



# Development and Application of a Prophage Integrase Typing Scheme for Group B *Streptococcus*

Chiara Crestani<sup>1\*</sup>, Taya L. Forde<sup>1</sup> and Ruth N. Zadoks<sup>1,2</sup>

<sup>1</sup> Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, United Kingdom,

<sup>2</sup> Sydney School of Veterinary Science, University of Sydney, Camden, NSW, Australia

## OPEN ACCESS

### Edited by:

Iain Sutcliffe,  
Northumbria University,  
United Kingdom

### Reviewed by:

Barbara Spellerberg,  
Ulm University Medical Center,  
Germany  
Rui Zhang,  
Xiamen University, China

### \*Correspondence:

Chiara Crestani  
c.crestani.1@research.gla.ac.uk

### Specialty section:

This article was submitted to  
Evolutionary and Genomic  
Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 08 April 2020

**Accepted:** 28 July 2020

**Published:** 31 August 2020

### Citation:

Crestani C, Forde TL and Zadoks RN  
(2020) Development and Application  
of a Prophage Integrase Typing  
Scheme for Group B *Streptococcus*.  
*Front. Microbiol.* 11:1993.  
doi: 10.3389/fmicb.2020.01993

Group B *Streptococcus* (GBS) is a gram-positive pathogen mainly affecting humans, cattle, and fishes. Mobile genetic elements play an important role in the evolution of GBS, its adaptation to host species and niches, and its pathogenicity. In particular, lysogenic prophages have been associated with a high virulence of certain strains and with their ability to cause invasive infections in humans. It is therefore important to be able to accurately detect and classify prophages in GBS genomes. Several bioinformatic tools for the identification of prophages in bacterial genomes are available on-line. However, genome searches for most of these programs are affected by the composition of their reference database. Lack of databases specific to GBS results in failure to recognize all prophages in the species. Additionally, performance of these programs is affected by genome fragmentation in the case of draft genomes, leading to underestimation of the number of phages. They also prove impractical when dealing with large genome datasets and they do not offer a quick way of classifying bacteriophages. We developed a GBS-specific method to screen genome assemblies for the presence of prophages and to classify them based on a reproducible typing scheme. This was achieved through an extensive search of a vast number of high-quality GBS sequences ( $n = 572$ ) originating from different host species and countries in order to build a database of phage integrase types, on which the scheme is based. The proposed typing scheme comprises 12 integration sites and sixteen prophage integrase types, including multiple subtypes per integration site and integrase genes that were not site-specific. Two putative phage-inducible chromosomal islands (PICI) and their insertion sites were also identified during the course of these analyses. Phages were common and diverse in all major clonal complexes associated with human disease and detected in isolates from every animal species and continent included in the study. This database will facilitate further work on the prevalence and role of prophages in GBS evolution, and identifies the roles of PICIs in GBS and of prophage in hypervirulent ST283 as areas for further research.

**Keywords:** prophage, bacteriophage, group B *Streptococcus*, *Streptococcus agalactiae*, integrase, phage typing, phage-inducible chromosomal island, PICI

## 1. INTRODUCTION

Group B *Streptococcus* (GBS)—also known as *Streptococcus agalactiae*—is a gram-positive bacterium with a wide host range (Brochet et al., 2006; Delannoy et al., 2016; Richards et al., 2019). Major hosts of interest from a public health and socio-economic point of view are humans (High et al., 2005; Le Doare et al., 2017; Seale et al., 2017), fishes (Jafar et al., 2008; Liu et al., 2014; Zamri-Saad, 2018), and cattle (Zadoks et al., 2011; Lyhs et al., 2016; Sørensen et al., 2019). In humans, GBS is a leading cause of neonatal invasive disease, and a pathogen of immunocompromised adults and elderly people (Farley and Strasbaugh, 2001; High et al., 2005; Skoff et al., 2009). The epidemiology and clinical manifestations of GBS disease in humans continue to evolve, as exemplified by the recent emergence of hypervirulent GBS in adults without underlying comorbidities (Barkham et al., 2019). Phages and other mobile genetic elements (MGE) play an important role in the evolution of GBS, its adaptation to different hosts and niches and its virulence profile (Richards et al., 2019). In human isolates, a high prevalence of prophages has been associated with greater pathogenicity, particularly the ability to cause invasive infections (van der Mee-Marquet et al., 2006; Domelier et al., 2009; Salloum et al., 2010, 2011). A number of these phages carry genes associated with virulence and host adaptation, suggesting that lysogeny (the process of integration of temperate bacteriophages into the bacterial genome as lysogenic prophages), may play an important role in the biological success of the strains (van der Mee-Marquet et al., 2018). Likewise, host adaptation of a cattle-associated lineage is thought to have been driven by the acquisition of mobile genetic content, including prophages (Richards et al., 2011). This may include transfer of prophages between streptococcal species, as demonstrated between *Streptococcus pyogenes* and *Streptococcus equi* subsp. *equi* (Holden et al., 2009), *Streptococcus dysgalactiae* subsp. *equisimilis* (Davies et al., 2005), or *S. dysgalactiae* subsp. *dysgalactiae* (Suzuki et al., 2011), and suspected between *S. pyogenes* and GBS (Bai et al., 2013).

Considering the association of prophage carriage with virulence and host adaptation, there is a need for a method to screen isolates for the presence of prophages, and to classify these phages based on a reproducible typing scheme. Several bioinformatic tools for the identification of prophages in bacterial genomes are available on-line (Bose and Barber, 2006; Lima-Mendez et al., 2008; Arndt et al., 2016). Most of these tools, however, are based on databases of known prophage sequences, the composition of which can influence their performance and ability to detect prophages (Javan et al., 2019). If the database does not include phages that are specific to the bacterial species being examined, the program may only identify parts of the prophage structure and under-report the total number of prophages per genome (Javan et al., 2019). Prophage detection can also be hampered by assembly of the prophage sequence across multiple contigs (Jamrozny et al., 2017), which can happen in the case of short-read sequencing technologies such as Illumina (Bennett, 2004), particularly when draft assemblies are not closed. Short-read sequencing is currently the most widely used method of

sequencing because of its accuracy of basecalling and low cost compared to most long-read sequencing techniques. Lastly, most prophage identification programs require at least some manual investigation of the output; this is impractical for large-scale genome studies, which are increasingly common.

A possible way to overcome these issues would be the adoption of a classification scheme based on host-specific prophage integrase types. This approach is already in use for other bacterial species, including *Staphylococcus aureus* (Goerke et al., 2009). These typing schemes are based on the concept that prophage integrases are site-specific (i.e., one type of integrase is usually found at only one chromosomal insertion site) through the recognition of attachment sites in the bacterial chromosome (*attB*) (Campbell, 1992), short nucleotide segments that are identical to attachment sites on the phage (*attP*). The *attB* corresponds to the insertion site where the phage recombinates to become an integrated lysogenic prophage. Once the prophage is integrated, the *att* site is usually found at both ends of the prophage. Integrase-based typing schemes also exist for phage-inducible chromosomal islands (PICIs), small molecular parasites that hijack phage packaging systems to be transferred to a new bacterial cell (Penadés and Christie, 2015; Martínez-Rubio et al., 2017; Fillo-Salom et al., 2018). No bioinformatic programs specifically designed for the detection of these mobile genetic elements are available to date and manual inspection of whole genome sequence data is currently the only strategy for *in silico* identification of PICIs.

A typing scheme for prophages based on full-prophage sequence diversity and insertion sites has been proposed for GBS (van der Mee-Marquet et al., 2018), however, a classification scheme based on site-specific integrases had yet to be developed for this species. Such a scheme would enable the rapid screening of large batches of full and draft genome sequences for the presence of prophages (based on integrase amino-acid sequences) and would be less affected by genome fragmentation. Additionally, as prophage integrases are thought to be site-specific, this typing method would allow for the unambiguous typing of the identified phages. To this end, we developed a GBS-specific prophage integrase typing scheme based on a large genomic dataset representing five continents and all major host species and clonal complexes of GBS.

