DOI:https://dx.doi.org/10.71373/PZDB5938
Submitted 10 November 2025
Accepted 26 Novenber 2025
Pulished 28 November 2025

Investigating the Role of Pangolin Pseudogenes in Host-Pathogen Immune Interactions: A BiLSTM-Based Approach

Tengcheng Que[1,2,3#], Zhining Zhang[4#], Yunlin He[4#], Qiuyu Wu[2], Jinying He[2], Xinni Yang[5], Panyu Chen[3], Hong Qiu[4], Yuankun Liu[1], Hua Zhang[2], Wenjian Liu[1*]

This study integrates a Bidirectional Long Short-Term Memory (BiLSTM) deep learning framework with bioinformatics approaches to elucidate the functional role of pangolin pseudogenes in the immune interactions between tick-borne pathogens and their hosts. We developed a novel BiLSTM-autoencoder model incorporating dynamic weight positional encoding and a multi-head self-attention mechanism, effectively capturing pseudogene sequence features while overcoming limitations inherent in traditional analytical methods. Genome-wide screening of the *Manis pentadactyla* (GCF_030020395.1) and *Manis javanica* (GCF_001685135.1) assemblies identified 3,209 and 2,035 pseudogenes, respectively. Subsequent filtration yielded 94 immune-related homologous genes, classified into ten distinct immune system pathways. Our analysis reveals significant species-specific variation and functional plasticity in pangolin pseudogenes: Chinese Pangolin LILRA6 Pseudogenes: Fourteen identified variants modulate bacterial recognition and inflammatory responses through specific amino acid deletions. Interferon Receptor Pseudogenes (e.g., XP_036870501.1): Critical mutations disrupt JAK-STAT signaling pathway functionality. Toll-like Receptor (TLR) Pseudogenes: Reconstruction error values (ranging from 28.149 to 68.957) correlate strongly with structural integrity. Notably, the disrupted LRR domain in the giant pangolin sequence KAK2502118.1 (error = 65.986) suggests an adaptive immune strategy. Furthermore, we identified a potential molecular interaction between a pangolin-derived 52 kDa sequence and the 8.9 kDa salivary protease of *Amblyomma javanense*, providing novel insights into tick-borne transmission mechanisms. This study represents the first application of deep learning to elucidate the functional role of pangolin pseudogenes, confirming their active involvement in immune resistance. It establishes a new paradigm for investigating host-pathogen interactions and provides a critical foundation for the analysis of underlying data in the surveillance and control of tick-borne diseases.

Introduction

Tick-borne pathogens pose a significant threat to global public health security; their intricate host interaction mechanisms have become a focal point of research. These pathogens encompass a wide range of organisms, including viruses, bacteria, and protozoan parasites[1]. In recent years, ecological shifts and the expansion of human activities have contributed to the emergence of new tick-borne pathogens. Notably, at the beginning of2025, a research team led by Dr. Cao Wuchun identified a novel tick-borne ortho- nairovirus in the northeastern region of China, known as the Xue-Cheng virus (XCV). Among the clinical cases resulting from XCV infection, 38% of patients exhibited severe symptoms such as fever with liver damage, necessitating hospitalization. The high pathogenicity of XCV notonly underscores the pathogen's direct invasiveness but also highlights the pivotal role of host immune response variability in disease progression[2].

At the molecular level of pathogen-host interactions, there is adynamic interplay between the host's immune resistanceand the adaptive evolution of viruses. On one hand, the genetic polymorphism of host innate immune factors influences susceptibility to tick-borne viruses. On the other hand, pathogens canachieve cross-species transmission by targeting host cell receptors. Research has shown that tick-borne flaviviruses from the Flaviviridae family can complete the host invasion process byutilizing the TIM-1 receptor (T-cell immunoglobulin and mucin domain-containing molecule 1, encoded by the HAVCR1 gene) on the surface of host cells[3]. This receptor utilization mechanism not only mediates the replication and spread of viruses within the host but also potentially serves as a bridge for cross-species trans- mission, enabling pathogens to infect humans or other mammalian hosts[4]. Notably, the *Manis javanica* has been confirmed to carry the tentative species "Candidatus Borrelia javanense," transmitted by the *Amblyomma javanense*.The genome of this pathogen exhibits significant genetic recombination characteristics, such as genetic plasticity may enhance its adaptability to the immune systems of different hosts by promoting antigenic variation or immune evasion capabilities[5].

The transmission of tick-borne pathogens involves a complex three-way interaction characterized by[6]: Immunosuppressive proteins in tick saliva, such as Salp15, which create an 'immune- privileged microenvironment' by inhibiting the activation of host CD4+ T cells and the complement system[7]; Pathogens that evade host immune clearance through antigenic variation,exemplified by the hemagglutinin glycoprotein of bunyaviruses and signal pathway interference, such as rickettsiae inhibiting host NF-κB activation[8]; Hosts that defend against tick-borne pathogens by relying on innate immunity,

1. Faculty of Data Science, City University of Macau, 999078, Macau, China 2. Youjiang Medical University for Nationalities, Baise, 533000, Guangxi, China 3. Terrestrial Wildlife Rescue and Epidemic Diseases Surveillance Center of Guangxi, Nanning, 530003, Guangxi, China 4. Guangxi Henbio Biotechnology Co., Ltd., Nanning, 530000, Guangxi, China 5. Department of Microbiology, School of Basic Medical Sciences, Guangxi Medical University, Nanning, 530021, Guangxi, China

# Tengcheng Que, Zhining Zhang, and Yunlin He contributed equally to this study

**Corresponding author*

E-mail addresses: ylau@cityu.edu.mo (Wenjian Liu)

including the TLR4/NF-κB pathway and adaptive immunity, exemplified by CD8+T cells[9]. However, T cell exhaustion in severely infected patients differs from the gradual dysfunction observed in chronicviral infections. Its mechanism may involve mitochondrial metabolic abnormalities or the over- expression of immune checkpoint molecules, such as PD-1[10]. Taking *Amblyomma javanense* as an example — a tick that parasitizes pangolins as key hosts — the pathogens it carries can invade host immune cells by binding to specific receptor proteins on the surface of host pangolin cells, including PSGL-1, integrins (αvβ3/β1), Toll-like receptors, and sialylated receptors. Variations in related genes, such as *SELPLG, ITGAV, ITGB3, TLR2/TLR4*, etc., may all affect the host's susceptibility or resistance[11~14].

Studies on the immune systems of host animals have shown that positive selection of genes in the TLR signaling pathway and potential functional remodeling of pseudogenes occur to resist pathogen invasion. For instance, unique variations in the TLR genes of pangolins may affect their ability to recognize Borrelia burgdorferi, while the inactivation of certain pseudogenes, such as glycosylation-related genes in the FUT family, may alter the glycosylation modification of cell surface receptors, such as PSGL-1, thereby influencing pathogen binding efficiency [12,15].Nevertheless, research on how pangolins regulate tick-borne pathogen infections through immune genes, including functional genes and pseudogenes, remains scarce.Therefore,exploring the role of non-coding elements, such as pseudogenes, in the host genome in anti-infection has become a key direction to overcome the limitations of existing research.

Given the intricate dynamics of pathogen infection and the body's immune resistance, the advent of AI technology offers a novel approach to tackling these complex challenges. Jiang et al.(2023) focused on the recognition characteristics of T cell receptors (TCRs) for foreign antigens and designed the TEI Net model. Employing transfer learning, this model transforms TCR sequences and epitope sequences into numerical vectors using two distinct pre-trained encoders. These vectors are then fed into a fully connected neural network to predict the binding specificity between TCRs and epitopes. The findings demonstrated that the TEI Net model can accurately forecast the specific binding between TCRs and epitopes utilizing solely the CDR3β sequences of TCRs and epitope sequences[16]. Considering the complexity of the inter- actions between pathogens and the host immune system, Kim et al.(2017) developed a method based on Support Vector Machine (SVM) to convert key features of viral and host proteins into fixed-length feature vectors. These key features encompass differences in the relative frequency of amino acid triplets, frequency disparities of amino acid triplets between viral and host proteins, and amino acid composition.The study indicated that this method is more effective in predicting heterogeneous protein- protein interactions between humans and viruses, such as hepatitis C virus (HCV) or human papillomavirus (HPV), outperforming ot- her methods in prediction accuracy[17]. Weiskopf et al.(2013) discovered that in the interaction between dengue virus and the host, T cells exhibit a memory function against the virus and proposed a protective correlation between CD8+ T cells and human leukocyte antigen (HLA)-like proteins[18].

To directly identify immunogenic peptides from sequences, Li et al.(2021) proposed a method grounded in the beta-binomial distribution. They researched three validated sets of immunogenic peptides (from dengue virus, cancer neoantigens, and SARS-CoV-2) and performed systematic benchmarking across five machine learning models (ElasticNet, KNN, SVM, RF,and AdaBoost) and three deep learning models (CNN, ResNet, and GNN). Ultimately, they identified CNN as the optimal prediction model. CNN can not only accurately predict the amino acid residues most critical for T cell antigens but also forecast the impact of SARS-CoV-2 variants. Furthermore, they employed a generative adversarial network (GAN) approach to accurately simulate immunogenic peptides with predicted physicoch emical properties and immunogenicity[19]. These studies provide crucial insights for a better understanding of the pivotal role of T cells in the immune response and for exploring prevention, control, and treatment strategies for tick-borne diseases.

Among mammalian hosts of tick-borne pathogens, pangolins rep- resent a particularly noteworthy species. Following the publication of the whole genome of the Malayan pangolin (*Manis javanica*) by Choo et al.(2016), it was proposed that the pseudogenization of the IFN-ε gene in pangolins might facilitate a low-damage coexistence with pathogens by attenuating the inflammatory response in the skin and mucous membranes[20]. Gene function analysis further reveals that the loss of function in the IFIH1 (melanoma differentiation- associated protein 5, MDA5) gene can diminish the ability to recognize double-stranded RNA of coronaviruses, thereby preventing tissue damagecaused by excessive immune activation[21]. Additionally, it hasbeen suggested that the pseudogene of pangolin ACE2 may influence viral infection efficiency by competitively binding to the viral RBD (receptor-binding domain) or regulating the expression of host cell surface receptors. This hypothesis has received indirect support from a mouse infection model using the pangolin-derived coronavirus GX/P2V/2017[22].

