



OPEN

Increasing incidence of group B streptococcus neonatal infections in the Netherlands is associated with clonal expansion of CC17 and CC23

Dorota Jamrozy¹✉, Merijn W. Bijlsma², Marcus C. de Goffau^{1,7}, Diederik van de Beek², Taco W. Kuijpers^{3,4}, Julian Parkhill^{1,7}, Arie van der Ende^{5,6} & Stephen D. Bentley¹

Group B streptococcus (GBS) is the leading cause of neonatal invasive disease worldwide. In the Netherlands incidence of the disease increased despite implementation of preventive guidelines. We describe a genomic analysis of 1345 GBS isolates from neonatal (age 0–89 days) invasive infections in the Netherlands reported between 1987 and 2016. Most isolates clustered into one of five major lineages: CC17 (39%), CC19 (25%), CC23 (18%), CC10 (9%) and CC1 (7%). There was a significant rise in the number of infections due to isolates from CC17 and CC23. Phylogenetic clustering analysis revealed that this was caused by expansion of specific sub-lineages, designated CC17-A1, CC17-A2 and CC23-A1. Dating of phylogenetic trees estimated that these clones diverged in the 1960s/1970s, representing historical rather than recently emerged clones. For CC17-A1 the expansion correlated with acquisition of a new phage, carrying gene encoding a putative cell-surface protein. Representatives of CC17-A1, CC17-A2 and CC23-A1 clones were identified in datasets from other countries demonstrating their global distribution.

Group B streptococcus (GBS), also known as *Streptococcus agalactiae*, is a commensal bacterium that colonizes the human gastrointestinal and genital tracts¹. It is also an opportunistic pathogen and a leading cause of neonatal invasive disease worldwide². GBS disease in neonates occurs predominantly during the first 90 days of life and the most common clinical syndromes are septicaemia and meningitis^{3,4}. The infection is defined as either early-onset disease (EOD, days 0–6) or late-onset disease (LOD, days 7–89)⁵. EOD results from intrapartum GBS transmission from a colonized mother whereas routes of GBS acquisition in LOD are less well defined and can involve vertical, nosocomial or community transmission⁶.

Since the 1990s many high-income countries have implemented policies for prevention of neonatal invasive GBS disease involving intrapartum antibiotic prophylaxis (IAP), which is effective in reducing the incidence of EOD⁶. Subsequently, the rates of EOD have declined substantially in a number of countries^{7–10}. However, the challenge of reducing the burden of GBS disease in newborns remains. IAP has no effect on the incidence of LOD rates, which have remained stable or increased^{7,9}. Moreover, increased incidence of both EOD and LOD has been recently reported in some countries despite their implementation of national disease prevention guidelines,

¹Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK. ²Department of Neurology, Amsterdam Neuroscience, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, Netherlands.

³Department of Immunopathology, Sanquin Research and Landsteiner Laboratory of the Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands. ⁴Department of Paediatric Haematology, Immunology and Infectious Diseases, Emma Children's Hospital, Amsterdam University Medical Center, Amsterdam, Netherlands.

⁵Department of Medical Microbiology, Amsterdam Infection and Immunity, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, Netherlands. ⁶Netherlands Reference Laboratory for Bacterial Meningitis, Center of Infection and Immunity Amsterdam, Amsterdam University Medical Center, Amsterdam, Netherlands.

⁷Present address: Department of Veterinary Medicine, University of Cambridge, Cambridge, UK. ✉e-mail: dorota.jamrozy@sanger.ac.uk

indicating that reassessment of their current prevention strategies is needed^{4,11}. This includes the Netherlands, the UK and Ireland, where IAP is offered based on the presence of clinical risk factors, in contrast to a more widely implemented culture-based screening for GBS carriage during pregnancy¹². The reasons for this rise in disease incidence are poorly understood. Possible explanations include changes in the host, medical practice or the pathogen's population. In the Netherlands, a shift in the prevalence of certain GBS clonal complexes (CC) was reported following the implementation of IAP policy in 1999, with an increase in the number of cases caused by isolates representing CC17¹¹. Isolates from CC17 are overrepresented among invasive GBS isolates from infections in newborns and are particularly associated with meningitis and LOD cases with CC17 often referred to as a hypervirulent clone³.

GBS isolates are commonly genotyped using multi-locus sequence typing (MLST), which allows to group GBS isolates into sequence types (STs) that can then be clustered into major CC groups. MLST revealed that the GBS population has a relatively low genetic diversity and most isolates belong to a small number of dominant STs¹³. This is replicated when inferring species population structure of GBS using whole genome sequence-based phylogeny, which splits GBS into well-resolved lineages that represent major CCs¹⁴. Genomic analysis allows a higher level of isolate discrimination, offering further insights into population structure of individual GBS lineages. Genome-based study of a geographically diverse GBS collection reported that most isolates within major CCs cluster into one or two dominant clonal groups, suggesting that the contemporary GBS population structure has been shaped by evolutionary bottlenecks and clonal expansion events¹⁴. Beyond molecular epidemiology, genomic data also offers insights into mechanisms of GBS evolution and host adaptation. In this study we examined the genomes of 1345 GBS isolates from neonatal invasive disease in the Netherlands collected between 1987 and 2016 to investigate changes in pathogen population during a previously reported rise in disease incidence¹¹.

Results

Genotypic characteristics of Dutch GBS from neonatal invasive disease. Clustering of isolates using hierBAPS was highly concordant with MLST and CC assignment correlated with GBS population structure revealed by a core genome-based phylogeny (Supplementary Fig. S1). Of the 1345 GBS isolates, the majority ($n = 1321$, 98%) clustered into five dominant human-associated lineages. The most common group was CC17 (526/1345; 39%), followed by CC19 (332/1345; 25%), CC23 (239/1345; 18%), CC10 (127/1345; 9%), and CC1 (97/1345; 7%). The remaining isolates (24/1345; 2%) were assigned to the hierBAPS bin group, a polyphyletic cluster of low-frequency genotypes. It consisted of isolates representing 7 different STs (Supplementary Data S1). Four of the STs were phylogenetically distinct and represented CC22, CC26, CC103 and CC130. The remaining three STs in the bin group were ST23, ST88 and ST280 that formed a monophyletic cluster most closely related to the CC23 clade.

The majority of GBS isolates were derived from EOD (830/1345; 62%). EOD isolates were most common in CC1 (79/97; 81%), followed by CC10 (97/127; 76%), CC23 (173/239; 72%) and CC19 (216/332; 65%). CC17 was represented by a significantly lower proportion of EOD isolates (248/526, 47%) in comparison with all other CCs (pairwise chi-squared test with Bonferroni correction; CC17: CC1, p value = 1.03×10^{-8} ; CC17: CC10, p value = 5.77×10^{-8} ; CC17: CC19, p value = 4.25×10^{-6} ; CC17: CC23, p value = 1.32×10^{-9}). Consequently, CC17 had the highest proportion of LOD isolates (278/526, 53%), followed by CC19 (116/332, 35%), CC23 (66/239, 28%), CC10 (30/127, 24%) and CC1 (18/97, 19%).

This was also reflected in proportion of different CCs within each disease onset group. More than half of LOD isolates were CC17 (278/515, 54%), followed by CC19 (116/515, 23%), CC23 (66/515, 13%), CC10 (30/515, 6%) and CC1 (18/515, 3%). In contrast, CC17 represented only around a third of EOD isolates (248/830, 30%), followed by CC19 (216/830, 26%), CC23 (173/830, 21%), CC10 (97/830, 12%) and CC1 (79/830, 10%).

A key marker for monitoring the molecular epidemiology of GBS disease is the capsular polysaccharide (CPS) serotype. The most common CPS serotype in this collection was III (817/1345; 61%), followed by Ia (262/1345; 19%). Other serotypes were represented by up to 6% of isolates, which included Ib (84/1345; 6%), II (76/1345; 6%), V (64/1345; 5%), IV (29/1345; 2%), VI (5/1345; 0.4%), IX (4/1345; 0.3%) and VII (2/1345; 0.1%) (Supplementary Data S1). All but one CC17 isolate carried CPS serotype III (525/526; 99%), which was also observed in the majority of CC19 isolates (278/332; 84%). All but two CC23 isolates carried serotype Ia (237/239; 99%). Isolates from CC1 and CC10 revealed high serotype heterogeneity with eight and six different serotypes, respectively (Supplementary Data S1).

Recent studies reported a rise in the prevalence of antimicrobial resistance among clinical GBS isolates, in particular to macrolides, which was associated mostly with CC1 and CC19^{15–17}. In our collection, 8% of isolates carried at least one macrolide-lincosamide-streptogramin (MLS) resistance gene (*ermB*, *ermTR*, *mefA*, *msrD*, *lsaC*, *lsaE* or *lnuB*; Supplementary Data S1). Carriage of MLS resistance genes was highest in CC1 (19/97; 20%), followed by CC19 (30/332; 9%), CC17 (43/526; 8%), CC23 (12/239; 5%) and CC10 (5/127; 4%). Also of concern is the emergence of GBS with reduced susceptibility to beta-lactams, the first line antibiotics for IAP and treatment of GBS infections¹⁷. Here, we identified two isolates with mutations in the *pbp2x* gene that were previously associated with reduced susceptibility to beta-lactams (T394I and Q557E)¹⁸, both representing CC19. Human-associated GBS isolates often show a high prevalence of tetracycline resistance genes due to clonal expansion of GBS lineages that acquired mobile elements carrying *tet* determinants¹⁴. In this study, 85% of all isolates carried at least one tetracycline resistance gene (*tetM*, *tetL*, *tetO* or *tetW*; Supplementary Data S1). Prevalence of *tet* genes varied between CCs and was highest in CC17 (504/526, 96%), followed by CC23 (221/239, 92%), CC19 (272/332, 82%), CC1 (72/97, 74%) and CC10 (61/127, 48%).

