Persistence of a dominant bovine lineage of group B S treptococcus reveals genomic signatures of host adaptation

Alexandre Almeida, Cinthia Alves-Barroco, Elisabeth Sauvage, Ricardo Bexiga, Pedro Albuquerque, Fernando Tavares, Ilda Santos-Sanches, Philippe Glaser

▶ **To cite this version:**

Alexandre Almeida, Cinthia Alves-Barroco, Elisabeth Sauvage, Ricardo Bexiga, Pedro Albuquerque, et al.. Persistence of a dominant bovine lineage of group B S treptococcus reveals genomic signatures of host adaptation. Environmental Microbiology, 2016, 18 (11), pp.4216-4229. 10.1111/1462-2920.13550 . pasteur-03153236

HAL Id: pasteur-03153236
https://pasteur.hal.science/pasteur-03153236

Submitted on 26 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution - NoDerivatives 4.0 International License

environmental microbiology

Environmental Microbiology (2016) **18**(11), 4216–4229 doi:10.1111/1462-2920.13550

Persistence of a dominant bovine lineage of group B *Streptococcus* reveals genomic signatures of host adaptation

Alexandre Almeida,[1,2,3] Cinthia Alves-Barroco,[4] Elisabeth Sauvage,[1,2] Ricardo Bexiga,[5] Pedro Albuquerque,[6] Fernando Tavares,[6,7] Ilda Santos-Sanches[4] and Philippe Glaser[1,2]*

[1]*Institut Pasteur, Unité Evolution et Ecologie de la Résistance aux Antibiotiques, Paris, France.*

[2]*CNRS UMR 3525, Paris, France.*

[3]*Université Pierre et Marie Curie, Paris, France.*

[4]*Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, UCIBIO – Unidade de Ciências Biomoleculares Aplicadas, Universidade NOVA de Lisboa, Lisboa, Portugal.*

[5]*Faculdade de Medicina Veterinária, Centro de Investigação Interdisciplinar em Sanidade Animal, Universidade de Lisboa, Lisboa, Portugal.*

[6]*CIBIO – Centro de Investigação em Biodiversidade e Recursos Genéticos, InBIO, Laboratório Associado, Universidade do Porto, Campus Agrário de Vairão, Vairão, Portugal.*

[7]*Faculdade de Ciências, Departamento de Biologia, Universidade do Porto, Porto, Portugal.*

Summary

Group B *Streptococcus* (GBS) is a host-generalist species, most notably causing disease in humans and cattle. However, the differential adaptation of GBS to its two main hosts, and the risk of animal to human infection remain poorly understood. Despite improvements in control measures across Europe, GBS is still one of the main causative agents of bovine mastitis in Portugal. Here, by whole-genome analysis of 150 bovine GBS isolates we discovered that a single CC61 clone is spreading throughout Portuguese herds since at least the early 1990s, having virtually replaced the previous GBS population. Mutations within an iron/manganese transporter were independently acquired by all of the CC61 isolates, underlining a key adaptive strategy to persist in the bovine host. Lateral transfer of bacteriocin production and antibiotic resistance genes also underscored the contribution of the microbial ecology and genetic pool within the bovine udder environment to the success of this clone. Compared to strains of human origin, GBS evolves twice as fast in bovines and undergoes recurrent pseudogenizations of human-adapted traits. Our work provides new insights into the potentially irreversible adaptation of GBS to the bovine environment.

Received 20 July, 2016; revised 20 September, 2016; accepted 26 September, 2016. *For correspondence. E-mail pglaser@pasteur.fr; Tel. (+33) 1 45 68 89 96

Introduction

Bovine mastitis is an inflammatory disease of the mammary gland, causing a significant burden for animal welfare and for the dairy industry worldwide (Heikkila *et al.*, 2012; Deb *et al.*, 2013). One of the major pathogens responsible for bovine intramammary infections is *Streptococcus agalactiae* (group B *Streptococcus*, GBS) (Keefe, 1997; Wyder *et al.*, 2011). Control strategies implemented since 1960 were able to reduce the incidence of GBS mastitis in several European countries. However, many dairy farms worldwide continue to be predominantly infected by GBS, or have observed a recent re-emergence of GBS mastitis (Kalmus *et al.*, 2011; Klimiene *et al.*, 2011; Mweu *et al.*, 2012; Bi *et al.*, 2016; Jorgensen *et al.*, 2016). Particularly in Portuguese herds, GBS is among the most frequently detected species in animals diagnosed with mastitis (Almeida *et al.*, 2013; Rato *et al.*, 2013). Besides infecting cattle, GBS asymptomatically colonizes the gastrointestinal tract of 10–30% of the human population (Schuchat, 1998), and is a leading cause of infections in neonates and in immunocompromised individuals (Dermer *et al.*, 2004).

Multilocus sequence typing (MLST) studies of GBS have described various clonal complexes (CCs), and corresponding sequence types (STs) with different host specificities. Phylogenetic studies showed that human GBS throughout the world encompass a generally conserved population of a few clones that were selected following the use of tetracycline in the 1950s onwards (Da Cunha *et al.*, 2014). They correspond most frequently to

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd. This is an open access article under the terms of the Creative Commons Attribution -NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited and no modifications or adaptations are made.

clonal complexes 1, 6-8-10, 17, 19 and 23 (Jones *et al*., 2003). Although fewer population analyses of bovine isolates have been performed, the distribution of CCs found in cattle is more variable across different studies. For instance, although CC61/67 is known as a bovine-specific lineage (Sorensen *et al*., 2010) and found to have been widely distributed across both the UK (Bisharat *et al*., 2004) and the USA (Springman *et al*., 2014), a recent study of Norwegian dairy farms did not identify any CC61/67 strains in the 19 herds that were infected by GBS (Jorgensen *et al*., 2016). Specifically in the south-western region of Portugal, genotyping of GBS strains isolated between 2002 and 2003 revealed that they clustered mainly into ST2, ST23, ST61, plus a novel ST61-related lineage, ST554 (Rato *et al*., 2013). In France, a high proportion of bovine GBS belonging to CC23 and CC61/67 were isolated from various geographical areas, while a smaller number of strains were found belonging to CC1, CC2, CC6, CC17 and CC19 (Haenni *et al*., 2010). Indeed, several clonal complexes, such as CC1, CC7, CC23 and CC26 comprise GBS strains adapted to both humans and bovines, suggesting the possibility of interspecies transmission (Oliveira *et al*., 2006; Sorensen *et al*., 2010; Zadoks *et al*., 2011). A PCR-based screening (Richards *et al*., 2011) and a transcriptome analysis of a bovine-specific strain (Richards *et al*., 2013) have looked at some of the genetic factors unique to GBS strains of bovine origin. However, population genomics studies are still lacking to understand the specific adaptation of GBS to the bovine host.

In this work, we performed an in-depth whole-genome analysis of 150 isolates, revealing that cattle throughout Portugal is infected almost exclusively by a single GBS clone. We investigated the adaptive strategies contributing to its evolutionary success, while unveiling new insights into the distinct dynamic between human- and bovine-adapted GBS.

Results

Genotyping of the bovine GBS population

We first analyzed a set of 197 GBS isolates collected between 2011 and 2014 from mastitic milk samples from 15 dairy farms throughout Portugal (Supporting Information Table S1). For a preliminary assessment of their genetic diversity, we sequenced the CRISPR1 locus due to its high discriminatory resolution (Lopez-Sanchez *et al*., 2012). Spacers identified in the leader end of the CRISPR1 locus presented a high degree of heterogeneity with a farm-specific distribution, suggesting that isolates from a single farm are genetically closer (Supporting Information Table S2). Surprisingly, spacers 7 and 147, characteristically found in CC61 strains at the trailer end of the locus (Lopez-Sanchez *et al*., 2012), were completely conserved across all but one isolate that showed an ST2-specific profile.

Based on the CC predicted for each isolate using this approach, a epidemiological analysis of these 197 isolates was performed in combination with earlier collections of GBS strains from the south-west of Portugal (Rato *et al*., 2013) and France (Haenni *et al*., 2010), and a set of 47 isolates from the north of Spain previously genotyped by our group (Lopez-Sanchez *et al*., 2012) (Fig. 1). GBS populations analyzed in these earlier collections were genetically more diverse, with a significant number of isolates belonging to CC2, CC23 or CC61 (Fig. 1). In particular, there was a considerable shift in the proportion of CC61 isolates found in the south-west of Portugal, rising from 53% in 2002–2003 to 98% in 2011–2013, and leading to a replacement of the GBS population (Fig. 1). Indeed, in 2002–2003, only two (33%) out of the six herds sampled were purely infected by CC61 (Rato *et al*., 2013). However, in the 2011–2013 collection, four (80%) out of the five farms analyzed were infected by this clone (Supporting Information Table S1).

