Contents

[Comments and questions about the data 2](#_Toc180584713)

[Comments and questions about Figures 3](#_Toc180584714)

[Figure 1 – Serotype and CC distribution across Age groups and time 3](#_Toc180584715)

[Table 1 – Association of Infant, ICU and Meningitis with Laboratory metrics 3](#_Toc180584716)

[Table 2 – Odds of severe disease by virulence factor 3](#_Toc180584717)

[Figure 2 – Global and BCH Phylogenies 3](#_Toc180584718)

[Supplementary Table 4 – ST genes 4](#_Toc180584719)

[Supplementary Table 5 – Distribution of vaccine targets across age groups 4](#_Toc180584720)

[Phylogeny Outgroup 5](#_Toc180584721)

[Phylogeny Distribution of Same Patient isolates 6](#_Toc180584722)

[Phylogeny Distribution of Twins 7](#_Toc180584723)

[Clinical Risk Factor Analysis Proposal 9](#_Toc180584724)

[Proposal 9](#_Toc180584725)

[Preliminary results 11](#_Toc180584726)

## Comments and questions about the data

* The original Raw dataset contains only 98 samples, although 100 patients are reported. Why ?
* In Carl’s dataset:
  + 35\_S35 is included, but should be treated as “contamination” and excluded (%GBS= 13)
  + 26\_S26 and 93\_S93 are excluded. Why ?
  + 86\_S86 has days\_at\_dx = 4651. It should be 4590, since date of Diagnosis is 2020-02-25 and DoB is 2007-08-02
  + 59\_S59 is isolated from “blood” instead of “both blood and meningitis”, and thus as “No Meningitis”, instead of “Meningitis”
  + 21\_S2, 34\_S34, and 51\_S51 are classified as not having Neutropenia although they have ANC <1.50 *\*10^3 cells/mm3* (1.32, 1.49 and 1.29, respectively)
  + 74\_S74: DoB missing but Days at dx is there.
  + I believe (TBC) Leukocytosis and Leucopenia were coded from ALC (Absolute Lymphocyte Count) rather than WBC.
* Any differences in genomic characteristics are due to my updating of Bioinformatics methods to meet gold standards.

## Comments and questions about Figures

### Figure 1 – Serotype and CC distribution across Age groups and time

* I propose a histogram showing serotype and CC frequency over time, not divided into ICU vs Non-ICU, because this is relevant only in the subsequent “Clinical characteristics” part.

### Table 1 – Association of Infant, ICU and Meningitis with Laboratory metrics

* Multiple testing correction – Use bonferonni ?
* The presumed causal direction of the associations investigated is from “Infant” and “Meningitis” to the various laboratory metrics, and from the various laboratory metrics to “ICU admission”. For the binary exposures of “Infant” and “Meningitis”, we used Wilcoxon test to investigate their association with continuous laboratory metrics and logistic regression for the binary metrics.For the binary outcome of “ICU admission”, we used logistic regression to investigate the effect of both continuous and binary laboratory metrics.
* Ideally, due to the different direction of the causal relationships presented in this table, it should be separated into 2 tables, and contain risk and odds rations + 95%

CIs.

### Table 2 – Odds of severe disease by virulence factor

* Multiple testing correction – Use bonferonni ?
* Do we want to add an adjusted OR ? i.e. to control for potential confounders ? (e.g. CC)
* Do we want to check for association with
  + Meningitis?
  + Age of disease onset ?
  + Neutropenia, leukopenia etc

### Figure 2 – Global and BCH Phylogenies

* Global phylo: Make Circular ?
* BCH phylo: GAS was used as an Outgroup, but is not shown on phylogeny because it shows up as long branch such that the diversity within GBS becomes difficult to see.

### Supplementary Table 4 – ST genes

* The table originally contained Sequence Cluster-Specific accessory genes, as identified by a pangenome analysis… I replaced it with the table I believe was intended. Is this what you had in mind ?

### Supplementary Table 5 – Distribution of vaccine targets across age groups

* Should I keep the SIP1a and SIP3a rows ?

