Glycobiology, 2020, vol. 00, no. 00, 1–15
doi: 10.1093/glycob/cwaa073
Advance Access Publication Date: 20 January 2001
Regular Manuscripts

Microbial Biology

Allelic polymorphisms in a glycosyltransferase gene shape glycan repertoire in the *O*-linked protein glycosylation system of *Neisseria*

Nelson Wang[2,3], Jan Haug Anonsen[2,4], Chris Hadjineophytou[2,3], William Brynildsen Reinar[5], Bente Børud[6], Åshild Vik[2,7] and Michael Koomey[1,2,3,5]

[2]Department of Biosciences, Section for Genetics and Evolutionary Biology, University of Oslo, 0371 Oslo, Norway , [3]Department of Biosciences, Centre for Integrative Microbial Evolution, University of Oslo, 0371 Oslo, Norway, [4]Norwegian Research Centre AS, 4072 Randaberg, Norway, [5]Department of Biosciences, Centre for Ecological and Evolutionary Synthesis, University of Oslo, 1066 Oslo, Norway, [6]Division for Infection Control and Environmental Health, Norwegian Institute of Public Health, 0403 Oslo, Norway, and [7]Research Council of Norway, 0283 Oslo, Norway

[1]To whom correspondence should be addressed: Tel: +47 22854091; e-mail: johnk@ibv.uio.no

Received 22 June 2020; Revised 24 July 2020; Accepted 27 July 2020

Abstract

Glycosylation of multiple proteins via *O*-linkage is well documented in bacterial species of *Neisseria* of import to human disease. Recent studies of protein glycosylation (*pgl*) gene distribution established that related protein glycosylation systems occur throughout the genus including nonpathogenic species. However, there are inconsistencies between *pgl* gene status and observed glycan structures. One of these relates to the widespread distribution of *pglG*, encoding a glycosyltransferase that in *Neisseria elongata* subsp. *glycolytica* is responsible for the addition of di-*N*-acetyl glucuronic acid at the third position of a tetrasaccharide. Despite *pglG* residing in strains of *N. gonorrhoeae*, *N. meningitidis* and *N. lactamica*, no glycan structures have been correlated with its presence in these backgrounds. Moreover, PglG function in *N. elongata* subsp. *glycolytica* minimally requires UDP-glucuronic acid (GlcNAcA), and yet *N. gonorrhoeae*, *N. meningitidis* and *N. lactamica* lack *pglJ*, the gene whose product is essential for UDP-GlcNAcA synthesis. We examined the functionality of *pglG* alleles from species spanning the *Neisseria* genus by genetic complementation in *N. elongata* subsp. *glycolytica*. The results indicate that select *pglG* alleles from *N. meningitidis* and *N. lactamica* are associated with incorporation of an *N*-acetyl-hexosamine at the third position and reveal the potential for an expanded glycan repertoire in those species. Similar experiments using *pglG* from *N. gonorrhoeae* failed to find any evidence of function suggesting that those alleles are missense pseudogenes. Taken together, the results are emblematic of how allelic polymorphisms can shape bacterial glycosyltransferase function and demonstrate that such alterations may be constrained to distinct phylogenetic lineages.

Key words: bacterial glycosylation, evolution, epistasis, *pglG*

Introduction

Glycoform diversity and phase variable expression are cardinal features of the broad—spectrum *O*-linked protein glycosylation (*pgl*) systems in the pathogenic *Neisseria* species *N. gonorrhoeae* and *N. meningitidis* (Vik et al. 2009; Børud et al. 2010; Børud et al. 2011). As defined there, intrastrain glycan variation results from phase-variable expression of extending glycosyltransferases, while inter-strain variation reflects differences in glycosyltransferase and ancillary gene content. The genus *Neisseria* also encompasses a large number of other human-associated species that are members of the commensal microbiota (Liu et al. 2015). Past and ongoing studies have begun to examine *O*-linked *pgl* systems in commensal species in order to better understand what roles protein glycosylation might play and if *pgl* gene content and function might be phylogenetically informative. A major finding of these studies was the discovery that *O*-linked protein glycosylation is conserved across the genus and that a subset of *pgl* genes are members of the genus-wide core genome (Hadjineophytou et al. 2019). Studies of protein glycosylation in the deeply branching species *N. elongata* subsp. *glycolytica* have been particularly insightful by identifying a unique tetrasaccharide and five previously unrecognized *pgl* genes (Anonsen et al. 2016; Wang et al. 2019). Among these are four genes required for the synthesis of UDP-di-*N*-acetyl-glucuronic acid used in glycan biosynthesis (*pglJ*, *pglK*, *pglM* and *pglN*) and one encoding a glycosyltransferase (*pglP*) responsible for addition of the terminal sugar to complete the mature tetrasaccharide structure (Hadjineophytou et al. 2019; Wang et al. 2019) (Figure 1). The former genes are limited in their distribution across the genus, while *pglP* is found in all species except *N. lactamica*. Of particular note, *pglP* has undergone parallel but independent pseudogenization in *N. gonorrhoeae* and *N. meningitidis* (Hadjineophytou et al. 2019). These results have led to a model in which the current *pgl* glycan repertoires in *N. gonorrhoeae, N. meningitidis* and *N. lactamica* reflect the ongoing replacement of a primordial glycan biosynthetic pathway with one encoding galactose-containing oligosaccharides.

Another unique finding from the *N. elongata* subsp. *glycolytica* studies involves identifying a glycosyltransferase encoded by *pglG* as being responsible for the incorporation of a di-*N*-acetyl-glucuronic acid glycan (GlcNAc(3NAc)A) as the third sugar residue of its tetrasaccharide (Anonsen et al. 2016). Intact alleles of *pglG* are found in all human neisserial species save for several isolates of *N. gonorrhoeae* and *N. meningitidis* that carry a partially deleted form of the gene (Kahler et al. 2001; Børud et al. 2011). In addition, *pglG* shows remarkably conserved synteny across the genus in that it invariably maps 5′ of *pglH/H2* (which encodes the glycosyltransferase generating the Und-PP-disaccharide oligosaccharide that serves as a target substrate for PglG) (Hadjineophytou et al. 2019). Mutant analyses have shown that PglG in *N. elongata* subsp. *glycolytica* shows relaxed specificity for its UDP-sugar donor and minimally requires UDP-glucuronic acid (GlcNAcA) (Wang et al. 2019). Thus, in *N. elongata* subsp. *glycolytica*, *pglJ* is epistatic to *pglG* while *pglK*, *pglM* or *pglN* mutants expressed mixtures of di-, tri- and tetrasaccharide glycoforms bearing glucuronic acid-related intermediates predicted from a disrupted di-*N*-acetyl-glucuronic acid biosynthetic pathway at the third position (see Figure 1).

The findings from *N. elongata* subsp. *glycolytica* were intriguing since despite its ubiquitous nature, *pglG* has not been implicated in glycoform-related phenotypes in other species examined to date. In particular, it is noteworthy that *pglJ* is not found within the genomes of many of those species (Hadjineophytou et al. 2019). Thus, a facile explanation for the lack of PglG-related phenotypes would be that PglG function is merely precluded by the absence of UDP-glucuronic acid. Alternatively, PglG in such backgrounds could have novel enzymatic activities whose effects have simply escaped detection.

To better understand the potential functionality of PglG encoded by alleles from across the genus, we sought to introduce such genes into an isolate of *N. elongata* subsp. *glycolytica* in which glycan phenotypes could be readily assessed. Here, we establish a modified method for allelic replacement and use this to uncover a unique PglG activity differing in UDP-sugar donor specificity as well as what appears to be a *pglG* pseudogene in *N. gonorrhoeae*.

Results

Examination of *pglG* phylogeny in *Neisseria*

In prior studies, *pglG* glycosyltransferase genes within the genus *Neisseria* were collectively defined by their levels of shared amino acid sequence identities and conserved synteny within what is termed the *pgl* core gene locus (Hadjineophytou et al. 2019). To examine the levels of *pglG* diversity both within and between species groups, 136 gene sequences were used in the construction of a neighbor-joining unrooted phylogenetic tree (Figure 2). *N. elongata, N. oralis, N. mucosa* and *N. subflava pglG* alleles show high degree of intraspecies conservation and have seemingly diverged to form distinct species-specific *pglG* groups. This is measured by the degree of genetic changes indicated by the tree (branch) scale (i.e. 0.1 = 10% in sequence divergence). All *N. cinerea* and seven *N. polysaccharea* alleles occupy the same clade. As *N. polysaccharea* represents a polyphyletic *Neisseria* group, the extreme diversity seen in *pglG* is not surprising as evidence of horizontal gene transfer and undetermined speciation events result in poor characterization of the current isolates in this group (Bennett et al. 2012). Two classes of *N. meningitidis* alleles emerged from this analysis as one class is more related to the *N. cinerea*/*N. polysaccharea* type, whereas the second class is more related to *N. lactamica*. *N. gonorrhoeae* alleles form its own subclade with a low-sequence diversity. Two major branching points were found with the first diverging into the pathogenic *Neisseria, N. lactamica, N. polysaccharea* and *N. cinerea* species, and the other into *N. subflava, N. mucosa, N. oralis* and *N. elongata* species. One *N. meningitidis* allele was found in the *N. subflava* clade node and could indicate recombination events between the two species. The *pglG* alleles from *N. elongata* and *N. oralis* have diverged the most from the pathogenic *Neisseria* in sequence relatedness as observed by the longer branch lengths and distance from those alleles. *N. gonorrhoeae, N. meningitidis, N. lactamica* and three *N. polysaccharea pglG* sequences show relatively higher interspecies relatedness compared to the commensal *Neisseria* species groupings as observed by the short, and at times indistinguishable branching distances within each grouping (Figure 2). Based on the relatively high conservation of the alleles within neisserial taxonomic groupings, an allele was selected from each taxonomic group and two each from *N. meningitidis* and *N. gonorrhoeae* isolates for further amino acid sequence similarity and identity analyses (Supplementary data, Figures S1 and S2).

