MHC Class I-like Gene in Squirrel Monkey Cytomegalovirus and the Evolution of the Virus

Eiichi Hondo[1], Tetsuya Mizutani[2], Hiroshi Shimoda[3], Atsuo Iida[1]

[1] Laboratory of Animal Morphology, Department of Animal Science, Graduate School of Bioagricultural Sciences, Nagoya University, 464-8603 Nagoya, Japan

[2]Center for Infectious Disease Epidemiology and Prevention Research, Tokyo University of Agriculture and Technology, Tokyo 183-8538, Japan

[3] Laboratory of Microbiology, Joint Faculty of Veterinary Medicine, Yamaguchi University, Yamaguchi 1677-1, Japan

Corresponce: Eiichi Hondo, DVM, PhD
e-mail: ehondo@agr.nagoya-u.ac.jp
Tel&Fax: +81-52-789-4185

Abstract

Homology searches between the entire viral peptides and the human peptides revealed a squirrel monkey cytomegalovirus (Saimiriine herpesvirus 4 strain [SBHV4]) S9 protein exhibiting 60.6% amino acid sequence identity across 86.3% of the full-length human HLA-A. This represents the most extensive and highest level of homology observed to date between a viral protein and an Major Histocompatibility Complex (MHC) class I molecule. Molecular modeling of the S9 and HLA-A proteins and subsequent structural comparison confirmed significant similarity, including the antigen-presenting region. Although cytomegalovirus exhibits high host specificity, the presence of the S9 protein in this virus warrants immunological caution in the event of cross-species transmission.

Phylogenetic analysis of MHC class I-like gene sequences corresponding to primate HLA-A and SBHV4 suggested that SBHV4-S9 was likely acquired via LINE-1-mediated horizontal gene transfer from the African Old World monkey *Papio anubis* or a closely related ancestral species approximately 12.9 to 22.2 million years ago. SBHV4-S9 displayed evidence of strong negative selection following its divergence from *Papio anubis*. Furthermore, phylogenetic analysis of the SBHV4-S1 gene and primate SLAMF6-like genes indicated that this gene was acquired after the divergence from the New World monkey *Saimiri boliviensis boliviensis*. Evaluation of the overall genetic similarity between the SBHV4 genome and cDNAs from individual primate species using hierarchical clustering based on Levenshtein distance and Wasserstein distance revealed a higher degree of similarity between the SBHV4 genome and *Papio anubis* cDNA compared to that with *Saimiri boliviensis boliviensis* cDNA. Taken together, these findings suggest that *Papio anubis* has a longer co-evolutionary history with SBHV4 than squirrel monkeys, and that SBHV4 likely spread from Africa to South America, acquiring novel genomic elements in new hosts.

Introduction

Vertebrate genomes harbor a diverse array of virus-like sequences [1]. These represent remnants of ancient viral integrations resulting from host-virus interactions, existing in a spectrum from static accumulations to structurally and functionally dynamic elements. The distinction between static and dynamic is likely a continuum influenced by the accumulation of mutations over time, rather than a clear dichotomy.

Endogenous retroviruses (ERVs), in particular, exhibit dynamic behavior. Including those that have become static due to mutations, ERVs are estimated to constitute approximately 8% of the human genome [2]. While implicated in human host defense and cellular functions [3], ERVs have also been associated with the pathogenesis of diseases such as cancer and Alzheimer's disease [4, 5]. Furthermore, considering retroelements such as SINEs, their suspected involvement in the formation of mouse genome boundaries [6] suggests that these elements have played a crucial role in the acquisition and maintenance of vertebrate morphology and function from ancient times to the present.

The significant proportion of retroelements in vertebrate genomes, the recent endogenization of a retrovirus [7], and the existence of a range of sequences from static to dynamic make retroelements a fascinating subject. While retroviruses possess enzymes that facilitate the insertion of their own genomes into the host genome, The integration of other viruses is currently understood to be largely contingent upon seemingly random and stochastic events. Consequently, non-mutated, dynamic non-retroviral sequences are seldom found within vertebrate genomes, leading to the inference that ancient ancestral viruses integrated and now exist in a static state or as quasi-dynamic sequences with limited transcription in extant vertebrate genomes..

Nevertheless, even static sequences, when present in substantial quantities, possess the potential to influence the three-dimensional architecture of vertebrate genomes, and any transcriptional activity originating from these non-retroviral elements could potentially modulate immune responses during infection with related viruses.

While initial compilations of non-retroviral integrations date back to the early 2000s [1], the recent accumulation of vast viral data and advancements in

bioinformatics have enabled comprehensive analyses [8]. This prior study identified 2,040 virus-like sequences within 295 vertebrate genomes by screening 24,478 viral proteins from the *Shotokuvirae* (16 families) and *Orthornavirae* (112 families) kingdoms (tblastn E-value 1e-5), and discussed the host phylogenetic relationships for major viral groups.

In this study, we performed a search using the FASTA program [9] against the entire human peptides set, specifically peptides derived from the human genome build GRCh38, with all viral proteins registered in the Virus-Host DB (https://www.genome.jp/virushostdb/) as of 2022 (RefSeq 1,860,785 entries). This analysis revealed a high degree of homology between the S9 protein of Simian beta-herpesvirus 4 (simian cytomegalovirus; SBHV4) and the human Major Histocompatibility Complex class I molecule, HLA-A, covering 86.3% of the HLA-A full length. Notably, this S9 protein is unique and was identified in a genomic region distinct from the herpesviral regions found in the aforementioned comprehensive search for virus-like proteins [8]. In this study, we discuss the evolution of this virus, focusing on the S9 protein, from a perspective slightly different from the hypothesis of ancient endogenization of a *Saimiriine beta-herpesvirus* ancestor in the human lineage.

Results＆Discussion

Homology searches using FASTA36 software between the entire viral peptide set and the human peptide set yielded 372 hits with ≥60% identity and ≥200 amino acid residues. Among these, we focused on a 315-residue overlap (86.3% of full-length HLA-A) between the membrane protein S9 (SBHV4-S9; 344 residues in length) of Saimiriine herpesvirus 4 strain (NC_016448.1) and human major histocompatibility complex class IA (HLA-A; 365 residues in length), showing 191 identical amino acids (60.6%) (Supplemental data 1). This represented the most extensive and highest homology observed thus far between a viral protein and an MHC class I molecule.

Further homology analysis of the SBHV4-S9 peptide (restricted to the HLA-A homologous region) against all other viral peptides revealed low similarity, with the highest identity being 43.2% (137/317) to a sequence from Murmansk poxvirus (Supplemental data 2), indicating the unique nature of SBHV4-S9. Given the low probability that the corresponding proteins in the viral species listed in Supplemental data 2 shared a recent common viral ancestor, phylogenetic analysis was not performed.

Structural analysis was conducted on the full-length peptide sequences of SBHV4-S9 and HLA-A. Based on subsequent phylogenetic analysis indicating the closest relationship of SBHV4-S9 to an MHC class I-like gene from the African *Papio anubis* and the most distant relationship to that of *Scleropages formosus*, we included structural analyses of the full-length peptides corresponding to these two species (Fig 1). Visual inspection revealed a high degree of structural similarity among the three (Fig 1, upper panel). Template Modeling (TM)-scores and Root Mean Square Deviation (RMSD) values between SBHV4-S9 and HLA-A, SBHV4-S9 and "MHC class I-like" (*Papio anubis*), and SBHV4-S9 and "MHC class I-like"(*Scleropages formosus*) were (0.83, 1.98 Å), (0.84, 2.55 Å), and (0.78, 2.86 Å), respectively, further demonstrating significant structural resemblance based on intermolecular distances. Furthermore, the SBHV4-S9 peptide exhibited greater similarity to HLA-A and the *Papio anubis* MHC class I-like peptide than to the *Scleropages formosus* MHC class I-like peptide. Superimposition of the SBHV4-S9 and HLA-A structures showed particularly high similarity in the α1, α2, and α3 domains of HLA-A, with SBHV4-S9 also possessing the peptide-binding pocket between α1 and α2. However, differences in the angles at the base of the branching towards the N- and C-termini and variations in individual molecular features were observed (Fig 1 A and B). Considering the potential for cross-species infection, the immunogenicity of this structure itself warrants consideration. In contrast, the human cytomegalovirus UL18 protein has been previously reported as the viral protein exhibiting the highest peptide homology to HLA, displaying approximately 25% sequence identity [10]. UL18 has been shown to bind inhibitory NK cell receptors with surprisingly high affinity (over 1000-fold greater than HLA), suggesting a potential mechanism for NK cell function suppression [11]. Although the strain examined in this study was isolated from a South American squirrel monkey, the immunogenicity of the SBHV4-S9 structure itself and its potential to mimic HLA necessitate careful attention in the context of potential cross-species transmission.

What are the implications of the uniqueness of the SBHV4-S9 protein among previously isolated or sequenced viruses, and its striking structural similarity to HLA-A? Beyond the integration of viral genomes into the host genome, as observed with endogenous retroviruses, there are several documented instances suggesting the acquisition of host genomic fragments by viruses. For example, in the *Pestivirus* genus, the genomes of three bovine viral diarrhea virus (BVDV) strains exhibit distinct

insertion sites of host-derived ubiquitin genes. Notably, the pathogenic CP1 strain and the non-pathogenic NCP1 strain, isolated from the same bovine host, showed the presence of a host ubiquitin gene insertion specifically in CP1, which was absent in NCP1. The potential role of this ubiquitin gene insertion in the differential pathogenicity has been discussed [12].

Furthermore, studies on poxviruses, which are DNA viruses, have revealed the presence of target site duplications (TSDs) generated by the host Long Interspersed Nuclear Element-1 (LINE-1) retrotransposon. This observation suggests a LINE-1-mediated molecular mechanism for the incorporation of host genomic fragments (DNA reverse-transcribed from mRNA) into the viral genome [13].

Regarding the squirrel monkey cytomegalovirus focused on in this study, its S1 protein (285 amino acids; NC_016448.1) has been shown to possess high homology to the signaling lymphocytic activation molecule (SLAM) family receptors of the Bolivian squirrel monkey (77% identity to sbSLAMF6) [14]. It has been proposed that this S1 protein was acquired through retrotransposition during the co-evolution of the virus and its host. Similarly, the absence of intron sequences characteristic of HLA-A within the homologous region of SBHV4-S9 may indicate a potential acquisition via reverse transcription. This specific S1 homologous sequence was also identified in our current investigation.

Initially, we conducted a search for TSDs flanking the SBHV4-S9 gene. Multiple sequences, suggestive of TSDs, were identified at both ends of the HLA-A homologous region within the S9 gene, as well as at both ends of the entire S9 gene. These flanking sequences were consistently observed regardless of whether the reference was the highly HLA-A homologous region (Fig 2, upper panel), the full-length S9 gene (Fig 2, middle panel), or the lifted-over region of the HLA-A gene (Fig 4, green bar). Furthermore, a lower GC content was consistently observed across these regions when calculated with a window size of 25 (Fig 2, lower panel). While TSDs mediated by LINE-1 are typically directly adjacent to the inserted DNA, the presence of pseudo-empty sites containing polyadenylation signals can serve as supplementary evidence for LINE-1 involvement [13]. In this study, some clear polyadenylation signals were observed within the low GC content regions; however, due to accumulated mutations, the authentic polyadenylation signal utilized remains undetermined. Consequently, while the precise length and integration site of the SBHV4 MHC class I-like sequence cannot be definitively established, these findings suggest a potential

transfer of genetic material from the host genome to this cytomegalovirus via a LINE-1-mediated mechanism, analogous to that reported in poxviruses.

