Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

All three parrot superfamilies have lost TLR21, the immune sensor birds use to detect bacterial and viral DNA

Babalo Reda

gbabaloreda@gmail.com

Independent researcher

Research Article

Keywords: TLR21, immunogenetics, gene loss, parrots, comparative genomics, innate immunity

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11600183/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

All three parrot superfamilies have lost TLR21, the immune sensor birds use to detect bacterial and viral DNA

Babalo Reda
Independent researcher, Toronto, Ontario, Canada
Correspondence: gbabaloreda@gmail.com

Running title: TLR21 loss across parrots

Original Paper | Immunogenetics

Keywords: TLR21; immunogenetics; gene loss; parrots; comparative genomics; innate immunity

Acknowledgements

I gratefully acknowledge the researchers, institutions and sequencing teams whose efforts in generating, preserving and openly sharing genomic and transcriptomic datasets made this study possible.

I also extend my sincere appreciation to the conservationists, wildlife researchers, veterinarians, wildlife carers, volunteers and organisations who work tirelessly to protect and conserve endangered parrots. Their dedication to preserving these remarkable birds, protecting their habitats and supporting their survival is invaluable. This research is offered in recognition of their efforts and in the hope that a deeper understanding of parrot genetics and immune evolution will contribute to their long-term conservation.

Funding

This study was entirely self-funded by the author and received no external funding.

Abstract

Parrots include some of the world's most endangered birds, yet important aspects of their immune evolution remain poorly understood. Here, we present evidence for the evolutionary loss of Toll-like receptor 21 (TLR21), an immune receptor involved in recognising bacterial and viral DNA, across all three parrot superfamilies. Using comparative genomic analysis of 34 parrot genome assemblies, supported by long-read sequencing, transcriptomic searches and detection-sensitivity analyses, we identified previously unresolved deletions involving TLR21 and neighbouring genes. In Psittacoidea, we discovered an inherited deletion removing three genes, PRR19, TLR21 and TMEM145, from the region between CD79A and MEGF8. Shared junction sequences and spanning long reads support inheritance of this structural change across divergent parrot families. In black cockatoos, we identified a structurally distinct deletion removing TLR21 and TMEM145, supported by seven and eleven spanning reads from Baudin's and Carnaby's black cockatoos, respectively. In the critically endangered kākāpō, long-read sequencing revealed an altered ancestral TLR21 locus, with two independent reads spanning the expected region without detectable TLR21 sequence. This extends the evidence for receptor loss into the third parrot superfamily. Complementary analysis of 534 million kea sequencing reads found no credible TLR21 sequence. In contrast, analysis of up to 52 non-parrot bird species revealed that TLR21 remains under purifying selection, while structural analysis identified a conserved cluster of positively charged surface residues representing a candidate DNA-interaction region. An exploratory comparison of 496 viral genomes representing 39 species revealed differences in viral DNA composition between parrot-associated viruses and those associated with other birds, with the strongest and most consistent contrast occurring among herpesviruses. Together, these findings identify substantial evolutionary restructuring of an ancestral immune-gene region and raise important questions about how parrots recognise microbial DNA without TLR21. The discoveries provide new directions for investigating the relationship between parrot immune evolution and viral diseases, including psittacine beak and feather disease, with potential relevance to endangered parrot conservation.

Introduction

Parrots include some of the world's most endangered birds, yet important aspects of their innate immune evolution remain poorly understood. Understanding their evolutionary history requires examining not only the genes they have retained but also those they have lost. Changes in immune genes are particularly important because these genes influence how animals recognise pathogens and initiate protective responses. One significant question concerns the evolutionary disappearance of Toll-like receptor 21 (TLR21), an immune receptor involved in recognising microbial DNA.

Toll-like receptors (TLRs) recognise microbial molecules and initiate innate immune responses. Their evolutionary history includes gene duplication, loss and pseudogenisation across vertebrates (Velová et al. 2018; Liu et al. 2020). Birds lack TLR9, the mammalian receptor for unmethylated CpG DNA, but retain TLR21, which recognises the same class of ligand. In chickens, TLR21 responds to synthetic CpG DNA and bacterial genomic DNA; introducing the receptor into unresponsive cells confers CpG responsiveness, whereas silencing it in macrophages reduces that response (Keestra et al. 2010). The loss of TLR21 in parrots therefore raises an important question about how these birds recognise microbial DNA without a receptor retained in other birds.

Establishing gene loss requires examining the physical genomic region, because GC-rich genes can be difficult to sequence and assemble. Genes once considered missing from birds have subsequently been recovered from transcriptomes (Botero-Castro et al. 2017). In the orange-bellied parrot, MHC class II genes absent from an assembly and raw HiFi reads were recovered using PCR and transcriptomic analyses (Silver et al. 2025). Reconstructing the region surrounding TLR21 and identifying long reads connecting surviving flanks can therefore reveal which genes were physically removed and which junction structures were inherited.

Parrots comprise three major superfamilies, Psittacoidea, Cacatuoidea and Strigopoidea, including the critically endangered kākāpō and threatened black cockatoos. Previous work reported that TLR21 could not be located in the orange-bellied parrot genome (Silver et al. 2025). The broader distribution of receptor loss, the physical structure of its ancestral locus and the relationship between deletion events remained important questions. Comparing the ancestral region, ARHGEF1–CD79A–PRR19–TLR21–TMEM145–MEGF8, across these lineages can distinguish shared derived junctions from structurally different outcomes. A common ancestral loss followed by further rearrangements and separate initial losses could both produce the surviving structures.

Materials and Methods

Study scope and genomes

We used public sequence data to investigate whether TLR21 was physically lost in parrots and to reconstruct changes in its surrounding genomic region. The main panel comprised 34 parrot genome assemblies from NCBI, covering all three superfamilies: Strigopoidea, Cacatuoidea and Psittacoidea. Sampling across these lineages allowed us to assess the distribution of TLR21 absence and compare the surviving locus across the principal branches of parrot evolution. Assembly quality is summarised in Online Resource 1, Fig. S7.

Five non-parrot species provided reference TLR21 sequences and neighbouring genes: chicken, zebra finch, great tit, peregrine falcon and lesser kestrel. These comparisons helped identify the expected locus and distinguish retained sequence from deletions or rearrangements. We included both the peregrine RefSeq assembly and a 2025 long-read assembly to assess how assembly completeness affected recovery of the region.

Gene and locus searches

We combined two complementary approaches to search for TLR21 independently of existing gene annotations. Miniprot aligned reference proteins to genomic DNA while allowing for introns, the non-coding intervals that separate coding exons, to reconstruct candidate gene models (Li 2023). DIAMOND blastx searched translated DNA in 100-kb windows for matching protein segments, allowing us to detect recognisable coding fragments even where a complete gene model could not be reconstructed (Buchfink et al. 2021). TLR21 proteins from chicken, zebra finch, great tit, lesser kestrel and peregrine falcon served as references.

Other chicken TLRs served as both detection controls and alternative matches. Their recovery tested whether the searches could identify related receptors in the same datasets. Comparing ambiguous sequences against these proteins helped distinguish TLR21 from other members of the TLR family. Matches confined to the low-complexity N-terminus, where repeated amino-acid patterns can resemble unrelated proteins, were not accepted as TLR21 recovery without supporting similarity elsewhere in the protein.

We located the expected TLR21 neighbourhood using ARHGEF1, CD79A, PRR19, TMEM145 and MEGF8. These neighbouring genes provided anchors for comparing gene order, orientation and retained sequence across species. Conservation of this local arrangement, or synteny, helped us identify the corresponding locus even where TLR21 itself was missing. We also searched with blastn for two non-repetitive junction markers, B and C, identified in the budgerigar, Amazon and monk parakeet loci. Comparing their positions with those in non-parrot reference loci distinguished ancestral flanking sequence from new sequence adjacencies associated with deletion. The informative feature was a new connection between sequences that were separated in the reference locus. Shared ancestral sequence helped locate the region; a shared new connection helped identify an inherited deletion.

Each species was classified according to the strongest available evidence (Online Resource 1, Table S1).

Raw-read tests

We searched the original sequencing reads to test whether TLR21 was present in DNA omitted from the genome assemblies. We then examined reads crossing the expected locus to determine whether individual DNA molecules connected the surviving flanks without TLR21 between them.

DIAMOND blastx searches included TLR21 reference proteins, neighbouring genes, GC-rich control genes and other TLRs. Each control addressed a different question. Neighbouring genes tested recovery from the same genomic region; GC-rich genes tested recovery of compositionally difficult sequence; and other TLRs tested receptor detection while providing alternative matches for sequences resembling TLR21.

PacBio continuous long reads (CLR) and Oxford Nanopore (ONT) reads were searched in frameshift mode (-F 15, range culling, $E < 10^{-5}$). This allowed alignments to accommodate sequencing insertions and deletions that disrupt the translated reading frame. Range culling retained matches from different regions of a long read, helping recover multiple genes on the same molecule. PacBio HiFi reads were searched without frameshift mode, using -k 10 --max-hsps 3.

We searched the complete Carnaby's black cockatoo ONT run, comprising 7.41 million reads and 81.8 Gb. Large runs were streamed in byte ranges and checked against archive checksums. For Carnaby's black cockatoo and the kea, each range was also decompressed and checked record by record; damaged ranges were downloaded again before analysis. These checks ensured that incomplete or damaged downloads were not mistaken for sequence absence.

Kea Illumina reads were searched individually with DIAMOND blastx in sensitive mode at $E < 10^{-3}$. Candidate TLR21 reads had to select a TLR21 reference as their best panel hit, align over at least 25 amino acids and reach at least 39% identity. The identity threshold was the 10th percentile of matches recovered for the TLR3 and TLR7 controls. Candidates were then compared with the full chicken proteome and accepted as TLR21 only when TLR21 remained their best match. This broader comparison tested whether another protein could explain a short apparent match. Kea control reads and Carnaby's TLR21-like reads were checked in the same way.

Reads matching locus genes were examined with blastn, tblastx and minimap2 overlaps(Li 2018)to compare gene order, orientation and low-copy sequence anchors. These comparisons tested whether the proposed structure had consistent sequence support. Overlaps restricted to repeats were not accepted as connections between flanks because similar repeats can occur at different genomic positions.

To test whether sequencing errors could conceal TLR21, we simulated 400 reads per condition from intact lesser-kestrel and peregrine loci. Read lengths were drawn from the observed CLR and ONT datasets, with random substitutions and indels introduced at approximately 14% and 20% error. Falcon TLR21 references were excluded from the search panel so that detection required similarity to other bird references. This measured detection sensitivity for reads containing TLR21. A separate calculation used CD79A read counts as a proxy for local coverage to estimate the chance of missing all TLR21-overlapping reads under that assumption. The two calculations addressed complementary parts of detection: recognising TLR21 within a read and obtaining reads that cover its expected position. The raw-read datasets are summarised in Online Resource 1, Table S2.

RNA

We searched RNA-sequencing data for expressed TLR21 that might have escaped genomic detection. The panel comprised five datasets from four parrot species: blood from Ducorps's corella (SRR10852853) and African grey (SRR10852952), and spleen from monk parakeet (SRR13480251 and SRR13480262) and orange-bellied parrot (SRR30410751). TLR3 and TLR7 served as controls to test whether the searches could recover expressed TLRs in each sample.

