UCLA
UCLA Previously Published Works

Title

Dispersal history and bidirectional human-fish host switching of invasive, hypervirulent Streptococcus agalactiae sequence type 283.

Permalink

Journal

Authors

Schar, Daniel

Zhang, Zhenyu

Pires, Joao

et al.

Publication Date

DOI

Peer reviewed

Dispersal history and bidirectional human-fish host switching of invasive, hypervirulent *Streptococcus agalactiae* sequence type 283

Daniel Schar[1]*, **Zhenyu Zhang**[2], **Joao Pires**[3], **Bram Vrancken**[1,4], **Marc A. Suchard**[2,5], **Philippe Lemey**[4], **Margaret Ip**[6], **Marius Gilbert**[1,7], **Thomas Van Boeckel**[3,8], **Simon Dellicour**[1,4]

1 Spatial Epidemiology Laboratory, Université Libre de Bruxelles, Brussels, Belgium, 2 Department of Biostatistics, Fielding School of Public Health, University of California, Los Angeles, Los Angeles, CA, United States of America, 3 Institute for Environmental Decisions, ETH Zurich, Zurich, Switzerland, 4 Department of Microbiology, Immunology and Transplantation, Rega Institute for Medical Research, KU Leuven, Leuven, Belgium, 5 Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California, Los Angeles, Los Angeles, CA, United States of America, 6 Department of Microbiology, Faculty of Medicine, Prince of Wales Hospital, The Chinese University of Hong Kong, Hong Kong SAR, China, 7 Fonds National de la Recherche Scientifique, Brussels, Belgium, 8 Center for Diseases Dynamics, Economics, and Policy, New Delhi, India

* dlschar@gmail.com, Daniel.Schar@ulb.be

Abstract

Human group B *Streptococcus* (GBS) infections attributable to an invasive, hypervirulent sequence type (ST) 283 have been associated with freshwater fish consumption in Asia. The origin, geographic dispersion pathways and host transitions of GBS ST283 remain unresolved. We gather 328 ST283 isolate whole-genome sequences collected from humans and fish between 1998 and 2021, representing eleven countries across four continents. We apply Bayesian phylogeographic analyses to reconstruct the dispersal history of ST283 and combine ST283 phylogenies with genetic markers and host association to investigate host switching and the gain and loss of antimicrobial resistance and virulence factor genes. Initial dispersal within Asia followed ST283 emergence in the early 1980s, with Singapore, Thailand and Hong Kong observed as early transmission hubs. Subsequent intercontinental dispersal originating from Vietnam began in the decade commencing 2001, demonstrating ST283 holds potential to expand geographically. Furthermore, we observe bidirectional host switching, with the detection of more frequent human-to-fish than fish-to-human transitions, suggesting that sound wastewater management, hygiene and sanitation may help to interrupt chains of transmission between hosts. We also show that antimicrobial resistance and virulence factor genes were lost more frequently than gained across the evolutionary history of ST283. Our findings highlight the need for enhanced surveillance, clinical awareness, and targeted risk mitigation to limit transmission and reduce the impact of an emerging pathogen associated with a high-growth aquaculture industry.

Introduction

Group B *Streptococcus* (GBS; *Streptococcus agalactiae*) is carried in the gastrointestinal and urogenital tracts and is well recognized as a cause of neonatal sepsis and meningitis as well as severe disease in pregnant adults and the immunocompromised [1]. Recent outbreaks of an invasive, hypervirulent GBS sequence type (ST) 283 responsible for severe disease in younger adults with fewer comorbidities have been associated with handling and consumption of freshwater fish [2–7]. Clinically, severe disease attributable to ST283 infection is characterized by sepsis, septic arthritis, meningitis, and infective endocarditis [7].

Retrospectively identified ST283 human infections were reported from Hong Kong from samples collected in 1993 and were exclusively associated with invasive sites in non-pregnant adults [7]. A 2015 Singapore epidemic of invasive GBS infections identified co-incident clonal ST283 isolates from human case-patients and fish, complementing a case-control study associating raw fish consumption with ST283 infection, and confirming this strain is transmissible as a freshwater fish foodborne disease [3, 4]. Outside of Asia, ST283 has also been reported sporadically from isolates collected in the United States, the United Kingdom, and from osteoarticular infections reported in France [8].

In aquaculture, GBS is responsible for substantial fish mortality and production loss [9], and both ST283 and a single locus variant (ST491) have been reported from diseased fish in southeast Asia [10, 11]. Isolation of ST283 in 2016 from fish in Brazil indicates this strain may currently be undergoing further global dissemination [12].

The foodborne transmission identified in Singapore as a conduit for ST283 invasive human disease suggests that freshwater fish consumption and contact may be under-appreciated as the source of at least some severe, invasive GBS disease globally [5, 13, 14]. Furthermore, whereas predominantly *tet*(M) gene carriage encoding tetracycline resistance is frequently reported in human-adapted GBS isolates [15], heterogeneous rates of tetracycline resistance gene carriage have been reported in ST283 from fish and human isolates, with putative loss of tetracycline resistance gene events [5, 16]. An enhanced understanding of the origin of ST283, its international distribution, host transitions and the acquisition and loss of antimicrobial resistance and virulence genes holds potential to inform targeted risk mitigation reducing morbidity and mortality associated with this emerging pathogen [13].

Here, we analyze the sequences of 328 genomes from ST283 isolates collected between 1998 and 2021 from eleven countries across four continents. We performed Bayesian phylodynamic and phylogeographic analyses to reconstruct the evolutionary and dispersal history of ST283. We further investigate host switching events and the gain and loss of antimicrobial resistance and virulence factor genes. These approaches provide insight into the evolutionary history of ST283 to inform surveillance and interventions for a pathogen closely associated with freshwater aquaculture as that industry experiences continued global growth.

Results

Time scale for the emergence of ST283

Root-to-tip distance was correlated with isolation date (Fig A in S1 Text) to assess the presence of a temporal signal within the genomic dataset. The time to the most recent common ancestor of ST283 inferred from the time-calibrated phylogeny was estimated as 1982 (95% highest posterior density (HPD) interval: 1976 to 1987; Fig 1 and Fig B in S1 Text).