The traditional view of pseudogenes as "genomic fossils" has been increasingly challenged with the deeper understanding of their functions. Since Tam et al. first proposed in 2010 that pseudogenes regulate gene expression by competitively binding to miRNAs[23], numerous studies have confirmed that pseudogenes can extensively participate in immune regulation through mechanisms such as competitively binding to miRNAs (the ceRNA mechanism) or encoding functional small peptides [24~26].The role of pseudogenes in immune resistance is thus undeniable. For instance, while humans possess 13 functional IFN-α genes[27], pangolins have significantly fewer, with the Malayan pangolin having only 3 and the Chinese pangolin (*Manis pentadactyla*) only 2[16].This reduction in the number of IFN family genes in pangolins suggests unique immune genomic characteristics that could provide insights into the mechanisms of host-pathogen coexistence. Although IFN-ε is crucial for skin and mucosal immunity in most mammals, it is pseudogenized in all pangolin epithelial cells. Whether this pseudogenization weakens IFN-mediated innate immunity to reduce the inflammatory response triggered by pathogen invasion and thus facilitates coexistence with pathogens remains an open question[24]. Zhang et al. identified TIM-1 as a functional receptor for tick-borne encephalitis virus (TBEV), noting that viral particles can enter host cells through co-internalization with the host animal's membrane protein TIM-1 receptor. Importantly, they found that in mice with interferon (IFN) deficiency, TIM-1-deficient mice exhibited attenuated TBEV infection and pathogenesis. Additionally, TIM-1 deficiency was found to reduce viral load and pathogenicity in tissues, demonstrating that the pseudogenization of the TIM-1 gene benefits the body's resistance to viruses[3]. The pseudogenization status of the TIM-1 receptor gene in pangolins has yet to be reported. However, during rescue efforts, it has been observed that pangolins are frequently infected with *Amblyomma javanense* (Java tick), which can carry a variety of pathogens, including Canine parvovirus (CPV), Jingmen tick virus (JMTV), *Rickettsia spp., Anaplasma spp., Ehrlichia spp., Borrelia spp., Babesia spp., and Colpodella spp.*[5,29,30]. The relationship between these pathogens and gene or pseudogene receptors has not been systematically explored.

In this study,we investigate the relationship between tick-borne pathogens and immune resistance in pangolins, with a focus on systematically analyzing the role of pangolin pseudogenes in immune resistance to elucidate the adaptive mechanisms of their immunity.

Methods

Collection of Immune Information on Tick Infections in Malayan Pangolins

From October 2017 to March 2025, we collected a total of 860 ticks with complete records and preserved samples. These ticks were obtained from Malayan pangolins and Chinese pangolins, and identified independently by the State Key Laboratory of Pathogen and Biosecurity, Guangxi Medical University, Guangzhou Zoo, and our research team. All ticks were confirmed to be Amblyomma javanense (Java tick), including 317 males, 431 females, and 112 nymphs (unsexed). Existing reports show that ticks found on Chinese pangolins include Amblyomma javanense[31] and ticks of the genus Haemaphysalis (*Haemaphysalis spp.*)[32]. Most scholars have reported that ticks parasitizing Malayan pangolins are also Amblyomma javanense[5,31,33–37], while some scholars have occ- asionally reported other tick species on Malayan pangolins, such as *Rhipicephalus spp.*[38] and other ticks of the genus Amblyomma (*Amblyomma spp.*)[35,39]. Our identification results and relevant reports confirm that Amblyomma javanense has a certain degree of host tropism toward pangolins in Asia.

Through detection, we found that some *Amblyomma javanensecarry* factors mediated by the synergistic effect between tick salivary proteases and pathogen invasion effectors. Considering the immune resistance of hosts like pangolins and the pseudogenization of key effector genes,we downloaded and conducted comparative analysis on key genes or proteins—including 8.9 kilodalton (kDa) salivary proteins of Amblyomma ticks, TIM-1,and Toll-like receptors (TLRs) to explore their immune interaction relationships.

Extraction of pseudogenes and homolo-gous genes

To ascertain the pseudogenization status of the Malayan pangolin and the Chinese pangolin, we initially retrieved FASTA reference genome sequences and GTF annotation files from NCBI: Manis pentadactyla (GCF_030020395.1) and Manis javanica (GCF_001685135.1). We employed the rtracklayer package(Version 1.60.1) within the R software environment to extract locus information for each pseudogene from the GTF files.Subsequently, we generated a BED file, which included the chromosome, start position, end position, and pseudogene ID for each pseudogene, and exported it for further use in extracting pseudogene sequences. To identify pseudogenes associated with anti-infection and immunity among the vast array of pseudogenes, we installed the bedtools software (Version 2.27.1) on a Linux-based server. Utilizing the getfasta command within bedtools, we extracted sequences for each pseudogene annotated with immune function-related information, as indicated in the BED file, from the FASTA reference genome sequence files. We then used the extracted pangolin pseudogenes as input for alignment against NCBI's non-redundant protein sequence database (nr database) using BLASTx software. We filtered for homologous genes from related species with an E-value threshold of less than 1e-5 and a similarity level exceeding 80%.When the number of sequences meeting these criteria surpassed 10, we selected the top 10 sequences with the highest scores. Finally, employing keywords such as "infection" and "immunity," we conducted automated screening and analysis of immunity-related pseudogenes using Python. Our methods included modular design, regular expression matching, data validation, and classification.

Bioinformatics analysis of pseudogenes

This study primarily employs conventional bioinformatics methodologies to execute a suite of analyses, including sequence trimming, multiple sequence alignment, phylogenetic tree construction, similarity heatmap generation, and variation statistics. Moreover, it delineates exons within gene sequences by leveraging key elements such as promoters and stop codons. Building upon these findings, the study further engages in functional optimization and conducts a comprehensive detection and analysis of conserved and specific sequences.

Deep learning of pseudogene sequences

Based on the genomic sequence characteristics, we developed a BiLSTM model[40] and an autoencoder, both grounded in the LSTM[41] deep learning algorithm. Figure 1 illustrates the schematic of the BiLSTM network structure. The BiLSTM model incorporates bidirectional LSTM coupled with an attention mechanism, designed to capture long-range dependencies and identify critical regions within the sequence. This approach facilitates the determination of pseudogenes and the conservation analysis of their functions. Concurrently, the autoencoder is employed for dimensionality reduction and unsupervised feature extraction, aiding in the identification of potential structural fe-atures.

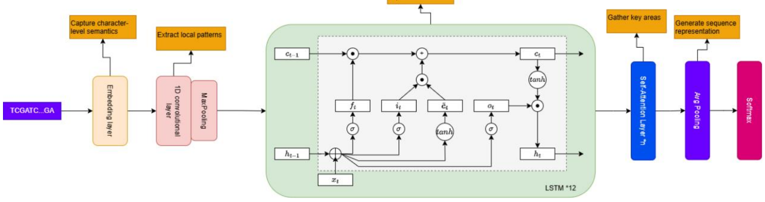

Figure 1: Network Structure of BiLSTM Mode.

(1)Improvement of Position Encoding Formula

Based on the BiLSTM-Autoencoder framework, we introduce a novel dynamic weight positional encoding (DWPE) mechanism to effectively capture the positional information of genomic and proteomic sequences. The proposed encoding formula is designed as follows:

PEpos, 2i = $\sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right)\times(1+\lambda\bullet \mathrm{sigmoid}(hpos\text{-}1\bullet Wp))$ ①

PEpos, 2i+1 = $\cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right)\times(1+\lambda\bullet \mathrm{sigmoid}(hpos-1\bullet Wp))$ ②

In this formulation, pos represents the sequence position, i denotes the dimension index, dmodel indicates the model's dimensionality, λ serves as a learnableparameter, hpos−1 corresponds to the hidden state of the preceding position, and W_p is the positional weight matrix. Through dynamic adjustment of the positional encoding based on the hidden state of the previous position hpos−1, themodel effectively captures the sequential dependencies and enhances its perception of positional relationships within the sequence.

(2) Integrate the formula of the multi-head self-attention bidirectional LSTM unit:

(2-1) Bidirectional LSTM Hidden State Calculation:

Forward LSTM hidden state:

$\overrightarrow{h}_t = \mathrm{LSTMf}\ (xt,\ \overrightarrow{h}_{t-1})$

Backward LSTM hidden state:

$\overleftarrow{h}_t = \mathrm{LSTMb}\ (xt,\ \overleftarrow{h}_{t+1})$

Integrating bidirectional hidden state:

ht = Concat($\overrightarrow{h}_t,\ \overleftarrow{h}_t$)

(2-2) Multi-head self-attention calculation:

Qk = h• W_k^Q

Kk = h• W_k^K

Vk = h• W_k^V

Attention weight:

Attentionk = sofmox($\frac{Q_k K_k^T}{\sqrt{d_k}}$) ③

Output:

Headk = Attentionk • Vk

Integrate multiple outputs:

MultiHeadAttn(h)=Concat(Head1, Head2, …, Headh)•W0 ④

Multi-head self-attention output is covertly fused with bidirectional LSTM:

Hnew = h +MultiHeadAttn(h)

(3) Loss Function

$$L = -\frac{1}{N}\sum_{i=1}^{N} w_i \cdot [y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)]$$ ⑤

Where N is the number of samples, Yi is the true label, $\hat{y}_i$ is the predicted probability, and the dynamic weight wi is calculated as follows:

$$w_i = \frac{freq(y_i)}{\min(freq(0), freq(1))}$$ ⑥

The process begins with the input of sequences such as "TCGA… & GAM…". These sequences are first processed through the "Embed layer", which transforms the discrete input characters into continuous vector representations and extracts "local patterns". Following this, the bidirectional LSTM layer operate-s: the forward LSTM processes the input sequentially (from t=1 to t=T), retaining the hidden state at each time step, while the backward LSTM processes the input in reverse order (from t=T to t=1), also retaining the hidden state. The output at each time step is the concatenation of the hidden states from both directions ([ht, ht]), allowing the model to simultaneously utilize contextual information from both preceding and subsequent positions. The notation "LSTM*12" indicates a multi-layer stack, yet the fundamental bidirectional logic remains consistent.

Post the bidirectional processing, the "Self-Attention Layer" is utilized to concentrate on pivotal regions. Subsequently, the "Decoder" generates a sequence representation, culminating in t-he "Softmax" classifier outputting the results. These methodologies effectively address the limitations of traditional methods in pseudogene analysis, particularly in capturing complex patterns such as nonlinear relationships or concealed features. This is particularly relevant as pseudogenes, while generally similar to functional genes, often contain variations like frameshift mutations or premature stop codons.

Results

The strategy of ticks for blood-sucking and indirect transmission of pathogens

The blood-sucking and indirect pathogen-assisting transmission strategies of the Ixodidae family

Tick salivary proteins are known for their diverse functions, such as inhibiting host coagulation, inflammatory responses, and complement mediated pathways. Key proteins in this category include Salp15, TSLPI, and the 8.9 kDa protein. Ticks belonging to the family Ixodidae predominantly secrete the Salp15 protease, which specifically binds to the $CD4^+$ receptor on the surface of host T cells. This interaction blocks the T cell's engagement with MHC class II molecules, thereby inhibiting the T cell receptor (TCR) signaling pathway.Consequently,this leads to a reduction in T cell proliferation and cytokine secretion, such as IL-2. Furthermore, Salp15 suppresses the antigen-presenting capacity of dendritic cells, thereby impairing the host's adaptive immune response[42,43]. Figure 2a illustrates the conserved region beginning from the methionine (Met) residue corresponding to the promoter's start co-don (AUG).The yellow region exhibits minimal variation, while the blue and cyan regions show variations that may enable Salp15 to adapt to different hosts. For instance, the ANA07190.1 sequence from Ixodes holocyclus and the AAK97817.1 sequence fro-m Ixodes scapularis display consistent color patterns in multiple conserved regions, suggesting analogous and stable functions between these two species. Figure 2b shows closely clustered sequences, such as ABU9365.1 and AAK57817.1, from the same clade,indicating a close evolutionary relationship and the potential inheritance of certain key characteristics from a common ancestor. Figure 2c highlights the peak region of the highly conserved domain in Salp15, representing its functional core. The conservation of amino acids at positions 10-40 varies,which may correspond to the boundaries of different structural domains. The decreased conservation at positions 90-110 suggests that this region could be the active site. Figure 2d reveals the strength of interactions betweenSalp15 and other proteins, with darker colors indicating stronger correlations. These proteins function in concert with Salp15 to suppress the host's inflammatory response and coagulation process, collectively fostering an environment conducive to tick survival and pathogen transmission.

