Research Paper

Adaptive selection in the evolution of programmed cell death-1 and its ligands in vertebrates

Hafiz Ishfaq Ahmad[1], Jiabin Zhou[1], Muhammad Jamil Ahmad[2], Gulnaz Afzal[3], Haiying Jiang[1], Xiujuan Zhang[1], Abdelmotaleb A. Elokil[3,4], Musarrat Abbas Khan[5], Linmiao Li[1], Huiming Li[1], Liu Ping[1], Jinping Chen[1]

[1]Guangdong Key Laboratory of Animal Conservation and Resource Utilization, Guangdong Public Laboratory of Wild Animal Conservation and Utilization, Guangdong Institute of Applied Biological Resources, Guangzhou, Guangdong, China
[2]College of Animal Science and Technology, Huazhong Agricultural University, Wuhan, China
[3]Department of Zoology, The Islamia University, Bahawalpur, Pakistan
[4]Animal Production Department, Faculty of Agriculture, Benha University, Moshtohor, Egypt
[5]Department of Animal Breeding and Genetics, Cholistan University of Veterinary and Animal Sciences, Bahawalpur, Pakistan

Correspondence to: Jinping Chen; **email:** chenjp@giabr.gd.cn
Keywords: PD1, PD-L1, adaptive selection, evolution, vertebrates
Received: December 4, 2019 **Accepted:** January 28, 2020 **Published:** February 11, 2020

ABSTRACT

Programmed cell death-1 (PD-1) and its ligands, particularly PD-L1 and PD-L2, are the most important proteins responsible for signaling T-cell inhibition and arbitrating immune homeostasis and tolerance mechanisms. However, the adaptive evolution of these genes is poorly understood. In this study, we aligned protein-coding genes from vertebrate species to evaluate positive selection constraints and evolution in the PD1, PD-L1 and PD-L2 genes conserved across up to 166 vertebrate species, with an average of 55 species per gene. We determined that although the positive selection was obvious, an average of 5.3% of codons underwent positive selection in the three genes across vertebrate lineages, and increased positive selection pressure was detected in both the Ig-like domains and transmembrane domains of the proteins. Moreover, the PD1, PD-L1 and PD-L2 genes were highly expressed in almost all tissues of the selected species indicating a distinct expression pattern in different tissues among most species. Our study reveals that adaptive selection plays a key role in the evolution of PD1 and its ligands in the majority of vertebrate species, which is in agreement with the contribution of these residues to the mechanisms of pathogen identification and coevolution in the complexity and novelties of vertebrate immune systems.

INTRODUCTION

The activation of mature peripheral B and T cells induces programmed cell death-1 (PD-1), a member of the CD28/CTLA-4 family [1, 2]. PD-1 and its ligands (PD-L1 and PD-L2) maintain peripheral tolerance by negatively regulating antigen receptor signaling [3]. PD-L1 is extensively distributed on non-lymphoid tissues, non-hematopoietic cells, leucocytes and pancreatic islets. PD-L2 expression is restricted to monocytes and dendritic cells (DCs) [4]. PD-L1 and PD-L2 ligation act as a secondary signal to T cells in combination with T-cell antigen receptor (TCR) signaling and results in the co-stimulation of a negative or inhibitory signal [5] that prevents the activation of TCR-mediated T cells and the production and proliferation of cytokines [6]. Recent studies have revealed that the PD-1/PD-L signaling pathway plays a significant role in autoimmunity and that

the abnormal signaling of PD-1/PD-L results in the loss of peripheral tolerance and autoimmune disorders [7]. The expression of PD-L1 is observed on DCs, mast cells, T cells, B cells, macrophages, and nonhematopoietic cells, including astrocytes, vascular endothelial cells, keratinocytes, pancreatic islet cells, and corneal endothelial and epithelial cells. Both PD-L1 and PD-L2 are also expressed on tumor stroma and tumor cells. The appearance of PD-L2 at tumor positions may contribute to T-cell restriction mediated by PD-1 [8].

Both PD-1 and PD-L1 belong to the immunoglobulin (Ig) superfamily and are type I transmembrane proteins. PD-1 contains cytoplasmic domains that possess two tyrosine signal motifs, a transmembrane domain, and an Ig-V-like extracellular domain [2]. PD-L1 comprises a cytoplasmic domain that does not consist of recognized signaling motifs, a transmembrane domain, and Ig-V-like and Ig-C-like extracellular domains [9, 10]. Communication between the extracellular domains of PD-1 and PD-L1 promotes a conformational modification in PD-1, which results in the phosphorylation of the tyrosine-based immunoreceptor switch motif (ITSM) and the cytoplasmic immunoreceptor inhibitory motif (ITIM) by Src kinases [11]. Furthermore, the interaction of PD-1 and PD-L1 can also affect CD80, which may transport inhibitory signals to activated T cells [12]. The activation of PD-1 by PD-L1 changes T-cell activities in various ways, such as cytokine production, survival, the inhibition of T-cell propagation, and the functions of other effectors [13].

The interaction of PD-1 with PD-L1 is an important factor in immune tolerance. Mice lacking PD-1 are susceptible to developing lethal autoimmune cardiomyopathy or lupus-like autoimmune problems due to altered thymic T-cell responses [14, 15], and PD-L1 blockage has been shown to weaken feto-maternal tolerance [16]. These studies reveal an important role of PD-1 and its ligands in immune tolerance at the cellular and molecular levels [17]. It has been revealed that the expression of PD-1 in newly triggered T cells is involved in monitoring the strength of the early T-cell response upon the detection of an antigen [18]. In agreement with the role of PD-1 in modulating T-cell activities, the interruption of the PD1-PD-L1 interaction enhances the immune response toward various pathogens [19, 20].

Additionally, in some invertebrate clades, the genes directly interacting with pathogens and receptor genes are the most frequent targets of positive selection [21, 22]. Early comparative genomic studies recognized immune system processes as mutual goals of natural selection in Drosophila, mammals, primates, ants, bees, and other organisms [23–25] that harbored candidate signatures of selection. This suggests that pathogens, which evoke an immune reaction, maybe potent and steady selection pressures across species [26]. Previous studies on mammals revealed that proteins that interact with pathogens undergo twice as many amino acid variations as proteins that do not [27], and proteins that interact with Plasmodium undergo relatively higher adaptation rates [28]. In the context of conservation genetics, variation analyses at functional genomic regions provide an enhanced understanding of the mechanisms by which inbreeding and population bottlenecks can influence the adaptive potential of endangered species [29]. Early genome-wide studies have focused on single taxonomic ancestries (e.g., mammals) or restricted subgroups of candidate genes across lineages, but the accessibility of several new genomes now permits thorough evaluations across extremely diverse clades to estimate the extent to which specific genes show conserved regions of positive selection over extended evolutionary periods [25]. However, the role of selective pressure in natural populations in driving the diversification of additional aspects (non-MHC) of the immune system, such as innate immunity, remains poorly understood [30, 31]. Evidence has revealed that pathogens are the main selective pressure that drive evolution, and several new genomes now permit comparisons among various lineages to identify the extent to which particular genes undergo positive selection. The purpose of this study is to analyze the genomic sequences of programmed cell death-1 and its ligands in vertebrate species to calculate the selection pressure on these genes, which can contribute to adaptive evolution. Here, we explore the evolutionary routes by thoroughly analyzing these genes from a set of diverse vertebrates, and confer the role of selection and diversification of this gene family. Our results revealed that positive selection acting on PD1, PD-L1 and PD-L2 genes drives adaptive changes for biological functions directly related to immunological tolerance in vertebrates. The analysis in this study is likely to deliver understandings into the functional inference of the gene in the development of vertebrate evolution.

RESULTS

This study was designed to analyze the genomic sequences of programmed cell death-1 and its ligands in vertebrate species to calculate the selection pressure on these genes, which can contribute to adaptive evolution. Our maximum-likelihood phylogenetic analyses of amino acid sequences from 166 vertebrate species revealed that PD1 genes evolved in a shared ancestor of vertebrates. In our study, we selected three genes (PD1, PD-L1, and PD-L2) that are involved in immune tolerance in vertebrates, and we also explored the evolutionary processes, phylogenetic relationships, and

resulting structural and functional characteristics of PD1, PD-L1 and PD-L2 homologs in vertebrates. We analyzed the CDS of the PD1, PD-L1, and PD-L2 genes. We identified the orthologs of human PD1, PD-L1, and PD-L2 through a comprehensive BLASTP search approach. We filtered 202, 179 and 181 orthologs for PD1, PD-L1 and PD-L2 genes to retrieve the sequences in vertebrate species. The sequences were omitted if absent in most of the species or absent in two of the three of the main taxonomic groups and had sequences in less than 10 taxa. Secondly, poorly aligned sequences were screened using a sliding-window comparison method. These CDS were compiled into a complete multiple sequence alignment, which was used as the input for both the construction of Bayesian phylogenetic trees and other subsequent analyses.

Protein domain analysis

To recognize the domains of these proteins, we used the SMART online tool to predict the protein domains. We identified two domains (Ig-V and TM) in PD1 (Figure 1A), three domains (Ig, Ig-like and TM) in PD-L1 (Figure 2A) and three domains (Ig, C2_set-2 and TM) in PD-L2 (Figure 3A). According to the literature, Ig-like domains are present in various protein families, which are related both in structure and sequence. Ig-like domains contribute to various functions, such as the immune system, cell surface receptors, cell recognition, and muscle structure formation [32]. The V-set domains are Ig-like and resemble the antibody variable region. The V-set domains are present in various protein families, including Ig heavy and light chains; T-cell receptors, such as the cluster of differentiation (CD2, CD4, CD80, and CD86); tyrosine-protein kinase receptors; myelin membrane adhesion molecules; and junction adhesion molecules (JAM) in PD1 [33]. C2-set domains are Ig-like domains that resemble the antibody constant domain and are mainly found in the mammalian T-cell surface antigens CD2, CD4 and CD80 and in intercellular and vascular cell adhesion molecules (ICAMs and VCAMs) [34].

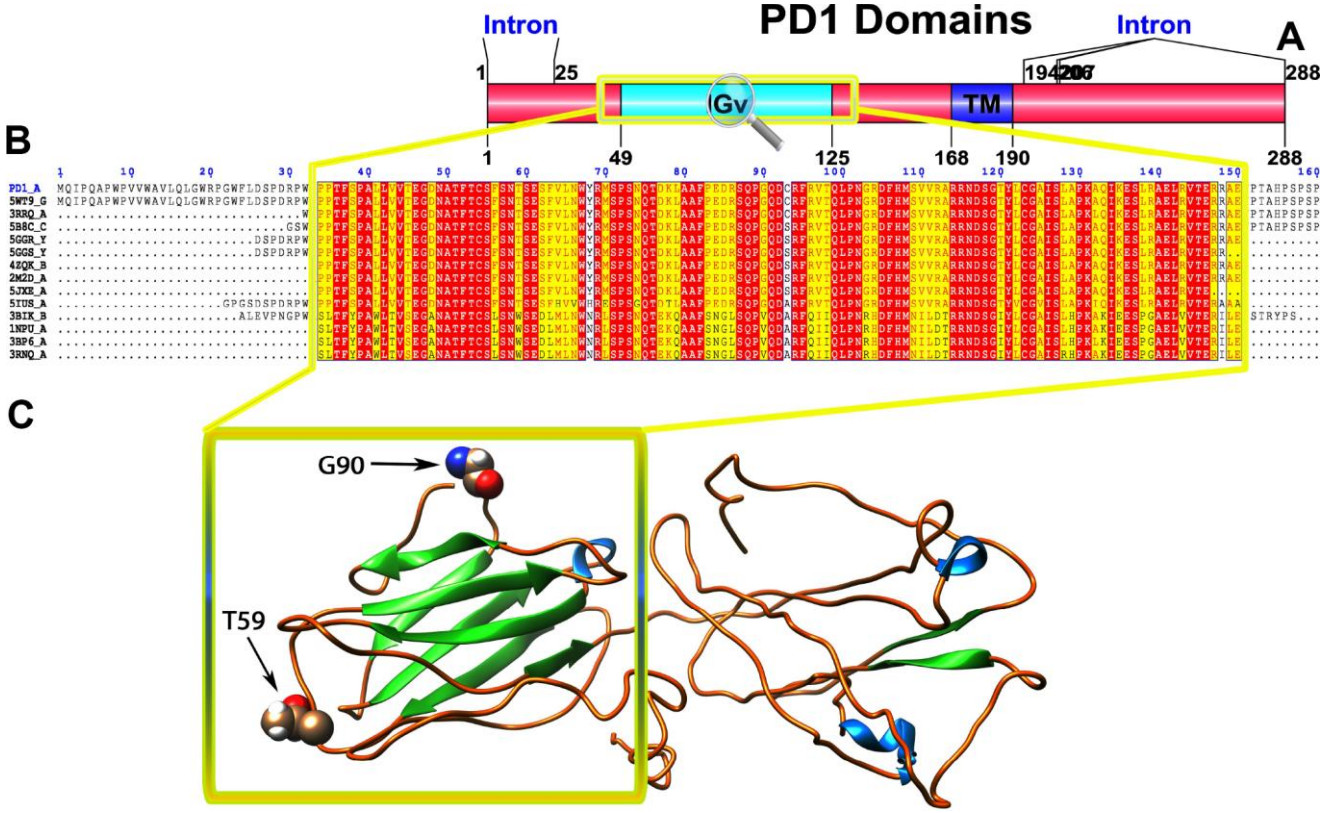


Figure 1. (**A**) Molecular structure of PD1 and Conserved domain analysis of PD1 protein. (**B**) Showing the MSA of the 20 most homologous proteins to PD1 (obtained with a BLAST+ search against the PDBAA database). Known secondary structure elements are displayed for all aligned sequences. Alternate residues are highlighted by gray. Identical and similar residues are boxed in red and yellow, respectively. (**C**) Location of positively selected amino acid sites identified PD1 conserved Ig domain. The crystal structure of human PD1 was used as a reference sequence and positively selected sites were drawn onto the crystal structure using Phyre tool (http://www.sbg.bio.ic.ac.uk/phyre2/html). Two residues identified under selection fall in the immunoglobulin-like domain containing the ligand-binding site. The sites which fall in the region identified as the ligand-binding site and another cluster in a region immediately following the signal sequence.

Positive selection analysis

We used different site models to recognize the genes under positive selection across the vertebrate species. We compared different models for the selected genes in the data set using the phylogenetic tree as input data. We performed a likelihood analysis that compared different models based on ω ratios to identify the codons under positive selection in the corresponding genes. These likelihood models included an additional ω parameter for some fraction of sites and models that do not include the additional ω parameter. The codeml program was used to compute the parameters related to gene selection among 55 species, and a positive selection test was analyzed by the two models for M1a and M2a and for M7 and M8. The results indicated that the test of the PD1 gene in M1a-M2a was not significant with the likelihood test value 2ΔlnL=0 ($p>0.05$), while in M7-M8, the likelihood ratio test (2ΔlnL) was 5.103. The test results of positive selection model M8 were significant ($p<0.05$); the PD1 gene was under positive selection indicating that the M8 was accepted, while M7 was rejected. In the PD-L1 gene, the codon sites were positively selected with P>95% and P>99%, based on the Naïve Bayesian (naïve empirical method based on Bayes, NEB) and Bias (Bayes empirical Bayes experience method, BEB) analysis. However, for the PD-L2 gene, both sets of models M1a: M2a and M7: M8

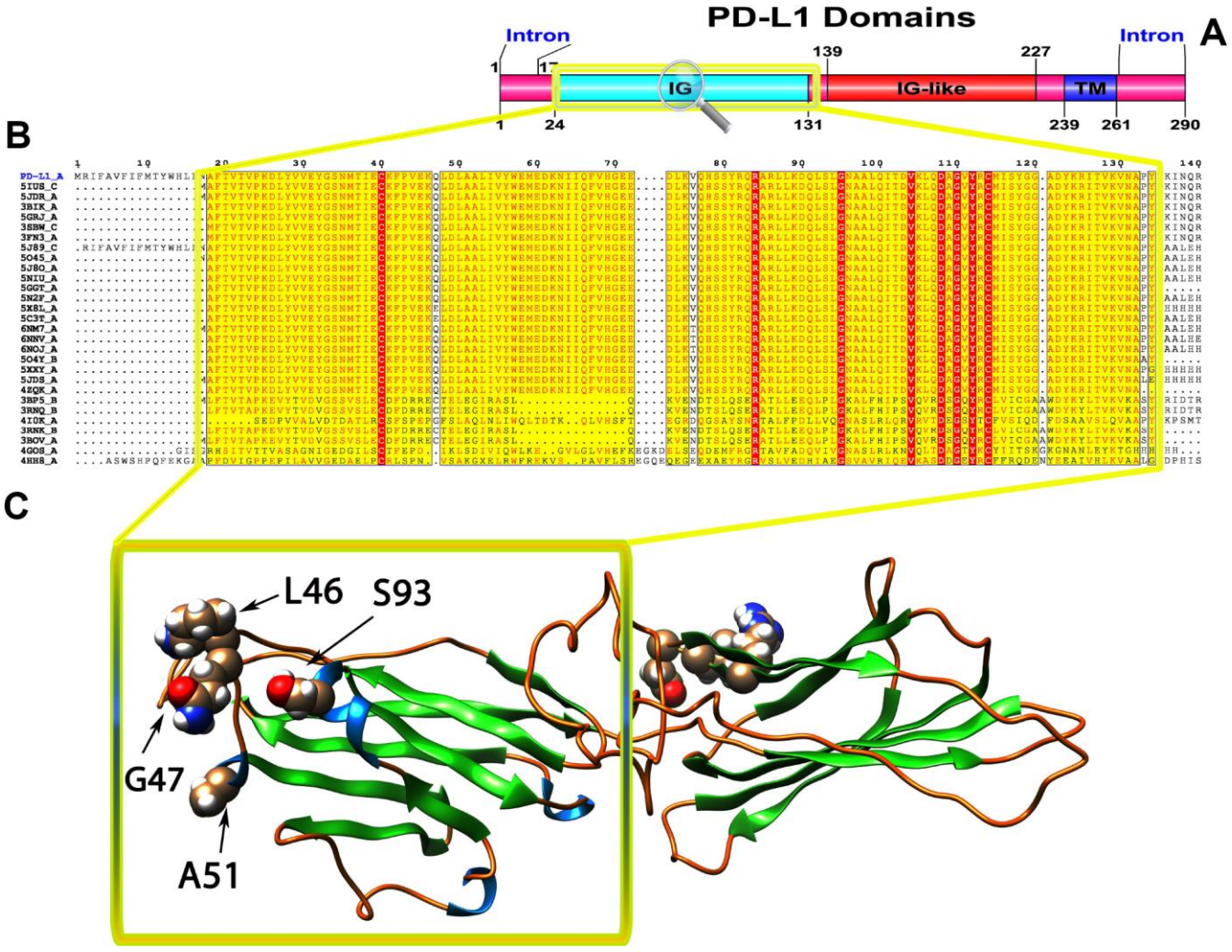


Figure 2. (**A**) Molecular structure and Conserved domain analysis of PD-L1 protein. (**B**) Showing the MSA of the 20 most homologous proteins to PD-L1 (obtained with a BLAST+ search against the PDBAA database). Known secondary structure elements are displayed for all aligned sequences. Alternate residues are highlighted by gray. Identical and similar residues are boxed in red and yellow, respectively. (**C**) Location of positively selected amino acid sites identified PD-L1 conserved Ig domain. The crystal structure of human PD-L1 was used as a reference sequence and positively selected sites were drawn onto the crystal structure using Phyre tool (http://www.sbg.bio.ic.ac.uk/phyre2/html). Four residues identified under selection fall in the immunoglobulin-like domain containing the ligand-binding site. The sites which fall in the region identified as the ligand-binding site and another cluster in a region immediately following the signal sequence.

was highly significant, with 2ΔlnL values of 137.60 and 827.14, respectively (Table 1). We evaluated the global ω values to further conclude the evolutionary signatures of positive selection by using MEME, FEL, and SLAC analyses. Our results revealed robust evidence of positive evolutionary selection for the PD1, PD-L1 and PD-L2 genes in the vertebrates. We applied the Bayesian method to identify the sites under selective pressure by calculating the posterior probabilities for each codon. The sites with higher probabilities are more likely to be under positive selection with ω>1 than sites with lower probabilities. We identified several sites under positive selection in the PD1 protein by using BEB analysis, with most sites exhibiting high posterior probabilities at 99% or 95%. To evaluate the false positive outcomes of PAML, we further confirmed the positive selection using the Selecton server that recognizes adaptive selection at individual amino acid sites in the protein. The MEC model identifies the variations among amino acid exchange rates. As a result, we identified an adaptive selection at various amino acid sites in PD1 (Supplementary Figure 1), PD-L1 (Supplementary Figure 2), and PD-L2 (Supplementary Figure 3). In our analysis, we found that there were few sites in the Ig-V-like domain region in these proteins, and this Ig-V-like domain had evolved. However, the proteins that

