432	59.49	2116	5	46	1	1
17	<i>T. pyogenes</i> SH03	Pig, <i>Sus scrofa</i>	China	2350892	59.58	2079	7	51	1	1
18	<i>T. pyogenes</i> SH01	Pig, <i>Sus scrofa</i>	China	2334225	59.49	2068	3	46	1	2
19	<i>T. pyogenes</i> DSM 20630	Pig, <i>Sus scrofa</i>	-	2187257	59.49	1958	9	45	1	1

**Figure 2.** Circular representation of nineteen *T. pyogenes* genomes with reference genome as *T. pyogenes* TP6375. From outside to inside, the figure depicts CDS, core-genome, pairwise alignment of investigated *T. pyogenes* strains against reference genome *T. pyogenes* TP6375, GC content, and GC skew in different colors.

3.4. Pan-Genome Calculations Reveals Open Pan-Genome and Strain-Specific Genes

The pan-genome calculated by the EDGAR 3.0 platform provided us with the entire gene repository of the investigated *T. pyogenes* genomes, which included the core-genome and accessory genome (Figure 3). The pan-genome repertoire of investigated *T. pyogenes* consists of 3214 CDS including a core-genome of 1520 CDS, dispensable genome of 1093 CDS and strain-specific genes in the range of 2–63 (Tables S3–S5). Significantly, a total of 307 CDS in the range of 2–63 were detected as strain-specific genes, also known as singletons in the analyzed genomes (Table S4). Notably, *T. pyogenes* MS249 harbored the highest number (63) of singletons. There were four strains of *T. pyogenes*, viz., strains TP3, TP6375, TP 4479, and TP 2849 without any unique gene detected in their genomes (Table S4). The list of CDS identified as components of pan- and core-genome along with strain-specific genes with their name and function are provided in Tables S3–S5.

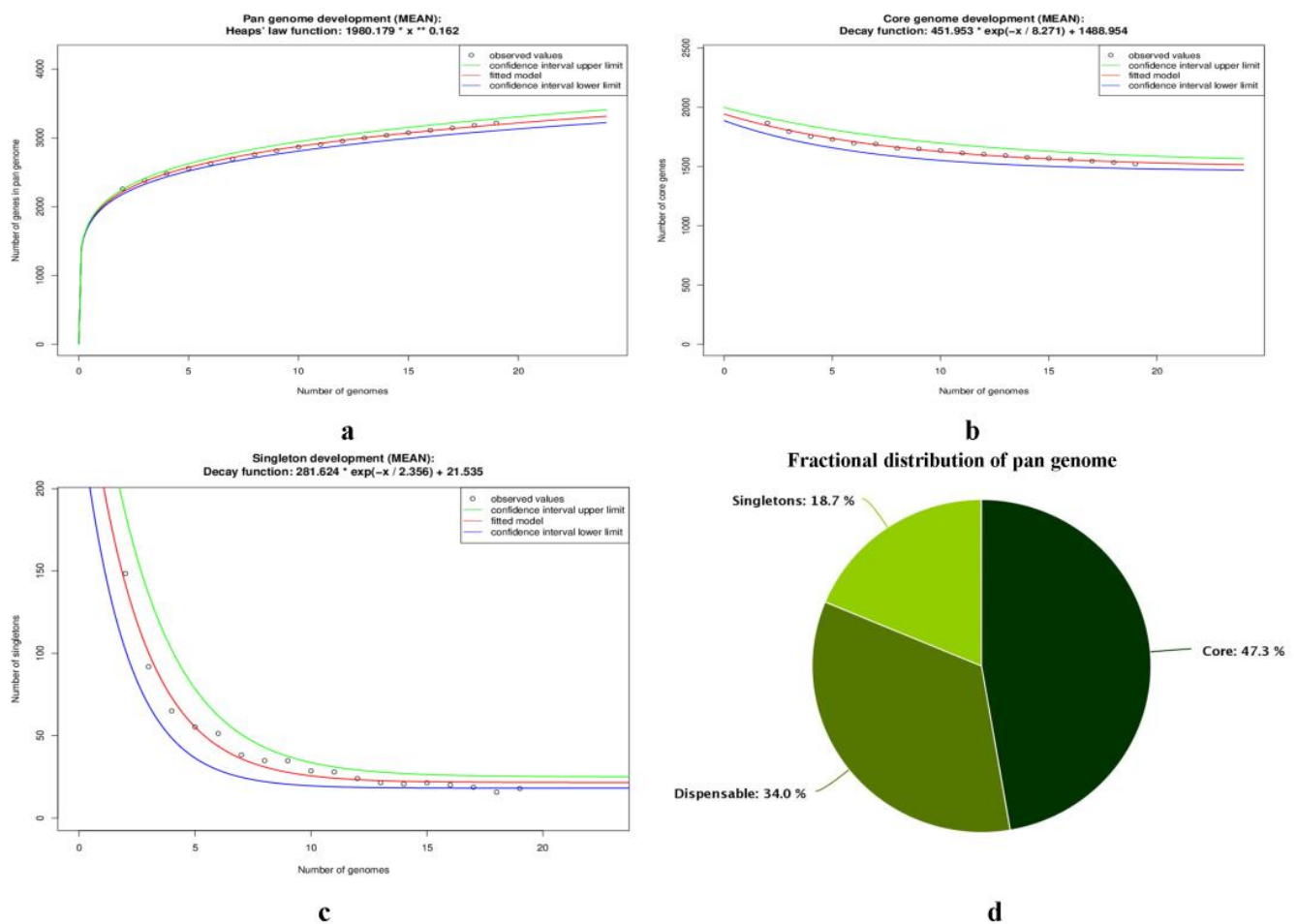


Figure 3. (a) The pan-genome development plot projections of nineteen investigated *T. pyogenes* genomes. The red curve illustrates the fitted exponential Heaps' law function; blue and green curves represent the upper and lower boundary of the 95% confidence interval; (b) the core-genome development projections of studied *T. pyogenes* genomes. The red curve represents the fitted exponential decay function; blue and green curves depict the upper and lower boundary of the 95% confidence interval; (c) singleton development plot projections of examined *T. pyogenes* genomes; (d) fractional distribution of pan-genome of nineteen investigated *T. pyogenes* genomes. * denotes multiplication operator, ** denotes exponentiation operator.

The pan-genome development plot depicted steady growth with adding up of each new genome and reached 3215 on addition of 19th genome (Figure 3a; Table S1). However, a core-genome development plot converged to 1488 genes, as with the addition of new genomes, a decrease in shared genes was observed (Figure 3, Table S1). On the other hand,

a singleton development plot depicted that 21 new genes could be found with the adding up of each new genome (Figure 3). The open pan-genome state of *T. pyogenes* species can be inferred by the analysis of pan-genome development plot (growth exponent value of 0.162 (95% confidence interval 0.157 to 0.167)), core-genome development plot, and the singleton genome development plot of investigated strains (Figure 3).

3.5. Functional Annotation of Core-Genome and Strain-Specific Gene Repertoire

Functional annotation of representative core-genome subset and complete set of strain-specific genes i.e., singletons, was carried by eggNOG-mapperv2 on the basis of pre-computed orthology assignment. The tool provided functional annotation, orthology assignment, and domain prediction, which included COG category, gene ontology (GO), E.C number, KO identifier, PFAMID, and general description, among others, for each of the analyzed CDS (Table S6a,b). 1392 CDS (91.57%) out of the total 1520 CDS that were part of the representative core-genome were queried by the Egnog mapper tool (Table S6a,b). The maximum number (220) of CDS of core-genome fell into the category of function unknown.

The functional classes that were enriched with the most CDS were carbohydrate metabolism and transport (139), translation (135), inorganic ion transport and metabolism (120), transcription (115), amino acid metabolism and transport (113), energy production and conversion (93), replication and repair (85), and coenzyme metabolism (74), among others (Figure 4A). On the other hand, 97 CDS (31.29%) out of the total 310 CDS of complete subset of singletons were queried by the tool. The repertoire of singletons that was queried by the tool fell into the categories of, replication and repair (23), function unknown (16), defence mechanism (12), transcription (10), and no functional class assigned (20) (Table S6a,b). A stacked bar graph representing functional roles of the genes unique to individual strain are depicted in Figure 4B and Table S6a,b.

3.6. Phylogenetic and Synteny Plot Analysis

A phylogenetic tree based on 1520 core genes was constructed to decipher the evolutionary relationship amongst the analyzed nineteen *T. pyogenes* genomes. The phylogenetic tree generated depicted three major divergent clades (Figure 5). The Clade I and clade II consist of only large ruminant isolates. Clade I has *T. pyogenes* strains Arash114 and MS249 with the origin from Iran and the UK, respectively. The clade II consisted of Chinese origin cattle strains TP1 and TP2. Notably, the majority (14) of strains were observed in clade III, which bifurcated into two clusters i.e., cattle strains TP6375 (USA) and UFV1 (Brazil), forming one small group and a main large cluster consisting of Asian origin strains (China (11), unknown (2) and India (Bu5)). The Bu5 strain of buffalo origin also makes a separate subgroup with high support, whereas the other subgroup is mainly dominated by the rest of the 10 Chinese origin porcine strains (Figure 5). The SH branch support values of core-genome based phylogenetic tree were very good in general, with only one value of 0.352 below the maximum of 1.00.

In order to compare the investigated complete genomes against reference strain *T. pyogenes* 6375 and visualize genome scale rearrangements, synteny plots were constructed by utilizing the EDGAR web interface. The synteny plots depicted synteny and large-scale genomic rearrangements such as inversion, duplication, relocation, and palindrome (Figure 6). Notably, the 2012CQ-ZSH genome showed the highest collinearity with the reference genome *T. pyogenes* 6375. Large inversion events were observed in the synteny visualization of 2012CQ-ZSH, TP3, TP8, TP2849, and TP 4479 against the reference genome (Figure S1A,F,H-J). In addition, relocation and duplication were also observed in the synteny plots of TP3 and TP8 genomes against the reference genome (Figure S1F,H). In comparison, smaller inversions along with deletions and duplications were observed in the synteny plots of jx18, TP1, and TP4 against the reference strain (Figure S1C,D,G). In particular, a large palindrome with duplication along with high synteny was also observed in the TP2 synteny plot against the reference genome (Figure S1E). The synteny plots of

each of the analyzed complete genomes against the reference genome *T. pyogenes* 6375 are depicted in Figure S1A–J.

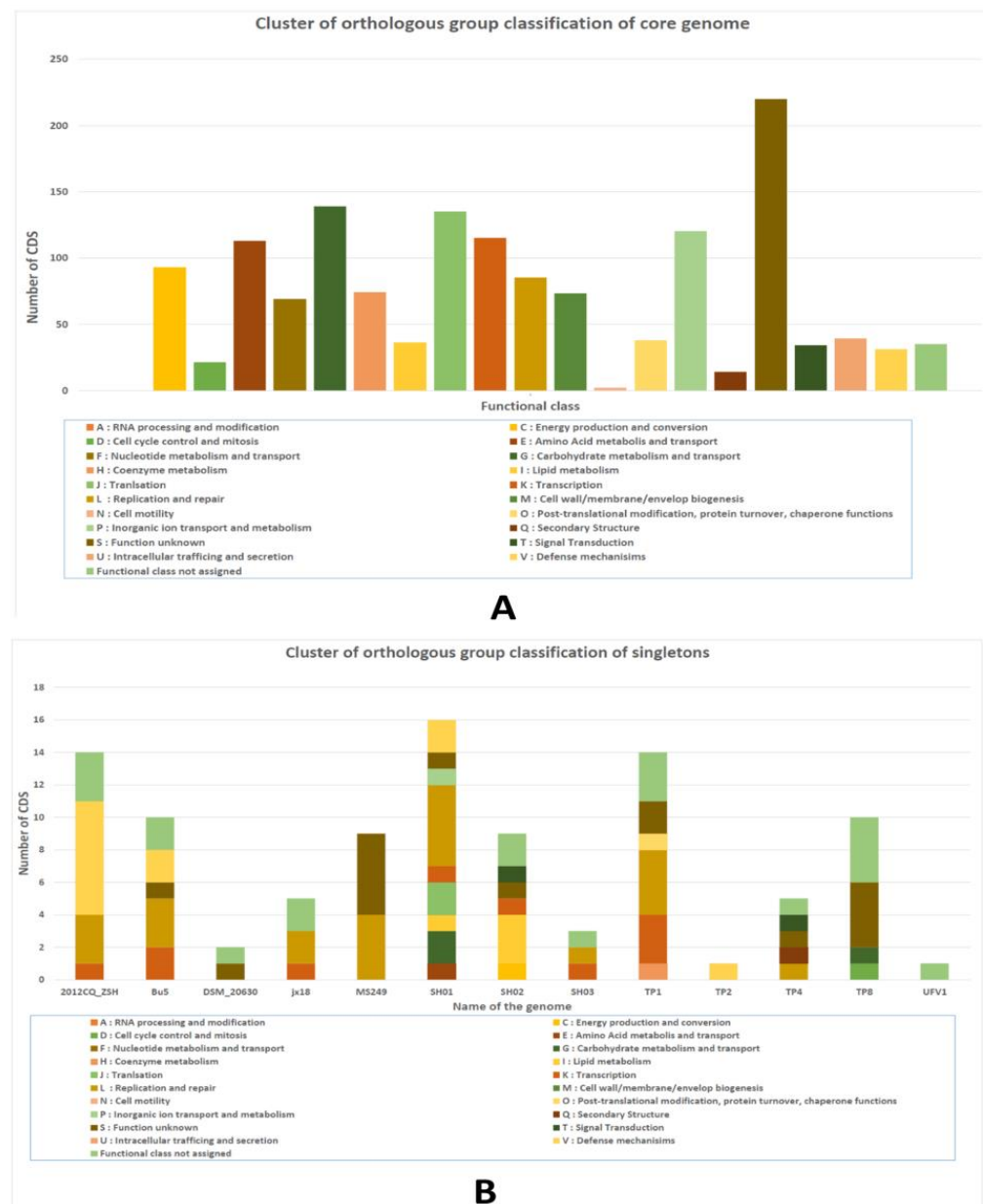


Figure 4. (A) COG functional classification of a representative core genome of nineteen investigated *T. pyogenes* genomes as obtained by eggNOG-mapper v2. The x-axis denotes the number of CDS, and the y-axis represents functional classes. The maximum number of CDS fell in the function unknown category (220 CDS) followed by carbohydrate metabolism and transport category (139 CDS); (B) COG functional classification of singletons detected in the investigated *T. pyogenes* genomes by eggNOG-mapper v2. 297 CDS (31.29%) out of the total 310 CDS of a complete subset of singletons were queried by the eggNOG-mapper v2. The x-axis denotes the number of CDS, and the y-axis represents the name of the genome.

