iScience

Article

Long read genome unravels MHC I genomic architecture, evolution, and diversity loss in *Gubernatrix cristata*

Highlights

- First high-quality long-read genome for the yellow cardinal
- Genome-wide data shows long-term decline
- Single copy and tandemly duplicated MHC-I genes
- Non-locus specific primers underestimate variation

Authors

Marisol Domínguez, Enrique Celemín, Nikolai Gusev, Binia De Cahsan, Katja Havenstein, Bettina Mahler, Ralph Tiedemann

Correspondence

dominguez@uni-potsdam.de

In brief

Evolutionary biology; Genomics; Ornithology

Domínguez et al., 2025, iScience *28*, 112301
May 16, 2025
https://doi.org/10.1016/j.isci.2025.112301

CellPress

iScience

CellPress
OPEN ACCESS

Article

Long read genome unravels MHC I genomic architecture, evolution, and diversity loss in *Gubernatrix cristata*

Marisol Domínguez,[1,4,*] Enrique Celemín,[1] Nikolai Gusev,[1] Binia De Cahsan,[2] Katja Havenstein,[1] Bettina Mahler,[3] and Ralph Tiedemann[1]

[1]Unit of Evolutionary Biology/Systematic Zoology, Institute for Biochemistry and Biology, University of Potsdam, Potsdam, Germany
[2]Globe Institute, University of Copenhagen, Copenhagen, Denmark
[3]IEGEBA, FCEN-UBA, Pabellón II, Ciudad Universitaria, Ciudad Autónoma de Buenos Aires 1428, Argentina
[4]Lead contact
*Correspondence: dominguez@uni-potsdam.de
https://doi.org/10.1016/j.isci.2025.112301

SUMMARY

The major histocompatibility complex (MHC) genes are vital for the adaptive immune response in vertebrates and are widely used in conservation genetics to represent adaptive variation. Accurate genotyping of MHC alleles is essential for effective conservation, particularly for endangered species like the yellow cardinal (*Gubernatrix cristata*). However, the absence of locus-specific primers and the highly repetitive nature of these genes present a technical limitation when using short-read sequencing technologies. We produced the first high-quality long-read reference genome for the yellow cardinal. This genome reveals sustained high genome-wide heterozygosity despite inbreeding, with homozygosity patterns and effective population size estimates indicating a long-term decline. We identified seven genomic MHC-I loci, while amplicon sequencing with non-locus specific primers had not confirmed more than two MHC-I loci. Our study also revealed mismatches in primer binding sites across multiple loci, emphasizing the need for high-quality, long-read genomic data to understand the genomic architecture of MCH and to accurately assess locus specific MHC variation.

INTRODUCTION

The major histocompatibility complex (MHC) genes have a central role in the immune defense of jawed vertebrates. They encode transmembrane proteins that bind and present peptides of either self or non-self origin, allowing T cell recognition and adaptive immune responses.[1] The protein binding region (PBR) of MHC molecules, essential for antigen presentation, exhibits high polymorphism enabling the recognition of diverse peptides. Hosts rely on MHC diversity to identify various pathogen proteins, while pathogens evolve to evade host defenses. This dynamic leads to an evolutionary arms race between hosts and their pathogens, which pushes MHC genes to be among the most polymorphic genes known in vertebrates. The notable polymorphism observed was attributed to the interplay of multiple forms of balancing selection, including heterozygote advantage, frequency-dependent selection, and fluctuating selection pressures.[2] From the host's point of view, reduced MHC diversity can jeopardize survival as it compromises the immune response.[3] Beyond their immune function, MHC genes may play a role in kin recognition and mate choice.[4–6] However, there is mixed evidence for these roles in natural populations.[7,8] Two main classes of MHC exist: MHC class I (MHC-I) molecules are monomeric proteins which present intracellular peptides (e.g., derived from pathogens from the cytoplasm of infected cells), while MHC class II (MHC-II) molecules are dimeric proteins that present extracellular peptides, hence dealing with extracellular infections.[9,10] MHC-I proteins are ubiquitously found in the membrane of all nucleated cells, including avian blood cells, where they play a crucial role in immune defense against various pathogens, including viruses and haemosporidian parasites.

Studies approach MHC polymorphism from two complementary directions: the number of alleles per locus at population level[11] or the number of alleles per individual.[12–14] From the latter, often, the minimal number of genes is derived. Extensive research of avian MHC-I diversity showed their allelic diversity to vary from low in falcons[15] to extremely high in the Scarlet Rosefinch (*Carpodacus erythrinus*).[16] However, accurately determining the exact number of MHC loci has been historically more challenging, mainly due to technological constraints,[17] since paralogs often have high sequence similarities and may vary in number (copy number variation). Like allelic diversity, the minimal number of genes required to explain the observed allele counts yielded vastly differential gene counts ranging between 1 and 7 MHC-I genes for Galliformes, Anserifomes, Ciconiiformes and Pelecaniformes,[18–21] and more than 30 MHC-I genes in Passeriformes (*Acrocephalus schoenobaenus*).[22]

Table 1. Reference genome assembly quality statistics

a) Assembly characteristics		
Sequencing technology		PacBio HiFi
Estimated haploid genome size		1.1 Gb
Number of contigs		519
Contig N50		26.5 Mb
Complete Aves BUSCO (%)		97.2%
b) Assembly Quality and Completeness		
Assembly	QV score	Kmer completeness
Primary	62.5	83.3%
Alternative	62.8	73.8%
Both	62.7	99.4%

Inaccurate MHC-I genotyping can lead to misinterpretation of the true MHC variability, which becomes especially important for endangered species as it could lead to inadequate conservation strategies. Long-read genomic approaches provide a way of accurately determining the number of MHC paralogs and overcome some of the limitations associated with traditional methods applying non-locus specific, often heterologous primers (e.g., allelic dropout), offering a more precise understanding of the genetic diversity and evolutionary dynamics of these crucial immune system genes.[23] Recent research leveraging genomic data has started to shed new light on MHC variability and its implications.[13,24–26]

In this study, we aim to overcome previous limitations in MHC characterization, and analyze MHC-I polymorphism in the yellow cardinal (*Gubernatrix cristata*), an endangered bird from the Thraupidae family endemic to South America with a highly discontinuous current distribution. Such fragmentation suggests that an interplay between differential selection pressures and genetic drift could have produced distinct adaptations to local environments that are important for the species' evolutionary potential. In light of this, a good assessment of genetic variation at MHC-I loci can bring insights into the species adaptive potential and inform conservation strategies. Accordingly, our aims were to: (1) provide the first whole genome assembly for this species, (2) infer genome-wide patterns of variation and demographic history that may have impacted the species' genetic makeup, (3) compare two methodological approaches (amplicon sequencing and PacBio HiFi whole genome sequencing) to investigate the genomic organization and polymorphism of MHC-I gene(s) in the yellow cardinal, and (4) assess diversity and signs of selection at MHC-I gene(s).

RESULTS

Yellow cardinal reference genome

We used our newly generated *de-novo* genome assembly to describe genome-wide diversity, infer ancestral demographic history and determine the number of MHC-I loci. We obtained ~35x mean genome coverage by sequencing the DNA of a female individual from management unit 1 (MU1)[27,28] using high-fidelity PacBio long reads. Contaminated contigs ptg000586L_1 and ptg000793L_1 which blasted to bacteria were excluded from further analysis. The final assembly (Gcri1.0) consisted of 519 contigs and a contig N50 of 26.5 Mb (Table 1). It had 96.3% complete single-copy bird orthologues, and 0.9% duplicated BUSCOs. ~85% of the genome was contained in 45 contigs > 7Mb, approaching the species' karyotype ($2n = 78$[29] (Table 1)).

Synteny alignment successfully placed 351 out of the 519 contigs assembled in the draft genome, resulting in 98.82% of the genome properly placed on 41 Zebra finch pseudochromosomes. The 168 contigs that did not map to any of the chromosomes had an average length of just 82,201 bp (~5,3 HiFi reads per contig) which may have hindered the ability to find their chromosomal location. 25.11% of the pseudoautosomes consisted of repeated elements (Table S4).

Heterozygosity, runs of homozygosity and demographic history of yellow cardinals

We identified 6,064,423 filtered variants (5,420,823 SNPs, 643,600 INDELs) in the yellow cardinal genome and estimated autosomal heterozygosity to be 0.0076. When studying heterozygosity on 1Mb windows across the genome, we observed intermediate variation across pseudochromosomes, with relatively uniform ups and downs distributed along the genome (Figure 1). The average 1Mb window had 7.4 SNPs per Kb with the maximum being 16.7 and the minimum 0 SNPs.

We identified 276 runs of homozygosity (ROHs) longer than 100,000 bp (Table S5), with a mean ROH size of 302,500 bp and covering ~83 Mb of the genome (FROH = 8). Pseudochromosomes exhibited from 0 to 50% of their content located in ROHs (Table S6). Yellow cardinal's effective population size (Ne) was inferred to have peaked at ~1.2 million individuals 800 kya, when the species started to experience a mild reduction in Ne. The Ne was inferred to have experienced a drastic reduction 100 kya, when the last glacial period started, from ~1 million individuals to less than 30,000 individuals, compatible with the Pleistocene glaciations having had a profound impact on the yellow cardinal's evolutionary history (Figure 1). Nevertheless, these results should be interpreted with caution, since PSMC is known to lose resolution at more recent timescales[30,31] and uncertainty in mutation rates and generation times could yield biased estimates.

