ARTICLE DETAILS

TITLE (PROVISIONAL)	Molecular characterisation of Neisseria gonorrhoeae associated with disseminated gonococcal infections in Queensland, Australia: a retrospective surveillance study
AUTHORS	Guglielmino, Christine; Sandhu, Sumeet; Lau, Colleen L.; Buckely, Cameron; Trembizki, Ella; Whiley, David; Jennison, Amy

VERSION 1 – REVIEW

REVIEWER	Mortimer, Tatum Harvard University T H Chan School of Public Health, Immunology and Infectious Diseases
REVIEW RETURNED	24-Feb-2022

GENERAL COMMENTS	In this manuscript, Guglielmino et al. analyzed cultured N. gonorrhoeae isolates from uncomplicated and disseminated infections in Queensland from 2010-2015 along with associated patient demographic data. They find that DGI occurs more often in particular demographic groups (women, men > 30 years). They also analyzed N. gonorrhoeae genetic factors that may be associated with DGI. Using NG-MAST, they confirm that the porB1a allele is associated with DGI. Intriguingly, they also find that particular NG-MAST are associated with DGI, suggesting that additional genetic factors contribute to DGI beyond the porB allele. The analysis includes whole genome sequencing of 16 isolates to further investigate additional loci; however, the sampling does not allow for conclusions to be drawn from the WGS data because isolates from sequence types that were not associated with DGI were not sequenced. As there are limited examples of WGS from DGI infections, these genomes are an important contribution to the field, and the association between NG-MAST and DGI suggests that further investigations into N. gonorrhoeae genetic loci contributing to DGI are warranted. Major comments: 1. WGS sampling: It is unclear how the isolates were chosen for WGS. The authors performed WGS to investigate additional genomic factors that might lead to increased risk for DGI. However, the sampling strategy does not allow these comparisons to be made. In Lines 102-103, the authors state that “two strains from each NG-MAST were selected, comprising both DGI and non-DGI strains”. To me, this implies that the WGS dataset should include 8 DGI isolates and 8 non-DGI isolates. However, in Figure 1, 12 isolates are annotated as DGI and 4 are annotated as non-DGI. Can the authors clarify how isolates were chosen for WGS? 2. Statistical analysis: In Line 117 of the methods, the authors state that $p<0.05$ was used to determine significance of association with DGI. However, the authors tested demographic factors (including several combinations of sex and age), porB

	types, and NG-MAST for association with DGI. Was multiple hypothesis test correction included? 3. Core genome phylogeny: The methods for the core genome phylogeny are not included in the manuscript. How was the core genome determined? What phylogenetic software was used? Were recombinant regions masked? In Figure 1, it is not clear how the phylogeny is rooted and what the scale bar indicates. 4. Locus specific analyses: The authors focused on analyzing the presence of the GGI and phase variable repeats in pglA. However, there are additional loci that have been shown to be involved in N. gonorrhoeae serum resistance (e.g. the lgt operon). Also, these loci are not examined in isolates from NG-MAST that are not associated with DGI in this study, so I'm not sure what conclusions can be drawn from the results presented here. For example, the results here are not necessarily inconsistent with Power et al (Line 208) or provide evidence one way or the other about whether DGI is associated with GGI (Line 208-209). Minor comments: 5. Line 33: In Figure 1, 3/12 DGI isolates are shown to have pglA phase variable allele, so this allele was in the minority in DGI isolates sequenced, and the sampling of non-DGI isolates is unlikely to be representative of all uncomplicated infections. 6. Lines 97-98: A phylogeny based on a concatenated alignment of porB and tbpB is unlikely to reflect genome-wide relationships between isolates due to the high recombination rate in these genes. This phylogeny is not actually included in the results, so I suggest that this be removed. 7. Line 106: Version numbers and parameters should be included for assembly software. 8. Line 109-110: WGS reads should be uploaded to NCBI or ENA, not just assemblies in PubMLST. 9. Line 130-131/Table 1: The number of infections reported in males < 30 is not the same in the text and Table 1. 10. Table 2: N. gonorrhoeae is misspelled in the second column 11. Lines 200, 205: These references appear to be incorrect. 12. Line 33,206: pglA instead of plgA 13. Supplementary Table 1 is great. Could a column be added to indicate accession numbers for the 16 isolates that were sequenced?

