



# PubMLST for Antigen Allele Mining to Inform Development of Gonorrhea Protein-Based Vaccines

Benjamin I. Baarda<sup>1†</sup>, Ryszard A. Zielke<sup>1†</sup>, Robert A. Nicholas<sup>2,3</sup> and Aleksandra E. Sikora<sup>1,4\*</sup>

<sup>1</sup> Department of Pharmaceutical Sciences, College of Pharmacy, Oregon State University, Corvallis, OR, United States,

<sup>2</sup> Department of Pharmacology, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, <sup>3</sup> Department of Microbiology and Immunology, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, <sup>4</sup> Vaccine and Gene Therapy Institute, Oregon Health and Science University, Beaverton, OR, United States

## OPEN ACCESS

### Edited by:

Kate Louise Seib,  
Griffith University, Australia

### Reviewed by:

Charlene Kahler,  
The University of Western Australia,  
Australia  
Keith Adrian Jolley,  
University of Oxford, United Kingdom

### \*Correspondence:

Aleksandra E. Sikora  
Aleksandra.Sikora@oregonstate.edu

<sup>†</sup> These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Infectious Diseases,  
a section of the journal  
Frontiers in Microbiology

**Received:** 25 September 2018

**Accepted:** 19 November 2018

**Published:** 07 December 2018

### Citation:

Baarda BI, Zielke RA, Nicholas RA  
and Sikora AE (2018) PubMLST  
for Antigen Allele Mining to Inform  
Development of Gonorrhea  
Protein-Based Vaccines.  
*Front. Microbiol.* 9:2971.  
doi: 10.3389/fmicb.2018.02971

*Neisseria gonorrhoeae* (Ng) is a human-specific pathogen and the etiological agent of gonorrhea, a sexually transmitted infection with a significant global health burden. While often asymptomatic, untreated gonorrhea can lead to pelvic inflammatory disease, ectopic pregnancy, infertility, and increased transmission/acquisition of HIV. A protective gonorrhea vaccine may be the only way to control disease transmission in the future due to the inexorable development of antibiotic resistance. Subunit antigens are proven candidates for vaccine development due to their safety, cost-effectiveness, and rapid preparation. To inform protein-based gonorrhea vaccine design by including different antigen variants, herein we present bioinformatics mining of alleles and single nucleotide/amino acid polymorphisms using DNA/protein sequences of all Ng isolates deposited into the PubMLST database and MtrE and BamA as model antigens. We also present phylogenetic analyses that can be performed using sequence data to gain insights into the evolutionary relationships between the polymorphisms found among the population of isolates using a convenient tool: Molecular Evolutionary Genetics Analysis (MEGA) software. Finally, we perform antigen polymorphism mapping onto the MtrE and BamA structures. This methodology can be applied for rational vaccine design to increase vaccine coverage and cross-protection by heteroligand presentation achieved via inclusion of diverse antigen variants and is relevant to over 100 different species and genera deposited into the PubMLST family of databases.

**Keywords:** *Neisseria gonorrhoeae*, vaccine, PubMLST, phylogenetic analysis, structural mapping, crystal structure, MtrE, BamA

## INTRODUCTION

Gonorrhea is a sexually transmitted infection caused by *Neisseria gonorrhoeae* (Ng), a human-specific pathogen that infects over 78 million people worldwide (Newman et al., 2015). Gonorrhea is a debilitating disease with clinical manifestations including cervicitis, urethritis, proctitis, conjunctivitis, or pharyngitis. Women are disproportionately affected by gonorrhea due to the high rate of asymptomatic infections and consequences of infection, which, if left untreated or if improperly treated, can lead to endometritis, pelvic inflammatory disease, pregnancy complications, and infertility. Antibiotics have been instrumental in treating infections in the past

eight decades, but *Ng* has gradually acquired resistance to all antimicrobials used in the clinic (Alirol et al., 2017). The currently recommended treatment is dual therapy with ceftriaxone and azithromycin, but *Ng* resistant to each of these antibiotics continue to arise worldwide (World Health Organization [WHO], 2011, 2012, 2015; Centers for Disease Control and Prevention [CDC], 2013, 2015; Newman et al., 2015; Alirol et al., 2017), including recently identified isolates that are resistant to both antibiotics (Fifer et al., 2016; Public Health England, 2018).

While options for developing effective antimicrobials are increasingly limited, an effective gonorrhea vaccine could significantly reduce infection burden and control gonorrhea globally, a goal that has been advocated by public health organizations and researchers (Zhu et al., 2011; ECDC, 2012; Jerse and Deal, 2013; Jerse et al., 2014; WHO, 2016; Zielke et al., 2016; CDC, 2017; Rice et al., 2017; Unemo and Sikora, 2017). Unsuccessful early attempts at using killed cells or purified pilin as vaccine candidates (Zhu et al., 2011), together with the realization that *Ng* readily undergoes phase and antigenic variation for many surface antigens and employs sophisticated immune evasion mechanisms, put vaccine research on hold for several decades. However, discovery of promising proteome-derived vaccine candidates (Zielke et al., 2014, 2016, 2018; Sikora et al., 2018), development of small animal models for systematic testing of vaccine formulations, and the recent success of the MenZB and 4CMenB vaccines for Group B meningitis, the major group of *N. meningitidis* (*Nm*) lacking a capsule-based vaccine (Petousis-Harris et al., 2017), indicates that developing an effective gonorrhea vaccine is within reach. Retrospective data suggested that MenZB had a 31% effectiveness against gonorrhea in the immunized cohort (Petousis-Harris et al., 2017). In support of this finding, immunization of mice with the similar 4CMenB vaccine significantly increased *Ng* clearance in the mouse model of gonorrhea ( $P < 0.0001$ ), and sera from 4CMenB-vaccinated mice cross-reacted with several proteins of the *Ng* outer membrane, including MtrE, BamA, Opa, and PilQ, suggesting that these are the cross-protective proteins (Connolly et al., 2018).

Subunit antigens are attractive candidates for vaccine development due to their safety, cost-effectiveness, and rapid preparation (Bae et al., 2009; Fischer et al., 2010). To inform protein-based gonorrhea vaccine design by including the most prevalent alleles of the target antigens to provide broad protection, herein we present bioinformatics mining of all *Ng* isolates deposited into the PubMLST database<sup>1</sup> to identify alleles and single nucleotide/amino acid polymorphisms of the model antigens MtrE and BamA. PubMLST is a curated database that hosts multi-locus sequence typing (MLST) data, isolate information, and progressively increasing numbers of whole genome sequences for many microbes including *Ng* and *Nm* (Maiden et al., 1998; Jolley et al., 2004; Jolley and Maiden, 2010). Additionally, allelic variation data are available for over 2500 loci, including the core *Neisseria* genome (Jolley et al., 2018). One limitation of the database is that gene and genome sequences are

deposited from specific studies that may have inherent sampling biases. As a result, the frequency of a particular variant in the database is not necessarily representative of the variant's global distribution. Nevertheless, the PubMLST database is the largest collection of *Neisseria* genetic and genomic information available and can therefore be instrumental in enabling rational vaccine design by identifying allelic variation of protein antigens in order to formulate vaccines with the broadest coverage.