## 2. MATERIALS AND METHODS

### 2.1. Datasets Included in This Study

Screening for prophages and integrase genes, as detailed in the subsequent sections, was initially carried out using a publicly available dataset consisting of closed genome sequences obtained from NCBI (dataset 1, **Table S1**). The use of closed genomes ensured that prophage detection would not be affected by genome fragmentation and that complete prophage sequences could be identified for subsequent use. Dataset 1 comprised 69 closed genome sequences representing major and minor host species (human,  $n = 15$ ; fish,  $n = 49$ ; cattle,  $n = 2$ ; camel,  $n = 1$ ; frog,  $n = 1$ ; unknown,  $n = 1$ ) and five continents (Africa,  $n = 1$ ; South America,  $n = 40$ ; North America,  $n = 4$ ; Asia,  $n = 22$ ; Europe,  $n = 1$ ; unknown,  $n = 1$ ). Automated and

manual screening of these isolates resulted in an initial database of prophages and integrase genes, similar to the existing *S. aureus* integrase typing scheme (Goerke et al., 2009).

To make the prophage and integrase database more comprehensive, a second dataset (dataset 2, **Table S2**) was subsequently screened, consisting of genomes that were of high quality, albeit not necessarily closed, and providing more in-depth coverage of major and minor GBS host species, geographic diversity, and GBS clades. Dataset 2 is a subset of 901 publicly available sequences included in the study by Richards et al. (2019) and includes all sequences with a maximum of 50 contigs ( $n = 503$ ). Isolates with more than 50 contigs were excluded from analysis because high genome fragmentation can lead to sub-optimal performance of bioinformatic programs. As for dataset 1, isolates in dataset 2 originated from major host species (humans,  $n = 486$ ; fishes,  $n = 8$ ; and cattle  $n = 5$ ) and minor host species (camel, dog, dolphin, and seal,  $n = 1$  per species). Geographical origins were diverse (Africa,  $n = 1$ ; the Americas,  $n = 353$ ; Asia,  $n = 10$ ; Europe,  $n = 117$ ; Oceania  $n = 18$ ; unknown,  $n = 4$ ). Fourteen clonal complexes (CC) and 53 sequence types (ST) were represented in dataset 2 (**Table S3**), with the most well-represented being common GBS clades from humans (CC1,  $n = 260$ ; CC17,  $n = 90$ ; CC23,  $n = 56$ ; CC12,  $n = 38$ ; CC19,  $n = 29$ ) or fishes (CC7,  $n = 6$ ).

## 2.2. Screening of Genomes and Development of a Phage Integrase Typing Scheme

### 2.2.1. Detection of Prophages and Integrase Genes in Dataset 1

To obtain the most complete database possible, and to assess agreement between methods, closed genomes from dataset 1 were analyzed with three methods, i.e. manual screening of GenBank files, PHASTER, and PhageMiner. GenBank files were used for manual screening of phage sequences starting from genes annotated as “site-specific integrase,” “integrase,” or “recombinase.” Manual inspection was also used to identify putative PICI, as there are no specific bioinformatic programs available for the detection of these MGE. Details on how PICI manual inspection was carried out can be found in **Supplementary Material, section 1.3**. PHASTER (PhAge Search Tool Enhanced Release) (Zhou et al., 2011; Arndt et al., 2016) is a widely used web-based integrated search and annotation tool for phage display. PhageMiner (Javan et al., 2019) is a user-supervised semi-automated computational tool that enables the identification of prophage sequences within complete or draft bacterial genomes. It allows for rapid identification, user inspection and curation of phage sequences from large numbers of genomes and has been validated on streptococci (Javan et al., 2019). For our study, PhageMiner was run locally on GenBank files annotated with one of the recommended annotation tools for this program, RAST v2.0 (Aziz et al., 2008) or Prokka v1.11 (Seemann, 2014). Using complete prophages identified with the three approaches, a database of phage integrase types was built. Incomplete prophages, whether due to genome fragmentation or lack of essential genes such as the integrase, were not included in the analysis. Integrases were classified based on

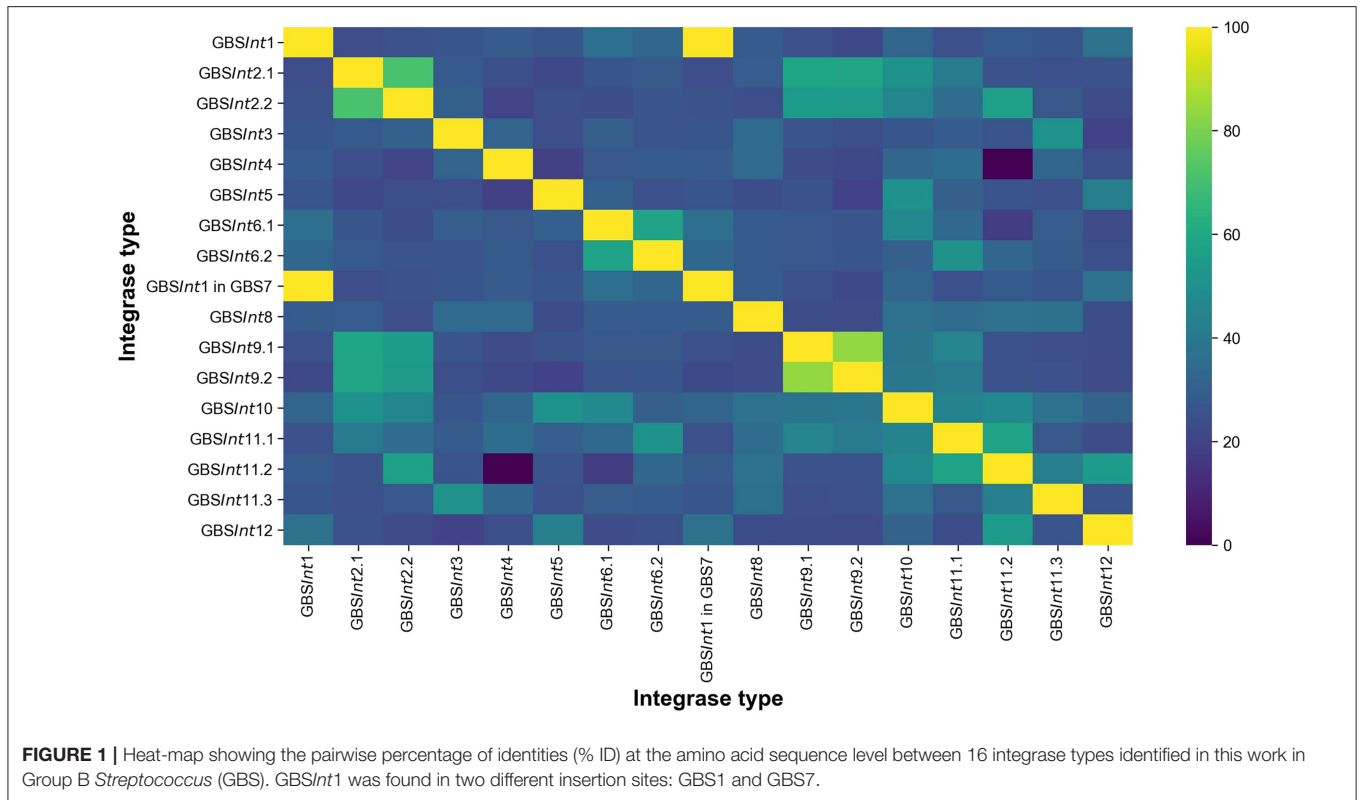
insertion site and percentage identity (% ID), using translated amino acid sequences, and numbered in order of detection. If a blastp (Camacho et al., 2009) comparison resulted in  $\geq 90\%$  ID (**Figure 1, Figure S1**) and  $\geq 95\%$  query cover (QC), integrases were considered to belong to the same type. When an integrase did not meet these thresholds but occupied the same integration site as an integrase that had already been classified, a subtype number was added (e.g., GBSInt2.1 and GBSInt2.2 represent integrases that both occupied integration site GBS2 but with  $< 90\%$  sequence similarity). The same set of prophages was identified with all three detection methods. Putative attachment sites were identified bioinformatically using blastn, through comparison of the site of integration in an empty genome (i.e., not harboring the prophage, chosen among closed genomes of ideally the same ST and host species) and the regions at both ends of the integrated prophage in a genome harboring the prophage.

### 2.2.2. Detection of Prophages and Integrase Genes in Dataset 2

Because all methods identified the same prophages in dataset 1, only PhageMiner was used for dataset 2. It can be run locally, eliminating waiting time for server queues that may affect analysis speed for server-based programs like PHASTER, which is particularly relevant for large batches such as dataset 2. In addition, PhageMiner can generate annotated maps of putative prophage sequences, allowing for almost instantaneous inspection, and it can automatically store the extracted prophage sequences. For complete prophages identified in dataset 2, the integrase amino acid sequence was compared against the phage integrase database derived from dataset 1 using blastp to classify the phage integrase type, as detailed for dataset 1. PhageMiner searches often recognized phages as partial rather complete, even for full prophages, e.g., due to annotation of integrase genes and other prophage-related genes as hypothetical proteins. To allow for visual differentiation between partial and full prophages, the inspection window was widened, generally by 15 genes on either side. Closed genomes in dataset 2 ( $n = 25$ ) were scanned manually for PICI identification, whereas draft genomes were screened for PICI presence with blast, searching for the integrase amino acid sequences of identified PICI.

