Open Peer Review on Qeios

Positive Selection and Adaptation Role of Gut Microbiota in the Evolution of Adaptive Immunity of Mammalian Species

Hafiz Ishfaq Ahmad[1]

1 Islamia University of Bahawalpur

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.

Abstract

Every mammalian species harbours a gut microbiota, and variation in the gut microbiota within mammalian species can have profound effects on host phenotypes. Understanding the consequences of gut microbiotas in mammalian evolution first requires testable hypotheses regarding the specific modes by which they alter the adaptive landscapes experienced by hosts. Mechanisms underlying adaptation to various gut microbiota during the evolutionary process remain poorly understood. This study examines how the immune system of the host influences the molecular evolution and adaptation of the gut microbiota in a variety of mammalian species. We assessed the evidence for the gut microbiota's influence on mammalian evolution and diversification. The maximum likelihood approach was used to identify evidence of positive selection in immune genes. To identify codons that underwent adaptive evolution, we looked for episodic and pervasive positive selection throughout all branches of the mammalian evolutionary tree. Our findings reveal intriguing co-evolutionary processes in which the host's immune system exerts selective pressure on immune genes, resulting in adaptive changes in microbial populations. Our findings suggest that, in the majority of mammalian species, episodic positive selection has played an important role in the genetic development of species-specific gene sequences and divergence. Furthermore, we found evidence of broad positive selection during the molecular evolution of immune genes on all branches of the mammalian phylogenetic tree. These results suggest that the gut microbiota plays a crucial role in influencing the way mammals adapt to their diet, their ability to change their physical characteristics, the structure of their gastrointestinal system, and their immune response.

Keywords: Molecular evolution; gut microbiota; humoral immunity; adaptation; selection.

Introduction

The gastrointestinal tract of mammals is a dynamic ecosystem with a high population of billions of bacteria known as the gut microbiota. Animal gut microbiota is a complex symbiotic ecosystem that undergoes continuous fluctuations. Environmental factors, including energy sources, and changes in the niche induced by microbial colonizers, influence

dynamic changes (Savage, 1977; Xu et al., 2007). Microorganism growth is supported by carbon sources in the host's food and via the shedding of epithelial cells. The quantity of viable microbial biomass is restricted by intestinal secretions and peristalsis (Sonnenburg et al., 2005). Understanding the boundaries of microbiota stability is essential for appropriately modeling biological systems and diseases in live organisms. Inbreeding can create animals with identical genes in the host, but the microbiota, the microorganism community in the host, varies based on factors like the supplier, housing facility, and specific cages used for experiments. Differences in the microbiota seen in various colonies of inbred or targeted mouse models can explain the differences in observed phenotypic outcomes across different research facilities (Mamantopoulos et al., 2017). In other cases, the dominance of potent traits from one hazardous microbe could surpass any differences in the microbiota's makeup. Yet, in other cases, variations in strains within a single organism might affect the interaction between the host and microbes in a mutually advantageous manner. Establishing a genetically homogeneous colony of mice can be achieved by standardizing the microbiota in their intestines, which has been comprehensively sequenced and encompasses a wide variety of microbial species. This aims to provide uniform and reproducible research projects across different time periods and research institutes. Existing studies on the microbiota have mostly been carried out for brief durations and have predominantly concentrated on certain species (Barroso-Batista et al., 2014; Ramiro et al., 2020; Sousa et al., 2017). The humoral immune response, comprising antibodies, cytokines, and other soluble proteins, is an essential component of the host immune system that interacts with the gut flora. It is crucial in protecting against infections. The reciprocal relationship between the host and the gut microbiota has garnered increasing attention in the scientific community as a co-evolutionary link. Genetic compositions of host and microbial communities have been shaped by internal and external evolutionary pressures over millions of years. This has resulted in a delicate equilibrium that improves the overall health and flexibility of the host organism. A thorough understanding of how the molecular evolution of the gut microbiota intersects with the selection pressures impacting the host immune system, particularly in respect to humoral immunity, is lacking despite distinct investigations on these topics. This topic continues to be intriguing and demanding for researchers. Humoral immunity is a crucial element of the adaptive immune system. It produces antibodies and orchestrates immunological reactions against various illnesses. The immune system faces continuous challenges, leading to an ongoing interaction between the host and its gut flora. The host's immune system influences the microbial populations in the gastrointestinal tract through selection forces. The gut microbiota influences the host's immune system through many mechanisms, contributing to the regulation of immunological equilibrium and tolerance. Advancements in high-throughput sequencing technology and bioinformatics tools have greatly enhanced our ability to analyze the intricate molecular processes involved in host-microbiota interactions. The utilization of these approaches provides extraordinary opportunities to examine the genetic traits of both hosts and the microorganisms that exist within them. This enables a more in-depth research of the mutually important dynamics that have altered the ecology of the mammalian gut. These methods allow for a thorough examination of how the gut microbiota's genetic development is affected by the host's humoral immune system. The objective of this study is to investigate a number of fundamental problems that pertain to the coevolution of the host and the microbiota. What molecular changes do gut microbiota undergo in response to the host's humoral immune responses? How do these adaptations differ among various mammalian species? Do preservative mechanisms or distinctive characteristics exist that establish the co-evolutionary connection between the host's humoral immunity and the gut microbiota? In order to

investigate these inquiries, we will conduct an extensive examination of the genetic variation present in the gut microbiota of several mammalian species. Our goal is to use advanced bioinformatics approaches to find genetic patterns that show positive selection. This will help us understand how the host's immune system affects the gut microbiota in terms of evolution. Furthermore, we will explore the practical consequences of these adaptations, aiming to decipher the fundamental mechanisms that propel the observed molecular evolution. The gut microbiota has been linked to a wide range of health issues, including metabolic disorders and autoimmune diseases. Comprehending the changes that occur in the microbiota due to the host's immune system can lead to the development of novel treatment approaches that try to manipulate the gut microbiota in order to improve the body's antibody-based immune responses. This research aimed to investigate the rapid evolution of GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1 genes and offer an explanation for the significant sequence divergence found in different animal species. We showed that these genes underwent rapid evolution due to positive selection. We examined that genes evolved rapidly due to positive selection.

Materials and Methods

Data collection

The sequences of the coding sections of the GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1 genes utilized in the analysis were obtained from NCBI. We used gene sequences from the genomes of representatives of various mammalian species (Supplementary Table S1). The whole-genome yak sequences were retrieved from the Ensembl database. The amino acid and nucleotide sequences of GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1, which are important for gut microbiota adaptation, are expressed in mammalian species. Ensembl was accessed to get the coding sequences for these genes, all of which cover various mammalian species. This was done based on the gene annotation (two conserved neighboring genes) [1]. The BLASTn v2.2.29+ algorithm was used to choose the most effective scaffold [2]. Checking for a start and stop codon was part of the annotation process that was carried out with MITOS [3]. MACSE v1.01b [4] and ClustalW v2 aligned the protein-coding and ribosomal genes[5]. We eliminated any genes that were less than a third of the length of the overall locus alignment.

Interspecific Sequences alignments

The nucleotide sequences of the GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1 genes were aligned separately using the ClustalW tool within the MEGA software v.7.0.14, using the default parameters. The software Gblocks v.0.91b (Castresana, 2000; Talavera and Castresana, 2007) was employed with standard settings to eliminate inaccurately aligned regions and differing segments.

Phylogenetic analysis

PartitionFinder v.1.1.1 (Lanfear et al., 2016) was utilized to determine the optimal partitioning scheme and substitution models for each partition prior to conducting phylogenetic analysis. This was based on the Akaike (AIC), corrected Akaike

(AICc), and Bayesian (BIC) information criteria. We found that the GTR + 0 + I model is the most suitable for molecular evolution. RAxML version 8.2.7 (Stamatakis, 2014) was utilized to generate the maximum probability unrooted tree with 10,000 bootstrap replicates. It was unable to designate an outgroup for tree construction since orthologs of the GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1 genes were not found in the genomes of other organisms. Phylogenetic trees were constructed using gene sequence data from the mammalian GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1 genes to show the evolutionary connections and alterations in these genes over time. Phylogenetic trees were created using MEGA (Molecular Evolutionary Genetics Analysis) version 10.0.5 [6] using a maximum likelihood method. The topology of the tree we built with the neighbor-joining method was evaluated by applying the maximum likelihood method to the Whelan and Goldman (WAG) substitution model {m/7/}. To further evaluate the stability of the tree structure, we conducted 1000 bootstrap repetitions. Gene trees and other phylogenetic trees can be compared and evaluated with precision using the TreeBeST-generated species tree as a benchmark (http://treesoft.svn.sourceforge.net/viewrc/treesoft/).

Codon-based positive selection analyses

The GARD program was utilized for the analysis of potential recombination events, as described by Kosakovsky Pond et al. in 2006. Statistical tests were conducted using the CODEML algorithm in the PAML software package v.4.8 to assess adaptive evolution in the GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1 genes due to positive selection. Site models (M8, M8a) and branch-site models (M1a, A, A1) were employed to identify the likelihood of positive selection occurring at particular sites throughout all branches of the phylogenetic tree or at specified branches (referred to as foreground branches) (Yang and Nielsen, 2002). Opposing models were compared using likelihood ratio tests (LRTs) to identify the best-fitting model among M8 vs. M8a, M1a vs. A, and A1 vs. A. The degrees of freedom (df) were calculated by subtracting the number of free parameters in the models being compared. Positive selection was shown by the identification of codons with a dN/dS ratio (ω) exceeding one.

The optimal value of the Codon parameter was determined in the M1 model, with Hominidae as a foreground clade, based on AIC, AICc, and BIC criteria. The parameter specifies the equilibrium codon frequencies in the codon substitution model. The correct branch lengths of the phylogenetic tree for the codon-based analysis of positive selection were calculated using model M0 with fix length = 0. The branch lengths were then fixed for all experiments with fix length = 2. Branch-site tests were conducted on 35 specific branches and clades of the mammalian phylogenetic tree using strict (χ 2-distribution of LRT statistics, $P < 0.01$) and relaxed conditions (50:50 mixture distribution of the χ 2-distribution and a point mass of zero, $P < 0.05$). The lenient settings were used to reduce the chance of a false-negative error, as the test is cautious under strict conditions (Zhang et al., 2005). The Bonferroni correction and the Benjamini-Hochberg procedure were used to minimize the chances of a false-positive error caused by multiple testing (Anisimova and Yang, 2007). The BEB technique was utilized over a major period of the LRT to identify codons likely to undergo positive selection, with PP criteria set at 0.9 and 0.95. The IBS program version 1.0 showed the localization of areas experiencing positive selection pressure in protein primary structures. We conducted further positive selection tests using the MEME program within the HyPhy software package v.2.2.4 (Murrell et al., 2012) to confirm the reliability of our results.