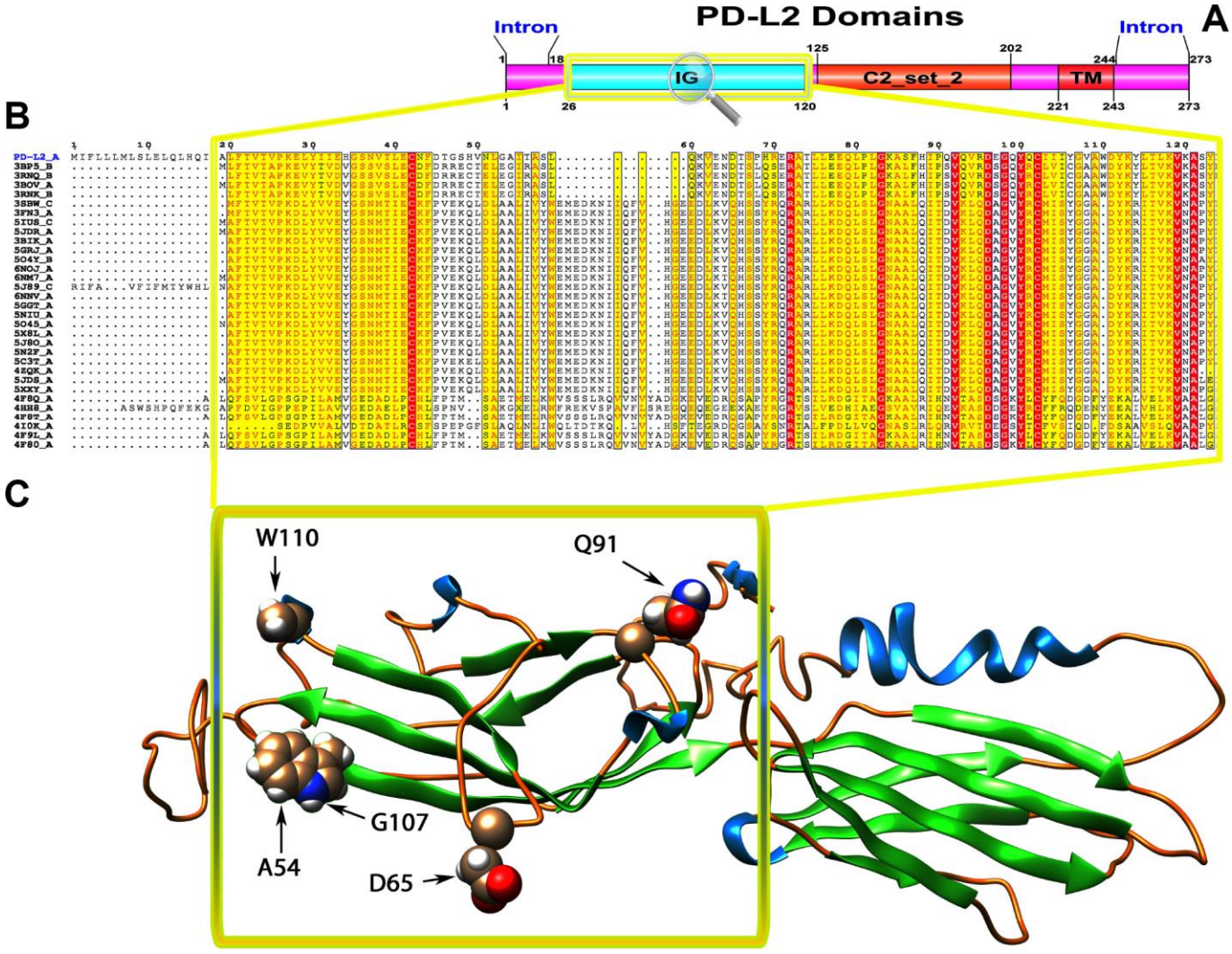


Figure 3. (**A**) Molecular structure of PD1 and Conserved domain analysis of PD-L2 protein. (**B**) Showing the MSA of the 20 most homologous proteins to PD-L2 (obtained with a BLAST+ search against the PDBAA database). Known secondary structure elements are displayed for all aligned sequences. Alternate residues are highlighted by gray. Identical and similar residues are boxed in red and yellow, respectively. (**C**) Location of positively selected amino acid sites identified PD-L2 conserved Ig domain. The crystal structure of human PD-L2 was used as a reference sequence and positively selected sites were drawn onto the crystal structure using Phyre tool (http://www.sbg.bio.ic.ac.uk/phyre2/html). Five residues identified under selection fall in the immunoglobulin-like domain containing the ligand-binding site. The sites which fall in the region identified as the ligand-binding site and another cluster in a region immediately following the signal sequence.

Table 1. Log-Likelihood Values and Test Statistics for PAML Site Models of positive selection.

Gene	Models	Parameter estimates	*lnL*	*LRTs*	Positively selected sites PAML	SLAC	FEL	MEME
PD1	M1a	p1: 0.51480 p2: 0.48520 ω1: 0.21574 ω2: 1.00000	-23577.93					
	M2a	p0: 0.51480 p1: 0.39898 p2: 0.08622 ω0: 0.21574 ω1: 1.00000 ω2: 1.00000	-23577.93					
	M7	P: 0.98018 q: 1.49822	-23414.34	5.103112*				
	M8	p0: 0.96818 p: 1.00330 q: 1.60757 p1: 0.03182) ω: 8.05118	-23411.79					
PD-L1	M1a	p1: 0.44199 p2: 0.55801 ω1: 0.18081 ω2: 1.00000	-21407.94	137.60***				
	M2a	p1: 0.39702 p2: 0.47651 p3: 0.12647 ω0: 0.18153 ω1: 1.00000 ω2: 3.26240	-21339.13					
	M7	P:0.66997 q: 0.64572	-21305.29	131.62***				
	M8	p0: 0.86159 p: 0.77177 q: 0.94085 (p1: 0.13841) ω: 2.57746	-21239.48					
PD-L2	M1a	p1: 0.48081 p2: 0.51919 ω1: 0.20239 ω2: 1.00000	-22727.73	137.60***				
	M2a	p0: 0.35589 p1: 0.38002 p2: 0.26409 ω0: 0.20533 ω1: 1.00000 ω2: 9.06683	-22297.1					
	M7	P:0.56058 q: 0.42402	-22649.48	827.14***				
	M8	p0: 0.73476 p: 0.70098 q: 0.70029 (p1: 0.26524) ω: 7.39740	-22235.91					

The proportion of sites under positive selection (*p*1), or under selective constraint (*p*0), and parameters p and q for the beta distribution. Parameters indicating positive selection are in bold. *p*: significant at 5% level; p: significant at 1% level. Sites potentially under positive selection identified under model M8 are listed according to the human sequence numbering. Positively selected sites with posterior probability .0.9 are underlined, 0.8–0.9 in bold, and 0.5– 0.7 in plain text. The test statistic *2Δl* is compared to a χ2 distribution with 2 degrees of freedom, critical values 5.99, 9.21, and 13.82 at 5%, 1%, and 0.1% significance, respectively. **: significant at 1% level; *: significant at 5% level.

experience positive selection might be conserved and exposed to purifying selection during adaptive evolution. This conservation derived from the evolution of each amino acid/nucleic acid position was identified by using the ConSurf server. The color-based representation of the ConSurf server allows the identification of regions with strong and weak conservation in the structure of the PD1, PD-L1 and PD-L2 proteins (Figures 1B, 2B and 3B, respectively). We found that several conserved residues had masked the signals of selection, and the residues that were buried or exposed according to the neural network algorithm were determined to be under positive selection on variable positions in the PD1, PD-L1 and PD-L2 proteins (Supplementary Figures 4–6, respectively).

To gain insight into the probable intermolecular interactions of these positively selected regions of the PD1 proteins with conserved functional domains, we generated 3D models of the proteins with a reported complex between the Ig region and the protein-coding region, which is the target of Ig-like proteins, as a models for homology modeling, assuming that this conserved Ig region could interact correspondingly with its targets. The 3D protein structure showed that T59 and G90 were the main PD1 protein-protein interaction residues under positive selection (Figure 1C), and L46, G47, A51 and S93 were the main interacting residues that were detected under strong selective pressure in the PD-L1 protein (Figure 2C). The residues A54, D65, Q91, G107 and W110 were the main interacting residues found under selection in the PD-L2 protein (Figure 3C). Motif analysis by MEME identified various species in our data set that shared high conservation in motifs 1 to 5 but differed in motif 1, which we determined was lacking in the PD1 protein of birds (Figure 4 and Supplementary Figure 7). Within the same subfamily, individuals had similar motif distributions, such as PD-L1 and PD-L2 lacking motif 5 in both avian and amphibian species, demonstrating that individuals of the same subfamily may have similar functions. All motifs were found in all protein sequences from diverse vertebrate species, excluding some of the mammalian species, including *Chlorocebus sabaeus*, which lacks motif 1 and motif 5 in PD1, *Monodelphis domestica*, which lacks motif 5 in PD1, and *Gorilla gorilla* and *Castor canadensis*, which lack motif 5 in PD-L1 (Figure 4 and Supplementary Figure 8). Different motif patterns were observed in PD-L2, where motif 5 or motifs 2-5 were missing in *Mus musculus*, *Canis familiaris* and *Oryctolagus cuniculus* (Figure 4 and Supplementary Figure 9). The lack of motifs in various species signifies the divergence of gene structural features concerning exon-intron relationships. These analyses revealed that the differences in motif distribution in PD proteins of vertebrate species might have diverged from the functions of these genes during adaptive evolution.

Lineage-specific selection analysis

The codon-based selection model can only classically confer positive selection signals when particular codons are under selection pressure in several lineages. We used an adaptive branch-site random effects likelihood (aBS-REL) model to relax this hypothesis to calculate the selection probability and identify selection restricted to specific lineages independently at each subgroup of the phylogeny. To further validate our site-model results, we used aBS-REL for each gene to identify the lineages that underwent positive selection during adaptive evolution. We noticed that the genes recognized as being under positive selection by BUSTED across mammalian lineages were also under significant positive selection in additional lineages according to the aBS-REL model (Figure 9; Supplementary Tables 1–3). Clades within avian, mammalian, and reptilian lineages showing considerable signals for positive selection ($p<.05$) were identified using the branch-site-REL (BSR) program executed in the Data Monkey Web Server. PD1 exhibited strong signatures of positive selection at various nodes of its mammalian and avian clades, including *C. sabaeus*, *Rhinopithecus roxellana*, *Acinonyx jubatus*, *Ovis aries* and *Meleagris gallopavo*, respectively (Figure 9). PD-L1 showed significant selection within a broad range of phylogeny. Positive selection signals for PD-L1 were identified at several nodes in two major clades: *O. aries*, *Pantholops hodgsonii*, *Sus scrofa*, *C. sabaeus*, *Macaca fascicularis*, *C. canadensis*, and *M. domestica* in the mammalian clade and *Apteryx australis mantelli*, *Falco peregrinus*, *Coturnix japonica*, *Taeniopygia guttata*, and *G. gallus* in the avian clade (Figure 9). However, for PD-L2, we obtained surprising results: all clades in the dataset showed strong signals of positive selection in vertebrate lineages (Figure 9).

Gene enrichment analysis

We used EnrichNet, which is a network-based gene enrichment analysis program, and gene IDs from the Ensembl database (PD1; ENSG00000188389, PD-L1; ENSG00000120217, and PD-L2; ENSG00000197646) as queries for the functional network analysis. We used the immune system regulation activity class of the Gene Ontology (GO) functional catalog to map the synteny genes and obtained networks including PD1, PD-L1, and PD-L2 as functionally associated genes (Figure 5). We selected the regulation of the immune response group because it includes the PD1, PD-L1 and PD-L2 genes in most of the functional databases. We expanded our enrichment analysis to various databases and all functional classes by generating functionally linked gene networks in the ConsensusPathDB obtained from the EnrichNet analysis. Using this technique, we identified an interaction network with 107 interactions and 62 unique nodes in the PD1 protein (Figure 5). The interaction between PDCD1 and PTPN6 was maintained through a transcription factor encoded by PTPN, which is the py223 gene that interacts with the PTPN11 and dmbx1 genes (Figure 5). We found 125 interactions and 55 unique nodes physically interacting with the PD-L1 protein (Figure 5). PD-L2 was identified in the conserved synteny map with 56 interactions and 22 distinct nodes, but its function was associated with KTN1, GALNT15 and TMEM147 after the enrichment analysis (Figure 5).

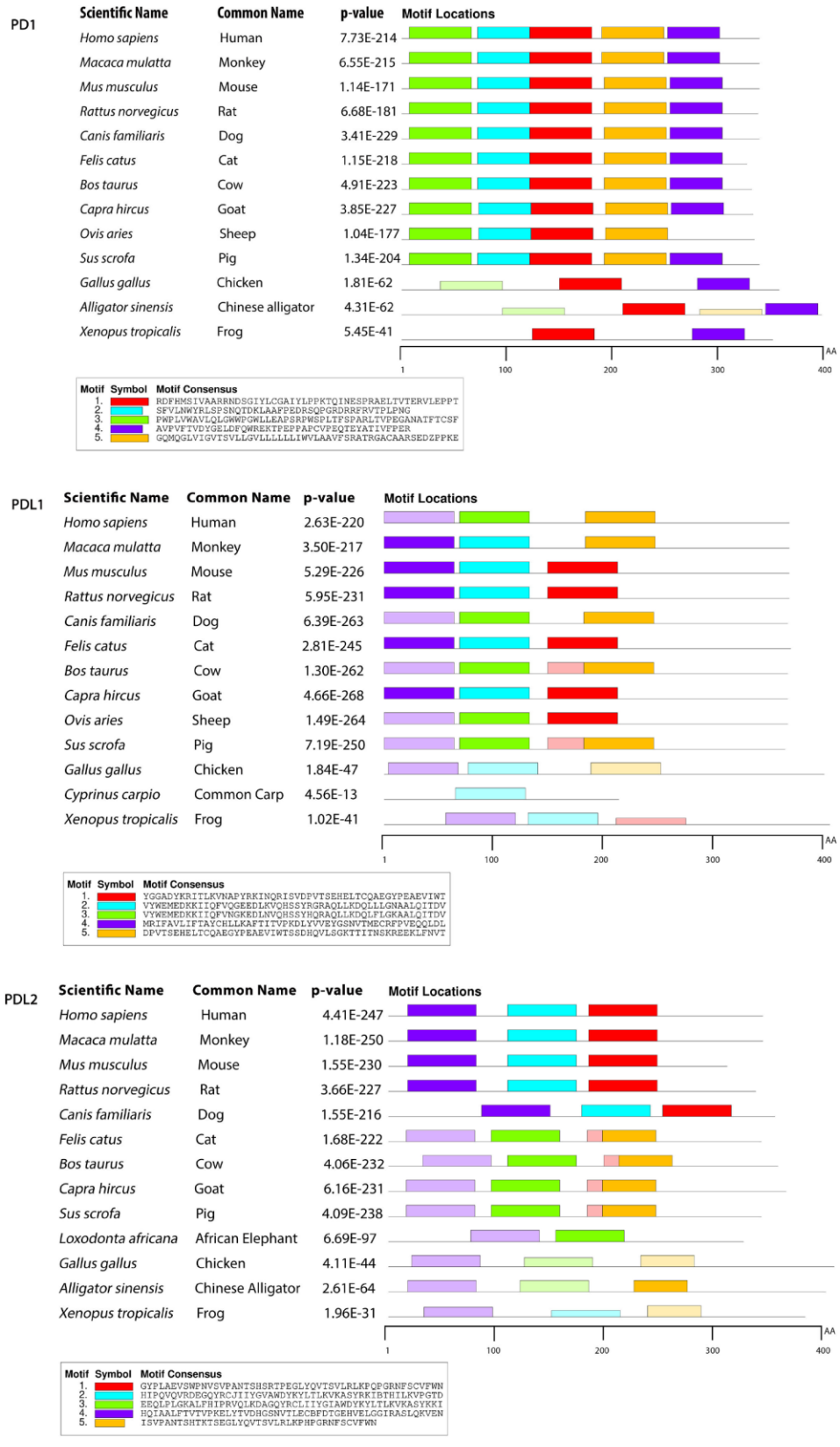


Figure 4. Motif distribution of PD1, PD-L1 and PD-L2 genes in representative vertebrate species. Motifs of these genes from representative species from each group are predicted using MEME suite (http://meme-suite.org/) based on amino acid sequences. All sequences are separated by 5 conservative motifs with colors, including motif 1 (red), motif 2 (cyan), motif 3 (green), motif 4 (purple) and motif 5 (brown).

Bioinformatics analysis

Homologous sequence analysis of PD1, PD-L1 and PD-L2 across several clades was used to predict related structural topographies of these proteins. The structure of the human PD1, PD-L1 and PD-L2 genes were used as a model for further analyses. In the case of PD1, there was higher variation among the homologous sequences than among the sequences corresponding to various domain regions. The prediction of solvent accessibility, coils and turns in the corresponding domain region (Figure 6A), regardless of hydrophobic clustering development, suggested that protein-protein interfaces were developed by the contribution of specific parts of this region, which are important for the interaction of PD1 with its targets. The possible development of these clusters was revealed by HCA in regions consistent with the Ig-like conserved domain (Figure 6B). The secondary structure prediction results display numerous helices and a reduced number of strands in the PD-L1 and PD-L2 protein sequences. Some of these secondary structures were confined within the conserved regions of PD-L1 and PD-L2, which supports the hypothesis that β-strands and helices are the most rigid types of secondary structures and that mutations might disrupt the secondary structure of proteins.

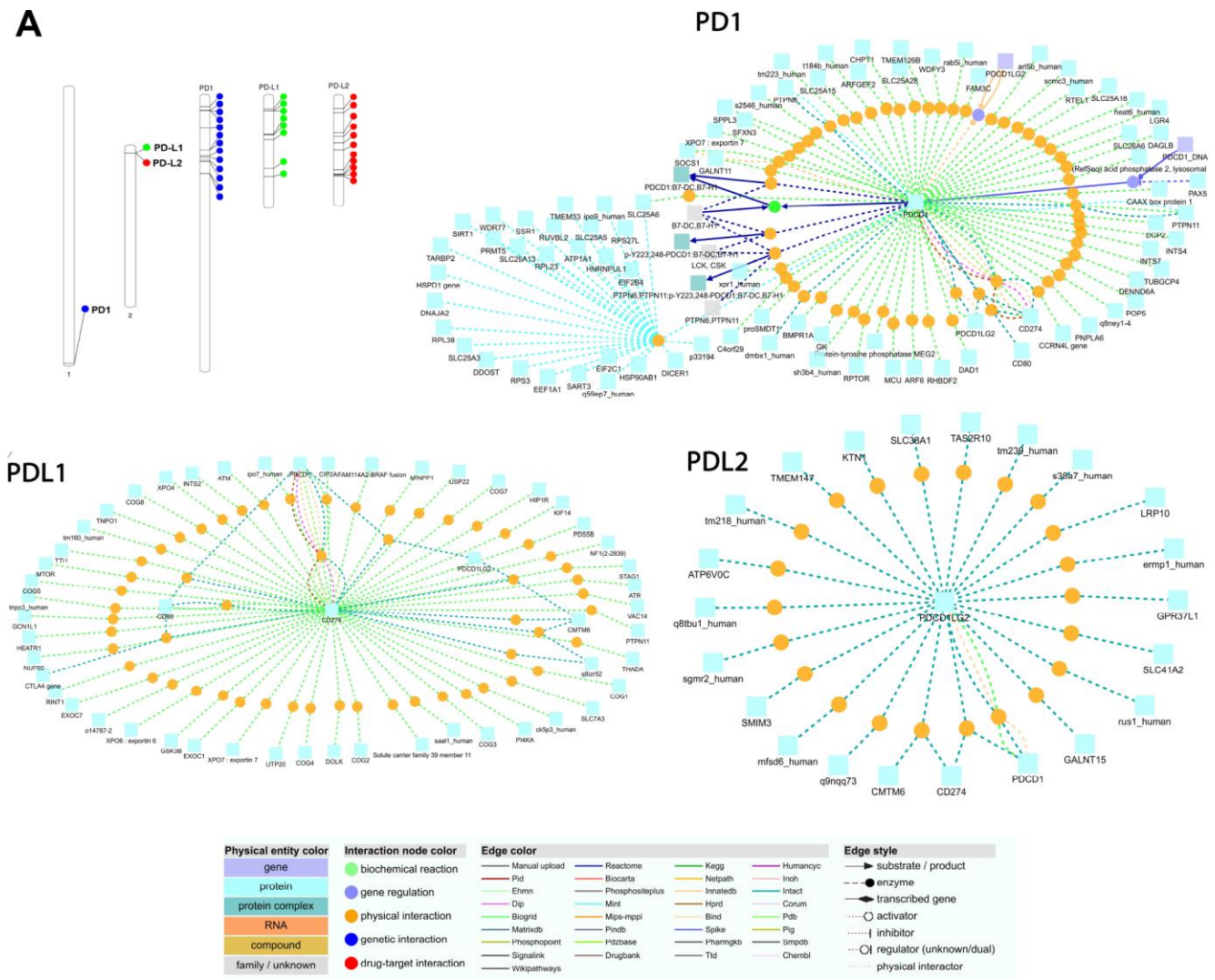


Figure 5. (**A**) Chromosomal locations and positively selected sites of PD1, PDL1, and PDL2 genes. The chromosome number is indicated above each bar. The chromosome size is indicated by its relative length using the information from NCBI. The scale of the chromosome is millions of base pairs (Mb). Functional interaction network of PD1, PDL1 and PDL2 genes generated by the visualization environment of Consensus Path DB meta-database, after conserved synteny and functional enrichment analysis. The network of the PD1 gene contains 107 interactions and 62 physical entity nodes. The network of PDL1 contains 125 interaction and 55 physical entity nodes. The network of PDL2 contains 56 interaction and 22 physical entity nodes. Each node represents a physical entity (gene, protein or compound). Each edge represents an interaction.