REVIEWER	Abad, Raquel Instituto de Salud Carlos III, National Centre for Microbiology
REVIEW RETURNED	28-Mar-2022

GENERAL COMMENTS	The study summarizes the molecular characterization by NG-MAST of 3953 Neisseria gonorrhoeae isolates from pathology laboratories servicing Queensland and surrounding areas from January 2010 to August 2015. Association of the NG-MAST data along basic demographic factors (age, sex, and geographical location) with disseminated gonococcal infection (DGI) were also analyzed. The manuscript addresses an important issue such as the monitoring and characterization of the circulating gonococcal strains, providing insight into the population structure of N. gonorrhoeae. Only several points would need to be revised/clarified: - Abstract, page 2, lines 33-34. The statement "WGS demonstrated that NG-MAST types having a plgA phase variation were more commonly detected in DGI" would not fully warranted from this study: on the one hand only 8 NG-MAST types, all of

	them detected in DGI, were analyzed by WGS, and only 2 NG-MAST types showed plgA phase variation; on the other hand, only 2 strains of each NG-MAST type were analyzed. - Methods, page 6, lines 96-97. “The porB sequence data from the NG-MAST was analysed to assign either PIA or PIB class”. Since the study associated strains harbouring PIA class of porB type with DGI, analysis to assign either PIA or PIB class from the porB sequence data should be more detailed. - Results, page 8, line 140: porB 4101 instead ST4101. - Results, page 8, “NG-MAST typing and Phylogenetic analysis” section. Although both in the section title as in the methodology a phylogenetic analysis from porB and tbpB data was included, results of the analysis have not been presented in the manuscript.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

Major comments:

1. WGS sampling: It is unclear how the isolates were chosen for WGS. The authors performed WGS to investigate additional genomic factors that might lead to increased risk for DGI. However, the sampling strategy does not allow these comparisons to be made. In Lines 102-103, the authors state that “two strains from each NG-MAST were selected, comprising both DGI and non-DGI strains”. To me, this implies that the WGS dataset should include 8 DGI isolates and 8 non-DGI isolates. However, in Figure 1, 12 isolates are annotated as DGI and 4 are annotated as non-DGI. Can the authors clarify how isolates were chosen for WGS?

Response: We were unfortunately unable to select more strains for WGS in this study due to budgetary constraints. We selected strains from NG-MAST that were prevalent in DGI, including types that were not associated with DGI, but the individual strain was one from a DGI patient, as well as strains from types associated with DGI but were from patients that did not have DGI. We acknowledge the limited strain selection is a limitation of this study (stated in text, line 231-232) and hope that future studies may be able to further address this.

2. Statistical analysis: In Line 117 of the methods, the authors state that $p<0.05$ was used to determine significance of association with DGI. However, the authors tested demographic factors (including several combinations of sex and age), porB types, and NG-MAST for association with DGI. Was multiple hypothesis test correction included?

Response: No mathematical correction was made for multiple comparisons because this study focussed on only a few comparisons based on scientific logic rather than every possible comparison (Article # 1390, GraphPad.com).

3. Core genome phylogeny: The methods for the core genome phylogeny are not included in the manuscript. How was the core genome determined? What phylogenetic software was used? Were recombinant regions masked? In Figure 1, it is not clear how the phylogeny is rooted and what the scale bar indicates.

Response: Methods are detailed in line 127-129. This is a core genome MLST scheme previously described in the reference given [25]. The scheme is set-up to screen for the core genes and it does compensate for the high recombination rates. In figure 1, the phylogeny is rooted at centre-point and the phylogenetic distance is indicated by the length of the horizontal lines.