Increasing incidence of GBS neonatal infections involved expansion of specific CC17 and CC23 clones. The number of invasive GBS isolates submitted to the Reference Laboratory between 1987 and 2015 increased significantly over time (two-sided Mann-Kendall trend test, p value = 1.19×10^{-7} ; Fig. 1), reflecting

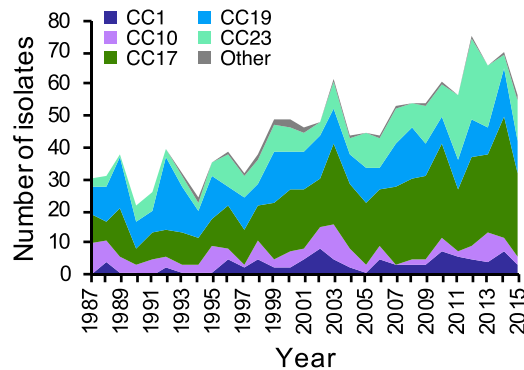


Figure 1. Number of invasive GBS isolates from neonatal infections submitted to the national Reference Laboratory per year stratified by CC. The exact number of isolates from each CC in each year is available in Supplementary Data S1. Number of isolates received in 2016 was excluded from the trend analysis as study sampling ended in May of that year.

a rise in the incidence of neonatal invasive GBS EOD and LOD in the Netherlands during the study period (Supplementary Fig. S2). This was mainly a result of a significant rise in a number of cases caused by isolates from CC17 and CC23 (two-sided Mann-Kendall trend test, p value $< 2.22 \times 10^{-16}$, p value $= 4.65 \times 10^{-6}$, respectively). Between 1987 and 1995, the dominant GBS lineage each year was predominantly CC17. In the mid-1990s CC17 began to outcompete CC19 to become the dominant CC each year from 1996 onwards. Isolates from CC23 were less prevalent than CC19 initially but it became the second most common lineage from 2009. The increase in CC17 isolates was significant among both EOD and LOD cases (two-sided Mann-Kendall trend test, p value $= 2.26 \times 10^{-6}$ and p value $= 1.58 \times 10^{-5}$, respectively) (Supplementary Fig. S3). CC23 isolates showed a more significant increase among EOD than LOD cases (two-sided Mann-Kendall trend test, p value $= 2.55 \times 10^{-5}$ and p value $= 0.008$, respectively) (Supplementary Fig. S3). There were no significant changes in the frequency of isolates from CC10 and CC19 (two-sided Mann-Kendall trend test, $p = 0.45$ and $p = 0.85$, respectively), but a minor rise in CC1 (two-sided Mann-Kendall trend test, p value $= 0.00076$).

To better understand the temporal trends of GBS lineages, in particular the increase in CC17 and CC23 isolates, we investigated the population structure of each CC to identify intra-lineage clusters that changed significantly in prevalence over time. For this, we reconstructed and then partitioned a phylogeny of each CC. The majority of CC17 isolates (428/526, 81%) belonged to a single dominant clade defined here as CC17-A (Fig. 2). The number of isolates from this clade increased significantly over time (two-sided Mann-Kendall trend test, p value $= 1.19 \times 10^{-7}$). The CC17-A clade was partitioned further, which revealed that the increasing temporal trend was associated with expansion of two distinct clones, a larger cluster CC17-A1 (217/526; 41% [of all CC17 isolates]) and a minor cluster CC17-A2 (57/526; 11% [of all CC17 isolates]) (Fig. 2). The remaining CC17-A population was defined as CC17-A0. Only CC17-A1 and CC17-A2 showed a significant increase in a number of isolates over time (two-sided Mann-Kendall trend test, $p = 1.19 \times 10^{-7}$ and $p = 0.0006$, respectively), with CC17-A1 largely responsible for the overall rise in the number of cases due to CC17 (Fig. 2). CC17-A1 isolates were infrequent between 1987 and 1998 but started to increase in 1999, which led to CC17-A1 becoming a dominant CC17 clone by 2007. The occurrence of CC17-A1 isolates increased significantly amongst cases of both EOD and LOD (two-sided Mann-Kendall trend test, $p = 3.1 \times 10^{-5}$, $p = 1.19 \times 10^{-7}$, respectively) (Supplementary Fig. S4). Similarly to CC17, the majority of CC23 isolates (177/239, 74%) belonged to a single clade CC23-A (Fig. 2), which showed a significant increase in the number of isolates over time (two-sided Mann-Kendall trend test, $p = 8.7 \times 10^{-6}$). A further phylogenetic clustering of CC23-A revealed that this was due to an expansion of a sub-cluster CC23-A1 (137/239; 57% [of CC23 isolates]; Fig. 2). After grouping of isolates by disease onset, the increasing CC23-A1 trend was more significant for EOD than LOD cases (two-sided Mann-Kendall trend test, $p = 6.79 \times 10^{-6}$ and $p = 0.026$, respectively) (Supplementary Fig. S4). Overall, the increase in the number of cases caused by isolates representing CC17-A1, CC17-A2 and CC23-A1 was responsible for the general rise in isolate submissions over time with the largest contribution from the CC17-A1 clone (Fig. 2).

Phylogenetic clustering was also performed for CC1, CC10 and CC19 to gain insights into their population structure and check for potential intra-lineage temporal trends. The phylogenetic tree of CC1 was partitioned into six clusters, while CC10 revealed five clusters but they showed no significant temporal changes (Fig. 3). CC19 clustered into four phylogenetic groups (Fig. 3). Isolates from the largest cluster CC19-A (183/332; 55% [of CC19 isolates]), showed a drop over time (two-sided Mann-Kendall trend test, $p = 0.01$), which was offset by a minor increase in the number of isolates from cluster CC19-B (two-sided Mann-Kendall trend test, $p = 0.03$) (Fig. 3).

CC17-A1 and CC23-A1 are not recently emerged clones. To investigate how the recent rise in frequency of isolates representing the CC17-A1 and CC23-A1 clones correlated with the phylogenetic history of CC17 and CC23, we inferred time-calibrated Bayesian phylogenies of each CC. The estimated mutation rates were 8.51×10^{-7} (95% confidence interval [CI] $8.1 \times 10^{-7} - 8.92 \times 10^{-7}$) and 6.21×10^{-7} (95% CI $5.69 \times 10^{-7} - 6.73 \times 10^{-7}$) substitutions per nucleotide site per year for CC17 and CC23, respectively.

For CC17 it was estimated that clades CC17-A and CC17-B emerged in parallel, around the turn of the 1950s and 1960s (Fig. 4). This was also the estimated time of emergence for the CC23-A clade while CC23-B emerged

