

Electronic Supplementary Information for:
**Comparative bioinformatics analysis of the mammalian and
bacterial glycomes**

**Alexander Adibekian,^a Pierre Stallforth,^a Marie-Lyn Hecht,^a Daniel B. Werz,^c Pascal Gagneux,^d
and Peter H. Seeberger^{*a,b}**

[a] Max Planck Institute of Colloids and Interfaces,
Department of Biomolecular Systems, Research Campus Golm,
D-14424 Potsdam, Germany.

E-mail: peter.seeberger@mpikg.mpg.de

[b] Freie Universität Berlin,
Institute for Chemistry and Biochemistry,
Arnimallee 22, 14195 Berlin, Germany

[c] Institut für Organische und Biomolekulare Chemie,
Georg-August-Universität Göttingen, Tammannstr. 2,
D-37077 Göttingen, Germany.

[d] Glycobiology Research Training Center
University of California, San Diego, School of Medicine
9500 Gilman Drive MC 0687
La Jolla, CA 92093-0687, U.S.A.

	Monosaccharide with glycosidic linkage	# found in the analyzed data set	Abundance (%)
1	β -D-Glc (terminal)	877	3.1
2	α -D-Glc (terminal)	707	2.5
3	β (1-3)-D-Gal	654	2.3
4	β (1-4)-D-Glc	633	2.3
5	β -D-Gal (terminal)	617	2.2
6	α (1-3,4)-L-Gro-D-Man-Hep	520	1.9
7	α -D-Gal (terminal)	447	1.6
8	α (1-3)-L-Rha	431	1.5
9	α (1-4,5)-D-Kdo	417	1.5
10	α -D-Kdo (terminal)	398	1.4
11	α -L-Rha (terminal)	393	1.4
12	α (1-2)-L-Rha	379	1.4
13	α -L-Gro-D-Man-Hep (terminal)	364	1.3
14	β (1-4)-D-GlcNAc	357	1.3
15	β -D-GlcNAc (terminal)	294	1.1
16	β (1-3)-D-GlcNAc	289	1.0
17	α (1-2,6)-D-Glc	261	0.9
18	α -D-Neu5Ac	251	0.9
19	β (1-6)-D-Glc	250	0.9
20	α (1-3)-D-Gal	248	0.9

Fig. S1: The 20 most abundant monosaccharides with specific glycosidic linkages in bacteria and their relative abundance.