Geographic expansion outside of Asia is recent

Heterogeneous sampling effort among locations inherent to our dataset precluded drawing any conclusion regarding the root state location. This limitation is confirmed by an analysis in
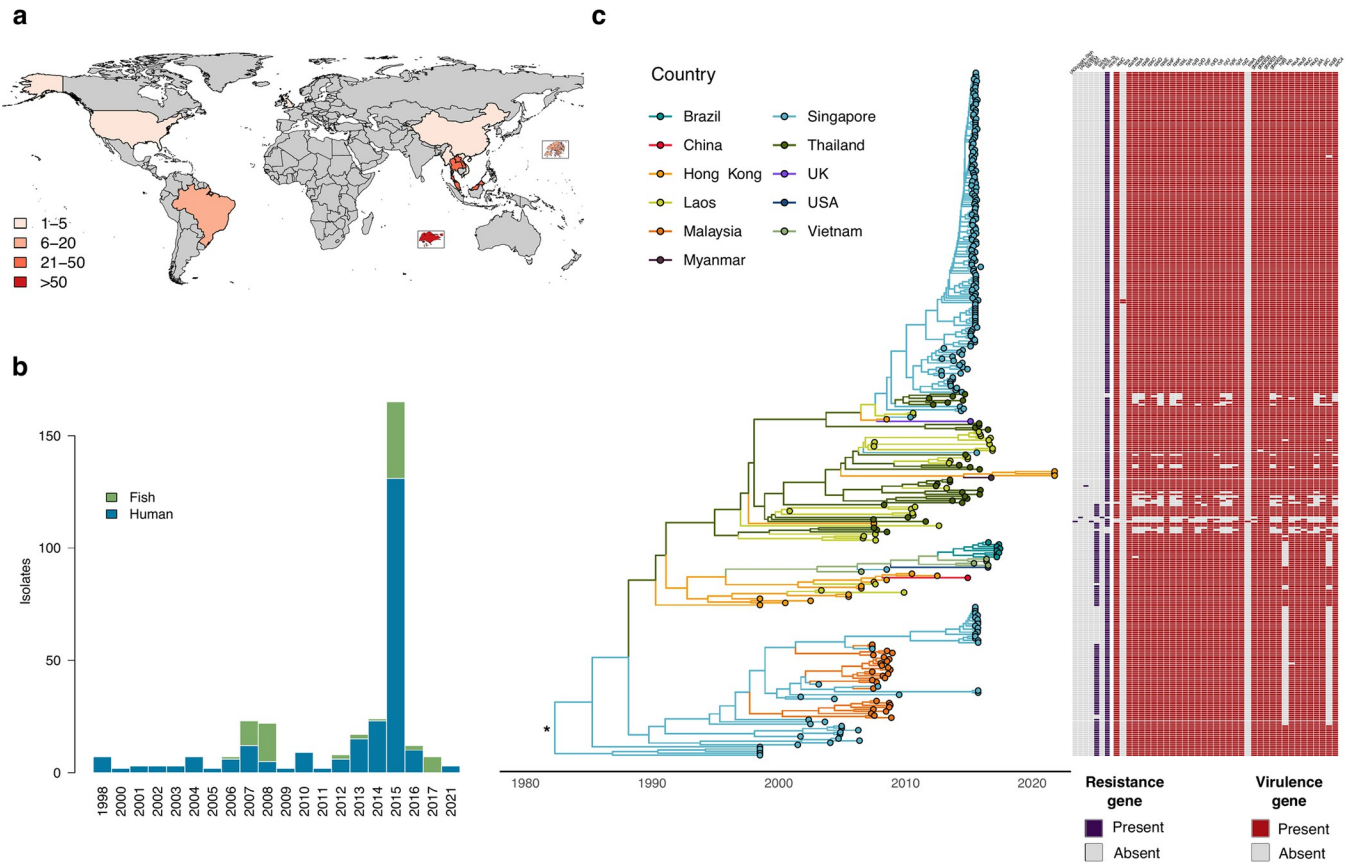
Fig 1. Spatio-temporal distribution of GBS ST283 isolates and inferred time-scaled phylogenetic tree. (a) Map of sampling location for ST283 isolates in the study (*n* = 328). Insets are Singapore (bottom) and Hong Kong (right). **(b)** Distribution of isolate sampling dates and host origin. **(c)** Annotated maximum clade credibility (MCC) tree resulting from a discrete phylogeographic analysis. Tip nodes and branches are colored according to the country of origin and the country inferred at ancestral nodes. Asterisk (*) indicates that, after having taken the heterogeneity of the sampling effort among sampled locations into account, no meaningful support was identified for ancestral root node location (see the Results section for further detail). A table of antimicrobial resistance (blue) and virulence factor (red) gene presence or absence is displayed for each isolate in the study. The base layer of the map is available at https://www.naturalearthdata.com/downloads/10m-cultural-vectors/10m-admin-0-countries/.

https://doi.org/10.1371/journal.pgph.0002454.g001

which the location states are randomly swapped among the tips of the tree during the phylogeographic reconstruction, which yielded a posterior probability for Singapore as the tree root location (*p* = 0.80) very similar to the posterior probability obtained through the standard phylogeographic analysis (*p* = 0.91; Table B in S1 Text). This result indicates that the finding of Singapore at the ancestral root node is not informed by the phylogenetic information but almost exclusively by the oversampling of that particular location.

Discrete phylogeographic analysis identified a posterior mean of 35 (95% HPD interval: 33 to 39) independent transition events between countries across the evolutionary history of ST283. The discrete phylogeographic reconstruction was analyzed in decadal bins, beginning in 1981 and continuing through the most recent sample collection in October 2021 (Fig 2). Following emergence, ST283 experienced an initial period of intracontinental expansion within Asia. Multiple transition events were inferred from Thailand to Hong Kong and Laos; Singapore to Malaysia; and from Hong Kong to Laos and Vietnam between the date of emergence and 2000, continuing through 2010. Thailand, Hong Kong, and Singapore appeared as central hubs for early dissemination of ST283, accounting for 62.1%, 21.5%, and 16.4% respectively of all supported transition events between 1981 and 2001. The earliest supported
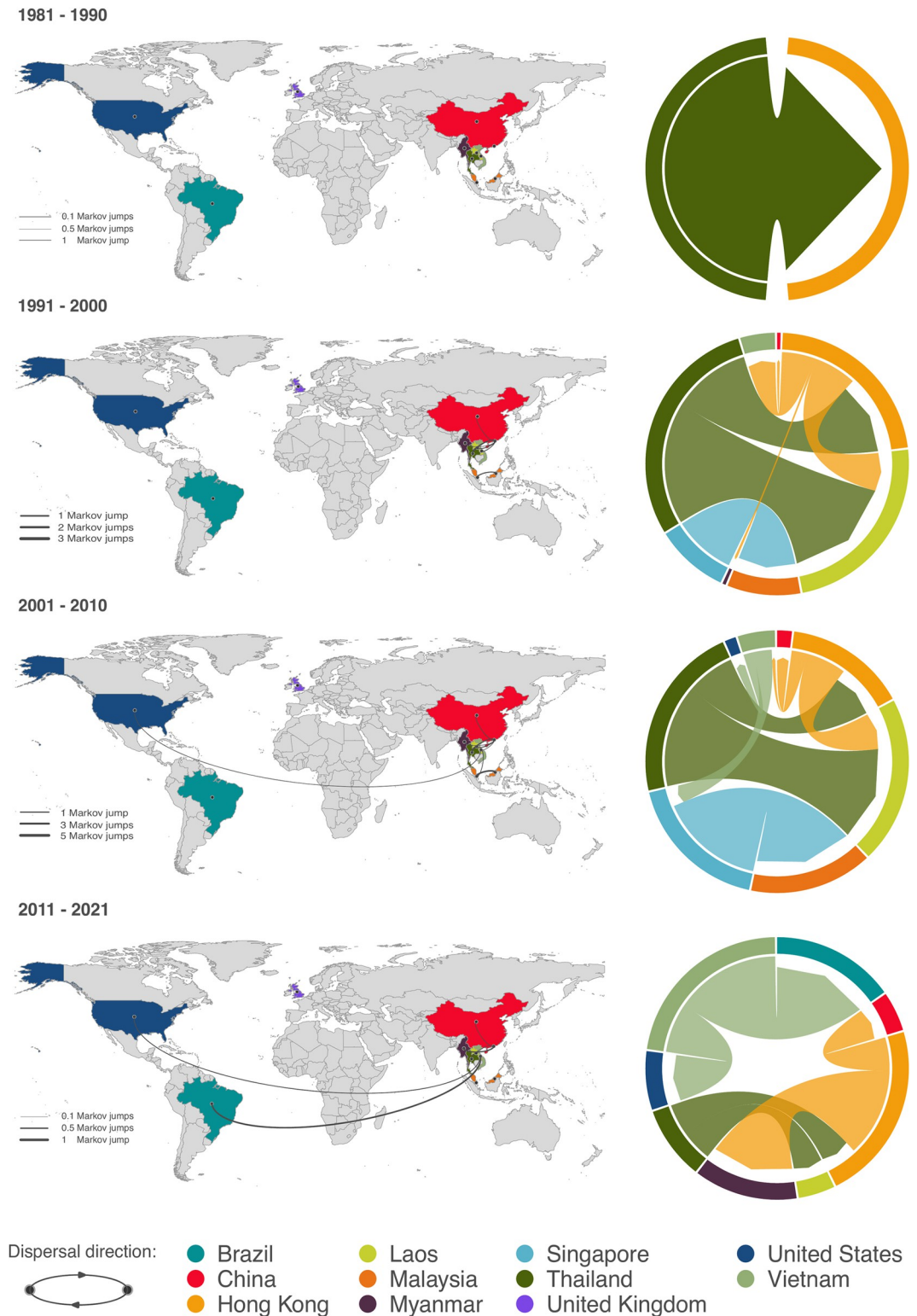
Fig 2. Discrete phylogeographic analysis of the dispersal history of GBS ST283 between 1981 and 2021. Intracontinental and intercontinental transition events are inferred as Markov jumps. Maps display transition events by decade and are accompanied by circular migration flow plots, in which transitions out of a country are represented by arrows originating at the outer ring and ending in an arrowhead offset from the destination country. Arrow width is proportional to the magnitude of the Markov jumps. Only transition events associated with an adjusted Bayes factor support > 20 are displayed, a threshold

intercontinental transition was inferred between 2001 and 2011 from Vietnam to the United States. In the most recent decade (2011–2021), Hong Kong and Thailand continued to seed international dissemination (34.3% and 18.7% of supported transition events, respectively), and Vietnam remained as a source of intercontinental ST283 movement. The rate of international dispersal of ST283 increased between 1996 and 2010 (Fig C in S1 Text).