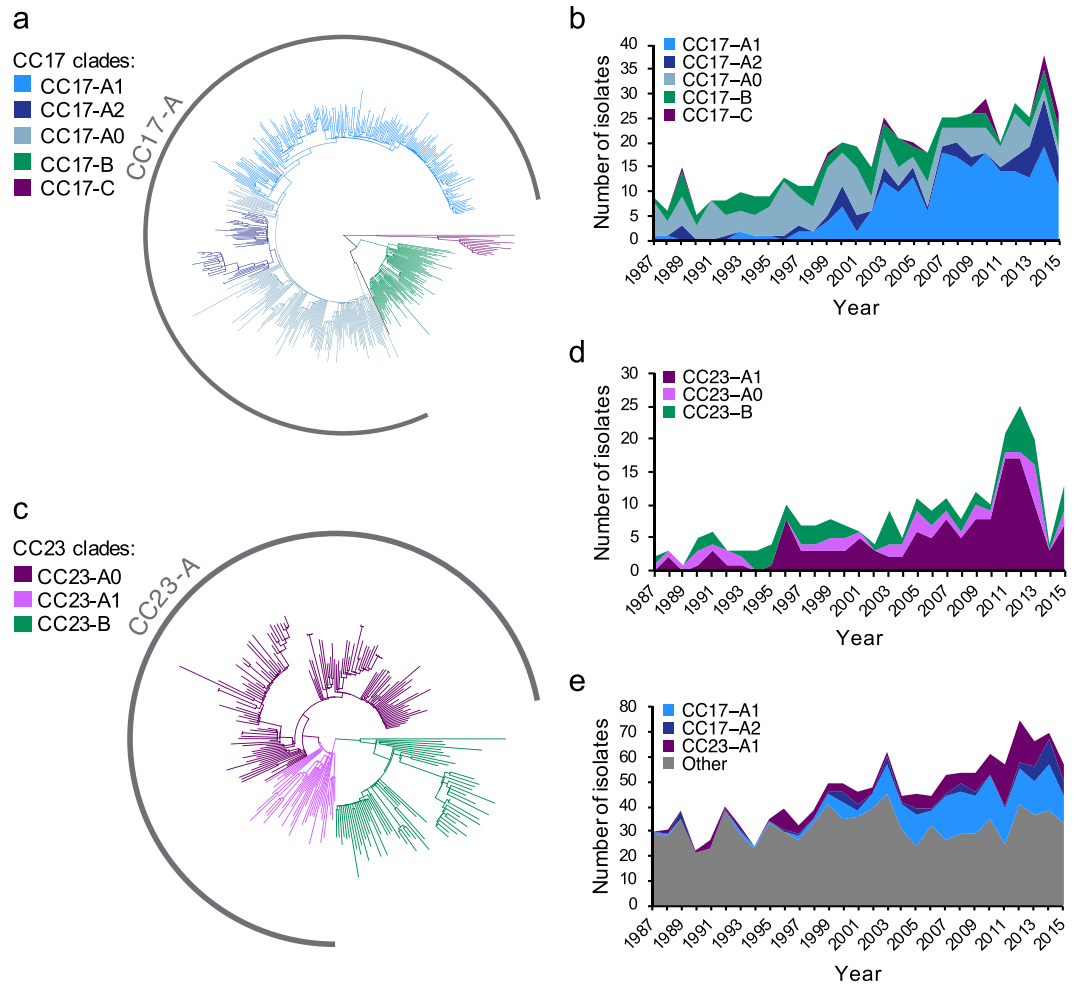


Figure 2. Phylogeny and temporal changes in population structure of CC17 and CC23 lineages. **(a)** Mid-point rooted maximum likelihood phylogenetic tree of CC17. Branches are coloured according to phylogenetic clade assignment. **(b)** Number of isolates from CC17 over time stratified by clade assignment. **(c)** Mid-point rooted maximum likelihood phylogenetic tree of CC23. Branches are coloured according to phylogenetic clade assignment. **(d)** Number of isolates from CC23 over time stratified by clade assignment. **(e)** Contribution of isolates from CC17-A1, CC17-A2 and CC23-A1 clades to the number of invasive GBS isolates from neonatal infections submitted each year. ‘Other’ – GBS isolates representing all other genotypes. The exact number of isolates in each year is available in Supplementary Data S1. Isolates from 2016 were excluded from trend analyses as study sampling ended in May of that year.

a decade earlier, around the turn of the 1940s and 1950s (Fig. 4). This parallel GBS clade divergence between the 1950s and 1970s is supported by the epidemiological history of GBS, which emerged at the time as a leading cause of infant invasive disease¹⁹. It is also in agreement with a previous study describing the emergence of adapted GBS clones in the mid-20th century¹⁴. The time of the most recent common ancestor (TMRCA) for CC17-A1 was estimated around 1968 (95% highest posterior density [HPD] 1965–1970). Similarly, the CC23-A1 clade diverged around 1968 (95% HPD 1965–1972). Therefore, the expanding clusters represent historical genotypes rather than recently emerged clones.

We also estimated the demographic changes over time for CC17-A1 and CC23-A1 using the Bayesian skyline plot model. The CC17-A1 clone showed a sharp rise in the relative genetic diversity (RGD) around the mid-1990s, making its past population changes highly concordant with the epidemiology of CC17-A1 isolates in our collection (Fig. 4). In contrast, the CC23-A1 clone experienced the most recent RGD increase around the mid-1980s (Fig. 4).

The CC17-A1 and CC23-A1 clones are globally distributed. To determine if the CC17-A1 and CC23-A1 clones are specific to the Dutch GBS population or globally distributed, we combined our sequence data with the published genomes of CC17 ($n = 722$) and CC23 ($n = 634$) (Supplementary Data S2)^{14,18,20–25} and reconstructed a global phylogeny of each lineage (Fig. 5). For CC17, we also included six publicly available reference genomes.

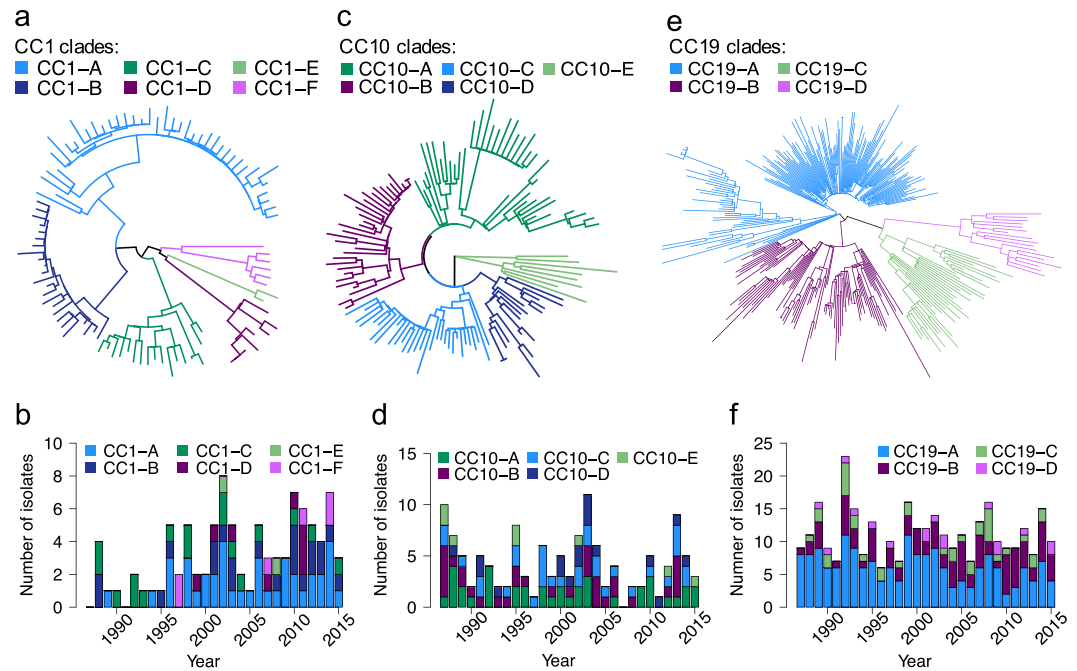


Figure 3. Phylogeny and temporal changes in population structure of CC1, CC10 and CC19 lineages. **(a)** Mid-point rooted maximum likelihood phylogenetic tree of CC1. Branches are coloured according to phylogenetic clade assignment. **(b)** Number of isolates from CC1 over time stratified by clade assignment. **(c)** Mid-point rooted maximum likelihood phylogenetic tree of CC10. Branches are coloured according to phylogenetic clade assignment. **(d)** Number of isolates from CC10 over time stratified by clade assignment. **(e)** Mid-point rooted maximum likelihood phylogenetic tree of CC19. Branches are coloured according to phylogenetic clade assignment. **(f)** Number of isolates from CC19 over time stratified by clade assignment. The exact number of isolates in each year is available in Supplementary Data S1. Isolates from 2016 were excluded from trend analyses as study sampling ended in May of that year.

The external dataset of CC17 contained isolates from 19 different countries although only seven countries were represented by more than 10 isolates (Supplementary Data S2). A total of 340 (47%) and 354 (49%) non-Dutch isolates represented clades CC17-A and CC17-B, respectively. More than half of global CC17-A isolates belonged to clone CC17-A1 (191/340; 56%), followed by CC17-A0 (116/340; 34%) and CC17-A2 (33/340; 10%). The CC17-A1 isolates were found in all external collections included in this analysis. The CC17-A1 clone was common among CC17 isolates from Canada (56/77, 73%), Italy (11/13, 85%), China (7/14, 50%), USA (70/159, 44%) and France (24/58, 41%). The reference genomes were distributed across the phylogeny, with the NGBS128 strain²⁰ representing CC17-A1 and the B105 and 874391 strains representing CC17-A2 (Fig. 5).

The global dataset of CC23 genomes was represented largely by isolates from Kenya and the USA (Fig. 5). The non-Dutch samples were dominated by isolates from the CC23-A1 clade (480/636; 75%), which represented 75% of CC23 isolates from the USA (275/367) and 82% of CC23 isolates from Kenya (180/219).

Expansion of the CC17-A1 clone correlates with acquisition of a novel phage. To investigate if expansion of the CC17-A1 clone can be attributed to presence of unique genetic markers, we investigated the core and accessory genome characteristics of the CC17-A1 cluster. Amongst core genome single nucleotide polymorphisms (SNPs) associated with the CC17-A1 clone (Supplementary Table S1), of particular interest is a double non-synonymous mutation in the *ciaH* gene (positions 954832 and 954833 in the *S. agalactiae* COH1 reference genome), which encodes the CiaH sensor histidine kinase. This mutation was found in all CC17-A1 isolates as well as 14 other CC17 isolates across the phylogeny (Supplementary Fig. S5), representing a homoplasic mutation that may indicate selective pressure acting on this locus.