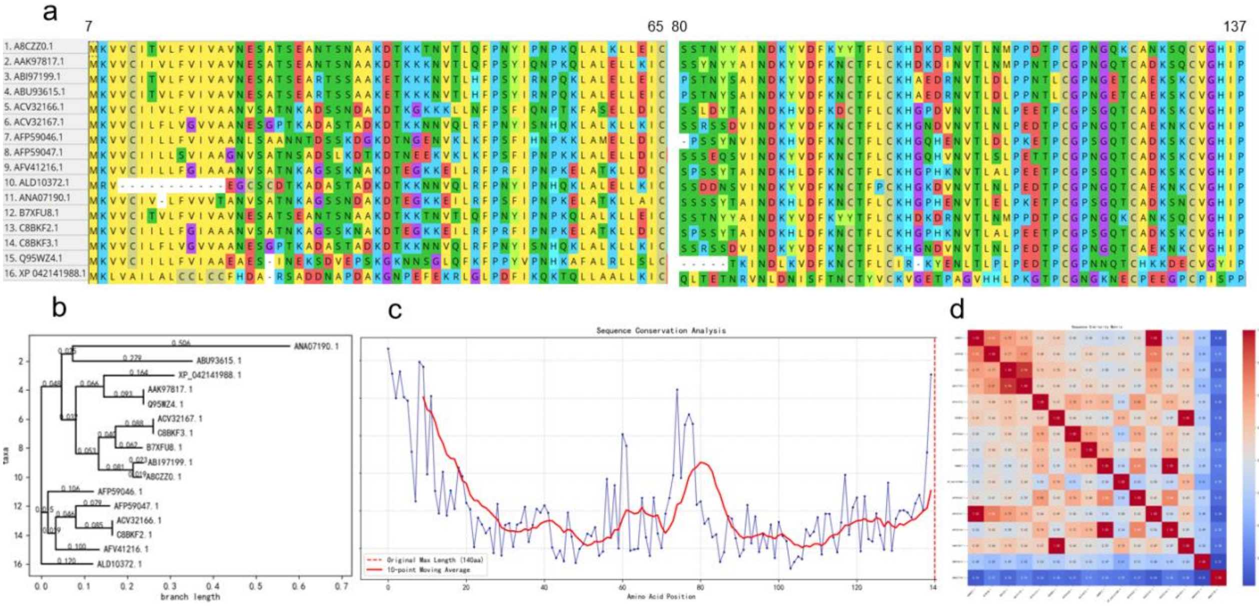

Figure 2: Evolutionary analysis of tick salivary proteinase Salp15 belonging to the Ixodidae family

The blood-sucking and indirect pathogen-assisting transmission strategies of the Ixodes genus

Salivary proteases from the genus Amblyomma are predominantly characterized by serine protease inhibitors of approximately 8.9 kDa, which are pivotal in facilitating pathogen transmission and evading host immune responses[44]. Figure 3a presents a phylogenetic tree analysis where the JAU02506.1 sequence from Amblyomma sculptum and the JAT91749.1 sequence from Amblyomma aureolatum, despite a significant genetic divergence, are grouped within the same clade. This suggests that these sequences may share analogous functional attributes or a common evolutionary lineage. In the sequence conservation analysis depicted in Figure 3b, variations in conservation across different positions are conspicuous. For example, the conservation score of the sequence diminishes progressively from positions 4 to 49 and then steadily increases from positions 70 to 108. This conservation pattern mirrors that observed in Figure 2c, suggesting a convergence in the functional conservation of these two protein classes. Nonetheless, the high variability and low sequence similarity among the 8.9 kDa inhibitors imply that such variations do not necessarily compromise the protein's overall functionality; rather, they might signify adaptations to diverse hosts or ecological niches. In the context of pangolin pseudogenes, two 52 kDa sequences were identified in Manis pentadactyla, with one each in Manis javanica and Smutsia gigantea. To date, no correlation has been established between these sequences and the 8.9 kDa protein found in ticks. However, the host specificity of Amblyomma javanense for Manis javanica is of particular interest. The potential role of these 52 kDa sequences in the pangolin's immune regulation or defense against tick-borne pathogen infections warrants further investigation.

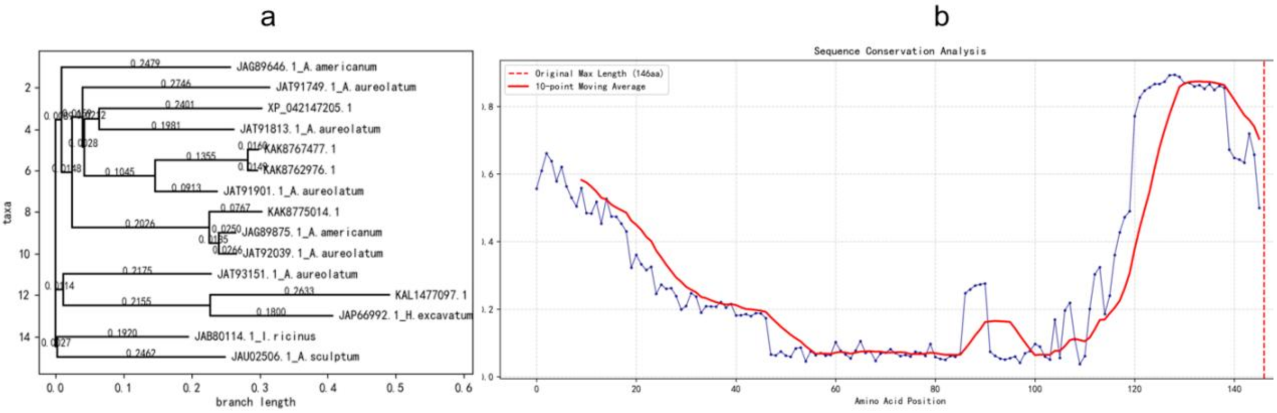

Figure 3.Evolutionary analysis of tick saliva proteinase 8.9 kDa belonging to the Ixodes genus

Pathogen invasion and adaptive evolution

Virus Invasion and Adaptive Evolution

In the intricate ecology of virus-host interactions, the TIM-1 receptor plays a pivotal role in numerous biological processes.To inIn the intricate ecology of virus-host interactions, the TIM-1 receptor plays a pivotal role in numerous biological processes. To investigate the mechanism of viral invasion, we conducted a comprehensive screening and analysis using the human TIM-1 protein sequence (accession number: AF066592.1) as a reference. Our findings revealed that human TIM-1, non-human primate TIM-1, and Hepatitis A Virus Cellular Receptor 1 (HAVCR1) – which are the same protein but named differently based on their distinct functional roles – exhibit a high degree of sequence similarity to the HAVCR1 sequence of pangolins. As illustrated in the sequence similarity matrix (Figure 4a), the similarity between the pangolin sequence (accession number: XP_036767594.1) and the human sequence (AF066592.1) is 0.555, while the similarity between the pangolin sequence and the western gorilla (Gorilla gorilla, accession number: BAJ61036.1) sequence is 0.544. Furthermore, a significant number of conserved amino acid fragments were identified across these species. Such high sequence similarity serves as critical evidence for gene or protein homology, suggesting a shared evolutionary origin from a common ancestral gene. Regarding receptor functionality, Zhang et al. demonstrated that the TIM-1 receptor facilitates the invasion of enveloped viruses by recognizing phosphatidylserine (PS) on the viral envelope[3]. Given the observed sequence similarity between pangolin HAVCR1 and TIM-1 from other species, it is plausible that pangolin HAVCR1 may also play a role in mediating viral infection or participating in immune responses.

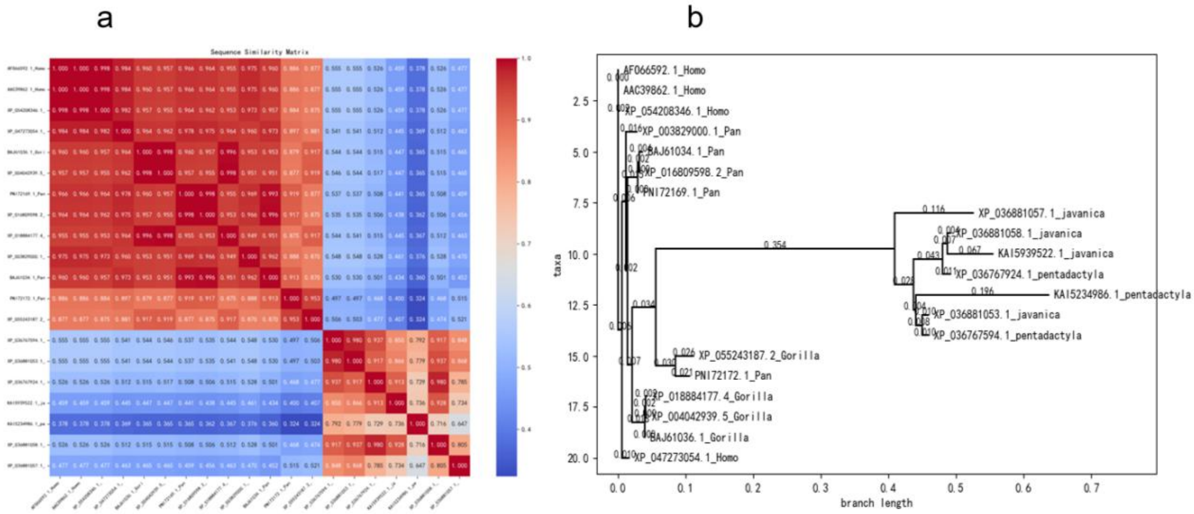

Figure 4.Diagram of the Relationship between Tick Virus Infection and Host Cell Receptors

A detailed analysis of the pangolin HAVCR1 receptor protein sequence revealed no amino acid variations indicative of gene pseudogenization. Considering the established role of TIM-1 in mediating viral invasion, this suggests that pangolins may employ additional, yet uncharacterized, defense mechanisms in response to tick-borne viral infections, beyond those potentially involving HAVCR1. The phylogenetic tree (Figure 4b) demonstrates that the positions of different species accurately reflect their genetic relationships. Although pangolins, humans, and gorillas exhibit considerable evolutionary divergence, they share specific genetic connections. This strongly supports the hypothesis that these species likely descended from a common ancestor early in their evolutionary history, with subsequent divergence leading to the retention of homologous characteristics.