Dairy herds in Portugal are almost exclusively infected by a single CC61 clone

To reconstruct the evolutionary history of this population and gain a broader overview of their phylogenetic structure, we selected 128 CC61 isolates from Portugal (118 collected between 2011 and 2014, and 10 between 2002 and 2003), together with the three CC61 isolates from France, to perform whole-genome analysis. The genome of SA111, used as a reference sequence, was completely assembled into a single contig of 2,275,139 base pairs (bp), using long-read sequencing (PacBio). A total of 2212 protein coding genes, 7 rRNA loci and 80 tRNAs genes were predicted and annotated with Prokka (Seemann, 2014). Whole-genome sequencing of 19 ST2 isolates from Portugal and Spain (Supporting Information Tables S1 and S2) was also performed to infer their genetic relatedness and search for analogous mechanisms of adaptation between the CC61 and ST2 bovine GBS clones.

The 131 CC61 isolates selected from Portugal and France were analyzed and compared with 25 CC61 genomes available on the NCBI database (Supporting Information Fig. S1). Phylogenetic analysis showed that all CC61 isolates from Portugal collected since 2002 correspond to one monophyletic clade (Supporting Information Fig. S1). Strikingly, this reveals that the entire CC61 population of GBS infecting dairy farms in Portugal resulted from the recent dissemination of a single clone. Likewise, whole-genome comparison of the 19 ST2 representative isolates, with 19 publicly available genomes, demonstrated a tight clustering of isolates from Portugal and the north of

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

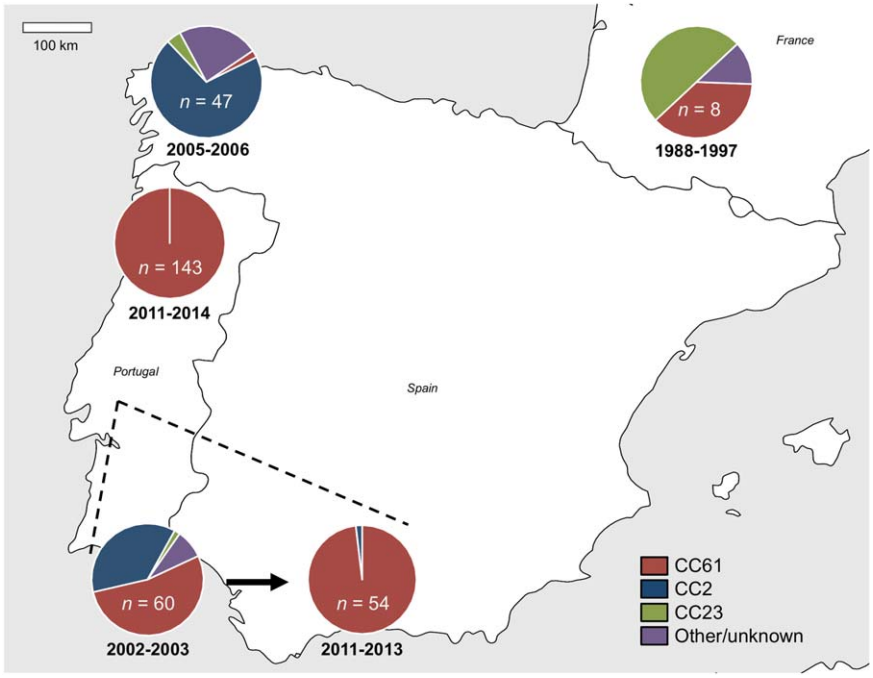

Fig. 1. Genotyping of GBS isolates identified in the south-west of Europe. Map depicting the total number of isolates identified in Portugal, the north of Spain and the south-west of France belonging to different clonal complexes (CCs) according to the figure key. Isolates included in the analysis correspond to those sampled in this work, in addition to other GBS collections isolated from the south-west of Portugal between 2002 and 2003 (Rato *et al.*, 2013), the south-west of France (Haenni *et al.*, 2010) and a collection from the north of Spain.

Spain (Supporting Information Fig. S2), with an average of only 20 SNPs relative to their most recent common ancestor (MRCA).

Recurrent infections are due to contagious transmission and resilience to treatment

For a more detailed analysis of the genetic variation among the CC61 GBS isolates from Portugal, their genomes were mapped and compared with the complete assembly of strain SA111, used as an internal reference. All CC61 GBS isolates from Portugal clustered into two major groups showing an average of 92 SNPs since their MRCA (Fig. 2). The well-defined and herd-specific structure of this population underscores the swift dissemination of GBS within individual farms. One exception was seen in farms C (Setúbal) and W (Centre-West), as isolates collected from these two herds are phylogenetically intermixed – possibly the result of a recent exchange of GBS-infected cattle (Fig. 2). Furthermore, aside from all Portuguese strains being monophyletic, there is no clear geographical distribution of the farm-specific clusters (Fig. 2).

To gain additional insight into the persistence of GBS mastitis in the country, we analyzed all isolates collected within farm PV2 during a 14-month timespan (Fig. 3). For one particular animal, seven isolates were collected at two different time points (t4 and t5) from the four quarters of the udder. Interestingly, isolates that were collected from the same quarter were genetically more similar than to the others obtained from this farm. This suggests that the same strain might have persisted in the udder between the two distinct time points. However, our limited sampling and the presence of within-host GBS heterogeneity in the udder does not exclude the possibility that the later isolates could have also been newly transmitted from other animals within this herd.

The CC61 epidemic clone has persisted in Portugal since at least the early 1990s

The low level of divergence between the CC61 isolates suggests that this epidemic clone disseminated recently. Taking advantage of the sampling of this population across 12 years (Supporting Information Table S1), together with older isolates from France collected between 1996 and 1997, we applied Bayesian phylogenetics to deduce the age of their MRCA (Supporting Information Fig. S3). BEAST (Drummond *et al.*, 2012) analyses estimated that this lineage started to expand in Portugal between 1960 and 1990 (Supporting Information Fig. S3), while diverging at a mean evolutionary rate of 1.69×10^{-6} substitutions/site/year (95% highest posterior densities [HPDs], $1.18–2.23 \times 10^{-6}$ substitutions/site/year). These results infer that the CC61 GBS strains colonizing and infecting cattle in Portugal have persisted since at least 1990. Furthermore, the divergence rate determined for this bovine-specific population is twice as high as that previously inferred for GBS of human origin ($0.56–0.93 \times 10^{-6}$ substitutions/site/year) (Da Cunha *et al.*, 2014).

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

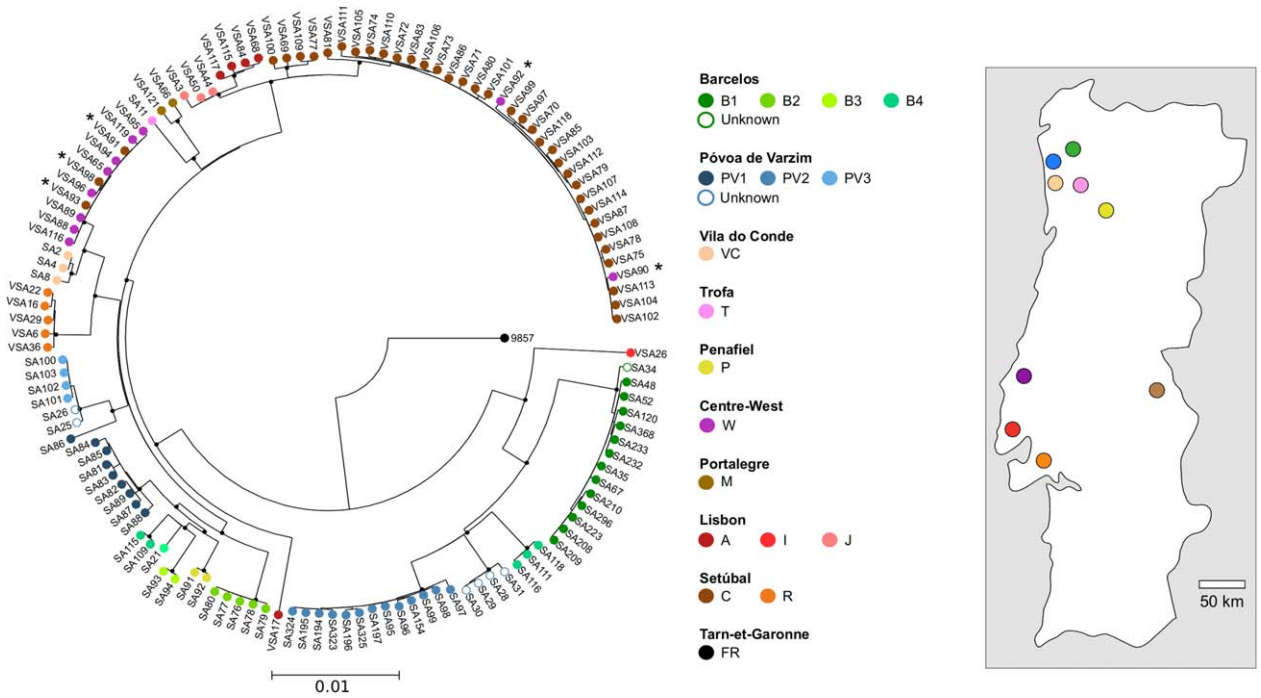

Fig. 2. Geographic distribution and phylogeny of the CC61 epidemic clone from Portugal. The ML phylogenetic tree was built using RAxML (Stamatakis, 2014), based on a total of 1340 core-genome SNPs over 1.51 Mb. Tree includes the 128 CC61 representative isolates from Portugal rooted with one CC61 genome from France (9857). Isolates are colour-coded according to the farm they were isolated from, as depicted in the figure key. Black dots in nodes denote a bootstrap support $>$ 95%. The five isolates that did not cluster exclusively with those within the same farm are indicated with an asterisk. The map of Portugal is depicted on the right with each location coloured as follows: Barcelos (green), Póvoa de Varzim (blue), Vila do Conde (beige), Trofa (pink), Penafiel (yellow), Centre-West (purple), Portalegre (brown), Lisbon (red) and Setúbal (orange).