Coevolution analysis

The structural and functional features of the positively selected residues were further examined through co-evolution analysis by identifying the coordinated contacts among these residues. This was performed by the identification of other residues that have co-varied with positively selected residues during evolution. This coevolving relationship among various amino acid sites within a protein could be the result of their structural or functional interactions. Hence, we conducted a coevolution analysis using homologs of PD1, PD-L1 and PD-L2 as inputs and identified various coevolving residue pairs that were identified as being under positive selection in the former analyses. A diagram showing the networks was built to recognize a link of significantly associated residues (Figure 7A–7C). We have identified that amino acids with a greater number of co-evolutionary contacts likely evolve more steadily than those with fewer co-evolutionary contacts. We also

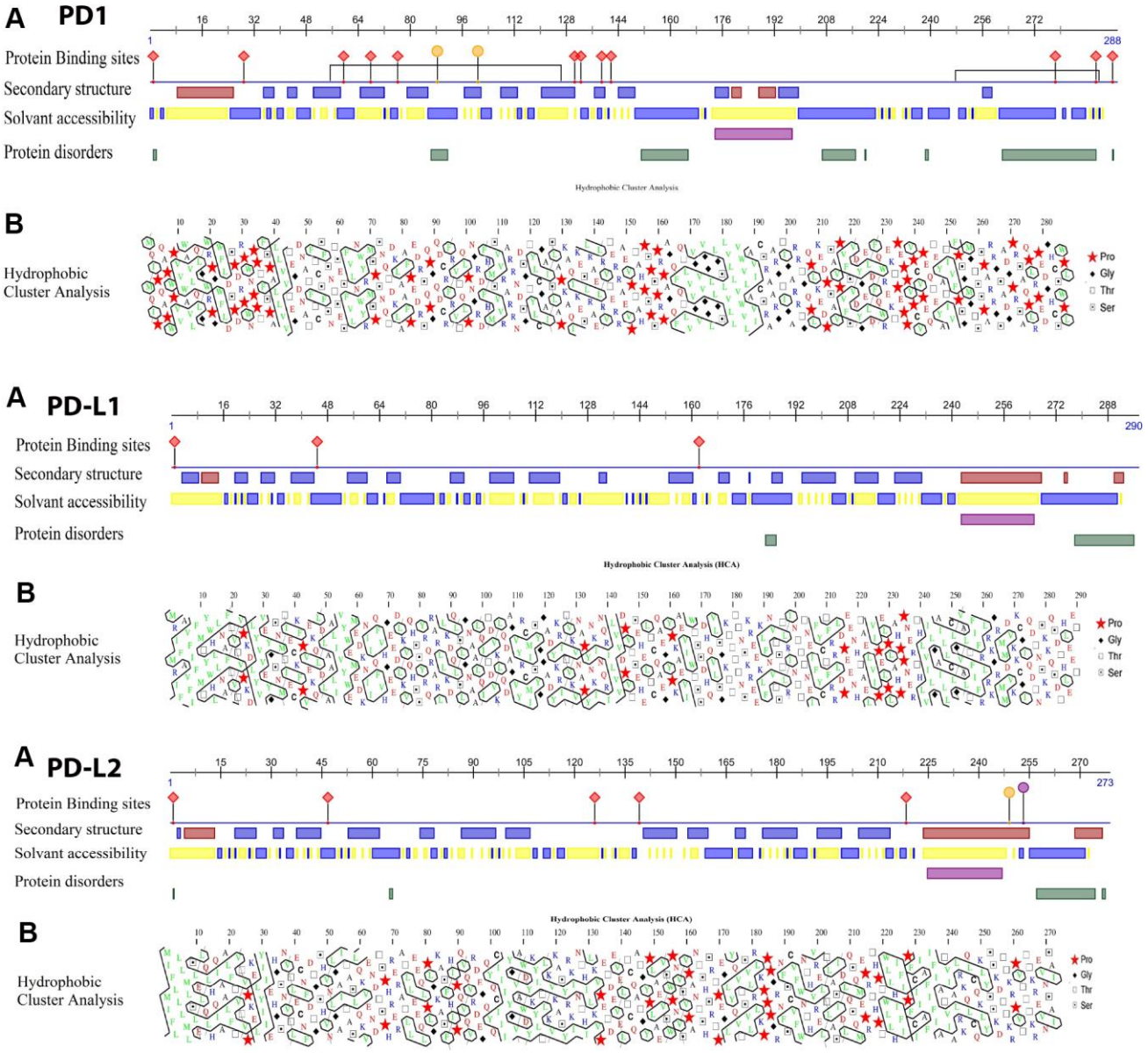


Figure 6. (**A**) Proteins analysis showing the results of the binding site, solvent accessibility and protein disorder predictions in the human PD1, PD-L1 and PD-L2 sequences. (**B**) Hydrophobic cluster analysis (HCA) plots of the human PD proteins. HCA plots were constructed with the HCA 1.0.2 program. HCA uses the standard one-letter amino acid abbreviations except for four amino acids, as shown in the key. Hydrophobic residues are outlined. Clusters of hydrophobic residues are usually associated with regular secondary structures (α helices or β sheets). Zigzagging vertical lines of hydrophobic residues indicate alternating hydrophobic and non-hydrophobic residues, typical of exposed β sheets (for example, β2, β3, β5, and β6). Continuous hydrophobic clusters are more common in internal β sheets.

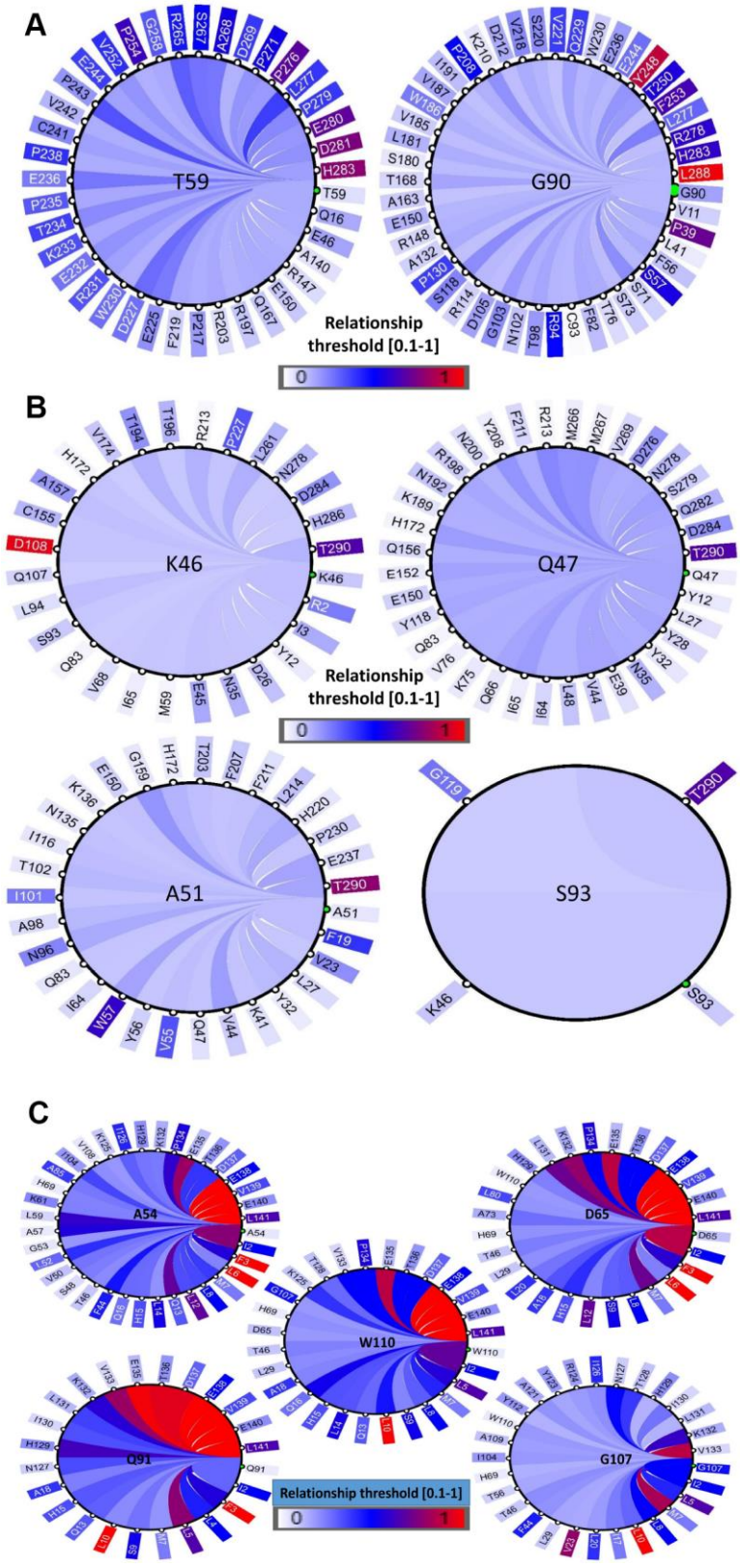


Figure 7. Coevolution analysis of positively selected conserved domain residues. The circular relation diagram centered on the residues with their top co-varying residues at cutoffs (**A**) PD1, (**B**) PD-L1 and (**C**) PD-L2. Labels on the diagram represent amino acid residues and their positions in the protein sequence. Colors of the arcs represent covariance scores between two given positions. Colors of the arcs represent covariance scores between two given positions.

determined that the residues with strong connectivity in the network, such as Y248 and L288 in PD1, D108 and T290 in PD-L1, and F3, L6, and L10 in PD-L2, were the residues with higher conservation, respectively (Figure 7A–7C). The positively selected residues were present in the nodes of a subnetwork that was constrained to amino acid residues present in the conserved Ig-like and Ig-V-like domains in the protein. Multidimensional scaling (MDS) scatterplots of co-varying residues in human PD1, PD-L1 and PD-L2 exhibited coevolving probabilities according to Pearson correlation (r). From this covariance analysis, we determined the distances and contacts of positively selected residues in the protein domains and clusters of functionally related residues (Supplementary Figure 10). The 3D viewing pane provides interactive zoom and rotation capabilities and labels selected residues. These outcomes support the proposal that the PD protein regions conforming to these conserved domains are structural-functional units.

Protein structure quality analysis

These positively selected residues were present in the loop, which can dock into the Ig-binding pocket. The extent of conservation of these amino acids was diverse, with only W110 and E136 being comparatively conserved in PD-L2 (Supplementary Figure 6). The low degree of conservation among the other residues can be conferred to neutral variations along with these sites according to the positive selection/purifying analysis (Supplementary Figures 1–3). Moreover, the complete quality of the predicted structures was evaluated by ERRAT, ProSA and VERIFY3D values. The Ramachandran plots were used to check the backbone conformation angles for the respective residue in the modelled protein crystal structures, which showed the empirical scattering and calculated energies of the residue coordinate system to display either plot of the conformation detected in the databank of identified 3D models or outlines or steric measures as a function. We observed that most of the residues were found in the allowed region of the graph that represented the conformational accuracy of all predicted models. For PD1, PD-L1 and PD-L2, the number of residues in the favored area was 95.9, 96.5, and 95.9%; the number of residues in the allowed region was 3.6%, 3.2%, and 3.6%; and the number of residues in the outlier region was 0.5% for all predicted structures, respectively (Supplementary Figure 11). The z-values show the complete model excellence of the predicted structures with values of -7.2, -7.8, and -6.7, respectively (Supplementary Figure 11). Additionally, the compatibility of all predicted structures and their correctness was validated using the VERIFY3D software package. The dynamics of the PD1, PD-L1 and PD-L2 protein model structures were computed using the Gaussian network model (GNM), which revealed that the residues that showed greater mobility were part of the binding pocket and could function as the protein-protein interaction region. The cross-correlation plots represented the variation in residues and their physical behavior. The predicted plot outcomes were analyzed based on the colors blue, dark red, yellow and cyan. We identified that the entirely correlated pairs were shown in dark red, whereas the anti-correlated pairs are were shown in blue. Additionally, the uncorrelated and moderately correlated regions were colored cyan and yellow, respectively (Supplementary Figure 11).

mRNA expression of the PD1, PD-L1 and PD-L2 genes

We investigated the mRNA expression patterns of the PD1, PD-L1 and PD-L2 genes in five vertebrate species, including *M. javanica, A. albiventris, G. gallus, A. schrenckii* and *S. crocodilurus*. We examined the mRNA expression profiles in heart, liver, spleen, lung, kidney, pancreas, brain, testis, and ovary and muscle tissues from male and female adults. We noticed that PD-L1 had increased expression in almost all species and that PD1 exhibited expression only in chicken tissues. PD-L2 showed increased expression in pangolin and fish in almost all tissues with some exceptions (Figure 8). The expression of PD-L1 was high in almost all tissues, indicating tissue-specific expression. The PD1, PD-L1 and PD-L2 genes were highly expressed in the heart, liver and spleen, indicating their predicted functions in immune development and the prevention of autoimmune disorders in organisms. These PD genes that were highly expressed in vertebrate tissues or organs may have important functions in signaling pathways that play a significant role in autoimmunity, and abnormal signaling in these pathways may result in the loss of peripheral tolerance. Hence, the tissue specificity of the PD genes identified here may be useful sources for further probing their biological functions in immune tolerance at the cellular and molecular levels.

DISCUSSION

Key evolutionary changes occur in the genomes of animals via divergence, the assimilation of genetic data from independent lineages, gene duplication, and epigenesis. Horizontal gene transfer and genome duplications lay the foundations for all main molecular mechanisms of adaptive immunity [35]. Previous studies on the evolution of immune genes in birds mainly focused on the coevolution of host-pathogen hotspots, including the MHC and TLRs [36, 37]. The positive selection signatures we identified were consistent with the adaptive selection here. PD-L1, PD1 and PD-L2

across vertebrate genes are first involved in regulating lymphocyte activation, immune system function, the promotion of T regulatory cell function and development, the progression of autoimmunity and immune tolerance as targets of positive selection. We identified numerous positions in three genes under positive selection in mammals, reptiles and birds (Table 1). We found that the genes with significantly increased overall distinctive expression in these lineages resulted in standards paralleled to those not under selection pressure or positively selected only in mammals. Across vertebrate clades, our results indicate that pathogens might be a stable selective pressure. Previous phylogenetic studies revealed uncertainties related to the first vertebrate phylogeny change in IRF, i.e., where we find an adaptive immune system. It was proposed that in the beginning, the IRF family divided into two branches that are found in all cnidarians and bilaterians [38]. Although facts and evidence continued to amass, the four antecedents probably advanced into the four classes of vertebrates, IRF1-G (IRF2IRF1), IRF4-G (IRF4, IRF9, IRF8, IRF10), IRFs, IRF3-G (IRF7, IRF3), and IRF5-G (IRF6, IRF5), following 2-fold whole-genome duplications [38, 39]. In the evolution of the PD proteins to know the comparative characters of positive selection, we used two approaches to compare the difference at the codon level: an evolutionary model (M8) that allowed positive selection, and another using the MCMC evolutionary model implemented in MrBayes implemented in the Selecton server [40, 41]. ω values for the individual site were calculated in both circumstances. Our consequences demonstrate the conservation of the Ig domain of PD-L1, PD1, and PD-L2 coding sequences attained after protein alignments MAFFT (Figures 2B, 1B, and 3B). The fallouts allowing for only the PD mature proteins have revealed that the normal ω values were 0.18153, 0.21574 and 0.20533 using M2 and M1 evolutionary models, correspondingly.

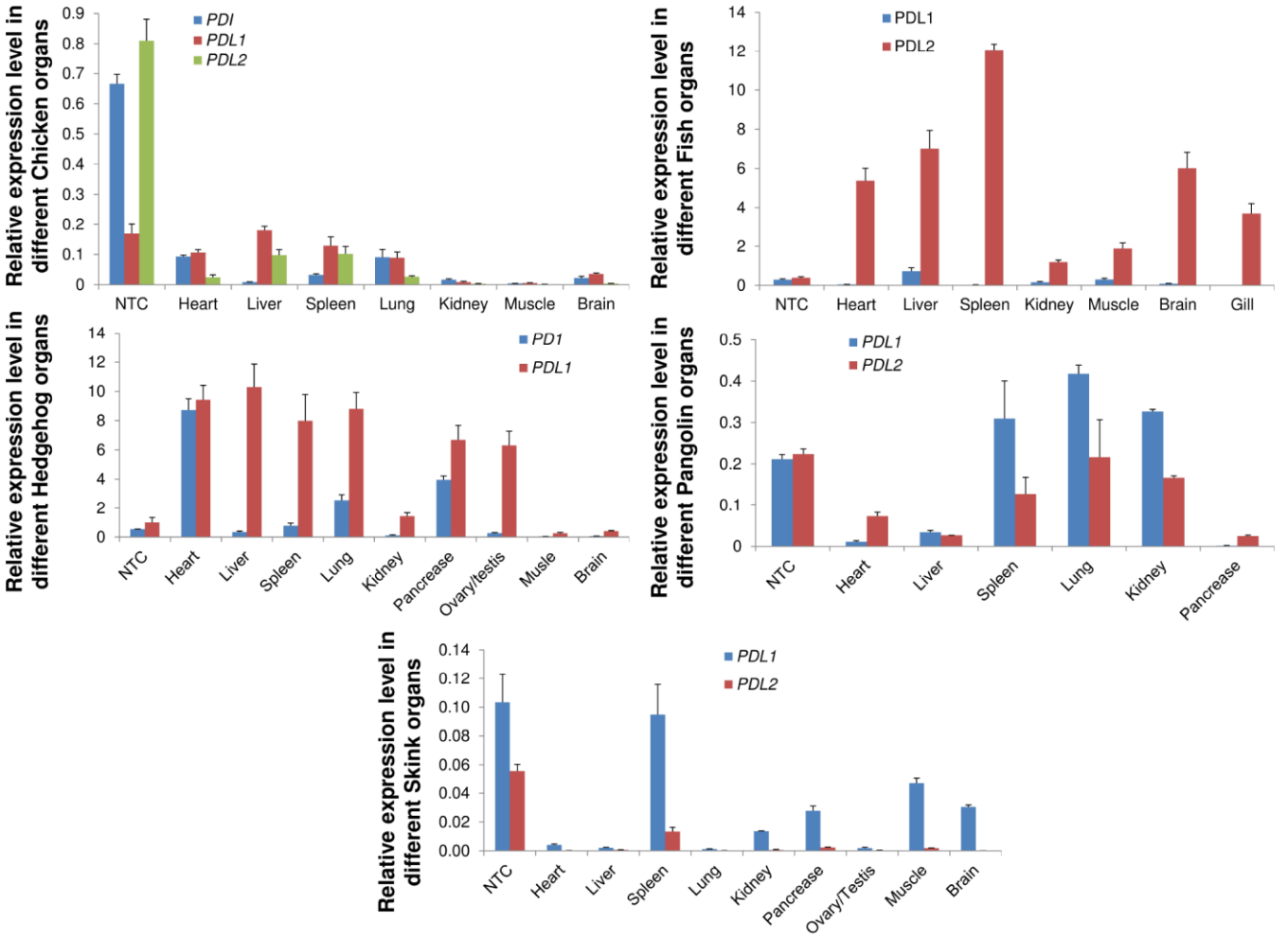


Figure 8. qRT-PCR analysis of PD1, PD-L1 and PD-L2 genes in different animal tissues. Expression patterns of genes in different tissues were examined. Heart, liver, spleen, lungs, kidney, pancreas, brain, were used for quantitative reverse transcription (qRT-PCR) polymerase chain reaction. Transcript levels are expressed relative to that of beta-actin. NTC: negative control.

