

## Supplemental Material

### Genotypic and phenotypic characterization of the *O*-linked protein glycosylation system reveals high glycan diversity in paired meningococcal carriage isolates

Running title: Glycan diversity in meningococcal carriage isolates

Bente Børud<sup>1, #</sup>, Guro K. Bårnes<sup>1,2</sup>, Ola Brønstad Brynildsrud<sup>1</sup>, Elisabeth Fritzsønn<sup>1</sup>, Dominique A. Caugant<sup>1,2</sup>

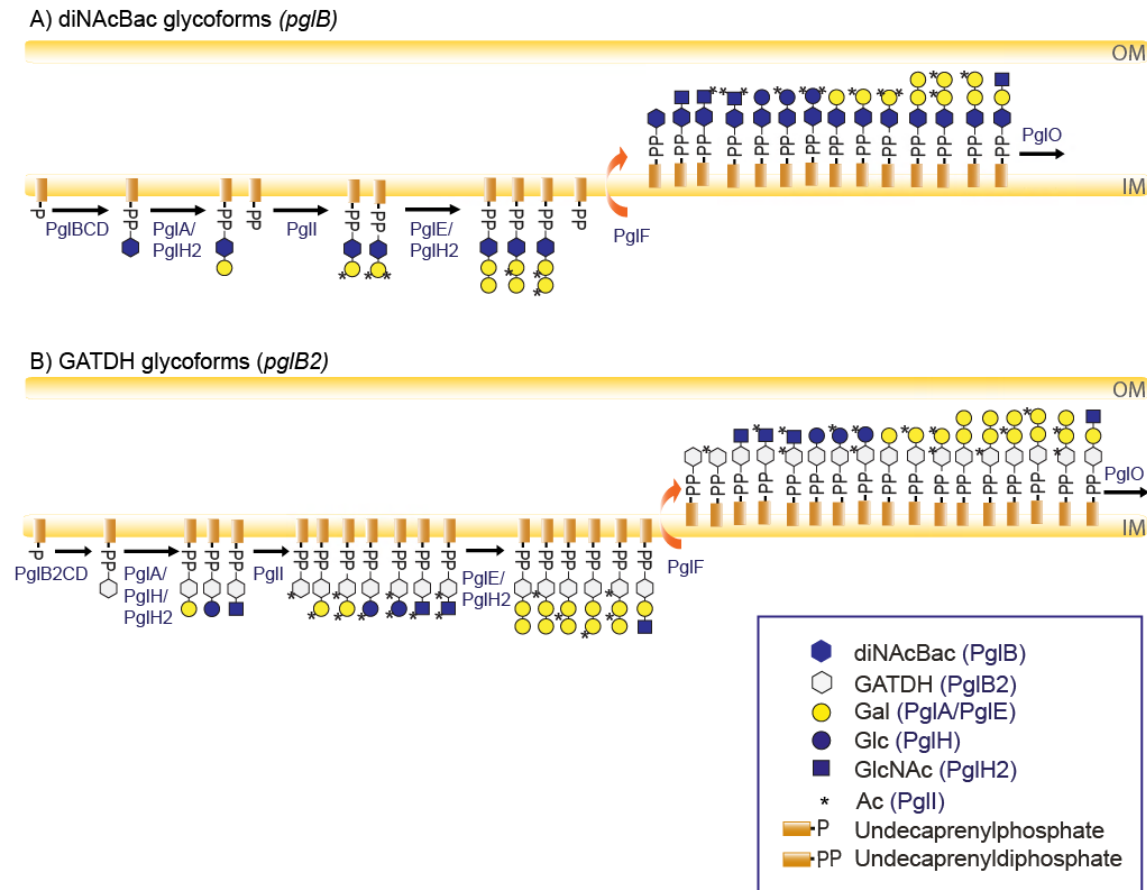
<sup>1</sup> Division for Infection Control and Environmental Health, Norwegian Institute of Public Health, Oslo, Norway

<sup>2</sup> Faculty of Medicine, University of Oslo, Norway

#Corresponding author: Bente Børud <bente.borud@fhi.no>

#### Keywords

*Neisseria meningitidis*, carriage, whole genome sequencing, *O*-linked protein glycosylation, glycan diversity, microheterogeneity.



**Figure S1. Overview of the *O*-linked protein glycosylation pathway in *Neisseria***

Current model of the broad spectrum *O*-linked glycosylation pathway expressed by species within the genus *Neisseria*. OM, outer membrane; IM, inner membrane.

**pg1A**

**ST-53** GCTTTGGGCGCAATCGACAAATCACGCGCGGGGGGGGGGGGGG-----CGATTGGAACGCCCTTGCCGCCCGCGATATT  
**ST-35** GCTTTGGGCGCAATCGACAAATCACGC--GGGGGGAAGGG-----CGATTGGAACGGCTTGCCGCCCGCGATATT  
**ST-198** GCTTTGGGCGCAATCGACAAATCACGC--GGGGGGGGGGGGGG-----CGATTGGAACGCCCTTATCGCCCGCGATATT  
**ST-192** GCTTTGGGCGCAATCGACAAATCACGC--GGGGGGGGGGGG-----CGATTGGAACGGCTTGCCGCCCGCGATATT  
**ST-11597** GCTTTGGGCGCAATCGACAAATCACGC--GGGGGGGGGGGGGG-----ATTTGGAACGCCCTTGCCGCCCGCGATATT  
**ST-11595** GCTTTGGGCGCAATCGACAAATCACGC--GGGGGGNNNNNNNN-----CGATTGGAACGCCCTTATCGCCCGCGATATT  
**ST-175** GCTTTGGGCGCAATCGACAAATCACGC--GGGGGGGGGGNNNNN-----CGATTGGAACGGCTTGCCGCCCGCGATATT