Convergent adaptations were detected in major transcriptional regulators and within an iron/manganese transporter operon

The clonal expansion of the CC61 population in Portugal allowed us to compare the parallel *in vivo* evolution of multiple lineages during adaptation to the bovine environment. Using an approach adapted from Lieberman and colleagues (2011), we searched for coding regions under natural selection and with recurrent patterns of mutations. We detected a signal of purifying selection at the genome level of all CC61 isolates from Portugal (dN/dS average of 0.56), which suggests that selection is predominantly removing genetic variation from this population. However, we hypothesized that loci independently mutated in different isolates would be under local positive selection. For this analysis, SNPs that were detected in multiple isolates as a result of a single mutational event in their common ancestor were counted as one independent SNP. A total of 1012 independent mutations were detected among coding genes, and 164 independent SNPs within intergenic sequences. After excluding regions with a low density of SNPs, we identified 417 genes and 36 intergenic regions with at least one mutation (Supporting Information Fig. S4). Under a neutral evolutionary model, the mutations we detected would be randomly distributed across the core genome of GBS (Supporting Information Fig. S4). Yet, the number of genes and intergenic regions containing more

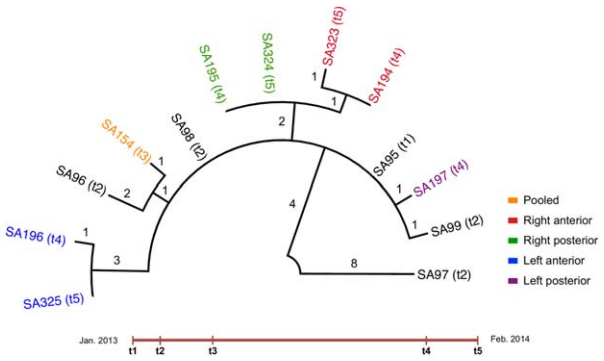

Fig. 3. ML phylogenetic tree built using RAxML (Stamatakis, 2014), highlighting longitudinal samples isolated from farm PV2. Eight isolates collected from one animal are colour-coded according to the quarter of the mammary gland they were isolated from, as indicated in the figure key. Black-coloured isolates correspond to those collected from other animals within the same farm. Digits match the number of SNPs underlying each branch of the tree. Time points t1 to t5 represent the following isolation dates: t1 – January 2013, t2 – February 2013, t3 – April 2013, t4 – December 2013, t5 – February 2014.

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

Table 1. Mutations in regulatory systems displaying a signal of positive selection.

Locus	Product	Length (bp)	NS[a]	S[b]	Isolates[c]
SA111_00141	HrcA family transcriptional regulator	1083	3	0	37
SA111_00245	Transcriptional antiterminator, BglG family	2037	4	0	56
SA111_01487	Two component system histidine kinase[d]	1230	3	0	45
SA111_01488	Two component system response regulator[d]	693	3	1	12
SA111_01781	Transmembrane histidine kinase CovS	1506	2	1	All
SA111_01923	Transcriptional regulator, GlnR	372	4	1	10
SA111_02131	Phosphate regulon sensor protein PhoR	1656	3	0	All
SA111_02142	Transcriptional regulator, MerR family	717	2	1	10
SA111_02249	TetR family transcriptional regulator	540	5	0	61

a. Number of nonsynonymous substitutions.
b. Number of synonymous substitutions.
c. Number of isolates, out of the 128 CC61 representatives from Portugal that were sequenced, affected by at least one of the mutations.
d. These two genes encode the sensor and regulator of the same two-component system.

than three independent mutations was significantly higher ($P < 0.001$) than expected by neutral drift (Supporting Information Fig. S4). The 85 genes that acquired at least three mutations are involved in a wide range of functions (Supporting Information Table S3). Among those, nine genes with 29 nonsynonymous SNPs and 4 synonymous substitutions correspond to regulatory systems involved in transcription control. dN/dS estimates further reinforced the hypothesis that this group of mutations is under positive selection (dN/dS = 2.25, CI = 1.84–2.47). One of the regulators affected is the CovS sensor histidine kinase (*SA111_01781*; Table 1), part of a two-component system known as CovRS, and a major regulator of virulence in GBS (Lamy *et al*., 2004; Santi *et al*., 2009). Two nonsynonymous mutations were detected in the C-terminal cytoplasmic kinase region of CovS, which is involved in the phosphorylation of its cognate response regulator CovR. An equivalent analysis of the ST2 bovine strains from Spain and Portugal uncovered three independent nonsynonymous substitutions also affecting the *covS* locus (data not shown). This further underscores the contribution of modifications within CovRS to virulence and colonization of the bovine host.

As for the intergenic regions, we saw a strong signal of convergent adaptation in a non-coding region located within a three-gene operon involved in iron and manganese transport (Bray *et al*., 2009) (Table 2 and Fig. 4). Indeed, 14 independent SNPs in 11 different positions were found downstream the first gene of the operon that encodes a metal-binding lipoprotein (Fig. 4A). Altogether, these mutations affected all of the CC61 isolates from Portugal (Table 2 and Fig. 4B). Furthermore, nine of these eleven positions were specifically located in a Rho-independent transcriptional terminator (Fig. 4A) (Rosinski-Chupin *et al*., 2015).

To assess the impact of these mutations, we quantified by RT-qPCR the expression of the genes flanking this internal terminator (*mtsA* and *mtsB*; Fig. 4C). We analyzed six bovine isolates (SA4, SA80, SA109, SA111, VSA66,

Table 2. Intergenic regions that acquired more than three independent mutations.

Region	Length (bp)	Locus[a]	Product	Strand[b]	SNPs	Isolates[c]
IR 1	178	*SA111_00084*	Alcohol/Acetaldehyde dehydrogenase	+	3	8
		SA111_00085	Alcohol dehydrogenase	+		
IR 2	36	*SA111_00149*	Hypothetical protein ywlG	+	3	2
		SA111_00150	Small-conductance channel	−		
IR 3	180	*SA111_00166*	Ribose operon repressor	−	3	39
		SA111_00167	Heme efflux system permease HrtB	+		
IR 4	109	*SA111_01084*	ABC transporter permease protein	+	4	3
		SA111_01085	DNA-binding response regulator	+		
IR 5	143	*SA111_01511*	tRNA-dependent ligase	−	4	3
		SA111_01512	Surface antigen-related protein	−		
IR 6	170	*SA111_01697*	Iron/manganese ABC transporter	−	14	All
		SA111_01698	Iron/manganese ABC transporter	−		
IR 7	138	*SA111_02248*	Phage infection protein	−	3	46
		SA111_02249	TetR family transcriptional regulator	+		

a. For each intergenic region, the first and second rows correspond to the genes identified upstream and downstream, respectively.
b. Gene direction. "+" denotes forward strand and "−" indicates reverse strand.
c. Number of isolates, out of the 128 CC61 representatives from Portugal that were sequenced, affected by at least one of the mutations.