ST283 exhibits frequent bidirectional host switching

Heterogeneous host sampling distribution in the dataset precluded inference of supported Markov jumps between hosts: fish-to-human and human-to-fish transition events were associated with an adjusted Bayes factor (BF_{adj}) support equal to and smaller than 1, respectively (see the Methods section for further detail on how Bayes factor supports were computed while considering heterogeneous sampling effort among host types). To account for this sampling bias, ancestral host state transitions were inferred by maximum likelihood estimation on downsampled phylogenies (see the Methods section for further detail). More human-to-fish transitions (median = 9; IQR: 8 to 9) were observed than fish-to-human (median = 2; IQR: 1 to 3) transitions across the evolutionary history of ST283 (Fig 3). This trend was consistent when controlling for the year of sampling as well as the phylogenetic diversity associated with each host type. Total human-to-fish transitions were greater than total fish-to-human transitions in each of 1,000 downsampled phylogenies.

Multi-drug resistance gene carriage is low in ST283

Seven antimicrobial resistance genes (ARGs) were identified, associated with phenotypic resistance to five classes of antibiotics: aminoglycosides, beta-lactams, dihydrofolate reductase inhibitors (trimethoprim), macrolides, and tetracyclines. Across all isolates, the *mre*(A) gene (321 isolates, 97.9%) and *tet*(M) gene (96 isolates, 29.3%) were most frequently carried, conferring resistance when expressed to macrolides and tetracyclines, respectively; all other ARGs were identified in single instances each. One isolate collected from a human bacteremic patient in Thailand in 2011 carried three ARGs—(AGly)*apH-Stph*, *dfrC*, and *mre*(A)—and was the only isolate in the dataset carrying the *esxA* gene, recently identified as encoding a pore-forming protein important in GBS pathogenesis [17].

Antimicrobial resistance and virulence genes are more frequently lost than gained across the evolutionary history of ST283

Discrete trait analyses of antimicrobial resistance genes revealed that the *tet*(M) and *mre*(A) genes were lost more frequently than they were gained across the ST283 evolutionary history. Similarly, GBS virulence genes associated with human adaptation—*hylB* (bacterial invasion and dissemination from initial site of infection), *lmb* (adherence), and *scpB* (neutrophil recruitment inhibition)—were lost more frequently than gained (Table 1 and Fig 4).

Trait correlations across the evolutionary history of ST283 were captured by applying a phylogenetic multivariate probit model capable of identifying conditional dependencies amongst any two traits after removing the effects of other traits–the so-called partial correlations. We identify a positive partial correlation between isolate origin from a human host and the *tet*(M) gene encoding tetracycline resistance (posterior median = 0.43). Regarding virulence factor genes, positive partial correlations were identified between human host and the *lmb* (posterior median = 0.25) and *cpsA* (posterior median = 0.31) genes (Fig 5). A partial
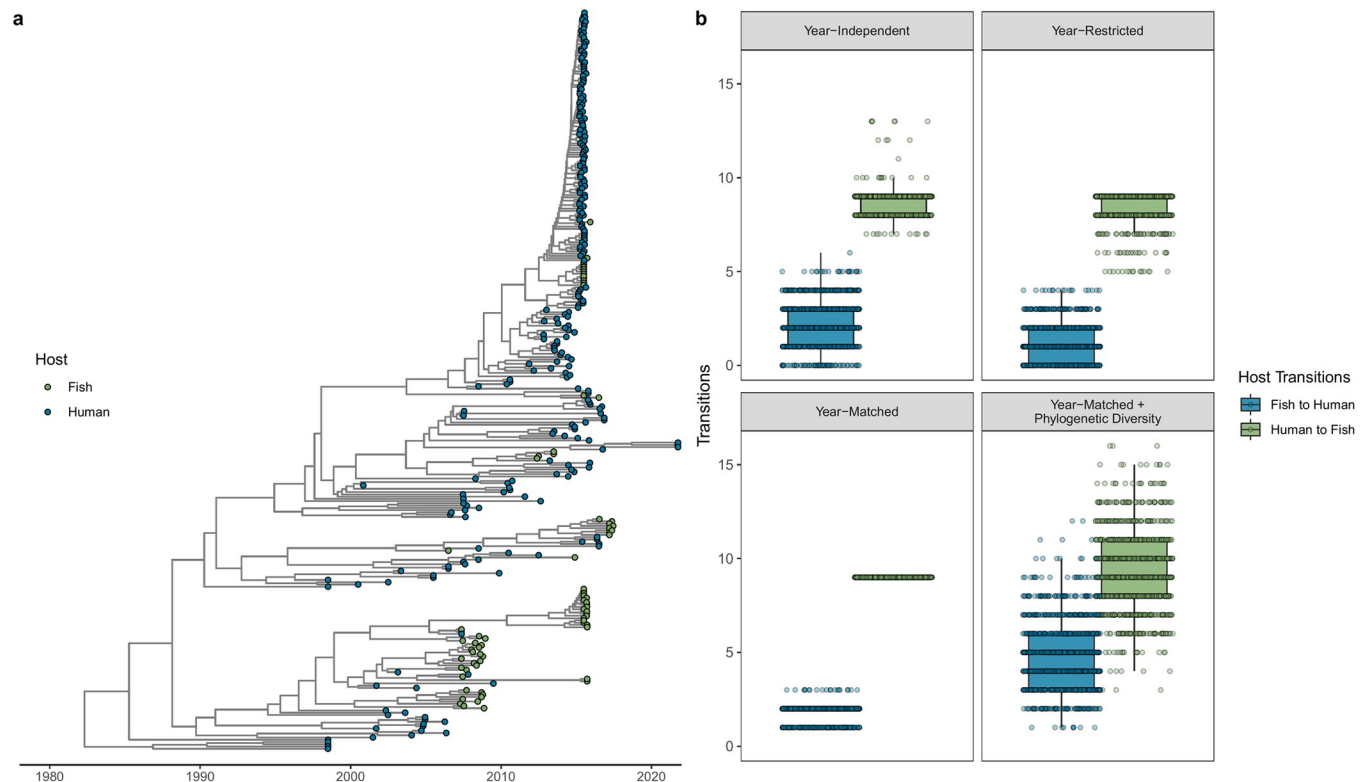
Fig 3. Bidirectional host switching in GBS ST283. (a) Host taxa (human or fish) are displayed in the time-scaled maximum clade credibility (MCC) tree resulting from the Bayesian phylogenetic analysis. **(b)** As inference of Markov jumps between hosts were not supported using the full dataset (see the text for further detail), ancestral host state transitions (human-to-fish and fish-to-human) were inferred by maximum likelihood estimate on 1,000 replicate downsampled phylogenies (see Methods for further detail). First, the full dataset was downsampled to produce equal numbers of human and fish origin isolates in the resulting phylogenies without any constraint on sampling year ('year-independent'; $n = 154$). To compare the effect of sampling year on this downsampling procedure, we conducted two year-constrained analyses: in the first, we worked with a downsampled dataset consisting of equal numbers of isolates originating from human and fish hosts using only isolates within a window of years where both human and fish host isolates were represented ('year-restricted'; $n = 140$). In the second analysis, we worked with the downsampled dataset from the year-restricted analysis, but further year-matched isolates to include equal numbers of human and fish origin isolates in each year as determined by the minimum number available for either host in that year ('year-matched'; $n = 116$). Finally, to account for uneven phylogenetic diversity associated with host type, a third analysis is conducted in which monophyletic clusters of sequences collected from the same host type are first subsampled to be represented by a single sequence prior to further subsampling as described in the 'year-matched' analysis (('year-matched + phylogenetic diversity'; $n = 32$). The horizontal box lines represent the first quartile, the median, and the third quartile. Whiskers denote the range of points within the first quartile $−1.5 \times$ the interquartile range and the third quartile $+1.5 \times$ the interquartile range. Ancestral host state transitions inferred in each of 1,000 replicate downsampled phylogenies are represented in their respective downsampling approach panel by a dot with horizontal jitter for visibility.

