Microbial Biology

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

Nelson Wang2,3, Jan Haug Anonsen2,4, Chris Hadjineophytou2,3, William Brynildsen Reinar5, Bente Børud6, Åshild Vik2,7 and Michael Koomey1,2,3,5

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

1To 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-acetylhexosamine 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; Borud et al. 2010; Borud et al. 2011). As defined there, intraspecies 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 tetrascarachide and five previously unrecognized pgl genes (Anøsen 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 tetrascarachide 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 tetrascarachide (Anøsen et al. 2016). Intact alleles of pglG are found in all human neisserial species save for several isolates of 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.

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 divergence. 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
Allelic polymorphisms in a glycosyltransferase gene shape glycan

Fig. 1. The O-linked protein glycosylation pathway in N. elongata spp. glycolytica. The products of genes pgIJ, pgIK, pgIM and pgIN are responsible for the synthesis of UDP-di-N-acetylgalactosamine that is used as a donor by the PglG glycosyltransferase in the extension of the UND-PP-disaccharide. PglG<sub>Nei</sub> has relaxed specificity and minimally requires a UDP-glucuronic acid and accordingly PglJ. Although pglG is present in all human-associated <i>Neisseria</i> species, pglJ is not found in <i>N. gonorrhoeae</i>, <i>N. meningitidis</i> and <i>N. lactamica</i>. 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 <i>N. elongata</i> subsp. <i>glycolytica</i> strain KS944 (aka ATCC 29315-Table II) using a previously detailed allelic replacement method (Hadjineophytou et al. 2019). The deletion of <i>pglG</i> 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<sub>F</sub>-protein bearing C-terminally localized polyhistidine tags. As seen previously (Wang et al. 2019), introduction of the <i>pglG</i> 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 <i>N. elongata</i> subsp. <i>glycolytica</i> (Nelgly) <i>pglG</i> 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 <i>pglG</i> alleles derived from other species yielded immunoblot results similar to the wildtype (Figure S3). Introduction of alleles from <i>N. oralis</i> (<i>Nora</i>) and <i>N. mucosa</i> (<i>Nmuc</i>) yielded immunoblot results similar to the wildtype background suggestive of complementing activities although the low, residual levels of pDAb2 reactivity associated with the <i>pglG<sub>Nmuc</sub></i> gene showed low level, residual incorporation of the diNAcBac-Glc glycan. The pDAb2 reactive-NirK forms seen there migrated faster than those seen in the <i>pglG</i> 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 <i>pglG</i> from <i>N. subflava</i> (<i>Nsu</i>), <i>N. cinerea</i> (<i>Nci</i>) and <i>N. polysaccarea</i> (<i>Npo</i>) strains. These all showed weak signals of pDAb2 reactivity as seen in the <i>pglG<sub>Nmuc</sub></i> strain but with altered patterns of ladder forms to those seen in the wildtype and other complemented backgrounds. Similar testing of <i>pglG</i> alleles

from <i>N. lactamica</i> (<i>Nla</i>), <i>N. meningitidis</i> (<i>Nme</i>) and <i>N. gonorrhoeae</i> (<i>Ngo</i>) 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 <i>N. lactamica</i> <i>pglG</i>) 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 <i>pglG</i> from the <i>N. lactamica</i> strain showed NirK laddering migration patterns distinct from that seen in the <i>pglG</i> 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 <i>N. meningitidis</i> strains although evidence for discrete, distinct NirK forms (seen as laddering) was less obvious. In the cases of the equivalent allele forms from <i>N. gonorrhoeae</i>, no evidence for either altered NirK mobility or pDAb2 antibody reactivity relative to the <i>pglG</i> null background was detected. Results derived for examining the effects of each of the species alleles on C<sub>F</sub> mobility corroborated those made for NirK (Supplementary data, Figure S3).

Mass spectrometric analysis of <i>pglG</i> allele effects on protein glycosylation and glycan composition

We next sought to examine functionality and activities of <i>pglG</i> 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 (*).
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 carrying disaccharides shown in the MS spectrum 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 pgfG null mutant displays only glycopeptides carrying disaccharide, illustrated by the overlapping reporter ions for diNAcBac (m/z 211.1) and diNAcBacHex (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 pgfG 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 pgfG (strain NW193) numerous HexNAC reporter ions were detected (Figure 4J). Hardly any structural ions representing the full-length tetrasaccharide or the diNAcHexA carrying trisaccharides 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<sub>nme</sub>NZ98/254 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 glycan (Figure 4L).

