AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

EVOLUTIONARY AND GENOMIC MICROBIOLOGY

Microevolution of *Streptococcus agalactiae* ST-261 from Australia Indicates Dissemination via Imported Tilapia and Ongoing Adaptation to Marine Hosts or Environment

Minami Kawasaki,[a] Jerome Delamare-Debouteville,[a] Rachel O. Bowater,[b] Mark J. Walker,[c] Scott Beatson,[c] Nouri L. Ben Zakour,[c]* Andrew C. Barnes[a]

[a]The University of Queensland, School of Biological Sciences and Centre for Marine Science, St. Lucia Campus, Brisbane, Queensland, Australia

[b]Biosecurity Queensland (North Region), Department of Agriculture and Fisheries, Townsville, Queensland, Australia

[c]The University of Queensland, School of Chemistry and Molecular Biosciences and Australian Infectious Diseases Research Centre, St. Lucia Campus, Brisbane, Queensland, Australia

ABSTRACT *Streptococcus agalactiae* (group B *Streptococcus* [GBS]) causes disease in a wide range of animals. The serotype Ib lineage is highly adapted to aquatic hosts, exhibiting substantial genome reduction compared with terrestrial conspecifics. Here, we sequence genomes from 40 GBS isolates, including 25 isolates from wild fish and captive stingrays in Australia, six local veterinary or human clinical isolates, and nine isolates from farmed tilapia in Honduras, and compared them with 42 genomes from public databases. Phylogenetic analysis based on nonrecombinant core-genome single nucleotide polymorphisms (SNPs) indicated that aquatic serotype Ib isolates from Queensland were distantly related to local veterinary and human clinical isolates. In contrast, Australian aquatic isolates are most closely related to a tilapia isolate from Israel, differing by only 63 core-genome SNPs. A consensus minimum spanning tree based on core-genome SNPs indicates the dissemination of sequence type 261 (ST-261) from an ancestral tilapia strain, which is congruent with several introductions of tilapia into Australia from Israel during the 1970s and 1980s. Pangenome analysis identified 1,440 genes as core, with the majority being dispensable or strain specific, with non-protein-coding intergenic regions (IGRs) divided among core and strain-specific genes. Aquatic serotype Ib strains have lost many virulence factors during adaptation, but six adhesins were well conserved across the aquatic isolates and might be critical for virulence in fish and for targets in vaccine development. The close relationship among recent ST-261 isolates from Ghana, the United States, and China with the Israeli tilapia isolate from 1988 implicates the global trade in tilapia seed for aquaculture in the widespread dissemination of serotype Ib fish-adapted GBS.

IMPORTANCE *Streptococcus agalactiae* (GBS) is a significant pathogen of humans and animals. Some lineages have become adapted to particular hosts, and serotype Ib is highly specialized to fish. Here, we show that this lineage is likely to have been distributed widely by the global trade in tilapia for aquaculture, with probable introduction into Australia in the 1970s and subsequent dissemination in wild fish populations. We report here the variability in the polysaccharide capsule among this lineage but identify a cohort of common surface proteins that may be a focus of future vaccine development to reduce the biosecurity risk in international fish trade.

KEYWORDS *Streptococcus agalactiae*, evolution, genome analysis, epidemiology, fish, aquaculture

Received 12 April 2018 **Accepted** 12 June 2018

Accepted manuscript posted online 18 June 2018

Citation Kawasaki M, Delamare-Debouteville J, Bowater RO, Walker MJ, Beatson S, Ben Zakour NL, Barnes AC. 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. https://doi.org/10.1128/AEM.00859-18.

Editor Johanna Björkroth, University of Helsinki

Address correspondence to Andrew C. Barnes, a.barnes@uq.edu.au.

* Present address: Nouri L. Ben Zakour, Westmead Institute for Medical Research, University of Sydney, Westmead, NSW, Australia.

Streptococcus agalactiae, or Lancefield group B *Streptococcus* (GBS), is a commensal and occasionally pathogenic bacterium with a very diverse host range. A common commensal in the urogenital tracts of humans, GBS is also a leading cause of morbidity in newborns causing meningitis, septicemia, and pneumonia (1–4). *S. agalactiae* can cause septicemic infections in cattle, domestic dogs and cats, camels, reptiles, and amphibians (5–8). In fish, disease outbreaks caused by *S. agalactiae* have substantial impact on the aquaculture industry, particularly on the production of warm freshwater species, such as tilapia (*Oreochromis* spp.) (9–12). Most outbreaks to date in freshwater farmed fish have resulted from infection by highly adapted strains of GBS with genomes that are 10 to 15% smaller than their terrestrial conspecifics (13). Unusually, *S. agalactiae* also causes significant mortality in wild aquatic animals, including grouper, stingrays, and mullet (5), suggesting further adaptation to marine and freshwater aquatic hosts.

Microevolution within a bacterial species can be driven by host or environmental adaptation (13, 14), permitting an inference of the epidemiology of disease outbreaks and how pathogens may have transferred within and between geographic regions (14–16). This requires an analysis of factors that evolve at a sufficiently rapid pace to be informative over relatively short timespans. In GBS, capsular serotyping either with antibodies or by molecular serotyping (i.e., sequencing of the capsular operon) has become a widely used method of typing for population studies (17–20), and currently, *S. agalactiae* can be divided into 10 capsular serotypes (Ia, Ib, and II to IX) (18, 21). Determining capsular serotypes is also critical for vaccine formulation, since capsular polysaccharide (CPS) is highly immunogenic, and antibodies against CPS can confer excellent protection against infections by the homologous CPS serotype (20, 22, 23). Further typing resolution is provided by multilocus sequence typing (MLST), a method that has been employed to great effect to conduct global population studies of isolates based on genetic variations among relatively slowly evolving housekeeping genes (17). Combining molecular serotyping and MLST in the analysis of *S. agalactiae* revealed that the majority of isolates associated with aquatic environments and hosts fall within serotypes Ia and Ib, in which Ia isolates belong to sequence type 7 (ST-7) in clonal complex 7 (CC-7) and ST-103 in CC-103 (12, 18, 24–28). Serotype Ib strains isolated in Central and South America are ST-260 and ST-552 in CC-552 (27, 29), and strains isolated in Australia, Israel, Belgium, China, Ghana, the United States, and Southeast Asia belong to ST-261 (5, 6, 9, 13, 29–31). Serotype III is commonly causative of disease in humans but has also been isolated from fish in Thailand, China, and recently, Brazil (26, 32–34).

While capsular serotyping and MLST have been useful in inferring the origins and dispersal of GBS subtypes, they do not display sufficient resolution to explore the spread and evolution within individual sequence types, nor can they reflect the complete genetic diversity of *S. agalactiae* (35). The rapid fall in the cost of whole-genome sequencing coupled with multiplexing and rapid development of open-source bioinformatics tools has permitted a much deeper analysis of evolution, host adaptation, and epidemiological modeling within a single bacterial species (36), including those from aquatic hosts (14, 15). Bacteria, such as *S. agalactiae*, that can colonize multiple host species often have greater genomic intraspecies diversity (37). In GBS, two major evolutionary trends have been implicated in rapid adaptation to new hosts, namely, the acquisition of new genes by lateral gene transfer and genome reduction via gene loss integral to host specialization (13, 38). For example, *S. agalactiae* Ia strains GD201008-001 and ZQ0910, isolated from tilapia in China, carry a 10-kb genomic island (GI) which is absent from their closely related human isolate A909. Moreover, this 10-kb GI bears many similarities with the *Streptococcus anginosus* SK52/DSM 20563 genome sequence, suggesting possible transfer from *S. anginosus* to GBS, with implications for virulence in tilapia (13, 39). During fish host adaptation, serotype Ib strains have undergone reductive evolution, resulting in 10 to 25% of their genomes being lost compared to terrestrial *S. agalactiae* isolates and serotype Ia piscine strains (13). The

evolution of *S. agalactiae* by genome reduction is an ongoing process, with a high number of pseudogenes present in GBS genomes from aquatic sources (13).

The evolution of *S. agalactiae* and adaptation to aquatic hosts is an incomplete and ongoing process; consequently, sequencing the genomes from a few isolates is insufficient to understand the full potential genetic diversity of *S. agalactiae* as a species (35). The pangenome or supragenome of a bacterial species defines the full complement of genes, or the union of all the gene sets, within the species (35). This pangenome is subdivided into its core genome, which includes all the genes that are present in all the strains of the same bacterial species and must therefore be responsible for essential biological functions to allow the species to survive, and the accessory genome, containing species-specific genes that are unique to single strains or constrained to a cohort of strains within the species; these genes contribute to the diversity makeup of the species. The pangenome of a species resolves the true genomic diversity of that species and permits the identification of gene cohorts that are essential to the species as a whole, along with gene complements in the accessory genome that permit host or habitat specialization (35). Moreover, by identifying potential antigens within the pangenome that are conserved across all strains that infect a particular host type, vaccine targets can be specified that are likely to cross-protect (35, 40). Indeed, the first multicomponent protein-containing universal vaccine against human *S. agalactiae* was developed using a pangenome reverse-vaccinology approach by analyzing eight human isolates to predict putative antigens that were conserved among those strains (41). Some antigens in this vaccine are in the accessory genome; consequently, it is important to analyze as much as possible of the dispensable genome for vaccine development (35, 41).

The *S. agalactiae* pangenome is now well advanced but still "open" (i.e., new genes continue to be added with more sequenced genomes) and geographically constrained (35, 40). In the present study, we sequenced the genomes of new aquatic *S. agalactiae* strains isolated from tilapia in Honduras and from wild and captive marine fish in Australia. We infer the potential epidemiological distribution of ST-261 in aquatic hosts in Australia and show continuing adaptation to saltwater fish. Moreover, we identify conserved surface proteins across ST-260 and ST-261 that may have potential for incorporation into vaccines for aquaculture of important food fish species, such as tilapia and grouper.

RESULTS AND DISCUSSION

GBS isolates from marine fish and rays in Queensland and tilapia in Honduras belong to differing host-adapted lineages. The average size of the draft genomes of isolates from fish and rays in Queensland was 1,801,022 bp, containing 1,881 genes on average, while the mean genome size of strains from Honduran tilapia was 1,801,133 bp, with an average of 1,869 genes, consistent with the small genomes associated with the host-adapted aquatic lineage (13) (Tables 1 and 2). Molecular serotyping indicated that all Queensland and Honduran aquatic strains belong to serotype Ib, but Queensland isolates belong to sequence type 261 (ST-261), while Honduran strains belong to ST-260. Both ST-260 and ST-261 have previously been identified as infecting aquatic animals. ST-260 isolates have been isolated from tilapia in Brazil and Costa Rica and occupy clonal complex 552 (CC552), along with ST-552 and ST-553 strains also isolated from tilapia in Latin America (27), ST-261 has been isolated from tilapia in the United States, China, Ghana, and Israel (9, 13, 30).

