Acta Scientific Microbiology (ISSN: 2581-3226)

Volume 7 Issue 2 February 2024

Research Article

Sequence Evolution and Copy Number Variation in *HINTW*

Jeet Sawant* and Bengt Hannson

Department of Biology, Lund University, Sweden

***Corresponding Author:** Jeet Sawant, Department of Biology, Lund University, Sweden.

DOI: 10.31080/ASMI.2024.07.1343

Received: December 11, 2023

Published: January 18, 2024

Abstract

The mammalian Y and avian W chromosomes share several features including being repeat rich and gene poor. However, while Y holds several multicopy ampliconic genes, only a single ampliconic gene is known on W, *HINTW. HINTW* has a role in avian sex determination, evolves through gene conversion, and has a homolog on Z, *HINTZ*. Studies in Galliformes found more *HINTW* copies in young than old individuals, but whether a similar aging effect occurs in other birds remains unexplored. We aligned avian *HINTZ* and *HINTW* exon-III sequences and designed primers to study the number of *HINTW* copies in passerines. The primers successfully amplified *HINTW* in DNA extracted from red blood cells of species representatives of three passerine families, and using qPCR we estimated 9-15 *HINTW* copies among passerines, in closely related species from superfamily *Sylvioidea*, which is fewer than the 18-40 copies reported for Galliformes. Finally, we uncovered a significant loss of *HINTW* copies in red blood cells of aging great reed warblers, and that the gene loss accelerated in birds infected with malaria parasites. A further study on migratory species infected with malaria suggests further loss of gene copies hampering the reproductive success in the great reed warbler population.

Keywords: *HINTW*; Ampliconic; Gene Conversion; Gene copies; qPCR; Passerines; Malaria; Great Reed Warblers

Introduction

Reproduction is an essential process for every organism present in nature, in order to propagate the survival of the particular species. The processes of sexual and asexual reproduction processes have been observed and studied in nature. Sexual reproduction is a biological process unique to eukaryotic species and is essential in the lifetime of multicellular eukaryotes such as animals, plants and fungi [1]. It is a biological process whereby gametes containing a single set of chromosomes (termed as haploid) from one sex fuse with gametes from the opposite sex, in a process termed as fertilization. This leads to a progeny (or offspring) individual with cells containing two sets of chromosomes (diploid), with each parent contributing to one of the sets. Haploid gametes (n) are produced in the parents by division of diploid cells through the process of meiosis involving genetic recombination between homologous pair of chromosomes, each cycle producing gametes with half the chromosome number as that of the parents. During fertilization, two haploid gametes fuse together to form a zygote (2n) which leads to a progeny (or offspring) individual with cells containing two sets of chromosomes (diploid), with each set resulting from contribution from one of the parents.

The evidence of the oldest reported occurrence of phenomenon of sexual reproduction was seen in fossilized *Bangiomorpha pubescens* (red algae). Evolutionary biologists have put forth several reasons for the preference of sexual mode of reproduction over asexual method to factors such as survival, better adaptability to changing environments due to periods of migration, DNA repair and preventing Muller's Ratchet, meaning accumulation of deleterious mutations leading to increased genetic load [2].

Sex determination

The sex of an individual organism is the trait that assigns the reproductive functioning of the individual in the nature. The sex of the organism constitutes an essential factor in reaping the maximum benefits associated with the individual [3]. In husbandry sectors like poultry and dairy farming, balancing the sex-ratio in a given population becomes important to increase the production efficiency [4,5]. The sex-determination system is an important tools in assigning sexual characteristics to the organism. In most species, the process is genetic sex-determination (GSD) where the male and female have distinct alleles or genes that help determine the sexual characteristics. However, some species have been reported to utilize different types of sex- determination mechanisms. Temperature-dependent sex determination (TSD) was first described by Charneir in 1966 in a lizard species *Agama agama* (common agama or rainbow agama) and is observed in reptiles and teleost fishes [6,7]. Location-based sex determination is observed in marine worm *Bonellia viridis* where larvae result in male on contact with a female and females result if the larvae attach to bottom of the marine floor, a system controlled by a female specific hormone called bonellin [8]. Varying levels of hermaphroditic behaviour (bi-

sexuality) is observed in the nature as well wherein individuals are capable of producing both male and female gametes [9,10].

In organisms exhibiting GSD, the chromosomal differences between a group of organisms give rise to several combinations of chromosome sets, which lead to the sex of the progeny. The mechanism for GSD is controlled by a set of chromosomes called sex-chromosomes or allosomes. These chromosomes are said to carry genes that control the expression of sex- specific traits in an individual [11]. Sex-determining regions (SDRs) are usually present on the heterogametic chromosome (e.g. Y chromosome in mammals and W chromosome in birds) exhibit two essential features that eventually lead to sex-specific traits, suppressed rate of recombination and a gene or set of genes specific to only the heterogametic chromosome. The XX/XY sex-determination system, as seen in humans among most mammals and insects, is the most familiar system of sex determination. Most females possess two sets of X-chromosomes (XX) while the males possess two distinct chromosomes (XY). The X and Y chromosomes are distinct from each other, with Y being smaller than the X- chromosome due to degeneration over the period of evolution owing to loss of genes on the heterogametic chromosome as a result of occurrence of tandem gene duplication events and accumulation of repetitive sequences [12]. Different variants of the XX/XY sex determination system are seen. In platypus species (*Ornithorhynchus anatinus*), there exist 10 sets of sex chromosome that determine the sex of the progeny due to lacking SRY (sex determining region Y) gene [13]. The sex-determination system in platypus is shown to resemble a pattern similar to that in birds, and evidence suggesting homologs of the master gene involved in avian sex determination have been found on X3 and X5 chromosomes of the platypus [14,15].

Sex determination in birds, similar to several other species, is determined by the sex chromosomes. However, unlike mammals where XX genotype leads to females and XY leads to males, homogametic sex (ZZ) is male while the heterogametic sex (ZW) is female [16]. The ZW sex-determination is also seen in crustaceans, reptiles such as snakes and Komodo dragon, some fish and insects [17-21]. It is postulated that the sex-determination in birds is W-dominant, mainly because of the fact that it is present in a single copy [22]. The W-chromosome has also been subjected to degradation, therefore consisting of only few genes, with regions rich in constitutive heterochromatin and presence of certain late replicating repetitive satellite DNA. The high heterochromatin content and existence of abundant ampliconic and repetitive DNA in heterogametic chromosomes has made it difficult to generate sequences of the W chromosome [23]. Despite the lack of detailed study on the W-chromosome, the avian sex chromosomes Z and W do share several similarities with the X and Y chromosomes, that both groups are metacentric; they pair during meiosis and formation of synaptonemal complex occurs at the short arm end of the two chromosomes. Therefore, there happens to be a pseudoautosomal region that exists [22].

Mechanism of sex determination in birds

Although the sex-determination systems have been well identified and studied for most species, the molecular mechanism behind the sex determination and gonadal differentiation has yet to be agreed upon. This in turn, affect the attempts of maintaining a sex-ratio balancing of the population in an environment [24-26]. There arises a need to identify essential genes involved in the process of sex-determination and understanding their functioning and molecular mechanism.

Although both the mammalian and avian sex-determination systems are based on chromosomal control, the molecular mechanism underlying them are extensively different, as both the systems are not sex-linked. For a long time, there was no evidence suggesting any set of genes common between XY and ZW chromosomes, with OTC (ornithine carbamoyltransferase), a Z-linked gene in birds being the only gene found on the human X-chromosome [27,28]. The lack of evidence suggesting homologies between the Z and W chromosome as in birds and X and Y chromosome in groups such as mammals might present a case for lack of similarities in the sex-determination mechanisms among both group of species. This claim is in all likelihood to be premature, as sequence homologies have been reported between the Z-chromosome and chromosome P9 in humans [29]. Despite the absence of a concrete proof suggesting a common origin between ZW and XY chromosomes, it is believed that both chromosome sets evolved sex-determining loci from the same protoancestral autosome which underwent suppressed recombination to lead to respective sex chromosomes [29].

The mechanism that leads to the determination of the sex of the progeny has been a subject of interest for many researchers [30]. In a few species, it's been a postulate that sex-determination is caused through a sex-specific gene, like Sox3 or SRY in mammals [31]. In groups such as birds, the theory that has been among the widely accepted phenomenon underlying the mechanism for sex-determination across species is the dosage compensation hypothesis. The dosage-compensation hypothesis suggests that expression of X/Z-linked genes is equalized between members of the different biological sexes in organisms with heteromorphic chromosome [31]. Different organisms have evolved to compensate for the same through different evolutionary pathways. X-chromosome inactivation is one of the methods of dosage compensation, observed in XX/XY system of sex-determination, as seen in humans and mice [30]. In contrast, birds have incomplete dosage compensation. In this scenario of sex-determination by genic balance, the ratio of autosome to Z-chromosome copies is crucial [32].

