Supplemental material S1

Plasmid and strain construction

The *pglC*::*kan* mutation in Nel_{gly} was generated by transforming KS944 with a PCR product of the *pglC*::*kan* locus from *Ngo* strain KS105 (1) generated using primers av2828 and av2829. Primer sequences are listed in Table S3.

KP65 is a plasmid that replaces Nel_{gly} *pglO* (Uniprot accession number D4DS59, *NEIELOOT_01902*) with a chloramphenicol resistance marker. Sequences flanking the *pglO* open reading frame were amplified by PCR splice overlap extension (SOE) from KS944 genomic DNA using primers av2890, av2891, av2892, and av2893, resulting in a PCR product in which *pglO* is deleted and a *Hinc*II site inserted. The resulting PCR product was cloned into the pCR4-TOPO TA vector (Invitrogen), resulting in the plasmid KP49. A *Hinc*II fragment carrying the chloramphenicol resistance marker from KP52 was introduced into the *Hinc*II site of KP49, resulting in plasmid KP65. KP52 is based on the pCRII-TOPO TA vector (Invitrogen) and contains a PCR product carrying the chloramphenicol resistance marker from pCM7 (Genbank accession umber U13846.1) flanked by *Hinc*II sites. KP65 was introduced into strain KS1000 by transformation, resulting in strain KS1006.

To make the *pglH*::*kan*, *pglG*::*kan* and *pglGH*::*kan* mutant strains, the *pglGH* sequence from Nel_{gly} was synthesized and cloned into pBluescript II SK+ (GenBank accession number X52327.1) by GeneCust. This plasmid, KP57, contained *pglG* (Uniprot accession number D4DNG4, *NEIELOOT_00594*) and *pglH* (GenBank accession number BankIt1835047 Neisseria KT206226) gene sequences as well as 535 bp upstream of *pglG*. To generate the *pglG*::*kan* allele,

the *Nco*I restriction enzyme was used to digest KP57 and the resulting ends were blunted using Mung Bean Nuclease (New England Biolabs). Next, a kanamycin gene cassette (the *Hinc*II fragment from pKan) was ligated into the blunted *NcoI* site. The resulting plasmid, KP54, was linearized with *Kpn*I and transformed into Nel_{gly} to generate strains carrying the *pglG*::*kan* allele. To make the *pglH*::*kan* allele, KP57 plasmid DNA was modified to contain a *Nru*I restriction site 13 bp into the *pglH* sequence (GeneCust), resulting in plasmid KP58. Subsequently, the *Nru*I site was used to open the KP58 plasmid, the ends were blunted using Mung Bean Nuclease and a kanamycin cassette (the *Hinc*II fragment from pKan) was ligated into the blunted *Nru*I site. The plasmid, KP53, was linearized with *Kpn*I and transformed into Nel_{gly} to generate strains carrying the *pglH*::*kan* allele. To generate the *pglGH*::*kan* allele, *Nco*I and *Nru*I restriction enzymes were used to digest the plasmid KP58 and the ends were blunted using Mung Bean Nuclease. Next, a kanamycin cassette (the *Hinc*II fragment from pKan) was ligated into the blunted *Nco*I - *Nru*I fragment. The plasmid, KP55, was linearized with *Kpn*I and transformed into Nel_{gly} to generate the strain carrying the *pglGH*::*kan* allele.

KP66 containing the *pglH*::*cm* construct was made from KP58 described above. A PCR product containing the chloramphenicol marker from pGCC6 (2, 3) was amplified using primers av2980 and av2981and cut with *Hinc*II before it was inserted into the blunted *Nru*I site in KP58. KP66 was transformed into strain KS944 to generate strain KS1049. This strain was again transformed with crude cell lysate from the *pglC*::*kan* strain KS945 to generate strain KS1050.