These results revealed that the protein changing gradually in the regions under purifying selection undergo non-identical switches, which are detrimental to health and thus have low chances of fixation during evolution [42]. The subsequent process favored positive selection and identified the amino acid residues with a ω>1 (Table 1). Rendering cross-correlations analysis between filtrate variations, this movement is connected in the section of the protein chain, which protects amino acids L288 and Y248 in PD1, D108 and T290 in PD-L1, and, L6, F3 and L10, are the filtrates with higher management, correspondingly (Figure 7A–7C). When the study deliberated the area consistent to the N-terminal, three positively selected sites L6, F3 and L10, were found in PD-L2 using the M8 evolutionary model (Figure 3C) with an ordinary *dN/dS* value of 7.39740. Our outcomes reveal that several sites in other proteins, which are under substantial positive selection have been developing more swiftly than the mature protein [43–45]. As a result, the dynamic selection forces its change which concerns to improve the protein secretion efficacy, which is true in case of PD-L2 protein (Figure 3C), which is unlike than the matured protein [46, 47].

We performed a branch-site test to determine the specific branches under selection in vertebrate clades, and we found few branches of mammalian clades under selection in the PD1 gene (Figure 9). In the case of PD-L1, positive selection was identified in mammalian and avian clades (Figure 9). However, surprising results were identified during the analysis of the PD-L2 gene, in which we found that there is positive selection in most of the branches of vertebrate clades in the data set. (Figure 9). Because branch-site analysis can lead to false implications of positive selection due to multi-nucleotide mutations, we further validated our results through the *aBS-REL* model, and similar patterns of selection were observed between the *aBS-REL* and site models. The results suggest that the overall selection patterns we

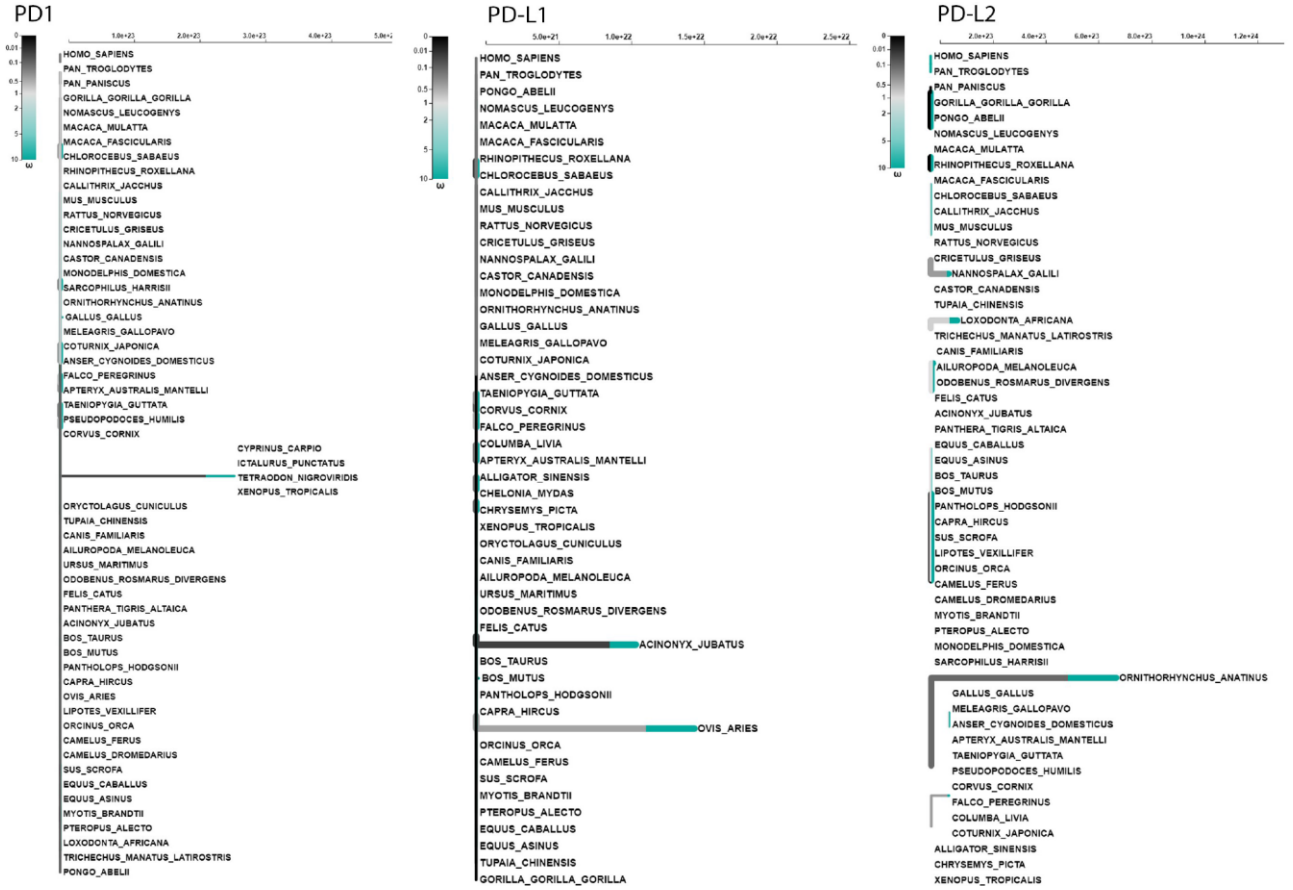


Figure 9. Adaptive branch-site REL test for episodic diversifying selection in PD1, PD-L1 and PD-L2 genes. The phylogenetic tree scaled on the expected number of substitutions/nucleotides. The hue of each color indicates the strength of selection, with primary red corresponding to ω > 5, primary blue to ω = 0 and grey to ω=1. The width of each color component represents the proportion of sites in the corresponding class. Thicker branches have been classified as undergoing episodic diversifying selection by the sequential likelihood ratio test at corrected ≤0.05.

observed were accurate with the alternate analysis. Generally, the PD1 protein in vertebrate lineages revealed no evidence of positive selection that suggested a constant selection pressure hampering genetic variation, mainly in the avian clade. Nearly a complete lack of positive selection, as observed by the M8 evolutionary model, has been entirely supported by *LRT* [48]. This may be associated with the concept that the evolutionary history of the PD1 gene has occurred without gene duplication events. Gene duplication is one of the evolutionary approaches that permit versatile advancement in genomes. It has been revealed for other proteins that positive selection happens after a duplication event that proposes an unwinding of the selective pressure supporting genetic variation [49, 50]. This relaxation was missing not only in avian but also in other vertebrates' lineages during PD1 evolution, conferring to the Bayesian phylogenetic approaches (Supplementary Figure 10). Overall, these findings suggest that PD1 molecular evolution has classically been determined by purifying selection. The synteny of PD1, PD-L1, and PD-L2 is conserved among vertebrate clades, and the variations found in some motifs could be described by the varying evolutionary backgrounds of these genes for homologous arrangements in mammals and other vertebrates. A comparing quality set enrichment analysis shown a functional relationship between a subset of the conserved synteny genes, PTPN6, KTN1, GALNT15 and TMEM147 might be supported through the transcription factor, encoded by the PTPN; py223 gene (Figure 5). The purifying/positive selection investigation did not distinguish positive selection in amino acid residues of PD1, showing that purifying choice, especially within the regions comparing to Ig like domain has driven PD1 molecular evolution. A co-evolution examination showed that residues within the regions comparing to each motif especially co-vary with each other during PD1 evolution.

The structure and function of proteins are reliant on synchronized connections among their amino acid residues. Therefore, the identification of structural characteristics of positively selected amino acid residues could resemble the detected residues that co-vary with each other during evolution. This co-evolutionary relationship among amino acid residues within proteins could be the consequence of their structural or functional interactions. Previous studies on protein coevolution have discovered roles in protein constancy and intermolecular interactions [51–54]. Therefore, the variation in PD1 and its ligands may occur during evolution. Here, we found that Ser93 of human PD-L1 and Gly107 in PD-L2 may have important roles in some of the previously revealed functions by estimating the *dN/dS* of mammalian PD1 sequences. PD1, PD-L1, and PD-L2 have domains that are highly conserved among vertebrates, indicating that their roles are not redundant and that this selective pressure may derive from PD specificity for their immune system regulation. However, proteins that directly interact with other molecules tend to have an increased chance of adaptation to fit each other's evolutionary changes. PD genes may have undergone such co-evolutionary traits in the context of their original function. Therefore, a coevolution analysis was performed via multiple sequence alignments generated for PD1, PD-L1 and PD-L2 homologs, and we found positively selected coevolving residues showing substantial variability (Figure 7). These results suggest that these regions corresponding to domains are structure-function modules within the PD proteins. The homology modelling of the PD protein regions corresponding to Ig-like domains and subsequent dynamics region was performed using a GNM (Supplementary Figure 11).

CONCLUSIONS

PD1, PD-L1, and PD-L2 have Ig-like domains that are highly conserved among vertebrates, which may enhance the understanding of their role in biological systems such as immunological tolerance. Our results revealed that positive selection acting on PD1, PD-L1 and PD-L2 genes drives adaptive changes for biological functions directly related to immunological tolerance in vertebrates. Our maximum-likelihood phylogenetic analyses of amino acid sequences from 166 vertebrate species revealed that PD1 genes evolved in a shared ancestor of vertebrates. According to the results, the major evolutionary processes causing the sequence variation observed in PD1 in vertebrates were adaptation and selection. Future studies integrating molecular data and pathogenicity evidence will help to determine the selective forces behind the long-term adaptation of programmed cell death genes, as well as to determine the genetic conflicts between immune system development pathways and immune tolerance.

MATERIALS AND METHODS

Ethics statement

This study was approved by the Ethics Board for Animal Trials at the Guangdong Institute of Applied Biological Resources (reference number: GIABR20170720) by following the basic ethical guidelines outlined by this committee.

Sample collection, RNA extraction and qRT-PCR

We selected five vertebrate species, *Manis javanica*, *Atelerix albiventris, Gallus, and Acipenser schrenckii* and *Shinisaurus crocodilurus* for tissue sampling;

among them, *M. javanica* and *S. crocodilurus* individuals died of their wounds. We collected the heart, liver, spleen, lung, kidney, pancreas, brain, testis, and ovary and muscle tissues from male and female adults. A total of 30 samples with an average of five individuals per species were used for RNA extraction using the RNAiso Pure RNA Isolation Kit (Takara, Japan). Total RNA was extracted from 0.25 g of tissue using the manual (TRIzol) method. We designed the primers by using references mRNA sequences retrieved from NCBI Genbank. The Prime Script™RT reagent kit with gDNA eraser was used to remove genomic DNA and to synthesize cDNA. cDNA samples from different tissue samples were assayed by quantitative real-time PCR (qRT-PCR) using specific set of primers (Supplementary Table 7) in the Thermal Cycler Dice® Real Time System (Bio-Rad, Hercules, CA, USA) using TB Green Premix Ex Taq™ II (Perfect Real Time, Cat. # PRO81A/B, Takara Co., Ltd.) with 96-well plates were used and each well contains a reaction mixture of 20µl containing 10µl of TB Green premix, 1.6µl primer mix (0.8µl of each primer) 2µl cDNA and 6.4ul of ddH_2O. The transcript level was normalized with beta-actin as a housekeeping gene of each representative species.

Identification, alignment, and filtering of vertebrate immune tolerance genes and orthologs

The orthologs of the PD1, PD-L1, and PD-L2 genes were identified, recovered, aligned, and filtered from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [55]. These genes were used to identify and retrieve the sequences of human genes from Ensembl BioMart [56]. The gene accession numbers were used to probe coding sequences (CDS) of vertebrate species in the NCBI and Ensembl databases (Supplementary Tables 4–6). Moreover, one-to-one orthologs in the vertebrate species were identified by performing tBLASTn and BLASTn searches [57]. In addition, the homology patterns among the protein-coding genes across the sequenced vertebrate genomes were determined by the OMA v.1.0.0 program [58]. The sequence alignments were made using MAFFT v.7.221 [59]. This aligned set of homologous proteins was used for further analyses.

Tests of selection

The proportion of sites in positively selected genes across vertebrate lineages was identified by comparison with the nearly neutral model of evolution to identify the signatures of positive selection. Positive selection sites were identified as those with higher nonsynonymous-synonymous substitution ratios ($\omega = dN/dS$) than expected under neutral evolution, $\omega = 1$ [45, 60, 61]. The genes under positive selection with high ω values at particular sites across vertebrate phylogenies were identified by using two different tools. Initially, the site models were used [62, 63] and executed in the PAML v4.8 package [61] to compute likelihood values and different constraint estimations for seven evolutionary models. Some genes contained copies; therefore, we performed all selection analyses on the gene trees from all species in the data set and individually for the genes with no duplicates in the species tree. The species tree constructed by OMA was used as the phylogenetic hypothesis [64]. First, we used M0, which evaluates a single ω for all positions in the alignment. The branch lengths predicted with M0 were used as fixed branch lengths for subsequent models to reduce computational time. We performed likelihood ratio tests between selection and neutral models ($\omega > 1$) to determine the genes under positive selection. The likelihood values from the M1a vs. M2a, M7 vs. M8, and M8 vs. M8a models were compared [62, 63], and the *p*-values were calculated, conferring a $\chi 2$ distribution with 2 degrees, 1 degree, and 1 degree of freedom, respectively [65].

Moreover, the genes with signatures of positive selection at a portion of sites were identified by using BUSTED [66], a modelling program executed in the HyPhy package [67]. BUSTED relies on a model that permits branch-to-branch variations across the whole tree [66]. Furthermore, the signals of positive selection for immune-tolerant genes were redetected by estimating the rates of nonsynonymous to synonymous substitutions at individual sites in the aligned sequences using various likelihood models, including the fixed effect likelihood (FEL), internal fixed effect likelihood (IFEL), single likelihood ancestor counting (SLAC), and maximum-likelihood estimation (MEME) methods [67–69]. To further confirm codon sites under selection pressure, aligned sequences of selected genes were tested in Selecton version 2.2 [41] (http://selecton.tau.ac.il/). Selecton allows the ω ratio to shift among different codons within the multiple sequence alignment, and this was estimated by the maximum-likelihood value via the Bayesian inference method [60, 70]. Moreover, the results from Selecton were visualized with color scales that indicated different types of selection.

Gene enrichment and conserved synteny analyses

The conserved synteny patterns of the PD1, PD-L1, and PD-L2 genes were determined using the Genomicus v.91.01 [71] and Ensembl [72] databases. We evaluated the conserved synteny for the genomic regions neighboring the PD1, PD-L1 and PD-L2 genes in vertebrate species. The evolutionary novelty of these genes was analyzed by using the Protein Historian program to recognize the taxon of sources of these genes [73]. Conserved syntenies might be linked with gene function and the corresponding gene expression [74, 75];

therefore, we used enrichment analysis to evaluated the biological significance of syntenic genes by searching these genes in various programs: EnrichNet, a network-based enrichment analysis [76], and ConsensusPathDB, a meta-database [77]. EnrichNet detects the genes in specific molecular systems and, using an arbitrary selection, estimates the intervals between the genes and pathways in a reference catalog [76]. ConsensusPathDB is a meta-database that contains an extensive assembly of human molecular interaction data linked to various public sources and has been used for reporting interaction network units and enrichment analyses [78]. The functional motifs and domains of the PD1, PD-L1 and PD-L2 proteins were predicted using the MEME tool (http://meme-suite.org/).

Three-dimensional (3D) protein modeling and structural analysis

The crystal structures of human PD1, PD-L1 and PD-L2 were generated using the Swiss model (https://swissmodel.expasy.org) online tool [79] and Phyre2 (http://www.sbg.bio.ic.ac.uk/phyre2/html). The homology modeling method was used to predict protein structure. The 3D structures of PD1, PD-L1, PD-L2 were predicted by the I-TESSAR and Swiss modeling approaches [80]. The assembled target proteins were minimized by using the Amber force field and the conjugate gradient algorithm in UCSF Chimera 1.10.1 [81]. Furthermore, the ProSA web server [82] was used to evaluate the stereochemical characteristics of the predicted structures.

Conservation analysis

The ConSurf server (consurf.tau.ac.il/) was used to evaluate the evolutionary conservation of amino acid residues of the human PD1, PD-L1 and PD-L2 proteins [83]. The amino acids are more conserved and are essential for protein interactions or are present within more enzymatic pockets than other amino acids of the protein. Therefore, the changes in the conserved amino acids are more lethal than polymorphisms located in flexible regions in a protein because they disrupt the protein function and structure [84, 85]. The conserved amino acids were predicted based on conservation values ranging from 1 to 9; a conservation value between 1–4 is considered variable, a value of 5–6 exhibits average conservation and a value ranging from 7 to 9 indicates very high conservation [86].

Bioinformatics analysis of protein sequences

The structures of the PD1, PD-L1 and PD-L2 proteins, including protein-binding positions, solvent accessibility, and disordered structures, were predicted using the Predict Protein server [87]. The secondary structures of the PD1, PD-L1, and PD-L2 proteins were predicted by the CFSSP program [54, 88]. Hydrophobic cluster analysis (HCA) was performed using the HCA 1.0.2 program [89] via the Mobyle@RPBS web portal and framework [90].

Coevolution analysis of protein residues

We used a web-based tool (CoeViz) [91] that allows the focused and rapid analysis of protein topographies, such as functional sites and structural domains, and that delivers a variable analysis and visualization of pairwise coevolution of amino acid residues. Full protein sequences of PD1, PD-L1 and PD-L2 were obtained from CoeViz analysis via $\chi 2$ covariance metrics [92] and were adjusted for phylogenetic bias in the MSA to predict the maps of covarying residues and the large overlaps with functional regions and the known domains of the protein. The visualization of the residue interactions was improved with circular drawings, and the residues were highlighted in the protein sequences and 3D structures.

AUTHOR CONTRIBUTIONS

Data curation, Hafiz Ishfaq Ahmad and Huiming Li; Formal analysis, Hafiz Ishfaq Ahmad, Xiujuan Zhang and Huiming Li; Funding acquisition, Hafiz Ishfaq Ahmad and Jinping Chen; Investigation, Jiabin Zhou and Xiujuan Zhang; Methodology, Hafiz Ishfaq Ahmad, Haiying Jiang, Abdelmotaleb A. Elokil, Gulnaz Afzal, and Liu Ping; Project administration, Haiying Jiang, Huiming Li and Linmiao Li; Resources, Jiabin Zhou, Haiying Jiang and Linmiao Li; Software, Hafiz Ishfaq Ahmad, Muhammad Jamil Ahmad, Abdelmotaleb A. Elokil and Liu Ping; Supervision, Linmiao Li and Jinping Chen; Validation, Jiabin Zhou and Xiujuan Zhang; Visualization, Liu Ping and Jinping Chen; Writing – original draft, Hafiz Ishfaq Ahmad and Gulnaz Afzal; Writing – review and editing, Hafiz Ishfaq Ahmad, Musarrat Abbas, Abdelmotaleb A. Elokil and Jinping Chen.