MHC class I polymorphism estimated by amplicon sequencing

We identified 9 alleles at nucleotide level (hereafter, "nucleotide alleles") spanning the entire MHC-I exon 3 of 56 individuals (Table S7). Amplicon alleles were 276 bp long, except allele MHC-Gucr*05 which carried a three nucleotides deletion retaining the correct reading frame. This allele occurred in 9% of the yellow cardinals and never occurred in homozygotic state (Tables S7 and S9). It was previously detected in other species,[32] which suggests that either the deletion emerged independently in different lineages (convergent evolution), or it persisted from a singular ancient event across several species (Figure S2). The nine alleles identified in cardinals were supported by a variable number of reads, but there was no correlation between the number of alleles per individual and sample coverage (Pearson r = 0.13, n.s., Figure S3). Detailed information about alleles detected per individual can be found in Table S8.

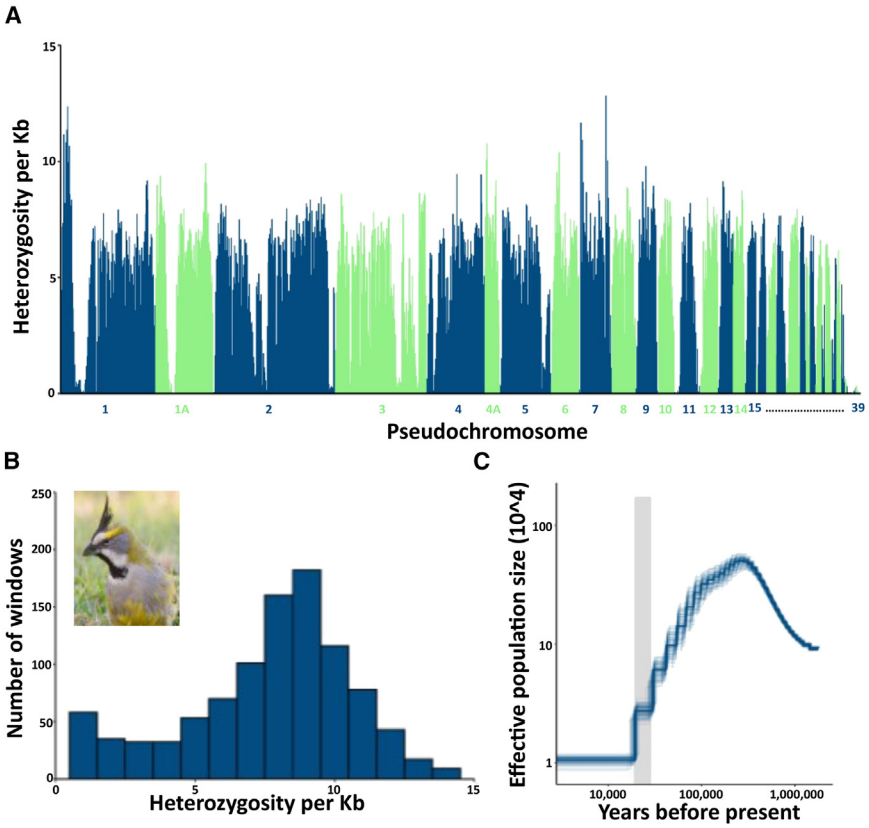

Figure 1. Genome-wide heterozygosity and demographic history of yellow cardinals

(A) Bar plot showing the distribution of heterozygosity across yellow cardinal pseudochromosomes in nonoverlapping 1-Mb windows. Pseudochromosomes are shown in alternating colors. (B) Histogram of the count of per-window heterozygosity levels. Credit photo: Carlos Figuerero. (C) Changes in effective population size of the yellow cardinal over time inferred from PSMC. Blurred trajectories represent 100 bootstraps, and the grey rectangle outlines the last glacial maximum (LGM) from ~19,000 to 29,000 years before present.

The alleles described here have 79 polymorphic sites and differ by 1–62 nucleotides (Table S7). Overall, the mean nucleotide diversity was Pi = 0.06 and the average number of nucleotide differences among alleles was k = 16.84. In MU1, only 3 alleles were identified, compared to all 9 alleles being found in both MU2 and MU3 (Table S1). No private alleles were found (Figures 2B and 2C). Despite the existence of homologous sequences in other species, the BLAST search evidenced that our sequences correspond to novel and previously undescribed alleles, which we thus named according to Klein et al. (1990)[33] and deposited into GenBank public domain.

The maximum likelihood phylogenetic reconstruction of yellow cardinal MHC-I amplicon alleles and from 21 other Passeriformes species with long-read genome assemblies (Table S3) provided evidence suggestive of transspecies polymorphism. Alleles did not cluster by species and yellow cardinal's alleles formed two paraphyletic clusters (Figure S2), which form monophyletic clades with alleles from other species.

After translation, we found 8 unique alleles at amino acid level (hereafter, "amino acid alleles") composed of 91 or 92 amino acids which differed by 1–32 amino acids (nucleotide alleles MHC-Gucr*02 and MHC-Gucr*07 translate into the same amino acid allele, Table S9). As we lack transcriptomic information, we are unable to confirm that alleles are effectively translated and functional; however, there is no evidence of premature stop codons, suggesting the absence of pseudogenes in our dataset. Sequence features typical of antigen-presenting MHC-I genes were conserved. As expected, highly conserved cysteines at positions 11 and 74 important for disulphide bridge formation and peptide binding were present in all alleles.[35]

A maximum of 4 nucleotide (or 3 amino acid) alleles were found per individual, suggesting the co-amplification of at least two MHC-I loci for the yellow cardinal assuming prevalent heterozygosity (Figure 2A). Overall, the low total number of amplicon MHC-I alleles ($N = 9$), the low number of suggested MHC-I loci ($N = 2$), and the many individuals with only one MHC-I allele identified ($N = 21$, only possible if different loci share the same allele, Figure 2A) suggest a case of extremely low MHC-I polymorphism. To exclude that the limited MHC-I polymorphism observed in the yellow cardinal is a technical artifact induced by the use of non-locus and non-species specific "universal" primers, we inferred the MHC-I genomic organization in the assembled reference genome.

MHC class I polymorphism in the high-fidelity genome assembly

To resolve if the limited MHC-I polymorphism observed in the yellow cardinal is caused by technical artifacts, as potentially introduced by the use of non-locus and non-species specific "universal" primers, we inferred the MHC-I genomic organization in the assembled reference genome.

The genome scan suggested the presence of 7 MHC-I loci in the reference genome's individual, all located in the same contig (Table 2). This contig was aligned to a microchromosome (pseudochromosome 35), as predicted for other avian species,[36,37] which coverage was 85X. The alleles located in loci (L) L2-L7 in the genome were newly identified, i.e., they had not been amplified in any of the 67 specimens using the universal MHC-I primers, whereas the allele in L1 was identical to amplicon allele MHC-Gucr*01 as found in the primary haplotig (Tables S10 and S11). Amplicon sequencing of this specimen confirmed amplification of L1 and yielded allele MHC-Gucr*02, the allele found in the alternative haplotig in the PacBio genome. Similarity at

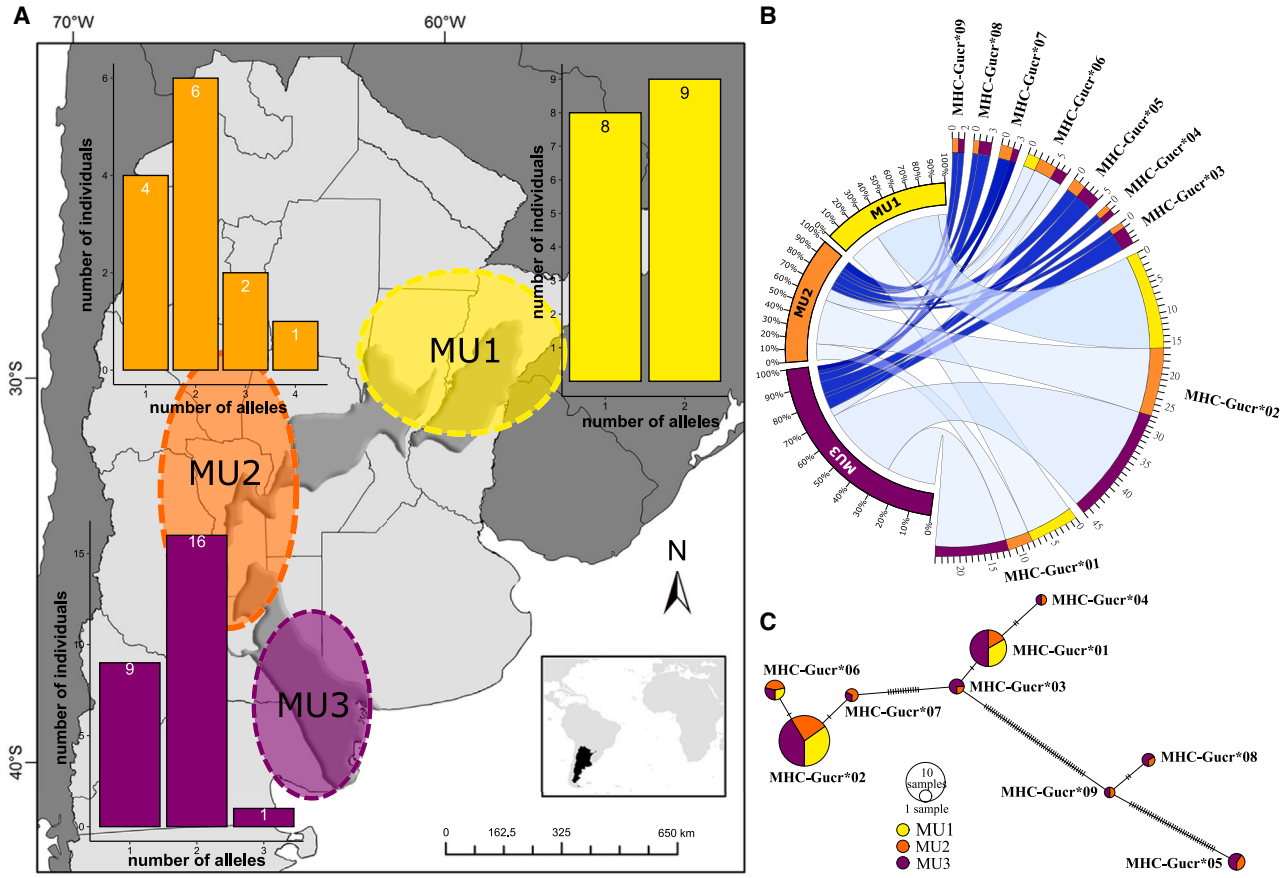

Figure 2. Variation of MHC class I amplicon alleles

(A) Map showing the location of the sampling sites (dashed circles) and the absolute frequency distributions of the inferred number of different nucleotide alleles per yellow cardinal by management unit (MU). The *Espinal* ecoregion in Argentina, as outlined by Brown and Pacheco (2006),[34] is illustrated in dark gray.