https://doi.org/10.1371/journal.pgph.0002454.g003

positive correlation was identified between fish host and the *gbs0632* gene (posterior median = 0.24). Other virulence factor gene trait correlations were considered as interactions governing gene function and expression. A positive partial correlation was identified between *lmb* and *scpB* (posterior median = 0.90), virulence factor genes involved in host cell adhesion and prevention of neutrophil recruitment, respectively. Positive partial correlation was identified between *cpsL*—a component of the capsular protein *cps* operon involved in molecular mimicry and immune evasion [18]—and *gbs0632*, which contributes to the GBS binding pilus architecture (posterior median = 0.61).

Discussion

Our findings confirm the emergence of GBS ST283 in Asia in the latter half of the twentieth century, corresponding with the start of expansive growth in freshwater aquaculture

Table 1. Antimicrobial resistance and virulence gene gains and losses. Median transitions from absence to presence (gained) and presence to absence (lost) with their 95% highest posterior density (HPD) intervals. Transitions are calculated from the discrete trait analysis post-burn-in posterior distribution implemented in BEAST.

		95% HPD Interval	
Gene	Median	Lower	Upper
tet(M)			
Gained	2	0	3.23
Lost	5	4.20	9.54
mre(A)			
Gained	0	0	0.07
Lost	4	3.95	4.05
hylB			
Gained	0	0	3.07
Lost	9	5.89	9.42
lmb			
Gained	0	0	1.05
Lost	5	4.95	5.05
scpB			
Gained	0	0	1.07
Lost	8	7.94	8.06

https://doi.org/10.1371/journal.pgph.0002454.t001

production driven primarily by Asia [5, 19]. Limited availability of ST283 genomes and a dataset characterized by heterogeneous sampling effort could however influence our findings. Future work to identify and sequence ST283 isolates from under-represented regions may elucidate missing diversity and transition events amongst locations. The time-calibrated phylogeny in this study infers a most recent common ancestor (MRCA) for ST283 in 1982, which aligns closely with a previous estimate of emergence in 1985 [5]. Our analysis yields an emergence date that precedes by more than a decade a MRCA estimate (1994) obtained by evaluating isolates sampled exclusively from humans [16]. The observation that isolates from humans and fish cluster together across multiple clades within the tree suggests that the difference in MRCA estimates is not due solely to the host origin. The short window measuring approximately a decade between the first reported human cases in 1993 and date of ancestral origin indicates that ST283 may have emerged from a proximal GBS lineage having nearly acquired capacity for–or being fully capable of–producing severe clinical disease in humans, although missing genetic diversity may affect MRCA estimates.

Evidence presented in this study suggests that ST283 may be following an expanding geographic range trajectory outside of Asia. The frequency of intercontinental transitions (Markov jumps from or to Asia) increased between 2006 and 2015 before declining between 2016 and 2021 (Fig 2 and Fig C in S1 Text). Whether this trend reflects an actual decline in intercontinental movement over the last five years or rather is influenced by comparatively fewer isolates in the dataset after 2016 ($n = 22$, 6.7%), and particularly from 2018 to 2021 ($n = 3$, 0.91%), will require further verification.

Laos, Hong Kong, and Singapore isolates are represented across clades, suggesting multiple introduction events, consistent with our discrete phylogeographic analysis showing active dispersal histories within Asia beginning in 1991 and accelerating into the decade from 2001 to 2010.

We observe more frequent human-to-fish than fish-to-human host switching–a trend that remains consistent when restricting the analysis to include the same number of sampled hosts in each year and when controlling for differences in phylogenetic diversity associated with
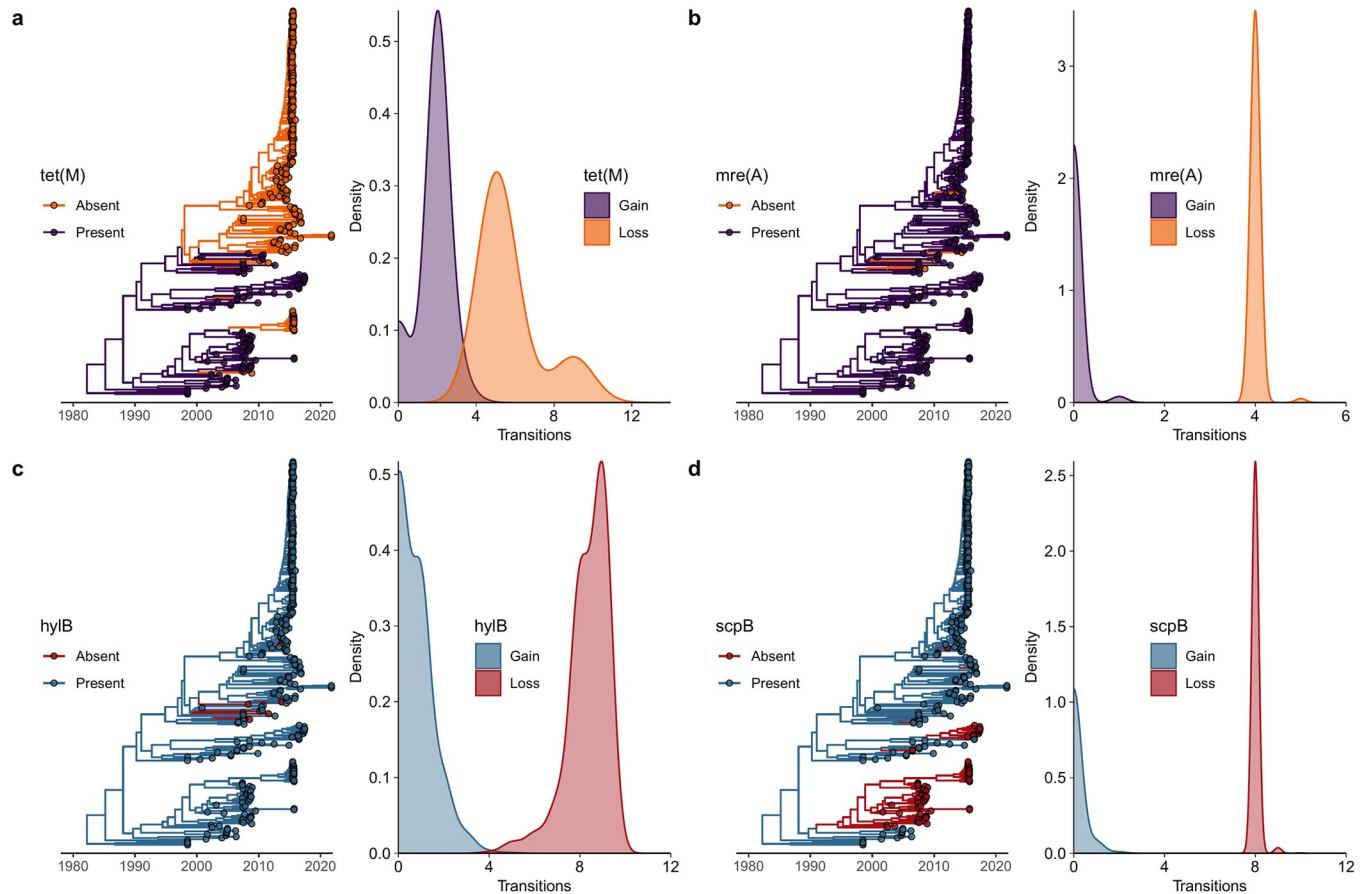
Fig 4. Ancestral trait reconstructions as well as corresponding gain/loss of antimicrobial resistance and virulence factor genes. Gene reconstructions are shown in the time-scaled maximum clade credibility (MCC) tree resulting from the Bayesian phylogenetic analysis and ancestral state reconstruction, with branches colored according to the most probable inferred gene state (absence or presence). The probability density of gene gains and losses are based on 1,000 trees sampled from the post burn-in posterior distribution. Displayed are **(a)** *tet*(M); **(b)** *mre*(A); **(c)** *hylB*; and **(d)** *scpB*.