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

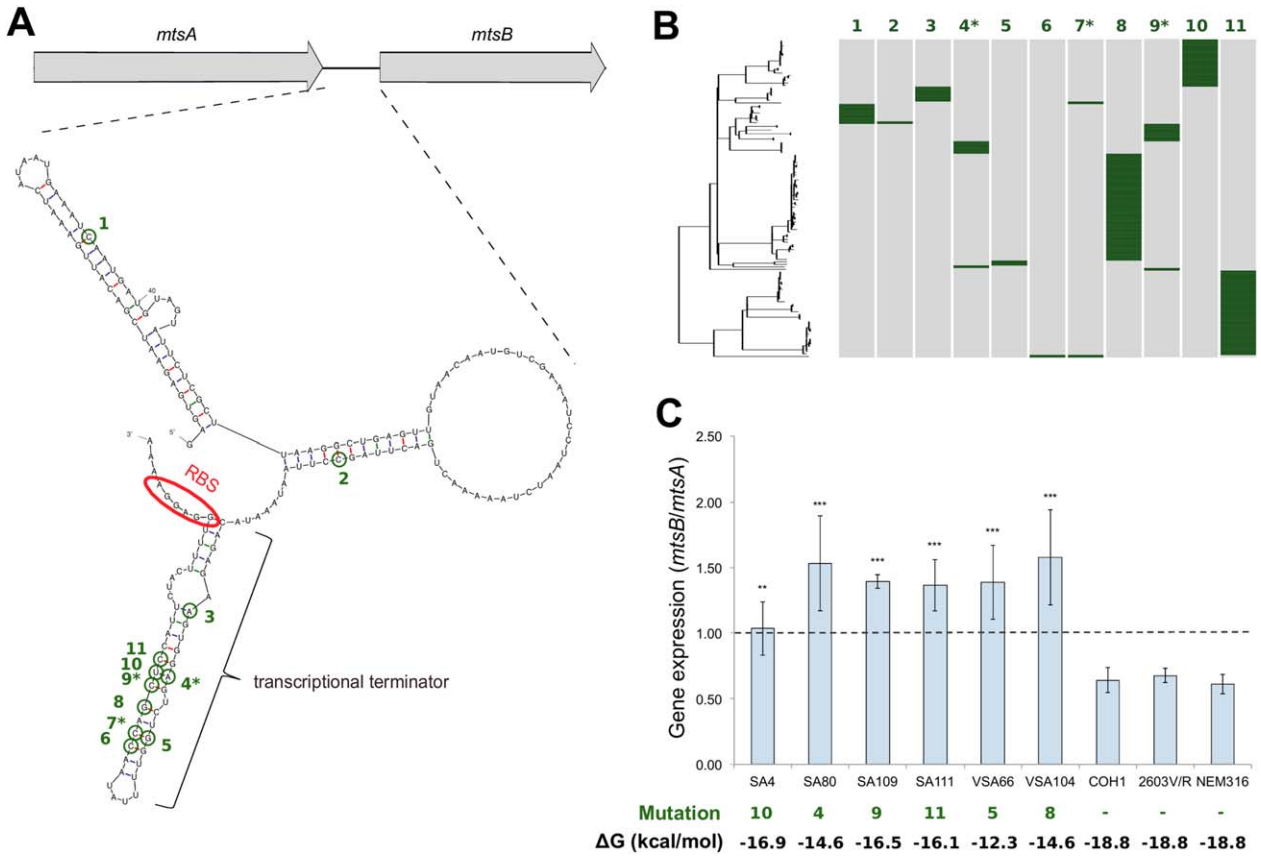

Fig. 4. Intergenic mutations within an operon for iron/manganese transport.
(**a**) Location of the independent mutations found between *mtsA* (SA111_01698) and *mtsB* (SA111_01697), relative to the mRNA secondary structure predicted with mfold (http://unafold.rna.albany.edu/?q = mfold) using the sequence of the MRCA of the CC61 isolates from Portugal. The ribosomal binding site (RBS) is indicated in red and all the positions where independent SNPs were found are circled in dark green and numbered from one to 11. Asterisks indicate that there were two independent mutational events in that position.
(**b**) Mutations acquired in each of the 11 sites (green-coloured boxes), in relation to the core-genome phylogeny of the 128 CC61 isolates from Portugal.
(**c**) RT-qPCR results obtained with six CC61 isolates from Portugal (SA4, SA80 SA109, SA111, VSA66 and VSA104) and three strains used as control (COH1, 2603V/R and NEM316). Gene expression is represented as a ratio between the transcription level of *mtsB* and *mtsA*. Isolate names are indicated below each graph, together with the mutation acquired, as shown in panel (a), and the Gibbs free energy (ΔG) of the terminator structure predicted with mfold. Experiments were performed in triplicate with at least three independent cultures. Error bars represent standard deviation (SD) +/−. A two-tailed *t* test was performed by comparing each isolate against NEM316 (***$P < 0.001$; **$P < 0.01$).

VSA104) that acquired six different mutations within the stem-loop structure of the terminator (Fig. 4C), together with three strains carrying the ancestral sequence (COH1, 2603V/R and NEM316). Although in these control strains tested there was a consistent 30–40% decrease in the expression of *mtsB* after the terminator, none of the bovine CC61 isolates showed any decline in transcription levels ($P < 0.01$; Fig. 4C). By analyzing the impact of each variant on the RNA secondary structure, all the mutations were predicted to reduce the thermodynamic stability of the terminator structure in relation to the parental sequence (Fig. 4C). These results suggest that the mutations present in the CC61 bovine isolates we tested affect the structure of the terminator and lead to an increased expression of the downstream gene (*mtsB*), which encodes the ATP-binding protein of the iron/manganese transporter.

Cohabiting streptococci have contributed to the adaptation of this clone

We used the software Roary (Page *et al.*, 2015) to characterize both the core and accessory genome (i.e., the pan-genome) of the CC61 population. Overall, among the 128 CC61 isolates, a total of 1744 genes were identified as core (present in $\geq$ 99% of the isolates) and 1551 as accessory (missing from at least two isolates). All of the core genes were also present in closely related genomes (9857, 10058, 10059 and LDS610; Supporting Information

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

Fig. S1) and, therefore, before the divergence of this epidemic lineage in Portugal.

The majority of accessory genes were related to bacteriophages, integrative and conjugative elements or of unknown origin, so we focused on those with functional relevance, specifically for drug resistance and sugar metabolism. We detected independent acquisitions of genes involved in resistance to tetracycline (*tetO*), macrolides (*ermB*), streptomycin (*aadE*) and lincosamides (*lnuC*), as well as of two gene clusters involved in the production and resistance to the lantibiotics macedocin and nisin (Fig. 5). The gene cluster responsible for the biosynthesis and immunity to macedocin was acquired by one of the major sublineages of the CC61 Portuguese population (Fig. 5). By reconstructing the mobile genetic elements (MGEs) carrying the drug-related genes, we found that they were not exclusively present with a significant similarity in GBS, but were also detected among *Streptococcus dysgalactiae*, *Streptococcus suis* and *Streptococcus uberis* (Supporting Information Table S4). Interestingly, *S. dysgalactiae* and *S. uberis* are frequently associated with bovine mastitis. In the ST2-specific isolates, a similar analysis of the accessory and drug-related genes also showed the acquisition of *tetO*, *ermB*, *aadE* and the nisin operon by their MRCA (Supporting Information Fig. S5). Focusing on sugar metabolism, the Lac.2-2 variant of the lactose operon comprising *lacABCDFEGX* was conserved among all of the isolates studied. Additionally, a second copy of the Lac.2 operon (Lac.2-1), also found in *S. dysgalactiae* and *S. uberis* (Supporting Information Table S4), was acquired by two clades of the CC61 population (Fig. 5), but not by the ST2 isolates (Supporting Information Fig. S5).

Recurrent pseudogenizations reveal that adaptation to the bovine host is likely irreversible

We combined the data obtained from the phylogenetic and pan-genome analyses to detect specific functions that may have been lost during the adaptation of GBS to the bovine host. Variant detection and subsequent prediction of their functional effect identified 114 frameshift and 33 nonsense mutations generating a total of 119 pseudogenes (with an average of 12 per isolate).

Intriguingly, several independent mutations were found reoccurring at different loci. A total of 43 isolates acquired altogether seven nonsense or frameshift mutations within the *secA2-Y2* locus (Table 3). Within the *cps* operon, involved in the biosynthesis of the type II capsular polysaccharides, seven inactivating events affected a total of 56 isolates (Table 3 and Fig. 5). In the ST2 population we analyzed, loss or truncation of three genes of the *cps* operon was also detected (Supporting Information Fig. S5). The synthesis of this polysaccharide poses a significant nutritional cost and has also been shown to inhibit biofilm formation (Qin *et al*., 2013; Smitran *et al*., 2013). Additionally, four pseudogenes were detected within a region involved in glycogen biosynthesis (Table 3). Presumably, while colonizing the bovine milk – a sugar-rich environment – GBS might not require an active storage and synthesis of glycogen to survive and grow. Similarly, we observed four independent nonsense or frameshift mutations in the *opuC* operon, involved in the uptake of osmoprotectants which are recruited in high osmolarity conditions (Sleator *et al*., 2001) (Table 3). The milk has a moderate and consistent osmotic concentration (Jackson and Rothera, 1914), so this operon might cause an unnecessary fitness burden during colonization of the udder. Supporting this hypothesis, we also detected an independent frameshift mutation of the *opuCA* gene in the ST2 bovine isolates additionally studied.

Looking at the pan-genome distribution (Fig. 5), two independent events of gene loss were detected within a segment known as the lactose operon Lac.1, which has already been shown to be dispensable for the metabolism of lactose (Loughman and Caparon, 2007). Moreover, a 5-kb region encompassing an ABC transporter coupled with a two-component regulatory system (*vexp*) showed a progressing decay among the CC61 bovine strains studied (Fig. 5). Also in the ST2 isolates, independent loss and inactivation of this genomic island was detected (Supporting Information Fig. S5). Lastly, we found that among the GBS strains that acquired a genomic island for macedocin production and immunity, several isolates have lost part of this genomic cluster (Fig. 5). This suggests that the evolutionary pressure to maintain the functional integrity of this element might have diminished after contributing to the selection of their ancestral clone.