(B) Alleles' distributions across yellow cardinal MUs (MU1,2,3 in yellow, orange and violet, respectively). The left outer track indicates the relative frequency of specific alleles by MU, while the right semicircle displays all discovered amplicon alleles. Ribbon thickness reflects the number of individuals carrying a certain allele. Light blue ribbons denote alleles common to all MUs, while dark blue ones represent alleles shared between MU2 and MU3.

(C) A minimum spanning network reveals the relationship and abundance of amplicon alleles. Circle size is proportional to the number of individuals and each dash represents a single base substitution.

nucleotide level was higher for alleles of the same locus (average similarity among alleles assigned to locus 1 was 96.5%), in relation to the similarity across all alleles at all MHCI genes (84.67%, Table S11).

Two genomic arrangements were found: one single copy gene (L1) separated by ~2.27 million bases from the other six MHC-I genes (L2-L7) which have comparatively short intergenic distances between them, and hence seem to have been duplicated in tandem (Figure 3). The 1Mb genomic windows where the MHC-I genes were identified (pseudochromosome_35) presented 8.3 SNPs per Kb in the window where MHC-I L1 was detected and just 1.8 SNPs per Kb for the window where the L2-L7 MHC-I genes were identified. While some amplicon alleles could be assigned to a genomic MHC locus, others had almost equal probabilities of assignment to several MHC loci (Table S10). Alleles were either assigned to L1 or had similar assignment probabilities to (some of) the tandemly duplicated loci L2-L7. Nucleotide similarity was higher between alleles assigned to tandemly duplicated genes (average 86.2%) than between the single copy vs. tandemly duplicated genes (average 79.6%, Table S11).

By retrieving the annotations from other passerine bird species, we identified almost all exons across the seven MHC-I loci in the yellow cardinal, except for exon 1 in locus 1, which was consistently absent, regardless of the query used (Table S12). Additionally, exons 7 and 8 were not found for some yellow cardinal's loci. These exons are small and challenging to detect via BLAST search,[25] hence we consider these loci also to be full-length.

Given that only 9 alleles were identified population-wide via amplicon sequencing, compared to the 7 different MHC-I alleles found in the high-quality genome of a single individual, we evaluated the possibility of a technical artifact associated with the published universal primers, leading to allelic dropout. Moreover, amplicon alleles MHC-Gucr*05,08,09 are absent in all individuals sampled from MU1, and the probability of getting this result by chance (assuming that the primers amplify these alleles equally

Table 2. Position, length, and distance to end position of exon 3 of previous locus between MHC-I exon 3 alleles identified in the reference genome

Locus	Allelic name	Start	End	Length	Distance
L1	MHC-Gucr*01	113,457	113,732	276	
L2	MHC-Gucr*10	2,385,245	2,385,520	276	2,271,513
L3	MHC-Gucr*11	2,474,851	2,475,126	276	89,331
L4	MHC-Gucr*12	2,599,263	2,599,538	276	124,137
L5	MHC-Gucr*13	2,686,867	2,687,142	276	87,329
L6	MHC-Gucr*14	2,773,232	2,773,504	273	86,090
L7	MHC-Gucr*15	2,858,903	2,859,178	276	85,399

well in all MUs, and that alleles are at equal frequencies in all MUs) is very low (2.7%). To investigate the efficacy of the published universal primers to amplify the different MHC-I identified in our reference genome, we aligned the primer sequences to the yellow cardinal's contig containing the MHC-I genes. Only one locus (L1) displayed perfect alignment for both forward and reverse primers (Figure 3). Despite these primers being designed to target invariant intronic regions,[38,39] perfect alignment at other loci was unattainable and at least one mismatch in one of the primers was identified (up to 94% perfect alignment). For loci L2-6, the reference genome's individual exhibited one base difference in the region where the reverse primer MHCpasCI binds, while locus L2 presented an additional mismatch in the region where the forward primer HN34 aligns (Figure 3B). Locus L7 exhibited a unique mismatch in the 3′ region where the forward primer aligns, potentially leading to unsuccessful PCR amplifications, even at a low annealing temperature. In essence, our finding that the universal primers only perfectly match for locus L1 was corroborated by our amplicon sequencing in 67 specimens, yielding consistent amplification only for this locus (no evidence of allelic dropout), with sporadic additional amplification of additional alleles not assigned to L1. Figure 3 further suggests that alleles of locus L1 have a distinct evolutionary history, relative to alleles assigned to other MHC-I loci. Consequently, for downstream selection and genetic diversity analysis only amplicon alleles assigned to locus L1 (MHC-Gucr*01–04, 06–07) were used.

Genetic diversity and selection at MHC-I locus 1

With regard to locus L1, allelic richness was similar in MU2 and MU3 while being lowest in MU1 (Table 3). Remarkably, the three alleles identified in MU1 are distinctively different from each other, as evidenced by the other genetic diversity metrics.

Through analysis of codon-specific signatures indicating historical selection, we identified 19 sites exhibiting evidence of positive selection (Table S9). Of these sites, all except codons 9, 32, and 69 have been previously found to be under selection in other passerine species. Notably, these three codons are located one or two residues adjacent to sites identified under putative selection in other bird species and are part of the human PBR.[32] Such evidence of selection at these sites suggests the potential of locus L1 to contribute to the functional diversity of the MHC.

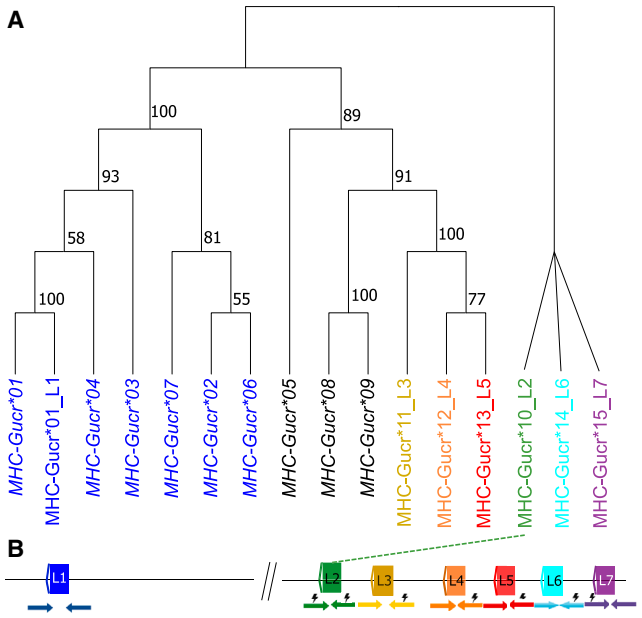

Figure 3. Genomic rearrangement of MHC-I genes in the *Guberna-trix cristata* reference genome

(A) Maximum likelihood tree showing the evolutionary relationships between the yellow cardinal's MHC amplicon (italics) and alleles from the reference genome. Only bootstrap values over 50% are shown. The polytomy on the far right reflects unresolved relationships due to limited phylogenetic signal.

(B) Amplicon alleles mapped to cardinal's contig *ptg000050L* (pseudochromosome 35). Contig orientation is reverse. Each lightning symbol denotes a mismatch between the "universal" primer and the genome. Alleles in black were not assigned to L1 but with almost equal probability to L2–L7 (see Table S10).

DISCUSSION

Accuracy of MHC genotyping

Accurate MHC genotyping is crucial in conservation genetics, where MHC variation serves as a proxy for adaptive variation. However, the reliable characterization of MHC loci has been a persistent challenge due to their complex repetitive nature and the limitations inherent to short-read sequencing technologies. Inaccuracies in MHC genotyping may lead to a misinterpretation of MHC variability, a concern of particular importance for endangered species as it could lead to inadequate conservation strategies. MHC-I variation plays a vital role in immune function and disease resistance, which are critical for the survival of endangered species. Maintaining genetic diversity within MHC loci enhances a species' ability to adapt to environmental changes and emerging diseases. Therefore, understanding and accurately characterizing MHC variation is essential for informing conservation efforts, such as improving breeding programs and safeguarding genetic diversity, to ensure the long-term survival and adaptability of endangered species.[3] By comparing two molecular methodological approaches, i.e., amplicon sequencing and high-fidelity long-read whole genome sequencing, we investigated the genomic organization and polymorphism of MHC-I genes in the endangered yellow cardinal.

Table 3. Locus-specific (L1) genetic diversity measures of yellow cardinal populations

Management Unit	Allelic richness	S	π	Haplotype diversity	K
MU1	2.950	16	0.025	0.542	7
MU2	6	17	0.024	0.719	6
MU3	4.699	17	0.028	0.657	8

Allelic richness (rarified to 26 alleles), the number of polymorphic sites (S), nucleotide (π) and haplotype diversity, as well as the average number of nucleotide differences among alleles (K) are reported.