To determine whether selective pressure acted on homologous nutritional pathway genes, we compared the ratio of synonymous to non-synonymous substitutions (dN/dS). The ω was calculated using the PAML codon-based ML approach, referred to as CODEML [7]. We used two different PAML models to determine whether there was a difference in the selective pressures exerted on the various grasshopper lineages. In this analysis, we focused on the ω values at the ends of the branches. We focused on the rate at which mutations have accumulated between modern species and their closest reconstructed relatives. According to the free-ratio model [8], the & values at each branch are predictably random. Initially, positive selection was detected using the branch-site model in PAML [9]. The parameters for testing the null hypothesis were ω=1. The level of statistical significance was determined by employing a chi-square distribution, with the difference in the number of parameters for the two models being equal to two times the difference in the log-likelihood values and the degrees of freedom. The identification of positive selection is frequently inconsistent due to differences between different approaches in terms of periods, assumptions, methodologies, and gene conversion bias [10]. The PAML site-branch model has been adjusted for multiple testing using Bonferroni's correction with various parameters. Furthermore, we validated these findings using various independent tools, including the HyPhy package [11]. We used site models (M1, M2, M8a, and M8) that allowed variation between sites to determine the chance of each site in each gene being under positive selection. This was done to assess the probability of each position within each gene. This model detected signs of positive selection within the gene at a few specific locations during brief periods of evolutionary time. Both the alternative model of positive selection (ω>1) and the null model of neutral evolution (ω=1) were employed in the branch-site test to assess if each branch experienced selective pressure. The alternative model of positive selection was selected for its prediction of higher selection levels on each branch compared to the null model. We used this methodology to identify examples of positive selection on a small number of genomic sites across all grasshopper lineages. We employed the likelihood ratio test (LRT) to assess each paired model and chose the one that most closely fit our data. We used site models (M1, M2, M8a, and M8) that allowed variation between sites to determine the chance of each site in each gene being under positive selection. This was done to assess the probability of occurrence for each location inside each gene. This model detected signs of positive selection within the gene at a few specific locations during brief periods of evolutionary time. Both the alternative model of positive selection (ω>1) and the null model of neutral evolution (ω=1) were employed in the branch-site test to assess selective pressure on each branch. The alternative model of positive selection was selected due to its prediction that each branch will experience greater degrees of selection compared to the null model. We used this methodology to identify examples of positive selection on a small number of genomic sites across all grasshopper lineages. We employed the likelihood ratio test (LRT) to assess each paired model and chose the one that most closely fit our data.

Protein domain and structure analysis

The positively selected sites from the previous stage were used for future structural analysis. We utilized the protein secondary structure prediction program PSIPRED 4.0 [12] and the AlphaFold2 protein structure database[13] to generate educated guesses about the degree of similarity between the mammalian proteins' predicted secondary and tertiary structures. SCANSITE 4.0 was used to develop predictions for the specific sites of kinase phosphorylation and binding

domains, given a database of 81 mammalian kinases/domains [14]. The output was then put through an additional filtering phase with the rigor level set to "high." After that, the linker sections and the domains were examined by hand. To further understand the functional significance of the putatively selected locations, we superimposed them on the 3D structures of the proteins. Using the homology modeling software made available by the I-TASSER server, we made predictions about the 3D gene structures [15]. The mammalian genome, received from GenBank, was used to deduce the protein sequences of positively chosen genes. From UniProt [16], we collected functional information regarding the presumptively recognized genes as being positively selected.

Functional analysis

The protein sequences were evaluated using two free tools found online; Clustal W was used for sequence alignment, and the LPIcom server was used to annotate amino acid similarities. This protein was analyzed with the help of the online LPIcom server [17]. We classified the detected proteins based on their projection at a particular gene ontology (GO) hierarchy level, emphasizing the GO 'Biological Process' (GOBP) class. To do this, we used the function 'groupGO,' which you can find here. The 'enrichGO' function was then used to execute enrichment tests for GOBP keywords based on a protein kinase distribution against a background list of all proteins in the relevant annotation database. These tests were conducted against a protein kinase distribution. To visualize all GO terms related to nutrient metabolism, we used g: GOSt, a web tool in the g: Profiler suite (http://biit.cs.ut.ee/gprofiler/), [18] in conjunction with Cytoscape's Enrichment Map program (http://www.baderlab.org/Software/EnrichmentMap) [19]. We integrated information from these large-scale transcriptome investigations with that from the Genotype-Tissue Expression (GTEx) database Release V8 (dbGaP Accession phs000424.v8.p2) [20]. This database offers information on gene-level associations that explain how gene expression levels test and mediate impacts on phenotypes [20].

Results

We employed the MirrorTree method (Ochoa and Pazos, 2010) to confirm the co-evolution of the GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1 genes. We computed the Pearson correlation coefficients for the evolutionary distance matrices of phylogenetic trees obtained from multiple sequence alignments of orthologous genes from several mammalian species. The Pearson correlation coefficient varied between 0.84 and 0.95, with a significance level of 0.001. Phylogenetic trees were compared in pairs, showing high Pearson correlation coefficients, which confirmed the co-evolution of the GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1 genes. We utilized concatenated gene sequences that were matched to create a phylogenetic tree. Employing concatenated gene sequences instead of individual gene sequences increased the statistical power of the molecular evolution study and improved the accuracy of the resulting phylogenetic tree by analyzing a greater number of substitutions. We generated an unrooted phylogenetic tree by merging the coding sections of the GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1 genes (Figure 1). The terminal nodes of the phylogenetic tree were strongly supported by bootstrap values and closely matched known mammalian evolutionary relationships, with minor discrepancies.

Codon-based positive selection analyses

Before performing positive selection tests, we examined the sequences for recombination events since recombination might lead to inaccurate positive outcomes. The maximum likelihood method was employed to study molecular evolution. Nucleotide sequences encoding proteins can help identify evolutionary events involving either episodic or persistent positive selection. Positive selection processes were tested during the molecular evolution of genes in the GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1 gene cluster. We utilized the CODEML tool to obtain log likelihood function values for site models M8 and M8a. We conducted a likelihood ratio test (LRT) to identify sites experiencing positive selection pressure ($\omega > 1$) across all branches of the mammalian evolutionary tree. The test's Likelihood Ratio Test (LRT) value of 175.19 with a p-value of 0.01 was statistically significant. In silico research showed that all branches of the mammalian phylogenetic tree, created using concatenated sequences of the GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1 genes, exhibit regions under positive selection pressure. We then employed the Bayes empirical Bayes (BEB) approach to identify the sites. Sites having a posterior probability (PP) > 0.9 were considered to have experienced positive selection in their evolution. Several amino acid positions have been identified as being under positive selection pressure across all branches of the evolutionary tree, with specific probabilities and values assigned to each. Most of the possible sites were identified inside the conserved domains regions of the proteins in the cluster (Figure 1). We observed widespread positive selection events in these genes' molecular history and also explored the potential impact of episodic positive selection on the genes' molecular evolution. Positive selection typically happens by influencing particular sites within specific clades and branches of a phylogenetic tree. After calculating the log likelihood values for two branch-site models, we conducted Likelihood Ratio Tests (LRT) on specific clades and branches of the mammalian phylogenetic tree under both strict and lenient conditions. To investigate whether certain sites are under positive selection ($\omega>1$) or under relaxed negative selection in specific branches (foreground branches) of the mammalian phylogeny compared to other branches, we initially used the branch-site test 1 (Zhang et al., 2005). Multiple verified phylogenetic branches and clades exhibited statistically significant likelihood ratio test (LRT) values. Under stringent criteria, selection events were identified in 20 test branches, but under lenient conditions, they were recognized in 27 test branches (Table 1). Even in lenient testing conditions, the likelihood ratio test (LRT) scores for the branches and clades of the phylogenetic tree were not statistically significant. For these branches, we did not find evidence of relaxed negative selection or positive selection. Test 1 could not distinguish between positive selection and relaxation of selective constraint, so we utilized test 2, developed by the authors, to directly assess the presence of positive selection in the lineages of interest. We tested the hypothesis that certain branches or groups of branches in the phylogenetic tree are under positive selection pressure ($\omega>1$) compared to other branches (A1 vs. A), for the branches and clades that passed test 1 (Zhang et al., 2005). We identified positive selection events in 15 out of 20 branches using strict criteria and in 26 out of 27 branches under less strict conditions (Table 1). An in silico investigation of the molecular evolution of the GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1 genes (Figure 2) revealed independent positive selection events in most branches of the mammalian phylogenetic tree.

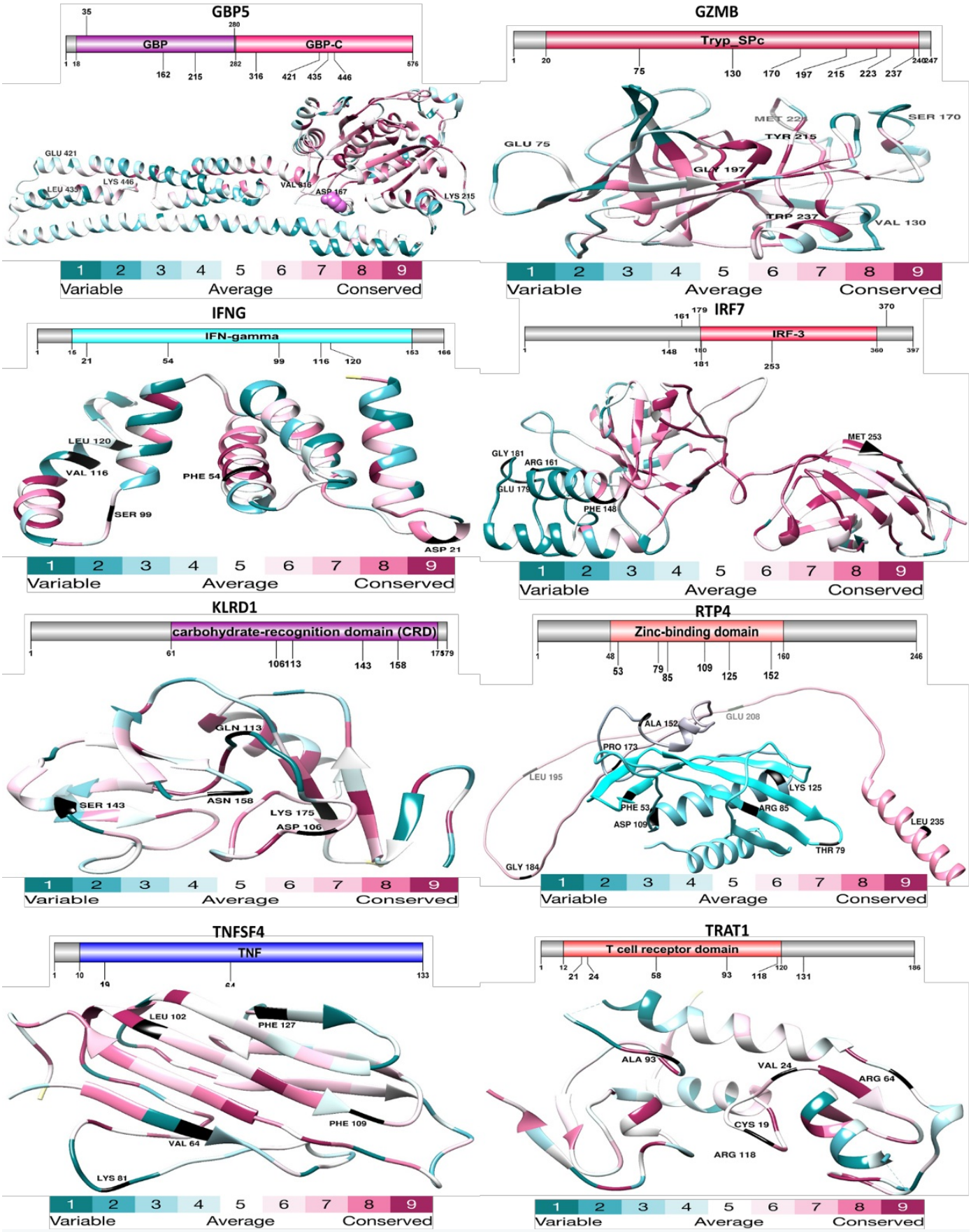

Figure 1. Analysis of the domain structure and selection of the proteins Trat1, Gbp5, Ifng, Irf7, Klrd1, Rtp4, and Tnfsf4. This diagram, generated using the DOG 1.0 illustrator, displays the structural organization of the GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1 proteins, with a focus on the examination of their conserved domains. Emphasis is given to the protein domains, specifically on the identification of regions

that are subject to positive selection. These sites are correlated with the three-dimensional configuration of proteins, exposing the adaptive evolution occurring at the molecular scale.