Phylogenetic analysis (maximum likelihood method) was performed on a total of 127 MHC class I-like gene sequences, including the finally obtained SBHV4-S9 and seven human variants. To initially ascertain the branching position of SBHV4-S9, a cladogram devoid of genetic distance information was generated to assess its placement and statistical support (Fig 3, left panel). Both SHaLRT and Ultra-fast bootstrap (UFbootstrap) values were calculated, with values ≥80% and ≥95%, respectively, considered reliable. As no integrated index for assessing both measures was available, a novel index, the S-value, based on their geometric mean, was calculated, with a threshold of 87 established for reliable support ($\sqrt{80\ X\ 95}$ = 87.18). Phylogenetic analysis (maximum likelihood) using Asian arowana as an outgroup confirmed robust branching leading to SBHV4-S9, as depicted in Fig 3 (left panel). While the resulting phylogeny generally congruent with primate species relationships, some discrepancies were observed. The clade containing SBHV4-S9 included *Papio anubis*, *Nomascus leucogenys*, *Cebus imitator*, *Aotus nancymaae*, *Callithrix jacchus*, and *Saimiri boliviensis*. This grouping was consistent with a Bayesian phylogenetic tree constructed using only primate sequences (excluding SBHV4) (Fig 4, right panel), which also exhibited high posterior probabilities (BPP) alongside the previously calculated S-value, supporting the reliability of this clade's topology based on our single-gene analysis. The genetic distance between SBHV4-S9 and *Papio anubis* estimated by maximum likelihood was substantial (Fig 3, right panel). Bayesian inference was performed excluding SBHV4-S9 for the following reasons: initial analysis including SBHV4-S9 placed its divergence basal to Asian arowana, a discrepancy likely arising from the substantial post-divergence distance of SBHV4-S9 as indicated by maximum likelihood analysis. Viruses can undergo rapid diversification during specific short periods, potentially driven by interspecies transmission and spread [15], a phenomenon likely experienced by SBHV4, distinct from vertebrate evolution. The high sequence similarity of SBHV4-S9 to primate peptides revealed by maximum likelihood analysis (Fig 3, right panel) and homology searches (Asian arowana was ranked outside the top hits in Supplemental data 3.), coupled with the greater structural similarity of SBHV4-S9 to HLA-A and *Papio anubis* MHC class I-like molecules compared to that of Asian arowana (Fig 1), supported the exclusion of a pre-arowana divergence for SBHV4-S9. Consequently, Bayesian inference was conducted without the SBHV4-S9 sequence. The divergence time between *Papio anubis* and *Nomascus leucogenys* was

considered the oldest likely divergence of SBHV4-S9 from *Papio anubis* (Fig 4, right panel), estimated to have occurred between 12.9 and 22.2 million years ago (Height: 17.3 million years). This suggests that SBHV4-S9 likely diverged from *Papio anubis* or a closely related ancestor within this timeframe.

Collectively, one hypothesis posits that SBHV4-S9 was transferred from the *Papio anubis* MHC class I-like gene to SBHV4 via LINE-1 approximately 12.9 to 22.2 million years ago. The current host range of SBHV4 includes the Bolivian squirrel monkey, a New World primate, while *Papio anubis* is an Old World primate, implying cross-species transmission across continents. To investigate the evolutionary dynamics of this process, we first calculated dN/dS (ω) values (Fig 4, left panel), which were generally neutral across vertebrates. Given the propensity for extreme ω values in closely related conspecifics (e.g., human intraspecies, canine inter-breed, porcine inter-breed, *Heterocephalus glaber* inter-sex), these were excluded. Potential positive selection was observed in *Macaca fascicularis*, *Macaca mulatta*, *Phocoena sinus*, and *Monodon monoceros*, while SBHV4-S9 exhibited evidence of purifying selection. A likelihood ratio test (LRT) using these five taxa yielded a statistic of 151.40 (5 degrees of freedom), with $p < 6.72 \times 10^{-31}$ (no instances exceeding this statistic in 10 million trials; $p < 1.0 \times 10^{-7}$). A similar LRT comparing SBHV4-S9 and *Papio anubis* (neutral) resulted in a statistic of 123.62 (2 degrees of freedom), $p < 1.43 \times 10^{-27}$ (no instances exceeding this statistic in 10 million trials; $p < 1.0 \times 10^{-7}$). These results led to the rejection of the null hypothesis, demonstrating that SBHV4-S9 has undergone purifying selection (Fig 4, left panel). While purifying selection is common in viruses [16], our calculation was performed within the context of human HLA-A and primate MHC class I-like sequences, not solely among viral strains. Considering the generally high host specificity of cytomegaloviruses, it is plausible that SBHV4-S9 has survived under strong selective constraints imposed by its host during co-existence.

Papio anubis is a primate species native to Africa, while SBHV4 was isolated from squirrel monkeys inhabiting the rainforests around Bolivia in South America. Direct transmission of a virus with the highest sequence similarity in SBHV4-S9 (*Papio anubis*) across such geographical and species barriers is improbable. Parallels could be drawn to the human introduction of yellow fever virus from Africa to South America [17]; however, historical records of human-mediated transport of *Papio anubis* to South America are lacking. Conversely, evidence suggests the translocation of *Chlorocebus sabaeus*, whose current range overlaps with *Papio anubis*, from Africa to the West Indies in the 16th century, where they remain extant [18]. Given the distant phylogenetic relationship between *Chlorocebus sabaeus* and *Papio anubis* observed in

our HLA-A phylogeny (Fig 3 and Fig 4), HLA-A was not a suitable index to establish a link between African *Papio anubis* and South American squirrel monkeys. Our hypothesis posits the transfer of an orthologous sequence with >60% identity to HLA-A from *Papio anubis* to SBHV4. As previously noted, such instances of host gene acquisition by viruses are not unprecedented; as seen with the S1 protein of squirrel monkey cytomegalovirus is proposed to have originated from a SLAMF receptor gene of the Bolivian squirrel monkey [14]. Thus, supporting both our current hypothesis and previous findings suggests the potential for squirrel monkey cytomegalovirus to have acquired cDNA sequences from other primate genomes as well.

To further investigate the overall similarity between the SBHV4 genome and cDNA sequences from various primate species, aiming to elucidate the potential missing link between *Papio anubis* and Bolivian squirrel monkeys, we performed FASTA36 homology searches of the entire SBHV4 genome against primate cDNA sequences. For each primate species, Levenshtein distances were calculated for all homologous SBHV4 sequences, followed by Wasserstein distance calculation and hierarchical clustering (Fig 5). The resulting dendrogram was visually partitioned into clusters 1-9. In this representation, cluster formation at lower Wasserstein distances indicates higher similarity in the distribution of Levenshtein distances, implying that if species A and species B cluster at a low distance, the overall similarity between their respective cDNAs and the SBHV4 genome is comparable. Given the relative nature of this similarity and the absence of an absolute criterion, statistical significance of similarity was assessed using Mann-Whitney U tests and permutation tests. Parenthetically, this original analytical approach, employing the cDNA set of a reference organism, holds potential for reconsidering the comprehensive phylogenetic relationships among closely related species beyond the scope of this study.

Using Cluster 5 (comprising *Papio anubis*) as a reference group, Mann-Whitney U and permutation tests revealed no significant differences in similarity to SBHV4 sequences across Clusters 3-7, but a significant difference with Cluster 9. Furthermore, when *Papio anubis* was the reference species, no significant difference in SBHV4 genome similarity was observed between *Papio anubis* and *Chlorocebus sabaeus* cDNAs, whereas a significant difference existed between *Papio anubis* and squirrel monkey cDNAs. The lower clustering position of *Papio anubis* and *Chlorocebus sabaeus* compared to the higher position of squirrel monkey clusters indicates a closer relationship between some *Papio anubis* and *Chlorocebus sabaeus* cDNAs and the SBHV4 genome than with some squirrel monkey cDNAs. Phylogenetic analysis (maximum likelihood) of the S1 gene (implicated in transfer from squirrel monkey

SLAMF6-like genes to SBHV4) and primate SLAMF6-like genes (Fig 6) confirmed the S1 gene's origin from Bolivian squirrel monkey. These findings suggest a longer co-evolutionary history of SBHV4 with *Papio anubis* and *Chlorocebus sabaeus* compared to the more recent co-evolution with Bolivian squirrel monkeys. While precise dating is challenging due to the single lineage (including subspecies) of the isolated virus, one hypothesis posits a long history of SBHV4 infection in *Papio anubis* and *Chlorocebus sabaeus*, with the co-evolutionary history with squirrel monkeys commencing more recently, potentially coinciding with the anthropogenic introduction of *Chlorocebus sabaeus* to South America. However, comparison of Wasserstein distances and the phylogeny highlights some species requiring further consideration. For instance, the New World common marmoset (Cluster 4) exhibits SBHV4 similarity closer to African Old World monkeys. Wasserstein distances show similar SBHV4 similarity for drill and squirrel monkey cDNAs despite their distant species phylogeny. Geographical separation and undocumented human or primate translocations may underlie these differing SBHV4 infection histories. While initial analyses explored SBHV4 genome similarity to primate genomes, the limited number of homologous genomic sequences (approximately 2-3 FASTA entries per species, insufficient for robust Wasserstein distance analysis compared to the extensive cDNA data in Supplemental data 4) precluded statistical evaluation. This disparity suggests that DNA transfer to viruses, including LINE-1 mediated events, predominantly involves mRNA-derived DNA rather than genomic DNA.

In conclusion, SBHV4 has likely co-evolved with vertebrates since ancient times, undergoing cycles of infection and extinction with African and Asian primates. Approximately 12.9 to 22.2 million years ago, SBHV4 acquired the SBHV4-S9 gene, homologous to HLA-A peptide (gene), from *Papio anubis* or a closely related ancestor. The newly acquired SBHV4 subsequently spread through *Chlorocebus sabaeus* or other primate species, eventually being introduced to South America by human activity. Following this introduction, a new evolutionary history of SBHV4 commenced through inter-primate transmission.

Materials&Methods

Entire Viral Peptide vs. Entire Human Peptide Set

The complete viral peptide set (release 225), updated August 6, 2024, was downloaded from GenomeNet (Virus-Host Database). The entire human peptide set, corresponding to the GRCh38 genome build, was obtained from NCBI. Homology

searches were performed using FASTA36 (-E 1e-20) [9], and results were screened for hits with ≥60% identity and ≥200 amino acid residues (Supplemental data 1).

SBHV4-S9 vs. All Viral Peptides

To investigate the relationship of SBHV4-S9 (YP_004940188.1; 344 residues, full length) to other viral peptides, a FASTA36 homology search was performed using the SBHV4-S9 peptide sequence (HLA-A homologous region only) and the aforementioned entire viral peptide dataset (Supplemental data 2).

SBHV4-S9 TSD Identification

To investigate the presence of TSDs flanking the HLA-A homologous region of SBHV4-S9, TSD searches were performed using the SBHV4 genome (NC_016448.1) and the high HLA-A homologous region with SBHV4-S9. The SBHV4-S9 locus within the SBHV4 genome was determined, and 200 bp flanking sequences were retrieved. The 5' 200 bp region was split into overlapping fragments (4-20 bp sliding window), and each fragment was used to query the 3' 200 bp region using the Needleman-Wunsch alignment algorithm [19] (Code 1). TSD sequences exhibiting ≥60% identity were output in FASTA format, converted to FASTQ format (with maximum quality scores), and subsequently mapped using HISAT2 [20] to a composite sequence ("the SBHV4 concatemer") comprising the 200 bp flanking the 5' end of the SBHV4-S9 gene, 50 'N's mimicking its HLA-A homologous region, and the 200 bp flanking its 3' end. This procedure was applied to both the HLA-A homologous region and the full-length DNA of SBHV4-S9. Mapping results were visualized using JBrowse 2 [21].