To check if the CC17-A1 clone features a unique accessory genome we investigated the pan-genome of CC17 and clustered isolates based on gene presence/absence using principal component analysis (PCA). Most of the accessory gene variation within CC17 could be attributed to the difference between clades CC17-A and CC17-B (91% of variance explained by PC1). An additional subdivision was visualized by plotting PC2 against PC3, which revealed three clusters, each composed of isolates from the different CC17-A phylogenetic groups (Fig. 6). This subdivision was therefore not due to accessory genes unique to the CC17-A1 clone. Correlation analyses of the accessory gene presence/absence within these clusters showed that the clustering pattern was due to variable carriage of two sets of genes that were found to be associated with two different mobile genetic elements (MGEs). One was identified as an integrative conjugative element (ICE), described previously as ICE_COH1_tRNA^{lys26} (hereafter referred to as ICE_COH1). This MGE encodes several virulence factors including antigen I/II²⁷, CAMP factor II²⁸, and manganese transporter MntH²⁹. The second element was a phage, designated here as phiStag1 (also found in the NGBS128 reference genome, coordinates: 689741–748064). It was found to carry a

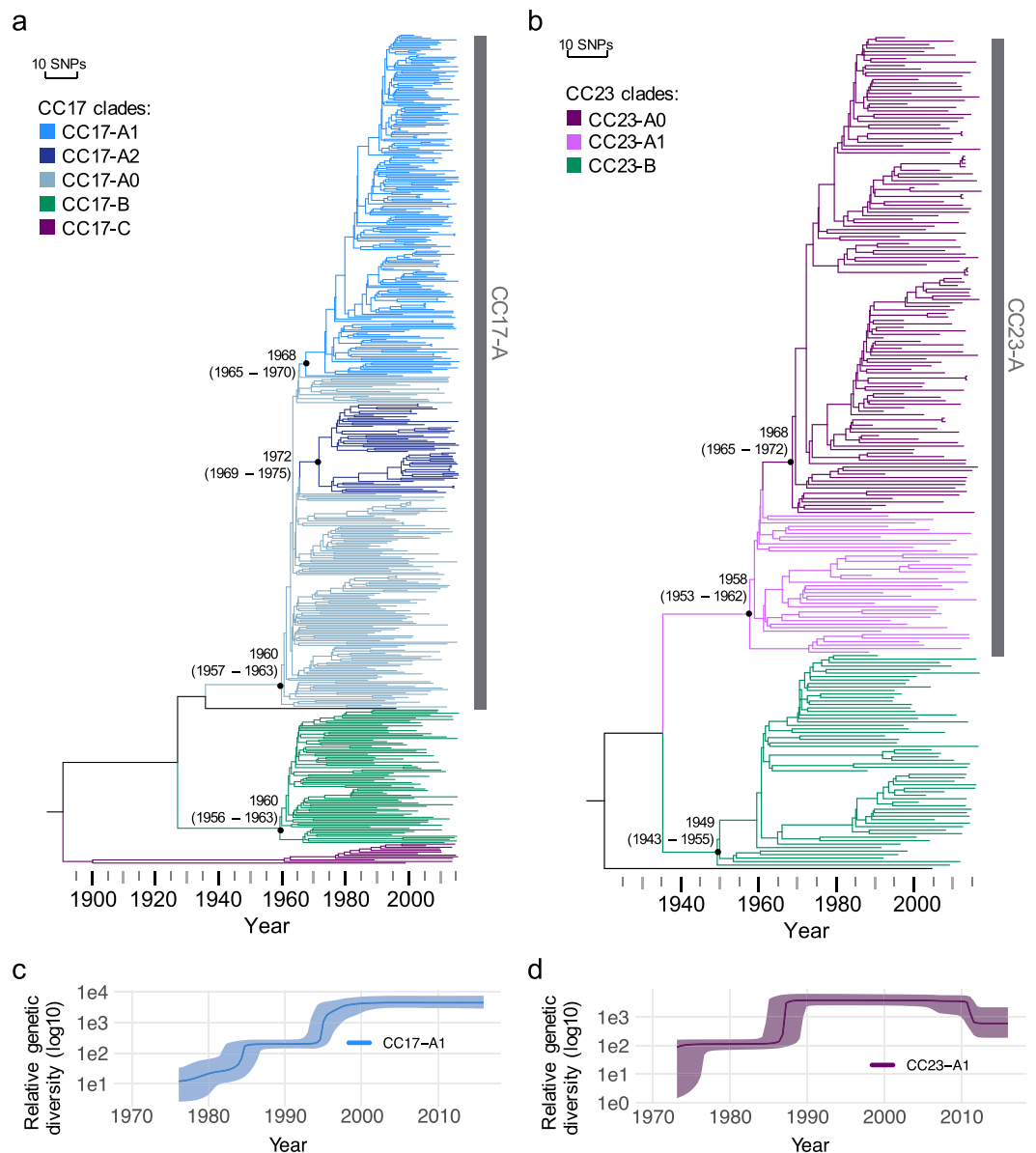


Figure 4. Evolutionary history of CC17 and CC23. **(a)** Bayesian maximum clade credibility tree of CC17. Tree branches are coloured in accordance with the CC17 clade assignment. **(b)** Bayesian maximum clade credibility tree of CC23. Tree branches are coloured in accordance with the CC23 clade assignment. **(c)** Bayesian coalescent skyline plot for CC17-A1 and **(d)** CC23-A1. Plot lines represent the median relative genetic diversity, over time. Band around each line represents the upper and lower 95% highest posterior density intervals.

gene (NGBS128 locus tag: AMM49_03725) encoding a putative cell-surface protein, containing an N-terminal YSIRK signal peptide, a C-terminal LPxTG cell-wall anchor motif and a L-type lectin domain with potential carbohydrate-binding properties.

Screening of all CC17 isolates showed that 45% (238/526) and 34% (179/526) carried ICE_COH1 and phiStag1 elements, respectively. Composition of PCA clusters reflected the distribution of these MGEs across the CC17 population, with isolates clustering based on carriage of both, either one or neither MGE (Fig. 6). Isolates carrying ICE_COH1 and phiStag1 were intermittently distributed across the CC17 phylogeny suggesting multiple horizontal acquisition events (Fig. 6). However, the prevalence of both MGEs was highest among CC17-A1 and CC17-A2 isolates. The phiStag1 was found in 60% (131/217) of CC17-A1 and 49% (28/57) of CC17-A2 isolates in contrast to just 7% (11/153) of CC17-A0 isolates. The CC17-A0 showed a higher prevalence of ICE_COH1 (56/153, 37%), although the element was still more common in CC17-A1 (130/217, 60%) and CC17-A2 (30/57, 53%) isolates. Analysis of temporal prevalence of the two elements among CC17 isolates showed that while ICE_COH1 occurred throughout the sampling period, isolates with phiStag1 were first detected in 1997 (Fig. 6), coinciding with the time of CC17-A1 clonal expansion. The number of CC17 phiStag1-positive isolates, predominantly from the CC17-A1 cluster, increased steadily from 2000.

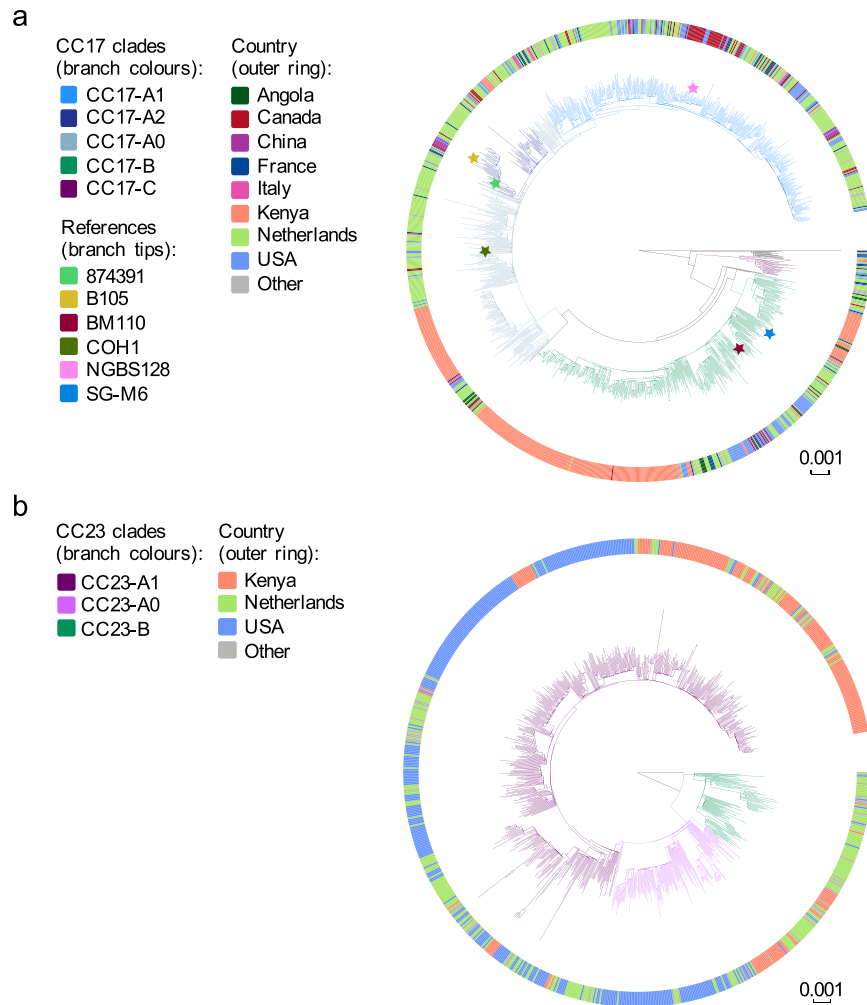


Figure 5. Global phylogeny of CC17 and CC23. Tree branches of each tree are coloured in accordance with CC clade assignment. Each tree is annotated with colour strips representing isolate's country of origin (unique colour for countries with at least 10 isolates, 'Other': countries represented by less than 10 isolates). Scale bars represent the number of nucleotide substitutions per site. **(a)** a maximum likelihood phylogenetic tree of 1254 CC17 isolates from the Netherlands and other countries, including six reference genomes (reference branch tips marked with a star): NGBS128, COH1, B105, 874391, BM110 and SG-M6. **(b)** a maximum likelihood phylogenetic tree of 873 CC23 isolates from the Netherlands and other countries.