## 2.3. Whole-Prophage and Integrase Gene Phylogenies

Two hundred eighty-two complete lysogenic prophages were detected in dataset 2. Using PhageMiner, all complete prophage sequences were extracted from the genomes (one complete prophage from 38% of genomes and two complete prophages from 9% of genomes) in dataset 2 and stored as GenBank files ( $n = 266$ ). Prophages that straddled two contigs were excluded from the phylogeny ( $n = 16$ ). Extracted prophages and 22 prophages identified by van der Mee-Marquet et al. (2018) were manually inspected and curated with Geneious v2020.1.2 (Biomatters Ltd, <https://www.geneious.com>). Sequences were reverse-complemented as needed to start with the integrase gene, and all integrase protein sequences were also stored separately. Multiple sequence alignments were performed for whole-prophage sequences and for integrase genes using ClustalW v2.1 (Thompson et al., 1994) with default settings (Gap opening



penalty = 10, Gap extension penalty = 0.20). Approximately-maximum-likelihood phylogenetic trees were constructed from the sequence alignments using FastTree v2.1.11 (Price et al., 2010) using the Jukes-Cantor model with default parameters. Figures were edited using Inkscape ([www.inkscape.org](http://www.inkscape.org)).

### 3. RESULTS

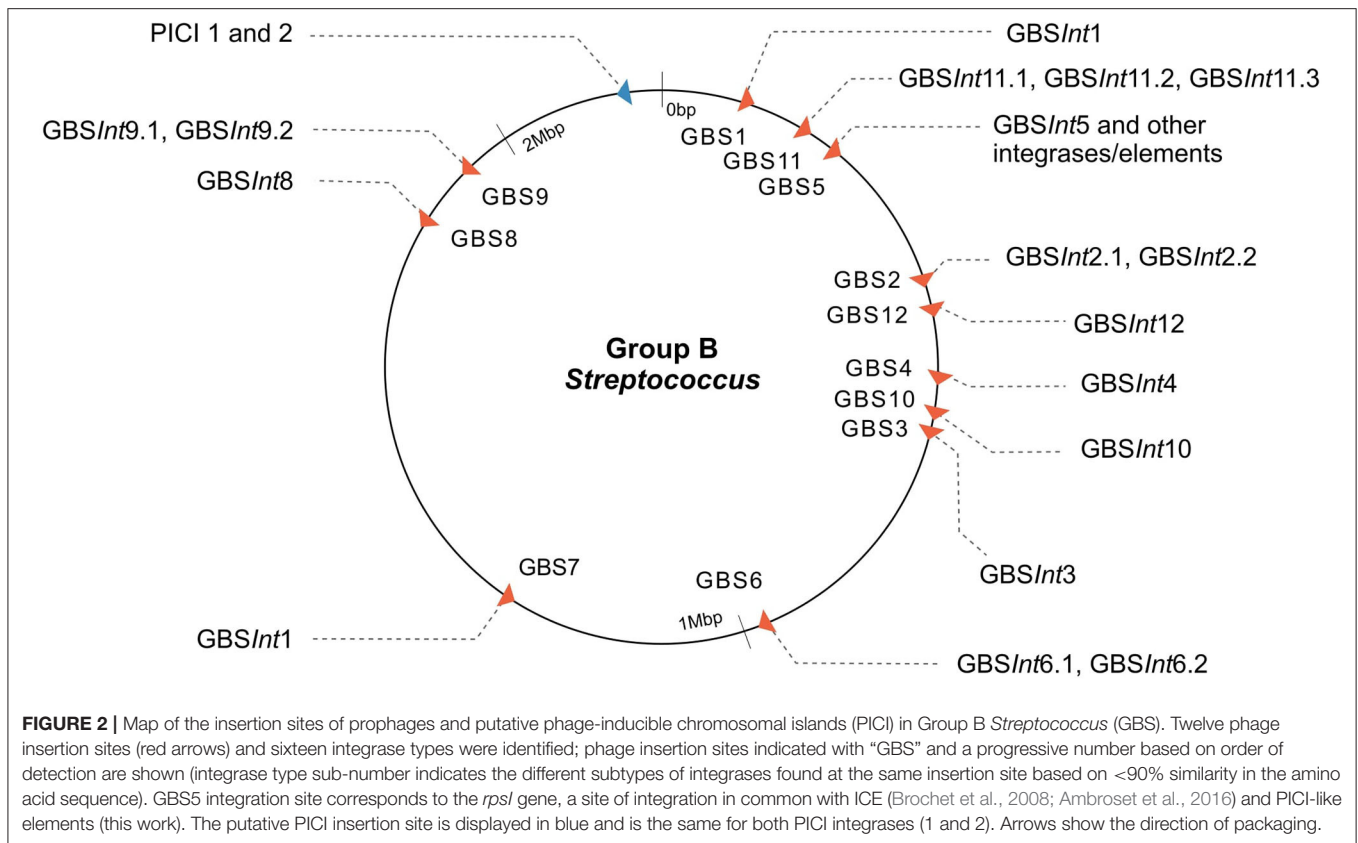
#### 3.1. Detection of Insertion Sites and Integrases Across the Prophage Phylogeny

Twelve integration sites were identified and progressively numbered as GBS1 to GBS12. The 12 integration sites were occupied by 16 integrase types, implying that there were subtypes for some integration sites (Figure 2, Table 1). Ten integrase types were identified in dataset 1, with two additional types and four subtypes identified in dataset 2. The complete database of integrase types can be found in **Supplementary Material, section 1.1** and at the GitHub online repository: ([https://github.com/chcrestani/GBS\\_prophage\\_integrase\\_typing](https://github.com/chcrestani/GBS_prophage_integrase_typing)). Putative attachment sites (Table 1) were identified bioinformatically for twelve integrase (sub)types, but the search was inconclusive for five (sub)types. Mean prophage integrase length was  $387 \pm 48$  AA (Table 1). Blast comparisons of integrase type % ID and QC can be found in Table S4. Integrase types and subtypes predominantly clustered with their respective prophages in the whole-prophage phylogeny (Figure 3). Major prophage groups were located at insertion sites GBS2, GBS3, GBS4, GBS9, and GBS11. Phages with GBSInt2.2 ( $n = 60$ ) were more common than those with GBSInt2.1 ( $n =$

2). For completeness, one representative sequence of prophage type GBSInt11.3, which had been identified in other analyses, was added to the phylogeny of phages, and its integrase was added to the integrase phylogeny. Minor prophage groups (GBS1, GBS5, GBS8, GBS10, GBS12) branched out from within major clusters. For some prophages, a mismatch between their integrase type and the integrase type of the surrounding prophage cluster was observed (red branches in Figure 3, Figure S2). This included all GBS6 prophages (GBSInt6.1 and GBSInt6.2), which were distributed across multiple branches of the prophage clades associated with GBS2 and GBS3 (Figure S2). GBSInt4 was associated with its own monophyletic phage clade and integration site, GBS4, but was also found on branches of the prophage clades associated with GBS2, GBS3 and GBS11. Likewise, GBSInt2.1 and GBSInt2.2, GBSInt3 and GBSInt11.1 were associated with their own clade and integration sites (GBS2, GBS3, and GBS11, respectively), as well as being detected in other prophage clades, i.e., GBS2 and/or GBS3. The integrase phylogeny (Figure S3) showed defined clusters, with varying levels of diversity within clusters. Integrases located at the same insertion site generally formed monophyletic clades, with the exception of GBSInt11.3, which was more closely related to GBSInt3 than to GBSInt11.1 or GBSInt11.2.

#### 3.2. Insertion Site Peculiarities and PIC1 Identification

GBSInt5 was identified in GBS5 (*rpsI* gene) in genome QMA0323, where the full prophage is present, preceded and followed by other genes with signatures of an ICE (integrative



conjugative element) (Figure S4). By contrast, in genome FSL\_S3-026, integrase GBSInt5 is present as a singleton, i.e., not followed by a full prophage. Rather, it was found inside what was classified as a putative ICE (~67,000 bp) by ICEFinder (Liu et al., 2018). This larger ICE showed partial similarity with a region of ~9,000 bp found after the prophage in QMA0323.