## Phylogeny Outgroup

A group of people's groups

Description automatically generated with medium confidenceA group of black lines with different colored dots

Description automatically generated

## Phylogeny Distribution of Same Patient isolates

A diagram of a patient

Description automatically generated

## Phylogeny Distribution of Twins

A diagram of a tree

Description automatically generatedA diagram of a company

Description automatically generated with medium confidence

## Clinical Risk Factor Analysis Proposal

### Proposal

Original, Carl looked at:

1. The association a) Disease in Infants, b) ICU admission, and c) Meningitis with various laboratory metrics (ANC, WBC, Platelet and Hemoglobin count; leukopenia, leukocytosis, Neutropenia)
   1. Why is it interesting to study this ?
   2. What is the directionality of these ?
      1. Age of patient is probably upstream of abnormal laboratory metrics
      2. ICU admission is probably downstream
      3. Meningitis could be both

A diagram of a patient flow

Description automatically generated

1. The association of various virulence genes with ICU admission.
   1. Should we control for any confounders ? Population structure MUST be controlled for (by controlling for CC for instance)
   2. Can we extend this to study other clinical outcomes
      1. ICU (Table 1)
      2. Meningitis (Table 2)
      3. Age of disease onset: Infant vs not infant (Table 3)
      4. Age of disease onset among Infants: LOD vs VLOD (Table 4)
      5. Current Abnormal laboratory metrics (leukopenia, leukocytosis, Neutropenia) (Table 5, 6, 7)
      6. Other Abnormal laboratory metrics ? (E.g. neutrophilia, Anemia/ Polycythemia, Lymphocytopenia/ Lymphocytosis, Thrombocytopenia/ Thrombocytosis… etc)
   3. Finally, re-iterating the possibilityrelevance of conducting a panGWAS.

A diagram of a cell phone

Description automatically generated

### Preliminary results

Table 1. ICU admission among infants

| Risk Factor | ICU (n=31) | Other (n=39) | Crude OR (95% CI) | P-value | Adjusted OR (95% CI) | Adjusted P-value | Significant |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ALP1 | 10 | 5 | 3.24 (0.97 - 10.79) | 0.056 | 4.54 (1.05 - 19.60) | 0.043 | No |
| ALP23 | 0 | 2 | 0.00 (0.00 - Inf) | 0.992 | 0.00 (0.00 - Inf) | 0.992 | No |
| ALPHA | 4 | 3 | 1.78 (0.37 - 8.61) | 0.475 | 1.68 (0.34 - 8.19) | 0.523 | No |
| HVGA | 17 | 29 | 0.42 (0.15 - 1.15) | 0.091 | 0.38 (0.13 - 1.05) | 0.063 | No |
| PI1 | 18 | 31 | 0.36 (0.12 - 1.03) | 0.056 | 0.35 (0.12 - 1.03) | 0.057 | No |
| PI2A1 | 14 | 7 | 3.76 (1.28 - 11.10) | 0.016 | 3.69 (1.25 - 10.91) | 0.018 | No |
| PI2A2 | 3 | 3 | 1.29 (0.24 - 6.86) | 0.769 | 1.26 (0.24 - 6.76) | 0.783 | No |
| PI2B | 13 | 25 | 0.40 (0.15 - 1.06) | 0.067 | 0.37 (0.14 - 0.99) | 0.049 | No |
| RIB | 17 | 12 | 2.73 (1.02 - 7.29) | 0.045 | 3.01 (1.11 - 8.16) | 0.031 | No |
| SRR1 | 14 | 26 | 0.41 (0.16 - 1.09) | 0.073 | 0.38 (0.14 - 1.01) | 0.053 | No |
| SRR2 | 3 | 5 | 0.73 (0.16 - 3.32) | 0.682 | 0.87 (0.18 - 4.23) | 0.868 | No |
| Sip.1a | 28 | 34 | 1.37 (0.30 - 6.25) | 0.682 | 1.14 (0.24 - 5.53) | 0.868 | No |
| Sip.3a | 14 | 26 | 0.41 (0.16 - 1.09) | 0.073 | 0.38 (0.14 - 1.01) | 0.053 | No |
| lmb | 30 | 39 | 0.00 (0.00 - Inf) | 0.991 | 0.00 (0.00 - Inf) | 0.991 | No |
| HVGA | 30 | 38 | 0.79 (0.05 - 13.15) | 0.869 | 0.86 (0.05 - 14.57) | 0.920 | No |
| scpB | 28 | 37 | 0.50 (0.08 - 3.23) | 0.470 | 0.51 (0.08 - 3.28) | 0.480 | No |
| hylB | 24 | 31 | 0.88 (0.28 - 2.78) | 0.834 | 0.81 (0.25 - 2.61) | 0.723 | No |
| fbsB | 24 | 31 | 0.88 (0.28 - 2.78) | 0.834 | 0.81 (0.25 - 2.61) | 0.723 | No |