KP68 is a plasmid that replaces Nel_{gly} *pglD* with a chloramphenicol resistance marker. Sequences flanking *pglD* were amplified by PCR SOEing from KS944 genomic DNA using the primers av2952, av2953, av2954 and av2955, resulting in a PCR product in which *pglD* was replaced by

an *Eco*RV site. The resulting PCR product was TA cloned into the pCR4-TOPO TA vector (Invitrogen), resulting in the plasmid KP67. A *Hinc*II fragment carrying the chloramphenicol resistance marker from plasmid KP52 was introduced into the *Hinc*II site of KP67, resulting in the plasmid KP68.

KP70 is a plasmid that replaces the Nel_{gly} ORF *NEIELOOT_02177* (Uniprot accession number D4DSY1) with a chloramphenicol resistance marker. Sequences flanking the ORF were amplified by PCR SOEing from KS944 genomic DNA using the primers av2940, av2941, av2942 and av2943, resulting in a PCR product in which *NEIELOOT_02177* was replaced by a *Sma*I site. The resulting PCR product was cloned into the pCR2.1-TOPO TA vector (Invitrogen), resulting in the plasmid KP69. A *Hinc*II fragment carrying the chloramphenicol resistance marker from plasmid KP52 was introduced into the *Sma*I site of KP69, resulting in the plasmid KP70.

KP72 is a plasmid that replaces Nel_{gly} *pglBa* (Uniprot accession number D4DQH0, *NEIELOOT_01309*) with a chloramphenicol resistance marker. Sequences flanking *pglBa* were amplified by PCR SOEing from KS944 genomic DNA using the primers av2944, av2945, av2946 and av2947, resulting in a PCR product in which *pglBa* was replaced by an *Eco*RV site. The resulting PCR product was cloned into the pCR4-TOPO TA vector (Invitrogen), resulting in the plasmid KP71. A *Hinc*II fragment carrying the chloramphenicol resistance marker from plasmid pGCC6 PCR amplified using primers av2890 and av2891 was introduced into the *Hinc*II site of KP71, resulting in the plasmid KP72.

KP74 is a plasmid that replaces Nel_{gly} *pglBb* (Uniprot accession number D4DQH1, *NEIELOOT_01310*) with a chloramphenicol resistance marker. Sequences flanking *pglBb* were

amplified by PCR SOEing from KS944 genomic DNA using primers av2948, av2949, av2950 and av2951, resulting in a PCR product in which *pglBb* was deleted in conjunction with insertion of an *Eco*RV site. The resulting PCR product was cloned into the pCR4-TOPO TA vector (Invitrogen), resulting in the plasmid KP73. A *Hinc*II fragment carrying the chloramphenicol resistance marker from plasmid pGCC6 PCR amplified using primers av2890 and av2891 was introduced into the *Hinc*II site of KP73, resulting in the plasmid KP74.

A C-terminal 6-His tag in *ccoP* (4) was constructed by use of a PCR-SOEing reaction. A pair of PCR fragments containing *ccoP* sequences, an in-frame 6-His tag, a stop codon and an *Eco*NI site was created using the primer FE2485 in combination with FE2483, and the primer FE2487 in combination with FE2484. The resulting overlapping PCR fragments were spliced together using the flanking primers FE2485 and FE2487. The resulting PCR fragment was then cloned into the pCR4-TOPO TA vector (Invitrogen). The plasmid DNA was sequenced to verify the introduction of the 6-His tag, stop codon and the *Eco*NI site and the absence of any other alterations. Next, the *tetM* gene from plasmid pAM120 (GenBank accession number U49939.1) was amplified by PCR using the primers TetF3 and TetR2. The resulting PCR product was digested with *Eco*NI and cloned into the *Eco*NI site downstream of ccoP-His in pCR4-TOPO, resulting in plasmid KP75 Finally, the construct consisting of most of *ccoP* with a C-terminal 6-His tag followed by a tetracycline resistance gene and *ccoP* downstream sequence, was amplified from KP75 by PCR using the flanking primers FE2485 and FE2487. The resulting PCR product was used to transform Nel_{gly} and transformants were selected for on tetracycline containing media.