Discussion

Bovine mastitis caused by GBS continues to be a major veterinary and economic issue worldwide, representing a particularly prevalent problem in Portugal. We have performed a whole-genome analysis of GBS strains colonizing Portuguese dairy herds and discovered that they are infected almost exclusively by a single clone belonging to CC61. A Bayesian phylogenetic inference allowed us to estimate that this particular CC61 clone has been expanding in Portugal since at least the early 1990s (Supporting Information Fig. S3). The development of the agriculture sector, concurrent with Portugal's entry into the European Economic Community (EEC) in 1986, led to the expansion of dairy herds and the implementation of milking parlours in the country. This might have been conducive to the spread of the CC61 GBS clone we uncovered, and to its replacement of the existing GBS population.

Our whole-genome epidemiological analysis tracked bovine infections across multiple farms and within a single

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

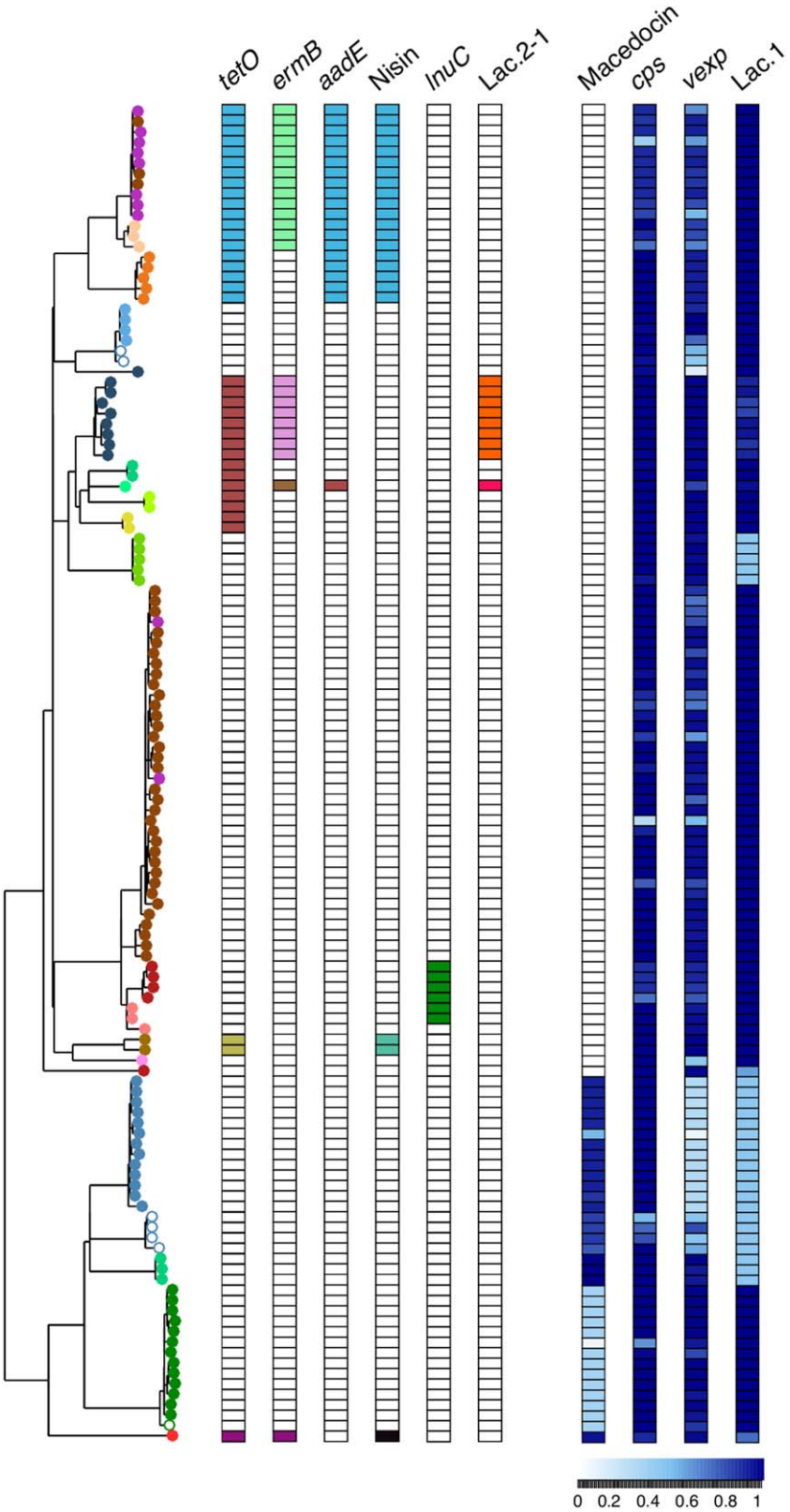

Fig. 5. Distribution of the most functionally relevant mobile genetic elements, involved in drug resistance (*tetO*, *ermB*, *aadE*, Nisin, *lnuC* and Macedocin) and sugar transport/metabolism (Lac.2-1 and Lac.1) in relation to the core-genome phylogeny of the 128 CC61 isolates from Portugal colour-coded as in Fig. 2. The ongoing degradation of the capsule operon (*cps*) and the *vexp* region are depicted on the right. White denotes absence, and for *tetO*, *ermB*, *aadE*, *nisin*, *lnuC* and Lac.2-1, equally coloured boxes mean they were detected within the same contiguous sequence. The genetic decay of the *cps*, *vexp* and the Lac.1 loci is depicted by a gradient from white to dark blue corresponding to a blast score ratio ranging from 0 to 1, as indicated in the figure key.

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

Table 3. Genomic regions most frequently inactivated.

Locus	Product	NS[a]	FS[b]	Isolates[c]
SA111_00292	L-carnitine/choline ABC transporter, OpuCD	0	1	11
SA111_00293	L-carnitine/choline ABC transporter, OpuCC	0	1	
SA111_00295	L-carnitine/choline ABC transporter, OpuCA	0	2	
SA111_00962	Glycogen debranching protein	0	2	17
SA111_00963	Glycogen branching enzyme, GlgB	0	1	
SA111_00964	Glucose-1-phosphate adenylyltransferase, GlgC	1	0	
SA111_00965	Glycogen biosynthesis protein GlgD	Missing[d]		
SA111_01329	Capsular polysaccharide repeat unit transporter CpsL	Missing[d]		56
SA111_01330	Capsular polysaccharide biosynthesis protein CpsK	0	1	
SA111_01333	Polysaccharide biosynthesis glycosyl transferase CpsJ	Missing[d]		
SA111_01334	Capsular biosynthesis protein CpsI	Missing[d]		
SA111_01335	Polysaccharide biosynthesis protein CpsH	Missing[d]		
SA111_01337	Polysaccharide biosynthesis protein CpsF	0	1	
SA111_01338	Galactosyl transferase CpsE	Missing[d],[e]		
SA111_01616	Glycosyltransferase GftA	1	0	43
SA111_01617	Protein export cytoplasm protein SecA2	1	1	
SA111_01619	Accessory secretory protein Asp2	0	1	
SA111_01621	Preprotein translocase SecY2 subunit	0	2	
SA111_01622	Glycosyl transferase, putative	1	0	

a. Number of nonsense mutations.
b. Number of frameshift mutations.
c. Number of isolates, out of the 128 CC61 representatives from Portugal that were sequenced, affected by at least one of the mutations.
d. Missing = loss/truncation of the gene.
e. Insertion sequence (IS) inserted within the gene in 34 isolates.

bovine host (Figs 2 and 3). Based on these observations, we suggest that the prevalence of this CC61 GBS lineage in Portugal is due both to persistent colonization of the same animal, as well as to its dissemination throughout the herd by cross-contamination. While GBS may be most frequently transferred during the milking process – through contamination of the milking apparatus – a recent study also found alternative environmental routes of transmission (Jorgensen *et al.*, 2016).

Distinguishing neutral from adaptive mutations is essential to understand how a pathogen adapts to its environment. dN/dS analyses of the CC61 isolates that expanded in Portugal underlined a general level of strong purifying selection. This suggests that the bovine environment imposes significant selective constraints and is quickly purging stochastic mutations that diminish the fitness of the CC61 population. Therefore, we reason that the ubiquitous prevalence of this epidemic clone in Portugal is most likely a result of its competitive advantage in the bovine host, rather than a product of chance. Conversely, the frequent occurrence of nonsynonymous mutations among major transcriptional regulators suggests these specific changes might be providing a competitive advantage during colonization of the bovine udder and milk (Table 1). Indeed, regulatory systems like CovRS are known to affect pathogenesis by controlling the expression of virulence-associated genes in response to different environmental stimuli (Lamy *et al.*, 2004; Santi *et al.*, 2009; Almeida *et al.*, 2015). Of special interest were the functionally significant mutations detected within an iron/manganese transporter operon that independently affected all of the isolates studied (Fig. 4). Trace metals are naturally found in low concentrations in the bovine milk (Lonnerdal *et al.*, 1981). Additionally, during the dry period of the bovine lactation cycle, the increased concentration of lactoferrin in the mammary gland further depletes the milk of metal nutrients essential for bacterial growth and survival (Hurley and Rejman, 1993). Thus, the mutations we observed might contribute to a higher uptake of iron and manganese by GBS, potentially underlying an adaptive strategy for this dominant clone to persist in more challenging conditions. Supporting this hypothesis, the acquisition of manganese during growth in milk was shown to be essential for infection of the bovine mammary gland by *S. uberis* (Smith *et al.*, 2003).