Blood reads, reported as paired-end 50-bp sequences, were searched with DIAMOND blastx against the TLR reference panel. Candidate matches were then compared against the chicken proteome to assess whether another protein provided a better explanation for the similarity. Spleen reads were searched against avian TLR21, TLR3 and TLR7 proteins, followed by a best-hit comparison against UniProtKB/Swiss-Prot. These broader comparisons helped distinguish genuine receptor matches from short sequences shared among related or repetitive proteins.

The RNA searches provided an expression-based comparison with the genomic and spanning-read results. TLR3 and TLR7 confirmed recovery of expressed receptors in each sample. Because the panel lacked a TLR21-positive non-parrot sample, TLR21-specific transcript-search sensitivity remained unmeasured.

Selection on TLR21 in other birds

We examined selection on retained TLR21 to establish the evolutionary context of its loss in parrots. Comparisons with TLR3, TLR4, TLR5 and TLR7 tested whether TLR21 differed from other avian TLRs in selective constraint and protein divergence.

RefSeq candidates for TLR3, TLR4, TLR5 and TLR7 were obtained from chicken-centred ortholog collections. TLR21 candidates were collected through searches for TLR21 and TLR13 names and a broader search of avian proteins annotated as Toll-like receptors, reducing dependence on consistent gene naming. Parrots were excluded from this analysis.

To reduce errors caused by incomplete or corrected gene models, we retained the longest valid coding sequence per gene that began with ATG, ended with a stop codon, contained no internal stops and encoded a protein 80–125% of the chicken reference length. Models labelled “LOW QUALITY PROTEIN” or modified relative to the genomic sequence were excluded. Reciprocal best-hit comparisons against a chicken TLR panel checked receptor identity in both directions: the candidate had to match the intended chicken receptor best, and that receptor had to select the candidate as its best match in the species. This reduced the risk of including a related TLR in the wrong gene alignment. Species with suspected multiple TLR21 copies were excluded from the single-copy analysis.

We used three species sets. The core set contained 22 species with all five receptors, allowing direct comparisons among genes in the same species. A second set contained 37 species with TLR21, TLR3, TLR4 and TLR5. A larger TLR21-only set contained 52 species, allowing us to assess whether its estimated constraint persisted with broader sampling.

Proteins were aligned with MAFFT L-INS-i(Katoh & Standley 2013)and back-translated to their coding sequences, using amino-acid correspondence to guide the alignment while preserving codons. Codon columns with more than 50% gaps were removed. An Open Tree of Life topology was pruned to each species set and held constant across genes within that set.

We used HyPhy(Kosakovsky Pond et al. 2020) to estimate the overall nonsynonymous-to-synonymous substitution-rate ratio, dN/dS, under an MG94×REV codon model(Kosakovsky Pond & Frost 2005). This compares substitutions that change an amino acid with those that do not; a ratio below one is consistent with purifying selection against protein-altering changes. FEL identified individual codons showing evidence of purifying or positive selection, using a site-level threshold of $p < 0.1$. BUSTED tested for episodic positive selection affecting at least one site on at least one tested branch.

Pairwise estimates from PAML yn00(Yang & Nielsen 2000)provided a separate methodological check on dN, dS and their ratio. Comparisons with $dS \geq 3$ were excluded from the reported pairwise summaries because repeated substitutions at the same sites make highly divergent comparisons less reliable. We also measured amino-acid identity and GC3, the proportion of guanine and cytosine at third codon positions. These measures helped distinguish low dN/dS from low protein divergence. The ratio measures amino-acid-changing

substitutions relative tosynonymous substitutions, whereas amino-acid identity describes how similar the proteins remain. Examining both was especially important for TLR21 because of its high GC3 and synonymous divergence.

To compare the extracellular and intracellular portions of each receptor, we divided the alignment using the transmembrane region predicted from the chicken protein's hydrophobicity with a 19-residue window. This scan identifies a stretch of hydrophobic amino acids consistent with the segment that crosses the cell membrane. For TLR21, the predicted transmembrane segment covered residues 753–771. It provided an approximate boundary between the ectodomain, the portion outside the cytoplasm, and the cytoplasmic region containing the TIR signalling domain.

Selection in parrots

We tested whether parrot branches differed from non-parrot branches in selection on other DNA-sensing genes, shared signalling partners and the neighbouring gene CD79A. Genes were rebuilt with miniprot from chicken protein references to reduce differences arising from species-specific annotation pipelines. HyPhy RELAX compared parrot branches with non-parrot reference branches (Wertheim et al. 2015). Its parameter K measures changes in selection intensity: $K < 1$ moves the distribution of dN/dS values towards neutrality, whereas $K > 1$ moves it away from neutrality. The comparison asked whether selection differed across the sampled parrot branches as a group.

The DNA-sensing and signalling set comprised cGAS, STING, TBK1, IRF7, MYD88 and DDX41. We also examined the set without MYD88. Ten partner genes provided a comparison group: CNPY3, HSP90B1, LGMN, CTSB, CTSK, SLC15A4, IRAK2, TOLLIP, IRAK4 and TRAF6. The sets cover different steps in DNA recognition and signalling, from receptors and adaptors to enzymes and transcription factors.

Initial fits were sensitive to starting values: the same alignment could yield different parameter estimates depending on where the optimisation began. We therefore applied three fitting procedures to each target and all 504 random control genes: a default fit, a five-start fit and a ten-start fit with a finer grid. Trying several starting points helped the fitting procedure find a better solution, and the fit with the highest likelihood was retained. Applying the same procedure to targets and controls made their comparison consistent. Each target's K was ranked against the control distribution, and group mean ranks were evaluated by permutation. This placed the target genes against the background variation in genes analysed through the same pipeline. Comparisons with covariate-matched controls tested whether the group pattern persisted among controls with comparable characteristics; a direct comparison with partner genes tested whether the pattern was concentrated in the DNA-sensing group.

UNC93B1 was examined separately in an expanded species panel that recovered clean models from nine parrots. That analysis used the 59-control expanded panel and is distinguished from the 504-control multi-start analysis.

For CD79A, locus-based gene models were reconstructed from budgerigar and zebra-finch proteins. The selection alignment contained 217 codons from four parrots and two non-parrots. RELAX tested for a change in selection intensity, and aBSREL tested for episodic positive selection on individual branches. To examine the adjacent non-coding region, 15 kb downstream of the stop codon were compared among species with low-complexity-masked dc-megablast at $E < 10^{-5}$. The purpose was to identify conserved non-coding sequence that might have been affected by the adjacent deletion.

Copy number

We searched for additional copies of immune genes that might provide an evolutionary route to compensation for TLR21 loss. Miniprot aligned a panel of 26 chicken immune proteins to 14 parrot and five non-parrot

genomes, retaining up to 30 alignments per query. A candidate locus required at least 50% query coverage and 35% identity. Overlapping matches were counted once and assigned to the best-matching panel protein. An intact model additionally required at least 90% coverage and no frameshift or premature stop.

The panel included nucleic-acid sensors and related immune proteins. These searches assessed gene-copy expansion as one possible route to compensation. Assembly structure affects the interpretation: fragmentation can split one gene across contigs, duplicated haplotypes can resemble extra copies, and collapsed sequence can merge distinct copies. Apparent extra loci on short or fragmented contigs were therefore treated as candidates for verification.

Repeat context

We tested whether repeat enrichment at the surviving parrot junctions was also present in intact outgroup loci. Tandem repeats were detected with Tandem Repeats Finder (Benson 1999), using identical settings throughout: 2 7 7 80 10 50 2000. For assembly-based comparisons, each locus was compared with 300 windows of the same length, matched to its GC fraction within ±0.03. Windows contained less than 1% ambiguous bases and came from contigs longer than 1 Mb.

For Baudin's black cockatoo, the background was drawn from 40,000 reads from the same sequencing run, allowing the locus to be compared with sequences affected by the same sequencing technology. The kākāpō comparison used assembly windows with simulated errors; because that background under-represents repeats, its percentile was not used to infer enrichment. Intact falcon loci provided the comparison relevant to whether repeat enrichment preceded the parrot deletions.

Structural conservation

We mapped amino-acid conservation across 52 birds onto the AlphaFold model of chicken TLR21 (AF-A0A8V0ZYL3-F1, version 6; Jumper et al. 2021) to identify conserved surface regions. AlphaFold predicts the protein's three-dimensional shape, allowing residues far apart in the sequence to be examined together if folding brings them close in space. The baseline analysis retained residues conserved in at least 90% of species, with predicted local confidence pLDDT ≥70 and a surface-exposure criterion of fewer than 24 neighbouring residues. The pLDDT threshold selected positions with greater local prediction confidence, while the neighbour-count criterion selected residues exposed on the protein surface. We focused on arginine, lysine and histidine as a set of basic amino acids relevant to a possible DNA-interaction surface.

For the selected residues, we measured the physical distance between every pair in the predicted structure and calculated the mean. A smaller mean distance means the residues form a more compact cluster. We compared that value with two random-sampling reference distributions, each containing the same number of residues. Random sets were matched by exposure tertile and ectodomain third: residues were drawn from comparable levels of surface exposure and comparable positions along the ectodomain. This accounted for the fact that surface residues are distributed unevenly across the protein. The first comparison sampled exposed basic residues, asking whether the conserved subset clustered more tightly than comparable basic residues overall. The second sampled conserved exposed non-basic residues, asking whether the pattern was distinctive to conserved basic residues or also typical of other conserved surface residues.

We repeated the analysis across 35 combinations of conservation, model-confidence and surface-exposure thresholds to test whether the clustering depended on the criteria used to select residues. This sensitivity analysis showed how the result changed when those criteria were made more or less restrictive.

Checking apparent ZAP frameshifts

Protein-to-genome alignments initially suggested shared frameshifts in ZAP (ZC3HAV1). We tested whether these represented coding disruptions or a missed intron by examining the region corresponding to chicken

residues 142–297 for canonical GT–AG splice boundaries. Candidate introns were removed in silico to check whether splicing restored a stop-free reading frame. The same procedure was checked against known chicken introns.

Blood RNA reads from Ducorps's corella and African grey were mapped with minimap2 using the splice:sr preset to test whether the proposed intron was spliced in expressed RNA. Corella reads were compared with the little-corella assembly and African grey reads with its assembly. Spliced alignments and coding-sequence agreement were used to distinguish an alignment artefact from a genuine disruption.

Dating and alternative histories

We compared retained genes, flanking sequences and junction markers across the parrot phylogeny to identify histories compatible with the observed loci. We distinguished the initial loss or inactivation of TLR21 from later deletions involving neighbouring genes. For example, TLR21 could have been lost first, with a later deletion removing its remaining sequence together with neighbouring genes. The present-day junction would then record the later event.

Divergence times from Smith et al. (2023) provided a temporal framework for events supported by shared structures. A shared derived junction places an event before the divergence of the sampled lineages carrying it; its absence from another lineage constrains the history only in conjunction with that lineage's retained structure. We compared alternative histories by the number of initial TLR21 losses they required and by whether their sequence of events could produce the retained genes and junctions observed today. The dating framework is illustrated in Online Resource 1, Fig. S5.

Exploratory viral CpG comparison

We compared CpG composition in parrot-associated DNA viruses with that of viruses associated with other birds in the same genus. The panel was defined from Virus-Host DB and RefSeq metadata before calculating CpG values (Mihara et al. 2016). Viral species with mixed or unresolved host-group assignments were excluded, and each genus had to contain both host groups. An initial panel of 42 viral species was expanded with GenBank genomes from those same species. To limit the influence of heavily sampled species, at most 50 additional records per species were selected using a fixed accession-independent hash rule. This yielded 554 GenBank records plus the 42 reference records. Online Resource 1, Supporting Methods S1 gives the complete selection and filtering procedure.

