

Makroevolucijska povijest glikozilacije i biostatistička analiza utjecaja okolišnih čimbenika na N-glikozilaciju imunoglobulina G

Kifer, Domagoj

Doctoral thesis / Disertacija

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Pharmacy and Biochemistry / Sveučilište u Zagrebu, Farmaceutsko-biokemijski fakultet**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:163:802817>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom](#).

Download date / Datum preuzimanja: **2024-11-25**



Repository / Repozitorij:

[Repository of Faculty of Pharmacy and Biochemistry University of Zagreb](#)





Sveučilište u Zagrebu

Farmaceutsko-biokemijski fakultet

Domagoj Kifer

**Makroevolucijska povijest glikozilacije i
biostatistička analiza utjecaja okolišnih
čimbenika na *N*-glikozilaciju
imunoglobulina G**

DOKTORSKI RAD

Zagreb, 2022.



Sveučilište u Zagrebu

Farmaceutsko-biokemijski fakultet

Domagoj Kifer

**Makroevolucijska povijest glikozilacije i
biostatistička analiza utjecaja okolišnih
čimbenika na *N*-glikozilaciju
imunoglobulina G**

DOKTORSKI RAD

Mentori:

prof. dr. sc. Gordan Lauc

doc. dr. sc. Erim Bešić

Zagreb, 2022.



University of Zagreb

Faculty of Pharmacy and Biochemistry

Domagoj Kifer

**Macroevolutionary History of
Glycosylation and Biostatistical
Analysis of Environmental Effects on
N-Glycosylation of Immunoglobulin G**

DOCTORAL DISSERTATION

Supervisors:
Prof. Gordan Lauc, PhD
Asst. Prof. Erim Bešić, PhD

Zagreb, 2022

INFORMACIJE O MENTORIMA

Doktorski je rad predan na ocjenu Fakultetskom vijeću Farmaceutsko-biokemijskog fakulteta Sveučilišta u Zagrebu radi stjecanja akademskog stupnja doktora znanosti iz područja biomedicine i zdravstva, polje farmacija, grana farmacija.

Rad je izrađen pod mentorstvom prof. dr. sc. Gordana Lauca i doc. dr. sc. Erima Bešića, na Zavodu za biofiziku i Zavodu za biokemiju i molekularnu biologiju Farmaceutsko-biokemijskog fakulteta Sveučilišta u Zagrebu u suradnji s Institutom Ruđer Bošković i Medicinskim fakultetom Sveučilišta u Splitu, u sklopu doktorskog studija „Farmaceutske znanosti“ Farmaceutsko-biokemijskog fakulteta Sveučilišta u Zagrebu.

ZAHVALE

Veliko hvala mentorima prof. dr. sc. Gordanu Laucu i doc. dr. sc. Erimu Bešiću na odбору teme, stručnom vodstvu, nesebičnoj pomoći i brojnim savjetima prilikom izrade ovog rada. Hvala vam na pruženim prilikama i ukazanom povjerenju. Hvala vam što ste mi i kroz brojne savjete i rad prenijeli dio svog velikog znanja ne samo u svijetu znanosti, nego i o životu općenito.

Hvala izv. prof. dr. sc. Tomislavu Domazet-Loši i prof. dr. sc. Goranu Klobučaru na ugodnoj suradnji, stručnom vodstvu i brojnim raspravama u području evolucijske genetike.

Hvala prof. dr. sc. Ozrenu Polašku na dopuštenju da se podaci prikupljeni projektom „10001 Dalmatinac“ koriste u ovoj disertaciji.

Hvala Frani Vučkoviću što je uvijek bio spreman prokomentirati razne statističke metode i otvoriti mi pogled na nove ideje i jednostavnija rješenja.

Olgi Gornik Kljaić i Maji Šegvić Klarić hvala na ukazanom povjerenju prilikom uključivanja u razne projekte! Uvijek ću rado biti dio vašeg tima.

Hvala svim članovima Zavoda za biokemiju i molekularnu biologiju, Zavoda za opću i anorgansku kemiju, Zavoda za mikrobiologiju, „mlađariji“ sa Zavoda za analitiku i kontrolu lijekova te višim stručnim savjetnicima za projekte na podršci i radnoj atmosferi. Zbog vas su mi radni dani lijepi!

Posebno se zahvaljujem Sanji Dabelić kao najopasnijoj kritičarki i mrziteljici! Sanja, tvoji nemilosrdni komentari su najbolja motivacija da se stvari učine boljima.

Na kraju, zahvaljujem svim onim osobama iz privatnog života koji možda nisu izravno pomogli pri stvaranju ovog rada, ali su svojim poticanjem i podrškom njegovu izradu učinili lakšom. Ovdje prvenstveno mislim na svoju obitelj, osobito roditelje i Nastasiu.

SAŽETAK

Glikozilacija je prisutna u svim živućim organizmima – od bakterija i arheja pa sve do čovjeka. Struktura glikana nije izravno zapisana u genomu nego je rezultat precizno kontrolirane aktivnosti proteina zaduženih za biosintezu glikana. Strukturne razlike glikana između vrsta su odgovor na selektivni pritisak, a svaka promjena koja je postala konzervirana u evoluciji nosi važnu ulogu. S druge strane, uočena je varijabilnost glikana unutar organizama jedne vrste što je dijelom rezultat okolišnih čimbenika i stanja pojedine jedinke.

Cilj ovog rada je odrediti makroevolucijske trendove glikozilacije i analizirati utjecaj okolišnih čimbenika na *N*-glikom imunoglobulina G (IgG).

Genomskom je filostratigrafijom određena starost gena koji kodiraju za glikoproteine i gena koji kodiraju za proteine koji sudjeluju u glikozilaciji, dobiveni rezultati su pokazali da proteini koji sudjeluju u glikozilaciji vuku korijene još iz zajedničkog pretka svih staničnih organizama, uz pojedine inovacije u zajedničkom pretku svih eukariota, te manje u zajedničkom pretku pravih životinja. Unatoč starosti, glikozilacija u svoje zlatno doba ulazi razvojem višestaničnih životinja s ciljem razvoja međustanične komunikacije i interakcije stanica s okruženjem kada u većoj mjeri nastaju novi geni koji kodiraju za polipeptidne lance glikoproteina.

S ciljem procjene utjecaja raznih čimbenika i stanja analizirana je povezanost podataka dostupnih u biobazi „10001 Dalmatinac“ s različitim deriviranim svojstvima *N*-glikoma IgG-a, kao najproučavanijeg glikoproteina. Rezultati dobiveni analizom preko 3500 ljudi su pokazali da postoji utjecaj prehrane (učestalost konzumacije voćnih derivata i slatke hrane), hipertenzije, artritisa, regularnosti menstrualnog ciklusa, debljine, krvnog tlaka i plazmatske koncentracije gliciranog hemoglobina, fibrinogena i inzulina na sintezu agalaktoziliranih glikana IgG-a. Zatim, digalaktozilacija IgG-a je povezana s hipertenzijom, artritismom, redovitošću menstrualnog ciklusa, debljinom, krvnim tlakom, mineralnom gustoćom kostiju, prehranom (voćni derivati, slatka hrana), koncentracijama urata, triglicerida, fibrinogena, inzulina, kolesterola i HDL kolesterola u plazmi. Sijalinizacija IgG-a je povezana s regularnošću menstrualnog ciklusa, debljinom, mineralnom gustoćom kostiju, krvnim tlakom, koncentracijama triglicerida i inzulina u plazmi. Udio struktura s račvajućim *N*-acetilglukozaminom u *N*-glikomu IgG-a ovisi o mineralnoj gustoći kostiju, redovitošću menstrualnog ciklusa, pušenju, te koncentracijama triglicerida i kolesterola u plazmi.

KLJUČNE RIJEČI

N-glikom, genomski filostratigrafija, evolucija glikozilacije, evolucija glikoproteina, utjecaj okoliša

SUMMARY

Introduction: Glycosylation is the enzymatic process that transfers an oligosaccharide (glycan) to an aglycon; i.e., protein, lipid or even RNA. In *Homo sapiens*, glycans are synthesized and attached to proteins or lipids through one of the 16 glycosylation pathways consisting of about 700 genes encoding enzymes transporters, and other proteins found in cellular glycosylation machinery. Protein glycosylation is the most complex post-translational modification occurring almost completely in the endoplasmic reticulum (ER) and the Golgi by sequential reactions catalysed by glycotransferases and glycoside hydrolases. It is known for a long time that more than a half of all human proteins are glycosylated.

N-glycosylation biosynthesis – the most studied type of protein glycosylation in eukaryotes – occurs at the cytoplasmic side of the ER membrane where monosaccharides are sequentially added to the activated lipid carrier until a glycan with $\text{Man}_5\text{GlcNAc}_2$ structure is formed. This glycan structure is then translocated by a flippase to the luminal side of the membrane of the ER, where it is further elongated to the $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ structure common to all eukaryotes and attached to a polypeptide. Glycan processing continues in the Golgi, where under control of specific glycosyltransferases and glycoside hydrolases, the final structure is formed. The final structure varies within populations and between species.

In prokaryotes N-glycosylation biosynthetic pathway starts at cytosolic side of plasma membrane involving lipid carrier and a flippase, which flips the oligosaccharide to the periplasmic side of the plasma membrane where it is transferred to the protein. However, prokaryotic glycans are structurally very different from those found in eukaryotes, with more diverse monosaccharide types, some of which could be exclusive to specific clades.

Comparing human N- and O- glycans with those found in other organisms seems to uncover the expected pattern of similarity decrease with the increase of the distance across the evolutionary tree. Glycosaminoglycan structures exhibit similar patterns, but they more depend on the habitat (i.e. marine organisms have higher degree of sulfation).

To date, several studies explored the evolutionary origin of some parts of glycosylation pathways – mostly N-glycosylation pathway. These studies suggest that the ER part of the N-glycosylation pathway has bacterial and archaeal origin, whereas the part of the pathway localized in Golgi uses newer enzymes that originated in eukaryotes.

Intraspecies variability also exists. Unlike protein synthesis, glycosylation is not template driven process, but rather a response to the environment by finely tuned expression of glycosyltransferases and glycoside hydrolases.

Numerous studies have shown that plasmatic *N*-glycome changes in different diseases, with age, and between genders, but it also depends on some environmental factors such as body mass index, dyslipidaemia, smoking or eating habits. However, the change in plasma *N*-glycome could arise due to a change in relative abundances of plasma proteins. Analysis of the *N*-glycome of a single protein overcomes the weaknesses of plasma *N*-glycome analysis. The most explored glycoprotein in plasma is immunoglobulin G (IgG). In recent years, a big step is made in finding correlations between *N*-glycans of IgG and various diseases. *N*-glycans also showed big variability within healthy people, as well as great stability over time which qualifies *N*-glycome as a potential biomarker in personalized medicine. Numerous studies found a relationship between IgG *N*-glycome and various diseases, however very few of them tried to explain the effect of environmental factors.

Therefore, the goal of this doctoral thesis is to reconstruct macroevolutionary trends in glycosylation and to evaluate the effects of environmental factors on IgG *N*-glycome in humans.

Methods: Genomic phylostratigraphy was used for the reconstruction of macroevolutionary trends of genes coding for proteins involved in glycosylation and genes coding for glycoproteins. The process of this method can be divided into two parts: 1) Finding homologues of human proteins in a database consisting of proteomes of other organisms; and 2) mapping the human proteins to reliable phylogeny based on the results of the first part of the method. Proteomes were downloaded from the Ensemble and NCBI database. These protein sequences were compared using the BLAST algorithm with a non-redundant database containing about 25,000 proteomes (108 protein sequences) with a cut-off value of 10^{-3} .

The results obtained by searching local databases (BLAST) was used in the assignment of human genes to phylogeny divided into 29 levels (phylostrata) from cellular organisms (first phylostratum) to *Homo sapiens* (last phylostratum), where each phylostratum represents a group of organisms that share a common ancestor with human. The resulting distribution was a null-distribution (control) and was compared with the equally obtained distribution of genes encoding proteins involved in glycan biosynthesis and genes encoding glycoproteins. The frequency comparison was expressed as the odds ratio, and the probability of deviation from the *null*-distribution (*p* value) was calculated for each phylostratum using hypergeometric test. Adjustment of *p* values for multiple comparisons was performed using the FDR (False discovery rate) method according to the Benjamini-Hochberg procedure. An adjusted *p* value of less than 0.05 was considered statistically significant. Statistical analysis was done using Python and R.

Environmental effect on *N*-glycome of the IgG. The data used in this research were obtained from the Croatian Biobank developed within the project "10001 Dalmatians", which aimed to create a comprehensive resource for research into genetic, environmental, and social determinants of health and disease, especially chronic diseases. The research results of this project to date have helped to understand uric acid metabolism, blood pressure levels, lung function, vision, blood fat levels, menstruation and menarche, sleep regulation, smoking and glycan levels. Data were sampled in 4 populations in the period from 2003 to 2013 on the island of Vis (N = 1012), Korčula (town of Korčula, N = 962; village and surroundings of Smokvica, N = 710) and the city of Split and surroundings (N = 987).

Data included A) anthropometrical data (height, weight, waist circumference, hips, lung capacity, bone density, blood pressure) obtained by appropriate measurements and results of cognitive tests (standard progressive matrices, symbol substitution test, verbal fluency tests), standard biochemical tests (serum concentrations of uric acid, glucose, cholesterol, triglycerides, HDL, LDL, VLDL, calcium, albumin, HbA1c, insulin and fibrinogen) and the results of chromatographic analysis of *N*-glycome of IgG. B) Demographic data (age, sex, duration of education, economic status of the household), data on medical history, data on fertility (only for women, age of first menstruation, regularity of menstruation, number of children, age of menopause, number of miscarriages, number of surgical abortions), data on alcohol consumption, data on smoking, data on physical activity, data on sociological behaviour, and data on intake and preference of food and dietary supplements obtained by the questionnaire. C) Results of validated questionnaires: World Health Organization (WHO) Rose angina questionnaire, WHO Questionnaire on intermittent claudication, European Community Respiratory Health Survey, Munich chronotype questionnaire, the General health questionnaire, the Epworth sleepiness scale, the Berlin questionnaire for assessing the risk of obstructive sleep apnea, Revised Eysenck Personality Questionnaire. Data were analysed using mixed modelling with glycan trait as dependent variable, and environmental factor as independent. Sex and age were included as covariates. Sampling place was used as random intercept. Before modelling both IgG *N*-glycome data and environmental factor data were transformed to the standard normal distribution by inversion transformation of ranks. All observed *p* values were adjusted by Benjamini-Hochberg's method to control the false discovery rate. Statistical analyses were done in R.

Results and discussion: Phylostrata have been successfully assigned to all sampled genes encoding glycoproteins. Among them, the largest part, as much as 38%, originated in the common ancestor of all cellular organisms (cellular organisms, ps1), which is statistically

significantly more than expected. In addition, a statistically significant signal of the emergence of new genes is visible in the evolutionary period from the common ancestor of Apoicozoa to the common ancestor of Euteleostoma (ps17). During this period, a total of 37% of genes encoding polypeptide chains of glycoproteins were created, with the odds ratio being statistically significantly higher than 1 and ranging from 1.26 to 2.32 per phyostratum. The exceptions were phyostrata Bilateralia (ps13) and Chordata (ps15), when more genes have occurred than expected, but this increase was not statistically significant. These results point to the fact that glycosylation in human evolution began to be intensively used with the appearance of the first organisms that could be grouped into colonies, for which it was necessary to develop a system of communication between cells. Accelerated development of glycoprotein polypeptides continues all the way to vertebrates with a real skeleton (Euteleostoma, ps17), which includes development from simple to complex animals.

Evolutionary age has also been successfully assigned to all genes encoding proteins involved in glycosylation (PIG). A total of 56% of these genes are traceable to the common ancestor of all cellular organisms (ps1). The rest of the genes encoding the proteins involved in the glycosylation process have their evolutionary origin mostly in the common ancestor of all eukaryotes – 24 %, and from the common ancestor of Apoicozoa (ps10) to the common ancestor of all Eumetazoa (ps12) – 12 %. However, those observations were not statistically significant.

In addition, a statistically significantly higher proportion of new genes occurred in sets of genes encoding for PIG labelled with *O-glycosylation*, *N-glycosylation*, and *GPI anchors* in the common ancestor of all Eukaryotes (ps6). There was also a statistically significantly higher proportion of newly formed genes among those labelled with tags *glycosaminoglycans* and *glycan-binding proteins* in the common ancestor of all Eumetazoa (ps12).

Interestingly, a significant proportion of human genes associated with glycosaminoglycans and glycan-binding proteins were innovations of the common ancestor of all Eumetazoans (ps12) – the ancestor of all animals except sponges and Placozoa (the simplest animals). One of the main characteristics of Eumetazoans includes tissues organized into cell layers. Glycosaminoglycans play a major role in building the extracellular matrix and endothelial glycocalyx layer – which are the basis for the structure and functionality of tissues. Glycan-binding proteins, on the other hand, are involved in cell adhesion and are important in many functions of multicellular organisms. Apparently, glycosaminoglycans and glycan-binding proteins appear to play an important role in the evolution of complex animals.

In the part of the *N*-glycan biosynthetic pathway that takes place in the endoplasmic reticulum, from the addition of the first monosaccharide to dolichol phosphate to the last reaction on the cytosolic side of the endoplasmic reticulum membrane, evolutionarily older enzymes take part phylostratum: cellular organisms (ps1). Flipase is evolutionarily newer with eukaryotic origin, as well as all subsequent enzymes that catalyse reactions localized in the lumen of the endoplasmic reticulum to a reaction catalysed by oligosaccharyltransferase (OST). The OST protein complex consists of a catalytic subunit found in the common ancestor of all cellular organisms (ps1). Other, regulatory subunits have eukaryotic origin (ps6).

It has been observed that these enzymes are grouped into blocks and that evolutionarily older enzymes are oriented towards the cytosolic side, while evolutionarily newer enzymes are oriented towards the lumen of the endoplasmic reticulum.

In the part of the *N*-glycan biosynthetic pathway that takes place in the Golgi apparatus, proteins predominate, the origin of which we successfully trace back to the common ancestor of all eukaryotes. Mannosidases of the cis cisternae of the Golgi apparatus, as well as mannosidases found in the medial cisternae of the Golgi apparatus, were successfully traced to the common ancestor of all cellular organisms (ps1). Other enzymes found in the medial cisternae of the Golgi apparatus, which are exclusively glycosyltransferases, are an innovation of the common ancestor of all eukaryotes (ps6) with a few exceptions.

The pronounced binary distribution of the evolutionary age of the genes between ps1 and ps6 with the pattern of evolutionarily older enzymes located on the cytosolic side of the endoplasmic reticulum membrane and evolutionarily newer enzymes located on the luminal side of the endoplasmic reticulum membrane is somewhat similar to the results published in Lombard's paper. Lombard concluded that the origin of enzymes is diverse: proteoarcheal, bacterial and eukaryotic. By comparing Lombard's results with the results presented in this dissertation, one can see that they coincide with each other for a part of the genes encoding enzymes that catalyse reactions in the lumen of the endoplasmic reticulum.

On the other hand, genes encoding enzymes that catalyse reactions on the cytosolic side of the endoplasmic reticulum membrane, Lombard has mostly characterized as proteoarcheal, except for ALG13 and ALG14 which are suggested to have bacterial origin. At the same time, the results from this dissertation show that these genes are present in both archaea and bacteria.

Given that the results showed that the genes encoding the enzymes of the first part of the biosynthetic *N* glycosylation pathway (with the catalytic site facing the cytosol) and originating from the common ancestor of all cellular organisms are almost completely present in both bacteria and archaea, suggesting that the development of the endomembrane system came by

invagination of the host membrane. The evolutionary age distribution of genes concerning the membrane topology of the catalytic site of the enzyme they encode also supports the relatively recently proposed “inside-out” eukaryogenesis hypothesis proposed by Baum and Baum who have suggested archaeal host. Thus, the former periplasmic space between extracellular protrusions is enclosed in the lumen of the endoplasmic reticulum in which part of the *en bloc* transfer of glycans to protein is retained. On the other hand, results also support revised syntrophy hypothesis where eukaryogenesis is based on two endosymbiotic events: (1st) Asgard-like archeon (future nucleus) engulfed by the δ -proteobacterium (future cytoplasm and endomembrane system) followed by (2nd) phagocytosis of α -proteobacterium (future mitochondria).

The evolutionary origin of genes encoding enzymes that catalyse *N*-glycan processing in the Golgi apparatus also showed a pronounced binary distribution between ps1 and ps6, indicating that glycosyltransferases present were mostly innovations of the common ancestor of all eukaryotic organisms. Unlike the case in the endoplasmic reticulum, enzymes encoding genes originating from the common ancestor of all cellular organisms are not found in archaea proteomes suggesting their bacterial origin.

Analysis of IgG *N*-glycome by two chromatography methods showed that the abundance of agalactosylated (G0) structures in IgG *N*-glycome decreases with more frequent consumption of fruit derivatives (dried fruits, fruit compotes, nuts) and sweet foods (cakes, chocolate, candies), and it is also reduced in subjects suffering from hypertension, arthritis, and in women who have a regular menstrual cycle. In contrast, the proportion of agalactosylated (G0) structures increases with obesity, blood pressure, serum urate concentration, serum glycosylated haemoglobin concentration, serum fibrinogen concentration, and serum insulin concentration. It was also observed that the proportion of monogalactosylated (G1) structures in IgG *N*-glycome decreases with increasing serum glycosylated haemoglobin concentration and increases with increasing serum albumin concentration.

The abundance of digalactosylated (G2) structures in IgG *N*-glycome decreases with increasing obesity, blood pressure, serum urate concentration, serum triglyceride concentration, serum fibrinogen concentration, serum insulin concentration, and is also lower in subjects with hypertension, arthritis. In contrast, it increases with increasing bone mineral density, HDL cholesterol concentration, the number of points achieved on the cognitive coding test and more frequent consumption of fruit derivatives (dried fruit, fruit compotes, nuts) and sweet foods (cakes, chocolate, candy, ...), and is higher in women who have a regular menstrual cycle and in subjects who take iron supplements.

The abundance of monosialylated (S1) structures in IgG *N*-glycome decreases with increasing obesity, blood pressure, serum triglyceride concentration, and serum insulin concentration. Additionally, it increases with increasing bone mineral density, the number of children, and is also higher in women who have a regular menstrual cycle.

The abundance of glycan structures with branching *N*-acetylglucosamine (B) in the IgG *N*-glycome decreases with increasing bone mineral density and is lower in women who have a regular menstrual cycle. In contrast, the proportion of structures with branching *N*-acetylglucosamine (B) in the IgG *N*-glycome increases with increasing serum cholesterol concentration, serum triglyceride concentration, and is higher in smokers.

Conclusion: Applying genomic phylostratigraphy it is shown that glycosylation found in *Homo sapiens* is ancient, originated mainly in the last common ancestor of all cellular organisms, with some innovations introduced by the common ancestor of Eukaryotes and less by the common ancestor of Metazoans. Despite its ancient history, glycosylation machinery usage enters its golden age with the development of complex multicellular organisms in order to achieve self-nonsel self recognition and as a universal language for coordination between the differentiated cells.

Analysing the *N*-glycome of IgG it is shown that environmental effects mostly affect agalactosylation, digalactosylation, sialylation and bisecting traits. *N*-glycome is correlated with obesity, blood pressure, various biochemical measurements, regularity of menstrual cycle, food intake and smoking.

KEYWORDS

N-glycome, genomic phylostratigraphy, glycosylation evolution, glycoprotein evolution, environmental effect

KRATICE

ABPI – gležanjnski indeks (*engl.* ankle-brachial pressure index)

Asn – asparagin

BMI – indeks tjelesne mase (*engl.* body mass index)

CAZy – Carbohydrate-Active enZYmes

DPANN – arhealna supergrupa koje uključuje Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, Nanohaloarchaeota, Woesearchaeota i Pacearchaeota

DNA – deoksiribonukleinska kiselina (*engl.* deoxyribonucleic acid)

Dol-P – dolikol-fosfat

EPQ-R – revidirani Eysenckov test osobnosti (*engl.* Revised Eysenck Personality Questionnaire)

ER – endoplazmatski retikulum

ECRHS – upitnik Europske zajednice o respiratornom zdravlju (*engl.* European Community Respiratory Health Survey)

ESS – Epworthova ljestvica pospanosti (*engl.* Epworth sleepiness scale)

FFQ – upitnik o učestalosti konzumacije hrane i pića (*engl.* food frequency questionnaire)

GAG glikozaminoglikan

Gal – galaktoza

GDP – gvanozin-difosfat

GHQ – opći zdravstveni upitnik (*engl.* general health questionnaire)

GL – glikolipidi

Glc – glukoza

GlcNAc *N*-acetilglukozamin

GP – glikoproteini

GPI – glikozilfosfatidinozitol

HbA1c – glicirani hemoglobin

HDL – lipoprotein visoke gustoće (*engl.* high density lipoprotein)

HGNC – *engl.* HUGO Gene Nomenclature Committee

HILIC-UPLC, UPLC – tekućinska kromatografija visoke djelotvornosti temeljena na hidrofилnim interakcijama (*engl.* hydrophilic interaction ultra-performance liquid chromatography)

IgG – imunoglobulin G

KeGG – Kyoto Encyclopedia of Genes and Genomes

KOPB – kronična opstruktivna plućna bolest

LacdiNAc – *N,N*-diacetillaktozamin

LacNAc – *N*-acetillaktozamin

LC-ESI-MS, LCMS – tekućinska kromatografija na obrnutim fazama spregnuta s elektrosprej ionizacijskom masenom spektroskopijom (*engl.* liquid chromatography coupled with electrospray ionization mass spectrometry)

LDL – lipoprotein niske gustoće (*engl.* low density lipoprotein)

Man – manaza

MAP – srednji arterijski tlak (*engl.* mean arterial pressure)

MCTQ – Münchenski upitnik o kronotipu (*engl.* Munich chronotype questionnaire)

MM – metabolizam monosaharida

NCBI – National Center for Biotechnology Information

NG – *N*-glikozilacija

OG – *O*-glikozilacija

OST – oligosahariltransferaza

ps – filostratum (*engl.* phylostratum)

PSG – proteini koji sudjeluju u procesu glikozilacije

PVG – proteini koji vežu glikane

RNA – ribonukleinska kiselina (*engl.* ribonucleic acid)

Ser – serin

TACK – superkoljeno koje uključuje Thaumarchaeota, Aigarchaeota, Crenarchaeota i Korarchaeota

Thr – treonin

UDP – uridin-difosfat

VLDL – lipoprotein vrlo niske gustoće (*engl.* very low density lipoprotein)

WAQ – upitnik Svjetske zdravstvene organizacije o angina pectoris (*engl.* WHO Rose angina questionnaire)

WCQ – upitnik Svjetske zdravstvene organizacije o intermitentnoj klaudikaciji (*engl.* WHO Questionnaire on intermittent claudication)

WHO – Svjetska zdravstvena organizacija (*engl.* World Health Organization)

Xyl – ksiloza

SADRŽAJ

1	Uvod	1
1.1	Pregled literature	1
1.2	Obrazloženje istraživanja	8
2	Metode	10
2.1	Određivanje makroevolucijske povijesti glikozilacije	10
2.2	Analiza utjecaja okolišnih čimbenika na N-glikozilaciju imunoglobulina G.....	25
3	Rezultati i rasprava	44
3.1	Makroevolucijska povijest glikozilacije.....	44
3.2	Utjecaj okolišnih čimbenika na N-glikozilaciju imunoglobulina G.....	58
4	Zaključak	87
5	Popis literature	90
6	Životopis autora s popisom objavljenih radova.....	100

1 UVOD

1.1 PREGLED LITERATURE

1.1.1 Glikani i glikozilacija

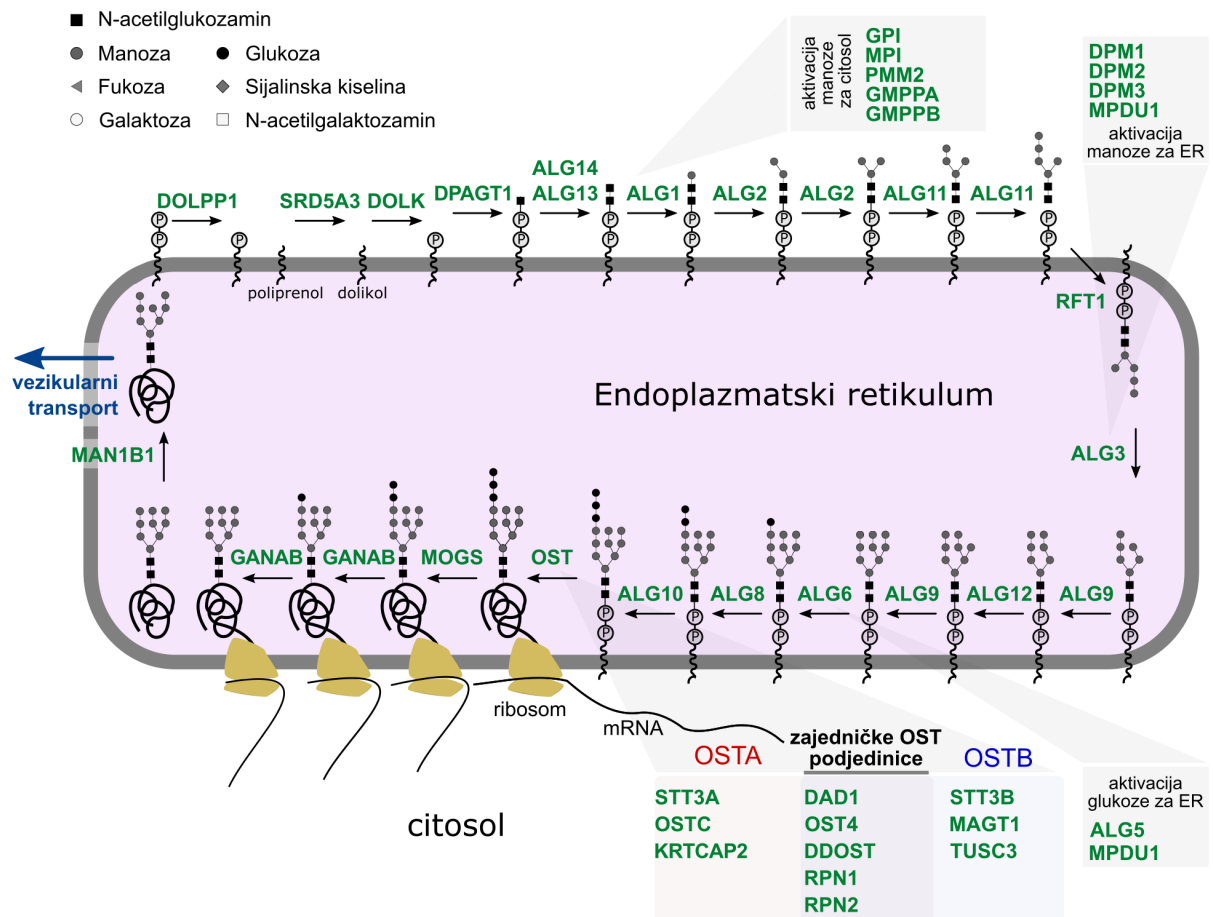
Glikozilacija je enzimatski proces kojim se ugljikohidrati (glikani) kovalentno vežu na aglikone: proteine, lipide ili RNA (1) pri čemu tvore glikoproteine, glikolipide i glikoRNA. Glikoproteini mogu nositi jedan ili više, najčešće *N*- ili *O*-vezanih glikana i najproučavaniji su glikokonjugati. Glikolipidi se sastoje od glikana koji je najčešće preko glukoze ili galaktoze vezan na hidroksilnu grupu ceramida koji se sastoji od dugolančane baze – sfingozina i masnih kiselina. Najčešće ih nalazimo na plazmatskoj membrani eukariotskih stanica gdje održavaju stabilnost membrane i posreduju interakciju između stanica, a često služe i kao vezna mjesta za viruse i druge patogene (2). Glikozilfosfatidilinozitolna sidra (GPI sidra) se sastoje od fosfatidilinozitola, glikanskog mosta i fosfoetanolamina koji veže C-terminalni dio proteina. Glavna uloga mu je sidrenje proteina u membranu, najčešće enzima, adhezijskih molekula, receptora, proteina koji reguliraju komplement ili histokompatibilnih antigena (3). Mucini su veliki glikoproteini koji nose velik broj gusto pakiranih *O*-vezanih glikana. Ovi glikoproteini su eksprimirani na epitelnim membranama gdje održavaju homeostazu i štite od fluktuirajućih uvjeta poput promjena pH, ionske jakosti, oksigenacije i hidratacije (4). Glikozaminoglikani (GAG) su linearni kopolimeri kiselih disaharidnih jedinica od kojih svaka najčešće sadrži heksozamin i heksozu ili heksuronsku kiselinu. Ponavljajuća jedinica definira glikozaminoglikane i razvrstava ih u glavne 4 skupine: heparin/heparan sulfat, kondroitin sulfat/dermatan sulfat, keratan sulfat i hijaluronan. U tijelu pronalaze različite funkcije koje se primarno temelje na njihovoj strukturnoj ulozi kao glavnih sastojaka izvanstaničnog matriksa i glikokaliksa stanica svih tkiva. Kroz interakcije s proteinima, GAG posreduju interakcije između stanica i izvanstaničnog matriksa, ili između dviju stanica, te modeliraju vezanje liganada na receptor. Glikozaminoglikani, vežući se na sržni protein, tvore proteoglikane (5). Proteini koji vežu glikane (*engl.* glycan-binding proteins) obuhvaćaju dvije skupine: lektine i proteine koji vežu sulfatirane glikozaminoglikane. Lektini su dalje klasificirani u evolucijski srodne obitelji na temelju domene za prepoznavanje glikana (*engl.* carbohydrate-recognition domain), koja prepoznaje terminalne strukture na glikanima. Proteini koji vežu glikozaminoglikane čine skupinu evolucijski nesrodnih proteina koji na temelju pozitivno nabijenih aminokiselina specifičnog rasporeda vežu karboksilne i sulfatne skupine glikozaminoglikanskih lanaca (6).

Funkcije glikana su brojne, a ugrubo se dijele na 4 skupine: (I) strukturne i modularne funkcije (npr. organizacija izvanstaničnog matriksa, osiguravanje topljivosti makromolekula, organizacija membrana, utjecaj na funkciju glikoproteina grananjem struktura), (II) ekstrinzično prepoznavanje glikana (između vrsta, npr. protutijela domaćina koja prepoznaju patogene, prepoznavanje, prikupljanje i obrada antigena, različite adhezivne molekule patogenih organizma), (III) intrinzično prepoznavanje glikana (unutar vrste, npr. unutarstanična kontrola kvalitete smatanja glikoproteina, stanična signalizacija, usmjeravanje proteina nakon sinteze, interakcije stanica s izvanstaničnim matriksom, oplodnja) i (IV) molekularna mimikrija glikana domaćina (npr. konvergentna evolucija glikana sličnih onima u domaćina, krađa glikana domaćina). Funkcije glikana su detaljnije sažete u radu Ajit Varkija (7).

1.1.1.1 N-vezani glikani

N-vezani glikani glikoproteina su najproučavaniji glikani. Njihova sinteza se ne provodi dodavanjem monosaharidnih jedinica na protein, nego su sintetizirani na lipidnom nosaču dolikol-fosfatu (Dol-P) s kojeg se kasnije prebacuju *en bloc* na protein koji još uvijek može biti u procesu translacije (8).

U ljudskim stanicama, biosinteza započinje na citosolnoj strani membrane endoplazmatskog retikuluma prebacivanjem *N*-acetilglukoamina (GlcNAc) sa UDP-GlcNAc na dolikol Dol-P tvoreći Dol-P-P-GlcNAc (katalizirano Dol-P-GlcNAc-1-fosfotransferazom koja je kodirana genom DPAGT1). Produljivanje glikana na Dol-P se nastavlja dodatkom drugog GlcNAc, koji se veže na prethodni β 1-4 vezom, također prijenosom s UDP-GlcNAc (katalizirano UDP-GlcNAc-transferazom čije su dvije podjedinice kodirane genima ALG13 i ALG14). Nakon toga, dodaje se pet manozna transferom sa GDP-Man na oligosaharid vezan na Dol-P (katalizirano hitobiosildifosfodolikol β -manoziltransferazom, α 1,3/1,6-manoziltransferazom i α 1,2-manoziltransferazom koje su kodirane genima ALG1, ALG2 i ALG11). Nakon sinteze strukture Man₅GlcNAc₂-P-P-Dol, dolazi do translokacije navedene strukture preko membrane na luminalnu stranu membrane endoplazmatskog retikuluma. Reakcija je katalizirana flipazom koja je kodirana RFT1 genom. Relokalizacijom u lumen endoplazmatskog retikuluma nastavlja se elongacija strukture manozama do ukupno devet manoznih jedinica. Reakcije su katalizirane α 1,3-manoziltransferazom, α 1,2-manoziltransferazom i α 1,6-manoziltransferazom, koje su kodirane genima ALG3, ALG9 i ALG12. Zatim, slijedi dodavanje triju glukočnih (Glc) jedinica do konačne strukture – reakcije su katalizirane α 1,3-glukočiltransferazama i α 1,2-glukočiltransferazom (kodirane genima ALG6, ALG8 i ALG10). Konačna struktura, Glc₃Man₉GlcNAc₂-P-P-Dol je evolucijski visokokonzervirana i zajednička svim eukariotima. Ovu strukturu od 14 monosaharida proteinski kompleks oligosahariltransferaza (OST) prenosi

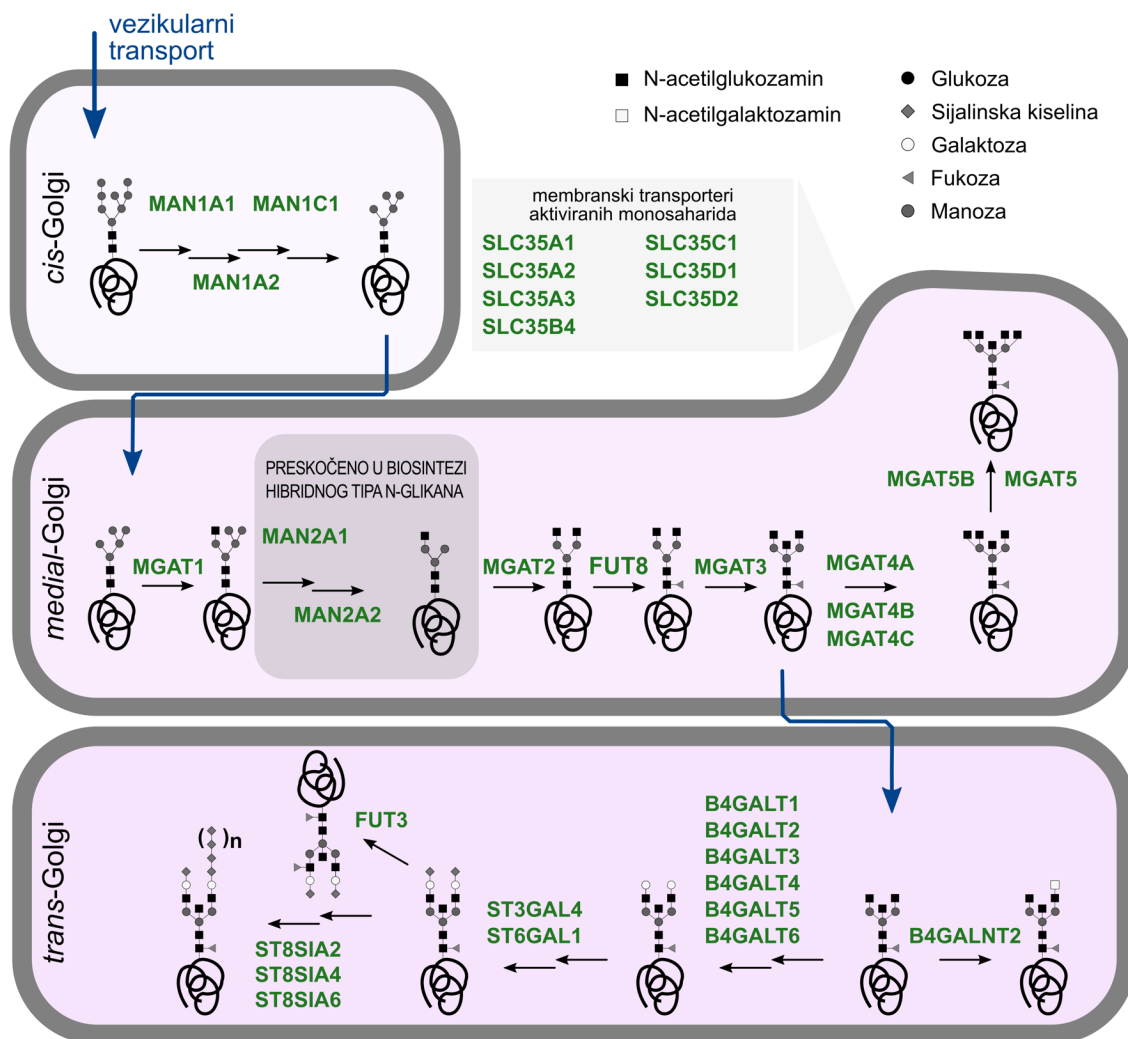


Slika 1. Biosinteza *N*-glikana čovjeka u endoplazmatskom retikulumu. Zelenom su bojom prikazani HGNC identifikatori gena koji kodiraju za enzime/podjedinice enzima koji kataliziraju pojedine reakcije.

s dolikol-fosfata na asparagin u tzv. glikozilacijskom sekvonu, Asn-X-Ser/Thr, gdje je X bilo koja aminokiselina osim prolina (Slika 1) (9). OST može biti jedna od dviju izoformi, pri čemu jedna sadrži STT3A katalitičku podjedinicu i primarno je uključena u kotranslacijsku glikozilaciju, te druga izoforma koja sadrži STT3B katalitičku podjedinicu koja je primarno uključena u post-translacijsku glikozilaciju (10).

Procesiranje *N*-glikana, dodanog na protein, započinje uklanjanjem terminalnih glukozičnih jedinica (katalizirano α -glukozidazama I i II koje su kodirane genima MOGS i GANAB). Nakon uklanjanja glukoze, uklanja se i terminalna α 1,2-Man s centralne grane (katalizirano s ER α -manozidazom I koja je kodirana genom MAN1B1). Time nastaje oligosaharidna struktura Man₈GlcNAc₂ koja vezikularnim transportom napušta endoplazmatski retikulum te nastavlja procesiranje u Golgijevom aparatu (9).

U *cis* odjeljku Golgijevog aparata dolazi do uklanjanja α 1-2 vezanih manozama katalizom α 1-2 manozidaza IA i IB (kodirane genima MAN1A1 i MAN1A2) pri čemu uklanjanje manozama ne mora biti potpuno. *N*-glikani s potpuno uklonjenim α 1-2 manozama, strukture Man₅GlcNAc₂, nastavljaju procesiranje u medijalnom dijelu Golgijevog aparata, gdje pod katalizom



Slika 2. Procesiranje *N*-glikana čovjeka u Golgijevom aparatu. Zelenom su bojom prikazani HGNC identifikatori gena koji kodiraju za enzime/podjedinice enzima koji kataliziraju pojedine reakcije.

N-acetilglukozaminiltransferaze (kodirana genom MGAT1) dolazi do dodavanja GlcNAc na C2 atom α 1-3 vezane manoze. Ta je reakcija preduvjet za uklanjanje terminalne α 1-3 i α 1-6 vezane manoze uz katalizu α -manozidaza II enzimima (koji su kodirani genima MAN2A1 i MAN2A2). Toj reakciji podliježu skoro svi glikani, no mali dio njih ipak ne – time tvoreći takozvani hibridni tip *N*-glikana. Glikanima kojima su uklonjene α 1-3 i α 1-6 vezana manozza, pod katalizom *N*-acetilglukozaminiltransferaze II (kodirane genom MGAT2) se dodaje GlcNAc na C2 atom α 1-6 vezane manoze. Dobivena je struktura prekursor za sintezu takozvanih *N*-glikana kompleksnog tipa, a dva novotransferirana *N*-acetilglukozamina će dati strukturu s dvije antene/grane. Kod kompleksnih je glikana moguće i dodatno grananje do maksimalno četiri antene, a postiže se vezivanjem GlcNAc-a na C4 atom sržne α 1-3 vezane manoze (katalizirano enzimom *N*-acetilglukozaminiltransferaza IV koji je kodiran genima MGAT4A-C), te na C6 atom sržne α 1-6 vezane manoze (katalizirano *N*-acetilglukozaminiltransferazom VI koja je kodirana genom MGAT5). Kompleksne, ali i hibridne strukture *N*-glikana također

moгу imati i „račvajući“ (*engl.* bisecting) GlcNAc vezan na β -vezanu sržnu manozu, no taj računajući GlcNAc se dodatnim procesiranjem nikada ne produljuje u antenu (Slika 2) (9).

Daljnje preinake glikana visokodiferenciraju do sada spomenutih par struktura, a uključuju daljnje produljenje antena, modifikaciju sržne strukture *N*-glikana te „ukrašavanje“ terminalnih struktura (*engl.* capping, decorating). Sve spomenute reakcije se odvijaju u *trans* odjeljku Golgijevog aparata.

Daljnje produljenje antena najčešće uključuje dodavanje galaktoze na postojeći GlcNAc. Spomenuti su monosaharidi vezani β 1-4 vezom i tvore disaharid *N*-acetilgalaktozamin (LacNAc) koji se može ponavljati nekoliko puta, rezultirajući poli-LacNAc lancem. Umjesto galaktoze, antene nekih glikoproteina se produljuju *N*-acetilgalaktozaminom, čime tvore LacdiNAc strukturu. Modifikacija sržne strukture *N*-glikana kod čovjeka podrazumijeva vezanje fukoze u hibridnim i kompleksnim tipovima struktura α 1-6 vezom na GlcNAc koji je izravno vezan na asparagin (katalizom α 1-6 fukoziltransferaze koja je kodirana genom FUT8). Najvažnije terminalno ukrašavanje uključuje dodavanje sijalinske kiseline (katalizirano sijaliltransferazama koje su kodirane genima ST3GAL4, ST6GAL1 i ST8SIA -2, -4, -6), fukoze (katalizirano fukoziltransferazom koja je kodirana genom FUT3), galaktoze (katalizirano β 1,4-galaktoziltransferazama kodiranim genima B4GALT1-6), *N*-acetilglukozamina (katalizirane β 1,4-*N*-acetilgalaktozaminiltransferazom koja je kodirana genom B4GALNT2) i sulfatnih skupina na grane kompleksnih *N*-glikana (nije prikazano na slici). Terminalne su strukture posebno važne budući da se upravo one nalaze na samoj površini glikoproteina i da one stupaju u interakciju s lektinima (proteinima koji vežu glikane) i protutijelima (Slika 2).

1.1.1.2 Razlike u glikanima između vrsta

N-glikani su prisutni i u prokariotima, kod kojih obrazac sinteze može biti vrlo sličan onome u eukariotima. Glikani se sintetiziraju na citosolnoj strani plazmatske membrane gdje se nukleotidima aktivirani monosaharidi koriste kao gradivni blokovi pri produljenju oligosaharida vezanog za lipidni nosač (11). Sintetizirani oligosaharid, vezan na lipidni nosač, se translocira preko plazmatske membrane u periplazmatski prostor, gdje se bez daljnjeg produljenja prenosi na protein s glikozilacijskim sekvonom uz pomoć oligosahariltransferaze koja katalizira *en bloc* transfer s lipidnog nosača na amidnu skupinu asparagina. Strukture koje se prebacuju s lipidnog nosača na protein pokazuju najveću raznolikost u sastavu monosaharida kod arheja s trendom smanjenja kod eukariota. Sastav monosaharida, eukarioti kompenziraju cijelim nizom dodatnih enzima u endoplazmatskom retikulumu i Golgijevom aparatu koji dodatno procesiraju *N*-vezane glikane u razgranatije strukture (11).

U različitim skupinama organizama dominiraju različite glikanske strukture koje često sadrže specifično vezane monosaharide karakteristične za određenu filogenetsku granu.

Svi kralježnjaci pokazuju donekle slične strukture *N*-vezanih glikana koje čine kompleksni glikani s dvije do četiri antene s terminalnim galaktozama i sijalinskim kiselinama. Insekti često tvore oligomanozne glikane, te skraćene glikane s malim brojem manozama, a često imaju i dvije sržne fukoze. Filamentozne gljive sintetiziraju varijante glikana s velikim brojem manozama, a među terminalnim strukturama mogu sadržavati i galaktofuranozu. Kvasci također proizvode *N*-vezane glikane s velikim brojem manozama, no taj se broj penje čak i do 200 manozama u strukturi. Biljke mogu proizvoditi hibridne, skraćene i kompleksne *N*-glikane, a često sadrže i β 1,2-vezane ksiloze i α 1,3-vezane fukoze. Najčešća struktura kompleksnog tipa ima α 1,4-vezanu fukozu na GlcNAc u Gal- β 1,3-GlcNAc, što je vrlo slično strukturi koja se pronalazi u kralježnjaka (11–13).

Bakterije i arheje tvore glikanske strukture koje ne nalikuju puno na one pronađene u eukariotima. Broj korištenih monosaharida je znatno veći (više od 700) i često specifičan za pojedinu vrstu. Literaturni su podaci relativno oskudni no čini se da postoji trend smanjenja kompleksnosti glikanskih struktura s razvojem višestaničnosti i razvojem organskih sustava (11).