The increased positive selection rates on these sequences may be due to dS saturation or inadequate taxon sampling, impacting the reconstruction of the ancestral sequence and the calculation of several model parameters. It is widely known that this problem can sometimes yield inaccurately positive outcomes. The positions in the primary protein structure of the genes GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1 were identified using the corresponding figures and table (Figure 3). Closely related populations and species of animals appear to have had rapid evolutionary repercussions triggered by their microbiota, and it's possible that these effects have even impacted the recent evolution of humans. The gut microbiota has helped in the adaptive evolution of mammalian gut shapes designed to accommodate helpful microbes. The gut microbiota probably contributed to the development of both innate and adaptive immune systems in mammals.

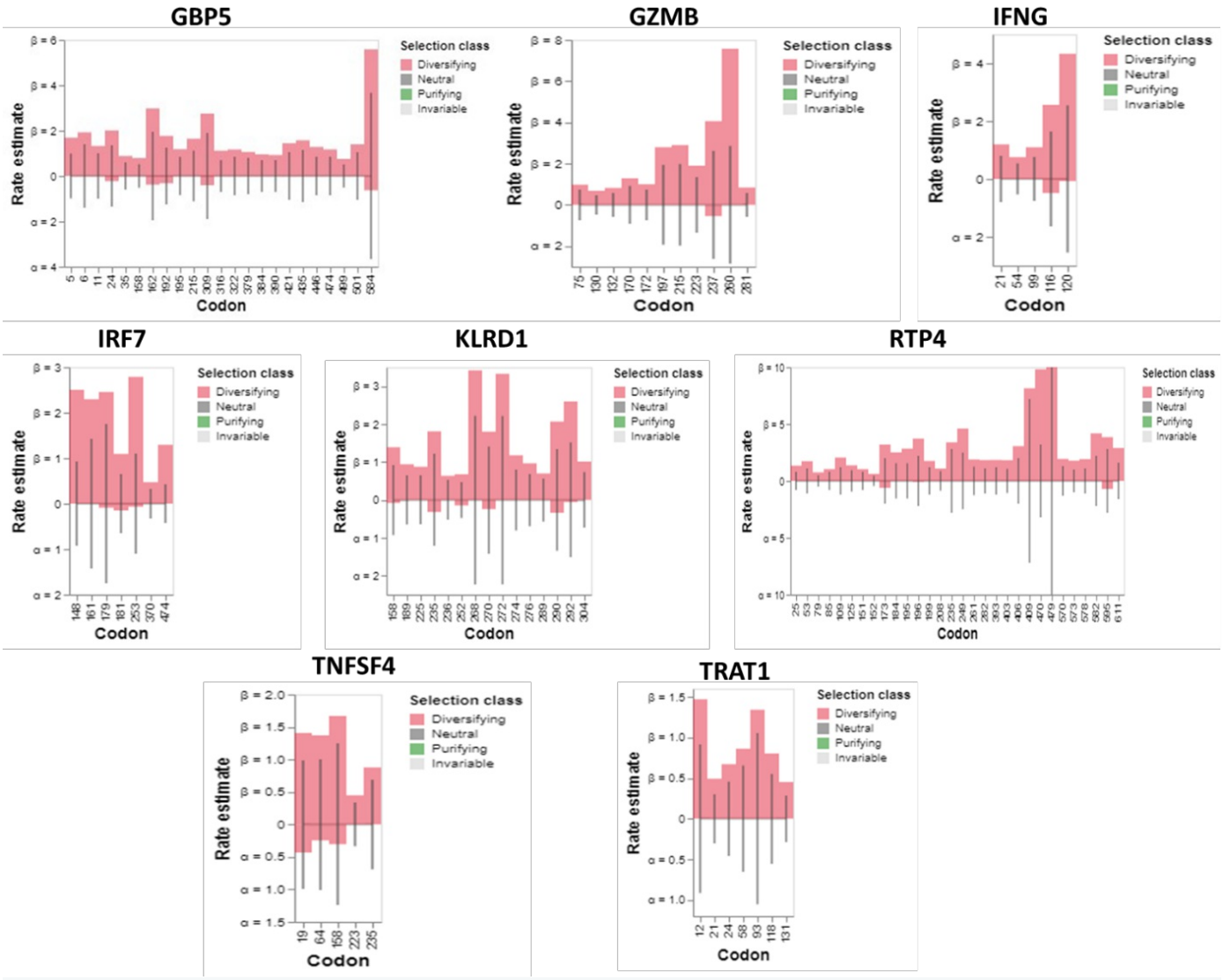

Figure 2. Results for synonymous (α) and non-synonymous (β) rates at each site are displayed as bars, representing maximum probability estimates. The estimations are shown by the line under the null model (α=β). This value censors estimates that are more than 10.

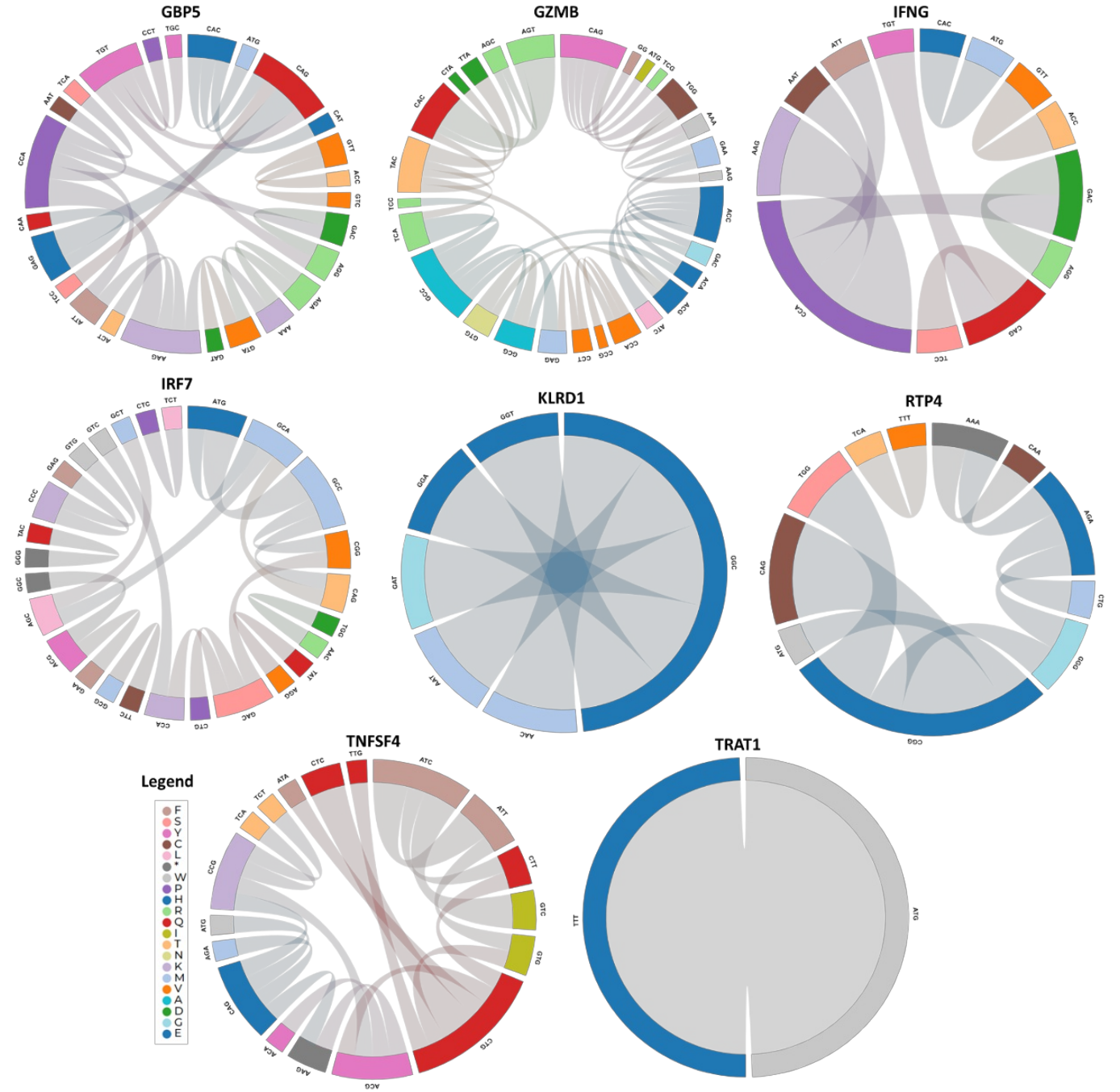

Figure 3. Three-hit replacements frequently occur in non-synonymous substitutions. Three-hit replacements with 3H+ support are substitutions that take place at sites with an ER (3H+:2H) configuration. Three-hit substitutions with a 2H but not 3H+ support are defined as replacements that happen at locations where the ratio of ER (3H+:2H) is less than 1 and the ratio of ER (2H: 1H) is not specified. The histogram displays the branch lengths where the two types of substitutions are estimated to take place.

Adaptation selection analysis

aBSREL detected evidence of episodic diversifying selection on two out of the 35 branches in the GBP5 phylogeny. There were a total of 35 branches that underwent official testing for diversifying selection. The significance of the results was

evaluated using the Likelihood Ratio Test at a significance level of $p < 0.05$, after adjusting for multiple comparisons (Figure 4). The comprehensive findings table provides information on the significance and number of rate categories inferred at each branch (Table 1). aBSREL detected episodic diversifying selection on 9 out of 29 branches in the GZMB phylogeny (Figure 4); 29 branches underwent formal testing for diversifying selection. The significance was evaluated using the Likelihood Ratio Test at a threshold of $p < 0.05$, following adjustment for multiple testing. The full findings table contains information about the significance and number of rate categories inferred at each branch. aBSREL detected episodic diversifying selection on two branches out of a total of 35 branches in the IFNG phylogeny (Figure 4). There were 35 branches that underwent official testing for diversifying selection. The significance of the results was evaluated using the Likelihood Ratio Test at a significance level of $p < 0.05$, with adjustments made for multiple testing (Figure 5). The comprehensive findings table contains information on the significance and number of rate categories inferred at each branch (Table 1). aBSREL detected episodic diversifying selection on 7 out of the 24 branches in the IRF7 phylogeny (Figure 4). There were 24 branches that underwent official testing for diversifying selection. The significance of the results was evaluated using the Likelihood Ratio Test at a significance level of $p < 0.05$, after adjusting for multiple comparisons. The comprehensive findings table provides information on the significance and number of rate categories inferred at each branch (Figure 4). aBSREL detected evidence of episodic diversifying selection on four out of 35 branches in the KLRD1 phylogeny (Figure 4). There were a total of 35 branches that underwent official testing for diversifying selection. The significance of the results was evaluated using the Likelihood Ratio Test at a threshold of $p < 0.05$, with adjustments made for multiple testing. The full findings table contains information about the significance and number of rate categories inferred at each branch. aBSREL detected episodic diversifying selection on 8 out of the 29 branches in the RTP4 phylogeny (Figure 4). 29 branches underwent formal testing for diversifying selection. The significance was evaluated using the Likelihood Ratio Test at a threshold of $p < 0.05$, following adjustment for multiple testing. The full findings table contains information about the significance and number of rate categories inferred at each branch. aBSREL detected episodic diversifying selection on two out of the 58 branches in the TNFSF4 phylogeny (Figure 4). A grand total of 58 branches underwent formal testing to assess the presence of diversifying selection. The significance was evaluated using the Likelihood Ratio Test with a threshold of $p < 0.05$, after adjusting for multiple testing. The comprehensive findings table provides information on the significance and number of rate categories inferred at each branch. aBSREL detected evidence of episodic diversifying selection on two out of the 41 branches in the TRAT1 phylogeny (Figure 4). 41 branches underwent formal testing for diverse selection. The significance of the results was evaluated using the Likelihood Ratio Test at a significance level of $p < 0.05$, while accounting for multiple testing (Figure 5). The full findings table contains information about the significance and number of rate categories inferred at each branch (Table 1).

