Reviewer’s report

Title: Genomic comparative analysis and gene function prediction in infectious diseases: Application to the investigation of a meningitis outbreak

Version: 2 Date: 27 August 2013

Reviewer: John Kriesel

Reviewer’s report:

Lavezzo et al present an interesting vignette describing comparative genotyping of strains isolated during before, during and after a N. meningitidis outbreak. They discover that one of their isolates, temporally associated with the outbreak, was of a different genotype as by deep sequencing, a finding missed by conventional typing methods. The differences found in the strains allowed generation of markers of the outbreak strain, which in turn allowed a streamlined analysis of more samples. This manuscript is a useful synthesis and application of methods that are not novel in and of themselves. However the ID community may benefit from this particular presentation.

Major requested revisions here are mainly for clarity. They are not intended to be critical about the scientific validity of the work.

Major Revisions
1) the use of the BLASTp cutoff of 90%/90% is not justified, and, as it is a critical parameter for everything that comes after it, some treatment in the Results (and a supplementary figure) that justifies the 90/90 choice and explains how relaxing these parameters would affect the choice of diagnostic genomic regions and the # of genes with "unknown function" would be helpful for some readers, embarking on similar studies. Also, the parameters used in the BLASTp runs should be specified.

2) It appears that only 2 strains, the 1st and last in the outbreak, were fully sequenced. The rest were analyzed by specific PCR. Why weren't all 9 meningococcal isolates deep sequenced? What were the constraints preventing this strategy? The specific sequencing and PCR strategy employed, including N's, should be clearly explained in the Abstract.

3) Figure 2: It is not clear that the FAM18 panels are important to
the results of this paper. And an ID audience would benefit from a simplified explanation of the chromosomal inversion in the text when figure 2 is introduced. Mentioning how this is also shown in blue in Figure 3, may help the reader to link these 2 ways of looking at the inversion. Also eliminating the FAM18 panels will help focus on what I believe the authors are trying to communicate.

4) Figure 3: It is essential to label the tracks with the strain identities.

Minor Revisions

5) The sequencing results belong in the results, not the methods, section. (i.e. number of reads, etc.)

6) The Table 1 legend does not specify which strains are outbreak strains and which are not. It also does not specify which strains were sequenced and which were analyzed by PCR. The legend should "stand alone".

7) N. menigitidis strain alpha275 (Genbank AM889138) - The authors say this is fully sequenced, but it is actually in contig form.

8) The authors reasonably suggest that horizontal gene transfer is implicated in generation of the K1207 outbreak strain and go on to say: "conventional genotyping methods that are used to investigate outbreaks and to monitor the circulation of new variants of hyper-virulent strains may not be accurate enough" These important issues of horizontal gene transfer should be addressed in somewhat more depth.

9.1) Are there any known strains of related bacteria with syntenic regions containing sacB, e.g.

9.2) Would addition of alpha 275 to Figure 3 help illustrate sacBand neighboring genes another context?

10) The authors should discuss the relative benefits and costs of their approach as opposed to simply sequencing all the isolates.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**
I declare that I have no competing interests