To make a C-terminal 6-His tag in Nel_{gly} c_5 (UniProt accession number D4DQS1, *NEIELOOT_01411*), the DNA sequence was synthesized (GeneCust) as follows: The last 585 bp

of the *Nel*$_{gly}$ c_5 ORF followed by an in-frame 6-His tag, a stop codon, a 1026 bp fragment from pACYC184 (GenBank accession number X06403.1) containing the chloramphenicol resistance gene, and another 634 bp of genomic sequence downstream of the c_5 ORF was amplified by PCR. The synthesized DNA fragment was then cloned by GeneCust into the unique BamHI/HindIII sites in pUC57 (GenBank accession number Y14837.1), generating plasmid KP76. Finally, we used the resulting plasmid to transform *Nel*$_{gly}$ using selection for chloramphenicol resistance.

A 6-His tag was introduced into the C-terminus of an otherwise wild-type allele of *Nel*$_{gly}$ *nirK* (UniProt accession number D4DP53, *NEIELOOT_00838*) by use of SOE-ing PCR. A pair of PCR fragments containing the 6-His tag was created using primers aniAF2 in combination with aniAHisR, and aniAHisF2 in combination with aniAR2. The resulting overlapping PCR fragments were spliced together using primers aniAF2 and aniAR2. This resulted in a PCR product with a C-terminal in-frame 6-His tag, a stop codon and an *Eco*RV restriction site. The PCR fragment was cloned into the plasmid pUP6 (5). The resulting plasmid was sequenced to verify the introduction of the 6-His tag, stop codon and *Eco*RV site and the absence of any other alterations. Subsequently the *cat* gene cassette was cloned into the *Eco*RV site located after the stop codon in *nirK*, generating the plasmid KP77. Finally, the *nirK* fragment linked to the *cat* cassette was amplified from KP77using primers aniAF2 and aniAR2. This PCR product was used to genetically transform *Nel*$_{gly}$.

References

1. **Vik A, Aas FE, Anonsen JH, Bilsborough S, Schneider A, Egge-Jacobsen W, Koomey M.** Proc Natl Acad Sci U S A **106:**4447-4452, 2009.

2. **Mehr IJ, Long CD, Serkin CD, Seifert HS. 154:**523-532, 2000.

3. **Mehr IJ, Seifert HS.** Mol Microbiol **30:**697-710, 1998.

4. **Aas FE, Li X, Edwards J, Hongro Solbakken M, Deeudom M, Vik A, Moir J, Koomey M, Aspholm M.** Environ Microbiol **17:**2114-2132, 2015.

5. **Wolfgang M, van Putten JP, Hayes SF, Koomey M.** Mol Microbiol **31:**1345-1357, 1999.