4. Locus specific analyses: The authors focused on analyzing the presence of the GGI and phase variable repeats in pglA. However, there are additional loci that have been shown to be involved in N. gonorrhoeae serum resistance (e.g. the lgt operon). Also, these loci are not examined in isolates from NG-MAST that are not associated with DGI in this study, so I'm not sure what conclusions can be drawn from the results presented here. For example, the results here are not necessarily inconsistent with Power et al (Line 208) or provide evidence one way or the other about whether DGI is associated with GGI (Line 208-209).

Response: Power *et al* found pglA was not associated with DGI, whereas we suggest it could be associated with certain NG-MAST types that are associated with DGI. This could be inconsistent because they suggest the pglA is the associated factor, whereas we suggest pglA could be associated with certain Ng-MAST types and is therefore a correlation not a causation. We understand we did not WGS enough isolates to prove this, and we acknowledge in text that this is a limitation of this study. In addition, an attempt was made to examine the opacity factor genes (*opa*), but like Ogbebor et al., 2020 found it impossible with short reads. The current study did not analyse either *opa* or *lgt* but a long-read study in the future could look at these.

Minor comments:

5. Line 33: In Figure 1, 3/12 DGI isolates are shown to have pglA phase variable allele, so this allele was in the minority in DGI isolates sequenced, and the sampling of non-DGI isolates is unlikely to be representative of all uncomplicated infections.

Response: The authors acknowledge the reviewer's comment and the statement has been removed.

6. Lines 97-98: A phylogeny based on a concatenated alignment of porB and tbpB is unlikely to reflect genome-wide relationships between isolates due to the high recombination rate in these genes. This phylogeny is not actually included in the results, so I suggest that this be removed.

Response: This statement is remnant, hence removed.

7. Line 106: Version numbers and parameters should be included for assembly software.

Response: Version numbers have been included in line 126-128.

8. Line 109-110: WGS reads should be uploaded to NCBI or ENA, not just assemblies in PubMLST.
Response: In progress; ENA submission details: Project Accession- PRJEB52601; Submission Accession- ERA13386019. This information added to line 130-132.

9. Line 130-131/Table 1: The number of infections reported in males < 30 is not the same in the text and Table 1.
Response: This has been fixed in the Table 1.

10. Table 2: N. gonorrhoeae is misspelled in the second column
Response: Spelling corrected in Table 2

11. Lines 200, 205: These references appear to be incorrect.
Response: References corrected

12. Line 33,206: pglA instead of plgA
Response: abbreviation corrected

13. Supplementary Table 1 is great. Could a column be added to indicate accession numbers for the 16 isolates that were sequenced?
Response: Accession numbers for the 16 isolates sequenced added to the supplementary table.

Reviewer: 2
Dr. Raquel Abad, Instituto de Salud Carlos III Comments to the Author:
The study summarizes the molecular characterization by NG-MAST of 3953 Neisseria gonorrhoeae isolates from pathology laboratories servicing Queensland and surrounding areas from January 2010 to August 2015. Association of the NG-MAST data along basic demographic factors (age, sex, and geographical location) with disseminated gonococcal infection (DGI) were also analyzed. The manuscript addresses an important issue such as the monitoring and characterization of the circulating gonococcal strains, providing insight into the population structure of N. gonorrhoeae. Only several points would need to be revised/clarified:

-Abstract, page 2, lines 33-34. The statement “WGS demonstrated that NG-MAST types having a plgA phase variation were more commonly detected in DGI” would not fully warranted from this study: on the one hand only 8 NG-MAST types, all of them detected in DGI, were analyzed by WGS, and only 2 NG-MAST types showed plgA phase variation; on the other hand, only 2 strains of each NG-MAST type were analyzed.

Response: this statement is removed as per reviewer’s suggestion

- Methods, page 6, lines 96-97. “The porB sequence data from the NG-MAST was analysed to assign either PIA or PIB class”. Since the study associated strains harbouring PIA class of porB type with DGI, analysis to assign either PIA or PIB class from the porB sequence data should be more detailed.

Response: this information about PIA or PIB assignment has been included in line 116-117.

- Results, page 8, line 140: porB 4101 instead ST4101.

Response: this has been changed as per reviewer’s comment.

- Results, page 8, “NG-MAST typing and Phylogenetic analysis” section. Although both in the section title as in the methodology a phylogenetic analysis from porB and tbpB data was included, results of the analysis have not been presented in the manuscript.