The model vaccine antigens used in our study, MtrE and BamA, are integral  $\beta$ -barrel outer membrane proteins (OMPs) with excellent properties for subunit vaccine candidates for *Ng*: (i) they both have surface-exposed loops that are highly conserved, (ii) the proteins are not phase or sequence variable, (iii) the proteins are expressed in all *Ng* isolates examined to date, (iv) they elicit broad cross-reactive bactericidal antibodies, and (v) MtrE and BamA play pivotal roles in bacterial pathogenesis and physiology, respectively (Delahay et al., 1997; Lee and Shafer, 1999; Veal et al., 2002; Jerse et al., 2003; Rouquette-Loughlin et al., 2005; Zhu et al., 2011; Lei et al., 2014; Zielke et al., 2014, 2016; Bakelar et al., 2016; Rice et al., 2017). MtrE is the outer membrane channel for three *Ng* efflux pumps: MtrCDE, the major antibiotic efflux pump; FarAB, which exports fatty acids; and MacAB, which exports macrolides (Delahay et al., 1997; Lee and Shafer, 1999; Veal et al., 2002; Rouquette-Loughlin et al., 2005; Lei et al., 2014). MtrE is intimately involved in antibiotic resistance; e.g., knockout of MtrE in penicillin- or ceftriaxone-resistant strains markedly decreased the MIC of penicillin and ceftriaxone (Veal et al., 2002; Golparian et al., 2014), indicating that the pump is crucial in resisting the effects of antibiotics. Furthermore, MtrE is important for resistance to innate effectors, including antimicrobial peptides, fatty acids, bile salts, progesterone, and long-chain fecal lipids (Unemo and Shafer, 2014), and crucial for experimental infection in the mouse model of gonorrhea (Jerse et al., 2003). BamA is the central component of the  $\beta$ -barrel assembly machinery (BAM) responsible for the biogenesis of  $\beta$ -barrel OMPs, including BamA itself and MtrE (Knowles et al., 2009; O'Neil et al., 2015; Bakelar et al., 2016; Noinaj et al., 2017). In contrast to the *Escherichia coli* BAM protein complex, BAM in *Ng* contains the cell-surface displayed lipoprotein BamE, the accessory RmpM protein, and no BamB ortholog (Volokhina et al., 2009; Zielke et al., 2016; Sikora et al., 2018). BamA is essential for *Ng* cell viability (Zielke et al., 2016).

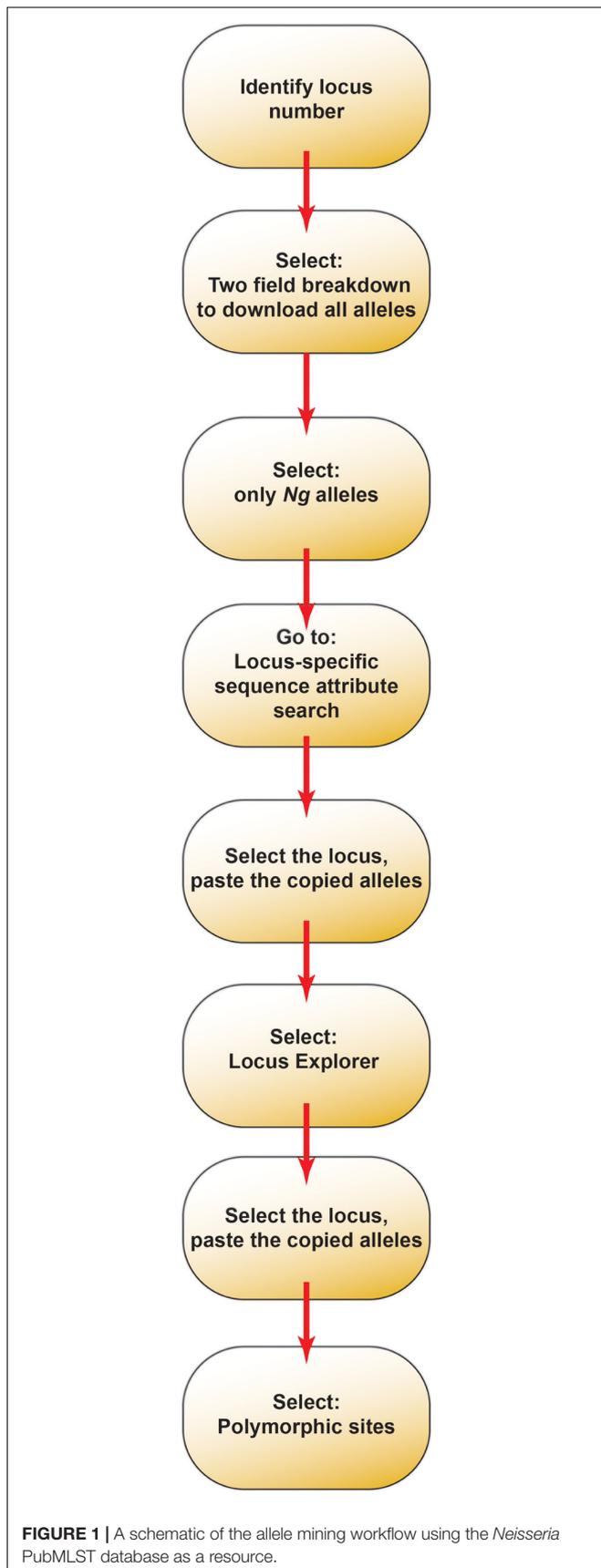
In this study, we present methodology for antigen allele mining that can be applied to rational vaccine design to increase vaccine coverage and cross-protection by inclusion of the most prevalent antigen alleles, particularly at extracellular regions exposed on the surface of the bacteria.

## METHODS

### Use of PubMLST *Neisseria* Database for Allele Mining Locus Identification

A schematic of the allele mining workflow can be found in **Figure 1**, with procedural details in the following sections.