**ST-11**

11B GCTTTGGGCGCAATCGACAAATCACGCGCGGGGGGGGGGGGGG-----CGATTGGAACGCCCTTGCCGCCCGCGATATT  
4A GCTTTGGGCGCAATCGACAAATCACGCGCGGGGGGGGGGGGGG-----CGATTGGAACGCCCTTGCCGCCCGCGATATT  
4B GCTTTGGGCGCAATCGACAAATCACGCGCGGGGGGGGGGGGGG-----CGATTGGAACGCCCTTGCCGCCCGCGATATT  
37A GCTTTGGGCGCAATCGACAAATCACGCGCGGGGGGGGGGGGGG-----CGATTGGAACGCCCTTGCCGCCCGCGATATT  
37B GCTTTGGGCGCAATCGACAAATCACGCGCGGGGGGGGGGGGGGGGGGGGG-----CGATTGGAACGCCCTTGCCGCCCGCGATATT  
12A GCTTTGGGCGCAATCGACAAATCACGCGCGGGGGGGGGGGGG-----CGATTGGAACGCCCTTGCCGCCCGCGATATT  
12B GCTTTGGGCGCAATCGACAAATCACGCGCGGGGGGGGGGGGG-----CGATTGGAACGCCCTTGCCGCCCGCGATATT  
16A GCTTTGGGCGCAATCGACAAATCACGCGCGGGGGGGGGGGGG-----CGATTGGAACGCCCTTGCCGCCCGCGATATT  
16B GCTTTGGGCGCAATCG-----GGGGGGGGGGGGGGGGGGGGGGGGGGGGCGATTGGAACGCCCTTGCCGCCCGCGATATT

**ST-2880**

2A GCTTTGGGCGCAATCGACAAATCACGC--GGGGGGGGGGGG-----CGATTGGAACGGCTTGCCGCCCGCGATATT  
2B GCTTTGGGCGCAATCGACAAATCACGC--GGGGGGGGGGGGGG-----CGATTGGAACGGCTTGCCGCCCGCGATATT  
5A GCTTTGGGCGCAATCGACAAATCACGC--GGGGGGGGGGGGGG-----CGATTGGAACGGCTTGCCGCCCGCGATATT  
5B GCTTTGGGCGCAATCGACAAATCACGC--GGGGGGGGGGGGGG-----CGATTGGAACGGCTTGCCGCCCGCGATATT  
10A GCTTTGGGCGCAATCGACAAATCACGC--GGGGGGGGGGGG-----CGATTGGAACGGCTTGCCGCCCGCGATATT  
10B GCTTTGGGCGCAATCGACAAATCACGC--GGGGGGGGGGGGGG-----CGATTGGAACGGCTTGCCGCCCGCGATATT  
44A GCTTTGGGCGCAATCGACAAATCACGC--GGGGGGGGGGGGGG-----CGATTGGAACGGCTTGCCGCCCGCGATATT  
44B GCTTTGGGCGCAATCGACAAATCAC-----GGGGGGGGGGGGGG-----CGATTGGAACGGCTTGCCGCCCGCGATATT  
43A GCTTTGGGCGCAATCGACAAATCACGC--GGGGGGGGGGGGGG-----CGATTGGAACGGCTTGCCGCCCGCGATATT  
43B GCTTTGGGCGCAATCGACAAATCACGC--GGGGGGGGGGGGGG-----CGATTGGAACGGCTTGCCGCCCGCGATATT  
48B GCTTTGGGCGCAATCGACAAATCACGC--GGGGGGGGGGGGGG-----CGATTGGAACGGCTTGCCGCCCGCGATATT  
15A GCTTTGGGCGCAATCGACAAATCACGC--GGGGGGGGGGGGGG-----CGATTGGAACGGCTTGCCGCCCGCGATATT  
15B GCTTTGGGCGCAATCGACAAATCAC-----GGGGGGGGGGGGGG-----CGATTGGAACGGCTTGCCGCCCGCGATATT  
17A GCTTTGGGCGCAATCGACAAATCACGC--GGGGGGGGGGGGGG-----CGATTGGAACGGCTTGCCGCCCGCGATATT  
17B GCTTTGGGCGCAATCGACAAATCACGC--GGGGGGGGGGGG-----CGATTGGAACGGCTTGCCGCCCGCGATATT

**ST-11372**

26A GCTTTGGGCGCAATCGACAAATCACGCGCGGGGGGGGGGGGG-----ATTTGGAACGCCCTTGCCGCCCGCGATATT  
26B GCTTTGGGCGCAATCGACAAATCACGCGCGGGGGGGGGGGGG-----CGATTGGAACGCCCTTGCCGCCCGCGATATT

**Figure S2. Alignment of the polyG tract and flanking sequences in *pglA*.** The figure shows an alignment of the polyG tract, shown as blue nucleotide sequence when equal in all isolates within an ST, and the extended guanines (Gs) are shown in grey. When there are additional differences within the ST (ST-11, ST-2880 and ST-11372), all isolates are shown. Other nucleotide variants and deletions are shown in red. N represent unknown nucleotide.