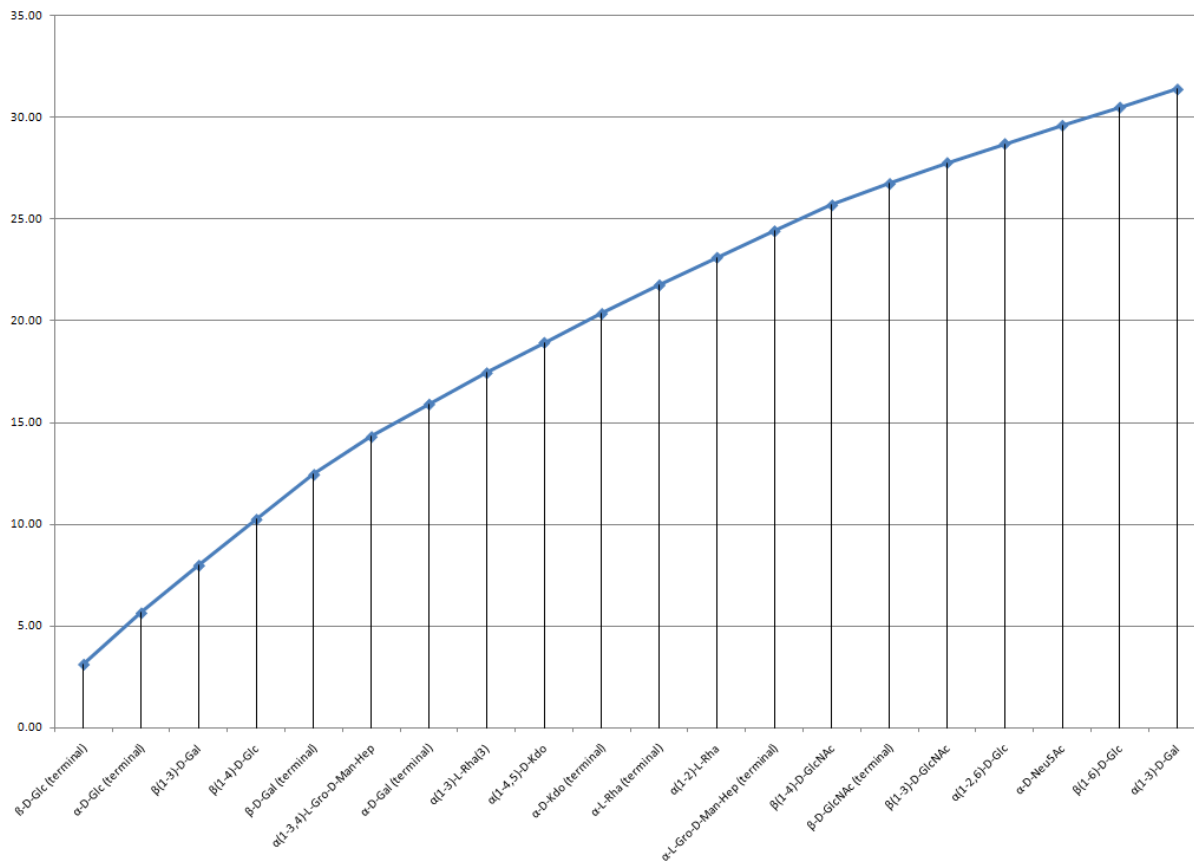


Diagram S1: Coverage of the bacterial glycome by 20 most abundant monosaccharides with specific glycosidic linkages.