This study represents the first report of a highly continuous and complete reference genome for the yellow cardinal. In this genome, we have identified 7 MHC class I loci, a finding that contrasts with the results from amplicon sequencing, which detected only 1 to 4 MHC-I amplicon alleles per cardinal, suggesting the presence of merely two MHC-I loci under the assumption of prevalent heterozygosity. Our amplicon sequencing approach involved frequently used non-locus specific primers, designed based on highly conserved sequences flanking the MHC-I exon 3 found in birds and reptiles.[16,39–41] Our findings indicate that such primers can lead to biased allelic dropout leading to potential underestimation of diversity and divergence. A detailed inspection of the primer's affinity to our *de-novo* genome assembly revealed mismatches in all but one locus. This suggests that these widely used primers only reliably detect variation at a single locus (L1), with sporadic co-amplification of additional loci, leading to incomplete allele identification. Given that allelic dropout is notably problematic in genotyping hypervariable regions like the MHC,[42] this could account for the observed inconsistencies in MHC-I polymorphism estimations between amplicon sequencing and long-read genome assembly in this study. The nucleotide differences in introns surrounding MHC-I exon 3 that we found in the reference genome underpin the importance of designing locus and species-specific primer pairs for each locus. This approach is necessary because the effectiveness of universal primers in non-coding regions depends on how much the paralogs have diverged - a factor influenced by time since duplication, species-specific mutation rates, recombination/gene conversion rates, and genetic drift.[43]

While recent studies have also highlighted the benefits of long-read sequencing technologies in identifying the number of MHC paralogs in humans[44] and other bird species,[13,25] very few have used high fidelity long reads. Genome assembly in the MHC region is notoriously challenging due to its high copy number and the high similarity among loci and short sequencing (i.e., Illumina) precludes reliable assignment of alleles to loci. Therefore, long reads that span entire exons and even genes are crucial for studies aiming to disentangle the MHC architecture. Moreover, low sequencing error rates, such as those reported in HiFi long reads,[45] are essential for assessing MHC polymorphism. We detected a total of 7 MHC-I loci in the yellow cardinal's genome assembled with PacBio HiFi reads, a count higher than the estimated number for *Thraupis episcopus* ($N = 4$ loci), another species within the same family,[46] based on the maximum number of alleles detected per individual with amplicon sequencing. However, as our analysis demonstrates, these inferences of gene copy numbers are not directly comparable. Moreover, further investigation is necessary to determine the functional importance of all 7 identified MHC-I loci in the cardinal's genome, as not all MHC-I gene copies are expressed in other species.[12,24]

The organization of MHC-I genes in the yellow cardinal genome comprises both single copy and tandemly duplicated MHC genes. In the broader context of passerine MHC studies, our findings align with those previously reported in other passerine species,[12,21,22] confirming the complexity and diversity of MHC genomic architecture at least across passerine birds. When conducting a detailed analysis of nucleotide similarities between the tandemly duplicated loci, we found higher observed nucleotide similarity between locus L2 and loci L6-L7, rather than between L2 and L3-L5. We hypothesize that this observation may reflect an ancestral inversion event, which suggests that loci L2, L6, and L7 were originally contiguous before being rearranged, leading to their current configuration (Table S11; Figure 3).

Locus-specific MHC-I variation

Despite the MHC genes being among the most polymorphic genes described in vertebrates,[47] amplicon sequencing of yellow cardinals revealed only six alleles at MHC-I locus 1, a count lower compared to other passerine species.[23,48,49] Also, both the number of polymorphic sites and the average number of nucleotide differences among alleles were smaller than those found in another Thraupidae species, the blue-grey tanager (*Thraupis episcopus*).[39] In our analysis of adaptive variation at the MHC-I locus 1, unlike previously published findings on neutral variation for mtDNA and microsatellites,[27] no private alleles were detected.

MU1 was the population with the lowest allelic richness, yet it maintained a nucleotide diversity similar to the other populations. This pattern of low allele numbers despite high nucleotide diversity at MHC-I genes is compatible with an evolutionary scenario involving the interplay between long-term balancing selection on MHC genes and genetic drift affecting the declining and fragmented cardinals' populations. Divergent MHC allele lineages may detect a complementary set of antigen peptides and provide, in combination, a broader range of immunity.[50] Our finding that the SNP density around the MHC-I L1 region is higher than the average genome-wide heterozygosity also supports a scenario of balancing selection acting on these MHC genes.

Through time, these lineages will experience further mutations, potentially giving rise to a set of more closely related MHC alleles within a lineage. How many of these alleles may be retained in a population may however depend on its effective population size, as genetic drift may purge alleles in small populations. Indeed, we see in the smallest population (MU1 in Table 3) a greatly reduced allelic richness and haplotype diversity, while nucleotide diversity is at the same levels as in larger populations (MU2, MU3). This is compatible with our proposed evolutionary scenario in which balancing selection maintains major MHC allele lineages, while drift may reduce allelic richness. As a caveat, this finding is only based on one locus (L1) and it remains to be evaluated in how far this pattern appears at other MHC loci.

Evolutionary perspectives inferred from the yellow cardinal's reference genome

The yellow cardinal's reference genome revealed high level of genome-wide heterozygosity (0.0076), which exceeds that observed in 16 of 17 other Passeriformes species (average heterozygosity = 0.0023).[51] Notably, other endangered species like the noisy scrubbird (*Atrichornis clamosus*, 0.0018) and the Seychelles magpie-robin (*Copsychus sechellarum*, 0.0005), and critically endangered species like the yellow-breasted bunting (*Emberiza aureola*, 0.0046) and the Bali starling (*Leucopsar rothschildi*, 0.0005) presented considerably lower heterozygosity levels than the yellow cardinal. Additionally, our analysis - for the analyzed specimen - revealed a low inbreeding level with most ROHs being short. With all caution due to the fact that these inferences are based on a single specimen, our finding is compatible with a scenario where – after a Pleistocene decline – populations may have been stable for a prolonged period, until a fairly recent decline due to illegal trapping.[52]

Conclusion

This is the first report of a high-quality genome of *Gubernatrix cristata*. We used this resource to shed light into the MHC-I genomic organization and variation in passerine birds. Through a comparative analysis of methods, we uncovered critical limitations of employing non-locus specific universal primers for amplicon sequencing of MHC-I, as they potentially underestimate diversity due to allelic dropout. The integrated genome-wide locus-specific approach presented in this study allows for a holistic understanding of the MHC-I genetic landscape in the yellow cardinal, its evolutionary dynamics, and implications for species conservation.

Limitations of the study

Our study relies on a single reference genome for MHC characterization, which may not adequately capture potential copy number variation in MHC genes across yellow cardinal individuals. Furthermore, as RNA sequencing data were not available, we are unable to ascertain which MHC genes are actively expressed. A broader analysis incorporating multiple whole genomes, coupled with transcriptomic data, could unveil a more complete picture of MHC variability and expression patterns within this globally threatened passerine. Such studies could utilize the genomic resources presented here to directly target the different MHC-I loci, as well as other relevant genes of interest.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Marisol Dominguez (dominguez@uni-potsdam.de).

Materials availability

All unique/stable reagents generated in this study are available from the lead contact with a completed materials transfer agreement.

Data and code availability

- The yellow cardinal genome assembly Gcri1.0.pri has been deposited at GenBank under the accession JBEIAG000000000. The MHC-I DNA sequences corresponding to the alleles described here are available at GenBank with accession numbers PP916322- PP916330. Both are publicly available as of the date of publication. Raw data reported in this paper will be shared by the lead contact upon request.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank the public education system of Argentina, and the National Scientific and Technical Research Council (CONICET) for their financial support to MD at the early stages of this study. Special thanks are also due to the Alexander von Humboldt Foundation for the postdoctoral fellowship and to the University of Potsdam for facilitating the latter stages of this research. This work was supported by CONICET, University of Buenos Aires and Aves Argentinas/AOP.

AUTHOR CONTRIBUTIONS

Conceptualization, R.T., B.M., and M.D.; methodology, R.T., B.M., M.D., K.H., and B.D.C.; formal analysis, M.D., E.C., and N.G.; Investigation, B.D.C. and M.D.; writing – original draft, M.D., E.C., and N.G.; writing – review and editing, M.D., E.C., N.G., R.T., B.M., K.H., and B.D.C.; funding acquisition, R.T. and B.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Animal materials
- METHOD DETAILS
  - Reference genome sequencing and assembly
  - SNP calling and heterozygosity estimation
  - Amplicon sequencing in cardinals' populations
  - MHC-I amplicon alleles
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Runs of homozygosity and demographic analysis
  - Phylogenetic tree for passerines MHC-I exon 3
  - MHC-I exon 3 alleles in Passeriformes
  - Assessing MHC-I polymorphism via amplicon sequencing
  - MHC-I manual annotation in genome assembly
  - Locus-specific diversity and historical selection

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2025.112301.

Received: June 15, 2024
Revised: November 24, 2024
Accepted: March 24, 2025
Published: March 27, 2025

REFERENCES

1. Rowe, G., Beebee, T.J.C., and Sweet, M. (2017). An Introduction to Molecular Ecology (Oxford University Press).
2. Spurgin, L.G., and Richardson, D.S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc. Biol. Sci. *277*, 979–988. https://doi.org/10.1098/rspb.2009.2084.

3. Sommer, S. (2005). The importance of immune gene variability (MHC) in evolutionary ecology and conservation. Front. Zool. *2*, 16. https://doi.org/10.1186/1742-9994-2-16.

4. Rymešová, D., Králová, T., Promerová, M., Bryja, J., Tomášek, O., Svobodová, J., Šmilauer, P., Šálek, M., and Albrecht, T. (2017). Mate choice for major histocompatibility complex complementarity in a strictly monogamous bird, the grey partridge (*Perdix perdix*). Front. Zool. *14*, 9. https://doi.org/10.1186/s12983-017-0194-0.