We required records labelled as complete genomes, at least 1,000 bases long and containing only A, C, G and T. We excluded explicit construct records and records with a discordant or nonbird source host, then collapsed exact sequence duplicates within each viral species. Source-host qualifiers were checked for every record, and the original references and nearest comparators were audited against their source literature. We excluded NC_076425.1 because its underlying sequence came from pooled parrot and mammal material, leaving its host unresolved (Duarte et al. 2019). A separate sensitivity analysis retained only records with an explicit source-host label matching the assigned group.

CpG observed/expected (O/E) was calculated as the observed frequency of adjacent C–G pairs divided by the product of the separate C and G frequencies. This measures CpG abundance relative to the expectation from nucleotide composition. Circular genomes included the last-to-first pair. Each species was represented by its median across distinct genomes, and the two host-group medians were compared within each genus. For an overall summary, we averaged genus log-ratios within families and then gave each family equal weight. We repeated the calculation after selecting one genome per species 2,000 times to assess sensitivity to reference choice and after omitting each family in turn.

To assess whether the result depended on the choice of relatives, we also matched each parrot-associated species to the nearest available non-parrot reference using a conserved protein. These were adenovirus DNA polymerase, herpesvirus UL30 polymerase and circovirus Rep. FAMSA alignments were used to infer maximum-likelihood trees with IQ-TREE 2.4.0 where at least four references were available; smaller panels used aligned protein distances (Deorowicz et al. 2016; Minh et al. 2020). Bootstrap trees assessed the stability of the nearest-comparator choice (Hoang et al. 2018). We compared GpC, TpA, GC fraction and CpG/GpC, and repeated CpG measurements separately in annotated coding and noncoding sequence. These checks tested whether the observed contrast was specific to CpG and whether it extended across genome compartments. The analysis files are supplied in Online Resource 2.

Use of AI-assisted tools

During October–November 2026, ChatGPT/Codex (OpenAI), Claude (Anthropic) and DeepSeek were used as assistive tools for methodological advice, coding, analysis checks, figure preparation, and manuscript drafting and editing. Their use supported work across the manuscript under the author's direction. AI-assisted outputs were evaluated against source data, computational results and relevant literature. The author developed and wrote the manuscript with AI assistance and takes responsibility for its accuracy and interpretation. The study used publicly available research data; no confidential or human participant data were supplied to these tools.

Results

Discovery of a shared three-gene deletion in Psittacoidea

Our comparative genomic analysis identified a shared deletion in Psittacoidea removing PRR19, TLR21 and TMEM145 from the region between CD79A and MEGF8. This brings CD79A and MEGF8 into direct proximity. The strongest physical evidence comes from the budgerigar: both haplotype assemblies contain the junction, supported by 13 spanning PacBio reads, four maternal and nine paternal. No analysed read supports the intact arrangement. The locus lies beside ARHGEF1, consistent with its position in falcon references (Figs. 4 and 5).

The yellow-crowned Amazon has the same compact CD79A–MEGF8 arrangement, supported by one HiFi read aligning at 99.3% identity across the junction. In the monk parakeet, the genes are joined without an assembly gap on chromosome 23, although no spanning raw read was established. The sun-conure assembly is consistent with this arrangement but contains a 6.6-kb gap. In the ring-necked parakeet, the CD79A-side junction markers are retained while MEGF8 lies approximately 0.3 Mb away, consistent with a subsequent rearrangement.

The gene-anchored junction intervals in budgerigar, Amazon and monk parakeet are approximately 3.1, 5.7 and 7.8 kb. Shared non-repetitive segments occur in the same order, and marker B lies beside CD79A in nine sampled Psittacoidea species spanning Psittaculidae and Psittacidae. Marker C also occurs beside MEGF8 in falcons, making it ancestral flanking sequence. Its adjacency to CD79A in the parrot junction is derived. The shared sequence arrangement supports inheritance of the deletion within the sampled Psittacoidea. Additional locus comparisons are shown in Online Resource 1, Fig. S2.

Western ground parrot reads also match this junction type. The single CD79A-containing read extends into budgerigar-type junction sequence, and two MEGF8-containing reads carry corresponding sequence on the other side. No read spans both genes. An apparent 1.2-kb overlap at 92% identity was rejected as repeat similarity, leaving this species classified as probable absence consistent with the shared deletion.

The junction's occurrence in both Psittaculidae and Psittacidae places its origin before the divergence of the sampled carriers. Its shared sequence arrangement provides evidence that the deletion was inherited across these lineages.

A distinct two-gene deletion in black cockatoos

Our analysis identified a distinct deletion removing TLR21 and TMEM145 from the ancestral region between CD79A and MEGF8 in black cockatoos. Seven ONT reads from Baudin's black cockatoo (Zanda baudinii), ranging from 15 to 31 kb, span the locus without either gene. The surviving arrangement includes ARHGEF1, CD79A and a short proline-rich open reading frame before MEGF8. The CD79A–MEGF8 interval is approximately 9.0 kb.

The intervening open reading frame is not securely identified as PRR19. It contains 57% low-complexity sequence, and its masked match to kestrel PRR19 covers 143 amino acids at 27% identity. In the gene tree it groups with chicken proline-rich look-alikes rather than the falcon and songbird PRR19 orthologs. We therefore describe it as a PRR19-like ORF of unresolved orthology. The deletion of TLR21 and TMEM145 is supported independently of that assignment.

The cockatoo junction lacks marker B. Comparisons with Psittacoidea detected shared ancestral flanking sequence and tandem repeats, but no unique derived junction sequence connecting the two deletion structures. The surviving cockatoo arrangement is therefore structurally distinct from the Psittacoidea junction.

Carnaby's black cockatoo (Zanda latirostris) carries the same junction type. The complete ONT run contained 7.41 million reads and 81.8 Gb, approximately 65-fold genome coverage. Eleven reads of 12–27 kb connect CD79A to MEGF8 without TLR21 or TMEM145 between them; several also contain the PRR19-like ORF. Seventy-two reads whose best panel hit was a TLR21 reference were rechecked against the full chicken proteome. All matched another protein better, most often TIRAP. CD79A was recovered in 86 reads and MEGF8 in approximately 1,600. As in Baudin's black cockatoo, marker B is absent and marker C is retained. Distances between CD79A and MEGF8 overlap between the species, and the Carnaby's reads align in the same order to the Baudin's consensus. This shared structure supports inheritance of the deletion within Zanda.

Long-read evidence extends TLR21 loss to the endangered kākāpō

Our analysis identified an altered ancestral TLR21 region in the critically endangered kākāpō, extending evidence for receptor loss into Strigopoidea. Two CLR reads, SRR25655832.70172 and SRR25655839.18027, extend from CD79A across approximately 9–11 kb to sequence matching the TMEM145 side of the falcon locus. The reads measure 20.7 and 21.1 kb, and their overlapping sequences support the same local arrangement. Neither contains detectable TLR21 between the anchors, including in translated searches that recover the distal flanking sequence.

The distal anchor contains a partial, spliced TMEM145-like model with approximately 65% identity to kestrel. It has no alternative unique placement in the available kākāpō assembly. This places the distal end of the reads on the TMEM145 side of the reference locus. No read joins CD79A to MEGF8, and no marker B was detected in the reads or assembly. The recovered structure therefore differs from the compact Psittacoidea junction.

These observations support probable, read-supported loss of TLR21 in kākāpō. The interval retains approximately the falcon spacing but lacks recognisable TLR21. Its complete sequence remains unresolved because only two noisy reads cross the interval, and neither links the TMEM145-side sequence to MEGF8. The structure is compatible with either an early shared loss followed by later changes elsewhere in parrots or an independent loss in the kākāpō lineage.

In the kea (Nestor notabilis), searches of 534 million Illumina reads (53.4 Gb) found no credible TLR21 sequence. The same searches recovered 573 TLR3 reads and 1,169 TLR7 reads at 60% identity or more, with almost all retaining the corresponding receptor as their best chicken-proteome match. Of 2,478 reads whose best panel hit was a TLR21 reference, 2,465 matched other leucine-rich-repeat proteins better. The remaining 13 mapped to one 100-bp genomic site covering a single leucine-rich repeat at about 50% identity. CD79A and MEGF8 reads were also recovered. The kea is therefore classified as "not detected in raw short reads". Its non-

detection is consistent with loss in Strigopoidea, while the short-read data leave the local arrangement unresolved.

TLR21 is absent from all 34 sampled parrot genome assemblies

Our analysis found no detectable TLR21 sequence in any of the 34 sampled parrot genome assemblies, although other Toll-like receptors were recovered in the same datasets. In contrast, intact TLR21 was recovered in chicken, zebra finch, great tit and lesser kestrel. The assembly and raw-read evidence are summarised in Table 2 and Fig. 3.

The TLR21 neighbourhood was nevertheless incompletely represented in many parrot assemblies. In falcons, the genes occur in the order ARHGEF1–CD79A–PRR19–TLR21–TMEM145–MEGF8. Much of this block was missing or fragmented in parrots, with CD79A at least 95% complete in only four of the 34 assemblies. The peregrine provides a direct example of assembly dropout: TLR21 is missing from its older RefSeq assembly but present in a 2025 long-read assembly. The searched subset of its PacBio long reads (SRR31755801) recovered only two TLR21 reads against 29 for TLR3.

The difference between assemblies and raw-read recovery made the local sequence connections decisive. We therefore examined raw reads and physical connections across the locus wherever suitable data were available.

Raw-read searches and detection controls provide further evidence of TLR21 loss

Our raw-read searches consistently recovered neighbouring genes and other immune receptors without credible TLR21 sequence. In kākāpō, CD79A was recovered in nine reads and MEGF8 in 45, despite neither appearing in its assembly. Other TLRs and GC-rich control genes were also detected. Apparent TLR21 matches were confined to a GA-repeat read or the receptor's low-complexity N-terminus.

In the black cockatoo searches, re-examined TLR21-like matches were short TIR-domain sequences that matched TIRAP more closely. Western ground parrot reads likewise produced no best match to TLR21, while other TLRs and GC-rich controls were recovered.

Simulations showed that the search could recover TLR21 despite substantial read error. Among simulated reads containing more than 1 kb of TLR21, recovery was 93% for CLR-like reads and 98% for ONT-like reads at approximately 14% error. At 20% error, recovery was 53% and 79%, respectively. These values measure the chance of recognising TLR21 when a read contains enough of the gene to search. Calculations using CD79A as a proxy for local coverage also supported detection under that coverage assumption. The spanning reads addressed the structural question directly by connecting the surviving flanks on individual DNA molecules. Simulation results are shown in Online Resource 1, Fig. S4.

No credible TLR21 transcript was detected in the five blood and spleen RNA datasets. TLR3 and TLR7 were detected in every dataset. Apparent TLR21 matches were short, ambiguous or better assigned to other proteins after broader comparison. The RNA observations agree with the genomic evidence within the sampled tissues and conditions.

Distinct genomic deletions reveal the evolutionary history of TLR21 loss in parrots

Our analysis identified distinct genomic changes at the ancestral TLR21 locus across all three parrot superfamilies. The lesser-kestrel and peregrine references share ARHGEF1–CD79A–PRR19–TLR21–TMEM145–marker C–MEGF8. Relative to this arrangement, sampled Psittacoidea have lost PRR19, TLR21 and TMEM145; black cockatoos have lost TLR21 and TMEM145; and the kākāpō retains TMEM145-side sequence without detectable TLR21.