*pglI*

```
ST-192      GATGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGG-----AGGTTTCCACTGAAGACATTTTATATCCGCCGCATTAAGCGG
ST-198      GATGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGG-----AGGTTTTCCCTGAAGGCATTTTATATCCGCCGCATCAAGCGG
ST-11597    GATGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGGGG----AGGTTTTCCCTGAAGGCATTTTATATCCGCCGCATCAAGCGG
ST-35       GATGACGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGG-----AGGTTTTCCCTGAAGGCATTTTATATCCGCCGCATCAAGCGG
ST-175      GATGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGG-----AGGTTTTCCCTGAAGGCATTTTATATCCGCCGCATCAAACGG
ST-11372    GATGACGAGATCCTCCTTCGCGAAATGTCGGGGGGGGGGGGG-----TTCCCTGAAGGCATTTTATATCCGCCGCATCAAACGG
ST-11595    GATGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGG-----AGGTTTTCCCTGAAGGCATTTTATATCCGCCGCATCAAGCGG
ST-11       GATGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGGGG---AGGTTTTCCCTGAAGGCATTTTATATCCGCCGCATCAAACGG

ST-2880
2A          GATGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGG-----AGGTTTTCCCTGAAGGCATTTTATATCCGCCGCATCAAGCGG
2B          GATGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGGGGGGGG---TTTTCCCTGAAGGCATTTTATATCCGCCGCATCAAGCGG
5A          GATGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGGGG----AGGTTTTCCCTGAAGGCATTTTATATCCGCCGCATCAAGCGG
5B          GATGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGGGGGG---TTTTCCCTGAAGGCATTTTATATCCGCCGCATCAAGCGG
10A         GATGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGGGG----AGGTTTTCCCTGAAGGCATTTTATATCCGCCGCATCAAGCGG
10B         GATGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGGGGGG---TTTTCCCTGAAGGCATTTTATATCCGCCGCATCAAGCGG
44A         GATGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGGGG---AGGTTTTCCCTGAAGGCATTTTATATCCGCCGCATCAAGCGG
43A         GATGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGGGG---AGGTTTTCCCTGAAGGCATTTTATATCCGCCGCATCAAGCGG
43B         GATGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGGGG----AGGTTTTCCCTGAAGGCATTTTATATCCGCCGCATCAAGCGG
48B         GATGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGGGG----AGGTTTTCCCTGAAGGCATTTTATATCCGCCGCATCAAGCGG
15A         GATGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGGGGGG---TTTTCCCTGAAGGCATTTTATATCCGCCGCATCAAGCGG
15B         GATGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGGGGGG---TTTTCCCTGAAGGCATTTTATATCCGCCGCATCAAGCGG
17A         GATGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGGGG---TTTTCCCTGAAGGCATTTTATATCCGCCGCATCAAGCGG
17B         GATGACGATGATTCTTCACCGCGAAATGTCGGGGGGGGGGGGGGG---TTTTCCCTGAAGGCATTTTATATCCGCCGCATCAAGCGG
```

**Figure S3. Alignment of the polyG tract and flanking sequences in *pglI*.** The figure shows an alignment of the polyG tract, shown as blue nucleotide sequence when equal in all isolates within an ST, and the extending guanines (Gs) are shown in grey. When there are additional differences within the ST (ST-2880), all isolates are shown. Other nucleotide variants and deletions are shown in red.

***pglH***

ST-53 ATGAACATCACCATAGTCGCCCCCTACTGCTCGCTGCCGTCCGAACCGTATTTCAACCGCTTTTGGTATCTGGCAGAAC**TG**TTGTCGCAATCGCACGACG 100  
ST-53 (*pglH2*) ATGAACATCACCATAGTCGCCCCCTACTGCTCGCTGCCGTCCGAACCGTATTTCAACCGCTTTTGGTATCTGGCAGAACGATTGTCGCAATCGCACGACG  
ST-2880 ATGAACATCACCATAGTCGCCCCCTACTGCTCGCTGCCGTCCGAACCGTATTTCAACCGCTTTTGGTATCTGGCAGAACGATTGTCGCAATCGCACGACG  
ST-11 ATGAACATCACCATAGTCGCCCCCTACTGCTCGCTGCCGTCCGAACCGTATTTCAACCGCTTTTGGTATCTGGCAGAAC**TG**TTGTCGCAATCGCACGACG  
ST-192 ATGAACATCACCATAGTCGCCCCCTACTGCTCGCTGCCGTCCGAACCGTATTTCAACCGCTTTTGGTATCTGGCAGAACGATTGTCGCAATCGCACGACG  
ST-11595 ATGAACATCACCATAGTCGCCCC**TTA**TTGCTCGCTGCCGTCCGAACCGTATTTCAACCGCTTTTGGTATCTGGCAGAACGATTGTCGCAATCGCACGACG  
ST-35 ATGAACATCACCATAGTCGCCCC**TTA**TTGCTCGCTGCCGTCCGAACCGTATTTCAACCGCTTTTGGTATCTGGCAGAACGATTGTCGCAATCGCACGACG  
ST-11597 ATGAACATCACCATAGTCGCCCC**TTA**TTGCTCGCTGCCGTCCGAACCGTATTTCAACCGCTTTTGGTATCTGGCAGAACGATTGTCGCAATCGCACGACG