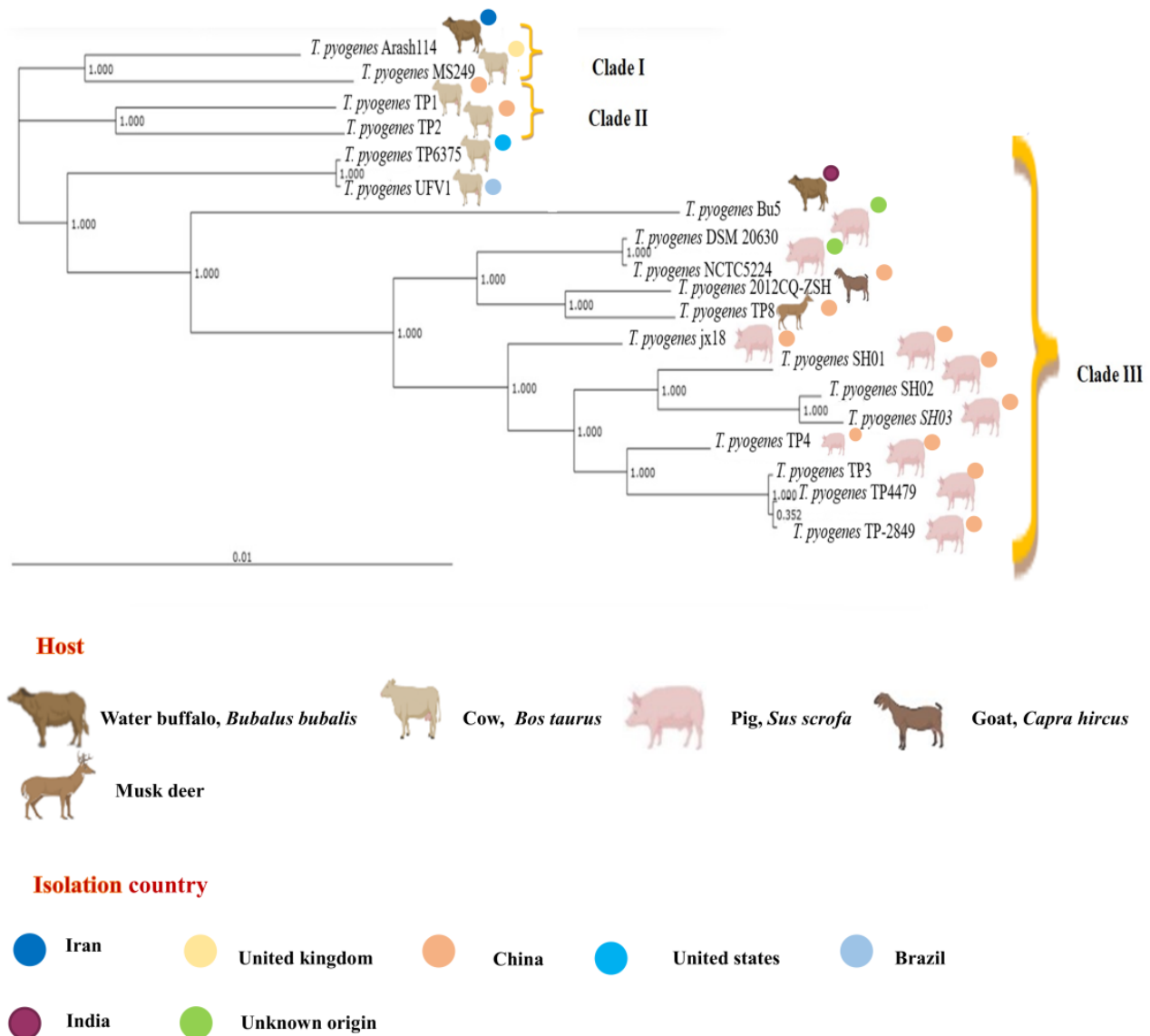


Figure 5. Phylogenetic tree constructed on the basis of core-genome of 19 investigated *T. pyogenes* genomes by EDGAR. The hosts associated with the strains and isolation countries are depicted. The phylogenetic tree can be divided into three clades i.e., clade I, clade II, and clade III, wherein clade III is harboring the maximum number (11) of strains. The majority of the internal branches have maximum branch support value.

3.7. Decoding of Virulome: Candidate Virulence Genes Identified

An inventory of potential virulence factors was identified by the combinatorial usage of the VFAnalyzer tool, BLASTN searches, literature mining, and manual curation in the investigated *T. pyogenes* genomes.

Firstly, we searched and analyzed homologues of principal virulence factors pyolysin (*plo*) along with other putative virulence factors such as neuraminidases (*nanH* and *nanP*), extracellular matrix-binding proteins (*cbpA*), and fimbriae (*fimA*, *fimC*, *fimE*, and *fimJ*), which are involved in adherence and colonization of the host tissue in the studied *T. pyogenes* strains by employing literature mining, BLASTN searches, and manual curation. The information related to BLASTN search results of detected homologues, which includes reference sequence used, query coverage, percent identity, start position, and end position, is tabulated in Table S7a–c. The homologs of cholesterol dependent pyolysin gene (*plo*) of *T. pyogenes* 6375 were observed to be part of the core-genome of all of the investigated

T. pyogenes genomes with a relatively high sequence identity in the range of 80.8% in Bu5 to 100% in UFV1 (Table S7a–c).

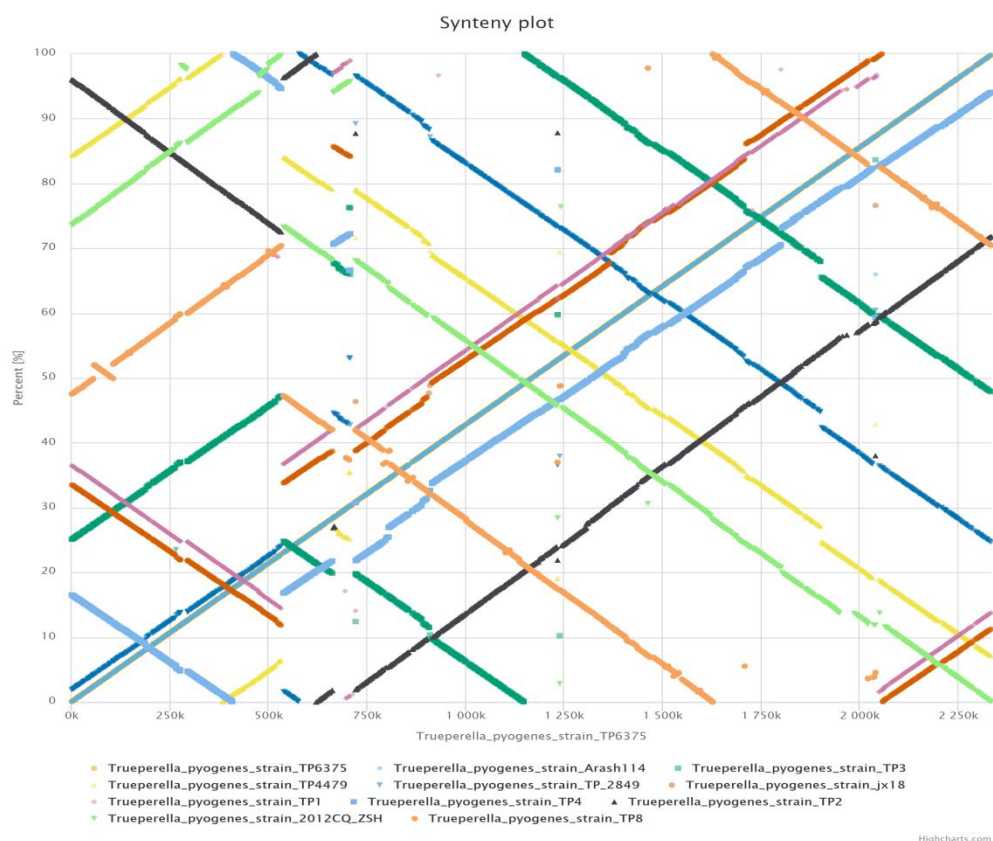


Figure 6. Synteny plot visualization of 10 complete *T. pyogenes* genomes compared against reference genome *T. pyogenes* TP6375 by the EDGAR webserver.

On the other hand, homologues of *T. pyogenes cbpA* were located and analyzed in all of the investigated *T. pyogenes* strains and that varied in the range of 53% QC and sequence similarity 97.66% in the SH01 strain to 100% QC and sequence similarity 96.21% in TP1 (Table S7a–c). The *cbpA* homologues in strains such as TP2, MS249, UFV1, and SH01 were observed to be truncated as depicted by the multiple sequence alignment of their corresponding protein sequences. (Table S7a–c, Figure S2B). Gene truncation at N terminal was observed in strains MS249, UFV1, and SH01, whilst C terminal truncation was observed in *cbpA* of TP2.

Similarly, homologues of *T. pyogenes nanH* were identified in all of the studied genomes with varying sequence identity in the range of 100% QC and sequence similarity 87.62% in TP2 to 78% QC and 87.59% sequence similarity in UFV1. Multiple sequence alignment of protein homologues of *nanH* depicted missing N terminal fragments in TP6375, jx18, DSM20630, SH01, SH03, 2012CQ-ZSH, and UFV1, and a missing C terminal fragment with gaps in reference protein, TP2, TP8, MS249, Bu5, and NCTC5224 (Figure S2C). However, both N terminal and C terminal fragments were observed to be missing in the UFV1 strain. On the other hand, *T. pyogenes nanP* homologues were identified in only 12 out of the 19 investigated *T. pyogenes* genomes with varying sequence identity in the range of 100% QC and 99.37% sequence similarity in Bu5 to 55% QC and 98.62% sequence similarity in MS249. Multiple sequence alignment depicted *nanP* homologues to be highly conserved with the exception of missing N terminal and C terminal fragment in TP8 and missing C terminal fragment in MS249 (Figure S2D).

Next, we searched the homologues of TP6375 fimbrial genes i.e., *fimA*, *fimC*, *fimE*, and *fimJ* in the *T. pyogenes* genomes. Amongst the four TP6375 fimbrial genes investigated, *fimC* and *fimE* were detected to be highly conserved in all of the investigated strains.

The homologues of TP6375 *fimA* were observed to be highly conserved in most of the investigated *T. pyogenes* strains, with the exception of few strains wherein truncation of *fimA* was observed, which includes strains 2012CQ-ZSH, DSM20630, NCTC5224, and UFV1 (Table S7a–c; Figure S2E). The homologues of TP6375 *fimA* varied in the range of QC 21% and sequence similarity 95.92 in DSM20630 and NCTC524 to QC 100% and sequence similarity 98.99% in TP1 (Table S7a–c). Truncation of C terminal fragment along with a gap was observed in the *fimA* of TP6375 and UFV1 and truncation of N terminal fragment was found in NCTC5224, DSM20630, and 2012CQ-ZSH in the multiple sequence alignment of *fimA* protein homologues (Figure S2E). Notably, homologues of TP6375 *fimC* were observed to be conserved in all of the investigated *T. pyogenes* genomes with sequence identity ranging from QC 99% and 70.01% sequence similarity in TP3, TP4479, and TP2849 to 100% QC and 100% sequence similarity in UFV1 (Table S7a–c; Figure S2F). Similarly, homologues of *fimE* TP6375 were observed in all of the investigated *T. pyogenes* genomes with sequence coverage and sequence identity in the range of sequence coverage of 87% and sequence similarity of 74.2% in MS249 to sequence coverage and sequence similarity of 100% in UFV1 (Table S7a–c; Figure S2G). On the contrary, homologues of TP6375 *fimJ* were observed in all of the investigated genomes with the exception of UFV1 and Bu5. The homologues varied in the range of query coverage 17% and sequence similarity 99.26% in SH01 and 2012CQ-ZSH to sequence coverage 100% and sequence identity 99.06% in Arash114 (Table S7a–c; Figure S2H).