Bacterial invasion and adaptive evolution

Host infection by tick-borne bacterial pathogens involves intricate immune evasion mechanisms. Research has demonstrated that tick-borne bacteria, such as Rickettsia, often evade host immune responses by inhibiting T cell activation signaling pathways. The T cell receptor (TCR) signaling pathway is critical for T cell activation, proliferation, and the execution of immune functions. These bacteria can secrete specific virulence factors that disrupt key molecules in TCR signal transduction, thereby preventing normal T cell activation. This impairment of the host's cellular immune function creates favorable conditions for bacterial survival and replication[45]. In the analysis of the leukocyte immunoglobulin-like receptor *(LILR)* subfamily in pangolins, it was observed that the Chinese pangolin (*Manis pentadactyla*) possesses 19 protein factors belonging to *LILRA5, LILRA6*, and *LILRB3*. Among these, *LILRA6* is particularly abundant, comprising 14 of these factors, whereas the Malayan pangolin (*Manis javanica*) has only one *LILRA6* protein factor. Notably, the LILR profiles of these two pangolin species differ significantly from those of humans. The human LILR family consists of 11 functional proteins, encoding five activating receptors (*LILRA1, LILRA2, LILRA4–LILRA6*), five inhibitory receptors (*LILRB1–LILRB5*), and one soluble receptor (*LILRA3*). These receptors modulate immune cell responses by recognizing pathogen-secreted ligands: activating receptors enhance antibacterial immunity by promoting phagocytosis, cytokine secretion, and oxidative burst, while inhibitory receptors suppress excessive immune responses to prevent tissue damage[46]. Given the unique presence of only one LILRA6 protein factor in the Malayan pangolin, its amino acid sequence was used as a template to retrieve and align sequences from other species with the highest coverage and similarity. The phylogenetic tree in Figure 5a illustrates the evolutionary relationships among different *LILR* proteins. The branching structure reveals distinct evolutionary trajectories for *LILR* proteins in Chinese and Malayan pangolins. Notably, the Malayan pangolin's *LILRA6* is most closely related to the XP_057352609.1 protein factor of the Chinese pangolin. In terms of clustering, it groups within the same clade as the sequence KAK2491512.1 from the giant pangolin (Smutsia gigantea) and the sequence XP_059941245.1 from Blainville's beaked whale (Mesoplodon densirostris), suggesting functional similarities among these proteins.

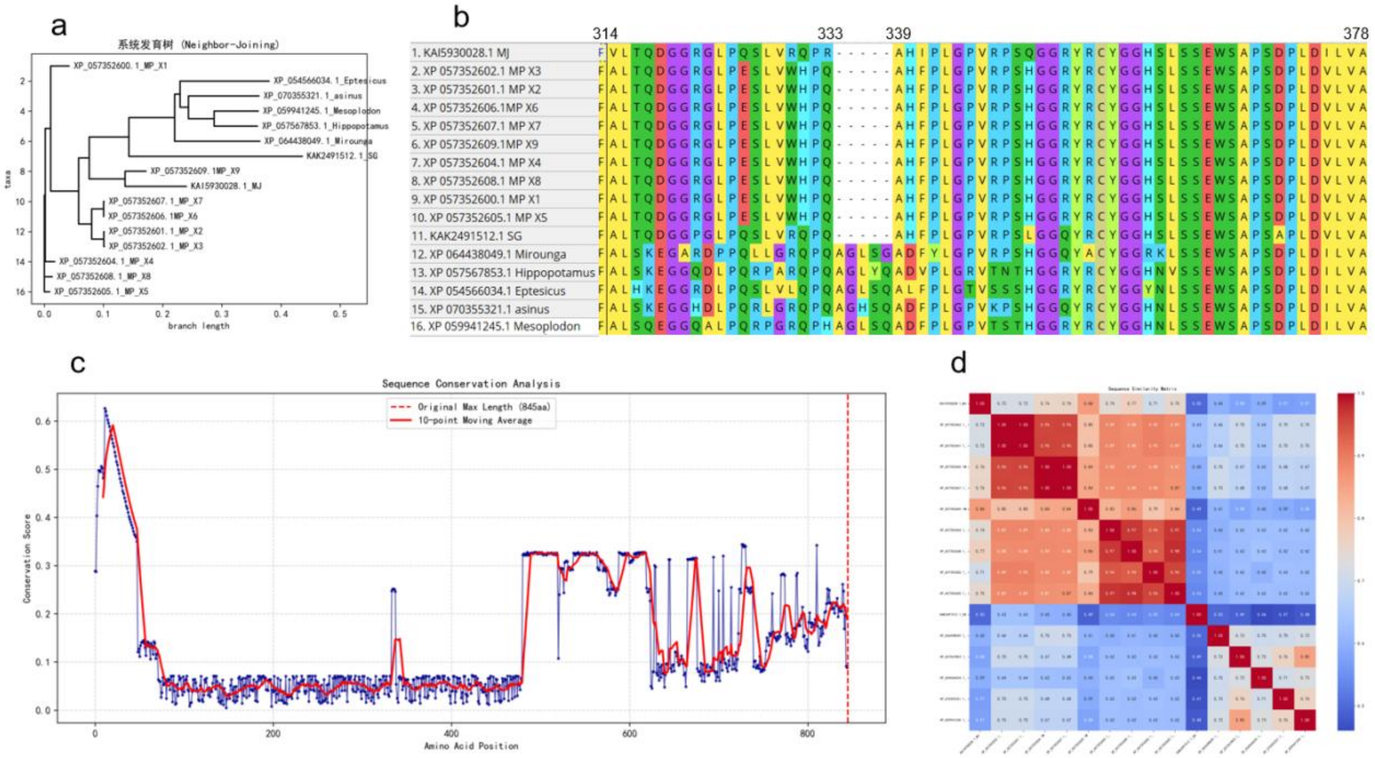

Figure5.Analysis of amino acid sequence of the LILRA6 protein of the Malayan pangolin (KAI5930028.1)

Figure 5b presents the results of multiple sequence alignment, with color-coded regions indicating the conservation of amino acids at specific positions. All pangolin sequences analyzed lack amino acids at positions 334–338. In other species, variations were observed: glutamine (Q) at position 338 in the big brown bat (Eptesicus fuscus) is replaced by glycine (G); serine (S) at position 337 in the hippopotamus is replaced by threonine (T); and leucine (L) at position 336 in the African wild ass (Equus asinus) is replaced by histidine (H). These findings suggest that this amino acid segment is not only prone to deletion but also highly variable, indicating that species undergo adaptive mutations to mitigate excessive inflammatory responses triggered by bacterial invasion. The sequence conservation analysis in Figure 5c quantifies the degree of conservation at each amino acid position. It highlights that the conservation score of the five missing amino acids (-AGLSQ-) between positions 333–339 in pangolins is approximately 0.26, underscoring the unique role of pangolin *LILR* proteins in immune regulation. The heatmap correlation analysis in Figure 5d reveals strong expression or interaction among proteins of the pangolin *LILR* family, explaining how the abundant *LILRA5, LILRA6*, and *LILRB3* protein factors in Chinese pangolins collaboratively regulate immune responses. However, their correlation with non-related species is minimal, emphasizing the specificity of the pangolin *LILR* system in immune regulation mechanisms.

Screening of pangolin pseudogenes and analysis of their immune association

Using the rtracklayer package in R software, we identified 3,209 and 2,035 pseudogenes from the whole-genome sequences of the Chinese pangolin (*Manis pentadactyla*, assembly ID: GCF_030020395.1) and the Malayan pangolin (*Manis javanica*, assembly ID: GCF_001685135.1), respectively. Subsequently, we employed the pangolin pseudogenes as templates to perform sequence alignment against the non-redundant protein sequence database (nr database) of the National Center for Biotechnology Information (NCBI) using BLASTx software. A total of 62,124 homologous gene receptors (including various isoforms) were initially screened. After modular filtering, 413 homologous gene receptors were retained, accounting for 0.67% of the total. Following the exclusion of isoforms, 94 homologous gene sequences were ultimately identified, which belonged to 10 distinct immune systems (see Table 1 for details). In Table 1, the rows corresponding to peptidyl-prolyl cis-trans isomerases, small inducible cytokine subfamilies, and inhibitors of protein kinases involve relatively few sequences and species. Notably, the pseudogenes of leukocyte immunoglobulin receptors and interferon receptors exhibited high consistency: the top 10 sequences with the highest coverage and similarity were all derived from Chinese pangolins and Malayan pangolins. This suggests that these two types of immune receptor genes may possess species-specific characteristics or unique evolutionary features in pangolins. Analysis of all immune pseudogenes revealed an imbalance in the host's defense mechanisms against infections, with antibacterial receptors typically playing a dominant role. This is clearly illustrated in Pie Chart 6a (Figure 6a) . The host species immune factor network (Figure 6b) demonstrates the sharing of immune receptor families or genes in anti-infection immunity, implying that related species may exhibit similarities or connections in terms of immune function, disease susceptibility, and other aspects.

Table 1: List of Automatically Retrieved Homologous Genes of Pangolin Immune Receptor Pseudogenes

No.	Sequence name	Immune Receptor Name	Representative species	Receptor category	Receptornumber	Involving the species number
1	NW_016532664.1	toll-like receptor	Manis pentadactyla; Tursiops truncatus	_1,7,12; X3	21	19
2	NW_016538650.1; NW_016545677.1	leukocyte immunoglobulin-like receptor	Manis pentadactyla	A_5,6; B3; X_1~19	20	2 (pangolin)
3	NW_016538972.1	immunoglobulin superfamily	Manis pentadactyla	Superfamily 22	9	8
4	NW_016545790.1	interferon-inducible	Manis javanica	GTPase 5-like	11	11
5	NW_016547093.1	peptidyl-prolyl cis-trans	Bos taurus	FKBP3	1	2
6	NW_016557009.1	small inducible cytokine subfamily	Homo	E_1	2	2
7	NW_016562793.1	interferon alpha/beta receptor	Manis javanica	_2; X1_5	10	2 (pangolin)
8	NW_016563014.1	Immunoglobulin heavy variable	Manis javanica	3_23	9	4
9	NW_016569945.1	suppressor cytokine signaling	Manis javanica	signaling 5	10	11
10	NW_016591780.1	52 kDa	Manis javanica	TPA	1	2

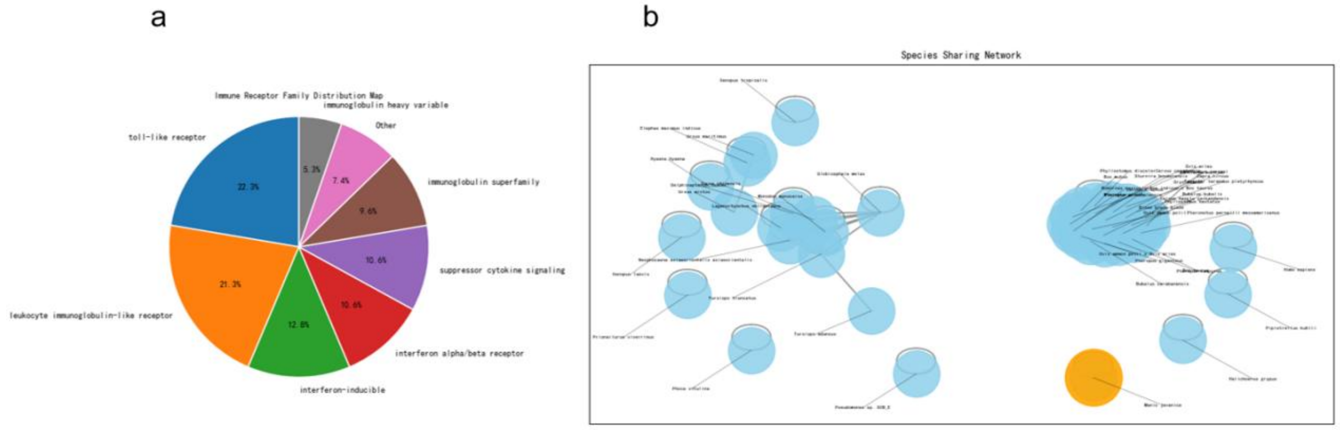

Figure 6.Pie chart and network diagram of immune-related receptors in the host organism