ST-53 TGTTGCTGATTACCAGCCGTTTCCGCCACTACGACAAATCCTTCAGACGGCATGAAGATGCTGCCGCCACTTCAAACGGCAGGCTGCGCGTCAAGCTGCT 200  
ST-53 (*pglH2*) TGTTGCTGATTACCAGCCGTTTCCGCCACTACGACAAATCCTTCAGACGGCATGAAGATGCTGCCGCCACTTCAAACGGCAG-----  
ST-2880 TGTTGCTGATTACCAGCCGTTTCCGCCACTACGACAAATCCTTCAGACGGCATGAAGATGCTGCCGCCACTTCAAACGGCAGGCTG**CA**CGTCAAGCTGCT  
ST-11 TGTTGCTGATTACCAGCCGTTTCCGCCACTACGACAAATCCTTCAGACGGC**CC**CGAAGATGCTGCCGCC**CC**CTG**CA**AGG**CC**CTG**CA**AA**AG**T**CA**TGCTG**TT**  
ST-192 TGTTGCTGATTACCAGCCGTTTCCGCCACTACGACAAATCCTTCAGACGGCATGAAGATGCTGCCGCCACTTCAAACGGCAGGCTGCGCGTCAAGCTGCT  
ST-11595 TGTTGCTGATTACCAGCCGTTTCCGCCACTACGACAAATCCTTCAGACGGCATGAAGATGCTGCCGCCACTTCAAACGGCAGGCTG**CA**CGTCAAGCTGCT  
ST-35 TGTTGCTGATTACCAGCCGTTTCCGCCACTACGACAAAT**CA**TTTCAGACGGCATGAAGATGCTGCCGCCACTTCAAACGGCAGGCTGCGCGTCAAGCTGCT  
ST-11597 TGTTGCTGATTACCAGCCGTTTCCGCCACTACGACAAATCCTTCAGACGGCATGAAGATGCTGCCGCCACTTCAAACGGCAGGCTG**CA**CGTCAAGCTGCT

ST-53 GGACGAACCCGGCTACCGAAAAACGCTCCCTCGCCCGCTGCCAG**T**CACCGCGTATTCGTCGCCAATTTGGCAGCTGGCTGCACAGCCCGCAGGCG 300  
ST-53 (*pglH2*) -----GCTGCACAGCCCGCAGGCG  
ST-2880 GGACGAACCCGGCTACCGAAAAACGCTCCCTCGCCCGCTGCCAGCCACCGCGTATTCGTCGCCAATTTGGCAGCTGGCTGCACAGCCCGCAGGCG  
ST-11 GGAG**AA**AGCGGCTAC**GG**CAAAAACGTGCT**TT**TAG**AA**CG**TG**TTGCCAGCCA**TT**ACC**GC**TT**CG**T**CA**AAAATTT**T**AG**CA**ATGGCTG**AAA**AC**TG**CC**GG**CCG  
ST-192 GGATGAACCCGGCTACCGAAAAACGCTCCCTCGCCCGCTGCCAGCCACCGCGTATTCGTCGCCAATTTGGCAGCTGGCTGCACAGCCCGCAGGCG  
ST-11595 GGACGAACCCGGCTACCGAAAAACGCTCCCTCGCCCGCTGCCAGCCACCGCGTATTCGTCGCCAATTTGGCAGCTGGCTGCACAGCCCGCAGGCG  
ST-35 GGACGAAC**CC**GGCTACCGAAAAACGCTCCCTCGCCCGCTGCCAGCCACCGCGTATTCGTCGCCAATTTGGCAGCTGGCTGCACAGCCCGCAGGCG  
ST-11597 GGACGAACCCGGCTACCGAAAAACGCTCCCTCGCCCGCTGCCAGCCACCGCGTATTCGTCGCCAATTTGGCAGCTGGCTGCACAGCCCGCAGGCG

ST-53 GCAGAACAGGACATCGTCTATTCCGCCATCCGCTGATGGCGACCAACCTGCT**TTT**AGGCAGGCACAAAGCGCGTTTGGGTACAAACT**CAT**CATCGACG 400  
ST-53 (*pglH2*) GCGGAACAGGACATCGTCTATT**TG**CCATCCGCTGAT**TG**CCACCAACCTGCTGTTGGGCAACACAAAGCGCGTTTGGGT**T**TACAAACTGATTGTCGATG  
ST-2880 GCGGAACA**AG**ATATCGTCTATTCCGCCATCCGCTGATGGC**AA**CCAACCTGCTGTTGGGCAAAACACAAAGCGCGTTTGGGTACAA**AT**TGATGTCGATG  
ST-11 GCGGAACA**AG**ATATCGTCTATTCCGCCATCCGCTGAT**TG**CCACCAACCTGCTGTTGGGCAAAACACAAAGCGCGTTTGGGTACAAACTGATTGTCGATG  
ST-192 GCGGAACA**AG**ATATCGTCTATTCCGCCATCCGCTGATGGC**AA**CCAACCTGCTGTTGGGCAAAACACAAAGCGCGTTTGGGTACAA**AT**TGATGTCGATG  
ST-11595 GCGGAACAGGACATCGTCTATTCCGCCATCC**ACT**GATGGCGACCAACCTGCT**TTT**AGGCAGGCAT**AA**AGCGCGTTTGGGTACAAACTGATT**AT**CGACG  
ST-35 GCGGAACAGGACATCGTCTATT**TG**CCATCCGCTGATGGCGACCAACCTGCTGTTGGGCAAAACACAAAGCGCGTTTGGGTACAAACTGATTGTCGATG  
ST-11597 GCGGAACAGGACATCGTCTATTCCGCCATCC**ACT**GATGGCGACCAACCTGCT**TTT**AGGCAGGCAT**AA**AGCGCGTTTGGGTACAAACTGATT**AT**CGACG