The label GBSInt7 is not used because the site-specific integrase at insertion site GBS7 was identical to GBSInt1 at insertion site GBS1 (Figure 1, Table 1). GBSInt1 at site GBS7 was only observed in this location when the GBS1 site was occupied by a prophage and it was uniquely observed in ST283, the only known hypervirulent GBS in human adults (Barkham et al., 2019).

At insertion site GBS11 (Table 1), the full prophage immediately followed *gspF* ( $n = 18$  genomes), or it was separated from *gspF* by a few genes encoding small proteins ( $n = 17$  genomes). The latter included competence proteins, type II secretion system proteins, and hypothetical proteins (Figure S5). There was no clear correlation between the integrase subtype (GBSInt11.1, GBSInt11.2) and any of these GBS11 site variants, but there was correlation between prophage subcluster and integrase type (Figure S6). For 26 prophages with either GBSInt11.1 or GBSInt11.2, it was not possible to assess the integration site because the prophage was found at the end of a contig.

In addition to prophages, two putative PIC1 sequences (PIC1 and PIC12) were identified using manual screening

(Figures 2, 4). Both integrases were 398 AA long and shared the same integration site (*rpsD* gene). Amino acid sequences for PIC1Int and PIC12Int can be found in Supplementary Material, section 1.2. PIC12 was uniquely identified in dataset 2. PIC1-like MGE were also detected in the integration site corresponding to the *rpsI* gene, i.e., in the same location as GBSInt5 (Figure S7). However, it was not possible to classify these PIC1-like elements with certainty, as they could have been fragments of other elements such as prophages or ICE (see section 4).

### 3.3. Detection of Prophages Across Host Species, Countries, and GBS Clades in Dataset 2

To create as complete an integrase database as possible, GBS genomes representing a wide variety of host species, countries, and GBS clades were included in the analysis. The study was not designed to be an epidemiologically representative survey of prophage or integrase distributions, so calculation of prophage prevalences is not meaningful, but some qualitative observations about the association with genome origin can be made.

Complete prophages were detected across isolates from most host species, including 47% of human GBS genomes ( $n = 230$  out of 486 isolates, with a total of 274 complete prophages) and three fish GBS genomes (Figure S8) but with the exception of bovine and canine GBS genomes ( $n = 5$  and 1, respectively). PIC1

**TABLE 1** | Insertion sites of 16 integrase types for phages identified in Group B *Streptococcus* isolated from across multiple host species and countries.

Insertion site	Phage Integrase type	Phage integrase length (AA)	Gene	Putative attachment site
GBS1	GBSInt1	369	<i>comX</i> (sigma-70 family RNA polymerase sigma factor)- 3' end	<i>attL</i> TTTTTTGTATAATATA <b>AAGA</b> <i>attR</i> TTTTTTGTATAATATA <b>ATA</b>
GBS2	GBSInt2.1	360	tRNA methyltransferase—3' end	ATCCCTCTCTCTCTTAAT
	GBSInt2.2	363		
GBS3	GBSInt3	368	<i>rplS</i> —3' end	<i>attL</i> GATTCC <b>GGC</b> AGGGGACAT <i>attR</i> GATTCC <b>AGC</b> AGGGGACAT
GBS4	GBSInt4	389	HU, histone-like DNA-binding protein	CTCTAAAGACGCTGTTAAATA ATTCGTCTAGAAAAACCTTGTC ATATCAATGTTTATTGATAGCGAC AAGGTTCT
GBS5	GBSInt5	331	<i>rpsI</i> —3' end (within ICE)	—
GBS6	GBSInt6.1	366	CatB-related O-acetyltransferase—5' end	TGGAGCCGGTGGGAGT
	GBSInt6.2	366		
GBS7	GBSInt1	369	<i>hylB</i> —5' end	<i>attL</i> TTTTTTGTATAATAT <b>AAGA</b> <i>attR</i> TTTTTTGTATAATAT <b>GAGA</b>
GBS8	GBSInt8	382	YbaB/EbfC family nucleoid-associated protein	TTTTGCATATTCATCATA
GBS9	GBSInt9.1	360	<i>nhaK</i> (sodium/proton antiporter)—3' end	AAGCGGTAGACGGATTGAA
	GBSInt9.2	359		
GBS10	GBSInt10	476	DNA-binding protein WhiA—3' end	—
GBS11	GBSInt11.1	489	<i>gspF</i> or <i>gspF</i> +competence proteins/type II	CTTTTGAATGTTTGGTA—
	GBSInt11.2	486	secretion system proteins—3' end	
	GBSInt11.3	367		
GBS12	GBSInt12	386	5-formyltetrahydrofolate cyclo-ligase—5' end	—

With the exception of GBSInt1, integrases are site specific. Putative attachment sites (*att*) are shown when known. For *att* sites that differed slightly at the two ends of the lysogenic prophages, left and right (*attL* and *attR*) sequences are specified and differences are highlighted in bold.

were also detected across GBS from most host species, with PIC11 found in a total of 328 GBS genomes from humans, fish, cattle, a dog and a dolphin, and PIC12 found in a camel GBS genome from Kenya.

Prophages and integrases were detected in most GBS clades, with the exception of certain clades represented by 3 or fewer isolates (CC22, CC67, and CC130, **Table S5**), and the majority of integrases were detected in multiple clades (**Figure 5**). The number of integrase types per CC ranged from 1 to 10 (**Figure 5**). All major ST in dataset 2 (ST1, ST17, ST19, ST23, ST459) harbored at least four prophage types (**Table S6**).

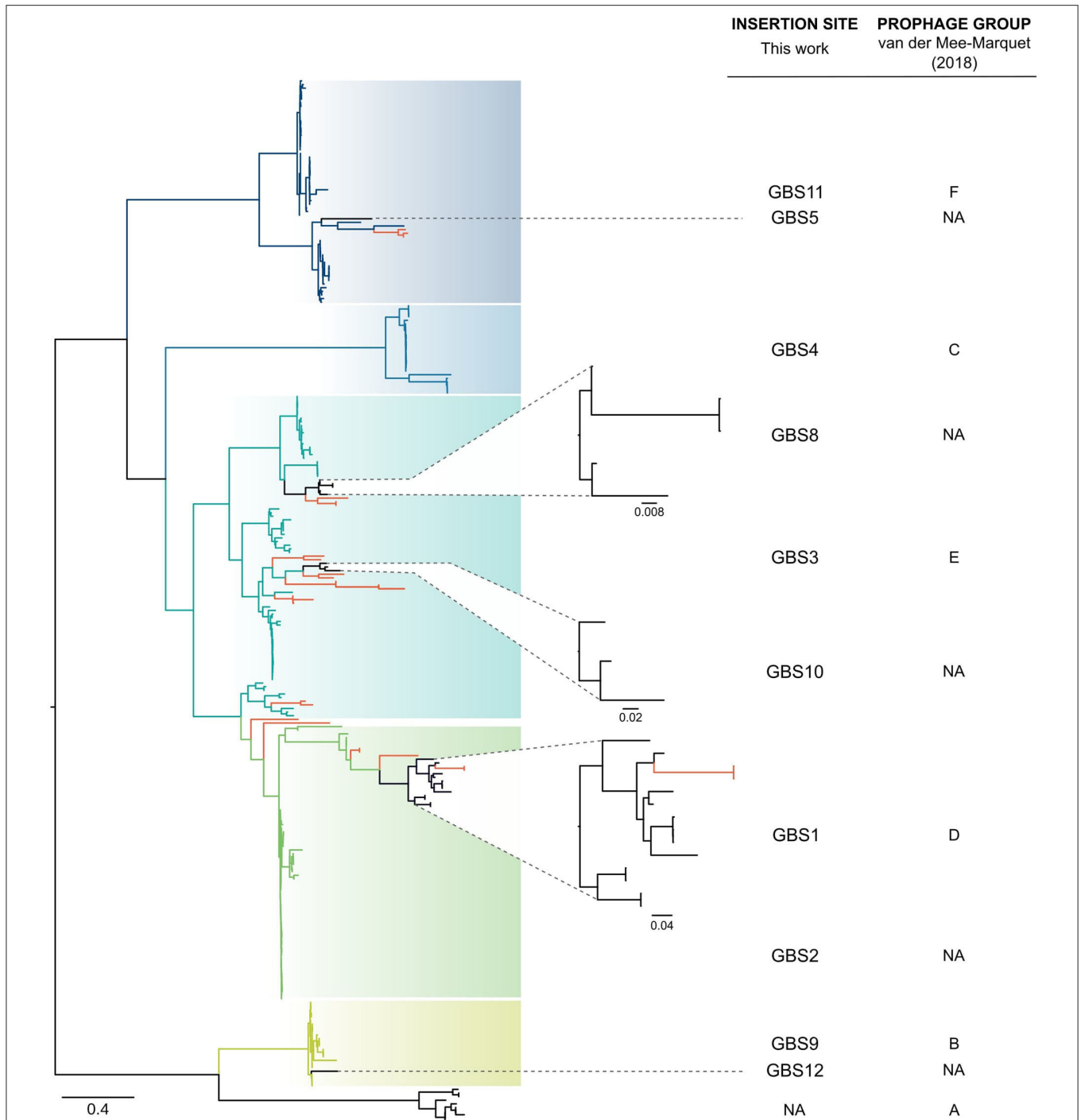
Complete prophages were identified in GBS isolates from all continents except for South America. The number of discovered prophages tended to reflect the total number of genomes per continent, whereby more prophages were detected in continents with more genome sequences (**Table S7**).

