

## Article

# How GBS Got Its Hump: Genomic Analysis of Group B *Streptococcus* from Camels Identifies Host Restriction as well as Mobile Genetic Elements Shared Across Hosts and Pathogens

Chiara Crestani <sup>1,2,\*</sup>, Dinah Seligsohn <sup>3</sup>, Taya L. Forde <sup>2</sup> and Ruth N. Zadoks <sup>4</sup><sup>1</sup> Department of Global Health, Institut Pasteur, 75015 Paris, France<sup>2</sup> Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Garscube Campus, Glasgow G61 1AF, UK<sup>3</sup> National Veterinary Institute, SE-751 89 Uppsala, Sweden<sup>4</sup> Sydney School of Veterinary Science, Faculty of Science, The University of Sydney, Camden, NSW 2570, Australia

\* Correspondence: chiara.crestani@pasteur.fr

† Current address: Department of Global Health, Institut Pasteur, rue du Docteur Roux 25-28, 75015 Paris, France.

**Abstract:** Group B *Streptococcus* (GBS) literature largely focuses on humans and neonatal disease, but GBS also affects numerous animals, with significant impacts on health and productivity. Spill-over events occur between humans and animals and may be followed by amplification and evolutionary adaptation in the new niche, including changes in the core or accessory genome content. Here, we describe GBS from one-humped camels (*Camelus dromedarius*), a relatively poorly studied GBS host of increasing importance for food security in arid regions. Genomic analysis shows that virtually all GBS from camels in East Africa belong to a monophyletic clade, sublineage (SL)609. Capsular types IV and VI, including a new variant of type IV, were over-represented compared to other host species. Two genomic islands with signatures of mobile elements contained most camel-associated genes, including genes for metal and carbohydrate utilisation. Lactose fermentation genes were associated with milk isolates, albeit at lower prevalence in camel than bovine GBS. The presence of a phage with high identity to *Streptococcus pneumoniae* and *Streptococcus suis* suggests lateral gene transfer between GBS and bacterial species that have not been described in camels. The evolution of camel GBS appears to combine host restriction with the sharing of accessory genome content across pathogen and host species.

**Keywords:** group B *Streptococcus*; *Streptococcus agalactiae*; camel; milk; GWAS; host adaptation

**Citation:** Crestani, C.; Seligsohn, D.; Forde T.L.; Zadoks R.N. How GBS Got Its Hump: Genomic Analysis of Group B *Streptococcus* from Camels Identifies Host Restriction as well as Mobile Genetic Elements Shared Across Hosts and Pathogens. *Pathogens* **2022**, *11*, 1025. <https://doi.org/10.3390/pathogens11091025>

Academic Editors: Kirsty Le Doare and Konstantinos Karampatsas

Received: 25 July 2022

Accepted: 5 September 2022

Published: 8 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**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

Group B *Streptococcus* (GBS), or *Streptococcus agalactiae*, is primarily known as a cause of human neonatal and maternal disease [1]. However, it is a versatile multi-host pathogen that also inflicts significant damage in adults [2–4] and in several animal species [5–7]. The different host species can be seen as different niches for GBS, whilst niches can also be differentiated within host species. For example, hypervirulent clonal complex (CC) 17 is a human GBS clade specifically associated with neonates. In cattle, GBS is specifically associated with the mammary gland, where it causes infection and inflammation (mastitis). The ability of bacterial pathogens to thrive in different niches can be driven by adaptation of the core genome, such as point mutations in genes of particular functional relevance, or changes in accessory genome content [8,9]. Host-adapted lineages of GBS, such as CC17 in humans, CC61/67 in cattle and CC552 in poikilothermic species, are identified based on multilocus sequence typing (MLST), which represents the core genome. However, the accessory genome appears to be particularly important in GBS: the ability of host generalist lineages such as CC23 (humans and cattle) or CC283 (humans and fishes) to infect

multiple host species is enabled by the acquisition of host-associated accessory genome content [6,7,10]. This occurs predominantly through the horizontal gene transfer (HGT) of mobile genetic elements (MGE), including transposons, plasmids, phages, genomic islands (GEI), and phage inducible chromosomal islands (PICI) [11,12]. The ensemble of MGE carried by a given species, also known as the mobilome [13], plays a fundamental role in shaping the adaptability of bacteria, and may affect the host range as well as tissue tropism [14].

The colonisation or infection of a new niche provides access to new accessory genetic material and creates opportunities for the acquisition of niche-associated genes that may facilitate amplification and subsequent onward transmission [7]. This is postulated to have happened to GBS CC283, which expanded in fishes around the time of the intensification of aquaculture, and might be explained by initial human-to-animal transfer with the subsequent acquisition of fish-associated MGE that allowed for amplification in the new niche [3,6]. Likewise, the acquisition of a lactose operon is thought to explain the expansion of CC1 and CC103 in cattle [15]. The spill over of GBS between host species happens in multiple directions, including between its three major host groups, i.e., humans, cattle, and fishes, which contain different microbiomes and hence provide an opportunity for exposure to and acquisition of a diversity of MGE [7]. Thus, the GBS pangenome and its evolution can only be understood by studying its ecology and genomics across host species. Increasingly, this is being done for GBS in humans, cattle, and fishes [16,17], but an important emerging host species has received limited attention: the one-humped camel (*Camelus dromedarius*).

GBS has been reported in camels in East Africa [18–20] and the Middle East [21]. In camels, as in humans, GBS is both a commensal and a pathogen, with carriage in close to a quarter of the camel population. Carriage in camels occurs primarily in the nasopharynx [22–25], in contrast to the situation in humans, where the gastrointestinal and urogenital tract are the most common carriage sites. As in dairy cattle, GBS in camels is commonly associated with mastitis (i.e., inflammation of the mammary gland). In camels, however, GBS is not limited to the mammary gland and may also cause skin and soft tissue infections [18]. The distinct ecology of camel GBS could be expected to reflect evolutionary adaptations that are unique to camels. Based on limited genomic data and multi-locus sequence typing, camel GBS appears to be distinct from GBS of humans or other animals [19,26], except for one report in the camel milk of a sequence type (ST) 1 isolate that showed genetic features of adaptation to the human host, pointing to a possible human-to-camel transmission event [25,27]. Analysis of a single camel-derived GBS genome suggested distinct gene content compared to GBS from other host species, and biochemical enrichment of carbohydrate metabolism [7]. As camel husbandry in East Africa is on the cusp of intensification [25], it is quite possible that new strains of GBS in camels will emerge in this setting. Thus, now is an opportune time to study the adaptation of GBS to a hitherto poorly represented host species, including different organ systems within this host species, and to explore the potential for HGT between GBS clades or spill-over of GBS strains into or out of the camel population.

To gain insight into potential genetic mechanisms for the host adaptation of GBS in camels, we investigated the population structure of camel GBS. To identify host-associated core and accessory genome content, camel GBS was compared to representative genomes from major host groups and lineages. Because commonly used typing schemes for GBS have largely been developed for human isolates, they may miss animal-specific genome content. Therefore, we used genome-wide association studies (GWAS), which are an unbiased statistical approach to detect genes that are positively or negatively associated with a trait [28]. In addition to host as a trait, we considered mastitis as a trait because of the importance of this disease in terms of camel health and welfare, as well as the contribution of camels to nutrition security in the Horn of Africa.

## 2. Methods

### 2.1. Dataset Selection

A total of 680 GBS genomes were used for this study, including all publicly available camel GBS genomes ( $n = 139$ ), and 541 GBS genomes from the three major GBS host groups for comparison (human,  $n = 285$ ; bovine,  $n = 155$ ; piscine,  $n = 101$ ). The latter were selected to represent all major sublineages (SL) and clonal groups (CG), and a wide range of geographical and temporal origins (35 countries; 1953–2019). We use the terms SL and CG to refer to phylogenetic clusters that include several sequence types (ST), with the latter being at a finer scale of resolution. To avoid redundancy in the genomic data, which would have increased computational requirements without providing additional information on gene presence/absence, we selected one isolate per host/year/serotype/ST combination for each epidemiological unit, e.g., hospital or farm. Camel GBS originated from six sample types or sampling sites (mammary gland/milk,  $n = 81$ ; nose,  $n = 38$ ; oropharynx,  $n = 8$ ; abscesses,  $n = 6$ ; rectum,  $n = 3$ ; skin/wounds,  $n = 3$ ). A complete list of camel and GBS genomes from other species used in this study can be found in the Supplementary Material, Supplementary File S1. Genomes originating from public repositories ( $n = 633$ ) were downloaded in the form of raw reads (either paired-end or single-end) when possible ( $n = 462$ ), or in the form of assembled genomes when raw data were not available ( $n = 171$ ). Data were newly generated for human ( $n = 4$ ), bovine ( $n = 1$ ) and piscine ( $n = 42$ ) GBS isolates from Thailand ( $n = 33$ ) and Vietnam ( $n = 14$ ) by Prof. Swaine Chen (Genome Institute of Singapore, Singapore) and used with permission from Dr. Wanna Sirimanapong (Mahidol University, Nakhon Pathom, Thailand) and Dr. Nguyen Ngoc Phuoc (Hue University of Agriculture and Forestry, Hue City, Vietnam), respectively.

### 2.2. Sequencing, Assembly and Typing

The sequencing of the 47 new genomes was carried out with Illumina HiSeq technology at MicrobesNG (Birmingham, UK) ( $n = 11$ ) and at the Genome Institute of Singapore ( $n = 36$ ). Sequence data have been made publicly available on ENA (PRJEB53664). *De novo* assembly of reads was conducted for 509 genomes. The Wellcome Sanger Institute assembly pipeline with velvet v1.2.10 was used for genomes that were already present within the Sanger server ( $n = 159$ ) and SPAdes v3.13.0 for the remaining genomes with available raw reads ( $n = 350$ , including the 47 newly sequenced isolates). The assembly quality for all 680 genomes was checked with QUAST v5.0.2 [29] (Figure S1), and species identity was confirmed with KmerFinder v3.2 [30].