The aquatic isolates analyzed herein extend the known host range of GBS ST-261 to rays and marine finfish and expand the environmental distribution from freshwater to marine habitats, in addition to those previously reported (5, 6, 42). GBS strains isolated from humans and terrestrial animals in Queensland and Northern Territory, Australia, were also sequenced and have larger genomes of 2,072,596 bp, comprising 2,067 genes on average, suggesting that recent possible local transfer from terrestrial origin to Australian fish is highly unlikely, although probable transfer between humans and fish has been reported for ST-7 GBS elsewhere (8, 39, 43). Nonhuman mammalian strains

TABLE 1 *S. agalactiae* isolates and sequences used in this study

Isolate	Origin[a]	Yr	Host	Serotype	ST[b]	Accession no.	Assembly level
QMA0264	QLD, AU	2008	*Epinephelus lanceolatus*	Ib	261	QGSK00000000	Contig
QMA0266	QLD, AU	2008	*E. lanceolatus*	Ib	261	QGSJ00000000	Contig
QMA0267	QLD, AU	2008	*E. lanceolatus*	Ib	261	QGSI00000000	Contig
QMA0268	QLD, AU	2009	*Pomadasys kaakan*	Ib	261	QGSH00000000	Contig
QMA0271	QLD, AU	2009	*Arius thalassinus*	Ib	261	CP029632	Complete
QMA0273	QLD, AU	2009	*A. thalassinus*	Ib	261	QGSG00000000	Contig
QMA0274	QLD, AU	2009	*Liza vaigiensis*	Ib	261	QGSF00000000	Contig
QMA0275	QLD, AU	2009	*Aptychotrema rostrata*	Ib	261	QGSE00000000	Contig
QMA0276	QLD, AU	2009	*Himantura granulata*	Ib	261	QGSD00000000	Contig
QMA0277	QLD, AU	2009	*Dasyatis fluviorum*	Ib	261	QGSC00000000	Contig
QMA0280	QLD, AU	2010	*E. lanceolatus*	Ib	261	QGSB00000000	Contig
QMA0281	QLD, AU	2010	*E. lanceolatus*	Ib	261	QGSA00000000	Contig
QMA0284	QLD, AU	2010	*E. lanceolatus*	Ib	261	QGRZ00000000	Contig
QMA0285	QLD, AU	2010	*E. lanceolatus*	Ib	261	QGRY00000000	Contig
QMA0287	QLD, AU	2010	*P. kaakan*	Ib	261	QGRX00000000	Contig
QMA0290	QLD, AU	2010	*A. thalassinus*	Ib	261	QGRW00000000	Contig
QMA0292	QLD, AU	2010	*A. rostrata*	Ib	261	QGRV00000000	Contig
QMA0294	QLD, AU	2010	*E. lanceolatus*	Ib	261	QGRU00000000	Contig
QMA0320	QLD, AU	2010	*Dasyatis fluviorum*	Ib	261	QGRT00000000	Contig
QMA0321	QLD, AU	2010	*D. fluviorum*	Ib	261	QGRS00000000	Contig
QMA0323	QLD, AU	2010	*D. fluviorum*	Ib	261	QGRR00000000	Contig
QMA0326	QLD, AU	2010	*D. fluviorum*	Ib	261	QGRQ00000000	Contig
QMA0347	QLD, AU	2010	*D. fluviorum*	Ib	261	QGRT00000000	Contig
QMA0368	QLD, AU	2010	*E. lanceolatus*	Ib	261	QGRO00000000	Contig
QMA0369	QLD, AU	2011	*E. lanceolatus*	Ib	261	QGRN00000000	Contig
QMA0485	Honduras	2014	*Oreochromis niloticus*	Ib	260	QGRG00000000	Contig
QMA0487	Honduras	2014	*O. niloticus*	Ib	260	QGRF00000000	Contig
QMA0488	Honduras	2014	*O. niloticus*	Ib	260	QGRE00000000	Contig
QMA0489	Honduras	2014	*O. niloticus*	Ib	260	QGRD00000000	Contig
QMA0494	Honduras	2014	*O. niloticus*	Ib	260	QHHT00000000	Contig
QMA0495	Honduras	2014	*O. niloticus*	Ib	260	QGRC00000000	Contig
QMA0496	Honduras	2014	*O. niloticus*	Ib	260	QGRB00000000	Contig
QMA0497	Honduras	2014	*O. niloticus*	Ib	260	QGRA00000000	Contig
QMA0499	Honduras	2014	*O. niloticus*	Ib	260	QGQZ00000000	Contig
QMA0355	QLD, AU	2011	*Homo sapiens*	Ia	23	QGRM00000000	Contig
QMA0357	QLD, AU	2011	*H. sapiens*	Ia	23	QGRL00000000	Contig
QMA0336	NT, AU	2005	*Crocodylus porosus*	Ia	23	QGRK00000000	Contig
QMA0300	QLD, AU	2008	*Canis lapis familiaris*	V	1	QGRJ00000000	Contig
QMA0303	QLD, AU	2009	*Felis catus*	V	1	QGRI00000000	Contig
QMA0306	QLD, AU	2005	*Bos taurus*	V	1	QGRH00000000	Contig
GS16-0008	Ghana	2016	*O. niloticus*	Ib	261	SRX2698682	Contig
GS16-0031	Ghana	2016	*O. niloticus*	Ib	261	SRX2698681	Contig
GS16-0035	Ghana	2016	*O. niloticus*	Ib	261	SRX2698680	Contig
GS16-0046	Ghana	2016	*O. niloticus*	Ib	261	SRX2698679	Contig
ND2-22	Israel	1988	*O. niloticus*	Ib	261	FO393392	Complete
138P	USA	2007	*O. niloticus*	Ib	261	CP007482	Complete
138spar	USA	2011	*O. niloticus*	Ib	261	CP007565.1	Complete
GX026	China	2011	*O. niloticus*	Ib	261	CP011328.1	Complete
S13	Brazil	2015	*O. niloticus*	Ib	552	CP018623.1	Complete
S25	Brazil	2015	*O. niloticus*	Ib	552	CP015976.1	Complete
SA20	Brazil		*O. niloticus*	Ib	552[c]	CP003919.2	Complete
GD201008-001	China	2010	*O. niloticus*	Ia	7	NC_018646.1	Complete
HN016	China	2011	*O. niloticus*	Ia	7	NZ_CP011325.1	Complete
WC1535	China	2015	*O. niloticus*	Ia	7	NZ_CP016501.1	Complete
A909	USA		*H. sapiens*	Ia	7	NC_007432.1	Complete
GBS85147	Brazil		*H. sapiens*	Ia	103	NZ_CP010319.1	Complete
Sag37	China	2014	*H. sapiens*	Ib[d]	12	NZ_CP019978.1	Complete
GBS1-NY	USA	2012	*H. sapiens*	II	22	NZ_CP007570.1	Complete
GBS2_NM	USA	2012	*H. sapiens*	II	22	NZ_CP007571.1	Complete
GBS6	USA	2009	*H. sapiens*	II	22	NZ_CP007572.1	Complete
FDAARGOS_254	USA	2014	*H. sapiens*	II[d]	22	CP020449.1	Complete
COH1	USA		*H. sapiens*	III	17	NZ_HG939456.1	Complete
NEM316	France		*H. sapiens*	III	23	NC_004368.1	Complete
CU_GBS_08	Hong Kong	2008	*H. sapiens*	III	283	NZ_CP010874.1	Complete
CU_GBS_98	Hong Kong	1998	*H. sapiens*	III	283	NZ_CP010875.1	Complete

(Continued on next page)

TABLE 1 (Continued)

Isolate	Origin[a]	Yr	Host	Serotype	ST[b]	Accession no.	Assembly level
NGBS128	Canada	2010	*H. sapiens*	III	17	NZ_CP012480.1	Complete
SG-M1	Singapore	2015	*H. sapiens*	III	283	NZ_CP012419.2	Complete
H002	China	2011	*H. sapiens*	III	736	NZ_CP011329.1	Complete
Sag158	China	2014	*H. sapiens*	III[d]	19	NZ_CP019979.1	Complete
NGBS061	Canada	2010	*H. sapiens*	IV	459	NZ_CP007631.1	Complete
NGBS572	Canada	2012	*H. sapiens*	IV	452	NZ_CP007632.1	Complete
2603V/R	USA		*H. sapiens*	V	110	NC_004116.1	Complete
CNCTC10/84	USA		*H. sapiens*	V	26	NZ_CP006910.1	Complete
NGBS357	Canada	2011	*H. sapiens*	V	1	NZ_CP012503.1	Complete
SS1	USA	1992	*H. sapiens*	V	1	NZ_CP010867.1	Complete
GBS-M002	China	2014	*H. sapiens*	VI[d]	1	NZ_CP013908.1	Complete
SA111	Portugal	2013	*H. sapiens*	II[d]	61[e]	NZ_LT545678.1	Complete
FWL1402	China	2014	*Hoplobatrachus chinensis*	III[d]	739[e]	NZ_CP016391.1	Complete
09mas018883	Sweden		*B. taurus*	V[d]	1	NC_021485.1	Complete
GBS ST-1	USA	2015	*C. lapis familiaris*	V	1	NZ_CP013202.1	Complete
ILRI005	Kenya		*Camelus dromedarius*	V[d]	609	NC_021486.1	Complete
ILRI112	Kenya		*C. dromedarius*	VI[d]	617	HF952106.1	Complete

[a]QLD, Queensland; AU, Australia; NT, Northern Territory.
[b]ST, sequence type.
[c]Gap in *glcK*.
[d]Serotype was detected with Kaptive.
[e]ST was determined via the Center for Genomic Epidemiology.

from Australia sequenced in this study, QMA0300 and QMA0303, belong to ST-1 serotype V, and QMA0306 belongs to ST-67 serotype III, whereas human isolates QMA0355 and QMA0357 and crocodile strain QMA0336 belong to ST-23 serotype Ia. Indeed, the high sequence identity between the human ST-23 serotype Ia isolates and those from farm-raised crocodiles supports the idea of probable human transfer to these animals, as previously implied (44).

To identify the possible origin of the ST-261 isolates from marine fish in Australia, a phylogenetic tree was constructed by maximum likelihood from 29,689 nonrecombinant core-genome SNPs and short indel positions derived by the alignment of whole-genome sequences of 25 Queensland fish isolates, 9 Honduran isolates, 6 Queensland terrestrial isolates, and 42 genomes from public databases (Fig. 1A). Two distinct groups were resolved, the first entirely composed of aquatic isolates (including serotype Ib isolates from ST-552, ST-260, and ST-261), while the second major group comprised various isolates of terrestrial origin and some fish isolates from ST-7 serotype Ia that may have infected fish via transfer from terrestrial sources (Fig. 1A). The serotype Ib aquatic group branched into three distinct lineages based on ST (Fig. 1A). One lineage comprised all Honduran strains of ST-260, which were derived from a lineage comprising ST-552 isolates from Brazil (Fig. 1A). This is supportive of previous observations in which an extended typing system based on MLST, virulence genes, and serotype indicated geographic endemism within fish isolates from differing regions of Brazil (27). The isolates belonging to ST-261 from the United States, Israel, Ghana, China, and Queensland clustered together (Fig. 1A). The second major division, containing serotype Ia fish strains along with terrestrial isolates, was more complex, but the isolates largely clustered in line with serotype and ST (Fig. 1A). The aquatic serotype Ia isolates clustered with human isolates of ST-7 (Fig. 1A). These fish isolates have acquired a 10-kb genomic island, putatively from *S. anginosus*, in contrast to their human ST7 serotype Ia relatives (39). Lineage 10 comprised serotype III ST-17 human isolates (Fig. 1A). Serotype Ia strains were divided among several additional ST groups, where three Australian serotype Ia ST-23 strains (QMA0336, QMA0355, and QMA0357) isolated from humans and a saltwater crocodile clustered together with strains of serotype III ST-23 (NEM316) and serotype IV ST-452 (NGBS572), also of human origin (Fig. 1A). This lineage appears to be derived from NEM316, which is a frequent cause of late-onset disease in human infants (45). A further serotype Ia isolate was located in ST-103 and clustered together with two strains of serotype V from ST-609 and ST-617 isolated from camels