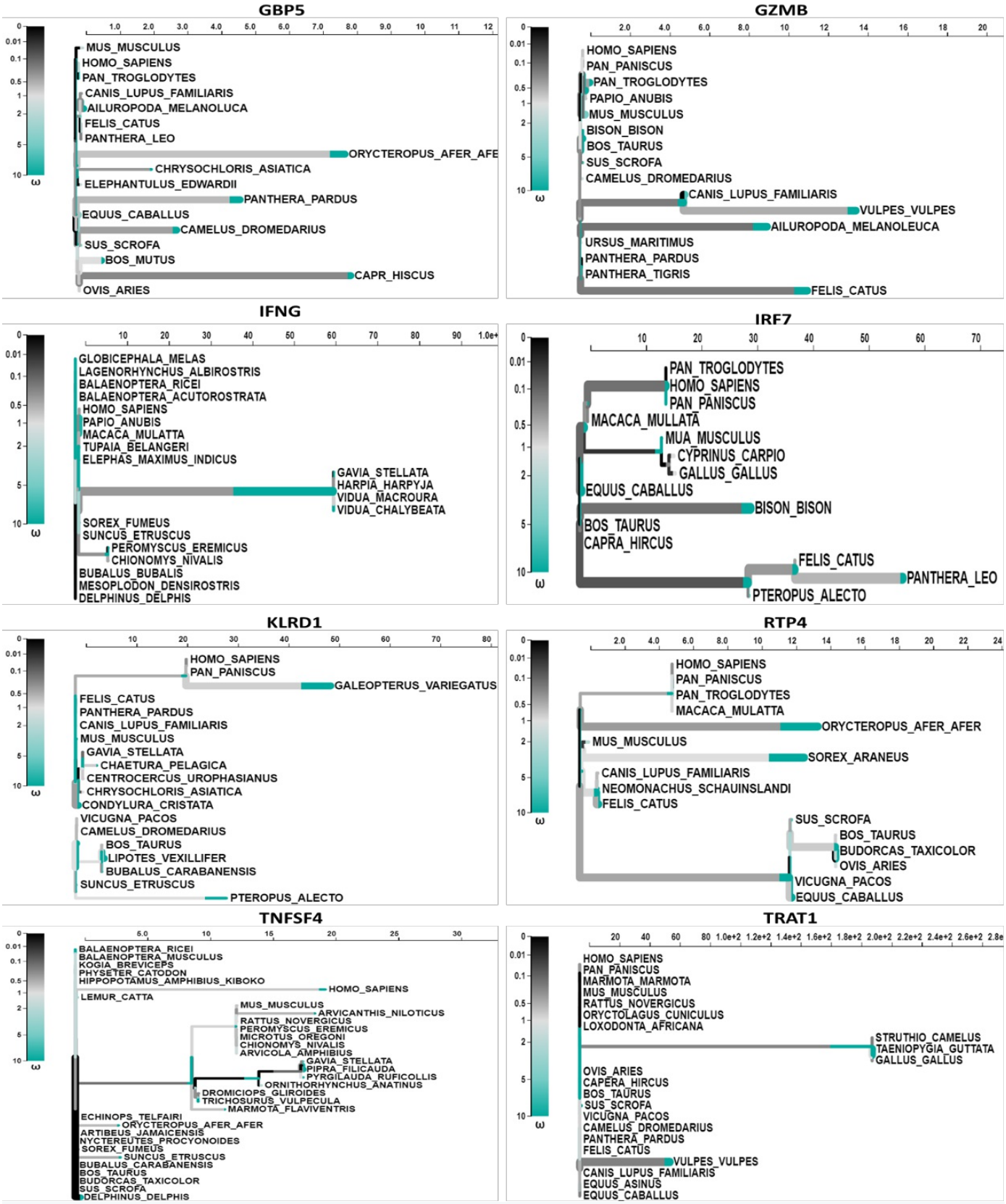

Figure 4. An aBSREL adaptive model tree was applied to analyze the full-length GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1 genes across mammalian species. The shade of the branches is determined by the inferred ω distribution. Alleles that have been determined to be under positive selection, with a statistical significance of less than 0.05 after adjustment, are visually represented by thick black branches.

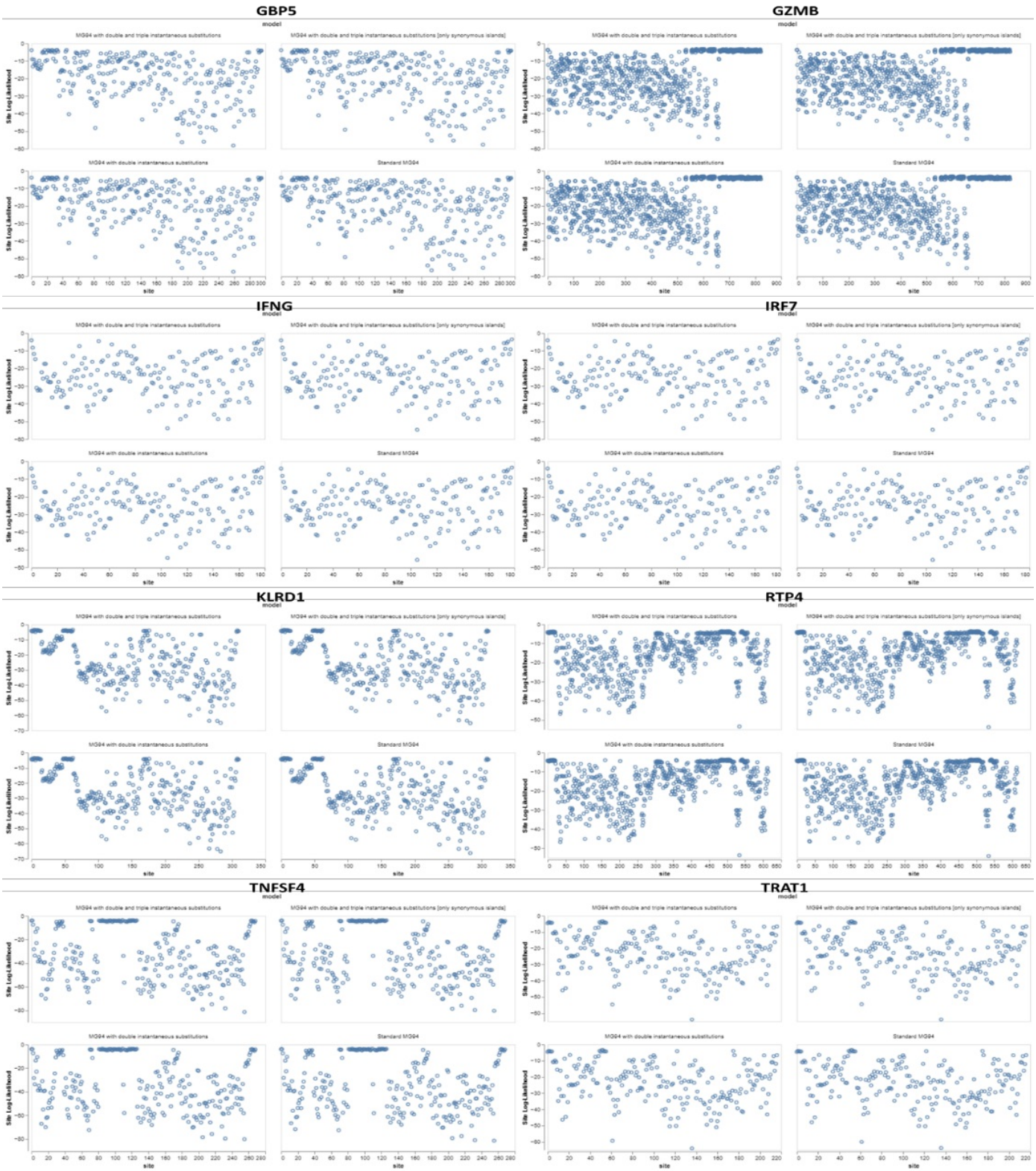

Figure 5. The Site Log-Likelihood analyses provide estimated ω rate distributions for the relative rate distribution (mean 1) for site-to-site non-synonymous rate variation that fits MG94 with double and triple instantaneous substitutions.

Table 1. Detailed site-by-site results from the FEL analysis

Gene	Codon	alpha	beta	alpha=beta	LRT	p-value	Total branch length	class
	5	0	1.69	0.968	7.776	0.0053	2.731	Diversifying

GBP5	6	0	1.926	1.376	4.873	0.0273	3.884	Diversifying
	11	0	1.322	0.989	3.856	0.0496	2.793	Diversifying
	35	0	0.885	0.577	4.096	0.043	1.63	Diversifying
	162	0.38	2.981	1.937	5.871	0.0154	5.467	Diversifying
	215	0	1.642	1.11	4.856	0.0276	3.133	Diversifying
	316	0	1.117	0.691	5.56	0.0184	1.951	Diversifying
	421	0	1.445	1.029	4.402	0.0359	2.905	Diversifying
	435	0	1.575	1.129	5.272	0.0217	3.188	Diversifying
	446	0	1.281	0.834	5.359	0.0206	2.353	Diversifying
GZMB	75	0	0.983	0.723	3.012	0.0826	1.952	Diversifying
	130	0	0.687	0.456	3.211	0.0731	1.23	Diversifying
	132	0	0.823	0.576	2.811	0.0936	1.555	Diversifying
	170	0	1.288	0.913	3.846	0.0499	2.463	Diversifying
	172	0	1.002	0.728	3.047	0.0809	1.965	Diversifying
	197	0	2.797	1.915	5.445	0.0196	5.168	Diversifying
	215	0	2.9	1.964	6.379	0.0115	5.3	Diversifying
	223	0	1.901	1.321	4.702	0.0301	3.564	Diversifying
	237	0.54	4.06	2.613	4.7	0.0302	7.051	Diversifying
	260	0	7.571	2.842	10.49	0.0012	7.669	Diversifying
	281	0	0.841	0.576	3.523	0.0605	1.555	Diversifying
IFNG	21	0	1.194	0.783	5.309	0.0212	2.876	Diversifying
	54	0	0.753	0.528	3.765	0.0523	1.94	Diversifying
	99	0	1.096	0.743	4.438	0.0351	2.731	Diversifying
	116	0.49	2.563	1.634	3.617	0.0572	6.006	Diversifying
	120	0.07	4.329	2.532	4.269	0.0388	9.306	Diversifying
IRF7	148	0	2.503	0.925	3.283	0.07	6.233	Diversifying
	161	0	2.295	1.421	3.074	0.0796	9.571	Diversifying
	179	0.09	2.456	1.742	3.37	0.0664	11.74	Diversifying
	181	0.15	1.093	0.648	3.67	0.0554	4.368	Diversifying
	253	0.07	2.787	1.1	2.921	0.0874	7.409	Diversifying
	370	0	0.471	0.323	3.1	0.0783	2.175	Diversifying
	474	0	1.294	0.412	11.68	0.0006	2.775	Diversifying
Klrd1	158	0.08	1.39	0.918	3.311	0.0688	5.077	Diversifying
	189	0	0.942	0.644	3.101	0.0782	3.558	Diversifying
	225	0	0.875	0.635	4.334	0.0374	3.508	Diversifying
	235	0.32	1.809	1.207	3.914	0.0479	6.67	Diversifying
	236	0	0.637	0.514	2.986	0.084	2.841	Diversifying
	252	0.15	0.677	0.468	2.725	0.0988	2.586	Diversifying
	268	0	3.421	2.215	5.524	0.0188	12.24	Diversifying
	270	0.24	1.801	1.405	4.09	0.0431	7.764	Diversifying
	272	0	3.33	2.213	7.117	0.0076	12.23	Diversifying