According to the dosage hypothesis, primary sex determination in birds is based on dosage compensation by the genes on the Z-chromosome [30]. A strong candidate among birds is the Z- linked gene DMRT1(Doublesex and mab-3 related transcription factor 1) which when present in two copies (ZZ) results in males and in one

copy (ZW) in females [33]. There are evidences suggesting that *DMRT-1* gene controls the primary sex-determination in bird species by the dosage compensation [16]. This has been observed to be more prevalent in males that in females belonging to the species. Studies showed that reduction in levels of *DMRT-1* leads to feminization of genetically male gonads, whereas overexpression leads to masculinization of female gonads [34]. The loss of even a single copy of *DMRT-1* promoted ovarian formation instead of testicular development brought about by increased expression levels of transcription factor SOX9 [33].

Though *DMRT-1* dosage is responsible for the primary sex-determination in birds, the birds have acquired cell-autonomous sex identity (CASI), meaning that the secondary sexual characteristics are determined by sex-chromosome content of individual cell or tissue types and are not influenced by the sex hormone environment [33]. Furthermore, despite the fact that all the other genes having an ortholog present on the W-chromosome, even in ratites, the absence of the *DMRT-1* copies on the W-chromosome suggest that the mechanism leading to the development of female progeny might be more than it being just controlled by the *DMRT-1* gene. The search for a candidate that influences the development of female specific characteristics in birds and how does it act as an influencer to the regulation of *DMRT-1* has led to researchers to focus on a particular gene named *HINTW* (Histidine Triad Nucleotide-Binding W), believing it to be a potential factor in female sex-determination in birds.

The mammalian Y and avian W have different evolutionary origin but have been affected by similar evolutionary processes and therefore share striking features including being extremely gene poor and heavily enriched on repetitive sequences [11]. This is an effect of their hemizygous status and their inability to recombine with another homologous chromosome, which increases linkage between genes and lowers the efficiency of selection, resulting in an inevitably accumulation of deleterious mutations and gene degeneration [35]. Genes that remain on old Y and W chromosomes typically have essential functions and are maintained by strong purifying selection [36]. However, another way to restrain this degenerative process at least partly is to evolve other mechanisms of recombination than typical homologous crossover. In mammals, several Y linked genes occur in multiple copies, a feature that is hypothesised to facilitate palindrome pairing and gene conversion between ampliconic gene copies along the chromosome. Palindrome formation can also be facilitated by the rich repeat sequence content of these chromosomes. Indeed, 70- 200 ampliconic genes are found on the mice Y chromosome and their functional integrity is assumed to be maintained by gene conversion, resulting in concerted evolution and homogenization of alleles within species [37]. In birds the situation is drastically different with only a single W-linked ampliconic gene known to occur, *HINTW* (histidine triad nucleotide-binding, W-linked). In Galliformes, *HINTW* evolves through gene conversion and has a hypothesised role in sex determination [38].

A potential candidate in *HINTW*

Histidine triad nucleotide-binding protein (HINT) or formerly denoted as Protein Kinase C Inhibitor (PKCI) genes encode proteins associated with binding various substrates via a histidine triad motif. HINT proteins belong to a superfamily of HIT (histidine triad) proteins known for the presence of His-φ-His-φ-His-φ-φon the C-terminus end, where φ is indication of a hydrophobic amino acid residue [39,40]. Several homologs of the *HINT* gene exist such as the *HINT1*, *HINT2*, etc. found across species including humans. In birds, the gene *HINTW* (or WPKC-1) has been implicated in sex determination [41]. Studies have suggested that the *HINTW* is conserved among all carinate (flying) birds' embryos [42]. *HINTW* is located on the W chromosome and has a homolog on the Z chromosome, *HINTZ*. It has been observed to be expressed in a female- specific way and has been found in abundance in females while compared to the males [33]. Speculation is that *HINTW* participates in early secondary sex determination and differentiation in birds by altering the sexual characteristics by inducing morphological changes in the gonads [38]. This was observed in a study carried out where it was observed that overexpression of *HINTW* gene led to asymmetrical testes formation and inhibition of the same resulted in abnormal ovarian development [38]. On the other hand, interference in expression in males led to significant increase in levels of testosterone and decreased estrogen resulting in thinned gonadal cortex [38]. In females, increased expression of *HINTW* gene leads to thickening of gonads and collapsing of the seminiferous tubules, a pattern observed in ovarian morphology [38].

However, *HINTW* has some interesting distinct differences to its homolog on the Z- chromosome (*HINTZ*). Studies based on sequence comparison in *Gallus gallus* indicates of 65% similarity between *HINTW-HINTZ* copies while *HINTZ*-human *HINT* gene copies share 87% identity [41]. The study carried out in duck species may indicate occurrence of gene conversion across *HINTW* to homogenize the gene sequence among variants and is consistent with the results from studies seen in galliform species [43]. Sequence analysis of the gene in different species including *Coturnix coturnix* (Common quail), *Phasianus colchicus* (Ringed-neck pheasant) among others suggested that the *HINTW* shows low levels of nucleotide diversity and copies within the species are more likely to be identical to each other than to any copy in a related species [44].

Copy number variation (CNV)

The phenomenon of copy number variation is a recurring observation in genetic studies conducted across different species where certain sequences within the genome are repeated and the number of repetitions for a particular sequence varies substantially among individuals [45]. These variations have risen from events of gene duplications or deletions occurring at one or more sites in the sequence, leading to different variants of the same gene with >90% sequence identity [46]. Evolution of ampliconic gene families have occurred in genomes as a defence mechanism in response to avoiding Muller's Ratchet and deleterious mutation accumulation due

to the absence of recombination events between the sex chromosome [47].

HINTW happens to be the only gene found in multiple copies on the W-chromosome, evolving under positive selection. The pattern observed was very similar to that observed in the case of testis-specific genes involved in reproductive functions on the human Y-chromosome, lacking homologs with similar function on the X-chromosome [47]. Evidence about presence of Leu-Arg rich regions in the amino acid sequence of *HINTW*, a pattern absent in other HINTs might suggest that the *HINTW* copy evolved from an ancestral protosex bird chromosome and went on to acquire a female specific function [39,48-50]. The *HINTW-HINTZ* gene interaction is postulated to be functioning in the same pattern observed in mammalian SRY-SOX3/SOX9, wherein *HINTW-HINTZ* binding inhibits the *HINTZ* copy from regulating male specific function in the females [51].

However, the copy number was found to vary, where Pekin duck possessed a single copy of *HINTW* while Mallard duck had 18 ampliconic copies [11]. In a study based on *Gallus gallus*, *HINTW* gene was the only gene found in multiple copies, with up to 40 copies identified on the female-specific W-chromosome [47]. Copy number variation among species is also seen based on several influencing factors. With respect to species of chicken, the currently available marketable breeds of chicken have higher number of *HINTW* gene copies than their ancestor Red Junglefowl. This might be attributed to the domestication process of chickens, which may have allowed for a period of female specific selection for the process of selective breeding for purpose of egg-laying [52]. This is further supported by the observed trend that breeds selected for egg-laying purposes have a higher copy number than those bred for plumage characteristic and for cock-fighting, a postulated result of relaxed female specific selection in the case of the breeds belonging to the latter category [53]. However, argument in this case has been presented over the fact that this observation might only be true just in case of Black Minorca and not in the White Leghorn breed of chickens [43].

Aim of the study

In the present study, we quantify the number of *HINTW* copies in passerines (songbirds). We downloaded and aligned *HINTZ* and *HINTW* sequences from several species of Galliformes and Passeriformes and designed primers to amplify a portion of *HINTW* exon-III in passerines. We evaluated primer amplification success in three passerine families. These families belong to the species-rich superfamily Sylvioidea (≈1200 species) and diverged ca. 21 million years ago [54]. We focus on one Sylvioidea species, the great reed warbler (*Acrocephalus arundinaceus*), because our ongoing long-term study of individually marked birds in a Swedish population provides access to DNA samples from specific individuals over their entire lives. We apply non-invasive sampling of a small amount of blood of living birds being released after examination. This sampling scheme enables examining whether *HINTW* copy numbers are stable over time or whether individuals lose gene copies in their nucleated red blood cells as they age. Such a degeneration has been observed. We designed a qPCR protocol that amplifies the single-copy gene *HINTZ* as a standard to which the amplification rate of the ampliconic *HINTW* can be evaluated, and its copy number estimated. We demonstrate successful amplification of *HINTW*, uncover copy number variation between species within Sylvioidea, and a significant loss of *HINTW* copies over time in aging great reed warblers with insights for a further study on the effects of malarial infection on the genetic predisposition for female-specific reproductive success.

