Group B Streptococcal serotype III ST17 lineages associated with increases in late-onset disease incidence in the United States

B. Metcalf, S. Chochua, Y. Li, L. Zhongya, S. Mathis, T. Tran, J. Rivers, S. Nanduri, S. Schrag, B. Beall, L. McGee

Respiratory Diseases Branch, Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, Atlanta, Georgia, USA.

BACKGROUND

- Group B Streptococcus (GBS) late-onset disease (LOD) occurs between 7 to 89 days of life and is commonly associated with bacteremia and meningitis
- Centers for Disease Control and Prevention (CDC) uses the Active Bacterial Core surveillance (ABCs) program to perform active laboratory- and population-based surveillance of invasive GBS disease in 10 US states
- Previous work showed that LOD incidence caused by serotype III sequence type 17 (ST17) increased significantly from 2006 to 2015
- We sequenced 574 serotype III ST17 isolates from LOD cases (1995-2018) to identify genomic determinants associated with this increase

METHODS

- Identification of expanding lineages was carried out with the 'treeStructure v0.1.0' R package using time-scaled phylogeny generated by 'TreeTime v0.8.3.1'
- The 'Scoary v1.6.16' GWAS tool was used to find genes associated with the lineage L1 based on a pangenome calculated with 'Roary v1.007002' and the 'Prokka v1.14.5' software.
- The R program 'hclust' was used to cluster the gene presence/absence matrix from Roary to identify putative accessory elements
- The 'unitig-counter v1.1.0' program was used to call unitigs representing variation across the pan-genome and 'pyseer v1.3.3' identified unitigs associated with lineage L1
- Characterization of genes associated with this expanding lineage was performed using 'interproscan v5.30-69.0', 'PSORTb v3.0.4', and the VFDB VFanalyzer tool.
- BEAST v2.5.1 was used to generate a time-scaled phylogeny and the 'phylodyn v 0.9.02' R package estimated the effective population size trajectory

RESULTS

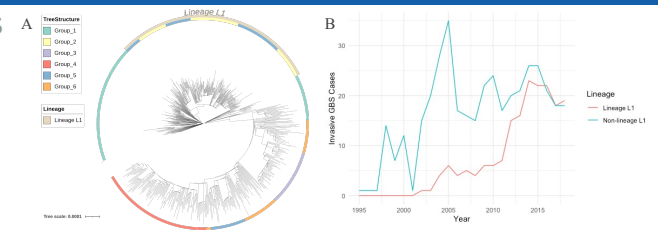

Figure 1. The increase of LOD among GBS Serotype ST17 strains is primarily driven by the expansion of a specific sublineage (designated as lineage L1). (A) Phylogeny of GBS serotype 3 ST17 isolates acquired through the ABC program clustered by groups defined using the TreeStructure program. Lineage L1 is comprised of alternating TreeStructure groups 2 and 5. (B) Plot of lineage L1 and non-lineage L1 invasive GBS serotype 3 ST17 cases over time indicate that cases from lineage L1 have increased significantly and particularly after 2010.

RESULTS

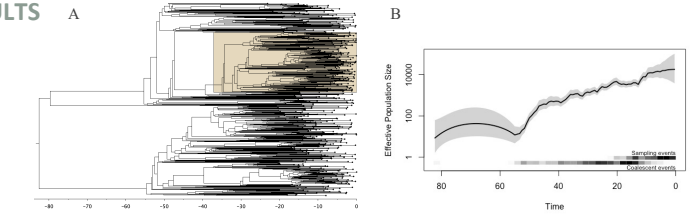

Figure 2. Time-scaled phylogeny and effective population size trajectory of GBS serotype 3 ST17. (A) A Beast time-measured phylogeny with lineage L1 highlighted in beige. The most recent common ancestor of L1 occurred ~35 years ago. Tree indicates that a particular sublineage is not responsible for the expansion of LOD over the past decade as isolates are spread across L1. (B) Plot of the effective population size of GBS serotype 3 ST17 over time using the phylodyn R package. The graph shows a rapid expansion occurring ~50 years ago followed by a steady increase and then a second expansion ~10 years ago.