TABLE 2 Genome assembly statistics

Strain	BioSample no.	Sequence yield (bp)	No. of contigs	Genome size (bp)	GC content (%)	N_{50} (bp)	Coverage (×)
QMA0264	SAMN07998566	383,903,016	26	1,800,255	35.27	181,173	213
QMA0266	SAMN07998567	383,903,016	23	1,799,604	35.26	181,173	213
QMA0267	SAMN07998568	822,647,364	22	1,805,121	35.33	181,785	456
QMA0268	SAMN07998569	541,821,504	22	1,796,834	35.25	130,961	302
QMA0271	SAMN07998570	488,654,460	1	1,802,470	35.28	1,802,470	271
QMA0273	SAMN07998571	488,596,752	26	1,799,223	35.26	181,173	272
QMA0274	SAMN07998572	666,932,868	33	1,795,949	35.23	91,064	371
QMA0275	SAMN07998573	444,271,968	28	1,798,598	35.26	129,867	247
QMA0276	SAMN07998574	595,443,828	30	1,801,706	35.28	101,410	330
QMA0277	SAMN07998575	532,021,392	30	1,794,975	35.23	125,723	296
QMA0280	SAMN07998576	507,086,244	26	1,798,991	35.26	130,981	282
QMA0281	SAMN07998577	947,517,396	28	1,795,694	35.26	130,961	528
QMA0284	SAMN07998578	870,597,252	27	1,804,787	35.3	129,871	482
QMA0285	SAMN07998579	946,715,952	25	1,798,637	35.26	181,149	526
QMA0287	SAMN07998580	744,380,280	27	1,799,347	35.27	130,969	414
QMA0290	SAMN07998581	1,059,786,168	24	1,804,000	35.3	181,153	587
QMA0292	SAMN07998582	383,835,312	22	1,800,208	35.26	181,172	213
QMA0294	SAMN07998583	807,742,572	24	1,800,272	35.26	181,172	449
QMA0300	SAMN07998584	703,661,784	33	2,109,161	35.25	115,466	334
QMA0303	SAMN07998585	949,167,156	42	2,123,226	35.05	94,602	447
QMA0306	SAMN07998586	745,047,660	106	2,165,975	35.52	40,135	344
QMA0320	SAMN07998587	445,338,936	30	1,804,060	35.3	181,347	247
QMA0321	SAMN07998588	406,041,552	22	1,800,073	35.26	181,173	226
QMA0323	SAMN07998589	487,351,368	22	1,800,031	35.26	181,173	271
QMA0326	SAMN07998590	847,046,508	22	1,800,125	35.26	181,173	471
QMA0336	SAMN07998591	437,035,536	37	2,022,886	35.26	87,595	216
QMA0347	SAMN07998592	942,292,680	21	1,795,434	35.23	181,173	525
QMA0355	SAMN07998593	903,514,248	66	1,996,200	35.31	63,524	453
QMA0357	SAMN07998594	564,675,720	35	2,004,249	35.23	92,563	282
QMA0368	SAMN07998595	506,471,448	26	1,803,560	35.3	104,282	281
QMA0369	SAMN07998596	1,056,496,560	28	1,804,404	35.3	130,985	586
QMA0485	SAMN07998597	490,192,752	28	1,803,278	35.27	109,741	272
QMA0487	SAMN07998598	467,765,088	28	1,800,682	35.25	109,741	260
QMA0488	SAMN07998599	356,791,512	29	1,800,616	35.25	109,741	198
QMA0489	SAMN07998600	432,449,472	28	1,801,015	35.25	109,741	240
QMA0494	SAMN07998601	363,434,736	37	1,792,992	35.28	71,935	203
QMA0495	SAMN07998602	412,230,672	28	1,803,007	35.27	109,741	229
QMA0496	SAMN07998603	499,707,096	29	1,801,879	35.27	109,737	277
QMA0497	SAMN07998604	364,009,464	28	1,800,976	35.25	109,741	202
QMA0499	SAMN07998605	428,636,208	28	1,799,426	35.24	109,741	238

(46) (Fig. 1A). Two newly sequenced Australian terrestrial strains, QMA0300 isolated from a dog and QMA0303 isolated from a cat, clustered with other serotype V ST-1 strains isolated from human, cattle, and dog hosts (Fig. 1A). This lineage also contained NGBS061 serotype IV ST-495 and GBS-M002 serotype VI ST-1 (Fig. 1A). ST-1 emerged as a significant cause of infection and disease in humans during the 1990s but was recently inferred to have evolved from strains causing mastitis in cattle in the 1970s (47). Moreover, QMA0306 ST-1 serotype V from cattle in Queensland was closely related to SA111 ST-61 serotype II, which represents a host-adapted lineage of *S. agalactiae* that is dominant in cattle in Europe (48) (Fig. 1A).

Our phylogeny based on whole-genome SNPs does not support a recent direct transfer of GBS from Australian human clinical or terrestrial animal sources to marine fish and stingrays, in spite of close proximity of many of the wild fish cases to human habitation (6). Consequently, we refined our analyses to the ST-261 aquatic host-adapted lineage to attempt to infer a possible route of introduction and subsequent evolution in Australian marine fish. A consensus minimum spanning tree based on a distance matrix composed of all core-genome SNPs derived from alignment of the ST-261 serotype Ib strains revealed a likely original introduction via tilapia from Israel, with only 63 core-genome SNPs separating an Australian stingray isolate from the type strain of *Streptococcus difficile* (reassigned as *S. agalactiae* serotype Ib [49]), isolated

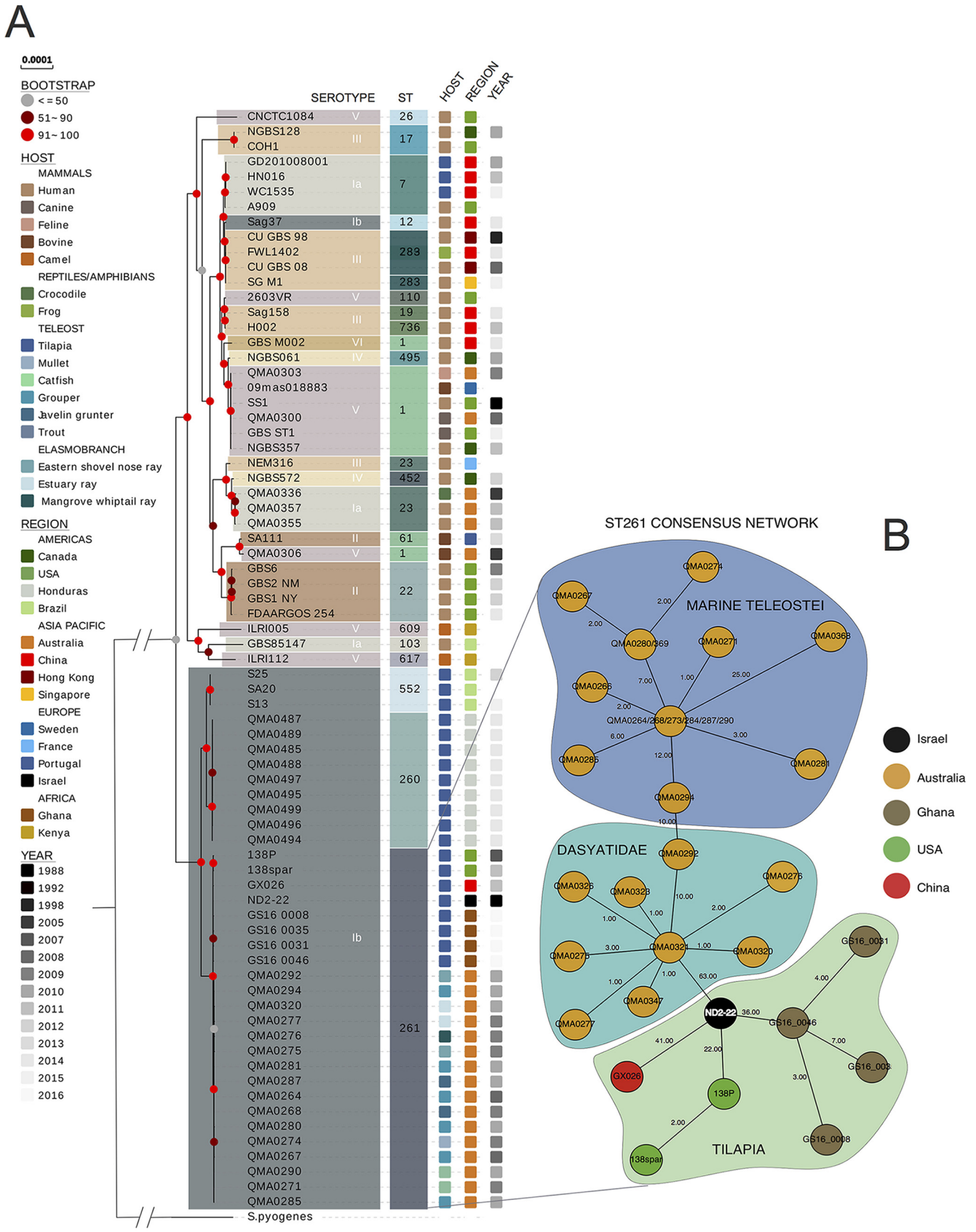

FIG 1 (A) Maximum likelihood phylogeny of 82 *S. agalactiae* strains. The tree was inferred from alignment of 6,050 nonrecombinant core-genome SNPs. Branch length was adjusted for ascertainment bias using Felsenstein's correction implemented in RAxML (76). Nodes are supported by 1,000 bootstrap replicates. The

(Continued on next page)

from tilapia in Israel in 1988 (Fig. 1B) (50). Tilapia were imported on a number of occasions during the 1970s and 1980s from Israel into North Queensland around Cairns and Townsville, and a number of strains and hybrids have since colonized rivers and creeks throughout Queensland (51). Globally, the aquatic ST-261 lineage appears to have been transferred through human movements of tilapia for aquaculture and other purposes over the last several decades. The U.S. and Chinese tilapia isolates also appear to derive from the early ND2-22 isolate, as do recent isolates from tilapia in Ghana (Fig. 1B). Indeed, phylogenetic analysis by maximum likelihood of draft whole-genome alignments suggests that the Ghanaian and Chinese isolates share a recent common ancestor that is derived from ND2-22, with only 60 SNPs separating ND2-22 and the Ghanaian strains (9). We identified only 36 core-genome SNPs separating the Ghanaian isolates from ND2-22, but this reflects the smaller core genome in our study as a result of the high number of GBS isolates analyzed (40 isolates in the present study compared with 9 isolates in the previous study) (9).

The minimum spanning tree implicates continued adaptation of the ST-261 lineage postintroduction and suggests that grouper (marine Teleostei family) may have been infected via estuary stingrays (Dasyatidae family) (Fig. 1B). Stingrays are occasional prey for adult giant Queensland grouper, and stingray barbs have been found in the gut of grouper postmortem (R. O. Bowater, unpublished data). ST-261 GBS has also caused mortality in captive stingrays in South East Queensland, translocated from Cairns in North Queensland (42).