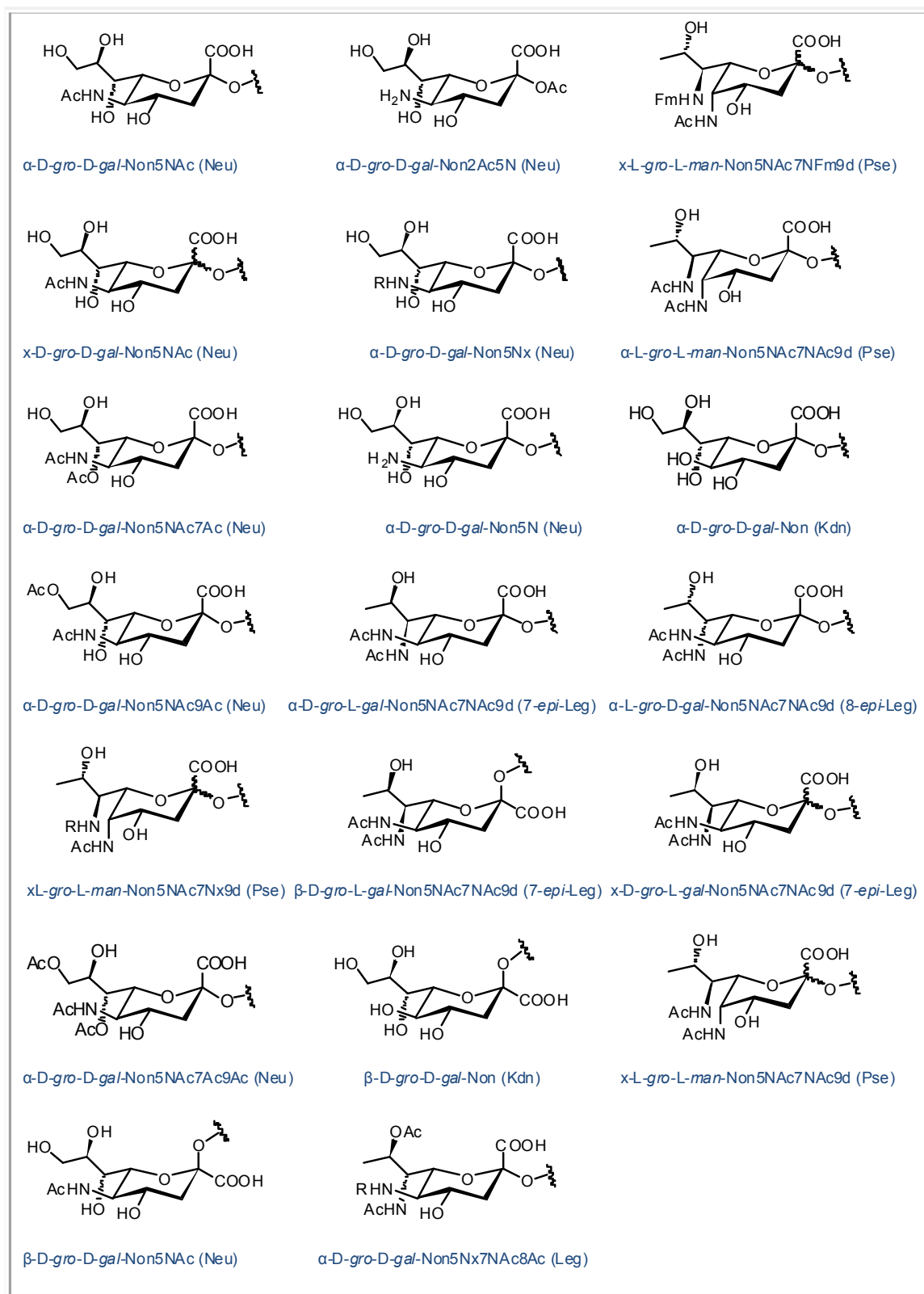


Fig. S2: Structures of 20 most abundant sialic acid derivatives.

Classes	Entries	Percentage (%)	Classes	Entries	Percentage [%]
Mammalia	4739	34.4	Gammaproteobacteria	3625	26.32
- Primates (2174)			- Enterobacteriales (e.g. <i>E. coli</i> , <i>Salmonella</i> , <i>Shigella</i> , <i>Yersinia</i> <i>pestis</i>) (2279)		
- Rodentia (1107)			- Pseudomonadales (e.g. <i>Pseudomonas</i> <i>aeruginosa</i>) (590)		
- Rest (237)			- Rest (e.g. <i>Vibrio</i> <i>cholerae</i> , <i>Haemophilus</i> <i>influenza</i> , <i>Legionella</i>) (756)		
Liliopsida (e.g. lily plants)	308	2.24	Bacilli (e.g. <i>Streptococcus</i> , <i>Staphylococcus</i> , <i>Bacillus</i> <i>anthracis</i>)	706	5.13
Aves (birds)	294	2.13	Actinobacteria (e.g. <i>Streptomyces</i> , <i>Mycobacterium tuberculosis</i>)	515	3.74
Saccharomycetes (e.g. yeasts)	284	2.06	Betaproteobacteria (e.g. <i>Neisseria meningitidis</i> , <i>N.</i> <i>gonorrhoea</i>)	386	2.80
Actinopterygii (ray-finned fishes)	156	1.13	Alphaproteobacteria (e.g. <i>Sphingomonas</i> , <i>Rickettsia</i>)	373	2.71
Insecta	91	0.66	Epsilonproteobacteria (e.g. <i>Helicobacter</i> , <i>Campylobacter</i>)	282	2.05
Eurotiomycetes (e.g. <i>Penicillium</i>)	87	0.63	Clostridia (e.g. <i>Clostridium</i> <i>tetani</i>)	73	0.53
Heterobasidiomycetes (jelly fungi)	75	0.54	Chlamydiae (e.g. <i>Chlamydia</i> <i>trachomatis</i>)	51	0.37
Chondrichthyes (cartilaginous fishes)	45	0.33	Bacteroidetes (e.g. <i>Bacteroides fragilis</i>)	39	0.28
Coniferopsida (conifers)	27	0.20	Fibrobacteres	11	0.08
Total	6106	44.32	Total	6061	44.01

Fig. S3: The 10 eu- and prokaryotic classes with highest number of oligo- and polysaccharide entries in the *BCSDB* and *Glycosciences.de* databases and their relative percentage to all assigned records in both databases.

Abbreviations	
A	uronic acid
Ac	acetyl
Ara	Arabinose
d	deoxy
<i>f</i>	furanose
Fm	formyl
Fuc	Fucose
<i>gal</i>	<i>galacto-</i>
Gal	Galactose
GalA	Galacturonic acid
<i>Galf</i>	Galactofuranose
GalNAc	<i>N</i> -Acetyl-galactosamine
Glc	Glucose
GlcA	Glucuronic acid
GlcN	Glucosamine
GlcNAc	<i>N</i> -Acetyl-glucosamine
<i>gro</i>	<i>glycero-</i>
GulNAc	<i>N</i> -Acetyl-gulosamine
Hep	Heptose
Kdn	3-Deoxy-D-glycero-D-galacto-nonulosonic acid
Kdo	3-Deoxy-D-manno-oct-2-ulosonic acid
Leg	Legionaminic acid
<i>man</i>	<i>manno-</i>
Man	Mannose
Me	methyl
Neu	Neuraminic acid
Neu5Ac	<i>N</i> -Acetyl-neuraminic acid
Non	Nonose
p	phosphate
<i>p</i>	pyranose
Pse	Pseudaminic acid
QuiNAc	<i>N</i> -Acetyl-quinovosamine
Rha	Rhamnose
Xyl	Xylose

List S1: Carbohydrate-related abbreviations.