5. Grieves, L.A., Gloor, G.B., Bernards, M.A., and MacDougall-Shackleton, E.A. (2019). Songbirds show odour-based discrimination of similarity and diversity at the major histocompatibility complex. Anim. Behav. *158*, 131–138. https://doi.org/10.1016/j.anbehav.2019.10.005.

6. Song, X., Wang, L., Yang, Y., Wei, X., Yu, J., Gong, Y., and Wang, H. (2021). Mate choice for major histocompatibility complex (MHC) complementarity in the Yellow-rumped Flycatcher (*Ficedula zanthopygia*). Avian Res. *12*, 27. https://doi.org/10.1186/s40657-021-00261-w.

7. Kamiya, T., O'Dwyer, K., Westerdahl, H., Senior, A., and Nakagawa, S. (2014). A quantitative review of MHC-based mating preference: the role of diversity and dissimilarity. Mol. Ecol. *23*, 5151–5163. https://doi.org/10.1111/mec.12934.

8. Stervander, M., Dierickx, E.G., Thorley, J., Brooke, M.L., and Westerdahl, H. (2020). High MHC gene copy number maintains diversity despite homozygosity in a Critically Endangered single-island endemic bird, but no evidence of MHC-based mate choice. Mol. Ecol. *29*, 3578–3592. https://doi.org/10.1111/mec.15471.

9. Eltschkner, S., Mellinger, S., Buus, S., Nielsen, M., Paulsson, K.M., Lindkvist-Petersson, K., and Westerdahl, H. (2023). The structure of songbird MHC class I reveals antigen binding that is flexible at the N-terminus and static at the C-terminus. Front. Immunol. *14*, 1209059. https://doi.org/10.3389/fimmu.2023.1209059.

10. Blum, J.S., Wearsch, P.A., and Cresswell, P. (2013). Pathways of antigen processing. Annu. Rev. Immunol. *31*, 443–473. https://doi.org/10.1146/annurev-immunol-032712-095910.

11. Lenz, T.L., Mueller, B., Trillmich, F., and Wolf, J.B.W. (2013). Divergent allele advantage at MHC-DRB through direct and maternal genotypic effects and its consequences for allele pool composition and mating. Proc. Biol. Sci. *280*, 20130714. https://doi.org/10.1098/rspb.2013.0714.

12. O'Connor, E.A., and Westerdahl, H. (2021). Trade-offs in expressed major histocompatibility complex diversity seen on a macroevolutionary scale among songbirds. Evolution *75*, 1061–1069. https://doi.org/10.1111/evo.14207.

13. Mellinger, S., Stervander, M., Lundberg, M., Drews, A., and Westerdahl, H. (2023). Improved haplotype resolution of highly duplicated MHC genes in a long-read genome assembly using MiSeq amplicons. PeerJ *11*, e15480. https://doi.org/10.7717/peerj.15480.

14. Radwan, J., Babik, W., Kaufman, J., Lenz, T.L., and Winternitz, J. (2020). Advances in the Evolutionary Understanding of MHC Polymorphism. Trends Genet. *36*, 298–311. https://doi.org/10.1016/j.tig.2020.01.008.

15. Gangoso, L., Alcaide, M., Grande, J.M., Muñoz, J., Talbot, S.L., Sonsthagen, S.A., Sage, G.K., and Figuerola, J. (2012). Colonizing the world in spite of reduced MHC variation. J. Evol. Biol. *25*, 1438–1447. https://doi.org/10.1111/j.1420-9101.2012.02529.x.

16. Promerová, M., Albrecht, T., and Bryja, J. (2009). Extremely high MHC class I variation in a population of a long-distance migrant, the Scarlet Rosefinch (*Carpodacus erythrinus*). Immunogenetics *61*, 451–461. https://doi.org/10.1007/s00251-009-0375-x.

17. Stipoljev, S., Bužan, E., Rolečková, B., Iacolina, L., and Šprem, N. (2020). MHC Genotyping by SSCP and Amplicon-Based NGS Approach in Chamois. Animals *10*, 1694. https://doi.org/10.3390/ani10091694.

18. Shiina, T., Shimizu, C., Oka, A., Teraoka, Y., Imanishi, T., Gojobori, T., Hanzawa, K., Watanabe, S., and Inoko, H. (1999). Gene organization of the quail major histocompatibility complex (MhcCoja) class I gene region. Immunogenetics *49*, 384–394. https://doi.org/10.1007/s002510050511.

19. Moon, D.A., Veniamin, S.M., Parks-Dely, J.A., and Magor, K.E. (2005). The MHC of the duck (*Anas platyrhynchos*) contains five differentially expressed class I genes. J. Immunol. *175*, 6702–6712. https://doi.org/10.4049/jimmunol.175.10.6702.

20. Chaves, L.D., Krueth, S.B., and Reed, K.M. (2009). Defining the turkey MHC: sequence and genes of the B locus. J. Immunol. *183*, 6530–6537. https://doi.org/10.4049/jimmunol.0901310.

21. Chen, L.-C., Lan, H., Sun, L., Deng, Y.-L., Tang, K.-Y., and Wan, Q.-H. (2015). Genomic organization of the crested ibis MHC provides new insight into ancestral avian MHC structure. Sci. Rep. *5*, 7963. https://doi.org/10.1038/srep07963.

22. Biedrzycka, A., O'Connor, E., Sebastian, A., Migalska, M., Radwan, J., Zając, T., Bielański, W., Solarz, W., Ćmiel, A., and Westerdahl, H. (2017). Extreme MHC class I diversity in the sedge warbler (*Acrocephalus schoenobaenus*); selection patterns and allelic divergence suggest that different genes have different functions. BMC Evol. Biol. *17*, 159. https://doi.org/10.1186/s12862-017-0997-9.

23. Vekemans, X., Castric, V., Hipperson, H., Müller, N.A., Westerdahl, H., and Cronk, Q. (2021). Whole-genome sequencing and genome regions of special interest: Lessons from major histocompatibility complex, sex determination, and plant self-incompatibility. Mol. Ecol. *30*, 6072–6086. https://doi.org/10.1111/mec.16020.

24. Westerdahl, H., Mellinger, S., Sigeman, H., Kutschera, V.E., Proux-Wéra, E., Lundberg, M., Weissensteiner, M., Churcher, A., Bunikis, I., Hansson, B., et al. (2022). The genomic architecture of the passerine MHC region: High repeat content and contrasting evolutionary histories of single copy and tandemly duplicated MHC genes. Mol. Ecol. Resour. *22*, 2379–2395. https://doi.org/10.1111/1755-0998.13614.

25. He, K., Minias, P., and Dunn, P.O. (2021). Long-Read Genome Assemblies Reveal Extraordinary Variation in the Number and Structure of MHC Loci in Birds. Genome Biol. Evol. *13*, evaa270. https://doi.org/10.1093/gbe/evaa270.

26. Pikus, E., and Minias, P. (2022). Using de novo genome assembly and high-throughput sequencing to characterize the MHC region in a non-model bird, the Eurasian coot. Sci. Rep. *12*, 7031. https://doi.org/10.1038/s41598-022-11018-w.

27. Domínguez, M., Tiedemann, R., Reboreda, J.C., Segura, L., Tittarelli, F., and Mahler, B. (2017). Genetic structure reveals management units for the yellow cardinal (*Gubernatrix cristata*), endangered by habitat loss and illegal trapping. Conserv. Genet. *18*, 1131–1140. https://doi.org/10.1007/s10592-017-0964-4.

28. Domínguez, M., Reboreda, J.C., and Mahler, B. (2016). Effects of fragmentation and hybridization on geographical patterns of song variation in the endangered Yellow Cardinal *Gubernatrix cristata*. Ibis *158*, 738–746. https://doi.org/10.1111/ibi.12388.

29. Bülau, S.E., Kretschmer, R., Furo, I.d.O., de Oliveira, E.H.C., and de Freitas, T.R.O. (2021). Karyotype Organization of the Endangered Species Yellow Cardinal (*Gubernatrix cristata*). DNA (N.Y.) *1*, 77–83. https://doi.org/10.3390/dna1020008.

30. Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. Nature *475*, 493–496. https://doi.org/10.1038/nature10231.

31. Nadachowska-Brzyska, K., Konczal, M., and Babik, W. (2022). Navigating the temporal continuum of effective population size. Methods Ecol. Evol. *13*, 22–41. https://doi.org/10.1111/2041-210X.13740.

32. Minias, P., Pikus, E., Whittingham, L.A., and Dunn, P.O. (2018). A global analysis of selection at the avian MHC. Evolution *72*, 1278–1293. https://doi.org/10.1111/evo.13490.

33. Klein, J., Bontrop, R.E., Dawkins, R.L., Erlich, H.A., Gyllensten, U.B., Heise, E.R., Jones, P.P., Parham, P., Wakeland, E.K., and Watkins, D.I. (1990). Nomenclature for the major histocompatibility complexes of different species: a proposal. Immunogenetics *31*, 217–219.

34. Brown, A.D., and Pacheco, S. (2006). Propuesta de actualizacion del mapa de ecorregiones de la Argentina (Fundacion Vida Silvestre).

35. Peaper, D.R., and Cresswell, P. (2008). Regulation of MHC class I assembly and peptide binding. Annu. Rev. Cell Dev. Biol. *24*, 343–368.

36. He, K., Liang, C.H., Zhu, Y., Dunn, P., Zhao, A., and Minias, P. (2022). Reconstructing macroevolutionary patterns in avian MHC architecture with genomic data. Front. Genet. *13*, 823686. https://doi.org/10.3389/fgene.2022.823686.

37. Kaufman, J., Milne, S., Göbel, T.W., Walker, B.A., Jacob, J.P., Auffray, C., Zoorob, R., and Beck, S. (1999). The Chicken B Locus Is a Minimal Essential Major Histocompatibility Complex. Nature *401*, 923–925. https://doi.org/10.1038/44856.