Apart from searching and analysis of homologues of *plo*, *fimA*, *fimC*, *fimE*, *fimJ*, *cbpA*, *nanH*, and *nanP*, we also utilized the VFAnalyzer tool to search other candidate virulence genes (Figure 1). The VFAnalyzer tool of VFDB along with BLASTN searches and manual curation detected 500 potential virulence genes that were observed to be homologues of known virulence genes of bacterial pathogens belonging to genera such as *Mycobacterium*, *Klebsiella*, *Haemophilus*, and *Francisella*, among others, as detailed in Table S7a–c. The potential virulence factors identified in the study appear to be associated with a wide range of virulence related functions such as adherence, iron uptake, regulation, toxin, and amino-acid and purine metabolism, anti-apoptosis factor, lipid and fatty acid metabolism, phagosome arresting, biofilm formation, protease, and stress adaptation, among others (Figures 7 and 8, Table S7a–c). The current investigation revealed the presence of homologues of genes related to iron uptake and the siderophore biosynthesis system i.e., *ciuA*, *ciuB*, *ciuC*, and *ciuD* along with a homolog of ABC-type heme transporter related gene *hmuU* in all of the investigated *T. pyogenes* genomes. Additionally, homologues of mycobacterial regulatory proteins such as *relA*, *regX3*, *sigH*, and *sigA/rpoV* were also observed in each of the investigated genomes. Similarly, homologues of mycobacterial genes related to lysine synthesis (*lysA*), glutamine synthesis (*glnA1*), anti-apoptosis factor (*nuoG*), protease (*mpa*, *zmp1*), stress adaptation (*sodA*), and anti-phagocytosis (*rmlA*) were also found by the usage of a VFAnalyzer tool, NCBI BLASTN searches, and manual curation in all of the genomes under study. In a similar manner, homologues of genes related to the secretion system (T6SS-II from *Klebsiella*), capsule (*rmlB* from *Streptococcus*), and pyrimidine biosynthesis (*Francisella*) were also detected in all of the investigated genomes. On the contrary, genes related to adherence (*srtB*), ABC transporter (*fagC*), ABC-type heme transporter (*hmuV*), regulation (*phoP*, *whiB3*), lysine synthesis (*lysA*), pantothenate synthesis (*panC*, *panD*), capsule (*gnd*), trehalose-recycling ABC transporter (*sugC*), exopolysaccharide (*galE*), pyrimidine biosynthesis (*carB*), and exopolysaccharide (*galE* and *pgi*) showed differential distribution in the investigated genomes (Figures 7 and 8, Table S7a–c).

VFclass	Virulence factors	Related genes	2012CQ-ZSH	Arash114	js18	TP1	TP2	TP3	TP4	TP8	TP4479	TP6375	TP-2849	Ba5	DSM 20630	MS249	NCTC5224	SH01	SH02	SH03	UFV1		
Adherence	Collagen-binding protein	<i>cbpA</i>																					
	SpaA-type pili	<i>spaA, spaB, spaC, srtA</i>																					
		<i>spaD, spaE, spaF</i>																					
	SpaD-type pili	<i>srtB</i>																					
		<i>srtC</i>																					
SpaH-type pili	<i>spaG, spaH, spaI, srtD, srtE</i>																						
	Surface-anchored pilus proteins	<i>sapA, sapD, sapE</i>																					
Iron uptake	ABC transporter	<i>fagA, fagB</i>																					
		<i>fagC</i>																					
		<i>fagD</i>																					
	ABC-type heme transporter	<i>hmuT</i>																					
		<i>hmuU</i>																					
		<i>hmuV</i>																					
	Ciu iron uptake and siderophore biosynthesis system	<i>ciuA</i>																					
		<i>ciuB</i>																					
<i>ciuC</i>																							
<i>ciuD</i>																							
Siderophore-dependent iron uptake system	<i>irp6A, irp6B, irp6C</i>																						
	Diphtheria toxin repressor DtxR	<i>dtxR</i>																					
Regulation	(p)ppGpp synthesis and hydrolysis (Mycobacterium)	<i>relA</i>																					
	PhoP/R(Mycobacterium)	<i>phoP</i>																					
	RegX3(Mycobacterium)	<i>regX3</i>																					
	Sigma H(Mycobacterium)	<i>sigH</i>																					
	Sigma A(Mycobacterium)	<i>sigA/rpoV</i>																					
	WhiB3(Mycobacterium)	<i>whiB3</i>																					

Yellow coloured boxes represents homologues detected by VFAnalyzer tool of VFDB database.
 Green coloured boxes represent homologues detected by BLASTN searches and manual curation.
 White coloured boxes represents no homologue detected by VFAnalyzer tool as well as by BLASTN searches and manual curation.

Figure 7. Presence/absence profile of virulence genes homologues aka candidate virulence factors (CVF's) across the *T. pyogenes* genomes by utilization of the VFAnalyzer tool of the VFDB database, BLASTN searches, and manual curation. Yellow colored boxes represent CVF detected by the VFAnalyzer tool of the VFDB database. Green colored boxes represent homologues detected by BLASTN searches and manual curation. White colored boxes represent no homologue detected by VFAnalyzer tool or BLASTN search.

VFclass	Virulence factors	Related genes	2012CQ-ZSH	Arash114	js18	TP1	TP2	TP3	TP4	TP8	TP4479	TP6375	TP-2849	Ba5	DSM 20630	MS249	NCTC5224	SH01	SH02	SH03	UFV1	
Toxin	Diphtheria toxin (DT)	<i>tox</i>																				
	Phospholipase D	<i>pld</i>																				
Amino acid and purine metabolism	Lysine synthesis(Mycobacterium)	<i>lysA</i>																				
	Glutamine synthesis(Mycobacterium)	<i>glnA1</i>																				
Anti-apoptosis factor	NuoG(Mycobacterium)	<i>nuoG</i>																				
Lipid and fatty acid metabolism	Pantothenate synthesis(Mycobacterium)	<i>panC</i>																				
		<i>panD</i>																				
Phagosome arresting	Nucleoside diphosphate kinase(Mycobacterium)	<i>ndk</i>																				
Protease	Proteasome-associated proteins(Mycobacterium)	<i>mpa</i>																				
		<i>zmp1</i>																				
Stress adaptation	Iron-cofactored SOD(Mycobacterium)	<i>sodA</i>																				
Antiphagocytosis	Capsule(Klebsiella)	<i>gnd</i>																				
Cell surface components	Trehalose-recycling ABC transporter (Mycobacterium)	<i>sugC</i>																				
Immune evasion	Exopolysaccharide(Haemophilus)	<i>galE</i>																				
Nutritional virulence	Pyrimidine biosynthesis(Francisella)	<i>carB</i>																				
Secretion system	T6SS-II(Klebsiella)																					
Antiphagocytosis	GPL locus(Mycobacterium)	<i>rmlA</i>																				
Cell surface components	Capsule(Streptococcus)	<i>rmlB</i>																				
Immune evasion	Exopolysaccharide(Haemophilus)	<i>galE</i>																				
Nutritional virulence	Pyrimidine biosynthesis(Francisella)	<i>pgi</i>																				

Yellow coloured boxes represents homologues detected by VFAnalyzer tool of VFDB database.
 Green coloured boxes represent homologues detected by BLASTN searches and manual curation.
 White coloured boxes represents no homologue detected by VFAnalyzer tool as well as by BLASTN searches and manual curation.

Figure 8. Distribution of CVFs in the nineteen investigated *T. pyogenes* strains belonging to functional classes of toxin, amino acid and purine metabolism, anti-apoptosis factor, lipid and fatty acid metabolism, phagosome arresting, protease, and stress adaptation, anti-phagocytosis, cell surface components, immune evasion, nutritional virulence, and secretion system.

3.8. Genomic Island Detection

A total of 206 GIs were detected in the investigated *T. pyogenes* genomes. The detected genomic islands varied in the range of 14–25 and 4.00–82.09 kb in terms of number and size, respectively. The highest numbers (25) of GIs were found in *T. pyogenes* SH02, and the largest sized GI (82.093 kb) was observed in *T. pyogenes* Arash114. However, the lowest number (12) of GIs were detected in *T. pyogenes* TP8 and the smallest size (4 kb) GI in *T. pyogenes* TP2 (Figure 9a). Data related to identified GI's such as starting position, end position, and size are listed in Table S8a–s.

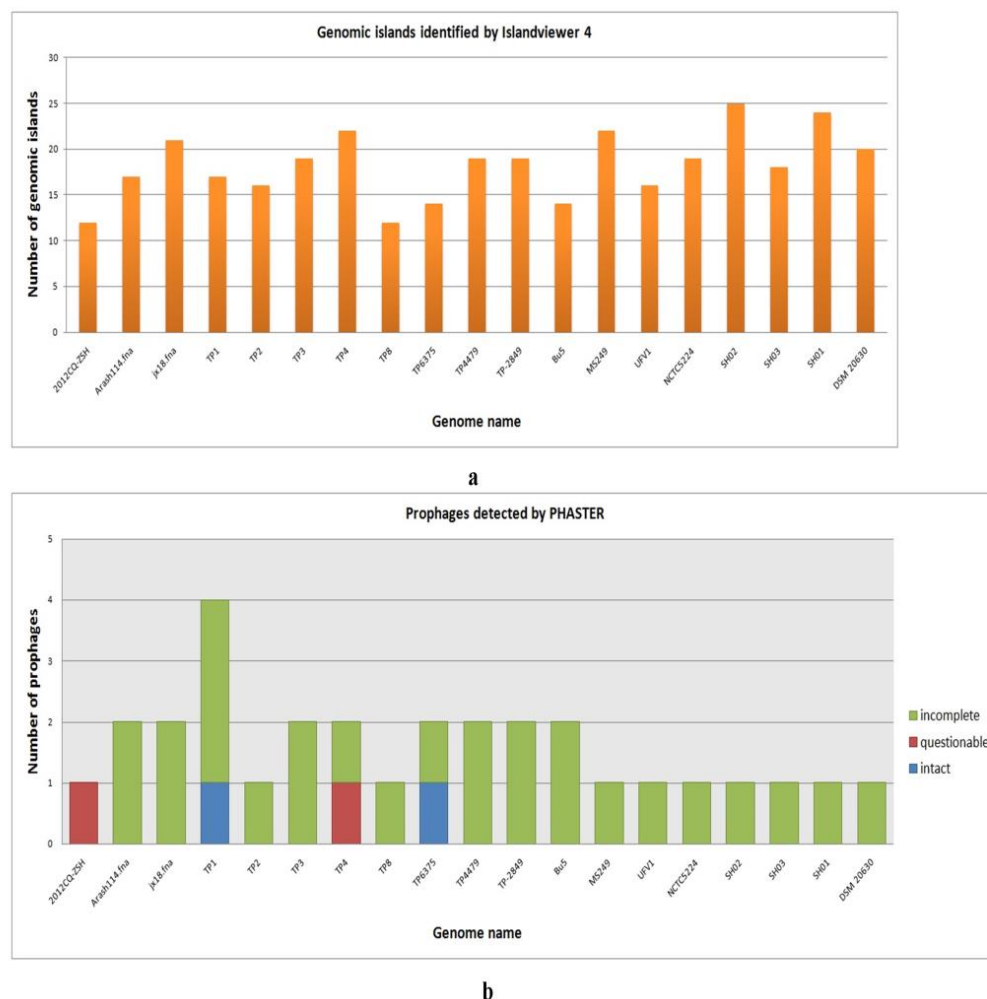


Figure 9. (a) Prevalence of genomic islands (GIs) in the investigated 19 *T. pyogenes* genomes as detected by Islandviewer4. The strains harbored GIs in the range of 12–25 with the maximum observed in strain SH01; (b) number of prophages detected by PHASTER in the investigated 19 *T. pyogenes* strains. The identified prophages were further classified into intact, incomplete, and questionable categories as depicted by blue, green, and red colors, respectively. Prophages were harbored in the range of 1–4 in the investigated strains with maximum prophages being observed in *T. pyogenes* strain TP1.

3.9. Prophage Detection

A total of 30 prophage sequences were detected in all of the investigated *T. pyogenes* genomes in which two were classified as intact, 26 as incomplete and two as questionable by PHASTER (Table S9). The identified prophage sequences varied in the range of 1–4 and 5.2–47.7 kb in terms of number and size, respectively, in the investigated genomes. The GC% of identified prophages varied in the range of 52.20–64.49%. The average size of identified prophages is 19.79 kb. The highest number (4) of prophage sequences were found in the *T. pyogenes* TP1 genome in which three were classified as incomplete prophage

sequences and one as an intact prophage sequence. Only one prophage sequence was detected in the genomes of *T. pyogenes* strains 2012CQ-ZSH, TP2, TP8, DSM 20630, MS249, SH01, SH02, SH03, UFV1, and NCTC5224 by PHASTER, and all of them were classified in incomplete categories (Figure 9b). The detailed information of each identified prophage sequence such as region length, completeness, score, region position, most common phage and GC% are listed in Table S9.

3.10. Antibiotic Resistance Genes (ARG) Detection

A total of 40 ARGs were detected in the investigated genomes by CARD in which 35 were classified into strict and five into perfect category according to RGI criteria (Figure 10). The ARGs which were detected mainly conferred resistance against aminoglycosides (*APH(3′)*-Ia, *APH(6)*-Id, *rmtB*, *ANT(3′)*-Ia, *ANT(2′)*-Ia) tetracyclines (*tet(W/N/W)*, *TetZ*, *Tet33*), phenicols (*cmlA6*), sulphonamides (*sul1*), and streptogramin, macrolides, and lincosamide antibiotics (*ermX*) apart from disinfectants and antiseptics (*qacE delta1*).