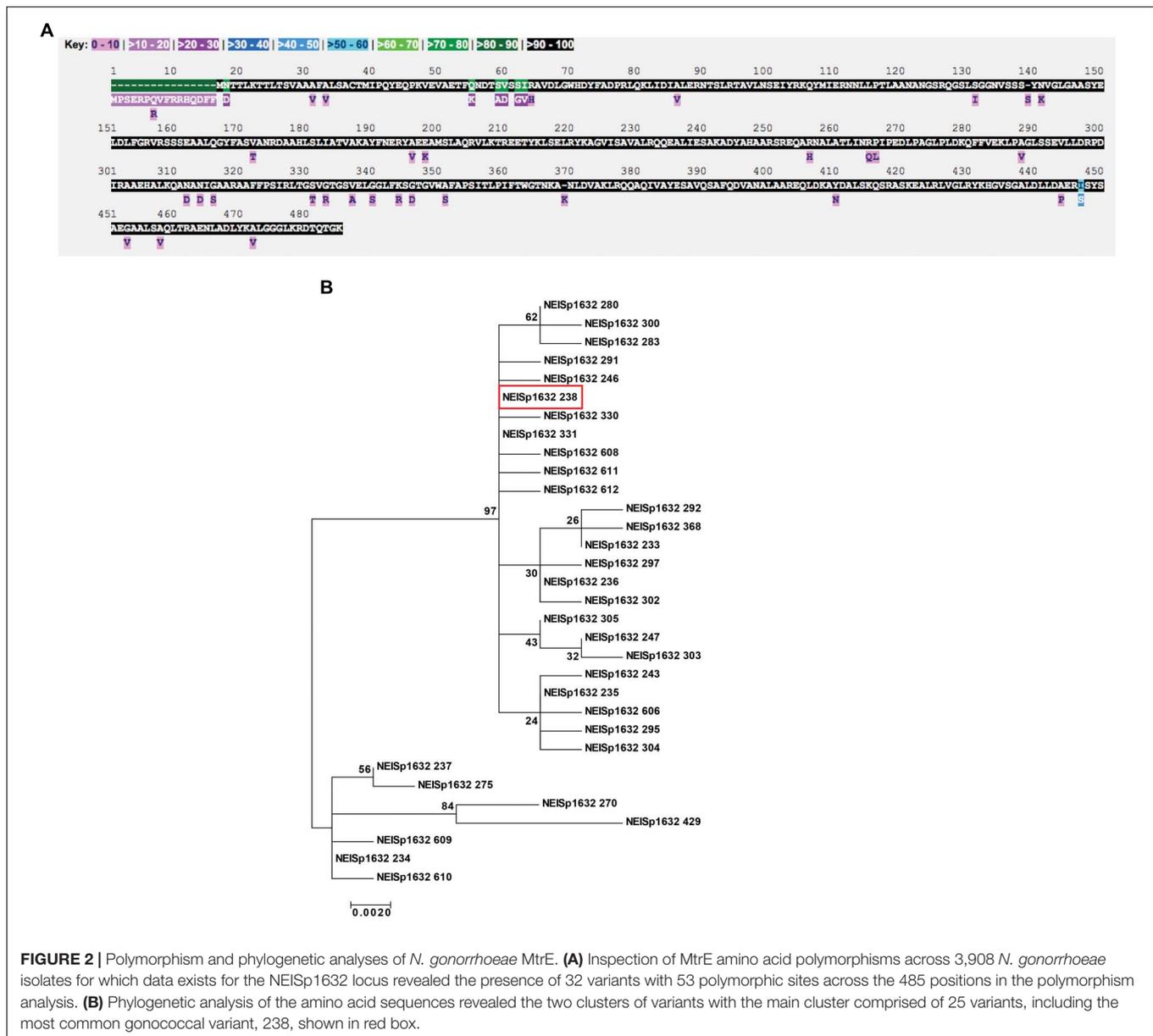
<sup>1</sup><https://pubmlst.org/neisseria/>



It should be noted that the website interface may change at some point after publication of this article. However, an up-to-date user's manual can be found at <http://bigsd.readthedocs.org/>, which should be referenced if difficulties arise. The steps provided here can serve as a guideline when using the manual and can identify the broader topics under which more specific information can be found. From the *Neisseria* PubMLST homepage<sup>1</sup> (**Supplementary Figure S1A**), determine the locus identifier for each antigen by navigating to the "Sequence and profile definitions" link, then clicking on the "Sequence query" link to search the database. Select "All loci" from the drop-down menu, and order results by locus (default parameters; **Supplementary Figure S1B**). Copy and paste the nucleotide sequence for each antigen into the query sequence box, and submit the query. If exact matches for the sequence are found, the results page will identify the locus and the allele that matches the query sequence. If exact matches are not present in the database, the closest alignments will be returned. In the example used in **Supplementary Figure S1** (*N. gonorrhoeae* MtrE using the FA1090 *mtrE* gene sequence to query the database), two exact matches were returned: "NEIS1632 (mtrE)," which represents the nucleotide sequence, and "NEISp1632," which represents the amino acid sequence of the translated protein (**Supplementary Figure S1C**). Translated protein loci are available for only a subset of loci in the database, but are available for both antigens investigated in the use case presented here. All following steps were performed with the translated protein locus, but the steps are identical for examining nucleic acid-only loci, with the exception that nucleic acid sequences should be translated prior to allele mapping and phylogenetic analyses.

### Selection of Species-Specific Alleles

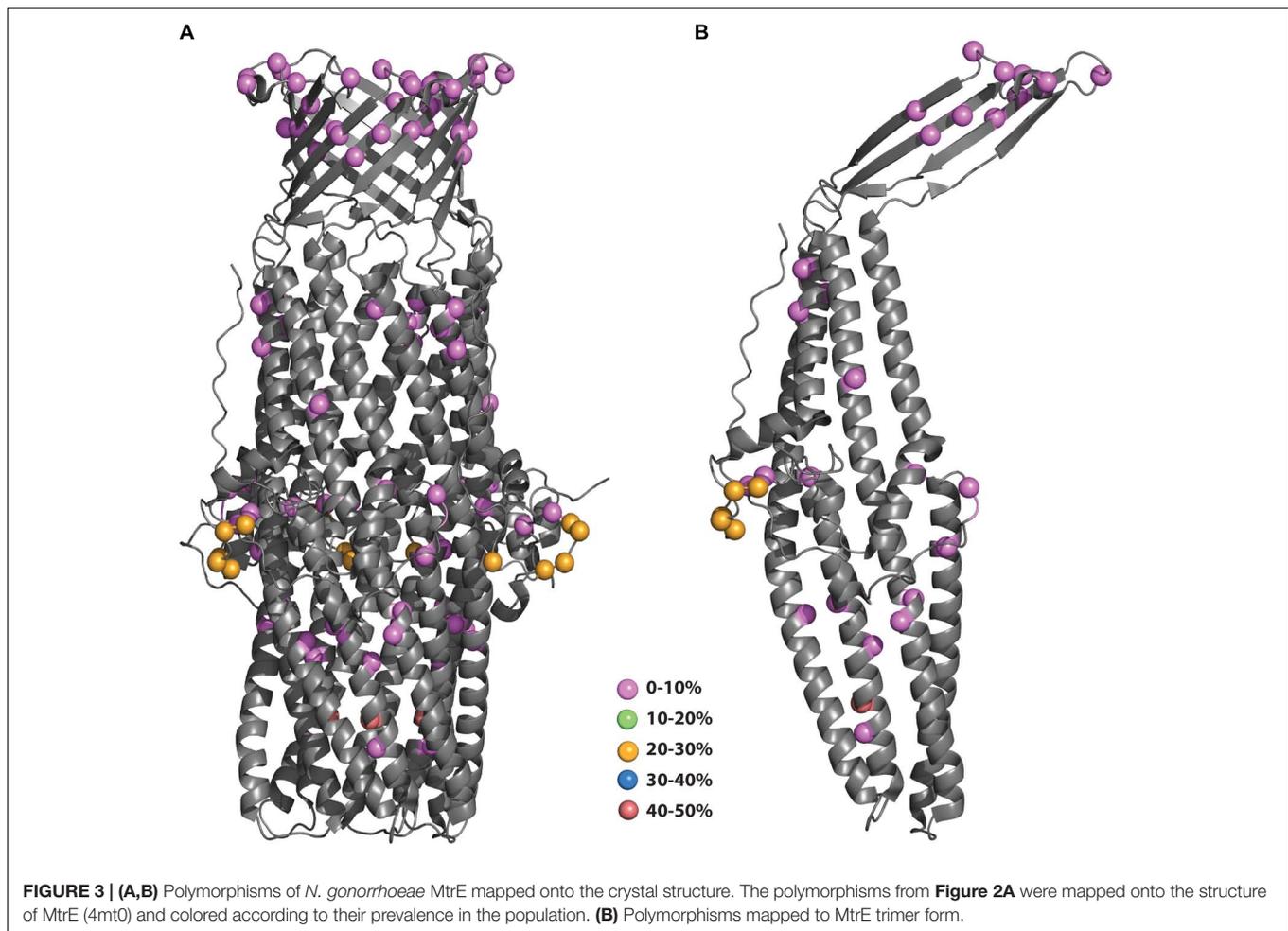
A two-field breakdown can be used to identify all alleles present in *Ng*, rather than in all *Neisseria* species existing in the database. From the homepage, select the "Isolates" page, and navigate to the "Two field" link under the Breakdown section of the Isolates page (**Supplementary Figure S2A**, red box). To perform the two-field breakdown of the dataset, list all isolates (**Supplementary Figure S2B**, red arrow), set Field 1 to the locus identifier (NEISp1632 in the use case), and set Field 2 to "species." All other options should be maintained at their default settings. After submission, a table listing each allele for each species in the database, as well as the number of isolates associated with the alleles, will be generated. At the bottom of the page, select "Excel format" to download the table as an Excel spreadsheet (**Supplementary Figure S2C**, orange box). In the spreadsheet, select the row with species names, click "Sort & Filter" in the Excel toolbar (**Supplementary Figure S2D**, blue arrow), and select "Filter" to enable sorting by species. Click on the filter arrow in the right corner of the filtered cell of the species of interest and select "Sort Descending" (**Supplementary Figure S2D**, orange arrow) to acquire a list of all alleles found for the selected species, sorted from highest prevalence to lowest. Note that in the results table for both MtrE and BamA, allele 0 is returned, which is associated with 23 and 5 isolates, respectively. Allele 0 represents a null value where the specific sequence was not discovered during automated scanning of the genome;