Complementation analyses of *pglG* alleles in *N. elongata* subsp. *glycolytica*

To examine the functionality and potential activities of *pglG* gene products, representative alleles from across the *Neisseria* genus were

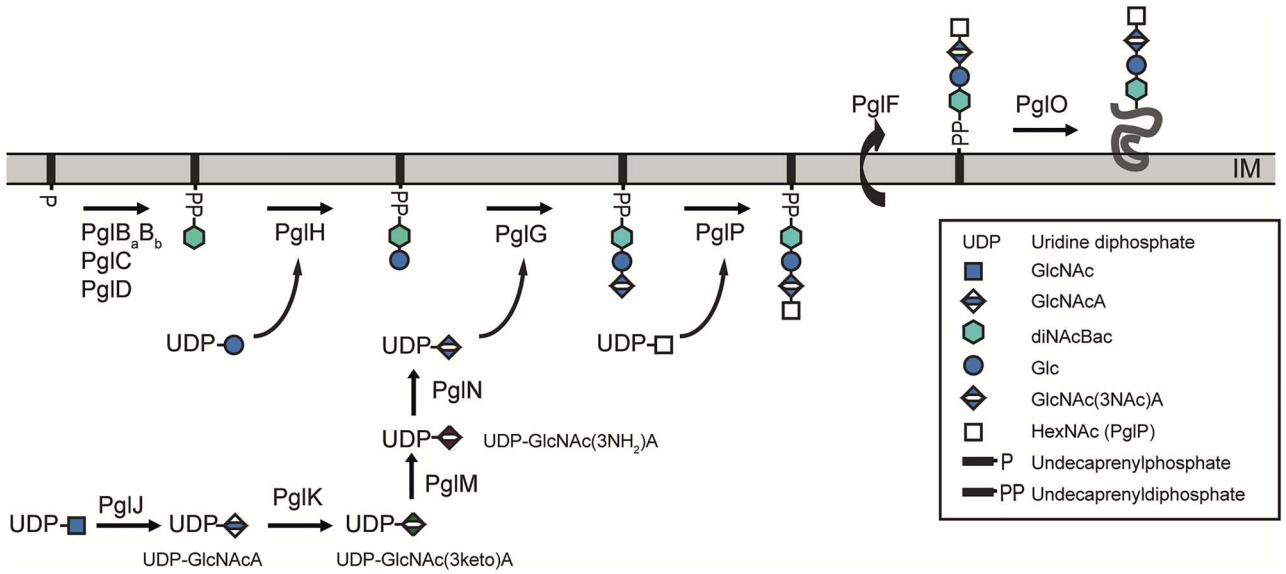


Fig. 1. The *O*-linked protein glycosylation pathway in *N. elongata spp. glycolytica*. The products of genes *pglJ*, *pglK*, *pglM* and *pglN* are responsible for the synthesis of UDP-di-*N*-acetylglucuronic acid that is used as a donor by the PglG glycosyltransferase in the extension of the UND-PP-disaccharide. PglG$_{Nelgly}$ has relaxed specificity and minimally requires a UDP-glucuronic acid and accordingly PglJ. Although *pglG* is present in all human-associated *Neisseria* species, *pglJ* is not found in *N. gonorrhoeae*, *N. meningitidis* and *N. lactamica*. This situation provided the incentive to address the status of *pglG* in the latter species. OM, outer membrane; IM, inner membrane.

introduced into the well-characterized *N. elongata* subsp. *glycolytica* strain KS944 (aka ATCC 29315-Table II) using a previously detailed allelic replacement method (Hadjineophytou et al. 2019). The deletion of *pglG* in this background leads to the expression of a truncated diNAcBac-Glc protein-associated disaccharide whose presence on glycoproteins is detectable by reactivity with the pDAb2 antiserum in immunoblotting (Børud et al. 2011). To facilitate the analyses, the recipient strains expressed either the NirK nitrite reductase or cytochrome C_5-protein bearing *C*-terminally localized polyhistidine tags. As seen previously (Wang et al. 2019), introduction of the *pglG* deletion led to the loss of a broad ladder of NirK forms whose variable mobilities reflect macroheterogeneity (i.e. differences in the number of occupied glycan attachment sites per protein) and appearance of a more compressed and faster migrating set of NirK forms (Figure 3). Based on pDAb2 antibody reactivity and MS analyses, these changes in NirK occur in conjunction with the acquisition of diNAcBac-Glc glycoforms. Reintroduction of *N. elongata* subsp. *glycolytica* (*Nel$_{gly}$*) *pglG* into this background led to the restoration of the wildtype NirK mobility pattern and absence of pDAb2 antibody reactivity seen for the wildtype background. Complementation using *pglG* alleles derived from other species led to varying results. Introduction of alleles from *N. oralis* (*Nora*) and *N. mucosa* (*Nmuc*) yielded immunoblot results similar to the wildtype background suggestive of complementing activities although the low, residual levels of pDAb2 reactivity associated with the *pglG$_{Nmuc}$* gene showed low level, residual incorporation of the diNAcBac-Glc glycan. The pDAb2 reactive-NirK forms seen there migrated slower than those seen in the *pglG* null background suggesting the presence of individual NirK molecules bearing mixtures of glycoforms. Analogous but slightly varied results were seen for the strains complemented with *pglG* from *N. subflava (Nsu)*, *N. cinerea (Nci)* and *N. polysaccharea* (*Npo*) strains. These all showed weak signals of pDAb2 reactivity as seen in the *pglG$_{Nmuc}$* strain but with altered patterns of ladder forms to those seen in the wildtype and other complemented backgrounds. Similar testing of *pglG* alleles from *N. lactamica (Nla)*, *N. meningitidis (Nme)* and *N. gonorrhoeae* (*Ngo*) were complicated due to the presence of phase variable tract segments of polyC mononucleotide repeats. Specifically, some alleles are in out-of-frame configurations (e.g. all *N. lactamica pglG*) and confirming the "ON" versus "OFF" phase status of these alleles can be technically challenging. We therefore chose to assess the function of wildtype alleles bioinformatically predicted to be in an ON configuration and mutants where the phase-variable segment has been locked into an ON configuration by altering codon usage to preclude hypermutability (Supplementary data, Figure S2B). Both such versions of *pglG* from the *N. lactamica* strain showed NirK laddering migration patterns distinct from that seen in the *pglG* null background. However, the wildtype allele retained significant pDAb2 reactivity while that in locked ON configuration had clearly reduced pDAb2 reactivity. These patterns were mirrored by the equivalent allele forms from two *N. meningitidis* strains although evidence for discrete, distinct NirK forms (seen as laddering) was less obvious. In the cases of the equivalent allele forms from *N. gonorrhoeae*, no evidence for either altered NirK mobility or pDAb2 antibody reactivity relative to the *pglG* null background was detected. Results derived for examining the effects of each of the species alleles on C_5 mobility corroborated those made for NirK (Supplementary data, Figure S3).

Mass spectrometric analysis of *pglG* allele effects on protein glycosylation and glycan composition

We next sought to examine functionality and activities of *pglG* gene products in more detail using MS analyses. Here, trypsin- and chymotrypsin-derived peptides of affinity-purified NirK were subjected to liquid chromatography–tandem mass spectrometry (LC–MSMS) and the generated MSMS spectra investigated for the presence of glycan reporter ions. As previously shown (Anonsen et al. 2016), LC–MSMS extracted-ion chromatograms (XIC) of trypsin-digested NirK-His from the wildtype background showed

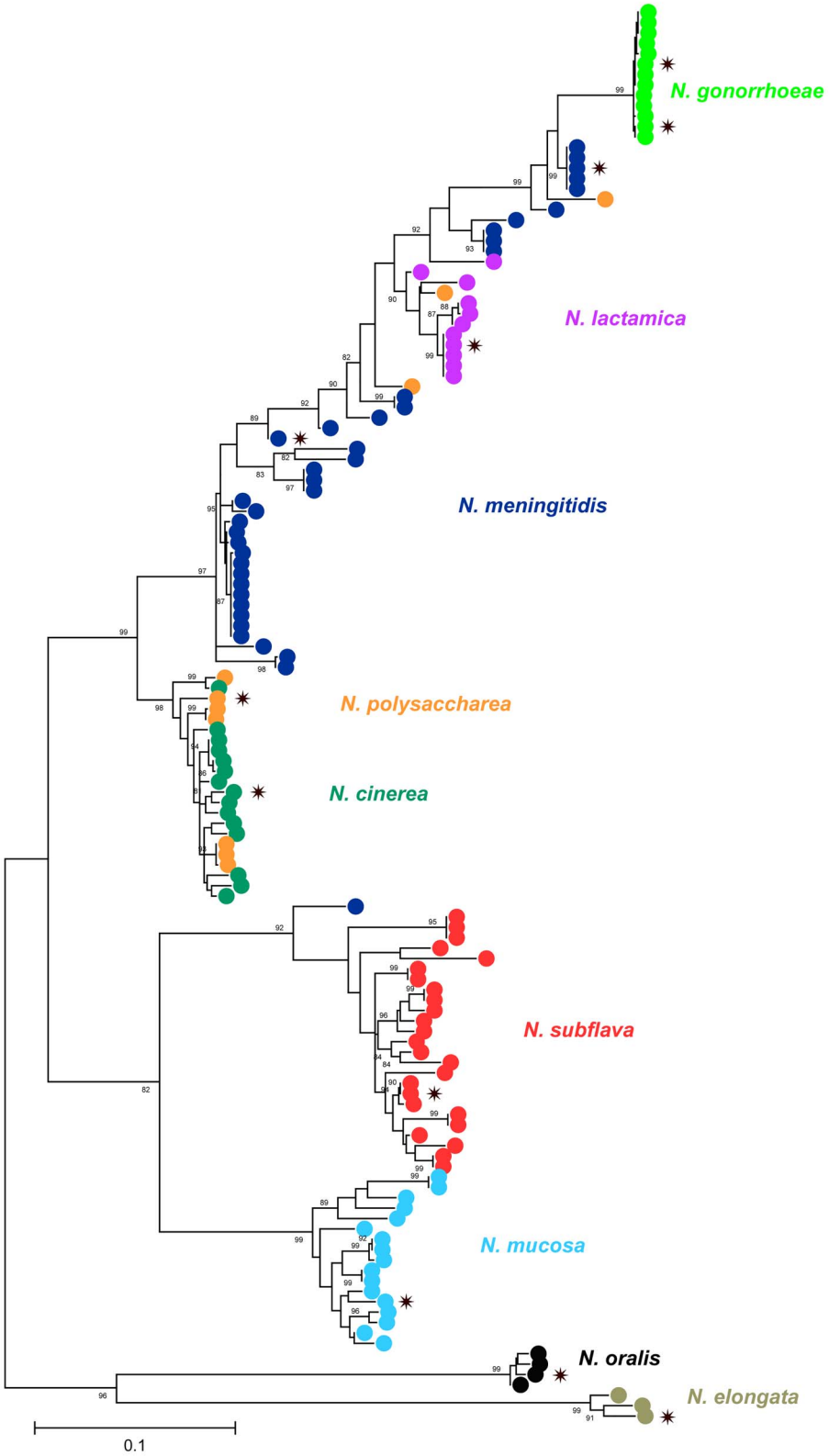


Fig. 2. Phylogenetic analysis of *pglG* alleles. Maximum likelihood phylogenetic tree of *pglG* nucleotide sequences. Consensus support values ≥80% are shown above the branches. *PglG* alleles that were selected as species representatives to be used in this study are marked with a star (*).

