

(a)	gB-1	1	MESRIWCLVVCVNLICIVCLGAAVSSSST- RGTS -ATHSH SSHTT SA HS RSGSV SQRVTSSQTVSHGVNETI YNTTLKYGDVVGVTNTKYPYRVCSMAQ	98
	gB-2	1	MESRIWCLVVCVNLICIVCLGAAVSSSST SHATS STHNGSH TSRTT SA-QTRSV-Y SQHV TSS EAVSHRANETI YNTTLKYGDVVGVTNTKYPYRVCSMAQ	98
	gB-3	1	MESRIWCLVVCVNLICIVCLGAAVSSSST SHATS STHNGSH ASRTT SA-QTRSV- SSQHV TSS KAVSHRANETI YNTTLKYGDVVGVTNTKYPYRVCSMAQ	98
	gB-4	1	MESRIWCLVVCVNLICIVCLGAAVSSS TS - RATS -THNGH TSHTT SA-QTR SI - SSQRV TSS EAVSHRANETI YNTTLKYGDVVGVTNTKYPYRVCSMAQ	96
	gB-5	1	MESRIWCLVVCVNLICIVCLGAAVSSS ST TGSRG TS IHNH SSHTT SA ARSQ SGSV SHHV TSSQ TVSHD VNETIYNTTLKYGDVVGVTNTKYPYRVCSMAQ	100
(b)	gB-1	398	KLQQIFNTSYNQTYEKYGNVS V FET TGGLV VFWQGI KQKSL VELERLANR SS LN LTHN RTRK ST DG NN AT HL SNMESVHN L VYAQLQFTYD TL RGYINRA	498
	gB-2	398	KLQQIFNTSYNQTYEKYGNVS V FET SGGLV VFWQGI KQKSL VELERLANR SS LN LITH -RTR ST SD NN THLS SM ESVHN L VYAQLQFTYD TL RGYINRA	497
	gB-3	398	KLQQIFN AS YNQTYEKYGNVS V FET TGGLV VFWQGI KQKSL LELERLAN SS GV NSTR -RTRK ST GN-T TL SP ES SVR N VYAQLQFTYD TL RSYINRA	496
	gB-4	396	KLQQIFN AS YNQTYEKYGNVS V FET TGGLV VFWQGI KQKSL VELERLANR SS LN LTHS TRR ST DG TN THLS NMD SVHN L VYAQLQFTYD TL RGYINRA	496
	gB-5	400	KLQQIFN AS YNQTYEKYGNVS V FET TGGLV VFWQGI KQKSL LELERLAN SS GV NATR -R SK R ST N NT TL SL -ENDSVRS V YAQLQFTYD TL RNYINRA	498
(c)	gN-1	1	MEWNTLVGLLVLSVA ESS GN NS ST ST SAT T - SK SS AS SV ST KL TT VAT TS AT TT TTTT LT ST TK LS ST THDPNV MRR HAN DD FYKA HCT SHMYELSL	99
	gN-2	1	MECKTLVLGLL LI SA VT GSS SS NS ST ST S ST-- TP SP SV ST SK PT TS VT TT ST TT TTTT TT ST ST RL SS THDPNV MRR HAN DD FYKA HCT SHMYELSL	98
	gN-3a	1	MEWKT VIL GLFVLSVA AG SSGN SS ST ST SAT LK SS SV ST SK L TT AT TT TTTT TT - MS T RS ST TD NP VM KR TH DDFYKA HCT SHMYELSL	99
	gN-3b	1	MEWKT RV LSFVLSVA AG SYGN SS ST ST SAT TP SP SV ST SV KT ST SV TT ST PT -TTTT LT ST TK PG ST THNP VM KR HD DDFYNA HCT SHMYELSL	99
	gN-4a	1	MEWNTLVGLLVLSVA ASS -- NNT ST V ST P SP-- SS TR LT TVKAT T AT TT ST TT TR ST ST ST TK LS ST TH DPNV MRR HAN DD FYKA HCT SHMYELSL	97
(d)	gN-4b	1	MEWNTLVGLLVLSVA ASS -- NHT ST A ST P SP-- SS ST HT ST TV KAT T AT TT ST TT VT ST ST ST TK PG ST TH DPNV MR PHAN DD FYKA HCT SHMYELSL	97
	gN-4c	1	MEWNTLVGLLVLSVA ASS -- NNT ST A ST P SP-- SS ST RT ST TV KSTAVAT TT ST TA - T ST ST ST SA KPG ST THDPNV MR PHAN DD FYNA HCT SHMYELSL	96
(e)	gO-1a	1	MGRKEM-MVRD V PK MF LIS IS FL LV SF IN CK V MSKA-LYN R W R GL V LS K IG KY KL D QL K LE IL Q LE T-T-- I ST KY - NV - S K P V K N L T M N M TE FF Q Y ILAG P IQ N	103
	gO-1b	1	MGRKGE-M-RGV F N L FL ML LT FL LS F IN C RA AV RL-SV G RY W SG V LS T IG K Q RL DK F KE IL K Q LE K -D-- I Y T K Y F N M-T R Q H I K N L T M N M TE FF Q Y ILAG P IQ N	103
	gO-1c	1	MGRKGD-M-RS I SK L F FI IS L T V LL FS I IN CK V VR-- PP G RY W L G V LS T IG K Q LD KE F KE IL K Q LE R -E-- P Y T K Y F N M-T R Q H V K N L T M N M T Q FF Q Y IL AG P IR N	101
	gO-2a	1	MGRKGE--MRG V N L - L FL I W L T FL LF FF IN CK SARS Q -- RA - P FR P RI W HT V L K KL Q KL LE IL Q LE P -I-- P Y I K Y P Q I-N T RV Q SL T V N MTE FF Q Y ILAG P IR N	100
	gO-2b	1	MGKKKI-LVR G V P R- I FM V - S T FL LI FL IN CK GAL N - P RG P W I G V PS L K W KL KE Q Q L K IE IL K Q LS Q -D-- I Y T K Y P Q I-T K N Y T Q IT EL K K F FL Y IL AG P IR N	102
(f)	gO-3	1	MGRKGE-M-RGV F N L FL LI SL T FL FL SL LN CK SA RV-FRL FP PY GR VL SK NG RL AE IK W Q EL L K Q IGAS Q -- D Y K FT TI - P T K Q GL NA V T ME R FF DN Y ILAG P IQ N	104