therefore, we have treated this allele as though it is included with the isolates with “No value” for the locus and removed the isolates associated with it from the prevalence calculations in the use case below. BamA is absolutely required for gonococcal viability (Zielke et al., 2016). While MtrE is not essential (Delahay et al., 1997), MtrE-deficient mutants are significantly attenuated *in vivo* (Jerse et al., 2003). If the 23 isolates with a null value for the MtrE locus are in fact lacking MtrE, they are unlikely to contribute substantially to a productive infection. In light of these lines of evidence, isolates with a null value for either protein are likely to be mis-annotated. The analysis described here provides information on the overall conservation of the antigen, based on the number of alleles, as well as each allele’s distribution within the population of isolates present in the database.

### Species-Specific Polymorphism Analysis

From the PubMLST *Neisseria* database homepage, navigate to the “Sequence and profile definitions” page and click the “Locus-specific sequence attribute search” link. Select the locus of interest from the drop-down menu (**Supplementary Figure S3A**). This action will reload the page; data should therefore not be entered prior to locus selection. To perform an analysis of the alleles of interest, click the “Modify form options” box in the top right corner of the form page (**Supplementary Figure S3A**, red box) and click the plus sign next to the “Allele id list box” option (**Supplementary Figure S3B**, red arrow). This preference can be saved by clicking the save icon in the bottom left corner of the option menu so that the option will not need to be selected every time the analysis is performed. From the Excel spreadsheet generated in the previous section, copy the list of alleles for which



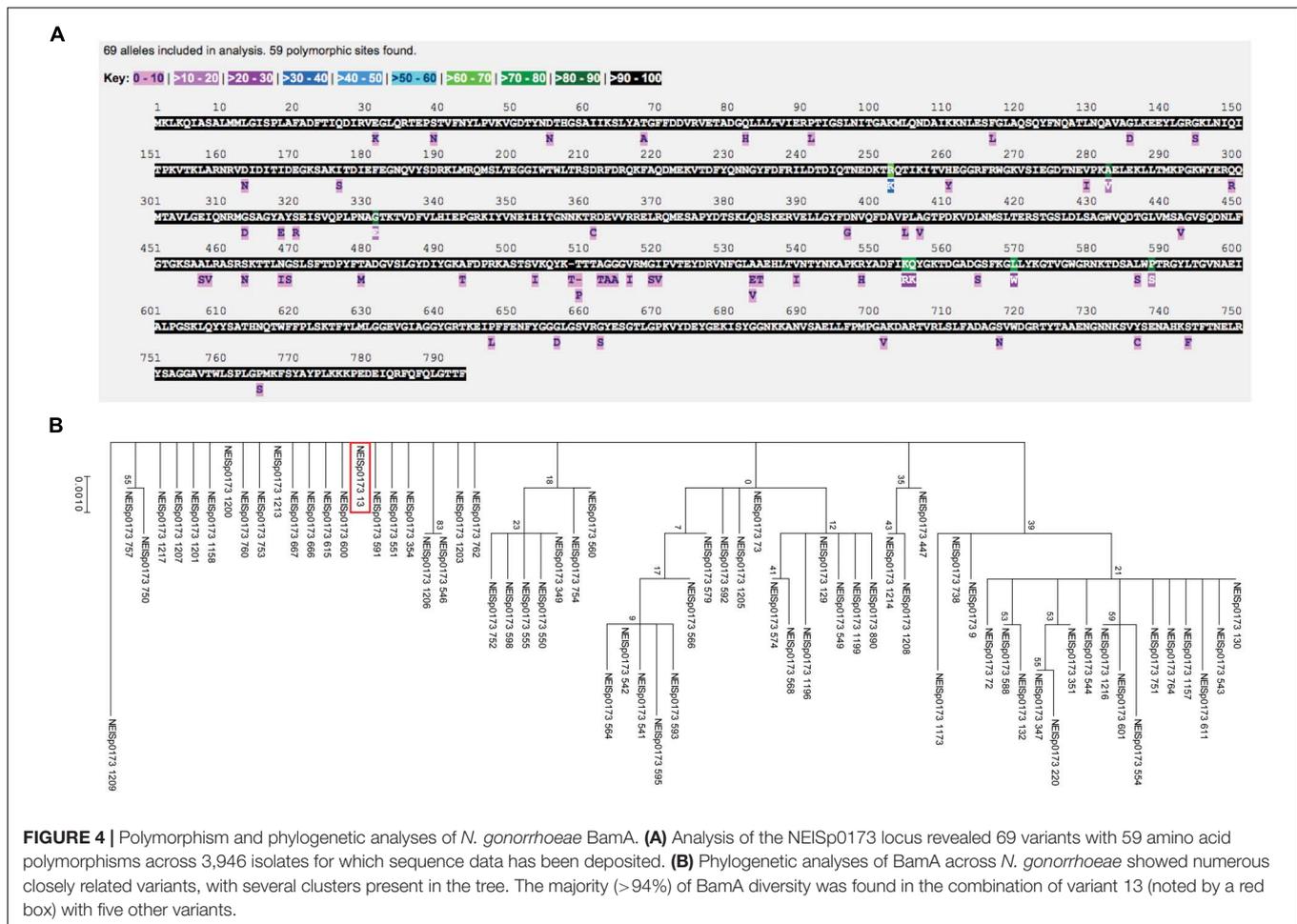
the analysis is desired, ignoring the “No value” entry. Paste the allele list into the “Allele id list” box on the locus analysis form (**Supplementary Figure S3C**) and click submit. A table of results will be generated that lists information on each allele, including any comments entered for a particular allele (**Supplementary Figure S3D**). From this point, the sequences (nucleotide or amino acid) can be exported in FASTA format for alignment and phylogenetic analysis, as described below, or the table can be exported to Excel or in a tab-delimited format.