SBHV4-S9 vs. Peptides from Other Animal Species

To identify peptide sequences homologous to SBHV4-S9 in diverse species, the SBHV4-S9 peptide (HLA-A homologous region only) was queried against the comprehensive peptide datasets of all species available at https://ftp.ensembl.org/pub/release-112 using a homology search (-b 1 -d 1 -E 1e-10 -m BB). Results were filtered for sequences exhibiting ≥300 amino acid residues, ≥50% identity, and a query start position <3 amino acid residues in NCBI BLAST output format. A tabular summary of the homology observed for each species was compiled (Supplemental data 3).

Structural Analysis of Full-Length SBHV4-S9 and MHC Class I Peptides

Structural analysis of the full-length sequences of SBHV4-S9 (YP_004940188.1), HLA-A (AZU89775.1), the *Papio anubis* MHC class I-like peptide (African origin; closest homolog identified by FASTA36), and the *Scleropages formosus* MHC class I-like peptide (XP_029104321.1; most distant homolog identified by FASTA36) was performed using AlphaFold 2.3.2 [22]. Full-length sequences were employed to explore potential functional similarities beyond high-homology subregions. TM-scores for pairwise comparisons were normalized by the length of the shorter sequence. The resulting four PDB files were used to assess pairwise structural similarity via the TM-align web server [23], and visualized using ChimeraX software [24].

Molecular Phylogenetic Analysis of MHC Class I-like and SBHV4-S9 Genes

To elucidate the molecular evolutionary history of SBHV4-S9 in relation to MHC class I-like genes, we performed phylogenetic analysis.

The sequence corresponding to the high-homology region of SBHV4-S9 within HLA-A peptide was manually extracted from the Ensembl entry ENST00000706898.1 (HLA-A-216). For SBHV4-S9, the homologous region to HLA-A was extracted from the SBHV4 genome (NC_016448.1).

Subsequently, CDS data for all species available on the Ensembl FTP site (https://ftp.ensembl.org/pub/release-112) were downloaded and subjected to homology searches (-b 1 -d 1 -E 1e-10 -m BB). Resulting BLAST-formatted .txt files within this directory were filtered to retain sequences ≥900 bp in length, initiating with a query start position ≤15 (to ensure data integrity), yielding FASTA36 results for 126 animal species + 1 virus species. All MHC class I-like sequences for these 126 animal species + 1 virus species were compiled into a single FASTA file, to which the SBHV4-S9 DNA FASTA data was manually appended.

To investigate associations with human variants, HLA-A haplotype sequences were collected from human genome CRAM files available from the International Genome Sample Resources (IGSR; https://www.internationalgenome.org/data-portal/data-collection/30x-grch38). A comprehensive list of available CRAM files (igsr_30x GRCh38.tsv.tsv) was

downloaded from the IGSR website (Available data tab, as of July 2024). A list of up to four individuals per cohort with non-redundant Population designations was generated (igsr_30x GRCh38_rev.tsv; 104 individuals). CRAM files for these 104 individuals were retrieved. FASTQ files extracted from these CRAM files were mapped to HLA-A_gene_GRCh38.fa (EU445472.1). De novo assembly was performed on the resulting .sam files using SPAdes software (-k 31) [25]. Following careful examination of open reading frames and homologies, and excluding redundant HLA-A haplotypes, six human individuals exhibiting ≥99% homology to HLA-A CDS datum (GRCh38) were selected as candidates for phylogenetic analysis (IDs specified in Fig 3 and Fig 4). The FASTA data for these human individuals, previously recovered other species (126 animal species including humans + 1 virus species), and SBHV4-S9 were concatenated into a final file (containing 133 FASTA entries; Supplemental data 5). The presence of three motifs characteristic of MHC class I molecules was confirmed across all FASTA data using MEME suite (https://meme-suite.org/meme/index.html) [26] (Supplemental data 6). Finally, this dataset underwent global alignment with MAFFT [27] and trimming with ClipKIT [28], followed by maximum likelihood phylogenetic tree inference using IQ-TREE 2 [29] (substitution model selection: -m MFP+LM; UFbootstrap with 7 million replicates including SHaLRT calculation). Branch support reliability was assessed using a composite S-value, calculated as the geometric mean of SHaLRT and UFbootstrap values output by IQ-TREE 2.

$$S\text{-}value=\sqrt{SHaLRT\ X\ UFbootstrap}$$

Nodes with S-values ≥87% were considered reliably supported and visualized using FigTree (https://github.com/rambaut/figtree). Initially, a cladogram was generated to illustrate branching patterns. Subsequently, a standard phylogenetic tree, including primates and selected outgroup species, was constructed to delineate the genetic distance of SBHV4-S9's divergence from *Papio anubis*.

Divergence Time Estimation of SBHV4-S9

From the final FASTA file used for maximum likelihood analysis (Supplemental data 5), FASTA data for primates and outgroups (*Monodelphis domestica*, *Cavia porcellus*, *Ornithorhynchus anatinus*, *Oryctolagus cuniculus*) were extracted. Following global alignment with MAFFT and parameter acquisition with IQ-TREE 2, an XML file (Supplemental data 7) was generated using BEAUti [30] and BEAST 2 was executed

[30] (1 billion MCMC iterations; ESS values ≥5000 for all parameters). From the resulting treefile, a Maximum Clade Credibility tree with Mean heights was generated after discarding the initial 20% burn-in using TreeAnnotator [30]. Numerical values pertaining to the divergence time estimation of *Papio anubis* and *Nomascus leucogenys* were extracted from this Nexus format file.

dN/dS Calculation

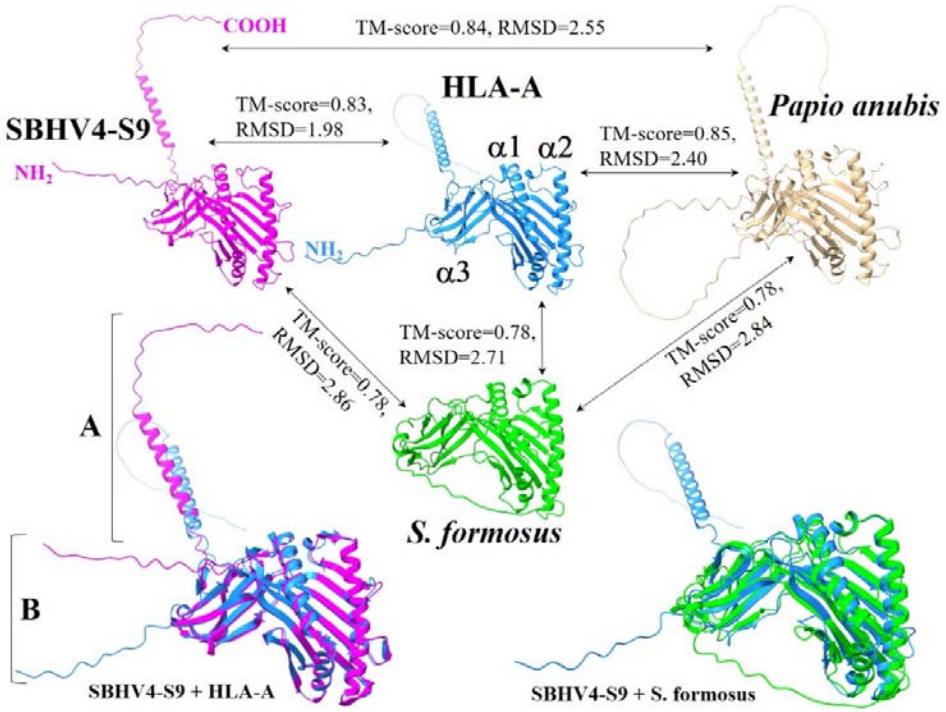
dN/dS (ω) values were calculated using CODEML in PAML [31]. For codon alignment, the FASTA file from Supplemental data 5 was aligned using PRANK software [32], employing the phylogenetic tree from Fig 3. CODEML was configured (model = 1, NSsites = 0) to estimate a single ω value for all branches, assuming a uniform ω across sites, to infer the ω value of SBHV4-S9 within the broader vertebrate context (details in Supplemental data 8). To facilitate visualization given the range of ω values, log□ □(ω) was calculated, and its distribution is presented in Supplemental data 9. To statistically assess negative selection on SBHV4-S9, ω values for SBHV4-S9 and *Papio anubis* were estimated separately under the same site-homogeneous ω assumption (model = 2; all other settings identical to Supplemental data 8). A likelihood ratio test (LRT) was subsequently performed using the results from a null model where all ω values were fixed (model = 0).

Molecular Phylogenetic Analysis of SLAMF6-like and Primate SBHV4-S1 Genes

To elucidate the co-evolutionary history of SBHV4-S1 with primate SLAMF6-like genes, the corresponding DNA sequences for the high-homology region between SBHV4-S1 and human SLAMF6 were obtained. Using the acquired human SLAMF6 cDNA, a FASTA36 search against the Ensembl release 112 coding sequence database yielded all homologous sequences (SLAMF6-like sequences from 99 animal species). From this dataset, only primate-derived FASTA entries (26 species) were extracted. The previously obtained SBHV4-S1 DNA was manually appended to this file (Supplemental data 10). To assess preliminary orthology, motif searches were performed using MEME suite [26], as previously described. All sequences commonly exhibited two motifs characteristic of SLAMF6 (Supplemental data 6). Following global alignment with MAFFT, maximum likelihood phylogenetic inference was conducted using IQ-TREE 2, employing -m MFP+LM for substitution model selection and *Carlito syrichta* as the outgroup, with Ultra-fast bootstrap (3 million replicates) including SHaLRT.

SBHV4 Genome vs. Primate cDNA

Genomic sequences for all available primate species were downloaded from Ensembl release 113. Homology searches between the SBHV4 genome (NC_016448.1) and these primate genomes were performed using FASTA36 (-b 1 -d 1 -m BB), and the resulting aligned animal and SBHV4 sequences were saved as separate FASTA files (total of 27 paired files, with viral sequences identified by sequential numerical suffixes _1, _2 appended to the ID). Subsequently, for one FASTA file of each pair (animal query), redundant sequences with <80% identity were removed using CD-HIT [33]. The corresponding SBHV4 sequences for the remaining non-redundant animal sequences were then extracted (synchronized 27 paired files). Normalized Levenshtein distances were calculated between the nucleotide sequences of each animal species and their corresponding SBHV4 sequences (Levenshtein distance divided by the length of the nucleotide sequence for normalization). As the number of high-homology FASTA entries identified by FASTA36 varied across animal species, the distribution of normalized Levenshtein distances for each species was divided into 50 bins to create a histogram, approximating the probability distribution for each species. The similarity between these approximated probability distributions was measured using the Wasserstein distance (Earth Mover's Distance), normalized by the 50 bins, and hierarchical clustering was performed based on the normalized Wasserstein distances using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA), also known as average linkage (implemented in Python as 'average'), which uses the average distance between all data points within probability distributions (Code 2). The consistency between the hierarchical clustering dendrogram and the approximated probability distributions was evaluated by calculating the cophenetic correlation coefficient and silhouette coefficient. Furthermore, to statistically evaluate the clustering of animal species, Mann-Whitney U tests and permutation tests were performed on the normalized Wasserstein distances to calculate p-values for the hypotheses that the normalized Wasserstein distances of the *Papio anubis*-containing cluster differ from other clusters, or that the normalized Wasserstein distance of *Papio anubis* differs from other individual animal species.