Name of the genome	ARO term											
Strain name	ErmX	tet(W/N/W)	ANT(2′)-Ia	qacEdelta1	aadA	sul1	Mtb rpsL mutations conferring resistance to Streptomycin	tet(33)	tet(Z)	cmlA6	APH(6)-Id	APH(3′)-Ia
<i>T. pyogenes</i> 2012CQ-ZSH	Yellow											
<i>T. pyogenes</i> Arash114		Yellow										
<i>T. pyogenes</i> jx18	Yellow											
<i>T. pyogenes</i> TP2							Yellow					
<i>T. pyogenes</i> TP1		Yellow	Green	Green	Yellow	Yellow						
<i>T. pyogenes</i> TP3	Yellow	Yellow						Yellow				
<i>T. pyogenes</i> TP4	Yellow	Yellow										
<i>T. pyogenes</i> TP8							Yellow					
<i>T. pyogenes</i> TP6375		Yellow										
<i>T. pyogenes</i> TP4479	Yellow	Yellow						Yellow				
<i>T. pyogenes</i> TP-2849	Yellow	Yellow						Yellow				
<i>T. pyogenes</i> Bu5												
<i>T. pyogenes</i> SH02		Yellow	Green	Green	Yellow				Yellow			
<i>T. pyogenes</i> SH01		Yellow			Green	Yellow				Yellow	Yellow	Yellow
<i>T. pyogenes</i> MS249		Yellow					Yellow					
<i>T. pyogenes</i> SH03		Yellow										

Figure 10. Antimicrobial resistance genes (ARGs) detected in the investigated *T. pyogenes* strains by the CARD database. The yellow and green boxes in the table depict ARGs classified into perfect and strict categories, respectively, by the RGI software of CARD database.

The highest numbers of ARGs were found in *T. pyogenes* SH01 (6), *T. pyogenes* SH02 (6), and *T. pyogenes* TP1 (5), respectively. Notably, no ARG were detected in genomes of *T. pyogenes* strains DSM20630, NCTC5224, Bu5, and UFV1 by CARD (Table S10).

4. Discussion

Trueperella pyogenes is a resident commensal of skin biota and mucous membrane of the upper respiratory and genital tracts of domestic and wild animals including humans [1]. The bacterium has also been detected in bovine rumen and the gastrointestinal tract of swine [50]. However, in response to a precipitating injury or infection, the bacterium manifests as one of the most common opportunistic pathogens, resulting in a range of purulent infections such as mastitis, cutaneous and liver abscessation, metritis, endometritis, and pneumonia in domestic and wild animals, which results in huge eco-

nomic losses [1,11,51–53]. Its versatility can be gauged from the range of animals it has been isolated from, including wild ungulates (antelopes, bison, deer, musk deer, gazelles, wildebeest), avian species (chicken, macaws, turkeys) elephants, reindeer, and companion animals (dogs, cats, and horses). Even though the species has been recognized for a long time, still the underlying mechanisms of disease pathogenesis, reservoirs as well as routes of transmission of bacteria, are incompletely understood [1]. The rise in drug resistance to available antibiotics poses a significant challenge to control the disease effectively [23,52,54]. Vaccination can be an important and effective measure for the control of pathogen, but no commercial vaccine with adequate protection is available till date [55–58]. Consequently, the understanding of genomic architecture of prevalent *T. pyogenes* strains is crucial for development of the strategies to control the thriving of pathogenic strains. Significantly, only a few virulence factors, namely, pyolysin(*ply*), fimbriae (*fimA*, *fimC*, *fimE*, and *fimJ*), collagen-binding protein A (*cbpA*), and neuraminidases (*NanH* and *NanP*), which contribute to the pathogenic potential, have been recognized [1,11,13]. However, the complete role of such virulence factors can only be understood after elucidation of core-genomic features of this pathogen.

In the current investigation, we have compared all the available 19 *T. pyogenes* strains originating from distinct geographical regions i.e., China (11), India (1), Iran (1), US (1), Brazil (1), and Australia (1) to elucidate the gene repertoire as well as their distinct genomic features (Table S2a,b). The study encompasses eleven completely sequenced genomes and eight good quality draft genomes isolated from different host species (Table S2a,b). In order to obtain insight into the genetic repertoire of the analyzed strains, pan-genome calculations were performed. The pan-genome investigation of *T. pyogenes* genomes revealed a pan-genome repertoire of 3214 CDS, a core-genome of 1520 CDS (47.3%), a dispensable genome of 1093 CDS (34%), and strain-specific genes in the range of 2–63 (18.7%), respectively (Figure 3d). The core-genome, which is nearly 47.3% of the pan-genome, reveals that a high level of intra-species diversity exists at the genomic level among *T. pyogenes* strains included in the study. The pliant genomic subset comprised of dispensable genome and singletons is nearly 52.7% of the total pan-genome in this study, which indicates the capacity and propensity of *T. pyogenes* to adapt to the challenges posed by a variety of warm-blooded host cell surface and environmental stressors such as antimicrobial compounds. Notably, no strain-specific genes were observed in the genomes of TP3, TP6375, TP4479, and TP2849; however, it is notable that 3 strains out of 4, i.e., TP3, TP4479 and TP2849, are almost identical genomically with average nucleotide identity (ANI) of 100% (Figure S3). The fewer number of singletons detected in our study has similarly been observed in pan-genome analysis of 42 *Arcanobacterium phocae* strains which contained 73 unique genes [59]. The *Arcanobacterium* spp. and *Trueperella* spp. are phylogenetically close taxa having been recently separated [60].

The pan-genome development graph of investigated *T. pyogenes* strains depicts steady growth with the addition of each genome, which reflects the genomic diversity of the investigated strains as well as the capacity of *T. pyogenes* to acquire exogenous DNA (Figure 3). The core-genome development plot converges to about 1488 genes (Figure 3). This set of genes shared by all analyzed strains play an important role in bacterial survival [33,35,37] and in the case of *T. pyogenes*, this gene set may be helpful in defining its unique ability of a commensal with an ability to cause opportunistic pyogenic infections in a wide variety of mammals. The singleton development plot depicted the possibility of finding 21 new genes with the addition of a new *T. pyogenes* genome (Figure 3). Notably, the steady growth in the pan-genome development plot with a growth exponent value of 0.162, convergence in the core-genome development plot, and the possibility of finding ≈ 21 new genes, with the addition of a newly sequenced genome, led to significant inference that *T. pyogenes* genomes harbor an open pan-genome state (Table S1, Figure 3). However, the number of genomes used in our study is limited to 19, which may be a limiting factor in true estimation of genome openness. For example, *A. baumannii* strains were estimated to be closed by Chan et al. (2015), when they took 249 genomes into account, which was previously

estimated to be open on analysis of the pan-genome size of 16 strains [61,62]. However, two observations make our conclusion of pan-genome open status as significant, i.e., one in which 19 *T. pyogenes* strains have a wide origin on the basis of geography, host, and core-genome phylogeny (Table 1; Figure 5), and secondly, the ANI value of all 19 strains have been calculated to be $\geq 97.5\%$ (Figure S3) [63].

Next, we carried functional annotation of representative core-genome subset and a complete subset of strain-specific genes of investigated *T. pyogenes* genomes by utilizing an eggNOG-mapperv2tool. The tool provided functional annotation, orthology assignment, and domain assignment to the input genomic subsets. Notably, 91.57% CDS of the core-genome subset and 31.59% CDS of strain-specific genes were assigned into different COG functional classes by the eggNOG-mapperv2. The maximum CDS (220) of core-genome subset fell into the function unknown category. Next, the functional classes that were enriched with most CDS of core-genome after the function unknown category includes carbohydrate metabolism and transport (139), translation (135), inorganic ion transport and metabolism (120), transcription (115), amino acid metabolism and transport (113), and energy production and conversion (93), replication and repair (85), defence mechanism (31), among others (Table S6a,b; Figure 4A). On the other hand, 31.29% CDS of strain-specific genes subset were assigned into functional categories of replication and repair (23), function unknown (16), defence mechanism (12), and transcription (10), among others (Table S6a,b; Figure 4B). The functional analysis underlines the metabolic versatility of *T. pyogenes*, as it encompasses most of the basic categories of gene ontology [64].

In order to assess the evolutionary relationship among the investigated TP genomes, core-genome was utilized for phylogenetic tree construction. The strains of *T. pyogenes* were found to be falling into different phylogenetic clusters with statistically significant high SH branch support values (Figure 5). The phylogenetic tree built on the basis of 1520 core genes depicted three distinct clades, wherein clade I comprised of strains Arash114 (Iran) and MS249 (Australia) and clade II consisted of Chinese origin strains TP1 and TP2. Notably, clade III harbored the maximum number of strains (15) which bifurcated into a small group consisting of TP6375 (USA) and UFV1 (Brazil) and a large subclade comprising of Asian origin (India- 1, China- 11, and unknown origin-2) strains (Figure 5). The phylogenetic tree gives a picture of high degree of genomic diversity and evolution within the core-genome level of species, which agrees with the propensity of *T. pyogenes* to cause opportunistic infections in a large variety of warm-blooded animals [4].

Interestingly, the three clades can be divided into two groups based on the animal host associated with the strain. Clades I and II consist of domestic large ruminants including cattle and buffalo, whereas clade III is almost entirely derived from the porcine host. Although, within clade III, extensive branching has divided the strains into multiple subgroups with high SH values, the strain Bu5, which is an Indian water buffalo origin strain, is forming a separate clade (Figure 5). The core-genome phylogeny assignment of *A. phocae*, which is a close relative of *T. pyogenes*, also divided the 42 investigated strains into three clusters, with the different clusters predominated by different hosts [59]. The genetic variation in *Escherichia coli* among strains of different phylogroups is believed to support fitness in different ecological habitats, leading to niche preference [65]. The benefit of strains identity at clonal or phylogroup level has also been underlined as a function of its environmental niche, mode of living, and ability to cause disease [66,67]. The core-genome phylogeny in the current investigation shows clear distinction between the ancestral ruminant strains and porcine strains. However, in order to draw better inferences of evolutionary history, the addition of more completely sequenced genomes of *T. pyogenes* isolated from various hosts and geographical regions is required. Additionally, to understand the genomic architecture and possible evolutionary events in *T. pyogenes* strains, synteny plots were created with a reference genome taken as *T. pyogenes* 6375. The synteny plot of investigated complete genomes depicted high synteny with large scale genomic rearrangements such as inversions, insertion, deletion, and relocations

(Figure 6; Figure S1A–J). In particular, 2012CQ-ZSH genome showed a high degree of synteny conservation with the reference genome *T. pyogenes* 6375 (Figure S1).

Although reports of isolation of *T. pyogenes* from bovine rumen, and gastrointestinal tract of swine are there [50], the ecological niche of *T. pyogenes* is considered to be a mucus membrane and skin wherein the bacterium gains entry as an opportunistic pathogen after violence in the mucous membrane of the respiratory and genital tract [68,69]. Apart from this property, the synergistic collaboration of *T. pyogenes*, in conjunction with bacteria such as with anaerobe *Fusobacterium necrophorum* in inter digital pyogenic infections, and with *E. coli* in urogenital infections, shows the strategy of pathogenic synergy with many different bacteria [5]. *Trueperella pyogenes* is a highly versatile and adaptable pathogen, as it is able to produce many factors which aid its adherence and colonization in different body locations, both internally and externally. Therefore, decoding of virulence factors associated with *T. pyogenes* is crucial for the understanding of molecular pathogenesis and development of therapeutics for prevention and control [70,71]. In the current investigation, we utilized BLASTN searches, the VFAnalyzer tool, and manual curation to identify and analyze putative virulence genes in the investigated *T. pyogenes* genomes. First, we utilized BLASTN searches, and manual curation to identify and analyze putative virulence genes which includes pyolysin (*plo*), collagen-binding protein A (*cbpA*), neuraminidases (*NanH* and *NanP*), and fimbriae (*fimA*, *fimC*, *fimE*, and *fimJ*) in the investigated *T. pyogenes* genomes. BLASTN searches revealed the *plo* gene to be a highly conserved component of the core-genome, as homologs of *T. pyogenes* 6375 *plo* were observed in all of the investigated *T. pyogenes* strains with high sequence identity ranging from 80.8% (Bu5) to 100% (UFV1) (Table S7a–c). Multiple sequence alignment of the *plo* homologues suggested high sequence identity with gaps being observed only at two residues (alignment position 17th harboring glycine and 58th possessing threonine) in the Bu5 *plo* sequence (Figure S2A). In corroboration with our results, the presence of *plo* in all of the investigated wild type *T. pyogenes* strains has been previously reported [9,10,21]. *plo* is considered to be the sole haemolysin and crucial for *T. pyogenes* survival, which makes it a highly attractive drug and vaccine target. A number of attempts have been previously made for exploitation of *plo* gene as a vaccine target by utilizing different strategies but has failed to provide adequate protection against the lethal pathogen [57,72]. Therefore, new strategies involving novel targets or altered approaches to utilize *plo* are vital to counter the pathogen.