To view polymorphism data in a graphical format, click the “Locus Explorer” button (**Supplementary Figure S3D, bottom**). On the subsequent page, the correct alleles will automatically be selected. **DO NOT** click the “All” button, as that will select all alleles for the locus, not just the alleles for the species of interest (**Supplementary Figure S4A**). After submission, a graphic of polymorphism data will be generated, as well as a table with amino acid or nucleotide frequencies at each position of the sequence. A portion of this table is shown in **Supplementary Figure S4B**. If the results are not automatically displayed, click the “Locus schematic (HTML format)” link to navigate to the results page. Not only does the polymorphism analysis provide data on the level of conservation of each antigen under consideration, but the information gathered from the table of

nucleotide/amino acid frequencies can also be employed, as in the use case below, to map polymorphism data to protein structure where available, thus facilitating structure-based vaccinology.

### Phylogenetic Analysis of Antigen Variants

Phylogenetic analyses can be performed using the sequence data to gain insights into the evolutionary relationships between the polymorphisms found among the population of isolates. A convenient tool for these analyses is the Molecular Evolutionary Genetics Analysis (MEGA) software version 7, available as a free download from <https://www.megasoftware.net> (Kumar et al., 2016). MEGA can be downloaded in either a command-line format or a graphical user interface, the latter of which will be used in this report. From the previous section, after submitting the list of alleles on the locus attribute form, click the “Export: FASTA” box (**Supplementary Figure S3D**, above “Locus Explorer” button), then click the “Download” link on the subsequent page. All allele sequences will be returned in FASTA format. In MEGA, click the “Align” button in the main application window, click Edit/Build Alignment (**Supplementary Figure S5A**), select the “Create a new alignment” option on the pop-up window, click “OK,” and select whether the alignment will use DNA or protein sequences.



To build the alignment, select and copy all of the allele sequences generated in FASTA format, and paste them into the MEGA Alignment Explorer window (if using keyboard shortcuts, Mac users must use ctrl+v to paste, rather than the typical command+v keystrokes). The window will be populated with the sequences of interest, including the locus identifier and the allele number in the “Species/Abbrv” field (**Supplementary Figure S5B**). If the sequences are in the nucleic acid format, a tab will be present above the sequence data labeled “Translated Protein Sequences.” Click this tab to generate protein sequence data for phylogenetic analyses, then select “Yes” to use the standard genetic code for translation. When translation is complete, the sequences can be aligned with either the ClustalW or the MUSCLE algorithms either by selecting the desired algorithm from the Alignment menu on the menu bar, or by clicking one of the icons that look like a “W” or an arm in the toolbar (**Supplementary Figure S5B**, Red box). In the use case presented in this report, we have aligned the sequences using ClustalW. When the desired algorithm has been chosen, use the subsequent dialog box to select all sequences (this dialog box will appear only if no sequences have been selected for alignment; all sequences are selected by default after pasting). MEGA will next provide the option of changing the alignment parameters (**Supplementary**

**Figure S5C**). We have maintained the default settings in the use case presented here. To align the sequences, select “OK.” The software will generate an initial pairwise alignment and will then perform a multiple alignment. After the process is complete, the software will use asterisks to designate which residues are conserved in all sequences (**Supplementary Figure S6A**, row of gray boxes immediately above alignment data), and the display preferences can be altered to highlight residues conserved in 50, 60, 70, 80, 90, or 100% of sequences (click “Display” on the menu bar, then “Toggle Conserved Sites” and select the desired level of conservation).

To perform phylogenetic analyses, click “Data” in the menu bar, and select the “Phylogenetic Analysis” option (**Supplementary Figure S6A**). The software will process the data to enable the analysis. In the main window, click “Phylogeny” in the toolbar, select “Construct/Test Maximum Likelihood Tree” (**Supplementary Figure S6B**), and click “Yes” in the subsequent dialog box to use the currently active data. MEGA will again provide the option to change parameters. We have kept the default settings for this analysis as well, with the exception that we selected to test the phylogenies with 500 Bootstrap replicates (**Supplementary Figure S6C**). After “Compute” is selected, the software will generate the maximum likelihood tree, which can

then be exported as an enhanced metafile image for use in Adobe Illustrator or in other formats for manipulation in tree-editing software, which is not discussed here.

### Antigen Polymorphism Structural Mapping

Using the polymorphism data generated after clicking on “Locus Explorer,” the polymorphisms (which are color-coded as shown in **Figure 2**) can be mapped to the crystal structure of the protein provided that the structure is known. The following description is for MtrE, and uses PyMol<sup>2</sup> to visualize the structure. The coordinates for MtrE (4mt0; Lei et al., 2014), which can be downloaded from PubMed, are opened in PyMol. Under “H” in the right panel, select “Hide Everything,” then under “S,” select “Show as cartoon.” The color of the entire structure can be chosen under “C” (in the use case below the structure was colored as “gray50”). To mark polymorphic residues, display the sequence (Display > Sequence On), select all of the amino acids from one group (e.g., 0–10%), rename selection (i.e., 0–10), and color these according to your color scheme (in the use case below the color violet was used for this group). Repeat for the other groupings. To display spheres, right-click on each of the colored residues in the structure and select atom > show > spheres (one can also do this in the command line with “show

spheres, /4mt0//A/res1+res2+. . ./CA,” where res is the number of the residue). Finally, to obtain the trimer, under “A,” select generate > show symmetry mates > within 4 A. Remove all of the subunits except for those that complete the trimer by selecting “delete object” under “A.” To generate a high resolution.png file, use the command “ray 2400, 2400;png filename.png” and a .png file is created on your home page. The background can be changed to white under Display > Background > White, and can be removed by toggling off “Opaque Background.” The .png file can be imported into Adobe Illustrator for creation of figures.