#### 4. DISCUSSION AND CONCLUSIONS

We describe the development of a typing scheme for GBS prophages based on site-specific integrase genes and insertion sites, similar to the scheme used for prophage typing of *S. aureus* (Goerke et al., 2009; Valentin-Domelier et al., 2011; Jamrozny et al., 2017). The scheme is intended for detection of putative prophages through BLAST searches (tblastn) using one or several genome assemblies (nucleotide sequences) as subject sequences, and the database of integrase protein types

presented here as query sequence as detailed in the section 2, **Supplementary Material**, and online ([https://github.com/chcrestani/GBS\\_prophage\\_integrase\\_typing](https://github.com/chcrestani/GBS_prophage_integrase_typing)). This approach enables the rapid screening of large datasets of complete and draft genomes for the presence of GBS prophages, overcoming some of the limitations associated with existing phage detection programs, and enabling detection of phage content in fragmented genome assemblies. Additionally, BLAST-based searches of integrases can be performed by those with little computational experience, as BLAST is available as an online platform.

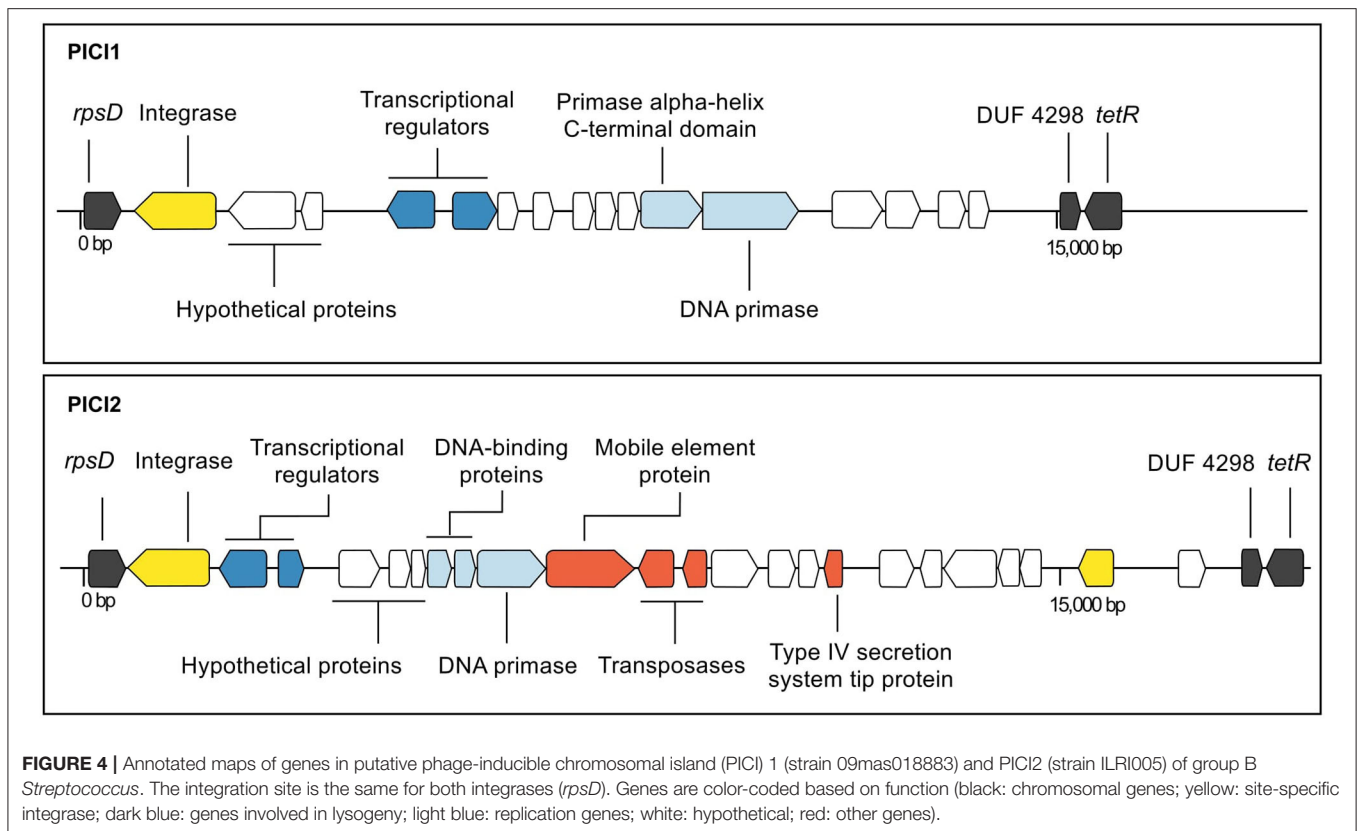
Phage integrase typing agreed with full-length prophage genome-based phylogenetic clusters, with a few exceptions. This is reminiscent of the relationship between the GBS whole-genome phylogeny and capsular serotypes, where serotypes tend to match phylogenetic clusters but capsular switching may occur (Martins et al., 2010; Bellais et al., 2012; Neemuchwala et al., 2016). We propose that integrase switching may also occur, leading to “mismatches” between prophage genome phylogeny and integrase phylogeny, and conferring to the prophage the ability to integrate in a different location in the GBS genome. This genome plasticity may impact on the function of the prophage, and on packaging of GBS genome content. There is growing evidence that prophages contribute to emergence, niche adaptation and spread of virulent GBS, especially in CC1 and CC17 (van der Mee-Marquet et al., 2018; Renard et al., 2019; Jamrozny et al., 2020). This may include transfer of prophage content between GBS from different host species, in



**FIGURE 3 |** Approximately-maximum-likelihood phylogenetic tree of 266 complete prophages identified in group B *Streptococcus* (GBS) in this study and 22 prophages identified by van der Mee-Marquet et al. (2018). In most cases, full-prophage phylogenetic clusters are concordant with insertion sites and their corresponding integrase types or subtypes (GBS2, GBS3, GBS4, GBS9, GBS11, blue to green branches), with smaller clusters (GBS1, GBS5, GBS8, GBS10, GBS12, black branches) embedded in the larger ones. Some mismatches between prophages and their integrase type or insertion site and major clusters were identified (red branches), which is suggestive of integrase switching events. Tree was rooted at midpoint. NA, not applicable.

agreement with the detection of prophage types and integrase types across GBS from different host species in our dataset. We discovered a potential contribution of prophages to the

emergence of hypervirulent ST283, which has recently been recognized as a major cause of adult invasive disease in Southeast Asia (Rajendram et al., 2016; Kalimuddin et al., 2017; Barkham