38. Westerdahl, H., Wittzell, H., von Schantz, T., and Bensch, S. (2004). MHC class I typing in a songbird with numerous loci and high polymorphism using motif-specific PCR and DGGE. Heredity *92*, 534–542. https://doi.org/10.1038/sj.hdy.6800450.

39. Alcaide, M., Liu, M., and Edwards, S.V. (2013). Major histocompatibility complex class I evolution in songbirds: universal primers, rapid evolution and base compositional shifts in exon 3. PeerJ *1*, e86. https://doi.org/10.7717/peerj.86.

40. Alcaide, M., Edwards, S.V., Cadahía, L., and Negro, J.J. (2009). MHC class I genes of birds of prey: isolation, polymorphism and diversifying selection. Conserv. Genet. *10*, 1349–1355. https://doi.org/10.1007/s10592-008-9653-7.

41. Miller, H.C., Belov, K., and Daugherty, C.H.; SMBE Tri-National Young Investigators (2006). Proceedings of the SMBE Tri-National Young Investigators' Workshop 2005. MHC Class I genes in the Tuatara (*Sphenodon spp.*): evolution of the MHC in an ancient reptilian order. Mol. Biol. Evol. *23*, 949–956. https://doi.org/10.1093/molbev/msj099.

42. Rekdal, S.L., Anmarkrud, J.A., Johnsen, A., and Lifjeld, J.T. (2018). Genotyping strategy matters when analyzing hypervariable major histocompatibility complex-Experience from a passerine bird. Ecol. Evol. *8*, 1680–1692. https://doi.org/10.1002/ece3.3757.

43. Cloutier, A., Mills, J.A., and Baker, A.J. (2011). Characterization and locus-specific typing of MHC class I genes in the red-billed gull (*Larus scopulinus*) provides evidence for major, minor, and nonclassical loci. Immunogenetics *63*, 377–394. https://doi.org/10.1007/s00251-011-0516-x.

44. Chin, C.-S., Wagner, J., Zeng, Q., Garrison, E., Garg, S., Fungtammasan, A., Rautiainen, M., Aganezov, S., Kirsche, M., Zarate, S., et al. (2020). A diploid assembly-based benchmark for variants in the major histocompatibility complex. Nat. Commun. *11*, 4794. https://doi.org/10.1038/s41467-020-18564-9.

45. Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.-C., Hall, R.J., Concepcion, G.T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N.D., et al. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat. Biotechnol. *37*, 1155–1162. https://doi.org/10.1038/s41587-019-0217-9.

46. Minias, P., Pikus, E., Whittingham, L.A., and Dunn, P.O. (2019). Evolution of Copy Number at the MHC Varies across the Avian Tree of Life. Genome Biol. Evol. *11*, 17–28. https://doi.org/10.1093/gbe/evy253.

47. Murphy, K.M., and Weaver, C. (2017). Janeway's Immunobiology (Garland Science).

48. Drews, A., and Westerdahl, H. (2019). Not all birds have a single dominantly expressed MHC-I gene: Transcription suggests that siskins have many highly expressed MHC-I genes. Sci. Rep. *9*, 19506. https://doi.org/10.1038/s41598-019-55800-9.

49. Huang, W., Wang, X., Liu, B., Lenz, T.L., Peng, Y., Dong, L., and Zhang, Y. (2022). Evaluation of Genetic Diversity and Parasite-Mediated Selection of MHC Class I Genes in *Emberiza godlewskii* (Passeriformes: Emberizidae). Diversity *14*, 925. https://doi.org/10.3390/d14110925.

50. Lenz, T.L. (2011). Computational prediction of MHC II-antigen binding supports divergent allele advantage and explains trans-species polymorphism. Evolution *65*, 2380–2390. https://doi.org/10.1111/j.1558-5646.2011.01288.x.

51. Wang, P., Hou, R., Wu, Y., Zhang, Z., Que, P., and Chen, P. (2022). Genomic status of yellow-breasted bunting following recent rapid population decline. iScience *25*, 104501. https://doi.org/10.1016/j.isci.2022.104501.

52. Pessino, M., and Tittarelli, F. (2006). The Yellow Cardinal (*Gubernatrix cristata*): a diagnosis of its situation in the province of La Pampa, Argentina. Gest. Ambient. *12*, 69–76.

53. Baid, G., Cook, D.E., Shafin, K., Yun, T., Llinares-López, F., Berthet, Q., Belyaeva, A., Töpfer, A., Wenger, A.M., Rowell, W.J., et al. (2023). DeepConsensus improves the accuracy of sequences with a gap-aware sequence transformer. Nat. Biotechnol. *41*, 232–238. https://doi.org/10.1038/s41587-022-01435-7.

54. Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics *27*, 764–770. https://doi.org/10.1093/bioinformatics/btr011.

55. Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods *18*, 170–175. https://doi.org/10.1038/s41592-020-01056-5.

56. Cheng, H., Jarvis, E.D., Fedrigo, O., Koepfli, K.-P., Urban, L., Gemmell, N.J., and Li, H. (2022). Haplotype-resolved assembly of diploid genomes without parental data. Nat. Biotechnol. *40*, 1332–1335. https://doi.org/10.1038/s41587-022-01261-x.

57. Guan, D., McCarthy, S.A., Wood, J., Howe, K., Wang, Y., and Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics *36*, 2896–2898. https://doi.org/10.1093/bioinformatics/btaa025.

58. Laetsch, D.R., and Blaxter, M.L. (2017). BlobTools: Interrogation of genome assemblies. F1000Res. *6*, 1287. https://doi.org/10.12688/f1000research.12232.1.

59. Smit, A., Hubley, R., and Green, P. RepeatMasker Open-4.0. 2013–2015. http://www.repeatmasker.org.

60. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. USA *117*, 9451–9457. https://doi.org/10.1073/pnas.1921046117.

61. Grabherr, M.G., Russell, P., Meyer, M., Mauceli, E., Alföldi, J., Di Palma, F., and Lindblad-Toh, K. (2010). Genome-wide synteny through highly sensitive sequence alignment: Satsuma. Bioinformatics *26*, 1145–1151. https://doi.org/10.1093/bioinformatics/btq102.

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

63. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. GigaScience *10*, giab008. https://doi.org/10.1093/gigascience/giab008.

64. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110.

65. Magoč, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics *27*, 2957–2963. https://doi.org/10.1093/bioinformatics/btr507.

66. Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. Bioinformatics *27*, 863–864. https://doi.org/10.1093/bioinformatics/btr026.

67. Sebastian, A., Herdegen, M., Migalska, M., and Radwan, J. (2016). AMPLISAS: a web server for multilocus genotyping using next-generation

amplicon sequencing data. Mol. Ecol. Resour. *16*, 498–510. https://doi.org/10.1111/1755-0998.12453.

68. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.
69. Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. *41*, 95–98.
70. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
71. Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods *9*, 772. https://doi.org/10.1038/nmeth.2109.
72. Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. *52*, 696–704. https://doi.org/10.1080/10635150390235520.
73. Edler, D., Klein, J., Antonelli, A., and Silvestro, D. (2021). raxmlGUI 2.0: A graphical interface and toolkit for phylogenetic analyses using RAxML. Methods Ecol. Evol. *12*, 373–377. https://doi.org/10.1111/2041-210X.13512.
74. Rambaut, A. (2018). Figtree Version 1.4.4, Institute of Evolutionary Biology (Edinburgh: University of Edinburgh).
75. Leigh, J.W., and Bryant, D. (2015). POPART: full-feature software for haplotype network construction. Methods Ecol. Evol. *6*, 1110–1116.
76. Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., and Sánchez-Gracia, A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Mol. Biol. Evol. *34*, 3299–3302. https://doi.org/10.1093/molbev/msx248.
77. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.
78. Okonechnikov, K., Golosova, O., and Fursov, M.; UGENE team (2012). Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics *28*, 1166–1167. https://doi.org/10.1093/bioinformatics/bts091.
79. Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics *26*, 419–420. https://doi.org/10.1093/bioinformatics/btp696.
80. Kosakovsky Pond, S.L., Poon, A.F.Y., Velazquez, R., Weaver, S., Hepler, N.L., Murrell, B., Shank, S.D., Magalis, B.R., Bouvier, D., Nekrutenko, A., et al. (2020). HyPhy 2.5-A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. Mol. Biol. Evol. *37*, 295–299. https://doi.org/10.1093/molbev/msz197.
81. Wheeler, T.J., Clements, J., Eddy, S.R., Hubley, R., Jones, T.A., Jurka, J., Smit, A.F.A., and Finn, R.D. (2013). Dfam: a database of repetitive DNA based on profile hidden Markov models. Nucleic Acids Res. *41*, D70–D82. https://doi.org/10.1093/nar/gks1265.
82. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. *2010*, pdb.prot5448. https://doi.org/10.1101/pdb.prot5448.
83. Fortes, G.G., and Paijmans, J.L.A. (2015). Analysis of Whole Mitogenomes from Ancient Samples. Methods Mol. Biol. *1347*, 179–195. https://doi.org/10.1007/978-1-4939-2990-0_13.
84. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. j. *17*, 10. https://doi.org/10.14806/ej.17.1.200.
85. Dodge, T.O., Farquharson, K.A., Ford, C., Cavanagh, L., Schubert, K., Schumer, M., Belov, K., and Hogg, C.J. (2023). Genomes of two Extinct-in-the-Wild reptiles from Christmas Island reveal distinct evolutionary histories and conservation insights. Mol. Ecol. Resour. Epub ahead of print. https://doi.org/10.1111/1755-0998.13780.
86. Bird, J.P., Martin, R., Akçakaya, H.R., Gilroy, J., Burfield, I.J., Garnett, S.T., Symes, A., Taylor, J., Şekercioğlu, Ç.H., and Butchart, S.H.M. (2020). Generation lengths of the world's birds and their implications for extinction risk. Conserv. Biol. *34*, 1252–1261. https://doi.org/10.1111/cobi.13486.
87. Bergeron, L.A., Besenbacher, S., Zheng, J., Li, P., Bertelsen, M.F., Quintard, B., Hoffman, J.I., Li, Z., St Leger, J., Shao, C., et al. (2023). Evolution of the germline mutation rate across vertebrates. Nature *615*, 285–291. https://doi.org/10.1038/s41586-023-05752-y.
88. Kumar, S., Suleski, M., Craig, J.M., Kasprowicz, A.E., Sanderford, M., Li, M., Stecher, G., and Hedges, S.B. (2022). TimeTree 5: An Expanded Resource for Species Divergence Times. Mol. Biol. Evol. *39*, msac174. https://doi.org/10.1093/molbev/msac174.
89. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645. https://doi.org/10.1101/gr.092759.109.
90. Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K., and Kosakovsky Pond, S.L. (2012). Detecting individual sites subject to episodic diversifying selection. PLoS Genet. *8*, e1002764. https://doi.org/10.1371/journal.pgen.1002764.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Taeniopygia guttata reference genome	Vertebrate Genomes Project	GenBank: GCF_003957565.2
Gallus gallus MHC-I sequence		GenBank: KJ094479.1
Dromaius novaehollandiae MHC-I sequence		GenBank: XM_026122857.1
MHC annotated zebra finch's genome		GenBank: GCA_002008985.2
Acrocephalus arundinaceus genome	Westerdahl et al.[24]	GenBank: PRJNA765537
Gubernatrix cristata reference genome	This study	GenBank: GCA_041475435.1
MHC-Gucr*01-09	This study	GenBank: PP916322-PP916330
Gallus gallus MHC-I sequence		KJ094479.1
Dromaius novaehollandiae MHC-I sequence		XM_026122857.1
MHC annotated zebra finch's genome		GCA_002008985.2
Acrocephalus arundinaceus genome	Westerdahl et al.[24]	PRJNA765537
Gubernatrix cristata reference genome	This study	GCA_041475435.1
MHC-Gucr*01-09	This study	PP916322-PP916330
Oligonucleotides		
primer HN34	Westerdahl et al.[38]	
primer MHCPasCI-RV	Alcaide et al.[39]	
Software and algorithms		
DeepConsensus v0.2.0	Baid et al. 2023[53]	https://github.com/google/deepconsensus
Jellyfish v2.2.10	Marçais & Kingsford, 2011[54]	https://github.com/gmarcais/Jellyfish
Genomescope2.0	Ranallo-Benavidez et al. 2020[53]	https://github.com/tbenavi1/genomescope2.0
Hifiasm v0.16.1-r375	Cheng et al. 2021,2022[55,56]	https://github.com/chhylp123/hifiasm
purge_dups v1.2.5	Guan et al. 2020[57]	https://github.com/dfguan/purge_dups
BlobTools v2.5.0	Laetsch & Blaxter 2017[58]	https://github.com/blobtoolkit/blobtoolkit
RepeatMasker v4.1.5	Smit et al.[59]	
RepeatModeler v2.0.2	Flynn et al.[60]	
SATSUMA v3.1.03	Grabherr et al.[61]	
minimap2	Li[62]	
SAMtools v1.1.15	Danecek et al.[63]	
Picard v2.27.2		https://broadinstitute.github.io/picard/
gatk4 v4.2.0.0	McKenna et al.[64]	
PSMC	Li and Durbin[30]	
BCFTools v1.11	Danecek et al.[63]	
FLASH v1.2.10	Magoč et al.[65]	
PRINSEQ	Schmieder and Edwards[66]	
AmpliSAS	Sebastian et al.[67]	
CLUSTALW	Thompson et al.[68]	
BioEdit v7.2.5	Hall[69]	
BLAST	Altschul et al.[70]	https://blast.ncbi.nlm.nih.gov/Blast.cgi
jModelTest	Darriba et al.[71]; Guindon and Gascuel[72]	
RAxML	Edler et al.[73]	
FigTree v1.4.4	Rambaut[74]	
PopArt	Leigh and Bryant[75]	
DNAsp	Rozas et al.[76]	

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
BEDTools v2.30.0	Quinlan and Hall[77]	
UGENE	Okonechnikov et al.[78]	
pegas R package	Paradis[79]	
HyPhy v2.5	Kosakovsky Pond et al.[80]	
DeepConsensus v0.2.0	Baid et al. 2023[53]	https://github.com/google/deepconsensus
Jellyfish v2.2.10	Marçais & Kingsford, 2011[54]	https://github.com/gmarcais/Jellyfish

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Animal materials

We generated a *de-novo* genome assembly from a female Yellow cardinal, previously assigned to management unit 1,[27,28] to describe genome-wide diversity, infer ancestral demographic history and determine the number of MHC-I loci. A total of 67 adult yellow cardinals spanning most of the species' range were captured using mist nets during different field campaigns conducted between 2011 and 2014, with the authorizations of the corresponding governmental offices, which granted all collection permits needed. The study was conducted following ethical guidelines for wildlife research. The gender of the sampled individuals is not reported in the dataset. As a result, potential sex-based differences in genetic diversity and MHC-I loci number could not be assessed, which represents a limitation in the generalizability of our findings.

METHOD DETAILS

Reference genome sequencing and assembly

High molecular weight genomic DNA was isolated from blood stored in 96% ethanol using the Nanobind CBB Big DNA kit (Circulomics, Inc.). A PacBio HiFi 15-kilobase (Kb) SMRT cell DNA library was prepared and sequenced in one cell on a Sequel2 platform in the MPI DRESDEN-concept Genome Center.

DeepConsensus v0.2.0[53] was used to generate high quality circular consensus sequencing (CCS) reads, which were later trimmed using cutadapt v1.12 to remove adapters. Haploid genome size was estimated with Jellyfish v2.2.10[54] and Genomescope2.0[53] using a *k*-mer value of 31 (Figure S1). Hifiasm v0.16.1-r375[55,56] was used to assemble the long reads, and purge_dups v1.2.5[57] to purge false duplications and overlaps from the scaffolds, as heterozygous regions can be mistakenly interpreted as duplicate regions. Following the VGP pipeline 2.0, two rounds of purging were applied to obtain the primary and alternative assemblies. The quality of assemblies obtained with different parameters were compared using a beta version of GEP (Genome Evaluation Pipeline, https://git.imp.fu-berlin.de/cmazzoni/GEP), and the best assembly (hifiasm parameters -k31, -l2) was chosen based on the number of contigs generated, contig N50, BUSCO completeness and duplicated rates, qv-score, and *k*-mer completeness. A contamination check was carried out using BlobTools v2.5.0,[58] which identifies and categorizes contaminants based on read coverage and taxonomic annotations.

Repeat annotation was performed with RepeatMasker v4.1.5[59] using the National Center for Biotechnology Information (NCBI) search engine with a library of repeats identified *de-novo* in the Yellow cardinal's genome by RepeatModeler v2.0.2[60] combined with the open Dfam 3.7 global database[81] consisting of 3.4 million transposable elements and repetitive DNA families across 2346 taxa.

We further processed the draft assembly by mapping our draft genome to the Zebra finch (*Taeniopygia guttata*) chromosome-level genome (accession no. GCF_003957565.2) using the Chromosemble pipeline from SATSUMA v3.1.0.[61] With this approach we assigned putative chromosome coordinates to our contigs via synteny, while preserving re-arrangements to the maximum degree possible. Only autosomes were retained for subsequent genome-wide analysis since sex chromosomes have different recombination rates and different effective population sizes that might impact the analysis.

SNP calling and heterozygosity estimation

We mapped the HiFi reads back to our reference genome, with both sex chromosomes and repetitive regions filtered, in order to study the genome-wide heterozygosity and to infer the demographic history of the Yellow cardinal. For mapping we used minimap2[62] with the option --map-hifi, SAMtools v1.1.15[63] for sorting and indexing the alignments, and Picard tools v2.27.2 (https://broadinstitute.github.io/picard/) to remove duplicates (MarkDuplicates) and add read groups (AddOrReplaceReadGroups). Then we called variants using gatk4 v4.2.0.0[64] with HaplotypeCaller and GenotypeGVCFs tools and with the –all-sites option to genotype every single position of the genome. Finally, we classified the sites into SNPs, INDELs, and INVARIANT SITES, and calculated statistics to inform the filtering process and applied GATK best practices for filtering each type of site. Specific thresholds were used for SNPs (QUAL<60, SOR>3, FS>60, MQ<40, MQRankSum<-12.5, ReadPosRankSum<-8) and for INDELs (QUAL<60, FS>200,

ReadPosRankSum<-20.0). Additionally, we applied depth filters of 2x average coverage and 1/3x average coverage, to exclude SNPs in regions with mapping artifacts, paralogs, unmasked repetitive sequences, and other unreliable sites. After filtering, we calculated genome-wide heterozygosity with the following formula: ((SNPs) / (SNPs + INDELs + INVARIANT SITEs)). Additionally, using BEDTools we divided our reference genome into non-overlapping windows of 1 million bins and calculated the number of SNPs per Kb in each window to study how genetic diversity is distributed across the entire genome.