Materials and Methods

Data collection

Study species

Chicken (CH, *Gallus gallus*), Common quail (CQ, *Coturnix coturnix*), Ringed-neck pheasant (RNP, *Phasianus colchicus*), turkey (TU, *Meleagris gallapavo*) are classified as Galliformes species and have been studied greatly for sequence diversity and copy number variation for *HINTW* gene. Zebra finch (ZF, *Taenopygia guttata*), Barn swallow (BS, *Hirundo rustica*), Great reed warbler (GRW, *Acrocephalus arundinaceus*), Great tit (GT, *Parus major*) belong to order Passeriformes. GRW, Marsh warbler (MW, *Acrocephalus palustris*), Common Whitethroat (WT, *Sylvia communis*), Eurasian skylark (SL, *Alauda arvensis*), African reed warbler (ARW, *Acrocephalus baeticatus*), Eurasian reed warbler (ERW, *Acrocephalus scirpaceus*) are Old-World songbirds further classified into the superfamily *Sylvioidea*.

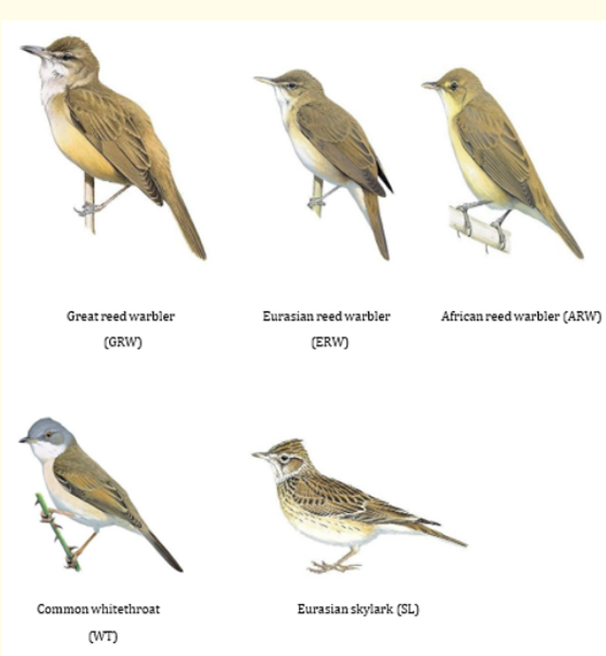

Figure a

Sequence data

The sequence data for Galliformes (chromosome, *HINTZ* and *HINTW*) and chromosome data and *HINTZ* sequences for the Passeriformes were downloaded from NCBI database, GenBank and Genome Data Viewer. The *HINTW* data for the Passeriformes species was obtained using by mapping *HINTW* sequences from related species onto the W-chromosome of the Passeriformes using Geneious Prime 2022.1.1 software to obtain a putative sequence *location* of the *HINTW* gene (Supplementary S1). *HINTW* dataset for Great reed warbler and *HINTZ* data for the great tit was obtained through dataset available with Bengt Hansson's lab acquired by assembling reads from PacBio Assembly sequencing using GRW sequences from Hannah Sigeman's study on sex chromosomes in Sylvioidea songbirds [55].

DNA samples

The DNA was extracted from blood samples using Phenol-chloroform extraction protocol. The members of superfamily Sylvioidea included in the study were obtained from Bengt Hansson's lab work on a previous study on the same. The DNA samples from female individuals were suspended in 1X TE buffer and stored at - 80°C till use. The dilutions were made with 1X TE buffer at the necessary concentration for use in experimental analysis.

Phylogenetic analysis

Phylogenetic tree construction for species under study

The phylogenetic trees for clustering of species were performed with the help of iTOL: Interactive Tree of Life online tool for the annotation, management and visualization of phylogenetic trees [56]. The species input was based on the NCBI taxID of the species provided by the NCBI Taxonomy Database [57] (Supplementary S2). For assessing evolutionary history, mitochondrial gene cytochrome-b sequences were used to construct cladograms for Galliformes and Passeriformes species using human cytochrome- b as an outgroup.

Sequence alignment of Galliformes and Passeriformes for *HINTW* and *HINTZ* sequences

The sequences alignment was performed using Clustal Omega 1.2.2 with Geneious Prime 2022.1.1 and necessary modifications were made to the sequences, when necessary, by eye. Region's exon-II and exon-III of *HINTZ* and *HINTW* genes were analyzed in detail to obtain properly aligned regions within the gene. Intron-II region of the genes were included in constructing phylogenetic trees for selection analysis and inferring evolutionary status.

Phylogenetic tree building for HINTW and HINTZ sequences

The evolutionary history of the sequences was inferred by constructing phylogenetic trees using Neighbour-Joining method with Tamura-Nei Genetic distance model using Geneious Tree Builder function in Geneious Prime 2022.1.1 software. Phylogenetic tree of the study species to be analyzed was constructed. An unrooted tree of sequences of *HINTZ* and *HINTW* from both Galliformes and Passeriformes. A repeat analysis was performed using the Neighbor-Joining method with the Nei-Gojobori method to compute evolutionary distances and are in the units of the number of synonymous substitutions per synonymous site) [58]. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analysed [59]. All ambiguous positions were removed for each sequence pair (pairwise deletion option). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [59]. Evolutionary analyses were conducted in MEGA11 [60].

Figure S1: Sequence alignment for exon 3 of 3 region of *HINTZ* and *HINTW* copies of Passeriformes and Galliformes for designing primer sites. Sequences 1-4 depict galliform *HINTZ* ,5-13 galliform *HINTW*, 13-17 passeriform *HINTZ* and 18-24 Passeriform *HINTW*. Primers were designed using Hirundo rustica as reference for *HINTW* and Parus major for *HINTZ*.

Primer compatibility analysis

Primer design

Primers specific to exon-III region of *HINTW* and *HINTZ* genes were designed using Geneious Prime version 2022.1.1 software (Biomatters Ltd., Auckland, New Zealand) (Figure S1). The primers designed were such that both the forward and reverse primers were in regions specific to only Passeriformes for amplification of both *HINTZ* and *HINTW* copies (Sequences S3). Different sets of forward and reverse primers were designed to find the most optimal pair for sequence amplification. The primer sequences were sent to Eurofins Scientific for synthesis HPSF® Purity grade primers.

Quantification of DNA

Genomic DNA samples extracted from blood samples from different species was available in the lab. To estimate accurate quantity of sample DNA to be added to the reactions, quantification was carried out using a nanodrop (Thermo Scientific ND-2000 spectrophotometer, USA) in order to measure absorbance at 260 nm and estimate the purity of the sample by calculating the OD_{260}/OD_{280} nm absorption ratio which is an indicator of the protein contamination in the sample of interest. 1 µL of the sample was used for quality check in the nanodrop.

Touchdown PCR reaction

Primer pairs designed were tested to find the optimum pair for analysis of *HINTW* and *HINTZ* copies in Passeriformes. Different primer combinations for both *HINTW* and *HINTZ* were tested in male and female samples of great reed warbler, common whitethroat and Eurasian skylark. Touchdown polymerase chain reaction (PCR) is a method used to find optimum annealing and prevent off-target priming for obtaining specific PCR products (Green & Sambrook, 2018). In touchdown PCR, the temperature profile for the reaction is such that the annealing step temperature is initially set the same as around the T_m (melting temperature) of the primers. The temperature is reduced gradually in subsequent cycles such that the annealing temperature ends up around 2°C - 5°C below the calculated T_m of the primers. PCR reaction was carried out using the primers (Table S1), 2 µl of DNA (10 ng/µl), 0.1 µl of Enzyme Taq DNA-Polymerase (5 U/µl), 2.5 µl of PCR 10x buffer, 2.5 µl of dNTP's,1.5 µl of $MgCl_2$ (25 µM), 1 µL of Forward primer (10 µM), 1 µl of Reverse primer (10 µM), 14.4 µl of nuclease-free water in a total reaction volume of 25 µL.