Seven-gene multi locus sequence typing (MLST) was carried out for all 680 genomes, using SRST2 v0.2.0 [31] on raw reads or MLST v2.0 [32] when only assemblies were available. The capsular type was detected in silico using a BLAST-based standard method [33], which was developed for human GBS and previously validated for bovine GBS [15]. To test this method in camels, we compared our results with previously reported capsular types obtained with a different method [34] when these data were available ( $n = 17$ ) [18]. For the entire dataset, we compared capsular types obtained with the method from Metcalf et al. [33] with those obtained using a new tool for capsular typing, GBS-SBG [35], which allows for differentiation of type III subtypes (1–4). For camel GBS only, capsular type loci (*cps*) were extracted from the genomes with in silico PCR within SnapGene v6.0.2 (from Insightful Science; available at [snapgene.com](http://snapgene.com)) (forward primer: *cps*-F 5'-ATGTCTAATCATTCGCGCCGTCAACAAAAGAAA-3'; reverse primer: *cps*-R 5'-TTATAAGGTTTTAACTTCGTCTACAAATAATTG-3'). When *cps* loci were found separated into two contigs ( $n = 18$ ), the sequences were manually extracted, and the *cps* locus was scaffolded by adding a series of 100 'Ns' between the two segments. Pseudogenisation of the capsular genes was checked manually by comparing the length and number of the open reading frames (ORFs) (as predicted by SnapGene) among sequences belonging to the same *cps* type. Prokka v1.14.5 [36] was used for genome annotation.

### 2.3. Core Genome Analysis of Camel Isolates

A phylogenetic tree of the whole dataset was estimated from the core gene alignment obtained from Roary (as described in the next section) with IQtree v1.6.10 [37], with a general time-reversible (GTR)+G model. Fastbaps v1.0.4 (fast hierarchical Bayesian analysis of population structure) [38] was run on the core genome alignment of camel isolates belonging to SL609 within RStudio v1.3.1093 [39], R v4.0.3 [40].

### 2.4. Genome-Wide Association Studies (GWAS)

Two pan-GWAS were performed with Scoary v1.6.16 [28]: (i) markers of host association were identified through comparison of camel against non-camel GBS genomes; (ii) markers of mastitis-association within camels were identified through comparison between GBS from camel milk (obtained from mammary glands with mastitis) and extra-mammary camel body sites or fluids (including nose, oropharynx, abscesses, rectum and skin/wounds) [25,27]. Gene presence/absence matrices were generated with Roary v3.13.0 [41] from annotation files created with Prokka. For each genetic variant, the Scoary output includes *p*-values, sensitivity (SE) and specificity (SP) measures. Sensitivity indicates which proportion of isolates from the niche of interest (here: camel or milk) contains the specified feature. Specificity indicates which proportion of isolates from other niches (here: other hosts, or other sample types) does not contain the specified feature. For example, perfect sensitivity and specificity at host level would mean that all camel GBS isolates contain the sequence of interest, but none of the GBS isolates from other hosts do. A flow chart of the methods from sample origin through to GWAS can be found in the Figure S2.

### 2.5. Analysis of Mobile Genetic Elements in Camel Isolates

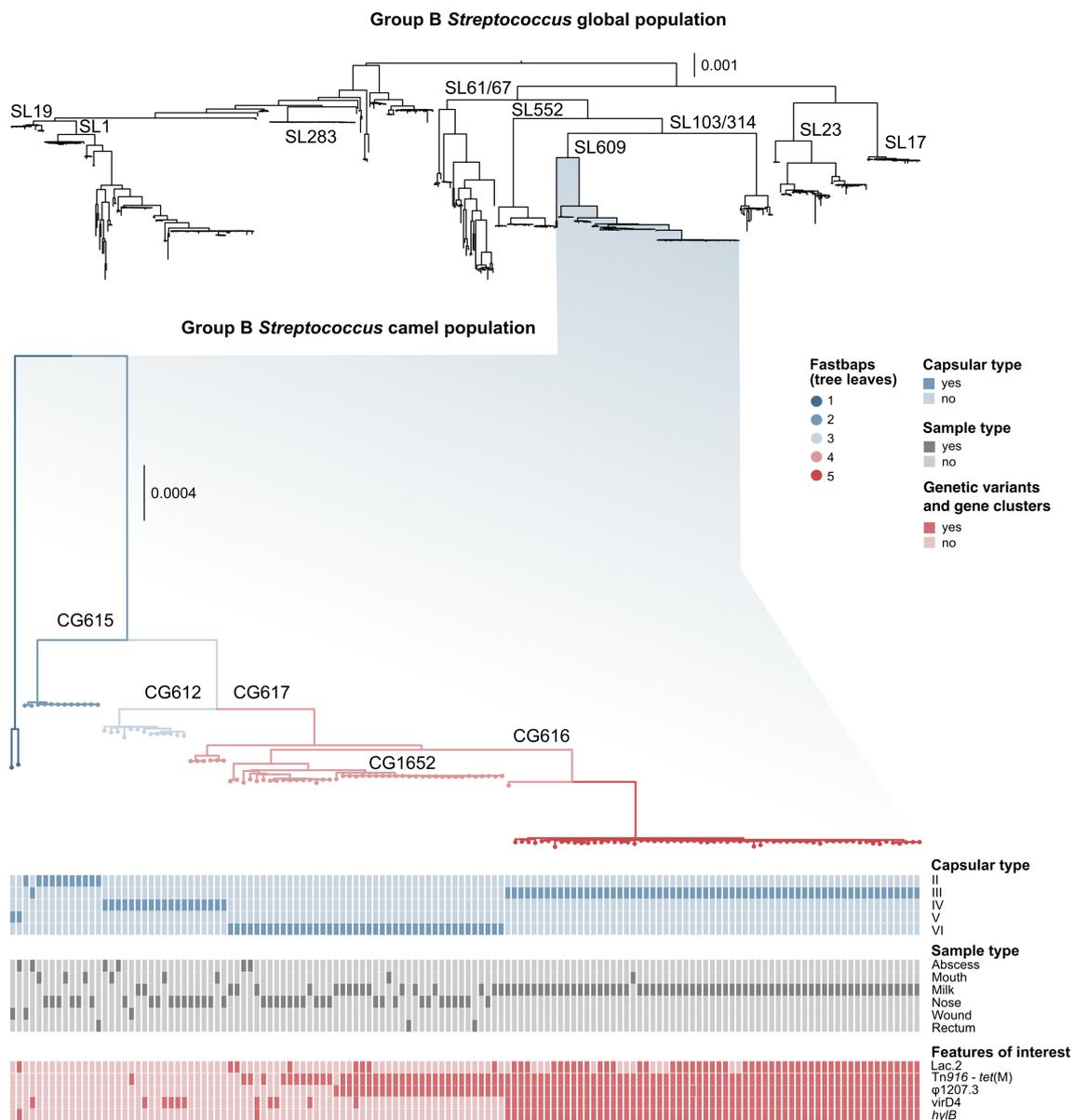
Presence of prophages and PICI was assessed by screening genomes for integrase gene types with a BLAST-based method, as previously described [12]. Thresholds were set at 90% for prophages and 96% for PICI for percentage of identity (ID), and at 99% for query coverage. Considering that only one camel genome was available when this method was developed, we also screened these assemblies with PHASTER [42] to ensure all prophages and PICI were detected by BLAST. Prophages and PICI which carried an integrase gene detected by PHASTER were classified based on integrase type [12]. We then compared results from the two analyses.

Figures were created with Phandango v1.3.0 [43], Circos v0.69-5 [44], BRIG v0.95 [45], and Easyfig v 2.2.2 [46], and modified with Inkscape v1.2.1 ([www.inkscape.org](http://www.inkscape.org)).