Analysis of amino acid sequence characteristics and functional domain mutations of interferon interference in hosts

Upon viral invasion, host cells secrete interferon-alpha/beta (IFN-α/β), which binds to interferon receptors (IFNAR1/IFNAR2) on the cell membrane. This binding activates intracellular JAK1/TYK2 kinases, leading to the phosphorylation of STAT1/STAT2. The phosphorylated STAT1/STAT2 proteins form a complex that induces the expression of antiviral genes, thereby inhibiting viral replication[47]. Conservation analysis (Figure 7c) reveals that the amino acid region spanning positions 100–300 exhibits high conservation (moving average ≥ 0.20). This region corresponds to the binding site between IFN-α/β and the D1 domain of the IFNAR2 receptor. Relying on a “WSXWS”-like motif, this interaction upregulates the expression of MHC-I molecules, enhancing the recognition of infected cells by CD8+ T cells. Simultaneously, it activates NK cells to promote the clearance of virus-infected cells[48]. As shown in Figure 7a, the Malayan pangolin pseudogene XP_036870501.1 harbors a mutation in this region, altering the sequence from “LYAIVYISLV” to “LYPMVYISLV” (highlighted in red), which reduces ligand-binding affinity. For the KA15941445.1_MJ sequence, the amino acids at positions 50–60 (“MLVSQNASAIRPR”, highlighted in blue) are identical to those of functional receptors. However, insertions/deletions are present in the C-terminal extension region (beyond position 300), which impairs receptor dimerization. Since the tyrosine residue “YVIDKLIPNT” in the intracellular domain requires dimerization to be phosphorylated by JAK kinases, this impairment also affects the recruitment of STAT proteins[49]. Additionally, the short isoform pseudogene XP_036870502.1 (Figure 7a) contains only 200 amino acids, resulting in the loss of the intracellular domain and thus the inability to activate downstream signaling.

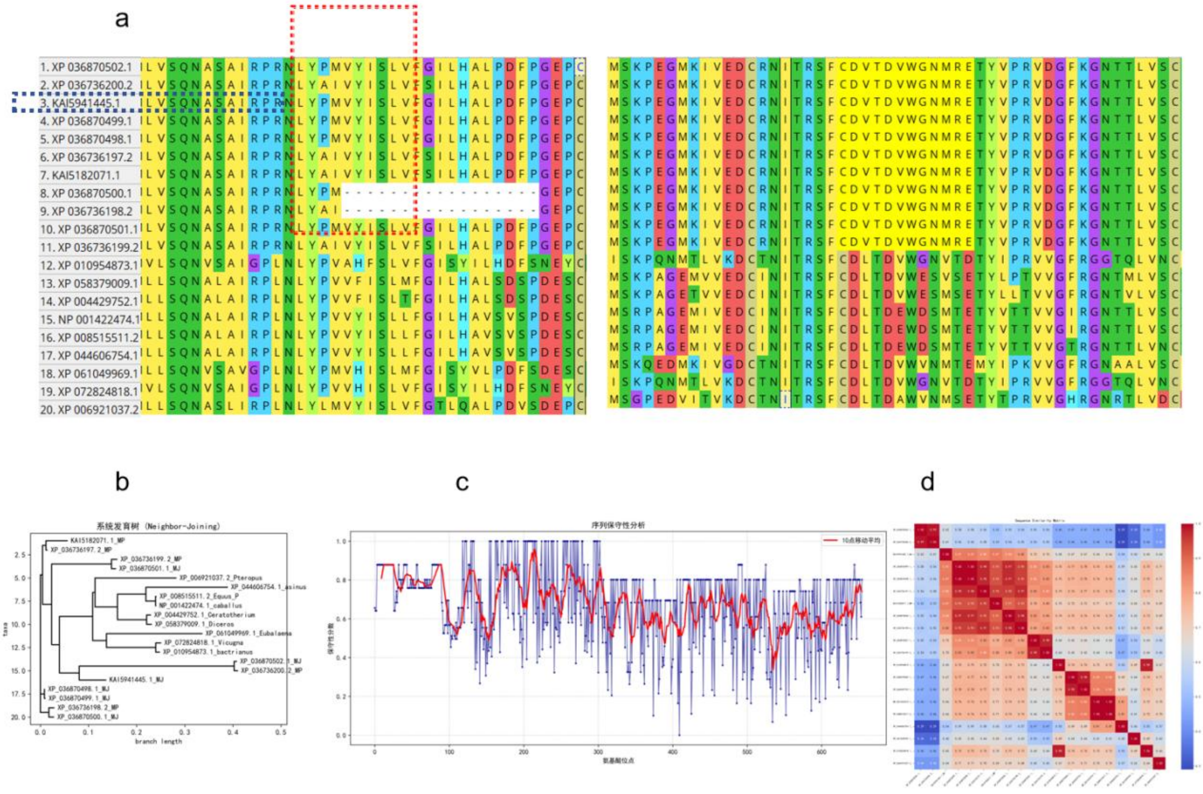

Figure 7.Analysis diagram of interferon pseudogenes of pangolins and related species

Functional genes typically contain a conserved “C-XX-C” motif (with cysteines separated by 8 amino acids), which forms a disulfide bond to stabilize the protein structure and contributes to high conservation[50]. The sequence XP_036736197.2 retains an intact “CDV TDVWGNMRE” motif (highlighted in yellow); therefore, it cannot be classified as a pseudogene based solely on ligand-binding potential. Nevertheless, pseudogenes may exhibit cysteine deletions. For instance, variations in the “EQSGRIVKKHKPK” motif—the binding site for JAK1/TYK2 kinases in the intracellular domain — can lead to pseudogenization[48]. An example is XP_044606754.1, where amino acid substitution in this region results in the sequence “EKSGSIVNLHRPK”, weakening its ability to bind kinases. The sequence XP_044606754.1 from the African wild ass (Equus asinus) contains a unique insert (“MNALGPEACWG PPPVSCPHPSLRWALEK”) in the first 40 amino acids of the N-terminal region, which differs significantly from the typical N-terminal sequence of interferon receptors (starting with “MLVSQN...”), indicating a loss of function. For XP_036870501.1, the deletion of “GWVFPE” in the “NFCNRSGWVFPE” region may disrupt the receptor conformation or cause reading frame abnormalities due to frameshift mutations. XP_036870500.1 lacks the “VYISLVFGIL” sequence in the N-terminal “LYPMGEPCV LK” region, suggesting the loss of functional domains. Furthermore, in the intracellular signaling domain of KAI5182071.1, “EVIH VNR” is replaced by “EVIHINR”; this substitution blocks JAK-STAT signal transduction. If this pseudogene encodes a secreted receptor, it may exhibit interferon-neutralizing activity in bodily fluids.

Another important function of pseudogenes is to form CpG islands that mediate regulatory processes. CpG islands are located in the 5′ UTR and promoter regions of genes; their high GC content corresponds primarily to regions enriched with glycine (G) and proline § in the N-terminal of amino acid sequences, where repeats of “GPGPC” and “GPGPW” are common[51]. Figure 7a shows that KA15941445.1 has enriched sequences “PGEPCVLK” and “GNTTLVSCTGS” in the N-terminal region, indicating the presence of key CpG island sites upstream. Additionally, XP_044606754.1 contains a unique insert in its N-terminal region, which may correspond to expression silencing caused by mutations in the promoter region. Moreover, the level of sequence similarity reflects differences in gene function. Figure 7b demonstrates that the functional interferon receptors XP_036736197.2 and XP_036870502.1 (from congeneric pangolin species) cluster closely in the phylogenetic tree, with a similarity ≥ 0.99—reflecting high conservation during evolution. In the similarity matrix (Figure 7d), KA15941445.1 shares a similarity of 0.87 with XP_036870499.1 but only 0.64 with XP_006921037.2 (from a distantly related Pteropus species), suggesting that KA15941445.1 is likely a pseudogene with functional degradation.

Analysis of TLR-like receptor gene mutations and the elucidation of the host pathogen resistance mechanism

TLR1 pseudogene cluster exhibits the highest homology with pangolins of the Manis genus, with an amino acid sequence similarity > 95% (see Figure 8c). However, the leucine-rich repeat (LRR) domain—a region rich in leucine repeat sequences—shows disruptions in key functional regions (amino acids 10–69, Figure 8a), resulting in an overall structural modeling error as high as 65.986 (see Table 2). This error is significantly higher than that of the KAI5188340.1 gene from a congeneric pangolin species (modeling error: 28.149). Modeling error reflects the structural stability of the protein encoded by a gene; a high error value indicates potential disorganization of the protein's three-dimensional structure, leading to a significant reduction in ligand-binding ability. Nevertheless, even within the same species, highly variable TLR1 genes exist. For example, the XP_036764368.2 sequence from the Chinese pangolin (Manis pentadactyla) has a modeling error of 68.25. This confirms significant sequence variation and differences in structural stability of TLR1 genes within the same species. Such gene variations with high modeling errors may result from functional differentiation, adaptive selection, or the accumulation of neutral mutations during evolution, reflecting the molecular strategy by which species maintain genetic diversity to cope with environmental pressures or pathogen evolution. Marine mammals, such as the blue whale (Balaenoptera musculus, XP_036709041.1) and the minke whale (Balaenoptera acutorostrata, XP_007178885.2), exhibit highly conserved TLR1 sequences (modeling error < 50, similarity > 0.95). This reflects the need for TLRs with relatively stable functions in the marine environment, where pathogen diversity is relatively low.

Analysis of the TLR12 receptor cluster (Figure 8b) shows that the XP_013375166.1 sequence from the chinchilla (Chinchilla lanigera) and the XP_037005439.2 sequence from the Jamaican fruit bat (Artibeus jamaicensis) have relatively low modeling errors, at 32.31 and 35.233, respectively (see Table 2). These errors are significantly lower than those of the common bottlenose dolphin (Delphinus delphis). Additionally, the similarity heatmap (Figure 8d) shows that their similarity to the core TLR12 cluster is generally ≥ 0.6. For instance, the similarity between XP_036917395.1 (from the Honduran yellow-shouldered bat, Sturnira hondurensis) and XP_037005439.2 reaches 0.935. No frameshift mutations or domain deletions were detected in these sequences, suggesting relatively intact functionality. The XP_008155195.2 sequence from the big brown bat (Eptesicus fuscus) serves as a representative of TLR12 pseudogenes. Its LRR domain retains intact conserved segments, such as amino acids 6−11 and 13−19 (see Figure 8b). However, amino acid substitutions occur in the Toll/interleukin-1 receptor (TIR) domain—specifically in regions related to signal transduction—which may affect the binding efficiency of downstream MyD88. This observation is consistent with the "partial recognition-limited activation" immune strategy of bats.