No.	Sequence Name	Fasta ID
1	*Melagris gallopavo* HINTW pseudogene HINTW-1 allele	AY713486.1
2	*Melagris gallopavo* HINTW pseudogene HINTW-2 allele	AY713487.1
3	*Meleagris gallopavo* HINTW-3 allele	AY713488.1
4	*Coturnix coturnix* HINTW-1 allele	AY713489.1
5	*Coturnix coturnix* HINTW-2 allele	AY713490.1
6	*Coturnix coturnix* HINTW-3 allele	AY713491.1
7	*Coturnix coturnix* HINTW-4 allele	AY713492.1
8	*Phasianus colchicus* HINTW-1 allele	AY713493.1
9	Phhasianus colchicus HINTW-2 allele	AY713494.1
10	*Phasianus colchicus* HINTW-3 allele	AY713495.1
11	*Phasianus colchicus* HINTW-4 allele	AY713496.1
12	*Gallus gallus* HINW-1 allele	AY713497.1
13	*Gallus gallus* HINW-2 allele	AY713498.1
14	*Gallus gallus* HINW-3 allele	AY713499.1
15	*Gallus gallus* HINW-4 allele	AY713500.1
16	*Gallus gallus* HINTZ	NC_052572.1:45514110- 45518130
17	*Gallus gallus* isolate bGalGal1 chromosome Z	NC_052572.1
18	*Taeniopygia guttata* isolate Blue55 chromosome Z	NC_044241.2
19	*Taeniopygia guttata* isolate Blue55 chromosome W	NC_045028.1
20	*Hirundo rustica* isolate bHirRus1 chromosome W	NC_053487.1
21	*Hirundo rustica* isolate bHirRus1 chromosome Z	NC_053488.1
22	*Homo sapiens* HINT1	NC_000005.10:c131165256- 131159027
23	*Parus major* isolate Abel chromosome Z	NC_031799.1

Table S1: FASTA ID for sequences used for alignment.

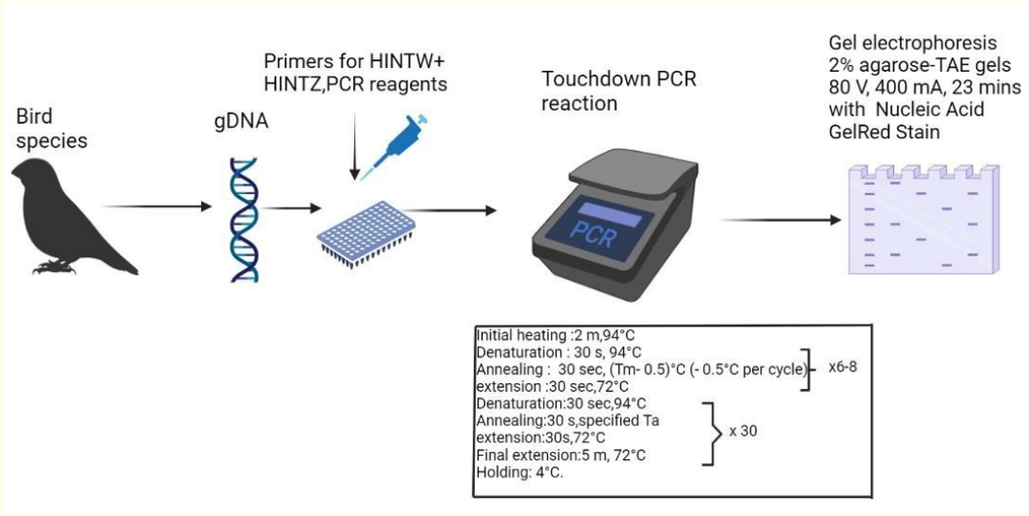

Figure 1: Workflow for primer specificity testing using Touchdown- PCR protocol created using BioRender online software. Designed primers for *HINTZ* and *HINTW* gene were tested on DNA samples extracted from female individuals of birds. Touchdown-PCR thermocycler reaction was used to accommodate optimal annealing temperature for each primer pair. PCR products were confirmed using gel electrophoresis run on 2% agarose-TAE gels, stained using Nucleic Acid GelRed Stain.

Thermocycler for the Touchdown PCR reaction with the following program

Initial heating for 2 min at 94°C, 6-8 cycles of denaturation at 94°C for 30 sec, annealing at temperature 0.5°C below primer melting temperature T_m (with 0.5°C at every cycle decrement up to temperature reaches to 3°C below T_m) for 30 sec, and extension at 72°C for 30 sec. This is followed by a further 30 cycles of denaturation at 94°C for 30 sec, annealing at the specified T_a for the primer pair for 30 sec and extension at 72°C for 30 sec. The reaction was then ended with a final extension for 5 min at 72°C and holding temperature of 4°C.

Gel electrophoresis

Following the reaction, the PCR products were confirmed by loading on 2% agarose-TAE gels and running at 80 V, 400 mA for 23 mins supplemented with 5 µl of Nucleic Acid GelRed Stain per 100 ml. with. The gels were photographed using BIO-RAD Gel Doc XR+ with Image Lab software (BIO-RAD, Hercules, California, USA) gel documentation system. 1kb Plus DNA ladder was used as a reference.

Copy Number analysis using qPCR reaction

Evidence presented by Hori., *et al*. 2000 suggested that the copy number for *HINTW* was seen to reduce between the embryo and adulting stages in different breeds of chicken and ducks. To determine if and how the copy number varies with age in great reed warblers, analysis of DNA samples from female individuals of great reed warblers were tested for *HINTW* gene copies using qPCR protocol mentioned previously. Age difference of 3 years was used as standard measurement for comparison.

qPCR reaction for confirming gene amplification and estimation of CNV

Quantitative data of gene amplification was obtained by using the CFX384 Touch Real-Time PCR Detection System together with the BIO-RAD CFX Maestro Software 1.1. The primer pairs used for analyzing the copy number were the same as used for PCR reaction (See Appendix S3). Based on results from the nanodrop, 1ng/µl dilutions of all the samples were prepared. Standard curve for melting curve analysis was performed using 8 ng, 4 ng, 2 ng, 1 ng, 0.5 ng and 0.25 ng of great reed warbler sample genomic DNA. 96-wells plates were used, each well contained the specific DNA of interest and a master mix with one of the primer pairs as well as Platinum SYBR Green qPCR SuperMix-UDG (Invitrogen) and ROX reference dye, with additions made as follows: 2 µl of DNA (1 ng/µl) was added to each well along with 12.5 µl of Platinum SYBR Green qPCR SuperMix-UDG, 0.1 µl of ROX reference dye, 1 µl of Forward primer (10 µM), 1 µl of Reverse primer (10 µM), 5.4 µl of nuclease-free water in total reaction volume of 25 µl.

Thermocycler for the qPCR reaction with the following program: initial heating for 10 min at 95°C, amplification and quantification steps were repeated for 30 cycles of denaturation (95°C, 30 sec), annealing (57°C,30 sec), elongation (72°C, 30 sec) and plate read. This was followed by Melting curve analysis which was done in 30 cycles of 57°C for 30 sec with + 0.5°C/cycle, Ramp 0.5°C/sec and plate read, at the end of which the plate was cooled down to and held at 4°C.

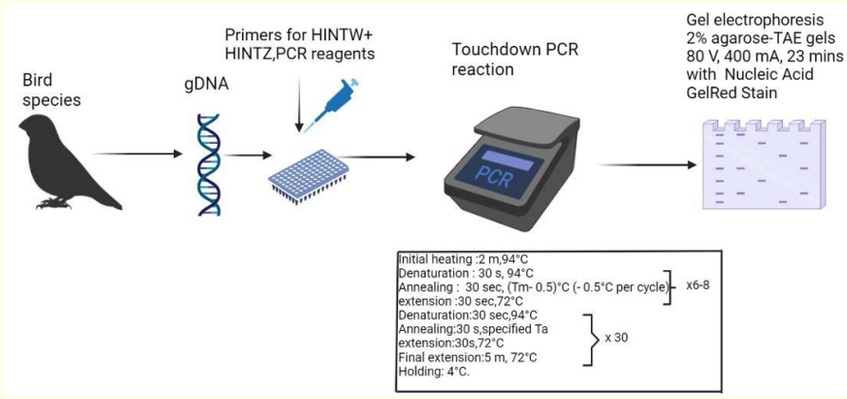

Figure 2: Workflow for copy number analysis using qPCR reaction created using BioRender online software. DNA samples extracted from female individuals of birds were analysed for CNV using qPCR analysis with CFX384 Touch Real-Time PCR Detection System with *HINTZ* as control gene. Melting curve analysis was performed BIO-RAD CFX Maestro Software 1.1 and ΔCq value between *HINTW* and *HINTZ* gene was quantified to estimate the copy number in the test sample.