Nekoliko je istraživanja do sada različitim metodama odredilo evolucijsku starost pojedinih gena koji kodiraju za proteine koji sudjeluju u *N*-glikozilaciji (14–16). Rezultati pokazuju da je prvi dio biosintetskog puta *N*-glikana mješovitog porijekla: iz zajedničkog pretka svih staničnih organizama i zajedničkog pretka svih eukariota, te da su sijaliltransferaze nastale u zajedničkom pretku svih eukariota.

O-glikozilacija se najčešće odvija na aminokiselinama serinu i treoninu, no moguća je i na tirozinu, hidroksilizinu i hidroksiprolinu. Kod kralježnjaka, postoji ukupno devet tipova *O*-glikozilacije i uglavnom su konzervirani u svim životinjama. Biosinteza započinje inicijacijom najčešće u endoplazmatskom retikulumu, osim za *O*-GlcNAc i *O*-Xyl tipove koji počinju u Golgijevom aparatu. Daljnja elongacija i dodavanje terminalnih struktura se odvija u Golgijevom aparatu (17).

Najčešća i najproučavanija *O*-glikozilacija je *O*-GlcNAc tipa koja kroz četiri najčešća tipa sržnih struktura tvori mucine koji se većinom nalaze na sekretornim proteinima i membranskim proteinima epitelnih stanica unutarnjih i vanjskih površina, posebno gastrointestinalnog i respiratornog trakta (18, 19). Glavne uloge *O*-GlcNAc glikana obuhvaćaju strukturne i modulatorne uloge, zaštitu od patogena, interakciju s receptorima i međustaničnu komunikaciju

pogotovo za vezanje i prizivanje stanica imunskog sustava, te također pruža zaštitu od endogenih hidrolizirajućih enzima.

Sržna struktura tipa 1 je osim kod kralježnjaka pronađena i kod insekata gdje također dolazi u obliku mucina koji oblažu gastrointestinalni trakt. Općenito su među svim životinjama prisutni mucini koji stvaranjem gela ovlažuju i štite epitel koji je u izravnom dodiru sa okolišem. Kod biljaka nije pronađen *O*-vezani GlcNAc, ali postoje drugi monosaharidi (arabinoza i galaktoza) koji su *O*-vezani na aminokiseline.

Bakterijska je *O*-glikozilacija oskudno istražena no uočeno je da postoji jedinstvena *en bloc* *O*-glikozilacija posredovana s OST-om, ali i sekvencijalni put sličan onome kod eukariota. Kao i kod *N*-glikana, prokarioti koriste širok spektar monosaharida rezultirajući vrlo raznolikim brojem *O*-glikana (11, 17).

Dio gena koji sudjeluje u procesiranju *N*-glikana u Golgijevom aparatu također sudjeluje i u procesiranju *O*-glikana (12, 17).

Do trenutka pisanja ovog doktorata, nisu objavljena istraživanja kojima je proučavana evolucijska starost gena koji kodiraju za proteine koji sudjeluju u *O*-glikozilaciji.

Glikozaminoglikani su nerazgranati polimeri ponavljajućih disaharidnih jedinica koje su često modificirane sulfatnim skupinama. U organizmu prekrivaju površinu stanica i tvore glavninu izvanstanične matrice. Velika strukturalna raznolikost im omogućava interakciju sa širokim spektrom bioloških molekula. Kroz te interakcije glikozaminoglikani sudjeluju u različitim biološkim procesima poput stanične adhezije, proliferacije i migracije, tvorbe izvanstaničnog matriksa, popravka tkiva, koagulacije i imunog odgovora (20).

Glikozaminoglikane nalazimo kod svih životinja, pa čak i kod spužvi koje imaju nešto drukčije glikozaminoglikane koji sadrže uroničnu kiselinu, no ne tvore uobičajena ponavljanja kondroitina i heparan sulfata. Biljke nemaju glikozaminoglikane kakve nalazimo kod životinja, ali imaju kisele pektine koji sadrže galakturonsku kiselinu. Bakterije imaju potpuno drugačije polisaharide, iako neki sojevi mogu uspješno imitirati glikozaminoglikane sisavaca (11, 21).

1.1.2 Utjecaj okoliša na *N*-glikozilaciju

Osim što se glikani razlikuju između vrsta, razlikuju se i unutar iste vrste. Analizom *N*-glikoma plazme utvrđeno je da su razine pojedinih glikanskih struktura dobrim dijelom nasljedne iako nisu izravno zapisane u genome (22, 23). Ostatak varijabilnosti *N*-glikoma plazme se pripisuje okolišnim čimbenicima, a do sada je utvrđeno da ovisi o dobi, pušačkom statusu, indeksu tjelesne mase (*engl.* body mass index, BMI), lipidnom profilu (razinama lipoproteina visoke gustoće (HDL), lipoproteina niske gustoće (LDL), ukupnog kolesterola i triglicerida), krvnom tlaku, koncentraciji glukoze natašte, prehrani i pojedinim lijekovima (24, 25). Prilikom

istraživanja *N*-glikoma plazme, nije sasvim moguće razlučiti je li došlo do promjena u glikozilaciji ili promjena u koncentraciji pojedinog proteina plazme kojem je zapravo *N*-glikom ostao nepromijenjen. Zbog toga, u svrhu proučavanja utjecaja okoliša na samu glikozilaciju, potrebno je promjene u glikomu korigirati na koncentracije pojedinih proteina plazme, ili jednostavnije, proučavati glikozilaciju na razini pojedinog proteina.

U kontekstu *N*-glikozilacije, najproučavaniji glikoprotein je imunoglobulin G. Skorašnja istraživanja pokazuju da se čak do 75% varijabilnosti za pojedina svojstva *N*-glikoma imunoglobulina G može objasniti genetikom (26). Ostatak je pod utjecajem okolišnih čimbenika, a dosadašnja su istraživanja pokazala da ovisi o različitim bolestima, dobi, pušačkom statusu, spolnim hormonima, trudnoći, profilu lipida, krvnom tlaku, jetrenim enzimima, serumskoj koncentraciji urata, ureje, inzulina, kalcija i fibrinogena (27–36).

1.1.3 Genomska filostratigrafija

Genomska je filostratigrafija statistička metoda kojom je moguće procijeniti porijeklo gena pretragom homologa u različitim vrstama. Metoda se temelji na činjenici da novi geni ne nastaju isključivo duplikacijom nego također i *de novo*, od negenskih dijelova DNA (37, 38). Novonastali geni u počeku imaju vrlo brzu stopu evolucije nakon čega vrlo vjerojatno poprime neku važnu ulogu nakon koje se više ne smiju bitno mijenjati (39, 40). Time je smisleno traženje homologa u genomima živućih organizama kojima je poznata povezanost u filogenetskom stablu života, te je uz pomoć tih informacija moguće odrediti evolucijsku starost pojedinog gena (38, 41–43).

1.2 OBRAZLOŽENJE ISTRAŽIVANJA

Različite glikanske strukture u različitim organizmima rezultat su prisutnosti/odsutnosti određenih enzima koji sudjeluju u biosintezi glikana. Očuvanost prvog dijela biosinteze *N*-vezanih glikana upućuje na njegovu izrazitu važnu ulogu u razvoju organizma, dok razvoj novih enzima sugerira razvoj novih struktura glikana, a time i novih uloga glikoproteina potencijalno ključnih u evolucijskom razvoju. Nekoliko je istraživanja različitim metodama odredilo evolucijsku starost pojedinih gena ili skupine gena koje sudjeluju u sintezi glikana, a rezultati upućuju na zaključak da je prvi dio *N*-glikozilacije proces koji je očuvan još od zajedničkog pretka svih eukariota ili čak svih staničnih organizama (14–16). Do danas nije provedeno istraživanje koje bi jednom metodom obuhvatilo evolucijski pregled svih poznatih gena koji sudjeluju u glikozilaciji.

Činjenica da glikanske strukture nisu izravno zapisane u genomu omogućava puno bržu evoluciju i veću raznolikost između različitih vrsta, ali i unutar iste vrste. Raznolikost

glikanskih struktura rezultat je razine ekspresije kompleksno i precizno kontroliranog skupa enzima i membranskih prijenosnika koji sudjeluju u biosintezi glikana. Iako je razina ekspresije tih proteina nekim dijelom epigenetski nasljedna (27, 44), ona bitno ovisi i o trenutnom stanju jedinke. Utvrđeno je da *N*-glikom plazmatskih glikoproteina ovisi o spolu, dobi i nekim okolišnim čimbenicima kao što su indeks tjelesne mase, lipidni status, pušački status ili prehrana (22, 25), a značajne su promjene uočene i kod različitih bolesti (27, 45–47). Veliki je problem što na glikom svih proteina plazme, osim strukturnih promjena u glikanima, utječe i promjena količina pojedinih proteina u plazmi, te je zbog toga strukturne promjene u glikozilaciji ispravnije pratiti proučavajući samo jedan protein. Jedan od najistraživanijih glikoproteina je imunoglobulin G (IgG), koji sadrži *N*-vezane glikane na konzerviranim asparaginima smještenim u Fc regijama teškog lanca (48). Ti glikani pokazuju značajnu varijabilnost, a budući da su odgovorni za vezanje na Fc γ receptore, ovisno o svojoj strukturi dovode do različitih efektorskih funkcija. Primjerice, uklanjanje sržne α 1-6 vezane fukoze dovodi do pojačane citotoksične aktivnosti ovisne o protutijelima, dok glikoforme bez galaktoze aktiviraju komplement preko manoza-vezujućeg lektina, a glikoforme sa sijalinskom kiselinom djeluju protuupalno (27, 49–51). Spomenute spoznaje su važne za poboljšanje ciljanih monoklonskih protutijela kao terapeutika (52).

Sukladno tome, u ovom istraživanju su postavljene dvije osnovne hipoteze:

1. Glikozilacija je većinom konzervirani proces koji se manjim dijelom nadograđivao kroz evolucijski razvoj čovjeka i omogućio intenzivniji razvoj glikoproteina u evolucijskim periodima kada su glikani bili ključni za evolucijski razvoj.
2. Kod čovjeka, razlike u glikanima Fc regije imunoglobulina G u različitoj mjeri ovise o pojedinim okolišnim čimbenicima.

Specifični ciljevi su:

1. Genomskom filostratigrafijom odrediti evolucijsku starost gena koji kodiraju proteine uključene u biosintezu glikana te gena koji kodiraju glikoproteine.
2. Statističkim modeliranjem odrediti utjecaj pojedinih okolišnih čimbenika dostupnih u bazi „10 001 Dalmatinac“ na *N*-glikozilaciju imunoglobulina G.

2 METODE

2.1 ODREĐIVANJE MAKROEVOLUCIJSKE POVIJESTI GLIKOZILACIJE

U svrhu određivanja makroevolucijske povijesti glikozilacije primijenjena je metoda genomske filostratigrafije (*engl.* genomic phylostratigraphy).

2.1.1 Uzorak

2.1.1.1 Proteini koji sudjeluju u procesu glikozilacije

Popis gena koji za kodiraju proteine koji sudjeluju u procesu glikozilacije je prikupljen iz različitih izvora: dopunska tablica znanstvenog rada Moremen i sur. (53), procesi biosinteze glikana vrste *Homo sapiens* dostupni u bazi podataka „KeGG“ (hsa00510, hsa00511, hsa00512, hsa00514, hsa00515, hsa00531, hsa00532, hsa00533, hsa00534, hsa00563, hsa00601, hsa00603, hsa00604) (54), unosi s definiranom vrstom *Homo sapiens* dostupni u bazi podataka „CAZy“ (55), knjiga izvršnog urednika Ajita Varkija: „Essentials of Glycobiology“ (56) i dopunska tablica znanstvenog rada Schjoldager i sur. (57).

Dopunska tablica rada Moremen i sur. (53) sadrži listu mišjih gena čiji su ortolozi kod čovjeka pronađeni pretraživanjem baze podataka „NCBI Gene“, ručno validirani i korišteni u ovoj analizi.

Nakon uklanjanja duplikata i pseudogena, ukupna je lista sadržavala 673 gena koji kodiraju za proteine koji sudjeluju u procesu glikozilacije s jedinstvenim „Ensembl gene ID“ identifikatorom.

Za svaki je gen ručno validirana povezanost s procesom glikozilacije pretraživanjem baze „UniProtKB/Swiss-Prot“ (58) pri čemu mu je pripisana jedna ili više oznaka: *N*-glikozilacija (NG), *O*-glikozilacija (OG), glikolipidi (GL), glikozilfosfatidilinozitolna sidra (GPI), glikozaminoglikani (GAG), proteini koji vežu glikane (PVG), metabolizam monosaharida (MM). Razdioba broja gena s pojedinom kombinacijom oznaka je prikazana *UpSet* dijagramom (59).

2.1.1.2 Glikoproteini

Popis gena koji kodiraju za polipeptidne lance glikoproteina su prikupljeni pretraživanjem baze podataka „UniProtKB/Swiss-Prot“ (58). Baza podataka je pretražena korištenjem upita: *annotation:(type:carbohyd) AND reviewed:yes AND organism:"Homo sapiens (Human) [9606]"*

Dobiveni su svi ljudski proteini koji na sebi nose kovalentno vezan glikan, pri čemu je taj unos u bazi podataka recenzirala neka osoba, a ne računalo.

Konačna je lista sadržavala 4565 gena koji kodiraju za polipeptidne lance glikoproteina s jedinstvenim „Ensembl gene ID“ identifikatorom.

2.1.2 Filogenija

Odabrana je filogenija temelj za određivanje evolucijske starosti pojedinih gena. Odabrana su ukupno 503 organizma koja su raslojena u 29 filostratumata – od zajedničkog pretka svih staničnih organizama (*engl. cellular organisms, ps1*), pa sve do čovjeka (*lat. Homo sapiens, ps29*) (Slika 3).

Svaki filostratum obilježava razvoj pojedinih karakteristika u evoluciji čovjeka, pri čemu su birani organizmi sa sekvenciranim genomom i pouzdano određenim položajem u evolucijskom stablu života na temelju recentne literature (60–63).

Potpuni je popis vrsta naveden u tablici (Tablica 1).

2.1.3 Baza podataka

Iako se metoda genomska filostratigrafija primjenjuje u svrhu određivanja evolucijske starosti gena, baza podataka koja se koristi u filostratigrafskoj analizi je sastavljena od proteinskih sekvenci. Proteomi organizama prisutnih u filogeniji su preuzeti iz baze podataka „Ensembl“ (64) ili „NCBI genome“ (65) u slučaju da nisu bili prisutni u bazi podataka „Ensembl“. Iznimno, proteomi organizma skupine Choenoflagellida su preuzeti iz dopunskih materijala znanstvenog rada Richter i sur. (63).

U svakom je proteomu izdvojena najdulja varijanta alternativnog prekrajanja za pojedini gen, a dobivene su sekvence korištene za izradu baze podataka koja se primjenjivala u filostratigrafskoj analizi.

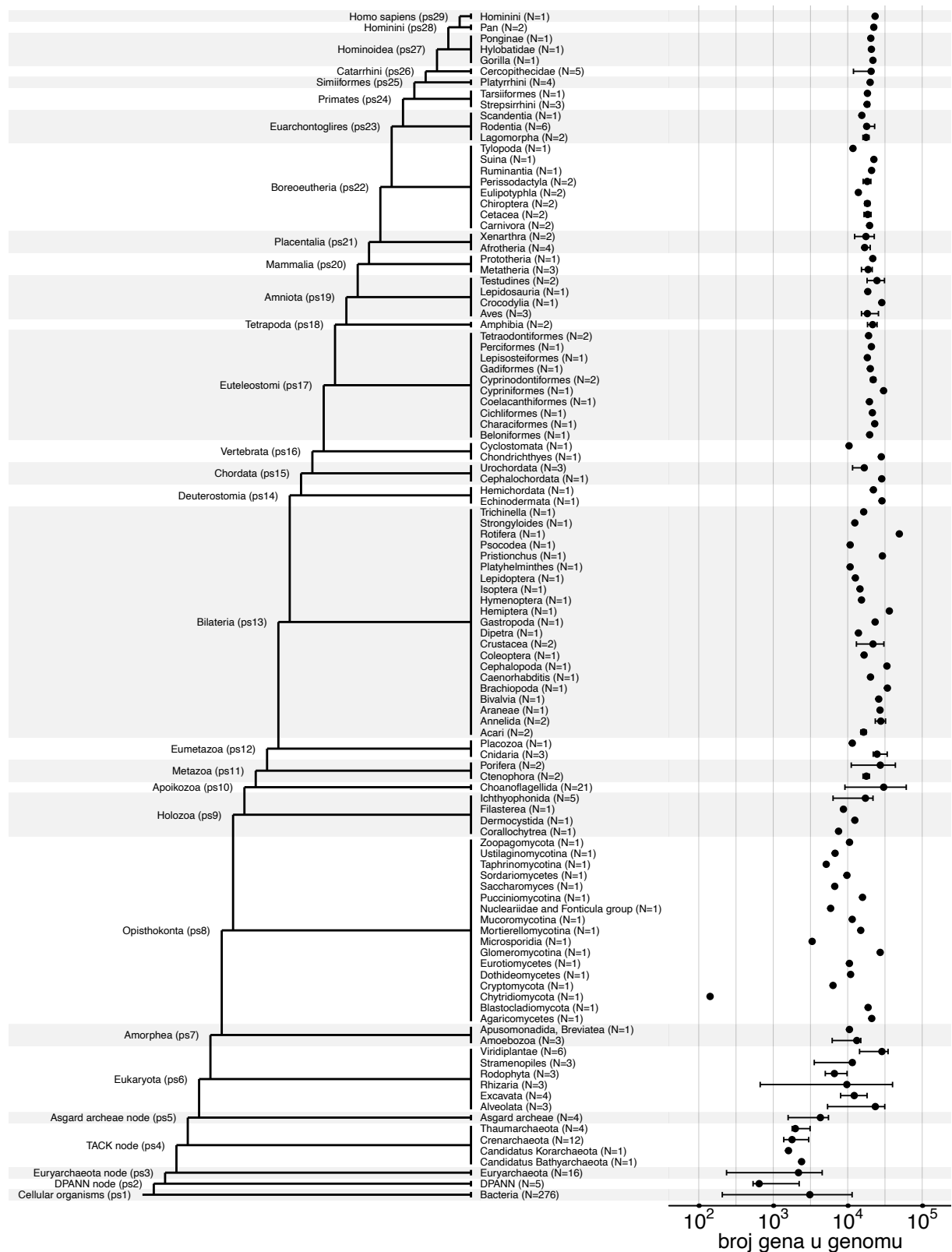
2.1.4 Filostratigrafska analiza

Za svaku najdulju varijantu proteina kodiranu pojedinim ljudskim genom pokrenut je „blastp“ algoritam (66) u svrhu pronalaska proteina slične sekvence u bazi podataka (dobivenoj prema poglavlju 2.1.3), uz graničnu e vrijednost od 10^{10} .

Dobiveni su rezultati, zajedno s filogenijom (Slika 3), korišteni za procjenu evolucijskog porijekla pojedinog gena. Ova je procjena definirana kao najmlađi filogenetski čvor (filostratum) koji sadrži zajedničkog pretka čovjeka i organizma u čijem je proteomu pronađen blastp rezultat s e vrijednosti manjom od 10^{-3} .

Razdiobe relativnih frekvencija gena koji kodiraju za proteine koji sudjeluju u procesu glikozilacije, odnosno gena koji kodiraju za polipeptidne lance glikoproteina po filostratumima su uspoređene s razdiobom relativnih frekvencija svih ljudskih gena po filostratumima (očekivana razdioba). Razlika u relativnim frekvencijama za svaki filostratum je testirana

dvostranim hipergeometrijskim testom. Rezultat je prikazan omjerom šansi (*engl.* odds ratio, *OR*), a izračunate p vrijednosti su prilagođene višestrukom testiranju prema Benjamini-Hochberg metodi. Budući da je odabir e vrijednosti podložan pristranosti, postupak nakon prikupljanja rezultata dobivenih „blastp“ algoritmom je ponovljen i za granične e vrijednosti: 1, 10^{-1} , 10^{-2} , 10^{-5} , 10^{-10} , 10^{-15} i 10^{-20} .



Slika 3. Filogenija korištena u genomskoj filostratigrafiji. Lijevo od dendrograma je popis naziv filostratuma i pripadajući redni broj (uz prefiks "ps"), desno od dendrograma je popis sestrinskih grupa koje, za pojedini filostratum, dijele zajedničkog pretka sa organizmima sljedećeg filostratuma. Broj u zagradi (prefiks "N=") označava broj organizama u skupini. U desnom dijelu slike prikazan je broj gena u genomima organizama pojedine skupine. Linija odstupanja se proteže od minimuma do maksimuma, a točka označava medijan. DPANN – arhealna supergrupa koje uključuje Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, Nanohaloarchaeota, Woearchaeota i Pacearchaeota; TACK – superkoljeno koje uključuje Thaumarchaeota, Aigarchaeota, Crenarchaeota i Korarchaeota.

Tablica 1. Popis organizama prisutnih u filogeniji korištenoj u genomskoj filostratigrafiji. DPANN – arhealna supergrupa koje uključuje Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota, Nanohaloarchaeota, Woesearchaeota i Pacearchaeota; TACK – superkoljeno koje uključuje Thaumarchaeota, Aigarchaeota, Crenarchaeota i Korarchaeota.

Cellular organisms (ps1)

Bacteria

Acidithrix ferrooxidans, *Acidobacteria bacterium Mor1*, *Acidobacterium capsulatum* ATCC 51196, *Acidovorax delafieldii* 2AN, *Acinetobacter baumannii* AYE, *Actinobacillus pleuropneumoniae* serovar 5b str. L20, *Advenella mimigardefordensis* DPN7, *Aeromonas hydrophila* subsp. *hydrophila* ATCC 7966, *Aggregatibacter actinomycetemcomitans* D11S-1, *Agrobacterium fabrum* str. C58, *Alloprevotella rava* F0323, *alpha proteobacterium* BAL199, *Anaerobaculum hydrogeniformans* ATCC BAA-1850, *Anaerolineaceae bacterium oral taxon 439*, *Anaplasma phagocytophilum* str. HZ, *Aquifex aeolicus* VF5, *Arcticibacter svalbardensis* MN12-7, *Arthrobacter enclensis*, *Arthrospira platensis* C1, *Asanoa ferruginea*, *Azotobacter vinelandii* DJ, *Bacillus anthracis*, *Bacillus subtilis* subsp. *subtilis* str. 168, *Bacillus subtilis* subsp. *subtilis* str. NCIB 3610, *bacterium* UASB270, *Bacteroides thetaiotaomicron* VPI-5482, *Balneola* sp. EhC07, *Bartonella henselae* str. Houston-1, *Bellilinea caldifistulae*, *Bifidobacterium longum* NCC2705, *Blastopirellula marina* DSM 3645, *Bordetella pertussis* Tohama I, *Borrelia burgdorferi* B31, *Borrelia parkeri* SLO, *Brachyspira suanatina*, *Bradyrhizobium diazoefficiens* USDA 110, *Brevibacillus parabrevis*, *Brucella abortus* bv. 1 str. 9-941, *Buchnera aphidicola* str. APS (*Acyrtosiphon pisum*), *Burkholderia pseudomallei* 1710b, *Caldicellulosiruptor hydrothermalis* 108, *Caldilinea aerophila* DSM 14535 = NBRC 104270, *Caldimicrobium thiodismutans*, *Caldithrix abyssi* DSM 13497, *Campylobacter jejuni* subsp. *jejuni* NCTC 11168 = ATCC 700819, *candidate division TM7 genomsp. GTL1*, *Candidatus Atelocyanobacterium thalassa* isolate ALOHA, *Candidatus Endomicrobium trichonymphae*, *Candidatus Izimaplasma* sp. HR1, *Candidatus Koribacter versatilis* Ellin345, *Candidatus Magnetobacterium bavaricum*, *Candidatus Melainabacteria bacterium* MEL.A1, *Candidatus Nomurabacteria bacterium* GW2011_GWA1_35_8, *Candidatus Peribacteria bacterium* GW2011_GWB1_54_5, *Candidatus Solibacter usitatus* Ellin6076, *Candidatus Synechococcus spongiarum* SP3, *Caulobacter crescentus* CB15, *Cecembia lonarensis* LW9, *Cephaloticoccus primus*, *Cesiribacter andamanensis* AMV16, *Cetobacterium somerae* ATCC BAA-474, *Chitinispirillum alkaliphilum*, *Chitinivibrio alkaliphilus* ACh1, *Chlamydia trachomatis* D/UW-3/CX, *Chlamydophila pneumoniae* AR39, *Chloracidobacterium thermophilum* B, *Chlorobaculum limnaeum*, *Chlorobium ferrooxidans* DSM 13031, *Chlorobium tepidum* TLS, *Chloroflexus aggregans* DSM 9485, *Chloroflexus aurantiacus* J-10-fl, *Chloroherpeton thalassium* ATCC 35110, *Chondromyces apiculatus* DSM 436, *Chromobacterium piscinae*, *Chthoniobacter flavus* Ellin428, *Chthonomonas calidirosea*, *Citrobacter freundii* 4_7_47CFAA, *Cloacibacillus porcorum*, *Clostridioides difficile* 630, *Clostridium botulinum* A str. Hall, *Coleofasciculus chthonoplastes* PCC 7420, *Corynebacterium glutamicum* ATCC 13032, *Coxiella burnetii* RSA 493, *Cyanothece* sp. PCC 8801, *Cylindrospermopsis* sp. CR12, *Cystobacter fuscus* DSM 2262,

Deferribacter desulfuricans SSM1, *Dehalogenimonas lykanthroporepellens* BL-DC-9, *Deinococcus radiodurans* R1, *Deinococcus* sp. RL, *Denitrovibrio acetiphilus* DSM 12809, *Desulfomicrobium baculatum* DSM 4028, *Desulfotalea psychrophila* LSv54, *Desulfovibrio vulgaris* str. Hildenborough, *Desulfurella acetivorans* A63, *Dethiosulfatarculus sandiegensis*, *Dialister microaerophilus* UPII 345-E, *Dictyoglomus thermophilum* H-6-12, *Dictyoglomus turgidum* DSM 6724, *Dietzia maris*, *Eggerthella* sp. YY7918, *Eggerthia catenaformis* OT 569 = DSM 20559, *Elusimicrobium minutum* Pei191, *Enterobacter cloacae* subsp. *cloacae* ATCC 13047, *Enterococcus faecalis* V583, *Erysipelatoclostridium ramosum* DSM 1402, *Erysipelothrix* sp. LV19, *Escherichia coli* str. K-12 substr. MG1655, *Fibrobacter succinogenes* subsp. *succinogenes* S85, *Fictibacillus arsenicus*, *Fimbriimonas ginsengisoli* Gsoil 348, *Finegoldia magna*, *Flavobacterium psychrophilum* JIP02/86, *Flexilinea flocculi*, *Francisella tularensis* subsp. *tularensis* SCHU S4, *Fretibacterium fastidiosum*, *Fusobacterium equinum*, *Fusobacterium nucleatum* subsp. *nucleatum* ATCC 25586, *Gardnerella vaginalis* 0288E, *Gemmatimonas aurantiaca* T-27, *Gemmatimonas phototrophica*, *Gemmatirosa kalamazooensis*, *Geobacter sulfurreducens* PCA, *Gloeobacter kilauensis* JS1, *Gloeobacter violaceus* PCC 7421, *Gordonia bronchialis* DSM 43247, *Gordonia otitidis* NBRC 100426, *Granulicella mallensis* MP5ACTX8, *Granulicella tundricola* MP5ACTX9, *Haemophilus influenzae* Rd KW20, *Helcococcus kunzii* ATCC 51366, *Helicobacter pylori* 26695, *Holdemania filiformis* DSM 12042, *Hydrogenivirga* sp. 128-5-R1-1, *Hydrogenobacter thermophilus* TK-6, *Ignavibacterium album* JCM 16511, *Ilumatobacter coccineus* YM16-304, *Ilyobacter polytropus* DSM 2926, *Jonquetella* sp. BV3C21, *Klebsiella pneumoniae* subsp. *pneumoniae* MGH 78578, *Kouleothrix aurantiaca*, *Ktedonobacter racemifer* DSM 44963, *Lactobacillus fabifermentans* T30PCM01, *Lactobacillus plantarum* WCFS1, *Lactococcus lactis* subsp. *lactis* III403, *Lawsonia intracellularis* N343, *Legionella pneumophila* str. Paris, *Lentisphaera araneosa* HTCC2155, *Leptospira interrogans* serovar Lai str. 56601, *Leptospira* sp. Fiocruz LV3954, *Leptospirillum ferriphilum*, *Leptotrichia goodfellowii* F0264, *Leuconostoc mesenteroides* subsp. *mesenteroides* ATCC 8293, *Limnochorda pilosa*, *Listeria floridensis* FSL S10-1187, *Listeria monocytogenes* EGD-e, *Lyngbya aestuarii* BL J, *Lyngbya confervoides* BDUI41951, *Lysinibacillus sphaericus* C3-41, *Magnetococcus marinus* MC-1, *Mannheimia haemolytica* serotype A2 str. OVINE, *Megasphaera micronuciformis* F0359, *Meiothermus silvanus* DSM 9946, *Melioribacter roseus* P3M-2, *Mesoplasma florum* L1, *Methylacidiphilum fumariolicum* SolV, *Methylobacillus flagellatus* KT, *Micrococcus luteus* NCTC 2665, *Microcystis aeruginosa* NIES-843, *Microlunatus phosphovorus* NM-1, *Microscilla marina* ATCC 23134, *Moorella thermoacetica* ATCC 39073, *Moraxella catarrhalis* 7169, *Mucispirillum schaedleri* ASF457, *Mycobacterium tuberculosis* H37Rv, *Mycoplasma pneumoniae* M129, *Myxococcus xanthus* DK 1622, *Neisseria meningitidis* Z2491, *Nitrolancea hollandica* Lb, *Nitrospina gracilis* 3/211, *Nitrospina* sp. SCGC_AAA799_A02, *Nitrospira* sp. SCGC AG-212-E16, *Nitrospirae bacterium* HCH-1, *Nocardioideae bacterium* Broad-1, *Nodularia spumigena* CCY9414, *Nostoc punctiforme* PCC 73102, *Olsenella profusa* F0195, *Opiritae bacterium* SCGC AG-212-L18, *Oscillochloris trichoides*

DG-6, *Oxalobacteraceae* bacterium IMCC9480, *Paenibacillus alvei* TS-15, *Paracoccus denitrificans* PD1222, *Pasteurella multocida* subsp. *multocida* str. Pm70, *Patulibacter medicamentivorans*, *Persephonella marina* EX-H1, *Phaeodactylibacter xiamenensis*, *Pirellula staleyi* DSM 6068, *Planktothricoides* sp. SR001, *Porphyromonas gingivalis* W83, *Prevotella intermedia* 17, *Prochlorococcus marinus* subsp. *marinus* str. CCMP1375, *Propionibacterium acnes* KPA171202, *Propionispora* sp. 2/2-37, *Proteus mirabilis* HI4320, *Pseudomonas aeruginosa* MPAO1/P2, *Pseudoramibacter alactolyticus* ATCC 23263, *Pseudothertomoga lettingae* TMO, *Pyramidobacter piscolens* W5455, *Ralstonia solanacearum* GMI1000, *Rhizobium leguminosarum* bv. *viciae* 3841, *Rhodobacter sphaeroides* 2.4.1, *Rhodopirellula baltica* SH 1, *Rhodospirillum rubrum* ATCC 11170, *Rhodovulum* sp. PH10, *Rickettsia prowazekii* str. Madrid E, *Rubidibacter lacunae* KORDI 51-2, *Rubrobacter xylanophilus* DSM 9941, *Ruminococcus bicirculans*, *Salinibacter ruber* DSM 13855, *Salinispira pacifica*, *Salmonella enterica* subsp. *enterica* serovar *Typhimurium* str. LT2, *Sebaldella termitidis* ATCC 33386, *Sediminispirochaeta smaragdinae* DSM 11293, *Shewanella oneidensis* MR-1, *Shigella dysenteriae* Sd197, *Sinorhizobium meliloti* 1021, *Smithella* sp. ME-1, *Sphaerobacter thermophilus* DSM 20745, *Sphaerochaeta pleomorpha* str. Grapes, *Sphingobacterium* sp. PM2-P1-29, *Spirochaeta lutea*, *Spiroplasma litorale*, *Staphylococcus aureus* subsp. *aureus* N315, *Stenotrophomonas maltophilia* K279a, *Streptobacillus moniliformis* DSM 12112, *Streptococcus pneumoniae* TIGR4, *Streptomyces coelicolor* A3(2), *Sulfuricella* sp. T08, *Sulfurihydrogenibium yellowstonense* SS-5, *Sutterella parvirubra* YIT 11816, *Synechocystis* sp. PCC 6803, *Terribacillus aindingensis*, *Terriglobus saanensis* SPIPR4, *Thauera phenylacetica* B4P, *Thermanaerovibrio acidaminovorans* DSM 6589, *Thermoanaerobaculum aquaticum*, *Thermocrinis albus* DSM 14484, *Thermodesulfatator autotrophicus*, *Thermodesulfatator indicus* DSM 15286, *Thermodesulfobacterium commune* DSM 2178, *Thermodesulfobacterium geofontis* OPF15, *Thermodesulfovibrio yellowstonii* DSM 11347, *Thermomicrobium roseum* DSM 5159, *Thermosiphon melanesiensis* BI429, *Thermosulfidibacter takaii* ABI70S6, *Thermosulfurimonas dismutans*, *Thermosynechococcus elongatus* BP-1, *Thermotoga maritima* MSB8, *Thermotoga* sp. RQ7, *Thermus* sp. NMX2.A1, *Thermus thermophilus* HB8, *Tissierella bacterium* S5-A11, *Tolypothrix bouteillei* VB521301, *Treponema pallidum* subsp. *pallidum* str. Nichols, *Truepera radiovictrix* DSM 17093, *Turicibacter* sp. HGF1, *Turneriella parva* DSM 21527, *Ureaplasma parvum* serovar 3 str. ATCC 700970, *Verrucomicrobiaceae* bacterium SCGC AG-212-N21, *Vibrio cholerae* O1 biovar El Tor str. N16961, *Vibrio fischeri* ES114, *Wolbachia* endosymbiont of *Drosophila melanogaster*, *Xanthomonas campestris* pv. *campestris* str. ATCC 33913, *Xenococcus* sp. PCC 7305, *Yersinia pestis* biovar *Microtus* str. 91001, *Yonghaparkia* sp. Soil809

DPANN node (ps2)

DPANN

Candidatus Haloredivivus sp. G17, *Candidatus Nanopusillus acidilobi*,
Candidatus Nanopusillus sp. Nst1, *Nanoarchaeum equitans* Kin4-M,
Nanohaloarchaea archaeon SG9

Euryarchaeota node (ps3)

Euryarchaeota

Archaeoglobus fulgidus DSM 4304, *Haloarcula marismortui* ATCC 43049, *Halobacterium salinarum* R1, *Haloferax volcanii* DS2, *Methanobacterium formicicum* DSM 3637, *Methanobrevibacter smithii* ATCC 35061, *Methanocaldococcus jannaschii* DSM 2661, *Methanococcus maripaludis* S2, *Methanopyrus kandleri* AV19, *Methanosarcina acetivorans* C2A, *Methanospirillum hungatei* JF-1, *Methanothermobacter thermoautotrophicus* str. Delta H, *Natronomonas pharaonis* DSM 2160, *Pyrococcus horikoshii* OT3, *Thermococcus kodakarensis* KOD1, *Thermoplasma acidophilum* DSM 1728

TACK node (ps4)
Candidatus Bathyarchaeota <i>Candidatus Bathyarchaeota archaeon BAI</i>
Candidatus Korarchaeota <i>Candidatus Korarchaeum cryptofilum OPF8</i>
Crenarchaeota <i>Thermococcus kodakarensis</i> KOD1, <i>Thermoplasma acidophilum</i> DSM 1728, <i>Candidatus Bathyarchaeota archaeon BAI</i> , <i>Candidatus Korarchaeum cryptofilum OPF8</i> , <i>Acidilobus saccharovorans</i> 345-15, <i>Aeropyrum pernix</i> K1, <i>Caldivirga maquilingensis</i> IC-167, <i>Fervidicoccus fontis</i> Kam940, <i>Hyperthermus butylicus</i> DSM 5456, <i>Ignicoccus islandicus</i> DSM 13165, <i>Pyrobaculum aerophilum</i> str. IM2, <i>Staphylothermus marinus</i> F1, <i>Sulfolobus solfataricus</i> P2, <i>Thermofilum pendens</i> Hrk 5, <i>Thermoproteus tenax</i> Kra 1, <i>Vulcanisaeta moutnovskia</i> 768-28
Thaumarchaeota <i>Cenarchaeum symbiosum</i> A, <i>Nitrosopumilus maritimus</i> SCM1, <i>Nitrososphaera viennensis</i> EN76, <i>Thaumarchaeota archaeon N4</i>
Asgard archeae node (ps5)
Asgard archeae <i>Candidatus Heimdallarchaeota archaeon LC_3</i> , <i>Candidatus Odinararchaeota archaeon LCB_4</i> , <i>Candidatus Thorarchaeota archaeon SMTZ1-45</i> , <i>Lokiarchaeum</i> sp. GC14 75
Eukaryota (ps6)
Alveolata <i>Perkinsus marinus</i> ATCC 50983, <i>Plasmodium falciparum</i> 3D7, <i>Stentor coeruleus</i>
Excavata <i>Bodo saltans</i> , <i>Leishmania donovani</i> BPK282A1, <i>Naegleria gruberi</i> , <i>Trypanosoma brucei</i>
Rhizaria <i>Lotharella oceanica</i> , <i>Plasmodiophora brassicae</i> , <i>Reticulomyxa filosa</i>
Rodophyta <i>Chondrus crispus</i> , <i>Cyanidioschyzon merolae</i> , <i>Galdieria sulphuraria</i>
Stramenopiles <i>Aureococcus anophagefferens</i> , <i>Nannochloropsis gaditana</i> CCMP526, <i>Phaeodactylum tricornutum</i>
Viridiplantae <i>Amborella trichopoda</i> , <i>Arabidopsis thaliana</i> , <i>Chlamydomonas reinhardtii</i> , <i>Physcomitrella patens</i> , <i>Selaginella moellendorffii</i> , <i>Vitis vinifera</i>
Amorphea (ps7)

Amoebozoa	<i>Acanthamoeba castellanii</i> str. Neff, <i>Dictyostelium discoideum</i> , <i>Entamoeba nuttalli</i> P19
Apusomonadida, Breviatea	<i>Thecamonas trahens</i> ATCC 50062
Opisthokonta (ps8)	
Agaricomycetes	<i>Laccaria amethystina</i> LaAM-08-1
Blastocladiomycota	<i>Allomyces macrogynus</i> ATCC 38327
Chytridiomycota	<i>Batrachochytrium dendrobatidis</i> JEL423
Cryptomycota	<i>Rozella allomycis</i> CSF55
Dothideomycetes	<i>Zymoseptoria tritici</i> IPO323
Eurotiomycetes	<i>Aspergillus nidulans</i>
Glomeromycotina	<i>Rhizophagus irregularis</i> DAOM 197198w (GCA_000597685)
Microsporidia	<i>Mitosporidium daphniae</i>
Mortierellomycotina	<i>Mortierella elongata</i> AG-77
Mucoromycotina	<i>Rhizopus microsporus</i> var. <i>microsporus</i>
Nucleariidae and Fonticula group	<i>Fonticula alba</i> (GCA_000388065)
Pucciniomycotina	<i>Puccinia graminis</i>
Saccharomyces	<i>Saccharomyces cerevisiae</i>
Sordariomycetes	<i>Neurospora crassa</i>
Taphrinomycotina	<i>Schizosaccharomyces pombe</i>
Ustilaginomycotina	<i>Ustilago maydis</i>
Zoopagomycota	<i>Conidiobolus coronatus</i> NRRL 28638
Holozoa (ps9)	
Corallochytreia	<i>Corallochytrium limacisporum</i>
Dermocystida	<i>Chromosphaera perkinsii</i>
Filasterea	<i>Capsaspora owczarzaki</i> ATCC 30864
Ichthyophonida	

Creolimax fragrantissima, *Sphaeroforma arctica* JP610, *Ichthyophonus hoferi*, *Abeoforma whisleri*, *Pirum gemmata*

Apoikozoa (ps10)

Choanoflagellida

Acanthoeca spectabilis, *Choanoeca perplexa*, *Codosiga hollandica*,
Diaphanoeca grandis, *Didymoeca costata*, *Hartaetosiga balthica*,
Hartaetosiga gracilis, *Helgoeca nana*, *Microstomoeca roanoka*, *Monosiga*
brevicollis MX1, *Mylnosiga fluctuans*, *Salpingoeca dolichothecata*,
Salpingoeca helianthica, *Salpingoeca infusionum*, *Salpingoeca kvevrii*,
Salpingoeca macrocollata, *Salpingoeca punica*, *Salpingoeca rosetta*,
Salpingoeca urceolata, *Savillea parva*, *Stephanoeca diplocostata* Australia

Metazoa (ps11)

Ctenophora

Mnemiopsis leidyi, *Pleurobrachia bachei*

Porifera

Amphimedon queenslandica, *Oscarella pearsei*

Eumetazoa (ps12)

Cnidaria

Acropora digitifera, *Hydra vulgaris*, *Nematostella vectensis*

Placozoa

Trichoplax adhaerens

Bilateria (ps13)

Acari

Strigamia maritima, *Tetranychus urticae*

Annelida

Capitella teleta, *Helobdella robusta*

Araneae

Stegodyphus mimosarum

Bivalvia

Crassostrea gigas

Brachiopoda

Lingula anatina

Caenorhabditis

Caenorhabditis elegans

Cephalopoda

Octopus bimaculoides

Coleoptera

Tribolium castaneum

Crustacea

Daphnia pulex, *Lepeophtheirus salmonis*

Dipetra

Drosophila melanogaster

Gastropoda

Lottia gigantea

Hemiptera

Acyrtosiphon pisum

Hymenoptera

Apis mellifera

Isoptera

<i>Zootermopsis nevadensis</i>
Lepidoptera
<i>Heliconius melpomene</i>
Platyhelminthes
<i>Schistosoma mansoni</i>
Pristionchus
<i>Pristionchus pacificus</i>
Psocodea
<i>Pediculus humanus corporis</i>
Rotifera
<i>Adineta vaga</i>
Strongyloides
<i>Strongyloides ratti</i>
Trichinella
<i>Trichinella spiralis</i>
Deuterostomia (ps14)
Echinodermata
<i>Strongylocentrotus purpuratus</i>
Hemichordata
<i>Saccoglossus kowalevskii</i>
Chordata (ps15)
Cephalochordata
<i>Branchiostoma floridae</i>
Urochordata
<i>Ciona intestinalis, Ciona savignyi, Oikopleura dioica</i>
Vertebrata (ps16)
Cyclostomata
<i>Petromyzon marinus</i>
Chondrichthyes
<i>Callorhynchus milii</i>
Euteleostomi (ps17)
Beloniformes
<i>Oryzias latipes</i>
Characiformes
<i>Astyanax mexicanus</i>
Cichliformes
<i>Oreochromis niloticus</i>
Coelacanthiformes
<i>Latimeria chalumnae</i>
Cypriniformes
<i>Danio rerio</i>
Cyprinodontiformes
<i>Poecilia formosa, Xiphophorus maculatus</i>
Gadiformes
<i>Gadus morhua</i>
Lepisosteiformes
<i>Lepisosteus oculatus</i>
Perciformes
<i>Gasterosteus aculeatus</i>

Tetraodontiformes	<i>Takifugu rubripes, Tetraodon nigroviridis</i>
Tetrapoda (ps18)	
Amphibia	<i>Nanorana parkeri, Xenopus tropicalis</i>
Amniota (ps19)	
Aves	<i>Ficedula albicollis, Gallus gallus, Melopsittacus undulatus</i>
Crocodylia	<i>Crocodylus porosus</i>
Lepidosauria	<i>Anolis carolinensis</i>
Testudines	<i>Chrysemys picta, Pelodiscus sinensis</i>
Mammalia (ps20)	
Metatheria	<i>Monodelphis domestica, Notamacropus eugenii, Sarcophilus harrisii</i>
Prototheria	<i>Ornithorhynchus anatinus</i>
Placentalia (ps21)	
Afrotheria	<i>Echinops telfairi, Loxodonta africana, Orycteropus afer, Procavia capensis</i>
Xenarthra	<i>Choloepus hoffmanni, Dasypus novemcinctus</i>
Boreoeutheria (ps22)	
Carnivora	<i>Felis catus, Mustela putorius furo</i>
Cetacea	<i>Physeter macrocephalus, Tursiops truncatus</i>
Chiroptera	<i>Myotis lucifugus, Pteropus vampyrus</i>
Eulipotyphla	<i>Erinaceus europaeus, Sorex araneus</i>
Perissodactyla	<i>Ceratotherium simum, Equus caballus</i>
Ruminantia	<i>Ovis aries</i>
Suina	<i>Sus scrofa</i>
Tylopoda	<i>Vicugna pacos</i>
Euarchontoglires (ps23)	
Lagomorpha	<i>Ochotona princeps, Oryctolagus cuniculus</i>
Rodentia	<i>Cavia porcellus, Chinchilla lanigera, Dipodomys ordii, Ictidomys tridecemlineatus, Jaculus jaculus, Mus musculus</i>
Scandentia	<i>Tupaia belangeri</i>

Primates (ps24)
Strepsirrhini
<i>Microcebus murinus, Otolemur garnettii, Propithecus coquereli</i>
Tarsiiformes
<i>Carlito syrichta</i>
Simiiformes (ps25)
Platyrrhini
<i>Aotus nancymae, Callithrix jacchus, Cebus capucinus, Saimiri boliviensis boliviensis</i>
Catarrhini (ps26)
Cercopithecidae
<i>Chlorocebus sabaeus, Colobus angolensis palliatus, Mandrillus leucophaeus, Papio hamadryas, Rhinopithecus roxellana</i>
Hominidae (ps27)
Gorilla
<i>Gorilla gorilla</i>
Hylobatidae
<i>Nomascus leucogenys</i>
Ponginae
<i>Pongo abelii</i>
Hominini (ps28)
Pan
<i>Pan paniscus, Pan troglodytes</i>
Homo sapiens (ps29)
Hominini
<i>Homo sapiens</i>

2.1.4.1 Filostratigram

Dobiveni rezultati filostratigrafske analize su prikazani filostratigramom. U gornjem je dijelu filostratigrama prikazana tablica s razdiobom apsolutnih frekvencija gena po filostratumima.

U srednjem je dijelu filostratigrama prikazana vrijednost omjera šansi linijskim dijagramom. Vrijednosti omjera šansi veće od 1 označavaju veći broj gena u filostratumu od onog koji se očekuje na temelju razdiobe svih ljudskih gena po filostratumima. Statistički značajno odstupanje od očekivanih frekvencija je naznačeno zvjezdicama ovisno o prilagođenoj p vrijednosti, prema sljedećem obrascu: $*** < 0,001 < ** < 0,01 < * < 0,05$.

U donjem je dijelu filostratigrama prikazan naziv i redni broj filogenetskih čvorova (filostratuma).

Osjenčanom su pozadinom naglašeni filostratumima u kojima je uočen statistički značajno veći broj gena od očekivanog.

2.1.4.2 Toplinska karta

Dok je filostratigramom prikazan trend nastanka novih gena, toplinskom je kartom (*engl.* heatmap) prikazana sličnost između ljudske proteinske sekvence i njoj najbližije sekvence za danu sestrinsku grupu pojedinog filostratuma. Ljudski su geni od interesa raspoređeni u redove pri čemu su grupirani po evolucijskom porijeklu određenom filostratigrafskom analizom. Stupci prikazuju sestrinsku granu svakog filostratuma, a intenzitetom boje pravokutnika na križanju reda i stupca kvantificirana je e vrijednost. Boja pravokutnika se proteže od blijedo crvene, preko bijele do tamno sive, pri čemu je čista bijela boja centrirana na e vrijednosti od 10^{-3} . Pravokutnici koji prikazuju e vrijednosti između 10^{-20} i 10^{-3} su obojeni bojama od tamno sive do bijele, a sve vrijednosti manje od 10^{-20} su prikazane fiksno, najtamnijom sivom bojom. Svi pravokutnici koji prikazuju e vrijednost manju od 10^{-3} imaju crni okvir. Pravokutnici koji prikazuju e vrijednosti između 10^{-3} i 10^5 su obojeni bojama od bijele do blijedo crvene, te imaju bijeli okvir.

2.1.5 Statistika gena biosintetskog puta N -glikozilacije

2.1.5.1 Niz proteina u biosintetskom putu

Nasumičnost evolucijskog porijekla gena koji kodiraju za proteine koji slijedno produljuju oligosaharid vezan na dolikol u endoplazmatskom retikulumu je procijenjena pomoću novouvedenog indeksa homogenosti (*engl.* homogeneity index, HI).