FIG S1

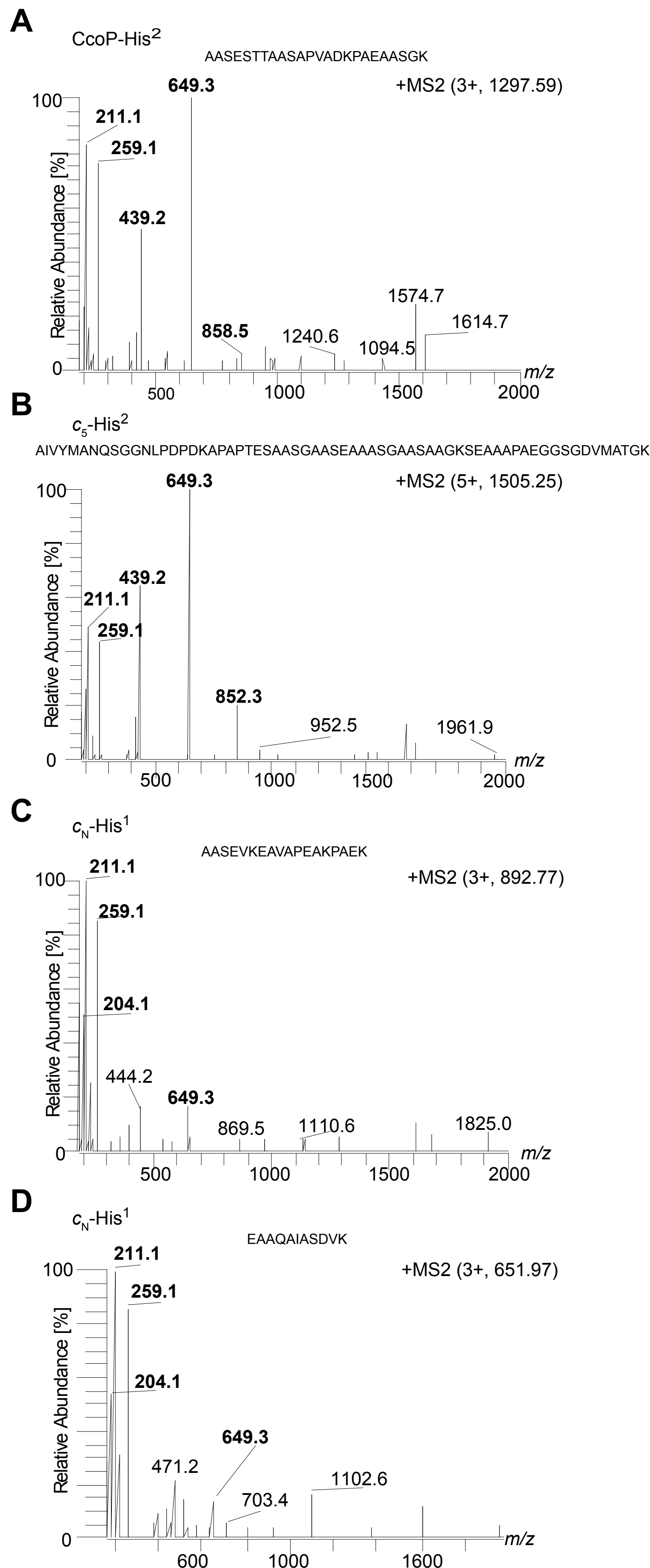

FIG S1. Identification of glycopeptides derived from trypsin cleavage of affinity-purified proteins using MS^2. A) MS^2 spectrum of $[M + 3H]^{3+}$ precursor at *m/z* 1297.59 derived from CcoP-His modified with two glycans. B) MS^2 spectrum of $[M + 5H]^{5+}$ precursor at *m/z* 1505.25 derived from c_5-His modified with two glycans. C) MS^2 spectrum of $[M + 3H]^{3+}$ precursor at *m/z* 892.77 derived from c_N-His modified with one glycan. D) MS^2 spectrum of $[M + 3H]^{3+}$ precursor at *m/z* 651.97 derived from c_N-His modified with one glycan. The glycan reporter ions derived from HCD fragmentation of glycopeptides at *m/z* 852.3, at *m/z* 649.3, at *m/z* 439.2, at *m/z* 259.1, at *m/z* 211.1 and at *m/z* 204.1 are shown in bold. Superscript numbering after protein designations indicates peptides modified with either 1(1) or 2 (2) glycans.