No.	Species	Name (Symbol)	NCBI TaxID
1	*Gallus gallus*	Chicken (CH)	9031
2	*Meleagris gallapavo*	Turkey (TU)	9103
3	*Coturnix coturnix*	Common Quail (CQ)	9091
4	*Phasianus colchicus*	Ring-necked Pheasant (RNP)	9054
5	*Taenopygia guttata*	Zebra finch (ZF)	59729
6	*Hirundo rustica*	Barn swallow (BS)	43150
7	*Parus major*	Great tit (GT)	9157
8	*Acrocephalus arundinaceus*	Great reed warbler (GRW)	39621
9	*Acrocephalus baeticatus*	African reed warbler (ARW)	68472
10	*Acrocephalus scirpaceus*	Eurasian reed warbler (ERW)	48156
11	*Sylvia communis*	Common whitethroat (WT)	216193
12	*Alauda arvensis*	Eurasian Skylark (SL)	88112
13	*Acrocephalus palustris*	Marsh warbler (MW)	68475

Table S2: NCBI Taxonomic ID for species using in cladogram.

No.	Name	Sequence	Length	Tm	GC%
1	F1-HINTW	GGGCATGTAATGATTGTTGGCG	22	60.3	50
2	R1-HINTW	CGGGTAGATGTATGCGATAGAC	22	60.3	50
3	F2-HINTW	GTTGGCGAGATGTGTGTTGCTC	22	62.1	54.5
4	R2-HINTW	ACGGCCACCCAGAACGGGT	19	63.1	68.4
5	R3-HINTW	GACGGCCACCCAGAACGG	18	62.8	72.2
6	F1-HINTZ	GGGCACTTAATGATTGTTGGCA	22	58.4	45.5
7	R1-HINTZ	CATGGAGATGTACGTGATAGAC	22	58.4	45.5
8	F2-HINTZ	GTTGGCAAGAAGTGTGCTGCTA	22	60.3	50
9	R2-HINTZ	GCGACCACCCAGAACATGG	19	61	63.1
10	R3-HINTZ	GGCGACCACCCAGAACAT	18	58.2	61.1

Table S3*:* Primer sequences designed for HINTW and HINTZ genes.

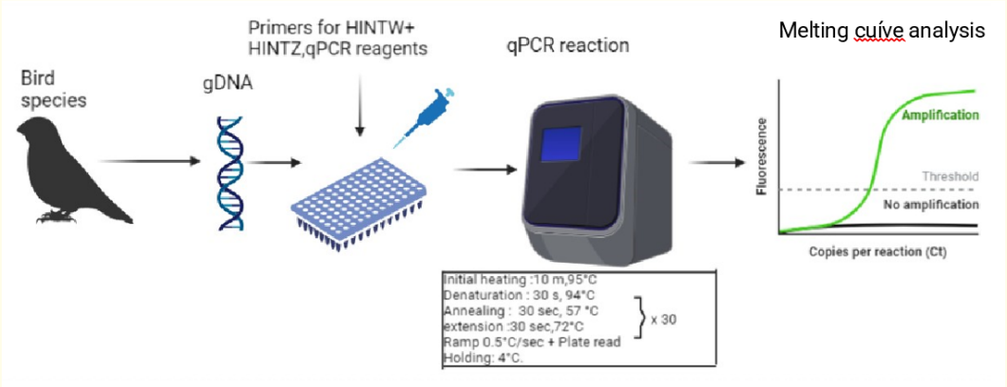

Figure 2: Workflow for copy number analysis using qPCR reaction created using BioRender online software. DNA samples extracted from female individuals of birds were analysed for CNV using qPCR analysis with CFX384 Touch Real-Time PCR Detection System with *HINTZ* as control gene. Melting curve analysis was performed BIO-RAD CFX Maestro Software 1.1 and ΔCq value between *HINTW* and *HINTZ* gene was quantified to estimate the copy number in the test sample.

Quantification of data

Copy number in the test samples was quantified by comparing quantification cycle or Cq value for amplification using *HINTZ* as control. Efficiency of the reaction within range 90-110% with R^2 value>0.98 was considered optimal. The copy number for *HINTW* gene was calculated by using simplified gene expression equation $2^{-\Delta Cq}$, where ΔCq stands for the difference in Cq values (Cq, *HINTW* − Cq, *HINTZ*) on the basis of knowledge that the female individuals (ZW) have a single Z-chromosome and *HINTZ* exits as a single copy gene per chromosome. We tested whether the number of *HINTW* copies changed over the lifetime of individual great reed warblers by linear effect mixed models using the lme4 package (v 1.1-33) in R studio (version 2023.03.1+446 and R version 4.3.0, Posit team 2023). The model included individual age as fixed factor and individual as random factor. We ran models including each sample replicate (most often 3 replicates per individual age-sample; see above) and models based on the mean copy number of the replicates (i.e., 1 value per individual age-sample). As the results from these models (using replicates or mean) did not differ qualitatively, we present the results of the models based on mean copy numbers. Plots were generated the ggplot2 package (v 3.4.2) to analyse age-related variation (1991-2004) in copy number, with consideration for effect of malarial infection status (infected or not).

Results

Sequence alignment for HINTW and HINTZ sequences

HINTZ and *HINTW* gene structure consists of three exons with two interspaced introns. Previous studies performed on the *HINTZ* and *HINTW* sequences in bird species such as chicken and common quail had suggested that the avian *HINTZ* is highly conserved and also shares high similarity in functional domains [41]. *HINTZ* and *HINTW* sequences from the study species were compiled together and aligned using Geneious Prime software version 2022.1 created by Biomatters which provides tools for performing multiple alignments on a single dataset such as MUSCLE, Clustal Omega and in-built Geneious global alignment with free end gaps (cost matrix 65% (5.0, -4.0)) [61]. The best alignment among the three generated ones was selected. Translation alignment of the sequences was also performed to check the sequence diversity at amino acid level.

For *HINTZ* alignment, the sequences showed great level of nucleotide similarity in the regions of exon-II and exon-III for study species in both orders (Figure 3A). The diversity between species exists in the two introns and in the exon-I region as it was not possible to generate a clean alignment. On comparing the translated sequences of *HINTZ* copies for exon-II and exon III region, the sequences were found to be well conserved (Figure 3B). The minor differences seen were that of a single exon in exon-III region of the zebra finch while codon differences are observed separating the species in both orders at position 62 and 78 in the translational alignment of the species.

However, in case of the *HINTW* copies, the nucleotide and amino acid diversity between species was observed to be amplified with each annotated region of the gene (Figure 3C). The only region indicating considerable level of similarity was exon III depicting considerable levels of nucleotide identity and also showing some levels of amino acid similarity. The alignment performed for the other regions (intron I, intron II and exon II) don't yield promising results as seen in case of *HINTZ* copies. The *HINTW* sequences displayed several differences on the nucleotide in the same regions compared in *HINTZ* analysis. Segregating regions separating species from the two orders were observed. Furthermore, comparison on nucleotide level showed that *HINTW* copies from Galliformes had 2 codons missing compared to the Passeriform *HINTW* copies, a TGG around position 97-99 and TAA/TAG towards end of sequence.

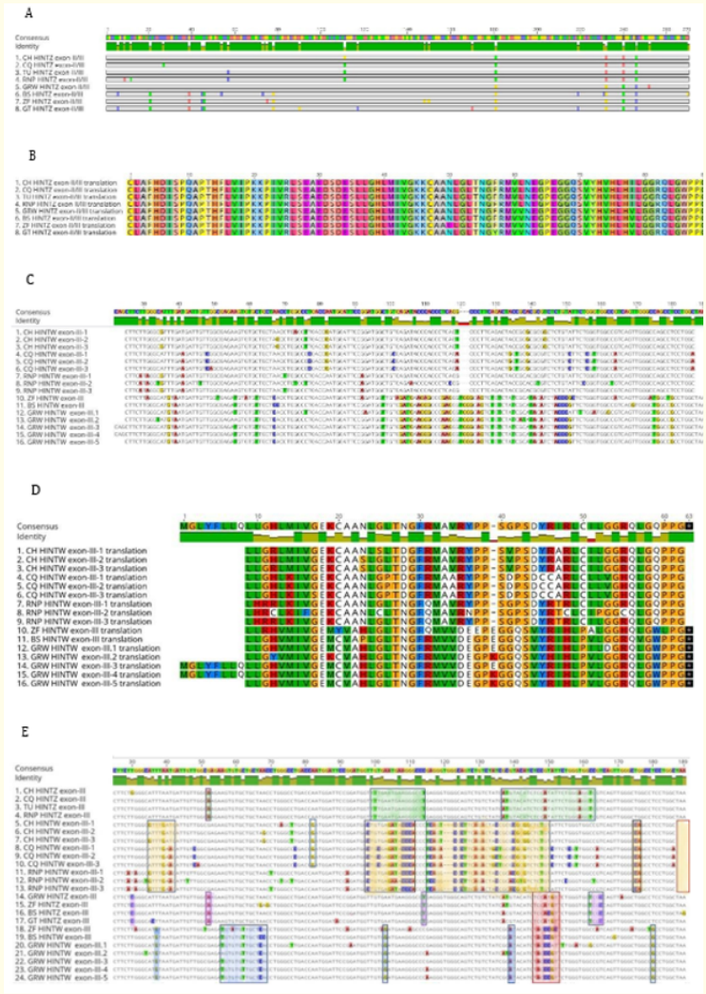

Figure 3: Alignment of HINTZ and HINTW in Galliformes and Passeriformes.