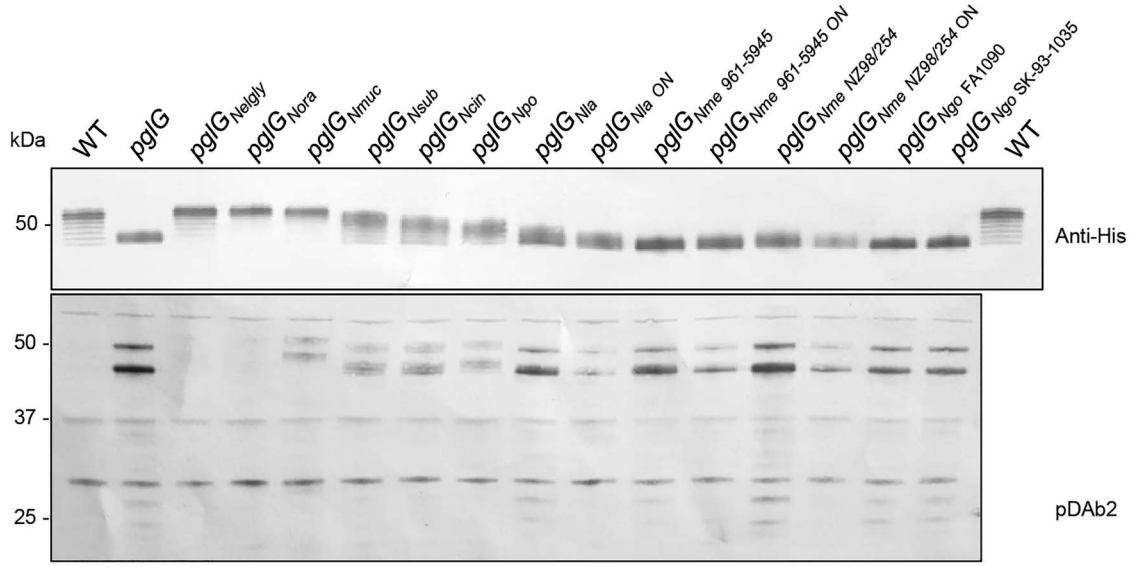


Fig. 3. Complementation panel of representative neisserial *pglG* alleles in *N. elongata* subsp. *glycolytica*. (**Top panel**) Immunoblot of whole cell lysates using a tetra-His epitope recognizing antibody (anti-His) from strains expressing NirK-His using wildtype *N. elongata* subsp. *glycolytica* (WT, NW37), *pglG* (NW39), $pglG_{Nelgly}$ (NW77), $pglG_{Nora}$ (NW83), $pglG_{Nmuc}$ (NW88), $pglG_{Nsub}$ (NW95), $pglG_{Ncin}$ (NW82), $pglG_{Npo}$ (NW107), $pglG_{Nla}$(NW160), $pglG_{Nla\ ON}$ (NW207), $pglG_{Nme961–5945}$ (NW142), $pglG_{Nme961–5945\ ON}$(NW165), $pglG_{Nme\ NZ98/254}$ (NW150), $pglG_{Nme\ NZ98/254\ ON}$ (NW193), $pglG_{Ngo\ FA1090}$ (NW47), $pglG_{Ngo\ SK\text{-}93\text{-}1035}$(NW70). (**Bottom panel**) Immunoblot of the strains using pDAb2, a polyclonal antibody recognizing the diNAcBac-Glc epitope.

overlapping glycan reporter ions peaks representing diNAcBac, diNAcBac-Hex, diNAcHexA and HexNAc (Figure 4A). Moreover, these carbohydrate component reporter ions translated into the full-length tetrasaccharide as the glycan structure ions for diNAcBac-Hex-diNAcHexA and diNAcBac-Hex-diNAcHexA-HexNAc individually completely overlap in the XIC (Figure 4B). To illustrate the connection between the presence of individual carbohydrate component ions and the glycan structure ions representing the tetrasaccharide, an MS spectrum of the trypsin-generated peptide ^{372}GTGAAAGAASGASGASAPAAPASSASGSSNPYGGEEGHHHH-HHH^{415} carrying five tetrasaccharides is presented in Figure 4C. The data confirm earlier findings that enzymatically derived glycopeptides were associated with the tetrasaccharide glycoform with no detectable microheterogeneity and only a small fraction of glycopeptides demonstrating macroheterogeneity. In contrast and as previously reported (Anonsen et al. 2016), the XIC of the *pglG* null mutant displays only glycopeptides carrying disaccharide, illustrated by the overlapping reporter ions for diNAcBac (*m/z* 211.1) and diNAcBac-Hex (*m/z* 391.2) (Figure 4D and E) and lack of ions representing diNAcHexA and HexNAc as well as glycan structure ions representing diNAcBac-Hex-diNAcHexA and diNAcBac-Hex-diNAcHexA-HexNAc. This is demonstrated in the representative MS spectrum of the trypsin-generated peptide carrying disaccharides (Figure 4F).

Interestingly, the XIC from complemented *pglG* from *N. lactamica* (strain NW207) showed a high abundance of HexNAc reporter ion (at *m/z* 204.086) (Figure 4G) but showed low levels of the tetrasaccharide structural ions (Figure 4H) indicating that the HexNAc carbohydrate was part of a novel glycan structure. Investigating an MS spectrum of the trypsin-generated peptide containing HexNAc reporter ion (Figure 4I) showed (in addition to the wildtype tetrasaccharide ion at *m/z* 852.3) a glycan reporter ion at *m/z* 594.252, consistent with an HexNAc terminating diNAcBac-Hex-HexNAc trisaccharide. In fact, the sextuple-charged precursor ion at *m/z* 905.885 (charge adjusted mass 5412.275 [M + H]+), shown in the MS spectrum Figure 4I, corresponds to the mass of the unmodified peptide (theoretical mass 3967.722 [M + H]+) carrying one tetrasaccharide (851.316 Da) and one trisaccharide (593.245 Da).

In addition, when exploring the chromatogram and MS spectra of NirK-His peptides purified from the *N. elongata* strain subsp. *glycolytica* expressing *N. meningitidis* NZ 98/254 *pglG* (strain NW193) numerous HexNAc reporter ions were detected (Figure 4J). Hardly any structural ions representing the full-length tetrasaccharide or the diNAcHexA carrying trisaccharide were found (Figure 4K). However, several peaks representing the trisaccharide diNAcBAc-Hex-HexNAc structure ion (*m/z* 594.252) were detected. This indicated that complementation with $PglG_{NmeNZ98/254}$ shifted the glycoform present on NirK-His from a diNAcHexA-based tetrasaccharide to a HexNAc-terminating trisaccharide. In fact, when investigating MS spectra of the generated peptides only two glycopeptides carrying full-length tetrasaccharides were detected (data not shown). However, several glycopeptide MS spectra showed the HexNAc reporter ion in addition to the *m/z* 594.3 tetrasaccharide structure ion as shown in the MS spectra of the trypsin-generated peptide carrying two diNAcBAc-Hex-HexNAc glycans (Figure 4L).

When complementing *N. elongata* subsp. *glycolytica* with *pglG* genes from across the genus *Neisseria*, some general observations were noted (summarized in Table I, Supplementary data, Figure S4). Overall, the MS data generated from complementation with the *pglG* alleles from *N. lactamica* and *N. meningitidis* strains indicated the presence of the diNAcBac-Glc-HexNAc terminating trisaccharide. The phase-variable *pglG* allele from *N. meningitidis* 961–5945 (expressed in strain NW142) displayed only disaccharide containing glycoforms, but when the allele was locked in an ON configuration, the HexNAc terminating trisaccharide glycoform was detected in this background (strain NW165). The MS spectra generated from the cross-complemented *N. gonorrhoeae pglG* alleles tested (strains