each host type. Heterogeneous effort inherent to this dataset in sampling practices (point prevalence, event-based, and opportunistic) could affect these results. As such, these findings reflect the data currently available, and we cannot ascertain to what extent a different sampling effort would yield different results, which calls for further investigation. Nevertheless, the trend raises a possibility that ST283 may be maintained in humans residing in areas where sub-standard sanitation, hygiene and wastewater management facilitate repeat introductions into fish populations. GBS is recognized as a colonizer of the gastrointestinal tracts of healthy, asymptomatic humans. However, both the overall prevalence of ST283 carriage in the human population and the potential for human-to-human transmission of ST283 are unknown [13]. Although none of the 82 sampled food handlers and fishmongers in the 2015 Singapore outbreak carried ST283 [4], a study from northeast Thailand identified ST283 human fecal carriage in 5/184 (2.7%) samples, indicating human carriers could be involved in transmitting ST283 [20]. Human-associated GBS ST23 and ST7 have been isolated from marine mammals and from farm-raised crocodiles (ST23) and fish and amphibians (ST7) respectively, implicating anthropogenic pollution of the environment and surface waters in the switching between human and animal hosts [10, 21, 22]. Genome plasticity in GBS may facilitate adaptation to new ecological and host niches [23], and the propensity for GBS host switching has been
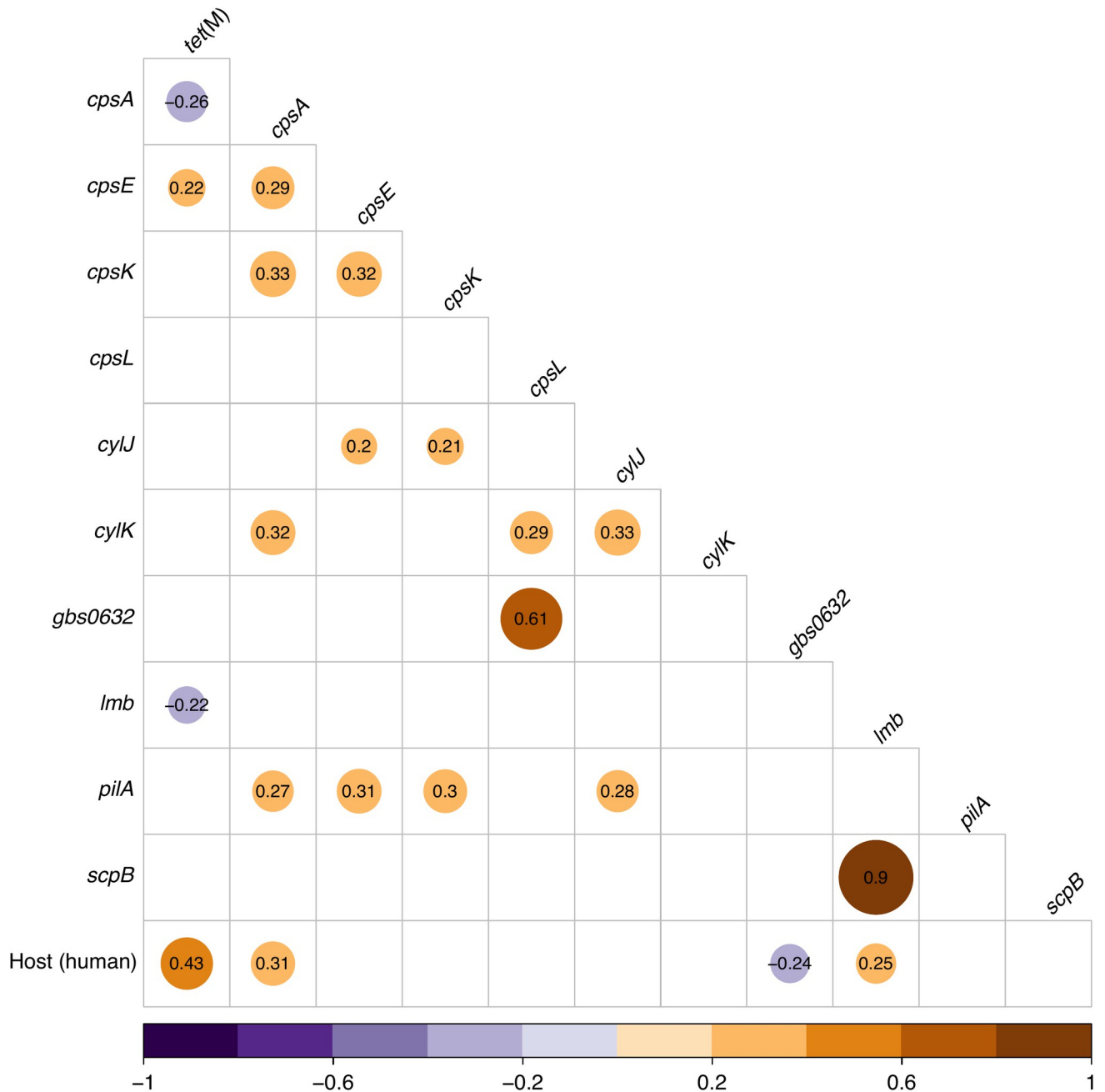
Fig 5. Host and gene correlations across the evolutionary history of GBS ST283. Gene-host and gene-gene partial correlations with a posterior median greater than 0.2 or less than– 0.2 (see Methods for further detail). Circle color represents the strength and direction of the correlation and circle size is proportional to the correlation magnitude. The host trait is coded as human = 1 and fish = -1.

recognized [23, 24]. Whether ST283 is maintained in fish or whether foodborne fish-to-human transmission is driven primarily by entry into the human food supply of fish during periodic outbreaks in these aquatic animals remains unknown. Maintenance in fish as a singular reservoir of ST283 could be expected to be associated with regular pulses of human food-borne infection in the absence of on-farm disease control measures. Such is the observed pattern for *Salmonella enterica* foodborne infection with continual human exposure through

the food chain associated with a *S. enterica* reservoir in laying hens [25]. In contrast, ST283 outbreaks in humans have been intermittent and stochastic, characterized by seasonal, short-duration episodes of case identification [4, 7]. Yet, if ST283 is maintained in fish, or if fish serve as an amplifying host subsequent to anthropogenic exposure, rapid growth in freshwater aquaculture in Asia beginning in the 1980s [19] could explain the observed increase in invasive GBS clinical disease first reported in Asia and highlight the significance of shifting dietary preferences in forewarning changes in foodborne disease incidence. Application of commercially available fish *Streptococcus agalactiae* vaccines may serve as a public health risk mitigation tool in limiting the potential role of fish as amplifying hosts and thus interrupting cyclical exchange of ST283 between humans and fish.

Human and fish exposure to a shared, as-yet undefined source must also be considered as GBS has a notoriously wide host range, the full extent of which for ST283 is undocumented [13]. Longitudinal surveillance of farmed fish and their rearing waters; establishing the prevalence of ST283 human carriage; and epidemiological investigation and source attribution of human outbreaks will be important in defining the cyclical movement of ST283 between humans and fish, the relative importance of each host as potential reservoirs, and the plausibility of a shared exposure source.