Za niz od N gena definiranog poretka u biosintetskom putu s poznatim evolucijskim porijeklom (filostratumom), HI se računa primjenom sljedeće jednadžbe:

$$HI = \sum_{i=2}^N f(i); f(i) = \begin{cases} 0, & ps_i = ps_{i-1} \\ 1, & ps_i \neq ps_{i-1} \end{cases} \quad (\text{Jednadžba 1})$$

gdje je i redni broj gena, N ukupan broj gena u nizu, a ps rang filostratuma pojedinog gena.

Nizovi u kojima su geni grupirani prema evolucijskom porijeklu će imati manji HI , dok nizovi u kojima evolucijsko porijeklo gena alterira duž članova niza će imati veći HI . Primjenom svih mogućih permutacija članova niza, izrađuje se razdioba indeksa homogenosti i prateća funkcija vjerojatnosti indeksa homogenosti iz koje se određuje vjerojatnost uočavanja niza gena s indeksom homogenosti manjim ili jednakim onom indeksu homogenosti opaženom u danom nizu gena.

2.1.5.2 Membranska topologija

Geni koji kodiraju za proteine koji sudjeluju u biosintezi N -glikana u endoplazmatskom retikulumu su podijeljeni u četiri grupe prema dva kriterija temeljena na: A) membranskoj topologiji (luminalna i citosolna strana) i B) evolucijskom porijeklu (prokariotsko, $ps < 6$ i eukariotsko porijeklo, $ps \geq 6$). Homogenost dobivene kontingencijske tablice, odnosno

nezavisnost membranske topologije i evolucijskog porijekla je testirana Fisherovim egzaktnim testom.

2.1.6 Statistički program

Manipulacija podacima je napravljena pomoću programskog jezika Python (verzija 3.9) (67, 68) i R (verzija 3.6.3) (69). Slike su napravljene u primjenom R paketa *ggplot2* (70) i *cowplot* (71). Shematski prikaz endoplazmatskog retikuluma i Golgijevog aparata je napravljen u računalnom programu Inkscape (verzija 1.1.1) (72).

2.2 ANALIZA UTJECAJA OKOLIŠNIH ČIMBENIKA NA NGLIKOZILACIJU IMUNOGLOBULINA G

2.2.1 Ispitanici: biobanka „10001 Dalmatinac“

Projekt „10001 Dalmatinac“ je trenutačno najbolje rangirani projekt u području biomedicine prema podacima Hrvatske znanstvene bibliografije (73). Projekt je do trenutka određivanja glikana nasumično uzorkovao 4884 ispitanika starija od 18 godina iz četiri populacije u periodu od 2003. do 2013. godine na otoku Visu i Korčuli (grad Korčula; selo i okolica Smokvice) te u gradu Splitu i okolici. Svi su ispitanici dobrovoljno pristali sudjelovati u istraživanju i potpisali su informirani pristanak sudjelovanju. Cijelo istraživanje je provedeno u skladu s etičkim smjernicama Helsinške deklaracije.

U svrhu izrade ove disertacije, u analizu su uključeni svi ispitanici kojima je izmjeren *N*-glikom imunoglobulina G.

2.2.1.1 Demografski podaci

Demografski su podaci prikupljeni putem upitnika. Ukupan broj ispitanika za koje postoje pojedini korišteni demografski podaci u biobazi je prikazan tablicom (Tablica 2).

Tablica 2. Ukupan broj ispitanika koji ima izmjerenu pojedinu demografsku varijablu među svim ispitanicima prisutnim u biobazi 10001 Dalmatinac kojima je određen *N*-glikom imunoglobulina G. Broj u zagradi prikazuje postotak ispitanika.

	otok Vis	grad Korčula	grad Split	selo i okolica Smokvice
Dob	975 (96)	961 (100)	979 (99)	695 (98)
Rod	971 (95)	961 (100)	979 (99)	694 (98)
Trajanje školovanja	964 (95)	941 (98)	971 (98)	686 (97)
Zanimanje	945 (93)	938 (98)	922 (93)	631 (89)
Zaposlenost	963 (94)	928 (97)	977 (99)	0 (0)
Broj članova kućanstva	0 (0)	927 (96)	976 (99)	680 (96)
Ekonomski status	961 (94)	926 (96)	974 (99)	655 (92)
Primanja kućanstva	0 (0)	926 (96)	971 (98)	666 (94)
Kompozitni indikator materijalnog statusa	958 (94)	924 (96)	976 (99)	684 (96)

Trajanje školovanja je definirano kao ukupan broj završenih godina obrazovanja. *Zaposlenost* je kategorijska varijabla sa osam razina: *zaposlen*, *samozaposlen*, *student*, *umirovljenik*, *kućanica*, *nezaposlen*, *uzdržavana osoba* i *ostalo*. *Ekonomski status* je definiran Likertovom ljestvicom sa 5 stupnjeva (lošiji od prosjeka – znatno bolji od prosjeka). *Primanja kućanstva* su definirana kao kategorijska varijabla sa četiri razine: *manje od 2000 HRK*, *2000-4000 HRK*, *4000-6000 HRK* i *više od 6000 HRK*. *Kompozitni indikator materijalnog statusa* je dobiven zbrajanjem odabranih materijalnih stvari koje ispitanik posjeduje: *vodovod*, *dva televizora*,

zahod s kotlićem, perilica posuđa, kupaonica, osobno računalo, plinsko/centralno grijanje, više od 100 knjiga, drveni pod, umjetnine (slike/vaze), telefon, automobil, video kamera, vikendica ili drugu nekretnina, zamrzivač i brodica.

Zbog specifične razdiobe, *trajanje školovanja* je kategorizirano u sljedeće kategorije: 0-4, 5-8, 9-12, 13-16, 17-19 i 20+ godina.

2.2.1.2 Antropometrijski podaci

Ukupan broj ispitanika za koje postoje pojedini podaci antropometrijski podaci u biobazi je prikazan tablicom (Tablica 3). Postupak prikupljanja pojedinih antropometrijskih podataka je detaljno opisan u znanstvenom radu Pulanić i sur. (74). Indeks tjelesne mase (*engl.* body mass index, BMI) je određen kao omjer tjelesne mase i kvadrata tjelesne visine. Debljina kožnih nabora je u svakom području mjerena tri puta. Spirometrija, u svrhu određivanja kapaciteta pluća je ponovljena dva puta. Krvni tlak je određen živinim manometrom, također dva puta. Mineralna gustoća kostiju je određena pomoću dvije metode: rendgenskom i ultrazvučnom denzitometrijom. U svrhu određivanja gležanjskog indeksa (*engl.* ankle-brachial pressure index, ABPI) izmjereni su arterijski tlakovi u brahijalnim arterijama, te stražnjim tibijalnim arterijama i dorzalnim arterijama stopala. ABPI je određen kao omjer najvišeg tlaka izmjenog na gležnjevima i višeg tlaka izmjenog na nadlakticama (75). Parametri arterijske elastičnosti su određeni dvjema metodama: aplanacijskom tonometrijom (SphygmoCor) i oscilometrijom (Arteriograph). Bioimpedancija je određena pomoću izmjenog otpora i reaktivnosti (kapaciteta) (76), te je procijenjena količina i udio masnog tkiva u tijelu prema Deurenberg i sur. (77). *Prisutnost Café-au-lait pjega* je binarna varijabla (da/ne).

Tablica 3. Ukupan broj ispitanika koji ima izmjerenu pojedinu antropometrijsku varijablu među svim ispitanicima prisutnim u biobazi 10001 Dalmatinac kojima je određen N-glikom imunoglobulina G. Broj u zagradi prikazuje postotak ispitanika.

	otok Vis	grad Korčula	grad Split	selo i okolica Smokvice
Tjelesna masa	957 (94)	945 (98)	979 (99)	691 (97)
Tjelesna visina	957 (94)	942 (98)	979 (99)	689 (97)
Polovica raspona ruku	365 (36)	907 (94)	978 (99)	686 (97)
Širina nadlaktice	954 (94)	916 (95)	976 (99)	0 (0)
Opseg				
struka	956 (94)	908 (94)	977 (99)	687 (97)
bokova	955 (94)	906 (94)	978 (99)	687 (97)
nadlaktice	956 (94)	904 (94)	978 (99)	688 (97)
glave	0 (0)	910 (95)	979 (99)	688 (97)
vrata	176 (17)	0 (0)	486 (49)	687 (97)
Debljina kožnih nabora u području				
bicepsa	955 (94)	914 (95)	978 (99)	0 (0)

	otok Vis	grad Korčula	grad Split	selo i okolica Smokvice
tricepsa	954 (94)	914 (95)	979 (99)	0 (0)
podlopatičnog mišića	955 (94)	913 (95)	979 (99)	0 (0)
abdomena	956 (94)	903 (94)	978 (99)	0 (0)
Kapacitet pluća				
forsirani ekspiratorni volumen	943 (92)	893 (93)	979 (99)	675 (95)
forsirani vitalni kapacitet	943 (92)	893 (93)	979 (99)	675 (95)
vršni ekspiratorni protok	0 (0)	893 (93)	979 (99)	685 (96)
Mineralna gustoća kostiju				
rendgen: gustoća	583 (57)	950 (99)	977 (99)	0 (0)
rendgen: T-vrijednost	583 (57)	777 (81)	976 (99)	0 (0)
ultrazvuk: T-vrijednost	0 (0)	121 (13)	79 (8)	681 (96)
ultrazvuk: brzina	0 (0)	119 (12)	79 (8)	679 (96)
ultrazvuk: prigušenje	0 (0)	119 (12)	79 (8)	679 (96)
Krvni tlak				
sistolčki	961 (94)	955 (99)	978 (99)	689 (97)
dijastolički	961 (94)	955 (99)	978 (99)	689 (97)
Gležanjski index, arterijski tlakovi				
desna brahijalna arterija	944 (93)	912 (95)	979 (99)	0 (0)
desna stražnja tibijalna arterija	940 (92)	910 (95)	979 (99)	0 (0)
desna dorzalna arterija stopala	939 (92)	910 (95)	978 (99)	0 (0)
lijeva brahijalna arterija	941 (92)	908 (94)	976 (99)	0 (0)
lijeva stražnja tibijalna arterija	937 (92)	909 (95)	978 (99)	0 (0)
lijeva dorzalna arterija stopala	934 (92)	909 (95)	977 (99)	0 (0)
Arterijska elastičnost				
SphygmoCor				
pAIx	0 (0)	0 (0)	476 (48)	555 (78)
cAIx	0 (0)	0 (0)	478 (48)	558 (79)
brzina pulsog vala	0 (0)	0 (0)	435 (44)	526 (74)
centralni sistolički krvni tlak	0 (0)	0 (0)	478 (48)	567 (80)
centralni dijastolički tlak	0 (0)	0 (0)	474 (48)	567 (80)
puls	0 (0)	0 (0)	463 (47)	568 (80)
Arteriograph				
pAIx	361 (35)	0 (0)	432 (44)	0 (0)
cAIx	305 (30)	0 (0)	431 (44)	0 (0)
brzina pulsog vala	306 (30)	0 (0)	436 (44)	0 (0)
centralni sistolički krvni tlak	305 (30)	0 (0)	435 (44)	0 (0)
Biološka impedancija				
otpor	917 (90)	0 (0)	0 (0)	0 (0)
reaktivnost	913 (90)	0 (0)	0 (0)	0 (0)
Prisutnost Café-au-lait pjega	0 (0)	113 (12)	476 (48)	613 (86)

2.2.1.3 Povijest bolesti

Podaci o povijesti bolesti su prikupljeni putem upitnika, a obuhvatili su status dijagnoze (da/ne), godinu dijagnoze, status terapije (da/ne) za hipertenziju, koronarnu bolest srca, cerebrovaskularni inzult, shizofreniju, maniju/depresiju, maligni tumor, dijabetes, giht, glaukom, artritis, bolest bubrega, čir i astmu/alergiju. Podaci o svakoj navedenoj bolesti postoje za 93 – 100 % ispitanika, ovisno o kohorti. Iznimno, podaci o astmi/alergiji su prisutni samo među ispitanicima koji su uzorkovani u selu i okolici Smokvice.

Dio ispitanika, u sklopu projekta 10001 Dalmatinac je praćen kroz period od 3 do 9 godina. U ovaj je doktorat uključen i podatak o razvoju hipertenzije u periodu praćenja za 401 ispitanika uzorkovanog na području otoka Visa, 125 ispitanika uzorkovanih u gradu Korčuli te 504 ispitanika uzorkovana u gradu Splitu.

2.2.1.4 Biokemijski testovi

Tablica 4. Ukupan broj ispitanika za koje postoji rezultati pojedinih biokemijskih testova među svim ispitanicima prisutnim u biobazi 10001 Dalmatinac kojima je određen N-glikom imunoglobulina G. Broj u zagradi prikazuje postotak ispitanika.

	otok Vis	grad Korčula	grad Split	selo i okolica Smokvice
N-glikom IgG-a				
LC-ESI-MS	831 (81)	951 (99)	988 (100)	710 (100)
HILIC-UPLC	890 (87)	915 (95)	0 (0)	0 (0)
Glukoza	965 (95)	956 (99)	967 (98)	690 (97)
Inzulin	960 (94)	951 (99)	0 (0)	0 (0)
Kolesterol	946 (93)	956 (99)	967 (98)	690 (97)
HDL kolesterol	946 (93)	953 (99)	963 (97)	690 (97)
LDL kolesterol	946 (93)	950 (99)	949 (96)	679 (96)
VLDL kolesterol	0 (0)	0 (0)	0 (0)	678 (95)
Trigliceridi	946 (93)	956 (99)	967 (98)	689 (97)
Kreatinin	964 (95)	956 (99)	967 (98)	691 (97)
Urati	964 (95)	956 (99)	967 (98)	690 (97)
Kalcij	965 (95)	956 (99)	966 (98)	690 (97)
Albumin	963 (94)	956 (99)	967 (98)	690 (97)
HbA1c	953 (93)	498 (52)	972 (98)	687 (97)
Fibrinogen	958 (94)	855 (89)	969 (98)	633 (89)

IgG – imunoglobulin G; LC-ESI-MS – *engl.* liquid chromatography coupled with electrospray mass spectrometry, tekućinska kromatografija spregnuta s masenom spektroskopijom; HILIC-UPLC – *engl.* hydrophilic interaction ultra-performance liquid chromatography, tekućinskom kromatografijom visoke djelotvornosti temeljenoj na hidrofilnim interakcijama; HDL – *engl.* high density lipoprotein, lipoprotein visoke gustoće; LDL – *engl.* low density lipoprotein, lipoprotein niske gustoće, VLDL – *engl.* very low density lipoprotein, lipoprotein vrlo niske gustoće, HbA1c – glicirani hemoglobin

U svrhu određivanja biokemijskih testova svim ispitanicima je uzorkovana puna krv u jutarnjim satima (između 8.30 i 9.30), pri čemu su ispitanici bili natašte.

Uzorci su pohranjeni na -70 °C do daljnje analize koja je za sve uzorke obavljena u istom laboratoriju u Zagrebu. Laboratorij odlikuje strogim pravilima unutarnje kontrole kvalitete, te posjeduje akreditaciju za provođenje analiza u sklopu Randox grupe, *engl.* Randox International Quality Assessment Scheme (RIQAS).

Iznimno, izolacija imunoglobulina G i analiza njegovog *N*-glikoma metodama tekućinske kromatografije na obrnutim fazama spregnute s elektrosprej ionizacijskom masenom spektroskopijom (*engl.* liquid chromatography coupled with electrospray ionization mass spectrometry, LC-ESI-MS) i tekućinskom kromatografijom visoke djelotvornosti temeljenoj na hidrofilnim interakcijama (*engl.* hydrophilic interaction ultra-performance liquid chromatography, HILIC-UPLC) su napravljene u Genos laboratoriju prema standardnim protokolima (78, 79). Detaljan opis kromatografskih vršaka je naveden u tablici (Tablica 5). Ukupan broj ispitanika za koje postoje rezultati pojedinih biokemijskih testova u biobazi je prikazan tablicom (Tablica 4).

Tablica 5. Opis struktura koje najvećim udjelom čine kromatografski vršak u kromatografskim analizama *N*-glikoma IgG-a.

Kromatografski vršak	Kratica strukture*	Opis strukture
LC-ESI-MS		
IgG1_G0F	FA2	sržno fukoziliran s dvije grane (biantenar) na IgG-u klase 1 (IgG1)
IgG1_G1F	FA2G1	sržno fukoziliran, monogalaktoziliran biantenar na IgG1
IgG1_G2F	FA2G2	sržno fukoziliran, digalaktoziliran biantenar na IgG1
IgG1_G0FN	FA2B	sržno fukoziliran, biantenar s račvajućim GlcNAc-om na IgG1
IgG1_G1FN	FA2BG1	sržno fukoziliran, monogalaktoziliran biantenar s račvajućim GlcNAc-om na IgG1
IgG1_G2FN	FA2BG2	sržno fukoziliran, digalaktoziliran biantenar s račvajućim GlcNAc-om na IgG1
IgG1_G1FS1	FA2G1S1	sržno fukoziliran, monogalaktoziliran, monosijaliniziran biantenar na IgG1
IgG1_G2FS1	FA2G2S1	sržno fukoziliran, digalaktoziliran, monosijaliniziran biantenar na IgG1
IgG1_G1FNS1	FA2BG1S1	sržno fukoziliran, monogalaktoziliran, monosijaliniziran biantenar s račvajućim GlcNAc-om na IgG1
IgG1_G2FNS1	FA2BG2S1	sržno fukoziliran, digalaktoziliran, monosijaliniziran biantenar s račvajućim GlcNAc-om na IgG1
IgG1_G0	A2	agalaktoziliran biantenar na IgG1
IgG1_G1	A2G1	monogalaktoziliran biantenar na IgG1
IgG1_G2	A2G2	digalaktoziliran biantenar na IgG1

Kromatografski vršak	Kratice strukture*	Opis strukture
IgG1_G0N	A2B	agalaktoziliran biantenar s račvujućim GlcNAc-om na IgG1
IgG1_G1N	A2BG1	monogalaktoziliran biantenar s račvujućim GlcNAc-om na IgG1
IgG1_G2N	A2BG2	digalaktoziliran biantenar s račvujućim GlcNAc-om na IgG1
IgG1_G1S1	A2G1S1	monogalaktoziliran, monosijaliniziran biantenar na IgG1
IgG1_G2S1	A2G2S1	digalaktoziliran, monosijaliniziran biantenar na IgG1
IgG1_G1NS1	A2BG1S1	monogalaktoziliran, monosijaliniziran biantenar s račvujućim GlcNAc-om na IgG1
IgG1_G2NS1	A2BG2S1	digalaktoziliran, monosijaliniziran biantenar s račvujućim GlcNAc-om na IgG1
IgG2_G0F	FA2	sržno fukoziliran biantenar na IgG klase 2 (IgG2)
IgG2_G1F	FA2G1	sržno fukoziliran, monogalaktoziliran biantenar na IgG2
IgG2_G2F	FA2G2	sržno fukoziliran, digalaktoziliran biantenar na IgG2
IgG2_G0FN	FA2B	sržno fukoziliran, biantenar s račvujućim GlcNAc-om na IgG2
IgG2_G1FN	FA2BG1	sržno fukoziliran, monogalaktoziliran biantenar s račvujućim GlcNAc-om na IgG2
IgG2_G2FN	FA2BG2	sržno fukoziliran, digalaktoziliran biantenar s račvujućim GlcNAc-om na IgG2
IgG2_G1FS1	FA2G1S1	sržno fukoziliran, monogalaktoziliran, monosijaliniziran biantenar na IgG2
IgG2_G2FS1	FA2G2S1	sržno fukoziliran, digalaktoziliran, monosijaliniziran biantenar na IgG2
IgG2_G1FNS1	FA2BG1S1	sržno fukoziliran, monogalaktoziliran, monosijaliniziran biantenar s račvujućim GlcNAc-om na IgG2
IgG2_G2FNS1	FA2G2S1	sržno fukoziliran, digalaktoziliran, monosijaliniziran biantenar na IgG2
IgG2_G0	A2	agalaktoziliran biantenar na IgG1
IgG2_G1	A2G1	monogalaktoziliran biantenar na IgG1
IgG2_G2	A2G2	digalaktoziliran biantenar na IgG1
IgG2_G0N	A2B	agalaktoziliran biantenar s račvujućim GlcNAc-om na IgG1
IgG2_G1N	A2BG1	monogalaktoziliran biantenar s račvujućim GlcNAc-om na IgG1
IgG2_G2N	A2BG2	digalaktoziliran biantenar s račvujućim GlcNAc-om na IgG1
IgG2_G1S1	A2G1S1	monogalaktoziliran, monosijaliniziran biantenar na IgG2
IgG2_G2S1	A2G2S1	digalaktoziliran, monosijaliniziran biantenar na IgG2
IgG2_G1NS1	A2BG1S1	monogalaktoziliran, monosijaliniziran biantenar s račvujućim GlcNAc-om na IgG2
IgG2_G2NS1	A2BG2S1	digalaktoziliran, monosijaliniziran biantenar s račvujućim GlcNAc-om na IgG2
IgG4_G0F	FA2	sržno fukoziliran biantenar na IgG klase 4 (IgG4)

Kromatografski vršak	Kratice strukture*	Opis strukture
IgG4_G1F	FA2G1	sržno fukoziliran, monogalaktoziliran biantenar na IgG4
IgG4_G2F	FA2G2	sržno fukoziliran, digalaktoziliran biantenar na IgG4
IgG4_G0FN	FA2B	sržno fukoziliran, biantenar s račvućim GlcNAc-om na IgG4
IgG4_G1FN	FA2BG1	sržno fukoziliran, monogalaktoziliran biantenar s račvućim GlcNAc-om na IgG4
IgG4_G2FN	FA2BG2	sržno fukoziliran, digalaktoziliran biantenar s račvućim GlcNAc-om na IgG4
IgG4_G1FS1	FA2G1S1	sržno fukoziliran, monogalaktoziliran, monosijaliniziran biantenar na IgG4
IgG4_G2FS1	FA2G2S1	sržno fukoziliran, digalaktoziliran, monosijaliniziran biantenar na IgG4
IgG4_G1FNS1	FA2BG1S1	sržno fukoziliran, monogalaktoziliran, monosijaliniziran biantenar s račvućim GlcNAc-om na IgG4
IgG4_G2FNS1	FA2BG2S1	sržno fukoziliran, digalaktoziliran, monosijaliniziran biantenar s račvućim GlcNAc-om na IgG4
UPLC		
GP1	FA1	sržno fukoziliran s jednom granom (monoantenar)
GP2	A2	agalaktoziliran biantenar
GP4	FA2	sržno fukoziliran, agalaktoziliran biantenar
GP5	M5	glikan s pet manoznih jedinica
GP6	FA2B	sržno fukoziliran, agalaktoziliran biantenar
GP7	A2G1	monogalaktoziliran biantenar
GP8	FA2[6]G1	sržno fukoziliran, (α 1-6) monogalaktoziliran biantenar
GP9	FA2[3]G1	sržno fukoziliran, (α 1-3) monogalaktoziliran biantenar
GP10	FA2[6]BG1	sržno fukoziliran, (α 1-6) monogalaktoziliran biantenar s račvućim GlcNAc-om
GP11	FA2[3]BG1	sržno fukoziliran, (α 1-3) monogalaktoziliran biantenar s račvućim GlcNAc-om
GP12	A2G2	digalaktoziliran biantenar
GP13	A2BG2	digalaktoziliran biantenar s račvućim GlcNAc-om
GP14	FA2G2	sržno fukoziliran, digalaktoziliran biantenar
GP15	FA2BG2	sržno fukoziliran, digalaktoziliran biantenar s račvućim GlcNAc-om
GP16	FA2G1S1	sržno fukoziliran, monogalaktoziliran, monosijaliniziran biantenar
GP17	A2G2S1	digalaktoziliran, monosijaliniziran biantenar
GP18	FA2G2S1	sržno fukoziliran, digalaktoziliran, monosijaliniziran biantenar
GP19	FA2BG2S1	sržno fukoziliran, digalaktoziliran, monosijaliniziran biantenar s račvućim GlcNAc-om
GP20		glikan nedefinirane strukture
GP21	A2G2S2	digalaktoziliran, disijaliniziran biantenar
GP22	A2BG2S2	digalaktoziliran, disijaliniziran biantenar s račvućim GlcNAc-om
GP23	FA2G2S2	sržno fukoziliran, digalaktoziliran, disijaliniziran biantenar

Kromatografski vršak	Kratice strukture*	Opis strukture
GP24	FA2BG2S2	sržno fukoziliran, digalaktoziliran, disijaliniziran biantenar s račvućim GlcNAc-om

*Početni *F* kratice označava prisutnost sržne fukoze; *A#* označava broj antena (grana); *B* označava prisutnost račvućeg GlcNAc-a; *G#* označava broj galaktoza; *S#* označava broj sijaliniskih kiselina; *[#]* označava ugljikov atom na kojem je vezana antena s galaktozom

2.2.1.5 Kognitivni testovi

Ukupan broj ispitanika za koje postoje podaci o rezultatima pojedinih kognitivnih testova u biobazi je prikazan tablicom (Tablica 6). Standardne progresivne matrice (*engl.* standard progressive matrices) su grupa testova koja neverbalnim putem procjenjuje inteligenciju (80). Test tečnosti (*engl.* verbal fluency test, F-A-S test) riječi odražava verbalnu fluentnost u kojoj se ispitanik u zadanom vremenu pokušava prisjetiti riječi koje počinju sa zadanim slovom (F, A i S) (81). Test kodiranja (*engl.* digital-symbol coding, DSC) se sastoji od 9 parova brojeva i simbola, te zadanim nizom znakova i brojeva kojima ispitanik mora pridružiti upareni simbol ili broj. Rezultat testa temelji se na broju uspješno pridruženih znakova-brojeva u zadanom vremenu (82). Reyov test auditivno-verbalnog učenja (*engl.* Rey's Auditory Verbal Learning Test) procjenjuje sposobnost učenja i pamćenja ispitanika na način da se ispitaniku da lista A od 15 riječi i traži se od njega da ponovi što više riječi sa liste. Postupak se ponavlja pet puta, nakon čega se uvodi nova lista (lista B). Po završetku liste B, nakon 5 i 30 minuta od ispitanika se traži da ponovi riječi s liste A (83).

Tablica 6. Ukupan broj ispitanika za koje postoji rezultati pojedinog kognitivnog testa među svim ispitanicima prisutnim u biobazi 10001 Dalmatinac kojima je određen *N*-glikom imunoglobulina G. Broj u zagradi prikazuje postotak ispitanika.

	otok Vis	grad Korčula	grad Split	selo i okolica Smokvice
Standardne progresivne matrice	0 (0)	649 (68)	950 (96)	498 (70)
Tečnost riječi s početnim slovom				
F	564 (55)	643 (67)	953 (96)	499 (70)
A	564 (55)	643 (67)	953 (96)	499 (70)
S	564 (55)	641 (67)	953 (96)	498 (70)
Kodiranje	474 (46)	726 (76)	954 (97)	496 (70)
Reyov test auditivno-verbalnog učenja				
neposredni rezultat	0 (0)	639 (66)	952 (96)	499 (70)
učenje	0 (0)	639 (66)	952 (96)	499 (70)
zaboravljanje	0 (0)	611 (64)	945 (96)	382 (54)
relativno zaboravljanje	0 (0)	611 (64)	945 (96)	382 (54)
Mill Hill test	0 (0)	768 (80)	943 (95)	480 (68)

Mill Hill test mjeri verbalnu inteligenciju kroz 88 riječi podijeljenih u dvije grupe po 44. Od ispitanika se traži da ispravno definira riječi u prvoj grupi kroz pitanja otvorenog tipa, a drugoj grupa kroz pitanja zatvorenog tipa – pitanja višestrukog odabira (84).

2.2.1.6 Podaci o plodnosti

Podaci o plodnosti se prvenstveno odnose na dio populacije ženskog spola, a prikupljeni su putem upitnika. Tablicom su prikazani brojevi ispitanika za uzorkovane varijable korištene u sklopu ove disertacije (Tablica 7). *Postojanje regularnog menstrualnog ciklusa* je varijabla binarnog tipa s mogućim odgovorima *da* i *ne*.

Tablica 7. Ukupan broj ispitanika za koje postoje podaci o plodnosti među svim ispitanicima prisutnim u biobazi 10001 Dalmatinac kojima je određen N-glikom imunoglobulina G. Broj u zagradi prikazuje postotak ispitanika.

	otok Vis	grad Korčula	grad Split	selo i okolica Smokvice
Dob pri prvoj menstruaciji	545 (53)	566 (59)	592 (60)	407 (57)
Postojanje regularnog menstrualnog ciklusa	544 (53)	562 (58)	597 (60)	407 (57)
Dob nastupanja menopauze	356 (35)	365 (38)	317 (32)	238 (34)
Broj porođaja	562 (55)	588 (61)	591 (60)	403 (57)
Broj mrtvorodenčadi	562 (55)	588 (61)	591 (60)	405 (57)
Broj pobačaja				
spontanih	562 (55)	588 (61)	592 (60)	402 (57)
kirurških	562 (55)	588 (61)	592 (60)	402 (57)
Broj djece	960 (94)	910 (95)	968 (98)	648 (91)

2.2.1.7 Podaci o konzumaciji alkohola

Podaci o konzumaciji alkohola su prikupljeni putem upitnika u obliku tjednog unosa alkoholnih pića izraženih u litrama. Iz prikupljenih podataka izračunat je ukupan unos etanola u mililitrima.

Tablica 8. Volumni udio etanola u vrsti alkoholnog pića i ukupan broj ispitanika za koje postoje podaci o konzumaciji pojedinih alkoholnih pića među svim ispitanicima prisutnim u biobazi 10001 Dalmatinac kojima je određen N-glikom imunoglobulina G. Broj u zagradi prikazuje postotak ispitanika.

	etanol (vol. %)	otok Vis	grad Korčula	grad Split	selo i okolica Smokvice
Pivo	5,0	962 (94)	952 (99)	978 (99)	638 (90)
Vino	12,0	964 (95)	0 (0)	0 (0)	0 (0)
Bijelo vino	11,0	0 (0)	952 (99)	978 (99)	638 (90)
Crno vino	13,0	0 (0)	951 (99)	978 (99)	636 (90)
Vino razrijeđeno	6,0	964 (95)	952 (99)	978 (99)	0 (0)
bijelo	5,5	0 (0)	0 (0)	230 (23)	638 (90)
crno	6,5	0 (0)	0 (0)	231 (23)	636 (90)
Žestoka pića	40,0	963 (94)	952 (99)	978 (99)	638 (90)

Tablicom su prikazane vrste alkoholnih pića za koje su podaci prikupljeni u sklopu projekta 10001 Dalmatinac te volumni udio etanola u alkoholnim pićima korišten za izračun ukupnog unosa (Tablica 8).

2.2.1.8 Konzumacija duhanskih proizvoda

Podaci o pušenju su prikupljeni putem upitnika. Tablicom su prikazani brojevi ispitanika za uzorkovane varijable korištene u sklopu ove disertacije (Tablica 9). *Pušački status* je kategorijska varijabla s tri razine: *ne*, *bivši pušač* i *da*. *Vrsta duhanskog proizvoda* je također kategorijska varijabla s tri razine: *cigareta*, *lula* i *cigara*. Izračunat je broj cigareta-ekvivalentnih proizvoda dnevno, na način da je dnevni broj proizvoda kod ispitanika koje puše lulu ili cigare pomnožen sa 2,5 (85). Za pušače i bivše pušače izračunata je vrijednost tzv. *pušačke godine* koja prikazuje ukupan broj godina u kojima ispitanik popuši kutiju dnevno (20 cigareta) (*engl.* pack-year) prema jednadžbi:

$$PY = \frac{\text{broj cigareta dnevno}}{20} \times \text{pušački staž} \quad (\text{Jednadžba 2})$$

Tablica 9. Ukupan broj ispitanika za koje postoje podaci o konzumaciji duhanskih proizvoda među svim ispitanicima prisutnim u biobazi 10001 Dalmatinac kojima je određen N-glikom imunoglobulina G. Broj u zagradi prikazuje postotak ispitanika.

	otok Vis	grad Korčula	grad Split	selo i okolica Smokvice
Pušački status	963 (94)	940 (98)	975 (99)	688 (97)
Vrsta duhanskog proizvoda	273 (27)	216 (22)	254 (26)	203 (29)
Pušači				
broj cigareta dnevno	270 (26)	216 (22)	254 (26)	203 (29)
pušački staž	270 (26)	216 (22)	254 (26)	203 (29)
Bivši pušači				
broj cigareta dnevno	281 (28)	254 (26)	270 (27)	122 (17)
pušački staž	281 (28)	254 (26)	270 (27)	122 (17)
broj godina od prestanka	279 (27)	254 (26)	270 (27)	122 (17)

2.2.1.9 Tjelesna aktivnost

Podaci o tjelesnoj aktivnosti su prikupljeni putem upitnika. Tablicom su prikazani brojevi ispitanika za uzorkovane varijable korištene u sklopu ove disertacije (Tablica 10).

Tablica 10. Ukupan broj ispitanika za koje postoje podaci dobiveni o tjelesnoj aktivnosti među svim ispitanicima prisutnim u biobazi 10001 Dalmatinac kojima je određen N-glikom imunoglobulina G. Broj u zagradi prikazuje postotak ispitanika.

	otok Vis	grad Korčula	grad Split	selo i okolica Smokvice
Tjelesna aktivnost				
na poslu	967 (95)	944 (98)	975 (99)	676 (95)
kod kuće	967 (95)	944 (98)	975 (99)	676 (95)

Tjelesna aktivnost na poslu/kod kuće je definirana kao kategorijska varijabla s četiri razine: *bez aktivnosti (sjedenje), blaga aktivnost, srednja aktivnost i intenzivna aktivnost*. Odgovori su kodirani brojevima od 0 do 3, a ukupna tjelesna aktivnost je izračunata zbrajanjem kodiranih vrijednosti tjelesne aktivnosti na poslu i kod kuće.

2.2.1.10 Općeniti upitnik o prehrani

Općenitim upitnikom o prehrani prikupljeni su podaci od kojih su neki iskorišteni u ovoj disertaciji, a brojevi ispitanika za te podatke su prikazani tablicom (Tablica 11).

Tablica 11. Ukupan broj ispitanika za koje postoje podaci dobiveni općenitim upitnikom o prehrani među svim ispitanicima prisutnim u biobazi 10001 Dalmatinac kojima je određen N-glikom imunoglobulina G. Broj u zagradi prikazuje postotak ispitanika.

	otok Vis	grad Korčula	grad Split	selo i okolica Smokvice
Učestalost obroka				
zajutrak	923 (90)	926 (96)	944 (96)	0 (0)
doručak	944 (93)	929 (97)	971 (98)	0 (0)
ručak	935 (92)	937 (98)	976 (99)	0 (0)
užina	935 (92)	922 (96)	947 (96)	0 (0)
večera	943 (92)	933 (97)	970 (98)	0 (0)
Dnevni unos tekućine	962 (94)	942 (98)	978 (99)	0 (0)
Dnevni unos šećera	880 (86)	840 (87)	945 (96)	564 (79)
Soljenje hrane prije kušanja	373 (37)	0 (0)	480 (49)	684 (96)
Vegetarijanska prehrana	944 (93)	942 (98)	972 (98)	0 (0)
Učestalost uzimanja vitamina	940 (92)	941 (98)	972 (98)	0 (0)
Učestalost uzimanja minerala uz				
zajutrak	962 (94)	954 (99)	977 (99)	0 (0)
doručak	962 (94)	954 (99)	977 (99)	0 (0)
ručak	962 (94)	954 (99)	977 (99)	0 (0)
užinu	962 (94)	954 (99)	977 (99)	0 (0)
večeru	962 (94)	954 (99)	977 (99)	0 (0)
Uzimanje dodataka prehrani koji sadrže				
vitamin A	958 (94)	954 (99)	976 (99)	644 (91)
vitamin B	959 (94)	954 (99)	977 (99)	645 (91)
vitamin C	959 (94)	954 (99)	977 (99)	646 (91)
vitamin D	958 (94)	954 (99)	977 (99)	646 (91)
vitamin E	958 (94)	954 (99)	977 (99)	646 (91)
više vitamina (multivitamin)	959 (94)	954 (99)	977 (99)	646 (91)
kalcij	959 (94)	954 (99)	978 (99)	647 (91)
magnezij	958 (94)	954 (99)	978 (99)	649 (91)
željezo	959 (94)	954 (99)	978 (99)	644 (91)
druge minerale	958 (94)	954 (99)	977 (99)	647 (91)

Učestalost obroka, je kategorijska varijabla s tri razine: *uvijek, ponekad, nikad*. Na temelju odgovora, izračunat je broj konkretnih obroka u danu kao zbroj kodiranih odgovora, pri čemu

su odgovori kodirani slijedećim vrijednosti: *uvijek* = 1, *ponekad* = 1/2, i *nikad* = 0. *Vegetarijanska prehrana* i *uzimanje dodatka prehrani* su binarne varijable (da/ne). *Dnevni unos tekućine* je kategorijska varijabla definirana s tri razine: *manje od 1 L*, *između 1 i 2 L* i *više od 2 L*. *Dnevni unos šećera* je kategorijska varijabla definirana s četiri razine: *ne koristim šećer*, *jedna žličica*, *jedna žlica* i *više od jedne žlice*. *Soljenje hrane prije kušanja* je također kategorijska varijabla s četiri razine: *nikad*, *ponekad*, *često* i *gotovo uvijek*.

2.2.1.11 Unos prehrambenih proizvoda

Podaci o učestalosti unosa različitih prehrambenih proizvoda su prikupljeni kroz upitnik o učestalosti konzumacije hrane i pića (*engl.* food frequency questionnaire, FFQ). Upitnik je obuhvatio učestalost konzumacije mlijeka, jogurta, vrhnja, svježeg sira, punomasnog sira, tvrdog sira, svinjetine, teletine, junetine, janjetine, iznutrica (jetra, bubreg, pluća,...), slanine, kobasica, salame, konzerviranih mesnih derivata, divljači, piletine, puretine, pršuta (osim u kohorti uzorkovanoj na otoku Visu), bijele ribe, plave ribe, plodova mora (rakovi i školjke), lignji/hobotnice, sušene ribe/usoljene sardine, ribljih prerađevina, jaja, lisnatog povrća, korjenastog povrća, cvjetastog povrća, plodastog povrća, mahunarki, konzerviranog povrća, krumpira, gljiva, svježeg voća, orašastih plodova, voćnih kompoti, sušenog voća, bijelog kruha, integralnog kruha, muesli žitarica, tjestenine/riže, kolača, čokolade, keksa, bombona, džema/marmelade, instant juha, povrtnih sokova, voćnih sokova, Cedevite, gaziranih bezalkoholnih pića, žestokih pića, kave, čaja, slanih grickalica, piva, bijelog vina, crnog vina, bevande.

Upitnik koji su rješavali ispitanici uzorkovani u selu i okolici Smokvice nije sadržavao podatke o konzerviranim mesnim derivatima, divljači, ribljim prerađevinama, gljivama, voćnim kompotima, muesli žitaricama, instant juhi, povrtnim i voćnim sokovima.

Za svaku od ponuđenih namirnica ispitanici su definirali učestalost konzumacije kroz 5 razina: *dnevno*, *2-3 puta tjedno*, *jednom tjedno*, *povremeno*, *nikada*. Medijan udjela ispravno popunjenih pitanja je bio 94 %, uz prvi i deveti decil 89 – 98 %.

2.2.1.12 Preferencija prehrambenih proizvoda

Podaci o preferenciji različitih prehrambenih proizvoda su prikupljeni putem upitnika u kojem su za svaki prehrambeni proizvod ispitanici ispunili Likertovu ljestvicu s devet stupnjeva (*uopće ne volim – u potpunosti volim*). Upitnik je obuhvatio preferenciju alkohola, artičoka, avokada, banana, bijelog vina, bolonjez umaka, cikoriije, crnih maslina, crnog vina, cvjetače, čaja, češnjaka, čokolade s visokim udjelom kakaa, eggnoga, espresso kave bez šećera, espresso kave sa šećerom, gorgonzola sira, gorke hrane, graha, brokule, grejpa, incuna, janjetine, jetrica,

jogurta, kapara, kisele hrane, komorača, kozjeg sira, krastavca, kruha s maslacem, kuhanog mlijeka, kupusa, likera, limuna, luka, ljute hrane, ljute papričice, marcipana, masne hrane, mesa, mliječne čokolade, mliječnih proizvoda, mlijeka, mozzarella sira, nara, octa, oraha, panettone kolača, paprika, paradajza, patlidžana, povrća, repe, sardina, sladoleda, slanih grickalica, slanine, soka od naranče, svinjetine, šlaga, šparoga, špinata, šumskih gljiva, šunke, tamnog piva, voća i zelenih maslina.

Upitnik su popunjavali ispitanici uzorkovani u mjestu i okolici Korčula, otok Korčula (N (%) = 105 (11)), gradu Splitu (N (%) = 395 (40)) i selu i okolici Smokvice, otok Korčula (N (%) = 657 (93)).

2.2.1.13 Opći zdravstveni upitnik

Opći zdravstveni upitnik s 30 pitanja (*engl.* general health questionnaire, GHQ-30) je validirani upitnik kojim je moguće procijeniti mentalno zdravlje ispitanika, pri čemu rezultat s više bodova ukazuje na lošije stanje mentalnog zdravlja (86). Upitnik se sastoji od 30 pitanja vezanih uz tjelesne simptome, anksioznost i nesanicu, socijalnu disfunkciju i depresiju. Na svako je pitanje ponuđeno četiri odgovora koji prate ordinalnu mjernu ljestvicu i visoko su usporedivi sa Likertovom ljestvicom. Rezultat upitnika je dobiven na dva načina bodovanja: originalnom GHQ metodom u kojoj se prva dva ponuđena odgovora boduju s 0, a druga dva s 1; te Likertovom metodom gdje su odgovori bodovani redom od 0 do 3. Ukupni rezultat je suma bodova ostvarenih na svim pitanjima.

Upitnik je ispunilo N (%): 958 (94), 935 (97), 961 (97) i 608 (86) ispitanika uzorkovanih na području otoka Visa, grada i okolice Korčule, grada Splita, te sela i okolice Smokvice (otok Korčula).

2.2.1.14 Münchenski upitnik o kronotipu

Münchenski upitnik o kronotipu (*engl.* Munich chronotype questionnaire, MCTQ) je upitnik koji sadrži pitanja o rasporedu spavanja na radne i neradne dane s ciljem razumijevanja biološkog sata čovjeka (87). Kroz ovaj je upitnik uveden pojam *socijalni jetlag* jer uspješno mjeri razliku u dvije "vremenske zone" u kojima čovjek živi – unutarne biološke vremenske zone čovjeka na slobodne dane i vanjske sociološke vremenske zone na radne dane (87). Iz upitnika je izračunata prosječna duljina sna, prosječna dnevna izloženost danjem svjetlu, tjedni gubitak sna i socijalni jetlag prema jednadžbama autora upitnika (87–90). Tablicom su prikazani brojevi ispitanika za varijable korištene u sklopu ove disertacije (Tablica 12).

Tablica 12. Ukupan broj ispitanika za koje postoje podaci dobiveni Münchenskim upitnikom o kronotipu među svim ispitanicima prisutnim u biobazi 10001 Dalmatinac kojima je određen *N*-glikom imunoglobulina G. Broj u zagradi prikazuje postotak ispitanika.

	otok Vis	grad Korčula	grad Split	selo i okolica Smokvice
Duljina sna	0 (0)	323 (34)	541 (55)	361 (51)
Tjedni gubitak sna	0 (0)	323 (34)	541 (55)	361 (51)
Izloženost svjetlu	0 (0)	323 (34)	542 (55)	369 (52)
Socijalni jetlag	0 (0)	859 (89)	547 (55)	573 (81)
Relativni socijalni jetlag	0 (0)	859 (89)	547 (55)	573 (81)

2.2.1.15 Epworthova ljestvica pospanosti

Epworthova ljestvica pospanosti (*engl.* Epworth sleepiness scale, ESS) je razvijena s ciljem mjerenja dnevne pospanosti (91). Upitnik se sastoji od osam pitanja koja obuhvaćaju svakodnevne aktivnosti. Na svako pitanje je moguće odgovoriti kroz ordinalnu mjernu ljestvicu od četiri odgovora kojom ispitanik opisuje osjećaj potrebe za spavanjem (*ne osjećam potrebu za spavanjem – imam neodoljivu potrebu za spavanjem*), pri čemu se bodovanje provodi bodovima od 0 do 3. Ukupan rezultat se dobije zbrajanjem bodova svih pitanja.

Upitnik je ispunilo N (%): 0 (0), 929 (97), 971 (98) i 661 (93) ispitanika uzorkovanih na području otoka Visa, grada i okolice Korčule, grada Splita, te sela i okolice Smokvice (otok Korčula).

2.2.1.16 Berlinski upitnik za procjenu rizika opstruktivske apneje u spavanju

Berlinski upitnik za procjenu rizika opstruktivske apneje u spavanju (*engl.* Berlin questionnaire) je razvijen s ciljem identifikacije pojedinaca s opstruktivskom apnejom u spavanju (92, 93).

Rizik (niski/visoki) se procjenjuje na temelju dva seta pitanja (o hrkanju i pospanosti) te na temelju hipertenzije i indeksa tjelesne mase. Ukupan rezultat može poprimiti vrijednosti od nula do tri, pri čemu vrijednosti dva i tri sugeriraju na visoki rizik od opstruktivske apneje u spavanju (69).

Upitnik je ispunilo N (%): 0 (0), 936 (97), 975 (99) i 679 (96) ispitanika uzorkovanih na području otoka Visa, grada i okolice Korčule, grada Splita, te sela i okolice Smokvice (otok Korčula).

2.2.1.17 Upitnik Europske zajednice o respiratornom zdravlju

Kratki upitnik Europske zajednice o respiratornom zdravlju (*engl.* European Community Respiratory Health Survey, ECRHS) je razvijen s ciljem prikupljanja informacija o incidenciji i prognozi alergija, alergijskih bolesti i slabe funkcije pluća u odraslih (94, 95). Upitnik se sastoji od 13 pitanja s ponuđenim da-ne odgovorima. Ukupan rezultat je dobiven brojanjem pitanja na koje je ispitanik odgovorio s „da“.

Upitnik je ispunilo N (%): 965 (99), 956 (99), 978 (99) i 691 (97) ispitanika uzorkovanih na području otoka Visa, grada i okolice Korčule, grada Splita, te sela i okolice Smokvice (otok Korčula).

2.2.1.18 Upitnik Svjetske zdravstvene organizacije o angina pectoris

Upitnik Svjetske zdravstvene organizacije o *angina pectoris* (engl. WHO Rose angina questionnaire) je razvijen s ciljem detekcije ishemijske srčane boli (*angina pectoris* i infarkt miokarda) (96). Sastoji se od šest pitanja vezanih uz bol u predjelu prsnog koša, a ukupni rezultat se može vrednovati na tri načina bodujući prva tri pitanja, prvih pet pitanja ili svih šest pitanja (97). Za potrebe određivanja povezanosti rezultata ovog upitnika i N-glikoma IgG-a korišten je rezultat dobiven bodovanjem svih šest pitanja.

Upitnik je ispunilo N (%): 816 (80), 820 (80), 752 (76) i 532 (75) ispitanika uzorkovanih na području otoka Visa, grada i okolice Korčule, grada Splita, te sela i okolice Smokvice (otok Korčula).

2.2.1.19 Upitnik Svjetske zdravstvene organizacije o intermitentnoj klaudikaciji

Upitnik Svjetske zdravstvene organizacije o intermitentnoj klaudikaciji (engl. WHO Questionnaire on intermittent claudication) je razvijen s ciljem procjene rizika intermitentne klaudikacije kao znaka arterioskleroze (96, 98). Sastoji se od 7 pitanja. Na temelju 6 pitanja se procjenjuje postoji li rizik od intermitentne klaudikacije (da/ne), dok jedno pitanje razgraničava između prvog i drugog stupnja bolesti.

Upitnik su ispunili samo ispitanici uzorkovani na području otoka Visa, N(%) = 956 (94).

2.2.1.20 Sociološko ponašanje

Tablica 13. Ukupan broj ispitanika za koje postoje podaci dobiveni upitnikom o sociološkom ponašanju među svim ispitanicima prisutnim u biobazi 10001 Dalmatinac kojima je određen N-glikom imunoglobulina G. Broj u zagradi prikazuje postotak ispitanika.

	otok Vis	grad Korčula	grad Split	selo i okolica Smokvice
Skлонost riziku	362 (35)	0 (0)	478 (48)	653 (92)
Stopa preferencije vremena	355 (35)	0 (0)	250 (25)	679 (96)
Osjećaj povjerenja	361 (35)	0 (0)	477 (48)	679 (96)
Ekonomska poštenost	359 (35)	0 (0)	478 (48)	679 (96)

Podaci o sociološkom ponašanju su prikupljeni putem upitnika. Kroz blokove od po 3 pitanja su procijenjeni sklonost riziku, stopa preferencije vremena, osjećaj povjerenja, osjećaj za ekonomsku poštenost (99). Tablicom su prikazani brojevi ispitanika za varijable korištene u sklopu ove disertacije (Tablica 13).

2.2.1.21 Revidirani Eysenckov test osobnosti

Revidirani Eysenckov test osobnosti (*engl.* Revised Eysenck Personality Questionnaire, EPQ-R) je upitnik kojim se procjenjuju 4 dimenzije osobnosti: ekstravertizam, neuroticizam, psihoticizam i sklonost laganju (100). Tablicom su prikazani brojevi ispitanika za dimenzije osobnosti procijenjene ovim upitnikom (Tablica 14).