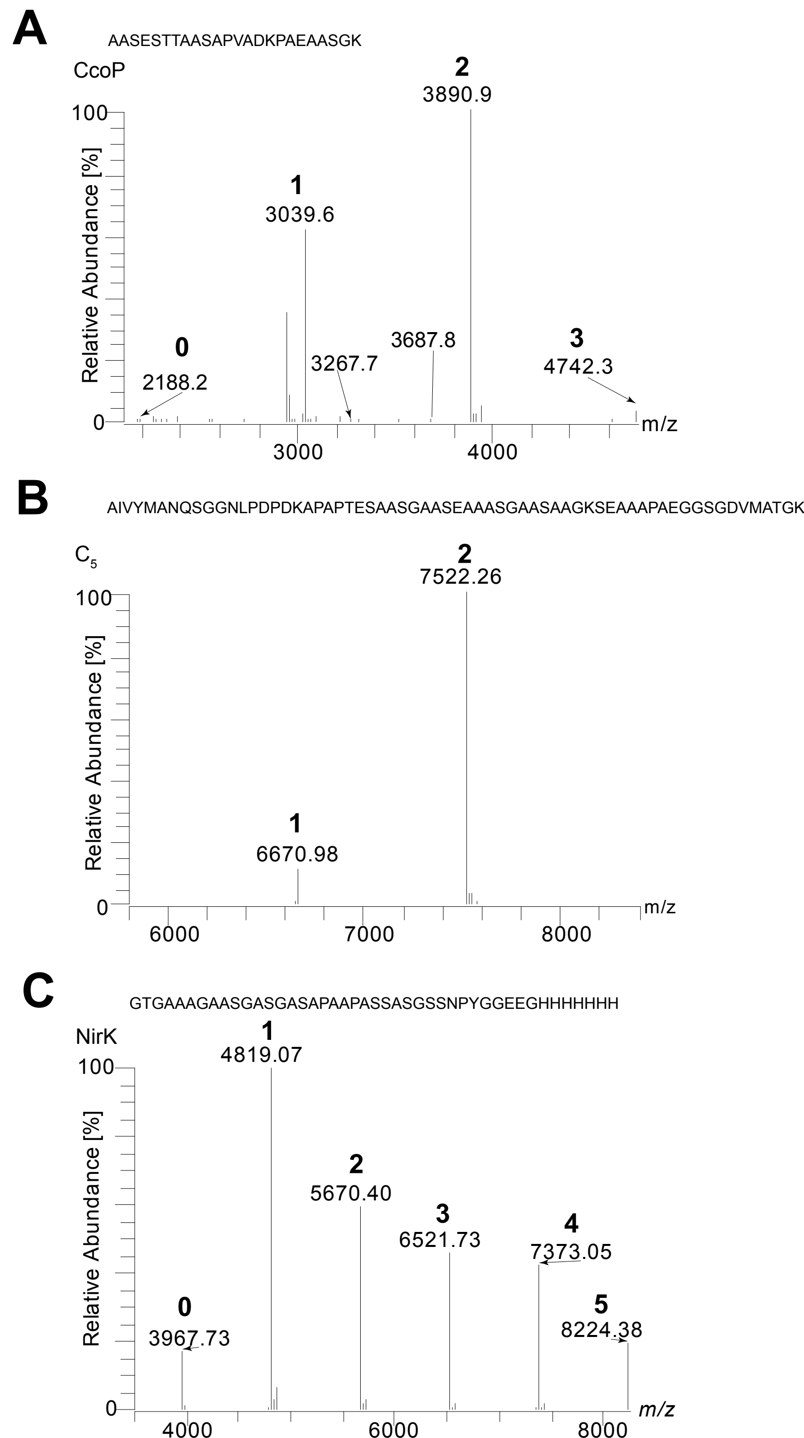

FIG S2 Deconvoluted molecular weight spectra of glycopeptides from affinity purified LC-MS^2 analysis show multiple attached glycans. (A) The deconvoluted molecular weight spectra of glycopeptide AASESTTAASAPVADKPAE-AASGK from CcoP-His. In the left side of the spectrum the species at *m/z* 2188.2 represent the unmodified peptide. In total, mass additions to the peptide equalling three full length glycan modifications were detected. The masses at *m/z* 3267.7 and at *m/z* 3687.8 represent peptide with a single full length glycan and additions of the monosaccharide diNAcBac and the tetrasaccharide diNAcBac-Hex-diNAcHexA, respectively. (B) The deconvoluted molecular weight spectra of glycopeptide AIVYMANQSGGNLPDPDKAPAPTESAASGAASEAAASGAASAAGKSEAAAPAEGGSGDV-MATGK from c_5-His. The mass of the unmodified peptide was not detected. In total, mass additions corresponding to two full length glycan modifications were detected. (C) The deconvoluted molecular weight spectra of glycopeptide GTGAAAGAASGASGASAPAAPASSASGSSNPYGGEEGHHHHHHH from NirK-His. In the left side of the spectrum, the species at *m/z* 3967.73 represents the unmodified peptide. In total, mass additions corresponding to five full length glycan modifications were detected. The deconvoluted molecular weight spectrum was generated from affinity purified NirK-His from a wild-type strain grown in liquid medium. Deconvoluted masses are slightly off theoretical masses due to the incorporation of all masses present in the LC-MS elution time of the peptides presented, which will slightly distort the exact mass of the ^{12}C isotope used for deconvolution. Bold numbers denote the number of full length glycan modifications present on deconvoluted peptide masses. Masses are presented as [MH+]$^+$.