Antimicrobial resistance and virulence factor genes were lost more frequently than gained throughout the evolutionary history of ST283. An active rate of gene change in ST283 is consistent with GBS as a pathogen under evolutionary pressure, with recombination and mobile elements contributing to a shifting genome composition generating new, niche-adapted lineages [11, 23, 26]. Gene loss may facilitate rapid adaptation to the functional requirements of host suitability, for example, reducing the metabolic burden of gene expression, enhancing fitness and conferring a selective advantage in such niche-adapted lineages [23]. Our findings demonstrate a moderate rate of gain and loss for the tetracycline resistance gene *tet*(M) (Table 1). Previous work has shown that *tet*(M) is carried in GBS by the integrative and conjugative elements (ICE) Tn*916* and Tn*5801* [15, 16], and that the expansion of several tetracycline resistant human GBS clones and their global dissemination in the mid-twentieth century followed establishment of *tet*(M) through acquisition of these ICEs [15]. We identified a positive correlation between human host and *tet*(M) across the ST283 phylogeny and discrete trait analysis confirmed the previous finding of *tet*(M) gene presence at the ancestral root [16], suggesting it emerged from these tetracycline resistant human GBS clones and that the evolutionary path of ST283 has involved the repeated loss—and, in select instances, subsequent regain—of the *tet*(M) gene (Fig 4). We observe the apparent loss of *lmb* and *scpB* genes important in human invasive disease in the transition from human to fish (Figs 1C and 4D), which may be mediated by the ISSag2 insertion sequence element harboring these genes [27]. The *lmb* and *scpB* genes were positively correlated, and *lmb* was correlated with isolates originating from human hosts, pointing to a possible complementary role for these genes in invasive human infections. Loss of human-associated virulence factor genes in transitioning to fish hosts could explain the stochastic pattern of foodborne fish consumption-associated ST283 outbreaks: within a cyclical human-to-fish spillover and fish-to-human spillback context, the loss of GBS genes critical for human invasive disease may limit potential for spillback to result in human clinical illness [11].

Our analyses are subject to limitations. First, an extensive search of public repositories and literature enabled us to assemble the largest dataset of ST283 to date. Yet, despite this, the isolates identified may reflect neither the complete genetic diversity nor geographic distribution of ST283. South Asia—particularly India and Bangladesh—contributes substantial and growing freshwater fish production volume, yet no isolates were identified from this region, possibly reflecting a combination of differing strain presence patterns, exposure pathways, and

global disparities in typing and sequencing capacity. Efforts to type existing GBS strains and genetic characterization of GBS isolates associated with severe, invasive disease in adults may yield additional insights on GBS as a foodborne pathogen and associated exposure risk. Second, our dataset is characterized by sampling heterogeneity. While we account for this heterogeneous sampling effort in our analyses, the disparity in host, location, and sampling date could influence our findings. Third, whole genome sequences were primarily obtained as assemblies. The quality of genomes generated across differing sequencing and assembly methods could not be assessed. While genome quality could influence our findings, these genomes have been utilized in prior analyses. Finally, we did not have access to antimicrobial susceptibility testing for the isolates, limiting the ability to correlate presence of resistance genes with antimicrobial susceptibility profiles.

Recent work reported that despite nearly uniform carriage of *mre*(A) in a collection of ST283 isolates, none expressed resistance to macrolides [16]. In *Streptococcus agalactiae*, the *mre*(A) gene may serve a metabolic function, whereas when cloned into *E.coli*, *mre*(A) conferred macrolide resistance [28]. Additional study is needed to elucidate the phenotypic resistance profiles of ST283 and correlation with antimicrobial resistance genes. GBS phenotypic resistance to macrolides and fluoroquinolones vary regionally but remain concerning given the burden of invasive GBS infection globally [29]. Although the prevalence of multi-drug resistance gene carriage in ST283 is currently low, increasing use of medically important antimicrobials in the freshwater aquaculture industry [30] and in humans—particularly in low- and middle-income countries [31]—risks the generation of expanded resistance profiles.

Our findings demonstrate that ST283 holds the potential for geographic expansion, underscoring the need for enhanced surveillance across sectors, clinical awareness, and targeted risk mitigation and messaging. These findings highlight the importance of hygiene and sanitation, particularly in the context of the aquatic environment, wastewater management and along the food chain to limit transmission and mitigate the impact from this emerging pathogen.

Methods

Bacterial isolates

GBS ST283 whole genome sequences (*n* = 328) with corresponding sampling date and location were gathered through searches of public repositories (NCBI, ENA, and PubMLST; *n* = 3), literature review (PubMed; *n* = 26), research networks (*n* = 3) and from the Singapore *Streptococcus agalactiae* BioProject PRJNA293392 (*n* = 296). PubMed was searched for records in English through December 20, 2021 using search parameters: ("Group B *Streptococcus*" OR "*Streptococcus agalactiae*") AND (ST283 OR CC283). Genome sequences were excluded if they were not accompanied by a sampling date, sampling location, or host from which the sample was collected. Sequences from France were unavailable [8]. The 328 isolates were reportedly collected from humans (*n* = 251) and fish (*n* = 77) in eleven countries between 1998 and 2021 (S1 Data).

Ethics statement

Ethics approval was obtained for three archived genome sequences from human clinical isolates from Hong Kong under the Joint Chinese University of Hong Kong–New Territories East Cluster Clinical Research Ethics Committee (CUHK-NTEC CREC) (ref. no.: 2018.509). The remaining 325 genome sequences were obtained from public repositories (accession numbers in S1 Data). The authors did not have access to individually identifiable data and no ethics committee approval for this study was sought.

Whole genome sequence analysis

A total of 317 sequences were obtained as assemblies, and eleven as whole genome sequencing reads. The eleven whole genome sequence paired-end reads were assembled using SPAdes v.3.15.2 [32] with the "—careful" option. Quality of whole genome sequence reads was assessed using FastQC v.0.11.9 [33]. Quality assessment of assemblies generated in this study was assessed using QUAST v5.2.0 [34]. Mean Phred scores and assembly statistics (N50, L50) are reported in S1 Data.

Phylogenetic analysis

GBS ST283 whole genome sequence assemblies were screened for antimicrobial resistance and virulence factor genes with Abricate [35] and AMRFinderPlus [36]. Assemblies were mapped to reference genome SG-M1 (accession CP012419.2) [37] and aligned using SKA v.1.0 [38]. The resulting 2.12 Mb whole genome alignment was analyzed using Gubbins [39] to identify and purge regions of recombination. Putative mobile genetic elements (MGE) were predicted and masked from the alignment. Single nucleotide polymorphisms (SNP) were called on this MGE-free and putatively non-recombinant alignment using SNP-sites [40]. A maximum likelihood (ML) phylogenetic tree was then inferred from the 1,214 SNP alignment using RAxML v8.2.12 [41] with a general time reversible (GTR) and gamma model of rate heterogeneity.

Spatio-temporal and discrete trait analyses

To circumvent convergence issues with a joint inference approach and reduce computational burden, we used a multi-step process to generate spatial reconstructions of GBS ST283 transitions and to infer the gain and loss of antimicrobial resistance and virulence factor genes. In the first step we inferred a time-scaled phylogenetic tree from the ML tree using Markov chain Monte Carlo (MCMC) simulations over 10 million generations in the R package "BactDating" [42]. In the second step, we used this time-scaled phylogenetic tree as a fixed tree topology to perform both a discrete phylogeographic reconstruction and to analyze gene transition rates using the discrete diffusion model [43] implemented in the software package BEAST v1.10 [44]. In the discrete phylogeographic reconstruction, country transitions were estimated as Markov jump counts and reported along with both standard (Fig D in S1 Text) Bayes factor (BF) support values (ratio of posterior over prior odds and interpreted as a measure of the strength of evidence for the alternate hypothesis) and an adjusted (Fig 2) Bayes factor (BF_{adj}) supports [45] accounting for sample size disparity. Standard BF and BF_{adj} values >20 were considered strong statistical support [46].