Amplicon sequencing in cardinals' populations

The 67 unrelated sampled individuals covered the three management units (MUs) previously identified by population structure analyses,[27,28] named from northeast to south as MU1, MU2, and MU3 (Table S1). Environmental conditions vary significantly across the Yellow cardinal's range. In MU1, the northernmost area, birds experience high annual precipitation (1737.3 mm) and extreme temperatures (42°C max, -4.3°C min, max mean 32°C, min mean 8°C). Conversely, MU3, at the southern end of the distribution, exhibits low annual rainfall (122.8 mm), while temperatures are similarly extreme, but on average lower (43.22°C max, -9.2°C min, max mean 28°C, min mean 2°C, Sources: SMN Argentina, INTA). Phytogeographic regions along the shrubland forest inhabited by cardinals (including mostly *del Espinal* in MU1, the *Chaqueña* in MU2, and *del Monte* in MU3, Cabrera 1976), show variation in vegetation cover, impacting the distribution of food resources. Blood samples (20–30 μL) were collected from the brachial vein of captured birds, stored in lysis buffer (100 mM Tris pH 8, 10 mM NaCl, 100 mM, EDTA and 2% SDS), and DNA was isolated using the DNeasy Tissue Extraction Kit (Qiagen, Hilden, Germany).

We amplified the entire exon 3 of MHC class I gene(s), due to its role in protein binding, using universal intronic primers HN34[38] and MHCPasCl-RV.[39] Polymerase chain reaction (PCR) reagents and conditions can be found in Table S2. We evaluated 1μL of the PCR products by electrophoresis on 1.5% agarose gel to detect positive amplifications. The individual amplicons were equimolarly pooled into a library following a previously published protocol,[82] with some minor modifications.[83] We additionally included 2 library blanks, and to balance out our low diversity libraries, we added 10% of the PhiX on the sequencing reaction. All libraries were sequenced with a 2x300 cycles MiSeq Reagent Kit v3 (Illumina) at the University of Potsdam, Germany. Amplicon analysis was additionally performed for the same individual from which the reference genome had been produced.

MHC-I amplicon alleles

Sequencing adapters, primers, and intron sequences were trimmed from raw reads, and too short sequences (< 80bp) were removed using cutadapt v1.12.[84] We merged overlapping paired end reads specifying a minimum overlap of 10 bp using FLASH v1.2.10[65] and PRINSEQ was used to remove any sequences with a Phred quality score below Q30.[66]

AmpliSAS[67] was used for allele calling, since it distinguishes artefacts from alleles by removing chimeras and PCR/sequencing errors, and thus enhances the coverage and reliability of the detected putative alleles. Initially, we examined the data for sequencing errors (chimeras, substitutions, and indels) and artefacts, using ampliCHECK. Since the frequency at which errors occurred in our sequencing dataset was 1.28%, we used this information to fit parameters for AmpliSAS genotyping analysis. AmpliSAS was run locally to avoid read coverage limitations. For the clustering step, we required that only variants in-frame with the expected length (276 bp) or without stop codons could be considered as dominant within a cluster, adjusted the minimum dominant frequency threshold to 10% (i.e., increasing the sensitivity to ensure inclusion of highly similar alleles) and used Illumina recommended settings (substitution threshold=1%, indel threshold=0.001%). For the filtering step, we discarded samples which yielded < 500 reads, variants with an intra-amplicon frequency < 1.3% or supported by < 25 reads, as well as variants not in-frame with theoretical lengths (allowing a length error of ± 3 bp). After running AmpliSAS, 11 samples were discarded due to low coverage, and two sequences marked as sequencing errors or artefacts were discarded. All retained sequences (putative alleles) were examined to confirm the absence of indels producing stop codons or shifts in the reading frame, and to confirm the presence of highly conserved cysteines important for disulphide bridge formation and peptide binding.[33] An allele alignment was performed using CLUSTALW[68] as implemented in BioEdit v7.2.5.[69]

QUANTIFICATION AND STATISTICAL ANALYSIS

Runs of homozygosity and demographic analysis

To study runs of homozygosity (ROH) we followed the approach of Dodge et al. 2023[85] with minor modifications. Using *makewindows* and *coverage* commands from BEDTools, we divided our reference genome into non-overlapping windows of 10 Kb and counted the number of SNPs in each window. Then, we manually identified stretches of more than 10 consecutive windows with less than two SNPs per window and annotated them as ROHs (ROH min length 100,000 bp). Additionally, to avoid breaking up long ROH due to potential erroneous calls caused by sequencing or mapping errors, annotated ROHs that were separated by only one window were merged. FROH, which refers to the fraction of the genome in ROH, was calculated as the total length of ROHs*100, divided by the total length of the autosomes.

On the same set of filtered SNPs, we inferred effective population size through time with the software PSMC[30] following authors' guidelines. Briefly, we called a consensus sequence with BCFTools v1.11,[63] converted to psmc input file with fq2psmcfa command

and run psmc with the following parameters: -N30 -t5 -r5 -p "4 + 30*2 + 4 + 6 + 10". Additionally, for plotting we used a generation time of 2.57 years,[86] a mutation rate of 1.01e-08 per site per generation[87] and performed 100 bootstrap replicates to assess variability in the estimates.

Phylogenetic tree for passerines MHC-I exon 3

We inferred molecular phylogenies using the MHC-I amplicon alleles from the Yellow cardinal, along with homologous sequences of similar length from the other Passeriformes. We used a general time reversible (GTR) model of nucleotide substitutions with a discrete Gamma distribution to account for varying rates of evolution at different sites as determined by jModelTest.[71,72] We constructed 100 independent ML tree searches in RAxML.[73] Support values for individual nodes were derived from 1000 bootstrap pseudoreplicates, and chicken (*Gallus gallus*) and emu (*Dromaius novaehollandiae*) MHC-I sequences (GenBank accession numbers KJ094479.1 and XM_026122857.1) were used as outgroups. We visualized the topology of the tree with the highest likelihood with FigTree v1.4.4[74] (Figure S2).

MHC-I exon 3 alleles in Passeriformes

The identified amplicon alleles were compared using the BLAST tool against the Genbank nucleotide database from NCBI to confirm the MHC-I exon 3 identity and to check if they have been previously described in other avian studies. We also downloaded the long-read genomes from 23 other Passeriformes species with high-quality assemblies (Figure S2; Table S3) and created a BLAST database for each. The included species diverged from the Yellow cardinal between ~14 and 36 MYA.[88] Using the nucleotide-nucleotide search option, BLASTN,[70] we identified homologous sequences using the putative alleles genotyped in the yellow cardinal as query. Hits were extracted using BEDtools and then aligned to the Yellow cardinal's MHC amplicon alleles. Sequences containing premature stop codons or shorter than our alleles were discarded.

Assessing MHC-I polymorphism via amplicon sequencing

MHC-I diversity per individual was calculated as the number of different alleles present, and frequency distributions were calculated per MU. The alleles' distribution per Yellow cardinal's MU was visualized in a Circos plot,[89] and the frequency and relationship between the MHC-I alleles with a minimum spanning network plotted with PopArt.[75] The number of segregating sites, of mutations, mean nucleotide diversity, and the average number of nucleotide differences were calculated in DNAsp.[76] The number of alleles and of private alleles within each MU is also reported. To evaluate whether the number of detected alleles per individual is driven by each individual's coverage we performed a correlation analysis between those metrics in R 3.6.3 (Figure S3).

MHC-I manual annotation in genome assembly

The number of MHC-I loci in the Yellow cardinal was unknown. Should there be recent duplications yielding similar alleles across loci, the use of non-locus specific primers could potentially inadvertently co-amplify multiple loci. Simultaneously, it increases the risk of allelic dropout. Thus, while additional alleles are detected, some true alleles might be overlooked, leading to an inaccurate representation of the individual's MHC genetic diversity. To bypass these complications, we mapped MHC-I amplicon alleles to our reference genome assembly using blastn v2.11.0+.[70] We considered hits with > 80 % percentage of identical positions and covering > 90% of the length of our query for all alleles. The genomic coordinates of the hits were extracted with BEDTools v2.30.0,[77] and maximum likelihood molecular phylogenies of amplicon and reference MHC alleles were inferred using MEGA in a similar fashion as described above. We also used the command *bedtools genomecov* to estimate the coverage of the pseudochromosome where the MHC I genes are located.

We checked for pseudogenes caused by indels producing frameshifts or mutations producing premature stop codons. Another source of pseudogenization in MHC-I genes is the absence of one or more exons. Therefore, we inspected if the MHC-I loci identified in the Yellow cardinal's genome are full-length by retrieving exons' annotations for the contigs containing MHC-I genes from the zebra finch (GCA_002008985.2) and great reed warbler (PRJNA765537[24]), and searching for the homologs of those 8 exons in the cardinal's genome. For the larger exons (2,3,4,5) we used the blastn tool and focused on matches where >75% of positions were identical, and the alignment covered >80% of the query sequence. For the smaller exons (1,6,7,8) we used blastn-short considering hits with >85% of identical positions and over 50% sequence overlap. Applying these criteria, we found that all hits were located in the vicinity of the annotated exon 3 in the Yellow cardinal's genome.

To investigate the "universality" of the primers, these were aligned to the contig containing all MHC-I genes in UGENE[78] allowing for up to 1 mismatch per primer. Coordinates and identity scores of each obtained hit were noted.

Locus-specific diversity and historical selection

Genetic diversity and selection analysis were performed using only amplicon alleles assigned to MHC-I locus 1, since the other loci were unreliably amplified by the universal primers (see below). The allelic richness of each MU (rarefied to 26 alleles) was calculated in R with the pegas package.[79] The number of polymorphic sites (S), nucleotide (π) and haplotype diversity, as well as the average number of nucleotide differences among alleles (K) of the different MUs were calculated in DNAsp.[76]

We combined the Yellow cardinals' amplicon alleles from MHC-I locus 1 (excluding MHC-Gucr*05, 08 and 09) with additional alleles extracted from other passerine species from which high quality long-read genomes were available (Figure S1) to test for positive selection, as expected for a classical MHC gene. Natural selection was inferred by calculating the ratio of non-synonymous (dN) to synonymous (dS) mutation rates. Codon specific signatures of episodic positive selection were tested using HyPhy v2.5[80] with the MEME algorithm.[90] A significance level of p-value < 0.05 was considered indicative of positively selected sites.