When complementing N. elongata subsp. glycolytica with pgfG 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 pgfG alleles from N. lactamica and N. meningitidis strains indicated the presence of the diNAcBac-Glc-HexNAc terminating trisaccharide. The phase-variable pgfG 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 pgfG alleles tested (strains...
Fig. 4. Effect of cross-complemented pglG on glycan structure. Shown are LC-MSMS chromatograms of the trypsin-derived peptide $^{32}$S-GTGAAAG AASGASGASAPAAPASSAGSSNYPYGGEEGHHHHHHH$^{415}$ 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: 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. 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. Representative MSMS spectrum of the trypsin-generated peptide from NirK-His from pglG$_{null}$ complemented background (strain NW207) with displayed XIC selected for glycan carbohydrate component reporter ions. The TIC intensity value was set at 1E10. (F) Panel G with displayed XIC selected for glycan structure ions. The TIC intensity value was set at 5E9. (G) LC-MSMS chromatogram of trypsin-derived peptides from affinity-purified NirK-His from a pglG$_{null}$ 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 D carrying three diNAcBac-Hex disaccharides. (J) LC-MSMS chromatogram of trypsin-derived peptides from affinity-purified NirK-His from pglG$_{null}$ complemented background (strain NW193) with displayed XIC selected for glycan carbohydrate component reporter ions.

| Carbohydrate Number represented | reporter ion m/z |
|---------------------------------|-----------------|
| 7                               | 594.252         |
| 6                               | 852.323         |
| 5                               | 649.255         |
| 4                               | 204.086         |
| 3                               | 259.091         |
| 2                               | 391.170         |
| 1                               | 211.107         |

Peptide
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\textsubscript{Nora}, pglG\textsubscript{Nmu}, pglG\textsubscript{Nir}, pglG\textsubscript{Nac} and pglG\textsubscript{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\textsubscript{Nelong} 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 pglF and pglK genes, while other species carry 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 C5 mobility patterns and reactivity with pDAb2 antibodies (Figure 5). In particular, levels of disaccharide identical to a pglK mutant in the otherwise wildtype N. elongata subsp. glycolytica background. Remarkably, the opposite result with regard to the glycoprotein pDaB2 reactivity was observed using the alleles from N. mucosa, N. subflava, N. cinerea and N. polysaccarea (Figure 5). In these strains, introducing the pglK mutation led to a loss of pDAb2 reactivity while clearly maintaining significant levels of NirK and C5 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 glycophenotypes 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\textsubscript{Nmu} 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 to the mutation of 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).
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_Nlgy           | diNAcBac-Glc-diNAcHexA-HexNAc                             |
| NW83 pglG_Neva           | diNAcBac-Glc-diNAcHexA-HexNAc                             |
| NW88 pglG_Nlmac          | diNAcBac-Glc-diNAcHexA-HexNAc                             |
| NW95 pglG_Nlub           | diNAcBac-Glc-diNAcHexA                                    |
| NW81 pglG_Nicin          | diNAcBac-Glc-diNAcHexA-HexNAc                             |
| NW107 pglG_Npo           | diNAcBac-Glc-diNAcHexA-HexNAc                             |
| NW160 pglG_Nla           | diNAcBac-Glc-diNAcHexA-HexNAc                             |
| NW207 pglG_Nla ON        | diNAcBac-Glc-diNAcHexA-HexNAc                             |
| NW142 pglG_Nme NZ98-5945 | diNAcBac-Glc                                             |
| NW165 pglG_Nme NZ98-5945 ON | diNAcBac-Glc-diNAcHexA-HexNAc                           |
| NW150 pglG_Nme NZ98254    | diNAcBac-Glc-diNAcHexA-HexNAc                             |
| NW193 pglG_Nme NZ98254 ON | diNAcBac-Glc-diNAcHexA-HexNAc                             |
| NW47 pglG Ngo FA1090      | diNAcBac-Glc                                             |
| NW70 pglG Ngo SK-93-1035 | diNAcBac-Glc                                             |
### 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 orthologues 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 subsp. 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

| kDa | WT | pglG | pglK | pglG+pglK | pglG+Nora | pglG+NirK | pglG+NirK+c5 | pglG+Nora | pglG+Nora+c5 | pglG+Nora+pglK | pglG+Nora+pglK+c5 | pglG+Nora+pglK+pglJ | pglG+Nora+pglK+pglJ+c5 | pglG+Nora+pglK+pglJ+pglM | pglG+Nora+pglK+pglJ+pglM+c5 |
|-----|----|------|------|----------|----------|----------|-------------|----------|-------------|---------------|----------------|-------------------|-------------------|-------------------|---------------------|
| 50  |    |      |      |          |          |          |             |          |             |               |                 |                   |                   |                   |                     |
| 37  |    |      |      |          |          |          |             |          |             |               |                 |                   |                   |                   |                     |
| 25  |    |      |      |          |          |          |             |          |             |               |                 |                   |                   |                   |                     |