Tablica 14. Ukupan broj ispitanika za koje postoje podaci dobiveni upitnikom o sociološkom ponašanju među svim ispitanicima prisutnim u biobazi 10001 Dalmatinac kojima je određen N-glikom imunoglobulina G. Broj u zagradi prikazuje postotak ispitanika.

	otok Vis	grad Korčula	grad Split	selo i okolica Smokvice
Ekstravertizam	929 (91)	0 (0)	0 (0)	609 (86)
Neuroticizam	932 (91)	0 (0)	0 (0)	649 (91)
Psihoticizam	931 (91)	0 (0)	0 (0)	644 (91)
Sklonost laganju	945 (93)	0 (0)	0 (0)	623 (88)

2.2.2 Statistička analiza podataka

2.2.2.1 Obrada N-glikoma prije analize

Prije početka analize rezultati kromatografskih analiza N-glikoma su normalizirani s obzirom na ukupnu površinu ispod cijelog kromatograma dobivenog HILIC-UPLC metodom, odnosno ispod dijela kromatograma koji odgovara pojedinoj klasi imunoglobulina G dobivenog LC-ESI-MS metodom. Normaliziranim podacima je uklonjena tehnička varijanca koja je rezultirala tzv. efektom šarže (*engl.* batch effect). Najprije su normalizirane vrijednosti logaritmom transformirane nakon čega je ComBat metodom (R paket sva (101)) uklonjen efekt šarže. Dobiveni podaci su transformirani natrag na ljestvicu udjela (*engl.* back-transformation). U svrhu lakšeg tumačenja pojedinih svojstava N-glikoma, izračunana su derivirana svojstva. Nazivi deriviranih svojstava i jednadžbe su navedene u tablicama (Tablica 15, Tablica 16).

2.2.2.2 Deskriptivna statistika

Radi konzistentnosti sve numeričke varijable, neovisno o normalnosti razdiobe, prikazane su kutijastim dijagramom (*engl.* boxplot). Medijan je naznačen debelom linijom unutar kutije, a rubovi kutije označavaju prvi (Q_1) i treći (Q_3) kvartil. Brkovi se protežu do zadnje vrijednosti unutarnjih granica definiranih kao $1,5 \times IQR$ ispod Q_1 i iznad Q_3 . Interkvartilni raspon (IQR) se definira kao razlika: $Q_3 - Q_1$. Ukoliko postoje vrijednost izvan unutarnjih granica (stršćeće vrijednosti, *engl.* outliers), prikazane su točkama. Relativne frekvencije, u svrhu sažimanja kategorijskih podataka su prikazane stupčastim dijagramima. Nedostajući podaci nisu ulazili u izračun relativnih frekvencija. Deskriptivna statistika je napravljena odvojeno za svako mjesto uzorkovanja (kohortu).

Tablica 15. Jednadžbe za izračun deriviranih svojstava *N*-glikoma imunoglobulina G određenog LC-ESI-MS metodom.

Derivirano svojstvo	Opis i jednadžba
IgG1_G0	Udio agalaktoziliranih glikana na IgG klase 1 (IgG1) IgG1_G0F + IgG1_G0FN + IgG1_G0 + IgG1_G0N
IgG1_G1	Udio asijaliniziranih monogalaktoziliranih glikana na IgG1 IgG1_G1F + IgG1_G1FN + IgG1_G1 + IgG1_G1N
IgG1_G2	Udio asijaliniziranih digalaktoziliranih glikana na IgG1 IgG1_G2F + IgG1_G2FN + IgG1_G2 + IgG1_G2N
IgG1_G	Udio asijaliniziranih galaktoziliranih glikana na IgG1 IgG1_G1F + IgG1_G2F + IgG1_G1FN + IgG1_G2FN + IgG1_G1 + IgG1_G2 + IgG1_G1N + IgG1_G2N
IgG1_CF	Udio glikana sa sržnom fukozom na IgG1 IgG1_G0F + IgG1_G1F + IgG1_G2F + IgG1_G0FN + IgG1_G1FN + IgG1_G2FN + IgG1_G1FS + IgG1_G2FS + IgG1_G1FNS + IgG1_G2FNS
IgG1_S	Udio sijaliniziranih glikana na IgG1 IgG1_G1FS + IgG1_G2FS + IgG1_G1FNS + IgG1_G2FNS + IgG1_G1S + IgG1_G2S + IgG1_G1NS + IgG1_G2NS
IgG1_B	Udio glikana sa račvujućim GlcNAc-om na IgG1 IgG1_G0FN + IgG1_G1FN + IgG1_G2FN + IgG1_G1FNS + IgG1_G2FNS + IgG1_G0N + IgG1_G1N + IgG1_G2N + IgG1_G1NS + IgG1_G2NS
IgG2_G0	Udio agalaktoziliranih glikana na IgG klase 2 (IgG2) IgG2_G0F + IgG2_G0FN + IgG2_G0 + IgG2_G0N
IgG2_G1	Udio asijaliniziranih monogalaktoziliranih glikana IgG2 IgG2_G1F + IgG2_G1FN + IgG2_G1 + IgG2_G1N
IgG2_G2	Udio asijaliniziranih digalaktoziliranih glikana IgG2 IgG2_G2F + IgG2_G2FN + IgG2_G2 + IgG2_G2N
IgG2_G	Udio asijaliniziranih galaktoziliranih glikana IgG2 IgG2_G1F + IgG2_G2F + IgG2_G1FN + IgG2_G2FN + IgG2_G1 + IgG2_G2 + IgG2_G1N + IgG2_G2N
IgG2_CF	Udio glikana sa sržnom fukozom na IgG2 IgG2_G0F + IgG2_G1F + IgG2_G2F + IgG2_G0FN + IgG2_G1FN + IgG2_G2FN + IgG2_G1FS + IgG2_G2FS + IgG2_G1FNS + IgG2_G2FNS
IgG2_S	Udio sijaliniziranih glikana na IgG2 IgG2_G1FS + IgG2_G2FS + IgG2_G1FNS + IgG2_G2FNS + IgG2_G1S + IgG2_G2S + IgG2_G1NS + IgG2_G2NS
IgG2_B	Udio glikana sa račvujućim GlcNAc-om na IgG2 IgG2_G0FN + IgG2_G1FN + IgG2_G2FN + IgG2_G1FNS + IgG2_G2FNS + IgG2_G0N + IgG2_G1N + IgG2_G2N + IgG2_G1NS + IgG2_G2NS
IgG4_G0	Udio agalaktoziliranih glikana na IgG klase 4 (IgG4) IgG4_G0F + IgG4_G0FN
IgG4_G1	Udio asijaliniziranih monogalaktoziliranih glikana na IgG4 IgG4_G1F + IgG4_G1FN
IgG4_G2	Udio asijaliniziranih digalaktoziliranih glikana na IgG4 IgG4_G2F + IgG4_G2FN
IgG4_G	Udio asijaliniziranih galaktoziliranih glikana na IgG4 IgG4_G1F + IgG4_G2F + IgG4_G1FN + IgG4_G2FN

Derivirano svojstvo	Opis i jednadžba
IgG4_S	Udio sijaliniziranih glikana na IgG4 IgG4_G1FS + IgG4_G2FS + IgG4_G1FNS + IgG4_G2FNS
IgG4_B	Udio glikana sa računjućim GlcNAc-om na IgG4 IgG4_G0FN + IgG4_G1FN + IgG4_G2FN + IgG4_G1FNS + IgG4_G2FNS

Tablica 16. Jednadžbe za izračun deriviranih svojstava N-glikoma imunoglobulina G određenog UPLC metodom.

Derivirano svojstvo	Opis i jednadžba
B	Udio glikana sa računjućim GlcNAc-om GP6 + GP10 + GP11 + GP13 + GP15 + GP19 + GP22 + GP24
CF	Udio glikana sa sržnom fukozom GP1 + GP4 + GP6 + GP8 + GP9 + GP10 + GP11 + GP14 + GP15 + GP16 + GP18 + GP19 + GP23 + GP24
HM	Udio oligomanoznih glikana GP5
G0	Udio agalaktoziliranih glikana GP1 + GP2 + GP4 + GP6
G1	Udio asijaliniziranih monogalaktoziliranih glikana GP7 + GP8 + GP9 + GP10 + GP11
G2	Udio asijaliniziranih digalaktoziliranih glikana GP12 + GP13 + GP14 + GP15
S0	Udio asijaliniziranih glikana GP1 + GP2 + GP4 + GP6 + GP7 + GP8 + GP9 + GP10 + GP11 + GP12 + GP13 + GP14 + GP15
S1	Udio monosijaliniziranih glikana GP16 + GP17 + GP18 + GP19
S2	Udio disijaliniziranih glikana GP21 + GP22 + GP23 + GP24
G	Udio asijaliniziranih galaktoziliranih glikan GP7 + GP8 + GP9 + GP10 + GP11 + GP12 + GP13 + GP14 + GP15
S	Udio sijaliniziranih glikana GP16 + GP17 + GP18 + GP19 + GP21 + GP22 + GP23 + GP24

2.2.2.3 Inferencijalna statistika

Povezanost N-glikoma i različitih varijabli prisutnih u biobazi „10001 Dalmatinac“ je napravljena mješovitim (u slučaju dostupnosti podataka sa više mjesta uzorkovanja) ili linearnim modelom (u slučaju kada su podaci dostupni sa samo jednog mjesta uzorkovanja). Za svaki par *derivirano svojstvo* N-glikoma (*glikan*) – *varijabla iz biobaze „10001 Dalmatinac“* napravljen je model u kojem je *razina glikana* modelirana kao zavisna varijabla, a varijabla iz biobaze kao nezavisna varijabla. *Spol* i *dob* su uključeni kao dodatni nezavisni faktori. U slučaju dostupnosti podataka iz više kohorti, kao nasumični odsječak uključeno je i *mjesto uzorkovanja* u model (102). Sve numeričke i ordinalne varijable su prije modeliranja

transformirane reverznom transformacijom rangova u standardnu normalnu razdiobu (rankit transformacija) (103, 104).

S ciljem smanjenja broja varijabli, kompletan set antropometrijskih podataka je podvrgnut faktorskoj analizi uz promax rotaciju s ciljem identifikacije faktora koji predstavljaju grupu varijabli koje su međusobno korelirane (105). Optimalni okvirni broj faktora je procijenjen na temelju grafičkog prikaza svojstvenih vrijednosti faktora (*engl.* eigenvalues) – *scree* test. Faktorska je analiza napravljena za svaku kohortu posebno za sve varijable koje su izmjerene kod barem 60 % ispitanika. Nakon identifikacije faktora, izračunat je istoimeni indeks za svaku kohortu posebno. Indeks je za svakog ispitanika izračunat kao prosječna vrijednosti svih dostupnih varijabli koju faktor sadrži, a te su varijable prethodno bile transformirane reverznom transformacijom rangova u standardnu normalnu razdiobu za svaku kohortu posebno. Jednakim je postupkom napravljena faktorska analiza i za upitnik o učestalosti unosa hrane i pića, te upitnik o preferenciji pojedinih prehrambenih proizvoda.

Analiza povezanosti *N*-glikoma imunoglobulina G s budućim razvojem hipertenzije napravljena je mješovitim Coxovim regresijskim modelom (106). Analiza preživljenja (*engl.* survival analysis) je sadržavala *vrijeme do razvoja hipertenzije* kao zavisnu varijablu te *razinu glikana* kao nezavisnu varijablu. *Spol, dob, indeks tjelesne mase i srednji arterijski tlak* su također uključeni kao dodatni nezavisni faktori, a *mjesto uzorkovanja* je uključeno kao nasumični odsječak. Glikani su prije modeliranja transformirani reverznom transformacijom rangova u standardnu normalnu razdiobu (103, 104). U analizi su korišteni samo ispitanici koji u trenutku mjerenja glikana nisu imali hipertenziju.

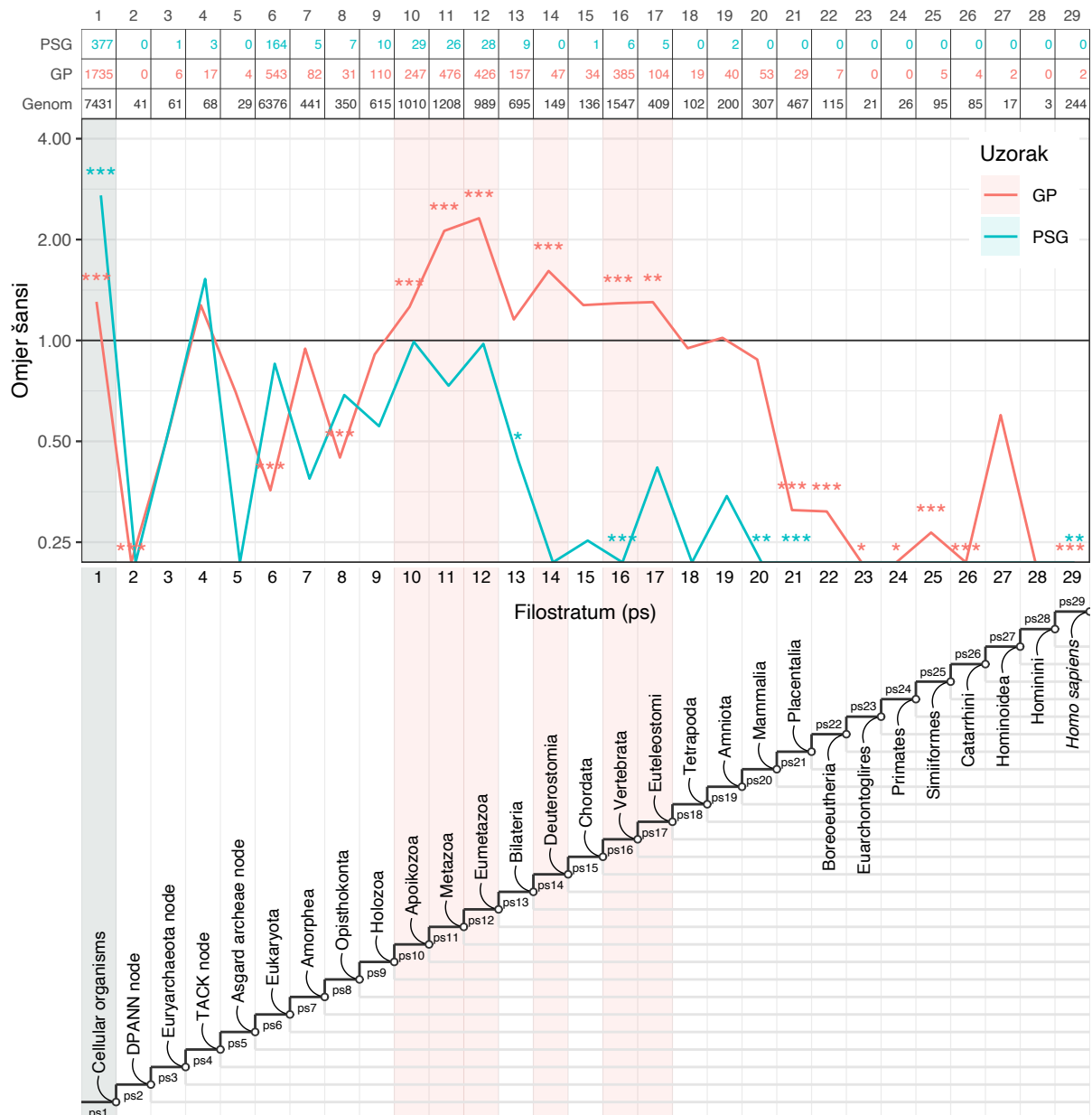
Razina statističke značajnosti je postavljena na $\alpha = 0,05$. Obzirom na velik broj testova koji se provodi očekivan je i velik broj lažno statistički značajnih rezultata, stoga su sve dobivene *p* vrijednosti prilagođene višestrukom testiranju prema Benjamini-Hochberg metodi (107).

3 REZULTATI I RASPRAVA

3.1 MAKROEVOLUCIJSKA POVIJEST GLIKOZILACIJE

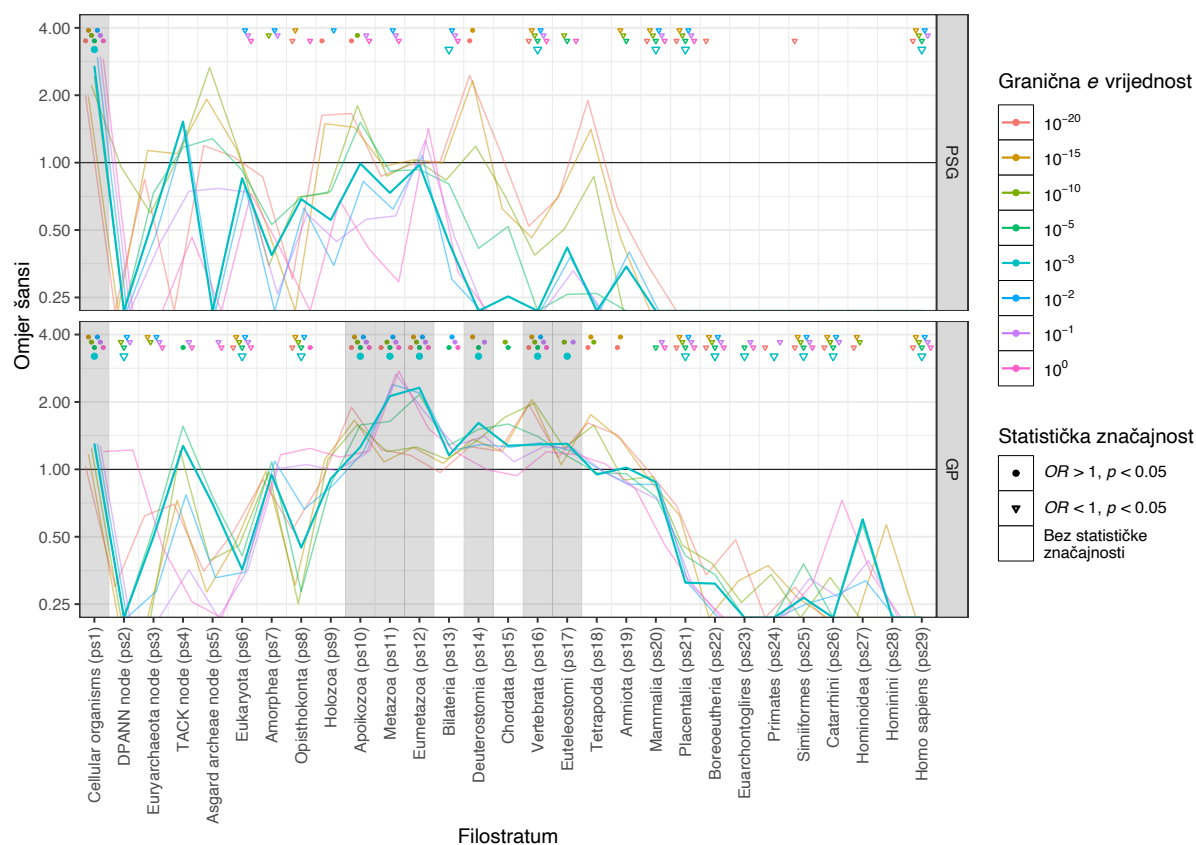
Genomska je filostratigrafija primijenjena za procjenu evolucijske starosti ljudskih gena koji kodiraju za proteine koji sudjeluju u procesu glikozilacije i gena koji kodiraju za polipeptidne lance glikoproteina.

Svim uzorkovanim genima koji kodiraju za polipeptidne lance glikoproteina (GP) je uspješno određen filostratum. Među njima, najveći je dio – čak 38 %, nastao još u zajedničkom pretku svih staničnih organizama (cellular organisms, ps1) što je statistički značajno više od očekivanih 32 % ($OR = 1,30$; $p < 0,001$). Osim toga, statistički značajni signal nastanka novih gena vidljiv je i u evolucijskom periodu od nastanka zajedničkog pretka svih opisthokonta koji se mogu grupirati u kolonije (Apoikozoa, ps10), do nastanka zajedničkog pretka euteleostoma (životinja sa pravim kosturom) (Euteleostomi, ps17). U tom je periodu nastalo ukupno 37% gena koji kodiraju za polipeptidne lance glikoproteina pri čemu je omjer šansi statistički značajno veći od 1 i za pojedini filostratum se kreće u rasponu 1,26 – 2,32. Iznimka su filostratumi ps13 (Bilateria, zajednički predak svih bilateralno simetričnih životinja) i ps15 (Chordata, zajednički predak svih životinja s unutrašnjim potpornim sustavom) kod kojih je nastalo više gena nego što je očekivano u tom periodu, no taj porast nije statistički značajan ($OR_{ps13} = 1,16$; $p_{ps13} = 0,072$; $OR_{ps15} = 1,27$; $p_{ps15} = 0,143$). Rezultati su detaljnije prikazani filostratigramom (Slika 4, crvena krivulja). Ovi rezultati upućuju na činjenicu da se glikozilacija u evoluciji čovjeka intenzivno počela primjenjivati pojavom prvih organizama koji su se mogli grupirati u kolonije, za što je bilo potrebno razviti sustav komunikacije između stanica. Ubrzani razvoj polipeptidnih lanca glikoproteina se nastavlja sve do kralježnjaka s pravim kosturom (Euteleostoma, ps17) čime je obuhvaćen razvoj od jednostavnih do kompleksnih životinja. Dobivene rezultate treba pažljivo tumačiti budući da prikazuju evolucijski razvoj polipeptidnih lanaca proteina koji su danas, kod čovjeka glikozilirani, a pritom nije moguće sasvim točno znati *kada* su oni postali glikozilirani. Ipak, u prilog govori činjenica da su glikozilacijska mjesta visokokonzervirana te da glikoproteini imaju znatno nižu stopu promjene u evolucijskom razvoju vrsta u usporedbi s neglikoziliranim proteinima (108–110).



Slika 4. Filostratigram gena koji kodiraju za glikoproteine (GP), i gena koji kodiraju za proteine koji sudjeluju u procesu glikozilacije (PSG). Opis pojedinog dijela filostratigrama je naveden u poglavlju 2.1.2.

Evolucijska starost je također uspješno pripisana i svim genima koji kodiraju za proteine koji sudjeluju u procesu glikozilacije (PSG). Ukupno 56 % ovih gena je sljedivo do zajedničkog pretka svih staničnih organizama (cellular organisms, ps1), što je statistički značajno više od očekivanih 32 % ($OR = 2,71; p < 0,001$). Ostatak gena koji kodiraju za proteine koji sudjeluju u procesu glikozilacije vuče korijene najvećim dijelom iz zajedničkog pretka svih eukariota – čak 24 %, te iz perioda od nastanka zajedničkog pretka Apoikozoa (ps10) do nastanka zajedničkog pretka svih Eumetazoa (ps12) – ukupno 12 %, ali to je i u skladu s očekivanjem obzirom na udio svih gena čovjeka koji su nastali u tim periodima (Slika 4).

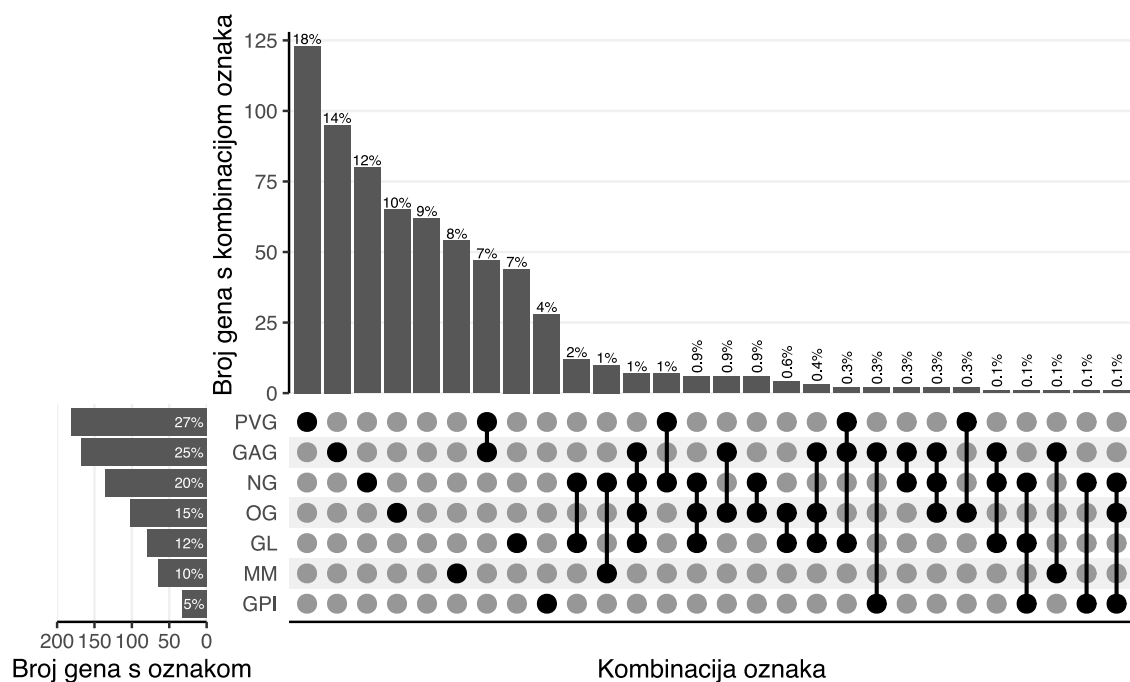


Slika 5. Stabilnost signala uočenih na filostratigramu gena koji kodiraju za glikoproteine (GP, donji panel) i gena koji kodiraju za proteine koji sudjeluju u procesu glikozilacije (PSG, gornji panel) (Slika 4). Različitim bojama su naglašene različite granične e vrijednosti, a statistička značajnost je naznačena krugom ili trokutom ovisno o vrijednosti omjera šanse.

Spomenuti su rezultati bazirani na e vrijednosti 10^{-3} , no uočeni signali oba uzoraka gena koji kodiraju za GP odnosno PSG ostaju stabilni i u području ostalih e vrijednosti intervala $10^{-20} - 1$ (Slika 5).

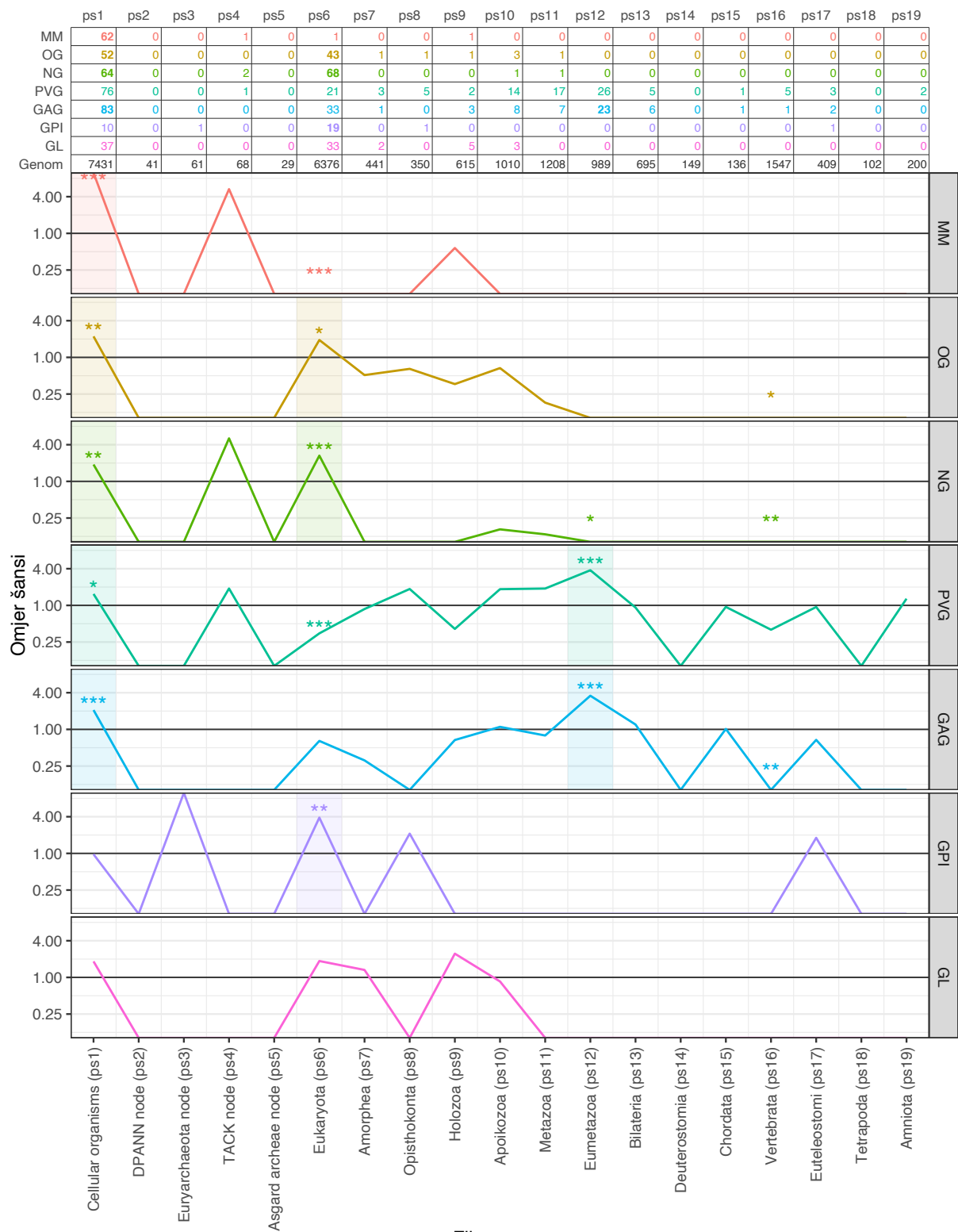
Kako je navedeno u uvodu, dosadašnja istraživanja evolucije proteina uključenih u glikozilaciju su obuhvatila samo manje dijelove glikozilacije, najčešće vezane uz dio biosintetskog puta N -glikozilacije ili pojedine vrste enzima. S ciljem detaljnijeg uvida u rezultate i u svrhu usporedivosti rezultata s drugim istraživanjima, promatranim genima koji kodiraju za PSG su pripisane oznake (prema poglavlju 2.1.1.1). Razdioba udjela gena po kombinacijama oznaka pokazuje da je 73 % gena koji kodiraju za PSG rezervirano samo za jednu oznaku. Istovremeno 18 % gena ima dvije ili više oznaka, a od toga najviše čine geni koji kodiraju za proteine koji vežu glikane (Slika 6).

S ciljem detaljnijeg uvida u evolucijsku starost gena koji kodiraju za PSG filostratigrafska analiza napravljena za svaku pojedinu skupinu gena kojoj je pripisana određena oznaka. Za pet od sedam oznaka (MM, OG, NG, PVG i GAG), uočen je očekivani signal u ps1 kao i za ukupan skup gena koji kodiraju za PSG (Slika 7).



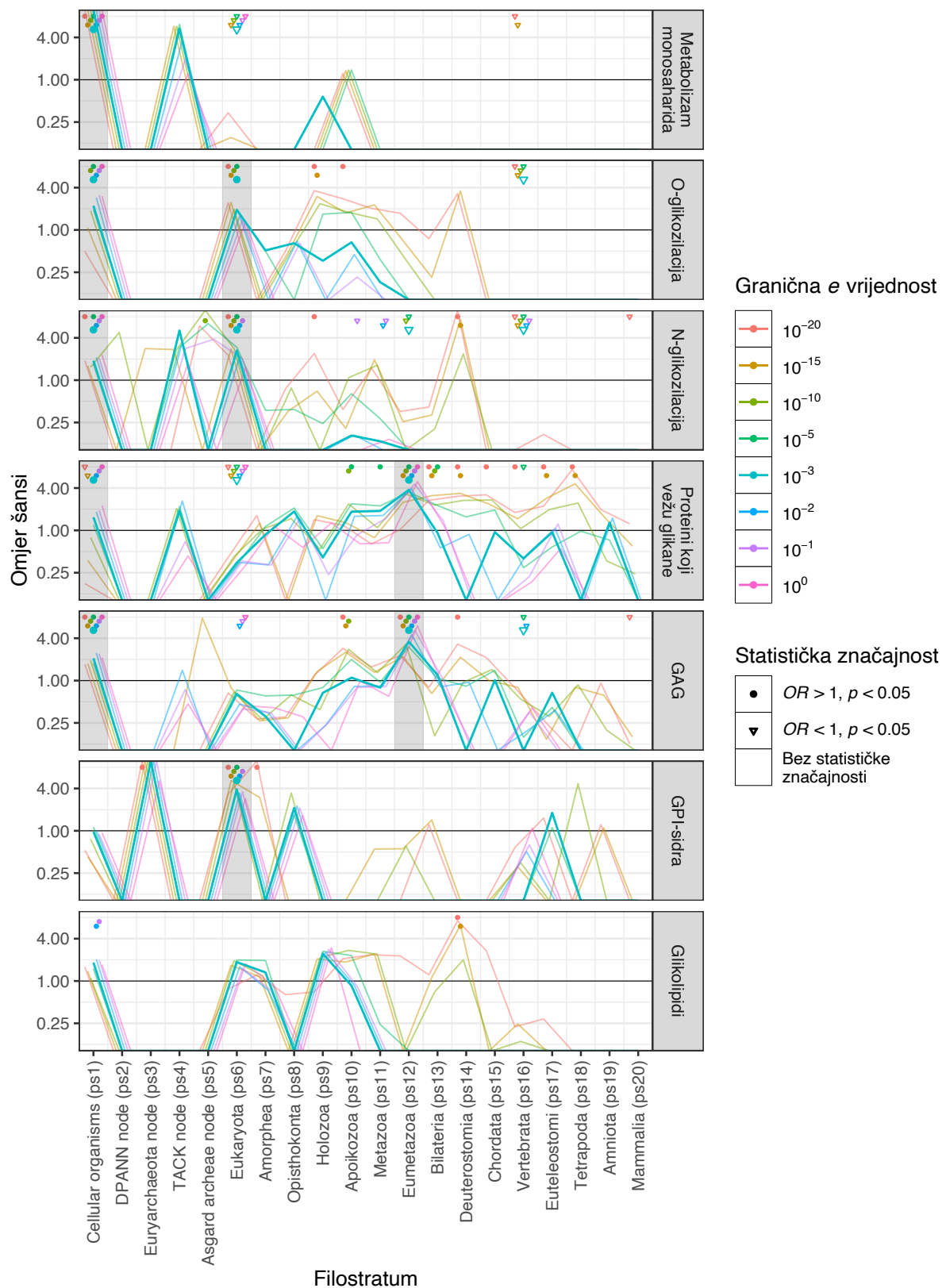
Slika 6. *UpSet* dijagram. Stupčasti dijagram na lijevoj strani prikazuje broj gena s pojednom oznakom, a navedeni postoci prikazuju udjele gena s pojednom oznakom u ukupnom skupu gena. Stupičasti dijagram na gornjoj strani prikazuje broj gena s kombinacijom oznaka, a navedeni postoci prikazuju udio gena s pojednom oznakom u ukupnom skupu gena. Dolje desno se nalazi točkasti dijagram iz kojeg je moguće iščitati kombinaciju oznaka. Kratice su kodirane sukladno poglavlju 2.1.1.1.

Dodatno, statistički značajno veći udio novonastalih gena od očekivanog je nastao u skupovima gena koji kodiraju za PSG s oznakom *O*-glikozilacije (OG, $OR = 1,93$; $p = 0,013$), *N*-glikozilacije (NG, $OR = 2,64$; $p = 3,89 \times 10^{-7}$) i glikozilfosfatidilinozitolnih sidra (GPI, $OR = 3,86$; $p = 2,04 \times 10^{-3}$) u zajedničkom pretku svih eukariota (Eukaryota, ps6). Zabilježen je i statistički značajno veći udio novonastalih gena od očekivanog među genima s oznakom glikozaminoglikana (GAG, $OR = 3,57$; $p = 1,17 \times 10^{-5}$) i proteina koji vežu glikane (PVG, $OR = 3,77$; $p = 8,09 \times 10^{-7}$) kod zajedničkog pretka svih životinja koje razvijaju prava tkiva organizirana u stanične slojeve (Eumetazoa, ps12). Skup gena s oznakom glikolipida (GL) nije pokazao nikakva odstupanja od evolucijskog razvoja ostalih gena čovjeka. Opaženi su rezultati su stabilni i pri drugim korištenim graničnim e vrijednostima u intervalu $10^{-20} - 10^{-3}$ (Slika 8). Zanimljivo je da je značajan udio ljudskih gena koji su povezani sa glikozaminoglikanima i proteinima koji vežu glikane inovacija zajedničkog pretka svih Eumetazoana (ps12) – pretka svih životinjskih organizama osim spužvi i Placozoa (najjednostavnije životinje). Jedna od glavnih karakteristika Eumetazona, koje još nazivamo i "pravim životinjama", uključuje tkiva organizirana u stanične slojeve.



Filostratum

Slika 7. Filostratigram svih gena koji kodiraju za proteine koji sudjeluju u procesu glikozilacije (PSG) raspoređenim u panele prema dodijeljenim oznakama: metabolizam monosaharida (MM), O-glikozilacija (OG), N-glikozilacija (NG), proteini koji vežu glikane (PVG), glikozaminoglikani (GAG), glikozilfosfatidilinozitolna sidra (GPI) i glikolipidi (GL). Iako korišteni u analizi, svi filostratumi koji sadrže organizme evolucijski mlađe od zajedničkog pretka svih reptila i sisavaca (Amniota, ps19) nisu prikazani jer u njima nije zabilježen nastanak gena koji kodiraju za PSG.



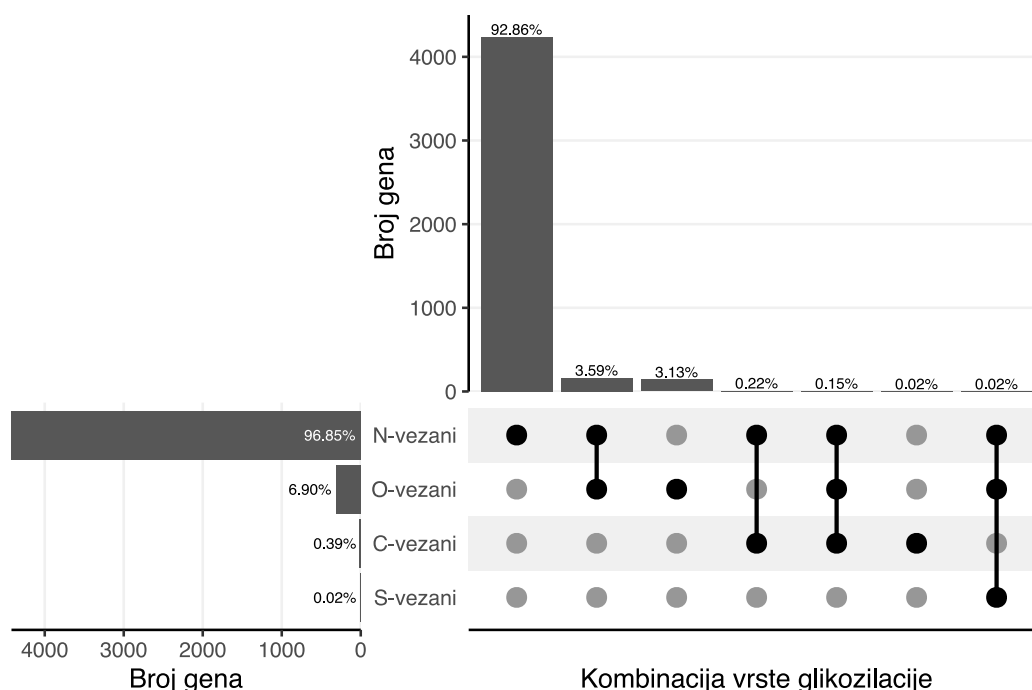
Slika 8. Filostratigram svih gena koji kodiraju za proteine koji sudjeluju u procesu glikozilacije (PSG) raspoređenim u panele prema dodijeljenim oznakama (Slika 7). Različitim bojama su naglašene različite granične e vrijednosti, a statistička značajnost je naznačena krugom ili trokutom ovisno o vrijednosti omjera šanse. GAG – glikozaminoglikani; GPI-sidra – glikozilfosfatidilinozitolna sidra.

Glikozaminoglikani igraju glavnu ulogu u izgradnji izvanstanične matrice i sloja endotelnog glikokaliksa (*engl.* endothelial glycocalyx layer) – koji su osnova za strukturu i funkcionalnost tkiva (111, 112). S druge strane, proteini koji vežu glikane sudjeluju u staničnoj adheziji, te su važni u brojnim funkcijama višestaničnih organizama poput: klirensa proteina, upravljanja interakcijama između stanica ili interakcijama stanica sa izvanstaničnim matriksom što također uključuje i imuni odgovor temeljen na razlikovanju vlastitog i tuđeg (*engl.* self-nonself recognition) (6).

Po svemu sudeći, čini se da glikozaminoglikani i proteini koji vežu glikane igraju važnu ulogu u evolucijskom razvoju kompleksnijih životinja. Kada su stanice ranih Eumetazoa uspostavile komunikaciju putem glikozaminoglikana i proteina koji vežu glikane, glikozilacija je postala važan alat, pogotovo u razlikovanju vlastitog i tuđeg.

Iako su biosintetski putevi glikozilacije relativno stari, rezultati vezani uz evolucijski razvoj glikoproteina sugeriraju da se glikozilacija počinje intenzivnije koristiti tek nakon razvoja višestaničnih životinja (Eumetazoa, ps12).

Među proučavanim glikoproteinima najčešća vrsta glikozilacije bila je N-glikozilacija (Slika 9).



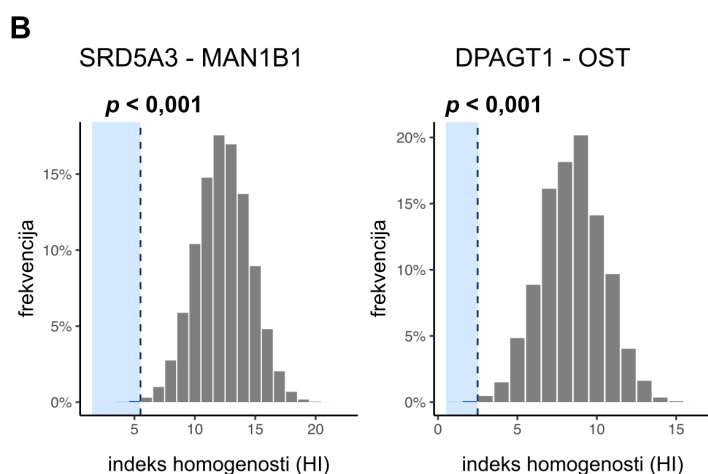
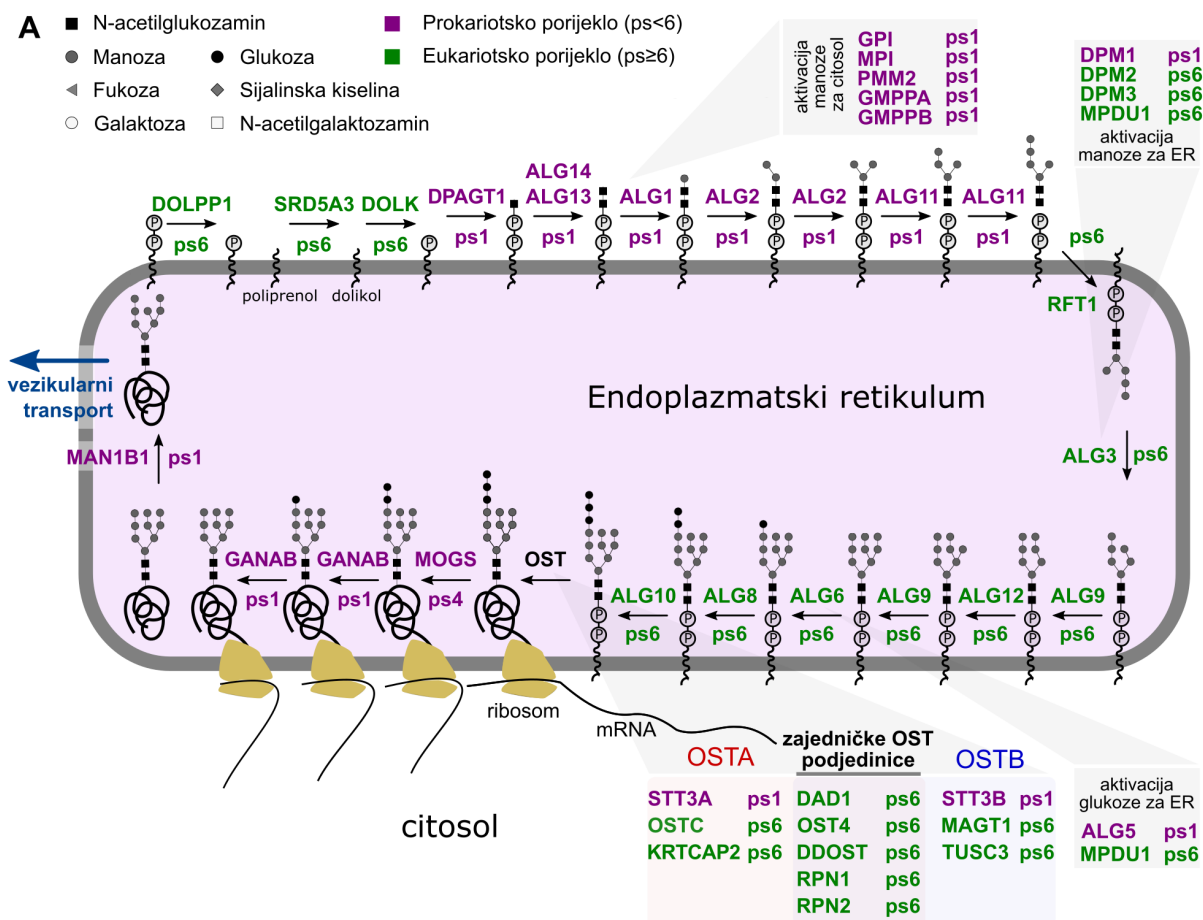
Slika 9. *UpSet* dijagram. Stupčasti dijagram na lijevoj strani prikazuje broj gena koji kodiraju za glikoproteine s pojednom vrstom glikozilacije (uzorak prema 2.1.1.2), a navedeni postoci prikazuju udjele s pojednom vrstom glikozilacije u ukupnom skupu proučavanih gena koji kodiraju za glikoproteine. Stupčasti dijagram na gornjoj strani prikazuje broj gena s kombinacijom vrsta glikozilacija, a navedeni postoci prikazuju udio gena s pojednom kombinacijom u ukupnom skupu gena. Dolje desno se nalazi točkasti dijagram iz kojeg je moguće iščitati kombinaciju vrsta glikozilacije.

S ciljem detaljnijeg uvida u evolucijsku starost gena koji kodiraju za proteine koji sudjeluju u biosintetskom putu *N*-glikozilacije, dobiveni su rezultati evolucijske starosti (filostratumi) naznačeni na shemi biosintetskog puta *N*-glikozilacije koja se odvija u endoplazmatskom retikulumu (Slika 10) te shemi procesiranja *N*-glikana koje odvija u Golgijevom aparatu (Slika 11).

Na prikazanim su shemama *N*-glikozilacije u endoplazmatskom retikulumu evolucijske starosti naglašene bojama, ovisno o tome jesu li geni prokariotskog ($ps < 6$, ljubičasto) ili eukariotskog ($ps \geq 6$, zeleno) porijekla.

U dijelu biosintetskog puta *N*-glikana koji se odvija u endoplazmatskom retikulumu je vidljivo da završne dijelove sinteze i obnovu dolikol-fosfata kataliziraju evolucijski noviji enzimi koje nalazimo u zajedničkom pretku svih eukariota ($ps6$) što je u suglasju s već objavljenim rezultatima istraživanja koja su proučavala evolucijsko porijeklo lipidnih nosača u *N*-glikozilaciji (113). U nastavku, od dodatka prvog monosaharida na dolikol-fosfat pa sve do zadnje reakcije koja se odvija na citosolnoj strani membrane endoplazmatskog retikuluma kataliziraju evolucijski stariji enzimi (kodirani genima DPAGT1, ALG13, ALG14, ALG1, ALG 2 i ALG11) koje nalazimo već u zajedničkom pretku svih staničnih organizama ($ps1$). Flipaza (kodirana genom RFT1) je evolucijski novija – eukariotskog je porijekla, kao i svi slijedeći enzimi (kodirani genima ALG3, ALG9, ALG12, ALG6, ALG8 i ALG10) koji kataliziraju reakcije lokalizirane u lumenu endoplazmatskog retikuluma do reakcije katalizirane oligosahariltransferazom (OST). Proteinski kompleks OST sastoji se od katalitičke podjedinice kodirane genom STT3A ili STT3B koje pronalazimo u zajedničkom pretku svih staničnih organizama ($ps1$). Ostale, regulatorne podjedinice su eukariotskog porijekla ($ps6$). Nakon vezanja oligosaharida na polipeptidni lanac, slijede reakcije dviju glikozidaza koje su uz kalretikulin i kalneksin (nisu prikazani na shemi) dio kontrole kvalitete pravilnog smatanja proteina (114). Kodirane su genom MOGS koji se nalazi u zajedničkom pretku svih organizama koji su obuhvaćenim superkoljenom TACK, te genom GANAB koji se nalazi u zajedničkom pretku svih staničnih organizama ($ps1$). Jednom kada je protein ispravno smotan, α -manozidaza uklanja manozu s centralne grane glikoproteina i priprema ga za prijenos izvan endoplazmatskog retikuluma. α -manozidaza je kodirana genom MAN1B1 koji vuče porijeklo iz zajedničkog pretka svih staničnih organizama ($ps1$) (Slika 10A).