The *S. agalactiae* pangenome comprises a small core of protein-coding genes and is open. To further elucidate adaptation among the fish-pathogenic GBS types, a pangenome was built from 39 complete genomes retrieved from GenBank and using our curated ST-261 grouper isolate QMA0271 as a high-quality reference seed genome. The resulting pangenome was 4,074,275 bp (Fig. 2A). All-versus-all BLAST analysis of the genomes implemented in BRIG clearly indicated the substantial reduction in genome size among the fish-pathogenic ST-261 cohort, as previously reported for a limited number of ST-261 isolates (13). Here, we find that ST-260 and ST-552 fish-pathogenic sequence types within serotype Ib are similarly reduced and that conservation among the serotype Ib strains is high (Fig. 2A). In total, 4,603 protein-coding genes were predicted in the *S. agalactiae* pangenome using Roary (Fig. 2B), which is consistent with previous research (39). The number of core genes was 1,440 (representing 35% of the pangenome), while previous studies reported 1,202 to 1,267 genes in the pangenome (39, 40). These differences may result from the methods being used to examine the pangenome, the difference in number of sequences being used to create the pangenome, and finally, the use of draft sequences in the analysis (39). A majority of protein-coding genes found in the pangenome belonged to both the dispensable and strain-specific genes. This could be a result of the inclusion of a high number of serotype Ib strain sequences, which were all significantly smaller (approximately 1.8 Mbp) than those of other isolates. Liu et al. (39) demonstrated that removing Ib piscine isolates from their analysis resulted in an increase in the number of core genes.

Frequency analysis of IGRs in *S. agalactiae* pangenome showed that the number of IGRs shared across all strains was smaller than core protein-coding genes, whereas strain-specific IGRs were much higher than protein-coding strain-specific genes (Fig. 2C). IGR analysis with Piggy excludes IGRs that are less than 30 bp, which may result in fewer IGRs than protein-coding genes in core regions (52). Most IGRs identified in the pangenome belonged to either core genes or strain-specific genes (Fig. 2C), in line with previous findings in *Staphylococcus aureus* ST-22 and *Escherichia coli* ST-131, where similar distributions of IGRs were detected (52). The gradients of the accumulation curves for

FIG 1 Legend (Continued)
tree was rooted using *S. pyogenes* M1 GAS (RefSeq accession number NC_002737.2) as an outgroup. (B) Minimum spanning tree showing relationship among ST-261 serotype Ib GBS isolates based on a distance matrix derived from nonrecombinant core-genome SNP alignment. Edge labels indicate the number of nonrecombinant core-genome SNPs between adjacent strains. The consensus network was computed in MSTgold (78).

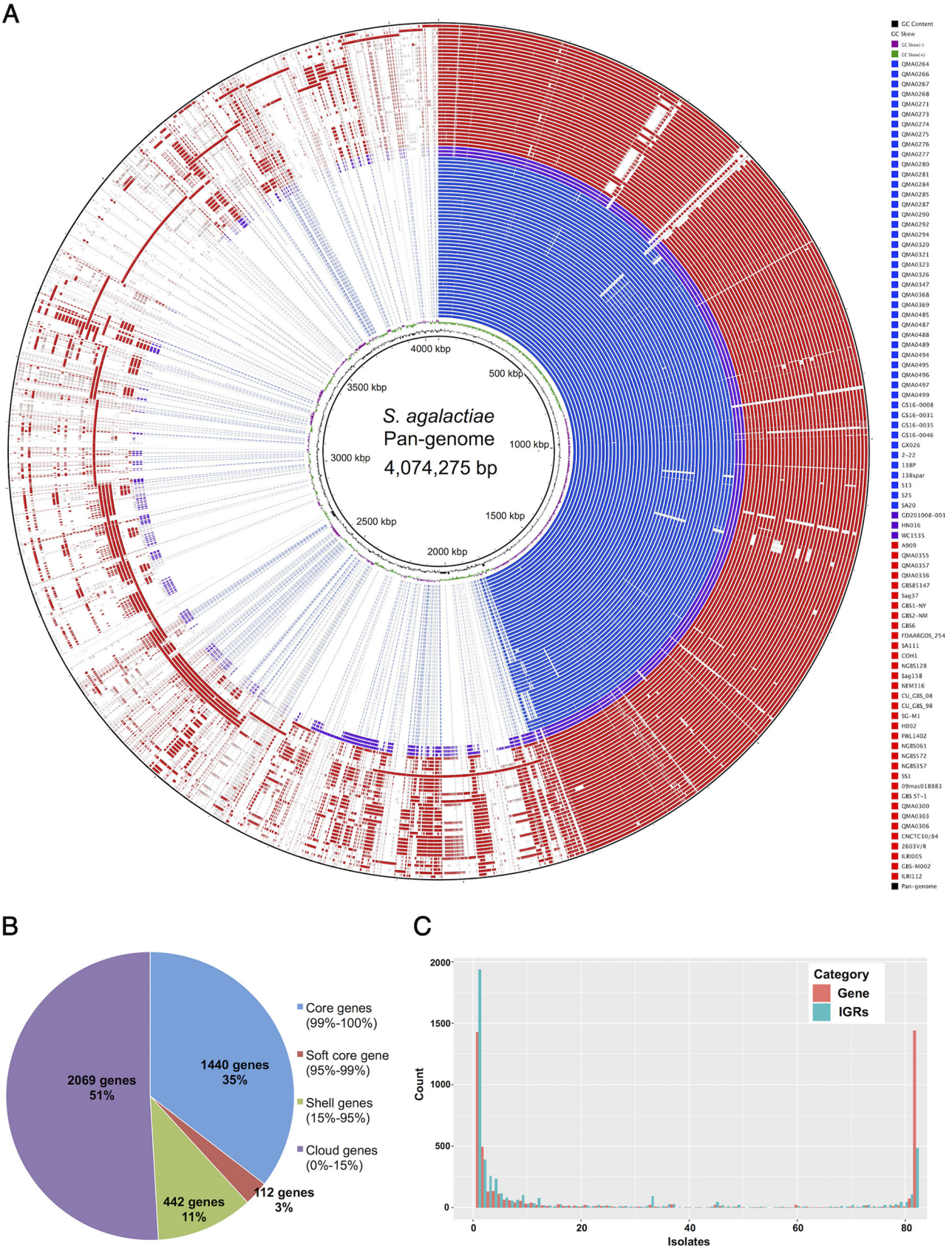

FIG 2 The pangenome of *S. agalactiae*. (A) BLASTN-based sequence comparison of 82 *S. agalactiae* genomes against the *S. agalactiae* pangenome as reference constructed with BRIG 0.95 (80). Rings from the innermost to the outermost show GC content and GC skew of the pangenome reference, and

(Continued on next page)

both protein-coding genes and IGRs are still strongly positive; therefore, the *S. agalactiae* pangenome remains open.

To determine whether strain-specific IGR sequences may be derived from genes (or pseudogenes) that have undergone erosion, we queried the gene sequences used to construct the pangenome by BLASTn using 1,539 strain-specific representative IGR sequence clusters (output from Piggy). A total of 1,862 hits were retrieved from 683/1,539 representative IGR sequence clusters (see Table S2 in the supplemental material) supportive of the gene origins of many IGRs. Overrepresentation of sequences from the QMA0271 reference genome among the hits (309/1,862 from 39 genomes employed in the pangenome) was expected, as it was employed as the seed for the pangenome assembly. Of the 241 unique hits among genes from QMA0271, 39 hits fell in genes that we annotated as pseudogenes (Table S2). Pseudogenes are predicted to be rapidly lost from the genome, evidenced by the lack of conserved pseudogenes across multiple strains of the same species (53, 54). The relatively high presence of pseudogenes or remnants thereof within the serotype Ib ST-261 QMA0271 is supportive of ongoing adaptation of this lineage to the aquatic host and habitat (13). As pseudogenes are often not annotated as such in database assemblies (54, 55), further manual inspection of the 849 unique BLAST hits was performed in Excel and by reference to the pangenome. Pseudogenes may arise through frameshift SNPs, resulting in early termination or interruption by insertion elements (55), and 239 of the 849 unique BLAST hits were identified as insertion element (IS) transposases and a further four as or phage/prophage proteins, suggesting possible gene disruptions by insertion (Table S2). Hypothetical proteins comprised 215 of the 849 unique sequences identified by BLAST of strain-specific IGR through the pangenome, while genes annotated as transcriptional regulators comprised 48 sequences (Table S2). These results lend preliminary support to the hypothesis that much of the IGR of bacteria comprises gene remnants and is worthy of in-depth exploration.

Aquatic serotype Ib isolates have a reduced repertoire of virulence factors. Almost all genes classed as adhesins were involved in immune evasion and host invasion, and most of the toxin-related genes found in terrestrial isolates were absent from serotype Ib aquatic isolates (Fig. 3). Rosinski-Chupin et al. (13) reported that ~60% of the virulence genes found in human strains were present in a serotype Ib GBS strain from fish. We found that the CAMP factor gene *cfa-cfb* was present in all strains, including serotype Ib ST-260, ST-261, and ST-552 isolates, but the CAMP reaction was previously reported to be negative for ST-260 and ST-261 strains (13, 28). These authors identified that the CAMP factor gene in ST-261 is disrupted, while the gene in ST-260 is unaltered, but the level of gene expression may be too low to be detected by the test (13). Most of the genes in the *cyl* locus have been lost from serotype Ib strains, and only *cylB*, an ABC ATP binding cassette transporter, was present in ST-260 and ST-552 (Fig. 3). The *cyl* locus is responsible for hemolytic activity and the production of pigment via cotranscription of *cylF* and *cylL* (56). In ST-261, the *cyl* gene cluster is replaced by a genomic island, resulting in a loss of hemolytic activity (13). Of particular relevance to virulence and antigenic diversity, the transmembrane immunoglobulin A-binding C protein beta-antigen gene *cba* was absent in all aquatic isolates (Fig. 3). This gene has been reported in type Ib and Ia GBS strains previously (57), is implicated in virulence in neonates, and is upregulated in the GBS serotype Ia strain A909 in response to human serum (57–59). In contrast, capsule-related genes were largely conserved and associated with serotype (Fig. 3), supporting the major role of capsule in virulence of fish-pathogenic streptococci (60, 61).

FIG 2 Legend (Continued)
then sequence similarities of each of the 82 strains listed in the legend; from top to bottom, rings are colored according to origin, with fish isolates belonging to serotype Ib ST-261, ST-260, and ST-552 in blue, fish strains belonging to serotype Ia in purple, and terrestrial strains in red. The outermost ring (black) represents the reference pangenome. (B) Proportion of protein-coding genes in the core, soft core, shell, and cloud of the pangenome of 82 *S. agalactiae* isolates determined with Roary. (C) Histogram indicates the frequency of genes (protein-coding) in red and IGRs (non-protein-coding intergenic regions) in blue-green from 82 *S. agalactiae* genomes analyzed by Piggy (52).