A) Sequence alignment for HINTZ copies, concatenated sequences for exon-II and exon-III region for Galliformes (1-4) and Passeriformes (4-8).

B) Translated alignment for HINTZ gene sequences for exon-II and exon-III regions concatenated.

C) Sequence alignment for HINTW copies for exon III region for Galliformes (1-9) and Passeriformes (9-16)

D) Translated alignment for HINTZ gene sequences for exon-III regions. All alignments were generated by Geneious Prime version 2022.1 created by Biomatters. Different numbered versions of certain sequences are result of different sequence variations of the gene obtained through GenBank and Genome Data Viewer databases.

E) Combined Sequence alignment for exon-III region of HINTZ and HINTW copies of Passeriformes and Galliformes. 1-4 depict galliform HINTZ ,5-13 galliform HINTW, 13-17 passeriform HINTZ and 18-24 Passeriform HINTW. Signatures of specific selection in the sequences have been highlighted using different color coding. Yellow represents selection specific to galliform HINTW, red represents Passerine specific, blue denotes specificity with Passerine HINTW, green color code used for denoting domain specific to Galliform HINTW while pink coloring represents regions specific to Passerine HINTZ.

Identification of several motifs specific to a group of sequences for analysis performed on exon- III region *HINTZ* and *HINTW* gene copies of Passeriformes and Galliformes are denoted in Figure 3E. *HINTW* gene copies from galliformes are found to differ greatly from their counterpart sequences (yellow color coding). The most notable difference observed was lack of TGG motif in *HINTW* copies in Galliformes at position 120-122 and a missing TAA motif towards the end of the sequence. It was observed that position 100-160 was found to depict high level of sequence diversity on combined analysis. Signatures of diversifying selection were found in this region for all the sequences.

Phylogenetic tree for *HINTW* and *HINTZ* sequences

The evolutionary history was inferred by using the Maximum Likelihood method and Tamura- Nei model [63]. The bootstrap consensus tree inferred from 500 replicates is taken to represent the evolutionary history of the taxa analysed. Branches corresponding

to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test of 500 replicates are shown next to the branches [59]. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 200.0000)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 0.00% sites).

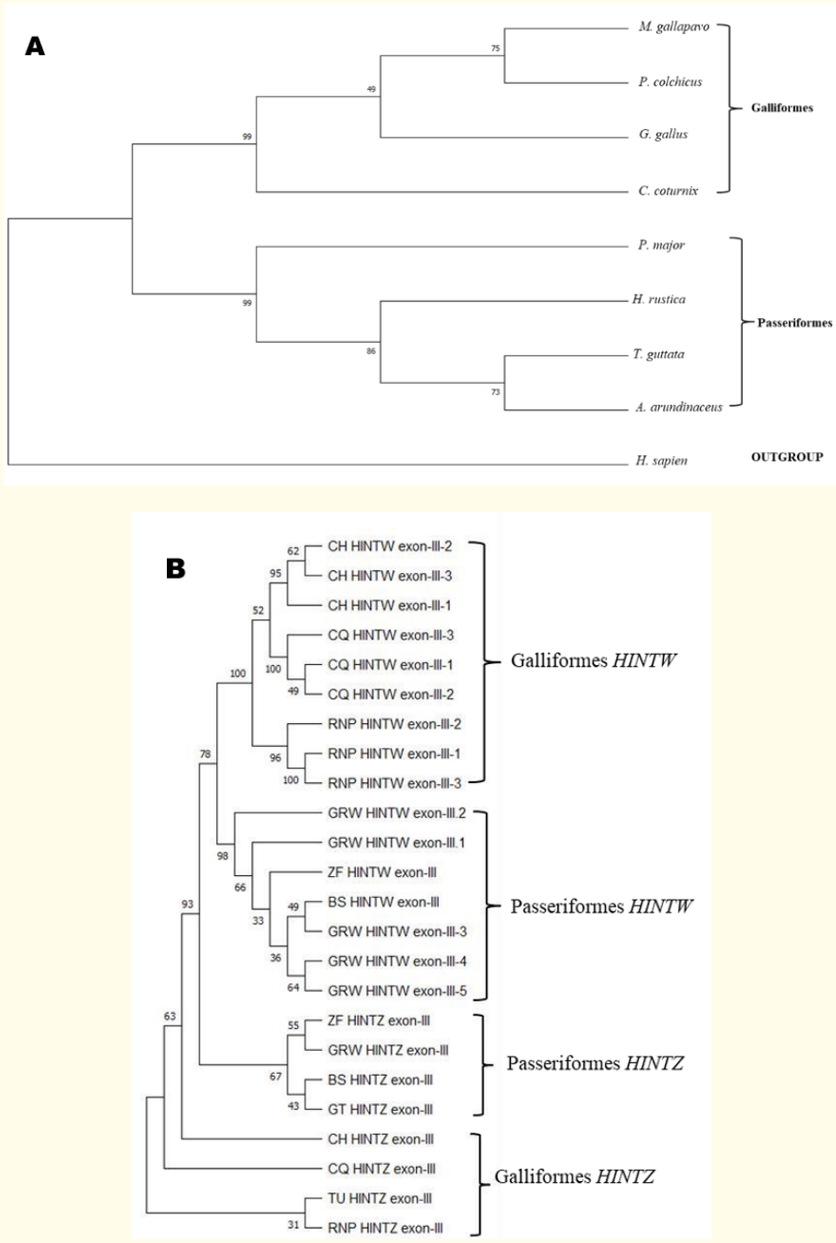

Figure 4: (A) Cladogram using for inferring evolutionary history of species using mitochondrial gene cyt-b, with human gene as outgroup. (B) Cladogram for inferring evolutionary history of HINTZ and HINTW genes in Passerines and Galliformes. Evolutionary history inferred using Maximum Likelihood method and Tamura-Nei model with MEGA11 software.

The cladograms depicted in Figure 4 provide insight on evolution of *HINTZ* and *HINTW* genes in reference to the evolutionary relationship in the respective species. The evolutionary analysis was performed using MEGA11 software, using mitochondrial gene cytochrome-b as basis of construction of phylogenetic tree on the species level (Tamura, Stecher and Kumar, 2021), with human gene as the specified outgroup for the species tree (Figure 4A). The result obtained showed a topology of the cytochrome-b gene tree that differed from the expected evolutionary relationship in case of *H. rustica* (BS) and *A. arundinaceus* (GRW). Both species are classified under the superfamily Sylvioidea and are more closely related to each other than to *T. guttata* (ZF). Therefore, it was expected that the gene tree would have shown clustering of barn swallow and great reed warbler, which was not the case (Figure 4A). A similar pattern was seen in case of *HINTZ* sequences as well (Figure 4B). However, the bootstrap values for these branches were low. For *HINTW*, it was noted that the single sequences for barn swallow and zebra finch clustered inside the five different sequences for the great reed warbler.

Finding optimal primer pair

The primers were designed in such a way so that the 3' end of the primer sequence was coinciding with variations in the sequence alignment for the species. This was done in order to increase specificity of the primers for both *HINTW* and *HINTZ* in the samples tested. The primer pairs were tested on samples from male and female individuals of great reed warbler, common whitethroat and Eurasian skylark to determine if amplification was successful for *HINTW* and *HINTZ* genes. Touchdown PCR thermocycler reaction was used to perform site- specific synthesis, followed by gel electrophoresis on 2 % agarose gel for 23 mins, 80 V, 500 mA. The reactions were performed on the same set of DNA samples using all possible primer combinations to determine the best pair of primers for both *HINTW* and *HINTZ*. Total primer combinations tested were 12 (6 each for *HINTW* and *HINTZ*). The expected result was that the primers should amplify regions, depicting band lying within 100-125 bp region. In case of *HINTW* primer sets, visible band should be seen for the female samples and not for the male ones, as the male (ZZ) doesn't possess W-chromosome which contains the *HINTW* gene. For *HINTZ* primer pairs, amplification should occur in both males and females (ZW).