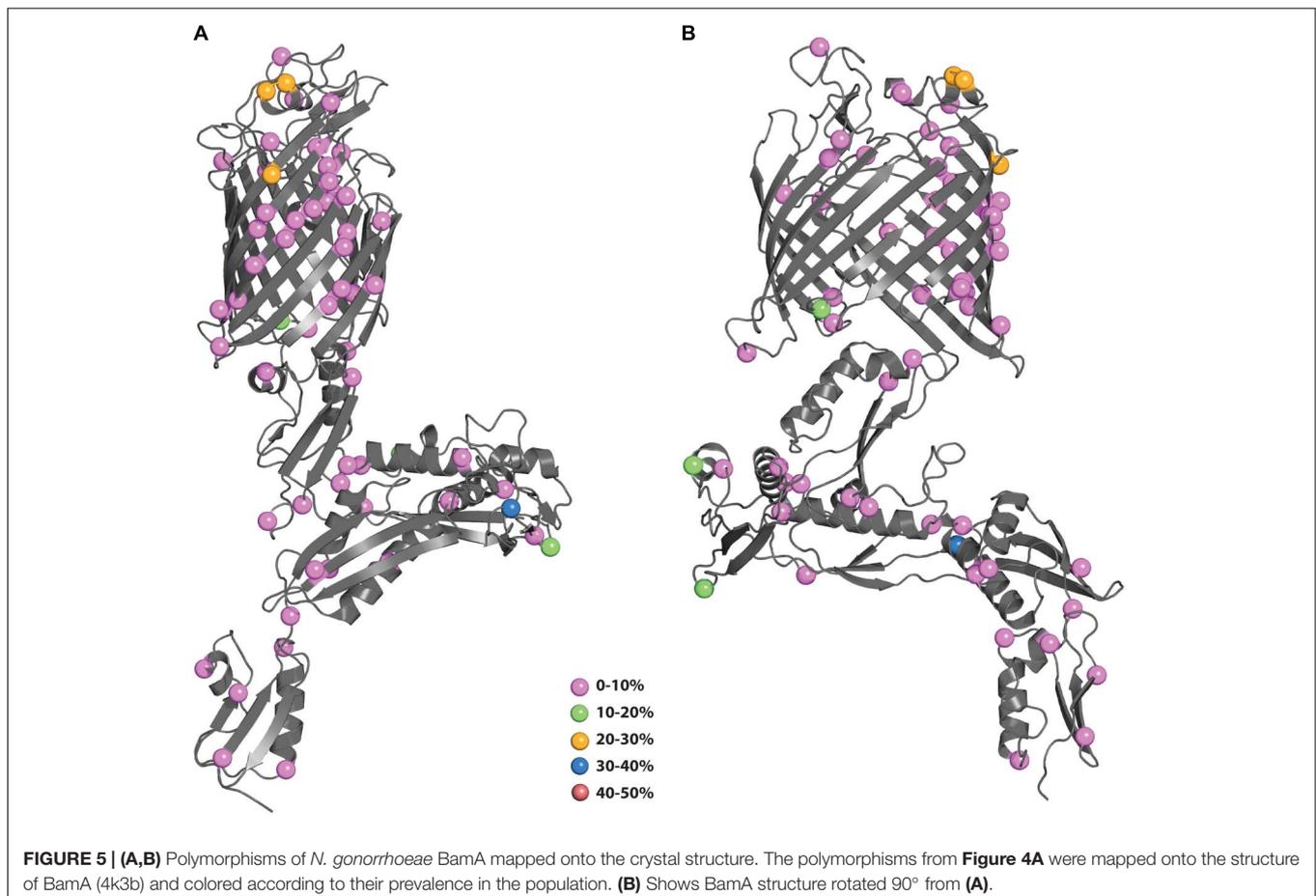
### USE CASE

In this section, we have used the methods described above to perform each of the analyses on two promising, broadly conserved vaccine candidates, MtrE and BamA.

### Phylogenetic Analysis and Structure Mapping of MtrE Polymorphisms

Examination of MtrE amino acid polymorphisms across 3,908 *N. gonorrhoeae* isolates for which data exists for the NEISp1632 locus revealed the presence of 32 variants with 53 polymorphic sites across the 485 positions in the polymorphism analysis.

<sup>2</sup>pymol.org



Large portions of the protein were conserved among all isolates in the database (**Figure 2A**). The most common variant was 238, which represents 30% of all *N. gonorrhoeae* MtrE sequences. Protein variant 238 and four others account for 98% of the diversity of gonococcal MtrE: variant 233, 21%; variant 235, 21%; variant 236, 13%; and variant 234, 13% (NEISp1632 two-field breakdown; **Supplementary Table S1**).

Phylogenetic analysis of the amino acid sequences revealed the existence of two clusters of antigen variants (**Figure 2B**). The main cluster was composed of 25 variants that appeared to be closely related, including the most common gonococcal variant, 238 (**Figure 2B**, red box). The secondary cluster was made up of six variants that contain a 16 amino acid insertion at the beginning of the protein sequence (variants 234, 237, 270, 429, 609, and 610) and variant 275, which does not possess the insertion but nonetheless contains portions of its sequence with similarities to the other six variants in the secondary cluster. It is possible that this “insertion” is actually a mis-annotation that includes sequence upstream of the translation start site, especially considering that all of the longer sequences also contain a methionine residue at the start site annotated for the other sequences. Automated annotation may have detected an alternate start site in the affected sequences that has not been manually curated. While it is possible that these variants represent true MtrE sequence diversity among the population of *N. gonorrhoeae* isolates present in the database, the result of this extension would be to increase the length of the signal sequence, which is removed during transport of the protein across the cytoplasmic membrane by the secretory machinery and thus not present in the mature protein.

Similar analyses performed on MtrE sequences from all *Neisseria* isolates in the database revealed the presence of 492 protein variants with 298 polymorphic sites. As observed for MtrE in *N. gonorrhoeae*, the majority of the sequence was conserved in 90–100% of variants, with numerous low-prevalence polymorphisms (**Supplementary Figure S7**). A phylogenetic tree constructed from all MtrE sequences again indicated that MtrE variants are closely related among *Neisseria* with the exception of variant 75, which originated from an isolate of *N. animalis* and formed an outgroup in the analysis (**Supplementary Figure S8**). Variant 2, exclusively associated with *N. meningitidis* isolates, was the most common *Neisseria* MtrE variant (2,467 of 16,363 isolates for which data exists for this locus; marked with a red box in **Supplementary Figure S8**).

Structure mapping is particularly important for rational vaccine design. The utility of subunit vaccines is that they generate antibodies that would recognize intact bacteria by binding to the surface-exposed regions of the protein. With the goal of increasing vaccine coverage and cross-protection, it is important to select polymorphisms that reside in surface-exposed regions. As shown in **Figure 3**, the low prevalence polymorphisms are scattered throughout the monomeric MtrE subunit, including 6 residues apparently surface exposed (residues 115, 122, 320, 323, 327, and 329). In contrast,

the more prevalent polymorphisms are all located in the periplasm, with all but one clustered on a loop about halfway down the long alpha helical bundles. The most common polymorphism is located on one of the helices near the bottom of the protein (PyMol session files showing the mapping of MtrE polymorphisms to the monomer and trimer forms of MtrE can be found in **Supplementary Files S1, S2**, respectively).

## Phylogenetic Analysis and Structure Mapping of BamA Polymorphisms

Assessment of the NEISp0173 locus revealed 69 variants with 59 amino acid polymorphisms across 3,946 isolates for which sequence data has been deposited into the database (**Figure 4A**). Large portions of the BamA amino acid sequence were conserved among 90–100% of protein variants, with low-frequency polymorphisms distributed throughout the sequence. Seven positions were conserved in only 60–80% of variants (green residues in **Figure 4A**). The most common BamA variant was number 13, which accounted for 67.6% of isolates. Over 94% of BamA diversity was found in the combination of variant 13 with five others: variant 349, 8.0%; variant 73, 7.4%; variant 130, 4.2%; variant 129, 3.7%; and variant 220, 3.3%. Additionally, 55 of the 69 *N. gonorrhoeae* BamA variants were associated with five or fewer isolates (NEISp0173 two-field breakdown; **Supplementary Table S2**).