ST-53 TGCAGGACGTGTGGCCGGAGTCTTCTCATCGGTGCTGCCGTTTTTGAATAAATCCCCACA**CC**TGCTGCCCTTTGCT**GAC**CGCGCC**GAC**CGCGCCGA 500  
ST-53 (*pglH2*) TGCAGGACGTATGGCCGGAGTCTTCTCATCGGTGCTACCGTTTTTGAATAAAGTACCGCATAAGCTGCTGCCCTTTGCT**C**TACGCGCCAAC**CA**AGCC--  
ST-2880 TGCAGGACGTATGGCCGGAGTCTTCTCCTCGGTGCTACCGTTTTTGAATAAAGTACCGCACAAGCTGCTGCCCTTTGCTT**C**ACGCGCCAACCGTGCC--  
ST-11 TACAGGACGTATGGCCGGAGTCTTCTCCTCGGTGCTACCGTTTTTGAATAAAGTACCGCACAAGCTGCTGCCCTTTGCTT**C**ACGCGCCAACCGTGCC--  
ST-192 TGCAGGACGTATGGCCGGAGTCTTCTCCTCGGTGCTACCGTTTTTGAATAAAGTACCGCACAAGCTGCTGCCCTTTGCTT**C**ACGCGCCAACCGTGCC--  
ST-11595 TACAGGACGTATGGCCGGAACTCTCTCCTCGGT**CA**TACCGTTTTTGAATAAATCCCGC**CC**CTGCTGCCCTTTGCT**CC**GT**CG**CGCC**GAC**CGTGCC--  
ST-35 TGCAGGACGTATGGCCGGAGTCTTCTCCTCGGTGCTACCATTTTTGAATAAAGTACCGCACAAGCTGCTGCCCTTTGCTT**C**ACGCGCCAACCGTGCC--  
ST-11597 TACAGGACGTATGGCCGGAACTCTCTCCTCGGT**CA**TACCGTTTTTGAATAAATCCCGC**CC**CTGCTGCCCTTTGCT**CC**GT**CG**CGCC**GAC**CGTGCC--

ST-53 CCGTTCCTACCGTTGCGCCGACGCACCTCGTTGCCGTCTCGCAAACCTACCTCGACCGTGCCAAAGAAGCCAATCCCGACGTCCCTGCCGAAGCCGTCTGT 600  
ST-53 (pglH2) -----TACCGCTACGCCGACGCGCTGGTTGCCGTATCGCAGACTTATCTCGACCGGCCAAAGAAACCAATCCGAACGTACCCGGCGAAGTCACTAT  
ST-2880 -----TACCGCTGCGCCGACGCGCTGATTGCCGTATCGCAGACTTATCTCGACCGGCCAAAGAAGCCAATCCGAACGTACCCGGCGAACCCTGTAT  
ST-11 -----TACCGCTGCGCCGACGCGCTGGTTGCCGTATCGCAGACTTATCTCGACCGGCCAAAGAAGCCAATCCGAACGTACCCGGCGAAGCCGTCTAT  
ST-192 -----TACCGCTGCGCCGACGCGCTGATTGCCGTATCGCAGACTTATCTCGACCGGCCAAAGAAGCCAATCCGAACGTACCCGGCGAACCCTGTAT  
ST-11595 -----TACCGTTGCGCCGACGCACCTCGTTGCCGTCTCGCAAACCTACCTCGACCGTGCCAAAGAAGCCAATCCCGATGTTCCCTGCCGAAGCCGTCTGT  
ST-35 -----TACCGCTGCGCCGACGCGCTGGTTGCCGTATCGCAGACTTATCTCGACCGGCCAAAGAAGCCAATCCGAACGTACCCGGCGAAGCCGTCTAT  
ST-11597 -----TACCGTTGCGCCGACGCACCTCGTTGCCGTCTCGCAAACCTACCTCGACCGTGCCAAAGAAGCCAATCCCGATGTTCCCTGCCGAAGCCGTCTGT

ST-53 ATCGGCGCGGATTTCCGCCCATAGCATCCGCGCCCGCCAAAGATTTCAGCGACAGCACAAACCCGATTTCTTCTATTTCCGGCAGCTCAGCCACAGCTACG 700  
ST-53 (pglH2) ATTTGGTGTGACTTTCCCAAACTCGATGCGCCACCTGCCAAAGATTTCAGCGACGACAAACCCGCTTTTTCTACTTGGGCACGCTCAGTTACAGCTATG  
ST-2880 ATCGGTACGGATTTTGGCCCATCGCCCCCCCCCCCC--GCGATTCC-GCTCCAAAACCGTCCGCCTTTTTCTACTTGGGCACGCTCAGTTACAGCTACG  
ST-11 ATCGGTACGGATTTTGGCCCATCGCCCCCCCCCCCC--GCGATTCC-GCTCCAAAACCGTCCGCCTTTTTCTACTTGGGCACGCTCAGTTACAGCTATG  
ST-192 ATCGGTACGGATTTTGGCCCATCGCCCCCCCCCCCCCGGCGATTCC-GCTCCAAAACCGTCCGCCTTTTTCTACTTGGGCACGCTCAGTTACAGCTACG  
ST-11595 ATCGGTACGGATTTTGGCCCATCGCCCCCCCCCCCC--CGGATTCC-GCTCCAAAACCGTCCGCCTTTTTCTACTTGGGCACGCTCAGTTACAGCTACG  
ST-35 ATCGGTACGGATTTTGGCCCATCGCCCCCCCCCCCC--GCGATTCC-GCTCCAAAACCGTCCGCCTTTTTCTACTTGGGCACGCTCAGTTACAGCTACG  
ST-11597 ATCGGCGCGGATTTCCGCCCATAGCATCCGCAACCGCCAAAGATTTCAGCGACAGCACAAACCCGATTTTCTATTTCCGGCAGCTCAGCCACAGCTACG