Out of the 12 combinations tested, only 3 combinations were found to give satisfactory result: two pairs for *HINTW* and one for *HINTZ*.

The primer pair combinations which didn't produce any results might have been as a result of differences in the T_m of the primers, resulting in a non-compatible temperature profile to give an appropriate T_a for the PCR thermocycler reaction (Information about primer T_m in Appendix B).

GENE	Primer sequence (5'-3')
HINTW	F1: GGGCATGTAATGATTGTTGGCG R1: CGGGTAGATGTATGCGATAGAC
HINTW	F2-*HINTW*: GTTGGCGAGATGTGTGTTGCTC R2-*HINTW*: ACGGCCACCCAGAACGGGT
HINTZ	F1-*HINTZ*: GGGCACTTAATGATTGTTGGCA R1-*HINTZ*: GCGACCACCCAGAACATGG

Table 1: List of primers for amplification of exon-III in *HINTW* and *HINTZ.*

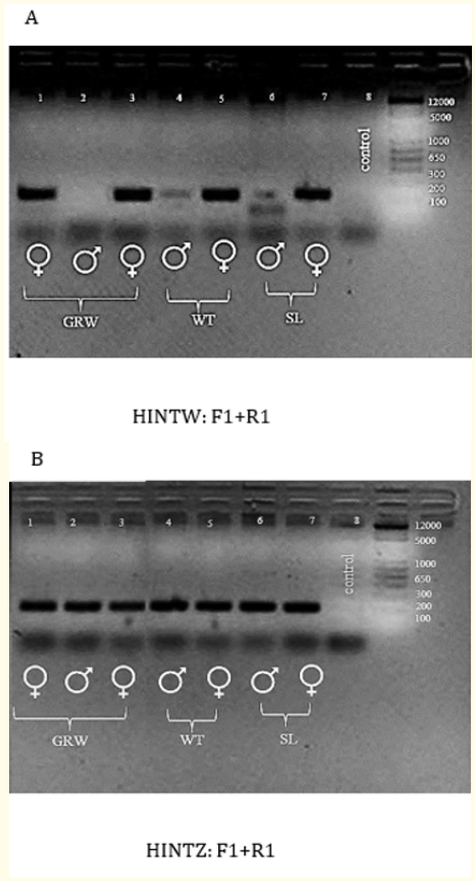

Figure 5: (A) Analysis of primer pairs F1-HINTW+R1-HINTW and (B) F1-HINTZ+R1-HINTZ DNA samples extracted from male and female individuals of passerines sample were amplified via PCR programme mentioned. Negative control with nuclease-free water was used to check for contamination. Six different primer combinations each were tested for both HINTZ and HINTW genes in three different passerines (Great reed warbler (GRW), Common whitethroat (WT), Eurasian Skylark (SL)) . Succesful amplification was seen in two combinations for HINTW and one combination for HINTZ. Amplification was observed only in female samples for HINTW as male genotype lacks a HINTZ copy.The product size in comparison to 1kb DNA ladder as reference was found to between 100-200 kb, which was as expected.

Gel electrophoresis on 2% agarose gel with Nucleic Acid GelRed Stain was performed and subsequently imaged using BIO-RAD Gel Doc XR+. Negative control with nuclease-free water was used for checking the purity of the samples (Lane 8). The images were optimised to give the best possible resolution possible. Figure 5A depicts results from primer combination F1- *HINTW* and R1-*HINTW.* The results for this primer combination were slightly uncertain as there were visible bands seen for male samples of common whitethroat and Eurasian skylark (faint bands in lane 4 and lane 6). However, as female samples of all the species had strong visible bands, this primer combination was considered for application in qPCR. The product size lies in the expected range of 100-150 bp in comparison to 1kb DNA ladder (Supplementary S4).

Figure 5B depicts the gel run for primer combination F1-*HINTZ* and R1-*HINTZ.* There were strong, visible bands seen for all the samples lying in the range of 100-200 kb product size. The results seen can be considered optimal to consider this primer pair as optimal for further analysis of *HINTZ* for Passeriformes.

Copy number analysis in Passerines

Primer pair gene amplification

The qPCR reaction was performed in order to test the primer efficiency in the DNA samples of species under study. MW, ERW, ARW from genus *Acrocephalus* were included in the study. The melting curve analysis and subsequent quantification was done to confirm the results of PCR reaction. The samples were tested in duplicates to obtain accurate readings and average Cq value used to calculate ΔCq. The data obtained was experimentally relevant as confirmed by R^2 value=0.995 and Eff_{GoI}=112.9% and $Eff_{control}$=99.8%.

Sample	Average Cq, *HINTW* (Target)	Average Cq,*HINTZ* (Control)	ΔCq (Cq, *HINTW* –Cq, *HINTZ*)	Copy number ($2^{-\Delta Cq}$)
GRW-1	21.685	24.85	-3.165	8.97
GRW-2	21.6	25.02	-3.42	10.70
GRW-3	20.61	24.6	-3.99	15.89
WT-1	22.17	25.22	-3.05	8.28
WT-2	21.04	24.305	-3.265	9.61
WT-3	21.11	25.02	-3.91	15.03
SL	22.25	25.12	-2.87	7.31
ARW	21.25	24.46	-3.19	9.19
ERW	21.35	24.545	-3.195	9.16
MW	21.025	24.42	-3.395	10.52

Table 2: The quantification data for qPCR reaction for finding out copy number variation between Passerine species.

The expected observation was that the gene amplification by *HINTW* primers should only be seen in the female samples of the respective species and the quantification cycle (Cq) value in case of *HINTW* (gene of interest) gene should be lower in comparison to *HINTZ* (control) gene. This is based on the assumption that *HINTW* is a multicopy gene as compared to the single copy *HINTZ* gene. For male samples, amplification by *HINTW* specific primers should not be seen due to lack of W-chromosome.

The results from the qPCR reaction were in accordance to the expected observations. No amplification was observed for male individual samples while testing for *HINTW*. The Cq value obtained for the female samples was lower in all the samples for *HINTW* specific amplification compared to that for *HINTZ* specific amplification (See Table 2). The copy number estimate in the three samples for great reed warbler varied greatly.

Copy number variation among species in superfamily Sylvioidea

Analysis on DNA samples of female individuals from different species belonging to superfamily Sylvioidea was performed to observe copy number variation for *HINTW* gene. The obtained values through qPCR experiment were reported as mean value with standard deviation among the species for each branch (Figure 6).

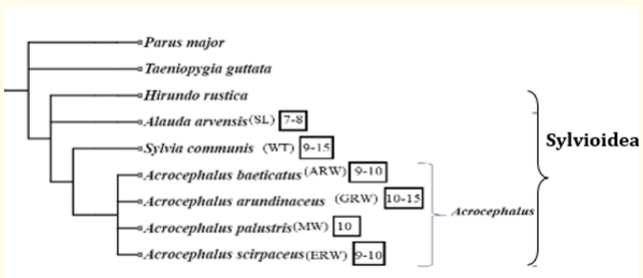

Figure 6: Phylogenetic tree for passerine constructed using iTOL online tree-building tool. The species name and copy number value observed via qPCR analysis and quantification with ΔCq values of the test sample using single-copy *HINTZ* gene as reference is annotated in front of the scientific name of the species according to taxonomical classification. Reported individuals with CNV are old-world songbirds classified into superfamily Sylvioidea and reported related value for copy number. GRW, ARW, MW, ERW belonging to genus Acrocephalus report similar values for CNV.

Species considered in the analysis include Great reed warbler, Marsh warbler, African reed warbler, Eurasian reed warbler, Common Whitethroat and Eurasian skylark. The number of individuals used in the analysis for the species was limited by the quantity of female DNA samples in the lab. Copy number for species with multiple individuals in the analysis was reported as mean value with standard deviation while test individuals have absolute value reported (Figure 6). GRW samples in the analysis were selected randomly from the pool of individuals.

Copy number variation in Great reed warblers

The results from preliminary analysis suggested that variation is observed in copy number of *HINTW* gene in great reed warblers. Two groups of GRW females were analyzed for copy number variation. The copy number among younger individuals aged T0 years (age 2 or 3 years) were compared in the respective individuals after 3 years (age 5 or 6). 4 of the 5 different individuals tested showed reduction in the number of *HINTW* gene copies between the younger and older individuals (Figure 7A). In individual GRW H8-37, a relative increase in the copy number was observed with age. The individual in question reported lowest value for copy number. In other individuals with lower copy number in older individuals, the reported average value was 9.35 ± 4.51 for the individuals aged T0 years and 5.78 ± 1.26 for the same set of individuals at 3 years older (Figure 7B).