Across the collection, phiStag1 occurred sporadically in isolates from other CCs, but no isolate carried phiStag1 before its first emergence within CC17 in 1997. The highest prevalence of phiStag1 outside of CC17 was among CC23 isolates (41/239, 17%), however, it was not overrepresented within the CC23-A1 cluster. As expansion of the CC23-A1 clone could not be attributed to carriage of phiStag1, a separate pan-genome analysis of the entire CC23 was conducted but this did not reveal any accessory genes or MGEs associated with the CC23-A1 clone.

Discussion

Genomic analysis of a national GBS collection from cases of neonatal invasive disease in the Netherlands between 1987 and 2016 revealed expansion of historical clones from two major lineages: CC17 and CC23. A significant increase in isolates representing these clones, defined here as CC17-A1, CC17-A2 and CC23-A1, was sufficient to account for the overall increase in a number of GBS infection cases during this period. The largest single contributor to the increasing trend was the CC17-A1 clone, and its clonal expansion was the key underlying mechanism for the previously reported rise in the number of CC17 isolates among cases of neonatal GBS infections in the Netherlands¹¹. It also resulted in CC17 becoming the most prevalent GBS lineage each year from the mid-1990s up until the end of the study period. This further underscores the importance of CC17 as a dominant GBS lineage associated with neonatal disease and offers new insights into its population structure and changes over time. The expansion of CC23-A1 led to CC23 gradually becoming the second most common lineage, alongside CC19 isolates. The rise in the number of isolates representing the CC23-A1 clone was most significant among cases of EOD, which might be addressed by a better adherence to or a revision to IAP guidelines. In contrast,

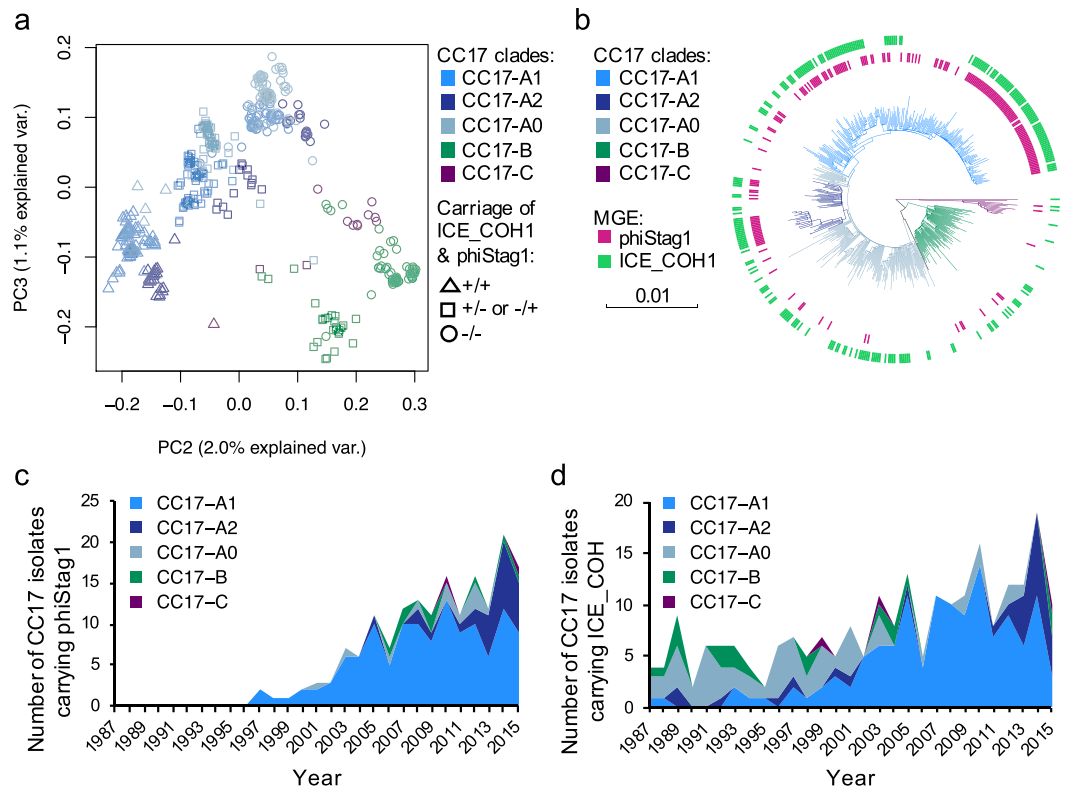


Figure 6. Distribution of phiStag1 and ICE_COH elements across CC17 population. **(a)** PCA of the accessory gene distribution in CC17. Each data point is colour-coded to show isolate clade assignment (CC17-A1, CC17-A2, CC17-A0, CC17-B or CC17-C). Data point shapes represent the pattern of ICE_COH1 and phiStag1 carriage to indicate if both MGEs are present (+/+), if only either one MGE is present (+/- or -/+), or if neither MGE is present (-/-). Isolates from CC17-A clade revealed all three patterns, while none of the CC17-B isolates carried both MGEs. **(b)** Distribution of ICE_COH1 and phiStag1 across the CC17 phylogeny. Tips of the phylogenetic tree are annotated with strips indicating presence of each MGE. Tree branches are coloured in accordance with CC17 clade assignment. Scale bar represents the number of nucleotide substitutions per site. **(c)** Number of CC17 isolates carrying phiStag1 each year. **(d)** Number of CC17 isolates carrying ICE_COH each year.

the CC17-A1 isolates increased significantly for both EOD and LOD cases, thus reviewing domestic IAP policy would only partly counteract the rising prevalence of CC17-A1-associated infections.

The driving forces behind the expansion of CC17-A1, CC17-A2 and CC23-A1 clones are unclear. A common driver of bacterial pathogen evolution is the antibiotic selective pressure which can lead to clonal expansion of pre-existing bacterial genotypes that already carry resistance determinants³⁰. Antimicrobial resistance may also play an important role in GBS epidemiology as expansion of a GBS clone showing macrolide resistance was recently reported in Portugal among isolates from invasive disease in non-pregnant adults³¹. There have also been reports of multidrug resistant clones of CC17 in Portugal and China that were found among isolates from neonatal disease^{16,23}. Phenotypic antimicrobial susceptibility data was not available for our GBS collection and isolates were characterised by genotypic resistance typing only. Overall, we observed a low prevalence of macrolide resistance genes in this collection in comparison to macrolide resistance rates among GBS isolates from neonatal disease in other countries such as France and Portugal^{3,31}. This might be due to a lower consumption of antibiotics in the Netherlands in comparison to other countries in Europe³². In the absence of phenotypic resistance data, genotypic resistance profiles might also underestimate the level of antimicrobial resistance in our GBS collection.

Genomic analysis did not reveal any association between the expansion of CC17 and CC23, and carriage of antimicrobial resistance determinants although the time of their clonal expansion coincided with the introduction of national GBS disease prevention guidelines in the Netherlands in 1999 that recommend antibiotic prophylaxis for women in labour with clinical risk factors for transmission of GBS disease. However, a link between IAP treatment and the rise in the frequency of isolates representing these GBS clones could not be established here as clinical data on whether mothers received the IAP was not available. It was therefore not possible to determine which disease cases were associated with a failed antimicrobial prophylaxis.

Expansion of pre-existing strains can be triggered by genetic changes within circulating clones such as acquisition of novel genes that alter bacterial virulence³³. An increase in prevalence of a pre-existing GBS clone after it acquired new genetic determinants was recently reported for GBS ST-1 serotype V^{34,35}. Our GBS collection showed an emergence of a novel phage phiStag1, which was first detected in a CC17-A1 isolate from 1997. This was followed by a year-on-year increase in a number of CC17 isolates carrying this phage, predominantly among the expanding CC17-A1 and CC17-A2 clones. The phiStag1 phage may have become more prevalent due to an

association with a successful clone. Alternatively, the acquisition of phiStag1 directly contributed to the expansion of CC17-A1 and CC17-A2. Phages have been shown to play an important role in the evolution, emergence and persistence of virulent *S. pyogenes* lineages by facilitating acquisition of novel virulence traits^{33,36}. In GBS, phage-associated genes represent around 10% of strain-specific genes and some might be involved in adaptation and virulence^{37,38}. No resistance determinants were identified in the genome of phiStag1 but it was found to carry gene encoding a putative cell surface protein that could be involved in interaction with host factors. In parallel with our findings, acquisition of a gene encoding a novel cell-surface protein was linked to the expansion of the GBS ST-1 serotype V clone³⁵, indicating that such genetic determinants may play an important role in the evolution of GBS virulence.