\* Adjusted OR is calculated by controlling for CC (a proxy of population structure)

\* Bonferonni Multiple testing correction: significance level is 0.05/ 17= 0.003

Table 2. Meningitis among infants

| **Risk Factor** | **Meningitis (n=13)** | **No Meningitis (n=57)** | **Crude OR (95% CI)** | **P-value** | **Adjusted OR (95% CI)** | **Adjusted P-value** | **Significant** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ALP1 | 4 | 11 | 1.86 (0.48 - 7.16) | 0.368 | 2.15 (0.53 - 8.76) | 0.286 | no |
| ALP23 | 0 | 2 | 0.00 (0.00 - Inf) | 0.993 | 0.00 (0.00 - Inf) | 0.993 | no |
| ALPHA | 1 | 6 | 0.71 (0.08 - 6.45) | 0.760 | 0.68 (0.07 - 6.20) | 0.728 | no |
| HVGA | 8 | 38 | 0.80 (0.23 - 2.78) | 0.726 | 0.76 (0.22 - 2.66) | 0.666 | no |
| PI1 | 7 | 42 | 0.42 (0.12 - 1.44) | 0.166 | 0.41 (0.12 - 1.45) | 0.167 | no |
| PI2A1 | 5 | 16 | 1.60 (0.46 - 5.63) | 0.463 | 1.58 (0.45 - 5.56) | 0.477 | no |
| PI2A2 | 2 | 4 | 2.41 (0.39 - 14.83) | 0.343 | 2.39 (0.39 - 14.69) | 0.348 | no |
| PI2B | 4 | 34 | 0.30 (0.08 - 1.09) | 0.068 | 0.29 (0.08 - 1.04) | 0.058 | no |
| RIB | 6 | 23 | 1.27 (0.38 - 4.26) | 0.702 | 1.33 (0.39 - 4.49) | 0.649 | no |
| SRR1 | 6 | 34 | 0.58 (0.17 - 1.95) | 0.378 | 0.55 (0.16 - 1.87) | 0.339 | no |
| SRR2 | 2 | 6 | 1.55 (0.27 - 8.70) | 0.621 | 1.80 (0.31 - 10.52) | 0.516 | no |
| Sip.1a | 11 | 51 | 0.65 (0.11 - 3.64) | 0.621 | 0.56 (0.10 - 3.26) | 0.516 | no |
| Sip.3a | 6 | 34 | 0.58 (0.17 - 1.95) | 0.378 | 0.55 (0.16 - 1.87) | 0.339 | no |
| lmb | 13 | 56 | 1336635.14 (0.00 - Inf) | 0.992 | 1394279.21 (0.00 - Inf) | 0.992 | no |
| scpB | 13 | 55 | 3699412.52 (0.00 - Inf) | 0.993 | 4003741.99 (0.00 - Inf) | 0.993 | no |
| hylB | 12 | 53 | 0.91 (0.09 - 8.85) | 0.932 | 0.91 (0.09 - 8.94) | 0.939 | no |
| fbsB | 10 | 45 | 0.89 (0.21 - 3.75) | 0.873 | 0.84 (0.20 - 3.58) | 0.818 | no |