## 3. Results

### 3.1. Genotyping and Serotyping

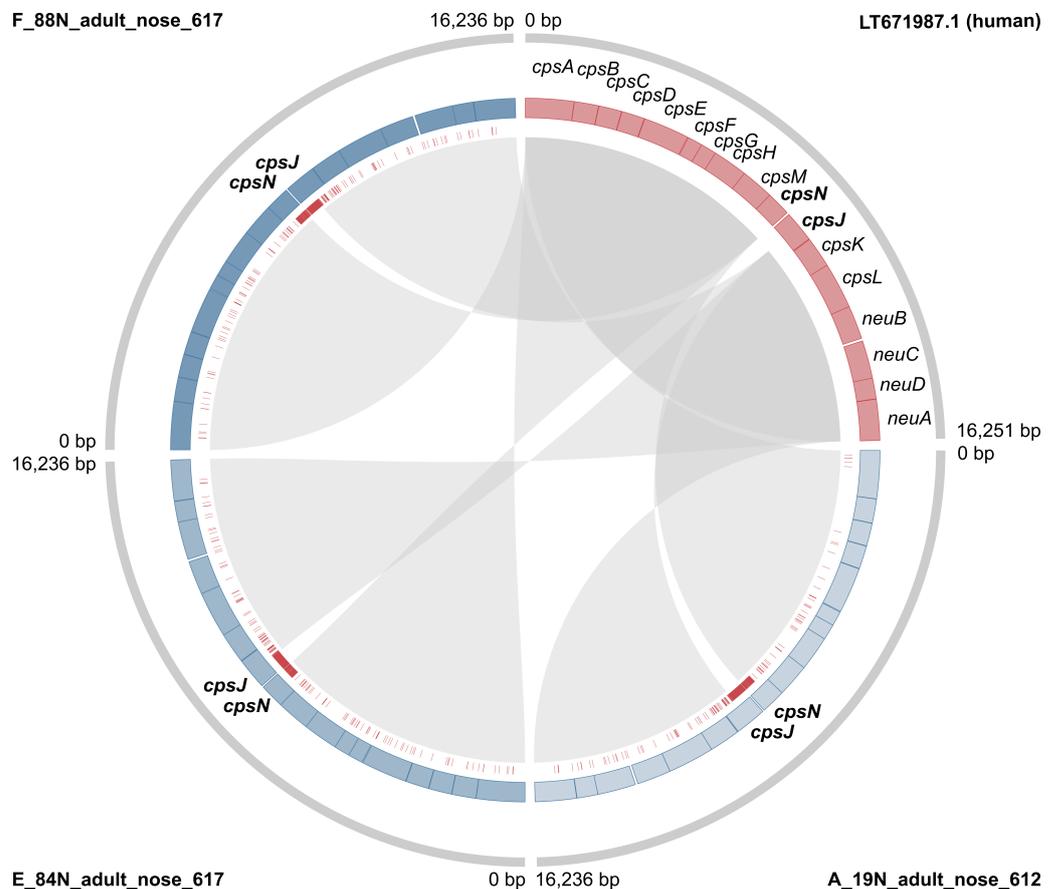
Our dataset included 113 ST belonging to 15 sublineages (see Supplementary File S1). All camel GBS belonged to SL609 (Figure 1), with the exception of a single ST1 isolate [27]. SL609, which did not include any GBS genomes from other host species, comprised 12 ST (see Supplementary File S1) and capsular types II–VI (Figure 1). Within SL609, fastbaps identified five subpopulations (Figure 1). These largely corresponded to clonal groups (CG), e.g., fastbaps population 2 corresponds to CG615 (including one ST612 and all ST615, all with capsular type II except for one type III) and fastbaps population 5 corresponds to CG616 (including ST616, ST1653 and ST1654, all with capsular type III).



**Figure 1.** Midpoint-rooted trees of the global group B *Streptococcus* (GBS) population, and of the camel-specific sublineage (SL)609. All but one sequence type (ST)1 camel genome described to date belong to this SL. Clonal groups (CG) within SL609 are indicated on the tree branches. Leaves and branches are coloured based on fastbaps populations, as indicated by the legend. Coloured blocks below show capsular type, sample type, and selected features of interest based on Scoary results.

Across the entire dataset, results obtained with the two *cps* typing methods largely agreed, but there were some inconsistencies and ambiguities. Among camel isolates, one ST615 had capsular type III based on the method developed by Metcalf et al. [33], but type II or possibly type IV (strong second best match) based on the GBS-SBG method from Tiruvayipati et al. [35]. This *cps* locus showed pseudogenisation (see below); in particular, there was an assembly gap upstream from the gene *cpsK*, possibly due to the presence of MGE-associated repeated sequences. ST616 isolates had capsular type III based on the Metcalf method, with nearly identical matches for subtypes III-4 and III-1 (Bit Score: 0.989 and 0.997, respectively) based on GBS-SBG. Finally, three isolates gave one best match with the first method, but two possible matches with the second (ST612 type IV vs. IV/V; ST614 type V vs. V/IX; ST617 VI vs. VI/III-1) (see Supplementary File S1). Discrepancies were also observed between results from the two methods used in the current study, and those

previously published for the same genomes: several isolates with serotype IV based on our analyses had previously been reported as capsular type Ia based on a multiplex PCR method [18,34]. In addition, an isolate with capsular type III based on our analyses was previously reported as type VI (ILRI021) with the same band-based method [18]. To explore the potential origin of those differences, the (*cps*) sequences of all camel capsular type IV isolates were aligned with a reference sequence from human GBS type IV (LT671987.1) using Geneious Prime v2021.1 [47], revealing high SNP density in the region straddling *cpsN* and *cpsJ* (Figure 2). Based on bioinformatic analysis, the SNPs likely prevent binding of one of the primers for *cpsJ*, which is used to distinguish type IV from Ia [34] (Figure S3). We also detected a series of InDels (insertions/deletions) in capsular type IV, leading to pseudogenisation in a minority of isolates ( $n = 4$ ). Pseudogenisation was also observed in other capsular types (type II,  $n = 1$ ; type III,  $n = 27$ ; type V,  $n = 2$ ; type VI,  $n = 3$ ), and involved multiple genes within the capsular operon, in particular *cpsI* ( $n = 13$  genomes), and *cpsE* ( $n = 7$  genomes). In total, 37 camel GBS genomes showed pseudogenisation of the capsular operon (Supplementary File S1).



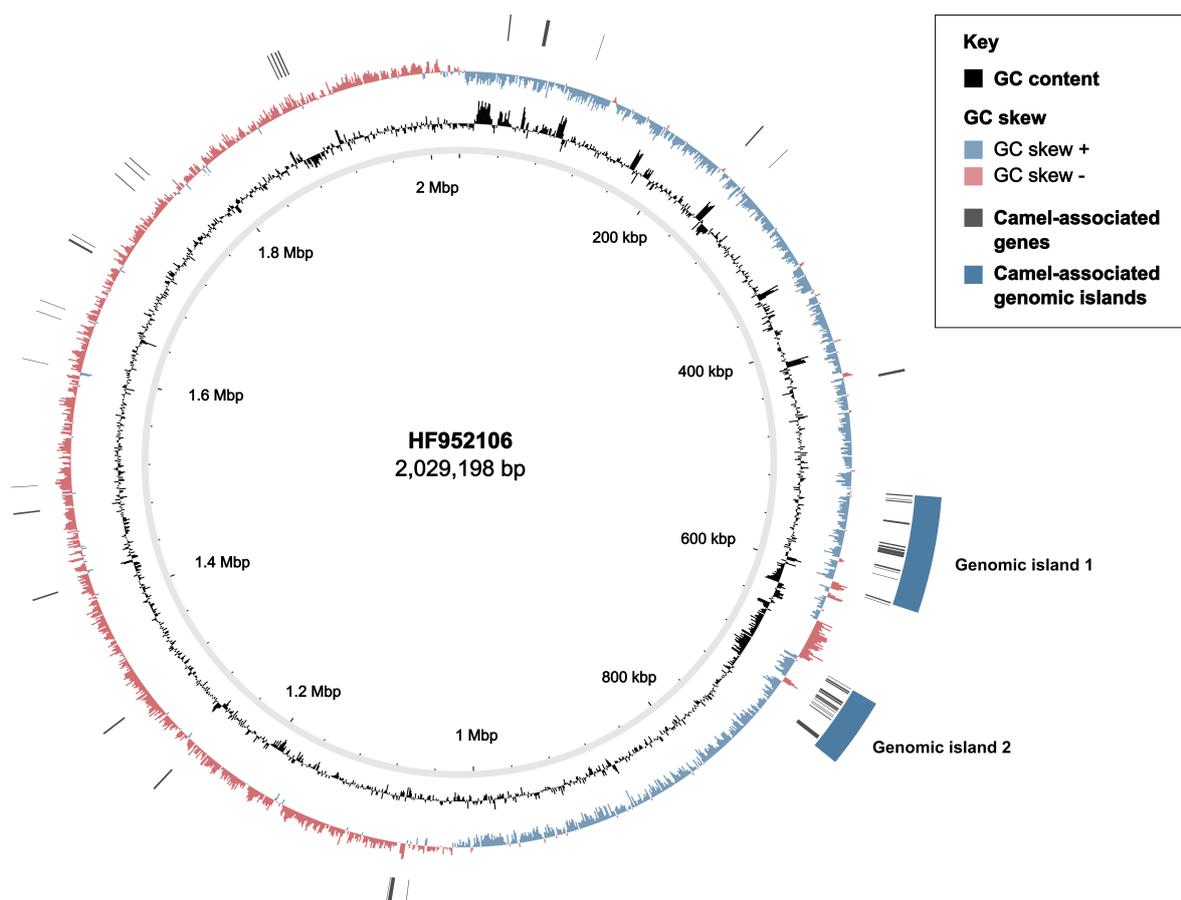
**Figure 2.** The capsular locus of the human reference genome for Group B *Streptococcus* capsular type IV is compared with three sequences that represent the overall phylogenetic diversity of type IV in GBS from camels (see Figure S4). External strips show the full *cps* locus (grey), internal strips the *cps* genes (pink: human, blue shades: camel), and the red bars indicate single nucleotide polymorphisms relative to the human GBS reference genome. Internal links show gene identity (default BLAST settings).

### 3.2. Markers of Camel-Association Within GBS

Only genes whose  $p$ -values reported by Scoary were  $<0.001$  are discussed in this and in further sections; the exact  $p$ -values can be found in the Supplementary Material.

Among the 100 best-scoring genes in the host-level GWAS, 28 were positively and uniquely associated with GBS from camels (camel-specific; SP = 100%) (Supplementary File S2).

High-scoring genes largely mapped to two genomic islands (GEI) (Figure 3). GEI1 included genes for carbohydrate metabolism (sugar phosphatase), DNA modification (DNA topoisomerase and DNA cytosine methyltransferase) and protein channels (voltage-gated chloride channel family *eriC*). In GEI2, high-scoring genes included a sensor histidine kinase for copper and silver (*cusS*), IMMA/IrrE family metallo-endopeptidase, and CPBP family intramembrane metalloproteases (highly sensitive and 100% specific for camel GBS, Supplementary File S2), as well as a transcriptional regulator (Rgg/GadR/MutR), an oxidoreductase (Gfo/Idh/MocA), a histidine phosphatase, an N-acetyltransferase (GNAT family), an ATP-dependent DNA helicase and a pyrimidine nucleotidase (YjjG). GEI2 showed lower GC content than the remainder of the GBS genome (GC GEI2: 31% vs. GC whole genome reference HF952106: 35%) (Figure 3). Alternating opposed GC skews were detected in this area of the genome and corresponded to several MGE (Figure S5), including GEI1 (positive GC skew, average GC content 35%), partial prophage *GBSInt1* immediately followed by partial *ICESp1108* from *Streptococcus pyogenes* [48] (opposed GC skew peaks, GC content 36% for the prophage and 30% for the ICE), *Tn916* (positive GC skew, GC content 38%), *ICESp1108* (negative GC skew, GC content 31%), and GEI2 (positive GC skew, GC content 31%). Other significant hits included a carbohydrate kinase (FGGY family), a PTS sugar transporter subunit IIC, three insertion sequences (IS) (*IS1562*, *ISLre2* and *ISSag9*) and a type I restriction modification system (RMS) (subunit M) associated with camel GBS.