FIG S3

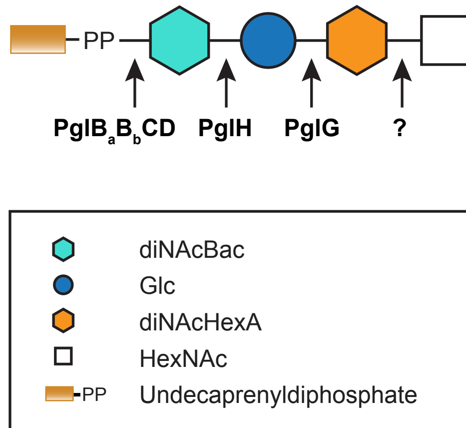

FIG S3 *Nel_{gly}* Pgl proteins and glycan structure

Table S1 Strains

Strain name	Relevant genotype*	Parent	Reference
KS944	*N. elongata* subsp. *glycolytica* ATCC29315		(1)
KS945	*pglC::kan*	KS944	(2)
KS1000	*pglH::kan*	KS944	This study
KS1001	*pglG::kan*	KS944	This study
KS1002	*pglGH::kan*	KS944	This study
KS1041	*pglH::kan pglBa::cm*	KS1000	This study
KS1043	*pglH::kan pglBb::cm*	KS1000	This study
KS1050	*pglH::cm pglC::kan*	KS1049	This study
KS1051	*pglH::kan pglD::cm*	KS1000	This study
KS1040	*pglH::kan pglF::cm*	KS1000	This study
KS1006	*pglH::kan pglO::cm*	KS1000	This study
KS951	*cycC-his::cat*	NK1840	(2)
KS1053	*cycC-his::cat pglH::kan*	KS951	This study
KS992	*nirK-His::cat*	KS944	This study
KS994	*nirK-His::cat pglC::kan*	NK1847	This study
KS1034	*nirK-His pglG::kan*	KS992	This study
KS1052	*nirK-His pglH::kan*	KS992	This study
KS996	*ccoP-His::cat*	KS944	This study
KS1003	*ccoP-His::cat pglC::kan*	KS997	This study
KS997	*cycB-His::cat*	NK1840	This study
KS1005	*cycB-His::cat pglC::kan*	KS997	This study

* All strains are in the *N. elongata* subsp. *glycolytica* ATCC 29315 background.

References

1. **Bovre K, Holten E.** J Gen Microbiol **60:**67-75, 1970

2. **Aas FE, Li X, Edwards J, Hongro Solbakken M, Deeudom M, Vik A, Moir J, Koomey M, Aspholm M.** Environ Microbiol **17:**2114-2132, 2015

Table S2 - *pgl* genes in *Ngo* FA1090 and Nel_{gly} ATCC29315

Protein	Ngo gene	Neg gene name/locus tag	Max identity (%)	Query coverage (%)	BLASTp E-value
PglBa* PglBb*	*ngo0085*	*NEIELOOT_01309* *NEIELOOT_01310*	67 57	48 51	6e-86 1e-70
PglC	*ngo0084*	*NEIELOOT_01311*	91	99	0
PglD**	*ngo0083*	sequenced complete gene	68	98	0
PglF	*ngo0088*	*NEIELOOT_02177*	49	96	8e-112
PglG	*ngo0087*	*NEIELOOT_00594*	63	100	2e-167
PglH***	*ngo0086*	sequenced complete gene	67	99	0
PglO	*ngo0178*	*NEIELOOT_01902*	39	96	7e-140
PglA****	*ngo1765*	-			0.001
PglE****	*ngo0207*	-			2e-08
PglI****	*ngo0065*	-			1.4

* The *pglB* gene is encoded as two separate orfs in Nel_{gly} (manuscript in preparation).