- Allelic polymorphisms in a glycosyltransferase gene shape glycan.
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<sub>Muc</sub> (NW88), *pglG<sub>Muc</sub>* Y112C, ΔL164 (NW428), *pglG<sub>Muc</sub>* Y112C (NW413), *pglG<sub>Muc</sub>* FY (111–112)IC (NW419), *pglG<sub>Muc</sub>* FY111C (NW419), *pglG<sub>Muc</sub>* ΔL164 (NW422), *pglG<sub>Muc</sub>* 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<sub>Muc</sub>* gene as it was possible that other missense-type substitutions might be present in *pglG<sub>Ngo</sub>*. Interestingly, only introduction of both substitutions into wildtype PglG<sub>Muc</sub> resulted
Allelic polymorphisms in a glycosyltransferase gene shape glycan

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 PglGNmuc 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 pglGNgo 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 KS944 | nirK-His::cat Smr | – | (Hadjineophytou et al. 2019) |
| NW39 | nirK-His pglG::kan/rpsL+ | NW39 | This study |
| RV731 | nirK-His pglK::kan | NW39 | (Wang et al. 2019) |
| NW47 | pglGN. gonorrhoeae FA1090 | NW39 | This study |
| NW70 | pglGN. gonorrhoeae SK-93-1035 | NW39 | This study |
| NW77 | pglGN. elongata | NW39 | This study |
| NW83 | pglGN. oralis | NW39 | This study |
| NW88 | pglGN. mucosa | NW39 | This study |
| NW95 | pglGN. subflava | NW39 | This study |
| NW81 | pglGN. cinerea | NW39 | This study |
| NW112 | pglGN. polysaccharea | NW39 | This study |
| NW160 | pglGN. lactamica | NW39 | This study |
| NW207 | pglGN. lactamica ON | NW39 | This study |
| NW142 | pglGN. meningitidis 961–5945 | NW39 | This study |
| NW165 | pglGN. meningitidis 961–5945 ON | NW39 | This study |
| NW150 | pglGN. meningitidis NZ98/254 | NW39 | This study |
| NW193 | pglGN. 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 c3-His background KS997 | cycB-His::cat | – | (Anonsen et al. 2016) |
| NK2239 | Smr | KS944 | (Hadjineophytou et al. 2019) |
| NW31 | pglG::kan/rpsL+ | NK2259 | This study |
| NW267 | cycB-His::cat | NW31 | This study |
| NW336 | pglG::kan/rpsL+ | NW267 | This study |
| NW282 | pglGN. oralis | NW267 | This study |
| NW344 | pglGN. mucosa | NW267 | This study |
| NW348 | pglGN. subflava | NW267 | This study |
| NW352 | pglGN. cinerea | NW267 | This study |
| NW300 | pglGN. polysaccharea | NW267 | This study |
| NW356 | pglGN. lactamica | NW267 | This study |
| NW308 | pglGN. lactamica PV ON | NW267 | This study |
| NW360 | pglGN. meningitidis 961–5945 ON | NW267 | This study |
| NW316 | pglGN. meningitidis 961–5945 ON | NW267 | This study |
| NW364 | pglGN. meningitidis NZ98/254 | NW267 | This study |
| NW368 | pglGN. meningitidis NZ98/254 ON | NW267 | This study |
| NW372 | pglGN. gonorrhoeae FA1090 | NW267 | This study |
| NW412 | pglG rescue | NW267 | This study |
| NW376 | cycB-His::cat pglK::kan | KS997 | This study |
| NW411 | NW336 pglK::kan | NW336 | This study |