Response: this statement is remnant, hence removed

VERSION 2 – REVIEW

REVIEWER	Mortimer, Tatum Harvard University T H Chan School of Public Health, Immunology and Infectious Diseases
REVIEW RETURNED	26-May-2022

GENERAL COMMENTS	Review #2 In their revision, the authors have addressed the majority of my previous comments. The statistical analysis has been clarified, the limitations of the sampling are more clearly stated, and whole genome sequencing reads have been uploaded to a repository. The methods used for the phylogeny in Figure 1 are still unclear. Which software was used for phylogenetic analysis? Also, the authors state in the figure legend that the phylogeny is rooted at the centre-point, but in the figure, it appears to be rooted on a branch leading to a group of ST-12040 DGI isolates. I have one additional minor comment, which is that line 82 (in the marked version) is the only use of pgtA in the manuscript now that line 87 has been edited. Would it make sense to somehow clarify that these are the same gene? Or perhaps just change line 82 to pglA as well?

REVIEWER	Abad, Raquel Instituto de Salud Carlos III, National Centre for Microbiology
REVIEW RETURNED	16-Jun-2022

GENERAL COMMENTS	All comments have been addressed.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 1

Comment 1: The methods used for the phylogeny in Figure 1 are still unclear. Which software was used for phylogenetic analysis? Also, the authors state in the figure legend that the phylogeny is rooted at the centre-point, but in the figure, it appears to be rooted on a branch leading to a group of ST-12040 DGI isolates.

Response: The software used for phylogenetic analysis was Ridom SeqSphere+ 4.1.0 (Ridom GmbH, Germany). This tool allows automatic processing and analysing of whole genome sequence data for microbial typing e.g., core genome MLST (cgMLST) or traditional MLST.

The Neisseria spp MLST scheme [26] was selected in Ridom SeqSphere+ to assign MLST to the isolate sequences included in this study.

In Figure 1, the phylogeny has been re-rooted to centre-point and the phylogenetic distance is indicated by the length of the horizontal lines. Ridom SeqSphere+ utilises Nei's DA distance

algorithm to calculate the genetic distance (Reference: Nei, M., F. Tajima, & Y. Tateno (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *J. Mol. Evol.* **19**:153-170).

Comment 2: I have one additional minor comment, which is that line 82 (in the marked version) is the only use of pgtA in the manuscript now that line 87 has been edited. Would it make sense to somehow clarify that these are the same gene? Or perhaps just change line 82 to pglA as well?

Response: Line 87 changed to *pgtA* again, explanation as follows and explained in manuscript text also:

pglA was the first gene identified to be involved in glycosylation of the pili in *Neisseria meningitidis* (Jennings et al., 1998). In 2002, Banerjee et al reported an equivalent gene in *Neisseria gonorrhoeae*, *pgtA* which was 96% identical to *pglA* at nucleotide level, due to this high level of similarity *pgtA* was later referred to as *pglA*. These references are listed below.

Jennings, M. P., M. Virji, D. Evans, V. Foster, Y. N. Srikhanta, L. Steeghs, P. van der Ley, and E. R. Moxon. 1998. Identification of a novel gene involved in pilin glycosylation in *Neisseria meningitidis*. Mol. Microbiol. 29**:**975-984.

Banerjee, A., R. Wang, S. L. Supernavage, S. K. Ghosh, J. Parker, N. F. Ganesh, P. G. Wang, S. Gulati, and P. A. Rice. 2002. Implications of phase variation of a gene (*pgtA*) encoding a pilin galactosyl transferase in gonococcal pathogenesis. J. Exp. Med. 196**:**147-162.

VERSION 3 – REVIEW

REVIEWER	Mortimer, Tatum Harvard University T H Chan School of Public Health, Immunology and Infectious Diseases
REVIEW RETURNED	30-Jun-2022

GENERAL COMMENTS	The authors have addressed my previous comments by re-rooting the phylogeny and clarifying that phylogenetic analysis was performed with the Ridom SeqSphere software.