**Figure 3.** Circular genome map showing camel-associated accessory genome content in Group B *Streptococcus* based on Scoary, mapped to reference genome HF952106 (camel GBS, ST609, capsular type V). From inner to outer, circles show: position of the genome, GC content, GC skew, camel-associated genes (black blocks) and camel-associated genomic islands (GEI) (blue blocks). A lower GC content can be appreciated in the area corresponding to island 2 (GC island 2: 31% vs. GC whole genome: 35%) and in between the two GEI. Opposed GC skew, corresponding to the partial sequence of *ICESp1108*, can be observed in this area of the genome. Plot was generated with BRIG v0.95.

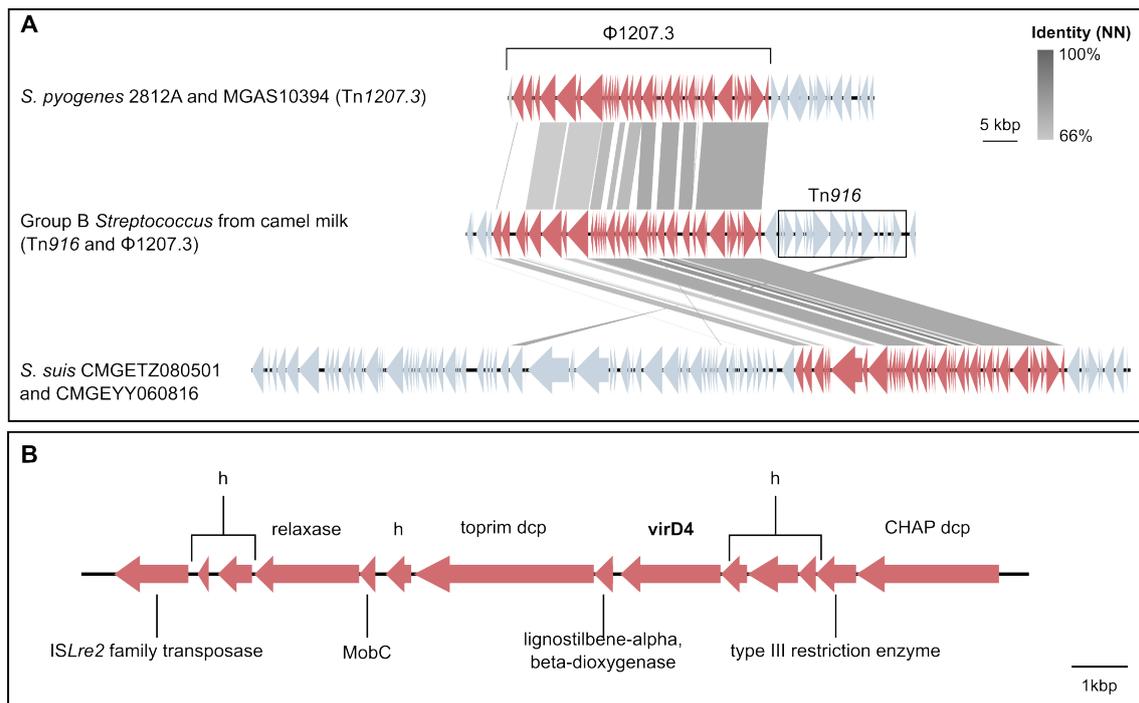
### 3.3. Markers of Mastitis-Association within Camel GBS

Genes that were strongly and positively associated with isolates from milk compared to those from other camel samples (Supplementary File S3) included a hyaluronate lyase (*hylB*) variant (Figure 1), genes coding for transporters and enzymes (ABC transporter, NAD-dependent malic enzyme, deacetylase, phosphatidate cytidylyltransferase, epoxyqueuosine reductase), and five *cps* gene variants (*cpsC*, *cpsH*, *cpsI*, *cpsJ* and *cpsK*). These *cps* variants were associated with capsular type III and ST616, which is the predominant serotype/ST combination among GBS isolates from camel milk [27].

Two GEI carrying milk-associated genes were identified (Figure 4): Tn916, which carries *tet(M)*, which was previously associated with milk isolates from camels [20,27], and a 14-gene cluster of ~15,000 bp that shows signatures of mobility (transposase, relaxase and mobilisation genes) and carries virulence gene (*virD4*). Tn916 was often linked to a long gene cluster (~43,000 bp, Figure 4A) that was also significantly associated with milk isolates. PHASTER identified this region as an incomplete prophage because it was lacking its integrase gene, indicating that it might be hijacking Tn916 to be mobilised. When the region surrounding the Tn916-prophage combination was blasted against the general ICEberg database [49] with multigene BLAST, it showed high sequence similarity with a segment of Tn1207.3 from *Streptococcus pyogenes* strains 2812A [50] and MGAS10394 (Figure 4), and with two composite MGE from *Streptococcus suis* (CMGETZ080501 and CMGEYY060816) (Figure 4A). This prophage, named  $\phi$ 1207.3 [51], carries a type II toxin/antitoxin system with a Phd/YefM family antitoxin followed by a Doc (death-on-curing) family toxin. Among the genomes that carried Tn916 ( $n = 101$ ), 12 did not carry  $\phi$ 1207.3, whereas only a single genome (ST617) was  $\phi$ 1207.3-positive but *tet(M)*/Tn916-negative (Supplementary File S1). Accuracy (combined sensitivity and specificity) for milk-associated genes was generally not as high as for host-associated genes (Supplementary File S3). *Lac.2* was detected in 70% of milk isolates, but sensitivity to lactose genes as an indicator of the milk-derived phenotype was low (1.2 to 58%) due to the existence of multiple variants of most *lac* genes: (*lacA*,  $n = 3$ ; *lacB*,  $n = 2$ ; *lacC*,  $n = 4$ ; *lacD*,  $n = 3$ ; *lacE*,  $n = 2$ ; *lacF*,  $n = 4$ ; *lacG*,  $n = 3$ ; *lacR*,  $n = 5$ ; *lacX*,  $n = 2$ ). This also explains the comparatively low Scoary rankings for this milk-associated operon (positions 104 *lacF\_1*, 150 *lacX\_1*, and 243-965 for *lacABCDEG*).

### 3.4. Mobile Genetic Elements in Camel GBS

Based on PHASTER, 16 (11.5%) of the camel GBS genomes carried complete prophages, 51 (36.7%) carried complete and incomplete (remnant) prophages, 61 (43.9%) carried remnant prophages only, and 11 (7.9%) carried no prophages. No more than two complete prophages were detected per GBS genome. The BLAST-based method identified prophage integrase types in 88 (63.3%) of the camel GBS genomes, including all genomes with complete prophages according to PHASTER, and with up to three hits per genome. The most common integrase types were *GBSInt11.1* ( $n = 55$ ) and *GBSInt10* ( $n = 35$ ), followed by *GBSInt1* ( $n = 15$ ), *GBSInt6.1* ( $n = 8$ ), *GBSInt11.2* ( $n = 5$ ) and *GBSInt5* ( $n = 3$ ). Two new PICI were detected (Figure S6): PICI3 (length = 13,723 bp), which shares its insertion site with PICI1 and PICI2 (*rpsD*—30S ribosomal protein S4), and PICI4 (length = 12,919 bp), which is found between a hypothetical gene and an aminoacyl-tRNA synthetase gene (amino acid sequences for *PICI3Int* and *PICI4Int* can be found in the Supplementary Material). Sixty-five genomes (46.8%) were positive for at least one PICI integrase. *PICI4Int*, which carries the virulence-associated protein E (*vapE*) was found across all fastbaps camel populations, particularly in population 4 (CG617 and CG1652), which includes multiple serotypes and sample types and comprises 72.4% of the 58 *PICI4Int*-positive camel GBS genomes.



**Figure 4.** Genomic islands (GEI) that were associated with milk-derived Group B *Streptococcus* from camels upon within-host comparison with extra-mammary isolates: Tn916- $\phi$ 1207.3 (**A**) and *virD4* GEI (**B**). (**A**) Visualisation of a BLASTn comparison between the Tn916- $\phi$ 1207.3 associated with camel milk isolates (centre, from isolate 1M) and mobile elements from *Streptococcus pyogenes* (top, Tn1207.3 *S. pyogenes* strain 2812A and MGAS10394) and *Streptococcus suis* (bottom, CMGETZ080501 and CMGEYY060816). Figure was obtained with Easyfig v2.2.2 [46] and modified with Inkscape v1.2.1 ([www.inkscape.org](http://www.inkscape.org)). (**B**) Visualisation of gene orientation and annotation of the *virD4* GEI from camel GBS isolate 1M (see Supplementary File S1). The contig on which this element was located ended at the right hand side of the CHAP domain containing protein (dcp). Hypothetical proteins are indicated with h.