Mobile genetic elements (MGEs) have been shown to play a major role in the evolution of GBS (Brochet *et al.*, 2008a; Richards *et al.*, 2011), while representing a genomic record of its interaction with other neighbouring bacteria. The accessory genes shared with other mastitis-causing streptococci reveal the contribution of the genetic repertoire of bacterial species circulating in the bovine environment to the evolutionary success of the CC61 clone (Fig. 5 and Supporting Information Table S4). In particular, the acquisition of genes involved in the synthesis and resistance to macedocin correlates with the expansion of one of the two major clades of the CC61 population in Portugal (Fig. 5). Macedocin is known to inhibit the growth of a wide range of lactic acid bacteria and several pathogenic

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

species (Georgalaki *et al*., 2002). The genomic island specifically acquired by the CC61 isolates is similar to that first described in *Streptococcus macedonicus* (Papadelli *et al*., 2007), a species normally isolated from fermented milk products. Therefore, macedocin production and resistance might have contributed to the selective advantage of this clone and to its replacement of the local GBS population.

How the adaptation of GBS differs within humans and bovines has been largely unknown, albeit crucial to understand the risk of interspecies transmission. The long-term persistence of this dominant CC61 clone allowed us to uncover the genomic footprint of GBS's adaptation to the bovine host. The faster evolutionary rate that was estimated for the bovine-adapted population probably stems from the different conditions encountered by GBS in humans and cattle. In the milk, a more nutrient-rich and bacteria-poor environment, GBS might display a faster growth rate, while the milking process promotes the continuous growth of the bacteria by regularly replenishing their natural medium. Evolutionary studies have shown that host specialization may be reflected by reductive evolutionary processes associated with gene loss and inactivation (Toft and Andersson, 2010). Bovine-specific strains analyzed in this work showed a directed and recurrent pseudogenization of multiple genes, such as the *secA2-Y2*, *cps* and *vexp* loci. The *secA2-Y2* system was shown in GBS to be involved in the synthesis and export of the serine-rich glycoprotein Srr1, which promotes bacterial adhesion to human lung epithelial cells (Mistou *et al*., 2009). Moreover, the *vexp* region is highly conserved exclusively among human GBS strains and was suggested to play a role in their adaptation to the human gut (Brochet *et al*., 2008b). Therefore, these data reflect distinct requirements for the adaptation of GBS to humans and bovines. The loss of human-adapted traits might not allow GBS to as efficiently invade and recolonize the human host, suggesting that interspecies transmission is more permissive from humans to cattle. In addition, our results were further supported by the occurrence of a similar evolutionary trend in the ST2 clone dominant in Galicia, which represents a recent adaptation to the bovine environment (Supporting Information Fig. S2).

Our study illustrates an intriguing example of the differential host adaptation of a generalist species, and how whole-genome analysis may be leveraged to understand its underlying potential for causing disease in different environments. These results will also be instrumental to eradicate this dominant CC61 clone and prevent its ongoing dissemination.

Experimental procedures

Sequencing, genome assembly and annotation

A total of 234 GBS isolates collected from Portugal, Spain and France were analyzed in this study (Supporting Information Table S1). The CRISPR1 locus of all GBS isolates was sequenced and analyzed as previously described (Lopez-Sanchez *et al*., 2012). One hundred and fifty GBS isolates (131 from CC61 and 19 from ST2) were then chosen for whole-genome sequencing, based on the diversity of CRISPR spacer profiles identified, as well as on their temporal and geographical distribution (Supporting Information Tables S1 and S2). Genomes were sequenced using the Illumina HiSeq 2000 platform with single- or paired-end read runs of 101 bp. Reads were filtered for quality and genome sequences were assembled using the Velvet software (Zerbino and Birney, 2008) with an optimized k-mer length, a minimum coverage of 10 and a contig length of at least 200 bp. Strain SA111 was additionally selected for single molecule real-time sequencing (PacBio RS II system). PacBio subreads were assembled with both PBcR (Berlin *et al*., 2015) and the RS_HGAP_Assembly.3 protocol from the SMRT analysis toolkit v2.3. A finished assembly was achieved by complementing the predictions obtained by both tools. Consensus accuracy was further polished using Quiver (Chin *et al*., 2013) and any remaining indels within homopolymer regions were manually corrected with the corresponding Illumina reads. Assembled genomes were annotated with Prokka (Seemann, 2014) and global pan-genome analyses were carried out using Roary (Page *et al*., 2015).

Sequencing reads and corresponding genome assemblies have been deposited in the EMBL nucleotide sequence database (http://www.ebi.ac.uk/ena) under study accession number PRJEB12926. For the accession numbers of the individual samples, see Supporting Information Table S1.

Genome mapping, variant calling and phylogenetic inference

Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009) was used to map reads of each CC61 isolate against the complete genome of SA111, and of each ST2 isolate against the draft assembly of VSA10. Variant calling was performed with Genome Analysis ToolKit (GATK) (McKenna *et al*., 2010) according to the published recommendations (Van der Auwera *et al*., 2013; DePristo *et al*., 2011).

Phylogenetic trees of the CC61 and ST2 isolates studied in combination with publicly available GBS genomes were built from the core-genome alignment of their assemblies with Parsnp (Treangen *et al*., 2014) (Supporting Information Figs S1 and S2). Genomes with no more than one allele difference in relation to the ST61 profile were considered CC61. Phylogeny of the CC61 isolates from Portugal (Fig. 2) was inferred from the polymorphic positions detected in the variant calling workflows. Recombinant sites were removed following identification with Gubbins (Croucher *et al*., 2015). Variants present in the accessory genome of these strains, determined using the filter_BSR_variome.py script from the Large Scale BLAST score ratio (LS-BSR) pipeline (Sahl *et al*., 2014), were also not considered. Maximum-likelihood (ML) phylogenies were inferred with RAxML (Stamatakis, 2014), using a General Time-Reversible (GTR) substitution model with a gamma-distributed rate across sites combined with an ascertainment bias to take into account the sole use of variable positions in the alignment. ML trees were bootstrapped with 1000 replicates.

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

To investigate the temporal evolution of the GBS isolates included in this study, the Bayesian phylogenetic software BEAST v1.8.2 (Drummond *et al.*, 2012) was used to calibrate the evolutionary rate with the corresponding sampling date of each isolate (Supporting Information Table S1). We used a GTR substitution model considering a gamma-distributed rate across sites with a proportion of invariant positions. To identify the most suitable tree and clock models, we compared the strict, uncorrelated lognormal relaxed and uncorrelated exponential relaxed clock models together with coalescent constant, exponential growth, expansion growth and Bayesian skyline tree models. BEAST runs for model testing were conducted in duplicate for 50 million Markov Chain Monte Carlo (MCMC) generations with samples taken every 1000 generations. The best model was deduced by marginal likelihood estimation with stepping-stone/path sampling. The constant size tree model and the uncorrelated exponential relaxed clock model were preferred. Final phylogenetic analysis was run for 100 million MCMC generations with 10% burn-in using the above parameters. Final run presented good mixing and convergence with effective sample size (ESS) values above 200 for all parameters. The presence of temporal structure was further validated by comparison with 10 simulated datasets with randomized sampling dates, as previously described (Firth *et al.*, 2010).

Variant annotation, parallel evolution and dN/dS

The functional effect of each SNP detected within coding sequences was classified with snpEff (Cingolani *et al.*, 2012) as either nonsynonymous (N) or synonymous (S). In the case of indels, only frameshift mutations were considered. To detect parallel evolution in a given number of mutations observed (m_t), 1000 simulations of m_t in the reference genome were performed to establish an expected distribution of the number of mutations per gene. One-tailed *P* values were calculated by assessing the frequency in which the number of mutations per gene observed was higher than the simulated expectation. Subsequently, the same analysis was performed while only taking into account intergenic regions. Genes and intergenes with a low SNP density, i.e. fewer than one SNP per their average length, were excluded. For dN/dS calculations, the observed spectrum of *N* and *S* mutations per nucleotide change was normalized by an expected frequency simulated for the whole genome, as previously described (Lieberman *et al.*, 2014). Values above 1 are indicative of positive selection. Clopper–Pearson confidence intervals (CI) were calculated by the binomial test.