COOH
TM-score=0.84, RMSD=2.55
HLA-A
SBHV4-S9
TM-score=0.83,
RMSD=1.98
Papio anubis
α1 α2
TM-score=0.85,
RMSD=2.40
NH2
NH2
α3
TM-score=0.78,
RMSD=2.71
TM-score=0.78,
RMSD=2.86
TM-score=0.78,
RMSD=2.84
A
S. formosus
B
SBHV4-S9 + HLA-A
SBHV4-S9 + S. formosus

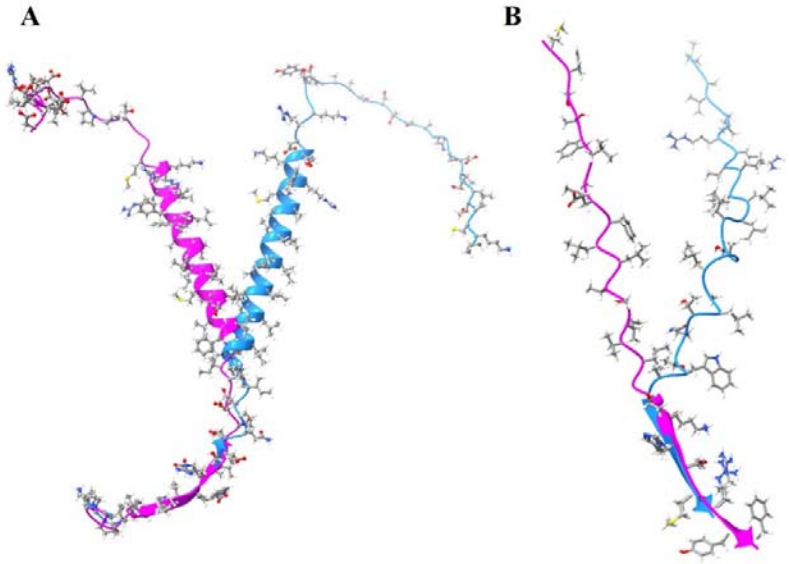

A
B

Fig1 **Molecular Models of MHC Class I-like Peptides (AlphaFold).** Magenta: SBHV4-S9; cyan: human HLA-A; gold: *Papio anubis* MHC class I-like peptide; green: *Scleropages formosus* MHC class I-like peptide. **Middle Left:** Superimposed structures of SBHV4-S9 and HLA-A. **Middle Right:** Superimposed structures of SBHV4-S9 and *Scleropages formosus* MHC class I-like peptide. TM-score and RMSD values quantifying structural similarity between the respective molecular models are indicated. α1, α2, α3 denote the HLA-A subunits. **Bottom:** Structural divergence between SBHV4-S9 (magenta) and HLA-A (cyan) at the C-terminus (A) and N-terminus (B).

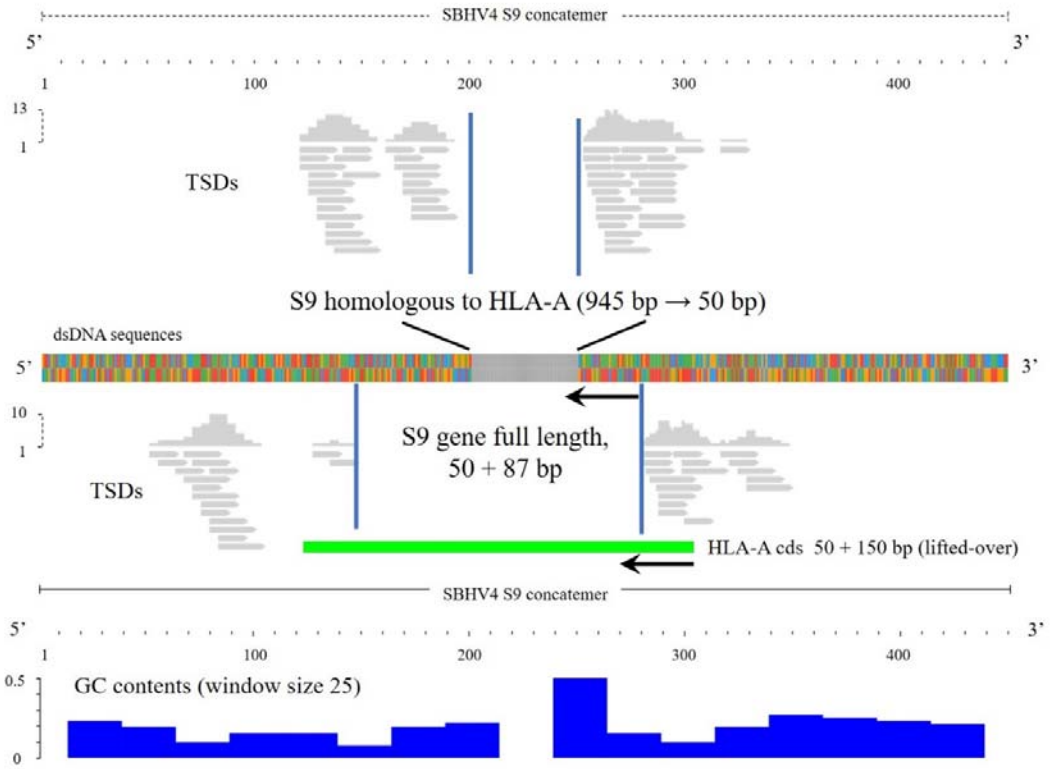


Fig2 **Predicted Target Site Duplications (TSDs) Flanking the SBHV4 S9 Gene.** The middle color-coded section illustrates the S9 gene (grey) and surrounding nucleotides. For visualization, the 945 bp S9 gene is compressed to 50 bp (referred to as SBHV4 S9 concatemer). Putative flanking short sequences, suggestive of TSDs, are observed at both ends of the compressed HLA-A homologous region within the S9 gene, as well as at both ends of the full-length S9 gene. Similar flanking sequences are also present at both ends of the lifted-over region (green) of the full-length HLA-A CDS, compressed to 50 bp to align with the SBHV4 S9 concatemer. The bottom panel displays a histogram (window size 25) of the GC content across this region, indicating values below 30% for the majority of the sequence.

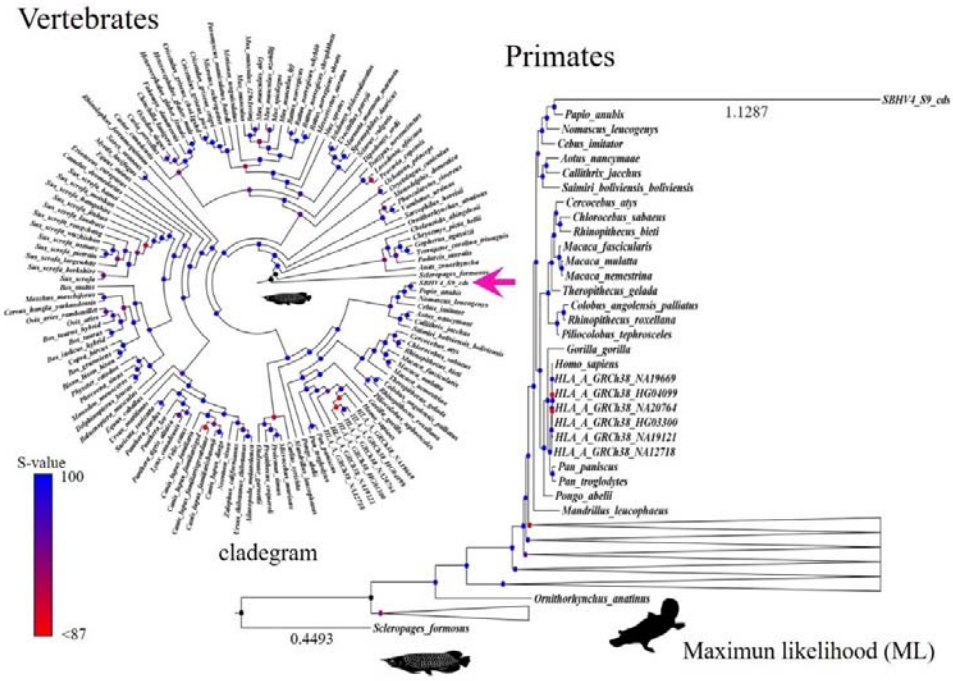


Fig3 **Phylogenetic Analysis of SBHV4-S9 Gene and Primate MHC Class I-like Sequences. Left:** Cladogram (maximum likelihood) illustrating branching patterns of 133 FASTA sequences, including SBHV4-S9 and seven human individuals, with the magenta arrow indicating the SBHV4-S9 gene's position. **Right:** Phylogenetic tree (maximum likelihood) depicting genetic distances among primates and selected outgroups. Node colors represent S-values (geometric mean of SHaLRT and UFbootstrap values), with ≥87 indicating reliable support. The tree shows SBHV4-S9 diverging after *Nomascus leucogenys* from the *Papio anubis* lineage, exhibiting a long terminal branch indicative of substantial genetic distance.

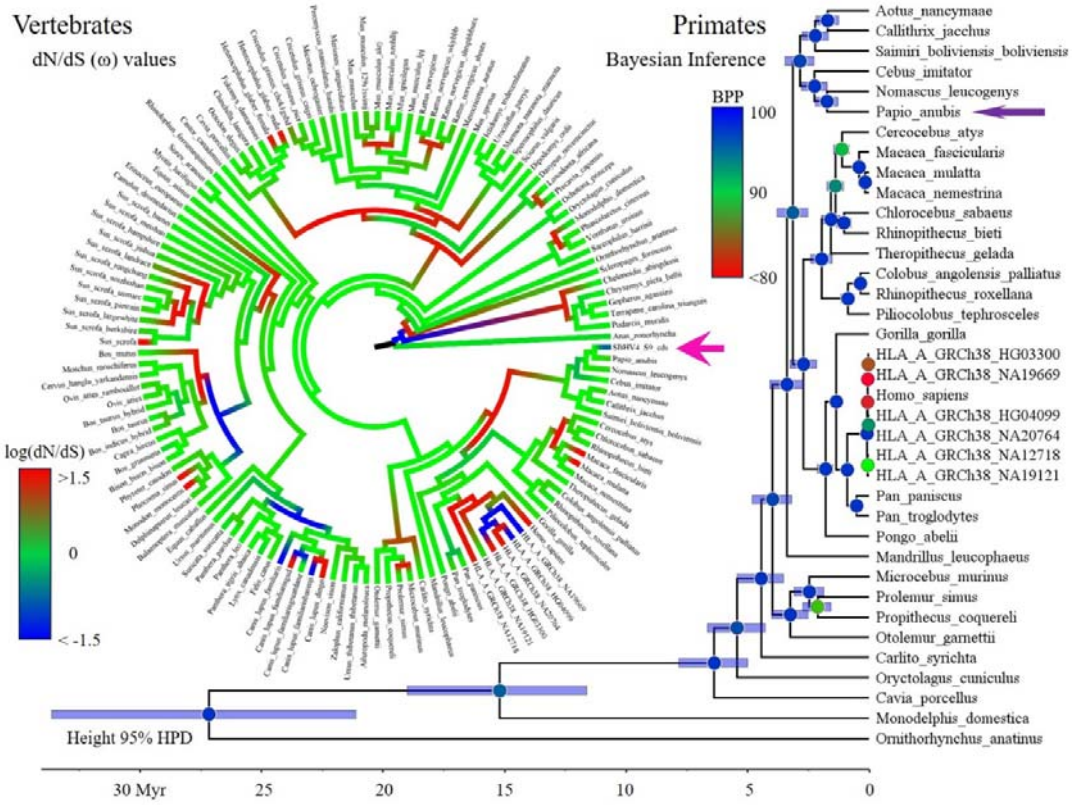


Fig4 **dN/dS (ω) Values and Bayesian Phylogenetic Tree. Left:** dN/dS (ω) values for each branch calculated from 133 FASTA sequences. The magenta arrow indicates the SBHV4-S9 gene. Values ≥1.5 (log□□(ω)) are red, 0 (log□□(ω) = 1) is green, and ≤-1.5 (log□□(ω)) are blue. Reliability within human, pig, and *Heterocephalus glaber* clades is low due to intraspecific and inter-breed comparisons; thus, negative selection is primarily inferred for SBHV4-S9. **Right:** Bayesian phylogenetic tree excluding SBHV4-S9. The purple arrow indicates *Papio anubis*. Node colors represent Bayesian Posterior Probabilities (BPP; threshold ≥80%): red (<80%), green (≥90%), blue (100%). Support for the lineage leading to *Papio anubis* is high. Bars on nodes indicate 95% Highest Posterior Density intervals (HPD).