ST-53 ATGTGCGTACTGCGTGCAGGGGGATACAAAACTTCTGGACAGCGGGAAAAACGTCGAACTGCACATCATGGGCGGCGGCCCGGACCTAGAAAAGCTGAA 800  
ST-53 (pglH2) ACGTGGAAACCGTGTGCAAAGCGTTTCGGAACTTTTGGACGACGGCGAAAAATGTAGAGCTGCACATCATGGGCGGCGGCCCGGATTTGGACAGGCTCAA  
ST-2880 ACGTGGAAACCGTGTGCAAAGCGTTTCGGAACTTTTAGACGACGGCGAAAAACGTCGAAATGCACATTTATGGGCGGAGGACCCGATTTGGAGAAACTCAA  
ST-11 ACGTGGAAACCGTGTGCAAAGGTGTTTCGGAACTTTTAGACGACGGCGAAAAACGTCGAAATGCACATTTATGGGCGGAGGACCCGATTTGGAGAAACTCAA  
ST-192 ACGTGGAAACCGTGTGCAAAGCGTTTCGGAACTTTTAGACGACGGCGAAAAACGTCGAAATGCACATTTATGGGCGGAGGACCCGATTTGGAGAAACTCAA  
ST-11595 ATGTGCGTACTGCGTGCAGGGGGATACAAAACTTCTGGACAGCGGGAAAAACGTCGAACTGCACATCATGGGCGGCGGCCCGGACCTAGAAAAGCTGAA  
ST-35 ACGTGGAAACCGTGTGCAAAGGTGTTTCGGAACTTTTAGACGACGGCGAAAAACGTCGAAATGCACATTTATGGGCGGAGGACCCGATTTGGAGAAACTCAA  
ST-11597 ATGTGCGTACTGCGTGCAGGGGGATACAAAACTTCTGGACAGCGGGAAAAACGTCGAACTGCACATCATGGGCGGCGGCCCGGACCTAGAAAAGCTGAA

ST-53 ACAGCATGAAAAACCGCCCATCAAGTTTTACGGCTACCTCCCTACGCGCATATGATTTCCATCGCCAAAGGCTGCGACATCGCCGTCACGCCATCCAT 900  
ST-53 (pglH2) ACAATACGCTGCGAGGGCATCAAGTTTTACGGCTATATCCCTACGCGAAATGATGTAGTTCGCCAAAGGCTGCGACATCTCCGTCAACGCCATCCAT  
ST-2880 ACAATATGAAAAATCGCCCATCAAGTTTTACGGTTATTTACCTATTTCTGAAATGATGTCGATTGCCAAAGCGTGCACATTTGCCGTCAACGCCATCCAT  
ST-11 ACAATATGAAAAATCGCCCATCAAGTTTTACGGTTATTTACCTATTTCTGAAATGATGTCGATTGCCAAAGCGTGCACATTTGCCGTCAACGCCATCCAT  
ST-192 ACAATATGAAAAATCGCCCATCAAGTTTTACGGTTATTTACCTATTTCTGAAATGATGTCGATTGCCAAAGCGTGCACATTTGCCGTCAACGCCATCCAT  
ST-11595 ACAGATGAAAAACCGCCCATCAAGTTTTACGGCTACCTCCCTACGCGCATATGATTTCCATCGCCAAAGGCTGCGACATCGCCGTCACGCCATCCAT  
ST-35 ACAATATGAAAAATCGCCCATCAAGTTTTACGGTTATTTACCTATTTCTGAAATGATGTCGATTGCCAAAGGCTGCGACATTTGCCGTCAACGCCATCCAT  
ST-11597 ACAGATGAAAAACCGCCCATCAAGTTTTACGGCTACCTCCCTACGCGCATATGATTTCCATCGCCAAAGGCTGCGACATCGCCGTCAACGCCATCCAT

ST-53 TCTTACGCCATGCAGTCGGTTACCAACAAACTTTCCGACTACATAGCATTTGCAAAAACCGTGCTGAACAGCCAGACCAATGCCGAAGTCCCTCGACCTCC 1000  
ST-53 (pglH2) TCCCTACTCCATGCAATCGATACCAACAAACTGTCCGACTACATGGCATTGCAAAAACCGATTTAACAGCCAGGTCAACGACGAAGTCCGCCGAGTCC  
ST-2880 AGCCACGCCATGCAGTCGGTTACCAACAAACTCTCCGACTATATAGCTTTGCAAAAACCGTGCTGAACAGCCAGACCAATGCCGAAGTCCCTCGACCTCA  
ST-11 AGCCACGCCATGCAGTCGGTTACCAACAAACTCTCCGACTATATGGCTTTGCAAAAACCGATTCTGAACAGTCAGAACAAATGCCGAAGTCCCTCGACCTCC  
ST-192 AGCCACGCCATGCAGTCGGTTACCAACAAACTCTCCGACTATATGGCTTTGCAAAAACCGATTCTGAACAGTCAGAACAAATGCCGAAGTCCCTCGACCTCC  
ST-11595 AGCCACGCCATGCAGTCGGTTACCAACAAACTTTCCGACTATATGGCTTTGCAAAAACCGATTTAACAGCCAGACCAATGCCGATGTCCTCGGCCTCT  
ST-35 AGCCACGCCATGCAGTCGGTTACCAACAAACTCTCCGACTATATGGCTTTGCAAAAACCGATTCTGAACAGCCAGACCAATGCCGAAGTCCCTAGCCTCT  
ST-11597 AGCCACGCCATGCAGTCGGTTACCAACAAACTTTCCGACTATATGGCTTTGCAAAAACCGATTTAACAGCCAGACCAATGCCGATGTCCTCGGCCTCT