Table 2: Overview of Artificial Retrieval and Reconstruction Errors of TLR Receptor Null Genes in Malayan Pangolins

NO.	error burst	sample number	sequence ID	ReconstructionError	Species name
1	50-70	14	XP_073093842.1	67.715	Manis javanica
2			XP_036764368.2	68.25	Manis pentadactyla
3			XP_013375166.1	59.187	Chinchilla lanigera
4			KAK2502118.1	65.986	Smutsia gigantea
5			XP_036709041.1	69.45	Balaenoptera musculus
6			XP_061047302.1	68.686	Eubalaena glacialis
7			XP_007178885.2	68.957	Balaenoptera acutorostrata
8			XP_035968401.1	69.113	Halichoerus grypus
9			XP_032256081.1	69.021	Phoca vitulina
10			AZY91593.1	50.08	Sousa chinensis
11			AZY91589.1	50.248	Tursiops truncatus
12			XP_033712817.1	69.92	Tursiops truncatus
13			XP_030736898.2	69.558	Globicephala melas
14			XP_004419033.1	68.152	Ceratotherium simum simum
15	<50	11	AZY91591.1	49.915	Delphinus delphis
16			KAI5188340.1	28.149	Manis pentadactyla
17			XP_036302090.1	35.934	Pipistrellus kuhlii
18			XP_008155195.2	36.014	Eptesicus fuscus

NO.	error burst	sample number	sequence ID	ReconstructionError	Species name
19			XP_003415338.1	34.854	Loxodonta africana
20			XP_020031889.1	31.055	Castor canadensis
21			XP_028369613.1	35.639	Phyllostomus discolor
22			XP_045715276.1	35.678	Phyllostomus hastatus
23			WKA14365.1	35.985	Equus przewalskii
24			XP_036917395.1	35.07	Sturnira hondurensis
25			XP_037005439.2	35.233	Artibeus jamaicensis

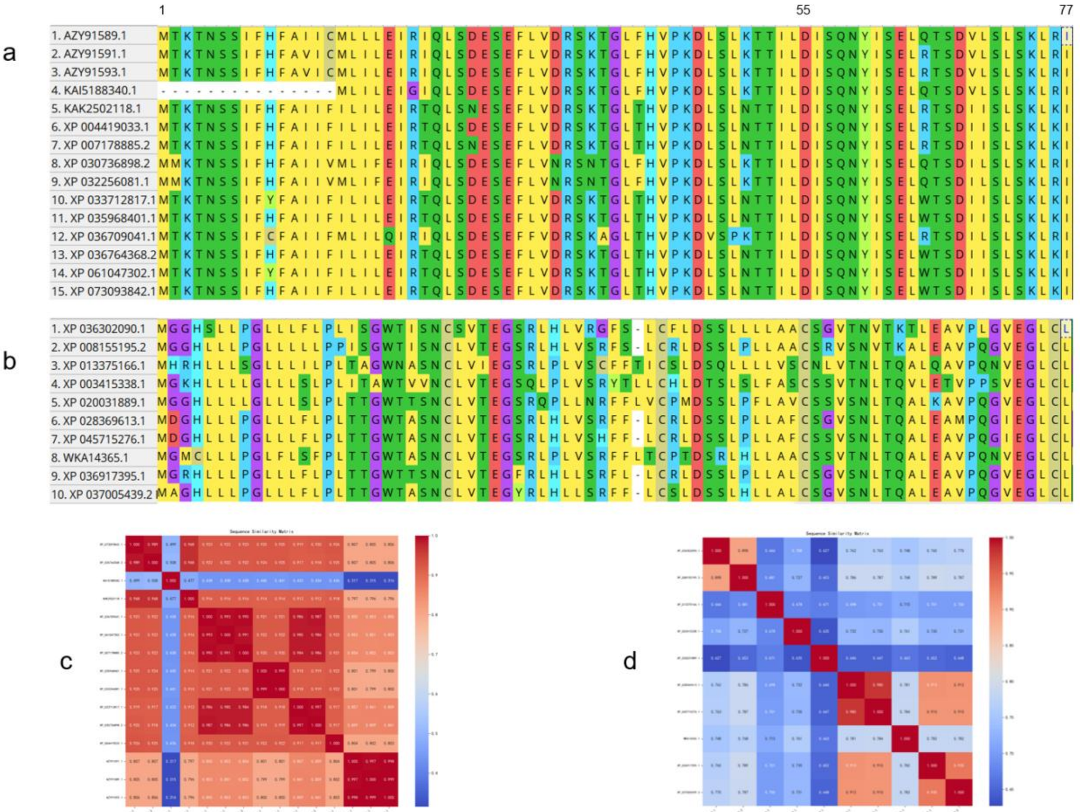

Figure 8.Sequence alignment and similarity diagram of pangolin pseudogene and homologous genes of related species' proteins

Discussion

The immunosuppressive mechanism of tick saliva proteins and the strategies for pathogen transmission

During their parasitic process, ticks secrete a variety of salivary proteins to suppress the host's immune response, thereby creating a favorable environment for blood-feeding and pathogen transmission. Among these, the salivary protein Salp15, secreted by ticks of the Ixodidae family, plays a pivotal role in this process. Salp15 specifically binds to CD4$^+$ receptors on the surface of host T cells, blocking the T cell receptor (TCR) signaling pathway and inhibiting T cell proliferation as well as cytokine secretion (e.g., IL-2). Additionally, Salp15 impairs the antigen-presenting capacity of dendritic cells, significantly reducing the host's adaptive immune response [7,9,42,43]. From an evolutionary perspective, studies have demonstrated that the conserved regions of Salp15 (e.g., amino acids 10–40) remain highly stable over long-term evolution. These conserved regions are likely closely associated with the protein's function in inhibiting the host's Toll-like receptor (TLR)-mediated innate immune signaling pathway (Figure 2c). Conversely, the amino acid region 90–110 represents the active center of Salp15 variation. It is hypothesized that proteins transcribed from these variable regions are involved in suppressing the host's TLR-mediated innate immune signaling pathway, thereby reducing the host's innate immune defenses and facilitating successful blood-feeding and pathogen transmission.

A study by Chaves-Arquero et al. on Salp15 from the black-legged tick (Ixodes scapularis) revealed that this protein is a long polypeptide composed of 135 residues. During secretion, the N-terminal signal sequence (positions 1–21) is cleaved. The mature Salp15 exists as a monomer and features a flexible N-terminal region, while its C-terminus contains three disulfide bridges and one free cysteine. This unique structural configuration enables precise interaction with the two outermost extracellular domains of CD4[52]. However, the tick salivary protease system is highly complex. In their study on Salp15 from Ixodes persulcatus, Jin et al. identified a 15 kDa protein, IpSAP, which directly interacts with the lymphotoxin β receptor (LTβR) and inhibits the host's LTβR signaling pathway. This interaction creates a localized immunosuppressive microenvironment, allowing Borrelia burgdorferi (the causative agent of Lyme disease) to evade immune surveillance during the early stages of infection and facilitating rapid pathogen dissemination[53].

The 8.9 kDa serine protease inhibitor, predominantly found in ticks of the Amblyomma genus, plays a critical role in suppressing host defense mechanisms. This inhibitor disrupts the host's coagulation and inflammatory responses through a complementary mechanism, significantly impairing the host's ability to mount an effective defense. A study by Esteves et al. on Ornithodoros sculptus (the sculpted soft tick) revealed that this species harbors a diverse array of immunity-related salivary proteins, including members of the 8.9 kDa superfamily. Furthermore, the sequence of its 8.9 kDa protein contains 10 cysteine residue sites, which exhibit a high degree of conservation. These abundant salivary proteins operate through a synergistic mechanism, not only facilitating the successful completion of the blood-feeding process but also indirectly creating an "immune escape pathway" for pathogens[54]. Interestingly, pangolins possess several 52 kDa sequences. To date, no research has been published on the functional role of the 52 kDa gene in pangolins. Whether this sequence is linked to the invasive process mediated by the 8.9 kDa salivary enzyme of ticks remains an open question and warrants further investigation.

Adaptive evolutionary game between pathogens and hosts

During the long-term co-evolutionary process, pathogens and their hosts have developed intricate and diverse adaptive strategies. For instance, a critical step in viral invasion involves the recognition and binding of the host's TIM-1 receptor by viral membrane proteins. Research has revealed that the pangolin HAVCR1 receptor exhibits significant homology with the TIM-1 receptor and HAVCR1 in humans and other non-primate animals, with sequence similarities ranging from 0.544 to 0.555. This suggests a shared ancestral origin. Zhang et al. demonstrated that the TIM-1 receptor facilitates the invasion of various enveloped viruses by recognizing phosphatidylserine (PS) on the viral envelope. In the absence of interferon (IFN), TIM-1 can mitigate the infection and pathogenesis of tick-borne encephalitis virus (TBEV). Furthermore, the absence of TIM-1 has been shown to reduce viral load and tissue pathogenicity, indicating that the pseudogenization of the TIM-1 gene enhances the host's antiviral defense[3]. However, sequence analysis did not detect pseudogenization in the pangolin HAVCR1 receptor. It is hypothesized that pangolins may employ alternative mechanisms to resist tick-borne virus invasion, warranting further investigation to elucidate these pathways.

Research on tick-borne bacteria interfering with the host's T cell receptor (TCR) signaling pathway has revealed that the leukocyte immunoglobulin-like receptor (LILR) family proteins, such as LILRA6, in pangolins exhibit significant species-specificity in immune regulation. Notably, deletions and variations in their amino acid sequences—for example, the deletion of the "-AGLSQ-" segment at positions 333–339—may impair bacterial recognition capabilities. However, this alteration also appears to mitigate excessive inflammatory responses (Figure 5c). Jones et al. (2011) demonstrated that various LILRs bind to human leukocyte antigen (HLA). Variations in HLA alleles alter LILR recognition, potentially contributing to the development of certain diseases. Their research also identified the weakest binding region between LILRs and HLA. Furthermore, they found that the activating receptor LILRA1 and the soluble LILRA3 protein exhibit a stronger binding affinity to the heavy (H) chain of free HLA-C[55]. Jilani et al. (2021) employed an algorithmic approach to develop a computer-generated mutant model. By evaluating the effects of amino acid insertion and deletion mutations, they observed that the model displays functional similarities to experimental mutants, both at the local regions of insertions/deletions and across the full protein scale [56].

This "immune balance" strategy is particularly evident in natural viral hosts such as bats. For instance, the Toll-like receptor (TLR) pseudogenes in bats retain partial ligand-binding ability while blocking signal transduction (e.g., through the deletion of the TIR domain). This mechanism not only limits pathogen replication but also prevents fatal inflammation, fostering a "coexistence" dynamic between pathogens and the host[12]. Similarly, pangolins may possess an analogous immune regulatory mechanism to adapt to the complex pathogenic environments they encounter over long-term evolution.