At least two distinct deletion configurations are physically supported, and the probable kākāpō loss occurs in a different retained interval. These differences can arise through more than one evolutionary history. An early

TLR21 loss could have been followed by larger deletions that removed different neighbouring sequences in different lineages. Alternatively, separate losses could have occurred within each lineage. Both routes can produce distinct present-day junctions.

One loss before the diversification of living parrots is the history requiring the fewest initial TLR21 losses. Histories with two or three initial losses remain compatible with the data. The one-loss history is therefore the most parsimonious for the initial loss of the receptor, while requiring later changes to account for the different locus structures. The evidence supports TLR21 loss across all three superfamilies, with the Strigopoidea inference resting on probable loss in kākāpō. Kea short-read non-detection provides additional evidence from Strigopoidea, while its locus structure remains unresolved. Alternative histories are summarised in Fig. 6 and Online Resource 1, Table S3.

Repeat enrichment at the deletion junctions

Our analysis revealed elevated repetitive DNA at several Psittacoidea deletion junctions. Tandem-repeat content at the budgerigar, Amazon and monk parakeet junctions ranked at the 98th, 85th and 99th percentiles of length- and GC-matched windows, compared with the 46th and 90th percentiles for intact peregrine and lesser-kestrel intervals. The cockatoo junction ranked at the 60th percentile against reads from the same run (Online Resource 1, Fig. S3).

The parrot junctions are repeat-rich, whereas repeat enrichment varies between the intact outgroup loci. Repeated sequence can make reconstruction difficult because similar copies offer several possible placements for a read or alignment. The comparison therefore helps explain the practical challenge of recovering the junctions. The contrasting outgroup results leave the ancestral repeat context and the mechanism of deletion unresolved.

Strong evolutionary conservation of TLR21 in other birds

Our analysis revealed that TLR21 remains under purifying selection across non-parrot birds, highlighting the evolutionary significance of its loss in parrots. In the matched 22-species set, its MG94 dN/dS was 0.201, compared with 0.363 for TLR3, 0.457 for TLR4, 0.518 for TLR5 and 0.306 for TLR7 (Table 1). The estimate remained similar with broader sampling: 0.221 in 37 species and 0.222 in 52. FEL identified purifying selection at 48.0% of TLR21 codons in the core set, rising to 55.0% and 58.8% in larger sets. BUSTED did not detect significant episodic positive selection ($p = 0.38$, 0.094 and 0.12). These comparisons are shown in Fig. 2.

TLR21 combined low dN/dS with substantial amino-acid divergence. Mean pairwise amino-acid identity was about 70%, compared with 81–86% for the other TLRs. TLR21 also had high GC3 (0.84–0.89) and high synonymous divergence. The distinction matters because dN/dS is a relative measure: a protein can accumulate amino-acid changes while those changes remain less frequent than synonymous substitutions. Greater divergence can also increase FEL's power to identify selected sites. These features support purifying selection on retained TLR21 while complicating comparisons of its strength with other receptors. In the 52-species set, FEL identified purifying selection at 58.0% of ectodomain codons and 60.7% of codons in the TIR-containing cytoplasmic region.

Conserved basic residues formed a compact cluster on the predicted TLR21 surface. Eighteen exposed arginine, lysine and histidine residues conserved in at least 90% of the 52 species were concentrated along the middle leucine-rich repeats. Their mean physical separation was 31.7 Å, compared with 38.4 Å for matched exposed basic residues ($p = 0.003$) and 42.4 Å for matched conserved non-basic surface residues ($p = 0.0002$). The comparison remained significant in 24 of 35 threshold settings against the first reference distribution and all 35 against the second. This identifies a conserved candidate DNA-interaction region. The smaller observed

spacing shows that these conserved basic residues occupy a more compact region than either matched comparison set. The structural prediction is shown in Online Resource 1, Fig. S1.

TLR9 was not detected in the 19 bird genomes examined, including 14 parrots. Together with the established absence of avian TLR9, this places the TLR21 finding in the context of losing a known endosomal CpG-DNA receptor.

Selection and copy number in other immune genes

Identical multi-start fitting of targets and 504 control genes did not support a robust relaxation of the DNA-sensing and signalling group. Four of its six genes had RELAX K <1, and their mean rank was at the 28th percentile of controls, but the group comparison gave $p = 0.061$ with all controls and $p = 0.13$ with covariate-matched controls. Excluding MYD88 gave $p = 0.092$ and 0.28. The sensing group also did not differ from the ten partner genes ($p = 0.42$). The selection and copy-number comparisons are shown in Fig. 7.

The fitting check materially changed the interpretation. DDX41 returned K values of approximately 0.80, 0.28 and 0.82 on the same alignment under different starting procedures, and many control genes reached better-likelihood solutions with multiple starts. The revised group result reflects the best-fitting estimates for both targets and controls, showing why consistent optimisation was important for this comparison.

The partner group showed no significant shift against either all controls or matched controls ($p = 0.27$ and 0.26). In the separate expanded analysis, UNC93B1 also showed no detected change in selection intensity ($K = 1.04$, $p = 0.67$).

CD79A retained an open reading frame and the paired tyrosine-containing motifs of its immunoreceptor tyrosine-based activation motif (ITAM) in the interpretable parrot models. Its small selection dataset showed no detected change in selection intensity (RELAX $K = 1.00$, $p = 1.0$). Other assemblies contained partial or disrupted models. No deeply conserved downstream non-coding block was recovered by the cross-species comparisons, leaving regulatory effects unresolved.

The copy survey found no consistent parrot-associated expansion among the surveyed nucleic-acid sensing genes. A tandem TLR2 pair occurred in ten of 14 parrots, providing a separate gene-family observation. Apparent additional loci on short contigs, including cGAS in the lorikeet, require assembly-level verification before being counted as genuine expansions.

The apparent shared ZAP frameshifts were explained by a missed intron. Removing the canonical 101–107-nt intron restored the reading frame, and blood RNA supported splicing in corella and African grey. The apparent coding disruption arose because the alignment had treated an intron as part of the coding sequence. The splice evidence is shown in Online Resource 1, Fig. S6.

Viral genome comparisons reveal differences associated with parrot hosts

Our comparative analysis identified differences in viral DNA composition between viruses associated with parrots and those associated with other birds. The dataset comprised 496 distinct genomes from 39 viral species, eight associated with parrots and 31 with other birds, distributed across five genera in three families. CpG O/E was higher among parrot-associated viruses in four genera: Iltovirus (+35.4%), Siadenovirus (+16.6%), Barthadenovirus (+12.4%) and Circovirus (+4.3%). Aviadenovirus showed the opposite direction (−18.7%; Fig. 8). Equal family weighting produced an overall increase of 13.0%, essentially unchanged after requiring explicit, concordant source-host labels. Genome-choice resampling yielded a 2.5th–97.5th percentile range of +10.0% to +14.6%. Four of five directions remained positive with the nearest protein-based comparators.

Iltovirus herpesviruses showed the most consistent contrast. Their CpG increase persisted with closer matching (+34.6%), in coding (+35.1%) and noncoding sequence (+37.9%), and relative to GpC (+42.5%). Other groups

varied across these checks. For Circovirus, represented by BFDV in the parrot arm, the small broad-panel CpG increase was accompanied by a similar GpC increase, leaving CpG/GpC essentially unchanged (−1.0%). Its closest comparator also varied across bootstrap trees (Online Resource 1, Fig. S8). The overall pattern was most influenced by herpesviruses: omitting that family reduced the summary increase to 3.2%. Detailed effects, host-source decisions and comparator assignments are provided in the analysis tables.

Discussion

TLR21 loss across parrots has distinct structural outcomes

This study moves the evidence for parrot TLR21 loss beyond missing annotation. In Psittacoidea and black cockatoos, individual long reads connect the surviving flanks across deleted sequence. In the kākāpō, reads cross the expected TLR21 interval without a recognisable receptor sequence, supporting probable loss. Together, these observations support loss across all three parrot superfamilies, with physically resolved deletions in Psittacoidea and black cockatoos and probable loss in kākāpō.

The surviving structures reveal different histories of locus remodelling. Psittacoidea share a junction joining CD79A to MEGF8, with PRR19, TLR21 and TMEM145 removed. Black cockatoos have a different junction, and the kākāpō retains sequence on the TMEM145 side of the interval. These differences establish lineage-specific restructuring of the locus. The timing of the initial TLR21 loss can be considered separately from that restructuring. A shared early loss could have preceded the larger lineage-specific deletions that now obscure its original boundaries.

A single early loss is an economical explanation for the distribution of receptor absence. Independent initial losses also remain compatible with the evidence because the later deletions may have removed the original lesion. This makes retained intermediate structures particularly valuable: a lineage that preserves more of the ancestral interval could show which sequences disappeared first. More accurate locus sequences and wider taxon sampling could identify those intermediate states and clarify the order of events. The present evidence establishes the broad distribution of TLR21 loss and the distinct structural outcomes at its locus.

The loss is notable because retained TLR21 remains under purifying selection across a broad sample of other birds. Chicken experiments establish that this receptor can mediate CpG recognition (Keestra et al. 2010). The genomic results identify the removal of a receptor with a known avian immune function.

What parrots retained

TLR21 loss occurred within a retained immune repertoire. Other nucleic-acid sensing genes were recovered, and TLR3 and TLR7 transcripts were detected in every sampled blood or spleen dataset. These observations establish the presence and, for those two receptors in the sampled tissues, expression of other recognition pathways. The loss of TLR21 therefore occurred alongside retention of other components of nucleic-acid recognition.

We found no consistent expansion of the surveyed nucleic-acid sensing genes and no robust shift towards relaxed selection in the DNA-sensing and signalling group after applying identical multi-start fitting to targets and controls. Shared partner genes likewise showed no significant group shift. Because these partners also serve other receptors, their retention is consistent with continued roles in the wider signalling network.

CD79A also retains its coding frame and signalling motif in the interpretable parrot models, despite the adjacent deletion. The small selection dataset detected no change in selection intensity. Together, the coding models and conserved motif support preservation of the neighbouring gene. Whether its expression was affected remains an open regulatory question. The overall pattern is a substantial local change in gene content within a retained immune repertoire.

What the locus and predicted structure reveal

Repeat-rich junctions help explain why this region is difficult to reconstruct. The intact falcon loci showed variable enrichment, and present-day junctions may differ from the sequence present before deletion. The key mechanistic question is therefore what the region looked like when TLR21 was lost. Retained ancestral sequence in additional lineages could help distinguish pre-existing repeat structure from sequence changes accompanying later rearrangement.

The structural analysis identifies features of retained TLR21 that are conserved across birds. Conserved basic residues form a compact region on the predicted ectodomain surface, and the pattern persists across matched comparisons and many threshold settings. This provides a defined candidate region for functional testing. Conservation and spatial clustering could reflect ligand interaction, receptor assembly or another structural requirement. Comparing the effects of changes in this region would help distinguish these possible roles.

The combination of purifying selection and a conserved surface feature gives biological context to the loss. Parrots lack a receptor whose retained avian orthologues preserve both coding constraint and a candidate functional surface. Testing the conserved surface would address how the receptor works, while restoring TLR21 in a suitable parrot cell system would address the cellular consequence of its absence. These are complementary next steps arising from the evolutionary result.