ST-53 TGAACCTGCTGCCGCACGAAAACACTACCGTTCCGGCGACATGGACAGCTTCGTTCAAGCCGCCAAAATATTTTGAACGC AAAACGACCCCGTTCAGTC 1100  
ST-53 (*pglH2*) TCA~~CCCT~~GCTGCCGCATGCGGACTACCGTTCCGGCGATGTGGACAGCTTCATCCAAAGCCGCCAAAGACATTCTGGCGCGCAAAAACGACCCCGTTCATC  
ST-2880 TGAACCTACTACCGCACGAAAACACTACCGTTCCGGCGACGTGGACAGCTTCGTTCAAGCCGCCAAAGATATTTTGAACGC AAAACGACCCCGTTCAGTC  
ST-11 TGAATCTGCTGCCGCATGAAAACACTACCGTTCCGGCGACGTGGACGGTTCGTTCAAGCCGCCAAAATATTTTGAACGC AAAACGACCCCGTTCAGTC  
ST-192 TGAACCTGCTGCCGCACGAAAACACTACCGTTCCGGCGATGTGGACAGCTTCGTTCAAGCCGCCAAAATATTTTGAACGC AAAAGACGACCCCGTTCAGTC  
ST-11595 TGAACCTACTACCGCACGAAAACACTACCGTTCCGGCGACGTGGACAGCTTCGTTCAAGCCGCCAAAGATATTTTGAACGC AAAACGACCCCGTTCAGTC  
ST-35 TGAACCTGCTGCCGCACGAAAACACTACCGTTCCGGCGACGTGGACAGCTTCGTTCAAGCCGCCAAAGATATTTTGAACGC AAAACGACCCCGTTCAGTC  
ST-11597 TGAACCTACTACCGCACGAAAACACTACCGTTCCGGCGACGTGGACAGCTTCGTTCAAGCCGCCAAAGATATTTTGAACGC AAAACGACCCCGTTCAGTC

ST-53 CGACGAAATCGTCCGCCGCTTAGGGCGGACATCTCCTATCAAAAAATCGTCAACCTGATTGAAAGATTAGCCCATGAGTAA  
ST-53 (*pglH2*) CGACGAAATCGTCCGCCGCTTCAAGCGGATGTAGCGTACCAAAAAATCGTCAACCTGATTGAAAGATTAGCTCATGAGTAA  
ST-2880 CGACGAAATCGTCCGCCGCTTCAAAACGCGACGTTCGCCTATCAAAAAATCGTCAACCTGATTGAAAGATTAGCTCATGA----  
ST-11 CGACGAAATCGTCCGCCGCTTCAGGGCGGACATTCCTATCGGAAAAATCGTCAACCTGATTGAAAGATTAGCTGATGATAA  
ST-192 CGACGAAATCGTCCGCCGCTTCAGGGCGGACATTCCTATCGGAAAAATCGTCAACCTGATTGAAAGATTAGCCCATGAGTAA  
ST-11595 CGACGAAATCGTCCGCCGCTTCAGGGCGGACATTCCTATCGGAAAAATCGTCAACCTGATTGAAAGATTGGCAAAATGAGTAA  
ST-35 CGATGAAATCGTCCGCCGCTTCAGGGCGGACATTCCTATCGGAAAAATCGTCAACCTGATTGAAAGATTGGCAAAATGAGTAA  
ST-11597 CGACGAAATCGTCCGCCGCTTCAGGGCGGACATTCCTATCGGAAAAATCGTCAACCTGATTGAAAGATTGGCAAAATGAGTAA

**Figure S4. Alignment of the *pglH* gene.** The figure shows an alignment of the *pglH* gene where the polyC tract is shown as blue nucleotide sequence when equal in all isolates within an ST, and the extending cytosines (Cs) are shown in grey. Other nucleotide variants and deletions are shown in red. The underlined red bold GCA sequence is unique to *pglH2* sequences (1).



*pglE*

<b>ST-11</b>	GAAGTGAAGATATATGTTATAAATATATAGATAAAAATATCAAA <b>CAACAAA</b> <sub>(31-50)</sub> -----CACCGGACACCAAATATC-AAATATATATATCAGG
<b>ST-35</b>	GAAGTGAAGATATATGTTATAAATATATAGATAAAAATATCAAA <b>CAACAAA</b> <sub>(10-11)</sub> -----CACCGGACACCAAATATC-AAATATATATATCAGG
<b>ST-53</b>	GAAGTGAAGATATATGTTATAAATATATAGATAAAA----- <b>CAACAAA</b> <sub>(9-33)</sub> ----- <b>CAACACCAGAT</b> ACCGAACAC <b>T</b> AAACATC-AAATATATATATCAGG
<b>ST-175</b>	GAAGTGAAGATATATGTTATAAATATATAGATAAAA----- <b>CAACAAA</b> <sub>(21-24)</sub> ----- <b>CAACACCAGAT</b> ACCGAACACCAA <b>C</b> ATC-AAATATATATATCAGG
<b>ST-198</b>	GAAGTGAAGATATATGTTATAAATATATAGATAAAAATATCAAA <b>CACCCAA</b> <sub>(12-14)</sub> ----- <b>CACCAGATACCCGAT</b> ACCGAACACCAA <b>C</b> ATC-AAATATATATATCAGG
<b>ST-2880</b>	GAAGTGAAGATATATGTTATAAATATATAGATAAAA----- <b>CAACAAA</b> <sub>(29-35)</sub> ----- <b>CCAGAT</b> ACCGAACACCAA <b>C</b> ATC-AAATATATATATCAGG
<b>ST-11372</b>	GAAGTGAAGATATATGTTATAAATATATAGATAAAA----- <b>CAACAAA</b> <sub>(17)</sub> -----CACCGGACACCAAATATC-AAATATATATATCAGG
<b>ST-11595</b>	GAAGTGAAGATATATGTTATAAATATATAGATAAAA----- <b>CAACAAA</b> <sub>(12)</sub> ----- <b>CAACACCAGAT</b> ACCGAACACCAA <b>C</b> ATC-AAATATATATATCAGG
<b>ST-11597</b>	GAAGTGAAGATATATGTTATAAATATATAGATAAAA----- <b>CAACAAA</b> <sub>(12)</sub> ----- <b>CAACACCAGAT</b> ACCGAACACCAA <b>C</b> ATC-AAATATATATATCAGG
<b>ST-192</b>	GAAGTGAAGATATATGTTATGAATATATAGATCAAATATCAAA <b>CAACAAA</b> <sub>(9-22)</sub> <b>CAATAAA</b> <b>CAACAAA</b> <sub>(4)</sub> -----CACCGG-ACACCAAATATC-AAATATATATATCAGG
23B	GAAGTGAAGATATATGTTACAAATATATAGATAAAAATATCAAA <b>CACCCGA</b> <sub>(10)</sub> ----- <b>CACCCAATACC</b> CAACACC <b>C</b> AAATATC-AAATATATATATCAGG
6A	GAAGTGAAGATATATGTTATGAATATATAGATCAAATATCAAA <b>CAACAAA</b> <sub>(14)</sub> <b>CAATAAA</b> <b>CAACAAA</b> <sub>(4)</sub> -----CACCGGACACCAAATATC-AAATATATATATCAGG
35AB/50B	GAAGTGAAGATATATGTTAT <b>G</b> AATATATAGATAAAAATATCAAA <b>CAACAAA</b> <sub>(17-18)</sub> <b>CAATAAA</b> <sub>(2)</sub> <b>CAACAAA</b> <sub>(2)</sub> -----CACCGGACACCAAATATC-AAATATATATATCAGG
36A	GAAGTGAAGATATATGTTAT <b>G</b> AATATATAGATAAAAATATCAAA <b>CAACAAA</b> <sub>(17)</sub> <b>CAATAAA</b> <sub>(2)</sub> <b>CAACAAA</b> <sub>(3)</sub> ----- <b>CACCGG</b> -ACACCAAATATC-AAATATATATATCAGG
24A	GAAGTGAAGATATATGTTATGAATATATAGATCAAATATCAAA <b>CAACAAA</b> <sub>(59)</sub> ----- <b>CACCGG</b> -ACACCAAATATC <b>A</b> AAATATATATATCAGG
24B	GAAGTGAAGATATATGTTATGAATATATAGATCAAATATCAAA <b>CAACAAA</b> <sub>(55)</sub> ----- <b>CACCGG</b> -ACACCAAATATC-AAATATATATATCAGG