The functions and evolutionary significance of host immune gene inactivation

The pseudogenization of host immune genes is a hallmark of long-term adaptive evolution driven by pathogenic pressures. A notable example is the TLR1 pseudogene in the Malayan pangolin (*Manis javanica*), where mutations in critical sites or regions of its leucine-rich repeat (LRR) domain result in reduced ligand-binding capacity while preserving its fundamental recognition function (Figure 8a). In contrast, the TLR12 pseudogene in bats utilizes a "decoy mechanism": it binds pathogenic RNA without initiating signal transduction. This mechanism effectively curtails excessive interferon production, thereby striking a balance between immune clearance and tissue protection. Dey et al. (2022) conducted a comprehensive analysis of nine Toll-like receptor (TLR) genes (tlr1–tlr9) across 36 mammalian species. Employing maximum likelihood and Bayesian inference methods, they identified two distinct clades, both of which retain the structural integrity of the LRR domain[57]. Within the TLR1 subfamily, members form heterodimers. Intriguingly, the dimerization and ligand-binding residues in the crystal structures of TLR1 and TLR6 are interchangeable, enabling the creation of chimeric proteins. These pattern recognition receptors play a pivotal role in mediating inflammatory and innate immune responses triggered by pathogen invasion. This finding offers a theoretical foundation for understanding the mutations observed in key sites of the LRR domain within the TLR1 pseudogene of the Malayan pangolin.

In the context of viral resistance, mutations in the functional domains of pangolin interferon pseudogenes (e.g., the "WSXWS" motif in XP_036870501.1) result in diminished ligand-binding affinity. However, the GC-enriched region of the CpG island at the N-terminus may influence the expression of adjacent functional genes through epigenetic regulatory mechanisms[51]. This alteration could modulate the host's immune response to viral infections to some extent. Moreover, the accumulation of pseudogenes is closely linked to the host's ecological niche. Marine mammals, such as cetaceans, inhabit a relatively less complex pathogenic environment and have transitioned from terrestrial to aquatic habitats, leading to adaptive changes in their innate immune mechanisms. In such scenarios, selective pressure is reduced, allowing Toll-like receptor (TLR) genes to evolve into highly conserved sequences. In contrast, terrestrial species, including pangolins and bats, face more diverse and intense pathogenic pressures. These species tend to eliminate redundant genes while preserving critical immune pathways, such as the TLR2/6 heterodimer, to maintain essential defensive capabilities[58].

This "functional screening" mechanism optimizes the host's immune resource allocation during evolution. However, it may also compromise the host's resistance to emerging pathogens, increasing susceptibility. To some extent, this explains why pangolins are often regarded as potential intermediate hosts for certain pathogens—likely due to specific variations in their leukocyte immunoglobulin-like receptor (LILR) system. Future research should further investigate the relationship between the pseudogenization of pangolin immune genes and pathogen infection. Such studies will enhance our understanding of pangolins' ecological roles and the mechanisms underlying disease transmission.

The advantages and challenges of integrating deep learning for infection resistance analysis

This study integrates traditional bioinformatics with deep learning approaches to analyze tick-borne pathogen-host interactions and the functional roles of pangolin pseudogenes, offering a novel paradigm for investigating complex immune mechanisms. Traditional bioinformatics relies on methods such as BLASTx sequence alignment and HMMER domain analysis. Screening pseudogenes based on stringent thresholds (e.g., amino acid deletions in the leucine-rich repeat [LRR] domain of the TLR1 pseudogene) provides a standardized framework for studying functional attenuation[59]. However, its reliance on manual thresholds limits its ability to identify low-homology pseudogenes (e.g., truncated pseudogenes associated with immune escape) and capture nonlinear relationships within pathogen-host interaction networks [60].

The BiLSTM (Bidirectional Long Short-Term Memory) + autoencoder deep learning model addresses these limitations. It effectively captures long-range sequence dependencies and accurately identifies the "ED??Y" motif in the Toll/interleukin-1 receptor (TIR) domain. For example, it reveals the potential role of the pangolin pseudogene KAI5188340.1 in regulating the NF-κB pathway by competitively binding to TRAF6[61]. Additionally, the autoencoder unsupervised extracts conserved short "GPGPC" repeat sequences, linking structural stability to the retained functions of pseudogenes and expanding the application of immunogenic peptide prediction methods in pseudogene analysis[62]. Current models are constrained by the scarcity of annotated pseudogene functional data. Although cross-species transfer learning mitigates data insufficiency, it still struggles with reduced prediction accuracy for low-homology pseudogenes (similarity < 50%)[63]. Furthermore, the decision logic of these models lacks traceability: while the attention mechanism highlights key regions, it cannot quantify molecular evolutionary events such as frameshift mutations and functional attenuation [64]. Future efforts should focus on integrating multi-omics data with functional experiments to establish a closed loop between computational predictions and experimental validation, thereby enhancing prediction accuracy [65]. Additionally, graph neural networks (GNNs) can be combined with AlphaFold2 modeling to analyze the three-dimensional structures of conserved sequences (e.g., "GPGPC") and protein interaction mechanisms, improving model interpretability[66,67]. A cross-species pseudogene function scoring system could also be developed to elucidate the adaptive evolution of host pseudogenes under tick-borne pathogen pressure. This can be further explained through the compensatory mechanism of high-frequency mutations in pangolin TLR pseudogenes, providing potential targets for anti-tick vaccine design.

Although the application of BiLSTM in pseudogene immune analysis faces challenges related to data availability and interpretability, deep learning has emerged as a core tool for deciphering the immune mechanisms of non-model species. It will drive the transformation of infectious disease prevention and control toward precise and predictive approaches.

Conclusion

This study reveals the degradation of pseudogenes in the immune system of pangolins, which may help attenuate immune responses and promote coexistence with pathogens. Pangolins suppress inflammation and coagulation responses, facilitating the transmission of tick-borne pathogens. In the future, we will further investigate the functional evolution of pangolin immune receptors and their mechanisms of resistance to pathogens.

References

1. De la Fuente, J., Antunes, S., Bonnet, S., et al.. (2017). Tick-Pathogen Interactions and Vector Competence: Identification of Molecular Drivers for Tick-Borne Diseases. Front Cell Infect Microbiol, 7: 114.
2. Zhang, M.-Z., Bian, C., Ye, R.-Z., et al.. (2025). Human Infection with a Novel Tickborne Orthonairovirus Species in China. The new england journal of medicine, 392: 200 - 202.
3. Zhang, X. - W., Liang, C. - Q., Wang, H. - Z., et al. (2022). T-Cell Immunoglobulin and Mucin Domain 1 (TIM-1) Is a Functional Entry Factor for Tick-Borne Encephalitis Virus. MBio., 13 (1): e0286021.
4. Moller-Tank, S., Kondratowicz, A. S., Davey, R. A., et al. (2013). Role of the phosphatidylserine receptor TIM-1 in enveloped-virus entry. J Virol., 87 (15): 8327 - 8341.
5. Jiang, B. - G., Wu, A. - Q., Jiang, J. - F., et al. (2021). Molecular Detection of Novel Borrelia Species, Candidatus Borrelia javanense, in Amblyomma javanense Ticks from Pangolins. Pathogens, 10: 728.
6. De la Fuente, J., Villar, M., Cabezas-Cruz, A., et al.. (2016). Tick - Host - Pathogen Interactions: Conflict and Cooperation. PLoS Pathog, 12(4): e1005488.
7. Anguita, J., Ramamoorthi, N., Hovius, J. W. R., et al. (2002). Salp15, an Ixodes scapularis salivary protein, inhibits CD4(+) T cell activation. Immunity, 6: 849 - 859.
8. Rana, V. S., Kitsou, C., Dumler, J. S., Pa, U.. (2023). Immune evasion strategies of major tick-transmitted bacterial pathogens. Trends Microbiol, 31 (1): 62 - 75.
9. Akoolo, L., Djokic, V., Rocha, S. C., Parveen, N.. (2021). Pathogenesis of Borrelia burgdorferi and Babesia microti in TLR4-Competent and TLR4-dysfunctional C3H mice. Cell Microbiol, 23 (9): e13350.