	274	0	1.177	0.797	6.915	0.0085	4.408	Diversifying
	276	0	0.966	0.682	6.166	0.013	3.768	Diversifying
	289	0	0.699	0.561	3.249	0.0715	3.102	Diversifying
	290	0.34	2.069	1.336	5.638	0.0176	7.384	Diversifying
	292	0.06	2.599	1.503	7.414	0.0065	8.308	Diversifying
	304	0	1.012	0.722	3.563	0.0591	3.99	Diversifying
Rtp4	25	0	1.342	0.792	2.842	0.0919	2.409	Diversifying
	53	0	1.731	1.075	5.869	0.0154	3.267	Diversifying
	79	0	0.753	0.498	3.14	0.0764	1.514	Diversifying
	85	0	1.043	0.766	2.946	0.0861	2.328	Diversifying
	109	0	2.056	1.188	3.257	0.0711	3.611	Diversifying
	125	0	1.369	0.92	3.1	0.0783	2.798	Diversifying
	151	0	1.04	0.771	3.073	0.0796	2.343	Diversifying
	152	0	0.622	0.424	3.04	0.0813	1.289	Diversifying
	173	0.6	3.197	1.97	2.799	0.0944	5.988	Diversifying
	184	0	2.512	1.525	8.134	0.0043	4.636	Diversifying
	195	0	2.839	1.525	9.264	0.0023	4.635	Diversifying
	196	0.07	3.725	2.159	3.891	0.0485	6.562	Diversifying
	199	0	1.754	1.166	3.689	0.0548	3.545	Diversifying
	208	0	1.088	0.829	3.178	0.0746	2.521	Diversifying
	235	0	3.386	2.776	3.551	0.0595	8.438	Diversifying
	249	0	4.603	2.443	4.928	0.0264	7.426	Diversifying
	261	0	1.882	1.226	3.569	0.0589	3.725	Diversifying
	282	0	1.833	1.101	2.833	0.0923	3.348	Diversifying
	393	0	1.843	1.18	2.729	0.0985	3.586	Diversifying
	403	0	1.823	1.029	2.932	0.0869	3.127	Diversifying
	406	0	3.058	1.953	3.004	0.0831	5.937	Diversifying
	409	0	8.143	7.189	3.489	0.0618	21.85	Diversifying
	470	0	9.818	3.175	2.895	0.0888	9.651	Diversifying
	479	0	22.38	18.44	4.826	0.028	56.06	Diversifying
	570	0	1.937	1.298	2.82	0.0931	3.947	Diversifying
	573	0	1.778	0.984	3.196	0.0738	2.991	Diversifying
	578	0	1.92	1.075	2.932	0.0868	3.267	Diversifying
	582	0	4.18	2.174	4.207	0.0402	6.607	Diversifying
	595	0.71	3.848	2.784	2.828	0.0926	8.461	Diversifying
	611	0	2.891	1.571	4.372	0.0365	4.774	Diversifying
TNFSF4	19	0.43	1.408	0.984	2.793	0.0947	7.414	Diversifying
	64	0.25	1.372	1.002	3.629	0.0568	7.546	Diversifying
	158	0.3	1.67	1.239	3.188	0.0742	9.331	Diversifying
	223	0	0.449	0.332	2.822	0.093	2.498	Diversifying
	235	0	0.877	0.684	5.097	0.024	5.152	Diversifying

TRAT1	12	0	1.471	0.91	4.67	0.0307	2.959	Diversifying
	21	0	0.492	0.299	2.895	0.0889	0.971	Diversifying
	24	0	0.671	0.454	3.187	0.0742	1.475	Diversifying
	58	0	0.862	0.651	2.756	0.0969	2.116	Diversifying
	93	0	1.341	1.05	3.21	0.0732	3.413	Diversifying
	118	0	0.804	0.55	3.787	0.0516	1.786	Diversifying
	131	0	0.451	0.282	2.736	0.0981	0.917	Diversifying

Recombination analysis

The GARD analysis detected recombination breakpoints in the GBP5 gene. GARD analyzed a total of 13,556 models at a speed of 21.42 models per second. The alignment consisted of 1183 possible breakpoints, resulting in a search space of 635810244030937500 models with a maximum of 7 breakpoints. However, the genetic algorithm only searched 0.00% of this search space (Figure 6). The AICc score of the best-fitting GARD model, which permits different topologies between segments (29983.2), is compared to that of the model assuming the same tree for all partitions inferred by GARD but allowing different branch lengths between partitions (30120.2). This suggests that the multiple tree model may be preferred over the single tree model by an evidence ratio of 100 or greater, indicating that at least one of the breakpoints represents a genuine topological incongruence. The GARD analysis detected recombination breakpoints within the GZMB gene. GARD analyzed a total of 9905 models at a speed of 50.54 models per second. The alignment consisted of 458 possible breakpoints, resulting in a search space of 166123556333 models with a maximum of 5 breakpoints. However, the genetic algorithm only searched 0.00% of this search space (Figure 6). The AICc score of the best-fitting GARD model, which permits different topologies between segments (10967.6), is compared to that of the model assuming the same tree for all partitions inferred by GARD but allowing different branch lengths between partitions (11723.4). This suggests that the multiple tree model may be preferred over the single tree model by an evidence ratio of 100 or greater, indicating that at least one of the breakpoints represents a genuine topological incongruence. GARD did not detect any signs of recombination in IFNG. GARD analyzed a total of 2630 models at a speed of 49.62 models per second. The alignment consisted of 409 possible breakpoints, resulting in a search space of 409 models with a maximum of 1 breakpoint (Figure 6). The genetic algorithm examined 643.03% of this search space. The comparison of the AICc scores between the best-fitting GARD model, which permits different topologies between segments (9309.1), and the model that assumes the same tree for all partitions inferred by GARD but allows different branch lengths between partitions (9309.1), indicates that the multiple tree model cannot be favored over the single tree model by an evidence ratio of 100 or more. This suggests that some or all of the breakpoints may be indicative of rate variation rather than topological incongruence. Notably, GARD detected evidence of recombination breakpoints in the IRF7 and KLRD1 genes. GARD analyzed a total of 12,831 models at a speed of 41.52 models per second. The alignment consisted of 1253 possible breakpoints, resulting in a search space of 25635663809007 models with a maximum of 5 breakpoints (Figure 6). However, the genetic algorithm only searched 0.00% of this search space. The AICc score of the best-fitting GARD model, which permits different topologies between segments (23104.9), is compared to that of the model assuming the same tree for all partitions inferred by GARD but allowing different branch lengths between partitions (23164.6). This suggests that the multiple tree

model may be preferred over the single tree model by an evidence ratio of 100 or more, indicating that at least one of the breakpoints represents a genuine topological incongruence. GARD analyzed a total of 1419 models at a speed of 43.00 models per second. The alignment consisted of 717 possible breakpoints, resulting in a search space of 257,403 models that might have up to 2 breakpoints (Figure 6). The genetic algorithm examined only 0.55% of this search area. Comparing the AICc score of the best-fitting GARD model, which permits different topologies between segments (16444.6), with that of the model assuming the same tree for all partitions inferred by GARD but allowing different branch lengths between partitions (16511.0), indicates that the multiple tree model is favored over the single tree model by an evidence ratio of 100 or more. This suggests that at least one of the breakpoints represents a genuine topological incongruence. The Genetic Algorithm for Recombination Detection (GARD) identified recombination breakpoints in the RTP4 gene. GARD analyzed a total of 11,981 models at a speed of 19.45 models per second. The alignment consisted of 1041 possible breakpoints, resulting in a search space of 10,138,915,336,889 models with a maximum of 5 breakpoints. However, the genetic algorithm only examined 0.00% of this search space. The Genetic Analysis and Recombination Detection (GARD) method did not detect any evidence of recombination in the TNFSF4 gene. GARD analyzed a total of 1793 models at a speed of 23.29 models per second. The alignment consisted of 559 possible breakpoints, resulting in a search space of 559 models with a maximum of 1 breakpoint (Figure 6). The genetic algorithm examined 320.75% of this search area. The GARD analysis detected recombination breakpoints in the TRAT1 gene. GARD analyzed a total of 2298 models at a speed of 69.64 models per second. The alignment consisted of 442 possible breakpoints, resulting in a search space of 97903 models with a maximum of 2 breakpoints. The genetic algorithm searched 2.35% of this search area.