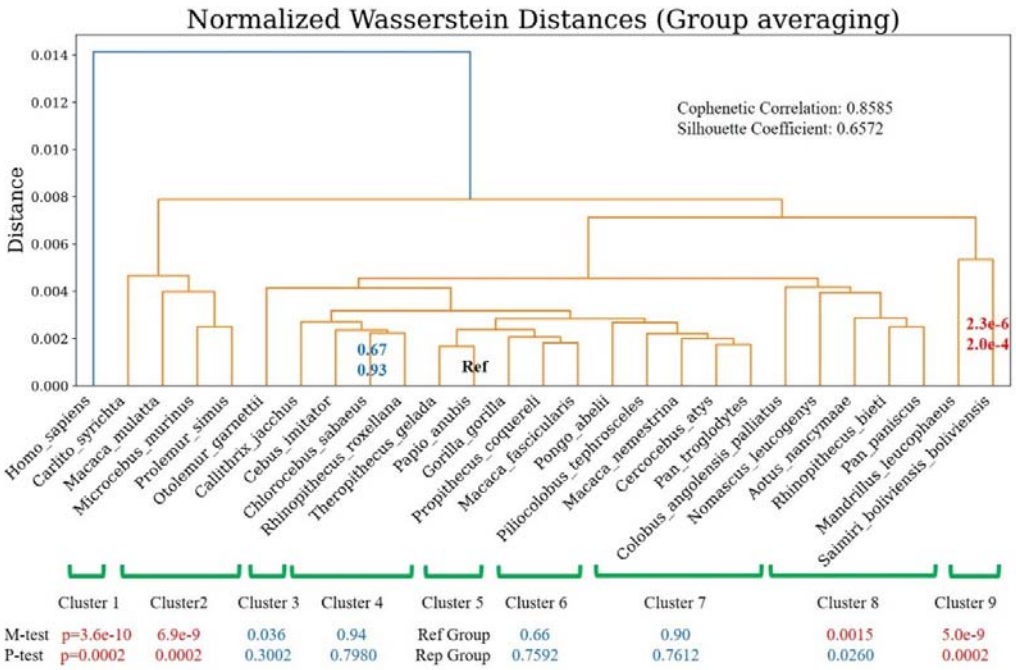


Fig5 **Comparison of SBHV4 Genome Similarity to Individual Animal Species cDNA via Wasserstein Distance.** Shorter Wasserstein distances indicate higher similarity between the SBHV4 genome and corresponding (high-homology) cDNA. Note the lack of direct correspondence with primate species phylogeny. Hierarchical clustering of normalized Wasserstein distances (detailed in Materials and Methods) was performed using UPGMA. Clusters were visually demarcated. Statistical comparisons between clusters were conducted using M-test (Mann-Whitney U test) and P-test (permutation test), with $p<0.01$ in red and $p\geq0.01$ in blue. *Papio anubis*-containing cluster (cluster 5) served as the reference group for significance testing. Numerical values on branches represent p-values from M-test and P-test comparing *Chlorocebus sabaeus* and *Saimiri boliviensis boliviensis* to *Papio anubis* (reference species). The cophenetic correlation coefficient and silhouette coefficient, displayed in the figure margins, indicate the reliability of the dendrogram.

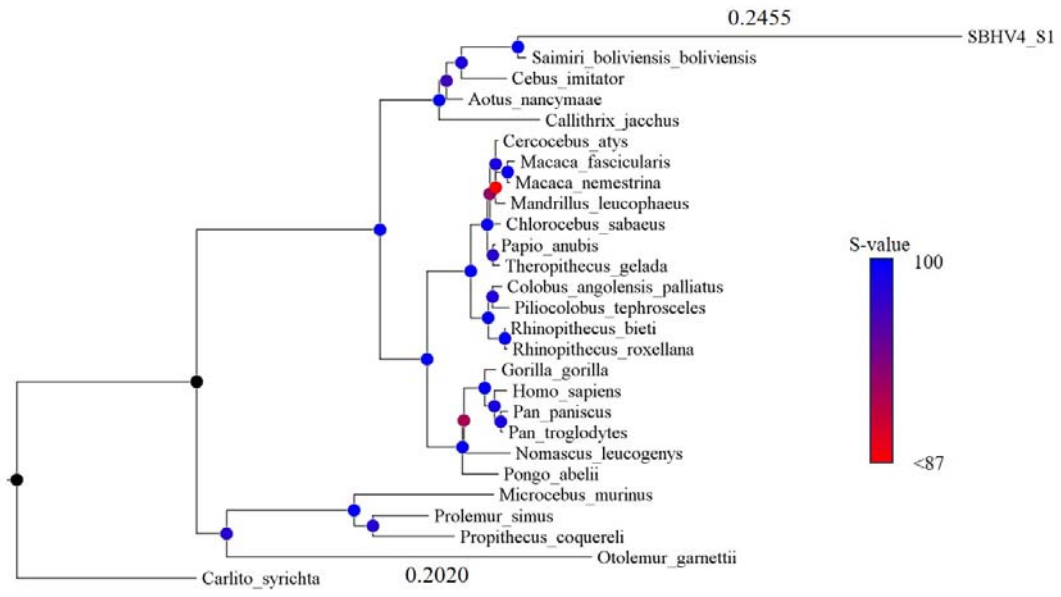


Fig6 **Maximum Likelihood Phylogenetic Analysis of Primate SLAMF6-like and SBHV4-S1 Genes.** Node colors represent S-values (geometric mean of SHaLRT and UFbootstrap values), consistent with Fig 3 and Fig 4. Nodes with missing values during tree construction (rooted with *Carlito syrichta*) are black. SBHV4-S1 branches from *Saimiri boliviensis boliviensis*, consistent with previous reports.

References

[1] Endogenous viruses: insights into viral evolution and impact on host biology
Cédric Feschotte, Clément Gilbert. Nat Rev Genet 2012 Mar 16;13(4):283-96. doi: 10.1038/nrg3199.

[2] Initial sequencing and analysis of the human genome
Lander E.S. et al. Nature 2001 Feb 15;409(6822):860-921. doi: 10.1038/35057062.

[3] Human Endogenous Retroviruses Are Ancient Acquired Elements Still Shaping Innate Immune Response
Nicole Grandi, Enzo Tramontano. Front Immunol 2018 Sep 10:9:2039. doi: 10.3389/fimmu.2018.02039. eCollection 2018.

[4] Human Endogenous Retrovirus-K (HML-2)-Related Genetic Variation: Human Genome Diversity and Disease
Wonseok Shin, Seyoung Mun, Kyudong Han. Genes (Basel) 2023 Nov 28;14(12):2150. doi: 10.3390/genes14122150.

[5] Awakening the dark side: Retrotransposon activation in neurodegenerative disorders
Elizabeth Ochoa Thomas, Gabbe Zuniga, Wenyan Sun, Bess Frost. Curr Opin Neurobiol. 2020 Feb 21;61:65–72. doi: 10.1016/j.conb.2020.01.012

[6] B2 SINE Copies Serve as a Transposable Boundary of DNA Methylation and Histone Modifications in the Mouse
Tomoko Ichiyanagi, Hirokazu Katoh, Yoshinobu Mori, Keigo Hirafuku, Beverly Ann Boyboy, Masaki Kawase, Kenji Ichiyanagi. Mol Biol Evol 2021 May 19;38(6):2380-2395. doi: 10.1093/molbev/msab033.

[7] Retroviral invasion of the koala genome
Rachael E. Tarlinton, Joanne Meers, Paul R. Young. Nature 2006 Jul 6;442(7098):79-81. doi: 10.1038/nature04841.

[8] Deep mining reveals the diversity of endogenous viral elements in vertebrate genomes
Jose Gabriel Nino Barreat, Aris Katzourakis. Nat Microbiol 2024 Nov;9(11):3013-3024. doi: 10.1038/s41564-024-01825-4. Epub 2024 Oct 22.

[9] Finding protein and nucleotide similarities with FASTA
William Pearson. Curr Protoc Bioinformatics 2004 Feb:Chapter 3:Unit3.9. doi: 10.1002/0471250953.bi0309s04.

[10] Human cytomegalovirus encodes a glycoprotein homologous to MHC class-I antigens
Beck S, Barrell B. G. Nature 1988 Jan 21;331(6153):269-72. doi: 10.1038/331269a0.

[11] Structure of UL18, a peptide-binding viral MHC mimic, bound to a host inhibitory receptor
Zhiru Yang, Pamela J Bjorkman. Proc Natl Acad Sci USA 2008 Jul 22;105(29):10095-100. doi: 10.1073/pnas.0804551105. Epub 2008 Jul 15.

[12] Viral cytopathogenicity correlated with integration of ubiquitin-coding sequences
Gregor Meyers, Norbert Tautz, Edward J Dubovi, Heinz-Jürgen Thiel. Virology 2004 Feb 9;180(2):602–616. doi: 10.1016/0042-6822(91)90074-L

[13] Poxviruses capture host genes by LINE-1 retrotransposition
Sarah M Fixsen, Kelsey R Cone, Stephen A Goldstein, Thomas A Sasani, Aaron R Quinlan, Stefan Rothenburg, Nels C Elde. eLife 2022 Sep 7;11:e63332. doi: 10.7554/eLife.63332

[14] Signaling Lymphocytic Activation Molecule Family Receptor Homologs in New World Monkey Cytomegaloviruses
Natàlia Pérez-Carmona, Domènec Farré, Pablo Martínez-Vicente, Cox Terhorst, Pablo Engel, Ana Angulo. J Virol 2015 Nov;89(22):11323-36. doi: 10.1128/JVI.01296-15. Epub 2015 Sep 2.

[15] Dispersal history of Miniopterus fuliginosus bats and their associated viruses in east Asia
Kimprasit T et al. PLoS One 2021 Jan 14;16(1):e0244006. doi: 10.1371/journal.pone.0244006. eCollection 2021.

[16] Evolutionary insights into host–pathogen interactions from mammalian sequence data

Manuela Sironi, Rachele Cagliani, Diego Forni, Mario Clerici. Nat Rev Genet 2015 Apr;16(4):224-36. doi: 10.1038/nrg3905.

[17] Out of Africa: a molecular perspective on the introduction of yellow fever virus into the Americas
Juliet E Bryant, Edward C. Holmes, Alan D. T Barrett. PLoS Pathog 2007 May 18;3(5):e75. doi: 10.1371/journal.ppat.0030075.

[18] Insights into the Geographical Origins of the Cabo Verde Green Monkey
Lara Almeida, Ivo Colmonero-Costeira, Maria J Ferreira da Silva, Cecilia Veracini, Raquel Vasconcelos. Genes (Basel) 2024 Apr 17;15(4):504. doi: 10.3390/genes15040504.

[19] A general method applicable to the search for similarities in the amino acid sequence of two proteins
Needleman S. B., Wunsch C. D. J Mol Biol 1970 Mar;48(3):443-53. doi: 10.1016/0022-2836(70)90057-4.

[20] HISAT: a fast spliced aligner with low memory requirements
Daehwan Kim, Ben Langmead, Steven L Salzberg. Nat Methods 2015 Apr;12(4):357-60. doi: 10.1038/nmeth.3317. Epub 2015 Mar 9.