ACKNOWLEDGMENTS

The authors are thankful to Springer Nature Author Services for providing English language editing support and critical review of the manuscript.

CONFLICTS OF INTEREST

There is no conflict of interest in the conduction of this study.

FUNDING

This work was supported by the GDAS project of Science and Technology Development (2019-GDASYL-0103059 and 2018GDASCX-0107).

REFERENCES

1. Ishida M, Iwai Y, Tanaka Y, Okazaki T, Freeman GJ, Minato N, Honjo T. Differential expression of PD-L1 and PD-L2, ligands for an inhibitory receptor PD-1, in the cells of lymphohematopoietic tissues. Immunol Lett. 2002; 84:57–62. https://doi.org/10.1016/S0165-2478(02)00142-6 PMID:12161284

2. Ishida Y, Agata Y, Shibahara K, Honjo T. Induced expression of PD-1, a novel member of the immunoglobulin gene superfamily, upon programmed cell death. EMBO J. 1992; 11:3887–95. https://doi.org/10.1002/j.1460-2075.1992.tb05481.x PMID:1396582

3. Latchman Y, Wood CR, Chernova T, Chaudhary D, Borde M, Chernova I, Iwai Y, Long AJ, Brown JA, Nunes R, Greenfield EA, Bourque K, Boussiotis VA, et al. PD-L2 is a second ligand for PD-1 and inhibits T cell activation. Nat Immunol. 2001; 2:261–68. https://doi.org/10.1038/85330 PMID:11224527

4. Yamazaki T, Akiba H, Iwai H, Matsuda H, Aoki M, Tanno Y, Shin T, Tsuchiya H, Pardoll DM, Okumura K, Azuma M, Yagita H. Expression of programmed death 1 ligands by murine T cells and APC. J Immunol. 2002; 169:5538–45. https://doi.org/10.4049/jimmunol.169.10.5538 PMID:12421930

5. Greenwald RJ, Freeman GJ, Sharpe AH. The B7 family revisited. Annu Rev Immunol. 2005; 23:515–48. https://doi.org/10.1146/annurev.immunol.23.021704.115611 PMID:15771580

6. Nie X, Chen W, Zhu Y, Huang B, Yu W, Wu Z, Guo S, Zhu Y, Luo L, Wang S, Chen L. B7-DC (PD-L2) costimulation of $CD4^+$ T-helper 1 response via RGMb. Cell Mol Immunol. 2018; 15:888–97. https://doi.org/10.1038/cmi.2017.17 PMID:28479601

7. Nie M, Liu Y, Li XX, Min YN, Yang DD, Li Q, Feng Q, Hou Y, Li GS, Sun JZ, Hou M, Shi Y. PD-1/PD-L Pathway Potentially Involved in ITP Immunopathogenesis. Thromb Haemost. 2019; 119:758–65. https://doi.org/10.1055/s-0039-1679909 PMID:30808044

8. Yearley JH, Gibson C, Yu N, Moon C, Murphy E, Juco J, Lunceford J, Cheng J, Chow LQ, Seiwert TY, Handa M, Tomassini JE, McClanahan T. PD-L2 expression in human tumors: relevance to anti-PD-1 therapy in cancer. Clin Cancer Res. 2017; 23:3158–67. https://doi.org/10.1158/1078-0432.CCR-16-1761 PMID:28619999

9. Zhang X, Schwartz JC, Guo X, Bhatia S, Cao E, Lorenz M, Cammer M, Chen L, Zhang ZY, Edidin MA, Nathenson SG, Almo SC. Structural and functional analysis of the costimulatory receptor programmed death-1. Immunity. 2004; 20:337–47. https://doi.org/10.1016/S1074-7613(04)00051-2 PMID:15030777

10. Lin DY, Tanaka Y, Iwasaki M, Gittis AG, Su HP, Mikami B, Okazaki T, Honjo T, Minato N, Garboczi DN. The PD-1/PD-L1 complex resembles the antigen-binding Fv domains of antibodies and T cell receptors. Proc Natl Acad Sci USA. 2008; 105:3011–6. https://doi.org/10.1073/pnas.0712278105 PMID:18287011

11. Zak KM, Kitel R, Przetocka S, Golik P, Guzik K, Musielak B, Dömling A, Dubin G, Holak TA. Structure of the complex of human programmed death 1, PD-1, and its ligand PD-L1. Structure. 2015; 23:2341–48. https://doi.org/10.1016/j.str.2015.09.010 PMID:26602187

12. Park JJ, Omiya R, Matsumura Y, Sakoda Y, Kuramasu A, Augustine MM, Yao S, Tsushima F, Narazaki H, Anand S, Liu Y, Strome SE, Chen L, Tamada K. B7-H1/CD80 interaction is required for the induction and maintenance of peripheral T-cell tolerance. Blood. 2010; 116:1291–98. https://doi.org/10.1182/blood-2010-01-265975 PMID:20472828

13. Butte MJ, Keir ME, Phamduy TB, Sharpe AH, Freeman GJ. Programmed death-1 ligand 1 interacts specifically with the B7-1 costimulatory molecule to inhibit T cell responses. Immunity. 2007; 27:111–22. https://doi.org/10.1016/j.immuni.2007.05.016 PMID:17629517

14. Zhang B, Chikuma S, Hori S, Fagarasan S, Honjo T. Nonoverlapping roles of PD-1 and FoxP3 in maintaining immune tolerance in a novel autoimmune pancreatitis mouse model. Proc Natl Acad Sci USA. 2016; 113:8490–95. https://doi.org/10.1073/pnas.1608873113 PMID:27410049

15. Sun C, Mezzadra R, Schumacher TN. Regulation and function of the PD-L1 checkpoint. Immunity. 2018; 48:434–52. https://doi.org/10.1016/j.immuni.2018.03.014 PMID:29562194

16. Zhang Y, Ma L, Hu X, Ji J, Mor G, Liao A. The role of the PD-1/PD-L1 axis in macrophage differentiation

and function during pregnancy. Hum Reprod. 2019; 34:25–36.
https://doi.org/10.1093/humrep/dey347
PMID:30500923

17. Bally AP, Tang Y, Lee JT, Barwick BG, Martinez R, Evavold BD, Boss JM. Conserved region C functions to regulate PD-1 expression and subsequent CD8 T cell memory. J Immunol. 2017; 198:205–17.
https://doi.org/10.4049/jimmunol.1601464
PMID:27895178

18. Honda T, Egen JG, Lämmermann T, Kastenmüller W, Torabi-Parizi P, Germain RN. Tuning of antigen sensitivity by T cell receptor-dependent negative feedback controls T cell effector function in inflamed tissues. Immunity. 2014; 40:235–47.
https://doi.org/10.1016/j.immuni.2013.11.017
PMID:24440150

19. Erickson JJ, Gilchuk P, Hastings AK, Tollefson SJ, Johnson M, Downing MB, Boyd KL, Johnson JE, Kim AS, Joyce S, Williams JV. Viral acute lower respiratory infections impair CD8+ T cells through PD-1. J Clin Invest. 2012; 122:2967–82.
https://doi.org/10.1172/JCI62860
PMID:22797302

20. Liu J, Zhang E, Ma Z, Wu W, Kosinska A, Zhang X, Möller I, Seiz P, Glebe D, Wang B, Yang D, Lu M, Roggendorf M. Enhancing virus-specific immunity in vivo by combining therapeutic vaccination and PD-L1 blockade in chronic hepadnaviral infection. PLoS Pathog. 2014; 10:e1003856.
https://doi.org/10.1371/journal.ppat.1003856
PMID:24391505

21. Netea MG, Schlitzer A, Placek K, Joosten LA, Schultze JL. Innate and adaptive immune memory: an evolutionary continuum in the host's response to pathogens. Cell Host Microbe. 2019; 25:13–26.
https://doi.org/10.1016/j.chom.2018.12.006
PMID:30629914

22. Ellis JS, Turner LM, Knight ME. Patterns of selection and polymorphism of innate immunity genes in bumblebees (Hymenoptera: apidae). Genetica. 2012; 140:205–17.
https://doi.org/10.1007/s10709-012-9672-7
PMID:22899493

23. Roux J, Privman E, Moretti S, Daub JT, Robinson-Rechavi M, Keller L. Patterns of positive selection in seven ant genomes. Mol Biol Evol. 2014; 31:1661–85.
https://doi.org/10.1093/molbev/msu141
PMID:24782441

24. Privman E, Cohen P, Cohanim AB, Riba-Grognuz O, Shoemaker D, Keller L. Positive selection on sociobiological traits in invasive fire ants. Mol Ecol. 2018; 27:3116–30.
https://doi.org/10.1111/mec.14767
PMID:29920818

25. Shultz AJ, Sackton TB. Immune genes are hotspots of shared positive selection across birds and mammals. eLife. 2019; 8:e41815.
https://doi.org/10.7554/eLife.41815 PMID:30620335

26. Daub JT, Moretti S, Davydov II, Excoffier L, Robinson-Rechavi M. Detection of pathways affected by positive selection in primate lineages ancestral to humans. Mol Biol Evol. 2017; 34:1391–402.
https://doi.org/10.1093/molbev/msx083
PMID:28333345

27. Enard D, Cai L, Gwennap C, Petrov DA. Viruses are a dominant driver of protein adaptation in mammals. eLife. 2016; 5:e12469.
https://doi.org/10.7554/eLife.12469 PMID:27187613

28. Ebel ER, Telis N, Venkataram S, Petrov DA, Enard D. High rate of adaptation of mammalian proteins that interact with Plasmodium and related parasites. PLoS Genet. 2017; 13:e1007023.
https://doi.org/10.1371/journal.pgen.1007023
PMID:28957326

29. Grueber CE, Wallis GP, King TM, Jamieson IG. Variation at innate immunity Toll-like receptor genes in a bottlenecked population of a New Zealand robin. PLoS One. 2012; 7:e45011.
https://doi.org/10.1371/journal.pone.0045011
PMID:23024782

30. Fornůsková A, Vinkler M, Pagès M, Galan M, Jousselin E, Cerqueira F, Morand S, Charbonnel N, Bryja J, Cosson JF. Contrasted evolutionary histories of two Toll-like receptors (Tlr4 and Tlr7) in wild rodents (MURINAE). BMC Evol Biol. 2013; 13:194.
https://doi.org/10.1186/1471-2148-13-194
PMID:24028551

31. Tschirren B, Andersson M, Scherman K, Westerdahl H, Mittl PR, Råberg L. Polymorphisms at the innate immune receptor TLR2 are associated with Borrelia infection in a wild rodent population. Proc Biol Sci. 2013; 280:20130364.
https://doi.org/10.1098/rspb.2013.0364
PMID:23554395

32. Potapov V, Sobolev V, Edelman M, Kister A, Gelfand I. Protein—protein recognition: juxtaposition of domain and interface cores in immunoglobulins and other sandwich-like proteins. J Mol Biol. 2004; 342:665–79.
https://doi.org/10.1016/j.jmb.2004.06.072
PMID:15327963

33. Fowler SB, Clarke J. Mapping the folding pathway of an immunoglobulin domain: structural detail from Phi value analysis and movement of the transition state.

Structure. 2001; 9:355–66.
https://doi.org/10.1016/S0969-2126(01)00596-2
PMID:11377196

34. Teichmann SA, Chothia C. Immunoglobulin superfamily proteins in Caenorhabditis elegans. J Mol Biol. 2000; 296:1367–83.
https://doi.org/10.1006/jmbi.1999.3497
PMID:10698639

35. Müller V, de Boer RJ, Bonhoeffer S, Szathmáry E. An evolutionary perspective on the systems of adaptive immunity. Biol Rev Camb Philos Soc. 2018; 93:505–28.
https://doi.org/10.1111/brv.12355 PMID:28745003

36. Alcaide M, Edwards SV. Molecular evolution of the toll-like receptor multigene family in birds. Mol Biol Evol. 2011; 28:1703–15.
https://doi.org/10.1093/molbev/msq351
PMID:21239391

37. Grueber CE, Wallis GP, Jamieson IG. Episodic positive selection in the evolution of avian toll-like receptor innate immunity genes. PLoS One. 2014; 9:e89632.
https://doi.org/10.1371/journal.pone.0089632
PMID:24595315

38. Nehyba J, Hrdlicková R, Bose HR. Dynamic evolution of immune system regulators: the history of the interferon regulatory factor family. Mol Biol Evol. 2009; 26:2539–50.
https://doi.org/10.1093/molbev/msp167
PMID:19638535

39. Ikushima H, Negishi H, Taniguchi T. The IRF family transcription factors at the interface of innate and adaptive immune responses. Cold Spring Harb Symp Quant Biol. 2013; 78:105–16.
https://doi.org/10.1101/sqb.2013.78.020321
PMID:24092468

40. Doron-Faigenboim A, Stern A, Mayrose I, Bacharach E, Pupko T. Selecton: a server for detecting evolutionary forces at a single amino-acid site. Bioinformatics. 2005; 21:2101–03.
https://doi.org/10.1093/bioinformatics/bti259
PMID:15647294

41. Stern A, Doron-Faigenboim A, Erez E, Martz E, Bacharach E, Pupko T. Selecton 2007: advanced models for detecting positive and purifying selection using a Bayesian inference approach. Nucleic Acids Res. 2007; 35:W506–11.
https://doi.org/10.1093/nar/gkm382 PMID:17586822

42. Ahmad HI, Liu G, Jiang X, Liu C, Fangzheng X, Chong Y, Ijaz N, Huarong H. Adaptive selection at agouti gene inferred breed specific selection signature within the indigenous goat populations. Asian-Australas J Anim Sci. 2017. [Epub ahead of print].
https://doi.org/10.5713/ajas.16.0994 PMID:28423875

43. Evans AG, Rothberg PG, Burack WR, Huntington SF, Porter DL, Friedberg JW, Liesveld JL. Evolution to plasmablastic lymphoma evades CD19-directed chimeric antigen receptor T cells. Br J Haematol. 2015; 171:205–09.
https://doi.org/10.1111/bjh.13562 PMID:26084925

44. Ovchinnikov S, Kamisetty H, Baker D. Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. eLife. 2014; 3:e02030.
https://doi.org/10.7554/eLife.02030 PMID:24842992

45. Ahmad HI, Ahmad MJ, Adeel MM, Asif AR, Du X. Positive selection drives the evolution of endocrine regulatory bone morphogenetic protein system in mammals. Oncotarget. 2018; 9:18435–45.
https://doi.org/10.18632/oncotarget.24240
PMID:29719616

46. Li YD, Xie ZY, Du YL, Zhou Z, Mao XM, Lv LX, Li YQ. The rapid evolution of signal peptides is mainly caused by relaxed selection on non-synonymous and synonymous sites. Gene. 2009; 436:8–11.
https://doi.org/10.1016/j.gene.2009.01.015
PMID:19393172

47. Neves F, Abrantes J, Esteves PJ. Evolution of CCL11: genetic characterization in lagomorphs and evidence of positive and purifying selection in mammals. Innate Immun. 2016; 22:336–43.
https://doi.org/10.1177/1753425916647471
PMID:27189425

48. Anisimova M, Bielawski JP, Yang Z. Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol. 2001; 18:1585–92.
https://doi.org/10.1093/oxfordjournals.molbev.a003945 PMID:11470850

49. Alvarez-Ponce D, Torres-Sánchez M, Feyertag F, Kulkarni A, Nappi T. Molecular evolution of DNMT1 in vertebrates: duplications in marsupials followed by positive selection. PLoS One. 2018; 13:e0195162.
https://doi.org/10.1371/journal.pone.0195162
PMID:29621315

50. Rodrigo G, Fares MA. Intrinsic adaptive value and early fate of gene duplication revealed by a bottom-up approach. eLife. 2018; 7:e29739.
https://doi.org/10.7554/eLife.29739 PMID:29303479

51. Beltrao P, Bork P, Krogan NJ, van Noort V. Evolution and functional cross-talk of protein post-translational modifications. Mol Syst Biol. 2013; 9:714.
https://doi.org/10.1002/msb.201304521
PMID:24366814

52. Ackerman SH, Tillier ER, Gatti DL. Accurate simulation and detection of coevolution signals in multiple sequence alignments. PLoS One. 2012; 7:e47108.

https://doi.org/10.1371/journal.pone.0047108 PMID:23091608

53. Zhao Y, Wang Y, Gao Y, Li G, Huang J. Integrated analysis of residue coevolution and protein structures capture key protein sectors in HIV-1 proteins. PLoS One. 2015; 10:e0117506. https://doi.org/10.1371/journal.pone.0117506 PMID:25671429

54. Solís-Calero C, Carvalho HF. Phylogenetic, molecular evolution and structural analyses of the WFDC1/prostate stromal protein 20 (ps20). Gene. 2019; 686:125–40. https://doi.org/10.1016/j.gene.2018.10.046 PMID:30423385

55. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000; 28:27–30. https://doi.org/10.1093/nar/28.1.27 PMID:10592173

56. Flicek P, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, Fitzgerald S, Gil L, Gordon L, Hendrix M, et al. Ensembl 2012. Nucleic Acids Res. 2012; 40:D84–90. https://doi.org/10.1093/nar/gkr991 PMID:22086963

57. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. Nucleic Acids Res. 2008; 36:W5–9. https://doi.org/10.1093/nar/gkn201 PMID:18440982

58. Altenhoff AM, Gil M, Gonnet GH, Dessimoz C. Inferring hierarchical orthologous groups from orthologous gene pairs. PLoS One. 2013; 8:e53786. https://doi.org/10.1371/journal.pone.0053786 PMID:23342000

59. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30:772–80. https://doi.org/10.1093/molbev/mst010 PMID:23329690

60. Ahmad HI, Liu G, Jiang X, Edallew SG, Wassie T, Tesema B, Yun Y, Pan L, Liu C, Chong Y, Yu ZJ, Jilong H. Maximum-likelihood approaches reveal signatures of positive selection in BMP15 and GDF9 genes modulating ovarian function in mammalian female fertility. Ecol Evol. 2017; 7:8895–902. https://doi.org/10.1002/ece3.3336 PMID:29177034

61. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007; 24:1586–91. https://doi.org/10.1093/molbev/msm088 PMID:17483113

62. Nielsen R, Yang Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics. 1998; 148:929–36. PMID:9539414

63. Yang Z, Nielsen R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol. 2002; 19:908–17. https://doi.org/10.1093/oxfordjournals.molbev.a004148 PMID:12032247

64. Sackton TB, Grayson P, Cloutier A, Hu Z, Liu JS, Wheeler NE, Gardner PP, Clarke JA, Baker AJ, Clamp M, Edwards SV. Convergent regulatory evolution and loss of flight in paleognathous birds. Science. 2019; 364:74–78. https://doi.org/10.1126/science.aat7244 PMID:30948549

65. Ahmad MJ, Ahmad HI, Adeel MM, Liang A, Hua G, Murtaza S, Mirza RH, Elokil A, Ullah F, Yang L. Evolutionary Analysis of Makorin Ring Finger Protein 3 Reveals Positive Selection in Mammals. Evol Bioinform Online. 2019; 15:1176934319834612. https://doi.org/10.1177/1176934319834612 PMID:31024214

66. Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner T, Martin DP, Smith DM, Scheffler K, Kosakovsky Pond SL. Gene-wide identification of episodic selection. Mol Biol Evol. 2015; 32:1365–71. https://doi.org/10.1093/molbev/msv035 PMID:25701167

67. Pond SL, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. Bioinformatics. 2005; 21:676–9. https://doi.org/10.1093/bioinformatics/bti079 PMID:15509596

68. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol. 2005; 22:1208–22. https://doi.org/10.1093/molbev/msi105 PMID:15703242

69. Pond SL, Frost SD, Grossman Z, Gravenor MB, Richman DD, Brown AJ. Adaptation to different human populations by HIV-1 revealed by codon-based analyses. PLoS Comput Biol. 2006; 2:e62. https://doi.org/10.1371/journal.pcbi.0020062 PMID:16789820

70. Yang JR, Liao BY, Zhuang SM, Zhang J. Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. Proc Natl Acad Sci USA. 2012; 109:E831–40. https://doi.org/10.1073/pnas.1117408109 PMID:22416125

71. Louis A, Nguyen NT, Muffato M, Roest Crollius H. Genomicus update 2015: KaryoView and MatrixView provide a genome-wide perspective to multispecies comparative genomics. Nucleic Acids Res. 2015;