Table 3. LOD vs VLOD among infants

| **Risk Factor** | **LOD (n=48)** | **VLOD (n=20)** | **Crude OR (95% CI)** | **P-value** | **Adjusted OR (95% CI)** | **Adjusted P-value** | **Significant** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ALP1 | 7 | 6 | 0.40 (0.11 - 1.39) | 0.148 | 0.03 (0.00 - 0.32) | 0.004 | no |
| ALP23 | 1 | 1 | 0.40 (0.02 - 6.80) | 0.529 | 0.55 (0.03 - 11.65) | 0.703 | no |
| ALPHA | 3 | 4 | 0.27 (0.05 - 1.32) | 0.106 | 0.29 (0.06 - 1.53) | 0.144 | no |
| RIB | 37 | 9 | 4.11 (1.36 - 12.46) | 0.012 | 5.00 (1.56 - 16.05) | 0.007 | no |
| PI1 | 36 | 13 | 1.62 (0.52 - 4.99) | 0.404 | 2.94 (0.72 - 12.07) | 0.135 | no |
| **PI2A1** | **8** | **11** | **0.16 (0.05 - 0.52)** | **0.002** | **0.14 (0.04 - 0.50)** | **0.002** | **yes** |
| PI2A2 | 6 | 0 | 20259434.45 (0.00 - Inf) | 0.992 | 19269895.97 (0.00 - Inf) | 0.992 | no |
| PI2B | 30 | 8 | 2.50 (0.86 - 7.28) | 0.093 | 2.91 (0.97 - 8.75) | 0.056 | no |
| SRR1 | 16 | 11 | 0.41 (0.14 - 1.19) | 0.100 | 0.35 (0.12 - 1.05) | 0.060 | no |
| SRR2 | 31 | 9 | 2.23 (0.77 - 6.44) | 0.139 | 2.62 (0.88 - 7.84) | 0.085 | no |
| Sip.1a | 7 | 1 | 3.24 (0.37 - 28.26) | 0.287 | 3.30 (0.33 - 32.54) | 0.306 | no |
| Sip.3a | 41 | 19 | 0.31 (0.04 - 2.69) | 0.287 | 0.30 (0.03 - 2.99) | 0.306 | no |
| HVGA | 31 | 9 | 2.23 (0.77 - 6.44) | 0.139 | 2.62 (0.88 - 7.84) | 0.085 | no |
| lmb | 47 | 20 | 0.00 (0.00 - Inf) | 0.992 | 0.00 (0.00 - Inf) | 0.992 | no |
| scpB | 47 | 19 | 2.47 (0.15 - 41.61) | 0.529 | 1.34 (0.04 - 41.52) | 0.867 | no |
| hylB | 43 | 20 | 0.00 (0.00 - Inf) | 0.992 | 0.00 (0.00 - Inf) | 0.992 | no |
| fbsB | 38 | 15 | 1.27 (0.37 - 4.33) | 0.706 | 1.11 (0.26 - 4.71) | 0.886 | no |