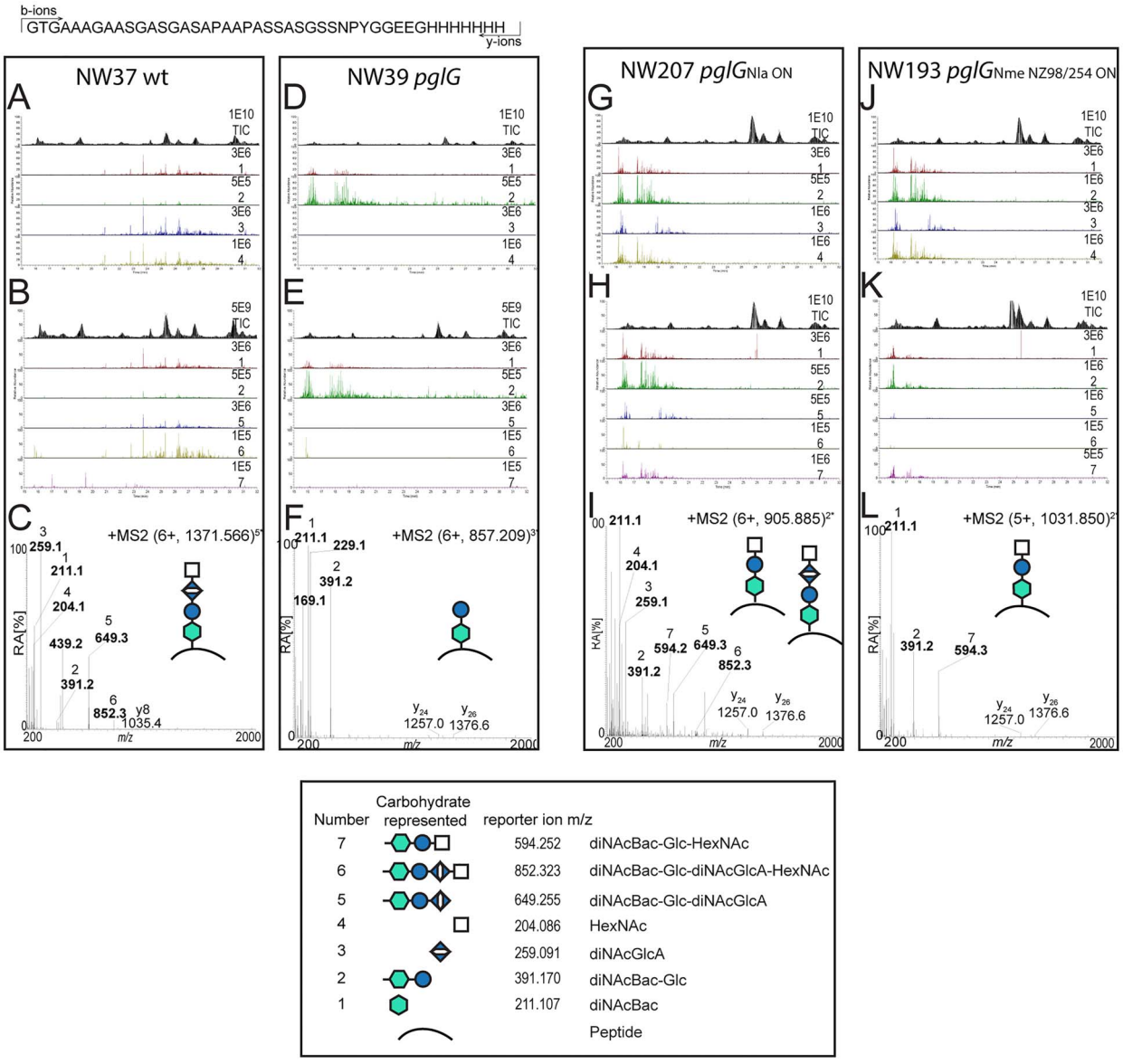


Fig. 4. Effect of cross-complemented *pglG* on glycan structure. Shown are LC–MSMS chromatograms of the trypsin-derived peptide ^{372}GTGAAAG AASGASGASAPAAPASSASGSSNPYGGEEGHHHHHHH^{415} from affinity-purified NirK-His from *N. elongata* subsp. *glycolytica* strains carrying cross-complemented *pglG* at between 20 and 32 min. The total ion chromatogram (TIC) intensity values represent the amount of peptides entering the mass spectrometer. The selected ion chromatograms (XIC) are of the four glycan reporter ions characteristic for a tetrasaccharide, as well as the glycan structure reporter ions representing the *N. elongata* subsp. *glycolytica* wildtype tetrasaccharide and the glycan structure ion representing the novel trisaccharide: diNAcBac at *m/z* 211.108 (1), diNAcBac-Hex at *m/z* 391.170 (2), diNAcHexA at *m/z* 259.093 (3), HexNAc at *m/z* 204.086 (4), diNAcBac-Hex-diNAcHexA at *m/z* 649.254 (5) diNAcBac-Hex-diNAcHexA-HexNAc at *m/z* 852.323 (6) and diNAcBac-Hex-HexNAc at *m/z* 594.252 (7). All glycan reporter ions were searched with a 10-ppm window. The MSMS spectrum demonstrates the presence of glycan reporter ions (marked in boldface type and numbered as described above). All XIC values were normalized to display all relevant glycan reporter ions to the most abundant glycan reporter ion, or for the less abundant reporter ion or to the abundance of the overall least abundant reporter ion in the chromatogram. (**A**) LC–MSMS chromatogram of trypsin-derived peptides from affinity-purified NirK-His from wildtype (WT) *N. elongata* subsp. *glycolytica* (strain NW37) with displayed XIC selected for glycan carbohydrate component reporter ions. The TIC intensity value was set at 1E10. (**B**) Panel A with displayed XIC selected for glycan structure ions. The TIC intensity value was set at 5E9. (**C**) Representative MSMS spectrum of the trypsin-generated peptide from NirK-His from wildtype (strain NW37) carrying five diNAcBAc-Hex-diNAcHexA-HexNAc tetrasaccharides. (**D**) LC–MSMS chromatogram of trypsin-derived peptides from affinity-purified NirK-His from a *N. elongata* subsp. *glycolytica pglG* null background (strain NW39) with displayed XIC selected for glycan carbohydrate component reporter ions. The TIC intensity value was set at 1E10. (**E**) Panel D with displayed XIC selected for glycan structure ions. The TIC intensity value was set at 5E9. (**F**) Representative MSMS spectrum of the trypsin-generated peptide from NirK-His selected from panel D carrying three diNAcBAc-Hex disaccharides. (**G**) LC–MSMS chromatogram of trypsin-derived peptides from affinity-purified NirK-His from a *pglG*$_{Nla}$ complemented background (strain NW207) with displayed XIC selected for glycan carbohydrate component reporter ions. The TIC intensity value was set at 1E10. (**H**) Panel G with displayed XIC selected for glycan structure ions. The TIC intensity value was set at 5E9. (**I**) Representative MSMS spectrum of the trypsin-generated peptide from NirK-His selected from panel G carrying a diNAcBAc-Hex-diNAcHexA-HexNAc tetrasaccharide and a diNAcBAc-Hex-HexNAc trisaccharide. (**J**) LC–MSMS chromatogram of trypsin-derived peptides from affinity-purified NirK-His from pglG$_{Nme}$NZ98/254 background (strain NW193) with displayed XIC selected for glycan carbohydrate component reporter ions.

NW47 and NW70) displayed only glycopeptides carrying a disaccharide identical to a *pglG* null background (Table I, Supplementary data, Figure S4), consistent with the pDAb2 immunoblotting data (Figure 3). In testing *pglG* alleles derived from the other commensal *Neisseria* species, less microheterogeneity was observed as assessed by overlapping glycan structure ions in the MS chromatograms when complemented with *pglG* alleles originating from species more closely related to *N. elongata* subsp. *glycolytica (*i.e. $pglG_{Nora}$, $pglG_{Nmuc}$, $pglG_{Nsu}$, $pglG_{Nci}$ and $pglG_{Npo}$—Supplementary data, Figures S1 and S4). The patterns of micro-and macroheterogeneity observed in the LC–MSMS analysis of NirK-His from complemented *N. elongata* subsp. *glycolytica* strains generally supports the observed pDAb2 immunoblot reactivity pattern/bands as seen in Figure 3. However, using this MS approach, we were not able to quantify the abundance of the individual glycoforms detected and as such, only the more abundant glycoforms detected are reported in Table I. Thus, there are instances where one sees reactivity with the pDAb2 serum but the corresponding diNAc-Hex disaccharide was not reported in the MS data.

Influence of *pglK* status on *pglG* activity and function

In *N. elongata* subsp. *glycolytica*, PglG functions as an extending glycosyltransferase to add GlcNAc(3NAc)A at the third position of the normal tetrasaccharide. UDP-GlcNAc(3NAc)A is synthesized sequentially by the products of four enzymes initiating with PglJ that acts as a UDP-*N*-acetyl-D-glucosamine 6-dehydrogenase to generate UDP-GlcNAcA. This glycan is further modified by the products of *pglK*, *pglM* and *pglN* to generate UDP-GlcNAc(3NAc)A. $PglG_{Nelgly}$ minimally requires a UDP-GlcNAcA donor and functions optimally using UDP-GlcNAc(3NAc)A (Wang et al. 2019). As such, a *pglJ* null mutant phenocopies a *pglG* null mutant in this system, and *pglK*, *pglM* and *pglN* mutants show increased degrees of glycan microheterogeneity as evidenced by the presence of diNAcBac-hex disaccharide and diNAcBac monosaccharide modified glycoproteins in those backgrounds (Wang et al. 2019). The assumption then is that given the lack of specificity of the putative PglF flippase and the PglL/O oligosaccharyltransferase for Und-PP-oligosaccharides, kinetic defects or otherwise reduced functionality of PglG with regard to donor availability or donor specificity would lead to a short-circuited pathway.

Previously, we reported that some species possessing *pglG* carry only *pglJ* or the *pglJ* and *pglK* genes, while others lack all genes for the UDP-GlcNAc(3NAc)A biosynthetic pathway (Hadjineophytou et al. 2019). We hypothesized accordingly that *pglG* alleles from such backgrounds might have differing UDP-sugar donor specificities. We, therefore, assessed glycan-related phenotypes in *N. elongata* subsp. *glycolytica* strains carrying various *pglG* alleles in a *pglK* background by examining both NirK and C_5 mobility patterns and reactivity with pDAb2 antibodies (Figure 5). In particular, levels of disaccharide modification (as seen by pDAb2 reactivity) were enhanced and the relative mobilities of the detected glycoproteins were reduced. The strain expressing $pglG_{Nora}$, showed phenotypes similar to those exhibited by *pglK* mutant in the otherwise wildtype *N. elongata* subsp. *glycolytica* background. Remarkably, the opposite result with regard to the glycoprotein pDAb2 reactivity was seen using the alleles from *N. mucosa*, *N. subflava*, *N. cinerea* and *N. polysaccharea* (Figure 5). In these strains, introducing the *pglK* mutation led to a loss of pDAb2 reactivity while clearly maintaining significant levels of NirK and C_5 with reduced mobility. As previously proposed, we surmise that the latter phenotype reflects the presence of individual proteins carrying both the diNAcBac-Glc disaccharide as well as larger glycans (Wang et al. 2019). We assume these results indicate the presence of larger glycoforms extended on the diNAcBac-Glc disaccharide backbone. Together, these findings begin to reveal subtle genetic interactions defining altered PglG activities and functions across the genus.

Evidence for pseudogenization of *pglG* in *N. gonorrhoeae*

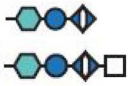
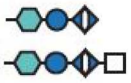
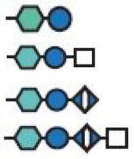
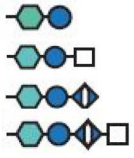
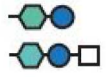
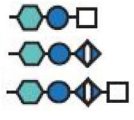
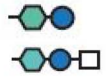
Currently, there is no evidence for *pglG*-related glycan phenotypes in either *N. gonorrhoeae* or in *N. elongata* subsp. *glycolytica* expressing gonococcal *pglG* alleles. We hypothesized these findings could be due to alleles that encode enzymatically inactive products and thus, that gonococcal *pglG* genes are pseudogenes due to function-ablating missense mutations. To examine this possibility in more detail, *pglG* ORF sequences from across the genus were aligned and scored for the presence of amino acid residues that were highly conserved in the majority of species but significantly deviating in *N. gonorrhoeae*. We noticed two such unique variants in the amino acid sequences of *N. gonorrhoeae* PglG when compared to other PglG sequences. These included a cysteine at residue 112 in place of a tyrosine seen in all other species, a threonine in place of an alanine at residue 163 and the absence of a leucine situated between residues 163 and 164 (Supplementary data, Figure S5). To expedite assessing the potential effects of these polymorphisms, we chose for practical reasons to alter the corresponding sites in the active $pglG_{Nmuc}$ gene and express the mutated alleles in *N. elongata* subsp. *glycolytica*. As shown in Figure 6, none of these single amino acid substitutions disrupted PglG function as measured by the relative mobility of NirK, although the isoleucine substitution (F111I) and the leucine deletion (ΔL164) mutation did increase glycoprotein pDAb2 reactivity indicating some compromised PglG function. We also tested a mutant in which both the cysteine substitution at position 112 and the deletion of the leucine were incorporated and in this case, the complemented strain phenocopied the *pglG* null mutant. We conclude with caveats that *N. gonorrhoeae pglG* alleles are missense pseudogenes whose products fail to function due minimally to two discrete alterations in primary structure.