Adhesion to epithelial cells is generally the first step in pathogenesis and is crucial for the ability of bacteria to colonize host mucosal surfaces [73–76]. Therefore, virulence factors associated with adhesion and immunological response such as neuraminidase (*nanH* and *nanP*), extracellular binding proteins (*cbpA*), and fimbriae (*fimA*, *fimC*, *fimE*, and *fimJ*) are considered significant for the establishment of *T. pyogenes* infection inside the host [1,11]. Therefore, next, we located and analyzed these potential virulence factors in the *T. pyogenes* genomes. *Trueperella pyogenes nanH* homologs were detected in all of the investigated genomes with varying sequence identity ranging from 100% QC and sequence similarity 87.62% in TP2 to 78% QC and 87.59% sequence similarity in UFV1 (Table S7a–c). The multiple sequence alignment of *nanH* protein homologues described truncated N-terminus fragments in the strains TP6375, jx18, DSM20630, SH01, SH03, 2012CQ-ZSH, as well as in UFV1 and truncated C-terminus fragments with gaps in reference protein in the strains TP2, TP8, MS249, Bu5, and NCTC5224, whereas both N-terminus and C-terminus fragments were detected to be truncated in UFV1 strains (Figure S2C). On the other hand, *T. pyogenes nanP* homologues were identified in only 12 out of the 19 investigated *T. pyogenes* genomes with varying sequence identity in the range of 100% QC and 99.37% sequence similarity in Bu5 to 55% QC and 98.62% sequence similarity in MS249 (Table S7a–c). The multiple sequence alignment of *nanP* protein homologues was observed to be highly conserved with the exemption of truncated amino and carboxy terminal fragments in TP8 and truncated carboxy terminal fragments in MS249 (Figure S2D). Neuraminidases (*nanH* and *nanP*) promote host cell adhesion by cleaving off terminal sialic acids and thus exposing host cell receptor molecules. Neuraminidase also decreases mucous viscosity, which facilitates

bacterial colonization in underlying tissues. Moreover, the enzyme has been reported to impair the host immune response as susceptibility of mucosal IgA to bacterial proteases is increased [1,11].

Additionally, *T. pyogenes* also harbors virulence proteins that aid in binding to extracellular matrix binding proteins such as collagen, fibrinogen, and fibronectin. However, only collagen binding ability and not fibrinogen and fibronectin binding ability has been characterized in *T. pyogenes*. CbpA, an MSCRAMM-like surface protein, exploits collagen types I, II, and IV for adherence and subsequent colonization in collagen rich tissue. CbpA is comprised of signal peptide, collagen binding domain, repetitive B domains, and a cell wall anchoring domain. Reduced binding to epithelial and fibroblast cell lines has been demonstrated in *cbpA* knockout mutant [1,11]. The homologues of *T. pyogenes cbpA* were positioned and analyzed in all of the investigated *T. pyogenes* strains and that showed variation in the range of (53% QC and sequence similarity 97.66%) in SH01 to (100% QC and sequence similarity 96.21%) in TP1 Table S7a–c. The *cbpA* homologues in strains such as TP2, MS249, UFV1, and SH01 were observed to be truncated as depicted by the multiple sequence alignment of their corresponding protein sequences (Table S7a–c, Figure S2B). Gene truncation at N terminus was observed in strains MS249, UFV1, and SH01, while C terminal truncation was observed in *cbpA* of TP2.

TP6375 *fimA* homologues displayed a high degree of sequence identity in all of the investigated genomes, with the exception of a few strains wherein gene truncation of *fimA* was observed, which includes strains 2012CQ-ZSH, DSM20630, NCTC5224, and UFV1 (Table S7a–c). Consequently, multiple sequence alignment of *fimA* homologues depicted a high degree of conservation with the exception of *fimA* of 2012CQ-ZSH, DSM20630, and NCTC5224, which displayed missing N terminal fragments and of UFV1 that displayed missing C terminal fragments. Homologues of *fimC* of TP6375 in all of the investigated *T. pyogenes* genomes were observed to be conserved with a sequence identity in the range of 100% in UFV1 to 70.01% in TP3, TP4479, and TP2849 (Table S7a–c). Similarly, homologues of *fimE* TP6375 were observed in all of the investigated *T. pyogenes* genomes with percent identity in the range of 98.23% in Arash114 to 68.55% in NCTC5224 and DSM20630. It is worth noting that homologues of TP6375 *fimJ* were observed in all of the investigated *T. pyogenes* strains with the exception of Bu5 and UFV1 strains in the range of 100% query coverage and sequence identity of 99.11% in Arash 114 to 17% query coverage and sequence identity of 99.26% in SH01 strains (Table S7a–c). The multiple sequence alignment of fimbrial protein sequences revealed that *fimC* and *fimE* harbored lesser mutation events in comparison with *fimA* and *fimJ* in the analyzed sequences (Figure S2E–H). The expression profile of putative fimbrial proteins i.e., *fimA*, *fimC*, and *fimE*, in cultured *T. pyogenes*, was determined by Liu et al., (2018) by cloning respective fimbrial proteins and then generating rabbit anti-rFimA, anti-rFim C, and anti-rFim E serum. The Western blot assay performed using these sera revealed that only *fimE* was constitutively expressed in *T. pyogenes* [24].

Out of all the investigated candidate virulence factors, *plo* appears to be strongly conserved among the investigated genomes. Riseti et al. (2017) investigated 71 *T. pyogenes* strains recovered from mastitis ($n = 35$), and non-mastitis and reported the presence of *plo* (100.00%), *fimA* (98.6%), *nanP* (78.9%), *fimE* (74.6%), *fimC* (64.8%), *nanH* (63.4%), *cbpA* (8.4%) and *fimG* (5.6%) in their studied strains [9]. The variability in gene content is the basis of bacterial evolution, and gene truncation is significant for shaping bacterial genomes [77]. The truncated candidate virulence genes detected in the study (Table S7a–c; Figure S2A–H) might be the result of small-scale lateral gene transfer events and could possibly encode shortened proteins with different functions which can be unravelled by experimental investigation [77,78]. On the other hand, the indels detected in the current study could also be present for reverse gene silencing i.e., for adaptations with respect to changing conditions (Figure S2). The indel frequency inversely correlates with linguistic complexity of genome, gene–position as well as gene–essentiality [79].

After careful and comprehensive analysis of homologues of *plo*, *fimA*, *fimC*, *fimE*, *fimJ*, *cbpA*, *nanH*, and *nanP*, we applied the VFAnalyzer tool and manual curation, which

led to identification of 500 candidate virulence genes (Figures 7 and 8). The identified virulence genes shared homology with known virulence genes of pathogens belonging to genera such as *Mycobacterium*, *Klebsiella*, *Haemophilus*, and *Francisella*, among others, as detailed in Table S7a–c. The candidate repertoire of virulence factors identified in the study seems to be associated with a wide range of virulence related functions such as adherence, iron uptake, biofilm formation, regulation, toxin, amino-acid and purine metabolism, anti-apoptosis factor, lipid and fatty acid metabolism, phagosome arresting, protease, and stress adaptation as listed in Table S7a–c. Notably, the homologues of genes related to iron uptake and siderophore biosynthesis system i.e., *ciuA*, *ciuB*, *ciuC*, and *ciuD* along with a homolog of ABC-type heme transporter related gene *hmuU* were present in all of the investigated *T. pyogenes* genomes. The bacterial systems for iron uptake, such as high-affinity siderophore uptake systems, are important virulence factors in many Gram-positive bacterial pathogens such as *C. diphtheriae* and *C. pseudotuberculosis* [80,81]. In addition, homologs of mycobacterial regulatory genes, such as *relA*, *regX3*, *sigH*, and *sigA/rpoV*, were also detected to be part of the core-genome of all of the investigated *T. pyogenes* genomes. These mycobacterial regulatory genes play a pivotal role in bacterial persistence (*relA*), nutrient sensing (*regX3*), encountering of thermal and oxidative stress (*sigH*), and growth promotion in macrophages (*sigA/rpoV*) [82–85].

Similarly, homologues of mycobacterial genes related to lysine synthesis (*lysA*), glutamine synthesis (*glnA1*), anti-apoptosis factor (*nuoG*), protease (*mpa*, *zmp1*), stress adaptation (*sodA*), and antiphagocytosis (*rmlA*) were also found by the usage of the VFAnalyzer tool, NCBI BLASTN searches, and manual curation in all of the genomes under study. In a similar manner, homologs of *Klebsiella* T6SS-II, mycobacterial GPL locus (*rmlA*), and streptococcal capsule (*rmlB*) were also identified in all of the studied genomes.

In contrast, genes related to adherence (*srtB*), ABC transporter (*fagC*), ABC-type heme transporter (*hmuV*), regulation (*phoP*, *whiB3*), lysine synthesis (*lysA*), pantothenate synthesis (*panC*, *panD*), capsule (*gnd*), trehalose-recycling ABC transporter (*sugC*), exopolysaccharide (*galE*), pyrimidine biosynthesis (*carB*), and exopolysaccharide (*galE* and *pgi*) showed differential distribution in the investigated genomes (Table S7a–c).

Microbial genomes are dynamic in nature as they evolve and adapt with time through horizontal gene transfer (HGT), mutations, and gene rearrangements [86,87]. These adaptations aid in better survival of bacterium in response to stresses encountered in varying conditions [88,89]. As mobile genetic elements (MGEs) add to the resilience of genome, we next searched for MGEs which include genomic islands, prophages, and antibiotic resistance genes by employing various bioinformatics tools in the investigated *T. pyogenes* genomes.

Genomic islands are the cluster of genes attained by HGT and known to be significant for pathogenesis, fitness, and antibiotic resistance [90]. A total of 346 genomic islands in the range of 12–25 in terms of number were found in the investigated *T. pyogenes* genomes by Islandviewer (Table S8a–s, Figure 9a). Highest and lowest numbers of GIs were detected in the strains SH02 and TP8, which comprise 16.5% and 2% of their genome, respectively, by Islandviewer. Genomic islands of *T. pyogenes* strain TP3 (TP3-GI1 and TP3-GI5) and *T. pyogenes* strain TP4 (TP4-GI5 and TP4-GI8), which are involved in multidrug resistance, were investigated by Dong et al., 2020. The investigators documented that TP3-GI1 and TP4-GI5 shared a common region of 20 kb, which is comprised of tetracycline resistance gene *tet(W)* along with a series of genes involved in type IV secretion systems [91]. On the other hand, TP3-GI5 and TP4-GI8 also shared homology, which includes a series of phage and DNA replication linked genes. The TP3-GI5 is comprised of two IS6100 located in the same orientation and one *tet(33)* forming a composite transposon along with macrolide resistance gene *erm(X)* located near the end, whereas TP4-GI8 harbors two copies of *erm(X)* present between two IS1634 elements placed in the same orientation which might have formed a composite transposon. Except for these four GIs of TP3 and TP4 strains, the structural composition and role of other GIs in infection pathogenesis have not been investigated yet.

In the current investigation, we have not further attempted to analyze the identified GIs, as it is beyond the scope of our study.

Bacteriophages are the most abundant genetic entities on the biosphere that influence bacterial virulence and evolution in a number of ways [92]. The lytic reproduction and temperate reproduction (prophage) of bacteriophage depict the extremes of parasitism and mutualism within a single phage genotype [93]. Prophages are significant genetic elements of bacterial genomes as they contribute to diversity, virulence, and fitness [94]. The current study identified 31 prophages in the range of 1–4 in terms of number and 5.2–47.7 kb in terms of size. Out of 31 prophages, two were classified as intact, two as questionable and 27 as incomplete (Table S9; Figure 9b). The average prophage number was detected to be 1.63 per genome. The two intact prophage elements which shared structural similarities with *Lactobacillus* phage iA2 (NC_028830) and *Staphylococcus* phage SPbeta-like (NC_029119) were observed in the strains *T. pyogenes* TP6375 and *T. pyogenes* TP1, respectively. However, the maximum number (4) of prophage elements was detected in the strain TP1 in which one was classified as complete and three as incomplete prophage elements. The most abundant phage type was found to be “PHAGE_Lactob_iA2”, “PHAGE_Coryne_Adelaide”, and “PHAGE_Salmon_118970_sal3” with each type being found in four investigated *T. pyogenes* strains. Interestingly, seven out of the 31 identified prophage elements were unique i.e., were not carried by any another analyzed *T. pyogenes* strains, which include PHAGE_Propio_E6_NC_041894 (jx18), PHAGE_Staphy_SPbeta_like_NC_029119 (TP1), PHAGE_Mycoba_Gaia_NC_026590 (TP4), PHAGE_Coryne_Stiles_NC_048789 (TP8), PHAGE_Synech_S_CAM8_NC_021530 (MS249), PHAGE_Coryne_Lederberg_NC_048790 (NCTC5224), and PHAGE_Brevib_Jimmer2_NC_041976 (Bu5). The unique prophage repertoire detected in the study depicts that the strains harboring them have encountered different environmental conditions and thus have acquired unique prophage sequences. The diverse prophage repertoire is maintained by evolutionary forces such as positive selection of host beneficial genes, negative selection of lytic function genes, and mutational bias of bacteria towards deletion leading to gene loss [93].