#### 4. Discussion

Genomic analysis of GBS in humans, cattle and fishes has shown that events in both the host population and the pathogen population may contribute to niche adaptation: the emergence of GBS in fishes is attributed to the global expansion and intensification of fish farming [3,52] combined with the acquisition of niche-associated MGE, notably locus 3, which includes genes for galactose metabolism [6,7], whilst adaptation of GBS to humans and cattle, respectively, is attributed to the acquisition of tetracycline resistance and the Lac.2 operon encoding lactose metabolism [53,54]. Expansion in a new host population may be followed by host specialisation, as exemplified by CC552, which is restricted to fish and other cold-blooded animals as a result of genome reduction [10], and by bovine-adapted lineage CC61/67, which is undergoing pseudogenisation of its capsule operon, resulting in loss of a key virulence factor for infection of humans [55]. Here, we present the first large-scale genomic study of GBS from camels. Camel husbandry in Africa is undergoing intensification due to expansion of the camel milk market [27], providing a unique opportunity to explore the current population composition and, subsequently, its evolution under changing selection pressures.

We show that the GBS population from camels is largely distinct from GBS found in other host species in terms of its core genome. With the exception of one isolate belonging to ST1 [27], all camel GBS genomes cluster in one SL (Figure 1). Earlier studies had identified two groups of camel GBS isolates based on MLST maximum-likelihood phylogeny [18], but when considering the entirety of the core genome and all known GBS SL from different hosts, camel GBS forms a single monophyletic lineage (SL609). Not only did camel

GBS constitute a unique lineage, but it also harboured a unique variant of capsular type IV. The divergence between capsular type IV in GBS from humans and camels affected capsular type prediction in camel isolates based on a PCR method developed for human isolates [34]. The high prevalence of serotype IV and VI among camel GBS, as described previously [25], and the novel variant of serotype IV may provide a reservoir of capsular variants that are not covered by human GBS vaccines that are currently under development. This could potentially lead to the evolution of vaccine escape mutants, similar to what has been observed in *Streptococcus pneumoniae* [56], especially if the transfer of capsular operons between strains—and the transfer of strains between host species—is possible. Although rare, the detection of ST1 in camels [25] and evidence for capsular switching in GBS [57–59] suggest that both processes can occur. An example of capsular switching in camel GBS is shown in Figure 1, where a single member of CG615 has capsular type III whilst the remaining eleven isolates in this CG have capsular type II. As in cattle GBS [55], there is evidence of pseudogenisation in camel GBS. This may limit animal-to-human transmission, as the GBS capsule is considered essential for pathogenicity in humans, where it limits complement deposition and phagocytosis and masks surface proteins to avoid stimulating the immune response [60,61]. Pseudogenisation, however, only affects 37% of the bovine-associated SL61/67 and 27% of camel-associated SL609, i.e., the capsular operon is intact in the majority of isolates from both SL so spillover may still be possible. Indeed, the detection of CC61/67 in humans, albeit being extremely rare, has already been described in China [62], whereas data from east Africa are too scarce to prove or rule out the possibility of human infection with SL609.

Using GWAS, we identified accessory genome content that is both uniquely associated with camel GBS (SP = 100%) and almost universally present in camel GBS (SE > 98%). As in bovine [54] and piscine [6] GBS, camel-associated accessory genome content was concentrated in GEI. These GEI were found in an area of the genome that shows signatures of multiple integration/recombination events, including alternating opposed GC skew and variation of the GC content. This genomic region also contained a partial ICE from *S. pyogenes*. Likewise, accessory genome content that was significantly associated with isolates from milk based on within-host GWAS included a prophage that has been described as part of other ICE/MGE in *S. pyogenes* and *S. suis*. Although *S. pyogenes* is thought to be an exclusively human pathogen, whereas *S. suis* is primarily associated with pigs [50,51], MGE with high homology to *S. pyogenes* has also been reported in bovine GBS [15]. In addition to ICE, prophages may contribute to HGT. Complete prophages of type GBS10 and GBS11 [12] were common among camel GBS. They were previously described only among human GBS isolates, and in one seal isolate that likely originated from the anthropogenic contamination of the aquatic environment [6]. Phages can facilitate DNA exchanges through transduction [63], including lateral transduction, which involves extended areas of the genome [64], and exceeds chromosomal mobility of classical MGE [65]. A third mechanism that may contribute to HGT between GBS across host species are PICIs. Previous work identified a low diversity of PICI in GBS [12] compared to other bacterial species [11], with only two types detected. PICI1 was widespread across isolates from multiple hosts (humans, fish, cattle, a dog and a dolphin), whilst PICI2 was found uniquely in a camel GBS genome from Kenya [12]. In this study, we identified two new PICI in camel GBS. PICI3 had the same integration site as PICI1 and PICI2, the *rpsD* gene, confirming this site is an important hotspot for recombination of PICI in GBS. The second novel PICI, PICI4, carried a virulence-associated protein (*vapE*) that has also been described in *Rhodococcus equi*, *S. suis* and *S. pneumoniae* [66–68], reinforcing the notion that pangenome evolution needs to be explored and understood beyond host- and pathogen species level.

Genes that were associated with the camel as a host included those involved in various metabolic processes, e.g., carbohydrate and metal utilisation. The utilisation of substrates that are present in certain niches is a well-described mechanism of bacterial adaptation [14]. In GBS, it manifests at the host level, as illustrated by the galactose genes associated with piscine GBS [6,7] and by sugar phosphatases associated with the camel host, and also at the

organ level, as demonstrated by the association of the lactose operon Lac.2 with isolates originating from milk and, hence, the mammary gland. In camel GBS, Lac.2 was found in 70% of milk isolates, compared to 98–100% of bovine milk isolates [2,15], and the presence of Lac.2 is not exclusive to milk isolates in camels. This suggests that Lac.2 is not necessary for GBS to successfully colonise and establish infections in the udder of camels, but that it offers an evolutionary advantage in adaptation to the mammary gland, similar to the situation reported for *Klebsiella* spp. and *E. coli* [69,70]. The role of the *cusS* gene, normally involved in sensing copper and silver, is unclear. The ability to uptake and utilise essential metals can counteract the host immune defences, which usually sequester these molecules to protect the host from infection [71], but *cusS* has also been described as a contributor to metal homeostasis or metal resistance, notably in *Klebsiella* spp. and *E. coli* [72,73]. In camels, there is a third albeit speculative possibility: copper levels in camel plasma have been reported to be lower than those in bovine plasma [74]. Thus, there may be an advantage to having the ability to acquire a scarce resource that supports cellular processes.

Another milk-associated gene was a gene variant of *hylB*, whose product is able to degrade certain components of the extracellular matrix. Although *hylB* is believed to contribute significantly to invasion [61,75], hyaluronidase activity is not essential for human GBS infections [76], and GBS associated with neonatal disease may show a disrupted *hylB* gene [77]. To our knowledge, there is no evidence of the functional relevance of *hylB* in the context of milk or mastitis. Considering the population structure within the camel-associated GBS lineage, the *hylB* variant may show a statistical association with camel milk through its association with CG616. Transposon Tn916 was linked to an incomplete prophage ( $\phi$ 1207.3), and both were significantly associated with camel milk in the within-host GWAS. Unlike the element described in *S. pyogenes*,  $\phi$ 1207.3 in GBS did not carry macrolide resistance genes. In our experience, camel owners in East Africa mostly use tetracycline, penicillin or streptomycin, whereas we are not aware of macrolide use or a potential selective advantage of macrolide resistance in camels. The incomplete prophage from camel GBS did encode the same toxin/antitoxin (TA) system (Phd/Doc) as in *S. pyogenes*. Phd/Doc TA systems have been described in various Gram-positive [78,79] and Gram-negative bacterial species [80,81]. Although the biological impact of TA systems in virulence and pathogenicity remains unclear [79], they seem to confer an advantage under stressful environmental conditions for the bacteria [79,82]. Finally, a gene cluster that was associated with camel milk and CG616 carried the *virD4* gene, a previously identified type IV secretion system (T4SS) [83]. This gene was recently described in GBS as significantly associated with CC19 [17], a lineage primarily associated with human infections. It is also found in other bacterial species, including *S. suis* [84], and it is associated with conjugation, the translocation of virulence factors [85,86], anti-phagocytic activity and a pro-inflammatory effect [84].

Although our study was limited geographically (all GBS genomes from camels available to date originated from Kenya and Somalia) and could not elucidate the functional relevance of camel- or milk-associated genes identified by GWAS, we clearly demonstrate the existence of a GBS lineage (SL609) that is both fully specific to camels, and encompasses all but one known camel GBS isolates. Although the camel GBS core genome is phylogenetically distinct, we identify commonalities in the accessory genome of human and camel GBS, e.g., Tn916 and shared prophage integrase sites, and of bovine and camel GBS, e.g., Lac.2. Moreover, we identified multiple MGE, e.g., ICESp1108 and prophage  $\phi$ 1207.3, that show high sequence similarity to MGE from other streptococcal species, including those that are largely limited to other host species, notably humans or pigs. Thus, camel GBS seems to have an expansive mobilome in terms of host and pathogen species of origin. As camel husbandry and milk production intensify in light of climate change, the animal's drought resilience, and the growing demand for camel milk, the emergence and expansion of new GBS strains in or derived from camels appears to be a distinct possibility. Ongoing surveillance of the prevalence and population composition of GBS in camels is recommended so that the emergence of resistant, virulent, or zoonotic strains can be detected early.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/pathogens11091025/s1>. The supplementary files include: (i) Supplementary File S1: accession numbers and metadata for all 680 genomes included in the study, results of bioinformatic analyses carried out on the whole dataset (assembly quality control, multilocus sequence-typing, attribution to sublineage, capsular locus typing with two bioinformatic methods—Metcalf et al. and Tiruvayipati et al.), and results of bioinformatic analyses carried out on camel genomes only (fastbaps, pseudo-genomisation of *cps* locus, PHASTER, phage integrase typing and PICI detection); (ii) Supplementary File S2: Scoary results from the comparison between GBS genomes from camels vs. GBS genomes of other host species; (iii) Supplementary File S3: Scoary results from the comparison between GBS genomes from camel milk vs. GBS genomes from other sample types in camels; (iv) Supplementary File S4: Additional figures and sequences for the integrase genes of PICI3 and PICI4.