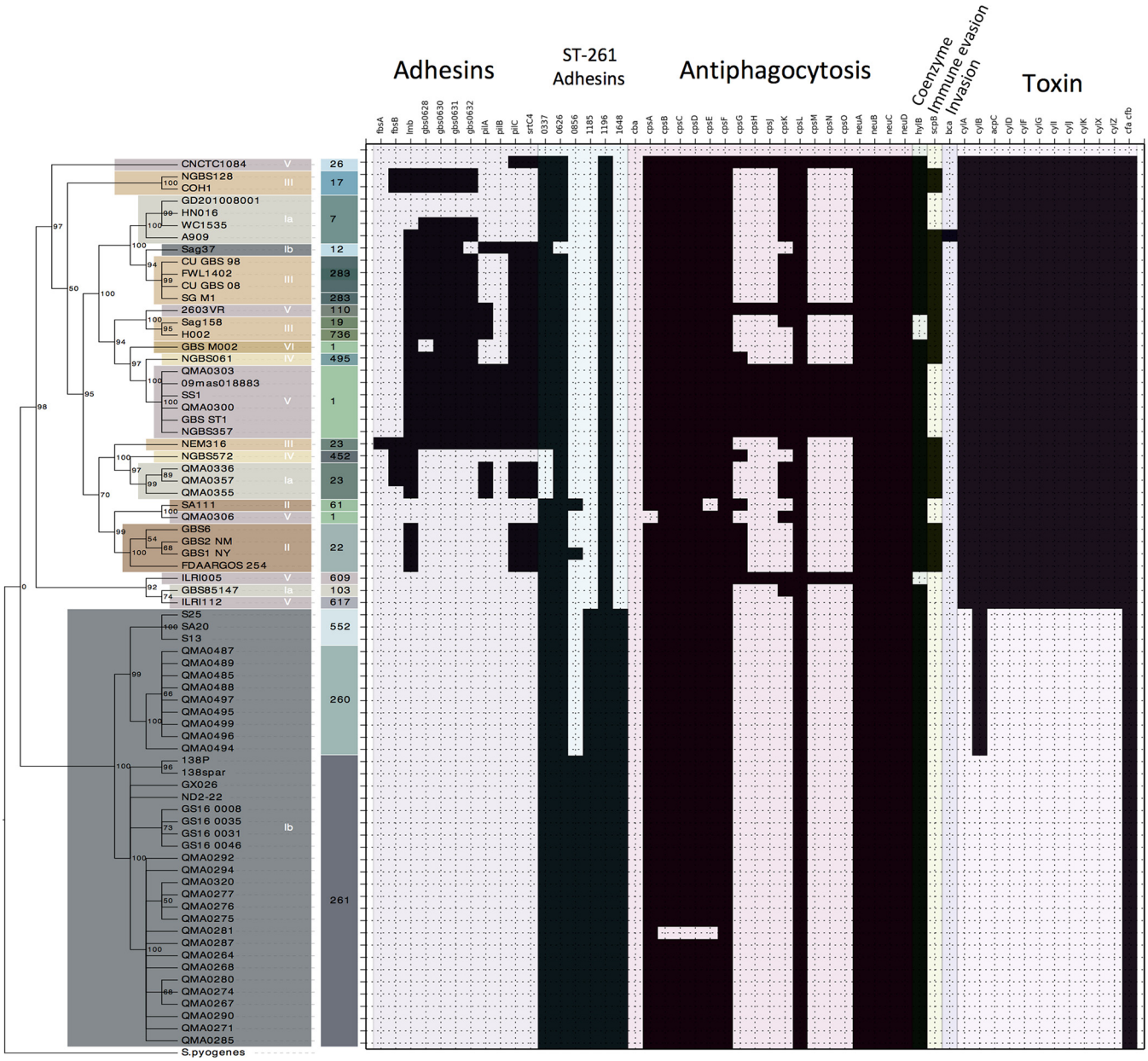

FIG 3 Virulence gene presence and absence in *S. agalactiae*. Genes were identified from VFDB to create an *S. agalactiae* database for comparison of 82 strains by BLAST using SeqFindr, with default settings (95% identity cutoff).

Although serotype Ib aquatic isolates have lost the majority of the virulence factors found in terrestrial isolates, most contained six sequences that were identified as probable adhesins by homology (Fig. 3). These ST-261 adhesins (named 0337, 0626, 0856, 1185, 1196, and 1648, based on positions in the annotated ST-261 genome from QMA0271) were fully conserved among aquatic serotype Ib ST-261 isolates (Fig. 3). Moreover, ST-261 adhesin 0856 was only present in ST-261 and two serotype II isolates, GBS1-NY and SA111 (Fig. 3). ST-261 adhesins 1185 and 1648 were shown to be unique to aquatic serotype Ib isolates, being absent from Ia fish isolates and all terrestrial strains (Fig. 3). The analysis indicated that ST-261 adhesins 0337, 0626, and 1196 were well conserved across most of the isolates, regardless of origin, but 1196 was the only adhesin-like gene present in all strains analyzed (Fig. 3). ST-261 adhesin 0626 was also ubiquitous but contained deletions in the Sag37 isolate (Fig. 3). Some terrestrial strains, such as QMA0355, QMA0357, QMA0336, and NGBS572, lacked adhesin 0337 (Fig. 3).

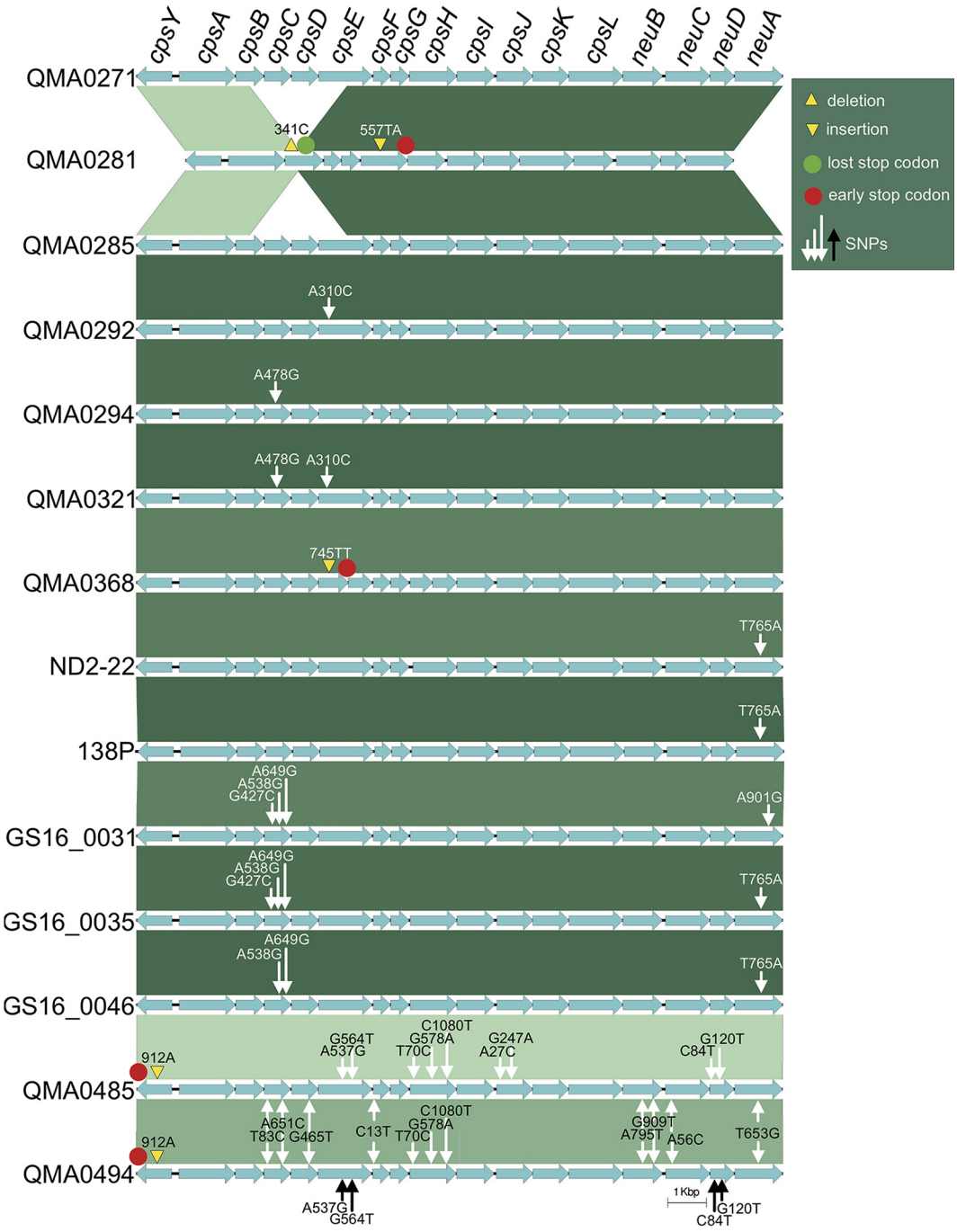

FIG 4 The capsular polysaccharide (*cps*) operon of the ST-261 lineage. The operon was identified in GenBank files manually and then compared by BLAST using Easyfig (91).

As capsular polysaccharide is a requirement for full virulence in several fish-pathogenic streptococci (60, 61) and is the major antigen in GBS (62–64), further analysis of the serotype Ib *cps* operon was conducted. Within the serotype Ib lineage, the capsular polysaccharide operon was well conserved (Fig. 4). However, QMA0281 from grouper had a deletion at position 341 in *cpsB*, encompassing *cpsC*, *cpsD*, and *cpsE* (Fig. 4), resulting in a chimeric open reading frame (ORF). Moreover, an insertion in a TA repeat at position 557 in *cpsH* resulted in an early stop codon, marginally reducing gene size (Fig. 4). QMA0368 had a TT insertion at nucleotide position 745 in *cpsE*, resulting in the insertion of an early stop codon and truncation of the gene (Fig. 4). Deletion of

this region that includes the priming glycosyltransferase for capsular biosynthesis results in a loss of capsule, attenuated virulence, and modified pathology in the fish pathogen *Streptococcus iniae* (65). Buoyant density analysis in Percoll indicated that QMA0281 is also deficient in capsule (not shown). *cps* operon SNPs among the Australian serotype Ib isolates were rare (Fig. 4). Indeed, only 1 SNP in *neuA* separated QMA0271 from the 1988 tilapia isolate from Israel, suggesting very little evolution of the *cps* operon since the introduction of the lineage (Fig. 4). This may reflect a well-adapted capsule for colonization-naive hosts, as the Australian isolates were from wild fish and captive stingrays recently after capture and transport, thus placing little selective pressure for novel capsular sequence types. Immunity in fish drives the evolution of novel capsular sequence types in *S. iniae*, but these reported cases were all in high-intensity aquaculture with occasional use of autogenous vaccination and an opportunity for the development of *cps*-specific immunity (65). This may explain the relatively high number of SNPs in genes encoding sugar- and sialic acid-modifying enzymes among the isolates from tilapia farmed in Honduras relative to the other isolates examined, as autogenous vaccination is occasionally used on farms where isolates were sourced and may apply selective pressure favoring modified polysaccharide capsule.

Conclusions. The clade of aquatic *S. agalactiae* serotype Ib, including ST-260, ST-261, and ST-552, is a highly adapted fish pathogen with a substantially reduced genome compared to all other serotypes from terrestrial mammalian, reptile, and fish hosts. These variants were originally identified as *S. difficile* due to their impoverished growth on laboratory media and hence difficulty of isolation from diseased fish (50). Isolates from fish that were previously identified as *S. difficile* were assigned to serotype Ib GBS a few years later (49), but the recent discovery that serotype Ib isolates have substantially smaller genomes than those of other GBS serotypes (13) explains the marked phenotypic difference that merited early phenotypic assignation of these strains to a separate species. Other serotypes have been isolated from fish, notably serotype Ia and serotype III, but these seem to arise from terrestrial transfer rather than being retained among the aquatic host population (28). In contrast, serotype Ib has only been isolated from fish and stingrays, appears to be well adapted, and is likely to have been transferred internationally via trade in domesticated tilapia, evidenced here by the very close relationship (only a few SNPs) between a strain isolated from tilapia in 1988 in Israel and those found in fish and stingrays in Australia since 2008 and in tilapia in the United States, China, and Ghana. The ST-261 lineage in Australia likely arrived with several introductions of tilapia in the 1970s and 1980s. Tilapia are classed as a noxious pest in Australia, and their import was banned, but not before several lines became established throughout tropical and subtropical freshwater habitats in Queensland (51). Although ST-261 serotype Ib GBS has not been isolated from farm fish in Queensland in spite of the proximity of freshwater farms to tilapia-infested creeks, this clade of GBS is a substantial problem in farmed tilapia globally. The cohort of putative adhesins identified here to be conserved throughout all fish-pathogenic serotypes (including Ia and III, in addition to Ib) may be promising candidates for cost-effective cross-serotype protective vaccines for aquaculture and are worthy of future research.