Distribution of phase-variable *pglG* alleles across the genus *Neisseria*

The *pglA*, *pglE* and *pglH* glycosyltransferase genes found in *N. gonorrhoeae*, *N. meningitidis* and *N. lactamica* are subject to high-frequency on–off expression as a result of hotspots for frame-shifting events in the ORFs (Børud et al. 2011). Analogous events have been proposed to take place within *pglG* genes of these species. Although an earlier work examined the phase variability of all four of these genes in neisserial species, the overall number of genomes examined was limited (Børud et al. 2011). We scanned in an unbiased fashion for the occurrence of DNA repeat elements capable of generating significant frequencies of phase variable expression (Figure 7,

Fig. 4. The TIC intensity value was set at 1E10. (**K**) Panel J with displayed XIC selected for glycan structure ions. The TIC intensity value was set at 1E10. (**L**) Representative MSMS spectrum of the trypsin-generated peptide from NirK-His selected from panel J carrying two diNAcBAc-Hex-HexNAc trisaccharides.

Table I. Summary of major glycoforms resulting from complementation with defined *pglG* alleles (represented by their respective monosaccharide symbols as in Figure 4)

Strain/Genotype	Glycoforms detected	
NW37 WT		diNAcBac-Glc-diNAcHexA-HexNAc
NW39 *pglG*		diNAcBac-Glc
NW77 $pglG_{Nelgly}$		diNAcBac-Glc-diNAcHexA-HexNAc
NW83 $pglG_{Nora}$		diNAcBac-Glc-diNAcHexA-HexNAc
NW88 $pglG_{Nmuc}$		diNAcBac-Glc-diNAcHexA-HexNAc
NW95 $pglG_{Nsub}$		diNAcBac-Glc-diNAcHexA diNAcBac-Glc-diNacHexA-HexNAc
NW81 $pglG_{Ncin}$		diNAcBac-Glc-diNAcHexA diNAcBac-Glc-diNAcHexA-HexNAc
NW107 $pglG_{Npo}$		diNAcBac-Glc-diNAcHexA diNAcBac-Glc-diNAcHexA-HexNAc
NW160 $pglG_{Nla}$		diNAcBac-Glc diNAcBac-Glc-HexNAc diNAcBac-Glc-diNAcHexA diNAcBac-Glc-diNAcHexA-HexNAc
NW207 $pglG_{Nla}$ ON		diNAcBac-Glc diNAcBac-Glc-HexNAc diNAcBac-Glc-diNAcHexA diNAcBac-Glc-diNAcHexA-HexNAc
NW142 $pglG_{Nme}$ 961-5945		diNAcBac-Glc
NW165 $pglG_{Nme}$ 961-5945 ON		diNAcBac-Glc diNAcBac-Glc-HexNAc
NW150 $pglG_{Nme}$ NZ98/254		diNAcBac-Glc-HexNAc diNAcBac-Glc-diNAcHexA diNAcBac-Glc-diNAcHexA-HexNAc
NW193 $pglG_{Nme}$ NZ98/254 ON		diNAcBac-Glc diNAcBac-Glc-HexNAc
NW47 $pglG_{Ngo}$ FA1090		diNAcBac-Glc
NW70 $pglG_{Ngo}$ SK-93-1035		diNAcBac-Glc

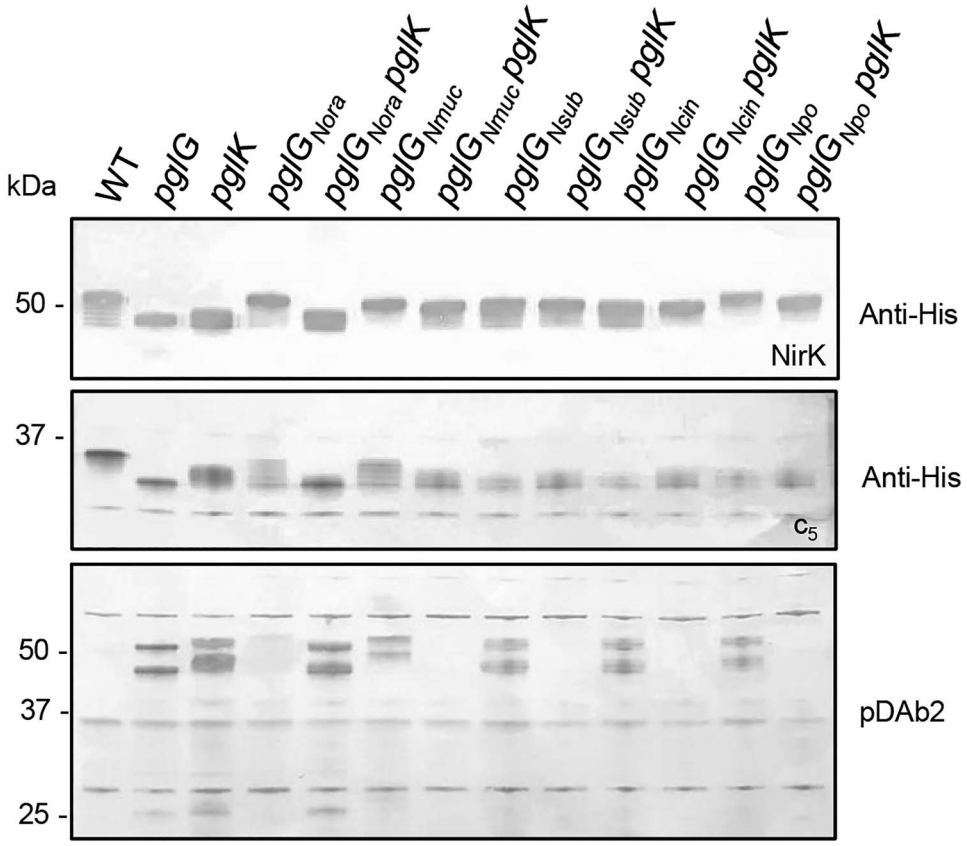


Fig. 5. PglG function is influenced by *pglK* status as determined by immunoblot analysis in *N. elongata* subsp. *glycolytica*. (**Top two panels**) Immunoblot of whole cell lysates using anti-His from strains expressing either NirK-His or c_5-His (strain names listed as corresponding pairs accordingly) using *N. elongata* subsp. *glycolytica* wildtype (NW37/KS997), *pglG* (NW39/NW267), *pglK* (RV731/NW376), $pglG_{Nora}$ (NW83/NW336), $pglG_{Nora}$ *pglK* (NW219/NW311), $pglG_{Nmuc}$ (NW88/NW282), $pglG_{Nmuc}$ *pglK* (NW223/NW388), $pglG_{Nsub}$ (NW95/NW344), $pglG_{Nsub}$ *pglK* (NW392), $pglG_{Ncin}$ (NW82/NW346), $pglG_{Ncin}$ *pglK* (NW238/NW396), $pglG_{Npo}$ (NW107/NW352), $pglG_{Npo}$ *pglK* (NW242/NW400). (**Bottom panel**) Immunoblot of the strains above using diNAcBac-Glc recognizing antibody (pDAb2).

see Materials and methods). These were present in over 94% of *N. meningitidis* and all the *N. lactamica* alleles. Those scored as not phase variable in these species were exclusively represented by short polyC tracts and likely reflect alleles in which repeat contraction leads to stretches too short to phase vary at significant rates. This same situation accounts for the approximately 69% of *N. gonorrhoeae* alleles scored as nonphase variable. The situation is different for *N. polysaccharea* where most alleles scored as nonphase variable in fact are more closely related to *pglG* alleles in *N. cinerea* (Figure 2). This situation is in line with earlier results showing that *N. polysaccharea* is polyphyletic with isolates bearing overall genome sequence identities with those of either *N. meningitidis* and *N. lactamica* or *N. cinerea*. No evidence for phase variability was seen for alleles from the remaining species (Data set 2).

Discussion

The *O*-linked *pgl* systems expressed by human–associated *Neisseria* species provide a unique opportunity to explore both the genetic basis for protein-associated glycoform diversification and the potential selective forces shaping glycan repertoire. Here, we discovered allelic polymorphisms in the *pglG* gene as a new genetic determinant underlying neisserial glycoform diversification. Specifically, we used trans-species genetic complementation to discover that allelic polymorphisms of *pglG* are correlated with the expression of glycosyltransferases with a range of functional activities. Specifically, different alleles are associated with the use of distinct UDP-sugar donors to generate a group of heretofore unidentified oligosaccharides due to unique sugar moieties at the third position. In addition to the incorporation of di-*N*-acetyl-glucuronic acid seen for the endogenous *N. elongata* subsp. *glycolytica* PglG, some *pglG* gene products such as those from *N. subflava*, *N. cinerea* and *N. polysaccharea* appear to function optimally in the incorporation of glucuronic acid rather than di-*N*-acetyl-glucuronic acid or its biosynthetic pathway intermediates. Interestingly, the latter such alleles are found in species whose genomes carry intact *pglJ* orthologous genes but for which genes orthologous to *pglK*, *pglM* and *pglN* are absent. Assuming such alleles function in their endogenous source strain as they do in *N. elongata* subspecies *glycolytica*, it would follow that such PglG alleles would have evolved specificity for a UDP-glucuronic acid donor.