A CpG composition pattern associated with parrot hosts

The viral comparison adds a host-associated composition pattern to the evolutionary evidence for receptor loss. Higher CpG O/E recurred in four of five genera after genome expansion, host-source filtering and closer matching. The clearest example was Iltovirus, where the increase extended across coding and noncoding sequence and exceeded the corresponding GpC contrast. This identifies a specific viral group in which to investigate the evolutionary pressures associated with parrot hosts. The other groups show that the pattern is uneven: aviadenoviruses run in the opposite direction, and the BFDV contrast depends on the comparator and composition measure.

Several evolutionary influences can shape viral nucleotide composition (Upadhyay and Vivekanandan 2015). The present comparison establishes an association with host group; determining the contribution of TLR21 requires separating receptor status from other host traits and viral ancestry. Most parrot-associated viral species in the panel still have only one distinct genome, and comparator hosts were not all independently checked for TLR21. The locus evidence supplies the evolutionary question, while the herpesvirus result provides a focused comparative observation for testing it.

Limitations and next tests

The strongest evidence is the physical connection of surviving flanks on individual long reads. This resolves deletions in the sampled Psittacoidea and black cockatoos. The kākāpō remains a probable loss because its inference rests on two noisy reads and short conserved anchors. In the western ground parrot, accurate reads support junction-related sequence, but local depth does not resolve a complete spanning molecule. Additional accurate reads across these loci would directly address those remaining structural questions.

Sampling also limits ancestral reconstruction. The two read-supported cockatoos are closely related, and the surviving junctions may have lost the sequence needed to distinguish an early shared TLR21 lesion from independent initial losses. Wider sampling can identify retained intermediate structures and refine the branches carrying particular rearrangements. The most informative loci will be those that preserve sequence from before the later rearrangements. In the kea, short-read non-detection extends the Strigopoidea evidence, but accurate spanning reads are needed to compare its locus directly with the kākāpō.

RNA non-detection provides complementary evidence from five blood or spleen samples, with expressed TLR3 and TLR7 as controls. A TLR21-positive non-parrot sample processed identically would provide a direct sensitivity control by showing that the search recovers the transcript when it is present. Sampling additional tissues and conditions would broaden the expression comparison. These additions would strengthen the connection between the genomic result and the expressed immune repertoire.

A functional comparison of a suitable parrot cell system with and without an active avian TLR21 would test whether receptor restoration changes CpG responsiveness. Establishing that the cells retain the necessary signalling capacity would be essential to interpret the outcome. Such work would extend the present study from a documented change in the immune gene repertoire to its cellular consequences.

The evidence distinguishing competing explanations is summarised in Online Resource 1, Table S4.

Statements and Declarations

Funding

This study was entirely self-funded by the author and received no external funding.

Competing interests

The author declares no conflict of interest.

Author contributions

Babalo Reda conceived and designed the study, formulated the research questions, selected the datasets, directed the computational analyses, evaluated and interpreted the genomic, structural and evolutionary findings, and developed the scientific conclusions. The author prepared, critically revised and approved the final manuscript and accepts full responsibility for the accuracy and integrity of the research.

Data and code availability

This study reanalysed publicly available genomic and transcriptomic data. Assembly and sequencing accessions are listed in ACCESSIONS.tsv in the reproducibility package. These include budgerigar PacBio CLR runs SRR25727928, SRR25727929 and SRR25727931; Amazon HiFi run SRR29949499; peregrine PacBio run SRR31755801 (searched subset); ONT runs ERR18017548 and ERR18018952; western ground parrot HiFi run SRR31744274; kea Illumina runs SRR959225, SRR959226, SRR959227, SRR13483825 and SRR8236591; and RNA-seq runs SRR10852853, SRR10852952, SRR13480251, SRR13480262 and SRR30410751. No new animal samples were collected.

The reproducibility package is publicly available in GitHub release v1.0.0 (Reda 2026; https://github.com/gbabaloreda-afk/TLR21_Parrots_Repo/releases/tag/v1.0.0). The release contains TLR21_parrots_repro_v1.1.zip, with scripts, software settings, alignments, trees, locus sequences, supporting read identifiers, model-quality decisions, figures and result tables. The separate CpG archive supplied as Online Resource 2 (ESM_2.zip) includes the sequence snapshot, accession-level host metadata, inclusion and exclusion decisions, species summaries, comparator assignments, protein trees and sensitivity analyses.

The study relies on the contributions of the original sample collectors, sequencing teams and data generators. Their datasets and source studies are credited in the accession manifest and associated references.

Ethics

This study analysed publicly available sequence data and involved no new animal sampling or experiments and no human participants.

Clinical trial number: not applicable.

Supplementary information

Online Resource 1 (ESM_1.pdf): supporting evidence tables, detailed viral CpG methods and supplementary figures S1–S8. Online Resource 2 (ESM_2.zip): CpG analysis data, scripts, accession-level audit, comparator assignments and sensitivity results.

References

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research 27:573–580. https://doi.org/10.1093/nar/27.2.573

Botero-Castro F, Figuet E, Tilak MK, Nabholz B, Galtier N (2017) Avian genomes revisited: hidden genes uncovered and the rates versus traits paradox in birds. Molecular Biology and Evolution 34:3123–3131. https://doi.org/10.1093/molbev/msx236

Buchfink B, Reuter K, Drost HG (2021) Sensitive protein alignments at tree-of-life scale using DIAMOND. Nature Methods 18:366–368. https://doi.org/10.1038/s41592-021-01101-x

Deorowicz S, Debudaj-Grabysz A, Gudyś A (2016) FAMSA: fast and accurate multiple sequence alignment of huge protein families. Scientific Reports 6:33964. https://doi.org/10.1038/srep33964

Duarte MA, Silva JMF, Brito CR, Teixeira DS, Melo FL, Ribeiro BM, Nagata T, Campos FS (2019) Faecal virome analysis of wild animals from Brazil. Viruses 11:803. https://doi.org/10.3390/v11090803

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2018) UFBoot2: improving the ultrafast bootstrap approximation. Molecular Biology and Evolution 35:518–522. https://doi.org/10.1093/molbev/msx281

Jumper J, et al (2021) Highly accurate protein structure prediction with AlphaFold. Nature 596:583–589. https://doi.org/10.1038/s41586-021-03819-2

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30:772–780. https://doi.org/10.1093/molbev/mst010

Keestra AM, de Zoete MR, Bouwman LI, van Putten JPM (2010) Chicken TLR21 is an innate CpG DNA receptor distinct from mammalian TLR9. Journal of Immunology 185:460–467. https://doi.org/10.4049/jimmunol.0901921

Kosakovsky Pond SL, Frost SDW (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. Molecular Biology and Evolution 22:1208–1222. https://doi.org/10.1093/molbev/msi105

Kosakovsky Pond SL, et al (2020) HyPhy 2.5: a customizable platform for evolutionary hypothesis testing using phylogenies. Molecular Biology and Evolution 37:295–299. https://doi.org/10.1093/molbev/msz197

Li H (2018) Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li H (2023) Protein-to-genome alignment with miniprot. Bioinformatics 39:btad014. https://doi.org/10.1093/bioinformatics/btad014

Liu G, Zhang H, Zhao C, Zhang H (2020) Evolutionary history of the Toll-like receptor gene family across vertebrates. Genome Biology and Evolution 12:3615–3634. https://doi.org/10.1093/gbe/evz266

Mihara T, Nishimura Y, Shimizu Y, Nishiyama H, Yoshikawa G, Uehara H, Hingamp P, Goto S, Ogata H (2016) Linking virus genomes with host taxonomy. Viruses 8:66. https://doi.org/10.3390/v8030066

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution 37:1530–1534. https://doi.org/10.1093/molbev/msaa015

Reda B (2026) Data and code for “All three parrot superfamilies have lost TLR21, the immune sensor birds use to detect bacterial and viral DNA”. GitHub release v1.0.0: https://github.com/gbabaloreda-afk/TLR21_Parrots_Repo/releases/tag/v1.0.0.

Silver LW, Farquharson KA, Peel E, Gilbert MTP, Belov K, Morales HE, Hogg CJ (2025) Temporal loss of genome-wide and immunogenetic diversity in a near-extinct parrot. Molecular Ecology 34:e17746. https://doi.org/10.1111/mec.17746

Smith BT, Merwin J, Provost KL, Thom G, Brumfield RT, Ferreira M, Mauck WM III, Moyle RG, Wright TF, Joseph L (2023) Phylogenomic analysis of the parrots of the world distinguishes artifactual from biological sources of gene tree discordance. Systematic Biology 72:228–241. https://doi.org/10.1093/sysbio/syac055

Upadhyay M, Vivekanandan P (2015) Depletion of CpG dinucleotides in papillomaviruses and polyomaviruses: a role for divergent evolutionary pressures. PLOS ONE 10:e0142368. https://doi.org/10.1371/journal.pone.0142368

Velová H, Gutowska-Ding MW, Burt DW, Vinkler M (2018) Toll-like receptor evolution in birds: gene duplication, pseudogenization, and diversifying selection. Molecular Biology and Evolution 35:2170–2184. https://doi.org/10.1093/molbev/msy119

Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K (2015) RELAX: detecting relaxed selection in a phylogenetic framework. Molecular Biology and Evolution 32:820–832. https://doi.org/10.1093/molbev/msu400

Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Molecular Biology and Evolution 17:32–43. https://doi.org/10.1093/oxfordjournals.molbev.a026236

Table 1 Selection and sequence composition in the matched core set

Gene	Codons	MG94 dN/dS	FEL purifying	Mean aa identity	GC3
TLR21	968	0.201	48.0%	0.703	0.885
TLR3	895	0.363	26.3%	0.860	0.384
TLR4	843	0.457	26.3%	0.814	0.573
TLR5	861	0.518	24.3%	0.823	0.397
TLR7	1047	0.306	32.8%	0.860	0.449

The matched core set contains 22 non-parrot species. FEL purifying sites use the nominal site-level threshold $p < 0.1$. Amino-acid identity is a proportion; GC3 is the proportion of G and C at third codon positions. Differences in synonymous divergence, nucleotide composition and statistical power limit interpretation of the FEL percentages as direct measures of relative functional constraint.