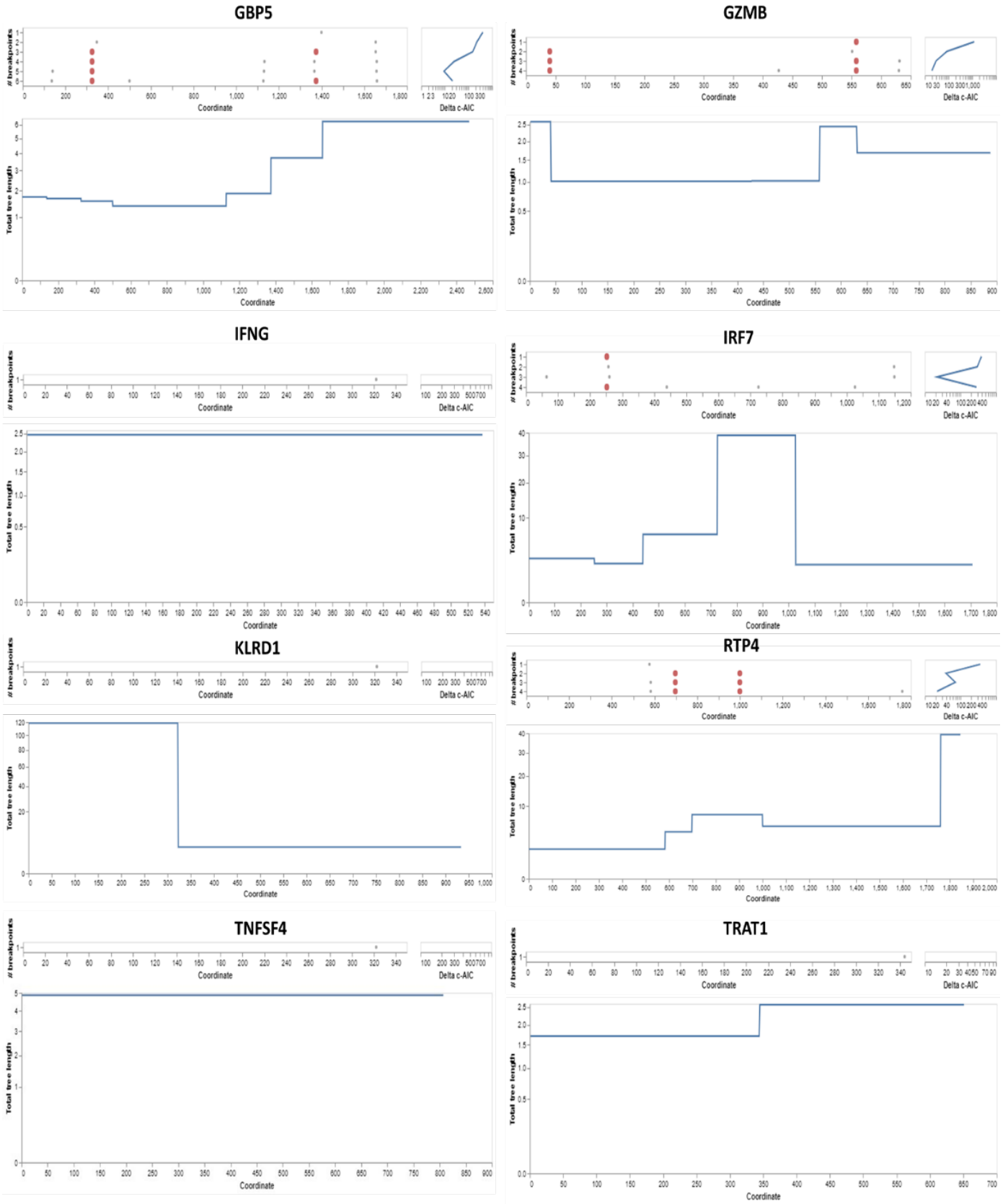

Figure 6. Left: the algorithm's best estimate of where to put breakpoints for each number of breakpoints taken into consideration. Correct: the increase in the c-AIC score (log scale) when breakpoint numbers increase.

Functional analysis

Initially, we identified all terms that were statistically enriched, such as GO/KEGG terms, canonical pathways, and hallmark gene sets. This was done based on either the default choices under Express Analysis or the choices made during Custom Analysis. We then calculated accumulative hypergeometric p-values and enrichment factors, which were used for filtering. The remaining important phrases were further organized into a hierarchical tree structure using Kappa-statistical similarities among their gene memberships, similar to the approach employed in the NCI DAVID site. A threshold of 0.3 kappa score was used to convert the tree into term clusters. The terms contained in each cluster are exported in the Excel worksheet titled "Enrichment Analysis". We extracted a subset of key phrases from the entire cluster and transformed them into a network layout. Each word is depicted as a circular node, with its size according to the number of input genes associated with that term. The color of the node indicates its cluster identification, meaning nodes of the same color belong to the same cluster. Terms that have a similarity score greater than 0.3 are connected by an edge, with the thickness of the edge representing the similarity score. The network is shown using Cytoscape, with a "force-directed" structure and edge bundling to enhance clarity. Positively chosen sites have been found in the GBP5 protein, which contains the Guanylate-binding protein (GBP) and the N-terminal domain of Interferon (IFN)-inducible GTPase. These pathogens are a wide variety of bacteria, viruses, and protozoa, and they play significant roles in innate immunity against them. After infection, it is drawn to bacteria that have escaped from vacuoles or contain pathogens, and it functions as a positive regulator of the assembly of inflammasomes by encouraging the release of ligands from the bacteria. This releases ligands that are recognized by inflammasomes, such as double-stranded DNA (dsDNA), that activates the AIM2 inflammasome or lipopolysaccharide (LPS), which activates the non-canonical CASP4/CASP11 inflammasome. The GZMB protein has a trypsin-like serine protease domain that contains the active site and is found in members of the trypsin family. The serine proteases from the trypsin family exhibit catalytic activity through a charge relay system. This system involves an aspartic acid residue that forms a hydrogen bond with a histidine, which in turn forms a hydrogen bond with a serine. The IFNG protein, which possesses the IFN-gamma domain, exhibits antiviral properties and plays a crucial role in regulating the immune system. This substance is highly effective at stimulating macrophages and has the ability to inhibit the growth of altered cells. It has the ability to enhance the antiviral and anticancer effects of type I interferons. The interferon-regulatory factor 7 includes the truncated CREB-binding protein domain. The DRAF1 (double-stranded RNA-activated factor 1) consists of these two subunits (Figure 8). The production of viral double-stranded RNA (dsRNA) during viral transcription or replication results in the activation of DRAF1. The DNA-binding specificity of DRAF1 is directly related to the transcriptional stimulation of ISGs (interferon-alpha, beta-stimulated genes). The protein IRF-3 is initially present in the cytoplasm of cells that have not been infected, but it moves to the nucleus after a viral infection occurs. The translocation of IRF-3 is accompanied by an elevation in phosphorylation of serine and threonine residues, and its interaction with the CREB co-activator only takes place following infection. The carbohydrate-recognition domain (CRD), often referred to as the C-type lectin domain (CTL), is composed of around 110 to 130 amino acid residues. There are four cysteines that are completely preserved and participate in the formation of two disulphide linkages. Lectins exhibit significant diversity in terms of both their structural composition and functional properties. The ability to bind carbohydrates may have separately and occasionally evolved in multiple unrelated families, with each family developing a structure that was conserved to serve a different activity and function. Animal lectins serve as recognition molecules in the immune system, playing roles in pathogen defense, cell movement, immunological control, and the prevention of

autoimmunity (Figure 8). The protein known as tumour necrosis factor comprises a domain called TNF. Cytokines that belong to a family can form complexes consisting of either three identical subunits or three different subunits. The p75 TNF receptor is responsible for mediating apoptosis produced by the mature T-cell receptor through TNF. The GTEx database includes a crucial tool known as an expression quantitative trait locus (eQTL) browser. This browser functions as a storage and graphical display of data collected from a nationwide research initiative that sought to discover links between genetic variants and high-throughput molecular-level expression phenotypes (Figure 9). It is worth mentioning that a considerable number of genes display connections with different tissues. Gbp5, Gzmb, Ifng, Irf7, Klrd1, Rtp4, and Trat1 genes exhibit significant expression in whole blood, whereas Klrd1, Rtp4, and Trat1 genes revealed expression in the spleen (Figure 7). Tnfsf4 and Rtp4 have shown expression in lymphocytes. Nevertheless, our examination of the mean expression levels of all (significant) genes using various enrichment techniques yielded inconclusive results about the tissues that are expected to have a higher prevalence of diseases and well-established biological processes (Table 2).

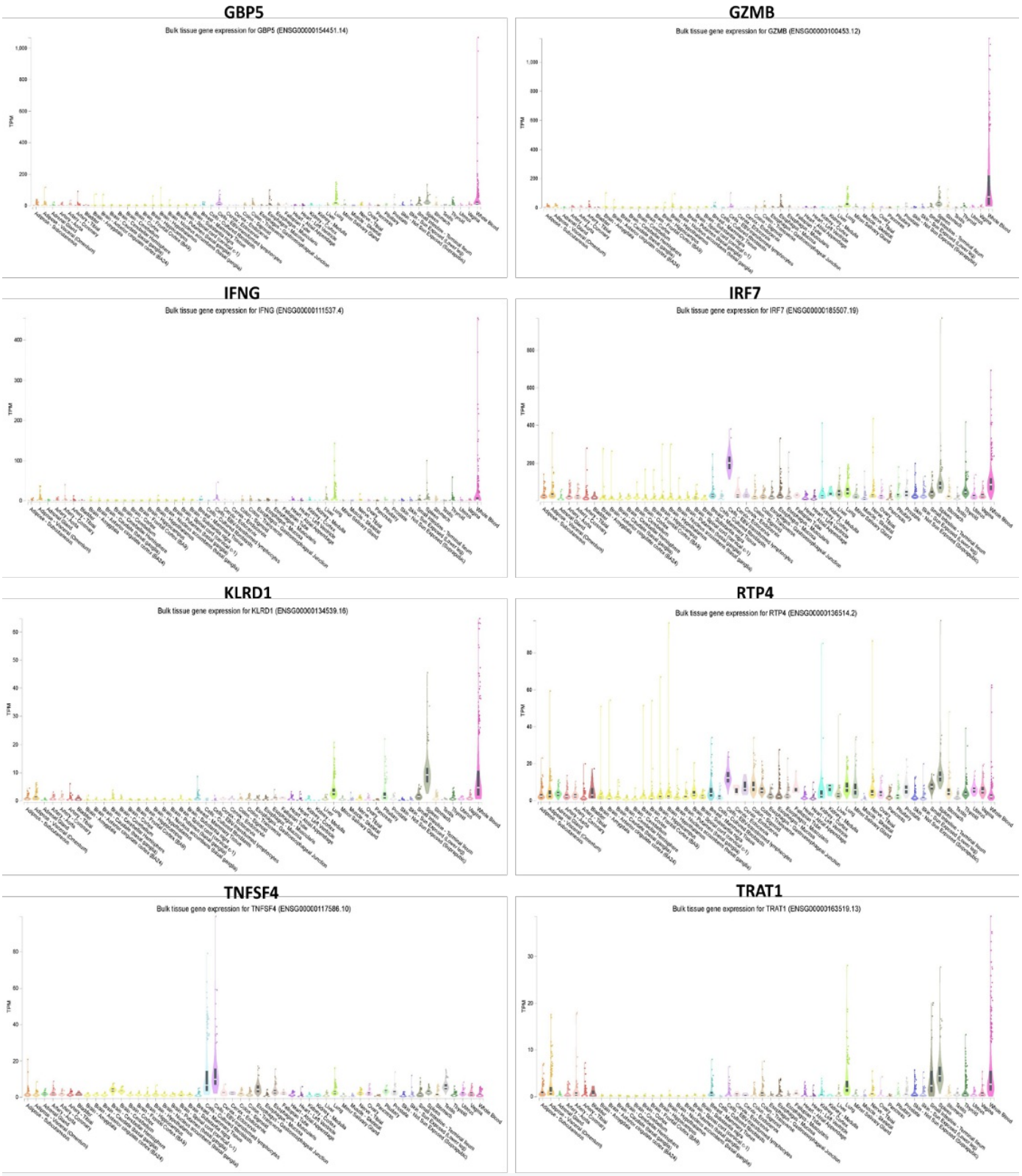

Figure 7. Tissue-specific expression of GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1 genes in humans used as a reference genome. The expression data of these genes were revealed across various tissues courtesy of the GTEx consortium.