[21] JBrowse 2: a modular genome browser with views of synteny and structural variation
Colin Diesh et al. Genome Biol 2023 Apr 17;24(1):74. doi: 10.1186/s13059-023-02914-z.

[22] Highly accurate protein structure prediction with AlphaFold
John Jumper et al. Nature 2021 Aug;596(7873):583-589. doi: 10.1038/s41586-021-03819-2. Epub 2021 Jul 15.

[23] TM-align: a protein structure alignment algorithm based on the TM-score
Yang Zhang, Jeffrey Skolnick. Nucleic Acids Res. 2005 Apr 22;33(7):2302–2309. doi: 10.1093/nar/gki524

[24] UCSF ChimeraX: Tools for structure building and analysis

Elaine C. Meng, Thomas D. Goddard, Eric F. Pettersen, Greg S. Couch, Zach J. Pearson, John H. Morris, Thomas E. Ferrin. Protein Sci. 2023 Nov 1;32(11):e4792. doi: 10.1002/pro.4792

[25] SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing
Anton Bankevich et al. J Comput Biol 2012 May;19(5):455-77. doi: 10.1089/cmb.2012.0021. Epub 2012 Apr 16.

[26] The MEME Suite
Timothy L. Bailey, James Johnson, Charles E. Grant, William S. Noble Nucleic Acids Res 2015 Jul 1;43(W1):W39-49. doi: 10.1093/nar/gkv416. Epub 2015 May 7.

[27] MAFFT version 5: improvement in accuracy of multiple sequence alignment.
Kazutaka Katoh, Kei-ichi Kuma, Hiroyuki Toh, Takashi Miyata. Nucleic Acids Res 2005 Jan 20;33(2):511-8. doi: 10.1093/nar/gki198. Print 2005.

[28] ClipKIT: A multiple sequence alignment trimming software for accurate phylogenomic inference
Jacob L. Steenwyk, Thomas J. Buida 3rd, Yuanning Li , Xing-Xing Shen, Antonis Rokas. PLoS Biol 2020 Dec 2;18(12):e3001007. doi: 10.1371/journal.pbio.3001007. eCollection 2020 Dec.

[29] IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era
Bui Quang Minh, Heiko A. Schmidt, Olga Chernomor, Dominik Schrempf, Michael D. Woodhams, Arndt von Haeseler, Robert Lanfear. Mol Biol Evol. 2020 May 1;37(5):1530-1534. doi: 10.1093/molbev/msaa015.

[30] BEAST 2: a software platform for Bayesian evolutionary analysis
Remco Bouckaert, Joseph Heled, Denise Kühnert, Tim Vaughan, Chieh-Hsi Wu, Dong Xie, Marc A. Suchard, Andrew Rambaut, Alexei J. Drummond. PLoS Comput Biol 2014 Apr 10;10(4):e1003537. doi: 10.1371/journal.pcbi.1003537. eCollection 2014 Apr.

[31] PAML 4: phylogenetic analysis by maximum likelihood
Ziheng Yang. Mol Biol Evol 2007 Aug;24(8):1586-91. doi: 10.1093/molbev/msm088.

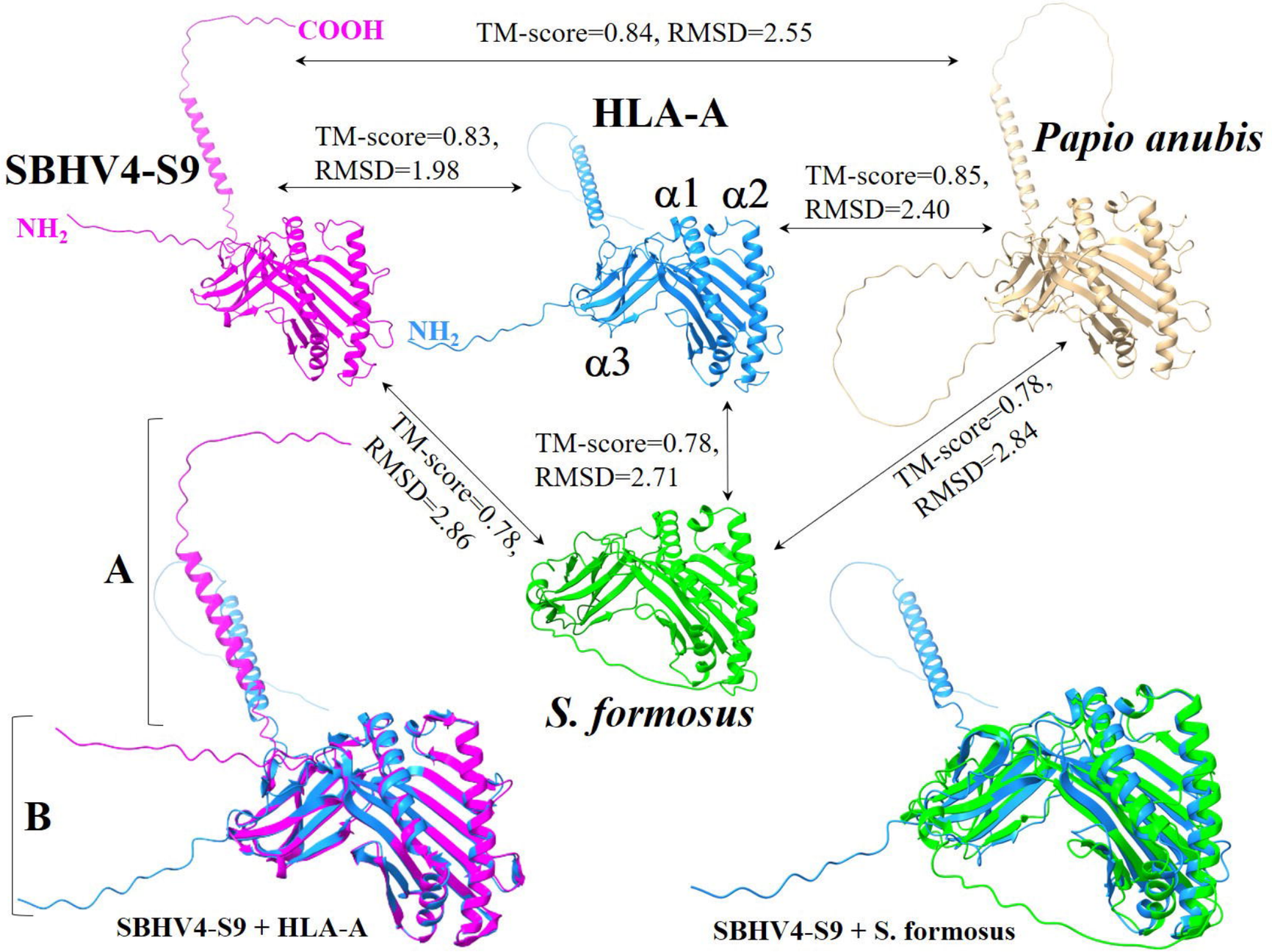
Epub 2007 May 4.

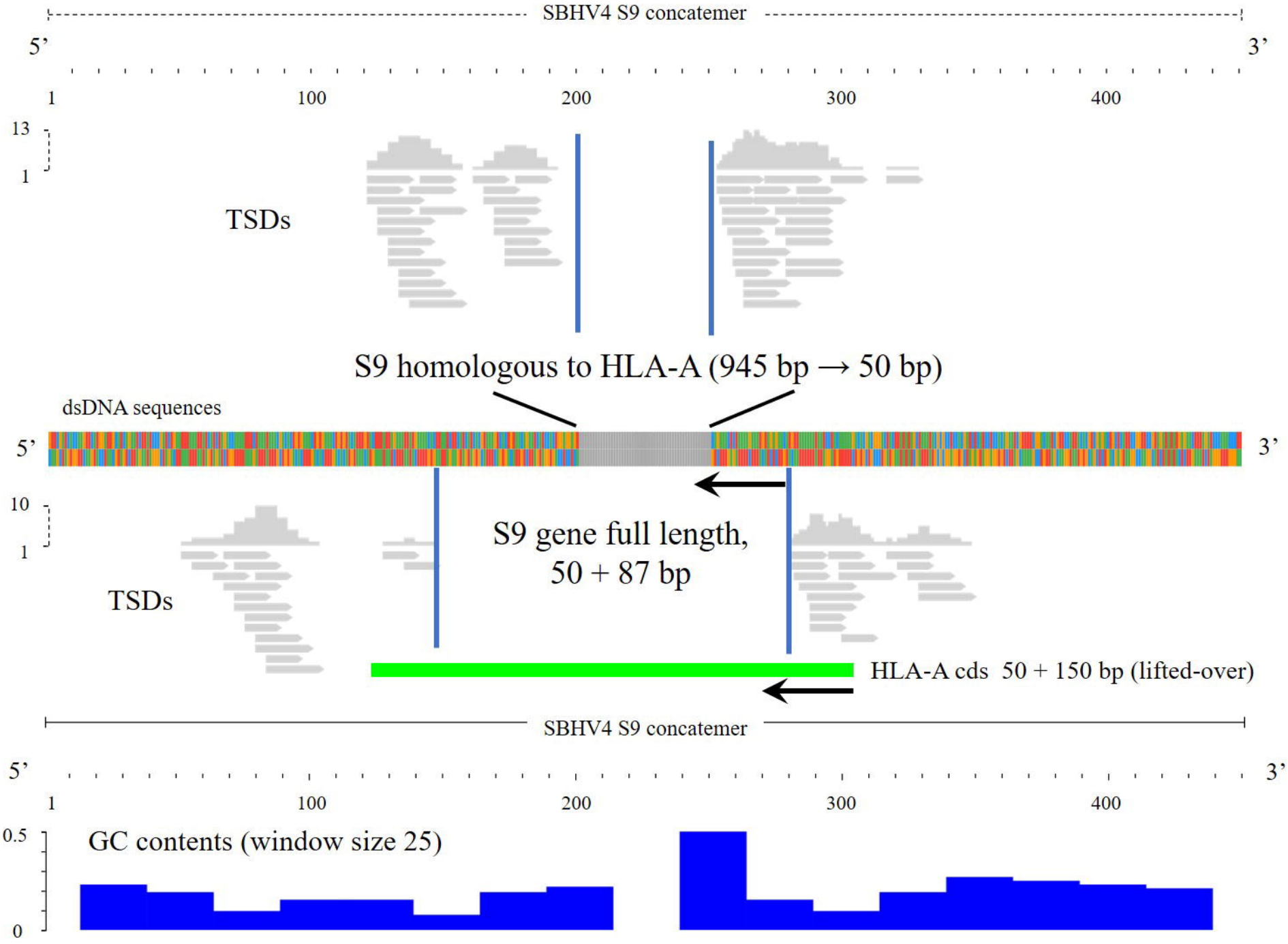
[32] Phylogeny-aware alignment with PRANK
Ari Löytynoja. Methods Mol Biol 2014:1079:155-70. doi: 10.1007/978-1-62703-646-7_10.