Table 2 Evidence for TLR21 status by species and dataset

Species / group	Assembly evidence	Read or transcript evidence	Classification
Budgerigar Psittacoidea	CD79A–MEGF8 junction in both haplotypes; markers B and C.	13 spanning CLR reads.	Structurally resolved deletion
Yellow-crowned Amazon Psittacoidea	Compact CD79A–MEGF8 junction; B and C.	One spanning HiFi read; reported 99.3% identity.	Structurally resolved deletion
Baudin's black cockatoo Cacatuoidea	Locus reconstructed from reads.	Seven spanning ONT reads; TLR21 and TMEM145 absent from the crossed interval; no marker B. PRR19-like ORF has unresolved origin.	Structurally resolved deletion
Carnaby's black cockatoo Cacatuoidea	Locus reconstructed from reads.	Eleven spanning ONT reads in the complete 81.8-Gb run; cockatoo-type junction without marker B.	Structurally resolved deletion
Monk parakeet Psittacoidea	CD79A–MEGF8 junction on one contig; B and C.	No spanning read reported. Two spleen samples: TLR21 not detected; TLR3 and TLR7 detected.	Probable absence
Ring-necked parakeet Psittacoidea	Shared junction sequence; MEGF8 subsequently rearranged approximately 0.3 Mb away.	Partial ONT run searched; no resolved spanning molecule reported.	Probable absence
Western ground parrot Psittacoidea	Locus incompletely assembled.	29.3 Gb HiFi; one CD79A and eight MEGF8 reads. Junction-related sequence, but no resolved full-span read.	Probable absence
Kākāpō Strigopoidea	CD79A and MEGF8 absent from the assembly but recovered in reads.	Two CLR reads cross approximately 9–11 kb to TMEM145-side sequence without recognisable TLR21.	Probable absence, read-supported
Kea Strigopoidea	TLR21 not detected; locus incompletely assembled.	534 million Illumina reads (53.4 Gb): TLR21 not detected; TLR3 (573) and TLR7 (1,169) recovered. Locus structure unresolved.	Not detected in raw short reads
African grey, Rüppell's parrot, Bourke's parrot and greater vasa parrot Psittacoidea	Some neighbours assembled; locus not recovered on one contig.	No spanning-read evidence reported. African grey blood: TLR21 not detected; TLR3 and TLR7 detected.	Assembly-only absence
Further Psittacoidea assemblies	TLR21 not detected; locus incompletely assembled.	No spanning-read evidence reported. Orange-bellied parrot spleen: TLR21 not detected; TLR3 and TLR7 detected.	Assembly-only absence
Palm cockatoo, galah, little corella, sulphur-crested cockatoo and cockatiel Cacatuoidea	TLR21 not detected; locus incompletely assembled.	No spanning-read evidence reported for these assemblies. Separate Ducorps's corella blood sample: TLR21 not detected; TLR3 and TLR7 detected.	Assembly-only absence

Species / group	Assembly evidence	Read or transcript evidence	Classification
Chicken, zebra finch, great tit and lesser kestrel Non-parrot references	Intact TLR21 recovered.	RNA not tested	Present
Peregrine falcon Non-parrot reference	Gap in older reference; intact locus in the 2025 long-read assembly.	PacBio long reads (SRR31755801; searched subset) recovered two TLR21 reads and 29 TLR3 reads.	Present

Classification reflects the strongest available locus evidence. Further Psittacoidea assemblies comprise lovebird, hanging parrot, macaw, golden parakeet, swift parrot, Pesquet's parrot, echo parakeet, crimson rosella, rainbow lorikeet, orange-bellied parrot, green-rumped parrotlet, sun parakeet, El Oro parakeet, red-masked parakeet, burrowing parrot, masked shining parrot, eclectus and Finsch's parakeet.

Figure legends

Fig. 1 TLR21 status across birds and the structure of the TLR21 locus in parrots

(A) Cladogram of the sampled lineages; branch lengths are not to scale. Blue dots indicate TLR21 presence and red crosses indicate non-detection. Filled red bars mark observed deletion structures: the shared Psittacoidea deletion of PRR19, TLR21 and TMEM145, and the TLR21–TMEM145 deletion supported by seven and eleven spanning Oxford Nanopore reads in Baudin's and Carnaby's black cockatoos, respectively. The hatched box marks probable, read-supported loss in kākāpō: two PacBio reads cross the CD79A–TMEM145-side interval without detectable TLR21. The dashed box denotes uncertainty about the number and position of initial TLR21 losses; one parrot-stem loss and separate lineage losses remain possible (Figure 6). Kea has no detectable TLR21 in raw short reads, but its locus structure is unresolved. Other assembly-only non-detections are not scored as deletions.

(B) Loci aligned on the start of CD79A and oriented from ARHGEF1 towards MEGF8. Peregrine falcon, black cockatoo, budgerigar, yellow-crowned Amazon and monk parakeet rows use measured coordinates. The black-cockatoo row shows Baudin's Nanopore read consensus; Carnaby's reads support the same layout. Kākāpō and ring-necked parakeet rows are schematic; in the latter, MEGF8 lies approximately 0.3 Mb away following rearrangement. Markers B and C are shared Psittacoidea junction markers; C also lies beside MEGF8 in falcons. Hatching denotes incomplete gene identity for the kākāpō TMEM145-side sequence and unresolved orthology for the cockatoo PRR19-like ORF. Arrows indicate gene orientation. Scale bar: 5 kb.

Fig. 2 TLR21 is a conserved, single-copy gene in other birds, with no endosomal backup

(A) dN/dS (MG94) for five bird TLRs across 22 non-parrot species.

(B) The share of codons under purifying selection (FEL).

(C) Presence of TLR9, TLR21, TLR7 and TLR3 across 19 genomes. TLR9 is not detected in any bird and TLR21 is not detected in any parrot. The peregrine RefSeq assembly has an assembly gap at TLR21; the gene is present in the 2025 long-read assembly.

Fig. 3 The absence of TLR21 in parrots is not an assembly artefact

(A) Locus matrix for 34 parrot assemblies, showing neighbour genes, junction markers, TLR21 and the evidence class. Evidence classes never call a deletion from assembly absence alone.

(B) Raw long-read hit counts for TLR21 against control genes in the kākāpō, Baudin's black cockatoo and western ground parrot. Hatched bars are TIR-domain matches reassigned to TIRAP.

(C) RNA read counts in five parrot samples: TLR3 and TLR7 detected, TLR21 not detected.

Fig. 4 Zooming in on the deletion

(A) The ancestral falcon locus (ARHGEF1–CD79A–PRR19–TLR21–TMEM145–segment C–MEGF8), drawn to scale above the budgerigar assembly, the Baudin's black cockatoo read consensus and a kākāpō read. The ~15 kb absent from Psittacoidea (PRR19, TLR21, TMEM145) is highlighted.

(B) The 3.1-kb budgerigar junction, which is 78% tandem repeat: CD79A exon 5, marker B, a tandem-repeat array, marker C and MEGF8 exon 1. The falcon interval (17.8 kb) is shown on a broken axis, and marker B has no detectable falcon counterpart.

(C) Base-level alignments to the peregrine at the CD79A side and MEGF8 side of the budgerigar junction, and at the MEGF8 side of the cockatoo junction. Breakpoints are approximate alignment ends, because the flanks have diverged for tens of millions of years.

Fig. 5 Long reads cross the TLR21 locus in four parrots. Read alignments for the budgerigar (13 spanning CLR reads on both haplotypes), Amazon (1 spanning HiFi read), Baudin's black cockatoo (7 spanning ONT reads) and kākāpō (2 CLR reads crossing the CD79A–TMEM145-side interval). No read contains TLR21

Fig. 6 Three structurally different outcomes and the histories consistent with them

(A) Gene and anchor retention across the three parrot superfamilies against the ancestral falcon order. Kākāpō TMEM145-side sequence is retained without a demonstrated complete gene.

(B) Three loss histories (H1–H3) on the parrot tree, and the contradicted single shared deletion. These are qualitative alternatives, not fitted ancestral-state models.

Fig. 7 No detected compensation or side effects around the loss

(A) RELAX K of DNA-sensing and partner genes against 504 calibrated control genes (best of three fits; bars show K within 2 log-likelihood units). The analysis is exploratory and shows no clear shift (sensors $p = 0.06$; sensors vs partners $p = 0.42$).

(B) Loci per genome for 26 nucleic-acid sensors across 19 genomes. Dots mark loci without an intact model, and the TLR2 tandem pair is present in 10 of 14 parrots.

(C) CD79A is intact, with both ITAM motifs conserved.

Fig. 8 CpG composition in parrot-associated and other avian viruses. (A) Percentage differences in CpG observed/expected after summarising distinct genomes within species. Circles compare the two host-group medians across all eligible species within each genus. Diamonds use the nearest available non-parrot reference identified from a conserved protein; multiple parrot comparisons are averaged on the log scale within genus. (B) Broad-panel comparisons for CpG, GpC, TpA, GC fraction and CpG/GpC. The panel contains 496 distinct genomes from 39 viral species in five genera and three families. Host groups were assigned from metadata and audited source records. Values describe composition differences; the groups are not independent TLR21-loss events

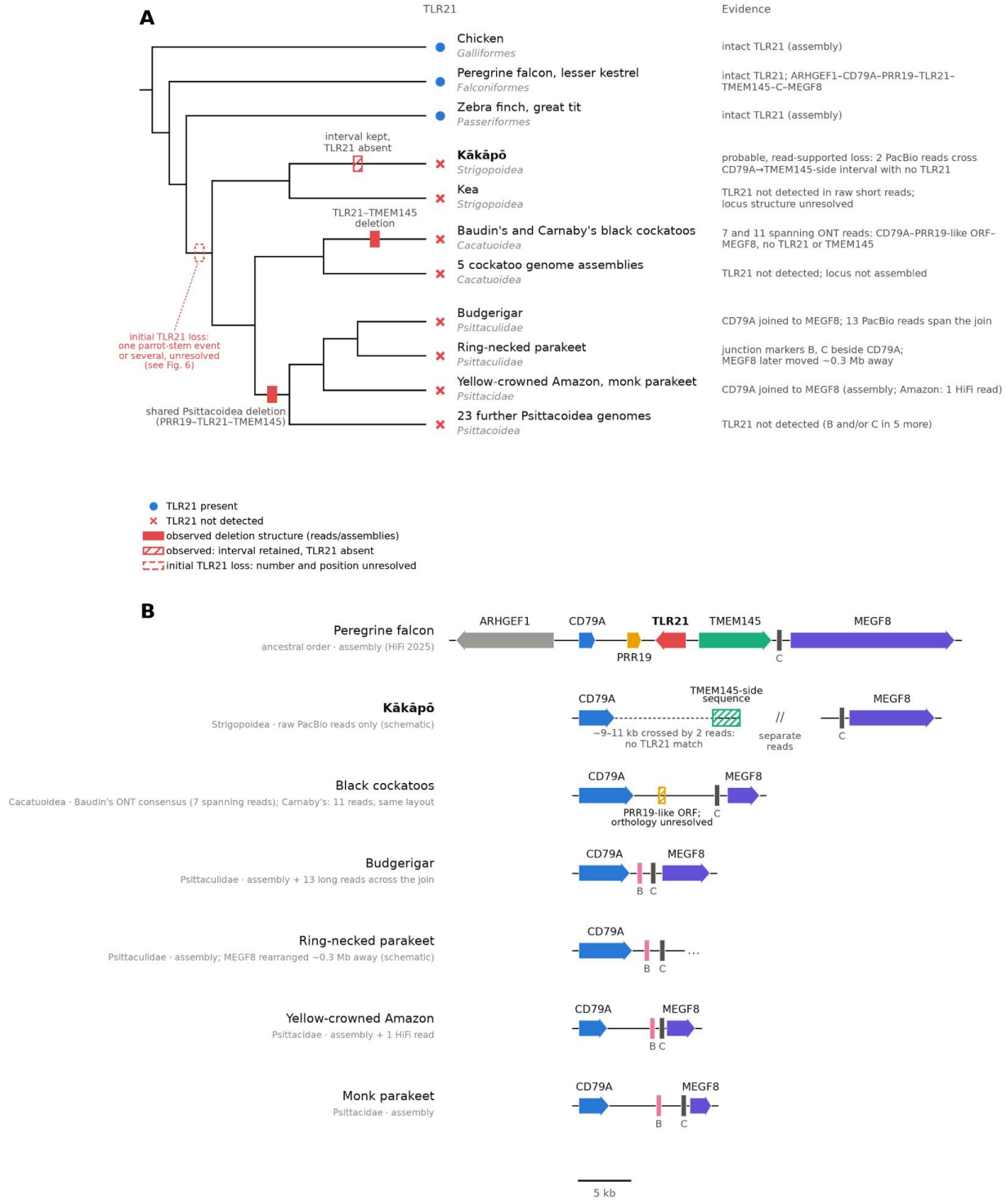


Figure 1. TLR21 status across birds and the TLR21 locus in parrots.