Vizualizacijom evolucijske starosti bojama, vidljivo je da su enzimi grupirani u blokove. Pod pretpostavkom da slijed enzima i evolucijska starost nisu povezani, vjerojatnost da se uoči niz gena koji kodira za enzime nekog biosintetskog puta tako uredno poredanog porijekla je vrlo mala: za niz od početka obnove dolikol fosfata (SRD5A3) do zadnje reakcije prije izlaska



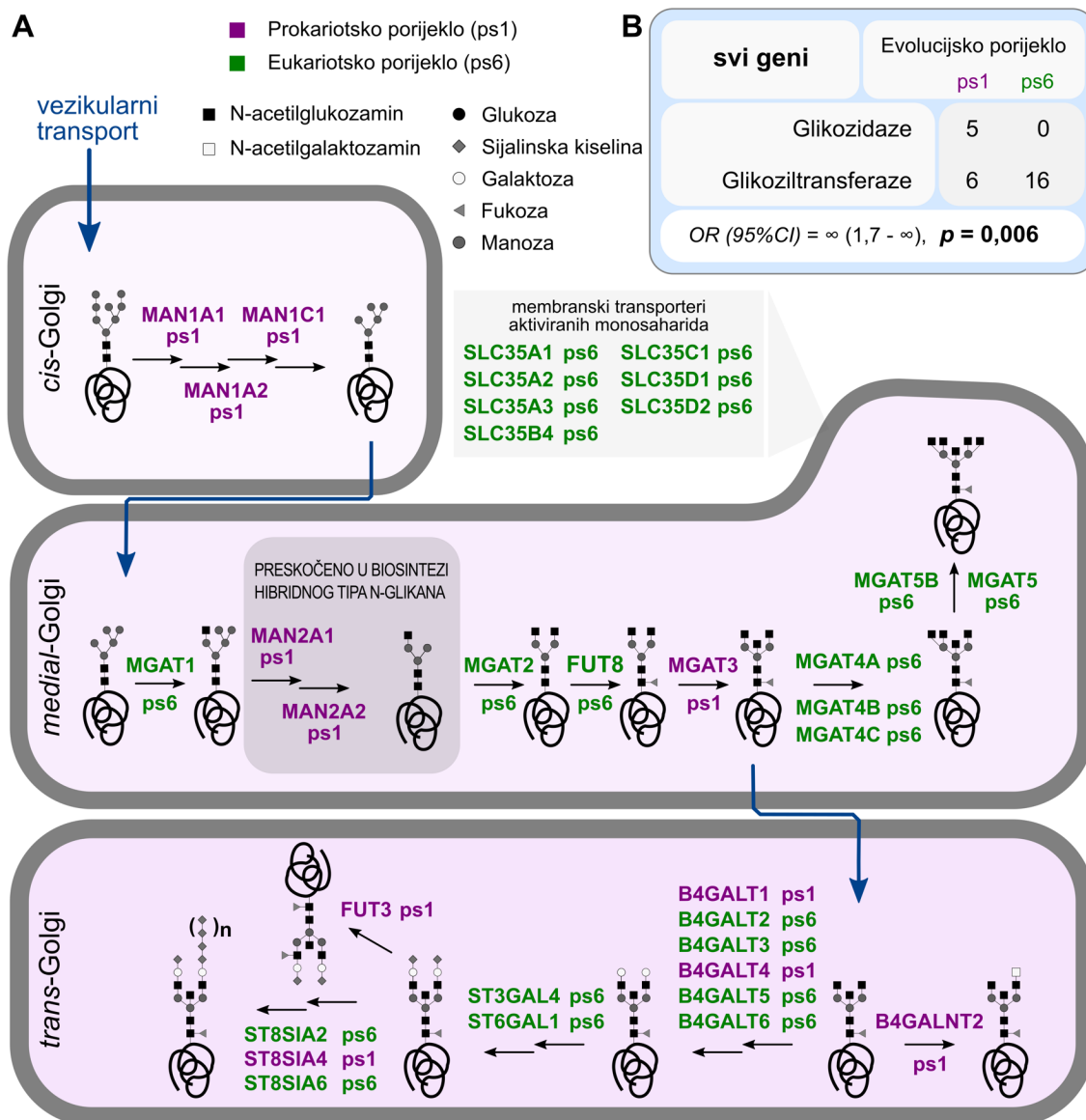
C

svi geni		Evolucijsko porijeklo	
		ps<6	ps6
Membranska topologija	citokol	13	5
	ER lumen	5	17
OR (95%CI) = 8,3 (1,8 - 47,6); $p = 0,004$			
DPAGT1 - OST		Evolucijsko porijeklo	
		ps1	ps6
Membranska topologija	citokol	6	0
	ER lumen	2	16
OR (95%CI) = ∞ (4,5 - ∞); $p < 0,001$			

Slika 10. (A) Shematski prikaz biosintetskog puta N-glikana u endoplazmatskom retikulumu čovjeka. Umjesto naziva enzima i transportera u biosintetskom putu, prikazani su HGNC identifikatori gena. Različitim je bojom označeno evolucijsko porijeklo gena. (B) Funkcija vjerojatnosti indeksa homogenosti dobivena permutacijom opaženih filostratuma za niz gena od SRD5A3 do MAN1B1, i od DPAGT1 - OST. Iscrtna linija prikazuje opaženi iznos indeksa homogenosti, a osjenčana površina obuhvaća opaženi i manji indeks homogenosti. Izračunata je p vrijednost navedena iznad iscrtnane linije (prema poglavlju 2.1.5.1). (C) Tablice kontingencije za sve gene i skup gena od DPAGT1 do OST kompleksa, koji su razdjeljeni prema membranskoj topologije proteina za kojeg kodiraju, te evolucijskog porijekla (prema poglavlju 2.1.5.2). HGNC - HUGO Gene Nomenclature Committee.

glikana iz endoplazmatskog retikuluma (MAN1B1), $p = 7 \times 10^{-4}$, te za niz od prve reakcije u kojoj sudjeluje glikan (DPAGT1) do prijenosa oligosaharida na polipeptidni lanac (OST), $p = 7 \times 10^{-4}$ (Slika 10B). Prilikom izračuna navedenih vjerojatnosti, od svih podjedinica OST korištena je samo katalitička podjedinica (STT3).

Razdvoje li se enzimi prema evolucijskom porijeklu i prema topologiji katalitičkog mjesta na membrani endoplazmatskog retikuluma također se može uočiti da su evolucijski stariji enzimi okrenuti prema citosolnoj strani, dok su evolucijski noviji enzimi okrenuti prema lumenu endoplazmatskog retikuluma.



Slika 11. (A) Shematski prikaz biosintetskog puta N-glikana u endoplazmatskom retikulumu čovjeka. Umjesto naziva enzima i transportera u biosintetskom putu, prikazani su HGNC identifikatori gena. Različitom je bojom označeno evolucijsko porijeklo gena. (B) Tablica kontingencije za sve enzime sa sheme koji su razdijeljeni prema vrsti katalizirane reakcije koju obavljaju proteini za koje kodiraju, te evolucijskog porijekla (prema poglavlju 2.1.5.2). HGNC - HUGO Gene Nomenclature Committee.

Pod pretpostavkom da ne postoji povezanost između membranske topologije i evolucijske starosti, vjerojatnost takve raspodjele je vrlo mala. Za gene koji kodiraju za sve proteine koje nalazimo na shemi (Slika 10) ona iznosi $p = 3,53 \times 10^{-3}$, a za gene koji kodiraju za proteine od prve reakcije u kojoj sudjeluje glikan do prijenosa oligosaharidu na polipeptidni lanac (DPAGT1 – OST) iznosi $p = 2,08 \times 10^{-4}$ (Slika 10C).

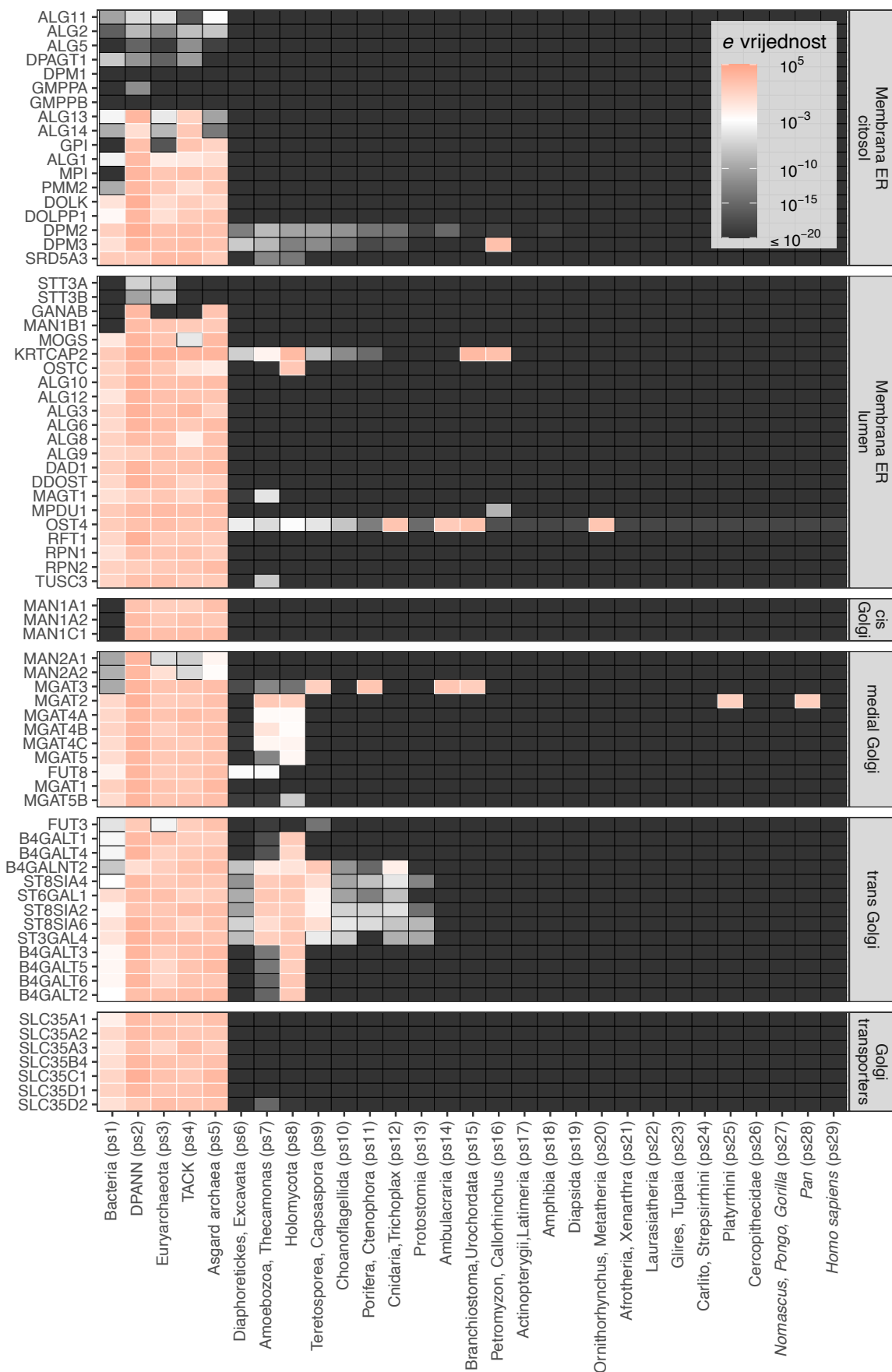
U dijelu biosintetskog puta *N*-glikana koji se odvija u Golgijevom aparatu prevladavaju većinom proteini čije porijeklo uspješno pratimo do zajedničkog pretka svih eukariota. Detaljnije je prikazano slikom (Slika 11A) na kojoj se, kao i kod endoplazmatskog retikuluma uočava razdioba gena u samo dva filostratuma (ps1 i ps6).

Manozidaze cis-odjeljka Golgijevog aparata, kao i manozidaze pronađene u medijalnom odjeljku Golgijevog aparata su uspješno praćene sve do zajedničkog pretka svih staničnih organizama (ps1). Ostali enzimi prikazani u medijalnom odjeljku Golgijevog aparata, koji su isključivo glikoziltransferaze, su inovacija zajedničkog pretka svih eukariota (ps6) uz iznimku *N*-acetilgalatozaminiltransferaze III (kodirane genom MGAT3), za koju je procijenjeno da je nastala još u zajedničkom pretku svih staničnih organizama (ps1). Glikoziltransferaze prisutne u trans-odjeljku Golgijevog aparata su podijeljene između ps1 i ps6: kao inovacije zajedničkog pretka možemo izdvojiti galaktoziltransferaze (osim onih kodiranih genima B4GALT1 i B4GALT4), sijaliltransferaze (osim one kodirane genom ST8SIA4), dok su *N*-acetilgalaktozaminiltransferaza (kodirana genom B4GALNT2) i fukoziltransferaza (kodirana genom FUT3) evolucijskog porijekla iz zajedničkog pretka svih staničnih organizama (ps1).

Grupiramo li enzime prema vrsti reakcije koju kataliziraju, uočavamo da oni koji uklanjaju monosaharide s oligosaharida (glikozidaze) pripadaju ps1, dok oni koji dodaju monosaharide na oligosaharid (glikoziltransferaze) pripadaju ps2. Tablica kontingencije je prikazana na slici (Slika 11B). Vjerojatnost takve razdiobe pod pretpostavkom da evolucijska starost gena nije povezana s vrstom reakcije koju enzim katalizira je mala ($p = 5,72 \times 10^{-3}$). To sugerira da je već zajednički predak svih staničnih organizama mogao razgrađivati glikane, ali glavnina gena odgovorna za procesiranje *N*-glikana je nastala u zajedničkom pretku svih eukariota.

Iznosi minimalnih *e* vrijednosti za sestrinske grupe pojedinih filostratuma su prikazane toplinskom kartom (Slika 12). Na karti su prikazani samo geni koje nalazimo na shematskim prikazima endoplazmatskog retikuluma i Golgijevog aparata.

Izražena binarna razdioba evolucijske starosti gena između ps1 i ps6 uz obrazac evolucijski starijih enzima smještenih na citosolnoj strani membrane endoplazmatskog retikuluma i evolucijski relativno novijih enzima smještenih na luminalnoj strani membrane



Slika 12. Toplinska karta minimalnih e vrijednosti po sestrijskim grupama filostratuma za gene prikazane slikama (Slika 10A, Slika 11A). Geni su raspoređeni po panelima ovisno o lokalizaciji proteina za koji kodiraju. Detalji toplinske karte su navedeni u poglavlju 2.1.4.2.

endoplazmatskog retikuluma je donekle slična rezultatima objavljenima u Lombardovom radu o evolucijskom porijeklu gena koji kodiraju za enzime početnog dijela biosintetskog puta *N*-glikana (16). Lombard je zaključio da je porijeklo enzima raznovrsno: ALG7, ALG11, ALG2 i STT3 su proteoarhealnog porijekla (superkoljeno TACK), ALG13 i ALG14 su bakterijskog porijekla, ALG3, ALG9, ALG12, ALG6, ALG8 i ALG10 su eukariotskog, a ostatak (ALG1, RFT1 i ALG9) su nepoznatog porijekla. Usporede li se njegove rezultati s rezultatima prikazanim u ovoj disertaciji, može se uočiti da se međusobno poklapaju za dio gena koji kodiraju za enzime koji kataliziraju reakcije u lumenu endoplazmatskog retikuluma.

Gene koji kodiraju za enzime koji kataliziraju reakcije na citosolnoj strani membrane endoplazmatskog retikuluma, Lombard je većinom okarakterizirao kao proteoarhealne, uz iznimku ALG13 i ALG 14 za koje tvrdi da su bakterijskog porijekla. Rezultati iz ove disertacije, prikazani slikom (Slika 12), pokazuju da su svi ti geni prisutni i u arhejama i u bakterijama, uz iznimku ALG1 kojeg nalazimo samo u bakterijama. Iako se korištene metode razlikuju, obje se temelje na jednakim ulaznim informacijama (proteomima).

S obzirom na to da su rezultati pokazali da su geni koji kodiraju za enzime prvog dijela biosintetskog puta *N*-glikozilacije (s katalitičkim mjestom okrenutim prema citosolu) i porijeklom iz zajedničkog pretka svih staničnih organizama skoro u potpunosti prisutni i u bakterijama i u arhejama (Slika 12), možemo pretpostaviti da su u prvu eukariotsku stanicu (tijekom eukariogeneze i razvoja endomembranskog sustava) došli invaginacijom membrane domaćina formacijom unutarstaničnih membranskih vezikula (115–118).

Raspored evolucijske starosti gena s obzirom na membransku topologiju katalitičkog mjesta enzima za koje kodiraju također donekle podupire relativno nedavno predloženu hipotezu eukariogeneze "inside-out" koju su predložili autori Baum i Baum (119). Prema ovoj hipotezi, eukariotska stanica je nastala fuzijom izvanstaničnih protruzija membrane arhealnog domaćina oko α -proteobakterija (proto-mitohondrija). Time je nekadašnji periplazmatski prostor između izvanstaničnih protruzija zatvoren u lumen endoplazmatskog retikuluma u kojem je zadržan dio *en bloc* transfera glikana na protein.

Evolucijsko porijeklo gena koji kodiraju za enzime koji kataliziraju procesiranje *N*-glikana u Golgijevom aparatu je također pokazalo izraženu binarnu razdiobu između ps1 i ps6, ukazujući da su prisutne glikoziltransferaze bile većinom inovacija zajedničkog pretka svih eukariotskih organizama. Za razliku od slučaja u endoplazmatskom retikulumu, enzimi koje kodiraju geni s porijeklom iz zajedničkog pretka svih staničnih organizama nisu pronađeni u proteomima arheja (Slika 12), što ukazuje na to da su oni nedvojbeno bakterijskog porijekla. Mehanizam

kojim su došli u Golgijev aparat ranih eukariotskih stanica nije moguće objasniti "inside out" hipotezom.

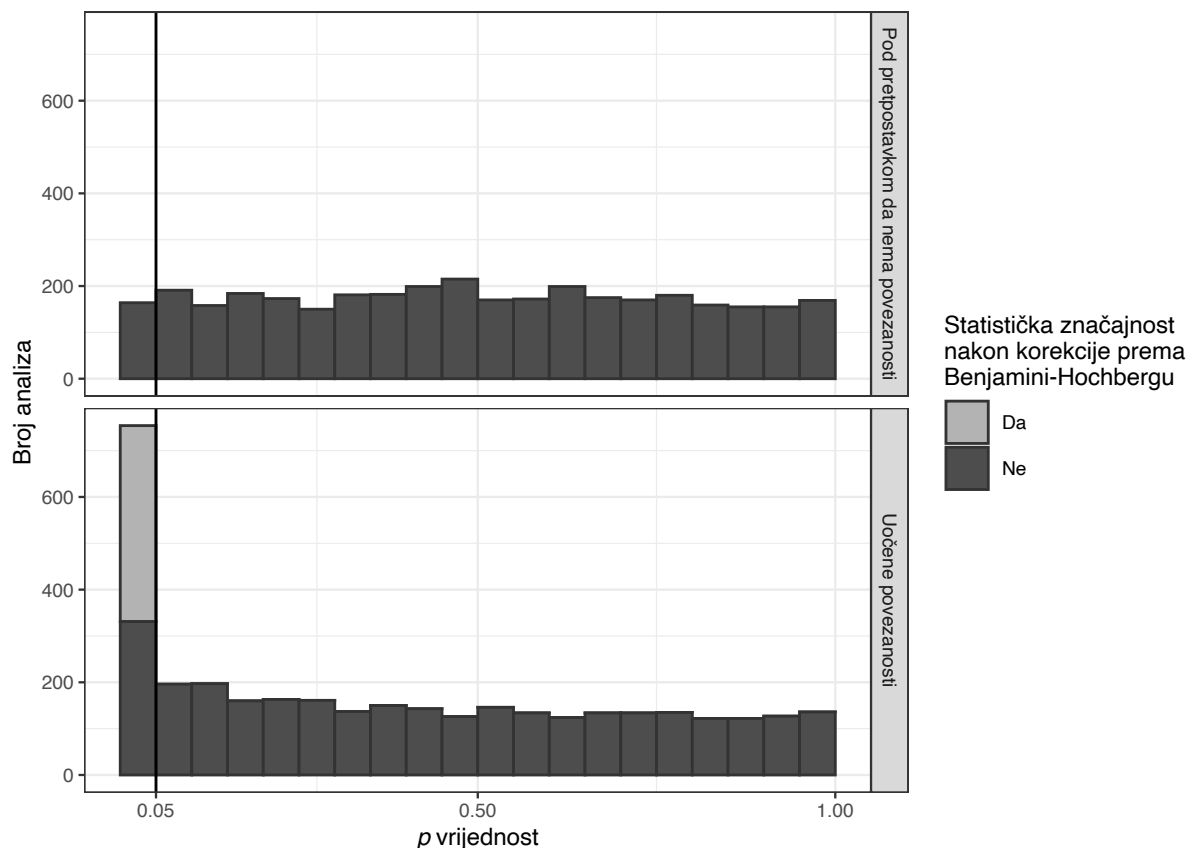
S druge strane, rezultati podupiru i predloženu hipotezu sintrofije (*engl.* syntrophy hypothesis) prema čijoj revidiranoj verziji (120) eukariogeneza obuhvaća tri organizma: (I) δ -proteobakterija koja ima sposobnost redukcije sulfata (domaćin), (II) endosimbiotski arheon sličan Asgard arhejama koji ima sposobnost generiranja vodika (buduća jezgra) i (III) metabolički svestrana, fakultativno anaerobna i sulfid-oksidirajuća α -proteobakterija (budući mitohondrij). Ovom hipotezom se pretpostavlja da je endomembranski sustav (ER i Golgijev aparat) nastao iz periplazmatskog prostora δ -proteobakterije, a prisutnost bakterijskih gena u genomu eukariota s jezgrom porijeklom iz arheja se objašnjava horizontalnim i endosimbiotskim genskim transferom (121).

Iako nije predmet istraživanja, u rezultatima (Slika 12) je uočen gubitak gena koji kodiraju za *N*-acetilglukozaminiltransferaze i galaktoziltransferaze u carstvu gljiva. Upravo tim gubitkom je moguće objasniti nedostatak procesiranja *N*-glikana elongacijom podjedinicama galaktoze i GlcNAc u gljiva. Istovremeno su u biljkama pronađeni homolozi gena koji kodiraju za te glikoziltransferaze što je vidljivo i u glikanskim strukturama pronađenim u tim carstvima (11, 12). Da je istraživanje napravljeno bez prisutnosti prokariotskih organizama i biljaka u bazi proteoma, nametao bi se zaključak da su glikoziltransferaze nastajale kroz evolucijski period od nastanka zajedničkog pretka gljiva i životinja pa sve do nastanka prvih višestaničnih životinja (ps8 – ps10) slično kao u istraživanju autora Tomono i suradnika (122). Temeljem rezultata dobivenih u ovoj disertaciji vjerojatniji je gubitak gena u carstvu gljiva.

Primjenom genomske filostratigrafije pokazano je da je glikozilacija prisutna kod ljudi drevna, porijeklom većinom iz zajedničkog pretka svih staničnih organizama, uz pojedine inovacije u zajedničkom pretku svih eukariota, te manje u zajedničkom pretku životinja. Unatoč starosti, glikozilacija u svoje zlatno doba ulazi razvojem višestaničnih životinja s ciljem razvoja međustanične komunikacije i interakcije stanica s okruženjem.

3.2 UTJECAJ OKOLIŠNIH ČIMBENIKA NA *N*-GLIKOZILACIJU IMUNOGLOBULINA G

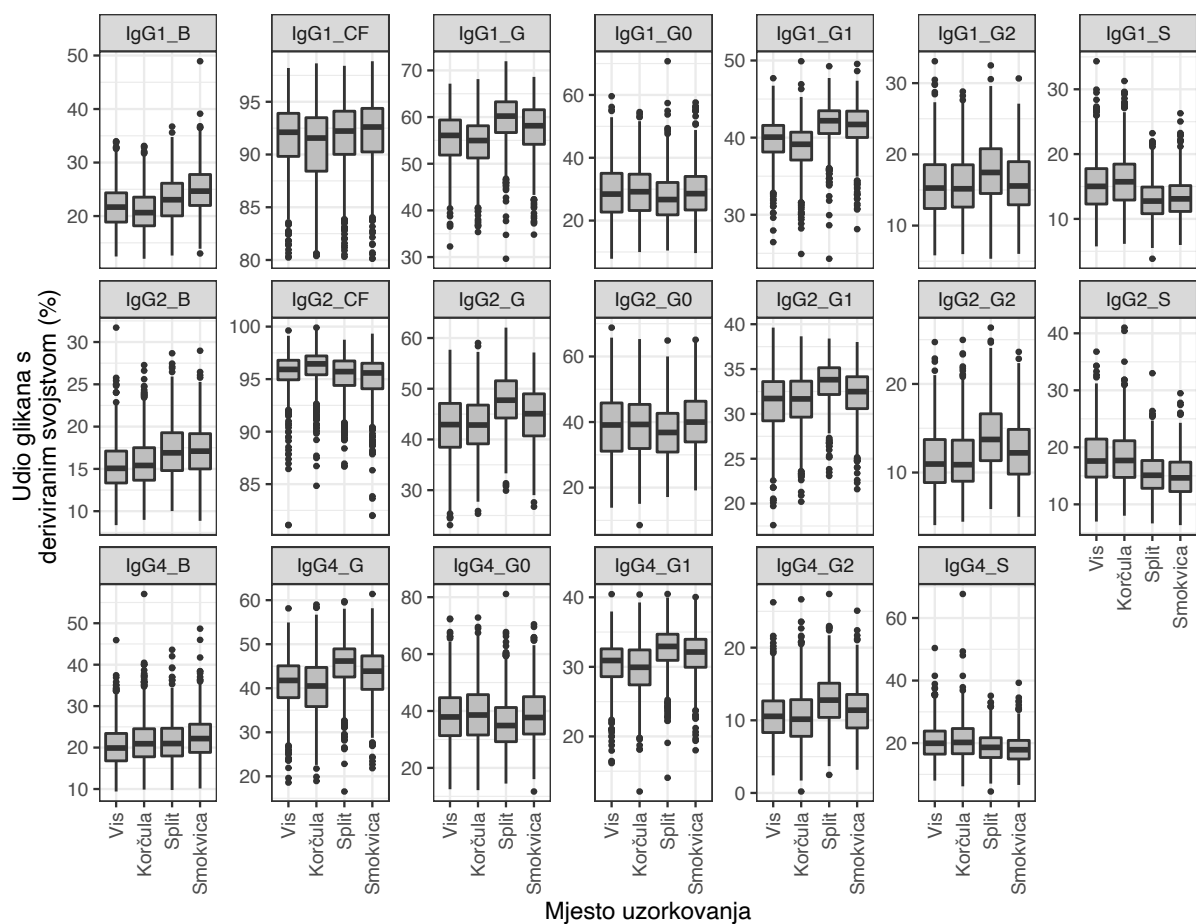
Utjecaj okolišnih čimbenika je procijenjen analizom dostupnih podataka iz biobaze „10001 Dalmatinac“. Napravljeno je ukupno preko 3500 usporedbi i na tako velikom broju testiranja povezanosti očekivano je oko 180 lažno statistički značajnih povezanosti pod pretpostavkom da ne postoji niti jedna povezanost (Slika 13).



Slika 13. Histogram p vrijednosti dobivenih analizom povezanosti dostupnih varijabli u studiji „10001 Dalmatinac“ s deriviranim svojstvima *N*-glikoma imunoglobulina G (dolje) i histogram p vrijednosti analiza povezanosti u kojima je detektirana povezanost rezultat slučajnosti (gore).

Ipak, analizom rezultata razdioba dobivenih p vrijednosti pokazuje veće frekvencije u području manjem od 0,05 u odnosu na razdiobu u kojima je frekvencija statistički značajnih povezanosti isključivo rezultat slučajnosti. Korekcijom p vrijednosti prema Benjamini-Hochbergu uspješno se uklanjaju statistički značajne lažne povezanosti koje su pronađene u simuliranoj studiji (82), a ista je primijenjena i na analizi stvarnih podataka.

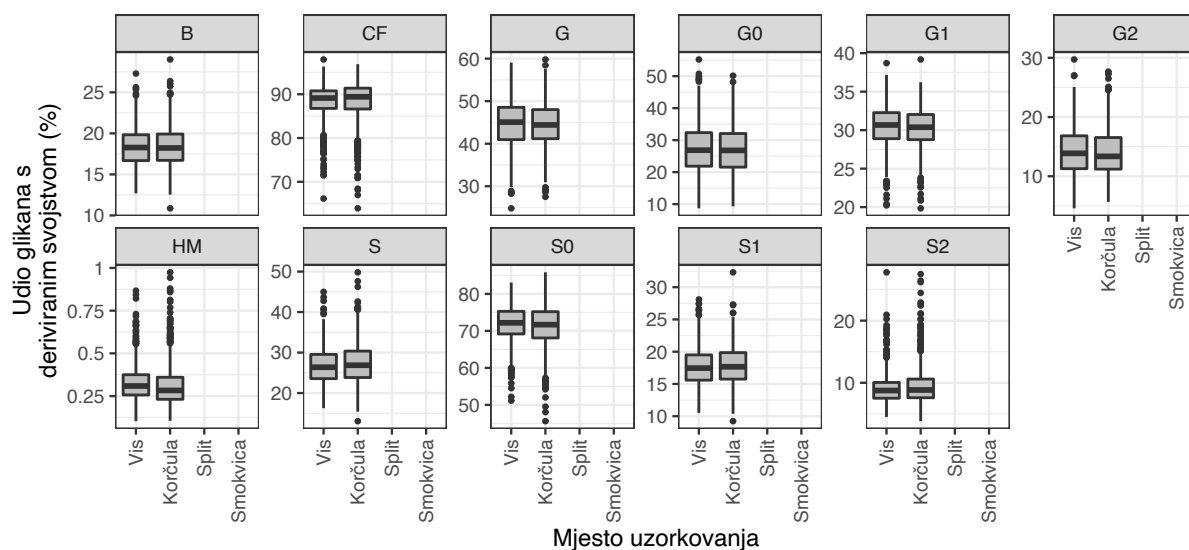
N-glikom imunoglobulina G je analiziran dvjema metodama kromatografije: LC-ESI-MS i HILIC-UPLC. Analizirani kromatografski vršci su korišteni za izračun deriviranih svojstva prema formulama navedenim u tablicama (Tablica 15 i Tablica 16) čija je deskriptivna statistika prikazana slikama (Slika 14 i Slika 15).



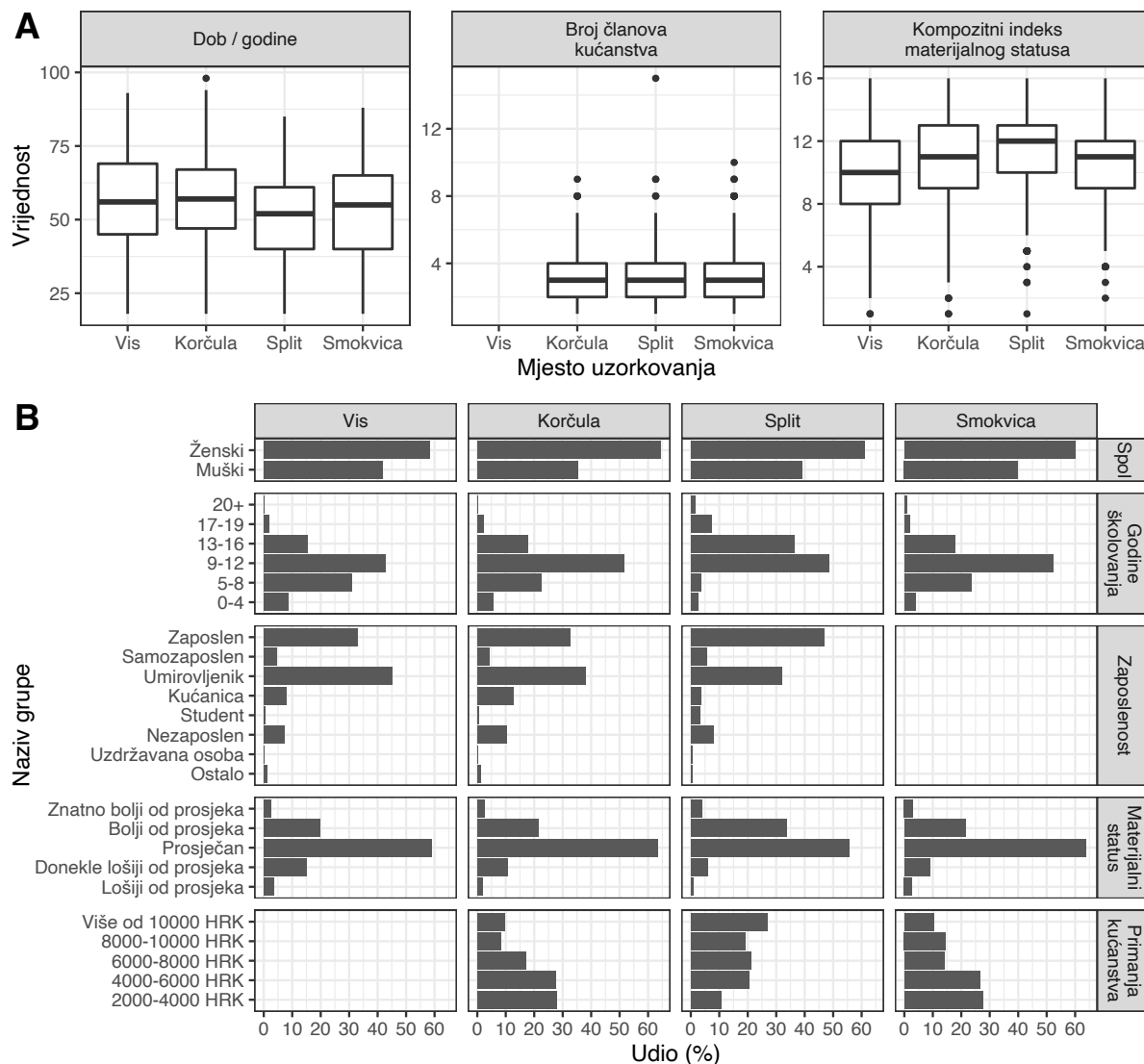
Slika 14. Deskriptivna statistika deriviranih svojstava N-glikoma IgG-a određenog LC-ESI-MS metodom. Objašnjenja pojedinih deriviranih svojstava N-glikoma su navedena u tablici (Tablica 15).

Metodom LC-ESI-MS analizirano je ukupno 3480 ispitanika uzorkovanih u sklopu sve četiri kohorte, dok ju UPLC metodom analizirano 1805 ispitanika u sklopu kohorti u kojima su ispitanici uzorkovani na otoku Visu i u gradu i okolici grada Korčule. Metode su analizirale iste uzorke i ne može ih se promatrati kao nezavisne događaje.

Deskriptivna statistika analiziranih demografskih podataka je prikazana slikom (Slika 16). Ispitanici ženskog spola su dominirali u svim kohortama u prosječnom omjeru 3:2. Prema dobi, sve četiri kohorte su relativno slične, iako su u splitskoj kohorti uzorkovani zanemarivo mlađi ispitanici. Broj članova kućanstva se nije razlikovao između kohorti. Ispitanici su najčešće srednjoškolskog obrazovanja, dok je osnovnoškolsko obrazovanje bilo drugo po redu učestalosti u kohortama u kojima su ispitanici uzorkovani u manjim sredinama (otok Vis, grad i okolica Korčule, te selo i okolica Smokvice). U splitskoj je kohorti uz srednjoškolsko obrazovanje dominiralo i visoko obrazovanje. Primanja kućanstva i subjektivna procjena materijalnog statusa su pratili obrazovanje, kao i kompozitni indeks materijalnog statusa, no u blažoj mjeri.



Slika 15. Deskriptivna statistika deriviranih svojstava *N*-glikoma IgG-a određenog UPLC metodom. Objašnjenja pojedinih deriviranih svojstava *N*-glikoma su navedena u tablici (Tablica 16).



Slika 16. Deskriptivna statistika demografskih numeričkih (A) i kategorijskih (B) varijabli.

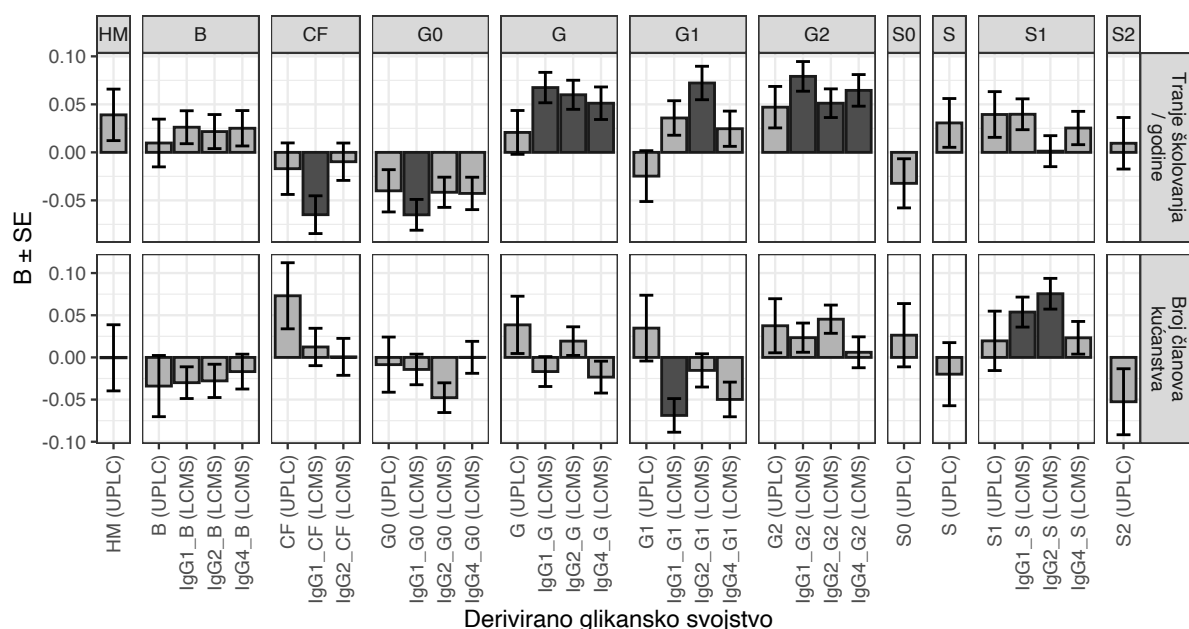
Dobiveni rezultati ovisnosti *N*-glikozilacije imunoglobulina G o demografskim varijablama su prikazane slikom (Slika 17). Osobe koje su se dulje školovale imaju veći udio galaktoziliranih glikana, pri čemu većinu efekta nose digalaktozilirane strukture. Uočene su promjene prisutne samo u LCMS analizi, dok isti trend postoji u UPLC analizi, ali bez statističke značajnosti. Istovremeno dolazi do smanjenja agalaktoziliranih struktura, no jedino kod IgG klase 1 u LCMS analizi uočavamo statistički značajne promjene.

Pojedinačna derivirana svojstva LCMS analize *N*-glikoma su pokazala povezanost s brojem članova kućanstva, no UPLC analizom nisu replicirala trendove.

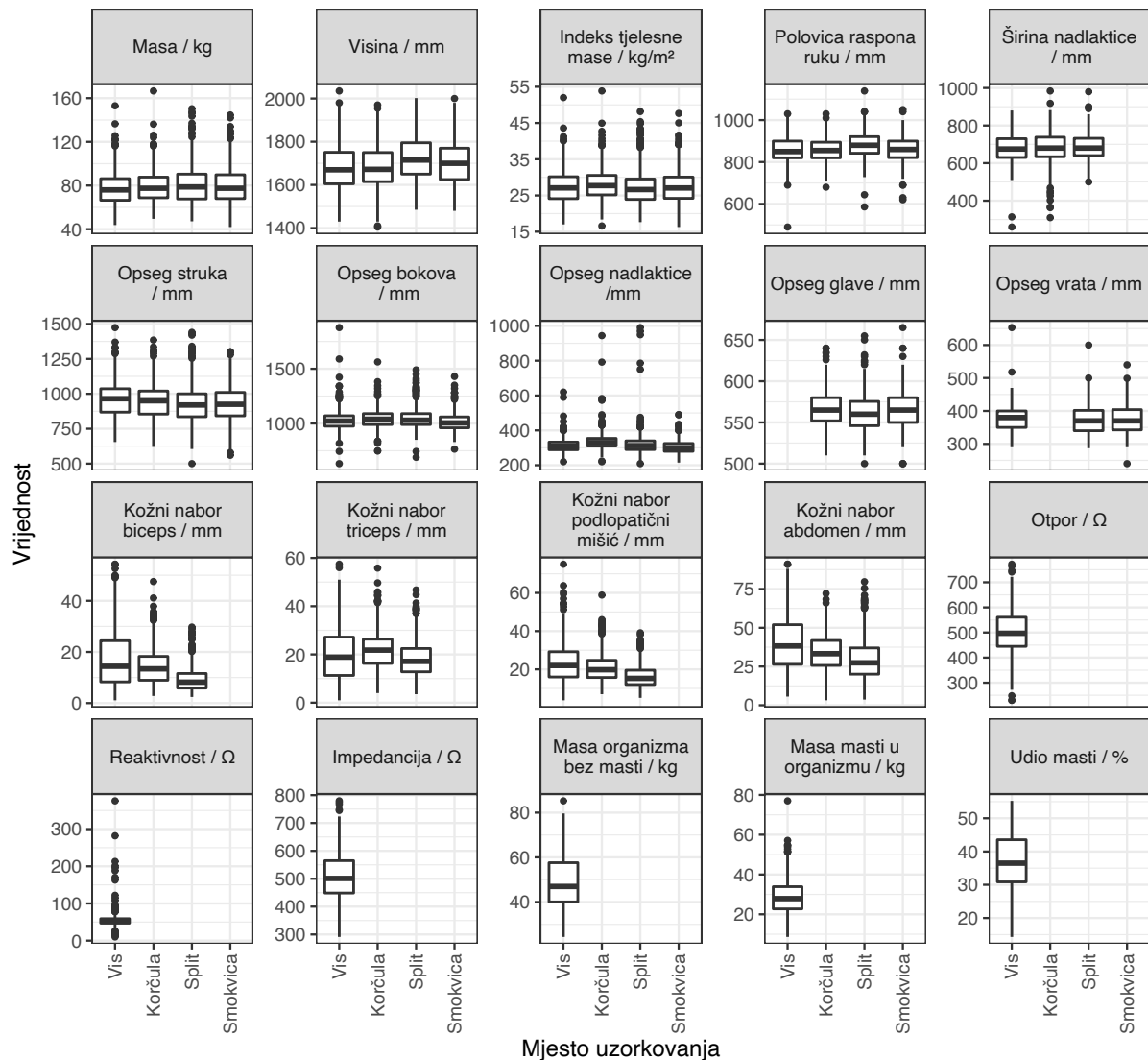
Ostale demografske varijable nisu pokazale povezanost niti s jednim deriviranim svojstvom *N*-glikoma imunoglobulina G (rezultati nisu prikazani).

Dosadašnja istraživanja također nisu pronašla povezanost između demografskih podataka (osim sa spolom i dobi) s *N*-glikomom imunoglobulina G.

Antropometrijski su podaci zbog svoje izražene međusobne korelacije podvrgnuti faktorskoj analizi s ciljem smanjenja dimenzionalnosti. Identificirano je pet faktora. Prvi, nazvan *veličina tijela*, je obuhvatio varijable koje su odraz dimenzija ljudskog tijela, a nisu povezane s pretilošću: visina, polovica raspona ruku, opseg glave, širina nadlaktice, masa tijela bez masti, forsirani ekspiratorni volumen, forsirani vitalni kapacitet i vršni ekspiratorni protok. Drugi, nazvan *pretilost*, je obuhvatio varijable koje su dobri pokazatelji prekomjerne tjelesne mase:

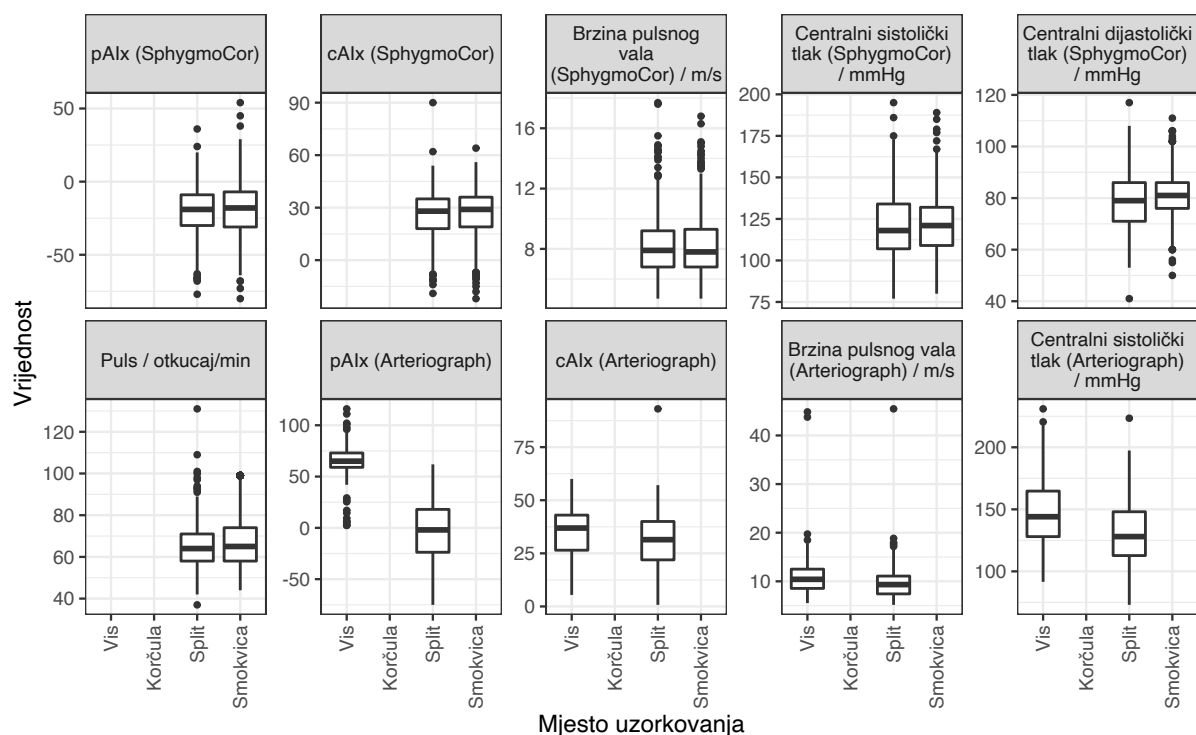


Slika 17. Rezultati regresijskih analiza povezanosti demografskih podataka s *N*-glikomom IgG-a. Na apscisi su prikazana derivirana svojstva *N*-glikoma, a na ordinati regresijski koeficijent (B) i pripadajuća standardna pogreška (SE). Navedeni brojevi prikazuju promjenu izraženu u standardnim devijacijama deriviranog svojstva za svaki porast od 1 standardne devijacije dane demografske varijable. Tamnije obojani stupići označavaju statistički značajnu povezanost. Kratice deriviranih svojstva su navedene u tablicama (Tablica 15 i Tablica 16).