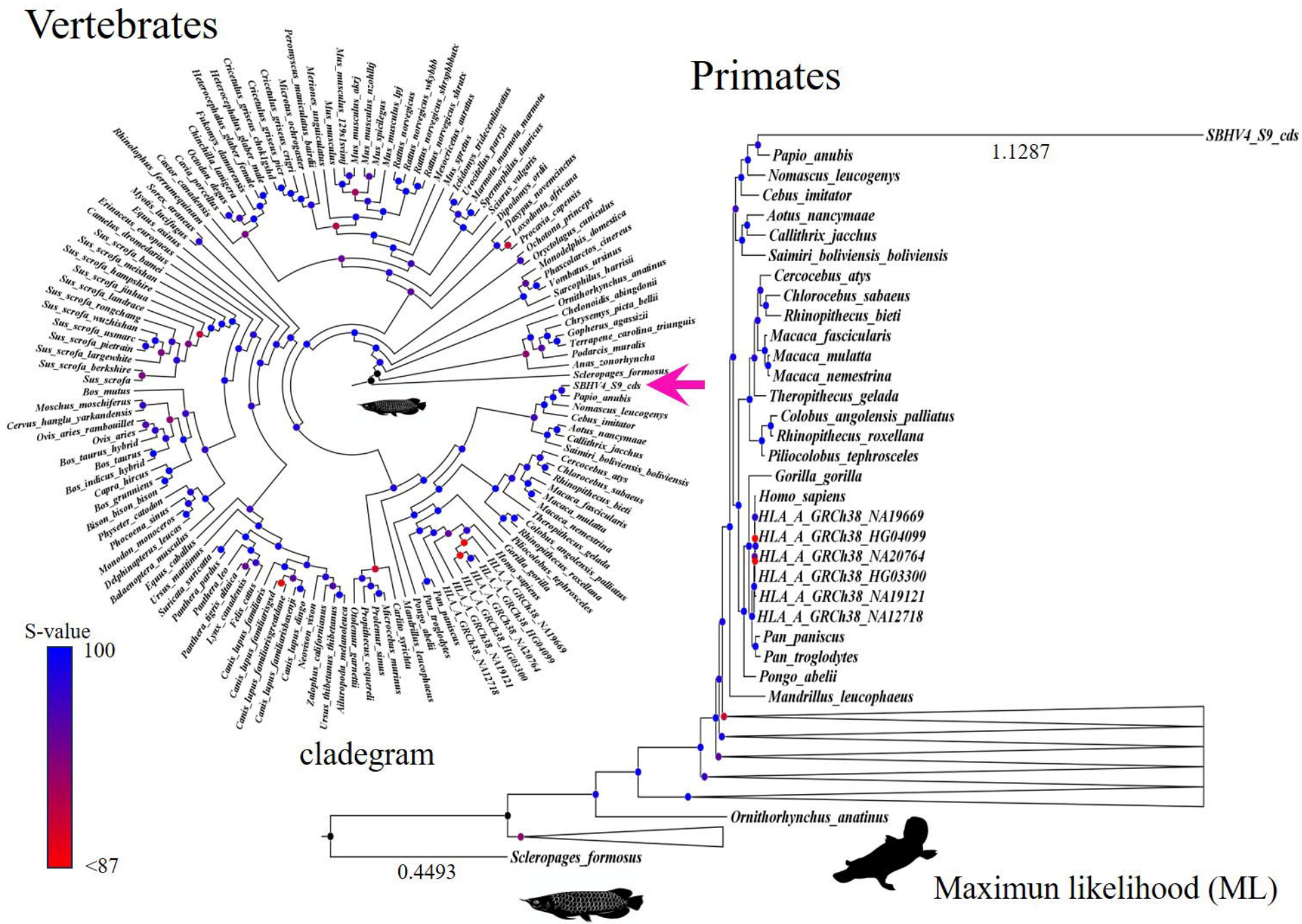
[33] Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences
Weizhong Li, Adam Godzik. Bioinformatics 2006 Jul 1;22(13):1658-9. doi: 10.1093/bioinformatics/btl158. Epub 2006 May 26.

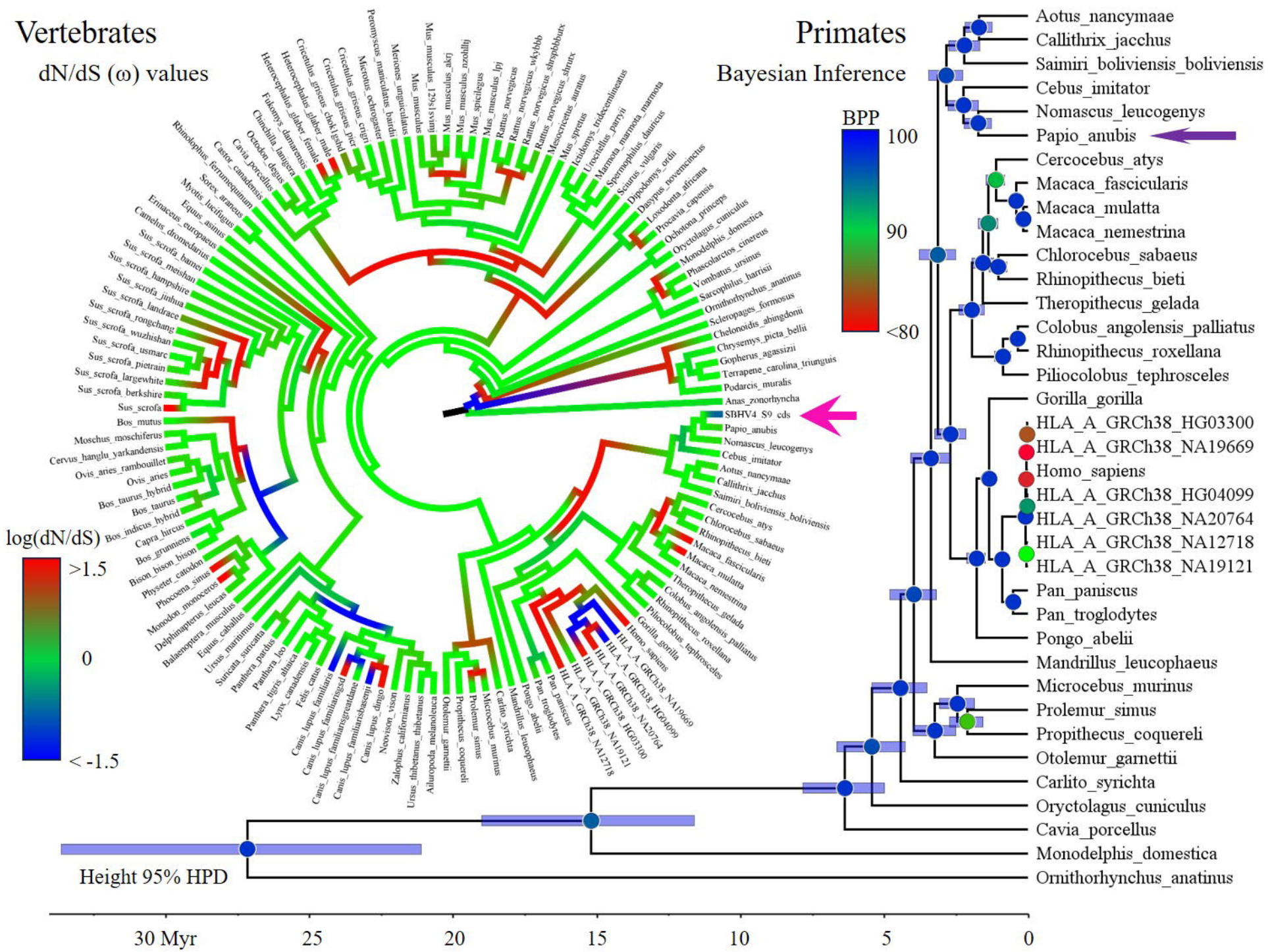
Acknowledgements

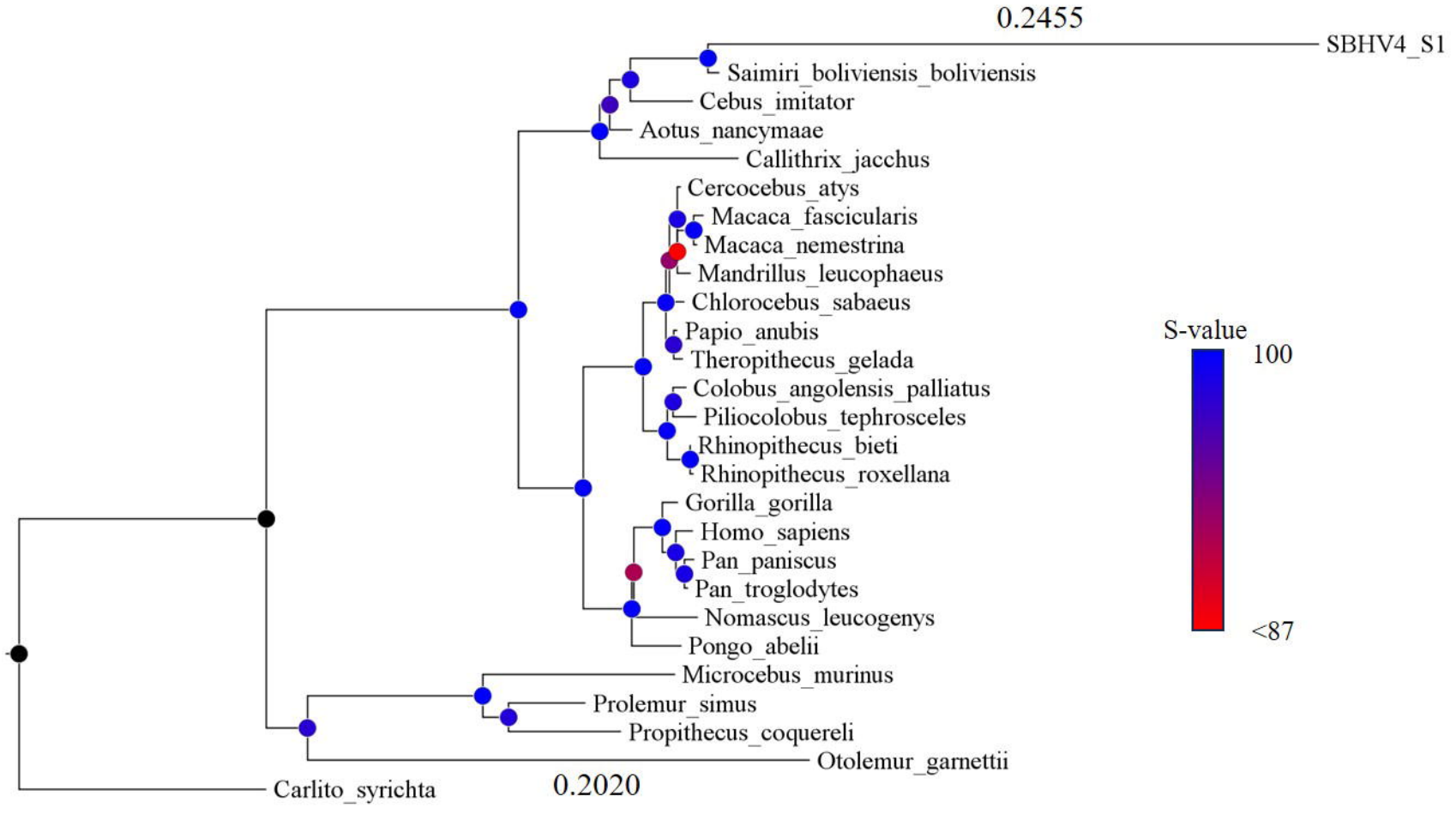
This work was partially supported by Grants-in-Aid for Scientific Research from the JSPS (24K21913 to E.H.). Computations were performed on the NIG supercomputer at ROIS National Institute of Genetics, Japan.