RESULTS

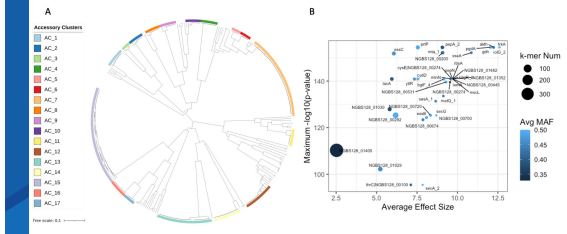

Figure 3. Overview of lineage L1-associated genes identified using the Scoary and Pyseer programs. (A) Using a p-value < 0.05 after Bonferroni correction for multiple tests, Scoary identified 323 genes associated with lineage L1. Clustering the Roary presence/absence matrix with hclust identified 17 clusters containing at least 5 genes which have been annotated on a phylogeny based on hclust derived distances. (B) Pyseer identified 2888 unique L1-associated unitigs (with Bonferroni correction) that mapped to 367 genes.

RESULTS

Gene ID	Software	Accessory Cluster	Mutation Type	PsortB Classification
sasA_1	Pyseer	-	Non-Syn	CytoplasmicMembrane
essB	Pyseer	-	Non-Syn	CytoplasmicMembrane
NGBS128_01352	Pyseer	-	Non-Syn	CytoplasmicMembrane
NGBS128_00200	Pyseer	-	Non-Syn	CytoplasmicMembrane
cydD	Pyseer	-	Non-Syn	CytoplasmicMembrane
bglF_4	Pyseer	-	Non-Syn	CytoplasmicMembrane
yagG	Pyseer	-	Non-Syn	CytoplasmicMembrane
mscL	Pyseer	-	Non-Syn	CytoplasmicMembrane
metQ_1	Pyseer	-	Non-Syn	CytoplasmicMembrane
NGBS128_01029	Pyseer	-	Non-Syn	Extracellular
group_1241	Scoary	AC_6	-	Extracellular
group_1613	Scoary	AC_7	-	CytoplasmicMembrane
group_183	Scoary	AC_1	-	CytoplasmicMembrane
group_240	Scoary	AC_15	-	CytoplasmicMembrane
group_1251	Scoary	AC_12	-	CytoplasmicMembrane
group_1361	Scoary	AC_8	-	CytoplasmicMembrane
group_2095	Scoary	AC_14	-	CytoplasmicMembrane
group_857	Scoary	AC_13	-	CytoplasmicMembrane

Table 1. A selection of L1-associated genes of interest using Scoary and Pyseer. For Pyseer, 10 genes were chosen based on a significant Bonferroni corrected p-value, a non-synonymous coding mutation, and a non-cytoplasmic PSORTb designation. Nine were cytoplasmic membrane-associated and one was extracellular. For Scoary, genes required a significant Bonferroni corrected p-value and a non-cytoplasmic PSORTb designation. A single gene was chosen if an accessory cluster contained multiple genes of interest. With these specifications, 8 genes were identified. Seven were cytoplasmic membrane-associated, and one was extracellular.

CONCLUSIONS

- An expanding lineage within serotype III ST17 (lineage L1) is primarily responsible for increasing invasive disease rates among serotype III LOD within ABCs. This result is consistent with previous work [1, 2]. Notably, there has been a significant increase in cases from L1 since approximately 2010.
- Using the Scoary program, we identified 323 genes associated with lineage L1. These genes belonged to 64 putative accessory clusters 17 of which contained at least 5 genes. PSORTb identified 37 genes with non-cytoplasmic localization sites.
- Using the Pyseer bacterial GWAS program, we identified 367 L1-associated genes. Some of these genes have also been associated with an expanding ST17 LOD lineage in previous work [1]. The PSORTb program classified 19 genes as having non-cytoplasmic localization sites though only 10 contained non-synonymous mutations.

REFERENCES

1. D. Jamrozy et al., Increasing incidence of group B streptococcus neonatal infections in the Netherlands is associated with clonal expansion of CC17 and CC23. Sci Rep 10, 9539 (2020).
2. S. Teatero 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 (2016).

ACKNOWLEDGEMENTS

Major funding for this work was provided through support from CDC's Advanced Molecular Detection (AMD) initiative and CDC's Emerging Infection Program.

The findings and conclusions in this report are those of the authors, and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Address correspondence to Ben Metcalf, ycm6@cdc.gov.

National Center for Immunization and Respiratory Diseases
Division Of Bacterial Diseases