**Author Contributions:** Conceptualisation, C.C. and R.N.Z.; methodology, data curation and formal analysis, C.C.; writing—original draft preparation, C.C.; writing—review and editing, C.C., R.N.Z., T.L.F. and D.S.; visualisation, C.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was funded by the University of Glasgow College of Medical, Veterinary and Life Sciences Doctoral Training Programme 2017–2021 (to CC). TLF was supported by a Biotechnology and Biological Sciences Research Council (BBSRC) Discovery Fellowship (BB/R012075/1). DS was supported by the Swedish Research Council, grant number VR2015-03583.

**Data Availability Statement:** Genomic data for this study are available in publicly accessible repositories (either at the National Centre for Biotechnology Information or at the European Nucleotide Archive web pages, please consult the Supporting Material for accession numbers). New genomic data were made available in the European Nucleotide Archive, project number PRJEB53664. Data resulting from the analyses described in this paper are all contained within the article or the Supplementary Material.

**Acknowledgments:** We acknowledge Swaine Chen, Nguyen Ngoc Phuoc and Wanna Sirimanapong for the contribution of sequence data from Asia, Gabriele Arcari for his contribution to some of the manuscript's graphics, and Erika Chenais for her collaboration and supervision of DS PhD project.

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

## Abbreviations

The following abbreviations are used in this manuscript:

CC	clonal complex
CG	clonal group
GBS	Group B <i>Streptococcus</i>
GEI	genomic islands
GWAS	genome-wide association study
MGE	mobile genetic elements
ML	maximum-likelihood
MLST	multilocus sequence-typing
NT	non-typeable
PICI	phage-inducible chromosomal islands
SL	sublineage
ST	sequence type
TcR	tetracycline resistance

## References

- Seale, A.C.; Bianchi-Jassir, F.; Russell, N.J.; Kohli-Lynch, M.; Tann, C.J.; Hall, J.; Madrid, L.; Blencowe, H.; Cousens, S.; Baker, C.J.; et al. Estimates of the burden of group B streptococcal disease worldwide for pregnant women, stillbirths, and children. *Clin. Infect. Dis.* **2017**, *65*, S200–S219. [[CrossRef](#)] [[PubMed](#)]
- Lyhs, U.; Kulkas, L.; Katholm, J.; Waller, K.P.; Saha, K.; Tomusk, R.J.; Zadoks, R.N. *Streptococcus agalactiae* Serotype IV Humans Cattle, Northern Europe. *Emerg. Infect. Dis.* **2016**, *22*, 2097. [[CrossRef](#)] [[PubMed](#)]
- Barkham, T.; Zadoks, R.N.; Azmai, M.N.A.; Baker, S.; Bich, V.T.N.; Chalker, V.; Chau, M.L.; Dance, D.; Deepak, R.N.; van Doorn, H.R.; et al. One hypervirulent clone, sequence type 283, accounts for a large proportion of invasive *Streptococcus agalactiae* isolated from humans and diseased tilapia in Southeast Asia. *PLoS Negl. Trop. Dis.* **2019**, *13*, e0007421. [[CrossRef](#)]

4. Paveenkittiporn, W.; Ungcharoen, R.; Kerdsin, A. *Streptococcus agalactiae* infections and clinical relevance in adults, Thailand. *Diagn. Microbiol. Infect. Dis.* **2020**, *97*, 115005. [[CrossRef](#)]
5. Brochet, M.; Couvé, E.; Zouine, M.; Vallaeys, T.; Rusniok, C.; Lamy, M.C.; Buchrieser, C.; Trieu-Cuot, P.; Kunst, F.; Poyart, C.; et al. Genomic diversity and evolution within the species *Streptococcus agalactiae*. *Microbes Infect.* **2006**, *8*, 1227–1243. [[CrossRef](#)] [[PubMed](#)]
6. Delannoy, C.M.; Zadoks, R.N.; Crumlish, M.; Rodgers, D.; Lainson, F.A.; Ferguson, H.; Turnbull, J.; Fontaine, M.C. Genomic comparison of virulent and non-virulent *Streptococcus agalactiae* Fish. *J. Fish Dis.* **2016**, *39*, 13–29. [[CrossRef](#)] [[PubMed](#)]
7. Richards, V.P.; Velsko, I.M.; Alam, T.; Zadoks, R.N.; Manning, S.D.; Pavinski Bitar, P.D.; Hasler, H.B.; Crestani, C.; Springer, G.H.; Probert, B.M.; et al. Population gene introgression and high genome plasticity for the zoonotic pathogen *Streptococcus agalactiae*. *Mol. Biol. Evol.* **2019**, *36*, 2572–2590. [[CrossRef](#)] [[PubMed](#)]
8. Viana, D.; Comos, M.; McAdam, P.R.; Ward, M.J.; Selva, L.; Guinane, C.M.; González-Muñoz, B.M.; Tristan, A.; Foster, S.J.; Fitzgerald, J.R.; et al. A single natural nucleotide mutation alters bacterial pathogen host tropism. *Nat. Genet.* **2015**, *47*, 361–366. [[CrossRef](#)] [[PubMed](#)]
9. Richardson, E.J.; Bacigalupe, R.; Harrison, E.M.; Weinert, L.A.; Lycett, S.; Vrieling, M.; Robb, K.; Hoskisson, P.A.; Holden, M.T.; Feil, E.J.; et al. Gene exchange drives the ecological success of a multi-host bacterial pathogen. *Nat. Ecol. Evol.* **2018**, *2*, 1468. [[CrossRef](#)] [[PubMed](#)]
10. Rosinski-Chupin, I.; Sauvage, E.; Mairey, B.; Mangenot, S.; Ma, L.; Da Cunha, V.; Rusniok, C.; Bouchier, C.; Barbe, V.; Glaser, P. Reductive evolution in *Streptococcus agalactiae* and the emergence of a host adapted lineage. *BMC Genom.* **2013**, *14*, 252. [[CrossRef](#)]
11. Penadés, J.R.; Christie, G.E. The phage-inducible chromosomal islands: A family of highly evolved molecular parasites. *Annu. Rev. Virol.* **2015**, *2*, 181–201. [[CrossRef](#)] [[PubMed](#)]
12. Crestani, C.; Forde, T.L.; Zadoks, R.N. Development and application of a prophage integrase typing scheme for Group B *Streptococcus*. *Front. Microbiol.* **2020**, *11*, 1993. [[CrossRef](#)]
13. Rouli, L.; Merhej, V.; Fournier, P.E.; Raoult, D. The bacterial pangenome as a new tool for analysing pathogenic bacteria. *New Microbes New Infect.* **2015**, *7*, 72–85. [[CrossRef](#)]
14. Sheppard, S.K.; Guttman, D.S.; Fitzgerald, J.R. Population genomics of bacterial host adaptation. *Nat. Rev. Genet.* **2018**, *19*, 549–565. [[CrossRef](#)] [[PubMed](#)]
15. Crestani, C.; Forde, T.L.; Lycett, S.J.; Holmes, M.A.; Fasth, C.; Persson-Waller, K.; Zadoks, R.N. The fall and rise of group B *Streptococcus* Dairy Cattle: Reintroduction due to human-to-cattle host jumps? *Microb. Genom.* **2021**, *7*, 000648. [[CrossRef](#)] [[PubMed](#)]
16. Botelho, A.C.N.; Oliveira, J.G.; Damasco, A.P.; Santos, K.T.; Ferreira, A.F.M.; Rocha, G.T.; Marinho, P.S.; Bornia, R.B.; Pinto, T.C.; Américo, M.A.; et al. *Streptococcus agalactiae* Carriage Pregnant Women Living Rio De Janeiro, Brazil, Over a Period of Eight Years. *PLoS ONE* **2018**, *13*, e0196925. [[CrossRef](#)]
17. Gori, A.; Harrison, O.B.; Mlia, E.; Nishihara, Y.; Chan, J.M.; Msefula, J.; Mallewa, M.; Dube, Q.; Swarthout, T.D.; Nobbs, A.H.; et al. Pan-GWAS of *Streptococcus agalactiae* Highlights Lineage-Specific Genes Associated with Virulence and Niche Adaptation. *mBio* **2020**, *11*, e00728–20. [[CrossRef](#)] [[PubMed](#)]
18. Fischer, A.; Liljander, A.; Kaspar, H.; Muriuki, C.; Fuxelius, H.H.; Bongcam-Rudloff, E.; de Villiers, E.P.; Huber, C.A.; Frey, J.; Daubenberger, C.; et al. Camel *Streptococcus agalactiae* Populations are Associated with Specific Disease Complexes and Acquired the Tetracycline Resistance Gene *tetM* via a Tn916-Like Element. *Vet. Res.* **2013**, *44*, 86. [[CrossRef](#)] [[PubMed](#)]
19. Zubair, S.; de Villiers, E.; Younan, M.; Andersson, G.; Tettelin, H.; Riley, D.; Jores, J.; Bongcam-Rudloff, E.; Bishop, R.P. Genome sequences of two pathogenic *Streptococcus agalactiae* isolates from the One-humped camel *Camelus dromedarius*. *Genome Announc.* **2013**, *1*, e00515–13. [[CrossRef](#)] [[PubMed](#)]
20. Seligsohn, D.; Nyman, A.; Younan, M.; Sake, W.; Persson, Y.; Bornstein, S.; Maichomo, M.; de Verdier, K.; Morrell, J.; Chenais, E. Subclinical mastitis in pastoralist dairy camel herds in Isiolo, Kenya: Prevalence, risk factors, and antimicrobial susceptibility. *J. Dairy Sci.* **2020**, *103*, 4717–4731. [[CrossRef](#)] [[PubMed](#)]
21. Iyer, A.P.; Albaik, M.; Baghallab, I. Mastitis in camels in African and Middle East countries. *J. Bacteriol. Parasitol.* **2014**, *5*, 1. [[CrossRef](#)]
22. Obied, A.; Bagadi, H.; Mukhtar, M. Mastitis in *Camelus dromedarius* Somat. Cell Content Camels' Milk. *Res. Vet. Sci.* **1996**, *61*, 55–58. [[CrossRef](#)]
23. Bekele, T.; Molla, B. Mastitis in lactating camels (*Camelus dromedarius*) in Afar region, north-eastern Ethiopia. *Berl. Und Münchener Tierärztliche Wochenschr.* **2001**, *114*, 169–172.
24. Younan, M.; Bornstein, S. Lancefield group B and C streptococci in East African camels (*Camelus dromedarius*). *Vet. Rec.* **2007**, *160*, 330–335. [[CrossRef](#)] [[PubMed](#)]
25. Seligsohn, D.; Crestani, C.; Gitahi, N.; Flodin, E.L.; Chenais, E.; Zadoks, R.N. Investigation of extramammary sources of Group B *Streptococcus* reveals its unusual ecology and epidemiology in camels. *PLoS ONE* **2021**, *16*, e0252973. [[CrossRef](#)]
26. Rothen, J.; Schindler, T.; Pothier, J.F.; Younan, M.; Certa, U.; Daubenberger, C.; Pflüger, V.; Jores, J. Draft genome sequences of seven *Streptococcus agalactiae* strains isolated from *Camelus dromedarius* at the horn of Africa. *Genome Announc.* **2017**, *5*, e00525–17. [[CrossRef](#)]