	gO-4	1	MGRKGE-M-RGV F N L FL ML SL T FL FL S F IN CK IT VA -- R ----- F R V KS Q KA KE ER Q L K L IL Q EL AS-K T G D Y Y FF TF PS Q Q K LY IT V EM K Q FP PS ILAG P IR N	99
	gO-5	1	MGKKEMIM V K G I P K IM L L IS IT FL LL SL IN CV LV NS R GT RS W PY T V LS Y R G KE IL KK Q ED IL K RL MS-TSSD G Y R FL MY PS Q Q K FA IV IS MD K FF Q Y ILAG P IR N	109
(g)	gO-1a	267	KEQFEKKAKKT Q ST TT PY FS YT TS AAL N VT T NV YS IT TA ARR V ST ST IA YR PD SS FM K S I MA	323
	gO-1b	267	KEQSEKKSKKS Q ST TT PY PS YT TT STAL N VT T NAT YS VT TT ARR V ST ST IA YR PD SS FM K S I MT	323
	gO-1c	265	KEQLEKKTKKS Q ST TT PY FS YT TT STAL N VT T NAT Y KV TT SA KR IP T ST IA YR PD SS FM K S IMA	321
	gO-2a	264	KEQLEKKTKKS Q ST TT PY LS YT TT STAL N VT T NV YS VT TT AK RV PT ST IA YR PD SS FM K S I MA	320
	gO-2b	266	KEQFEKKIKKS Q ST TT PY FS YT TT STAL N VT T NV YS VT TT ARR V PT ST IA YR PD SS FM K S I MA	322
(h)	gO-3	268	KEP-QKKSKKS Q ST TT PS F -Y TT FT FN V-TNV YS AT TT ARR T PT ST IA YR PD SS FM K S I MT	321
	gO-4	263	KSI-SK KSR --V ST TT PT SY ST -ST IF N V ST NT YS PI V- P TR IP T ST IG YR PD EN FM K S ILT	314
	gO-5	273	KEQPQ K N K G S QST TT PY LS YT TT STAF N VT T NV YS AT AA TR V AT ST GY R PD SN FM K S I MA	329
(i)	gH-1	1	MRPGLPPY L T V FT V Y LL SH L PS Q RY G AD A SE AL DP H AF H LL N TY G RP IR FL RE NT T Q CT YN SS LR N ST V V RE NA IS FN FF Q S YN Q Y V	90
	gH-2	1	MRPGLPSY L I V L AV CL L SH LL SS RY GA E AI SE PLD-KAF H LL N TY G RP IR FL RE NT T Q CT YN SS LR N ST V V RE NA IS FN FF Q S YN Q Y V	89

Figure S1. Peptide sequence variations that define the genotypes of glycoproteins. Peptide sequence alignments were constructed using T-coffee version 13.45.0.4846264 [32]. Sequence alignments are color-coded: red indicates a highly conserved residue, blue a weakly conserved residue, and black a non-conserved residue; dashes indicate insertions/deletions. **(a)** The most variable regions defining the five gB genotypes are codons 26-70 and codons 441-490. The GenBank accession numbers of published reference genotype sequences are: gB-1 (Towne strain, ACM48044.1), gB-2 (AD169 strain, DAA00160.1), gB-3 (Toledo strain, ADD39116.1), gB-4 (C194 strain, AAA45925), and gB-5 (saliva isolate, AZB53144). **(b)** The most variable regions defining the four gN genotypes and their subtypes are codons 1-87. The GenBank accession numbers of published reference genotype sequences are: gN-1 (AD169 strain, P16795.1), gN-2 (Can 2 strain, AAL77763.1), gN-3a (PS strain, AAL77773.1), gN-3b (A8-27F strain, AAO24841.1), gN-4a (ZV strain, AAL77779.1), gN-4b (Towne strain, AGT36491.1), and gN-4c (Toledo strain, AAS48964.1). **(c)** The most variable regions defining the five gO genotypes and their subtypes are codons 1-98 and codons 270-313. The GenBank accession numbers of published reference genotype sequences are: gO-1a (AD169 strain, ACL51143.1), gO-1b (Cincy 2 strain, ACS93309.1), gO-1c (Toledo strain, AAS48965.1), gO-2a (FUK19U strain, ABY48952.1), gO-2b (SW1102 strain, AAN40063.1), gO-3 (SW5 strain, AAN40074.1), gO-4 (Towne strain, AGT36493.1), and gO-5 (Merlin strain, YP_081522.1). **(d)** The most variable regions defining the two gH genotypes are codons 1-37. The GenBank accession numbers of published reference genotype sequences are: gH-1 (AD169 strain, ACL51144.1) and gH-2 (Towne strain, AGT36494.1).