An examination of BamA phylogeny across *N. gonorrhoeae* revealed that numerous variants were closely related, with several clusters present in the tree (**Figure 4B**). The six most common variants (variant 13 is noted by a red box in **Figure 4B**) were distributed throughout the tree, which could indicate they evolved independently. The most distantly related variant was 1209, which represents two isolates in the database.

Polymorphism analyses of BamA across all *Neisseria* revealed the presence of 744 variants with 622 polymorphic sites across the protein. The amino acid sequence was largely conserved, although numerous low-abundance polymorphisms were present at nearly every position (**Supplementary Figure S9**). Finally, phylogenetic analysis of all *Neisseria* BamA variants indicated that the majority of BamA variants were closely related. Variant 179, corresponding to an isolate of *N. shayegani*, formed an outgroup in this analysis. The most common *Neisseria* BamA variant was allele 13, found in 2,666 *N. gonorrhoeae* isolates out of 16,791 *Neisseria* isolates for which BamA sequence data has been deposited (**Supplementary Figure S10**, identified by a red box). This variant was not present in any other species in the database.

Mapping of the polymorphisms onto the *N. gonorrhoeae* BamA structure (4k3b) is shown in **Figure 5**. The structure of BamA consists of 5 POTRA domains at the *N*-terminus followed by a  $\beta$ -barrel domain at the *C*-terminus (Noinaj et al., 2013). As with MtrE, the polymorphisms are distributed throughout the different domains. There are 6 polymorphisms that map to surface-exposed regions: four of these (residues 547, 563, 735, and 742) are of low prevalence (0–10%), while the other two

(residues 553 and 554) are medium prevalence (20–30%). These latter two polymorphisms would be likely candidates to include in a subunit vaccine to increase coverage of strains harboring these existing polymorphisms (PyMol session file showing the mapping of BamA polymorphisms to the structure of BamA can be found in **Supplementary File S3**).

## AUTHOR CONTRIBUTIONS

AS, RN, BB, and RZ developed and designed the study, analyzed data, and wrote the manuscript. BB, RZ, and RN performed the analyses. AS and RN designed and executed the figures. BB prepared the supplementary figures.

## REFERENCES

- Alirol, E., Wi, T. E., Bala, M., Bazzo, M. L., Chen, X. S., Deal, C., et al. (2017). Multidrug-resistant gonorrhea: a research and development roadmap to discover new medicines. *PLoS Med.* 14:e1002366. doi: 10.1371/journal.pmed.1002366
- Bae, K., Choi, J., Jang, Y., Ahn, S., and Hur, B. (2009). Innovative vaccine production technologies: the evolution and value of vaccine production technologies. *Arch. Pharm. Res.* 32, 465–480. doi: 10.1007/s12272-009-1400-1
- Bakelar, J., Buchanan, S. K., and Noinaj, N. (2016). The structure of the beta-barrel assembly machinery complex. *Science* 351, 180–186. doi: 10.1126/science.aad3460
- Centers for Disease Control and Prevention [CDC] (2015). *Sexually Transmitted Disease Surveillance 2014*. Atlanta: CDC.
- Centers for Disease Control and Prevention [CDC] (2017). *Gonorrhea Statistics*. Available at: <https://www.cdc.gov/std/stats16/gonorrhea.htm>
- Centers for Disease Control and Prevention [CDC] (2013). *Gonorrhea Treatment Guidelines*. Available at: <https://www.cdc.gov/nchstp/newsroom/docs/factsheets/archive/gonorrhea-treatment-guidelines-factsheet.pdf>
- Connolly, K., Leduc, I., Rahman, N., Sempowski, G., and Jerse, A. E. (2018). “The group B meningococcal vaccine Bexsero induces antibodies that recognize several candidate gonorrhea vaccine targets and shows protective efficacy against experimental *Neisseria gonorrhoeae* genital tract infection in mice,” in *Proceedings of the International Pathogenic Neisseria Conference*, Pacific Grove, CA.
- Delahay, R. M., Robertson, B. D., Balthazar, J. T., Shafer, W. M., and Ison, C. A. (1997). Involvement of the gonococcal MtrE protein in the resistance of *Neisseria gonorrhoeae* to toxic hydrophobic agents. *Microbiology* 143(Pt 7), 2127–2133. doi: 10.1099/00221287-143-7-2127
- ECDC (2012). *Response Plan to Control and Manage the Threat of Multidrug-Resistant Gonorrhoea in Europe*. Solna Municipality: ECDC.
- Fifer, H., Natarajan, U., Jones, L., Alexander, S., Hughes, G., Golparian, D., et al. (2016). Failure of dual antimicrobial therapy in treatment of gonorrhea. *N. Engl. J. Med.* 374, 2504–2506. doi: 10.1056/NEJMc1512757
- Fischer, N. O., Infante, E., Ishikawa, T., Blanchette, C. D., Bourne, N., Hoeprich, P. D., et al. (2010). Conjugation to nickel-chelating nanolipoprotein particles increases the potency and efficacy of subunit vaccines to prevent West Nile encephalitis. *Bioconjug. Chem.* 21, 1018–1022. doi: 10.1021/bc100083d
- Golparian, D., Shafer, W. M., Ohnishi, M., and Unemo, M. (2014). Importance of multidrug efflux pumps in the antimicrobial resistance property of clinical multidrug-resistant isolates of *Neisseria gonorrhoeae*. *Antimicrob. Agents Chemother.* 58, 3556–3559. doi: 10.1128/AAC.00038-14
- Jerse, A. E., Bash, M. C., and Russell, M. W. (2014). Vaccines against gonorrhea: current status and future challenges. *Vaccine* 32, 1579–1587. doi: 10.1016/j.vaccine.2013.08.067
- Jerse, A. E., and Deal, C. D. (2013). Vaccine research for gonococcal infections: where are we? *Sex. Transm. Infect.* 89(Suppl. 4), iv63–iv68. doi: 10.1136/sextrans-2013-051225
- Jerse, A. E., Sharma, N. D., Simms, A. N., Crow, E. T., Snyder, L. A., and Shafer, W. M. (2003). A gonococcal efflux pump system enhances bacterial survival in a female mouse model of genital tract infection. *Infect. Immun.* 71, 5576–5582. doi: 10.1128/IAI.71.10.5576-5582.2003
- Jolley, K. A., Bray, J. E., and Maiden, M. C. J. (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* 3:124. doi: 10.12688/wellcomeopenres.14826.1
- Jolley, K. A., Chan, M. S., and Maiden, M. C. (2004). mlstdbNet - distributed multi-locus sequence typing (MLST) databases. *BMC Bioinformatics* 5:86. doi: 10.1186/1471-2105-5-86
- Jolley, K. A., and Maiden, M. C. (2010). BIGSdb: scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* 11:595. doi: 10.1186/1471-2105-11-595
- Knowles, T. J., Scott-Tucker, A., Overduin, M., and Henderson, I. R. (2009). Membrane protein architects: the role of the BAM complex in outer membrane protein assembly. *Nat. Rev. Microbiol.* 7, 206–214. doi: 10.1038/nrmicro2069
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Lee, E. H., and Shafer, W. M. (1999). The farAB-encoded efflux pump mediates resistance of gonococci to long-chained antibacterial fatty acids. *Mol. Microbiol.* 33, 839–845. doi: 10.1046/j.1365-2958.1999.01530.x
- Lei, H. T., Chou, T. H., Su, C. C., Bolla, J. R., Kumar, N., Radhakrishnan, A., et al. (2014). Crystal structure of the open state of the *Neisseria gonorrhoeae* MtrE outer membrane channel. *PLoS One* 9:e97475. doi: 10.1371/journal.pone.0097475
- Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., et al. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U.S.A.* 95, 3140–3145. doi: 10.1073/pnas.95.6.3140
- Newman, L., Rowley, J., Vander Hoorn, S., Wijesooriya, N. S., Unemo, M., Low, N., et al. (2015). Global estimates of the prevalence and incidence of four curable sexually transmitted infections in 2012 based on systematic review and global reporting. *PLoS One* 10:e0143304. doi: 10.1371/journal.pone.0143304
- Noinaj, N., Gumbart, J. C., and Buchanan, S. K. (2017). The beta-barrel assembly machinery in motion. *Nat. Rev. Microbiol.* 15, 197–204. doi: 10.1038/nrmicro.2016.191
- Noinaj, N., Kuzsak, A. J., Gumbart, J. C., Lukacik, P., Chang, H., Easley, N. C., et al. (2013). Structural insight into the biogenesis of beta-barrel membrane proteins. *Nature* 501, 385–390. doi: 10.1038/nature12521
- O’Neil, P. K., Rollauer, S. E., Noinaj, N., and Buchanan, S. K. (2015). Fitting the pieces of the beta-barrel assembly machinery complex. *Biochemistry* 54, 6303–6311. doi: 10.1021/acs.biochem.5b00852
- Petousis-Harris, H., Paynter, J., Morgan, J., Saxton, P., McArdle, B., Goodyear-Smith, F., et al. (2017). Effectiveness of a group B outer membrane vesicle meningococcal vaccine against gonorrhoea in New Zealand: a retrospective