Resistome analysis of genomes via CARD database revealed 40 Antibiotic resistance genes (ARGs) wherein five were classified in perfect and 35 in the strict category (Table S10; Figure 10). The identified ARGs in the current study are reported to show resistance to antibiotics such as macrolide, lincosamide, streptogramin, and tetracycline. Highest numbers (6) of ARGs were detected in the Chinese origin strains *T. pyogenes* SH01 and SH02 and that exhibited resistance to antibiotics—aminoglycoside, tetracycline, phenicol, sulphonamide, and acridine dye. Notably, 13 out of the 19 investigated genomes harboured the *tet* (W/N/W) gene which encodes for mosaic tetracycline-resistant ribosomal protection protein. Literature mining also revealed the resistance of *T. pyogenes* to tetracycline in dairy cows suffering from endometritis in Inner Mongolia and China [95]. Recently, 114 *T. pyogenes* isolates obtained from livestock and European bison were studied for the presence of tetracycline resistance determinants in their genomes, and the presence of *tetW* and *tetA*(33) was reported in 43.0% and 8.8% of the isolates, respectively, and the investigators concluded that wild ruminants can be a reservoir of tetracycline resistant *T. pyogenes* [96]. On the other hand, seven genomes in the current study harbored an *ermX* gene that protects ribosome inactivation and provides resistance to antibiotics macrolide, lincosamide, and streptogramin. Notably, eight out of the eleven investigated Chinese *T. pyogenes* strains that harbored ARGs were isolated from the pig samples. The presence of these ARGs may be attributed to intensive, large scale swine production wherein indiscriminate use of antimicrobials has become an integral part for not only treating critically ill animals but also for prophylaxis as well as for growth promotion. The high-density swine farms aggravate the hazard of rapid spread of infectious diseases [97]. Moreover, the increased availability of antimicrobials and lower awareness of pig farmers regarding risk and repercussions of antimicrobials and AMR in China is also one of the contributing factors for augmented prevalence [97]. Additionally, the ongoing globalization for trade and travel has also partly increased the dissemination of resistance determinants across

the world [98]. Notably, no ARG was detected by CARD in the genomes of DSM20630, NCTC5224, Bu5, and UFV1. The antibiotic resistance profile of the currently investigated genomes along with literature reports that show increased penetration of resistant genes in the *T. pyogenes* strains calls for prudent use of antibiotics in veterinary medicine. These findings emphasize the urgent need to explore novel drug targets and improved vaccines for *T. pyogenes* infection prevention and control. Because human health and livestock productivity is significantly associated with increased antimicrobial resistance, this makes it a global health concern [99]. Therefore, the usage of antimicrobials should be minimized and continuous monitoring of AMR resistance from the national level to a farm level should be carried out to follow the trend, so that veterinarians can choose the most optimal treatment [99,100].

The differential distribution of GIs, prophages, and ARGs in the investigated strains depicts intraspecific diversity amongst the *T. pyogenes* species and its ability to acquire exogenous DNA in response to various stresses encountered by the bacterium in order to survive inside different niche and hosts.

The study has unravelled an open pan-genome state of *T. pyogenes* and a range of candidate virulence genes, GIs, prophages, and ARGs in the investigated strains. To the best of our knowledge, this is the first descriptive comparative study of *T. pyogenes* genomes which can be used as a starting point for improved understanding of *T. pyogenes* pathogenesis, screening of drug and vaccine targets as well as for diagnostics. Nevertheless, we also emphasize the need of inclusion of greater number of completely sequenced *T. pyogenes* genomes from India as well as from across the world, as well as isolates from different animal hosts for expansion of the present investigation for the understanding of genetic diversity inherent in this opportunistic pathogen, which could enable us in better management of this emerging infectious pathogen.

5. Conclusions

Trueperella pyogenes is a significant opportunistic bacterium which affects domestic as well as wild animals, and rarely humans by causing pyogenic infections including metritis, mastitis, and pneumonia, which leads to silent economic losses. The increased antimicrobial drug resistance and lack of vaccine with a high degree of protection for infection prevention, and control necessitates detailed genome analysis of the available strains. Genome comparison is an indispensable tool for selection of promising gene target for reverse vaccinology. However, the *T. pyogenes* genomes have not been comprehensively analyzed apart from a few studies that focussed on genomic features of only few virulence determinants. The current investigation involved application of a pan-genomic approach on 19 *T. pyogenes* genomes, which also included an Indian water buffalo isolate Bu5. The investigation revealed a pan-genome repertoire of 3214 CDS, core-genome of 1520 CDS, indispensable genome of 1093, and singletons in the range of 2–63. The phylogenetic analysis utilizing the core-genome showed genetic variability among the analyzed strains with association of clustering of strains with animal hosts. Analysis of known candidate virulence genes such as pyolysin (*plo*), fimbriae (*fimA*, *fimC*, *fimE*, and *fimJ*), collagen-binding protein A (*cbpA*), and neuraminidases (*nanH* and *nanP*) revealed *plo* to be highly conserved amongst them and genetic variability was noted in the rest of the analyzed genes. Further analysis into virulome, mobiliome, and resistome of investigated strains revealed differential distribution and diversity in candidate virulence genes, genomic islands, prophages, and antimicrobial resistance genes. The identified catalogue of candidate virulence genes could be investigated for a nuanced understanding of *T. pyogenes* pathogenesis. Moreover, the identified differences can be further utilized for typing as discriminatory markers for epidemiological surveillance and tracking. In conclusion, the investigation has provided an insight into the genomic repertoire of *T. pyogenes* which can serve as a starting point for future studies related to drug targeting, bacterial typing, diagnostics as well as surveillance purposes.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/antibiotics12010024/s1>, Figure S1: (A–J) Synteny plot of investigated *T. pyogenes* complete genomes against the reference *T. pyogenes* TP6375 genome. Figure S2: Multiple sequence alignment of protein sequences of crucial virulence factors of *T. pyogenes* genomes which includes pyolysin (*plo*), collagen-binding protein A (*cbpA*), neuraminidases (*nanH* and *nanP*), and fimbriae (*fimA*, *fimC*, *fimE*, and *fimJ*) by NCBI COBALT. Figure S3: The Average Nucleotide Identity (ANI) calculation of nucleotide-level genomic similarity among the investigated *T. pyogenes* genomes and generation of an all versus all comparison matrix using a customized setup at the EDGAR web interface. Table S1: The pan-genome and core-genome development projections for investigated nineteen *T. pyogenes* strains. Table S2:a,b: (a) The table lists investigated *T. pyogenes* genomes with details about their geographical origin, status, accession number, host animal, etc. and genome quality parameters. Table S3: The table lists genes identified to be part of the pan-genome of investigated *T. Pyogenes* genomes with their accession number and function. Table S4: The table lists genes identified to be part of core-genome of investigated *T. pyogenes* genomes with their name and function. Table S5:a–s: Strain-specific genes detected in the investigated *T. pyogenes* genomes. Table S6:a,b Functional annotation of core-genome and singletons of investigated *T. pyogenes* genomes by eggNOG-mapper. Table S7:a–c The table lists putative virulence factors belonging to the classes of adherence, iron-uptake, regulation, etc. as identified by a VFAnalyzer tool of VFDB database as well as by manual curation in the investigated *T. pyogenes* genomes, BLASTN search results of virulence factors, namely pyolysin (*plo*), collagen binding protein (*cbpA*), Neuraminidase (*NanH* and *NanP*) and fimbrial genes (*fimA*, *fimC*, *fimE*, and *fimJ*) in the investigated *T. pyogenes* genomes. and BLASTN search results of virulence factors in the investigated genome *T. pyogenes* genomes that showed differential distribution in VFAnalyzer results belonging to the classes of adherence, iron uptake, regulation, etc. Table S8:a–s The table lists genomic islands identified in the investigated *T. pyogenes* genome by an Islandviewer 4 tool. Table S9: The table lists prophages identified in the investigated *T. pyogenes* genomes by the PHASTER tool. Table S10: The table lists antibiotic resistance genes (ARG's) identified in the investigated *T. pyogenes* genomes by the CARD database.

Author Contributions: Z.T. and R.K.V. conceived, designed, and performed the study. R.K.V. contributed to the dataset and supervised the project. Z.T. analyzed the data. Z.T. and R.K.V. wrote and edited the paper. Z.T., R.K.V., T.A. and B.N.T. reviewed and edited the manuscript. All authors contributed to the article and approved the submitted version. All authors have read and agreed to the published version of the manuscript.

Funding: This work under the Institute Project 'Phenotypic and genotypic authentication and preservation of network bacterial isolates' (IXX11884) was supported by the Director, National Research Centre on Equines, Indian Council of Agricultural Research, National Centre for Veterinary Type Cultures, Hisar, Haryana, India.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available in a publicly accessible repository that does not issue DOIs. Publicly available datasets were analyzed in this study. These data can be found in the NCBI genome (<https://www.ncbi.nlm.nih.gov/data-hub/genome/?taxon=1661> accessed on 10 April 2021).

Acknowledgments: The facilities provided by the Director, National Research centre on Equines, Sirsa Road, Hisar are thankfully acknowledged. The authors wish to posthumously remember and acknowledge fellow colleague Neeraj Rana, Principal Scientist, ICAR-Central Institute for Research on Buffaloes, Hisar (Haryana) India for his contribution to the isolation of strain Bu5.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Rzewuska, M.; Kwiecień, E.; Chrobak-Chmiel, D.; Kizerwetter-Świda, M.; Stefańska, I.; Gieryńska, M. Pathogenicity and virulence of *Trueperella pyogenes*: A review. *Int. J. Mol. Sci.* **2019**, *20*, 2737. [[CrossRef](#)] [[PubMed](#)]
2. Nagib, S.; Glaeser, S.P.; Eisenberg, T.; Sammra, O.; Lämmle, C.; Kämpfer, P.; Schauerte, N.; Geiger, C.; Kaim, U.; Prenger-Berninghoff, E.; et al. Fatal infection in three Grey Slender Lorises (*Loris lydekkerianusnordicus*) caused by clonally related *Trueperella pyogenes*. *BMC Vet. Res.* **2017**, *13*, 1–9. [[CrossRef](#)] [[PubMed](#)]
3. Wickhorst, J.P.; Hassan, A.A.; Sheet, O.H.; Eisenberg, T.; Sammra, O.; Alssahen, M.; Lämmle, C.; Prenger-Berninghoff, E.; Zschöck, M.; Timke, M.; et al. *Trueperella pyogenes* isolated from a brain abscess of an adult roebuck (*Capreoluscapreolus*). *Folia Microbiol.* **2018**, *63*, 17–22. [[CrossRef](#)] [[PubMed](#)]
4. Ribeiro, M.G.; Riseti, R.M.; Bolaños, C.A.D.; Caffaro, K.A.; De Moraes, A.C.B.; Lara, G.H.B.; Zamprogna, T.O.; Paes, A.C.; Listoni, F.J.P.; Franco, M.M.J. *Trueperella pyogenes* multispecies infections in domestic animals: A retrospective study of 144 cases (2002 to 2012). *Vet. Q.* **2015**, *35*, 82–87. [[CrossRef](#)]
5. Pillai, D.K.; Amachawadi, R.G.; Baca, G.; Narayanan, S.; Nagaraja, T.G. Leukotoxic activity of *Fusobacterium necrophorum* of cattle origin. *Anaerobe* **2019**, *56*, 51–56. [[CrossRef](#)]
6. Deliwala, S.; Beere, T.; Samji, V.; McDonald, P.J.; Bachuwa, G. When zoonotic organisms cross over—*Trueperella pyogenes* endocarditis presenting as a septic embolic stroke. *Cureus* **2020**, *12*, e7740. [[CrossRef](#)]
7. Kavitha, K.; Latha, R.; Udayashankar, C.; Jayanthi, K.; Oudeacoumar, P. Three cases of *Arcanobacterium pyogenes*-associated soft tissue infection. *J. Med. Microbiol.* **2010**, *59*, 736–739. [[CrossRef](#)]
8. Belser, E.H.; Cohen, B.S.; Keeler, S.P.; Killmaster, C.H.; Bowers, J.W.; Miller, K.V. Epithelial presence of *Trueperella pyogenes* predicts site-level presence of cranial abscess disease in white-tailed deer (*Odocoileus virginianus*). *PLoS ONE* **2015**, *10*, e0120028. [[CrossRef](#)]
9. Riseti, R.M.; Zastempowska, E.; Twarużek, M.; Lassa, H.; Pantoja, J.C.F.; De Vargas, A.P.C.; Guerra, S.T.; Bolaños, C.A.D.; de Paula, C.L.; Alves, A.C.; et al. Virulence markers associated with *Trueperella pyogenes* infections in livestock and companion animals. *Letts. Appl. Microbiol.* **2017**, *65*, 125–132. [[CrossRef](#)]
10. Rogovskyy, A.S.; Lawhon, S.; Kuczumanski, K.; Gillis, D.C.; Wu, J.; Hurley, H.; Rogovska, Y.V.; Konganti, K.; Yang, C.Y.; Duncan, K. Phenotypic and genotypic characteristics of *Trueperella pyogenes* isolated from ruminants. *J. Vet. Diagn. Investig.* **2018**, *30*, 348–353. [[CrossRef](#)]
11. Jost, B.H.; Billington, S.J. *Arcanobacterium pyogenes*: Molecular pathogenesis of an animal opportunist. *Antonie Van Leeuwenhoek* **2005**, *88*, 87–102. [[CrossRef](#)] [[PubMed](#)]
12. Jost, B.H.; Songer, J.G.; Billington, S.J. An *Arcanobacterium* (Actinomyces) *pyogenes* mutant deficient in production of the pore-forming cytolysin pyolysin has reduced virulence. *Infect. Immun.* **1999**, *67*, 1723–1728. [[CrossRef](#)] [[PubMed](#)]
13. Bisinotto, R.S.; Oliveira Filho, J.C.; Narbus, C.; Machado, V.S.; Murray, E.; Bicalho, R.C. Identification of fimbrial subunits in the genome of *Trueperella pyogenes* and association between serum antibodies against fimbrial proteins and uterine conditions in dairy cows. *J. Dairy Sci.* **2016**, *99*, 3765–3776. [[CrossRef](#)] [[PubMed](#)]
14. Zhao, K.; Liu, M.; Zhang, X.; Wang, H.; Yue, B. In vitro and in vivo expression of virulence genes in *Trueperella pyogenes* based on a mouse model. *Vet. Microbiol.* **2013**, *163*, 344–350. [[CrossRef](#)]
15. Machado, V.S.; Bicalho, R.C. Complete genome sequence of *Trueperella pyogenes*, an important opportunistic pathogen of livestock. *Genome Announc.* **2014**, *2*, e00400-14. [[CrossRef](#)]
16. Ashrafi Tamai, I.; Mohammadzadeh, A.; GhalyanchiLangeroudi, A.; Mahmoodi, P.; ZiafatiKafi, Z.; Pakbin, B.; Zahraei Salehi, T. Complete genome sequence of *Trueperella pyogenes* strain Arash114, isolated from the uterus of a water buffalo (*Bubalus bubalis*) in Iran. *BMC Res. Notes* **2021**, *14*, 1–4. [[CrossRef](#)]
17. Azawi, O.I. A study on the pathological lesions of oviducts of buffaloes diagnosed at postmortem. *Vet. Res. Commun.* **2009**, *33*, 77–85. [[CrossRef](#)]
18. Galiero, G. Causes of infectious abortion in the Mediterranean buffalo. *Ital. J. Anim. Sci.* **2007**, *6* (Suppl. 2), 194–199. [[CrossRef](#)]
19. Wani, A.H.; Verma, S.; Sharma, M.; Wani, A. Infectious lameness among migratory sheep and goats in north-west India, with particular focus on anaerobes. *Rev. Sci. Tech. OIE* **2015**, *34*, 855–867. [[CrossRef](#)]
20. Fujimoto, H.; Shimoji, N.; Sunagawa, T.; Kubozono, K.; Nakajima, C.; Chuma, T. Differences in phenotypic and genetic characteristics of *Trueperella pyogenes* detected in slaughtered cattle and pigs with septicemia. *J. Vet. Med. Sci.* **2020**, *82*, 626–631. [[CrossRef](#)]
21. Hijazin, M.; Ülbegi-Mohyla, H.; Alber, J.; Lämmle, C.; Hassan, A.A.; Abdulmawjood, A.; Prenger-Berninghoff, E.; Weiss, R.; Zschöck, M. Molecular identification and further characterization of *Arcanobacterium pyogenes* isolated from bovine mastitis and from various other origins. *J. Dairy Sci.* **2011**, *94*, 1813–1819. [[CrossRef](#)] [[PubMed](#)]
22. Ashrafi Tamai, I.; Mohammadzadeh, A.; Zahraei Salehi, T.; Mahmoodi, P. Genomic characterisation, detection of genes encoding virulence factors and evaluation of antibiotic resistance of *Trueperella pyogenes* isolated from cattle with clinical metritis. *Antonie Van Leeuwenhoek* **2018**, *111*, 2441–2453. [[CrossRef](#)] [[PubMed](#)]
23. Rezanejad, M.; Karimi, S.; Momtaz, H. Phenotypic and molecular characterization of antimicrobial resistance in *Trueperella pyogenes* strains isolated from bovine mastitis and metritis. *BMC Microbiol.* **2019**, *19*, 1–9. [[CrossRef](#)] [[PubMed](#)]
24. Liu, M.; Wang, B.; Liang, H.; Ma, B.; Wang, J.; Zhang, W. Determination of the expression of three fimbrial subunit proteins in cultured *Trueperella pyogenes*. *Acta Vet. Scand.* **2018**, *60*, 1–10. [[CrossRef](#)]