Continued
 Allelic polymorphisms in a glycosyltransferase gene shape glycan

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       | pglGN. mucosa Y112C | NW39   | This study |
| NW416       | pglGN. mucosa FY111-112IC | NW39   | This study |
| NW419       | pglGN. mucosa F111I | NW39   | This study |
| NW422       | pglGN. mucosa Δ164L | NW39   | This study |
| NW425       | pglGN. mucosa A163T | NW39   | This study |
| NW428       | pglGN. mucosa Y112C, Δ164L | NW39   | This study |

Plasmids

pKP79

| pUC19-pglK::kan |
|----------------|
| ermC::kan/ermC::kan |
| pUC19 |
| pW55 |
| pglGN. meningitidis 961-5945 |
| pCR4-TOPO |
| pW63 |
| pglGN. meningitidis 961-5945 ON |
| pCR4-TOPO |
| pW66 |
| pglGN. lactamica |
| pW66 |
| PW85 |
| pglGN. lactamica ON |
| pW66 |
| pW58 |
| pglGN. meningitidis NZ98/234 |
| pCR4-TOPO |
| pW71 |
| pglGN. meningitidis NZ98/234 ON |
| pW58 |
| pW122 |
| pglGN. mucosa Y112C |
| pUC57-kan |
| pW125 |
| pglGN. mucosa F111I |
| pUC57-kan |
| pW129 |
| pglGN. mucosa Δ164L |
| pUC57-kan |
| pW131 |
| pglGN. mucosa A163T |
| pUC57-kan |
| pW134 |
| pglGN. mucosa Y112C, Δ164L |
| pUC57-kan |

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-Ge-database (UniProt) (The UniProt Consortium 2017). Protein family domains were determined using the protein family database (El-Ge-

databases and Universal Protein Resource database (UniProt) (The UniProt Consortium 2017). Protein family domains were determined using the protein family database (El-Ge-

databases and Universal Protein Resource database (UniProt) (The UniProt Consortium 2017). Protein family domains were determined using the protein family database (El-Ge-

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 glyco-protein 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 K5997 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 pgIG::kan::pLS+ locus with pgIG alleles from representative neisserial species, flanking regions to pgIG were Gibson assembled with the donor pgIG 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 pgIG 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 pgIG ORF at the endogenous pgIGNelgy initiation and termination codons of the endogenous pgIGNelgy. Gene such that levels of pgIG 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 pgIG 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 S1.

To correct for the phase-varying polyC tract and lock the allele in an “ON” configuration, N. meningitidis and N. lactamica pgIG alleles were PCR mutagenized to truncate the polyC tract (Supplementary data, Figure S2). pW55, pW58 and pW66 are plasmids containing the wildtype pgIG alleles from N. meningitidis 961–5945, N. lactamica ST-640 and N. lactamica ST-640 strains, respectively, with flanking Nel pgIG 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 pgIG allele of N. meningitidis 961–5945, overlapping PCR products were generated from pW55 using primer pairs nw165/nw159 and nw165/nw53 and assembled to create pW63 (961–5945 pgIG ON). To correct the pgIG allele of N. meningitidis NZ98/254 and N. lactamica ST-640, overlapping PCR products were generated from pW58 and pW66 using primer pairs nw2159 and nw160/53 and assembled to create pW71 (NZ98/254 pgIG ON) and pW85 (ST-640 pgIG ON), respectively. Plasmids were transformed into NW39 or NW267 as detailed above, and the pgIG genes from all resulting strains were verified by DNA sequencing as described above.

Inactivation of pgIK

To inactivate pgIK, the pUC19-pgIK::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 pgIG. The mutated N. mucosa pgIG allele is flanked with 589 bp of upstream and 556 bp downstream sequence immediately adjacent to pgIGNelgy. The following mutations were introduced: Y112C (TAC→TGC), FY111-112IC (TTCTAC→ATCTGC), F111H (TTCT→ATC), ΔA164L (ΔCTG), A163T (GCC→ACC) and Y112C, ΔA164L (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-acetylgalactosamine (GalNAc), 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–MS/MS analysis of proteolytically derived peptides

LC–MS/MS 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.
Allelic polymorphisms in a glycosyltransferase gene shape glycan

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 Res. 3:124.

Kahler CM, Martin LE, Tseng YL, Miller YK, Sharkey K, Davies DS. 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. Uniprot: 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, Ashley 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.

Zachara NE, Vosseller K, Hart GW. 2011. Detection and analysis of proteins modified by O-linked N-acetylgalactosamine. Curr Protoc Protein Sci. Chapter 12: Unit 12.18.