Slika 18. Deskriptivna statistika antropometrijskih podataka dobivenih mjerenjem dimenzija pojedinih dijelova tijela i bioimpedancije.

tjelesna masa, indeks tjelesne mase, količina masti u tijelu, opseg struka, opseg bokova, omjer opsega struka i bokova, opseg vrata, opseg nadlaktice, debljina kožnih nabora u području abdomena, bicepsa, tricepsa i podlopatičnog mišića. Treći, nazvan *krvni tlak*, obuhvatio je sve varijable koje mjere krvne tlakove: sistolički krvni tlak, dijastolički krvni tlak, sve krvne tlakove mjerene u lijevim i desnim, brahijalnim, te stražnjim tibijalnim arterijama i dorzalnim arterijama stopala, te rezultate mjerenja elastičnosti arterija (centralne i periferne augmentacijske indekse, centralni sistolički i dijastolički krvni tlak, i brzinu pulsog vala). Ovaj faktor, te niti jedan drugi faktor, nije obuhvatio brzinu otkucaja srca (puls) i gležanjski indeks (ABPI), pa su te varijable analizirane posebno. Četvrti faktor, nazvan *mineralna gustoća kostiju*, je obuhvatio sve varijable dobivene denzitometrijom, neovisno o principu rada (rendgen



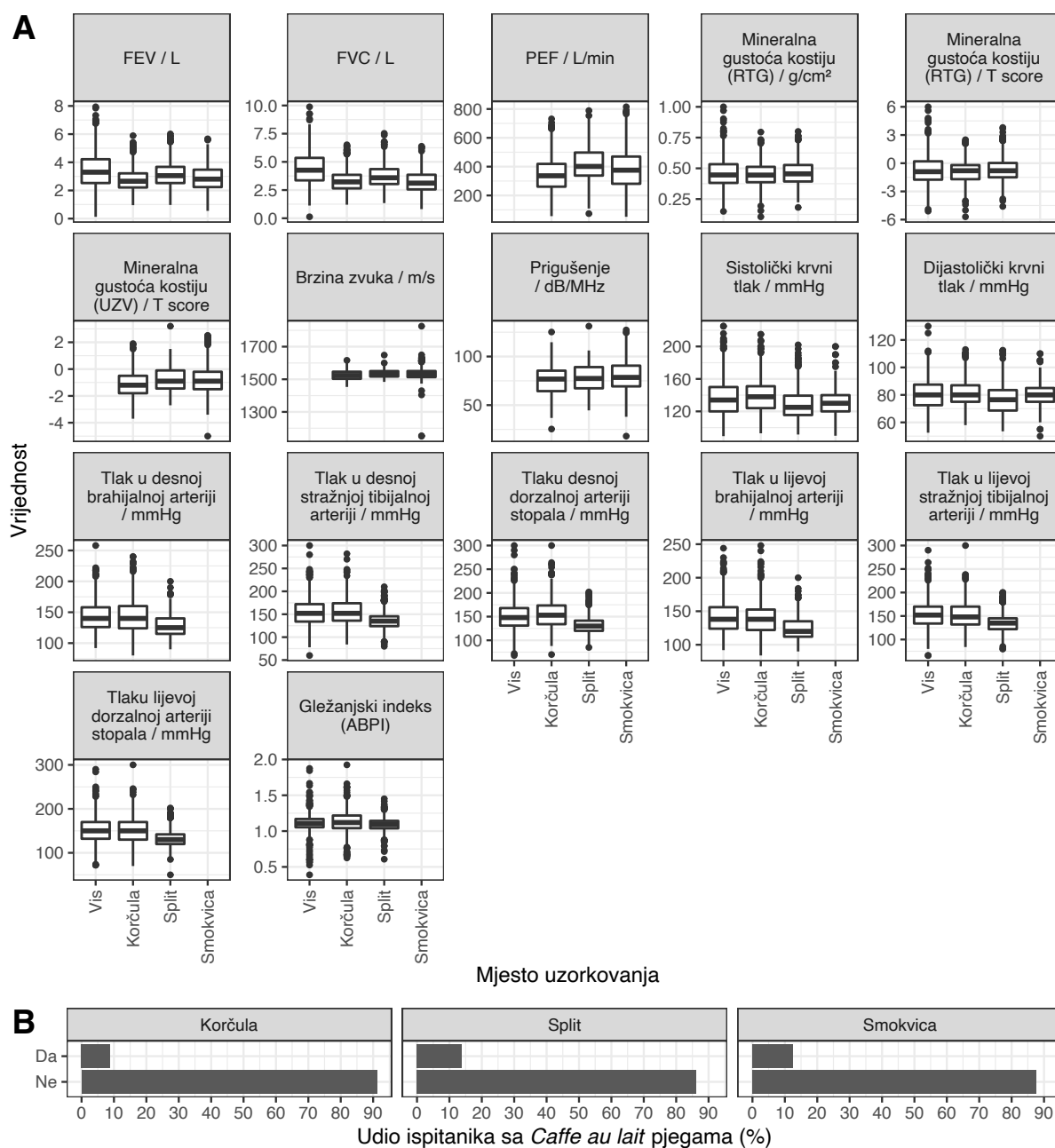
Slika 19. Deskriptivna statistika numeričkih antropoloških podataka vezanih uz elastičnost arterija prikupljenu uređajima SphygmoCor i Arteriograph. Kratice: cAIx - centralni augmentacijski indeks, pAIx - periferni augmentacijski indeks.

ili ultrazvuk). Peti faktor, nazvan *bioimpedancija*, je obuhvatio otpornost, reaktivnost i bioimpedanciju.

Rezultati povezanosti antropometrijskih faktora i varijabli s *N*-glikomom imunoglobulina G su prikazani slikom (Slika 21). Gležnjski indeks (ABPI) je pokazao stabilne trendove s galaktozilacijom IgG-a mjerenom s obje kromatografske metode, no statistički značajne povezanosti opažene su samo kod LCMS metode za IgG klase 1 i 4. Većini značajnog efekta doprinio je efekt promjene udjela digalaktoziliranih struktura. Pregledom literaturnih podataka nije pronađeno istraživanje koje je promatralo povezanost *N*-glikoma imunoglobulina G s gležnjskim indeksom.

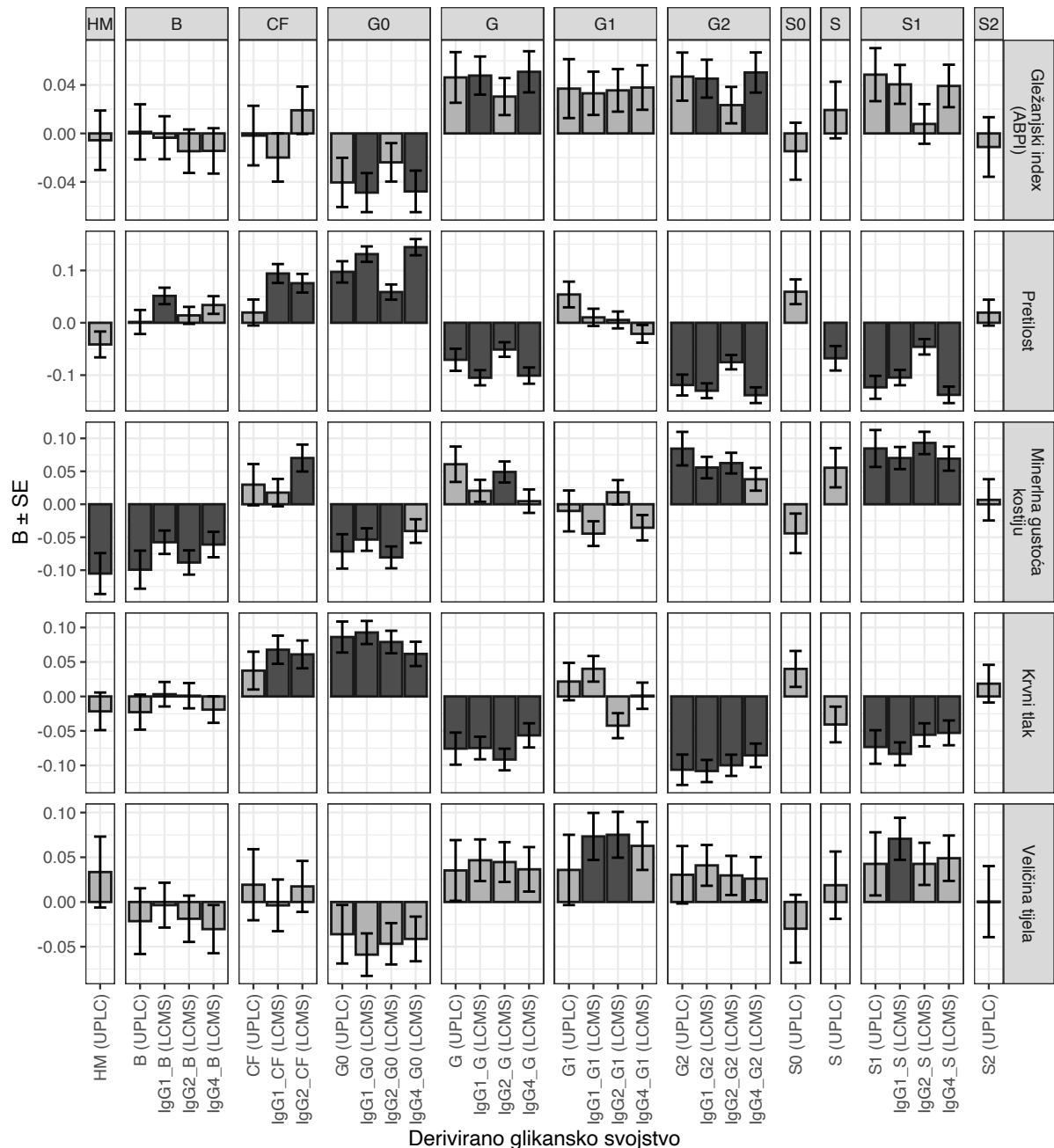
Pretilost je pokazala izrazito sustavnu povezanost s *N*-glikomom imunoglobulina G. Uočeno je da povećanje debljine uzrokuje statistički značajni rasta udjela agalaktoziliranih struktura, te smanjenje udjela digalaktoziliranih i monosijaliniziranih struktura, neovisno o kromatografskoj metodi. Uočeno je i povećanje udjela sržno fukožiliranih struktura s povećanjem debljine, ali samo u *N*-glikomu određenom LCMS metodom. Analizom povezanosti indeksa tjelesne mase s *N*-glikomom IgG-a uočeni su jednaki trendovi rasta agalaktoziliranih i pada digalaktoziliranih struktura na ispitanicima u TwinUK kohorti, te na pretilim ispitanicima koji su podvrgnuti operaciji želuca kombiniranom s izrazito restriktivnom dijetom (123).

Krvni tlak pokazuje identične obrasce povezanosti s *N*-glikomom IgG-a kao i pretilost. Ovi su rezultati objavljeni u sklopu suradnje sa University of Texas Southwestern Medical Center, u kojoj se pokazalo da smanjena sijalinizacija IgG-a uzrokuje hipertenziju zbog povećane tjelesne mase. Također, na miševima je pokazano da je unos prekursora sijalinske kiseline uzrokovao povećanje IgG sijalinizacije i smanjenja krvnog tlaka uzrokovanog pretilošću (124).



Slika 20. Deskriptivna statistika numeričkih antropoloških podataka vezanih uz plućni kapacitet, mineralnu gustoću kostiju i krvni tlak (A), te udio ispitanika u populaciji s *Caffe au lait* pjegama (B). Kratice: FEV – forsirani ekspiratorni volumen; FVC – forsirani vitalni kapacitet; PEF – vršni ekspiratorni protok (*engl.* peak expiratory flow); RTG – rendgen; UZV ultrazvuk; ABPI – *engl.* ankle brachial pressure index.

Mineralna gustoća kostiju je također pokazala povezanost s *N*-glikomom imunoglobulina G. Povećanje gustoće prati statistički značajno smanjenje udjela struktura s račvajućim GlcNAc-om, i smanjenje udjela agalaktoziliranih struktura, te povećanje udjela digalaktoziliranih (osim na IgG4) i monosijaliniziranih struktura.

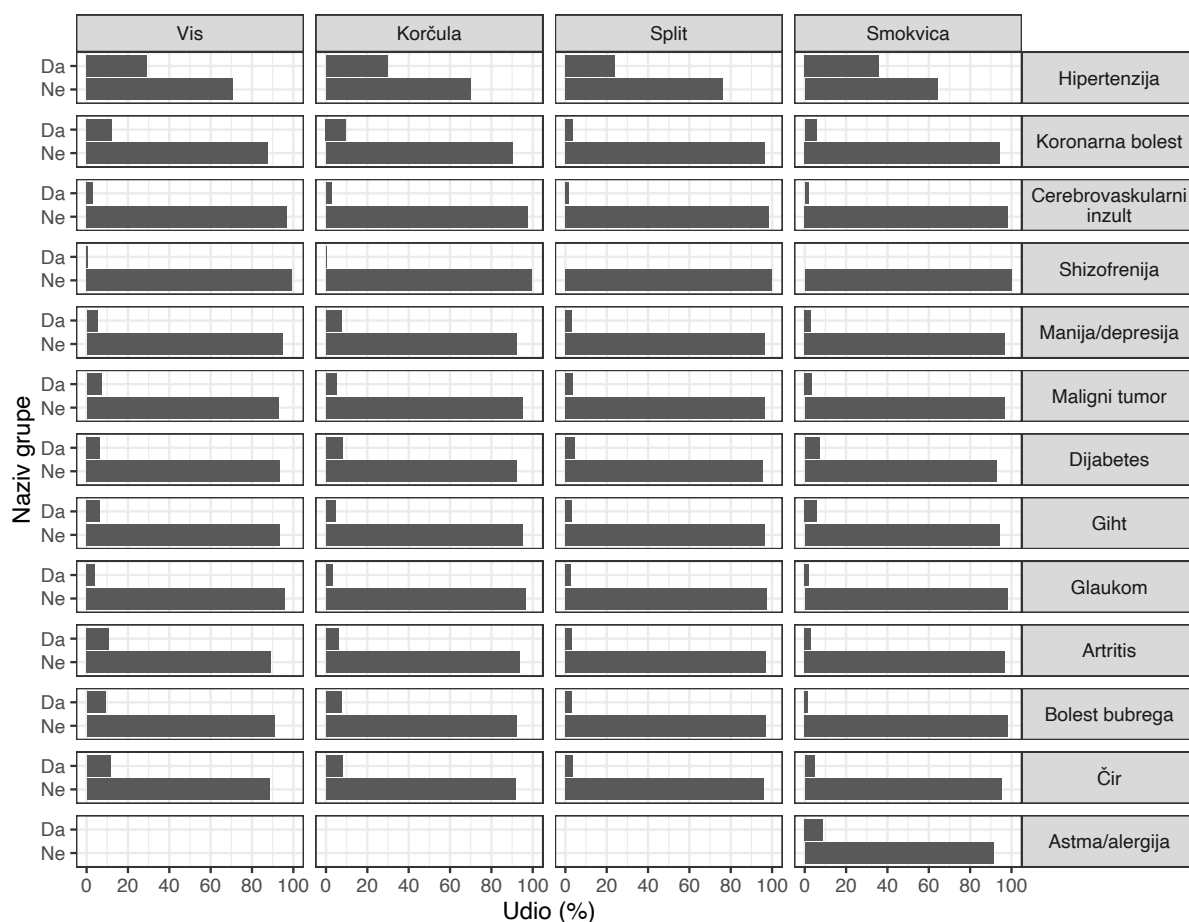


Slika 21. Rezultati regresijskih analiza povezanosti faktora identificiranih faktorskom analizom antropometrijskih podataka s *N*-glikomom IgG-a. Na apscisi su prikazana derivirana svojstva *N*-glikoma, a na ordinati regresijski koeficijent (B) i pripadajuća standardna pogreška (SE). Navedeni brojevi prikazuju promjenu izraženu u standardnim devijacijama deriviranog svojstva za svaki porast od 1 standardne devijacije dane antropometrijske varijable. Tamnije obojani stupići označavaju statistički značajnu povezanost. Kratice deriviranih svojstava su navedene u tablicama (Tablica 15 i Tablica 16).

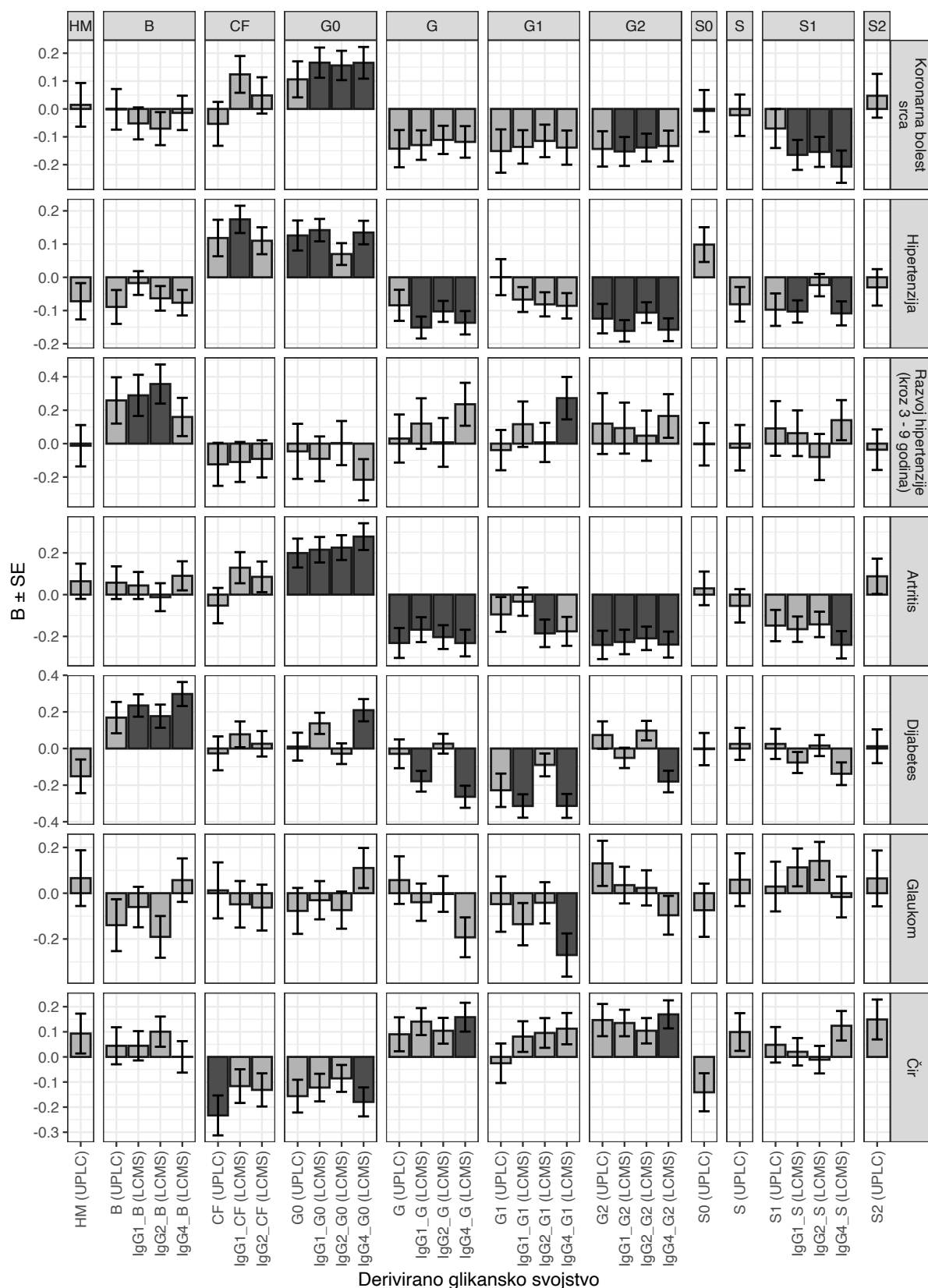
UPLC metoda uspješno razdvaja i mjeri udio strukture s visokim udjelom manoze (strukture s 5 manoznih jedinica), te je uočen statistički značajan smanjenje u udjelu te strukture kod ispitanika s većom mineralnom gustoćom kostiju.

Veličina tijela također pokazuje relativno sustavne obrasce povezanosti s *N*-glikomom imunoglobulina G, koji su suprotni onima koji su uočeni kod povezanosti s debljinom i krvnim tlakom. No statistički značajni regresijski koeficijenti su uočeni samo za porast monogalaktoziliranih struktura imunoglobulina G klasa 1 i 2, te monosijaliniziranih struktura IgG-a klase 1, bez statistički značajnih rezultata kod UPLC metode.

Pregledom literature nisu pronađeni radovi koji su proučavali povezanost *N*-glikoma IgG-a s mineralnom gustoćom kostiju ili dimenzijama ljudskog tijela koje ne karakteriziraju pretilost. Sažimanjem podatka o povijesti bolesti izračunan je udio ispitanika koji boluje od pojedinih bolesti (Slika 22). Analizirano je postoji li razlika u *N*-glikomu imunoglobulina G između zdravih i bolesnih. Rezultati za bolesti među kojima je uočena statistički značajna razlika u barem jednom deriviranom svojstvu su prikazani slikom (Slika 23).



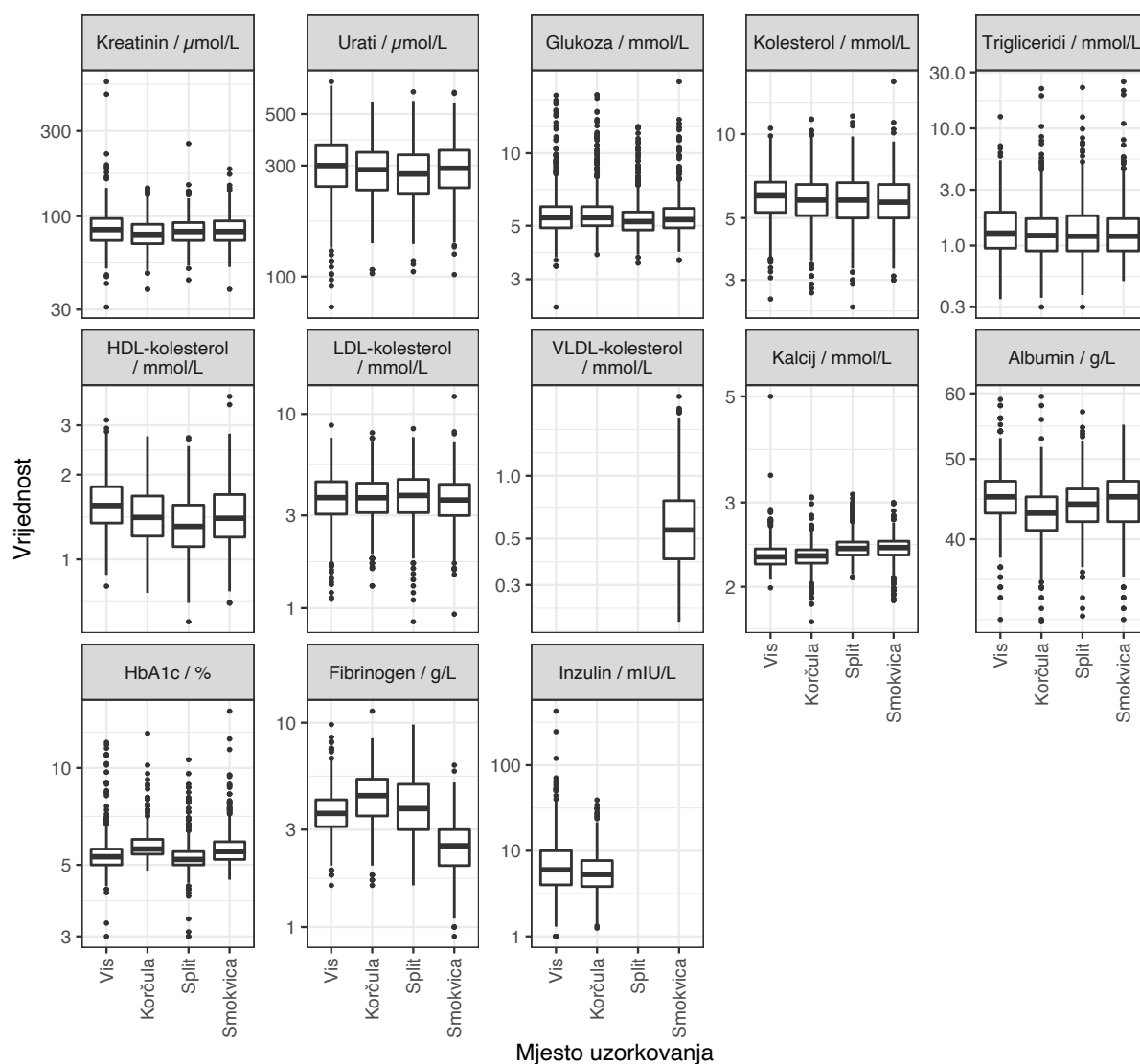
Slika 22. Deskriptivna statistika povijesti bolesti.



Slika 23. Rezultati regresijskih analiza povezanosti podataka dobiveni iz povijesti bolesti s *N*-glikomom IgG-a. Na apscisi su prikazana derivirana svojstva *N*-glikoma, a na ordinati regresijski koeficijent (B) i pripadajuća standardna pogreška (SE). Navedeni brojevi prikazuju razliku izraženu u standardnim devijacijama deriviranog svojstva između bolesnih i zdravih ispitanika. Iznimno za razvoj hipertenzije efekt prikazuje logaritam promjene rizika za svako povećanje od 1 standardne devijacije u razini deriviranog svojstva *N*-glikoma. Tamnije obojani stupići označavaju statistički značajnu povezanost. Kratice deriviranih svojstava su navedene u tablicama (Tablica 15 i Tablica 16).

Najkonzistentniji obrasci promjena su uočeni kod artritisa. Ispitanici koji boluju od artritisa imaju statistički značajno povećane udjele agalaktoziliranih struktura, te statistički značajno smanjene udjele digalaktoziliranih struktura. Udio monosijaliniziranih struktura je također smanjen u obje metode no promjena nije statistički značajna (osim za IgG4). Ove su promjene u literaturi opisane još 1988. godine i često se nalaze u knjigama kao školski primjer promjene *N*-glikoma IgG-a u bolestima (45).

Hipertenzija i koronarna bolest srca su popraćene jednakim obrascem promjene *N*-glikoma. Statistički značajne promjene u obje metode uočene su jedino kod hipertenzije u obliku povećanih udjela agalaktoziliranih te smanjenih udjela digalaktoziliranih struktura. Rezultati povezanosti *N*-glikoma izmjenjenog UPLC metodom s hipertenzijom su objavljeni u sklopu multinacionalne studije koja je obuhvatila ispitanike iz Kine, Škotske i Hrvatske (101). Čini se



Slika 24. Deskriptivna statistika rezultata biokemijskih testova. Ordinata je log transformirana. Kratice: HDL – lipoprotein visoke gustoće (*engl.* high density lipoprotein); LDL – lipoprotein niske gustoće (*engl.* low density lipoprotein); VLDL – lipoprotein vrlo niske gustoće (*engl.* very low density lipoprotein); HbA1c – glicirani hemoglobin.

da uočene promjene u *N*-glikomu nastaju kao posljedica hipertenzije budući da te iste promjene nisu uočene prije razvoja hipertenzije. Ipak, razvoju hipertenzije prethodi porast udjela glikanskih struktura s račvajućim GlcNAc. Dobiveni su rezultati objavljeni u sklopu suradnje s Kings Collage London (TwinsUK – 2617 ispitanika (125)) i Helmholtz Zentrum München (KORA – 729 ispitanika (126)), gdje je pokazano da dodatak informacije o udjelu struktura s račvajućim GlcNAc, te udjela izravno mjerenih glikanskih kromatografskih vršaka GP4, GP9 i GP21 u predikcijski model temeljen na dobi, indeksu tjelesne mase i srednjeg arterijskog tlaka (*engl.* mean arterial pressure, MAP; $MAP = \frac{2 \times DBP + SBP}{3}$) statistički značajno povećava prognozu (127). U citiranoj je publikaciji granica statističke značajnosti bila $\alpha = 0,1$, za razliku od ove disertacije gdje je $\alpha = 0,05$, pa se rezultat dobiven na povezanosti razvoja hipertenzije s količinom račvajućeg GlcNAc u *N*-glikomu IgG-a mjenog UPLC metodom smatra statistički značajnim.

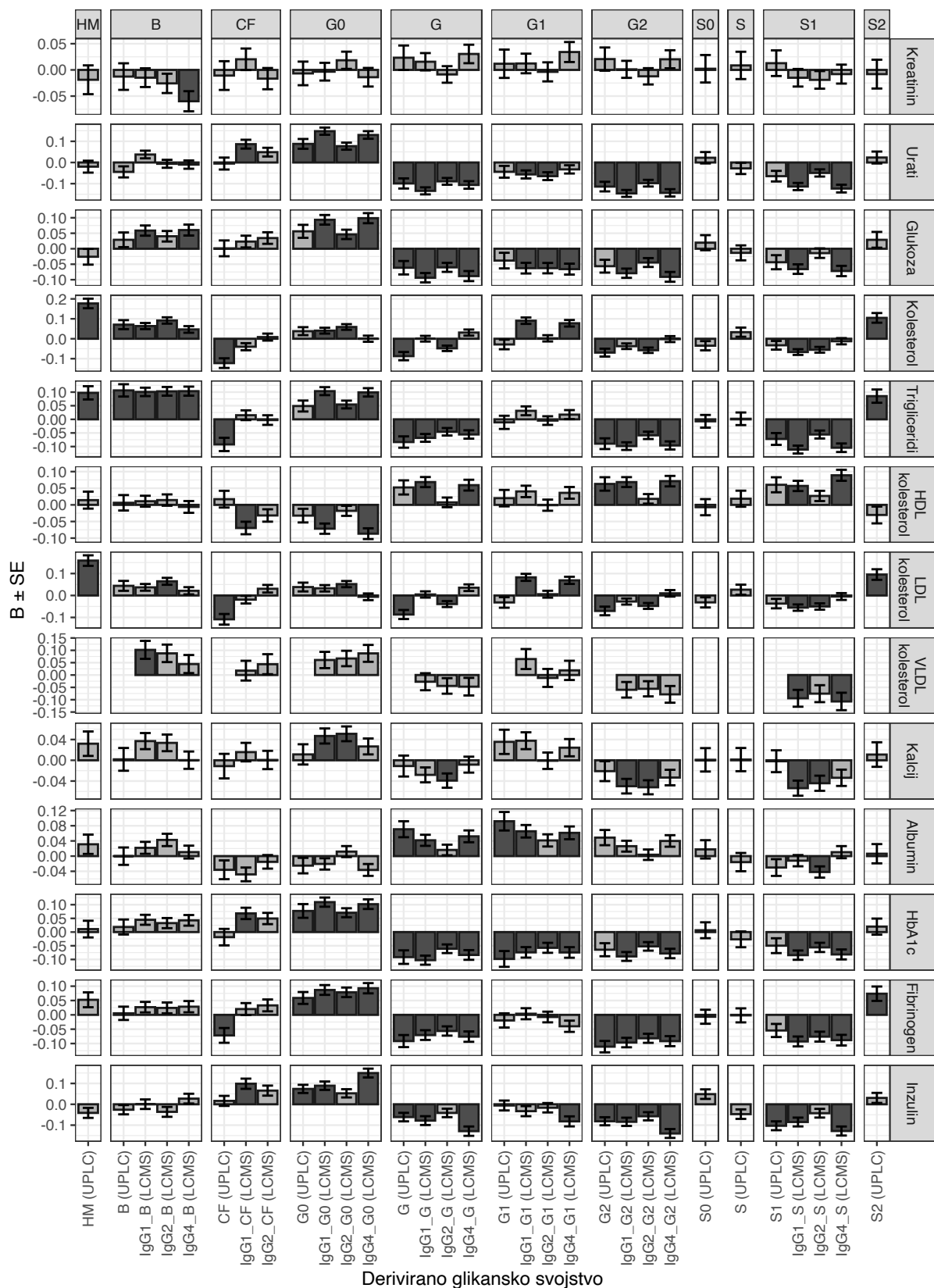
Kod ispitanika koji boluju od dijabetesa (nepoznatog tipa), uočeno je sustavno povećanje udjela struktura s račvajućim GlcNAc-om, te smanjenje udjela monogalaktoziliranih struktura, no uočene promjene nisu bile statistički značajne u UPLC analizi. Literaturni podaci pokazuju da je *N*-glikom IgG-a je promijenjen u dijabetesima različitih tipova (46, 128, 129).

Kod ispitanika koji boluju od čira je uočen smanjen udio sržno fukožiliranih struktura i agalaktoziliranih struktura, te povećanje digalaktoziliranih struktura, no navedene promjene nisu bile statistički značajne u obje metode.

Sporadične promjene u *N*-glikomu imunoglobulina G su uočene i kod ispitanika s glaukomom, no rezultati nisu pokazali konzistentne trendove između metoda.

Deskriptivna statistika biokemijskih testova je prikazana slikom (Slika 24). Rezultati analiza su pokazali brojne povezanosti *N*-glikoma imunoglobulina G sa različitim biokemijskim parametrima (Slika 25).

Udio struktura s račvajućim GlcNAc-om sustavno i statistički značajno se povećavao s povećanjem ukupnog kolesterola i triglicerida u plazmi. Među kolesterolom razdijeljenom po lipoproteinima različitih gustoća jedino su VLDL i LDL kolesterol pokazali jednak statistički značajan trend kod klase 1 IgG-a, odnosno klase 2 IgG-a. Istovremeno, VLDL kolesterol nije bio analiziran među ispitanicima kojima je *N*-glikom analiziran UPLC metodom. Uvidom u dosad objavljene literaturne podatke pronađeno je da su obrasce promjena *N*-glikoma IgG-a uočili i Liu i sur. (130).



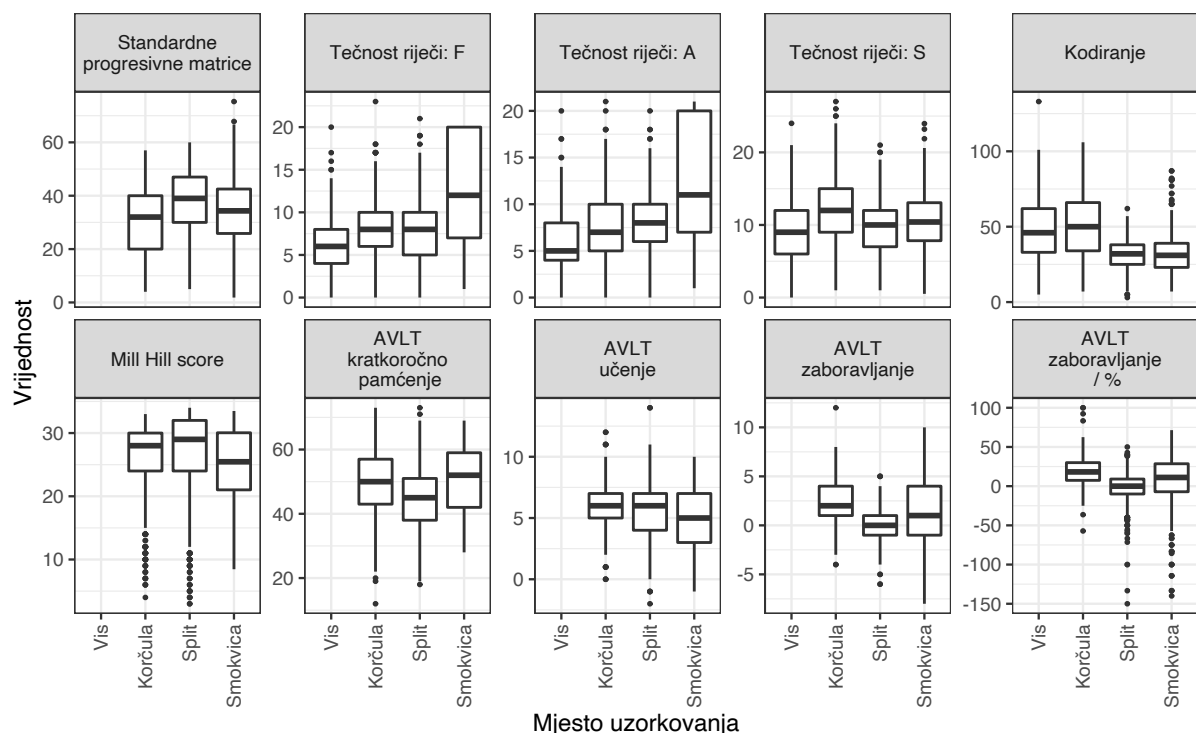
Slika 25. Rezultati regresijskih analiza povezanosti rezultata biokemijskih testova s N-glikomom IgG-a. Na apscisi su prikazana derivirana svojstva N-glikoma, a na ordinati regresijski koeficijent (B) i pripadajuća standardna pogreška (SE). Navedeni brojevi prikazuju promjenu izraženu u standardnim devijacijama deriviranog svojstva za svaki porast od 1 standardne devijacije dane biokemijske varijable. Tamnije obojani stupići označavaju statistički značajnu povezanost. Kratice deriviranih svojstava su navedene u tablicama (Tablica 15 i Tablica 16).

Povećanje udjela agalaktoziliranih struktura združeno sa smanjenjem udjela digalaktoziliranih i monosijaliniziranih struktura je uočeno kod povećanja koncentracije urata, glukoze, triglicerida, kalcija, gliciranog hemoglobina, fibrinogena i inzulina. Iako su trendovi svuda bili konzistentni, za koncentraciju glukoze, ukupnog kolesterola, triglicerida i kalcija nije pronađen statistički značajan rezultat kod *N*-glikoma određenog UPLC metodom. Suprotna je promjena glikoma uočena za koncentraciju HLD kolesterola pri čemu rezultat nije bio statistički značajan za *N*-glikom određen UPLC metodom.

Dodatno, s povećanjem koncentracije gliciranog hemoglobina, urata i glukoze uočeno je smanjenje udjela monogalaktoziliranih struktura, a povezanost s *N*-glikomom određenog UPLC-om je značajna samo za koncentraciju gliciranog hemoglobina. Suprotno tome, povećanje koncentracije albumina popraćeno je povećanjem udjela monogalaktoziliranih struktura.

Strukture s visokim sadržajem manoza kao i disijalinizirane strukture su mjerljive samo UPLC metodom. Među tim rezultatima, značajne povezanosti su uočene samo između udjela struktura s visokim sadržajem manoze s koncentracijama ukupnog kolesterola, triglicerida i LDL kolesterola.

Biokemijski testovi su često određivani u znanstvenim istraživanjima i za dobar dio njih je već potvrđena povezanost s *N*-glikomom imunoglobulina G. Jednaki obrasci promjena *N*-glikoma

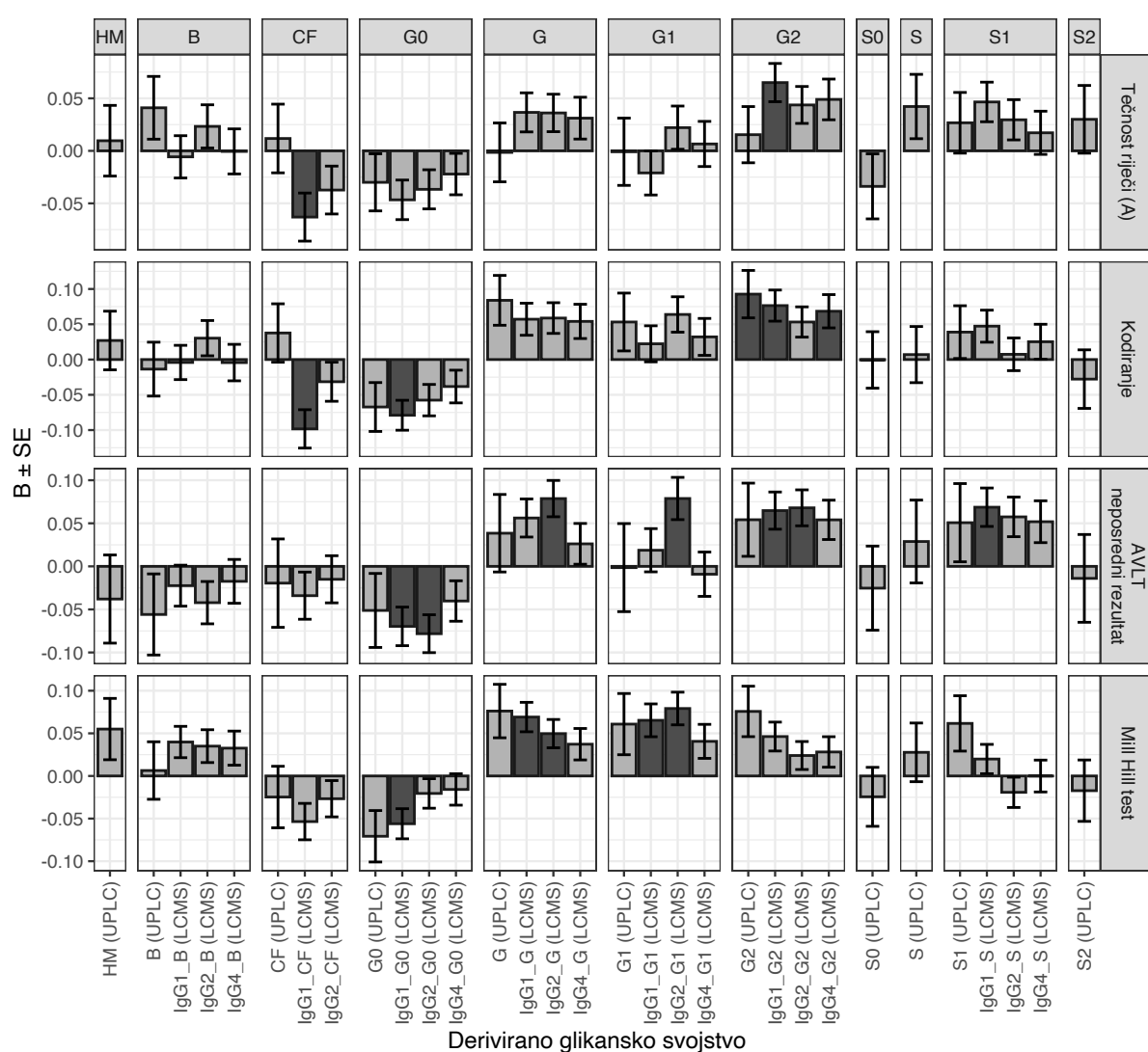


Slika 26. Deskriptivna statistika rezultata kognitivnih testova. Kratice: AVLT – test auditivno-verbalnog učenja (engl. Rey's Auditory Verbal Learning Test).

za koncentracije urata, inzulina, gliciranog hemoglobina i fibrinogena su već publicirani u znanstvenoj literaturi (131, 132).

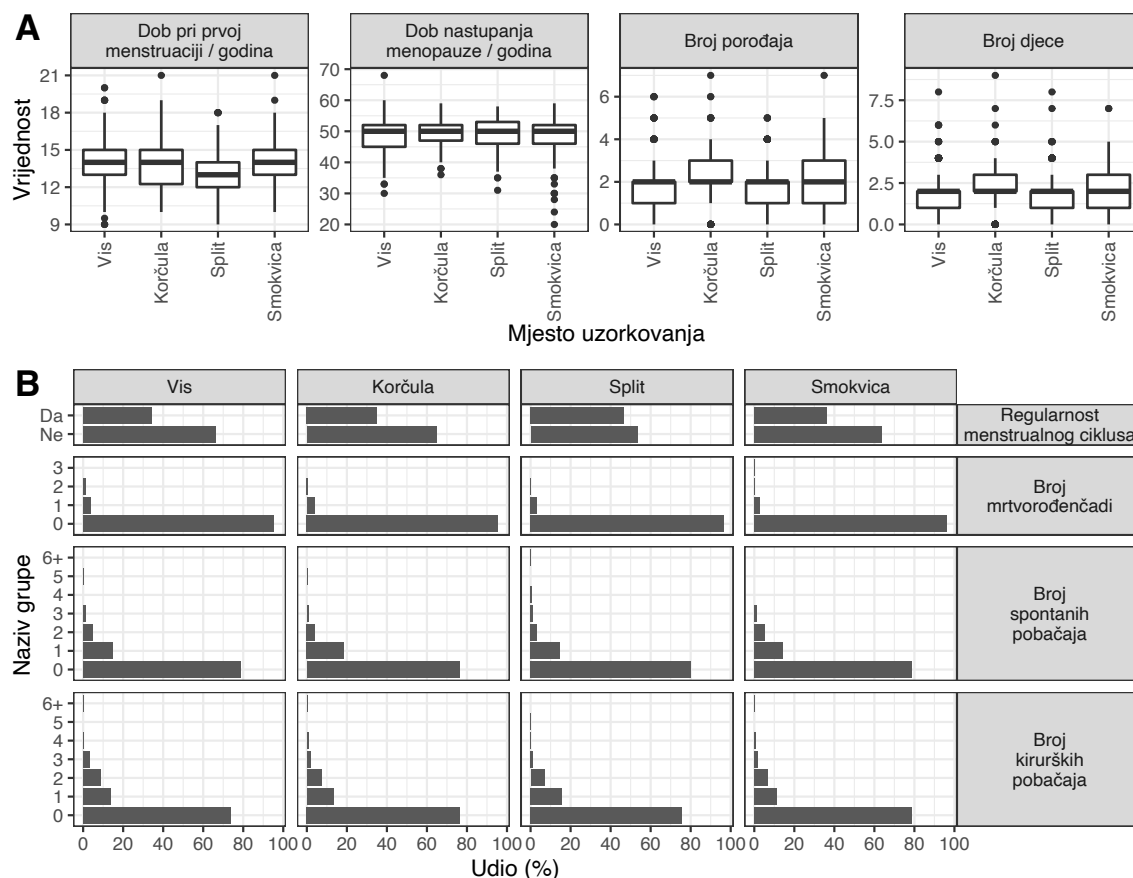
Deskriptivna statistika rezultata kognitivnih testova je prikazana slikom (Slika 26). Povezanost sa N-glikomom imunoglobulina G nije bila potpuno konzistentna. Općenito je kod svih testova za obje kromatografske metode uočeno povećanje udjela digalaktoziliranih struktura i smanjenje udjela galaktoziliranih struktura s povećanjem broja bodova ostvarenih na testu.

Navedene promjene su bile statistički značajne samo u pojedinačnim slučajevima, a detaljnije su prikazane slikom (Slika 27). Rezultati testova standardnih progresivnih matrica, tečnosti riječi koje počinju slovom F ili S, te procjena učenja i zaboravljanja putem testa auditivno-



Slika 27. Rezultati regresijskih analiza povezanosti rezultata kognitivnih testova s N-glikomom IgG-a. Na apscisi su prikazana derivirana svojstva N-glikoma, a na ordinati regresijski koeficijent (B) i pripadajuća standardna pogreška (SE). Navedeni brojevi prikazuju promjenu izraženu u standardnim devijacijama deriviranog svojstva za svaki porast od 1 standardne devijacije rezultata kognitivnih testova. Tamnije obojani stupci označavaju statistički značajnu povezanost. Kratice deriviranih svojstva su navedene u tablicama (Tablica 15 i Tablica 16).

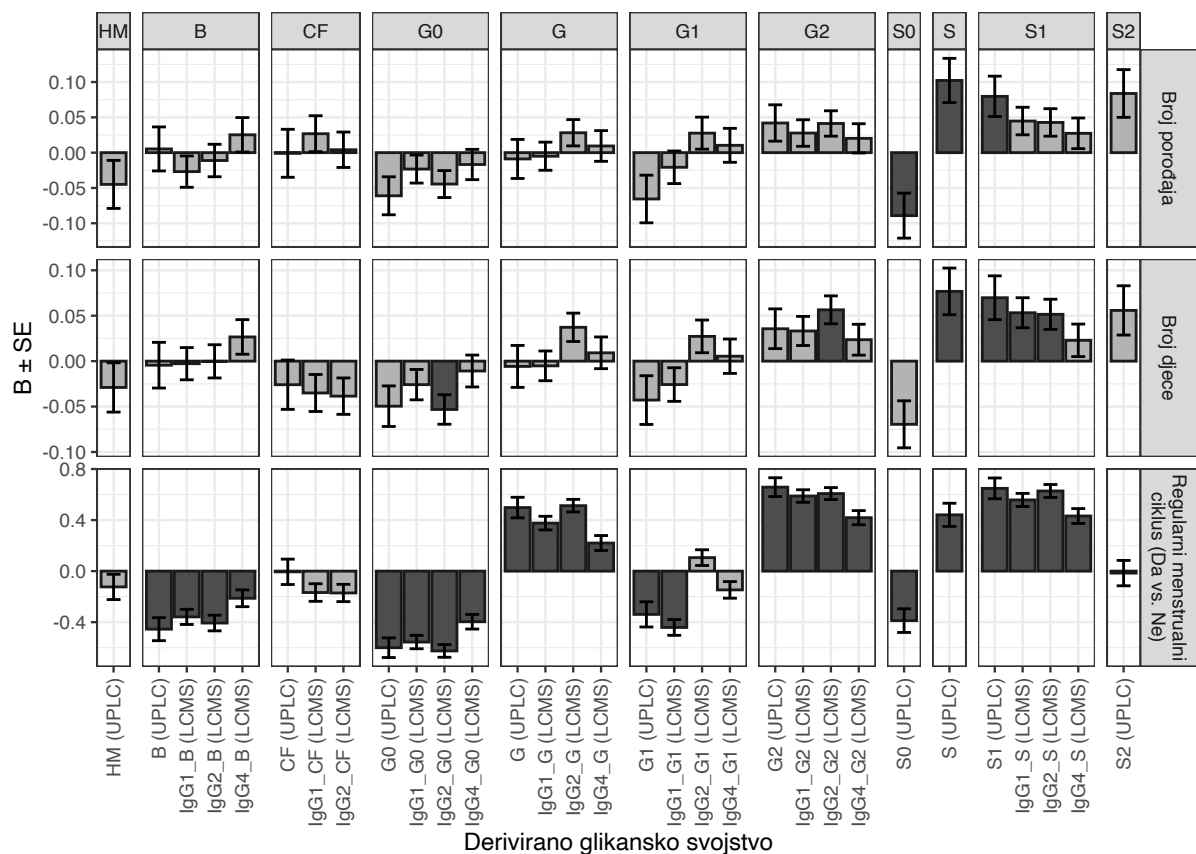
verbalnog učenja nisu pokazali nijednu statistički značajnu povezanost s deriviranim svojstvima *N*-glikoma imunoglobulina G.