25. Abd El-Aleam, R.H.; George, R.F.; Georgey, H.H.; Abdel-Rahman, H.M. Bacterial virulence factors: A target for heterocyclic compounds to combat bacterial resistance. *RSC Adv.* **2021**, *11*, 36459–36482. [[CrossRef](#)]
26. Kane, T.L.; Carothers, K.E.; Lee, S.W. Virulence factor targeting of the bacterial pathogen *Staphylococcus aureus* for vaccine and therapeutics. *Curr. Drug Targets* **2018**, *19*, 111–127.
27. Davis, J.J.; Wattam, A.R.; Aziz, R.K.; Brettin, T.; Butler, R.; Butler, R.M.; Chlenski, P.; Conrad, N.; Dickerman, A.; Dietrich, E.M.; et al. The PATRIC Bioinformatics Resource Center: Expanding data and analysis capabilities. *Nucleic Acids Res.* **2020**, *48*, D606–D612. [[CrossRef](#)]
28. Sayers, E.W.; Beck, J.; Bolton, E.E.; Bourexis, D.; Brister, J.R.; Canese, K.; Comeau, D.C.; Funk, K.; Kim, S.; Klimke, W.; et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **2021**, *49*, D10. [[CrossRef](#)]
29. Duarte, V.D.S.; Treu, L.; Campanaro, S.; Dias, R.S.; Silva, C.C.D.; Giacomini, A.; Corich, V.; de Paula, S.O. The complete genome sequence of *Trueperella pyogenes* UFV1 reveals a processing system involved in the quorum-sensing signal response. *Genome Announc.* **2017**, *5*, e00639–e17.
30. Machado, V.S.; Bicalho, M.L.D.S.; Meira Junior, E.B.D.S.; Rossi, R.; Ribeiro, B.L.; Lima, S.; Santos, T.; Kussler, A.; Foditsch, C.; Ganda, E.K.; et al. Subcutaneous immunization with inactivated bacterial components and purified protein of *Escherichia coli*, *Fusobacterium necrophorum* and *Trueperella pyogenes* prevents puerperal metritis in Holstein dairy cows. *PLoS ONE* **2014**, *9*, e91734. [[CrossRef](#)]
31. Jalili, V.; Afgan, E.; Gu, Q.; Clements, D.; Blankenberg, D.; Goecks, J.; Taylor, J.; Nekrutenko, A. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Res.* **2020**, *48*, W395–W402. [[CrossRef](#)]
32. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [[CrossRef](#)]
33. Tettelin, H.; Riley, D.; Cattuto, C.; Medini, D. Comparative genomics: The bacterial pan-genome. *Curr. Opin. Microbiol.* **2008**, *11*, 472–477. [[CrossRef](#)]
34. Costa, S.S.; Guimarães, L.C.; Silva, A.; Soares, S.C.; Baraúna, R.A. First steps in the analysis of prokaryotic pan-genomes. *Bioinform. Biol. Insights* **2020**, *14*, 1177932220938064.
35. Medini, D.; Donati, C.; Tettelin, H.; Massignani, V.; Rappuoli, R. The microbial pan-genome. *Curr. Opin. Genet. Dev.* **2005**, *15*, 589–594. [[CrossRef](#)] [[PubMed](#)]
36. Dieckmann, M.A.; Beyvers, S.; Nkouamedjo-Fankep, R.C.; Hanel, P.H.G.; Jelonek, L.; Blom, J.; Goesmann, A. EDGAR3.0: Comparative genomics and phylogenomics on a scalable infrastructure. *Nucleic Acids Res.* **2021**, *49*, W185–W192. [[CrossRef](#)] [[PubMed](#)]
37. Tettelin, H.; Massignani, V.; Cieslewicz, M.J.; Donati, C.; Medini, D.; Ward, N.L.; Angiuoli, S.V.; Crabtree, J.; Jones, A.L.; Durkin, A.S.; et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial “pan-genome”. *Proc. Natl. Acad. Sci. USA* **2005**, *10*, 13950–13955. [[CrossRef](#)] [[PubMed](#)]
38. Cantalapiedra, C.P.; Hernández-Plaza, A.; Letunic, I.; Bork, P.; Huerta-Cepas, J. eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **2021**, *38*, 5825–5829. [[CrossRef](#)]
39. Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernández-Plaza, A.; Forslund, S.K.; Cook, H.; Mende, D.R.; Letunic, I.; Rattei, T.; Jensen, L.J.; et al. eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **2019**, *47*, D309–D314. [[CrossRef](#)]
40. Blom, J.; Kreis, J.; Spänig, S.; Juhre, T.; Bertelli, C.; Ernst, C.; Goesmann, A. EDGAR 2.0: An enhanced software platform for comparative gene content analyses. *Nucleic Acids Res.* **2016**, *44*, W22–W28. [[CrossRef](#)]
41. Edgar, R.C. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* **2004**, *5*, 1–19. [[CrossRef](#)]
42. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* **2010**, *5*, e9490. [[CrossRef](#)]
43. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [[CrossRef](#)] [[PubMed](#)]
44. Liu, B.; Zheng, D.; Jin, Q.; Chen, L.; Yang, J. VFDB 2019: A comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res.* **2019**, *47*, D687–D692. [[CrossRef](#)] [[PubMed](#)]
45. Papadopoulos, J.S.; Agarwala, R. COBALT: Constraint-based alignment tool for multiple protein sequences. *Bioinformatics* **2007**, *23*, 1073–1079. [[CrossRef](#)]
46. Hentschel, U.; Hacker, J. Pathogenicity islands: The tip of the iceberg. *Microbes Infect.* **2001**, *3*, 545–548. [[CrossRef](#)] [[PubMed](#)]
47. Bertelli, C.; Laird, M.R.; Williams, K.P.; Simon Fraser University Research Computing Group; Lau, B.Y.; Hoard, G.; Winsor, G.L.; Brinkman, F.S. IslandViewer 4: Expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* **2017**, *45*, W30–W35. [[CrossRef](#)]
48. Arndt, D.; Grant, J.R.; Marcu, A.; Sajed, T.; Pon, A.; Liang, Y.; Wishart, D.S. PHASTER: A better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **2016**, *44*, W16–W21. [[CrossRef](#)]
49. Alcock, B.P.; Raphenya, A.R.; Lau, T.T.; Tsang, K.K.; Boucharad, M.; Edalatmand, A.; Huynh, W.; Nguyen, A.L.V.; Cheng, A.A.; Liu, S.; et al. CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **2020**, *48*, D517–D525. [[CrossRef](#)]
50. Jarosz, Ł.S.; Gradzki, Z.; Kalinowski, M. *Trueperella pyogenes* infections in swine: Clinical course and pathology. *Pol. J. Vet. Sci.* **2014**, *17*, 395–404. [[CrossRef](#)]