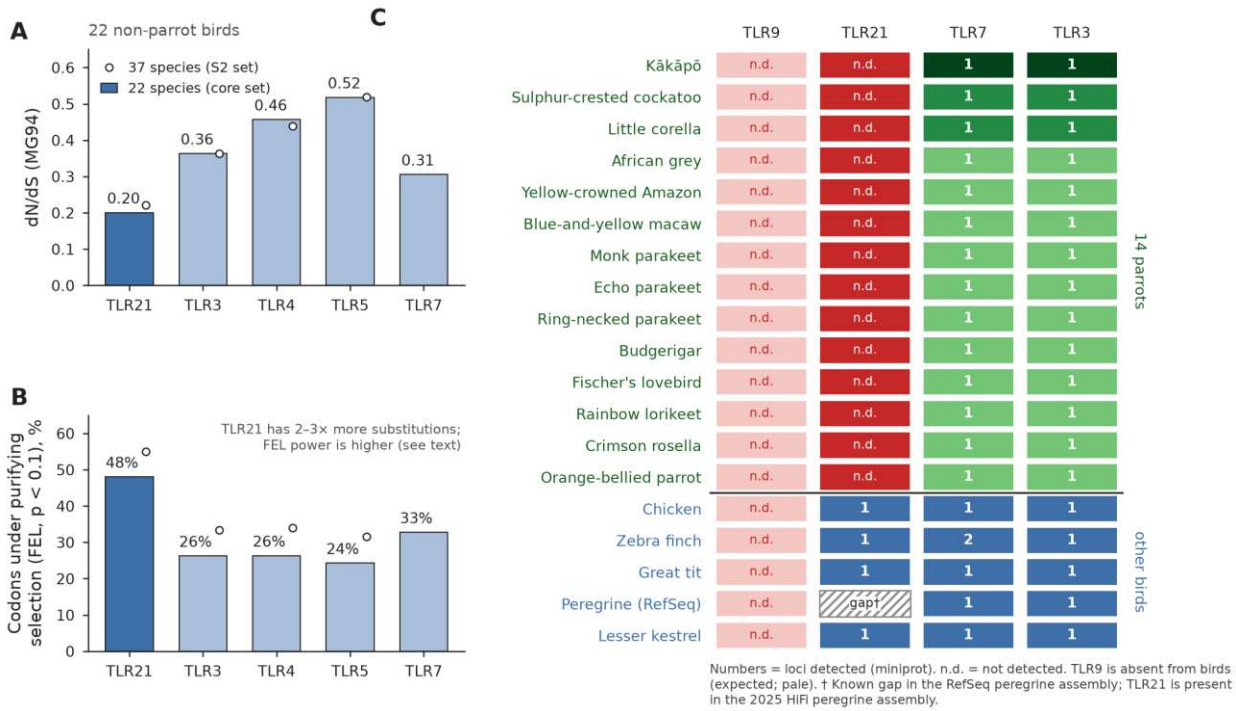


Figure 2. TLR21 is a conserved, single-copy gene in other birds, with no endosomal backup.



Figure 3. The absence of TLR21 in parrots is not an assembly artefact.

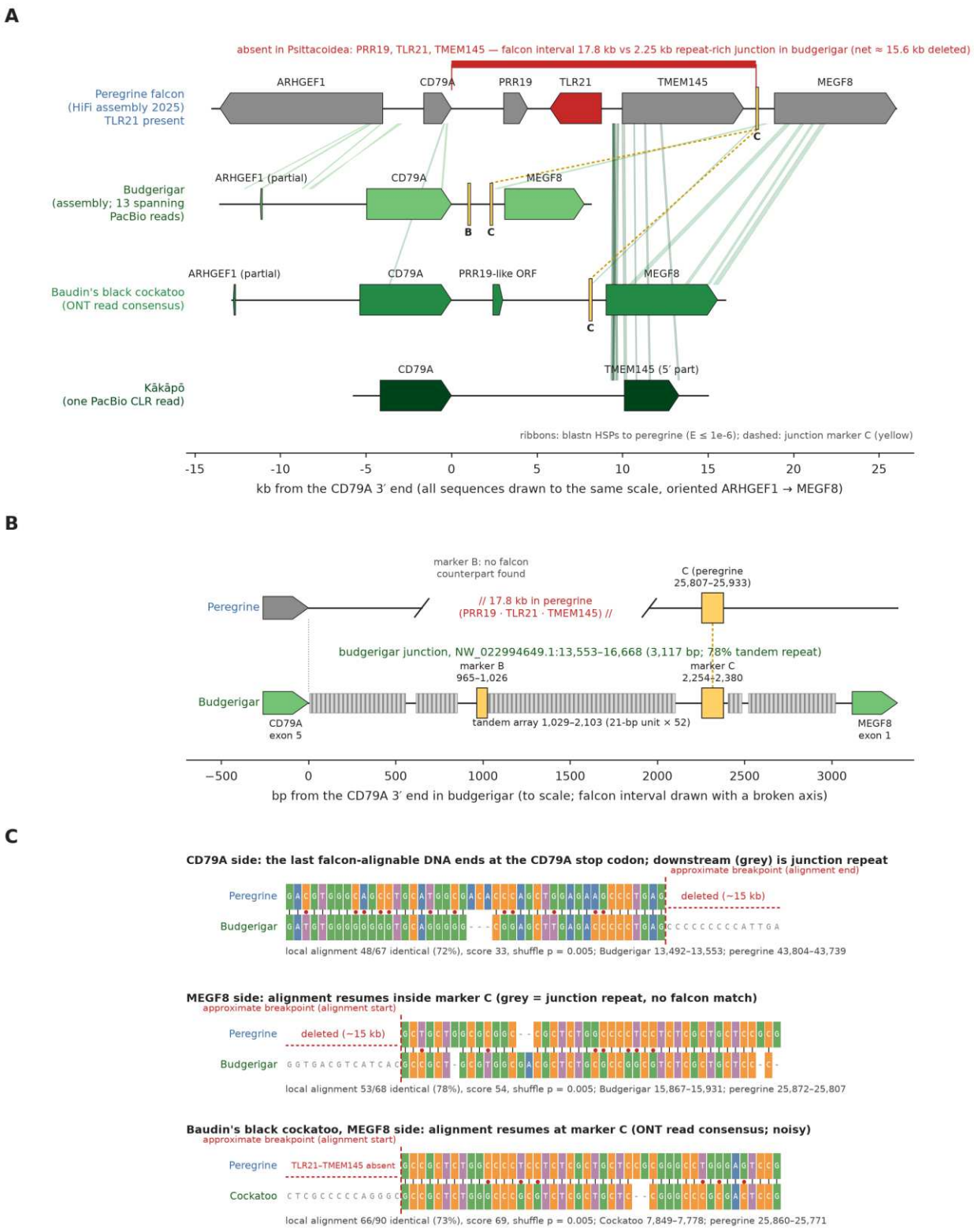
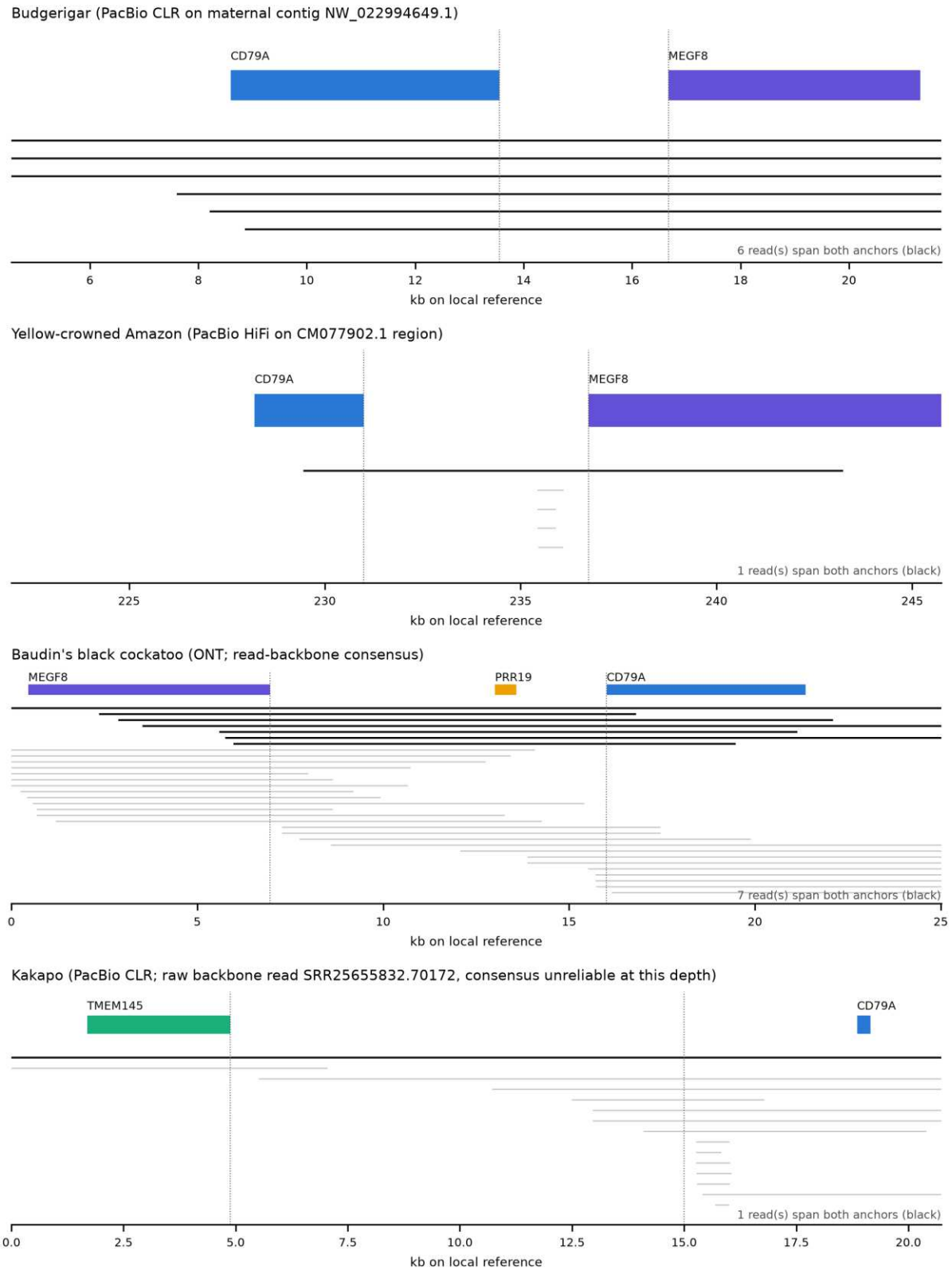


Figure 4. Zooming in on the deletion, from locus to single bases.