Two distinct functionalities are associated with *pglG* from species lacking a *pglJ* orthologue (and thus the ability to synthesize UDP-glucuronic acid). In the case of the *pglG* alleles tested from *N. meningitidis* and *N. lactamica*, a novel trisaccharide terminating in a HexNAc was present. These results can best be ascribed to PglG from these backgrounds having a new function in utilizing a UDP-HexNAc donor. To begin to examine the nature of the associated HexNAc moiety, a lectin-based detection assay using succinylated wheat germ agglutinin (sWGA) was used. This lectin shows a strong

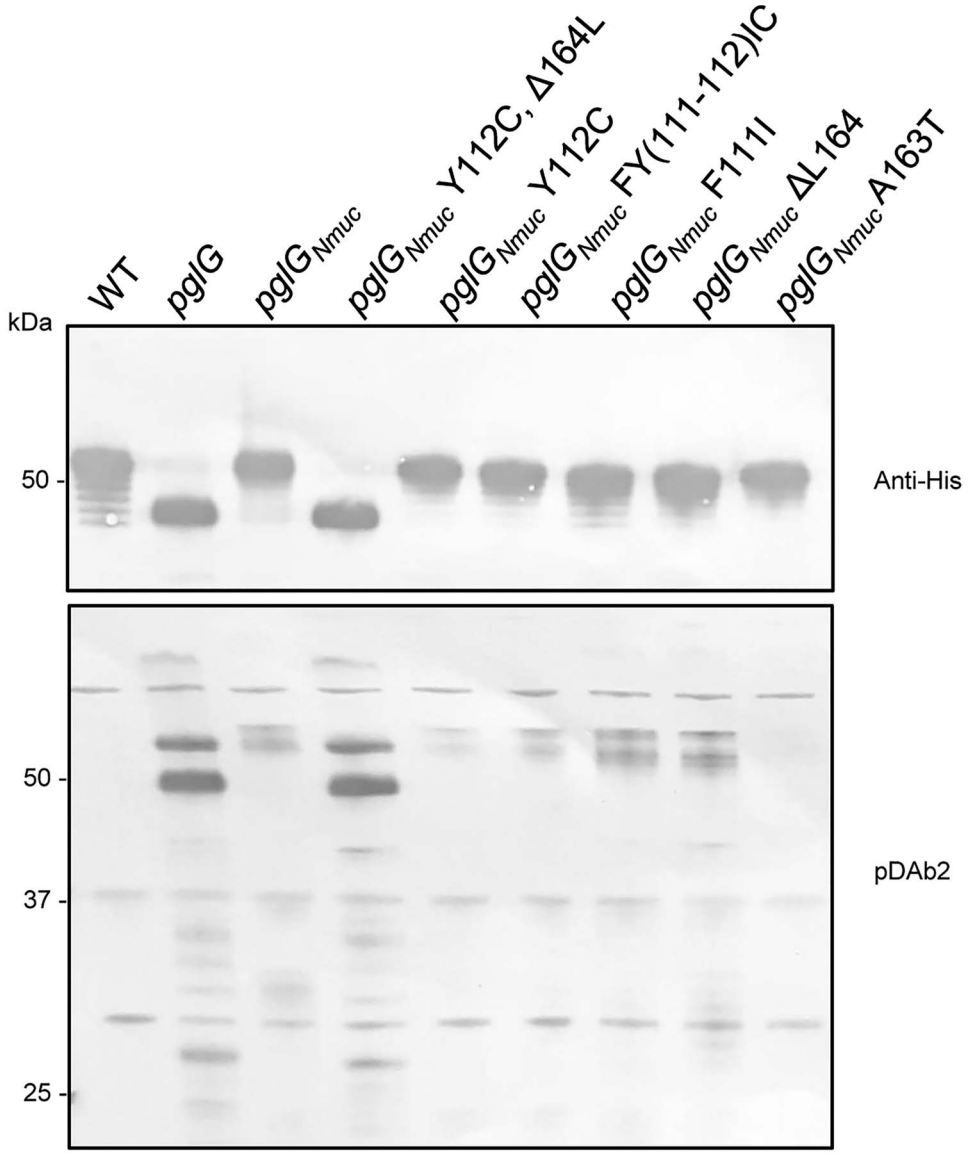


Fig. 6. Introduction of *N. gonorrhoeae pglG* specific mutations to a *N. mucosa pglG* allele results in loss of complementation. (**Top panel**) Immunoblot of whole cell lysates using anti-His from strains expressing NirK-His using *N. elongata* subsp. *glycolytica* wildtype (NW37), *pglG* (NW39), *pglG*$_{Nmuc}$ (NW88), *pglG*$_{Nmuc}$Y112C, ΔL164 (NW428), *pglG*$_{Nmuc}$ Y112C (NW413), *pglG*$_{Nmuc}$ FY(111–112)IC (NW419), *pglG*$_{Nmuc}$ FY111C (NW419), *pglG*$_{Nmuc}$ ΔL164 (NW422), *pglG*$_{Nmuc}$ A163T (NW425). (**Bottom panel**) Immunoblot of the strains above using diNAcBac-Glc recognizing antibody (pDAb2).

preference for terminal GlcNAc residues, and it selectively recognized the NirK glycoprotein from backgrounds expressing the *pglG* alleles of *N. meningitidis* and *N. lactamica*. (Supplementary data Figure S6). Although this result suggests that the HexNAc associated with these *pglG* is GlcNAc, confirming the identity of the sugar will require further biochemical characterization of these PglG enzymes and of the associated trisaccharide. Taken together with prior studies, the identification of the HexNAc-terminating trisaccharide increases the genetically defined repertoire capable of being expressed by a single *N. meningitidis* strain to 11 glycoforms and a total of 22 distinct glycoforms found across that species. The situation in *N. lactamica* remains less well characterized but based on the presence there of all of the glycosyltransferases and glycan-modifying genes found in *N. meningitidis*, a repertoire of glycoforms comparable to that established there can be foreseen.

In contrast, the findings for the *N. gonorrhoeae pglG* alleles suggest that the encoded protein is nonfunctional and that the genes have undergone pseudogenization. This scenario fits with the documented pseudogenization in this species of *pglP*, whose glycosyltransferase product functions downstream in the synthesis of the tetrasaccharide using the PglG-dependent Und-PP-trisaccharide as a substrate. In contrast to the situation for the gonococcal *pglP* pseudogene that encompasses three ORF terminating nucleotide alterations, lack of activity of gonococcal PglG seems to be associated with the presence of two nonsynonymous amino acid changes (relative to the other *pglG* alleles). We chose for the sake of expediency to assess the outcome of introducing the nonsynonymous substitutions into the wildtype *pglG*$_{Nmuc}$ gene as it was possible that other missense-type substitutions might be present in *pglG*$_{Ngo}$. Interestingly, only introduction of both substitutions into wildtype PglG$_{Nmuc}$ resulted

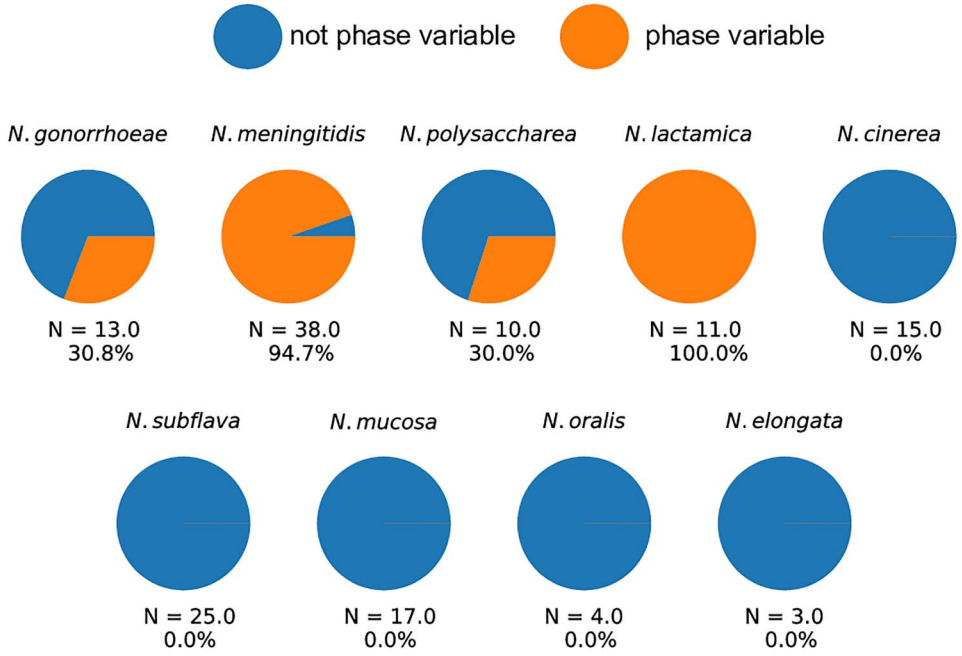


Fig. 7. Distribution of phase-variable *pglG* alleles in *Neisseria*. Phase variation in pglG is limited to *N. gonorrhoeae, N. meningitidis, N. lactamica* and *N. polysaccharea* species. Upto 30–100% of analyzed strains of these species contain polyC tracts with ≥9 repeats. Refer to Data Set 2 for more information on the specific lengths of the polyC tracts detected.

in a loss of function. Although it seems attractive to assume that the presence of the two amino acid substitutions in $PglG_{Nmuc}$ solely preclude function, it is possible that they perturb protein stability. Thus, we cannot discern whether the lack of activity correlates with qualitative versus quantitative alterations in PglG status. It would clearly be interesting to see if correcting either one or both of these nonsynonymous substitutions in $pglG_{Ngo}$ would be sufficient to reconstitute functionality.

A number of *pgl* genes in *N. gonorrhoeae*, *N. meningitidis* and *N. lactamica* have been shown to be contingency loci whose expression can undergo phase-variable expression due to changes in nucleotide repeat tracts. Prior work has shown that *pglH* alleles are predicted to be phase variable in the majority of isolates of *N. gonorrhoeae*, *N. meningitidis* and *N. lactamica* but not in other species (as the genes from the latter lack the nucleotide repeat tracts) (Børud et al. 2011). We carried out similar analyses searching for hypermutable repeat tracts in *pglG* and found that greater than 94% of *N. meningitidis* and *N. lactamica* alleles, less than 31% of gonococcal alleles and 30% of *N. polysaccharea* were capable of phase variation. The nonphase variable gonococcal alleles were exclusively due to repeat contraction (rather than absence of the repeat tract), and it is worth mentioning that gonococci in general have much less allele diversity at most genes. The situation in *N. polysaccharea* is more complex as there appears to be two classes of *pglG* alleles, one of whose members are more related to those of *N. meningitidis* and *N. lactamica* (and are phase variable) and a second whose members are more related to those of *N. cinerea* that lack hypermutable repeat tracts. In addition to *N. cinerea*, potential for phase-variable expression was not seen for the alleles from any of the other species. Expanding this analysis to include all *pgl* alleles from the genus has shown that phase variation is disproportionally found in species from the pathogenic clade (Supplementary data Figure S7). Overall, these results parallel those made for *pglH* and define a clear dichotomy in which phase variable glycan expression is a distinct property of the pathogenic species *N. gonorrhoeae* and *N. meningitidis* and its congeners *N. lactamica* and *N. polysaccharea*.