Analysis of genetic variants associated with the CC17-A1 clone also revealed a double non-synonymous mutation in the *ciaH* gene in all CC17-A1 isolates. Although this variant is not associated with a recent emergence as it occurred on a root of the CC17-A1 cluster, it represented a homoplastic mutation that was detected in other CC17 isolates and might therefore represent a pathoadaptive mutation. The *ciaH* gene encodes the CiaH sensor histidine kinase that belongs to the two-component regulatory system (TCRS) CiaRH. TCRSs have been shown to play an important role in GBS adaptation to changes in host environment, and can influence the fitness and pathogenesis of GBS^{39–41}. Although it is currently unknown what environmental parameters are monitored by CiaH, the CiaR regulator has been associated with GBS intracellular survival and enhanced virulence in a mouse model of infection⁴².

Bayesian inference of phylogeny and past population changes showed that the sudden rise in the number of CC17-A1 isolates among disease cases correlated with a sharp increase in its relative genetic diversity and clonal expansion, which coincided with phiStag1 acquisition. This was not observed for CC23-A1, for which a clonal expansion event based on phylogenetic reconstruction occurred in the mid-1980s, a decade before it started to rise in frequency among GBS isolates from neonatal disease. Analysis of the CC23 pan-genome did not show association between CC23-A1 and any particular gene or a subset of genes. As such it is currently unclear which CC23-A1 genetic variants might have contributed to its recent rise in prevalence. It is possible that CC23-A1 represents a fitter and better adapted clone due to a particular composition of core genome SNPs.

We found that both CC17-A1 and CC23-A1 are globally distributed, however, a wider genomic surveillance of GBS is needed to determine their contribution to the burden of GBS invasive disease in neonates worldwide. As the scope of our work was limited to the analysis of isolates from disease it is unclear if the expanding CC17-A1 and CC23-A1 clones have also become more prevalent in carriage. Analysis of genomic diversity of GBS from maternal carriage is therefore required to better understand the association between the CC17-A1 and CC23-A1 clones and the invasive disease in neonates.

Methods

Bacterial isolates. Isolates were derived from a nationwide surveillance of bacterial meningitis and infant bacteraemia conducted by the Netherlands Reference Laboratory for Bacterial Meningitis (NRLMB). The NRLMB receives approximately 90% of cerebrospinal fluid (CSF) isolates of all patients with bacterial meningitis in the Netherlands^{43,44}. In addition, they routinely receive blood isolates from infants with an invasive GBS infection from microbiology laboratories in the Netherlands⁴³. We included isolates received between 01.1987 and 05.2016 that were collected from patients aged 0–89 days ($n = 1345$). Isolates were excluded if the date of birth or date of collection were not recorded, they were non-invasive, or if they were found to represent a second isolate from the second episode in a patient with more than one episode. Although the majority of isolates were derived from blood or cerebrospinal fluid, three samples were isolated from other body sites (Supplementary Data S1). Patient age was calculated as the number of days between the date of birth and the earliest known date of the illness; mostly the first date of culture. EOD was defined as invasive infection between 0 and 6 days after birth whereas LOD was defined as invasive infection between 7 and 89 days of life.

Whole-genome sequencing, *de novo* genome assembly and annotation. Genomic DNA was extracted using Wizard[®] Genomic DNA Purification Kit from Promega. Tagged DNA libraries were created according to the Illumina protocol. Whole-genome sequencing was performed on the Illumina HiSeq. 2000 platform with 125 bp paired-end reads. Annotated assemblies were produced using a pipeline described previously⁴⁵. For each sample, sequence reads were used to create multiple assemblies using VelvetOptimiser v2.2.5 (<https://github.com/tseemann/VelvetOptimiser>) and Velvet v1.2⁴⁶. An assembly improvement step was applied to the assembly with the best N50 and contigs were scaffolded using SSPACE⁴⁷, with sequence gaps filled using GapFiller⁴⁸. Automated annotation was performed using PROKKA v1.11⁴⁹ and a *Streptococcus*-specific database from RefSeq⁵⁰.

Molecular typing. Isolates were assigned to a CC group based on MLST¹³, hierBAPS clustering⁵¹ and core-genome phylogeny data. MLST was performed using SRST2^{13,52}. For hierBAPS analysis a core genome SNP alignment was used that was generated using a reference-based approach. For this, sequence reads were mapped against the NGBS128 reference genome²⁰ using SMALT v0.7.4 (<https://www.sanger.ac.uk/science/tools/smalt-0>) and SNPs were called using SAMtools v0.1.19⁵³ and bcftools v0.1.19. Regions representing putative MGEs were masked using `remove_blocks_from_aln` (https://github.com/sanger-pathogens/remove_blocks_from_aln) and SNPs were identified with SNP-sites⁵⁴. Alignment positions with >5% uncalled variants were filtered out. This core genome alignment was also used to reconstruct an approximately maximum-likelihood tree using FastTree⁵⁵. Using MLST and hierBAPS clustering data isolates were assigned to a CC group if they represented the same hierBAPS cluster (except for the bin cluster) as the founder ST (confirmed with eBURSTv3⁵⁶). The CC assignment was confirmed by plotting clusters on a phylogenetic tree and visually inspecting if isolates from the same CC belong to a monophyletic cluster. Isolates representing hierBAPS bin group, a polyphyletic cluster of low-frequency genotypes, that represented a single ST and were phylogenetically distant from other STs/CCs had a CC defined based on their ST.

CPS serotype was determined by *in silico* PCR using previously described primers for the detection of serotypes Ia, Ib and II-IX^{57,58}. Carriage of antimicrobial resistance genes was checked with SRST2⁵² using the ARG-ANNOT database⁵⁹, supplemented with sequences of GBS resistance determinants described by Metcalf *et al.*¹⁸.

Phylogenetic reconstructions. To reconstruct the phylogeny of each major CC, a core genome SNP alignment was used that was generated using a reference-based method as described above. CC-specific *S. agalactiae* reference genomes were used for mapping: SS1 (CC1)³⁵, Sag37 (CC10), NGBS128 (CC17)²⁰, H002 (CC19)⁶⁰, NEM316 (CC23)⁶¹. In addition to masking of regions representing putative MGEs, each CC was also screened for presence of variable sites associated with recombination using Gubbins⁶². The observed impact of recombination on genetic variation varied between CCs. Based on the recombination to mutation ratio (*r/m*) it was highest in CC10 (*r/m* = 6), followed by CC1 (*r/m* = 4.53), CC19 (*r/m* = 1.08), CC23 (*r/m* = 0.87) and was the lowest in CC17 (*r/m* = 0.24). For each CC, variable sites due to recombination were removed from the genome alignment. As a result, size of the core genome varied between the CCs and consisted of 2499 SNPs for CC1, 4572 for CC10, 18595 for CC17, 11979 for CC19 and 6739 for CC23. Maximum likelihood (ML) phylogenetic tree of each CC was generated with RAxML v8.2.8⁶³ based on a generalised time reversible model with GAMMA method of correction for among site rate variation and 100 bootstrap replications.

The ML phylogenies were partitioned using fastbaps⁶⁴. The CC17 and CC23 phylogenies were additionally partitioned using a method by Prospero *et al.*⁶⁵ based on pairwise SNP distances and node reliability of $\geq 90\%$ (bootstrap support). For this, increasing pairwise SNP distance thresholds were tested, ranging from 150 to 300, with increments of 10. To identify expanding clusters, for each threshold that resulted in unique clustering pattern resulting clusters were checked for temporal trend by measuring the slope of a regression line based on the number of isolates from a cluster over time. Clusters with the highest positive slope value were selected.

To reconstruct time-calibrated phylogenies of CC17 and CC23, each sequence in the CC-specific core genome alignment was annotated with the date of isolation based on day, month, and year. A correlation between root-to-tip distance and the date of sampling was checked using TempEst⁶⁶. The estimated correlation coefficient was $R^2 = 0.455$ and $R^2 = 0.389$ for CC17 and CC23, respectively. Bayesian inference of phylogeny and past population dynamics was performed with BEAST 2⁶⁷ under a GTR + Γ nucleotide substitution model. Strict clock and relaxed lognormal clock models were compared, in combination with the constant, exponential and Bayesian skyline population models. Runs involving the relaxed lognormal clock model failed to converge. The strict molecular clock gave the best fit when used with the Bayesian skyline model. The MCMC chains were run for 50 million generations, sampling every 1000 states. Log files from six independent runs were combined after removal of burn-in (5% of samples) using LogCombiner and analysed with Tracer v1.5. The effective sample size (ESS) was 360 and 3251 for CC17 and CC23, respectively. Maximum clade credibility (MCC) tree was generated with TreeAnnotator. All phylogenetic trees were annotated using Evolview^{68,69}.