51. Drillich, M. An update on uterine infections in dairy cattle. *Slov. Vet. Res.* **2006**, *43*, 11–15.
52. Galán-Relaño, Á.; Gómez-Gascón, L.; Barrero-Domínguez, B.; Luque, I.; Jurado-Martos, F.; Vela, A.I.; Sanz-Tejero, C.; Tarradas, C. Antimicrobial susceptibility of *Trueperella pyogenes* isolated from food-producing ruminants. *Vet. Microbiol.* **2020**, *242*, 108593. [[CrossRef](#)] [[PubMed](#)]
53. Nagaraja, T.G.; McVey, D.S.; Kennedy, M.; Chengappa, M.M. *Arcanobacterium*. In *Veterinary Microbiology*; John Wiley & Sons: Hoboken, NJ, USA, 2013; pp. 203–205.
54. Alkansir, R.; Wang, J.; Gao, J.; Ali, T.; Zhang, L.; Szenci, O.; Bajcsy, Á.C.; Han, B.O. Properties and antimicrobial susceptibility of *Trueperella pyogenes* isolated from bovine mastitis in China. *Acta Vet. Hung.* **2016**, *64*, 1–12. [[CrossRef](#)] [[PubMed](#)]
55. Huang, T.; Song, X.; Jing, J.; Zhao, K.; Shen, Y.; Zhang, X.; Yue, B. Chitosan-DNA nanoparticles enhanced the immunogenicity of multivalent DNA vaccination on mice against *Trueperella pyogenes* infection. *J. Nanobiotechnology* **2018**, *16*, 1–15.
56. Jost, B.H.; Trinh, H.T.; Songer, J.G.; Billington, S.J. Immunization with genetic toxoids of the *Arcanobacterium pyogenes* cholesterol-dependent cytolysin, pyolysin, protects mice against infection. *Infect. Immun.* **2003**, *71*, 2966–2969. [[CrossRef](#)]
57. Yang, L.; Liang, H.; Wang, B.; Ma, B.; Wang, J.; Zhang, W. Evaluation of the potency of two pyolysin-derived recombinant proteins as vaccine candidates of *Trueperella pyogenes* in a mouse model: Pyolysin oligomerization and structural change affect the efficacy of pyolysin-based vaccines. *Vaccines* **2020**, *8*, 79. [[CrossRef](#)]
58. Zhang, W.; Wang, P.; Wang, B.; Ma, B.; Wang, J. A combined *Clostridium perfringens/Trueperella pyogenes* inactivated vaccine induces complete immunoprotection in a mouse model. *Biologicals* **2017**, *47*, 1–10. [[CrossRef](#)]
59. Aaltonen, K.J.; Kant, R.; KvistNikolaisen, N.; Lindegaard, M.; Raunio-Saarnisto, M.; Paulin, L.; Vapalahti, O.; Sironen, T. Comparative Genomics of 42 *Arcanobacterium phocae* Strains. *Antibiotics* **2021**, *10*, 740. [[CrossRef](#)]
60. Yassin, A.F.; Hupfer, H.; Siering, C.; Schumann, P. Comparative chemotaxonomic and phylogenetic studies on the genus *Arcanobacterium* Collins et al. 1982 emend. Lehnert et al. 2006: Proposal for *Trueperella* gen. nov. and emended description of the genus *Arcanobacterium*. *Int. J. Syst. Evol. Microbiol.* **2011**, *61*, 1265–1274. [[CrossRef](#)]
61. Chan, A.P.; Sutton, G.; DePew, J.; Krishnakumar, R.; Choi, Y.; Huang, X.Z.; Beck, E.; Harkins, D.M.; Kim, M.; Lesho, E.P.; et al. A novel method of consensus pan-chromosome assembly and large-scale comparative analysis reveal the highly flexible pan-genome of *Acinetobacter baumannii*. *Genome Biol.* **2015**, *16*, 143. [[CrossRef](#)]
62. Liu, F.; Zhu, Y.; Yi, Y.; Lu, N.; Zhu, B.; Hu, Y. Comparative genomic analysis of *Acinetobacter baumannii* clinical isolates reveals extensive genomic variation and diverse antibiotic resistance determinants. *BMC Genom.* **2014**, *15*, 1163. [[CrossRef](#)] [[PubMed](#)]
63. Goris, J.; Konstantinidis, K.T.; Klappenbach, J.A.; Coenye, T.; Vandamme, P.; Tiedje, J.M. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **2007**, *57*, 81–91. [[CrossRef](#)] [[PubMed](#)]
64. Zhao, K.; Li, W.; Kang, C.; Du, L.; Huang, T.; Zhang, X.; Wu, M.; Yue, B. Phylogenomics and evolutionary dynamics of the family *Actinomycetaceae*. *Genome Biol. Evol.* **2014**, *6*, 2625–2633. [[CrossRef](#)] [[PubMed](#)]
65. NandaKafle, G.; Huegen, T.; Potgieter, S.C.; Steenkamp, E.; Venter, S.N.; Brözel, V.S. Niche preference of *Escherichia coli* in a peri-urban pond ecosystem. *Life* **2021**, *11*, 1020. [[CrossRef](#)]
66. Picard, B.; Garcia, J.S.; Gouriou, S.; Duriez, P.; Brahimi, N.; Bingen, E.; Elion, J.; Denamur, E. The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect. Immun.* **1999**, *67*, 546–553. [[CrossRef](#)]
67. Walk, S.T.; Alm, E.W.; Gordon, D.M.; Ram, J.L.; Toranzos, G.A.; Tiedje, J.M.; Whittam, T.S. Cryptic lineages of the genus *Escherichia*. *Appl. Environ. Microbiol.* **2009**, *75*, 6534–6544. [[CrossRef](#)]
68. Queen, C.; Ward, A.C.; Hunter, D.L. Bacteria isolated from nasal and tonsillar samples of clinically healthy Rocky Mountain bighorn and domestic sheep. *J. Wildl. Dis.* **1994**, *30*, 1–7. [[CrossRef](#)]
69. Silva, E.; Gaivão, M.; Leitão, S.; Jost, B.H.; Carneiro, C.; Vilela, C.L.; da Costa, L.L.; Mateus, L. Genomic characterization of *Arcanobacterium pyogenes* isolates recovered from the uterus of dairy cows with normal puerperium or clinical metritis. *Vet. Microbiol.* **2008**, *132*, 111–118. [[CrossRef](#)]
70. Leitão, J.H. Microbial virulence factors. *Int. J. Mol. Sci.* **2020**, *21*, 5320. [[CrossRef](#)]
71. Sousa, S.; Mesquita, F.S.; Cabanes, D. Old war, new battle, new fighters! *J. Infect. Dis.* **2015**, *211*, 1361–1363. [[CrossRef](#)]
72. Huang, T.; Zhao, K.; Zhang, Z.; Tang, C.; Zhang, X.; Yue, B. DNA vaccination based on pyolysin co-immunized with IL-1 β enhances host antibacterial immunity against *Trueperella pyogenes* infection. *Vaccine* **2016**, *34*, 3469–3477. [[CrossRef](#)] [[PubMed](#)]
73. Haiko, J.; Westerlund-Wikström, B. The role of the bacterial flagellum in adhesion and virulence. *Biology* **2013**, *2*, 1242–1267. [[CrossRef](#)] [[PubMed](#)]
74. Haines-Menges, B.L.; Whitaker, W.B.; Lubin, J.B.; Boyd, E.F. Host sialic acids: A delicacy for the pathogen with discerning taste. *Metab. Bact. Pathog.* **2015**, *3*. [[CrossRef](#)]
75. Lewis, W.G.; Robinson, L.S.; Gilbert, N.M.; Perry, J.C.; Lewis, A.L. Degradation, foraging, and depletion of mucus sialoglycans by the vagina-adapted *Actinobacterium Gardnerella vaginalis*. *J. Biol. Chem.* **2013**, *288*, 12067–12079. [[CrossRef](#)]
76. Ribet, D.; Cossart, P. How bacterial pathogens colonize their hosts and invade deeper tissues. *Microbes Infect.* **2015**, *17*, 173–183. [[CrossRef](#)] [[PubMed](#)]
77. Hao, W.; Golding, G.B. Inferring bacterial genome flux while considering truncated genes. *Genetics* **2010**, *186*, 411–426. [[CrossRef](#)] [[PubMed](#)]
78. Holmes, D.E.; Dang, Y.; Walker, D.J.; Lovley, D.R. The electrically conductive pili of *Geobacter* species are a recently evolved feature for extracellular electron transfer. *Microb. Genom.* **2016**, *2*, e000072. [[CrossRef](#)] [[PubMed](#)]

79. Gupta, A.; Alland, D. Reversible gene silencing through frameshift indels and frameshift scars provide adaptive plasticity for *Mycobacterium tuberculosis*. *Nat. Commun.* **2021**, *12*, 1–11. [[CrossRef](#)]
80. Ibraim, I.C.; Parise, M.T.D.; Parise, D.; Sfeir, M.Z.T.; de Paula Castro, T.L.; Wattam, A.R.; Ghosh, P.; Barh, D.; Souza, E.M.; Góes-Neto, A.; et al. Transcriptome profile of *Corynebacterium pseudotuberculosis* in response to iron limitation. *BMC Genom.* **2019**, *20*, 1–24. [[CrossRef](#)]
81. Kunkle, C.A.; Schmitt, M.P. Analysis of a DtxR-regulated iron transport and siderophore biosynthesis gene cluster in *Corynebacterium diphtheriae*. *J. Bacteriol.* **2005**, *187*, 422–433. [[CrossRef](#)]
82. Dutta, N.K.; Mehra, S.; Martinez, A.N.; Alvarez, X.; Renner, N.A.; Morici, L.A.; Pahar, B.; MacLean, A.G.; Lackner, A.A.; Kaushal, D. The stress-response factor SigH modulates the interaction between *Mycobacterium tuberculosis* and host phagocytes. *PLoS ONE* **2012**, *7*, e28958. [[CrossRef](#)] [[PubMed](#)]
83. Franceschi, V.; Mahmoud, A.H.; Abdellrazeq, G.S.; Tebaldi, G.; Macchi, F.; Russo, L.; Fry, L.M.; Elnaggar, M.M.; Bannantine, J.P.; Park, K.T.; et al. Capacity to Elicit Cytotoxic CD8 T Cell Activity Against *Mycobacterium avium* subsp. *paratuberculosis* Is Retained in a Vaccine candidate 35 kDa peptide modified for expression in mammalian cells. *Front. Immunol.* **2019**, *10*, 2859. [[CrossRef](#)] [[PubMed](#)]
84. Pei, J.F.; Qi, N.; Li, Y.X.; Wo, J.; Ye, B.C. RegX3-mediated regulation of methylcitrate cycle in *Mycobacterium smegmatis*. *Front. Microbiol.* **2021**, *12*, 119. [[CrossRef](#)] [[PubMed](#)]
85. Wu, S.; Howard, S.T.; Lakey, D.L.; Kipnis, A.; Samten, B.; Safi, H.; Gruppo, V.; Wizel, B.; Shams, H.; Basaraba, R.J.; et al. The principal sigma factor sigA mediates enhanced growth of *Mycobacterium tuberculosis* in vivo. *Mol. Microbiol.* **2004**, *51*, 1551–1562. [[CrossRef](#)]
86. Arnold, B.J.; Huang, I.; Hanage, W.P. Horizontal gene transfer and adaptive evolution in bacteria. *Nat. Rev. Microbiol.* **2021**, *20*, 206–218. [[CrossRef](#)]
87. Juhas, M.; Van Der Meer, J.R.; Gaillard, M.; Harding, R.M.; Hood, D.W.; Crook, D.W. Genomic islands: Tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.* **2009**, *33*, 376–393. [[CrossRef](#)]
88. Li, W.; Wang, A. Genomic islands mediate environmental adaptation and the spread of antibiotic resistance in multiresistant Enterococci—evidence from genomic sequences. *BMC Microbiol.* **2021**, *21*, 1–10. [[CrossRef](#)]
89. Wang, X.; Kim, Y.; Ma, Q.; Hong, S.H.; Pokusaeva, K.; Sturino, J.M.; Wood, T.K. Cryptic prophages help bacteria cope with adverse environments. *Nat. Commun.* **2010**, *1*, 1–9. [[CrossRef](#)]
90. Rao, R.T.; Sharma, S.; Sivakumar, N.; Jayakumar, K. Genomic islands and the evolution of livestock-associated *Staphylococcus aureus* genomes. *Biosci. Rep.* **2020**, *40*, BSR20202287. [[CrossRef](#)]
91. Dong, W.L.; Xu, Q.J.; Atiah, L.A.; Odah, K.A.; Gao, Y.H.; Kong, L.C.; Ma, H.X. Genomic island type IV secretion system and transposons in genomic islands involved in antimicrobial resistance in *Trueperella pyogenes*. *Vet. Microbiol.* **2020**, *242*, 108602. [[CrossRef](#)]
92. Fortier, L.C.; Sekulovic, O. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* **2013**, *4*, 354–365. [[CrossRef](#)]
93. Khan, A.; Burmeister, A.R.; Wahl, L.M. Evolution along the parasitism-mutualism continuum determines the genetic repertoire of prophages. *PLoS Comput. Biol.* **2020**, *16*, e1008482. [[CrossRef](#)]
94. Costa, A.R.; Monteiro, R.; Azeredo, J. Genomic analysis of *Acinetobacter baumannii* prophages reveals remarkable diversity and suggests profound impact on bacterial virulence and fitness. *Sci. Rep.* **2018**, *8*, 1–11. [[CrossRef](#)]
95. Zhang, D.; Zhao, J.; Wang, Q.; Liu, Y.; Tian, C.; Zhao, Y.; Yu, L.; Liu, M. *Trueperella pyogenes* isolated from dairy cows with endometritis in Inner Mongolia, China: Tetracycline susceptibility and tetracycline-resistance gene distribution. *Microb. Pathog.* **2017**, *105*, 51–56. [[CrossRef](#)]
96. Kwiecień, E.; Stefańska, I.; Chrobak-Chmiel, D.; Kizerwetter-Świda, M.; Moroz, A.; Olech, W.; Spinu, M.; Binek, M.; Rzewuska, M. *Trueperella pyogenes* isolates from livestock and European bison (*Bison bonasus*) as a reservoir of tetracycline resistance determinants. *Antibiotics* **2021**, *10*, 380. [[CrossRef](#)]
97. Yang, H.; Paruch, L.; Chen, X.; Van Eerde, A.; Skomedal, H.; Wang, Y.; Liu, D.; Liu Clarke, J. Antibiotic application and resistance in swine production in China: Current situation and future perspectives. *Front. Vet. Sci.* **2019**, *6*, 136. [[CrossRef](#)]
98. Ström, G.; Boqvist, S.; Albiñ, A.; Fernström, L.L.; Andersson Djurfeldt, A.; Sokerya, S.; Sothya, T.; Magnusson, U. Antimicrobials in small-scale urban pig farming in a lower middle-income country—arbitrary use and high resistance levels. *Antimicrob. Resist. Infect. Control* **2018**, *7*, 1–11. [[CrossRef](#)]
99. Aarestrup, F.M.; Duran, C.O.; Burch, D.G. Antimicrobial resistance in swine production. *Anim. Health Res. Rev.* **2008**, *9*, 135–148. [[CrossRef](#)]
100. Sanders, P.; Vanderhaeghen, W.; Fertner, M.; Fuchs, K.; Obritzhauser, W.; Agunos, A.; Carson, C.; BorckHøg, B.; Dalhoff Andersen, V.; Chauvin, C.; et al. Monitoring of farm-level antimicrobial use to guide stewardship: Overview of existing systems and analysis of key components and processes. *Front. Vet. Sci.* **2020**, *7*, 540. [[CrossRef](#)]

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.