It should be acknowledged that the findings generated here are almost exclusively based on genetic data (allelic exchange, mutant analyses, glycan phenotyping) as opposed to biochemical studies. In fact, we lack evidence that PglG functions as a classical bacterial glycosyltransferase using UDP sugars as donors to modify a growing Und-PP-substrate, formally proving that will require further biochemical analyses as has been done for glycosyltransferases PglA, PglE and PglH. Despite these limitations, the work here assessing PglG function by genetic complementation emphasizes the potential strengths of studies assessing function and activity in a natural, in vivo context. In particular, this approach provides additional information on the overall integrity of the protein glycosylation system that cannot otherwise be achieved in "one-pot" chemoenzymatic synthesis schemes.

In summary, these data document the extreme versatility capable of being generated in a bacterial glycosyltransferase and expand the level of glycan complexity of the neisserial *pgl* system. The results should be useful in defining both the structure–function relationships of PglG and assessing the evolution of a locus that appears to be under differing selective constraints in species spanning the *Neisseria* genus.

Materials and methods

Bacterial strains and culture conditions

Bacterial strains used in this study are listed in Table II and were grown on conventional GC medium (Difco) as described previously (Anonsen et al. 2016). Antibiotic selection for *N. elongata* subsp. *glycolytica* transformants were at the following concentrations: streptomycin (750ug/mL), kanamycin (50ug/mL) and chloramphenicol

Table II. List of bacterial strains and plasmids used

Strain name	Relevant genotype	Parent	Source
KS944—*N. elongata* subsp. *glycolytica* ATCC 29315	–	–	Lab collection (Anonsen et al. 2016)
N. oralis F0314	–	–	Lab collection
N. mucosa ATCC 29256	–	–	Lab collection
N. subflava CCUG 23930	–	–	Lab collection
N. cinerea ATCC 14685	–	–	Lab collection
N. polysaccharea CCUG 4790	–	–	Lab collection
N. lactamica 020–06	–	–	Lab collection
N. meningitidis 961–5945	–	–	Bente Børud, Norwegian Institute of Public Health
N. meningitidis NZ98/254	–	–	Bente Børud, Norwegian Institute of Public Health
N. gonorrhoeae FA1090	–	–	Lab collection
N. gonorrhoeae SK-93-1035	–	–	Lab collection
N. elongata pglG complementation NirK-His background			
NW37	*nirK-His::cat* Smr	–	(Hadjineophytou et al. 2019)
NW39	*nirK-His pglG::kan/rpsL+*	–	This study
RV731	*nirK-His pglK::kan*	NW39	(Wang et al. 2019)
NW47	$pglG_{N.\ gonorrhoeae\ FA1090}$	NW39	This study
NW70	$pglG_{N.\ gonorrhoeae\ SK\text{-}93\text{-}1035}$	NW39	This study
NW77	$pglG_{N.\ elongata}$	NW39	This study
NW83	$pglG_{N.\ oralis}$	NW39	This study
NW88	$pglG_{N.\ mucosa}$	NW39	This study
NW95	$pglG_{N.\ subflava}$	NW39	This study
NW81	$pglG_{N.\ cinerea}$	NW39	This study
NW112	$pglG_{N.\ polysaccharea}$	NW39	This study
NW160	$pglG_{N.\ lactamica}$	NW39	This study
NW207	$pglG_{N.\ lactamica}$ ON	NW39	This study
NW142	$pglG_{N.\ meningitidis\ 961\text{–}5945}$	NW39	This study
NW165	$pglG_{N.\ meningitidis\ 961\text{–}5945}$ ON	NW39	This study
NW150	$pglG_{N.\ meningitidis\ NZ98/254}$	NW39	This study
NW193	$pglG_{N.\ meningitidis\ NZ98/254}$ ON	NW39	This study
NW219	NW83 *pglK::kan*	NW83	This study
NW223	NW88 *pglK::kan*	NW88	This study
NW227	NW95 *pglK::kan*	NW95	This study
NW238	NW81 *pglK::kan*	NW81	This study
NW242	NW112 *pglK::kan*	NW112	This study
N. elongata pglG complementation c_5-His background			
KS997	*cycB-His::cat*	–	(Anonsen et al. 2016)
NK2259	Smr	KS944	(Hadjineophytou et al. 2019)
NW31	*pglG::kan/rpsL+*	NK2259	This study
NW267	*cycB-His::cat* *pglG::kan/rpsL+*	NW31	This study
NW336	$pglG_{N.\ oralis}$	NW267	This study
NW282	$pglG_{N.\ mucosa}$	NW267	This study
NW344	$pglG_{N.\ subflava}$	NW267	This study
NW348	$pglG_{N.\ cinerea}$	NW267	This study
NW352	$pglG_{N.\ polysaccharea}$	NW267	This study
NW300	$pglG_{N.\ lactamica}$	NW267	This study
NW356	$pglG_{N.\ lactamica}$ PV ON	NW267	This study
NW308	$pglG_{N.\ meningitidis\ 961\text{–}5945}$	NW267	This study
NW360	$pglG_{N.\ meningitidis\ 961\text{–}5945}$ ON	NW267	This study
NW316	$pglG_{N.\ meningitidis\ NZ98/254}$	NW267	This study
NW364	$pglG_{N.\ meningitidis\ NZ98/254}$ ON	NW267	This study
NW368	$pglG_{N.\ gonorrhoeae\ FA1090}$	NW267	This study
NW372	$pglG_{N.\ gonorrhoeae\ SK\text{-}93\text{-}1035}$	NW267	This study
NW412	*pglG rescue*	NW267	This study
NW376	*cycB-His::cat pglK::kan*	KS997	This study
NW 411	NW336 *pglK::kan*	NW336	This study

Continued

Table II. Continued

Strain name	Relevant genotype	Parent	Source
NW388	NW282 *pglK::kan*	NW282	This study
NW392	NW344 *pglK::kan*	NW344	This study
NW396	NW346 *pglK::kan*	NW346	This study
NW400	NW352 *pglK::kan*	NW352	This study
N. mucosa pglG mutants: NirK-His background			
NW413	$pglG_{N.\ mucosa}$ Y112C	NW39	This study
NW416	$pglG_{N.\ mucosa}$ FY111-112IC	NW39	This study
NW419	$pglG_{N.\ mucosa}$ F111I	NW39	This study
NW422	$pglG_{N.\ mucosa}$ Δ164L	NW39	This study
NW425	$pglG_{N.\ mucosa}$ A163T	NW39	This study
NW428	$pglG_{N.\ mucosa}$ Y112C, Δ164L	NW39	This study
Plasmids			
pKP79	pFLOB4300 *ermC'::kan/rpsL+*	–	(Hadjineophytou et al. 2019)
pUC19-*pglK::kan*	pUC19	–	(Wang et al. 2019)
pW55	$pglG_{N.\ meningitidis\ 961–5945}$	pCR4-TOPO	This study
pW63	$pglG_{N.\ meningitidis\ 961–5945}$ ON	pW55	This study
pW66	$pglG_{N.\ lactamica}$	pCR4-TOPO	This study
PW85	$pglG_{N.\ lactamica}$ ON	pW66	This study
pW58	$pglG_{N.\ meningitidis\ NZ98/254}$	pCR4-TOPO	This study
pW71	$pglG_{N.\ meningitidis\ NZ98/254}$ ON	pW58	This study
pW122	$pglG_{N.\ mucosa}$ Y112C	pUC57-kan	This study
pW125	$pglG_{N.\ mucosa}$ FY111-112IC	pUC57-kan	This study
pW129	$pglG_{N.\ mucosa}$ F111I	pUC57-kan	This study
pW131	$pglG_{N.\ mucosa}$ Δ164L	pUC57-kan	This study
pW134	$pglG_{N.\ mucosa}$ A163T	pUC57-kan	This study
pW137	$pglG_{N.\ mucosa}$ Y112C, Δ164L	pUC57-kan	This study

(5ug/mL). Plasmid transformations were performed in *Escherichia coli* TOP10 (Invitrogen) cells and selected on kanamycin (50ug/mL). Details of all strains and plasmids used in this study are found in Table II and Supplementary data, Table SI.

Bioinformatics, genome and phase-variation analyses

Neisseria genome analysis was done by using deposited sequences located in the *Neisseria spp.* Bacterial Isolate Genome Sequence Database (BIGSdb) (Jolley et al. 2018), National Center for Biotechnology Information database servers and Universal Protein Resource database (UniProt) (The UniProt Consortium 2017). Protein family domains were determined using the protein family database (El-Gebali et al. 2019). The isolates selected for the *pglG* phylogenetic analysis was based on similar selection criteria established previously (Hadjineophytou et al. 2019). Thirteen *N. gonorrhoeae* isolates were chosen as representatives of the genomic diversity of the species and 38 *N. meningitidis* isolates, which represent disease-causing isolates from the end of the 20th century (BIGSdb database collection). For the remaining species, all human-colonizing neisserial isolates with complete *pglG* sequences listed in BIGSdb at the time of this study were included in this analysis and are listed in Data Set 1. The phylogenetic tree of *pglG* was based on a MUltiple Sequence Comparison by Log-Expectation alignment and was constructed in MEGA X (Kumar et al. 2018; Stecher et al. 2020) using the Tamura-Nei model (Tamura and Nei 1993). The phylogenetic tree was inferred using 136 *pglG* sequences at 1093 sites and resampled with 500 bootstrap replicates. To assess the distribution of phase-variable *pgl* genes, the satellite DNA search tool Phobos v.3.3.12 (Mayer 2006–2010) was used to detect short tandem repeats (STR) with unit sizes 1–10 in the open reading frames (ORF) of selected *pgl* genes. *Pgl* alleles were designated as phase variable using specific cut-offs for STRs based on similar criteria used in a previous study (Wanford et al. 2018). Specifically, genes were analyzed for the presence of homopolymers ≥9 repeats, dinucleotides or trinucleotides of ≥8 repeats, tetranucleotides or pentanucleotides of ≥5 repeats, and we required ≥3 repeats for STRs with unit sizes between 5 and 9. The strain list of the *pgl* alleles used for this analysis can be found in Data set 2.