RT-qPCRs

Bacterial cultures were grown to the exponential growth phase (OD_{600} = 0.4–0.5) in 15 ml Todd Hewitt (TH) broth at 37°C. RNA extraction, reverse transcription and RT-qPCRs were performed as previously described (Lamy *et al.*, 2004), using the primers indicated in Supporting Information Table S5. Relative gene expression was quantified with a standard curve-based method, in which a regression analysis was performed using serial dilutions of genomic DNA from strain VSA104. Each assay was performed with three experimental replicates starting from at least three independent cultures. A two-tailed *t*-test was carried out to determine whether the expression differences were statistically significant. RNA secondary structures and thermodynamic stabilities were predicted with mfold (http://unafold.rna.albany.edu/?q = mfold).

Acknowledgements

This work was supported by ANR-LabEx project IBEID. Partial support was also obtained through UCIBIO (UID/Multi/04378/2013 and POCI-01-0145-FEDER-007728) and Project PTDC/CVT-EPI/6685/2014. Sequencing was performed at the Pasteur Genopole, a member of France Génomique (ANR10-IBNS-09-08). A.A. is a scholar in the Pasteur – Paris University (PPU) International PhD program and received a stipend from the ANR-LabEx IBEID. Authors would like to thank Laurence Ma for her help in performing the Illumina sequencing. They also thank SEGALAB, the Laboratorio de Sanidade e Produción Animal de Galicia, Xunta de Galicia and Sophie Payot for providing isolates used in this study. Authors also thank Isabelle Rosinski-Chupin, Pierre-Emmanuel Douarre, Niza Ribeiro, Bruno Gonzalez-Zorn and Helena Madeira for fruitful discussions.

References

Almeida, A., Albuquerque, P., Araujo, R., Ribeiro, N., and Tavares, F. (2013) Detection and discrimination of common bovine mastitis-causing streptococci. *Vet Microbiol* **164:** 370–377.

Almeida, A., Villain, A., Joubrel, C., Touak, G., Sauvage, E., Rosinski-Chupin, I., *et al.* (2015) Whole-genome comparison uncovers genomic mutations between group B streptococci sampled from infected newborns and their mothers. *J Bacteriol* **197:** 3354–3366.

Berlin, K., Koren, S., Chin, C.S., Drake, J.P., Landolin, J.M., and Phillippy, A.M. (2015) Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nat Biotechnol* **33:** 623.

Bi, Y.L., Wang, Y.J., Qin, Y., Vallverdu, R.G., Garcia, J.M., Sun, W., *et al.* (2016) Prevalence of bovine mastitis pathogens in bulk tank milk in China. *PLos One* **11:** 13.

Bisharat, N., Crook, D.W., Leigh, J., Harding, R.M., Ward, P.N., Coffey, T.J., *et al.* (2004) Hyperinvasive neonatal group B *Streptococcus* has arisen from a bovine ancestor. *J Clin Microbiol* **42:** 2161–2167.

Bray, B.A., Sutcliffe, I.C., and Harrington, D.J. (2009) Expression of the MtsA lipoprotein of *Streptococcus agalactiae* A909 is regulated by manganese and iron. *Antonie Van Leeuwenhoek* **95:** 101–109.

Brochet, M., Couve, E., Glaser, P., Guedon, G., and Payot, S. (2008a) Integrative conjugative elements and related elements are major contributors to the genome diversity of *Streptococcus agalactiae*. *J Bacteriol* **190:** 6913–6917.

Brochet, M., Rusniok, C., Couve, E., Dramsi, S., Poyart, C., Trieu-Cuot, P., *et al.* (2008b) Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of *Streptococcus agalactiae*. *Proc Natl Acad Sci USA* **105:** 15961–15966.

Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., *et al.* (2013) Nonhybrid, finished microbial

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

genome assemblies from long-read SMRT sequencing data. *Nat Methods* **10:** 563.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Tung, N., Wang, L., *et al*. (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* **6:** 80–92.

Croucher, N.J., Page, A.J., Connor, T.R., Delaney, A.J., Keane, J.A., Bentley, S.D., *et al*. (2015) Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* **43.**

Da Cunha, V., Davies, M.R., Douarre, P.E., Rosinski-Chupin, I., Margarit, I., Spinali, S., *et al*. (2014) *Streptococcus agalactiae* clones infecting humans were selected and fixed through the extensive use of tetracycline. *Nat Commun* **5:** 11.

Deb, R., Kumar, A., Chakraborty, S., Verma, A.K., Tiwari, R., Dhama, K., *et al*. (2013) Trends in diagnosis and control of bovine mastitis: A review. *Pakistan J Biol Sci* **16:** 1653–1661.

DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., *et al*. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43:** 491.

Dermer, P., Lee, C., Eggert, J., and Few, B. (2004) A history of neonatal group B *Streptococcus* with its related morbidity and mortality rates in the United States. *J Pediatric Nurs* **19:** 357–363.

Drummond, A.J., Suchard, M.A., Xie, D., and Rambaut, A. (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **29:** 1969–1973.

Firth, C., Kitchen, A., Shapiro, B., Suchard, M.A., Holmes, E.C., and Rambaut, A. (2010) Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. *Mol Biol Evol* **27:** 2038–2051.

Georgalaki, M.D., Van den Berghe, E., Kritikos, D., Devreese, B., Van Beettmen, J., Kalantzopoulos, G., *et al*. (2002) Macedocin, a food-grade lantibiotic produced by *Streptococcus macedonicus* ACA-DC 198. *Appl Environ Microbiol* **68:** 5891–5903.

Haenni, M., Saras, E., Bertin, S., Leblond, P., Madec, J.Y., and Payot, S. (2010) Diversity and mobility of integrative and conjugative elements in bovine isolates of *Streptococcus agalactiae*, *S. dysgalactiae subsp. dysgalactiae*, and *S. uberis*. *Appl Environ Microbiol* **76:** 7957–7965.

Heikkila, A.M., Nousiainen, J.I., and Pyorala, S. (2012) Costs of clinical mastitis with special reference to premature culling. *J Dairy Sci* **95:** 139–150.

Hurley, W.L., and Rejman, J.J. (1993) Bovine lactoferrin in involuting mammary tissue. *Cell Biol Int* **17:** 283–289.

Jackson, L.C., and Rothera, A.C. (1914) Milk – Its milk sugar, conductivity and depression of freezing point. *Biochem J* **8:** 1–27.

Jones, N., Bohnsack, J.F., Takahashi, S., Oliver, K.A., Chan, M.S., Kunst, F., *et al*. (2003) Multilocus sequence typing system for group B *Streptococcus*. *J Clin Microbiol* **41:** 2530–2536.

Jorgensen, H.J., Nordstoga, A.B., Sviland, S., Zadoks, R.N., Solverod, L., Kvitle, B., and Mork, T. (2016) *Streptococcus agalactiae* in the environment of bovine dairy herds – Rewriting the textbooks? *Vet Microbiol* **184:** 64–72.

Kalmus, P., Aasmaee, B., Kaerssin, A., Orro, T., and Kask, K. (2011) Udder pathogens and their resistance to antimicrobial agents in dairy cows in Estonia. *Acta Vet Scand* **53.**

Keefe, G.P. (1997) *Streptococcus agalactiae* mastitis: A review. *Can Vet J* **38:** 429–437.

Klimiene, I., Ruzauskas, M., Spakauskas, V., Matusevicius, A., Mockeliunas, R., Pereckiene, A., *et al*. (2011) Antimicrobial resistance patterns to beta-lactams of gram-positive cocci isolated from bovine mastitis in Lithuania. *Polish J Vet Sci* **14:** 467–472.

Lamy, M.C., Zouine, M., Fert, J., Vergassola, M., Couve, E., Pellegrini, E., *et al*. (2004) CovS/CovR of group B *Streptococcus*: A two-component global regulatory system involved in virulence. *Mol Microbiol* **54:** 1250–1268.

Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25:** 1754–1760.

Lieberman, T.D., Michel, J.B., Aingaran, M., Potter-Bynoe, G., Roux, D., Davis, M.R., Jr, *et al*. (2011) Parallel bacterial evolution within multiple patients identifies candidate pathogenicity genes. *Nat Genet* **43:** 1275.

Lieberman, T.D., Flett, K.B., Yelin, I., Martin, T.R., McAdam, A.J., Priebe, G.P., and Kishony, R. (2014) Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. *Nat Genet* **46:** 82.

Lonnerdal, B., Keen, C.L., and Hurley, L.S. (1981) Iron, copper, zinc, and manganese in milk. *Annu Rev Nutr* **1:** 149–174.

Lopez-Sanchez, M.J., Sauvage, E., Da Cunha, V., Clermont, D., Hariniaina, E.R., Gonzalez-Zorn, B., *et al*. (2012) The highly dynamic CRISPR1 system of *Streptococcus agalactiae* controls the diversity of its mobilome. *Mol Microbiol* **85:** 1057–1071.

Loughman, J.A., and Caparon, M.G. (2007) Comparative functional analysis of the *lac* operons in *Streptococcus pyogenes*. *Mol Microbiol* **64:** 269–280.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., *et al*. (2010) The Genome Analysis ToolKit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20:** 1297–1303.