MATERIALS AND METHODS

Bacterial strains and culture conditions. Forty *S. agalactiae* isolates comprising strains collected from fish in Honduras and Australia, along with reptiles, humans, and other terrestrial mammals from Australia, were chosen for sequencing (Table 1). Of these 40 isolates, 25 strains were collected from several species of fish in Queensland, Australia, two human clinical strains were from Queensland, Australia, one strain was isolated from saltwater crocodile (*Crocodylus porosus*) in the Northern Territory, Australia, and three isolates were collected from domestic animals, including cats, dogs, and cattle in Australia. Additionally, nine isolates originating from disease in farmed tilapia in Honduras were sequenced during this study. All isolates were maintained at −80°C in Todd-Hewitt broth (THB; Oxoid) containing 25% glycerol as frozen stock. The isolates were recovered from stock on Columbia agar containing 5% sheep blood (Oxoid) for 24 h at 28°C. For liquid culture, the isolates were grown in THB for 18 h with low agitation at 28°C.

DNA extraction and sequencing. Genomic DNA (gDNA) was extracted from 10-ml early stationary-phase THB cultures with the DNeasy minikit (Qiagen), according to the manufacturer's instructions. The

TABLE 3 Reference sequences used for molecular capsule serotyping

Serotype	Accession no.	Size (bp)	Reference
Ia	AB028896.2	25,021	85
Ib	AB050723.1	9,987	86
II	EF990365.1	12,864	87
III	AF163833.1	17,276	88
IV	AF355776.1	17,596	89
V	AF349539.1	18,239	89
VI	AF337958.1	16,448	89
VII	AY376403.1	14,202	90
VIII	AY375363.1	12,637	90

quantity of extracted DNA was measured by Qubit fluorimetry (Invitrogen), and the quality was checked by agarose gel electrophoresis. To confirm the purity of the gDNA, the 16S rRNA gene was amplified by PCR using universal primers 27F and 1492R (66), and the PCR products were sent to Australian Genome Research Facility (AGRF, Brisbane, Australia) for Sanger sequencing. The 16S amplicon sequences were assembled in Sequencher version 5.2.2 and analyzed by BLAST. Once identity and purity were confirmed, Nextera XT paired-end libraries were generated using gDNA from each isolate and sequenced on the Illumina HiSeq 2000 platform system (AGRF, Melbourne, Australia).

***De novo* assembly and annotation.** Illumina sequencing yielded between 5,288,952 and 12,577,340 read pairs for each strain. Read quality control and contaminant screening were performed using FastQC (67). Reads were trimmed using the clip function in Nesoni (https://github.com/Victorian-Bioinformatics-Consortium/nesoni) and then assembled *de novo* with SPAdes assembler version 3.11 (68). Assemblies were quality checked using Quast version 4.6 (69). The assemblies of fish isolates in Queensland comprised about 1.8 Mbp of assembled sequence, while terrestrial strains comprised 2 Mbp. The assembled contigs for all Queensland strains were reordered against an internal curated reference genome from *S. agalactiae* strain QMA0271, using the Mauve contig ordering tool (70). Automated annotation was performed using Prokka 1.12 (71).

Molecular serotyping and MLST. Reference sequences for the nine CPS serotypes (Table 3) (21) were retrieved from GenBank to generate a database for the prediction of capsular serotype from the draft genomes with Kaptive, using default settings (72). To determine multilocus sequence types (MLST), all draft assemblies were analyzed using the Center for Genomic Epidemiology web-tools MLST version 1.8, using an *S. agalactiae* configuration (see https://cge.cbs.dtu.dk//services/MLST/) (73).

Phylogenetic analysis. To estimate approximate phylogenetic relationships among our strains and other isolates with whole-genome sequences available in GenBank (Table 1), a core-genome single nucleotide polymorphism (SNP)-based phylogenetic tree was constructed. Whole-genome sequences of Queensland and Honduras tilapia isolates, terrestrial isolates, and the genomes obtained from GenBank were aligned with Parsnp in the Harvest Tools suite version 1.2 (74). The genome of *Streptococcus pyogenes* M1 (group A *Streptococcus* [GAS]) (RefSeq accession number NC_002737.2) was also included as an outgroup for tree rooting. Hypothetical recombination sites in the core-genome alignment were detected and filtered out with Gubbins (75). Maximum likelihood phylogenetic trees were inferred with RAxML version 8.2.9 (76) based on nonrecombinant core-genome SNPs under a general time-reversible (GTR) nucleotide substitution model, with 1,000 bootstrap replicates. Ascertainment bias associated with analysis of only variable sites was accounted for using Felsenstein's correction implemented in RAxML (76). The resulting tree was exported and rooted, nodes with low bootstrap support were collapsed with Dendroscope, and the resulting tree/cladogram was annotated with Evolview version 2 (77).

In order to infer possible phylogenetic relationships within the aquatic ST-261 clade of GBS, minimum spanning trees were constructed in MSTgold 2.4 (78) from a distance matrix based on core-genome SNPs derived by the alignment of 27 ST-261 genomes in Geneious version 9.1 (Biomatters, Inc.). A consensus tree was constructed based on inference of 2,400 trees, and only those edges occurring in greater than 50% of trees were included in the consensus.

Pangenome analysis. A reference pangenome was built with the GView server (79) using our curated genome of the Queensland ST-261 serotype Ib grouper isolate QMA0271 as a seed and 38 complete genomes from public databases added sequentially (Table 1). Only complete genomes were included into the reference pangenome to avoid incomplete genes associated with the high fragmentation of draft sequences. For visualization, draft and complete genomes were compared with the reference pangenome using BRIG 0.95 (80). To investigate core and accessory protein-coding genes, sequences from all strains were analyzed with Roary (81) using default settings, and gene presence-absence tables are provided in Table S1. Since Roary only considers protein-coding sequences, we used Piggy (52) to identify non-protein-coding intergenic regions (IGRs) in each strain, which comprise about 15% of the GBS genomes. The 1,539 strain-specific IGR sequences identified in Piggy were then extracted from the representative merged clusters of IGR sequences and used as query sequences in a local BLASTn analysis (E value 1e−10) against all the coding sequences in the pangenome. The output from the BLAST search was tabulated and analyzed further in Microsoft Excel (Table S2).

Identification and comparison of virulence factors. Virulence factor screening was performed using SeqFindR (82) by comparing the assembled genomes of all strains in this study along with the genomes available through GenBank (Table 1) against a list of 51 *S. agalactiae*-specific virulence factor sequences collected from the Virulence Factors Database (VFDB) (83), complemented by six additional sequences identified in the sequenced ST-261 *S. agalactiae* strain ND2-22 (Table 1).

Analysis of effects of SNPs in ST-261 clade. Putative effects of SNPs among and between the ST-261 clade were determined using SnpEff version 4.3p (84). First, a new *S. agalactiae* database was constructed from the GenBank-formatted curated reference genome for QMA0271, in accordance with the manual (http://snpeff.sourceforge.net/SnpEff_manual.html#databases). Then, a VCF file generated from curated SNPs generated by an alignment of the complete genomes of 27 ST-261 isolates in Geneious version 9.1 was annotated for SNP effect with SnpEff using the database from QMA0271 as a reference (see the supplemental material).

Accession number(s). The accession numbers for genome assemblies used in this study are presented with strain metadata in Table 1. The assembly statistics and BioSample numbers are in Table 2.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AEM.00859-18.

SUPPLEMENTAL FILE 1, XLSX file, 1.8 MB.

SUPPLEMENTAL FILE 2, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS

This work was funded by the Fisheries Research and Development Corporation Aquatic Animal Health Subprogram Project 2010/34 awarded to R.O.B. and A.C.B., and the Australian Research Council Linkage Program LP130100242 with cosponsorship by Elanco Australia awarded to A.C.B., M.J.W., and S.B.