43:D682–89.
https://doi.org/10.1093/nar/gku1112 PMID:25378326

72. Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D, Cummins C, Clapham P, Fitzgerald S, Gil L, Girón CG, Gordon L, Hourlier T, et al. Ensembl 2016. Nucleic Acids Res. 2016; 44:D710–16.
https://doi.org/10.1093/nar/gkv1157
PMID:26687719

73. Capra EJ, Perchuk BS, Ashenberg O, Seid CA, Snow HR, Skerker JM, Laub MT. Spatial tethering of kinases to their substrates relaxes evolutionary constraints on specificity. Mol Microbiol. 2012; 86:1393–403.
https://doi.org/10.1111/mmi.12064
PMID:23078131

74. Nash AJ, Lenhard B. A novel measure of non-coding genome conservation identifies genomic regulatory blocks within primates. Bioinformatics. 2019; 35:2354–61.
https://doi.org/10.1093/bioinformatics/bty1014
PMID:30535005

75. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci USA. 2004; 101:6062–67.
https://doi.org/10.1073/pnas.0400782101
PMID:15075390

76. Glaab E, Baudot A, Krasnogor N, Schneider R, Valencia A. EnrichNet: network-based gene set enrichment analysis. Bioinformatics. 2012; 28:i451–57.
https://doi.org/10.1093/bioinformatics/bts389
PMID:22962466

77. Kamburov A, Stelzl U, Lehrach H, Herwig R. The ConsensusPathDB interaction database: 2013 update. Nucleic Acids Res. 2013; 41:D793–800.
https://doi.org/10.1093/nar/gks1055
PMID:23143270

78. Herwig R, Hardt C, Lienhard M, Kamburov A. Analyzing and interpreting genome data at the network level with ConsensusPathDB. Nat Protoc. 2016; 11:1889–907.
https://doi.org/10.1038/nprot.2016.117
PMID:27606777

79. Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc. 2009; 4:363–71.
https://doi.org/10.1038/nprot.2009.2
PMID:19247286

80. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. Nat Methods. 2015; 12:7–8.
https://doi.org/10.1038/nmeth.3213 PMID:25549265

81. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem. 2004; 25:1605–12.
https://doi.org/10.1002/jcc.20084 PMID:15264254

82. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res. 2007; 35:W407–10.
https://doi.org/10.1093/nar/gkm290
PMID:17517781

83. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 2010; 38:W529–33.
https://doi.org/10.1093/nar/gkq399
PMID:20478830

84. Schönichen A, Webb BA, Jacobson MP, Barber DL. Considering protonation as a posttranslational modification regulating protein structure and function. Annu Rev Biophys. 2013; 42:289–314.
https://doi.org/10.1146/annurev-biophys-050511-102349 PMID:23451893

85. Williamson K, Schneider V, Jordan RA, Mueller JE, Henderson Pozzi M, Bryk M. Catalytic and functional roles of conserved amino acids in the SET domain of the S. cerevisiae lysine methyltransferase Set1. PLoS One. 2013; 8:e57974.
https://doi.org/10.1371/journal.pone.0057974
PMID:23469257

86. Mayrose I, Graur D, Ben-Tal N, Pupko T. Comparison of site-specific rate-inference methods for protein sequences: empirical Bayesian methods are superior. Mol Biol Evol. 2004; 21:1781–91.
https://doi.org/10.1093/molbev/msh194
PMID:15201400

87. Hanson J, Paliwal K, Litfin T, Yang Y, Zhou Y. Improving prediction of protein secondary structure, backbone angles, solvent accessibility and contact numbers by using predicted contact maps and an ensemble of recurrent and residual convolutional neural networks. Bioinformatics. 2019; 35:2403–10.
https://doi.org/10.1093/bioinformatics/bty1006
PMID: 30535134

88. Kumar A, Kaur H, Jain A, Nair DT, Salunke DM. Docking, thermodynamics and molecular dynamics (MD) studies of a non-canonical protease inhibitor, MP-4, from Mucuna pruriens. Sci Rep. 2018; 8:689.
https://doi.org/10.1038/s41598-017-18733-9
PMID:29330385

89. Bitard-Feildel T, Callebaut I. HCAtk and pyHCA: A Toolkit and Python API for the Hydrophobic Cluster

Analysis of Protein Sequences. bioRxiv. 2018. https://doi.org/10.1101/249995

90. Alland C, Moreews F, Boens D, Carpentier M, Chiusa S, Lonquety M, Renault N, Wong Y, Cantalloube H, Chomilier J, Hochez J, Pothier J, Villoutreix BO, et al. RPBS: a web resource for structural bioinformatics. Nucleic Acids Res. 2005; 33:W44–9. https://doi.org/10.1093/nar/gki477 PMID:15980507

91. Baker FN, Porollo A. CoeViz: a web-based tool for coevolution analysis of protein residues. BMC Bioinformatics. 2016; 17:119. https://doi.org/10.1186/s12859-016-0975-z PMID:26956673

92. Baker FN, Porollo A. CoeViz: A Web-Based Integrative Platform for Interactive Visualization of Large Similarity and Distance Matrices. Data (Basel). 2018; 3:4. https://doi.org/10.3390/data3010004 PMID:29423399

SUPPLEMENTARY MATERIALS

Supplementary Figures

Legend:

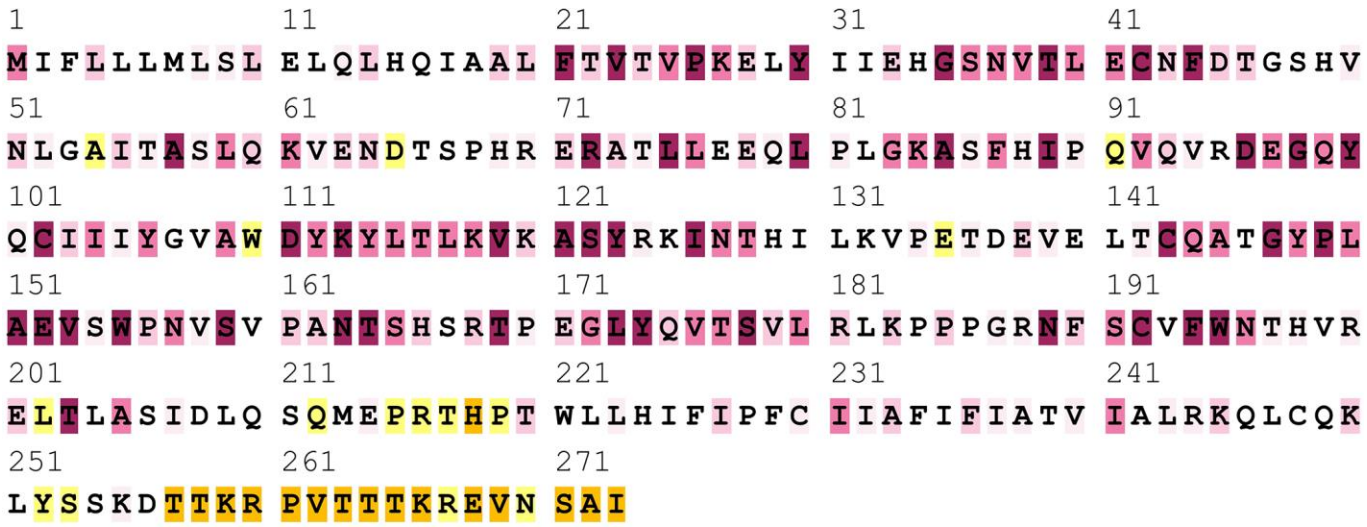
The selection scale:

1 2 3 4 5 6 7

Positive selection Purifying selection

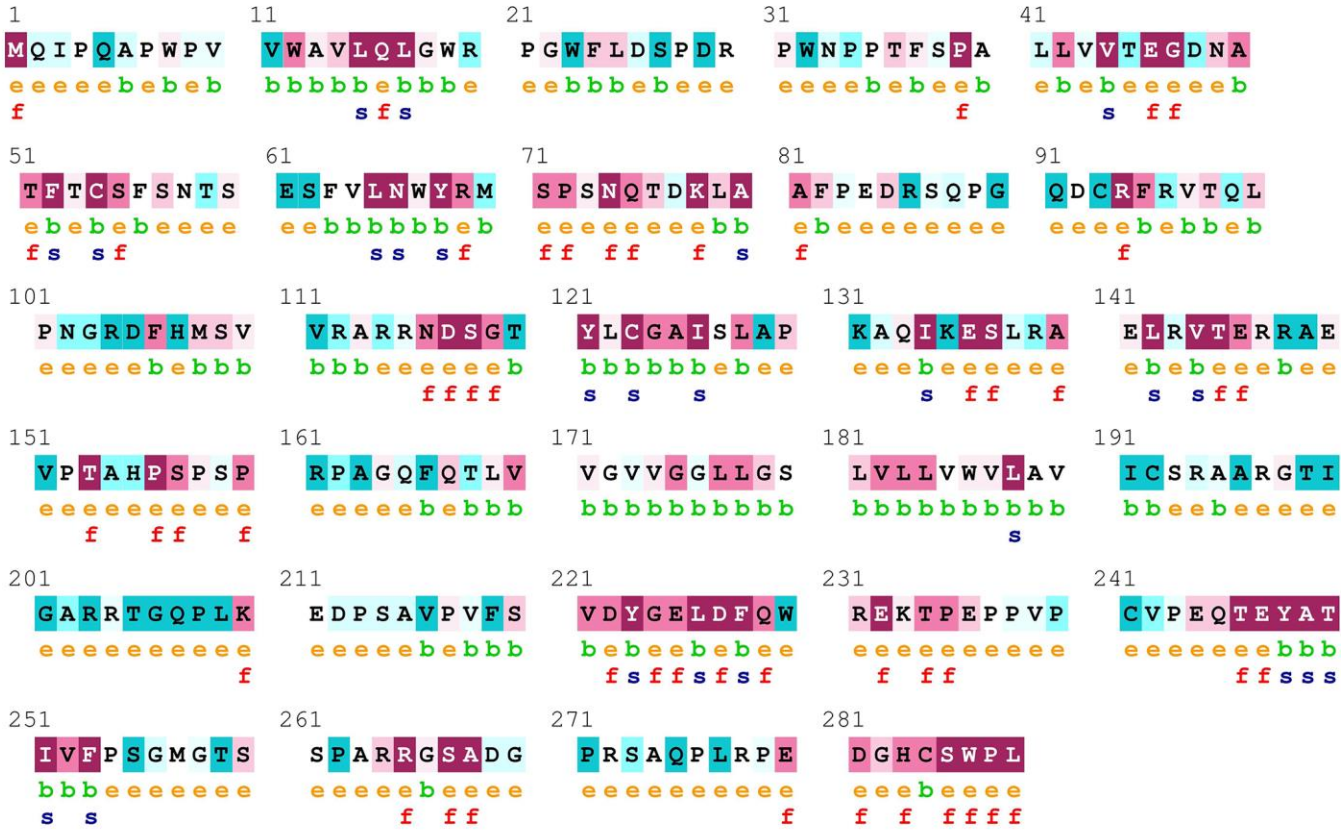
Supplementary Figure 1. Detection of positive/purifying selection in PD1 homologous sequences from vertebrates. Color-coded results of Selecton analyses of human PD1, compared to sequences from 52 aligned nucleotide coding sequences. Selection pressure was measured on sequences using mechanistic-empirical combination (MEC) model of Selecton version 2.2. Yellow and brown highlights represent positive selection, grey and white highlights represent a neutral selection and purple highlight represent purifying selection on codons.

Supplementary Figure 2. Detection of positive/purifying selection in PD-L1 homologous sequences from vertebrates. Color-coded results of Selecton analyses of human PD-L1, compared to sequences from 58 aligned nucleotide coding sequences. Selection pressure was measured on sequences using mechanistic-empirical combination (MEC) model of Selecton version 2.2. Yellow and brown highlights represent positive selection, grey and white highlights represent a neutral selection and purple highlight represent purifying selection on codons.



Supplementary Figure 3. Detection of positive/purifying selection in PD-L2 homologous sequences from vertebrates. Color-coded results of Selecton analyses of human PD-L2, compared to sequences from 56 aligned nucleotide coding sequences. Selection pressure was measured on sequences using mechanistic-empirical combination (MEC) model of Selecton version 2.2. Yellow and brown highlights represent positive selection, grey and white highlights represent a neutral selection and purple highlight represent purifying selection on codons.

ConSurf Results



The conservation scale:

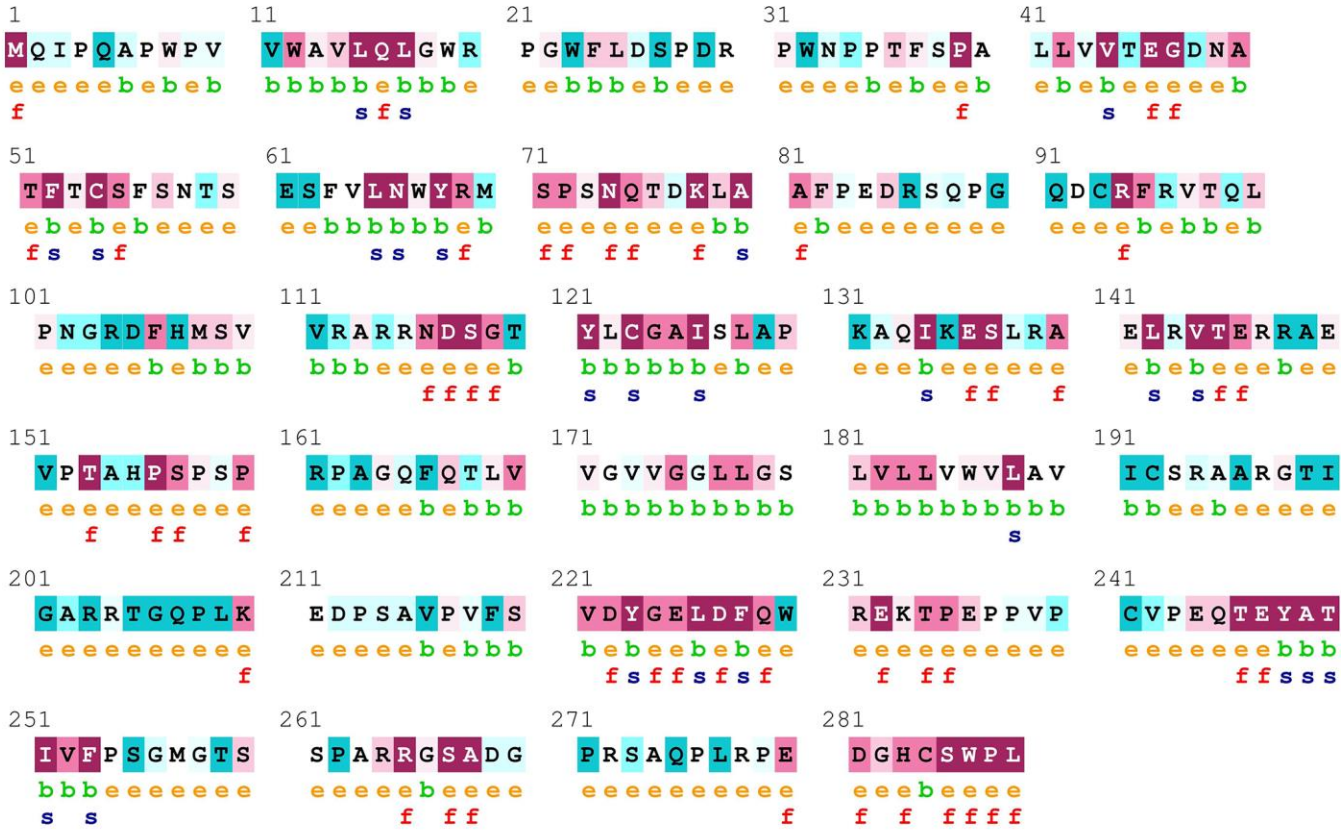
1 2 3 4 5 6 7 8 9

Variable **Average** **Conserved**

e – An exposed residue according to the neural-network algorithm.
b – A buried residue according to the neural-network algorithm.
f – A predicted functional residue (highly conserved and exposed).
s – A predicted structural residue (highly conserved and buried).

Supplementary Figure 4. ConSurf output of PD1, using the UniRef90 protein database. Colors of the ConSurf output indicate the level of sequence conservation. Purple indicates conservation and blue indicates variability. Residues are predicted to be exposed (e), buried (b), highly conserved and exposed (f), or highly conserved and buried, (s).

ConSurf Results



The conservation scale:

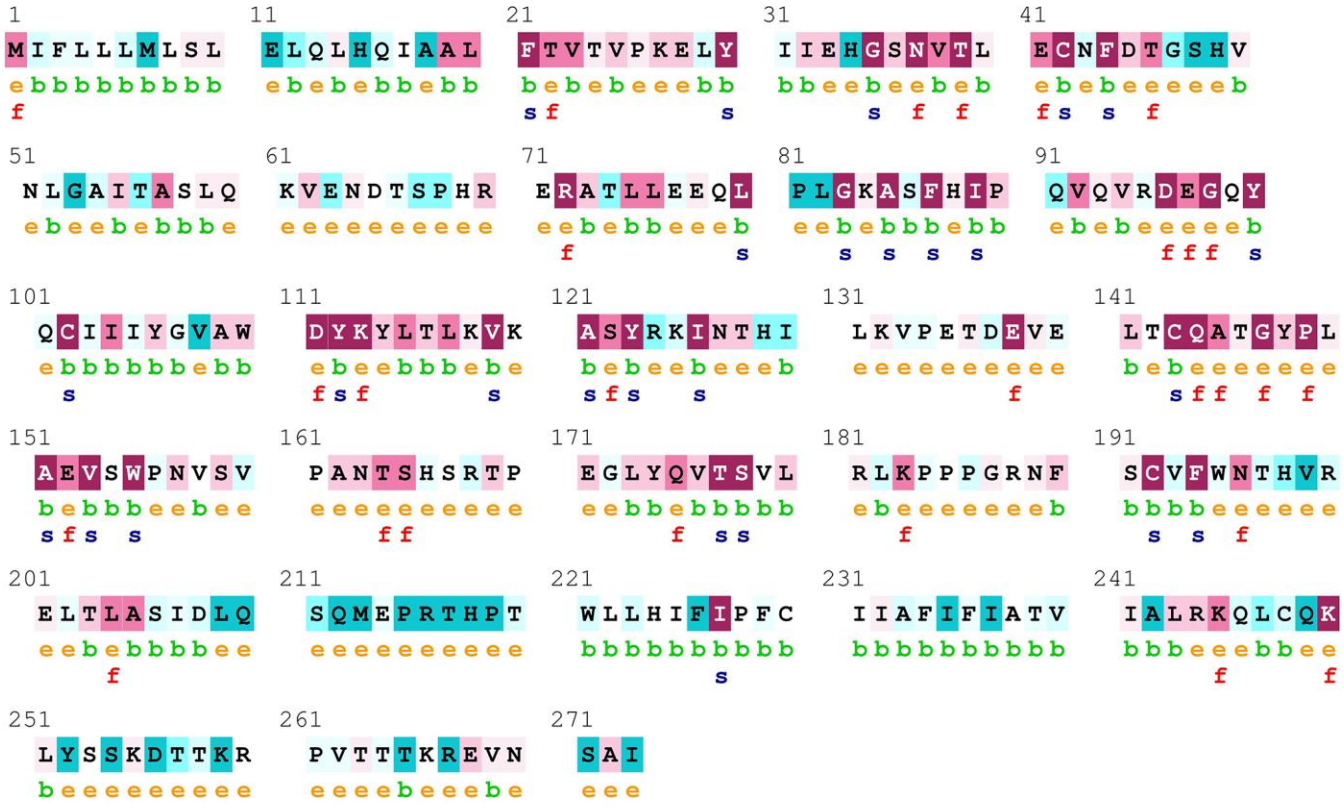
1 2 3 4 5 6 7 8 9

Variable **Average** **Conserved**

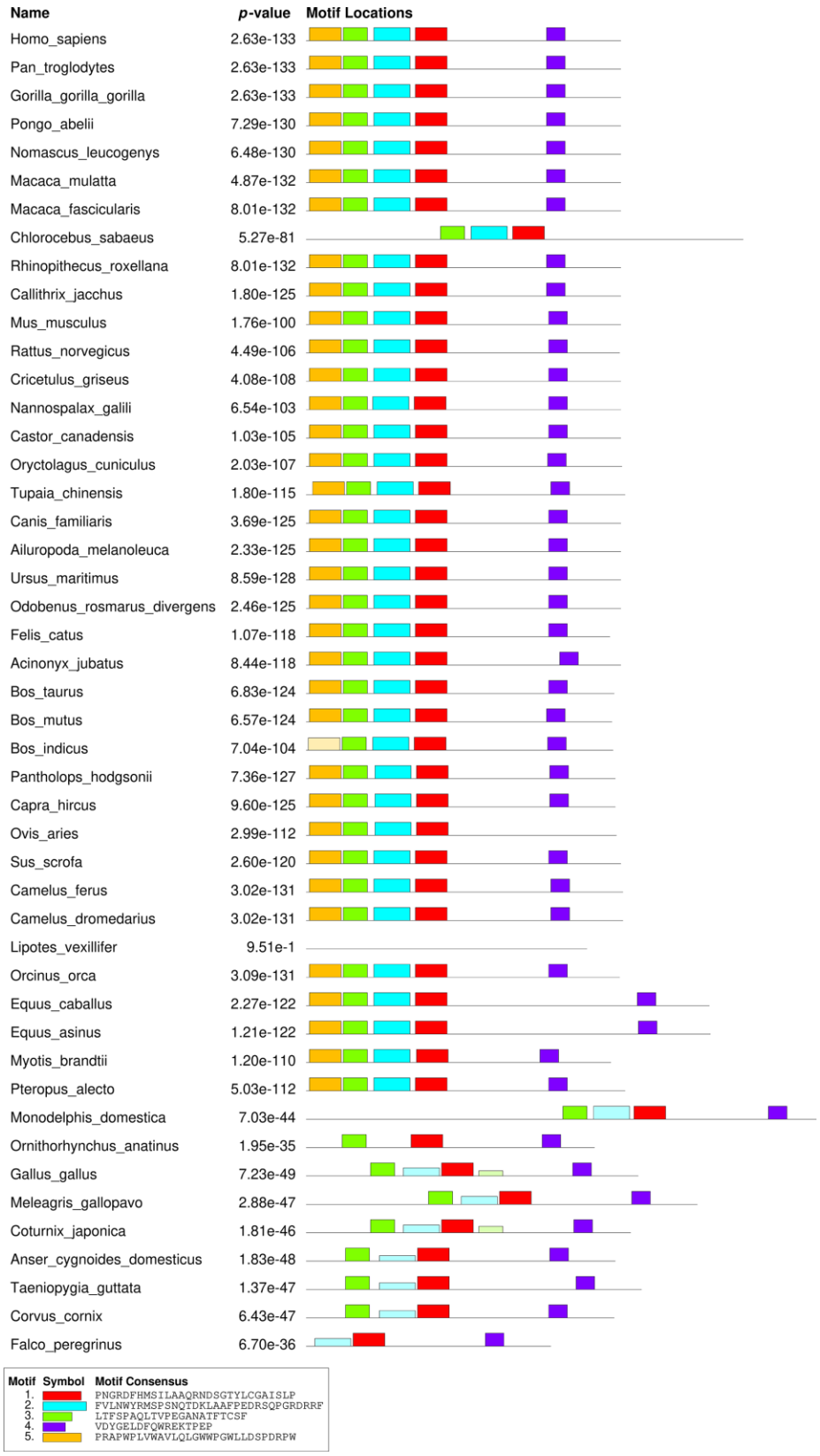
e – An exposed residue according to the neural-network algorithm.
b – A buried residue according to the neural-network algorithm.
f – A predicted functional residue (highly conserved and exposed).
s – A predicted structural residue (highly conserved and buried).

Supplementary Figure 5. ConSurf output of PD-L1, using the UniRef90 protein database. Colors of the ConSurf output indicate the level of sequence conservation. Purple indicates conservation and blue indicates variability. Residues are predicted to be exposed (e), buried (b), highly conserved and exposed (f), or highly conserved and buried, (s).

ConSurf Results

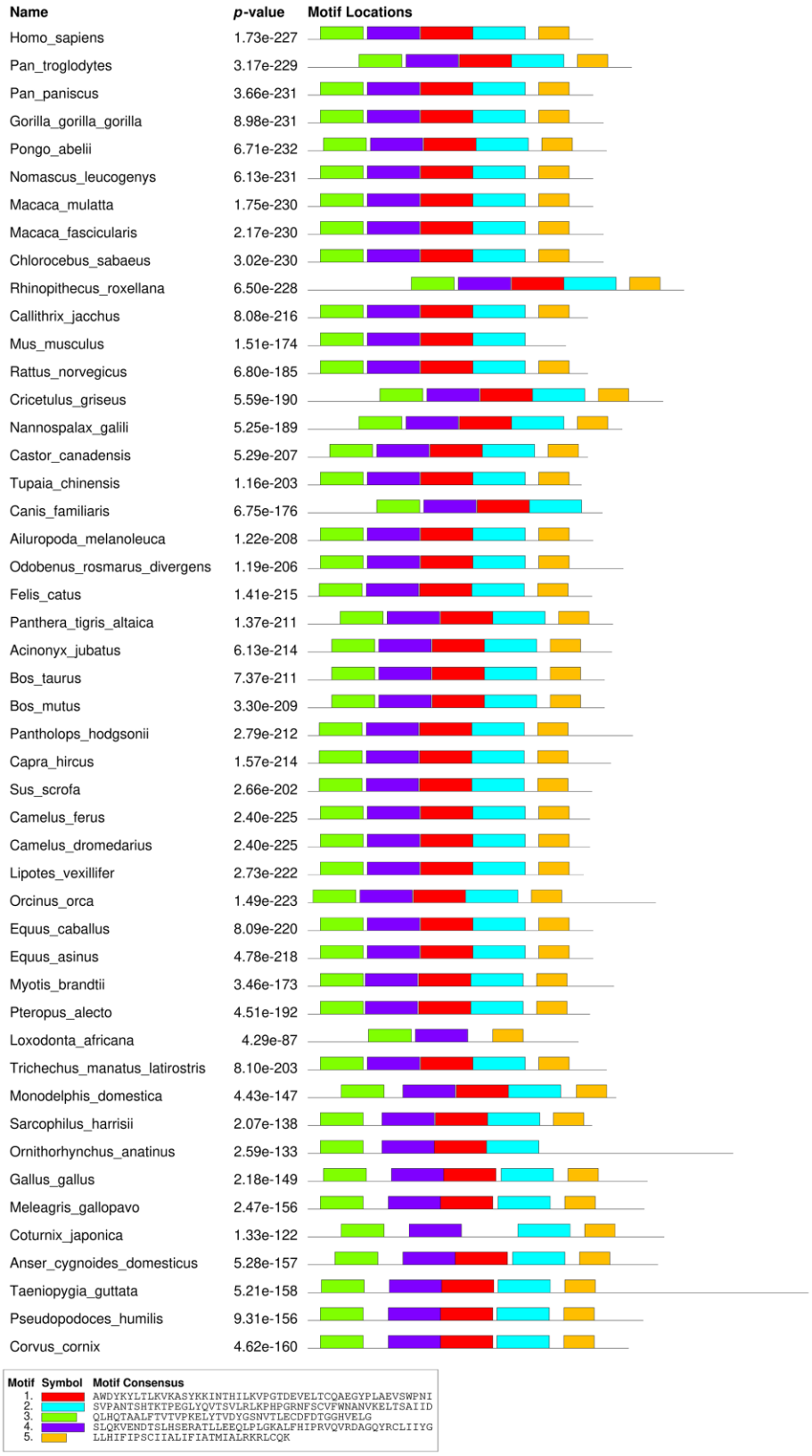


Supplementary Figure 6. ConSurf output of PD-L2, using the UniRef90 protein database. Colors of the ConSurf output indicate the level of sequence conservation. Purple indicates conservation and blue indicates variability. Residues are predicted to be exposed (e), buried (b), highly conserved and exposed (f), or highly conserved and buried, (s).

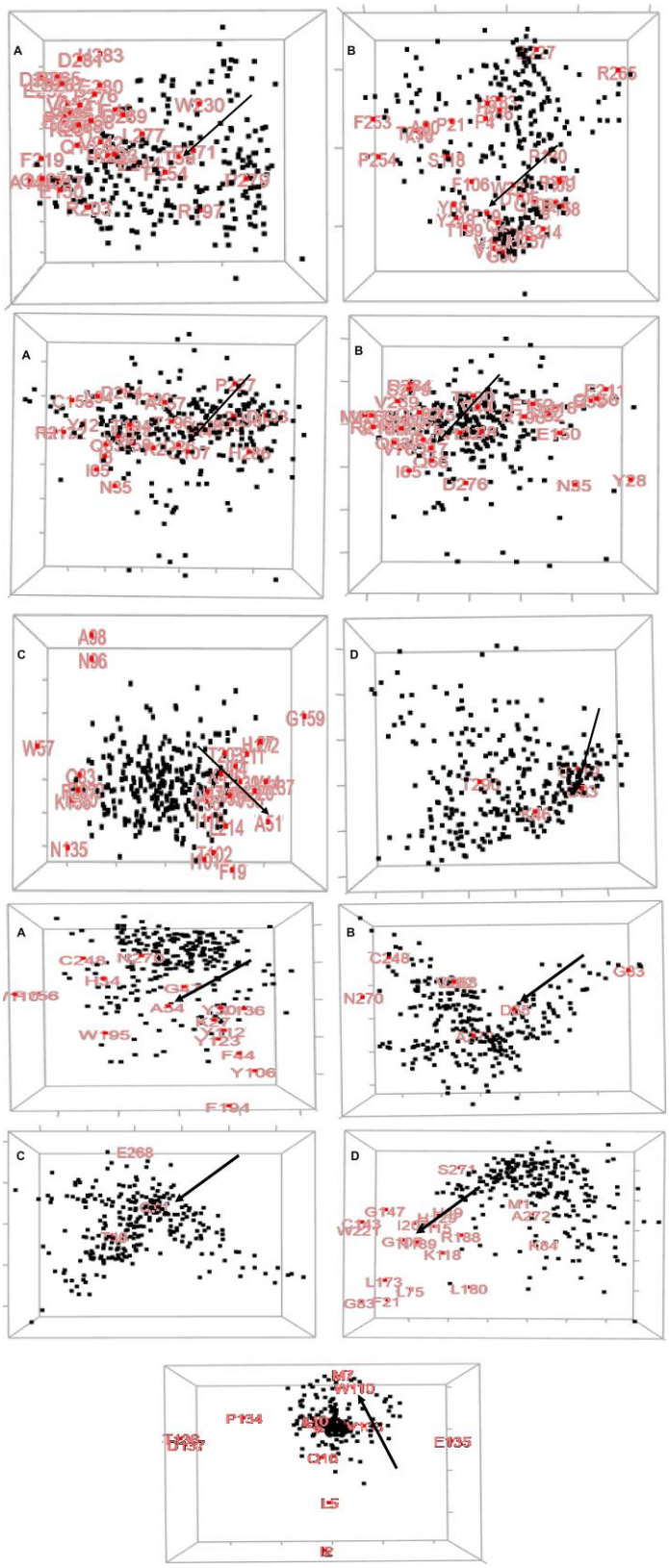


Supplementary Figure 7. Motif distribution of PD1 gene in the vertebrate species. Motifs of PD1 gene from 52 species are predicted using MEME suite (http://meme.nbcr.net/meme/) based on amino acids sequences. All sequences are separated by 5 conservative motifs with colors, including motif 1 (red), motif 2 (cyan), motif 3 (green), motif 4 (purple) and motif 5 (brown).

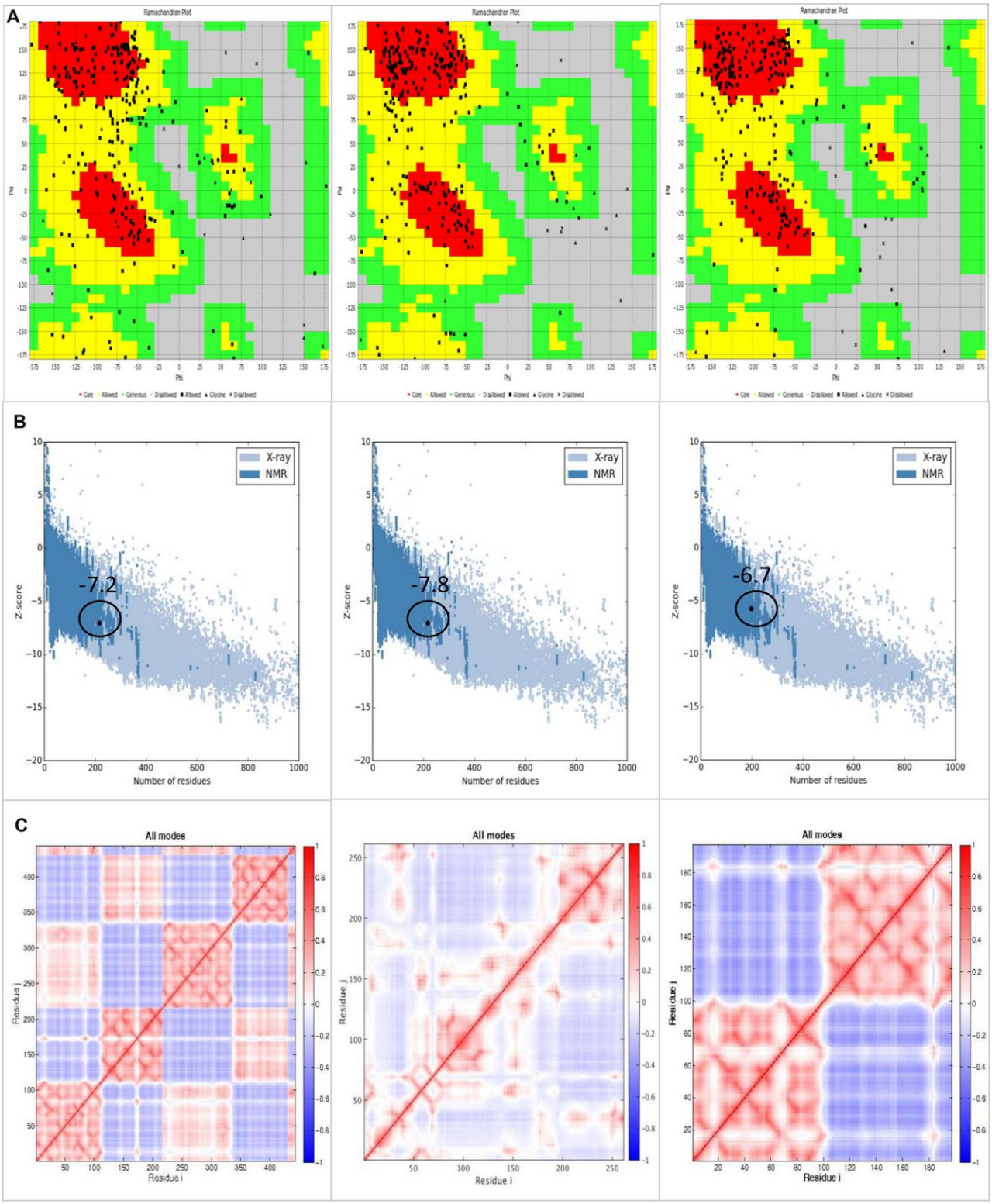
Supplementary Figure 8. Motif distribution of PD-L1 gene in the vertebrate species. Motifs of PD-L1 gene from 58 species are predicted using MEME suite (http://meme.nbcr.net/meme/) based on amino acids sequences. All sequences are separated by 5 conservative motifs with colors, including motif 1 (red), motif 2 (cyan), motif 3 (green), motif 4 (purple) and motif 5 (brown).



Supplementary Figure 9. Motif distribution of PD-L2 gene in the vertebrate species. Motifs of PD-L2 gene from 56 species are predicted using MEME suite (http://meme.nbcr.net/meme/) based on amino acids sequences. All sequences are separated by 5 conservative motifs with colors, including motif 1 (red), motif 2 (cyan), motif 3 (green), motif 4 (purple) and motif 5 (brown).



Supplementary Figure 10. 3D multidimensional scaling (MDS) scatterplots of co-varying residues in human PD1, PD-L1, and PD-L2. (**A**) Highlighted red are the residues corresponding to the positively selected residue. Both black and red dots can be optionally labeled with the residue identifiers.



Supplementary Figure 11. (**A**) Ramachandran plot analysis of PD1, PDL1 and PDL2 proteins. The backbone conformation angles for respective residue in the modeled protein crystal structures. Red color indicates the core region, yellow indicates allowed region, green is allowed region, and grey is disallowed region. (**B**) Z-score displays the quality analysis of predicted structures. (**C**) The cross-correlation analysis of all predicted structures. Positive correlations are plotted in the upper left triangle and negative correlations in the lower right triangle. Open squares: functional hydrophobic and negative residues.

Supplementary Tables

Supplementary Table 1. An adaptive branch-site (aBSREL) test for episodic diversification of the PD1 gene.

Name	B	LRT	Test p-value	Uncorrected p-value	ω distribution over sites
CHLOROCEBUS_SABAEUS	0.221	151.8573	0	0	ω1 = 0.0287 (72%) ω2 = 197 (28%)
ACINONYX_JUBATUS	0.1127	134.3588	0	0	ω1 = 0.0326 (84%) ω2 = 229 (16%)
OVIS_ARIES	0.1765	148.3166	0	0	ω1 = 0.732 (79%) ω2 = ∞ (21%)
TAENIOPYGIA_GUTTATA	0.0778	57.4098	0	0	ω1 = 0.361 (93%) ω2 = 745 (6.9%)
CHRYSEMYS_PICTA	0.1014	84.2906	0	0	ω1 = 0.205 (84%) ω2 = ∞ (16%)
Node47	0.0143	37.3204	0	0	ω1 = 0.00 (92%) ω2 = 131 (7.8%)
Node38	0.1042	25.0463	0.0001	0	ω1 = 0.441 (89%) ω2 = 5000 (11%)
ALLIGATOR_SINENSIS	0.2851	22.3465	0.0004	0	ω1 = 0.288 (72%) ω2 = 20.3 (28%)
Node37	0.1629	18.77	0.0025	0	ω1 = 0.114 (72%) ω2 = 56.3 (28%)
CORVUS_CORNIX	0.0603	17.9611	0.0037	0	ω1 = 0.109 (95%) ω2 = 35.9 (4.6%)
Node36	0.0821	15.1483	0.0151	0.0002	ω1 = 0.00 (74%) ω2 = 93.5 (26%)
APTERYX_AUSTRALIS_MANTELLI	0.1045	14.5227	0.0205	0.0002	ω1 = 0.244 (95%) ω2 = ∞ (5.5%)

B; Optimized branch length
LRT: Likelihood ratio test statistic for ω+ = 1 (null) versus ω+ unrestricted (alternative)
Test p-value: The p-value corrected for multiple testing.
Uncorrected P-value: The uncorrected p-value for the LRT test.

Supplementary Table 2. An adaptive branch-site (aBSREL) test for episodic diversification of PD-L1 gene.

Name	B	LRT	Test p-value	Uncorrected p-value	ω distribution over sites
COTURNIX_JAPONICA	0.0481	28.7232	0	0	ω1 = 0.298 (96%) ω2 = 5710 (3.7%)
TAENIOPYGIA_GUTTATA	0.0734	33.334	0	0	ω1 = 0.190 (94%) ω2 = 283 (5.8%)
PSEUDOPODOCES_HUMILIS	0.0672	41.2926	0	0	ω1 = 0.474 (93%) ω2 = 90.4 (6.8%)
FALCO_PEREGRINUS	0.1009	63.2226	0	0	ω1 = 0.691 (89%) ω2 = 212 (11%)
APTERYX_AUSTRALIS_MANTELLI	0.0976	28.7185	0	0	ω1 = 0.376 (92%) ω2 = 187 (8.5%)
ANSER_CYGNOIDES_DOMESTICUS	0.1042	22.369	0.0005	0	ω1 = 0.340 (95%) ω2 = 5710 (5.3%)
Node41	0.0217	19.3186	0.0022	0	ω1 = 0.00 (92%) ω2 = 5710 (8.4%)
CHLOROCEBUS_SABAEUS	0.0247	16.2851	0.0102	0.0001	ω1 = 0.385 (98%) ω2 = 1660 (1.6%)
Node33	0.1354	15.022	0.0191	0.0002	ω1 = 0.295 (88%) ω2 = 57.4 (12%)
SARCOPHILUS_HARRISII	0.1369	13.2103	0.0471	0.0005	ω1 = 0.216 (94%) ω2 = 30.1 (5.7%)

B; Optimized branch length
LRT: Likelihood ratio test statistic for ω+ = 1 (null) versus ω+ unrestricted (alternative)
Test p-value: The p-value corrected for multiple testing.
Uncorrected P-value: The uncorrected p-value for the LRT test.