817

818 Figure 5. Long reads cross the TLR21 locus in four parrots.

819

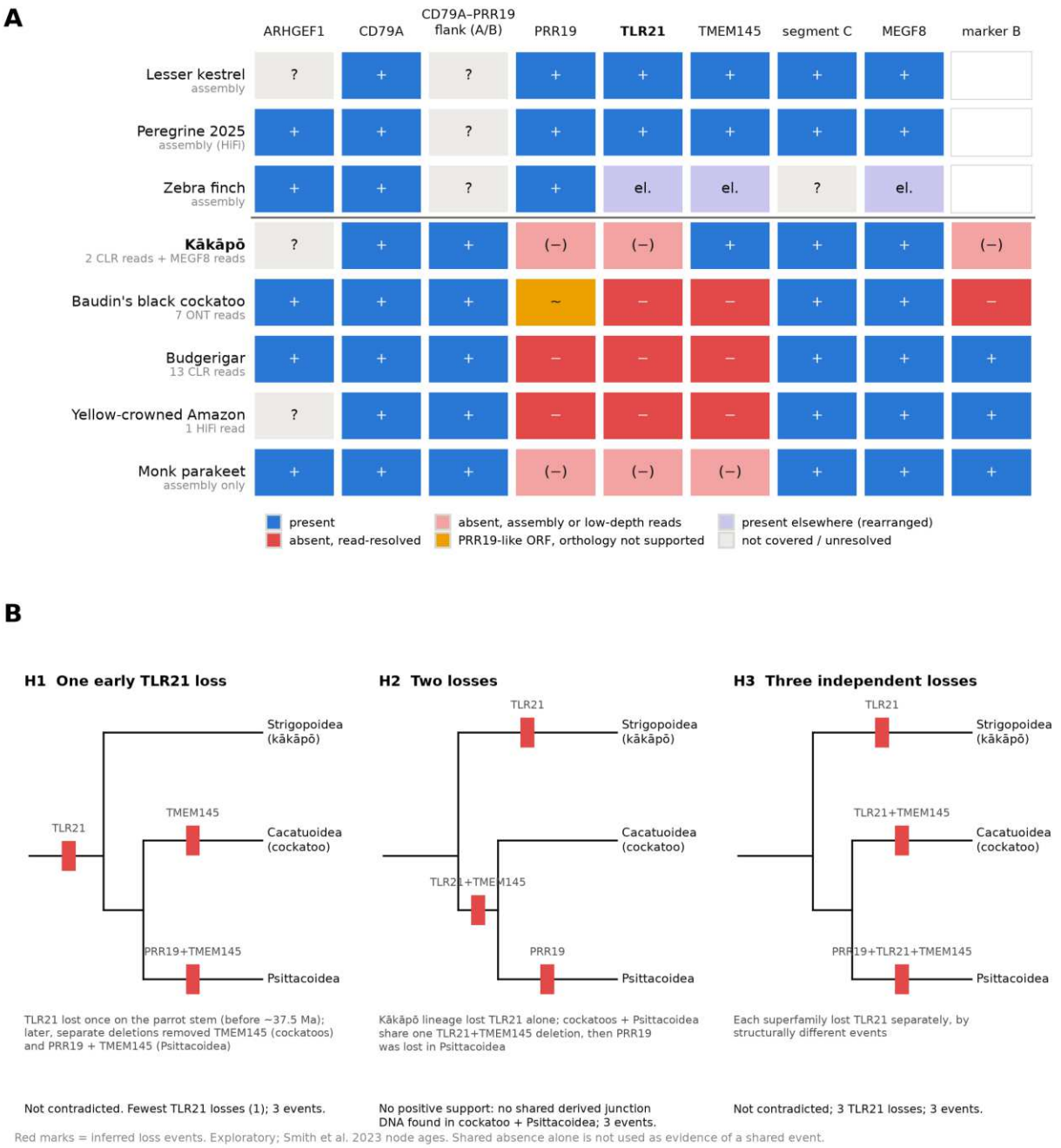


Figure 6. Three structurally different outcomes and the histories consistent with them.

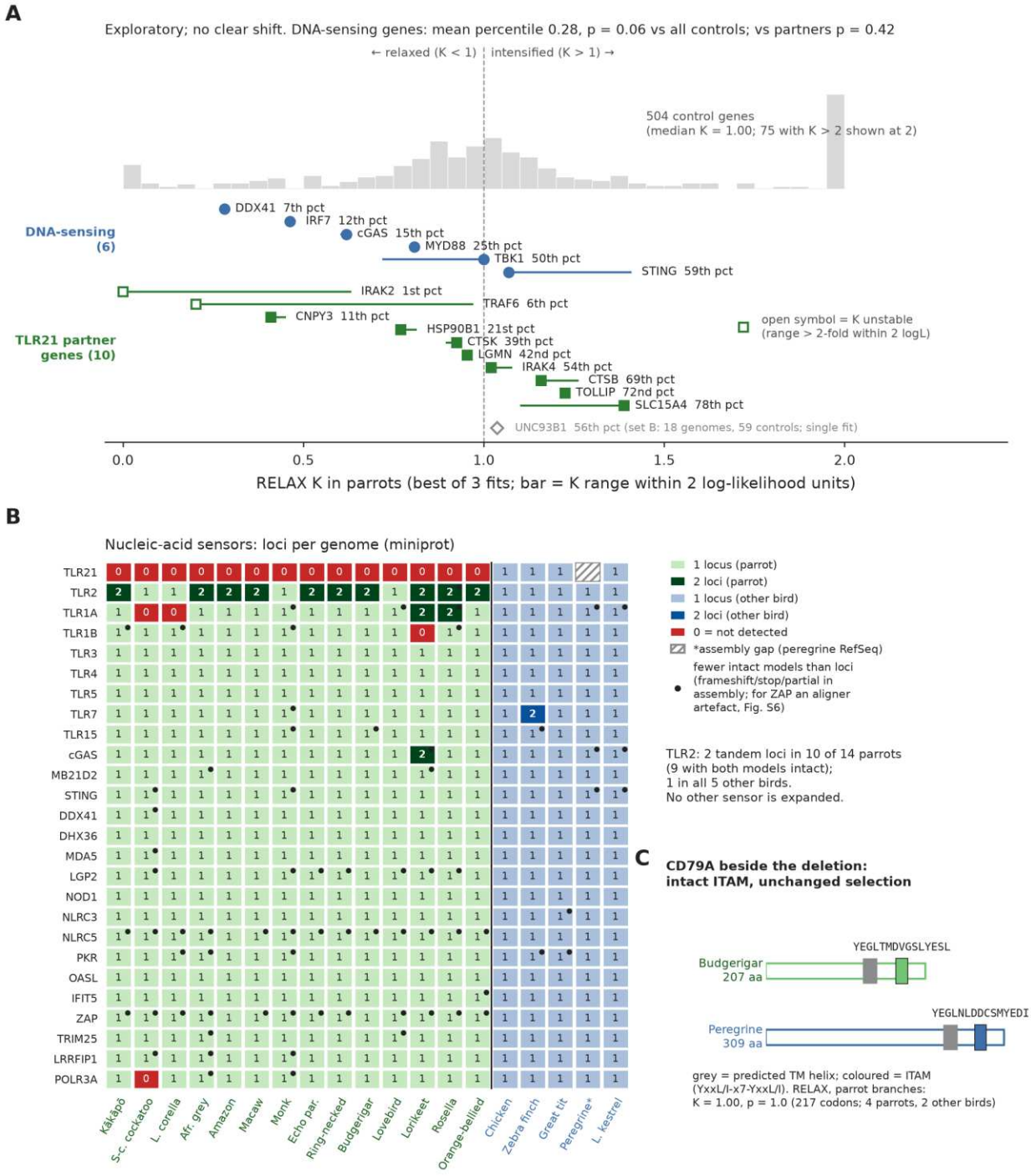
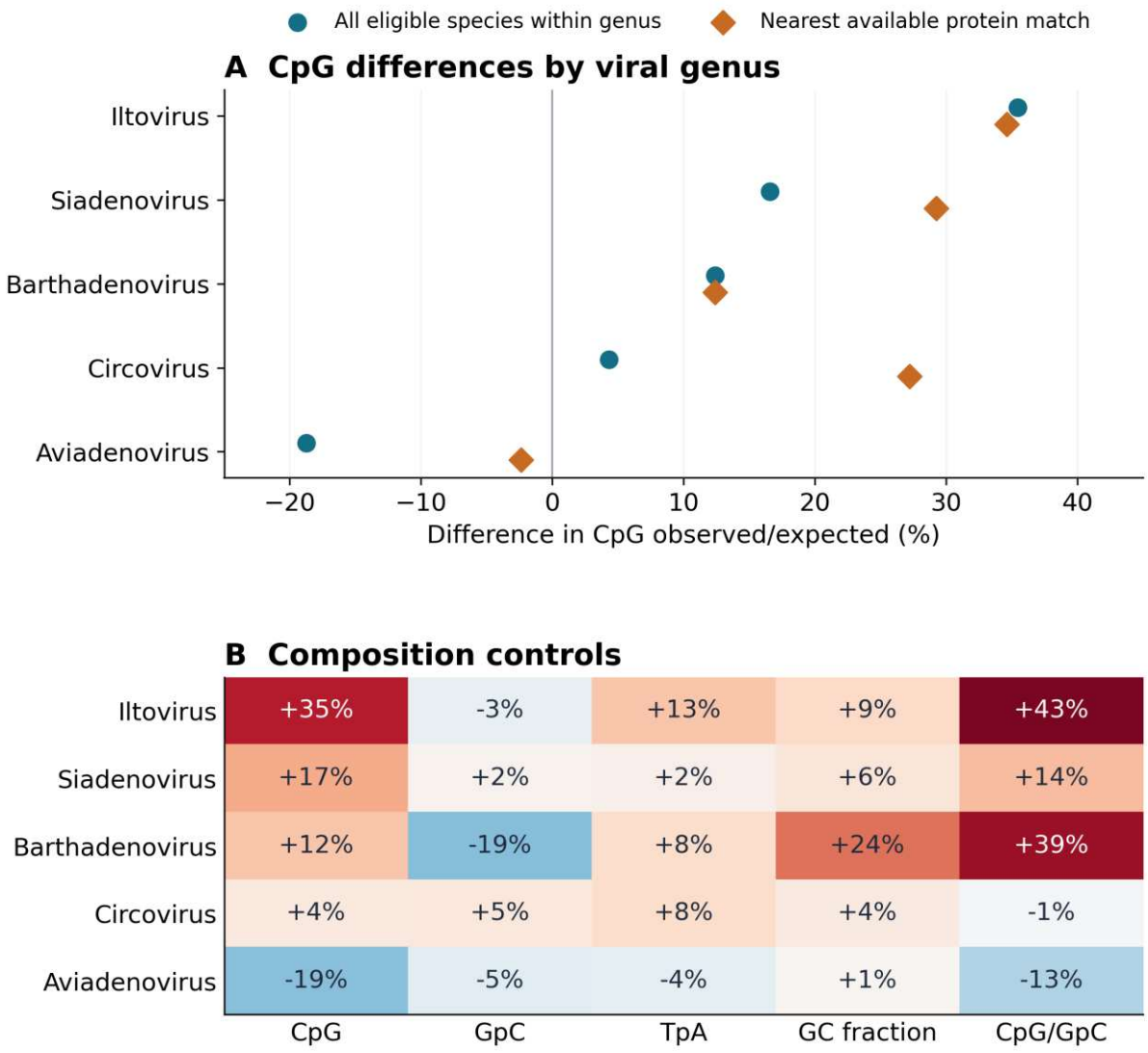


Figure 7. No detected compensation or side effects around the loss.



826

827 Figure 8. CpG composition after genome expansion and host audit.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigS11.png
- FigS21.png
- FigS31.png
- FigS41.png
- FigS51.png
- FigS61.png
- FigS71.png
- FigS81.png