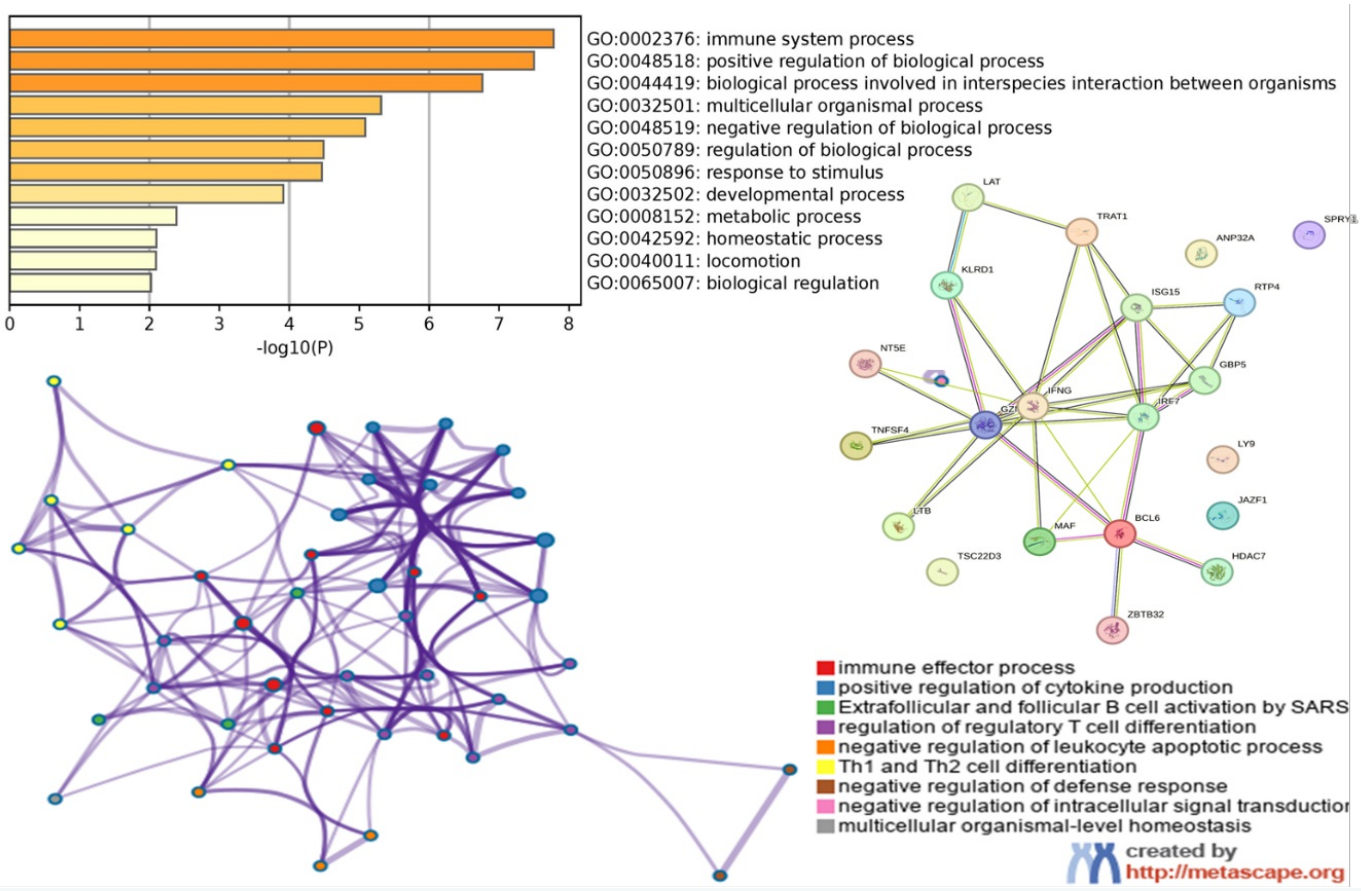

Figure 8. Detected all significantly enriched terms, including GO/KEGG terms, canonical pathways, and gene sets for GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1. Selected key phrases from the entire cluster and transformed them into a network arrangement. A protein-protein network was built by extracting connections among the genes GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1 from a data source of protein-protein interactions.

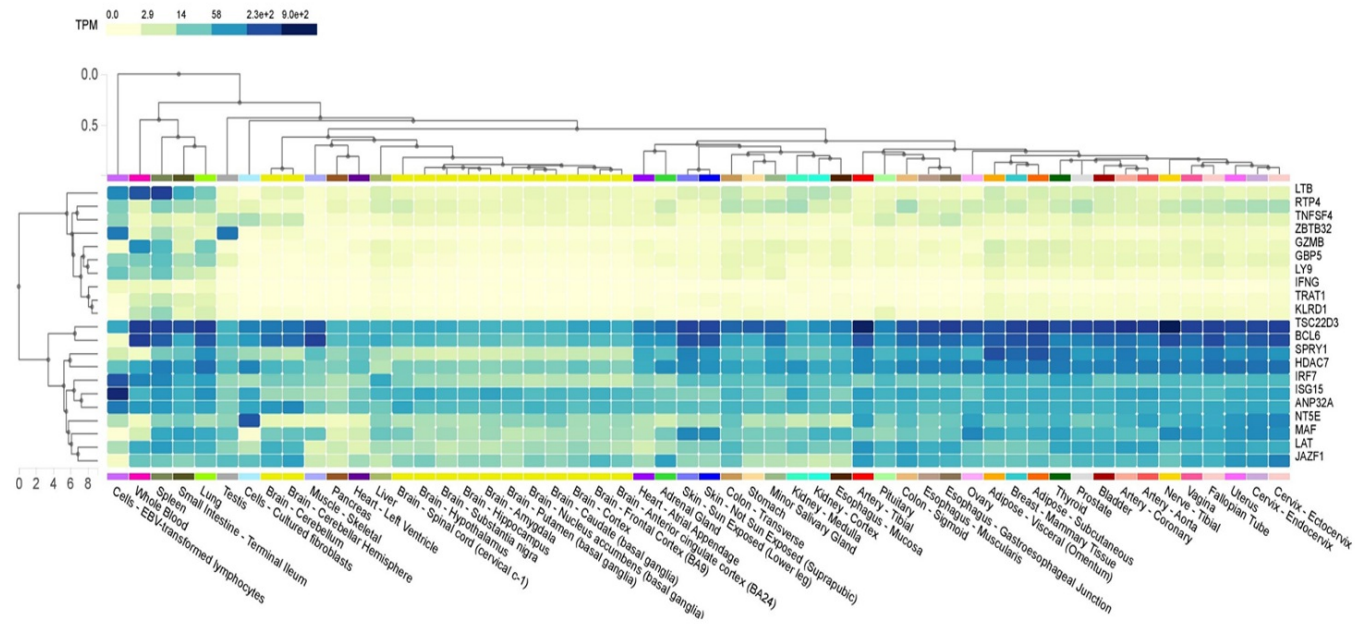

Figure 9. Expression analysis of GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1 proteins in tissues.

Table 2. The network performed GO enrichment analysis to identify the underlying "biological meanings".

GO	Category	Description	Count	%	Log10(P)	Log10(q)
GO:0002252	GO Biological Processes	immune effector process	7	33.33	-7.78	-3.46
GO:0001819	GO Biological Processes	positive regulation of cytokine production	7	33.33	-7.50	-3.46
WP5218	WikiPathways	Extrafollicular and follicular B cell activation by SARS CoV 2	4	19.05	-6.67	-3.11
GO:0045589	GO Biological Processes	regulation of regulatory T cell differentiation	3	14.29	-5.69	-2.50
GO:2000107	GO Biological Processes	negative regulation of leukocyte apoptotic process	3	14.29	-5.09	-2.07
hsa04658	KEGG Pathway	Th1 and Th2 cell differentiation	3	14.29	-4.46	-1.72
GO:0031348	GO Biological Processes	negative regulation of defense response	3	14.29	-3.00	-0.56
GO:1902532	GO Biological Processes	negative regulation of intracellular signal transduction	3	14.29	-2.13	0.00
GO:0048871	GO Biological Processes	multicellular organismal-level homeostasis	3	14.29	-2.10	0.00

Discussion

The gut mucosal immune system functions as the interface between the internal body and the external environment. The microorganisms present in the gut environment have a continuous impact on the immune system, and in return, the immune system of the host has an influence on the makeup of the microbiome. The intricate balance is maintained by a complex interplay between microbial activity, the intestinal epithelium, and elements of the innate and adaptive immune system. The mucosal surface's formation and development are crucial milestones in maintaining mammalian life from an evolutionary standpoint. We analyzed a group of 8 protein-coding orthologs (GBP5, GZMB, IFNG, IRF7, KLRD1, RTP4, TNFSF4, and TRAT1) that are present in the genomes of humans, monkeys, dogs, cats, cows, mice, and domestic yaks. Our goal was to identify signs of positive selection in these genes. Within these genes exhibiting statistically significant signals ($P < 0.05$ corrected), further analysis of branch sites indicated the presence of positive selection specifically along the mammalian lineages. The M8 model, which employs positive selection, was utilized to detect variations at the codon level. A Markov Chain Monte Carlo (MCMC) model, implemented in MrBayes on the Selecton server, was utilized to ascertain the disparity at the codon level. Values were calculated for each place in both cases. The results of our study demonstrate that the coding sequences of eight genes exhibit domain conservation when analyzed using MAFFT protein alignments (Figure 1). These findings indicate that non-identical protein switches in areas undergoing purifying selection are detrimental to health and therefore unlikely to become established during the process of evolution. Genetic transfers and duplications are fundamental in the development of all major adaptive immune molecular systems on a horizontal scale. Prior research on the evolution of immune genes in birds mostly examined the co-evolution of disease hotspots, such as MAVS, in the context of influenza virus infection. They play a role in activating lymphocytes, regulating the immune system, stimulating T regulatory cells, and influencing the development and tolerance of autoimmunity. The

impact of selection on host organisms in regulating gut microbiota during the adaptive evolution of mammalian species remains inadequately understood, despite the intricate mechanisms developed by our ancestors' predecessors. Immune genes in mammalian genomes are evolving rapidly, suggesting the presence of pathogens and co-evolutionary dynamics known as red-queen dynamics. The extent to which genetic variation in immune genes is affected by differences in gut microbiota among mammalian species remains uncertain. Bacteria can change rapidly during an evolutionary arms race, making it difficult for mammalian hosts to constantly adapt to control the microbiota, which also evolves to compete among itself. Multiple verified phylogenetic branches and clades exhibited statistically significant likelihood ratio test (LRT) values. Under stringent criteria, selection events were identified in 20 test branches, but under lenient conditions, they were recognized in 27 test branches (Table 1). Even in lenient testing conditions, the likelihood ratio test (LRT) scores for the branches and clades of the phylogenetic tree were not statistically significant. For these branches, we did not find evidence of relaxed negative selection or positive selection. Test 1 could not distinguish between positive selection and relaxation of selective constraint, so we utilized test 2, developed by the authors, to directly assess the presence of positive selection in the lineages of interest. We tested the hypothesis that certain branches or groups of branches in the phylogenetic tree are under positive selection pressure ($\omega>1$) compared to other branches (A1 vs. A), for the branches and clades that passed test 1 (Zhang et al., 2005). The functional repertoires in mammalian gut microbiotas have likely supported the evolution and diversification of chitin-eating and herbivory, the specialization of mammalian species and communities on hazardous diets, and potentially even recent dietary changes in human evolution. Furthermore, there is increasing evidence that animals have adapted to depend on signals from the specific gut bacteria of their hosts during postnatal development and functioning. House mice colonized with the gut microbiota of rats or humans did not exhibit fully differentiated T cell repertoires, unlike those colonized with the gut microbiota of other house mice. Mammals have evolved to rely on the specific gut microbiota of their hosts for guidance during postnatal growth and function, as supported by a growing body of research. When house mice were colonized with the gut microbiota of rats or humans, their T cell repertoires did not properly differentiate compared to when they were colonized with the gut microbiota of other house mice. These findings suggest that the immunological development of house mice has evolved to include components of their particular gut flora since the divergence of mice and rats. When germ-free house mice were inoculated with gut microbiota from different species within the genus Mus (Mus spretus and Mus pahari), they exhibited reduced differences in body composition between males and females, larger livers, and slower growth rates compared to those inoculated with their own gut microbiota. House mice have adapted to integrate their distinct gut microbiota into their postnatal growth process since diverging from other Mus species, as indicated by these results. The wild mouse gut microbiota has been shown to enhance disease resistance and improve fitness in laboratory mice when compared to the altered gut microbiome frequently present in laboratory mice. Additionally, mice injected with the gut microbiotas of several Peromyscus species exhibited reduced digestive efficiency compared to P. polionotus animals treated with their own gut microbiota. The adaptation of a species to a new food is a significant driving force in the evolution of that species. The dietary modifications that occurred throughout the development of several primate species, including humans, have been extensively recorded across time (61,62). Additionally, some genes have been identified as being involved in positive selection driven by nutrition. An extensively researched instance can be observed in a pancreatic ribonuclease present in old-world monkey species (63). Due to the monkey's dietary changes, the protein in this species has developed