27. Seligsohn, D.; Crestani, C.; Gitahi, N.; Flodin, E.L.; Chenais, E.; Zadoks, R.N. Genomic analysis of Group B *Streptococcus* from milk demonstrates the need for improved biosecurity: A cross-sectional study of pastoralist camels in Kenya. *BMC Microbiol.* **2021**, *21*, 217. [CrossRef]
28. Brynildsrud, O.; Bohlin, J.; Scheffer, L.; Eldholm, V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* **2016**, *17*, 238. [CrossRef]
29. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [CrossRef] [PubMed]
30. Hasman, H.; Saputra, D.; Sicheritz-Ponten, T.; Lund, O.; Svendsen, C.A.; Frimodt-Møller, N.; Aarestrup, F.M. Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. *J. Clin. Microbiol.* **2014**, *52*, 139–146. [CrossRef]
31. Inouye, M.; Dashnow, H.; Raven, L.A.; Schultz, M.B.; Pope, B.J.; Tomita, T.; Zobel, J.; Holt, K.E. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med.* **2014**, *6*, 90. [CrossRef]
32. Larsen, M.V.; Cosentino, S.; Rasmussen, S.; Friis, C.; Hasman, H.; Marvig, R.L.; Jelsbak, L.; Sicheritz-Pontén, T.; Ussery, D.W.; Aarestrup, F.M.; et al. Multilocus sequence typing of total-genome-sequenced bacteria. *J. Clin. Microbiol.* **2012**, *50*, 1355–1361. [CrossRef]
33. Metcalf, B.; Chochua, S.; Gertz, R., Jr.; Hawkins, P.; Ricaldi, J.; Li, Z.; Walker, H.; Tran, T.; Rivers, J.; Mathis, S.; et al. Short-read whole genome sequencing for determination of antimicrobial resistance mechanisms and capsular serotypes of current invasive *Streptococcus agalactiae* recovered in the USA. *Clin. Microbiol. Infect.* **2017**, *23*, 574.e7–574.e14. [CrossRef]
34. Imperi, M.; Pataracchia, M.; Alfarone, G.; Baldassarri, L.; Orefici, G.; Creti, R. A multiplex PCR assay for the direct identification of the capsular type (Ia to IX) of *Streptococcus agalactiae*. *J. Microbiol. Methods* **2010**, *80*, 212–214. [CrossRef] [PubMed]
35. Tiruvayipati, S.; Ying, T.W.; Barkham, T.; Chen, S.L. GBS-SBG—GBS serotyping by genome sequencing. *Microb. Genom.* **2021**, *7*, 000688. [CrossRef] [PubMed]
36. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [CrossRef]
37. Kozlov, A.M.; Darriba, D.; Flouri, T.; Morel, B.; Stamatakis, A. RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **2019**, *35*, 4453–4455. [CrossRef] [PubMed]
38. Tonkin-Hill, G.; Lees, J.A.; Bentley, S.D.; Frost, S.D.; Corander, J. Fast hierarchical Bayesian analysis of population structure. *Nucleic Acids Res.* **2019**, *47*, 5539–5549. [CrossRef] [PubMed]
39. RStudio Team. *RStudio: Integrated Development Environment for R*; RStudio, PBC: Boston, MA, USA, 2020. Available online: <http://www.rstudio.com/> (accessed on 1 September 2022).
40. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2018. Available online: <https://www.R-project.org/> (accessed on 1 September 2022).
41. Page, A.J.; Cummins, C.A.; Hunt, M.; Wong, V.K.; Reuter, S.; Holden, M.T.; Fookes, M.; Falush, D.; Keane, J.A.; Parkhill, J. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **2015**, *31*, 3691–3693. [CrossRef] [PubMed]
42. 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] [PubMed]
43. Hadfield, J.; Croucher, N.J.; Goater, R.J.; Abudahab, K.; Aanensen, D.M.; Harris, S.R. Phandango: An interactive viewer for bacterial population genomics. *Bioinformatics* **2018**, *34*, 292–293. [CrossRef] [PubMed]
44. Krzywinski, M.; Schein, J.; Birol, I.; Connors, J.; Gascoyne, R.; Horsman, D.; Jones, S.J.; Marra, M.A. Circos: An information aesthetic for comparative genomics. *Genome Res.* **2009**, *19*, 1639–1645. [CrossRef] [PubMed]
45. Alikhan, N.F.; Petty, N.K.; Zakour, N.L.B.; Beatson, S.A. BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC Genom.* **2011**, *12*, 402. [CrossRef] [PubMed]
46. Sullivan, M.J.; Petty, N.K.; Beatson, S.A. Easyfig: A genome comparison visualizer. *Bioinformatics* **2011**, *27*, 1009–1010. [CrossRef] [PubMed]
47. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef] [PubMed]
48. Brenciani, A.; Tiberi, E.; Bacciaglia, A.; Petrelli, D.; Valardo, P.E.; Giovanetti, E. Two distinct genetic elements are responsible for *Erm*(TR)-Mediat. Erythromycin Resist. Tetracycline-Susceptible Tetracycline-Resist. Strains *Streptococcus pyogenes*. *Antimicrob. Agents Chemother.* **2011**, *55*, 2106–2112. [CrossRef] [PubMed]
49. Liu, M.; Li, X.; Xie, Y.; Bi, D.; Sun, J.; Li, J.; Tai, C.; Deng, Z.; Ou, H.Y. ICEberg 2.0: An updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res.* **2018**, *47*, D660–D665. [CrossRef] [PubMed]
50. Santagati, M.; Iannelli, F.; Cascone, C.; Campanile, F.; Oggioni, M.R.; Stefani, S.; Pozzi, G. The novel conjugative transposon Tn1207.3 Carries Macrolide Efflux Gene *mef*(A) *Streptococcus pyogenes*. *Microb. Drug Resist.* **2003**, *9*, 243–247. [CrossRef] [PubMed]
51. Iannelli, F.; Santagati, M.; Santoro, F.; Oggioni, M.R.; Stefani, S.; Pozzi, G. Nucleotide sequence of conjugative prophage  $\Phi$ 1207.3 (formerly Tn1207.3) carrying the *mef*(A)/*msr*(D) genes for efflux resistance to macrolides in *Streptococcus pyogenes*. *Front. Microbiol.* **2014**, *5*, 687. [CrossRef] [PubMed]
52. Kawasaki, M.; Delamare-Deboutteville, J.; Bowater, R.O.; Walker, M.J.; Beatson, S.; Zakour, N.L.B.; Barnes, A.C. Microevolution of *Streptococcus agalactiae* ST-261 from Australia indicates dissemination via imported tilapia and ongoing adaptation to marine hosts or environment. *Appl. Environ. Microbiol.* **2018**, *84*, e00859-18. [CrossRef]