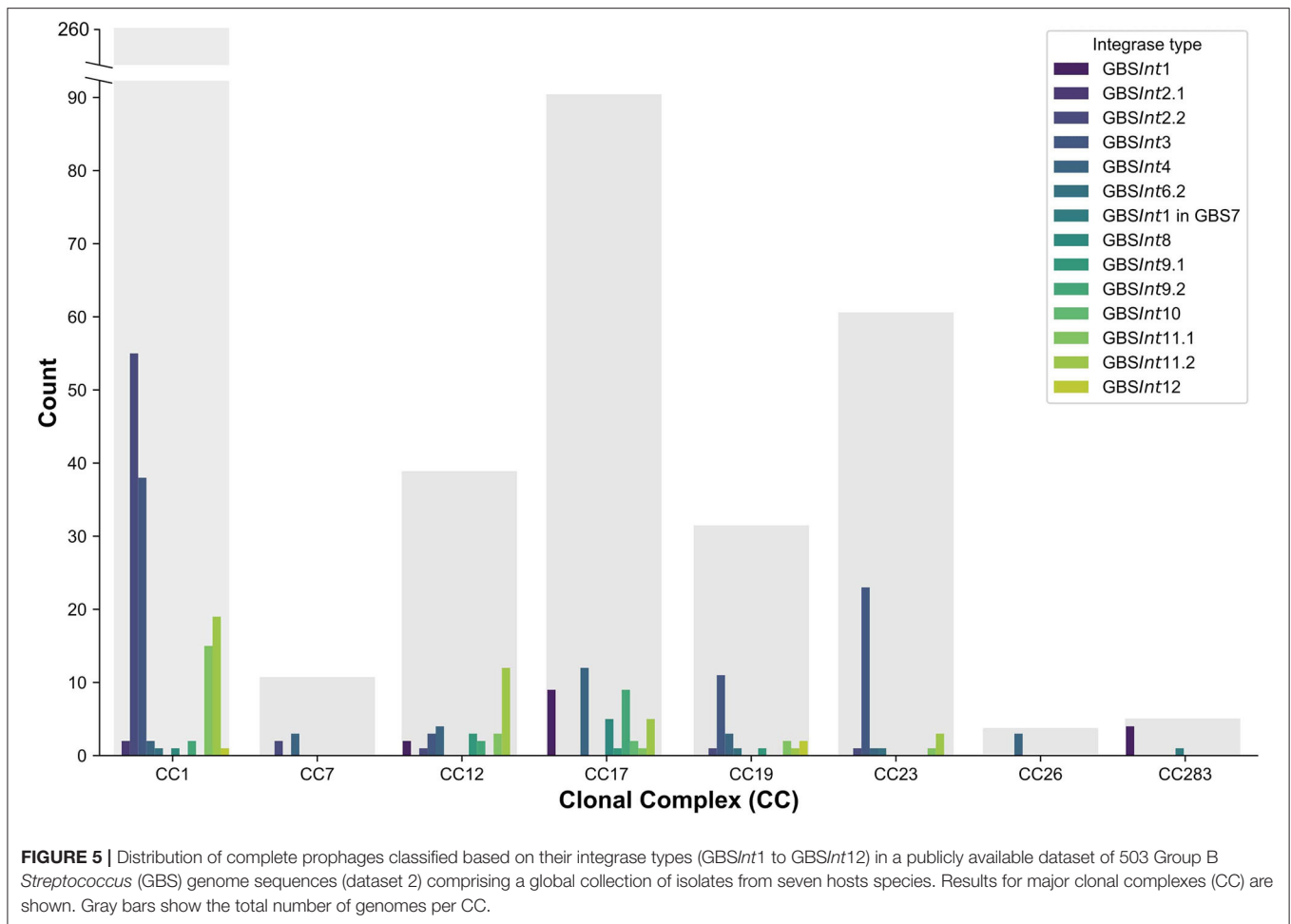
et al., 2019). Contradicting the dogma that phage integrase genes are site-specific (Frost et al., 2005), the integrase at insertion site GBS7 (5' end of *hylB*), was identical to the integrase at GBS1. Prophages in GBS7-*hylB* were only present when GBS1 was also occupied by a prophage, and they were unique to ST283. The virulence gene *hylB* codes for hyaluronate lyase, an enzyme that degrades extracellular matrix components and is believed to contribute significantly to invasion (Herbert et al., 2004; Maisey et al., 2008). We hypothesize that this prophage plays a role in regulation of the transcription and expression of *hylB* and contributes to the hypervirulence of ST283.

Our analysis of 572 GBS genomes extends previous work by van der Mee-Marquet and colleagues—who identified prophages in 14 GBS genome sequences to subsequently screen by PCR a larger collection of isolates ( $n = 275$ )—by increasing the number of known prophages, insertion sites and integrase types. Our major full prophage clades matched previously defined prophage groups (Prophage group B = GBS9, C = GBS4, D = GBS1, E = GBS3, F = GBS11) (van der Mee-Marquet et al., 2018), whilst other clusters, including those located at GBS2, GBS5, GBS8, GBS10, and GBS12, are described here for the first time. As the GBS genome database expands, the typing scheme will need to be updated with emerging integrase types and subtypes. This is also illustrated by results for insertion site GBS11 (3' end of *gspF*), which is located within an operon involved with host competence (*com* operon). GBS11 had previously been classified as two separate insertion sites, F1 and F2, based on

variations observed between three prophage genomes at this site (van der Mee-Marquet et al., 2018). Based on our analysis, the bifurcation of this group of prophages, which was also observed in our whole-prophage phylogeny, correlated with different integrase types (GBSInt11.1 and GBSInt11.2), rather than with different insertion sites. In many cases the insertion site for those integrases could not be confirmed because they were located at the edge of a contig. This suggests that sequence assembly tools struggle to assemble this region of the GBS genome, an issue that could be overcome by long read sequencing.

Our typing scheme does not include type A prophages (van der Mee-Marquet et al., 2018) because they are defective rather than whole prophages, lacking the integrase gene. Although this could be considered a false negative result in our typing scheme, lack of an integrase gene renders type A prophages incapable of horizontal gene transfer so that they can only be spread through vertical transmission, limiting their contribution to the evolution of virulence or niche adaptation. Based on integrase typing, false positive results may also occur, as demonstrated for GBSInt5, which was found once as part of a full prophage (isolate QMA0323, piscine ST261; Kawasaki et al., 2018) and once as a singleton within a larger ICE (isolate FSL S3-026, bovine ST67; Richards et al., 2011). A BLAST search of genomes with more than 50 contigs showed the presence of GBSInt5 as a singleton within an ICE rather than as part of a full prophage in nine bovine GBS genomes from bovine-associated lineage CC67 (Richards et al., 2019). This phenomenon was





only observed for GBSInt5, possibly because its insertion site, *rpsI* (30S ribosomal protein S9), is a hotspot for recombination of ICE in streptococcal species (Brochet et al., 2008; Ambroset et al., 2016). When this integrase is identified within a dataset, further analyses need to be performed to determine whether a full prophage is present. The GBS5 insertion site also contained PICI-like elements. Because of these multiple integration events, PICI-like elements in this position could not be classified with certainty, as they could have been fragments of prophages or ICE.

PICI1 and PICI2, which are reported here for the first time, were integrated into *rpsD*, which encodes 30S ribosomal protein S4. This gene had previously been described as the site of integration of an *S. agalactiae* chromosomal island (SagCI) in 3 of 9 complete GBS genomes (Nguyen and McShan, 2014), but not as a PICI. GBS differs from other members of the pyogenes group, including *S. pyogenes*, *S. canis*, *S. dysgalactiae* subsp. *equisimilis*, and *S. parauberis*, in that their chromosomal islands are primarily integrated in *mutL* rather than *rpsD*, which may affect their functional impact. The structure of both PICI1 and PICI2 includes typical PICI features such as transcriptional divergence, a size of around 15,000 bp, and presence of a DNA primase (Martínez-Rubio et al., 2017), whereas the variable portion of their organization and content

resembles the structure of SpnCIST556 in *S. pneumoniae* and SpyCI6180 in *S. pyogenes*, respectively (Penadés and Christie, 2015). Streptococcal PICI may have roles in gene regulation (Nguyen and McShan, 2014) or gene transfer (Martínez-Rubio et al., 2017). The high prevalence of PICI1 across GBS genomes from different host species, geographic origins and clades suggests that its function warrants further investigation. By contrast, PICI2 was exclusively found in one isolate from CC609, which is a camel-specific clade from East Africa (Fischer et al., 2013). Further research and experimentation would be needed to understand if and how PICI play a role in GBS evolution and virulence.

In summary, we propose a new typing scheme for rapid prophage identification in large datasets of GBS genomes based on site-specific integrase types and their insertion site. This method provides a practical way of identifying potential prophage presence with a BLAST-based approach in full and draft genomes, overcoming detection issues related to genome fragmentation and making it user-friendly for researchers with any level of computational experience. We show that multiple prophages and integrase types occur across GBS from a wide range of host species, geographic origins and clades, and that a secondary prophage was uniquely present in hypervirulent GBS

ST283. In addition, we report the high prevalence of putative PICI in GBS, opening up a new area of GBS research.

## DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

## AUTHOR CONTRIBUTIONS

RZ and CC conceived the study. CC conducted the data analysis with guidance from TF. CC drafted the manuscript. All authors edited the manuscript.