**Figure S5. Alignment of the heptanucleotide repeat tract and flanking sequences in *pglE*.** The figure shows an alignment of the heptanucleotide repeat tract, shown as blue nucleotide sequence when a CAACAAA repeat. Other repeat variants (CACCCAA and CAATAAA) are shown in green. Other nucleotide variants and deletions are shown in red.

*pglO*

```
40A      GCCAACGGCGGTTTCACAGACTTGCCGCGCCAAATCGAATGGAATAAAGCCCTTGCCGCCTTCCAGTCCGCCCC
40B      GCCAACGGCGGTTTCACAGACTTTCGCGCCAAATCGAATGGAATAAAACCCCTTGCCGCCTTCCAGTCCGCCCC
```

**Figure S6. Homologous recombination in *pglO*.** The figure shows an alignment of the *pglO* gene region from two isolates isolated from individual no. 40 with homologous recombination. Nucleotide variants are shown in red.

***pglG***

```
7A      ACGCCTACAACGCACTCCCCCCCCCCCCC-----GGAATTTCTTTCTGGAAAACCGTCAAAGGCTGCGTCATGCTCGGCACGCTGTCGCGTCTGGACACCGT
7B      ACGCCTACAACGCACTCCCCCCCC-----GGAATTTCTTTCTGGAAAACCAAAAAGAATGCGCCGTACTCGGCACGCTGTCGCGTCTGGACACCGT

46A     ACGCCTACAACGCACTCCCCC / CCCCCC-----GGAATTTCTTTCTGGAAAACCGTCAAAGGCTGCGTCATGCTCGGCACGCTGTCGCGTCTGGACACCGT (incomplete)
46B     ACGCCTACAACGCACTCCCCCCCCCCCCCCCCCCCCTCC-TTCTTTT-----CCCCCAAAGGCTGCGTCATGCTCGGCACGCTGTCGCGTCTGGACACCGT
```

**Figure S7. Homologous recombination in *pglG*.** The figure shows an alignment of the *pglG* gene region from two isolates isolated from individuals nos. 7 and 46 with homologous recombination. Nucleotide variants and deletions are shown in red.

**Table S2. *Neisseria gonorrhoeae* strains used in this study**

<b>Strain</b>	<b>Relevant genotype</b>	<b>References</b>
KS100	<i>recA6</i>	(2)
KS104	<i>pglC</i>	(3)
KS141	<i>pglA</i>	(4)
KS142	<i>pglE<sub>on</sub></i>	(4)
KS966	<i>pglA pglI lct::pglH2<sub>SK-03-1035</sub></i>	(5)

## References

1. 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 U S A* 108:9643-8.
2. Tonjum T, Freitag NE, Namork E, Koomey M. 1995. Identification and characterization of *pilG*, a highly conserved pilus-assembly gene in pathogenic *Neisseria*. *Mol Microbiol* 16:451-64.
3. Hegge FT, Hitchen PG, Aas FE, Kristiansen H, Lovold C, Egge-Jacobsen W, Panico M, Leong WY, Bull V, Virji M, Morris HR, Dell A, Koomey M. 2004. Unique modifications with phosphocholine and phosphoethanolamine define alternate antigenic forms of *Neisseria gonorrhoeae* type IV pili. *Proc Natl Acad Sci U S A* 101:10798-803.
4. Aas FE, Vik A, Vedde J, Koomey M, Egge-Jacobsen W. 2007. *Neisseria gonorrhoeae* O-linked pilin glycosylation: functional analyses define both the biosynthetic pathway and glycan structure. *Mol Microbiol* 65:607-24.
5. 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* 94:688-99.