an enhanced capacity to break down bacterial DNA. Another instance is lysozyme, which facilitates the breakdown of intestinal microorganisms. This protein has demonstrated positive selection in various primate groups, including humans (64,65). The langur monkey, a species that has developed a foregut fermentation mechanism of digesting comparable to ruminants, provides the clearest understanding of the nature of this selection (64,66). The cause for positive selection in humans is not well understood, although it has been hypothesized that the transition to a diet mostly consisting of meat, which would have necessitated adaptations in bacterial digestion, may have been a contributing factor (67,68). Studies of alanine-glyoxylate aminotransferase (AGT), a gene that has different functions in herbivores and carnivores, provide support for this notion. Additionally, there is evidence of positive selection for this gene among simian primates (69). aBSREL detected evidence of episodic diversifying selection on two out of the 35 branches in the GBP5 phylogeny. There were a total of 35 branches that underwent official testing for diversifying selection. The significance of the results was evaluated using the Likelihood Ratio Test at a significance level of $p < 0.05$, after adjusting for multiple comparisons (Figure 4). While episodic diversifying selection was found on 9 out of 29 branches in the GZMB phylogeny, aBSREL detected episodic diversifying selection on two branches out of a total of 35 branches in the IFNG phylogeny and 7 out of the 24 branches in the IRF7 phylogeny (Figure 4). aBSREL detected evidence of episodic diversifying selection on four out of 35 branches in the KLRD1 phylogeny and 8 out of the 29 branches in the RTP4 phylogeny (Figure 4). Furthermore, the mammalian innate and adaptive immune systems are examples of evolutionary adaptations that were most likely motivated, at least in part, by the need to control the gut microbiota's composition in ways that enhance host fitness. The immune system offers mechanisms for distinguishing and getting rid of harmful germs while allowing healthy or commensal microbes to coexist. Studies have demonstrated that the body's lack of immunological components might have detrimental effects on the composition of the gut microbiota in hosts. For instance, it has been demonstrated that the removal of Toll-like receptors from the host genome causes disruptions in the makeup of the gut microbiota in mice, which in turn modifies the host's energy harvesting and metabolism in likely maladaptive ways {m/60/}. The GARD analysis detected recombination breakpoints in the GBP5, GZMB, IRF7, KLRD1, RTP4, and TRAT1 genes. GARD analyzed a total of 13,556 models at a speed of 21.42 models per second. The alignment consisted of 1183 possible breakpoints, resulting in a search space of 635810244030937500 models with a maximum of 7 breakpoints. However, the genetic algorithm only searched 0.00% of this search space (Figure 6). The AICc score of the best-fitting GARD model, which permits different topologies between segments (29983.2), is compared to that of the model assuming the same tree for all partitions inferred by GARD but allowing different branch lengths between partitions (30120.2). This suggests that the multiple tree model may be preferred over the single tree model by an evidence ratio of 100 or greater, indicating that at least one of the breakpoints represents a genuine topological incongruence. GARD did not detect any signs of recombination in the IFNG and TNFSF4 genes. GARD analyzed a total of 2630 models at a speed of 49.62 models per second. The alignment consisted of 409 possible breakpoints, resulting in a search space of 409 models with a maximum of 1 breakpoint (Figure 6). The genetic algorithm examined 643.03% of this search space. The comparison of the AICc scores between the best-fitting GARD model, which permits different topologies between segments (9309.1), and the model that assumes the same tree for all partitions inferred by GARD but allows different branch lengths between partitions (9309.1), indicates that the multiple tree model cannot be favored over the single tree model by an evidence ratio of 100 or more. This suggests that some or all of the breakpoints may be indicative of rate variation rather than topological incongruence. Similarly, the

makeup of the gut microbiota is different in RAG1−/− mice, who do not have adaptive immune systems. While all mammalian species have gut microbiota, the evolutionary consequences of interacting with a gut microbiota probably vary from mammalian taxon to mammalian taxon. Mammalian orders exhibit varying degrees of concordance between the makeup of the gut microbiota and the phylogenetic history of the host, according to recent investigations [21]. While the gut microbiotas of certain animal orders—like bats—show relatively minor relationships with host phylogeny, the gut microbiotas of most other orders exhibit robust evidence of phylogenetic signal. Mammalian species may be spared the risk of evolutionary reliance on a particular gut microbiota if there is no species-specific gut microbiome. In these circumstances, hosts might only evolve to incorporate signals from ambient or non-specific microorganisms into their growth, as opposed to signals from particular bacteria or groups of microbes. On the other hand, hosts might completely stop depending on microbes for development. These theories highlight the necessity for manipulation studies including a greater variety of mammalian species with gut microbiotas that differ in terms of phylogenetic signal.

Conclusion

The mammalian gut microbiota is a crucial component of the host's biotic environment and is susceptible to host modification. Interacting with gut microbiota has altered the environment in which mammals adapt, leading to a wider range of dietary habits and incorporating the gut microbiota into the host's ability to change physical characteristics by enhancing signals from the surroundings and internal growth processes. The identification of positively selected genes is occurring at an ever-increasing rate as a result of the rapid expansion of genomic data and the availability of increasingly powerful analysis methods. Based on the data that is currently available, it would appear that the majority of these genes are associated with a small number of functional domains. The identification of positively selected genes is occurring at an ever-increasing rate as a result of the rapid expansion of genomic data and the availability of increasingly powerful analysis methods. Based on the data that is currently available, it would appear that the majority of these genes are associated with a small number of functional domains. It is possible that the identification of genes that have been positively selected, particularly those that are associated with the development of gut microbiota, could provide molecular proof for the extraordinary strength of the selection pressure that has been driving the evolution of our species. Furthermore, there are still numerous unresolved inquiries concerning the ways in which the evolutionary effects of gut microbiota are observed throughout the mammalian phylogeny. It is also unclear how these evolutionary effects are influenced by and contribute to the specificity of relationships between hosts and gut microbial lineages. Additionally, it remains uncertain if any mammalian species have managed to overcome or avoid dependence on gut microbiota for metabolism or signals to initiate adaptive phenotypic plasticity. While a considerable number of genes have been recognized as subjects of positive selection throughout human and/or primate evolution, these probably represent only a small portion of the total. The inclusion of more genes and the mapping of gene sequence variations to functional changes may lead to the widespread adoption of studying positively selected genes as a primary method for understanding human biology and disease.

References

1. ^*Kikuta, H.; Laplante, M.; Navratilova, P.; Komisarczuk, A.Z.; Engström, P.G.; Fredman, D.; Akalin, A.; Caccamo, M.; Sealy, I.; Howe, K. Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. Genome research 2007, 17, 545-555.*
2. ^*Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. Journal of molecular biology 1990, 215, 403-410.*
3. ^*Bernt, M.; Donath, A.; Jühling, F.; Externbrink, F.; Florentz, C.; Fritzsch, G.; Pütz, J.; Middendorf, M.; Stadler, P.F. MITOS: improved de novo metazoan mitochondrial genome annotation. Molecular phylogenetics and evolution 2013, 69, 313-319.*
4. ^*Ranwez, V.; Harispe, S.; Delsuc, F.; Douzery, E.J. MACSE: Multiple Alignment of Coding SEquences accounting for frameshifts and stop codons. PloS one 2011, 6, e22594.*
5. ^*Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R. Clustal W and Clustal X version 2.0. bioinformatics 2007, 23, 2947-2948.*
6. ^*Kumar, S.; Tamura, K.; Nei, M.; Lewis, P.O.; Lewis, L.A. MEGA: molecular evolutionary genetics analysis, version 1.02. Systematic Biology 1995, 44, 576-577.*
7. ^*Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology and evolution 2007, 24, 1586-1591.*
8. ^*Yang, Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Molecular biology and evolution 1998, 15, 568-573.*
9. ^*Ahmad, H.I.; Afzal, G.; Iqbal, M.N.; Iqbal, M.A.; Shokrollahi, B.; Mansoor, M.K.; Chen, J. Positive selection drives the adaptive evolution of mitochondrial antiviral signaling (MAVS) proteins-mediating innate immunity in mammals. Frontiers in Veterinary Science 2022, 8, 814765.*
10. ^*Ahmad, H.I.; Khan, F.A.; Khan, M.A.; Imran, S.; Akhtar, R.W.; Pandupuspitasari, N.S.; Negara, W.; Chen, J. Molecular evolution of the bactericidal/permeability-increasing protein (BPIFA1) regulating the innate immune responses in mammals. Genes 2022, 14, 15.*
11. ^*Ahmad, H.I.; Afzal, G.; Sadia, S.; Haider, G.; Ahmed, S.; Saeed, S.; Chen, J. Structural and evolutionary adaptations of Nei-like DNA glycosylases proteins involved in base excision repair of oxidative DNA damage in vertebrates. Oxidative medicine and cellular longevity 2022, 2022.*
12. ^*Buchan, D.W.; Jones, D.T. The PSIPRED protein analysis workbench: 20 years on. Nucleic acids research 2019, 47, W402-W407.*
13. ^*Bagdonas, H.; Fogarty, C.A.; Fadda, E.; Agirre, J. The case for post-predictional modifications in the AlphaFold Protein Structure Database. Nature structural & molecular biology 2021, 28, 869-870.*
14. ^*Obenauer, J.C.; Cantley, L.C.; Yaffe, M.B. Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic acids research 2003, 31, 3635-3641.*
15. ^*Zhang, Y. I-TASSER server for protein 3D structure prediction. BMC bioinformatics 2008, 9, 1-8.*
16. ^*Bairoch, A.; Apweiler, R.; Wu, C.H.; Barker, W.C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.;*

Magrane, M. The universal protein resource (UniProt). Nucleic acids research 2005, 33, D154-D159.

17. ^*Singh, H.; Srivastava, H.K.; Raghava, G.P. A web server for analysis, comparison and prediction of protein ligand binding sites. Biology direct 2016, 11, 1-14.*
18. ^*Zafeiropoulos, H.; Paragkamian, S.; Ninidakis, S.; Pavlopoulos, G.A.; Jensen, L.J.; Pafilis, E. PREGO: a literature and data-mining resource to associate microorganisms, biological processes, and environment types. Microorganisms 2022, 10, 293.*
19. ^*Isserlin, R.; Merico, D.; Voisin, V.; Bader, G.D. Enrichment Map–a Cytoscape app to visualize and explore OMICs pathway enrichment results. F1000Research 2014, 3.*
20. a, b*Consortium, G. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 2020, 369, 1318-1330.*
21. ^*Whelan, S.; Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Molecular biology and evolution 2001, 18, 691-699.*