Mistou, M.Y., Dramsi, S., Brega, S., Poyart, C., and Trieu-Cuot, P. (2009) Molecular dissection of the *secA2* locus of group B *Streptococcus* reveals that glycosylation of the Srr1 LPXTG protein is required for full virulence. *J Bacteriol* **191:** 4195–4206.

Mweu, M.M., Nielsen, S.S., Halasa, T., and Toft, N. (2012) Annual incidence, prevalence and transmission characteristics of *Streptococcus agalactiae* in Danish dairy herds. *Prev Vet Med* **106:** 244–250.

Oliveira, I.C.M., de Mattos, M.C., Pinto, T.A., Ferreira-Carvalho, B.T., Benchetrit, L.C., Whiting, A.A., *et al*. (2006) Genetic relatedness between group B streptococci originating from bovine mastitis and a human group B *Streptococcus* type V cluster displaying an identical pulsed-field gel electrophoresis pattern. *Clin Microbiol Infect* **12:** 887–893.

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., *et al*. (2015) Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31:** 3691–3693.

Papadelli, M., Karsioti, A., Anastasiou, R., Georgalaki, M., and Tsakalidou, E. (2007) Characterization of the gene

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

cluster involved in the biosynthesis of macedocin, the lantibiotic produced by *Streptococcus macedonicus*. *FEMS Microbiol Lett* **272:** 75–82.
Qin, L., Kida, Y., Imamura, Y., Kuwano, K., and Watanabe, H. (2013) Impaired capsular polysaccharide is relevant to enhanced biofilm formation and lower virulence in *Streptococcus pneumoniae*. *J Infect Chemother* **19:** 261–271.
Rato, M.G., Bexiga, R., Florindo, C., Cavaco, L.M., Vilela, C.L., and Santos-Sanches, I. (2013) Antimicrobial resistance and molecular epidemiology of streptococci from bovine mastitis. *Vet Microbiol* **161:** 286–294.
Richards, V.P., Lang, P., Bitar, P.D.P., Lefebure, T., Schukken, Y.H., Zadoks, R.N., and Stanhope, M.J. (2011) Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted *Streptococcus agalactiae*. *Infect Genet Evol* **11:** 1263–1275.
Richards, V.P., Choi, S.C., Bitar, P.D.P., Gurjar, A.A., and Stanhope, M.J. (2013) Transcriptomic and genomic evidence for *Streptococcus agalactiae* adaptation to the bovine environment. *BMC Genom* **14:** 15.
Rosinski-Chupin, I., Sauvage, E., Sismeiro, O., Villain, A., Da Cunha, V., Caliot, M.E., *et al*. (2015) Single nucleotide resolution RNA-seq uncovers new regulatory mechanisms in the opportunistic pathogen *Streptococcus agalactiae*. *BMC Genom* **16.**
Sahl, J.W., Caporaso, J.G., Rasko, D.A., and Keim, P. (2014) The large-scale blast score ratio (LS-BSR) pipeline: A method to rapidly compare genetic content between bacterial genomes. *Peerj* **2.**
Santi, I., Grifantini, R., Jiang, S.M., Brettoni, C., Grandi, G., Wessels, M.R., and Soriani, M. (2009) CsrRS regulates group B *Streptococcus* virulence gene expression in response to environmental pH: A new perspective on vaccine development. *J Bacteriol* **191:** 5387–5397.
Schuchat, A. (1998) Epidemiology of group B streptococcal disease in the United States: Shifting paradigms. *Clin Microbiol Rev* **11:** 497.
Seemann, T. (2014) Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30:** 2068–2069.
Sleator, R.D., Wouters, J., Gahan, C.G.M., Abee, T., and Hill, C. (2001) Analysis of the role of OpuC, an osmolyte transport system, in salt tolerance and virulence potential of *Listeria monocytogenes*. *Appl Environ Microbiol* **67:** 2692–2698.
Smith, A.J., Ward, P.N., Field, T.R., Jones, C.L., Lincoln, R.A., and Leigh, J.A. (2003) MtuA, a lipoprotein receptor antigen from *Streptococcus uberis*, is responsible for acquisition of manganese during growth in milk and is essential for infection of the lactating bovine mammary gland. *Infect Immunity* **71:** 4842–4849.
Smitran, A., Opavski, N.V., Eric-Marinkovic, J., Gajic, I., and Ranin, L. (2013) Adherence and biofilm production of invasive and non-invasive isolates of *Streptococcus pyogenes* after hyaluronidase treatment. *Arch Biol Sci* **65:** 1353–1361.
Sorensen, U.B., Poulsen, K., Ghezzo, C., Margarit, I., and Kilian, M. (2010) Emergence and global dissemination of host-specific *Streptococcus agalactiae* clones. *mBio* **1.**
Springman, A.C., Lacher, D.W., Waymire, E.A., Wengert, S.L., Singh, P., Zadoks, R.N., *et al*. (2014) Pilus distribution among lineages of group B *Sreptococcus*: An evolutionary and clinical perspective. *BMC Microbiol* **14.**
Stamatakis, A. (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30:** 1312–1313.
Toft, C., and Andersson, S.G.E. (2010) Evolutionary microbial genomics: Insights into bacterial host adaptation. *Nat Rev Genet* **11:** 465–475.
Treangen, T.J., Ondov, B.D., Koren, S., and Phillippy, A.M. (2014) The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol* **15.**
Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., *et al*. (2013) From fastQ data to high-confidence variant calls: The Genome Analysis ToolKit best practices pipeline. In *Current Protocols in Bioinformatics*. Alex Bateman, Sorin Draghici, Ekta Khurana, Sandra Orchard and William R. Pearson (eds.). Wiley: Hoboken, NJ, USA. pp. 11.10.1-11.10.33.
Wyder, A.B., Boss, R., Naskova, J., Kaufmann, T., Steiner, A., and Graber, H.U. (2011) *Streptococcus* spp. and related bacteria: Their identification and their pathogenic potential for chronic mastitis – A molecular approach. *Res Vet Sci* **91:** 349–357.
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.
Zerbino, D.R., and Birney, E. (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18:** 821–829.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Core-genome phylogeny of CC61. The ML phylogeny was built using RAxML (Stamatakis, 2014), based on a total of 4524 polymorphic positions. Tree depicts all the publicly available CC61 genomes, as well as the 128 CC61 representative isolates from Portugal analyzed in this work and the three additional isolates sampled from the southwest of France (9857, 10058, 10059). Sequence type (ST) and location of each isolate (if available) are plotted in the tree as depicted in the figure key. All genomes depicted are of bovine origin, except for ATCC13813, which was isolated from the oral cavity of a human host. All tree nodes presented a bootstrap support $>75\%$. Tree was rooted with the CC1 genome of GBS 09mas01883 (not shown).
Fig. S2. Core-genome phylogeny of ST2 genomes. ML phylogenetic tree built using RAxML (Stamatakis, 2014) based on a total of 1403 SNPs. Tree depicts all the publicly available ST2 genomes, together with the 19 ST2 isolates from Portugal and Spain analyzed in this work. Host origin and location of each isolate (if available) are plotted in the tree as depicted in the figure key. All tree nodes presented a bootstrap support $>75\%$. Tree was rooted with the CC19 genome of GBS 2603V/R (not shown).
Fig. S3. Timed phylogeny of the CC61 isolates. Bayesian phylogenetic inference of the 128 CC61 isolates sequenced from Portugal and the three additional CC61 isolates from France, performed with BEAST (Drummond *et al*., 2012). The CC61 isolates are colour-coded as in Fig. 2. Interval

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,

dates in brackets represent 95% HPDs of the estimated date of each node.

Fig. S4. Parallel evolution in genes and in intergenic regions. Number of genes (**a**) or intergenic regions (**b**) with at least *m* mutations, as a function of *m*. Expected values correspond to the average of 1000 random distributions of the number of mutations observed. Error bars represent standard deviation (SD) +/−. $^{***}P < 0.001$; $^{**}P < 0.01$; $^{*}P < 0.05$.

Fig. S5. Analysis of most functionally relevant accessory genes identified in the CC61 population, against the core-genome phylogeny of the 19 ST2 isolates from Portugal and Spain. Isolates are colour-coded as in Supporting Information Fig. S2. White denotes absence, and for *tetO*, *ermB* and *aadE*, different coloured boxes mean they were detected within distinct contiguous sequences. Similarities with the nisin operon, *cps*, *vexp* and Lac.1 loci are depicted by a gradient from white to dark blue corresponding to a blast score ratio ranging from 0 to 1, as indicated in the figure key.

Table S1. *S. agalactiae* strains used in this study.

Table S2. CRISPR spacer profiles of the *S. agalactiae* strains used in this study.

Table S3. Genes that acquired more than three independent mutations.

Table S4. BLASTn results of the drug resistance and metabolism genes acquired by the CC61 clone.

Table S5. Primer pairs used in the RT-qPCRs.

© 2016 The Authors. Environmental Microbiology published by Society for Applied Microbiology and John Wiley & Sons Ltd.,