Table 4. Infant vs older

| **Risk Factor** | **Infant (n=70)**  <dbl> | **Older (n=17)**  <dbl> | **Crude OR (95% CI)** | **P-value** | **Adjusted OR (95% CI)** | **Adjusted P-value** | **Significant** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ALP1 | 15 | 7 | 0.39 (0.13 - 1.20) | 0.100 | 0.65 (0.17 - 2.40) | 0.516 | no |
| **ALP23** | **2** | **5** | **0.07 (0.01 - 0.41)** | **0.003** | **0.05 (0.01 - 0.30)** | **0.001** | **yes** |
| ALPHA | 7 | 1 | 1.78 (0.20 - 15.51) | 0.603 | 1.42 (0.16 - 12.60) | 0.752 | no |
| RIB | 46 | 4 | 6.23 (1.83 - 21.20) | 0.003 | 5.03 (1.42 - 17.77) | 0.012 | no |
| PI1 | 49 | 12 | 0.97 (0.30 - 3.11) | 0.962 | 1.22 (0.37 - 4.06) | 0.746 | no |
| PI2A1 | 21 | 8 | 0.48 (0.16 - 1.42) | 0.186 | 0.33 (0.10 - 1.07) | 0.065 | no |
| PI2A2 | 6 | 3 | 0.44 (0.10 - 1.96) | 0.281 | 0.35 (0.08 - 1.63) | 0.183 | no |
| PI2B | 38 | 1 | 19.00 (2.39 - 151.22) | 0.005 | 15.97 (1.98 - 128.94) | 0.009 | no |
| SRR1 | 29 | 15 | 0.09 (0.02 - 0.44) | 0.003 | 0.11 (0.02 - 0.55) | 0.007 | no |
| SRR2 | 40 | 1 | 21.33 (2.68 - 169.91) | 0.004 | 17.97 (2.22 - 145.24) | 0.007 | no |
| Sip.1a | 8 | 9 | 0.11 (0.03 - 0.38) | 0.000 | 0.15 (0.04 - 0.55) | 0.004 | no |
| Sip.3a | 62 | 8 | 8.72 (2.62 - 29.06) | 0.000 | 6.73 (1.82 - 24.86) | 0.004 | no |
| HVGA | 40 | 1 | 21.33 (2.68 - 169.91) | 0.004 | 17.97 (2.22 - 145.24) | 0.007 | no |
| lmb | 69 | 17 | 0.00 (0.00 - Inf) | 0.992 | 0.00 (0.00 - Inf) | 0.992 | no |
| scpB | 68 | 14 | 7.29 (1.11 - 47.72) | 0.038 | 10.15 (1.50 - 68.50) | 0.017 | no |
| hylB | 65 | 15 | 1.73 (0.31 - 9.81) | 0.534 | 2.10 (0.37 - 12.11) | 0.405 | no |
| fbsB | 55 | 7 | 5.24 (1.71 - 16.09) | 0.004 | 4.02 (1.22 - 13.25) | 0.022 | no |

Table 5. Neutropenia among Infant

| **Risk\_Factor**  <chr> | **Infant (n=70)**  <dbl> | **Older (n=17)**  <dbl> | **Crude OR (95% CI)**  <chr> | **P-value**  <dbl> | **Adjusted OR (95% CI)**  <chr> | **Adjusted P-value**  <dbl> | **Significant**  <chr> |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ALP1 | 4 | 11 | 1.12 (0.30 - 4.12) | 0.866 | 1.29 (0.33 - 5.06) | 0.710 | no |
| ALP23 | 0 | 2 | 0.00 (0.00 - Inf) | 0.993 | 0.00 (0.00 - Inf) | 0.993 | no |
| ALPHA | 2 | 5 | 1.23 (0.22 - 6.99) | 0.818 | 1.16 (0.20 - 6.68) | 0.872 | no |
| RIB | 11 | 33 | 1.00 (0.32 - 3.15) | 1.000 | 0.94 (0.30 - 2.99) | 0.916 | no |
| PI1 | 12 | 35 | 1.10 (0.33 - 3.64) | 0.880 | 1.09 (0.33 - 3.65) | 0.885 | no |
| PI2A1 | 6 | 15 | 1.31 (0.41 - 4.19) | 0.650 | 1.29 (0.40 - 4.13) | 0.670 | no |
| PI2A2 | 3 | 3 | 3.43 (0.62 - 18.91) | 0.157 | 3.39 (0.61 - 18.73) | 0.161 | no |
| PI2B | 6 | 30 | 0.38 (0.12 - 1.19) | 0.098 | 0.36 (0.11 - 1.13) | 0.080 | no |
| SRR1 | 8 | 21 | 1.27 (0.42 - 3.83) | 0.671 | 1.35 (0.44 - 4.09) | 0.601 | no |
| SRR2 | 8 | 30 | 0.62 (0.21 - 1.88) | 0.399 | 0.59 (0.19 - 1.78) | 0.346 | no |
| Sip.1a | 3 | 5 | 1.97 (0.42 - 9.30) | 0.391 | 2.40 (0.48 - 12.02) | 0.286 | no |
| Sip.3a | 14 | 46 | 0.51 (0.11 - 2.39) | 0.391 | 0.42 (0.08 - 2.08) | 0.286 | no |
| HVGA | 8 | 30 | 0.62 (0.21 - 1.88) | 0.399 | 0.59 (0.19 - 1.78) | 0.346 | no |
| lmb | 17 | 50 | 1957656.39 (0.00 - Inf) | 0.992 | 2078351.00 (0.00 - Inf) | 0.992 | no |
| scpB | 17 | 49 | 5430063.84 (0.00 - Inf) | 0.993 | 6193694.55 (0.00 - Inf) | 0.993 | no |
| hylB | 15 | 48 | 0.47 (0.07 - 3.07) | 0.430 | 0.47 (0.07 - 3.11) | 0.436 | no |
| fbsB | 12 | 41 | 0.59 (0.17 - 2.05) | 0.402 | 0.55 (0.15 - 1.94) | 0.351 | no |