** The genome sequence available from the NCBI database suggested that the Nel_{gly} *pglD* gene was encoded as two separate orfs (annotated as *NEIELOOT_00842* and *NEIELOOT_00844*). We sequenced the gene and found that the database sequence was incorrect and contained a frame shift. The gene is encoded as a single orf.

*** The gonococcal gene *pglH* has homology to two loci (*NEIELOOT_00592* and *NEIELOOT_01305*) in the Nel_{gly} genome sequence that is available from the BIGsdb database. These loci are both located at the ends of contigs. We therefore sequenced the Nel_{gly} *pglH* gene. This sequence shows that there is one *pglH* gene in Nel_{gly} and that the two loci corresponded to the two ends of the *pglH* gene.

**** Neither BLASTp nor BLASTn against the Nel_{gly} draft genome returned any significant homologs based on a combination of e-value, query coverage and maximum identity.

Table S3. Primers used in this study.

Primer name	Sequence
Av2828	5’ cggcttcaaactgcccgttctg 3’
Av2829	5’ attcgcgcatctgtttgacggc 3’
Av2890	5’ ttcagacggcctgtccgcaaccg 3’
Av2891	5’ gcctcgattttgtattgcgcgtcgacgagtacgctcccgaacccgtttc 3’
Av2892	5’ acgggttcgggagcgtactc gtcgac gcgcaatacaaaatcgaggccgtc 3’
Av2893	5’ caactggtcgattcggcactgac 3’
Av2940	5’ ataccatccaaggcacacatacg 3’
Av2941	5’ ccgccgtattcggaagccagcag 3’
Av2942	5’ ttcttcccgccatttcctttcccgggtaatgcacgtccttttgctgccttc 3’
Av2943	5’ cagcaaaaggacgtgcattacccgggaaaggaaatggcgggaagaaacc 3’
Av2944	5’ ccgaagtcatcctgaaagaagcc 3’
Av2945	5’ gcgcctatgcctatccaagagcc 3’
Av2946	5’ ataaaggccgtctgaaaaccgatatcaaccggccgtccgaaacccgtttcc 3’
Av2947	5’ cgggtttcggacggccggttgatatcggttttcagacggcctttatttggc 3’
Av2948	5’ ggcatccccttcgtgatgtac 3’
Av2949	5’ gccaagtccatgatgccgtcc 3’
Av2950	5’ ctccaatcttggaaacacccgatatctaaatccgccgccgtattcaaacagg 3’
Av2951	5’ tgaatacggcggcggatttagatatcgggtgtttccaagattggaggcggcag 3’
Av2952	5’ cgtaccctgtctgacttaatc ‘3
Av2953	5’ tcgttccgccgcgtggcagtaac 3’
Av2954	5’ cgcggcagccgattgcccattcgatatctagatctcctatttttcagacggcctc 3’
Av2955	5’ ccgtctgaaaaataggagatctagatatcgaatgggcaatcggctgccgcgtac 3’

Av2980	5’ gaagcgttgacgtaagaggttccaactttcacc 3’
Av2981	5’ gcttagttgactcaggcatttgagaagcacacgg 3’
FE2485	5′ cggcgaattcttaggcgattttggcggtat 3′
FE2483	5′ ttaccttcggcaggtcagtgatggtgatggtgatggaatttaccaccggattcattc 3′
FE2487	5′ tctcggatcccgcacatcaaacatcaatcaat 3′
FE2484	5′ ccatcaccatcactgacctgccgaaggtaaggctgcacaacaaaaggg 3′
TetF3	5′ ttaacctgccgaaggagtttaagaatacctttatcatgtg 3′
TetR2	5′ ttaaccttcggcaggaggacacaatatccacttgtag 3′
*aniA*F2	3´ cggaattcgagctcaagctggcctgtacatctatc 5´
*aniA*HisR	5′ gatatcttaatggtgatggtgatggtgatgaccttcttcaccgccataag 3′
*aniA*HisF2	5′ caccatcaccatcaccattaagatatctagttaatttggttttatagagcc 3′
*aniA*R2	3´ gcgggatccgagctctccattgacggtaatttgtccg 5´