Host distribution in the dataset was associated with a heterogeneous sampling effort, with more isolates of human than fish origin (Fig 1). Sampling bias can introduce artifacts into discrete trait reconstructions [47], as confirmed by our BF_{adj} support computation for the Markov jumps estimated between hosts (see the Results section). To account for this heterogeneous sampling orientation, we inferred host switching by working on downsampled time-scaled phylogenetic trees obtained by randomly sampling equal numbers of isolates originating from human and fish hosts in the tree tips. Three additional analyses were performed to assess the impact of sampling date and phylogenetic diversity associated with host type on host switching (Fig 3 and Methods in S1 Text) [48]. In each analysis, we generated 1,000 downsampled phylogenies, performing in each tree a maximum likelihood ancestral host estimation using an equal rates model implemented in the R package "ape" and counting host state transitions (human-to-fish and fish-to-human). The transition counts in each tree were taken as the distribution from which the median and interquartile range (IQR) were calculated.

Across-trait host-gene and gene-gene correlations

To further investigate the dependencies between traits, we apply a recently developed phylogenetic multivariate probit model [49, 50], implemented in BEAST v1.10 [44], which can efficiently learn the correlation between discrete traits while adjusting for across-taxa covariation inherent to the phylogenetic tree. We report the across-trait partial correlations describing conditional dependencies between any two traits without confounding from other considered traits [49].

Details of the methods are provided in Methods in S1 Text.

Supporting information

S1 Text. File containing supplementary methods, tables, figures, and references.
(PDF)

S1 Data. File containing sampling date, location, host, genome characteristics and sequence accession numbers.
(CSV)

Author Contributions

Conceptualization: Daniel Schar, Simon Dellicour.

Data curation: Daniel Schar, Margaret Ip.

Formal analysis: Daniel Schar, Zhenyu Zhang, Joao Pires, Bram Vrancken, Marc A. Suchard, Philippe Lemey, Margaret Ip, Marius Gilbert, Thomas Van Boeckel, Simon Dellicour.

Investigation: Daniel Schar.

Methodology: Daniel Schar, Zhenyu Zhang, Joao Pires, Bram Vrancken, Simon Dellicour.

Supervision: Thomas Van Boeckel, Simon Dellicour.

Visualization: Daniel Schar, Zhenyu Zhang, Simon Dellicour.

Writing – original draft: Daniel Schar.

Writing – review & editing: Daniel Schar, Zhenyu Zhang, Joao Pires, Bram Vrancken, Marc A. Suchard, Philippe Lemey, Margaret Ip, Marius Gilbert, Thomas Van Boeckel, Simon Dellicour.

References

1. Raabe VN, Shane AL. Group B *Streptococcus* (*Streptococcus* agalactiae). Fischetti VA, Novick RP, Ferretti, Portnoy DA, Rood JI, editors. Microbiol Spectr. 2019; 7: 7.2.17. https://doi.org/10.1128/microbiolspec.GPP3-0007-2018 PMID: 30900541

2. Chau ML, Chen SL, Yap M, Hartantyo SHP, Chiew PKT, Fernandez CJ, et al. Group B *Streptococcus* Infections Caused by Improper Sourcing and Handling of Fish for Raw Consumption, Singapore, 2015–2016. Emerg Infect Dis. 2017;23. https://doi.org/10.3201/eid2312.170596 PMID: 29148967

3. Rajendram P, Mar Kyaw W, Leo YS, Ho H, Chen WK, Lin R, et al. Group B *Streptococcus* Sequence Type 283 Disease Linked to Consumption of Raw Fish, Singapore. Emerg Infect Dis. 2016; 22: 1974–1977. https://doi.org/10.3201/eid2211.160252 PMID: 27767905

4. Kalimuddin S, Chen SL, Lim CTK, Koh TH, Tan TY, Kam M, et al. 2015 Epidemic of Severe Streptococcus agalactiae Sequence Type 283 Infections in Singapore Associated With the Consumption of Raw Freshwater Fish: A Detailed Analysis of Clinical, Epidemiological, and Bacterial Sequencing Data. Clin Infect Dis. 2017; 64: S145–S152. https://doi.org/10.1093/cid/cix021 PMID: 28475781

5. Barkham T, Zadoks RN, Azmai MNA, Baker S, Bich VTN, Chalker V, 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. Torres AG, editor. PLoS Negl Trop Dis. 2019; 13: e0007421. https://doi.org/10.1371/journal.pntd.0007421 PMID: 31246981

6. Ip M, Cheuk ESC, Tsui MHY, Kong F, Leung TN, Gilbert GL. Identification of a *Streptococcus agalactiae* Serotype III Subtype 4 Clone in Association with Adult Invasive Disease in Hong Kong. J Clin Microbiol. 2006; 44: 4252–4254. https://doi.org/10.1128/JCM.01533-06 PMID: 17005749

7. Ip M, Ang I, Fung K, Liyanapathirana V, Luo MJ, Lai R. Hypervirulent Clone of Group B *Streptococcus* Serotype III Sequence Type 283, Hong Kong, 1993–2012. Emerg Infect Dis. 2016; 22: 1800–1803. https://doi.org/10.3201/eid2210.151436 PMID: 27648702

8. Salloum M, van der Mee-Marquet N, Domelier A-S, Arnault L, Quentin R. Molecular Characterization and Prophage DNA Contents of *Streptococcus agalactiae* Strains Isolated from Adult Skin and Osteoarticular Infections. J Clin Microbiol. 2010; 48: 1261–1269. https://doi.org/10.1128/JCM.01820-09 PMID: 20181908

9. Zhang D, Li A, Guo Y, Zhang Q, Chen X, Gong X. Molecular characterization of Streptococcus agalactiae in diseased farmed tilapia in China. Aquaculture. 2013; 412–413: 64–69. https://doi.org/10.1016/j.aquaculture.2013.07.014

10. Delannoy CM, Crumlish M, Fontaine MC, Pollock J, Foster G, Dagleish MP, et al. Human Streptococcus agalactiae strains in aquatic mammals and fish. BMC Microbiol. 2013; 13: 41. https://doi.org/10.1186/1471-2180-13-41 PMID: 23419028

11. Sirimanapong W, Phước NN, Crestani C, Chen S, Zadoks RN. Geographical, Temporal and Host-Species Distribution of Potentially Human-Pathogenic Group B Streptococcus in Aquaculture Species in Southeast Asia. Pathogens. 2023; 12: 525. https://doi.org/10.3390/pathogens12040525 PMID: 37111411

12. Leal CAG, Queiroz GA, Pereira FL, Tavares GC, Figueiredo HCP. *Streptococcus agalactiae* Sequence Type 283 in Farmed Fish, Brazil. Emerg Infect Dis. 2019; 25: 776–779. https://doi.org/10.3201/eid2504.180543 PMID: 30882311

13. Risk profile—Group B Streptococcus (GBS)–Streptococcus agalactiae sequence type (ST) 283 in freshwater fish. FAO; 2021. https://doi.org/10.4060/cb5067en

14. Luangraj M, Hiestand J, Rasphone O, Chen SL, Davong V, Barkham T, et al. Invasive Streptococcus agalactiae ST283 infection after fish consumption in two sisters, Lao PDR. Wellcome Open Res. 2022; 7: 148. https://doi.org/10.12688/wellcomeopenres.17804.2 PMID: 36324702

15. DEVANI Consortium, Da Cunha V, Davies MR, Douarre P-E, Rosinski-Chupin I, Margarit I, et al. Streptococcus agalactiae clones infecting humans were selected and fixed through the extensive use of tetracycline. Nat Commun. 2014; 5: 4544. https://doi.org/10.1038/ncomms5544 PMID: 25088811

16. Aiewsakun P, Ruangchai W, Thawornwattana Y, Jaemsai B, Mahasirimongkol S, Homkaew A, et al. Genomic epidemiology of Streptococcus agalactiae ST283 in Southeast Asia. Sci Rep. 2022; 12: 4185. https://doi.org/10.1038/s41598-022-08097-0 PMID: 35264716

17. Spencer BL, Tak U, Mendonça JC, Nagao PE, Niederweis M, Doran KS. A type VII secretion system in Group B Streptococcus mediates cytotoxicity and virulence. Wessels MR, editor. PLOS Pathog. 2021; 17: e1010121. https://doi.org/10.1371/journal.ppat.1010121 PMID: 34871327