Allelic exchange of the *pglG* locus in *N. elongata* subsp. *glycolytica*

We previously detailed the use of a two-gene cassette containing both a selectable (*kanR*) and counter-selectable marker (*rpsL+*) for use in *N. elongata* subsp. *glycolytica* (Hadjineophytou et al. 2019). We used two strains, each carrying a 6xHis C-terminal tag to a known glycoprotein in *N. elongata* subsp. *glycolytica* (NW37 *nirK-His* and KS997 *cycB-His::cat*, Table II) as the parental strains for complementation analyses of *pglG* allelic replacement. To create the counter-selectable *pglG* locus in a *cycB-His* background, the kanR/rpsL+ cassette was PCR amplified from pKP79 with primer pairs nw46/47 and Gibson assembled with approximately 700-bp homologous flanking regions to *pglG* (primer pairs nw44/nw50 and nw49/nw51). The assembled

fragment was transformed into NK2259, selected on kanamycin and counter-selected for streptomycin sensitivity, generating strain NW31. Whole cell lysate DNA from KS997 was used to transform NW31 to introduce the *cycB-His::cat* locus and selecting on chloramphenicol, to generate strain NW267. Transformation of NW37 with the same DNA fragment generated strain NW39.

To replace the *pglG::kan/rpsL+* locus with *pglG* alleles from representative neisserial species, flanking regions to *pglG* were Gibson assembled with the donor *pglG* allele and either cloned into a pCR4-TOPO vector first (Thermo Fisher Scientific), or transformed directly into the strain of interest. Plasmids were sequenced, linearized and transformed into either NW267 or NW39 (Supplementary data, Figure S2, Table II). Deposited *pglG* sequences in BIGSdb with the locus tag NGO0087/NEIS0401 corresponding to each strain were retrieved and used to design the PCR primers. The strategy employed relied on translational fusion of a *pglG* ORF at the endogenous $pglG_{Nelgly}$ initiation and termination codons of the endogenous $pglG_{Nelgly}$ gene such that levels of PglG expression were as equivalent as possible in all strains. Genetic transformation into both NW267 or NW39 was selected on streptomycin and scored for kanamycin sensitivity. Confirmation of the *pglG* alleles from all strains was verified by sequencing of PCR products at Eurofins Genomics (Germany). Details of the primer pairs used are listed in Table SI.

To correct for the phase-varying polyC tract and lock the allele in an "ON" configuration, *N. meningitidis* and *N. lactamica pglG* alleles were PCR mutagenized to truncate the polyC tract (Supplementary data, Figure S2). pW55, pW58 and pW66 are plasmids containing the wildtype *pglG* alleles from *N. meningitidis* 961–5945, NZ98/254 and *N. lactamica* ST-640 strains, respectively, with flanking *Nel pglG* sequences. Using these plasmids as the DNA donor source, overlapping PCR fragments containing the truncation were Gibson assembled and cloned into pCR4-TOPO4 vectors. To correct the *pglG* allele of *N. meningitidis* 961–5945, overlapping PCR products were generated from pW55 using primer pairs nw52/nw159 and nw165/nw53 and assembled to create pW63 (961–5945 $_{pglG\ ON}$). To correct the *pglG* allele *of N. meningitidis* NZ98/254 and *N. lactamica* ST-640, overlapping PCR products were generated from pW58 and pW66 using primer pairs nw52/159 and nw160/53 and assembled to create pW71 (NZ98/254 $_{pglG\ ON}$) and pW85 (ST-640 $_{pglG\ ON}$), respectively. Plasmids were transformed into NW39 or NW267 as detailed above, and the pglG genes from all resulting strains were verified by DNA sequencing as described above.

Inactivation of *pglK*

To inactivate *pglK*, the pUC19-*pglK::kan* plasmid was used to transform relevant complemented strains followed by selection for kanamycin resistance and PCR verification as previously described (Wang et al. 2019).

De novo gene synthesis

Six genes were synthesized by Genecust (Luxembourg) carrying the six mutations of interest to *N. mucosa pglG*. The mutated *N. mucosa pglG* allele is flanked with 589 bp of upstream and 556 bp downstream sequence immediately adjacent to $pglG_{Nelgly}$. The following mutations were introduced: Y112C (TAC → TGC), FY111-112IC (TTCTAC → ATCTGC), F111I (TTC → ATC), Δ164L (ΔCTG), A163T (GCC→ACC) and Y112C, Δ164L (Supplementary data, Figure S5). Synthesized genes were inserted into the pUC57-kan vector, validated by DNA sequencing and transformed into NW39. Plasmids are listed in Table II.

SDS/PAGE, immunoblotting and affinity purification of glycoproteins

Whole-cell lysates from *N. elongata* subsp. *glycolytica* were prepared as previously described (Wang et al. 2019). Information about immunoblotting and purification of His-tagged proteins for mass spectrometric (MS)-based analysis have been described previously (Anonsen et al. 2016). To identify glycoproteins bearing a terminal *N*-acetylglucosamine (GlcNAc), alkaline-phosphatase-conjugated succinylated wheat germ agglutinin (sWGA) lectin (EY Labs, USA) was used as described previously (Zachara et al. 2011; Børud et al. 2014).

LC–MSMS analysis of proteolytically derived peptides

LC–MSMS of chymotrypsin or trypsin-derived peptides from purified glycoproteins was performed as previously described (Anonsen et al. 2016).

Supplementary data

Supplementary data for this article are available online at http://glycob.oxfordjournals.org/.

Funding

The Research Council of Norway (grant 214442) and the Center for Integrative Microbial Evolution at the Department of Biosciences, University of Oslo, and funds from the University of Oslo, Department of Mathematics and Natural Sciences. This publication made use of the *Neisseria* Multi Locus Sequence Typing website (http://pubmlst.org/neisseria/) developed by Keith Jolley and sited at the University of Oxford. The development of this site has been funded by the Wellcome Trust and European Union. This study also made use of the Meningitis Research Foundation Meningococcus Genome Library (http://www.meningitis.org/research/genome) developed by Public Health England, the Wellcome Trust Sanger Institute, and the University of Oxford as a collaboration. That project is funded by the Meningitis Research Foundation.

Conflict of interest statement

None declared

References

Anonsen JH, Vik Å, Børud B, Viburiene R, Aas FE, Kidd SWA, Aspholm M, Koomey M. 2016. Characterization of a unique tetrasaccharide and distinct glycoproteome in the O-linked protein glycosylation system of *Neisseria elongata* subsp. *glycolytica*. *J Bacteriol.* 198(2):256–267.

Bennett JS, Jolley KA, Earle SG, Corton C, Bentley SD, Parkhill J, Maiden MC. 2012. A genomic approach to bacterial taxonomy: An examination and proposed reclassification of species within the genus *Neisseria*. *Microbiology.* 158(Pt 6):1570–1580.

Børud B, Aas FE, Vik A, Winther-Larsen HC, Egge-Jacobsen W, Koomey M. 2010. Genetic, structural, and antigenic analyses of glycan diversity in the *O*-linked protein glycosylation systems of human *Neisseria* species. *J Bacteriol.* 192(11):2816–2829.

Børud B, Anonsen JH, Viburiene R, Cohen EH, Samuelsen AB, Koomey M. 2014. Extended glycan diversity in a bacterial protein glycosylation system linked to allelic polymorphisms and minimal genetic alterations in a glycosyltransferase gene. *Mol Microbiol.* 94(3):688–699.

Børud B, Viburiene R, Hartley MD, Paulsen BS, Egge-Jacobsen W, Imperiali B, Koomey M. 2011. Genetic and molecular analyses reveal an evolutionary trajectory for glycan synthesis in a bacterial protein glycosylation system. *Proc Natl Acad Sci.* 108(23):9643–9648.

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A *et al*. 2019. The pfam protein families database in 2019. *Nucleic Acids Res*. 47(D1):D427–d432.

Hadjineophytou C, Anonsen JH, Wang N, Ma KC, Viburiene R, Vik A, Harrison OB, Maiden MCJ, Grad YH, Koomey M. 2019. Genetic determinants of genus-level glycan diversity in a bacterial protein glycosylation system. *PLoS Genet*. 15(12):e1008532.

Jolley KA, Bray JE, Maiden MCJ. 2018. Open-access bacterial population genomics: Bigsdb software, the pubmlst.Org website and their applications. *Wellcome Open Research*. 3:124.

Kahler CM, Martin LE, Tzeng YL, Miller YK, Sharkey K, Stephens DS, Davies JK. 2001. Polymorphisms in pilin glycosylation locus of *Neisseria meningitidis* expressing class ii pili. *Infect Immun*. 69(6):3597–3604.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. Mega x: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 35(6):1547–1549.

Liu G, Tang CM, Exley RM. 2015. Non-pathogenic *Neisseria*: Members of an abundant, multi-habitat, diverse genus. *Microbiology*. 161(7):1297–1312 Phobos 3.3.11. 2006-2010. [accessed].

Stecher G, Tamura K, Kumar S. 2020. Molecular evolutionary genetics analysis (mega) for macOS. *Mol Biol Evol*. 37(4):1237–1239.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol*. 10(3):512–526.

The UniProt Consortium. 2017. Unipprot: The universal protein knowledgebase. *Nucleic Acids Res*. 45(D1):D158–D169.

Vik A, Aas FE, Anonsen JH, Bilsborough S, Schneider A, Egge-Jacobsen W, Koomey M. 2009. Broad spectrum *O*-linked protein glycosylation in the human pathogen *Neisseria gonorrhoeae*. *Proc Natl Acad Sci U S A*. 106(11):4447–4452.

Wanford JJ, Green LR, Aidley J, Bayliss CD. 2018. Phasome analysis of pathogenic and commensal *Neisseria* species expands the known repertoire of phase variable genes, and highlights common adaptive strategies. *PLoS One*. 13(5):e0196675–e0196675.

Wang N, Anonsen JH, Viburiene R, Lam JS, Vik A, Koomey M. 2019. Disrupted synthesis of a di-*N*-acetylated sugar perturbs mature glycoform structure and microheterogeneity in the *O*-linked protein glycosylation system of *Neisseria elongata* subsp. *glycolytica*. *J Bacteriol*. 201(1):e00522–18.

Zachara NE, Vosseller K, Hart GW. 2011. Detection and analysis of proteins modified by *O*-linked *N*-acetylglucosamine. *Curr Protoc Protein Sci*. Chapter 12: Unit12 18.