Supplementary Table 3. An adaptive branch-site (aBSREL) test for episodic diversification of PD-L2 gene.

Name	B	LRT	Test p-value	Uncorrected p-value	ω distribution over sites
CRICETULUS_GRISEUS	0.1001	40.3338	0	0	ω1 = 0.359 (88%) ω2 = 211 (12%)
PANTHOLOPS_HODGSONII	0.0216	56.1676	0	0	ω1 = 0.00 (92%) ω2 = 1650 (7.7%)
LOXODONTA_AFRICANA	0.2699	71.5397	0	0	ω1 = 0.831 (71%) ω2 = ∞ (29%)
TAENIOPYGIA_GUTTATA	0.0998	147.3444	0	0	ω1 = 0.437 (81%) ω2 = 5000 (19%)
CORVUS_CORNIX	0.0583	47.1431	0	0	ω1 = 0.565 (94%) ω2 = 204 (6.3%)
Node68	0.0732	37.2754	0	0	ω1 = 0.308 (96%) ω2 = 5000 (3.9%)
Node84	0.0156	43.9788	0	0	ω1 = 0.421 (89%) ω2 = 1940 (11%)
PANTHERA_TIGRIS_ALTAICA	0.016	24.5613	0.0001	0	ω1 = 0.762 (97%) ω2 = 11400 (2.7%)
Node23	0.03	24.5946	0.0001	0	ω1 = 0.133 (89%)

					ω2 = 5000 (11%)
Node41	0.0311	24.1356	0.0002	0	ω1 = 1.00 (89%)
					ω2 = ∞ (11%)
CHRYSEMYS_PICTA	0.1462	21.302	0.0007	0	ω1 = 0.457 (91%)
					ω2 = 40.8 (9.4%)
NANNOSPALAX_GALILI	0.1469	20.5087	0.0011	0	ω1 = 0.359 (91%)
					ω2 = ∞ (9.3%)
LIPOTES_VEXILLIFER	0.0136	20.2389	0.0013	0	ω1 = 0.450 (98%)
					ω2 = 1530 (1.6%)
Node55	0.0103	20.1447	0.0013	0	ω1 = 1.00 (97%)
					ω2 = 90200 (2.6%)
COTURNIX_JAPONICA	0.1582	19.8778	0.0015	0	ω1 = 0.602 (91%)
					ω2 = 118 (8.7%)
ACINONYX_JUBATUS	0.0158	19.1886	0.0021	0	ω1 = 0.891 (99%)
					ω2 = 5820 (0.69%)
Node75	0.1565	16.6122	0.0074	0	ω1 = 0.278 (87%)
					ω2 = 2180 (13%)
Node5	0.0019	15.603	0.0122	0	ω1 = 0.0169 (95%)
					ω2 = ∞ (5.1%)
Node82	0.09	15.1548	0.0151	0	ω1 = 0.706 (83%)
					ω2 = 289 (17%)
Node14	0.0014	15.0154	0.016	0	ω1 = 0.0638 (96%)
					ω2 = ∞ (3.8%)
Node38	0.008	14.2666	0.023	0	ω1 = 0.00 (93%)
					ω2 = ∞ (7.0%)
CALLITHRIX_JACCHUS	0.0396	14.0674	0.0252	0	ω1 = 0.331 (97%)
					ω2 = 158 (2.7%)
ORNITHORHYNCHUS_ANATINUS	0.3646	13.9868	0.0259	0	ω1 = 0.114 (74%)
					ω2 = ∞ (26%)

B; Optimized branch length
LRT: Likelihood ratio test statistic for ω+ = 1 (null) versus ω+ unrestricted (alternative)
Test p-value: The p-value corrected for multiple testing.
Uncorrected P-value: The uncorrected p-value for the LRT test

Supplementary Table 4. The detail information of PD1 gene in vertebrate species.

Species	Scientific Name	NCBI-Protein ID	Ensembl	Uniprot
Human	*Homo sapiens*	NP_005009	ENSG00000188389	A0A0M3M0G7
Chimpanzee	*Pan troglodytes*	XP_016806440		A0A2I3SUZ1
Western lowland gorilla	*Gorilla gorilla gorilla*	XP_004033550	ENSGGOG00000014312	G3REH1
Sumatran orangutan	*Pongo abelii*	XP_024099392	ENSPPYG00000013360	
Northern white-cheeked gibbon	*Nomascus leucogenys*	XP_003282018		G1R8C2
Rhesus monkey	*Macaca mulatta*	NP_001107830	ENSMMUG00000008592	B0LAJ2
Crab-eating macaque	*Macaca fascicularis*	NP_001271065		B0LAJ3
Green monkey	*Chlorocebus sabaeus*	XP_007965216	ENSCSAG00000005736	
Golden snub-nosed monkey	*Rhinopithecus roxellana*	XP_010375274		A0A2K6QL35
White-tufted-ear marmoset	*Callithrix jacchus*	XP_002750041	ENSCJAG00000019795	
Mouse	*Mus musculus*	NP_032824	ENSMUSG00000026285	Q544F3
Rat	*Rattus norvegicus*	NP_001100397	ENSRNOG00000049797	D3ZIN8
Chinese hamster	*Cricetulus griseus*	XP_003499314		
Mole Rat	*Nannospalax galili*	XP_008827391		
American beaver	*Castor canadensis*	XP_020013289		
Rabbit	*Oryctolagus cuniculus*	XP_017195300	ENSOCUG00000008079	G1SUF0
Chinese tree shrew	*Tupaia chinensis*	XP_006165239		
Dog	*Canis familiaris*	NP_001301026	ENSCAFG00000013184	A0A090BAM7
Giant panda	*Ailuropoda melanoleuca*	XP_002923181	ENSAMEG00000016373	G1MD70
Polar bear	*Ursus maritimus*	XP_008688483		A0A384C0X7
Pacific walrus	*Odobenus rosmarus divergens*	XP_004412778		A0A2U3WTR3
Domestic cat	*Felis catus*	NP_001138982		A9YUA6
Amur tiger	*Panthera tigris altaica*	XP_007094211		
Cheetah	*Acinonyx jubatus*	XP_026912119		
Cow	*Bos taurus*	NP_001076975	ENSBTAG00000011543	A4FV85
Wild yak	*Bos mutus*	XP_005901569		
Chiru	*Pantholops hodgsonii*	XP_005959703		
Goat	*Capra hircus*	XP_013818552		
Sheep	*Ovis aries*	XP_012031617	ENSOARG00000017587	
Pig	*Sus scrofa*	NP_001191308		
Wild Bactrian camel	*Camelus ferus*	XP_014412330		
Arabian camel	*Camelus dromedarius*	XP_010986792		
killer whale	*Orcinus orca*	XP_004262561		
Horse	*Equus caballus*	XP_023498581		
Ass	*Equus asinus*	XP_014718129		
Brandt's bat	*Myotis brandtii*	XP_005857730		
Black flying fox	*Pteropus alecto*	XP_006909981		
Opossum	*Monodelphis domestica*	XP_016280819	ENSMODG00000029268	
Platypus	*Ornithorhynchus anatinus*	XP_007667778		
Chicken	*Gallus gallus*	XP_422723	ENSGALG00000046559	
Turkey	*Meleagris gallopavo*	XP_010715066		
Japanese quail	*Coturnix japonica*	XP_015727433		
Swan goose	*Anser cygnoides domesticus*	XP_013050415		
Zebra finch	*Taeniopygia guttata*	XP_012431056		
Hooded crow	*Corvus cornix*	XP_019140093		

Peregrine falcon	*Falco peregrinus*	XP_013154882	
Rock pigeon	*Columba livia*	XP_005513327	A0A2I0LU23
Brown kiwi	*Apteryx australis mantelli*	XP_013814564	
Western clawed frog	*Xenopus tropicalis*	XP_004914438	
Chinese alligator	*Alligator sinensis*	XP_025058540	
Western painted turtle	*Chrysemys picta*	XP_005311833	
Green sea turtle	*Chelonia mydas*	XP_007057904	

List of species and NCBI GenBank accession numbers for sequences used to construct the datasets for hypothesis testing.

Supplementary Table 5. The detail information of PD-L1 gene in vertebrate species.

Species	*Scientific Name*	Ensembl	Uniprot	NCBI-Protein ID
Human	*Homo sapiens*	ENSG00000120217	Q9NZQ7	
Chimpanzee	*Pan troglodytes*	ENSPTRG00000020755	H2QWZ8	
Bonobo	*Pan paniscus*		A0A2R9B063	XP_003823426
Western lowland gorilla	*Gorilla gorilla gorilla*			XP_018889139
Sumatran orangutan	*Pongo abelii*	ENSPPYG00000019237	H2PS75	XP_002819859
Northern white-cheeked gibbon	*Nomascus leucogenys*		G1RCP1	XP_003273874
Rhesus monkey	*Macaca mulatta*	ENSMMUG00000012235	A4GW29	NP_001077358
Crab-eating macaque	*Macaca fascicularis*		G7PSE7	XP_005581836
Green monkey	*Chlorocebus sabaeus*	ENSCSAG00000008879		XP_007967522
Golden snub-nosed monkey	*Rhinopithecus roxellana*		A0A2K6PJ04	XP_010363385
White-tufted-ear marmoset	*Callithrix jacchus*	ENSCJAG00000011388	A4GW22	NP_001254676
Mouse	*Mus musculus*	ENSMUSG00000016496	Q9EP73	NP_068693
Rat	*Rattus norvegicus*	ENSRNOG00000016112	D4AE25	NP_001178883
Chinese hamster	*Cricetulus griseus*			XP_007626098
Mole Rat	*Nannospalax galili*			XP_008849411
American beaver	*Castor canadensis*			XP_020015185
Rabbit	*Oryctolagus cuniculus*	ENSOCUG00000008117	G1SUI3	XP_008253343
Chinese tree shrew	*Tupaia chinensis*			XP_006152480
Dog	*Canis familiaris*	ENSCAFG00000002120	E2RKZ5	NP_001278901
Giant panda	*Ailuropoda melanoleuca*	ENSAMEG00000009800		XP_011228696
Polar bear	*Ursus maritimus*		A0A384C2Q2	XP_008689127
Pacific walrus	*Odobenus rosmarus divergens*		A0A2U3ZNI8	XP_012420554
Domestic cat	*Felis catus*	ENSFCAG00000009047	M3WAP8	XP_006939101
Amur tiger	*Panthera tigris altaica*			XP_007094214
Pangolin	*Manis javanica*			XM_017657766.1
Cheetah	*Acinonyx jubatus*			XP_014919968
Cow	*Bos taurus*	ENSBTAG00000000095	C5NU11	NP_001156884
Wild yak	*Bos mutus*		L8IU19	XP_005891421
Zebu cattle	*Bos indicus*			XP_019821547
Chiru	*Pantholops hodgsonii*			XP_005983887
Goat	*Capra hircus*			XP_005683750
Sheep	*Ovis aries*	ENSOARG00000013509		XP_004004411
Pig	*Sus scrofa*	ENSSSCG00000005211	Q4QTK1	NP_001020392
Wild Bactrian camel	*Camelus ferus*			XP_014416016

Arabian camel	*Camelus dromedarius*			XP_010991731
Yangtze River dolphin	*Lipotes vexillifer*		A0A340X5J4	XP_007454584
killer whale	*Orcinus orca*			XP_004279158
Horse	*Equus caballus*	ENSECAG00000016312	F7DZ76	XP_001492892
Ass	*Equus asinus*			XP_014716218
Brandt's bat	*Myotis brandtii*			XP_005861392
Black flying fox	*Pteropus alecto*		L5K6N3	XP_006918439
African savanna elephant	*Loxodonta africana*			XP_010586356
Florida manatee	*Trichechus manatus latirostris*		A0A2Y9DG34	XP_004373536
Opossum	*Monodelphis domestica*	ENSMODG00000015352		XP_007499604
Tasmanian devil	*Sarcophilus harrisii*	ENSSHAG00000002682		XP_012399523
Platypus	*Ornithorhynchus anatinus*	ENSOANG00000010091		XP_001506048
Chicken	*Gallus gallus*			XP_424811
Turkey	*Meleagris gallopavo*			XP_019466012
Japanese quail	*Coturnix japonica*			XP_015704470
Swan goose	*Anser cygnoides domesticus*			XP_013052847
Zebra finch	*Taeniopygia guttata*			XP_012433182
Tibetan ground-tit	*Pseudopodoces humilis*			XP_014107713
Hooded crow	*Corvus cornix*			XP_010391160
Peregrine falcon	*Falco peregrinus*			XP_013158040
Brown kiwi	*Apteryx australis mantelli*			XP_013812986
Western clawed frog	*Xenopus tropicalis*			XP_017946448
Common carp	*Cyprinus carpio*			XP_018933702
Channel catfish	*Ictalurus punctatus*		A0A2D0SYD8	XP_017347687
Spotted green pufferfish	*Tetraodon nigroviridis*		Q4T1R6	CAF93166

List of species and NCBI GenBank accession numbers for sequences used to construct the datasets for hypothesis testing.

Supplementary Table 6. The detail information of PD-L2 gene in vertebrate species.

Species	*Scientific Name*	Ensembl	Uniprot	NCBI-Protein ID
Human	*Homo sapiens*	ENSG00000197646	Q9BQ51	NP_079515
Chimpanzee	*Pan troglodytes*	ENSPTRG00000020756		XP_001140776
Bonobo	*Pan paniscus*		A0A2R9AHT0	XP_003823428
Western lowland gorilla	*Gorilla gorilla gorilla*			XP_018889135
Sumatran orangutan	*Pongo abelii*	ENSPPYG00000019236		XP_024107734
Northern white-cheeked gibbon	*Nomascus leucogenys*		G1RCR0	XP_003273875
Rhesus monkey	*Macaca mulatta*		A4GW30	NP_001077068
Crab-eating macaque	*Macaca fascicularis*			XP_005581838
Green monkey	*Chlorocebus sabaeus*	ENSCSAG00000008883		XP_007967513
Golden snub-nosed monkey	*Rhinopithecus roxellana*			XP_010363386
White-tufted-ear marmoset	*Callithrix jacchus*	ENSCJAG00000011422	A0A2R8P957	XP_017828946
Mouse	*Mus musculus*	ENSMUSG00000016498	Q3U304	NP_067371
Rat	*Rattus norvegicus*	ENSRNOG00000016136	D4AAV6	NP_001101052
Chinese hamster	*Cricetulus griseus*			XP_003511103

Mole Rat	*Nannospalax galili*			XP_017658223
American beaver	*Castor canadensis*			XP_020015183
Rabbit	*Oryctolagus cuniculus*			XP_017202816
Pangolin	*Manis javanica*			XR_001852674.1
Dog	*Canis familiaris*	ENSCAFG00000002121	44336	XP_013973349
Giant panda	*Ailuropoda melanoleuca*	ENSAMEG00000009830		XP_011228695
Polar bear	*Ursus maritimus*		A0A384C2M8	XP_008689126
Pacific walrus	*Odobenus rosmarus divergens*		A0A2U3ZNI5	XP_012420555
Domestic cat	*Felis catus*	ENSFCAG00000009048	L7SSK5	NP_001277173
Amur tiger	*Panthera tigris altaica*			XP_007094211
Cheetah	*Acinonyx jubatus*			XP_014919966
Cow	*Bos taurus*		A0A024FBV6	NP_001278965
Wild yak	*Bos mutus*			XP_014332357
Chiru	*Pantholops hodgsonii*			XP_005983898
Goat	*Capra hircus*			XP_013821401
Pig	*Sus scrofa*	ENSSSCG00000026305	Q4QTK0	NP_001020391
Wild Bactrian camel	*Camelus ferus*			XP_014415932
Arabian camel	*Camelus dromedarius*			XP_010991730
Yangtze River dolphin	*Lipotes vexillifer*		A0A340X676	XP_007454585
killer whale	*Orcinus orca*			XP_004279193
Horse	*Equus caballus*	ENSECAG00000020578		XP_005605049
Ass	*Equus asinus*			XP_014716221
Brandt's bat	*Myotis brandtii*			XP_014391175
Black flying fox	*Pteropus alecto*			XP_006918437
African savanna elephant	*Loxodonta africana*			XP_023401133
Florida manatee	*Trichechus manatus latirostris*		A0A2Y9QY77	XP_023584363
Opossum	*Monodelphis domestica*	ENSMODG00000015349	F6WY92	XP_007499609
Tasmanian devil	*Sarcophilus harrisii*	ENSSHAG00000004176	G3VNK0	XP_012399521
Platypus	*Ornithorhynchus anatinus*			XP_007653790
Chicken	*Gallus gallus*			XP_004949124
Turkey	*Meleagris gallopavo*			XP_010723930
Japanese quail	*Coturnix japonica*			XP_015704465
Swan goose	*Anser cygnoides domesticus*			XP_013052854
Zebra finch	*Taeniopygia guttata*			XP_002193013
Tibetan ground-tit	*Pseudopodoces humilis*			XP_014107710
Hooded crow	*Corvus cornix*			XP_010391172
Peregrine falcon	*Falco peregrinus*			XP_013157997
Brown kiwi	*Apteryx australis mantelli*			XP_013812991
Western clawed frog	*Xenopus tropicalis*			XP_017946447
Rock pigeon	*Columba livia*		A0A2I0MG20	XP_005499831
Chinese alligator	*Alligator sinensis*		A0A1U8CUP6	XP_014372758
Western painted turtle	*Chrysemys picta*			XP_005296857

List of species and NCBI GenBank accession numbers for sequences used to construct the datasets for hypothesis testing.

Supplementary Table 7. List of primers used for mRNA expression in qPCR of representative vertebrate species.

Species		5'-3'	TM
	GgPD1-F	CAGCCACGTTAACTCGTCCT	57.5
	GgPD1-R	GTTCCGGATGATCCCAGCAA	57.8
	GgPDL1-F	CTCATTGTGAGTGCCCTCGT	57.6
	GgPDL1-R	CATGCGCGCCCTTATCTTTC	57.3
	GgPDL2-F	CGCAATGGGAAAGCACTCAC	57.4
	GgPDL2-R	CGCATCTGTGATCTTGACGC	56.9
	EcPD1-F	CTGTTTCTGGTGCTCCCAGT	61.6
	EcPD1-R	TGGTGGCATATTCGGTCTGG	61.7
	EcPDL1-F	TGTATTGGTGTCCTGTTGGCA	56.6
	EcPDL1-R	GGTCATTCACTGGAAACCTGC	56.6
	EcPDL2-F	CGCTACCAGGGAAGAGCAAC	58.9
	EcPDL2-R	TGAGACAGCGGTAAGACCCT	58.1
	AsPD1-F	CAGAGATGGCCACTGCGTAA	61.7
	AsPD1-R	GTCTGTCTGGGAACGGGATG	61.7
	AsPDL1-F	TAATCCCCGTCGCCCTTTTC	57.8
	AsPDL1-R	TGGTGATGCGCTCTGTTAGG	57.7
	AsPDL2-F	TTGTACCCTTTCACTGCGCT	56.9
	AsPDL2-R	CCATTAAGGGCACCGTCTCA	57.5
	MjPD1-F	GAGGACGAGCCTCTGAAGGA	61.5
	MjPD1-R	CGTGGCATACTCGGTCTGTT	62.5
	MjPDL1-F	GTGAAAGTGGAGGAAGA	47.7
	MjPDL1-R	AGGATGGTCAGGAATTG	47.9
	MjPDL2-F	CCATCGGACGGTCTTTCACTA	60.2
	MjPDL2-R	GCCAGCTCCACACTCTAGCA	59.2
	AaPD1-F	GGCTCTCTTGCTTCTGCCTG	59.2
	AaPD1-R	CGACTGTGAGATGTTGGGGG	58.6
	AaPDL1-F	GTGGTGCCGACTACAAGC	56.8
	AaPDL1-R	TGGGGTAGCCCTCAGCCT	59.1
	AaPDL1-F	GGAGGGCCTGTTGACTTAGG	61.5
	AaPDL1-R	TCTGACTTCCAGGGTCAGGT	61.7