18. Rajagopal L. Understanding the regulation of Group B Streptococcal virulence factors. Future Microbiol. 2009; 4: 201–221. https://doi.org/10.2217/17460913.4.2.201 PMID: 19257847

19. FAO. The State of World Fisheries and Aquaculture 2022. FAO; 2022. https://doi.org/10.4060/cc0461en

20. Barkham T, Tang WY, Wang Y-C, Sithithaworn P, Kopolrat KY, Worasith C. Human Fecal Carriage of *Streptococcus agalactiae* Sequence Type 283, Thailand. Emerg Infect Dis. 2023;29. https://doi.org/10.3201/eid2908.230098 PMID: 37486205

21. Bishop EJ, Shilton C, Benedict S, Kong F, Gilbert GL, Gal D, et al. Necrotizing fasciitis in captive juvenile *Crocodylus porosus* caused by *Streptococcus agalactiae*: an outbreak and review of the animal and human literature. Epidemiol Infect. 2007; 135: 1248–1255. https://doi.org/10.1017/S0950268807008515 PMID: 17445318

22. Kawasaki M, Delamare-Deboutteville J, Bowater RO, Walker MJ, Beatson S, Ben Zakour NL, et al. Microevolution of Streptococcus agalactiae ST-261 from Australia Indicates Dissemination via Imported Tilapia and Ongoing Adaptation to Marine Hosts or Environment. Björkroth J, editor. Appl Environ Microbiol. 2018; 84: e00859–18. https://doi.org/10.1128/AEM.00859-18 PMID: 29915111

23. Richards VP, Velsko IM, Alam MT, Zadoks RN, Manning SD, Pavinski Bitar PD, et al. Population Gene Introgression and High Genome Plasticity for the Zoonotic Pathogen Streptococcus agalactiae. Mol Biol Evol. 2019; 36: 2572–2590. https://doi.org/10.1093/molbev/msz169 PMID: 31350563

24. Crestani C, Forde TL, Lycett SJ, Holmes MA, Fasth C, Persson-Waller K, et al. The fall and rise of group B Streptococcus in dairy cattle: reintroduction due to human-to-cattle host jumps? Microb Genomics. 2021;7. https://doi.org/10.1099/mgen.0.000648 PMID: 34486971

25. De Knegt LV, Pires SM, Hald T. Attributing foodborne salmonellosis in humans to animal reservoirs in the European Union using a multi-country stochastic model. Epidemiol Infect. 2015; 143: 1175–1186. https://doi.org/10.1017/S0950268814001903 PMID: 25083551

26. Chen SL. Genomic Insights Into the Distribution and Evolution of Group B Streptococcus. Front Microbiol. 2019; 10: 1447. https://doi.org/10.3389/fmicb.2019.01447 PMID: 31316488

27. Franken C, Haase G, Brandt C, Weber-Heynemann J, Martin S, Lämmler C, et al. Horizontal gene transfer and host specificity of beta-haemolytic streptococci: the role of a putative composite transposon containing scpB and lmb: Horizontal gene transfer in streptococci. Mol Microbiol. 2002; 41: 925–935. https://doi.org/10.1046/j.1365-2958.2001.02563.x PMID: 11532154

28. Clarebout G, Villers C, Leclercq R. Macrolide Resistance Gene *mreA* of *Streptococcus agalactiae* Encodes a Flavokinase. Antimicrob Agents Chemother. 2001; 45: 2280–2286. https://doi.org/10.1128/AAC.45.8.2280–2286.2001

29. Wang J, Zhang Y, Lin M, Bao J, Wang G, Dong R, et al. Maternal colonization with group B Streptococcus and antibiotic resistance in China: systematic review and meta-analyses. Ann Clin Microbiol Antimicrob. 2023; 22: 5. https://doi.org/10.1186/s12941-023-00553-7 PMID: 36639677

30. Schar D, Klein EY, Laxminarayan R, Gilbert M, Van Boeckel TP. Global trends in antimicrobial use in aquaculture. Sci Rep. 2020; 10: 21878. https://doi.org/10.1038/s41598-020-78849-3 PMID: 33318576

31. Klein EY, Van Boeckel TP, Martinez EM, Pant S, Gandra S, Levin SA, et al. Global increase and geographic convergence in antibiotic consumption between 2000 and 2015. Proc Natl Acad Sci. 2018; 115: E3463–E3470. https://doi.org/10.1073/pnas.1717295115 PMID: 29581252

32. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler. Curr Protoc Bioinforma. 2020;70. https://doi.org/10.1002/cpbi.102 PMID: 32559359

33. Andrews S. FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. 2010. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

34. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics. 2018; 34: i142–i150. https://doi.org/10.1093/bioinformatics/bty266 PMID: 29949969

35. Seemann T. Abricate. Available: https://github.com/tseemann/abricate

36. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Sci Rep. 2021; 11: 12728. https://doi.org/10.1038/s41598-021-91456-0 PMID: 34135355

37. Mehershahi KS, Hsu LY, Koh TH, Chen SL. Complete Genome Sequence of Streptococcus agalactiae Serotype III, Multilocus Sequence Type 283 Strain SG-M1. Genome Announc. 2015; 3: e01188–15. https://doi.org/10.1128/genomeA.01188-15 PMID: 26494662

38. Harris SR. SKA: Split Kmer Analysis Toolkit for Bacterial Genomic Epidemiology. Genomics; 2018 Oct. https://doi.org/10.1101/453142

39. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res. 2015; 43: e15–e15. https://doi.org/10.1093/nar/gku1196 PMID: 25414349

40. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genomics. 2016;2. https://doi.org/10.1099/mgen.0.000056 PMID: 28348851

41. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014; 30: 1312–1313. https://doi.org/10.1093/bioinformatics/btu033 PMID: 24451623

42. Didelot X, Croucher NJ, Bentley SD, Harris SR, Wilson DJ. Bayesian inference of ancestral dates on bacterial phylogenetic trees. Nucleic Acids Res. 2018; 46: e134–e134. https://doi.org/10.1093/nar/gky783 PMID: 30184106

43. Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. PLoS Comput Biol. 2009; 5: e1000520. https://doi.org/10.1371/journal.pcbi.1000520 PMID: 19779555

44. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 2018; 4: vey016. https://doi.org/10.1093/ve/vey016 PMID: 29942656

45. Vrancken B, Mehta SR, Ávila-Ríos S, García-Morales C, Tapia-Trejo D, Reyes-Terán G, et al. Dynamics and Dispersal of Local Human Immunodeficiency Virus Epidemics Within San Diego and Across the

San Diego-Tijuana Border. Clin Infect Dis Off Publ Infect Dis Soc Am. 2021; 73: e2018–e2025. https://doi.org/10.1093/cid/ciaa1588 PMID: 33079188

46. Kass R, Raftery A. Bayes factors. J Am Stat Assoc. 1995; 90: 773–795.

47. De Maio N, Wu C-H, O'Reilly KM, Wilson D. New Routes to Phylogeography: A Bayesian Structured Coalescent Approximation. Pritchard JK, editor. PLOS Genet. 2015; 11: e1005421. https://doi.org/10.1371/journal.pgen.1005421 PMID: 26267488

48. Dellicour S, Baele G, Dudas G, Faria NR, Pybus OG, Suchard MA, et al. Phylodynamic assessment of intervention strategies for the West African Ebola virus outbreak. Nat Commun. 2018; 9: 2222. https://doi.org/10.1038/s41467-018-03763-2 PMID: 29884821

49. Zhang Z, Nishimura A, Trovão NS, Cherry JL, Holbrook AJ, Ji X, et al. Accelerating Bayesian inference of dependency between complex biological traits. 2022 [cited 31 Oct 2022]. https://doi.org/10.48550/ARXIV.2201.07291

50. Zhang Z, Nishimura A, Bastide P, Ji X, Payne RP, Goulder P, et al. Large-scale inference of correlation among mixed-type biological traits with phylogenetic multivariate probit models. Ann Appl Stat. 2021;15. https://doi.org/10.1214/20-AOAS1394