Table 6. Leukopenia among Infant

| **Risk Factor** | **Leukopenia (n=51)** | **No Leukopenia (n=17)** | **Crude OR (95% CI)** | **P-value** | **Adjusted OR (95% CI)** | **Adjusted P-value** | **Significant** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ALP1 | 13 | 2 | 2.57 (0.52 - 12.76) | 0.250 | 5.66 (0.65 - 49.05) | 0.115 | no |
| ALP23 | 1 | 1 | 0.32 (0.02 - 5.41) | 0.430 | 0.28 (0.02 - 4.85) | 0.384 | no |
| ALPHA | 3 | 4 | 0.20 (0.04 - 1.02) | 0.053 | 0.17 (0.03 - 0.90) | 0.037 | no |
| RIB | 34 | 10 | 1.40 (0.45 - 4.32) | 0.559 | 1.20 (0.37 - 3.85) | 0.764 | no |
| PI1 | 33 | 14 | 0.39 (0.10 - 1.55) | 0.182 | 0.40 (0.10 - 1.60) | 0.196 | no |
| PI2A1 | 17 | 4 | 1.63 (0.46 - 5.75) | 0.451 | 1.53 (0.43 - 5.46) | 0.512 | no |
| PI2A2 | 5 | 1 | 1.74 (0.19 - 16.03) | 0.625 | 1.66 (0.18 - 15.33) | 0.656 | no |
| PI2B | 27 | 9 | 1.00 (0.33 - 3.00) | 1.000 | 0.87 (0.28 - 2.69) | 0.805 | no |
| SRR1 | 21 | 8 | 0.79 (0.26 - 2.37) | 0.671 | 0.91 (0.29 - 2.83) | 0.866 | no |
| SRR2 | 29 | 9 | 1.17 (0.39 - 3.53) | 0.778 | 1.02 (0.33 - 3.16) | 0.978 | no |
| Sip.1a | 6 | 2 | 1.00 (0.18 - 5.49) | 1.000 | 1.87 (0.22 - 16.06) | 0.570 | no |
| Sip.3a | 45 | 15 | 1.00 (0.18 - 5.49) | 1.000 | 0.54 (0.06 - 4.61) | 0.570 | no |
| HVGA | 29 | 9 | 1.17 (0.39 - 3.53) | 0.778 | 1.02 (0.33 - 3.16) | 0.978 | no |
| lmb | 50 | 17 | 0.00 (0.00 - Inf) | 0.992 | 0.00 (0.00 - Inf) | 0.992 | no |
| scpB | 49 | 17 | 0.00 (0.00 - Inf) | 0.993 | 0.00 (0.00 - Inf) | 0.993 | no |
| hylB | 47 | 16 | 0.73 (0.08 - 7.06) | 0.789 | 0.77 (0.08 - 7.44) | 0.822 | no |
| fbsB | 42 | 11 | 2.55 (0.75 - 8.69) | 0.136 | 2.23 (0.62 - 7.98) | 0.217 | no |

Table 7. Leukocytosis among Infant

Not done because Leukocytosis (n=67 ) and No Leukocytosis (n=1)