Slika 28. Deskriptivna statistika numeričkih (A) i kategorijskih (B) podataka vezanih uz plodnost.

Deskriptivna statistika analiziranih podataka o plodnosti je prikazana slikom (Slika 28).

Statističkom je analizom utvrđena snažna povezanost *N*-glikoma s regularnošću menstrualnog ciklusa kod žena (Slika 29). Žene koje imaju regularan menstrualni ciklus, u odnosu na one koje to nemaju (zbog menopauze), imaju statistički značajno manje udjele struktura s račvujućim GlcNAc-om, te agalaktoziliranih, monogalaktoziliranih i asijaliniziranih struktura u *N*-glikomu IgG-a. Istovremeno su statistički značajno povećani udjeli digalaktoziliranih i monosijaliniziranih struktura. Navedeni su učinci vidljivi na svim klasama imunoglobulina G. Navedeni su rezultati objavljeni u *iScience* časopisu (133).



Slika 29. Rezultati regresijskih analiza povezanosti podataka o plodnosti s N-glikomom IgG-a. Na apscisi su prikazana derivirana svojstva N-glikoma, a na ordinati regresijski koeficijent (B) i pripadajuća standardna pogreška (SE). Navedeni brojevi prikazuju promjenu izraženu u standardnim devijacijama deriviranog svojstva za svaki porast od 1 standardne devijacije numeričke varijable o plodnosti, ili razliku između prikazanih grupa u zagradi. Tamnije obojani stupići označavaju statistički značajnu povezanost. Kratice deriviranih svojstava su navedene u tablicama (Tablica 15 i Tablica 16).

Sličan, ali znatno slabiji efekt plodnosti na N-glikom IgG-a je vidljiv i na broju porođaja i broju djece, no statistički značajne povezanosti su vidljive samo kod monosijaliniziranih struktura.

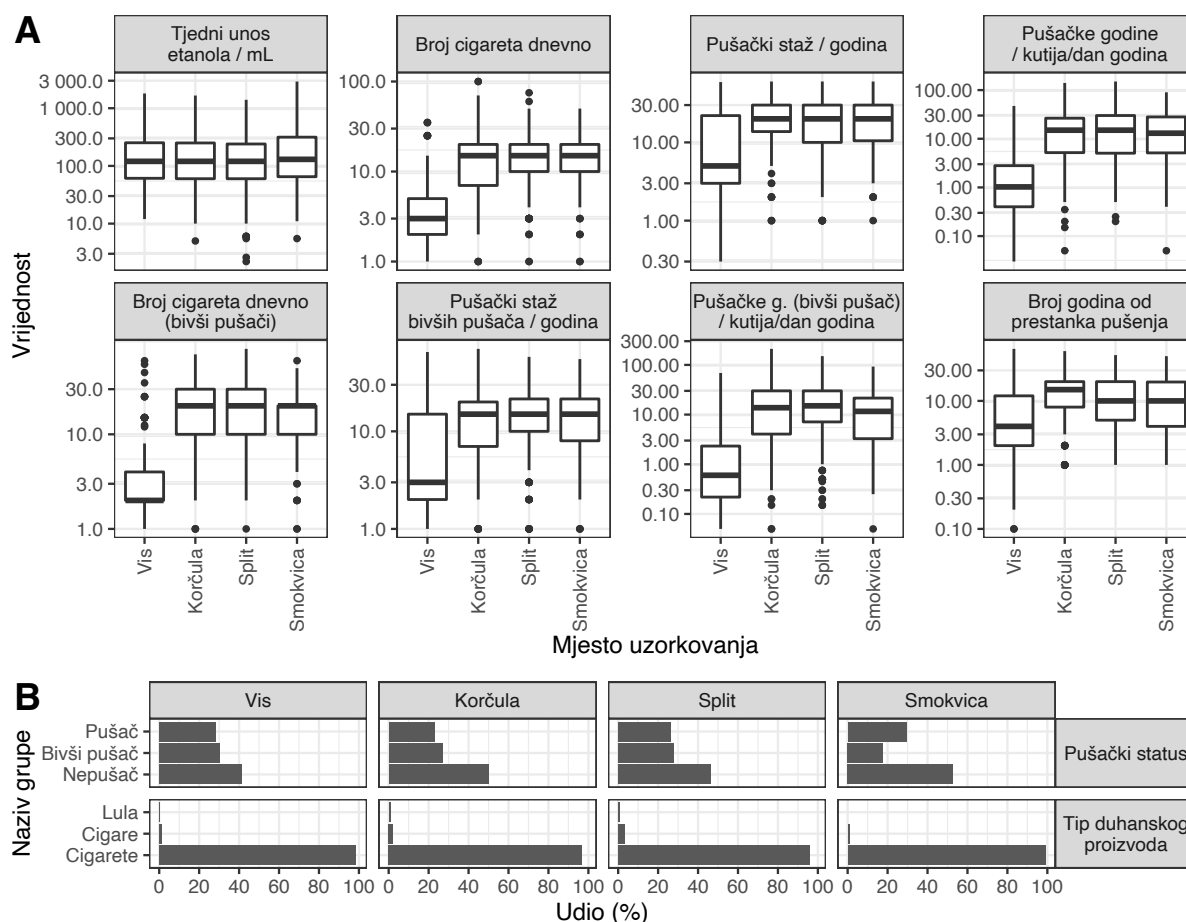
Ostale varijable povezane s plodnošću nisu pokazale statistički značajnu povezanost s nijednim deriviranim svojstvom N-glikoma imunoglobulina G.

Literaturni podaci pokazuju povezanost N-glikoma IgG-a sa koncentracijom estradiola (estrogen) kod žena koji odgovaraju opaženim rezultatima (28, 29).

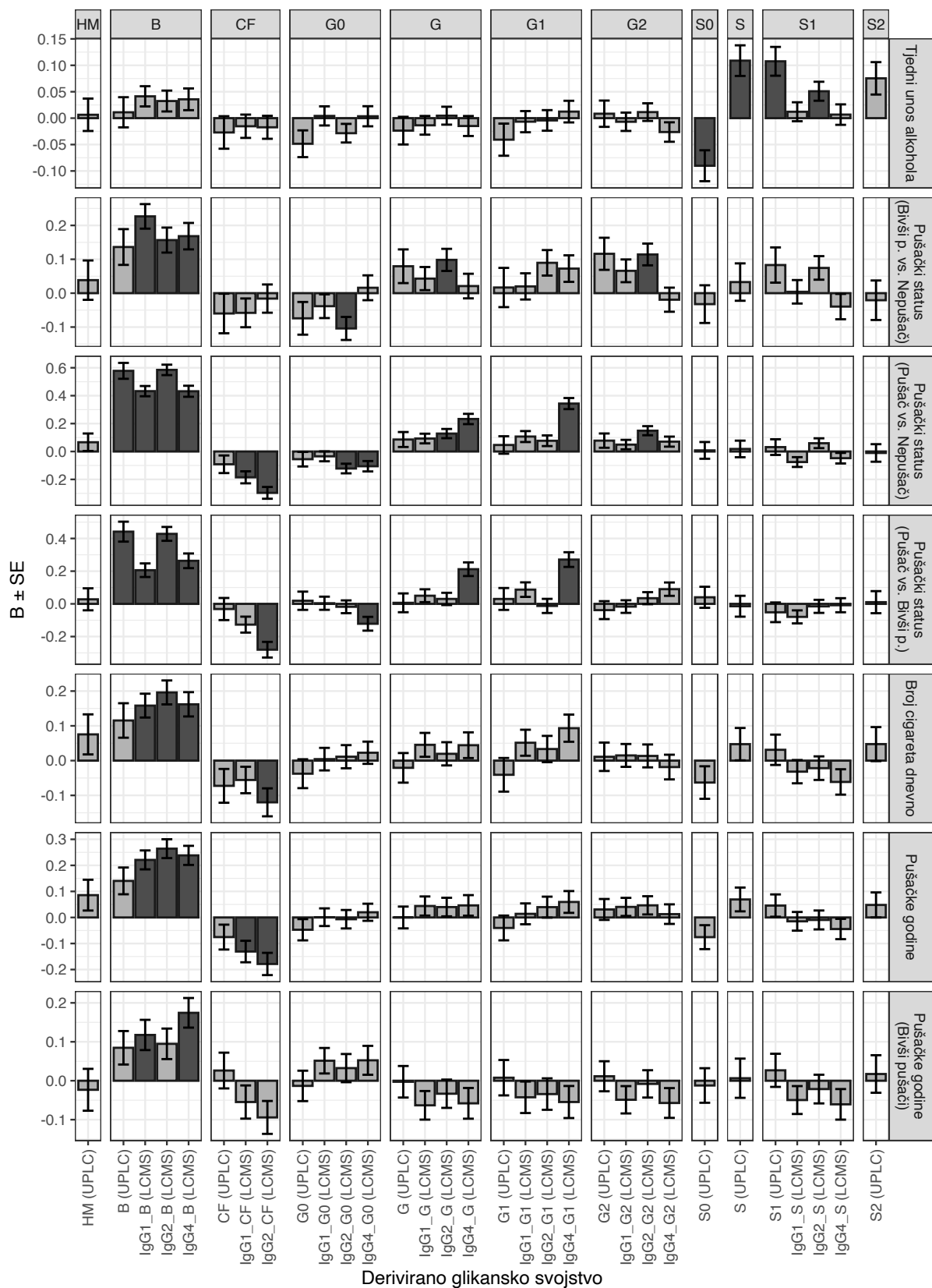
Deskriptivna statistika podataka o konzumaciji alkohola i duhanskih proizvoda je prikazana slikom (Slika 30).

Regresijskom analizom utvrđeno je da povećani unos etanola putem alkoholnih pića utječe na povećanje sijalinizacije glikana imunoglobulina G (Slika 31). *N*-glikom, dobiven UPLC analizom je pokazao da većim unosom etanola, dolazi do statistički značajnog smanjenja udjela asijaliniziranih struktura, te statistički značajnog povećanja udjela monosijaliniziranih struktura. U rezultatima LCMS analize *N*-glikoma opažena je statistički značajna promjena samo na klasi 2 IgG-a. Opaženo je i povećanje disijaliniziranih struktura s povećanjem unosa etanola, ali učinak nije statistički značajan.

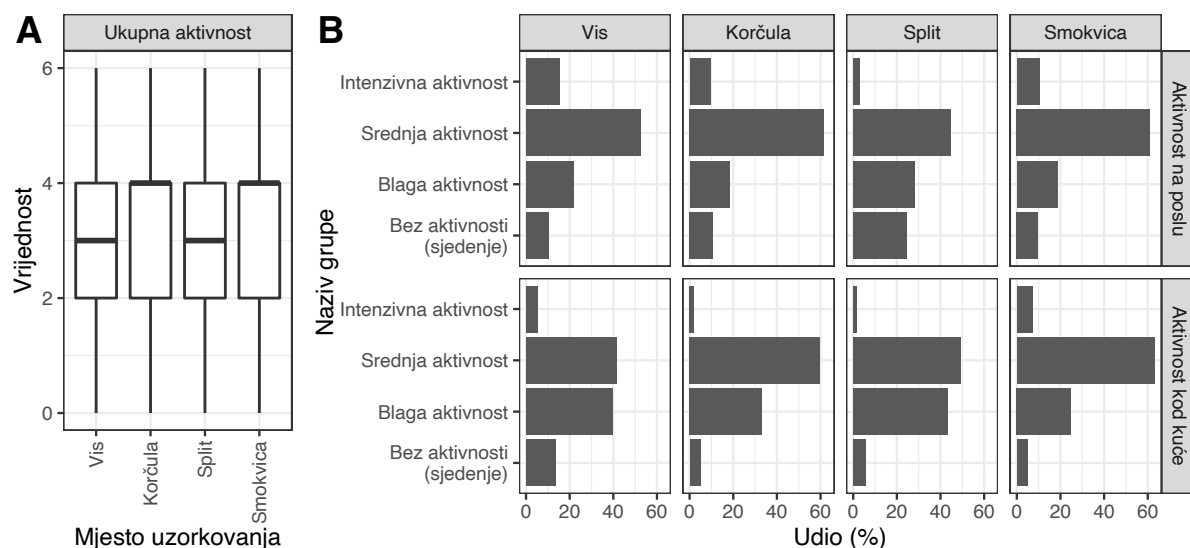
Pušenje utječe na *N*-glikom imunoglobulina G (Slika 31). Među pušačima je uočen statistički značajno veći udio glikana s računjućim GlcNAc-om, u odnosu na bivše pušače i nepušače. Prestankom pušenja, ta se razlika smanjuje, no još uvijek postoji kao statistički značajna razlika u *N*-glikomu određenom LCMS metodom. Broj popušenih cigareta dnevno, te pušačke godine (*engl.* pack-years) kod pušača, ali i bivših pušača pokazuju isti trend, no statistički značajni učinci su opaženi samo u *N*-glikomu analiziranom LCMS metodom.



Slika 30. Deskriptivna statistika numeričkih (A) i kategorijskih (B) podataka vezanih uz konzumaciju alkohola i duhanskih proizvoda. Na grafičkom prikazu numeričkih varijabli (A) ordinata je logaritamski transformirana.



Slika 31. Rezultati regresijskih analiza povezanosti podataka o konzumaciji alkoholnih i duhanskih proizvoda s *N*-glikomom IgG-a. Na apscisi su prikazana derivirana svojstva *N*-glikoma, a na ordinati regresijski koeficijent (B) i pripadajuća standardna pogreška (SE). Navedeni brojevi prikazuju promjenu izraženu u standardnim devijacijama deriviranog svojstva za svaki porast od 1 standardne devijacije numeričke varijable, ili razliku između prikazanih grupa u zagradi. Tamnije obojani stupići označavaju statistički značajnu povezanost. Kratice deriviranih svojstava su navedene u tablicama (Tablica 15 i Tablica 16).



Slika 32. Deskriptivna statistika ukupne tjelesne aktivnosti (A) dobivena zbrajanjem kodirane tjelesne aktivnosti na poslu i kod kuće (B).

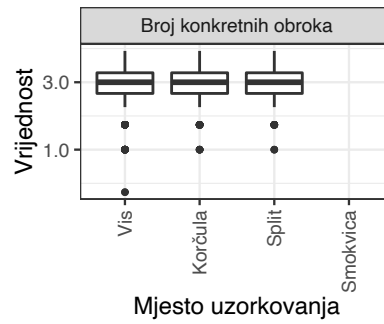
U dosadašnjim istraživanjima je također pokazana povezanost pušenja s udjelom račvajućeg GlcNAc u *N*-glikomu IgG-a na populaciji ispitanika s kroničnom opstruktivnom plućnom bolesti (KOPB), no u citiranom je istraživanju bilo nemoguće odvojiti učinak pušenja od učinka KOPB-a (36).

Deskriptivna statistika tjelesne aktivnosti prikazana je slikom (**Error! Reference source not found.**). Statističkom analizom nije utvrđena povezanost tjelesne aktivnosti i *N*-glikoma imunoglobulina G. S druge strane, literaturni podaci pokazuju da vrlo intenzivni treninzi ipak mogu uzrokovati smanjenje udjela agalaktoziliranih struktura i povećanje udjela digalaktoziliranih i monosijaliniziranih struktura u *N*-glikomu imunoglobulina G (134).

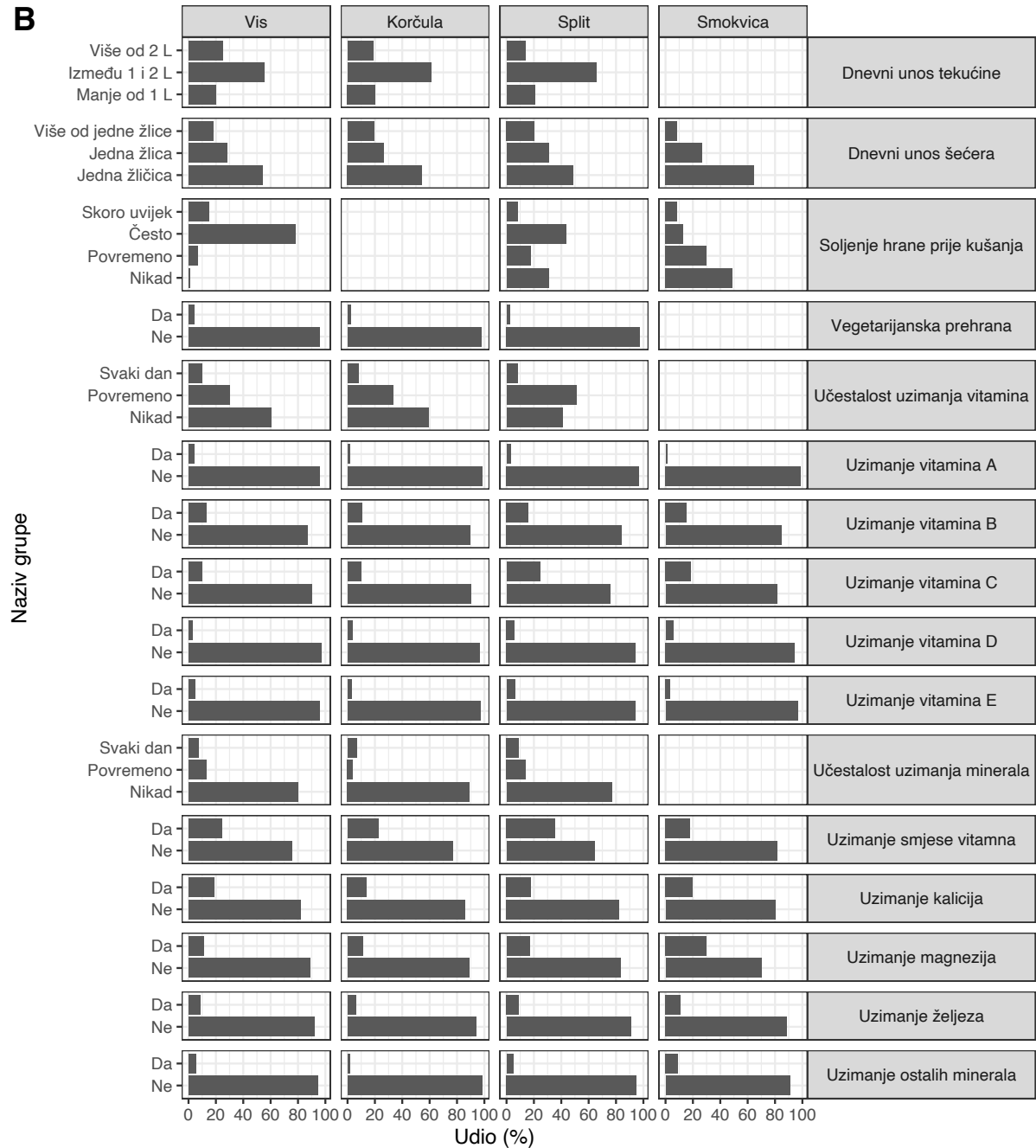
Deskriptivna statistika podataka dobivenih općenitim upitnikom o prehrani je prikazana slikom (Slika 33).

Uočena povezanost između uzimanja željeza kroz dodatke prehrani i povećanja udjela digalaktoziliranih struktura u *N*-glikomu imunoglobulina G klase 1 i 2 (Slika 34). Povećanje broja konkretnih obroka u danu je povezan sa smanjenjem udjela struktura s račvajućim GlcNAc-om, pri čemu taj trend nije bio statistički značajan u *N*-glikomu određenom UPLC metodom. Uočene su i druge sporadične povezanosti no zabilježene su samo jednom metodom određivanja *N*-glikoma, i/ili često ograničene samo na jednu klasu imunoglobulina G (Slika 34).

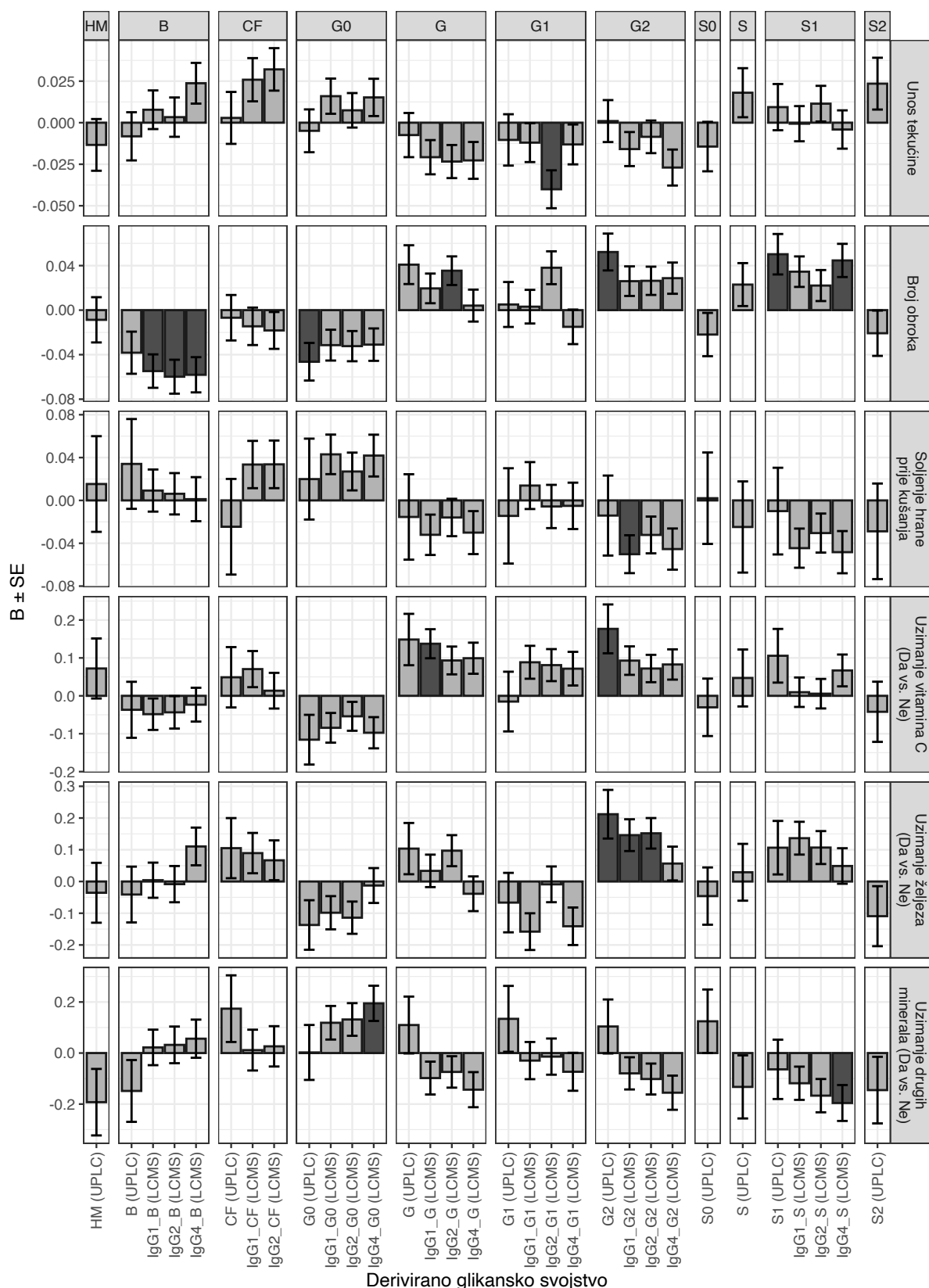
A



B



Slika 33. Deskriptivna statistika numeričkih (A) i kategorijskih (B) podataka prikupljenih općim upitnikom o prehrani.



Slika 34. Rezultati regresijskih analiza povezanosti podataka dobivenih općenitim upitnikom o prehrani s *N*-glikomom IgG-a. Na apscisi su prikazana derivirana svojstva *N*-glikoma, a na ordinati regresijski koeficijent (B) i pripadajuća standardna pogreška (SE). Navedeni brojevi prikazuju promjenu izraženu u standardnim devijacijama deriviranog svojstva za svaki porast od 1 standardne devijacije numeričke varijable, ili razliku između prikazanih grupa u zagradi. Tamnije obojani stupići označavaju statistički značajnu povezanost. Kratice deriviranih svojstava su navedene u tablicama (Tablica 15 i Tablica 16).



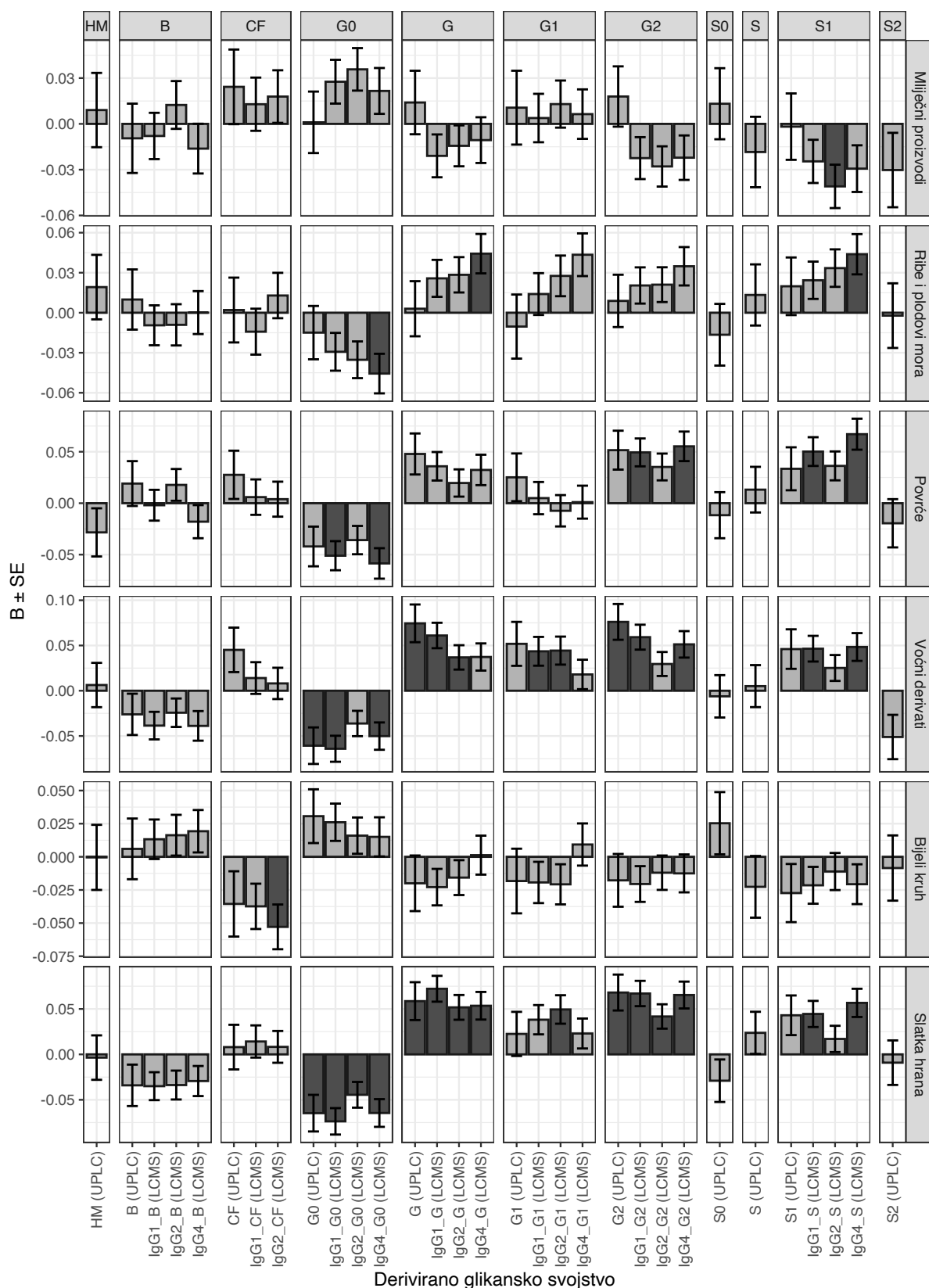
Slika 35. Učestalost konzumiranja širokog spektra namirnica.

Deskriptivna statistika upitnika o učestalosti konzumiranja pojedinih prehrambenih namirnica je prikazana slikom (Slika 35). S ciljem smanjenja dimenzionalnosti dobiveni su rezultati podvrgnuti faktorskoj analizi. Identificirano je ukupno devet faktora. Prvi faktor, nazvan *mliječni proizvodi*, je obuhvatio učestalost konzumacije mlijeka, jogurta, kiselog vrhnja, svježeg sira i punomasnog sira. Drugi faktor, nazvan *crveno meso*, je obuhvatio učestalost konzumacije svinjetine, teletine, junetine, janjetine, iznutrica (jetra, bubreg, pluća,...), slanine, kobasica, salame, konzerviranih mesnih derivata, pršuta i jaja. Treći faktor, nazvan *ribe i plodovi mora*, je obuhvatio učestalost konzumacije bijele i plave ribe, morskih plodova, lignji i hobotnica, te suhe ribe. Četvrti faktor, nazvan *povrće*, je obuhvatio učestalost konzumacije lisnatog, korjenastog, cvjetastog i plodastog povrća te mahunarki. Peti faktor, nazvan *voćni derivati*, je obuhvatio učestalost konzumacije orašastih plodova, voćnih kompota, suhog voća i muesli žitarica. Šesti faktor, nazvan *bijeli kruh*, je obuhvatio učestalost konzumacije bijelog i integralnog kruha pri čemu je kodiranje učestalosti konzumacije integralnog kruha invertirano (češća konzumacija je kodirana manjim brojevima). Sedmi faktor, nazvan *škrob*, je obuhvatio učestalost konzumacije tjestenine i riže, te krumpira. Osmi faktor, nazvan *slatka hrana*, je obuhvatio učestalost konzumacije kolača, čokolade, keksa, bombona, te džema i marmelade. Deveti faktor, nazvan *sokovi*, obuhvatio je učestalost konzumacije sokova od povrća, sokova od voća, Cedevite i gaziranih sokova. Prehrambene namirnice koje nisu pridružene niti jednom faktoru nisu analizirane.

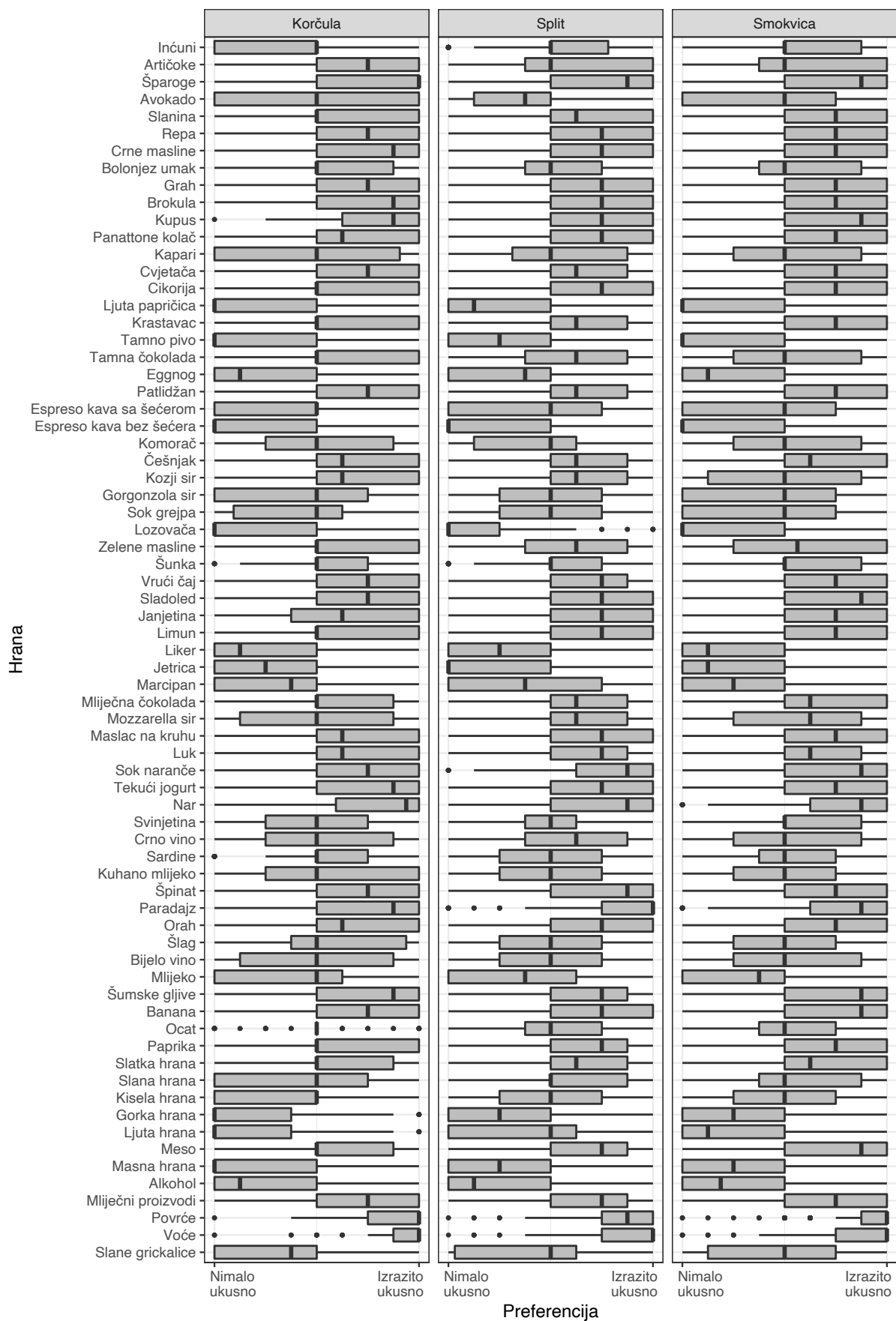
Rezultati povezanosti faktora s *N*-glikomom imunoglobulina G za one faktore kod kojih je barem jedno derivirano svojstvo *N*-glikoma bilo statistički značajno je prikazano slikom (Slika 36). Uočeni rezultati pokazuju da povećanim unosom slatke hrane se smanjuje udio agalaktoziliranih struktura, te povećava udio digalaktoziliranih struktura. Identičan obrazac je uočen i za voćne derivate pri čemu povezanost nije značajna za IgG klase 2, te za povrće gdje statistički značajna povezanost nestaje i za derivirana svojstva dobivena UPLC analizom. Osim galaktozilacije, postoji i povezanost češće konzumacije povrća, voćnih derivata i slatke hrane s povećanjem udjela monosijaliniziranih struktura na klasama 1 i 4 imunoglobulina G.

Ostale su detektirane povezanosti sporadične i nisu replicirane u obje kromatografske metode (Slika 36), a za faktore crveno meso, škrob i sokovi nije uočena nikakva statistički značajna povezanost.

Pretraživanjem literature nisu pronađeni radovi koji povezuju *N*-glikom IgG-a s vrstom prehrane.



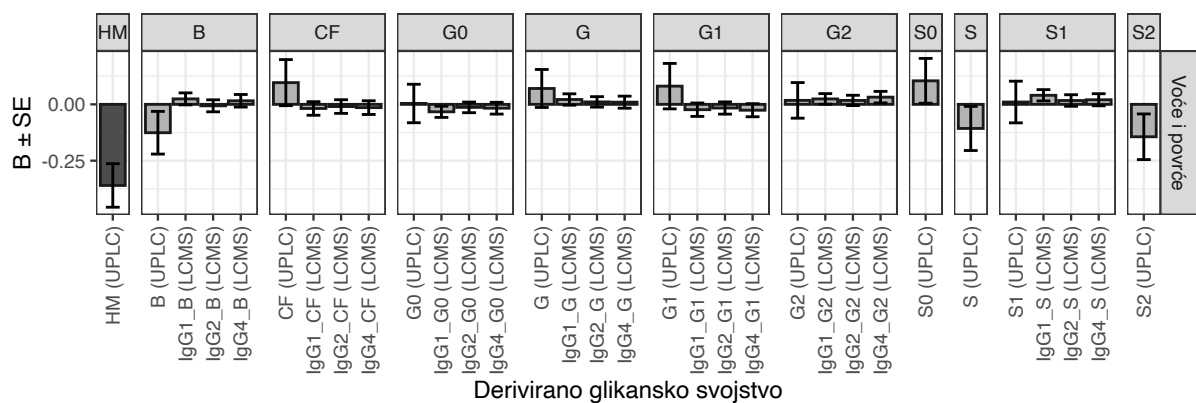
Slika 36. Rezultati regresijskih analiza povezanosti faktora identificiranih faktorskom analizom podataka dobivenih upitnikom o učestalosti konzumacije prehrambenih proizvoda s N-glikomom IgG-a. Na apscisi su prikazana derivirana svojstva N-glikoma, a na ordinati regresijski koeficijent (B) i pripadajuća standardna pogreška (SE). Navedeni brojevi prikazuju promjenu izraženu u standardnim devijacijama deriviranog svojstva za svaki porast od 1 standardne devijacije danog faktora. Tamnije obojani stupići označavaju statistički značajnu povezanost. Kratice deriviranih svojstava su navedene u tablicama (Tablica 15 i Tablica 16).



Slika 37. Deskriptivna statistika podataka dobivenih upitnikom o preferenciji hrane.

Deskriptivna statistika podataka dobivenih upitnikom o preferenciji pojedinih prehrambenih proizvoda je prikazana slikom (Slika 37). Dobiveni su rezultati podvrgnuti faktorskoj analizi i pritom je identificirano pet faktora. Prvi faktor, nazvan *voće i povrće*, obuhvatio je preferenciju različitih voća i povrća: artičoke, šparoge, crne i zelene masline, repa, grah, brokula, kupus, cvjetača, cikorija, krastavac, patlidžan, komorač, češnjak, sok grejpa, vrući čaj, nar, orah, špinat, šumske gljive, ocat, paradajz, paprika, općenito voće i općenito povrće. Drugi faktor, nazvan *specifični okusi*, obuhvatio je preferenciju namirnica jedinstvenog okusa: avokado, bolonjez umak, ljuta papričica, čokolada s visokim udjelom kakaa, kozji sir, kapari, tamno pivo, eggnog, espresso kava bez šećera, mozzarella sir, gorgonzola sir, liker, lozovača, jetrica, marcipan, puno mlijeko, alkohol. Treći faktor, nazvan *slatko*, obuhvatio je preferenciju namirnica većinom slatkog okusa: panettone kolač, sladoled, mliječna čokolada, slatka hrana, maslac na kruhu, šlag, banana, mliječni proizvodi i slane grickalice. Četvrti faktor, zvan *meso*, obuhvatio je preferenciju mesnate hrane: svinjetina, janjetina, šunka, slanina, bijelo vino i meso općenito. Peti faktor, nazvan *intenzivni okusi*, obuhvatio je pitanja o preferenciji općenitih okusa: slana hrana, gorka hrana, ljuta hrana, kisela hrana i masna hrana.

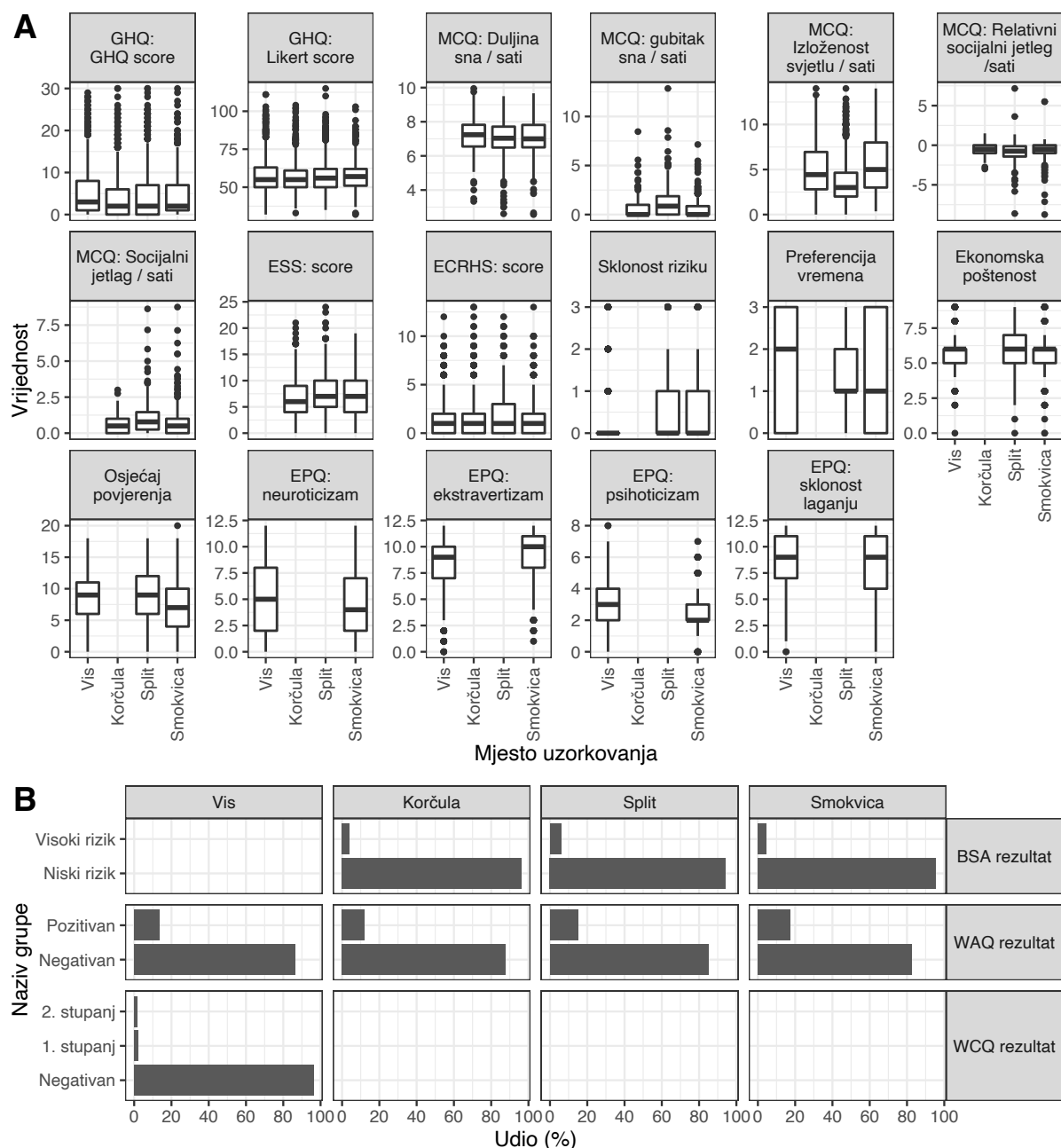
Regresijskim je analizama identificiranih faktora pronađeno da veća preferencija voća i povrća je povezana sa smanjenjem udjela struktura s visokim sadržajem manoze (Slika 38).



Slika 38. Rezultati regresijskih analiza povezanosti faktora identificiranih faktorskom analizom podataka dobivenih upitnikom o učestalosti konzumacije prehrambenih proizvoda s N-glikomom IgG-a. Na apscisi su prikazana derivirana svojstva N-glikoma, a na ordinati regresijski koeficijent (B) i pripadajuća standardna pogreška (SE). Navedeni brojevi prikazuju promjenu izraženu u standardnim devijacijama deriviranog svojstva za svaki porast od 1 standardne devijacije danog faktora. Tamnije obojani stupci označavaju statistički značajnu povezanost. Kratice deriviranih svojstava su navedene u tablicama (Tablica 15 i Tablica 16).

Regresijskim analizama ostalih faktora nije uočena statistički značajna povezanost ni sa jednim deriviranim svojstvom N-glikoma imunoglobulina G.

Deskriptivna statistika podataka dobivenih Općim zdravstvenim upitnikom, Mühenskim upitnikom o kronotipu, Epworthovom ljestvicom pospanosti, Berlinskim upitnikom za procjenu rizika od opstruktivne apneje u spavanju, Upitnikom Europske zajednice o

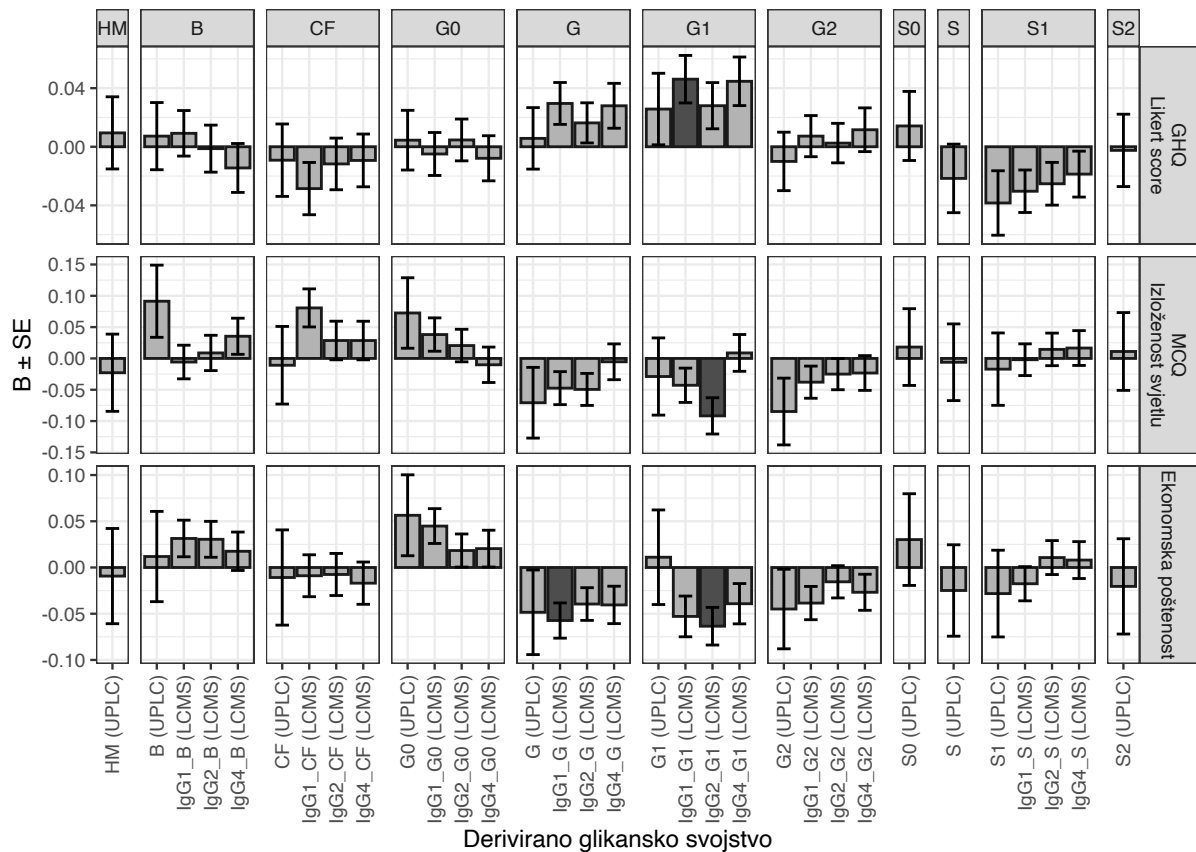


Slika 39. Deskriptivna statistika numeričkih (A) i kategorijskih (B) podataka dobivenih upitnicima. Kratice: GHQ – opći zdravstveni upitnik (*engl.* General Health Questionnaire); MCQ – Miichenski upitnik o kronotipu (*engl.* Munich chronotype questionnaire); ESS – Epworthova ljestvica pospanosti (*engl.* Epworth sleepiness scale); ECRHS – upitnik Europske zajednice o respiratornom zdravlju (*engl.* European Community Respiratory Health Survey); EPQ – revidirani Eysneckov test osobnosti (*engl.* Revised Eysenck Personality Questionnaire); BSA – Berlinski upitnik za procjenu rizika opstruktivne apneje u spavanju (*engl.* Berlin Sleep Apnoea Questionnaire); WAQ – upitnik Svjetske zdravstvene organizacije o *angina pectoris* (*engl.* WHO Rose angina questionnaire); WCQ – upitnik Svjetske zdravstvene organizacije o intermitentnoj klaudikaciji (*engl.* WHO Questionnaire on intermittent claudication).

respiratornom zdravlju, Upitnikom Svjetske zdravstvene organizacije o *angina pectoris*, Upitnikom Svjetske zdravstvene organizacije o intermitentnoj klaudikaciji, upitnikom o sociološkom ponašanju i revidiranim Eysneckovim testom osobnosti je prikazana slikom (Slika 39).

Regresijskim je analizama uočena povezanost tek malog broja rezultata upitnika i deriviranih svojstava *N*-glikoma IgG-a. Opažene statistički značajne povezanosti su pronađene kod monogalaktoziliranih struktura klase 1 ili 2 imunoglobulina G analiziranih LCMS analizom i nisu potvrđene UPLC analizom, detaljnije je prikazano slikom (Slika 40).

Rezultati ostalih upitnika nisu bili statistički značajno povezani ni sa jednim deriviranim svojstvom *N*-glikoma imunoglobulina G.



Slika 40. Rezultati regresijskih analiza povezanosti pojedinih rezultata Općeg zdravstvenog upitnika, Münchenskog upitnika o kronotipu i upitnika o sociološkom ponašanju s *N*-glikomom IgG-a. Na apscisi su prikazana derivirana svojstva *N*-glikoma, a na ordinati regresijski koeficijent (B) i pripadajuća standardna pogreška (SE). Navedeni brojevi prikazuju promjenu izraženu u standardnim devijacijama deriviranog svojstva za svaki porast od jedne standardne devijacije danog rezultata. Tamnije obojani stupići označavaju statistički značajnu povezanost. Kratice deriviranih svojstava su navedene u tablicama (Tablica 15 i Tablica 16). Ostale kratice: GHQ – opći zdravstveni upitnik (*engl.* general health questionnaire, GHQ-30); MCQ – Münchenski upitnik o kronotipu (*engl.* Munich chronotype questionnaire).