10. Sears, J. D., Waldron, K. J., Wei, J., & Chang, C. - H.. (2020). Targeting metabolism to reverse T-cell exhaustion in chronic viral infections. Immunology, 162 (2): 135 - 144.
11. Torina, A., Blanda, V., Villari, S., et al.. Immune Response to Tick-Borne Hemoparasites: Host Adaptive Immune Response Mechanisms as Potential Targets for Therapies and Vaccines. Int. J. Mol. Sci., 2020, 21: 8813.
12. Sharma, V., Hecker, N., Walther, F., et al.. Convergent Losses of TLR5 Suggest Altered Extracellular Flagellin Detection in Four Mammalian Lineages. Mol. Biol. Evol., 2020, 37 (7): 1847–1854.
13. Shi, W. - Q., Shio, M., Que, T. -C., et al.. Trafficked Malayan pangolins contain viral pathogens of humans. Nature Microbiology, 2022, 7: 1259 - 1269.
14. Tian, X. - C., Chen, L., Zhou, J. - F., et al.. Pangolin scales as adaptations for innate immunity against pathogens. BMC Biology, 2024, 22:234.
15. Torosin, N. S., Argibay, H., Webster, T. H., et al.. Comparing the selective landscape of TLR7 and TLR8 across primates reveals unique sites under positive selection in Alouatta. Molecular phylogenetics and evolution, 2020, 152: 106920.
16. Jiang YP， Huo MZ， Cheng Li S. TEINet： a deep learning framework for prediction of TCR-epitope binding specificity［J］. Brief Bioinfo m, 2023, 24 (2): 086.
17. Kim B，Alguwaizani S，Zhou X，et al. An improved method for predicting interactions between virus and human proteins［J］. J Bioinform Comput Biol，2017，15（1）:1650024.
18. Weiskopfa D., Angeloa M. A., Azeredoa E. L.D., et al.. Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. PNAS, 2013, 110 (22): E2046-2053.
19. Li G.Y., Iyer B., Prasath V.B.S., et al.. DeepImmuno: deep learning-empowered prediction and generation of immunogenic peptides for T-cell immunity. Briefings in Bioinformatics, 2021, 22(6): 1-10.
20. Choo, S. - W., Rayko, M., Tan, T. - K., et al.. Pangolin genomes and the evolution of mammalian scales and immunity. Genome Res., 2016, 26 (10): 1312 - 1322.
21. Fischer, H., Tschachler, E. & Eckhart, L..(2020). Pangolins lack IFIH1/MDA5, a cytoplasmic RNA sensor that initiates innate immune defense upon coronavirus infection. Front. In Immunol., 11: 939.
22. Niu, S., Wang, J., Bai, B., et al.. Molecular basis of cross-species TIM-1 interactions with SARS-CoV-2-like viruses of pangolin origin. EMBO J., 2021, 40 (16): e107786.
23. Tam, O. H., Aravin, A.i A., Stein, P., et al.. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature, 453: 534–538.
24. Xu, J. - R. & Zhang, J. - Z.. (2015). Are Human Translated Pseudogenes Functional? Molecular Biology and Evolution, 33 (3): 268.
25. Karreth, F. A., Ala, U., Provero, P. & Pandolfi, P. P.. (2021). Pseudogenes as Competitive Endogenous RNAs: Target Prediction and Validation. Methods Mol Biol., 2324:115-129.
26. Shi, C. - Y., Liu, F. - Z., Su, X. - W., et al. (2025). Comprehensive discovery and functional characterization of the noncanonical proteome. Cell Research, 35: 186–204.
27. Moreau, T. R. J., Bondet, V., Rodero, M. P., & Duffy, D.. (2023). Heterogeneity and functions of the 13 IFN-α subtypes - lucky for some? Eur J Immunol., 53 (8): e2250307.
28. Ahmad, H. I., Jameel, L., Zahra, Q., et al. (2023). Molecular Evolution of Interferon-Epsilon (IFN-ε) Pseudogene Modulates Innate and Specific Antiviral Immunity inManis javanica. Journal of Zoological Systematics and Evolutionary Research, 14: 2949008.
29. Li, B., Zhai, J. - Q., Wu, Y.-J., et al.. Molecular identification of tick-borne Rickettsia, Anaplasma, Ehrlichia, Babesia, and Colpodella in confiscated Malayan pangolins. PLoS Negl Trop Dis., 2024, 18 (11): e0012667.
30. Wei, X. (2022). Studies on the Ixodes metaviromics and phylogeny of Jingmen tick virus in Guangxi (Master’s thesis, Guangxi Medical University).
31. Kollars, T. M. & Sithiprasasna, R. New host and distribution record of Amblyomma javanense (Acari: Ixodidae) in Thailand. J Med Entomol, 2000, 37 (4): 640.
32. Khatri-Chhetri, R., Wang, H. - C., Chen, C. - C., et al.. Surveillance of ticks and associated pathogens in free-ranging Formosan pangolins (Manis pentadactyla pentadactyla). Ticks Tick Borne Dis., 2016, 7 (6): 1238-1244.
33. Li, Y., Chao, S., Zhang, F. - h.., et al.. Age structure and parasites of Malayan Pangolin (Manis javanica). J Econ Anim., 2010, 14 (1): 22 - 25.
34. Parola, P., Cornet, J., Sanogo, Y., et al.. Detection of Ehrlichia spp, Anaplasma spp, Rickettsia spp, and other eubacteria in ticks from the Thai-Myanmar border and Vietnam. J Clin Microbiol., 2003, 41: 1600 - 1608.
35. Hafiz, M. S., Marina, H., Afzan, A. W., & Chong, J. L.. Ectoparasite from confiscated Malayan pangolin (Manis javanica Desmarest) in peninsular Malaysia. UMT 11th International Annual symposiumon Sustainability Science and management, Terengganu, Malaysia, 09–11 Jul 2012, 1225 - 1229.
36. Hassan, M., Hafiz, M. S., Chong, J. L.. The prevalence and intensity of Amblyomma javanense infestation on Malayan pangolins (Manis javanica Desmarest) from Peninsular Malaysia. Acta Trop., 2013, 126: 142 - 145.
37. Dao, T. T. H., Takács, N., Tran. T. N.. et al.. Detection of tick-borne pathogens in the pangolin tick, Amblyomma javanense, from Vietnam and Laos, including a novel species of Trypanosoma. Acta Trop., 2024, 260: 107384.
38. Anonymous. Project report, Ecto and endo parasites of captive animals and birds of Nandankanan zoo. Nandankanan Zoological Park and Orissa Veterinary College. Bhubaneswar (Supported by Central Zoo Authority, New Delhi), 2010, 1 - 79.
39. Burridge, M. J.. Ticks (Acari: Ixodidae) spread by the international trade in reptiles and their potential roles in dissemination of diseases. Bull Entomol Res., 2001, 91: 3 - 23.
40. Schuster, M. & Paliwal, K. K.. Bidirectional Recurrent Neural Networks. IEEE Transactions on Signal Processing, 1997, 45: 2673 - 2681.
41. Hochreiter, S. & Schmidhuber, J.. Long Short-Term Memory. Neural Computation, 1997, 9: 1735 - 1780.
42. Anguita, J., Ramamoorthi, N., Hovius, J. W. R. et al.. Salp15, an ixodes scapularis salivary protein, inhibits CD4(+) T cell activation. Immunity, 2002, 16 (6): 849 - 859.
43. [43] Garg, R., Juncadella, I. J., Ramamoorthi, N., et al.. Cutting edge: CD4 is the receptor for the tick saliva immunosuppressor, Salp15. J Immunol, 2006, 177 (10): 6579 - 6583.
44. Costa, G. C. A., Ribeiro, I. C. T., Melo-Junior, O., et al.. Amblyomma sculptum Salivary Protease Inhibitors as

Potential Anti-Tick Vaccines. Front Immunol., 2021, 11: 611104.

45. Sahni, A., Fang, R., Sahni, S. K. & Walker, D. H.. Pathogenesis of Rickettsial Diseases: Pathogenic and Immune Mechanisms of an Endotheliotropic Infection. Annu Rev Pathol., 2019, 14: 127 - 152.
46. Hirayasu, K. & Arase, H.. Functional and genetic diversity of leukocyte immunoglobulin-like receptor and implication for disease associations. Journal of Human Genetics, 2015, 60: 703 - 708.
47. Sozzani, S., Bosisio, D., Scarsi, M. & Tincani, A.. Type I interferons in systemic autoimmunity. Autoimmunity, 2010, 43 (3): 196 - 203.
48. Kumaran, J., Wei, L. - h., Kotra, L. P. & Fish, E. N.. A structural basis for interferon-alpha-receptor interactions. FASEB J., 2007, 21 (12): 3288 - 3296.
49. Himpe, E. & Kooijman, R. Insulin-like growth factor-I receptor signal transduction and the Janus Kinase/Signal Transducer and Activator of Transcription (JAK-STAT) pathway. Biofactors, 2009, 35 (1): 76 - 81.
50. Ishimoto, N., Park, J. - H., Kawakami, K., et al.. Structural basis of CXC chemokine receptor 1 ligand binding and activation. Nature Communications, 2023, 14: 4107.
51. Bird, A. P.. CpG-rich islands and the function of DNA methylation. Nature, 1986, 321 (6067): 209 - 213.
52. Chaves-Arquero, B., Persson, C., Merino, N., et al.. Structural Analysis of the Black-Legged Tick Saliva Protein Salp15. Int J Mol Sci., 2022, 23 (6): 3134.
53. Jin, L., Jiang, B. - G., Yin, Y. - Z. & Lai, R.. Interference with LTβR signaling by tick saliva facilitates transmission of Lyme disease spirochetes. PNAS, 2022, 119 (47): e2208274119.
54. Esteves, E., Maruyama, S. R., Kawahara, R., et al.. Analysis of the Salivary Gland Transcriptome of Unfed and Partially Fed Amblyomma sculptum Ticks and Descriptive Proteome of the Saliva. Front. Cell. Infect. Microbiol., 2017, 7: 476.
55. Jones, D. C., Kosmoliaptsis, V., Apps, R., et al.. HLA class I allelic sequence and conformation regulate leukocyte Ig-like receptor binding. J Immunol, 2011, 186 (5): 2990-2997.
56. Jilani, M., Turcan, A., Haspel, N., et al.. Assessing the Effects of Amino Acid Insertion and Deletion Mutations. 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 9-12.
57. Dey, D., Dhar, D., Das, S., et al.. Structural and functional implications of leucine-rich repeats in toll-like receptor 1 subfamily. Biosci., 2022, 47: 59.
58. Ishengoma, E. & Agaba, M.. Evolution of toll-like receptors in the context of terrestrial ungulates and cetaceans diversification. BMC Evolutionary Biology, 2017, 17: 54.
59. Velová, H., Gutowska-Ding, M. W., Burt, D. W. & Vinkler, M. (2018). Toll-Like Receptor Evolution in Birds: Gene Duplication, Pseudogenization, and Diversifying Selection. Mol Biol Evol., 2018, 35 (9): 2170 - 2184.
60. Cheetham, S. W., Faulkner, G. J. & Dinger, M. E. (2020). Overcoming challenges and dogmas to understand the functions of pseudogenes. Nat Rev Genet., 2020, 21 (3): 191 - 201.
61. Hu, X. - h., Fernie, A. R., Yan J. - b.. (2023). Deep learning in regulatory genomics: from identification to design. Curr Opin Biotechnol., 79: 102887.
62. Wang, F., Dai, X. - w., Shen, L. - Y. & Chang, S. (2025). GraphEPN: A Deep Learning Framework for B-Cell Epitope Prediction Leveraging Graph Neural Networks. Appl. Sci., 2025, 15 (4): 2159.
63. Devlin, J., Chang, M. - W., Lee, K. & Toutanova, K.. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT 2019, 4171 - 4186.
64. Ribeiro, M., Singh, S. & Guestrin C.. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. NAACL-HLT 2016 (Demonstrations), 97 - 101.
65. Li, Z. - X., Gao, E., Zhou J. - X., et al.. (2023). Applications of deep learning in understanding gene regulation. Cell Reports Methods, 3: 100384.
66. Atwood, J. & Towsley, D. (2016). Diffusion-Convolutional Neural Networks. 30th Conference on Neural Information Processing Systems (NIPS 2016), 1 - 9.
67. Jumper, J., Evans, R., Pritzel, A., et al.. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596: 583 - 589.

Author Contributions

Tengcheng Que#: Methodology, Program development, Create chart & Writing. Zhining Zhang#: Software, Operational & Gene function analysis. Yunlin He#: Software, Validation & Genome analysis. The # symbols represent the first authors, who made the same contributions to the data analysis, charting, and writing of the paper. Qiuyu Wu and Jinying He: Data organization, thesis editing. Xinni Yang: Genome analysis. Panyu Chen and Hong Qiu: Sampling, submission for testing, and data collection. Yankun Liu: Code correction and verification. Hua Zhang: Editing and Submission. Wenjian Liu*: guide, Supervision, Writing - review & editing. The authors marked with * are the corresponding authors of the paper. They made the same contributions to the planning, writing, finalization of the paper, and the acquisition of experimental funds. All authors have read the final manuscript and approved it for publication.

Funding

This work is supported by the National Key Research and Development Program of China (2023YFC2605400); the Guangxi Natural Science Foundation Project, China (2023GXNSFAA 026517); and the Guangxi Key Research and Development Program (GuikeAB22035027, GuikeAB24010148).

Ethical Statement

All tick samples used in this study were collected from Malay pangolins and Chinese pangolins. The collection of these samples strictly adhered to relevant ethical guidelines and complied with local wildlife protection regulations. All tick samples were collected in a legal and ethical manner, ensuring no harm or disturbance to the pangolins or their habitats during the process. The sample collection activities were approved by the National Key Laboratory of Pathogen Microbiology Safety, Guangxi Medical University, Guangzhou Zoo, and other relevant institutions, and complied with both national and international animal ethics standards.

Consent Statement

This study does not involve human participants. All tick samples used in this study were provided under a collaborative framework, authorized by the National Key Laboratory of Pathogen Microbiology Safety, Guangxi Medical University, and Guangzhou Zoo. The use of data from public databases complies with their access terms.

Data Availability Statement

The data used in this study are sourced from the NCBI public database, and the relevant data can be accessed via the following link:https://www.ncbi.nlm.nih.gov/.The corresponding accession numbers are provided in the main text or appendix of the paper.

SUPPORTING INFORMATION

Additional supplementary information is available for download and review in the supplementary information section located on the right-hand side of this article's HTML page.

Pseudogene Alignment Results.xlsx

Protein Sequence Alignment Results.xlsx