## FUNDING

CC was supported by the University of Glasgow College of Medical, Veterinary, and Life Sciences Doctoral Training

## REFERENCES

- Ambroset, C., Coluzzi, C., Guédon, G., Devignes, M.-D., Loux, V., Lacroix, T., et al. (2016). New insights into the classification and integration specificity of *Streptococcus* integrative conjugative elements through extensive genome exploration. *Front. Microbiol.* 6:1483. doi: 10.3389/fmicb.2015.01483
- Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21. doi: 10.1093/nar/gkw387
- Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., et al. (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75. doi: 10.1186/1471-2164-9-75
- Bai, Q., Zhang, W., Yang, Y., Tang, F., Nguyen, X., Liu, G., et al. (2013). Characterization and genome sequencing of a novel bacteriophage infecting *Streptococcus agalactiae* with high similarity to a phage from *Streptococcus pyogenes*. *Arch. Virol.* 158, 1733–1741. doi: 10.1007/s00705-013-1667-x
- Barkham, T., Zadoks, R. N., Azmai, M. N. A., Baker, S., Bich, V. T. N., Chalker, V., et al. (2019). 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.* 13:e7421. doi: 10.1371/journal.pntd.0007421
- Bellais, S., Six, A., Fouet, A., Longo, M., Dmytruk, N., Glaser, P., et al. (2012). Capsular switching in group B *Streptococcus* CC17 hypervirulent clone: a future challenge for polysaccharide vaccine development. *J. Infect. Dis.* 206, 1745–1752. doi: 10.1093/infdis/jis605
- Bennett, S. (2004). Solexa ltd. *Pharmacogenomics* 5, 433–438. doi: 10.1517/14622416.5.4.433
- Bose, M., and Barber, R. D. (2006). Prophage Finder: a prophage loci prediction tool for prokaryotic genome sequences. *In Silico Biol.* 6, 223–227.
- Brochet, M., Couvé, E., Glaser, P., Guédon, G., and Payot, S. (2008). Integrative conjugative elements and related elements are major contributors to the genome diversity of *Streptococcus agalactiae*. *J. Bacteriol.* 190, 6913–6917. doi: 10.1128/JB.00824-08
- Brochet, M., Couvé, E., Zouine, M., Vallaes, T., Rusniok, C., Lamy, M.-C., et al. (2006). Genomic diversity and evolution within the species *Streptococcus agalactiae*. *Microbes Infect.* 8, 1227–1243. doi: 10.1016/j.micinf.2005.11.010
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421
- Campbell, A. M. (1992). Chromosomal insertion sites for phages and plasmids. *J. Bacteriol.* 174:7495. doi: 10.1128/JB.174.23.7495-7499.1992
- Davies, M. R., Tran, T. N., McMillan, D. J., Gardiner, D. L., Currie, B. J., and Sriprakash, K. S. (2005). Inter-species genetic movement may blur the epidemiology of streptococcal diseases in endemic regions. *Microbes Infect.* 7, 1128–1138. doi: 10.1016/j.micinf.2005.03.018
- Delannoy, C. M., Zadoks, R. N., Cruimlish, M., Rodgers, D., Lainson, F. A., Ferguson, H., et al. (2016). Genomic comparison of virulent and non-virulent *Streptococcus agalactiae* in fish. *J. Fish Dis.* 39, 13–29. doi: 10.1111/jfd.12319
- Domelier, A.-S., van Der Mee-Marquet, N., Sizaret, P.-Y., Héry-Arnaud, G., Lartigue, M.-F., Mereghetti, L., et al. (2009). Molecular characterization and lytic activities of *Streptococcus agalactiae* bacteriophages and determination of lysogenic-strain features. *J. Bacteriol.* 191, 4776–4785. doi: 10.1128/JB.00426-09
- Farley, M. M., and Strasbaugh, L. J. (2001). Group B streptococcal disease in nonpregnant adults. *Clin. Infect. Dis.* 33, 556–561. doi: 10.1086/322696
- Fillol-Salom, A., Martínez-Rubio, R., Abdulrahman, R. F., Chen, J., Davies, R., and Penadés, J. R. (2018). Phage-inducible chromosomal islands are ubiquitous within the bacterial universe. *ISME J.* 12:2114. doi: 10.1038/s41396-018-0156-3
- Fischer, A., Liljander, A., Kaspar, H., Muriuki, C., Fuxelius, H.-H., Bongcam-Rudloff, E., et al. (2013). Camel *Streptococcus agalactiae* populations are associated with specific disease complexes and acquired the tetracycline resistance *genetM* via a Tn916-like element. *Vet. Res.* 44:86. doi: 10.1186/1297-9716-44-86
- Frost, L. S., Leplae, R., Summers, A. O., and Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. *Nat. Rev. Microbiol.* 3, 722–732. doi: 10.1038/nrmicro1235
- Goerke, C., Pantucek, R., Holtfreter, S., Schulte, B., Zink, M., Grumann, D., et al. (2009). Diversity of prophages in dominant *Staphylococcus aureus* clonal lineages. *J. Bacteriol.* 191, 3462–3468. doi: 10.1128/JB.01804-08
- Herbert, M. A., Beveridge, C. J., and Saunders, N. J. (2004). Bacterial virulence factors in neonatal sepsis: group B *Streptococcus*. *Curr. Opin. Infect. Dis.* 17, 225–229. doi: 10.1097/00001432-200406000-00009
- High, K. P., Edwards, M. S., and Baker, C. J. (2005). Group B streptococcal infections in elderly adults. *Clin. Infect. Dis.* 41, 839–847. doi: 10.1086/432804
- Holden, M. T., Heather, Z., Paillot, R., Steward, K. F., Webb, K., Ainslie, F., et al. (2009). Genomic evidence for the evolution of *Streptococcus equi*: host restriction, increased virulence, and genetic exchange with human pathogens. *PLoS Pathogens* 5:e1000346. doi: 10.1371/journal.ppat.1000346
- Jafar, Q., Sameer, A.-Z., Salwa, A.-M., Samee, A.-A., Ahmed, A.-M., and Al-Sharifi, F. (2008). Molecular investigation of *Streptococcus agalactiae* isolates from environmental samples and fish specimens during a massive fish kill in Kuwait Bay. *Pakistan J. Biol. Sci.* 11, 2500–2504. doi: 10.3923/pjbs.2008.25.00.2504
- Jamrozny, D., Bijlsma, M. W., de Goffau, M. C., van de Beek, D., Kuijpers, T. W., Parkhill, J., et al. (2020). Increasing incidence of group B streptococcus neonatal infections in the Netherlands is associated with clonal expansion of CC17 and CC23. *Sci. Rep.* 10:9539. doi: 10.1038/s41598-020-66214-3

## ACKNOWLEDGMENTS

We thank Dr. Nuria Quiles-Puchalt and Prof. José Penadés (University of Glasgow) for their expertise and knowledge exchange on bacteriophages and PICI structure and ecology, as well as for early training of CC. We acknowledge Dr. Vincent Richards (Clemson University) for data access and Dr. Reza Rezaei Javan (University of Oxford) for granting early access to PhageMiner.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2020.01993/full#supplementary-material>