REFERENCES

1. Brochet M, Couvé E, Zouine M, Vallaeys T, Rusniok C, Lamy M-C, Buchrieser C, Trieu-Cuot P, Kunst F, Poyart C, Glaser P. 2006. Genomic diversity and evolution within the species Streptococcus agalactiae. Microbes Infect 8:1227–1243. https://doi.org/10.1016/j.micinf.2005.11.010.
2. Edwards MS, Gonik B. 2013. Preventing the broad spectrum of perinatal morbidity and mortality through group B streptococcal vaccination. Vaccine 31:D66–D71. https://doi.org/10.1016/j.vaccine.2012.11.046.
3. Six A, Bellais S, Bouaboud A, Fouet A, Gabriel C, Tazi A, Dramsi S, Trieu-Cuot P, Poyart C. 2015. Srr2, a multifaceted adhesin expressed by ST-17 hypervirulent group B Streptococcus involved in binding to both fibrinogen and plasminogen. Mol Microbiol 97:1209–1222. https://doi.org/10.1111/mmi.13097.
4. Doumith M, Mushtaq S, Martin V, Chaudhry A, Adkin R, Coelho J, Chalker V, MacGowan A, Woodford N, Livermore DM, BSAC Resistance Surveillance Standing Committee. 2017. Genomic sequences of *Streptococcus agalactiae* with high-level gentamicin resistance, collected in the BSAC bacteraemia surveillance. J Antimicrob Chemother 72:2704–2707. https://doi.org/10.1093/jac/dkx207.
5. Delamare-Debouteville J, Bowater R, Condon K, Reynolds A, Fisk A, Aviles F, Barnes AC. 2015. Infection and pathology in Queensland grouper, *Epinephelus lanceolatus*, (Bloch), caused by exposure to *Streptococcus agalactiae* via different routes. J Fish Dis 38:1021–1035. https://doi.org/10.1111/jfd.12273.
6. Bowater RO, Forbes-Faulkner J, Anderson IG, Condon K, Robinson B, Kong F, Gilbert GL, Reynolds A, Hyland S, McPherson G, Brien JO, Blyde D. 2012. Natural outbreak of Streptococcus agalactiae (GBS) infection in wild giant Queensland grouper, Epinephelus lanceolatus (Bloch), and other wild fish in northern Queensland, Australia. J Fish Dis 35:173–186. https://doi.org/10.1111/j.1365-2761.2011.01332.x.
7. Elliott JA, Facklam RR, Richter CB. 1990. Whole-cell protein patterns of nonhemolytic group B, type Ib, streptococci isolated from humans, mice, cattle, frogs, and fish. J Clin Microbiol 28:628–630.
8. Evans JJ, Klesius PH, Gilbert PM, Shoemaker CA, Al Sarawi MA, Landsberg J, Duremdez R, Al Marzouk A, Al Zenki S. 2002. Characterization of β-haemolytic group B *Streptococcus agalactiae* in cultured seabream, *Sparus auratus* L., and wild mullet, *Liza klunzingeri* (Day), in Kuwait. J Fish Dis 25:505–513. https://doi.org/10.1046/j.1365-2761.2002.00392.x.
9. Verner-Jeffreys D, Wallis T, Cano Cejas I, Ryder D, Haydon D, Domazoro JF, Dontwi J, Field T, Adjei-Boteng D, Wood G, Bean T, Feist SW. 2017. Streptococcus agalactiae multilocus sequence type 261 is associated with mortalities in the emerging Ghanaian tilapia industry. J Fish Dis 41:175–179. https://doi.org/10.1111/jfd.12681.
10. Li K, Liu L, Clausen JH, Lu M, Dalsgaard A. 2016. Management measures to control diseases reported by tilapia (Oreochromis spp.) and whiteleg shrimp (Litopenaeus vannamei) farmers in Guangdong, China. Aquaculture 457:91–99. https://doi.org/10.1016/j.aquaculture.2016.02.008.
11. Chen M, Li L-P, Wang R, Liang W-W, Huang Y, Li J, Lei A-Y, Huang W-Y, Gan X. 2012. PCR detection and PFGE genotype analyses of streptococcal clinical isolates from tilapia in China. Vet Microbiol 159:526–530. https://doi.org/10.1016/j.vetmic.2012.04.035.
12. Zhang D, Li A, Guo Y, Zhang Q, Chen X, Gong X. 2013. Molecular characterization of Streptococcus agalactiae in diseased farmed tilapia in China. Aquaculture 412:64–69. https://doi.org/10.1016/j.aquaculture.2013.07.014.
13. Rosinski-Chupin I, Sauvage E, Mairey B, Mangenot S, Ma L, Da Cunha V, Rusniok C, Bouchier C, Barbe V, Glaser P. 2013. Reductive evolution in Streptococcus agalactiae and the emergence of a host adapted lineage. BMC Genomics 14:252. https://doi.org/10.1186/1471-2164-14-252.
14. Brynildsrud O, Feil EJ, Bohlin J, Castillo-Ramirez S, Colquhoun D, McCarthy U, Matejusova IM, Rhodes LD, Wiens GD, Verner-Jeffreys DW. 2014. Microevolution of *Renibacterium salmoninarum*: evidence for intercontinental dissemination associated with fish movements. ISME J 8:746–756. https://doi.org/10.1038/ismej.2013.186.
15. Barnes AC, Delamare-Debouteville J, Brosnahan C, Gudkovs N, Morrison RN, Carson J. 2016. Whole genome analysis of Yersinia ruckeri isolated over 27 years in Australia and New Zealand reveals geographical endemism over multiple lineages and recent evolution under host selection. Microb Genom 2:e000095. https://doi.org/10.1099/mgen.0.000095.
16. Sabat AJ, Budimir A, Nashev D, Sá-Leão R, Van Dijl JM, Laurent F, Grundmann H, Friedrich AW, ESCMID Study Group of Epidemiological Markers (ESGEM). 2013. Overview of molecular typing methods for outbreak detection and epidemiological surveillance. Euro Surveill 18:20380. https://doi.org/10.2807/ese.18.04.20380-en.
17. Jones N, Bohnsack JF, Takahashi S, Oliver KA, Chan M-S, Kunst F, Glaser P, Rusniok C, Crook DWM, Harding RM, Bisharat N, Spratt BG. 2003. Multilocus sequence typing system for group B streptococcus. J Clin Microbiol 41:2530–2536. https://doi.org/10.1128/JCM.41.6.2530-2536.2003.
18. Carvalho-Castro GA, Silva JR, Paiva LV, Custódio DAC, Moreira RO, Mian GF, Prado IA, Chalfun-Junior A, Costa GM. 2017. Molecular epidemiology of Streptococcus agalactiae isolated from mastitis in Brazilian dairy herds. Braz J Microbiol 48:551–559. https://doi.org/10.1016/j.bjm.2017.02.004.
19. Kong F, Gowan S, Martin D, James G, Gilbert GL. 2002. Serotype identification of group B streptococci by PCR and sequencing. J Clin Microbiol 40:216–226. https://doi.org/10.1128/JCM.40.1.216-226.2002.
20. Kapatai G, Patel D, Efstratiou A, Chalker VJ. 2017. Comparison of molecular serotyping approaches of *Streptococcus agalactiae* from genomic

sequences. BMC Genomics 18:429. https://doi.org/10.1186/s12864-017-3820-5.
21. Sheppard AE, Vaughan A, Jones N, Turner P, Turner C, Efstratiou A, Patel D, Modernising Medical Microbiology Informatics Group, Walker AS, Berkley JA, Crook DW, Seale AC. 2016. Capsular typing method for *Streptococcus agalactiae* using whole-genome sequence data. J Clin Microbiol 54:1388–1390. https://doi.org/10.1128/JCM.03142-15.
22. Eldar A, Shapiro O, Bejerano Y, Bercovier H. 1995. Vaccination with whole-cell vaccine and bacterial protein extract protects tilapia against *Streptococcus difficile* meningoencephalitis. Vaccine 13:867–870. https://doi.org/10.1016/0264-410X(94)00067-W.
23. Evans JJ, Klesius PH, Shoemaker CA. 2004. Efficacy of *Streptococcus agalactiae* (group B) vaccine in tilapia (*Oreochromis niloticus*) by intraperitoneal and bath immersion administration. Vaccine 22:3769–3773. https://doi.org/10.1016/j.vaccine.2004.03.012.
24. Chong S, Wong W, Lee W, Tan Z, Tay Y, Teo X, Chee L, Fernandez C. 2017. *Streptococcus agalactiae* outbreaks in cultured golden pomfret, *Trachinotus blochii* (Lacépède), in Singapore. J Fish Dis 40:971–974. https://doi.org/10.1111/jfd.12570.
25. Sun J, Fang W, Ke B, He D, Liang Y, Ning D, Tan H, Peng H, Wang Y, Ma Y, Ke C, Deng X. 2016. Inapparent *Streptococcus agalactiae* infection in adult/commercial tilapia. Sci Rep 6:26319. https://doi.org/10.1038/srep26319.
26. Kannika K, Pisuttharachai D, Srisapoome P, Wongtavatchai J, Kondo H, Hirono I, Unajak S, Areechon N. 2017. Molecular serotyping, virulence gene profiling and pathogenicity of *Streptococcus agalactiae* isolated from tilapia farms in Thailand by multiplex PCR. J Appl Microbiol 122:1497–1507. https://doi.org/10.1111/jam.13447.
27. Godoy D, Carvalho-Castro G, Leal C, Pereira U, Leite R, Figueiredo H. 2013. Genetic diversity and new genotyping scheme for fish pathogenic Streptococcus agalactiae. Lett Appl Microbiol 57:476–483. https://doi.org/10.1111/lam.12138.
28. Evans JJ, Bohnsack JF, Klesius PH, Whiting AA, Garcia JC, Shoemaker CA, Takahashi S. 2008. Phylogenetic relationships among Streptococcus agalactiae isolated from piscine, dolphin, bovine and human sources: a dolphin and piscine lineage associated with a fish epidemic in Kuwait is also associated with human neonatal infections in Japan. J Med Microbiol 57:1369–1376. https://doi.org/10.1099/jmm.0.47815-0.
29. Delannoy CMJ, Crumlish M, Fontaine MC, Pollock J, Foster G, Dagleish MP, Turnbull JF, Zadoks RN. 2013. Human Streptococcus agalactiae strains in aquatic mammals and fish. BMC Microbiol 13:41. https://doi.org/10.1186/1471-2180-13-41.
30. Pridgeon JW, Zhang D. 2014. Complete genome sequence of a virulent streptococcus agalactiae strain, 138P, isolated from diseased Nile tilapia. Genome Announc 2:e00295-14. https://doi.org/10.1128/genomeA.00295-14.
31. Lusiastuti AM, Textor M, Seeger H, Akineden Ö, Zschöck M. 2014. The occurrence of *Streptococcus agalactiae* sequence type 261 from fish disease outbreaks of tilapia *Oreochromis niloticus* in Indonesia. Aquacult Res 45:1260–1263. https://doi.org/10.1111/are.12069.
32. Suanyuk N, Kong F, Ko D, Gilbert GL, Supamattaya K. 2008. Occurrence of rare genotypes of *Streptococcus agalactiae* in cultured red tilapia *Oreochromis* sp. and Nile tilapia *O. niloticus* in Thailand–relationship to human isolates? Aquaculture 284:35–40. https://doi.org/10.1016/j.aquaculture.2008.07.034.
33. Li L, Wang R, Liang W, Gan X, Huang T, Huang Y, Li J, Shi Y, Chen M, Luo H. 2013. Rare serotype occurrence and PFGE genotypic diversity of Streptococcus agalactiae isolated from tilapia in China. Vet Microbiol 167:719–724. https://doi.org/10.1016/j.vetmic.2013.09.001.
34. Chideroli RT, Amoroso N, Mainardi RM, Suphoronski SA, de Padua SB, Alfieri AF, Alfieri AA, Mosela M, Moralez ATP, de Oliveira AG, Zanolo R, Di Santis GW, Pereira UP. 2017. Emergence of a new multidrug-resistant and highly virulent serotype of *Streptococcus agalactiae* in fish farms from Brazil. Aquaculture 479:45–51. https://doi.org/10.1016/j.aquaculture.2017.05.013.
35. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, Deboy RT, Davidsen TM, Mora M, Scarselli M, Margarit y Ros I, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou L, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJ, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM. 2005. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome." Proc Natl Acad Sci U S A 102:13950–13955. https://doi.org/10.1073/pnas.0506758102.
36. Francis AR, Tanaka MM. 2012. Evolution of variation in presence and absence of genes in bacterial pathways. BMC Evol Biol 12:55. https://doi.org/10.1186/1471-2148-12-55.
37. Puymège A, Bertin S, Guédon G, Payot S. 2015. Analysis of Streptococcus agalactiae pan-genome for prevalence, diversity and functionality of integrative and conjugative or mobilizable elements integrated in the tRNALys CTT gene. Mol Genet Genomics 290:1727–1740. https://doi.org/10.1007/s00438-015-1031-9.
38. Richards VP, Palmer SR, Pavinski Bitar PD, Qin X, Weinstock GM, Highlander SK, Town CD, Burne RA, Stanhope MJ. 2014. Phylogenomics and the dynamic genome evolution of the genus *Streptococcus*. Genome Biol Evol 6:741–753. https://doi.org/10.1093/gbe/evu048.
39. Liu G, Zhang W, Lu C. 2013. Comparative genomics analysis of Streptococcus agalactiae reveals that isolates from cultured tilapia in China are closely related to the human strain A909. BMC Genomics 14:775. https://doi.org/10.1186/1471-2164-14-775.
40. He E-M, Chen C-W, Guo Y, Hsu M-H, Zhang L, Chen H-L, Zhao G-P, Chiu C-H, Zhou Y. 2017. The genome of serotype VI Streptococcus agalactiae serotype VI and comparative analysis. Gene 597:59–65. https://doi.org/10.1016/j.gene.2016.10.030.
41. Maione D, Margarit I, Rinaudo CD, Masignani V, Mora M, Scarselli M, Tettelin H, Brettoni C, Iacobini ET, Rosini R, D'Agostino N, Miorin L, Buccato S, Mariani M, Galli G, Nogarotto R, Nardi-Dei V, Vegni F, Fraser C, Mancuso G, Teti G, Madoff LC, Paoletti LC, Rappuoli R, Kasper DL, Telford JL, Grandi G. 2005. Identification of a universal group B streptococcus vaccine by multiple genome screen. Science 309:148–150. https://doi.org/10.1126/science.1109869.
42. Bowater RO, Dennis MM, Blyde D, Stone B, Barnes AC, Delamare-Deboutteville J, Horton MA, White M, Condon K, Jones R. 2017. Epizootics of *Streptococcus agalactiae* infection in captive rays from Queensland, Australia. J Fish Dis 41:223–232. https://doi.org/10.1111/jfd.12701.
43. Evans JJ, Klesius PH, Pasnik DJ, Bohnsack JF. 2009. Human Streptococcus agalactiae isolate in Nile tilapia (Oreochromis niloticus). Emerg Infect Dis 15:774–776. https://doi.org/10.3201/eid1505.080222.
44. Bishop EJ, Shilton C, Benedict S, Kong F, Gilbert GL, Gal D, Godoy D, Spratt BG, Currie BJ. 2007. Necrotizing fasciitis in captive juvenile *Crocodylus porosus* caused by *Streptococcus agalactiae*: an outbreak and review of the animal and human literature. Epidemiol Infect 135:1248–1255. https://doi.org/10.1017/S0950268807008515.
45. Herbert MA, Beveridge CJ, McCormick D, Aten E, Jones N, Snyder LA, Saunders NJ. 2005. Genetic islands of Streptococcus agalactiae strains NEM316 and 2603VR and their presence in other group B streptococcal strains. BMC Microbiol 5:31. https://doi.org/10.1186/1471-2180-5-31.
46. Fischer A, Liljander A, Kaspar H, Muriuki C, Fuxelius HH, Bongcam-Rudloff E, de Villiers EP, Huber CA, Frey J, Daubenberger C, Bishop R, Younan M, Jores J. 2013. Camel Streptococcus agalactiae populations are associated with specific disease complexes and acquired the tetracycline resistance gene tetM via a Tn916-like element. Vet Res 44:86. https://doi.org/10.1186/1297-9716-44-86.
47. Flores AR, Galloway-Pena J, Sahasrabhojane P, Saldana M, Yao H, Su X, Ajami NJ, Holder ME, Petrosino JF, Thompson E, Margarit YRI, Rosini R, Grandi G, Horstmann N, Teatero S, McGeer A, Fittipaldi N, Rappuoli R, Baker CJ, Shelburne SA. 2015. Sequence type 1 group B Streptococcus, an emerging cause of invasive disease in adults, evolves by small genetic changes. Proc Natl Acad Sci U S A 112:6431–6436. https://doi.org/10.1073/pnas.1504725112.
48. Almeida A, Alves-Barroco C, Sauvage E, Bexiga R, Albuquerque P, Tavares F, Santos-Sanches I, Glaser P. 2016. Persistence of a dominant bovine lineage of group B Streptococcus reveals genomic signatures of host adaptation. Environ Microbiol 18:4216–4229. https://doi.org/10.1111/1462-2920.13550.
49. Vandamme P, Devriese LA, Pot B, Kersters K, Melin P. 1997. Streptococcus difficile is a nonhemolytic group B, type Ib Streptococcus. Int J Syst Bacteriol 47:81–85. https://doi.org/10.1099/00207713-47-1-81.
50. Eldar A, Bejerano Y, Bercovier H. 1994. Streptococcus shiloi and Streptococcus difficile: two new streptococcal species causing a meningoencephalitis in fish. Curr Microbiol 28:139–143. https://doi.org/10.1007/BF01571054.
51. Mather PB, Arthington AH. 1991. An assessment of genetic differentiation among feral Australian tilapia populations. Austral J Mar Freshwater Res 42:721–728. https://doi.org/10.1071/MF9910721.
52. Thorpe HA, Bayliss SC, Sheppard SK, Feil EJ. 2017. Piggy: a rapid, large-

