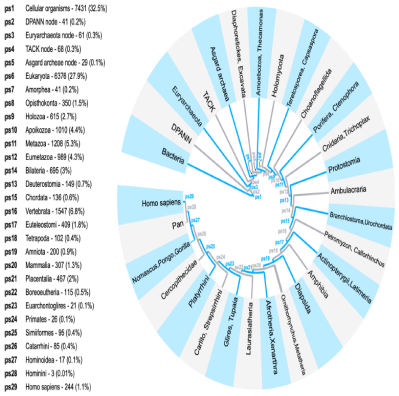
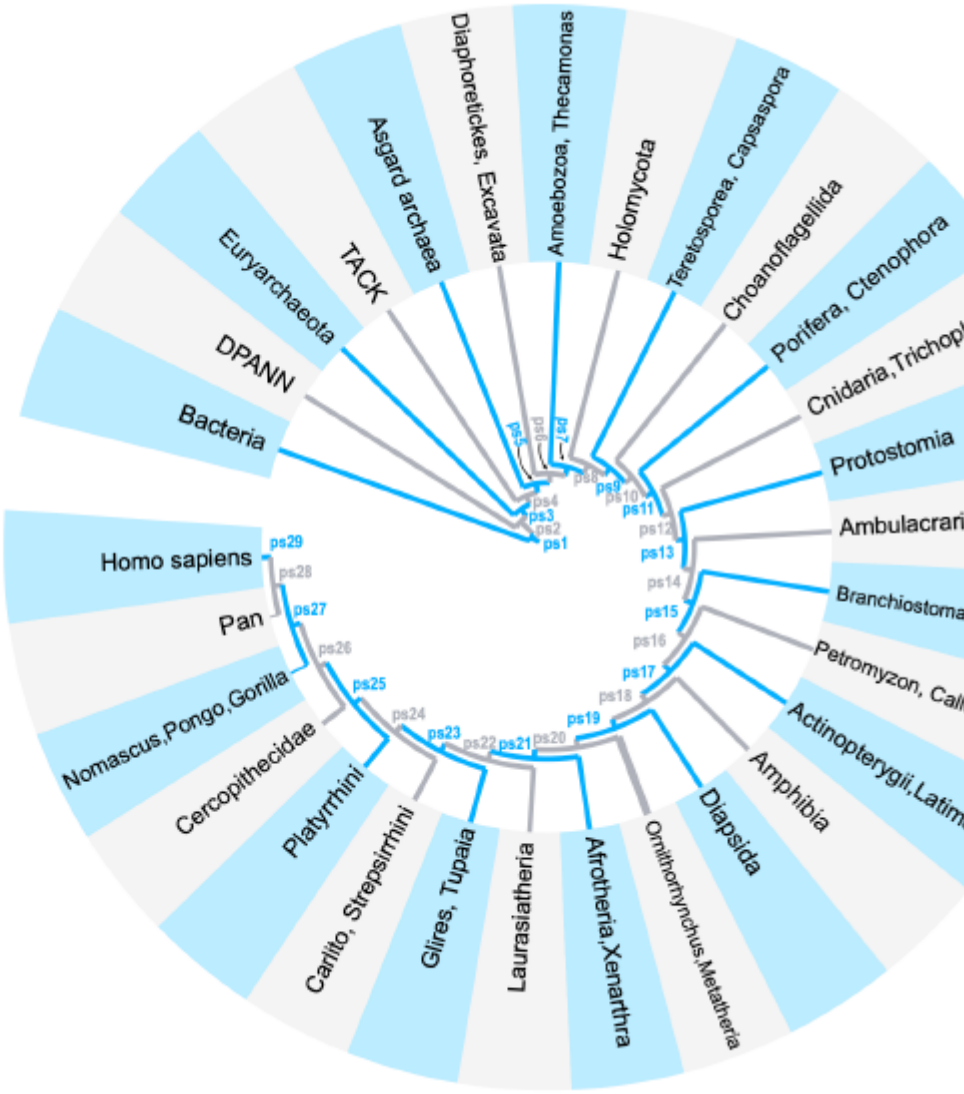


Contrasting macroevolutionary patterns in the human N-glycosylation pathway

Description

Based on coding mutations and splicing variants, post-translational modifications add a final layer to protein diversityat developmental and physiological timescales. Although protein glycosylation is one of the most common post-translational modifications, its evolutionary origin remains largely unexplored. Here, the authors performed a phylostratigraphic tracking of glycosylation machinery genes and their targets — glycosylated proteins — in a broad phylogenetic context. Their results show that most human glycosylation machinery genes trace back to two evolutionary periods: the origin of all cellular organisms and all eukaryotes.

ps1	Cellular organisms - 7431 (32.5%)
ps2	DPANN node - 41 (0.2%)
ps3	Euryarchaeota node - 61 (0.3%)
ps4	TACK node - 68 (0.3%)
ps5	Asgard archeae node - 29 (0.1%)
ps6	Eukaryota - 6376 (27.9%)
ps7	Amorphea - 41 (0.2%)
ps8	Opisthokonta - 350 (1.5%)
ps9	Holozoa - 615 (2.7%)
ps10	Apoikozoa - 1010 (4.4%)
ps11	Metazoa - 1208 (5.3%)
ps12	Eumetazoa - 989 (4.3%)
ps14	Bilateria - 695 (3%)
ps13	Deuterostomia - 149 (0.7%)
ps15	Chordata - 136 (0.6%)
ps16	Vertebrata - 1547 (6.8%)
ps17	Euteleostomi - 409 (1.8%)
ps18	Tetrapoda - 102 (0.4%)
ps19	Amniota - 200 (0.9%)
ps20	Mammalia - 307 (1.3%)
ps21	Placentalia - 467 (2%)
ps22	Boreoeutheria - 115 (0.5%)
ps23	Euarchontoglires - 21 (0.1%)
ps24	Primates - 26 (0.1%)
ps25	Simiiformes - 95 (0.4%)
ps26	Catarrhini - 85 (0.4%)
ps27	Hominoidea - 17 (0.1%)
ps28	Hominini - 3 (0.01%)
ps29	Homo sapiens - 244 (1.1%)



The consensus phylogeny used in the phylostratigraphic analysis. The condensed consensus tree covers divergence from the last common ancestor of cellular organisms to *Homo sapiens* as a focal organism. The tree is constructed by considering the importance of evolutionary transitions and the availability of reference genomes. The internodes (29 phylostrata) that lead from the tree's root to the focal species (*H. sapiens*) are marked by ps1-ps29. The numbers of *H. sapiens* genes traced to each phylostratum and corresponding percentages (in parentheses) are given after phylostrata names. A total 23,237 *H. sapiens* genes was mapped.

This indicates that protein glycosylation is an ancient process likely common to all life, further elaborated in early eukaryotes. In contrast, human glycoproteins exhibited prominent enrichment signals in more recent evolutionary periods, suggesting an essential role in the transition from metazoans to vertebrates. Focusing specifically on the N-glycosylation pathway, the authors noted that most N-glycosylation genes acting on the cytoplasmic side of the endoplasmic reticulum (ER) trace back to the origin of cellular organisms. This sharply contrasts with the rest of the N-glycosylation pathway, which is oriented toward the ER lumen, where genes of eukaryotic origin predominate. The authors also identified an analogous binary evolutionary origin of glycosylation machinery genes in the

Golgi. They discuss these findings in the context of the evolutionary emergence of the eukaryotic endomembrane system and propose that the ER evolved through the invagination of a prokaryotic cell membrane containing an N-glycosylation pathway.

Category

1. News