- Jamrozny, D., Coll, F., Mather, A. E., Harris, S. R., Harrison, E. M., MacGowan, A., et al. (2017). Evolution of mobile genetic element composition in an epidemic methicillin-resistant *Staphylococcus aureus*: temporal changes correlated with frequent loss and gain events. *BMC Genomics* 18:684. doi: 10.1186/s12864-017-4065-z
- Javan, R. R., Ramos-Sevillano, E., Akter, A., Brown, J., and Brueggemann, A. B. (2019). Prophages and satellite prophages are widespread in *Streptococcus* and may play a role in pneumococcal pathogenesis. *Nat. Commun.* 10, 1–14. doi: 10.1038/s41467-019-12825-y
- Kalimuddin, S., Chen, S. L., Lim, C. T., Koh, T. H., Tan, T. Y., Kam, M., et al. (2017). 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.* 64(Suppl. 2), S145–S152. doi: 10.1093/cid/cix021
- Kawasaki, M., Delamare-Deboutteville, J., Bowater, R. O., Walker, M. J., Beatson, S., Zakour, N. L. B., et al. (2018). Microevolution of *Streptococcus agalactiae* ST-261 from Australia indicates dissemination via imported tilapia and ongoing adaptation to marine hosts or environment. *Appl. Environ. Microbiol.* 84:e00859-18. doi: 10.1128/AEM.00859-18
- Le Doare, K., O’driscoll, M., Turner, K., Seedat, F., Russell, N. J., Seale, A. C., et al. (2017). Intrapartum antibiotic chemoprophylaxis policies for the prevention of group B streptococcal disease worldwide: systematic review. *Clin. Infect. Dis.* 65(Suppl. 2), S143–S151. doi: 10.1093/cid/cix654
- Lima-Mendez, G., Van Helden, J., Toussaint, A., and Leplae, R. (2008). Prophinder: a computational tool for prophage prediction in prokaryotic genomes. *Bioinformatics* 24, 863–865. doi: 10.1093/bioinformatics/btn043
- Liu, L., Li, Y., He, R., Xiao, X., Zhang, X., Su, Y., et al. (2014). Outbreak of *Streptococcus agalactiae* infection in barcoo grunter, *Scortum barcoo* (McCulloch & Waite), in an intensive fish farm in China. *Fish Dis.* 37, 1067–1072. doi: 10.1111/jfd.12187
- Liu, M., Li, X., Xie, Y., Bi, D., Sun, J., Li, J., et al. (2018). ICEberg 2.0: an updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res.* 47, D660–D665. doi: 10.1093/nar/gky1123
- Lyhs, U., Kulkas, L., Katholm, J., Waller, K. P., Saha, K., Tomusk, R. J., et al. (2016). *Streptococcus agalactiae* serotype IV in humans and cattle, northern Europe. *Emerg. Infect. Dis.* 22:2097. doi: 10.3201/eid2212.151447
- Maisey, H. C., Doran, K. S., and Nizet, V. (2008). Recent advances in understanding the molecular basis of group B *Streptococcus* virulence. *Expert Rev. Mol. Med.* 10:e27. doi: 10.1017/S1462399408000914
- Martínez-Rubio, R., Quiles-Puchalt, N., Martí, M., Humphrey, S., Ram, G., Smyth, D., et al. (2017). Phage-inducible islands in the Gram-positive cocci. *Int. Soc. Microb. Ecol. J.* 11:1029. doi: 10.1038/ismej.2016.163
- Martins, E. R., Melo-Cristino, J., and Ramirez, M. (2010). Evidence for rare capsular switching in *Streptococcus agalactiae*. *J. Bacteriol.* 192, 1361–1369. doi: 10.1128/JB.01130-09
- Neemuchwala, A., Teatero, S., Athey, T. B., McGeer, A., and Fittipaldi, N. (2016). Capsular switching and other large-scale recombination events in invasive sequence type 1 group B *Streptococcus*. *Emerg. Infect. Dis.* 22:1941. doi: 10.3201/eid2211.152064
- Nguyen, S. V., and McShan, W. M. (2014). Chromosomal islands of *Streptococcus pyogenes* and related streptococci: molecular switches for survival and virulence. *Front. Cell. Infect. Microbiol.* 4:109. doi: 10.3389/fcimb.2014.00109
- Penadés, J. R., and Christie, G. E. (2015). The phage-inducible chromosomal islands: a family of highly evolved molecular parasites. *Annu. Rev. Virol.* 2, 181–201. doi: 10.1146/annurev-virology-031413-085446
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2-approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5:e9490. doi: 10.1371/journal.pone.0009490
- Rajendram, P., Kyaw, W. M., Leo, Y. S., Ho, H., Chen, W. K., Lin, R., et al. (2016). Group B *Streptococcus* sequence type 283 disease linked to consumption of raw fish, Singapore. *Emerg. Infect. Dis.* 22:1974. doi: 10.3201/eid2211.160252
- Renard, A., Barbera, L., Courtier-Martinez, L., Dos Santos, S., Valentin, A.-S., Mereghetti, L., et al. (2019). phiD12-like livestock-associated prophages are associated with novel subpopulations of *Streptococcus agalactiae* infecting neonates. *Front. Cell. Infect. Microbiol.* 9:166. doi: 10.3389/fcimb.2019.00166
- Richards, V. P., Lang, P., Bitar, P. D. P., Lefebure, T., Schukken, Y. H., Zadoks, R. N., et al. (2011). Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted *Streptococcus agalactiae*. *Infect. Genet. Evol.* 11, 1263–1275. doi: 10.1016/j.meegid.2011.04.019
- Richards, V. P., Velsko, I. M., Alam, T., Zadoks, R. N., Manning, S. D., Pavinski Bitar, P. D., et al. (2019). Population gene introgression and high genome plasticity for the zoonotic pathogen *Streptococcus agalactiae*. *Mol. Biol. Evol.* 36, 2572–2590. doi: 10.1093/molbev/msz169
- Salloum, M., van der Mee-Marquet, N., Domelier, A.-S., Arnault, L., and Quentin, R. (2010). Molecular characterization and prophage DNA contents of *Streptococcus agalactiae* strains isolated from adult skin and osteoarticular infections. *J. Clin. Microbiol.* 48, 1261–1269. doi: 10.1128/JCM.01820-09
- Salloum, M., Van Der Mee-marquet, N., Valentin-Domelier, A.-S., and Quentin, R. (2011). Diversity of prophage DNA regions of *Streptococcus agalactiae* clonal lineages from adults and neonates with invasive infectious disease. *PLoS ONE* 6:e20256. doi: 10.1371/journal.pone.0020256
- Seale, A. C., Bianchi-Jassir, F., Russell, N. J., Kohli-Lynch, M., Tann, C. J., Hall, J., et al. (2017). Estimates of the burden of group B streptococcal disease worldwide for pregnant women, stillbirths, and children. *Clin. Infect. Dis.* 65, S200–S219. doi: 10.1093/cid/cix664
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Skoff, T. H., Farley, M. M., Petit, S., Craig, A. S., Schaffner, W., Gershman, K., et al. (2009). Increasing burden of invasive group B streptococcal disease in nonpregnant adults, 1990–2007. *Clin. Infect. Dis.* 49, 85–92. doi: 10.1086/599369
- Sørensen, U. B. S., Klaas, I. C., Boes, J., and Farre, M. (2019). The distribution of clones of *Streptococcus agalactiae* (group B streptococci) among herdspersons and dairy cows demonstrates lack of host specificity for some lineages. *Vet. Microbiol.* 235, 71–79. doi: 10.1016/j.vetmic.2019.06.008
- Suzuki, H., Lefebure, T., Hubisz, M. J., Pavinski Bitar, P., Lang, P., Siepel, A., et al. (2011). Comparative genomic analysis of the *Streptococcus dysgalactiae* species group: gene content, molecular adaptation, and promoter evolution. *Genome Biol. Evol.* 3, 168–185. doi: 10.1093/gbe/evr006
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673
- Valentin-Domelier, A.-S., Girard, M., Bertrand, X., Violette, J., François, P., Donnio, P.-Y., et al. (2011). Methicillin-susceptible ST398 *Staphylococcus aureus* responsible for bloodstream infections: an emerging human-adapted subclone? *PLoS ONE* 6:e28369. doi: 10.1371/journal.pone.0028369
- van der Mee-Marquet, N., Diene, S. M., Barbera, L., Courtier-Martinez, L., Lafont, L., Ouachée, A., et al. (2018). Analysis of the prophages carried by human infecting isolates provides new insight into the evolution of group B *Streptococcus* species. *Clin. Microbiol. Infect.* 24, 514–521. doi: 10.1016/j.cmi.2017.08.024
- van der Mee-Marquet, N., Domelier, A.-S., Mereghetti, L., Lanotte, P., Rosenau, A., van Leeuwen, W., et al. (2006). Prophagic DNA fragments in *Streptococcus agalactiae* strains and association with neonatal meningitis. *J. Clin. Microbiol.* 44, 1049–1058. doi: 10.1128/JCM.44.3.1049-1058.2006
- Zadoks, R. N., Middleton, J. R., McDougall, S., Katholm, J., and Schukken, Y. H. (2011). Molecular epidemiology of mastitis pathogens of dairy cattle and comparative relevance to humans. *J. Mamm. Gland Biol. Neoplasia* 16, 357–372. doi: 10.1007/s10911-011-9236-y
- Zamri-Saad, M. (2018). “GBS in fish and feed based vaccination,” in *Abstract for the 1st International Symposium on ‘Streptococcus agalactiae’ Disease* (Cape Town).
- Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J., and Wishart, D. S. (2011). PHAST: a fast phage search tool. *Nucleic Acids Res.* 39(Suppl. 2), W347–W352. doi: 10.1093/nar/gkr485

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Crestani, Forde and Zadoks. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.