4 ZAKLJUČAK

Analizom proteinskih sekvenci ljudskih gena koji kodiraju za proteine koji sudjeluju u glikozilaciji metodom genomske filostratigrafije utvrđeno je da je glavnina gena vezana uz glikozilaciju koje danas nalazimo kod čovjeka nastala još u zajedničkom pretku svih staničnih organizama.

Analizom proteinskih sekvenci ljudskih gena koji kodiraju za polipeptidne lance glikoproteina metodom genomske filostratigrafije utvrđeno je da su se geni koji kodiraju za polipeptidne lance glikoproteina kakve danas nalazimo kod čovjeka intenzivno razvijali prilikom razvoja kompleksnijih životinja.

Geni koji kodiraju za enzime koji kataliziraju reakcije biosintetskog puta *N*-glikana lokalizirane na luminalnoj strani membrane endoplazmatskog retikuluma su inovacija zajedničkog pretka svih eukariota.

Geni koji kodiraju za enzime koji kataliziraju reakcije biosintetskog puta *N*-glikana lokalizirane na citosolnoj strani membrane endoplazmatskog retikuluma su porijeklom iz zajedničkog pretka svih staničnih organizama.

Analizom *N*-glikoma imunoglobulina G dvjema metodama kromatografije je utvrđeno da:

- udio agalaktoziliranih (G0) struktura u *N*-glikomu imunoglobulina G:
 - se smanjuje s učestalijim konzumiranjem voćnih derivata (suhog voća, voćnih kompota, orašastih plodova) i slatke hrane (kolača, čokolade, bombona,...);
 - je manji kod ispitanika koji boluju od hipertenzije, artritisa i kod žena koje imaju regularni menstrualni ciklus;
 - se povećava s povećanjem debljine, krvnog tlaka, koncentracije urata, koncentracije gliciranog hemoglobina, koncentracije fibrinogena i koncentracije inzulina u plazmi.
- udio monogalaktoziliranih (G1) struktura u *N*-glikomu imunoglobulina G:
 - se smanjuje s povećanjem udjela gliciranog hemoglobina;
 - se povećava s povećanjem koncentracije albumina u plazmi.
- udio digalaktoziliranih (G2) struktura u *N*-glikomu imunoglobulina G:

- se smanjuje s povećanjem debljine, krvnog tlaka, koncentracije urata, koncentracije triglicerida, koncentracije fibrinogena, koncentracije inzulina u plazmi;
- je manji kod ispitanika koji boluju od hipertenzije, artritisa;
- se povećava s povećanjem mineralne gustoće kostiju, koncentracije HDL kolesterola, broja postignutog bodova na kognitivnom testu kodiranja i učestalijim konzumiranjem voćnih derivata (suhog voća, voćnih kompota, orašastih plodova) i slatke hrane (kolača, čokolade, bombona,..);
- je veći kod žena koje imaju regularni menstrualni ciklus, kod ispitanika koji uzimaju dodatke prehrani koji sadrže željezo.
- udio monosijaliniziranih (S1) struktura u *N*-glikomu imunoglobulina G:
 - se smanjuje s povećanjem debljine, krvnog tlaka, koncentracije triglicerida i koncentracije inzulina u plazmi;
 - povećava s povećanjem mineralne gustoće kostiju i brojem djece;
 - je veći kod žena koje imaju regularni menstrualni ciklus.
- udio struktura s račvajućim *N*-acetilglukozaminom (B) u *N*-glikomu imunoglobulina G:
 - se smanjuje s povećanjem mineralne gustoće kostiju;
 - je manji kod žena koje imaju regularni menstrualni ciklus ;
 - se povećava s povećanjem koncentracije kolesterola i triglicerida u plazmi;
 - je veći kod pušača u odnosu na nepušače i kod pušača u odnosu na bivše pušače.

Analizom *N*-glikoma imunoglobulina G dvjema metodama kromatografije nije utvrđeno da *N*-glikom imunoglobulina G sustavno ovisi o:

- materijalom statusu, obrazovanju i radnom statusu;
- brzini otkucaja srca, gležanjskom indeksu i bioimpedanciji;
- prisutnosti koronarne bolesti srca, glaukoma, čira, dijabetesa (nepoznatog tipa), cerebrovaskularnog infarkta, shizofrenije, manije/depresije, malignog tumora, gihta i bolesti bubrega;
- promjeni koncentracija kreatinina i kalcija;
- rezultatima kognitivnih testova: standardne progresivne matrice, testa tečnosti riječi, Mill Hill testa i testa auditivno-verbalnog učenja;
- dobi prve menstruacije, dobi nastupanja menopauze, broju porođaja, broju mrtvorođenčadi i broju spontanih ili kirurških pobačaja;

- tjednom unosu alkohola, broju popušenih cigareta dnevno, pušačkim godinama (*engl. pack-year*), pušačkom stažu i tjelesnoj aktivnosti;
- količini unosa tekućine, broju konkretnih obroka u danu, konzumaciji dodatne soli i šećera, uzimanju dodataka prehrane koji sadrže vitamine A, B, C, D, E i smjese vitamina, te minerale kalcija, magnezija i ostalih minerala koji ne sadrže željezo;
- učestalosti konzumacije mliječnih proizvoda, ribe i plodova mora, mesa, povrća i kruha;
- preferenciji hrane i pića;
- podacima dobivenim:
 - Općim zdravstvenim upitnikom;
 - Münchenskim upitnikom o kronotipu;
 - Epworthovom ljestvicom pospanosti;
 - Berlinskim upitnikom za procjenu rizika od opstruktivske apneje u spavanju;
 - Upitnikom Europske zajednice o respiratornom zdravlju;
 - Upitnikom Svjetske zdravstvene organizacije o *angina pectoris*;
 - Upitnikom Svjetske zdravstvene organizacije o intermitentnoj klaudikaciji;
 - Upitnikom o sociološkom ponašanju;
 - revidiranim Eysneckovim testom.

5 POPIS LITERATURE

1. Flynn RA, Pedram K, Malaker SA, Batista PJ, Smith BAH, Johnson AG, et al. Small RNAs are modified with N-glycans and displayed on the surface of living cells. *Cell* 2021;184:3109-3124.e22.
2. Merrill AH, Vu MN. Glycolipids. In: Bradshaw R, Stahl P, editors. *Encyclopedia of Cell Biology* 1st ed. Amsterdam: Elsevier; 2016:180–93.
3. Kinoshita T, Murakami Y, Morita YS. Diseases Associated with GPI Anchors. In: Kamerling H, editor. *Comprehensive Glycoscience* Amsterdam: Elsevier; 2007:393–419.
4. Perez-Vilar J, Hill RL. Mucin Family of Glycoproteins. In: Lennarz WJ, Lane MD, editors. *Encyclopedia of Biological Chemistry* Amsterdam: Elsevier; 2004:758–64.
5. Song Y, Zhang F, Linhardt RJ. Glycosaminoglycans. In: Lauc G, Trbojević-Akmačić I, editors. *The Role of Glycosylation in Health and Disease* New York City: Springer; 2021:103–16.
6. Taylor ME, Drickamer K, Schnaar RL, Etzler ME, Varki A. Discovery and Classification of Glycan-Binding Proteins. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al., editors. *Essentials of Glycobiology* 3rd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015.
7. Varki A. Biological roles of glycans. *Glycobiology* 2017;27:3–49.
8. Bieberich E. Synthesis, Processing, and Function of N-glycans in N-glycoproteins. In: Yu RK, Schengrund C-L, editors. *Glycobiology of the Nervous System* New York City: Springer; 2014:47–70.
9. Stanley P, Taniguchi N, Aebi M. N-Glycans. 3rd ed. Varki A, Cummings RD, Esko JD, Stanle P, Hart GW, Aebi M, et al., editors. *Essentials of Glycobiology* [Internet]. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015.
10. Pobre-Piza KFR, Hendershot LM. Protein Folding | Protein Folding in the Endoplasmic Reticulum. In: Jez J, editor. *Encyclopedia of Biological Chemistry III* 3rd ed. Amsterdam: Elsevier; 2021:127–39.
11. Gagneux P, Aebi M, Varki A. Evolution of Glycan Diversity. In: Varki A, Cummings RD, Esko JD, Stanle P, Hart GW, Aebi M, et al., editors. *Essentials of Glycobiology* [Internet] 3rd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2015:281–689.
12. Chung CY, Majewska NI, Wang Q, Paul JT, Betenbaugh MJ. SnapShot: N-

- Glycosylation Processing Pathways across Kingdoms. *Cell* 2017;171:258-258.e1.
13. Wang P, Wang H, Gai J, Tian X, Zhang X, Lv Y, et al. Evolution of protein N-glycosylation process in Golgi apparatus which shapes diversity of protein N-glycan structures in plants, animals and fungi. *Sci Rep* 2017;7:40301.
 14. Petit D, Teppa E, Cenci U, Ball S, Harduin-Lepers A. Reconstruction of the sialylation pathway in the ancestor of eukaryotes. *Sci Rep* 2018;8:2946.
 15. Petit D, Teppa E, Mir A-M, Vicogne D, Thisse C, Thisse B, et al. Integrative view of α 2,3-sialyltransferases (ST3Gal) molecular and functional evolution in deuterostomes: significance of lineage-specific losses. *Mol Biol Evol* 2015;32:906–27.
 16. Lombard J. The multiple evolutionary origins of the eukaryotic N-glycosylation pathway. *Biol Direct* 2016;11:36.
 17. Joshi HJ, Narimatsu Y, Schjoldager KT, Tytgat HLP, Aebi M, Clausen H, et al. SnapShot: O-Glycosylation Pathways across Kingdoms. *Cell* 2018;172:632-632.e2.
 18. Brockhausen I, Stanley P. O-GalNAc Glycans. In: Varki A, Cummings RD, Esko JD, Stanle P, Hart GW, Aebi M, et al., editors. *Essentials of Glycobiology* [Internet] 3rd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2017.
 19. Brockhausen I. Mucin-type O-glycans in human colon and breast cancer: Glycodynamics and functions. Vol. 7, *EMBO Reports*. 2006.
 20. Aquino RS, Lee ES, Park PW. Diverse Functions of Glycosaminoglycans in Infectious Diseases. In: Teplow D, editor. *Progress in Molecular Biology and Translational Science* Amsterdam: Elsevier; 2010:373–94.
 21. Yamada S, Sugahara K, Özbek S. Evolution of glycosaminoglycans. *Commun Integr Biol* 2011;4:150–8.
 22. Knezević A, Polasek O, Gornik O, Rudan I, Campbell H, Hayward C, et al. Variability, heritability and environmental determinants of human plasma N-glycome. *J Proteome Res* 2009;8:694–701.
 23. Zaytseva OO, Freidin MB, Keser T, Štambuk J, Ugrina I, Šimurina M, et al. Heritability of Human Plasma N -Glycome. *J Proteome Res* 2020;19:85–91.
 24. Saldova R, Huffman JE, Adamczyk B, Mužinić A, Kattla JJ, Pučić M, et al. Association of medication with the human plasma N -Glycome. *J Proteome Res* 2012;11:1821–31.
 25. Knežević A, Gornik O, Polašek O, Pučić M, Redžić I, Novokmet M, et al. Effects of aging, body mass index, plasma lipid profiles, and smoking on human plasma N-glycans. *Glycobiology* 2010;20:959–69.
 26. Frkatovic A, Zaytseva OO, Klaric L. Genetic Regulation of Immunoglobulin G

- Glycosylation. In: *Experientia supplementum* (2012) 2021:259–87.
27. Gudelj I, Lauc G, Pezer M. Immunoglobulin G glycosylation in aging and diseases. *Cell Immunol* 2018;333:65–79.
 28. Mijakovac A, Jurić J, Kohrt WM, Krištić J, Kifer D, Gavin KM, et al. Effects of Estradiol on Immunoglobulin G Glycosylation: Mapping of the Downstream Signaling Mechanism. *Front Immunol* 2021;12:1–13.
 29. Jurić J, Kohrt WM, Kifer D, Gavin KM, Pezer M, Nigrovic PA, et al. Effects of estradiol on biological age measured using the glycan age index. *Aging (Albany NY)* 2020;12:19756–65.
 30. Ercan A, Kohrt WM, Cui J, Deane KD, Pezer M, Yu EW, et al. Estrogens regulate glycosylation of IgG in women and men. *JCI Insight* 2017;
 31. Bondt A, Rombouts Y, Selman MHJ, Hensbergen PJ, Reiding KR, Hazes JMW, et al. Immunoglobulin G (IgG) fab glycosylation analysis using a new mass spectrometric high-throughput profiling method reveals pregnancy-associated changes. *Mol Cell Proteomics* 2014;13:3029–39.
 32. Wang Y, Klarić L, Yu X, Thaqi K, Dong J, Novokmet M, et al. The Association Between Glycosylation of Immunoglobulin G and Hypertension. *Medicine (Baltimore)* 2016;95:e3379.
 33. Wahl A, Kasela S, Carnero-Montoro E, van Iterson M, Štambuk J, Sharma S, et al. IgG glycosylation and DNA methylation are interconnected with smoking. *Biochim Biophys Acta - Gen Subj* 2018;1862:637–48.
 34. Nikolac Perkovic M, Pucic Bakovic M, Kristic J, Novokmet M, Huffman JE, Vitart V, et al. The association between galactosylation of immunoglobulin G and body mass index. *Prog Neuro-Psychopharmacology Biol Psychiatry* 2014;48:20–5.
 35. Krištić J, Vučković F, Menni C, Klarić L, Keser T, Beceheli I, et al. Glycans Are a Novel Biomarker of Chronological and Biological Ages. *Journals Gerontol Ser A* 2014;69:779–89.
 36. Pavić T, Dilber D, Kifer D, Selak N, Keser T, Ljubičić CDS, et al. N-glycosylation patterns of plasma proteins and immunoglobulin G in chronic obstructive pulmonary disease 11 *Medical and Health Sciences 1102 Cardiorespiratory Medicine and Haematology. J Transl Med* 2018;16:323.
 37. Neme R, Tautz D. Evolution: Dynamics of De Novo Gene Emergence. *Curr Biol* 2014;24:R238–40.
 38. Tautz D, Domazet-Lošo T. The evolutionary origin of orphan genes. *Nat Rev Genet*

- 2011;12:692–702.
39. Domazet-Lošo T, Tautz D. An Evolutionary Analysis of Orphan Genes in *Drosophila*. *Genome Res* 2003;13:2213–9.
 40. Long M, Betrán E, Thornton K, Wang W. The origin of new genes: glimpses from the young and old. *Nat Rev Genet* 2003;4:865–75.
 41. Domazet-Lošo T, Brajković J, Tautz D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet* 2007;23:533–9.
 42. Domazet-Lošo T, Tautz D. An Ancient Evolutionary Origin of Genes Associated with Human Genetic Diseases. *Mol Biol Evol* 2008;25:2699–707.
 43. Domazet-Lošo T, Carvunis A, Mar Albà M, Sebastijan Šestak M, Bakarić R, Neme R, et al. No evidence for phylostratigraphic bias impacting inferences on patterns of gene emergence and evolution. *Mol Biol Evol* 2017;msw284.
 44. Lauc G, Vojta A, Zoldoš V. Epigenetic regulation of glycosylation is the quantum mechanics of biology. *Biochim Biophys Acta - Gen Subj* 2014;1840:65–70.
 45. Rademacher TW, Parekh RB, Dwek RA, Isenberg D, Rook G, Axford JS, et al. The role of IgG glycoforms in the pathogenesis of rheumatoid arthritis. *Springer Semin Immunopathol* 1988;10:231–49.
 46. Rudman N, Gornik O, Lauc G. Altered N-glycosylation profiles as potential biomarkers and drug targets in diabetes. Vol. 593, *FEBS Letters*. 2019.
 47. Lauc G, Pezer M, Rudan I, Campbell H. Mechanisms of disease: The human N-glycome. *Biochim Biophys Acta - Gen Subj* 2016;1860:1574–82.
 48. Hayes JM, Cosgrave EFJ, Struwe WB, Wormald M, Davey GP, Jefferis R, et al. Glycosylation and Fc Receptors. In: Daeron M, Nimmerjahn F, editors. *Fc Receptors* New York City: Springer International Publishing; 2014:165–99.
 49. Shields RL, Lai J, Keck R, O'Connell LY, Hong K, Meng YG, et al. Lack of fucose on human IgG1 N-linked oligosaccharide improves binding to human Fcγ₃ and antibody-dependent cellular toxicity. *J Biol Chem* 2002;277:26733–40.
 50. Malhotra R, Wormald MR, Rudd PM, Fischer PB, Dwek RA, Sim RB. Glycosylation changes of IgG associated with rheumatoid arthritis can activate complement via the mannose-binding protein. *Nat Med* 1995;1:237–43.
 51. Kaneko Y, Nimmerjahn F, Ravetch J V. Anti-inflammatory activity of immunoglobulin G resulting from Fc sialylation. *Science* 2006;313:670–3.
 52. Mimura Y, Katoh T, Saldova R, O'Flaherty R, Izumi T, Mimura-Kimura Y, et al.

- Glycosylation engineering of therapeutic IgG antibodies: challenges for the safety, functionality and efficacy. *Protein Cell* 2018;9:47–62.
53. Moremen KW, Tiemeyer M, Nairn A V. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol* 2012;13:448–62.
 54. Kanehisa M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 2000;28:27–30.
 55. Cantarel BI, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res* 2009;37.
 56. Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al. *Essentials of Glycobiology* [Internet]. 3rd ed. Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, et al., editors. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2017.
 57. Schjoldager KT, Narimatsu Y, Joshi HJ, Clausen H. Global view of human protein glycosylation pathways and functions. *Nat Rev Mol Cell Biol* 2020;21:729–49.
 58. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, et al. UniProtKB/Swiss-Prot, the Manually Annotated Section of the UniProt KnowledgeBase: How to Use the Entry View. In: Edwards D, editor. *Plant Bioinformatics* New York City: Springer; 2016:23–54.
 59. Lex A, Gehlenborg N, Strobel H, Vuillemot R, Pfister H. UpSet: Visualization of intersecting sets. *IEEE Trans Vis Comput Graph* 2014;20:1983–92.
 60. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström Di, Juzokaite L, Vancaester E, et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 2017;541:353–8.
 61. Adam PS, Borrel G, Brochier-Armanet C, Gribaldo S. The growing tree of Archaea: New perspectives on their diversity, evolution and ecology. Vol. 11, *ISME Journal*. 2017:2407–25.
 62. Schoch CL, Ciufu S, Domrachev M, Hottton CL, Kannan S, Khovanskaya R, et al. NCBI Taxonomy: A comprehensive update on curation, resources and tools. *Database* 2020;2020.
 63. Richter DJ, Fozouni P, Eisen MB, King N. Gene family innovation, conservation and loss on the animal stem lineage. *Elife* 2018;7:1–43.
 64. Cunningham F, Achuthan P, Akanni W, Allen J, Amode MR, Armean IM, et al. *Ensembl* 2019. *Nucleic Acids Res* 2019;47:D745–51.

65. Sayers EW, Agarwala R, Bolton EE, Brister JR, Canese K, Clark K, et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res* 2019;47:D23–8.
66. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: Architecture and applications. *BMC Bioinformatics* 2009;10:421.
67. McKinney W. Data Structures for Statistical Computing in Python. In: *Proceedings of the 9th Python in Science Conference* 2010:56–61.
68. Van Rossum G, Drake FL. *Python 3 Reference Manual*; CreateSpace. Scotts Valley, CA. 2009:242.
69. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2020.
70. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag; 2016.
71. Wilke CO. *cowplot: Streamlined Plot Theme and Plot Annotations for “ggplot2.”* CRAN R Project; 2020.
72. Borgmann D, Lierop D van, Nick, Caldwell I, Marshall C, Mastrukov DG, et al. *Inkscape*. 2021.
73. Polašek O, Kolčić I, Kirin M, Gelemanović A, Miljković A, Relja A, et al. 10.001 Dalmatinac - Hrvatska biobanka [Internet]. 2022. Preuzeto s: <http://www.mefst.unist.hr/znanost/istrazivacke-skupine-i-laboratoriji/10-001-dalmatinac-hrvatska-biobanka/5035>: (pristupljeno: 10 siječnja 2022)
74. Pulanić D, Polašek O, Petrovečki M, Vorko-Jović A, Peričić M, Lauc LB, et al. Effects of isolation and inbreeding on human quantitative traits: An example of biochemical markers of hemostasis and inflammation. *Hum Biol* 2008;80:513–33.
75. Al-Qaisi M, Nott DM, King DH, Kaddoura S. Ankle Brachial Pressure Index (ABPI): An update for practitioners. *Vasc Health Risk Manag* 2009;5:833–41.
76. Khalil SF, Mohktar MS, Ibrahim F. The theory and fundamentals of bioimpedance analysis in clinical status monitoring and diagnosis of diseases. *Sensors (Switzerland)* 2014;14:10895–928.
77. Deurenberg P, Van Der Kooy K, Leenen R, Weststrate JA, Seidell JC. Sex and age specific prediction formulas for estimating body composition from bioelectrical impedance: A cross-validation study. *Int J Obes* 1991;15:17–25.
78. Trbojević-Akmačić I, Vilaj M, Lauc G. High-throughput analysis of immunoglobulin G glycosylation. *Expert Rev Proteomics* 2016;13:523–34.

79. Pucić M, Knezević A, Vidic J, Adamczyk B, Novokmet M, Polasek O, et al. High throughput isolation and glycosylation analysis of IgG-variability and heritability of the IgG glycome in three isolated human populations. *Mol Cell Proteomics* 2011;10:M111.010090.
80. Raven J. Raven's progressive matrices and vocabulary scales: CPM/CVS complete kit. Vol. 759. Oxford psychologists Press Oxford; 2004.
81. Borkowski JG, Benton AL, Spreen O. Word fluency and brain damage. *Neuropsychologia* 1967;5:135–40.
82. Joy S, Kaplan E, Fein D. Speed and memory in the WAIS-III Digit Symbol - Coding subtest across the adult lifespan. *Arch Clin Neuropsychol* 2004;19:759–67.
83. Moradi E, Hallikainen I, Hänninen T, Tohka J. Rey's Auditory Verbal Learning Test scores can be predicted from whole brain MRI in Alzheimer's disease. *NeuroImage Clin* 2017;13:415–27.
84. Raven J. The Raven's Progressive Matrices: Change and Stability over Culture and Time. *Cogn Psychol* 2000;41:1–48.
85. Mair A, Grant L, Wilson A, Thomas A, Gardiner C, Hind C, et al. Working out Cigarette Equivalents [Internet]. Scottish Community Pharmacy Smoking Cessation Programme. 2021. Preuzeto s: <http://www.smoking2.nes.scot.nhs.uk/module4/working-out-cigarette-equivalents.html>: (pristupljeno: 20 siječnja 2022)
86. BANKS MH, CLEGG CW, JACKSON PR, KEMP NJ, STAFFORD EM, WALL TD. The use of the General Health Questionnaire as an indicator of mental health in occupational studies. *J Occup Psychol* 1980;53:187–94.
87. Roenneberg T, Wirz-Justice A, Meroo M. Life between clocks: Daily temporal patterns of human chronotypes. *J Biol Rhythms* 2003;18:80–90.
88. Roenneberg T, Keller LK, Fischer D, Matera JL, Vetter C, Winnebeck EC. Human activity and rest in situ. *Methods Enzymol* 2015;552:257–83.
89. Roenneberg T, Allebrandt K V., Meroo M, Vetter C. Social jetlag and obesity. *Curr Biol* 2012;22:939–43.
90. Roenneberg T, Allebrandt K V., Meroo M, Vetter C. Erratum: Social jetlag and obesity (Current Biology (2012) 22 (939-943)). *Curr Biol* 2013;23:737.
91. Johns MW. A New Method for Measuring Daytime Sleepiness: The Epworth Sleepiness Scale. *Sleep* 1991;14:540–5.
92. Chung F, Yegneswaran B, Liao P, Chung SA, Vairavanathan S, Islam S, et al. Validation of the Berlin Questionnaire and American Society of Anesthesiologists Checklist as

- Screening Tools for Obstructive Sleep Apnea in Surgical Patients. *Anesthesiology* 2008;108:822–30.
93. Senaratna C V., Perret JL, Matheson MC, Lodge CJ, Lowe AJ, Cassim R, et al. Validity of the Berlin questionnaire in detecting obstructive sleep apnea: A systematic review and meta-analysis. *Sleep Med Rev* 2017;36:116–24.
 94. De Marco R, Zanolin ME, Accordini S, Signorelli D, Marinoni A, Bugiani M, et al. A new questionnaire for the repeat of the first stage of the European Community Respiratory Health Survey: A pilot study. *Eur Respir J* 1999;14:1044–8.
 95. Grassi M, Rezzani C, Biino G, Marinoni A. Asthma-like symptoms assessment through ECRHS screening questionnaire scoring. *J Clin Epidemiol* 2003;56:238–47.
 96. ROSE GA. The diagnosis of ischaemic heart pain and intermittent claudication in field surveys. *Bull World Health Organ* 1962;27:645–58.
 97. Biloglav Z, Ivanković D, Campbell H, Rudan I. Performance of WHO angina questionnaire in measuring burden of coronary heart disease in human isolate populations. *Coll Antropol* 2004;28:205–13.
 98. Rose G, McCartney P, Reid DD. Self administration of a questionnaire on chest pain and intermittent claudication. *Br J Prev Soc Med* 1977;31:42–8.
 99. Benjamin DJ, Cesarini D, van der Loos MJHM, Dawes CT, Koellinger PD, Magnusson PKE, et al. The genetic architecture of economic and political preferences. *Proc Natl Acad Sci* 2012;109:8026–31.
 100. Eysenck SBG, Eysenck HJ, Barrett P. A revised version of the psychoticism scale. *Pers Individ Dif* 1985;6:21–9.
 101. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012;28:882–3.
 102. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw* 2015;67:1–48.
 103. Solomon SR, Sawilowsky SS. Impact of rank-based normalizing transformations on the accuracy of test scores. *J Mod Appl Stat Methods* 2009;8:448–62.
 104. Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: An R library for genome-wide association analysis. *Bioinformatics* 2007;23:1294–6.
 105. Revelle MW. psych: Procedures for personality and psychological research (R package). CRAN R Project; 2021.
 106. Therneau TM. coxme: Mixed Effects Cox Models. R-package description. Vienna,

- Austria: CRAN R Project; 2015:1–14.
107. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B* 1995;57:289–300.
 108. Zielinska DF, Gnad F, Wiśniewski JR, Mann M. Precision Mapping of an In Vivo N-Glycoproteome Reveals Rigid Topological and Sequence Constraints. *Cell* 2010;141:897–907.
 109. Zielinska DF, Gnad F, Schropp K, Wiśniewski JR, Mann M. Mapping N-Glycosylation Sites across Seven Evolutionarily Distant Species Reveals a Divergent Substrate Proteome Despite a Common Core Machinery. *Mol Cell* 2012;46:542–8.
 110. Park C, Zhang J. Genome-Wide Evolutionary Conservation of N-Glycosylation Sites. *Mol Biol Evol* 2011;28:2351–7.
 111. Lindahl U, Couchman J, Kimata K, Esko JD. Proteoglycans and Sulfated Glycosaminoglycans. In: Varki A, Cummings RD, Esko JD, Stanle P, Hart GW, Aebi M, et al., editors. *Essentials of Glycobiology* New York, NY: Cold Spring Harbor Laboratory Press; 2017.
 112. Zeng Y, Ebong EE, Fu BM, Tarbell JM. The structural stability of the endothelial glycocalyx after enzymatic removal of glycosaminoglycans. *PLoS One* 2012;7:1–14.
 113. Lombard J. Early evolution of polyisoprenol biosynthesis and the origin of cell walls. *PeerJ* 2016;4:e2626.
 114. Kleizen B, Braakman I. Protein folding and quality control in the endoplasmic reticulum. *Curr Opin Cell Biol* 2004;16:343–9.
 115. Saier MH, Bogdanov M V. Membranous Organelles in Bacteria. *J Mol Microbiol Biotechnol* 2013;23:5–12.
 116. Grant CR, Wan J, Komeili A. Organelle Formation in Bacteria and Archaea. *Annu Rev Cell Dev Biol* 2018;34:217–38.
 117. Saier M. Membrane-Bound Compartments in Bacteria. *Microbe Mag* 2014;9:368–72.
 118. Saier MH. Microcompartments and Protein Machines in Prokaryotes. *J Mol Microbiol Biotechnol* 2013;23:243–69.
 119. Baum DA, Baum B. An inside-out origin for the eukaryotic cell. *BMC Biol* 2014;12:76.
 120. López-García P, Moreira D. The Syntrophy hypothesis for the origin of eukaryotes revisited. *Nat Microbiol* 2020;5:655–67.
 121. López-García P, Moreira D. Selective forces for the origin of the eukaryotic nucleus. *BioEssays* 2006;28:525–33.
 122. Tomono T, Kojima H, Fukuchi S, Tohsato Y, Ito M. Investigation of glycan evolution

- based on a comprehensive analysis of glycosyltransferases using phylogenetic profiling. *Biophys Physicobiology* 2015;12:57–68.
123. Greto VL, Cvetko A, Štambuk T, Dempster NJ, Kifer D, Deriš H, et al. Extensive weight loss reduces glycan age by altering IgG N-glycosylation. *Int J Obes* 2021;45:1521–31.
 124. Peng J, Vongpatanasin W, Sacharidou A, Kifer D, Yuhanna IS, Banerjee S, et al. Supplementation with the Sialic Acid Precursor N-Acetyl-D-Mannosamine Breaks the Link between Obesity and Hypertension. *Circulation* 2019;140:2005–18.
 125. Verdi S, Abbasian G, Bowyer RCE, Lachance G, Yarand D, Christofidou P, et al. TwinsUK: The UK Adult Twin Registry Update. *Twin Res Hum Genet* 2019;22:523–9.
 126. Holle R, Happich M, Löwel H, Wichmann HE. KORA - A research platform for population based health research. *Gesundheitswesen* 2005;67:19–25.
 127. Kifer D, Louca P, Cvetko A, Deriš H, Cindrić A, Grallert H, et al. N-glycosylation of immunoglobulin G predicts incident hypertension. *J Hypertens* 2021;39:2527–33.
 128. Lemmers RFH, Vilaj M, Urda D, Agakov F, Šimurina M, Klaric L, et al. IgG glycan patterns are associated with type 2 diabetes in independent European populations. *Biochim Biophys Acta - Gen Subj* 2017;1861:2240–9.
 129. Štambuk T, Kifer D, Duvnjak L, Vučić Lovrenčić M, Gornik O. Associations between plasma protein, IgG and IgA N-glycosylation and metabolic health markers in pregnancy and gestational diabetes. 2021;1–22.
 130. Liu D, Chu X, Wang H, Dong J, Ge SQ, Zhao ZY, et al. The changes of immunoglobulin G N-glycosylation in blood lipids and dyslipidaemia. *J Transl Med* 2018;16:235.
 131. Hou H, Xu X, Sun F, Zhang X, Dong H, Wang L, et al. Hyperuricemia is Associated with Immunoglobulin G N-Glycosylation: A Community-Based Study of Glycan Biomarkers. *Omi A J Integr Biol* 2019;23:660–7.
 132. Li X, Wang H, Russell A, Cao W, Wang X, Ge S, et al. Type 2 Diabetes Mellitus is Associated with the Immunoglobulin G N-Glycome through Putative Proinflammatory Mechanisms in an Australian Population. *Omi A J Integr Biol* 2019;23:631–9.
 133. Deriš H, Kifer D, Cindrić A, Petrović T, Cvetko A, Trbojević-Akmačić I, et al. Immunoglobulin G glycome composition in transition from premenopause to postmenopause. *iScience* 2022;103897.
 134. Tijardović M, Marijančević D, Bok D, Kifer D, Lauc G, Gornik O, et al. Intense Physical Exercise Induces an Anti-inflammatory Change in IgG N-Glycosylation Profile. *Front Physiol* 2019;10:1–10.

6 ŽIVOTOPIS AUTORA S POPISOM OBJAVLJENIH RADOVA

Domagoj Kifer je rođen 1991. godine u Vukovaru. Osnovnu školu je završio u Zagrebu nakon koje upisuje zagrebačku XV. gimnaziju koju završava 2010. godine. Iste godine upisuje studij Farmacije na Farmaceutsko-biokemijskom fakultetu Sveučilišta u Zagreba. Diplomirao je 2015. godine nakon čega se zapošljava kao asistent na projektu IntegraLife na Zavodu za biokemiju i molekularnu biologiju na Farmaceutsko-biokemijskom fakultetu Sveučilišta u Zagrebu gdje radi sve do kraja listopada 2015. godine. Od studenog 2015. godine radi kao zamjena na radnom mjestu asistenta na Zavodu za biokemiju i molekularnu biologiju na Farmaceutsko-biokemijskom fakultetu Sveučilišta u Zagrebu sve do kraja kolovoza 2016. godine. Nakon toga od rujna 2016. godine se zapošljava kao asistent na Zavodu za biofiziku na Farmaceutsko-biokemijskom fakultetu Sveučilišta u Zagrebu na kojem radi sve do trenutka pisanja ove disertacije.

Do sada je koautor na 26 znanstvenih radova i 42 priopćenja na domaćim i međunarodnim skupovima.

Popis znanstvenih radova u časopisima zastupljenim u bazi *Web of Science Core Collection*:

- Cvetko A, Mangino M, Tijardović M, **Kifer D**, Falchi M, Keser T, Perola M, Spector TD, Lauc G, Menni C, Gornik O. 2021. Plasma N-glycome shows continuous deterioration as the diagnosis of insulin resistance approaches. *BMJ Open Diabetes Research and Care*. 9(1):e002263.
- Cvetko A, **Kifer D**, Gornik O, Klarić L, Visser E, Lauc G, Wilson JF, Štambuk T. 2020. Glycosylation Alterations in Multiple Sclerosis Show Increased Proinflammatory Potential. *Biomedicines* 8:410.
- Dabelić S, **Kifer D**, Jakšić D, Kopjar N, Klarić MŠ. 2021. Sterigmatocystin, 5-Methoxysterigmatocystin, and Their Combinations are Cytotoxic and Genotoxic to A549 and HepG2 Cells and Provoke Phosphorylation of Chk2, but not FANCD2 Checkpoint Proteins. *Toxins* 13:464.
- Deriš H, **Kifer D**, Cindrić A, Petrović T, Cvetko A, Trbojević-Akmačić I, et al. Immunoglobulin G glycome composition in transition from premenopause to postmenopause. *iScience* 2022;103897.
- Dumić J, Cvetko A, Abramović I, Šupraha Goreta S, Perović A, Njire Bratičević M, **Kifer D**, et al. 2022. Changes in Specific Biomarkers Indicate Cardiac Adaptive and Anti-inflammatory Response of Repeated Recreational SCUBA Diving. *Front Cardiovasc Med* 9:1–20.

- Futo M, Opašić L, Koska S, Čorak N, Široki T, Ravikumar V, Thorsell A, Lenuzzi M, **Kifer D**, Domazet-Lošo M, et al. 2021. Embryo-Like Features in Developing *Bacillus subtilis* Biofilms. Perna N, editor. *Molecular Biology and Evolution* 38:31–47.
- Greto VL, Cvetko A, Štambuk T, Dempster NJ, **Kifer D**, Deriš H, Cindrić A, Vučković F, Falchi M, Gillies RS, et al. 2021. Extensive weight loss reduces glycan age by altering IgG N-glycosylation. *International Journal of Obesity* 45:1521–1531.
- Hlapčić I, Belamarić D, Bosnar M, **Kifer D**, Vukić Dugac A, Rumora L. 2020. Combination of Systemic Inflammatory Biomarkers in Assessment of Chronic Obstructive Pulmonary Disease: Diagnostic Performance and Identification of Networks and Clusters. *Diagnostics* 10:1029.
- Jakšić D, Čurtović I, **Kifer D**, Rašić D, Kopjar N, Micek V, Peraica M, Klarić MŠ. 2020. Single-Dose Toxicity of Individual and Combined Sterigmatocystin and 5-Methoxysterigmatocystin in Rat Lungs. *Toxins* 12:734.
- Jakšić D, Sertić M, **Kifer D**, Kocsubè S, Mornar Turk A, Nigović B, Šarkanj B, Krska R, Sulyok M, Šegvić Klarić M. 2021. Fungi and their secondary metabolites in water-damaged indoors after a major flood event in eastern Croatia. *Indoor Air* 31:730–744.
- Jakšić D, Sertić M, Kocsubé S, Kovačević I, **Kifer D**, Mornar A, Nigović B, Šegvić Klarić M. 2020. Post-Flood Impacts on Occurrence and Distribution of Mycotoxin-Producing *Aspergilli* from the Sections *Circumdati*, *Flavi*, and *Nigri* in Indoor Environment. *Journal of Fungi* 6:282.
- Jurić J, Kohrt WM, **Kifer D**, Gavin KM, Pezer M, Nigrovic PA, Lauc G, Juric J, Kohrt WM, et al. 2020. Effects of estradiol on biological age measured using the glycan age index. *Aging* 12:19756–19765.
- **Kifer D**, Bugada D, Villar-Garcia J, Gudelj I, Menni C, Sudre C, Vučković F, Ugrina I, Lorini LF, Posso M, et al. 2021. Effects of Environmental Factors on Severity and Mortality of COVID-19. *Frontiers in Medicine* 7.
- **Kifer D**, Jakšić D, Šegvić Klarić M. 2020. Assessing the Effect of Mycotoxin Combinations: Which Mathematical Model Is (the Most) Appropriate? *Toxins* 12:153.
- **Kifer D**, Louca P, Cvetko A, Deriš H, Cindrić A, Grallert H, Peters A, Polašek O, Gornik O, Mangino M, et al. 2021. N-glycosylation of immunoglobulin G predicts incident hypertension. *Journal of Hypertension* 39:2527–2533.
- **Kifer D**, Mužinić V, Klarić MŠ. 2016. Antimicrobial potency of single and combined mupirocin and monoterpenes, thymol, menthol and 1,8-cineole against *Staphylococcus aureus* planktonic and biofilm growth. *The Journal of Antibiotics* .

- **Kifer D**, Sulyok M, Jakšić D, Krska R, Šegvić Klarić M. 2021. Fungi and their metabolites in grain from individual households in Croatia. *Food Additives & Contaminants: Part B* 14:98–109.
- Kifer N, Sestan M, Frkovic M, **Kifer D**, Kozmar A, Padjen I, et al. 2022. 2019 ACR/EULAR classification criteria and therapy in predicting organ damage accrual in patients with childhood-onset systemic lupus erythematosus: A retrospective study over the last 29 years. *Lupus*; 096120332210947.
- Kuruc Poje D, **Kifer D**, Huys I, Miranda J, Jenzer H, Miljković N, Hoppe-Tichy T, Bochniarz M, Frontini R, Schwartz DG, et al. 2021. Patients perspectives on drug shortages in six European hospital settings – a cross sectional study. *BMC Health Services Research* 21:1–10.
- Kuruc Poje D, Mađarić V, Janeš Poje V, **Kifer D**, Howard P, Marušić S. 2019. Antimicrobial stewardship effectiveness on rationalizing the use of last line of antibiotics in a short period with limited human resources: a single centre cohort study. *BMC Research Notes* 12:531.
- Mijakovac A, Jurić J, Kohrt WM, Krištić J, **Kifer D**, Gavin KM, Miškec K, Frkatović A, Vučković F, Pezer M, et al. 2021. Effects of Estradiol on Immunoglobulin G Glycosylation: Mapping of the Downstream Signaling Mechanism. *Frontiers in Immunology* 12:1–13.
- Pavić T, Dilber D, **Kifer D**, Selak N, Keser T, Ljubičić Đ, Vukić Dugac A, Lauc G, Rumora L, Gornik O. 2018. N-glycosylation patterns of plasma proteins and immunoglobulin G in chronic obstructive pulmonary disease. *Journal of Translational Medicine* 16:323.
- Peng J, Vongpatanasin W, Sacharidou A, **Kifer D**, Yuhanna IS, Banerjee S, Tanigaki K, Polasek O, Chu H, Sundgren NC, et al. 2019. Supplementation With the Sialic Acid Precursor N-Acetyl-D-Mannosamine Breaks the Link Between Obesity and Hypertension. *Circulation* 140:2005–2018
- Rašić D, Jakšić D, Hulina A, **Kifer D**, Kopjar N, Rumora L, Davor Ž, Peraica M, Šegvić Klarić M. 2020. Sterigmatocystin moderately induces oxidative stress in male Wistar rats after short-term oral treatment. *Mycotoxin Research* 36:181–191.
- Rašić D, Želježić D, Kopjar N, **Kifer D**, Klarić MŠ, Peraica M. 2019. DNA damage in rat kidneys and liver upon subchronic exposure to single and combined ochratoxin A and citrinin. *World Mycotoxin Journal* 12:163–172.
- Tijardović M, Marijančević D, Bok D, **Kifer D**, Lauc G, Gornik O, Keser T. 2019. Intense Physical Exercise Induces an Anti-inflammatory Change in IgG N-Glycosylation Profile. *Frontiers in Physiology* 10:1–10.

Temeljna dokumentacijska kartica

Sveučilište u Zagrebu
Farmaceutsko-biokemijski fakultet
Zavod za biofiziku
A. Kovačića 1, 10000 Zagreb, Hrvatska

Doktorski rad

Makroevolucijska povijest glikozilacije i biostatistička analiza utjecaja okolišnih čimbenika na *N*-glikozilaciju imunoglobulina G

Domagoj Kifer

SAŽETAK

Glikozilacija je prisutna u svim živućim organizmima – od bakterija i arheja pa sve do čovjeka. Struktura glikana nije izravno zapisana u genomu nego je rezultat precizno kontrolirane aktivnosti proteina zaduženih za biosintezu glikana. Strukturne razlike glikana između vrsta su odgovor na selektivni pritisak, a svaka promjena koja je postala konzervirana u evoluciji nosi važnu ulogu. S druge strane, uočena je varijabilnost glikana unutar organizama jedne vrste što je dijelom rezultat okolišnih čimbenika i stanja pojedine jedinke. Cilj ovog rada je odrediti makroevolucijske trendove glikozilacije i analizirati utjecaj okolišnih čimbenika na *N*-glikom imunoglobulina G (IgG). Genomskom je filostratigrafijom određena starost gena koji kodiraju za glikoproteine i gena koji kodiraju za proteine koji sudjeluju u glikozilaciji. Dobiveni rezultati su pokazali da proteini koji sudjeluju u glikozilaciji vuku korijene još iz zajedničkog pretka svih staničnih organizama, uz pojedine inovacije u zajedničkom pretku svih eukariota, te manje u zajedničkom pretku pravih životinja. Unatoč starosti, glikozilacija u svoje zlatno doba ulazi razvojem višestaničnih životinja s ciljem razvoja međustanične komunikacije i interakcije stanica s okruženjem kada u većoj mjeri nastaju novi geni koji kodiraju za polipeptidne lance glikoproteina. S ciljem procjene utjecaja raznih čimbenika i stanja analizirana je povezanost podataka dostupnih u biobazi „10001 Dalmatinac“ s različitim deriviranim svojstvima *N*-glikoma IgG-a kao najproučavanijeg glikoproteina. Rezultati dobiveni analizom preko 3500 ljudi su pokazali da postoji utjecaj prehrane (učestalost konzumacije voćnih derivata i slatke hrane), hipertenzije, artritisa, regularnosti menstrualnog ciklusa, debljine, krvnog tlaka, koncentracije gliciranog hemoglobina, fibrinogena i inzulina na sintezu agalaktoziliranih glikana IgG-a. Zatim, digalaktozilacija IgG-a je povezana s hipertenzijom, artritismom, redovitošću menstrualnog ciklusa, debljinom, krvnim tlakom, mineralnom gustoćom kostiju, prehranom (voćni derivati, slatka hrana), koncentracijama urata, triglicerida, fibrinogena, inzulina, kolesterola i HDL kolesterola. Sijalinizacija IgG-a je povezana s regularnošću menstrualnog ciklusa, debljinom, mineralnom gustoćom kostiju, krvnim tlakom, koncentracijama triglicerida i inzulina. Udio struktura s račvajućim *N*-acetilglukozamin u *N*-glikomu IgG-a ovisi o mineralnoj gustoći kostiju, redovitošću menstrualnog ciklusa, pušenju, te koncentracijama triglicerida i kolesterola.

Rad je pohranjen u Središnjoj knjižnici Sveučilišta u Zagrebu Farmaceutsko-biokemijskog fakulteta.

Rad sadrži: 102 stranica, 40 grafičkih prikaza, 16 tablica i 134 literaturnih navoda. Izvornik je na hrvatskom jeziku.

Ključne riječi: *N*-glikom, genomski filostratigrafija, evolucija glikozilacije, evolucija glikoproteina, utjecaj okoliša

Mentori: **Dr. sc. Gordan Lauc**, redoviti profesor Sveučilišta u Zagrebu Farmaceutsko-biokemijskog fakulteta
Dr. sc. Erim Bešić, docent Sveučilišta u Zagrebu Farmaceutsko-biokemijskog fakulteta

Ocjenjivači: **Dr. sc. Olga Gornik Kljaić**, redovita profesorica Sveučilišta u Zagrebu Farmaceutsko-biokemijskog fakulteta
Dr. sc. Tomislav Domazet-Lošo, izvanredni profesor, Instituta Ruđer Bošković
Dr. sc. Kristian Vlahoviček, redoviti profesor Sveučilišta u Zagrebu Prirodoslovno-matematičkog fakulteta

Rad prihvaćen: 20. travanja 2022.

Basic documentation card

University of Zagreb
Faculty of Pharmacy and Biochemistry
Department of Biophysics
A. Kovačića 1, 10000 Zagreb, Croatia

Doctoral dissertation

Macroevolutionary History of Glycosylation and Biostatistical Analysis of Environmental Effects on N Glycosylation of Immunoglobulin G

Domagoj Kifer

SUMMARY

Glycosylation is present in all living organisms - from bacteria and archaea to humans. The structure of glycans is not directly written in the genome but is the result of the precisely controlled activity of proteins responsible for glycan biosynthesis. The structural differences of glycans between species are a response to selective pressure, and any change that has become conserved in evolution plays an important role. On the other hand, variability of glycans within the organism of one species was observed, which is partly the result of environmental factors and the state of individual individuals.

This thesis aims to determine the macroevolutionary trends of glycosylation and to analyse the influence of environmental factors on the *N*-glycome of immunoglobulin G.

Genomic phylostratigraphy determined the evolutionary age of genes encoding glycoproteins and genes encoding proteins involved in glycosylation eukaryotes. Obtained results pointed out that the genes involved in glycosylation which we find today in humans, were present in all cellular organisms, with some innovations in the last common ancestor of all eukaryotes and less in the common ancestor of eumetazoans. Despite its ancient evolutionary age, glycosylation in its golden era enters with the development of multicellular animals when new genes encode polypeptide glycoprotein chains are increasingly emerging, probably intending to develop tools for cell-cell and cell-environment interaction.

To assess the influence of various factors and conditions, the correlation of data available in the 10001 Dalmatian database with different derived properties of *N*-glycome immunoglobulin G (IgG), as the most studied glycoprotein, was analysed. The results obtained from the analysis of over 3500 people showed that there is an impact of diet (frequency of consumption of fruit derivatives and sugary foods), hypertension, arthritis, menstrual regularity, obesity, blood pressure and glycated haemoglobin, fibrinogen, and insulin concentration on IgG agalactosylation. Then, IgG digalactosylation is associated with hypertension, arthritis, menstrual regularity, obesity, blood pressure, bone mineral density, diet (fruit derivatives, sugary foods), urate, triglyceride, fibrinogen, insulin, cholesterol, and HDL cholesterol levels. Sialylation of IgG is associated with menstrual cycle regularity, obesity, bone mineral density, blood pressure, triglyceride, and insulin concentrations. The abundance of structures with branching *N*-acetylglucosamine in *N*-glycome IgG depends on bone mineral density, regularity of the menstrual cycle, smoking, and the concentration of triglycerides and cholesterol.

This thesis is deposited in the Central Library of the University of Zagreb Faculty of Pharmacy and Biochemistry.

Thesis includes: 102 pages, 40 figures, 16 tables and 134 references. Original is in Croatian language.

Keywords: N-glycome, genomic phylostratigraphy, glycosylation evolution, glycoprotein evolution, environmental effect

Mentors: **Gordan Lauc, PhD**, *Full Professor, University of Zagreb Faculty of Pharmacy and Biochemistry*
Erim Bešić, PhD, *Assistant Professor, University of Zagreb Faculty of Pharmacy and Biochemistry*

Reviewers: **Olga Gornik Kljaić, PhD**, *Full Professor, University of Zagreb Faculty of Pharmacy and Biochemistry*
Tomislav Domazet-Lošo, PhD, *Associate Professor, Ruđer Bošković Institute*
Kristian Vlahoviček, PhD, *Full Professor, University of Zagreb Faculty of Science*

The thesis was accepted: 20 April 2022