53. Da Cunha, V.; Davies, M.R.; Douarre, P.E.; Rosinski-Chupin, I.; Margarit, I.; Spinali, S.; Perkins, T.; Lechat, P.; Dmytruk, N.; Sauvage, E.; et al. *Streptococcus agalactiae* clones infecting humans were selected and fixed through the extensive use of tetracycline. *Nat. Commun.* **2014**, *5*, 5544. [[CrossRef](#)]
54. Richards, V.P.; Lang, P.; Bitar, P.D.P.; Lefébure, T.; Schukken, Y.H.; Zadoks, R.N.; Stanhope, M.J. Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted *Streptococcus agalactiae*. *Infect. Genet. Evol.* **2011**, *11*, 1263–1275. [[CrossRef](#)] [[PubMed](#)]
55. Almeida, A.; Alves-Barroco, C.; Sauvage, E.; Bexiga, R.; Albuquerque, P.; Tavares, F.; Santos-Sanches, I.; Glaser, P. Persistence of a dominant bovine lineage of group B *Streptococcus* reveals genomic signatures of host adaptation. *Environ. Microbiol.* **2016**, *18*, 4216–4229. [[CrossRef](#)] [[PubMed](#)]
56. Brueggemann, A.B.; Pai, R.; Crook, D.W.; Beall, B. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. *PLoS Pathog.* **2007**, *3*, e168. [[CrossRef](#)]
57. Neemuchwala, A.; Teatero, S.; Athey, T.B.; McGeer, A.; Fittipaldi, N. Capsular switching and other large-scale recombination events in invasive sequence type 1 group B *Streptococcus*. *Emerg. Infect. Dis.* **2016**, *22*, 1941. [[CrossRef](#)] [[PubMed](#)]
58. Bellais, S.; Six, A.; Fouet, A.; Longo, M.; Dmytruk, N.; Glaser, P.; Trieu-Cuot, P.; Poyart, C. Capsular switching in group B *Streptococcus* CC17 Hypervirulent Clone: A future challenge for polysaccharide vaccine development. *J. Infect. Dis.* **2012**, *206*, 1745–1752. [[CrossRef](#)] [[PubMed](#)]
59. Martins, E.R.; Melo-Cristino, J.; Ramirez, M. Evidence for rare capsular switching in *Streptococcus agalactiae*. *J. Bacteriol.* **2010**, *192*, 1361–1369. [[CrossRef](#)]
60. Maisey, H.C.; Doran, K.S.; Nizet, V. Recent advances in understanding the molecular basis of group B *Streptococcus* Virulence. *Expert Rev. Mol. Med.* **2008**, *10*, e27. [[CrossRef](#)] [[PubMed](#)]
61. Herbert, M.A.; Beveridge, C.J.; Saunders, N.J. Bacterial virulence factors in neonatal sepsis: Group B *Streptococcus*. *Curr. Opin. Infect. Dis.* **2004**, *17*, 225–229. [[CrossRef](#)]
62. Li, L.; Wang, R.; Huang, Y.; Huang, T.; Luo, F.; Huang, W.; Yang, X.; Lei, A.; Chen, M.; Gan, X. High incidence of pathogenic *Streptococcus agalactiae* ST485 Strain Pregnant/puerperal women and isolation of hyper-virulent human CC67 strain. *Front. Microbiol.* **2018**, *9*, 50. [[CrossRef](#)]
63. Frost, L.S.; Leplae, R.; Summers, A.O.; Toussaint, A. Mobile genetic elements: The agents of open source evolution. *Nat. Rev. Microbiol.* **2005**, *3*, 722. [[CrossRef](#)] [[PubMed](#)]
64. Chen, J.; Quiles-Puchalt, N.; Chiang, Y.N.; Bacigalupe, R.; Fillol-Salom, A.; Chee, M.S.J.; Fitzgerald, J.R.; Penadés, J.R. Genome hypermobility by lateral transduction. *Science* **2018**, *362*, 207–212. [[CrossRef](#)]
65. Humphrey, S.; Fillol-Salom, A.; Quiles-Puchalt, N.; Ibarra-Chávez, R.; Haag, A.F.; Chen, J.; Penadés, J.R. Bacterial chromosomal mobility via lateral transduction exceeds that of classical mobile genetic elements. *Nat. Commun.* **2021**, *12*, 6509. [[CrossRef](#)] [[PubMed](#)]
66. Takai, S.; Hines, S.A.; Sekizaki, T.; Nicholson, V.M.; Alperin, D.A.; Osaki, M.; Takamatsu, D.; Nakamura, M.; Suzuki, K.; Ogino, N.; et al. DNA sequence and comparison of virulence plasmids from *Rhodococcus equi* ATCC 33701 and 103. *Infect. Immun.* **2000**, *68*, 6840–6847. [[CrossRef](#)] [[PubMed](#)]
67. Ji, X.; Sun, Y.; Liu, J.; Zhu, L.; Guo, X.; Lang, X.; Feng, S. A novel virulence-associated protein, *VapE*, *Streptococcus suis* Serotype 2. *Mol. Med. Rep.* **2016**, *13*, 2871–2877. [[CrossRef](#)] [[PubMed](#)]
68. Shaw, D. The association between satellite prophages and pneumococcal carriage and disease. In Proceedings of the Abstract for the 31st European Congress of Clinical Microbiology and Infectious Diseases, Online Event, 9–12 July 2021.
69. Holt, K.E.; Wertheim, H.; Zadoks, R.N.; Baker, S.; Whitehouse, C.A.; Dance, D.; Jenney, A.; Connor, T.R.; Hsu, L.Y.; Severin, J.; et al. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in *Klebsiella pneumoniae*, an urgent threat to public health. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E3574–E3581. [[CrossRef](#)]
70. Blum, S.; Heller, E.; Krifucks, O.; Sela, S.; Hammer-Muntz, O.; Leitner, G. Identification of a bovine mastitis *Escherichia coli* subset. *Vet. Microbiol.* **2008**, *132*, 135–148. [[CrossRef](#)]
71. Mortensen, B.L.; Skaar, E.P. The contribution of nutrient metal acquisition and metabolism to *Acinetobacter baumannii* survival within the host. *Front. Cell. Infect. Microbiol.* **2013**, *3*, 95. [[CrossRef](#)]
72. Affandi, T.; McEvoy, M.M. Mechanism of metal ion-induced activation of a two-component sensor kinase. *Biochem. J.* **2019**, *476*, 115–135. [[CrossRef](#)]
73. Hanczvikkel, A.; Füzi, M.; Ungvári, E.; Tóth, Á. Transmissible silver resistance readily evolves in high-risk clone isolates of *Klebsiella pneumoniae*. *Acta Microbiol. Immunol. Hung.* **2018**, *65*, 387–403. [[CrossRef](#)] [[PubMed](#)]
74. Faye, B.; Bengoumi, M. Trace-element metabolism in camel. In Proceedings of the Third Annual Meeting for Animal Production Under Arid Conditions, Al-Ain, United Arab Emirates, 2–3 May 1998; Volume 1, pp. 9–35.
75. Rolland, K.; Marois, C.; Siquier, V.; Cattier, B.; Quentin, R. Genetic features of *Streptococcus agalactiae* strains causing severe neonatal infections, as revealed by pulsed-field gel electrophoresis and *hylB* gene analysis. *J. Clin. Microbiol.* **1999**, *37*, 1892–1898. [[CrossRef](#)] [[PubMed](#)]
76. Sukhnanand, S.; Dogan, B.; Ayodele, M.O.; Zadoks, R.N.; Craver, M.P.J.; Dumas, N.B.; Schukken, Y.H.; Boor, K.J.; Wiedmann, M. Molecular subtyping and characterization of bovine and human *Streptococcus agalactiae* isolates. *J. Clin. Microbiol.* **2005**, *43*, 1177–1186. [[CrossRef](#)] [[PubMed](#)]

77. Domelier, A.S.; van der Mee-Marquet, N.; Grandet, A.; Mereghetti, L.; Rosenau, A.; Quentin, R. Loss of catabolic function in *Streptococcus agalactiae* strains and its association with neonatal meningitis. *J. Clin. Microbiol.* **2006**, *44*, 3245–3250. [[CrossRef](#)] [[PubMed](#)]
78. Behrooz, S.K.; Lida, L.; Ali, S.; Mehdi, M.; Rasoul, M.; Elnaz, O.; Farid, B.T.; Gholamreza, I. Study of MazEF, sam, and phd-doc putative toxin–antitoxin systems in *Staphylococcus epidermidis*. *Acta Microbiol. Immunol. Hung.* **2018**, *65*, 81–91. [[CrossRef](#)]
79. Chan, W.T.; Yeo, C.C.; Sadowy, E.; Espinosa, M. Functional validation of putative toxin-antitoxin genes from the Gram-positive pathogen *Streptococcus pneumoniae*: Phd-Doc Is Fourth Bona-Fide Operon. *Front. Microbiol.* **2014**, *5*, 677. [[CrossRef](#)] [[PubMed](#)]
80. Guérout, A.M.; Iqbal, N.; Mine, N.; Ducos-Galand, M.; Van Melderen, L.; Mazel, D. Characterization of the phd-doc and ccd toxin-antitoxin cassettes from *Vibrio* Superintegrons. *J. Bacteriol.* **2013**, *195*, 2270–2283. [[CrossRef](#)]
81. Lehnher, H.; Maguin, E.; Jafri, S.; Yarmolinsky, M.B. Plasmid addiction genes of bacteriophage P1: Doc, which causes cell death on curing of prophage, and phd, which prevents host death when prophage is retained. *J. Mol. Biol.* **1993**, *233*, 414–428. [[CrossRef](#)]
82. Yang, Q.E.; Walsh, T.R. Toxin–antitoxin systems and their role in disseminating and maintaining antimicrobial resistance. *FEMS Microbiol. Rev.* **2017**, *41*, 343–353. [[CrossRef](#)]
83. Schulein, R.; Guye, P.; Rhomberg, T.A.; Schmid, M.C.; Schröder, G.; Vergunst, A.C.; Carena, I.; Dehio, C. A bipartite signal mediates the transfer of type IV secretion substrates of *Bartonella henselae* human cells. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 856–861. [[CrossRef](#)]
84. Jiang, X.; Yang, Y.; Zhou, J.; Zhu, L.; Gu, Y.; Zhang, X.; Li, X.; Fang, W. Roles of the putative type IV-like secretion system key component VirD4 and PrsA in pathogenesis of *Streptococcus suis* Type 2. *Front. Cell. Infect. Microbiol.* **2016**, *6*, 172. [[CrossRef](#)]
85. Wallden, K.; Rivera-Calzada, A.; Waksman, G. Microreview: Type IV secretion systems: Versatility and diversity in function. *Cell. Microbiol.* **2010**, *12*, 1203–1212. [[CrossRef](#)] [[PubMed](#)]
86. Alvarez-Martinez, C.E.; Christie, P.J. Biological diversity of prokaryotic type IV secretion systems. *Microbiol. Mol. Biol. Rev.* **2009**, *73*, 775–808. [[CrossRef](#)] [[PubMed](#)]