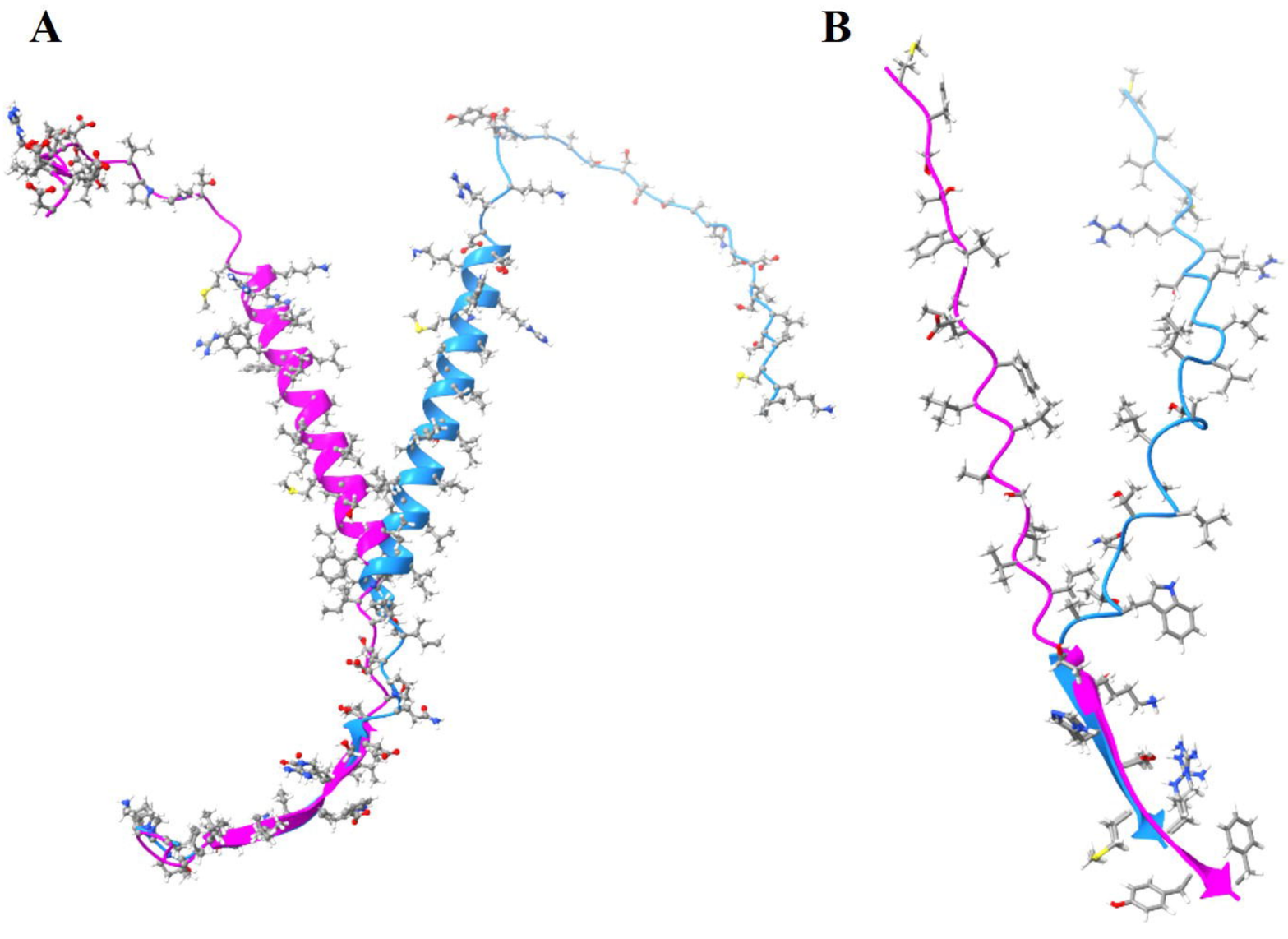

COOH
TM-score=0.84, RMSD=2.55
HLA-A
Papio anubis
SBHV4-S9
TM-score=0.83,
RMSD=1.98
α1 α2
TM-score=0.85,
RMSD=2.40
NH2
NH2
α3
TM-score=0.78,
RMSD=2.86
TM-score=0.78,
RMSD=2.71
TM-score=0.78,
RMSD=2.84
A
S. formosus
B
SBHV4-S9 + HLA-A
SBHV4-S9 + S. formosus

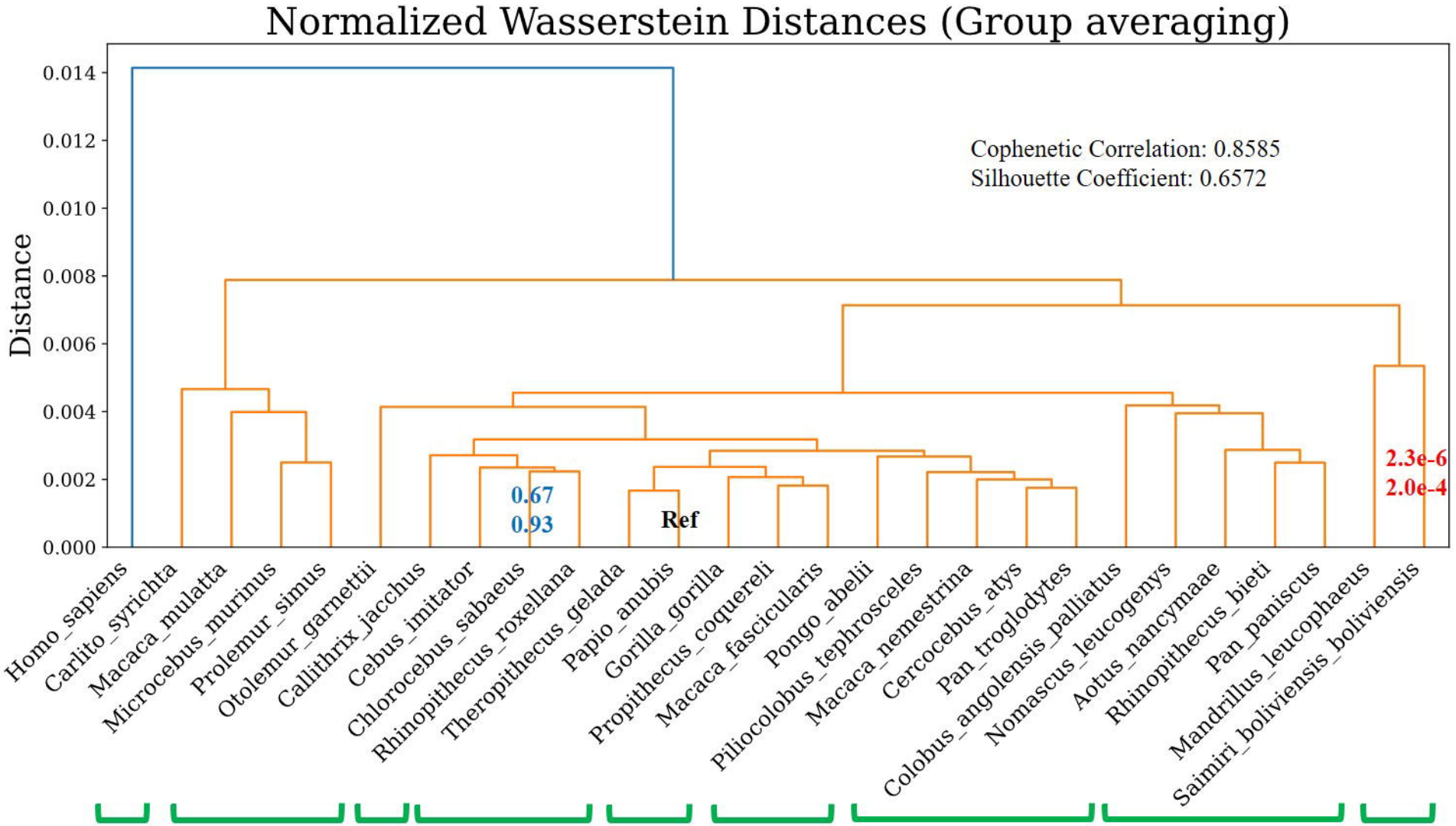

SBHV4 S9 concatemer
5'
3'
1
100
200
300
400
13
1
TSDs
S9 homologous to HLA-A (945 bp → 50 bp)
dsDNA sequences
5'
3'
10
1
TSDs
S9 gene full length,
50 + 87 bp
HLA-A cds 50 + 150 bp (lifted-over)
SBHV4 S9 concatemer
5'
3'
1
100
200
300
400
0.5
0
GC contents (window size 25)


Vertebrates
Primates
SBHV4_S9_cds
1.1287
Papio_anubis
Nomascus_leucogenys
Cebus_imitator
Aotus_nancymaae
Callithrix_jacchus
Saimiri_boliviensis_boliviensis
Cercocebus_atys
Chlorocebus_sabaeus
Rhinopithecus_bieti
Macaca_fascicularis
Macaca_mulatta
Macaca_nemestrina
Theropithecus_gelada
Colobus_angolensis_palliatus
Rhinopithecus_roxellana
Piliocolobus_tephrosceles
Gorilla_gorilla
Homo_sapiens
HLA_A_GRCh38_NA19669
HLA_A_GRCh38_HG04099
HLA_A_GRCh38_NA20764
HLA_A_GRCh38_HG03300
HLA_A_GRCh38_NA19121
HLA_A_GRCh38_NA12718
Pan_paniscus
Pan_troglodytes
Pongo_abelii
Mandrillus_leucophaeus
Ornithorhynchus_anatinus
Scleropages_formosus
0.4493
Maximun likelihood (ML)
cladegram
S-value
100
<87


Vertebrates
dN/dS (ω) values
Primates
Bayesian Inference
BPP
100
90
<80
log(dN/dS)
>1.5
0
< -1.5
Aotus_nancymaae
Callithrix_jacchus
Saimiri_boliviensis_boliviensis
Cebus_imitator
Nomascus_leucogenys
Papio_anubis
Cercocebus_atys
Macaca_fascicularis
Macaca_mulatta
Macaca_nemestrina
Chlorocebus_sabaeus
Rhinopithecus_bieti
Theropithecus_gelada
Colobus_angolensis_palliatus
Rhinopithecus_roxellana
Piliocolobus_tephrosceles
Gorilla_gorilla
HLA_A_GRCh38_HG03300
HLA_A_GRCh38_NA19669
Homo_sapiens
HLA_A_GRCh38_HG04099
HLA_A_GRCh38_NA20764
HLA_A_GRCh38_NA12718
HLA_A_GRCh38_NA19121
Pan_paniscus
Pan_troglodytes
Pongo_abelii
Mandrillus_leucophaeus
Microcebus_murinus
Prolemur_simus
Propithecus_coquereli
Otolemur_garnettii
Carlito_syrichta
Oryctolagus_cuniculus
Cavia_porcellus
Monodelphis_domestica
Ornithorhynchus_anatinus
SBHV4_S9_cds
Height 95% HPD
30 Myr
25
20
15
10
5
0


0.2455
SBHV4_S1
Saimiri_boliviensis_boliviensis
Cebus_imitator
Aotus_nancymaae
Callithrix_jacchus
Cercocebus_atys
Macaca_fascicularis
Macaca_nemestrina
Mandrillus_leucophaeus
Chlorocebus_sabaeus
Papio_anubis
Theropithecus_gelada
Colobus_angolensis_palliatus
Piliocolobus_tephrosceles
Rhinopithecus_bieti
Rhinopithecus_roxellana
Gorilla_gorilla
Homo_sapiens
Pan_paniscus
Pan_troglodytes
Nomascus_leucogenys
Pongo_abelii
Microcebus_murinus
Prolemur_simus
Propithecus_coquereli
Otolemur_garnettii
Carlito_syrichta
0.2020
S-value
100
<87


A
B



	Cluster 1	Cluster2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9
M-test	p=3.6e-10	6.9e-9	0.036	0.94	Ref Group	0.66	0.90	0.0015	5.0e-9
P-test	p=0.0002	0.0002	0.3002	0.7980	Rep Group	0.7592	0.7612	0.0260	0.0002