## FUNDING

Funding was provided to AS by grant R01-AI117235 from the National Institute of Allergy and Infectious Diseases, National Institutes of Health. The funding agency had no role in study design; collection, analysis or interpretation of data; writing the report; or the decision to submit the article for publication.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.02971/full#supplementary-material>

- case-control study. *Lancet* 390, 1603–1610. doi: 10.1016/S0140-6736(17)31449-6
- Public Health England (2018). UK case of *Neisseria gonorrhoeae* with high-level resistance to azithromycin and resistance to ceftriaxone acquired abroad. *Health Protect. Rep.* 12, 1–4.
- Rice, P. A., Shafer, W. M., Ram, S., and Jerse, A. E. (2017). *Neisseria gonorrhoeae*: drug resistance, mouse models, and vaccine development. *Annu. Rev. Microbiol.* 71, 665–686. doi: 10.1146/annurev-micro-090816-093530
- Rouquette-Loughlin, C. E., Balthazar, J. T., and Shafer, W. M. (2005). Characterization of the MacA-MacB efflux system in *Neisseria gonorrhoeae*. *J. Antimicrob. Chemother.* 56, 856–860. doi: 10.1093/jac/dki333
- Sikora, A. E., Wierzbicki, I. H., Zielke, R. A., Ryner, R. F., Korotkov, K. V., Buchanan, S. K., et al. (2018). Structural and functional insights into the role of BamD and BamE within the beta-barrel assembly machinery in *Neisseria gonorrhoeae*. *J. Biol. Chem.* 293, 1106–1119. doi: 10.1074/jbc.RA117.000437
- Unemo, M., and Shafer, W. M. (2014). Antimicrobial resistance in *Neisseria gonorrhoeae* in the 21st century: past, evolution, and future. *Clin. Microbiol. Rev.* 27, 587–613. doi: 10.1128/CMR.00010-14
- Unemo, M., and Sikora, A. E. (2017). Infection: proof of principle for effectiveness of a gonorrhoea vaccine. *Nat. Rev. Urol.* 14, 643–644. doi: 10.1038/nrurol.2017.139
- Veal, W. L., Nicholas, R. A., and Shafer, W. M. (2002). Overexpression of the MtrC-MtrD-MtrE efflux pump due to an mtrR mutation is required for chromosomally mediated penicillin resistance in *Neisseria gonorrhoeae*. *J. Bacteriol.* 184, 5619–5624. doi: 10.1128/JB.184.20.5619-5624.2002
- Volokhina, E. B., Beckers, F., Tommassen, J., and Bos, M. P. (2009). The beta-barrel outer membrane protein assembly complex of *Neisseria meningitidis*. *J. Bacteriol.* 191, 7074–7085. doi: 10.1128/JB.00737-09
- World Health Organization [WHO] (2011). “Prevalence and incidence of selected sexually transmitted infections. *Chlamydia trachomatis*, *Neisseria gonorrhoeae*, syphilis and *Trichomonas vaginalis*,” in *Proceedings of the Methods and Results used by WHO to Generate 2005 Estimates*, Geneva.
- World Health Organization [WHO] (2012). *Global Action Plan to Control the Spread and Impact of Antimicrobial Resistance in Neisseria gonorrhoeae*. Available at: [http://whqlibdoc.who.int/publications/2012/9789241503501\\_eng.pdf](http://whqlibdoc.who.int/publications/2012/9789241503501_eng.pdf)
- World Health Organization [WHO] (2015). *Antimicrobial Resistance Draft Global Action Plan on Antimicrobial Resistance*. Geneva: WHO.
- World Health Organization [WHO] (2016). *WHO | Global Health Sector Strategy on Sexually Transmitted Infections, 2016–2021*. Available at: <http://www.who.int/reproductivehealth/publications/rtis/ghss-stis/en/>
- Zhu, W., Chen, C. J., Thomas, C. E., Anderson, J. E., Jerse, A. E., and Sparling, P. F. (2011). Vaccines for gonorrhea: can we rise to the challenge? *Front. Microbiol.* 2:124. doi: 10.3389/fmicb.2011.00124
- Zielke, R. A., Le Van, A., Baarda, B. I., Herrera, M. F., Acosta, C. J., Jerse, A. E., et al. (2018). SliC is a surface-displayed lipoprotein that is required for the anti-lysozyme strategy during *Neisseria gonorrhoeae* infection. *PLoS Pathog.* 14:e1007081. doi: 10.1371/journal.ppat.1007081
- Zielke, R. A., Wierzbicki, I. H., Baarda, B. I., Gafken, P. R., Soge, O. O., Holmes, K. K., et al. (2016). Proteomics-driven antigen discovery for development of vaccines against gonorrhea. *Mol. Cell. Proteomics* 15, 2338–2355. doi: 10.1074/mcp.M116.058800
- Zielke, R. A., Wierzbicki, I. H., Weber, J. V., Gafken, P. R., and Sikora, A. E. (2014). Quantitative proteomics of the *Neisseria gonorrhoeae* cell envelope and membrane vesicles for the discovery of potential therapeutic targets. *Mol. Cell. Proteomics* 13, 1299–1317. doi: 10.1074/mcp.M113.029538

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Baarda, Zielke, Nicholas and Sikora. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.