scale pan-genome analysis tool for intergenic regions in bacteria. bioRxiv https://doi.org/10.1101/179515.
53. Kuo CH, Ochman H. 2010. The extinction dynamics of bacterial pseudogenes. PLoS Genet 6:e1001050. https://doi.org/10.1371/journal.pgen.1001050.
54. Lerat E, Ochman H. 2005. Recognizing the pseudogenes in bacterial genomes. Nucleic Acids Res 33:3125–3132. https://doi.org/10.1093/nar/gki631.
55. Lerat E, Ochman H. 2004. Psi-Phi: exploring the outer limits of bacterial pseudogenes. Genome Res 14:2273–2278. https://doi.org/10.1101/gr.2925604.
56. Spellerberg B, Martin S, Brandt C, Lutticken R. 2000. The *cyl* genes of *Streptococcus agalactiae* are involved in the production of pigment. FEMS Microbiol Lett 188:125–128. https://doi.org/10.1111/j.1574-6968.2000.tb09182.x.
57. Nagano N, Nagano Y, Taguchi F. 2002. High expression of a C protein beta antigen gene among invasive strains from certain clonally related groups of type Ia and Ib group B streptococci. Infect Immun 70:4643–4649. https://doi.org/10.1128/IAI.70.8.4643-4649.2002.
58. Yang Q, Zhang M, Harrington DJ, Black GW, Sutcliffe IC. 2011. A proteomic investigation of Streptococcus agalactiae reveals that human serum induces the C protein beta antigen and arginine deiminase. Microbes Infect 13:757–760. https://doi.org/10.1016/j.micinf.2011.03.001.
59. Yang Q, Zhang M, Harrington DJ, Black GW, Sutcliffe IC. 2010. A proteomic investigation of Streptococcus agalactiae grown under conditions associated with neonatal exposure reveals the upregulation of the putative virulence factor C protein beta antigen. Int J Med Microbiol 300:331–337. https://doi.org/10.1016/j.ijmm.2010.01.001.
60. Locke JB, Colvin KM, Datta AK, Patel SK, Naidu NN, Neely MN, Nizet V, Buchanan JT. 2007. Streptococcus iniae capsule impairs phagocytic clearance and contributes to virulence in fish. J Bacteriol 189:1279–1287. https://doi.org/10.1128/JB.01175-06.
61. Delannoy CMJ, Zadoks RN, Crumlish M, Rodgers D, Lainson FA, Ferguson HW, Turnbull J, Fontaine MC. 2016. Genomic comparison of virulent and non-virulent Streptococcus agalactiae in fish. J Fish Dis 39:13–29. https://doi.org/10.1111/jfd.12319.
62. Yao KH, Poulsen K, Maione D, Rinaudo CD, Baldassarri L, Telford JL, Sorensen UBS, DEVANI Study Group, Kilian M. 2013. Capsular gene typing of Streptococcus agalactiae compared to serotyping by latex agglutination. J Clin Microbiol 51:503–507. https://doi.org/10.1128/JCM.02417-12.
63. Pasnik DJ, Evans JJ, Panangala VS, Klesius PH, Shelby RA, Shoemaker CA. 2005. Antigenicity of Streptococcus agalactiae extracellular products and vaccine efficacy. J Fish Dis 28:205–212. https://doi.org/10.1111/j.1365-2761.2005.00619.x.
64. Seib KL, Zhao X, Rappuoli R. 2012. Developing vaccines in the era of genomics: a decade of reverse vaccinology. Clin Microbiol Infect 18(Suppl 5):109–116. https://doi.org/10.1111/j.1469-0691.2012.03939.x.
65. Millard CM, Baiano JC, Chan C, Yuen B, Aviles F, Landos M, Chong RS, Benedict S, Barnes AC. 2012. Evolution of the capsular operon of Streptococcus iniae in response to vaccination. Appl Environ Microbiol 78:8219–8226. https://doi.org/10.1128/AEM.02216-12.
66. Amann RI, Ludwig W, Schleifer KH. 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. Microbiol Rev 59:143–169.
67. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
68. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
69. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
70. Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. 2009. Reordering contigs of draft genomes using the Mauve aligner. Bioinformatics 25:2071–2073. https://doi.org/10.1093/bioinformatics/btp356.
71. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
72. Wyres KL, Wick RR, Gorrie C, Jenney A, Follador R, Thomson NR, Holt KE. 2016. Identification of Klebsiella capsule synthesis loci from whole genome data. Microb Genom 2:e000102. https://doi.org/10.1099/mgen.0.000102.
73. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 50:1355–1361. https://doi.org/10.1128/JCM.06094-11.
74. Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol 15:524. https://doi.org/10.1186/s13059-014-0524-x.
75. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res 43:e15. https://doi.org/10.1093/nar/gku1196.
76. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
77. He Z, Zhang H, Gao S, Lercher MJ, Chen W-H, Hu S. 2016. Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. Nucleic Acids Res 44:W236–W241. https://doi.org/10.1093/nar/gkw370.
78. Salipante SJ, Hall BG. 2011. Inadequacies of minimum spanning trees in molecular epidemiology. J Clin Microbiol 49:3568–3575. https://doi.org/10.1128/JCM.00919-11.
79. Petkau A, Stuart-Edwards M, Stothard P, Van Domselaar G. 2010. Interactive microbial genome visualization with GView. Bioinformatics 26:3125–3126. https://doi.org/10.1093/bioinformatics/btq588.
80. Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12:402. https://doi.org/10.1186/1471-2164-12-402.
81. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691–3693. https://doi.org/10.1093/bioinformatics/btv421.
82. Stanton-Cook M, Ben Zakour N, Alikhan N, Beatson S. 2013. SeqFindR. https://github.com/mscook/SeqFindR.
83. Chen L, Yang J, Yu J, Yao Z, Sun L, Shen Y, Jin Q. 2005. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res 33:D325–D328. https://doi.org/10.1093/nar/gki008.
84. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w^{1118}; iso-2; iso-3. Fly (Austin) 6:80–92. https://doi.org/10.4161/fly.19695.
85. Yamamoto S, Miyake K, Koike Y, Watanabe M, Machida Y, Ohta M, Iijima S. 1999. Molecular characterization of type-specific capsular polysaccharide biosynthesis genes of Streptococcus agalactiae type Ia. J Bacteriol 181:5176–5184.
86. Watanabe M, Miyake K, Yanae K, Kataoka Y, Koizumi S, Endo T, Ozaki A, Iijima S. 2002. Molecular characterization of a novel beta1,3-galactosyltransferase for capsular polysaccharide synthesis by Streptococcus agalactiae type Ib. J Biochem 131:183–191. https://doi.org/10.1093/oxfordjournals.jbchem.a003086.
87. Martins ER, Melo-Cristino J, Ramirez M. 2007. Reevaluating the serotype II capsular locus of Streptococcus agalactiae. J Clin Microbiol 45:3384–3386. https://doi.org/10.1128/JCM.01296-07.
88. Chaffin DO, Beres SB, Yim HH, Rubens CE. 2000. The serotype of type Ia and III group B streptococci is determined by the polymerase gene within the polycistronic capsule operon. J Bacteriol 182:4466–4477. https://doi.org/10.1128/JB.182.16.4466-4477.2000.
89. Chaffin DO, McKinnon K, Rubens CE. 2002. CpsK of *Streptococcus agalactiae* exhibits alpha2,3-sialyltransferase activity in *Haemophilus ducreyi*. Mol Microbiol 45:109–122. https://doi.org/10.1046/j.1365-2958.2002.02988.x.
90. Cieslewicz MJ, Chaffin D, Glusman G, Kasper D, Madan A, Rodrigues S, Fahey J, Wessels MR, Rubens CE. 2005. Structural and genetic diversity of group B streptococcus capsular polysaccharides. Infect Immun 73:3096–3103. https://doi.org/10.1128/IAI.73.5.3096-3103.2005.
91. Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. Bioinformatics 27:1009–1010. https://doi.org/10.1093/bioinformatics/btr039.