Core genome SNPs in CC17-A1. To analyse core genome SNPs in CC17 isolates, sequence reads were mapped using SMALT v0.7.4 (<https://www.sanger.ac.uk/science/tools/smalt-0>) to the COH1 reference genome¹⁴, and variants were called using SAMtools v0.1.19⁵³ and bcftools v0.1.19. Variants were grouped by CC17 clusters. Variants located in regions representing putative MGEs were excluded from the analysis.

Pan-genome of CC17. Pan-genome analysis of CC17 was performed with Roary⁷⁰ using 95% identity threshold for blastp with splitting of clusters that contain paralogs. PCA of the accessory gene content was performed in R v3.3.3. Genes that were correlated (based on Spearman's rank correlation coefficient) with selected PCA clusters were identified and analysed further. To determine their genomic context, genes were mapped back to whole genome assemblies using blastn pairwise sequence alignments⁷¹. Sequence similarity searches of genomic fragments found to contain the accessory genes of interest were performed using the NCBI BLAST and GenBank nucleotide sequence database⁷². MGE carriage was checked by short read mapping with SRST2⁵² using 90% sequence coverage cut-off.

Data availability

Sequence reads were deposited in the European Nucleotide Archive under study accession PRJEB14124. Individual isolate accession numbers are provided in Supplementary Data S1. GenBank accession numbers of reference genomes used in this study are: SS1: CP010867, Sag37: CP019978, NGBS128: CP012480, H002: CP011329, NEM316: AL732656, COH1: HG939456, B105: CP021773, 87439: CP022537, BM110: LT714196 and SG-M6: CP021869). To reconstruct global phylogenies of CC17 and CC23, we included published genomes of 722 CC17 and 634 CC23 isolates from studies by Da Cunha *et al.*¹⁴, Rosini *et al.*²¹, Seale *et al.*²², Teatero *et al.*²⁰, Campisi *et al.*^{23,24}, Almeida *et al.*²⁵ and Metcalf *et al.*¹⁸. Individual isolate accession numbers are provided in Supplementary Data S2.

Received: 6 February 2020; Accepted: 15 May 2020;

Published online: 12 June 2020

References

- Schuchat, A. Group B streptococcus. *The Lancet* **353**, 51–56 (1999).
- Madrid, L. *et al.* Infant group B streptococcal disease incidence and serotypes worldwide: systematic review and meta-analyses. *Clin. Infect. Dis.* **65**, S160–S172, <https://doi.org/10.1093/cid/cix656> (2017).
- Joubrel, C. *et al.* Group B streptococcus neonatal invasive infections, France 2007–2012. *Clin. Microbiol. Infect.* **21**, 910–916, <https://doi.org/10.1016/j.cmi.2015.05.039> (2015).

4. O'Sullivan, C. P. *et al.* Group B streptococcal disease in UK and Irish infants younger than 90 days, 2014–15: a prospective surveillance study. *The Lancet Infectious Diseases* **19**, 83–90 (2019).
5. Le Doare, K. & Heath, P. T. An overview of global GBS epidemiology. *Vaccine* **31**, D7–D12 (2013).
6. Schuchat, A. Epidemiology of group B streptococcal disease in the United States: shifting paradigms. *Clin. Microbiol. Rev.* **11**, 497–513 (1998).
7. Phares, C. R. *et al.* Epidemiology of invasive group B streptococcal disease in the United States, 1999–2005. *JAMA* **299**, 2056–2065 (2008).
8. Darlow, B. A., Voss, L., Lennon, D. R. & Grimwood, K. Early-onset neonatal group B streptococcus sepsis following national risk-based prevention guidelines. *Aust. N. Z. J. Obstet. Gynaecol.* **56**, 69–74, <https://doi.org/10.1111/ajo.12378> (2016).
9. Romain, A.-S. *et al.* Clinical and laboratory features of Group B *Streptococcus* meningitis in infants and newborns: study of 848 cases in France, 2001–2014. *Clin. Infect. Dis.* **66**, 857–864 (2017).
10. Hakansson, S., Lilja, M., Jacobsson, B. & Kallen, K. Reduced incidence of neonatal early-onset group B streptococcal infection after promulgation of guidelines for risk-based intrapartum antibiotic prophylaxis in Sweden: analysis of a national population-based cohort. *Acta Obstet. Gynecol. Scand.* **96**, 1475–1483, <https://doi.org/10.1111/aogs.13211> (2017).
11. Bekker, V., Bijlsma, M. W., van de Beek, D., Kuijpers, T. W. & van der Ende, A. Incidence of invasive group B streptococcal disease and pathogen genotype distribution in newborn babies in the Netherlands over 25 years: a nationwide surveillance study. *Lancet Infect. Dis.* **14**, 1083–1089, [https://doi.org/10.1016/S1473-3099\(14\)70919-3](https://doi.org/10.1016/S1473-3099(14)70919-3) (2014).
12. Le Doare, K. *et al.* Intrapartum antibiotic chemoprophylaxis policies for the prevention of group B streptococcal disease worldwide: systematic review. *Clin. Infect. Dis.* **65**, S143–S151, <https://doi.org/10.1093/cid/cix654> (2017).
13. Jones, N. *et al.* Multilocus sequence typing system for Group B *Streptococcus*. *J. Clin. Microbiol.* **41**, 2530–2536 (2003).
14. Da Cunha, V. *et al.* *Streptococcus agalactiae* clones infecting humans were selected and fixed through the extensive use of tetracycline. *Nature communications* **5**, 4544 (2014).
15. Castor, M. L. *et al.* Antibiotic resistance patterns in invasive group B streptococcal isolates. *Infect. Dis. Obstet. Gynecol.* 2008 (2008).
16. Martins, E. R., Pedroso-Roussado, C., Melo-Cristino, J. & Ramirez, M., Portuguese Group for the Study of Streptococcal, I. *Streptococcus agalactiae* causing neonatal infections in Portugal (2005–2015): diversification and emergence of a CC17/PI-2b multidrug resistant sublineage. *Front. Microbiol.* **8**, 499, <https://doi.org/10.3389/fmicb.2017.00499> (2017).
17. Chu, Y. W. *et al.* Invasive group B *Streptococcus* isolates showing reduced susceptibility to penicillin in Hong Kong. *J. Antimicrob. Chemother.* **60**, 1407–1409 (2007).
18. Metcalf, B. J. *et al.* Short-read whole genome sequencing for determination of antimicrobial resistance mechanisms and capsular serotypes of current invasive *Streptococcus agalactiae* recovered in the USA. *Clin. Microbiol. Infect.* **23**, 574 e577–574 e514, <https://doi.org/10.1016/j.cmi.2017.02.021> (2017).
19. Jones, H. E. & Howells, C. Neonatal meningitis due to *Streptococcus agalactiae*. *Postgrad. Med. J.* **44**, 549 (1968).
20. Teatero, S. *et al.* Clonal Complex 17 Group B *Streptococcus* strains causing invasive disease in neonates and adults originate from the same genetic pool. *Sci. Rep.* **6**, 20047, <https://doi.org/10.1038/srep20047> (2016).
21. Rosini, R. *et al.* Genomic analysis reveals the molecular basis for capsule loss in the Group B *Streptococcus* population. *PLoS One* **10**, e0125985, <https://doi.org/10.1371/journal.pone.0125985> (2015).
22. Seale, A. C. *et al.* Maternal colonization with *Streptococcus agalactiae* and associated stillbirth and neonatal disease in coastal Kenya. *Nat Microbiol* **1**, 16067, <https://doi.org/10.1038/nmicrobiol.2016.67> (2016).
23. Campisi, E. *et al.* Genomic analysis reveals multi-drug resistance clusters in Group B *Streptococcus* CC17 hypervirulent Isolates causing neonatal invasive disease in southern mainland China. *Front. Microbiol.* **7**, 1265, <https://doi.org/10.3389/fmicb.2016.01265> (2016).
24. Campisi, E. *et al.* Serotype IV *Streptococcus agalactiae* ST-452 has arisen from large genomic recombination events between CC23 and the hypervirulent CC17 lineages. *Sci. Rep.* **6**, doi:ARTN 2979910.1038/srep29799 (2016).
25. Almeida, A. *et al.* Parallel evolution of Group B *Streptococcus* hypervirulent clonal complex 17 unveils new pathoadaptive mutations. *mSystems* **2**, e00074–00017, <https://doi.org/10.1128/mSystems.00074-17> (2017).
26. Brochet, M., Couve, E., Glaser, P., Guedon, G. & Payot, S. Integrative conjugative elements and related elements are major contributors to the genome diversity of *Streptococcus agalactiae*. *J. Bacteriol.* **190**, 6913–6917, <https://doi.org/10.1128/JB.00824-08> (2008).
27. Chuzeville, S., Dramsi, S., Madec, J. Y., Haenni, M. & Payot, S. Antigen I/II encoded by integrative and conjugative elements of *Streptococcus agalactiae* and role in biofilm formation. *Microb. Pathog.* **88**, 1–9, <https://doi.org/10.1016/j.micpath.2015.07.018> (2015).
28. Chuzeville, S., Puymege, A., Madec, J. Y., Haenni, M. & Payot, S. Characterization of a new CAMP factor carried by an integrative and conjugative element in *Streptococcus agalactiae* and spreading in Streptococci. *PLoS One* **7**, e48918, <https://doi.org/10.1371/journal.pone.0048918> (2012).
29. Juttukonda, L. J. & Skaar, E. P. Manganese homeostasis and utilization in pathogenic bacteria. *Mol. Microbiol.* **97**, 216–228, <https://doi.org/10.1111/mmi.13034> (2015).
30. Keenan, J. D. *et al.* Evidence for clonal expansion after antibiotic selection pressure: pneumococcal multilocus sequence types before and after mass azithromycin treatments. *The Journal of infectious diseases* **211**, 988–994 (2014).
31. Lopes, E. *et al.* Increasing macrolide resistance among *Streptococcus agalactiae* causing invasive disease in non-pregnant adults was driven by a single capsular-transformed lineage, Portugal, 2009 to 2015. *Eurosurveillance* **23**, 1700473 (2018).
32. European Centre for Disease Prevention and Control. Consumption of Antibacterials for systemic use in the community in Europe, reporting year 2018, <https://www.ecdc.europa.eu/en/antimicrobial-consumption/database/rates-country> (2020).
33. Aziz, R. K. & Kotb, M. Rise and persistence of global MIT1 clone of *Streptococcus pyogenes*. *Emerg. Infect. Dis.* **14**, 1511–1517, <https://doi.org/10.3201/eid1410.071660> (2008).
34. Elliott, J., Farmer, K. & Facklam, R. Sudden increase in isolation of group B streptococci, serotype V, is not due to emergence of a new pulsed-field gel electrophoresis type. *J. Clin. Microbiol.* **36**, 2115–2116 (1998).
35. Flores, A. R. *et al.* Sequence type 1 group B *Streptococcus*, an emerging cause of invasive disease in adults, evolves by small genetic changes. *Proc. Natl. Acad. Sci. USA* **112**, 6431–6436, doi:10.1073/pnas.1504725112 (2015).
36. Davies, M. R. *et al.* Emergence of scarlet fever *Streptococcus pyogenes* emm12 clones in Hong Kong is associated with toxin acquisition and multidrug resistance. *Nat. Genet.* **47**, 84 (2015).
37. Tettelin, H. *et al.* Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc. Natl. Acad. Sci. USA* **102**, 13950–13955, <https://doi.org/10.1073/pnas.0506758102> (2005).
38. Van der Mee-Marquet, N. *et al.* Analysis of the prophages carried by human infecting isolates provides new insight into the evolution of Group B *Streptococcus* species. *Clin. Microbiol. Infect.* **24**, 514–521 (2018).
39. Lembo, A. *et al.* Regulation of CovR expression in Group B *Streptococcus* impacts blood-brain barrier penetration. *Mol. Microbiol.* **77**, 431–443, <https://doi.org/10.1111/j.1365-2958.2010.07215.x> (2010).
40. Klinzing, D. C. *et al.* The two-component response regulator LiaR regulates cell wall stress responses, pili expression and virulence in group B *Streptococcus*. *Microbiology* **159**, 1521–1534, <https://doi.org/10.1099/mic.0.064444-0> (2013).
41. Faralla, C. *et al.* Analysis of two-component systems in Group B *Streptococcus* shows that RgfAC and the novel FspSR modulate virulence and bacterial fitness. *Mbio* **5**, e00870–14, <https://doi.org/10.1128/mBio.00870-14> (2014).

42. Quach, D. *et al.* The CiaR response regulator in Group B Streptococcus promotes intracellular survival and resistance to innate immune defenses. *J. Bacteriol.* **191**, 2023–2032, <https://doi.org/10.1128/JB.01216-08> (2009).
43. Netherlands Reference Laboratory For Bacterial Meningitis (AMC/RIVM). Bacterial meningitis in The Netherlands: Annual Report 2016. Amsterdam: University of Amsterdam, <https://www.amc.nl/web/file?uuid=6c071e7d-8680-42df-b381-0bbb789f658f&owner=9423f858-411c-47f3-b97b-98d8118882c1>. (2017)
44. Bijlsma, M. W. *et al.* Epidemiology of invasive meningococcal disease in the Netherlands, 1960–2012: an analysis of national surveillance data. *The Lancet infectious diseases* **14**, 805–812 (2014).
45. Page, A. J. *et al.* Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. *Microb Genom* **2**, e000083, <https://doi.org/10.1099/mgen.0.000083> (2016).
46. Zerbino, D. R. & Birney, E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829, <https://doi.org/10.1101/gr.074492.107> (2008).
47. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**, 578–579, <https://doi.org/10.1093/bioinformatics/btq683> (2011).
48. Boetzer, M. & Pirovano, W. Toward almost closed genomes with GapFiller. *Genome Biol.* **13**, R56, <https://doi.org/10.1186/gb-2012-13-6-r56> (2012).
49. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069, <https://doi.org/10.1093/bioinformatics/btu153> (2014).
50. Pruitt, K. D., Tatusova, T., Brown, G. R. & Maglott, D. R. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.* **40**, D130–135, <https://doi.org/10.1093/nar/gkr1079> (2012).
51. Cheng, L., Connor, T. R., Siren, J., Aanensen, D. M. & Corander, J. Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Mol. Biol. Evol.* **30**, 1224–1228, <https://doi.org/10.1093/molbev/mst028> (2013).
52. Inouye, M. *et al.* SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med.* **6**, 90, <https://doi.org/10.1186/s13073-014-0090-6> (2014).
53. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
54. Page, A. J. *et al.* SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial Genomics* **2** (2016).
55. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490, <https://doi.org/10.1371/journal.pone.0009490> (2010).
56. Feil, E. J., Li, B. C., Aanensen, D. M., Hanage, W. P. & Spratt, B. G. eBURST: Inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J. Bacteriol.* **186**, 1518–1530, <https://doi.org/10.1128/Jb.186.5.1518-1530.2004> (2004).
57. Poyart, C. *et al.* Multiplex PCR assay for rapid and accurate capsular typing of group B streptococci. *J. Clin. Microbiol.* **45**, 1985–1988 (2007).
58. Kong, F. *et al.* Use of phenotypic and molecular serotype identification methods to characterize previously nonserotypeable group B streptococci. *J. Clin. Microbiol.* **46**, 2745–2750 (2008).
59. Gupta, S. K. *et al.* ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrob. Agents Chemother.* **58**, 212–220 (2014).
60. Wang, R. *et al.* Genome sequence of *Streptococcus agalactiae* strain H002, serotype III, isolated in China from a pregnant woman. *Genome announcements* **3**, e01109–01115 (2015).
61. Glaser, P. *et al.* Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease. *Mol. Microbiol.* **45**, 1499–1513 (2002).
62. Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15–e15 (2014).
63. Stamatakis, A. 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> (2014).
64. Tonkin-Hill, G., Lees, J. A., Bentley, S. D., Frost, S. D. & Corander, J. Fast hierarchical Bayesian analysis of population structure. *Nucleic Acids Res.* **47**, 5539–5549 (2019).
65. Prosperi, M. C. *et al.* A novel methodology for large-scale phylogeny partition. *Nat Commun* **2**, 321, <https://doi.org/10.1038/ncomms1325> (2011).
66. Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol* **2**, vew007, <https://doi.org/10.1093/ve/vew007> (2016).
67. Bouckaert, R. *et al.* BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537, <https://doi.org/10.1371/journal.pcbi.1003537> (2014).
68. Zhang, H., Gao, S., Lercher, M. J., Hu, S. & Chen, W. H. EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res.* **40**, W569–572, <https://doi.org/10.1093/nar/gks576> (2012).
69. He, Z. *et al.* Evolvview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res.* **44**, W236–241, <https://doi.org/10.1093/nar/gkw370> (2016).
70. Page, A. J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693, <https://doi.org/10.1093/bioinformatics/btv421> (2015).
71. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421, <https://doi.org/10.1186/1471-2105-10-421> (2009).
72. Coordinators, N. R. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **44**, D7 (2016).

Acknowledgements

The authors like to thank Mrs Sandra Kit-Bovenkerk for her technical assistance, Dr Simon R. Harris and the Sequencing and Pathogen Informatics groups at the Wellcome Sanger Institute for their bioinformatics support. This work was supported by grant from the Meningitis Research Foundation [Project 1502.0 – ‘Group B streptococcal Genome Library to AvdE’]. DJ, JP and SDB are funded by the Wellcome Trust grant 098051. DvdB is funded by grants from the European Research Council (ERC Starting Grant, proposal/contract 281156; <https://erc.europa.eu/>), the Netherlands Organization for Health Research and Development (ZonMw; NWO-Vidi grant 2010, proposal/contract 016.116.358) and the AMC (innovation grant).

Author contributions

A.v.d.E., T.W.K., M.W.B., D.v.d.B. and S.D.B. conceived the study. J.P. and S.D.B. supervised the project. A.v.d.E., T.W.K., M.W.B. and D.v.d.B. provided samples and clinical information. D.J. and M.d.G. analysed the data. D.J. and S.D.B. prepared the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-66214-3>.

Correspondence and requests for materials should be addressed to D.J.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020