The results obtained from the experiment were found to be uniform regardless of the family the individual belongs to in the taxonomical classification of species. The superfamily *Sylvioidea* reported to have 10.97 ± 3.17 value for *HINTW* gene copies. GRW, ERW, ARW, MW belonging to genus *Acrocephalus* were found to have mean copy number value 10.77 ± 3.24. The statistical significance was inferred using 2-way ANOVA (α=0.05) using parameters defined by GraphPad Prism 9.3.1 software. The p-values obtained for the samples analysed suggested the data was not significant.

Discussion

The aim of the study was to understand the sequence diversity in *HINTW* and *HINTZ* gene sequences in Galliformes and Passeriformes while understanding how the copy number varies in individuals closely related to GRW. In order to perform the desired analysis, set of primers were designed and certain pairs were found to work better than others in some other species as well. It has been discussed herewith about the reasons for putting forth a case for potential use of the primers in analysis of *HINTW* in passerine birds. The results reported were found to be in accordance to previously reported hypothesis seen in *HINTW* gene. The implications of the results and exploration of potential reasons underlying factors contributing to the variations in the observations for copy number variation for *HINTW* gene will be explored, elucidating distinction between related species.

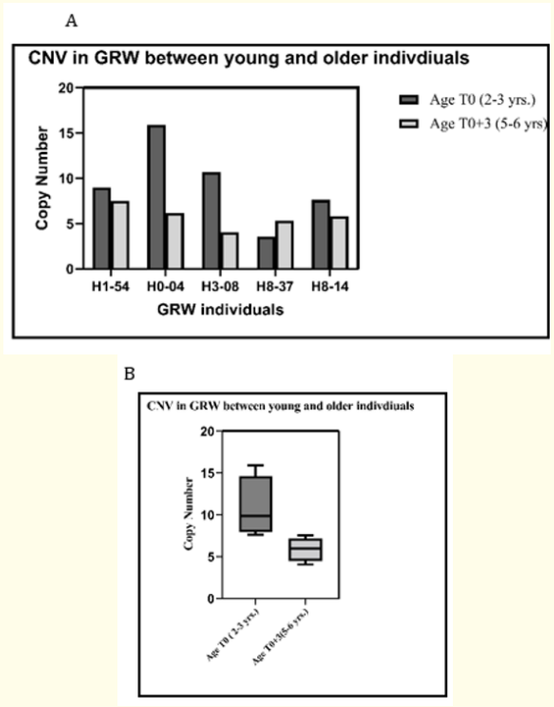

Figure 7: (A) Copy number variation in GRW individuals in two different age groups. DNA samples from five different GRW female individuals were tested for HINTW gene copies. Samples were classified into groups; age 2-3 years and age after 3 years (5-6 years old). Output from qPCR reaction was analysed using BIO-RAD CFX Maestro Software 1.1. Copy number was quantified using formula 2-ΔCq, where ΔCq indicates difference between Cq values of test and single-copy HINTZ as control gene. (B) CNV in the two age groups for all samples. Figure 7B depicts box-plot giving a combined analysis of CNV in both groups. Copy number was reported to decrease in four out of the five individual samples for older individuals as compared to the copy number observed at younger age. Average reported value for HINTW copies was 9.35 ± 4.51 for the individuals aged T0 years and 5.78 ± 1.26 for the same set of individuals at 3 years older. Graphical interpretation and statistical significance were done using GraphPad Prism 9.3.1 software.

Diversification in *HINTW* gene between Galliformes and Passeriformes

The combined sequence alignment performed on *HINTW* and *HINTZ* genes presents insight on the possible differentiation between the gene copies in Galliformes and Passeriformes. Whole gene alignments for both *HINTW* and *HINTZ* genes reported high levels of mismatches in the regions of exon-I and two introns. The region of exon-II and exon-III were found to be conserved to great extent for *HINTZ* genes in both orders. The conservation of the two exons is further supported by the observations from translational alignment of the sequences where despite certain substitutions observed in the alignment on sequence level, the translated sequences

showed high levels of similarity while comparing sequences from Galliformes and Passeriformes. The only observed difference was seen in Leucine and Isoleucine amino acids in Galliformes being substituted for Valine in Passeriformes (Figure 3B).

The *HINTW* gene present on the heterogametic chromosome undergoes series of degradation events and has been observed to exhibit female specific positive selection. From sequence analysis performed on *HINTW*, it was observed that a TGG region was absent in Galliformes but conserved within *HINTZ* sequences. From this observation, we hypothesized that absence in all of the Galliformes under study can be attributed to the evolutionary selection in the Galliformes occurring due to domestication events in the species classified under the taxa. The hypothetical reasoning can also be extended to the missing TAA/TAG region. However, the bases in question lies towards the end of the sequence. The sequence data for Passeriformes was obtained by alignment of similar *HINTW* copies from related species on available consortium of W-chromosome sequences lacking regions with sequenced bases. As a result, the missing bases can be attributed to discrepancies in sequence extraction.

Working primer pairs designed for *HINTW* and *HINTZ* gene copies in *Acrocephalus* species

The primers were designed using *HINTW* gene sequences from GRW, BS and ZF as reference and the region to be analyzed is subjected to high levels of sequence diversity even within individuals of the same species. Hence, there was a possibility that amplification of these regions via PCR reaction might not be as efficient in other test species included in the analysis. This was confirmed through the results from gel electrophoresis for the primer pairs depicted bands of varying intensity. As seen in primer pair F2+ R2-*HINTW*, the bands observed are solid for the GRW samples, but faint in case of WT and SL. Faint, visible bands are seen in male individuals for WT and SL samples for primer combination F1+R1-*HINTW*. These can be result of possible sequence similarities between *HINTZ* and *HINTW* gene copies for WT and SL. Among the possible 12 primer combinations tested, only 3 combinations were found to give satisfactory result. One of the possible reasons for the discrepancies can be posited as a result of the temperature profile for the PCR thermocycler reaction. The annealing temperature for each reaction was set 1-2 °C higher than the lowest Tm between the primer pair. The difference in Tm for the primer pairs in the reactions that didn't bear an outcome might be not have been optimal for efficient amplification of the required region in the DNA samples.

Irrespective of the concerns regarding specificity of the *HINTW* primers in WT and SL individuals for agarose gel electrophoresis experiment, the results from the qPCR reaction and subsequent melting curve analysis reported observations that remained true to the expected results. Amplification by *HINTW* gene specific primers designed using GRW, BS and ZF as the reference gene was observed in test samples from only female individuals in the warbler species (ARW, ERW, MW) belonging to genus *Acrocephalus*. This puts forth a case for the designed primer pairs to be used for analysis of *HINTW* and *HINTZ* gene copies in *Acrocephalus* species. Results also hold true for relatives of the GRW from the superfamily *Sylvioidea* when analysed for *HINTW* copies in case of whitethroat and skylark individuals. The performance of the selected primers in qPCR analysis for SL and WT might also suggest the use of the primers in superfamily *Sylvioidea*.

Copy number is conserved in species belonging to superfamily *Sylvioidea*

All genes located on the sex-determining locus on the chromosome are subjected to a considerably higher level of selection in comparison to those with copies on the other chromosomal regions. *HINTW* gene copies located on the female specific, non-recombing W- chromosome are further influenced by differential female-specific selection in individuals of different breeds of the same species based on reported observations pertaining to copy number variation. This has been documented in a study by Backstrom., *et al.* published in 2005, with evidence suggesting gene conversion (intrachromosomal recombination) leading to evolution among gene copies within species. Detailed analysis of ampliconic gene copies of *HINTW* in Galliformes had suggested that gene copies are independently amplified in different lineages or by being subjected to homogenisation mechanisms such as gene conversion or unequal crossing-over events [64].

The aim of the experiment was to understand copy number variation for *HINTW* gene in closely related passerine species. The study was performed on genomic DNA extracted from blood samples from female individuals for six species from superfamily *Sylvioidea*. The copy number of *HINTW* was reported to average ~9 -12 copies for most of the individuals analyzed. The reported copy number for all the species remained quite conserved for randomly selected samples. The copy number is observed to be more conserved in closely related species as seen with observations in whitethroats more uniform to species classified in genus *Acrocephalus* than the slightly distant relative in Skylarks.

The copy number in Passeriformes (9-15) is comparatively lower than that seen in Galliformes (~40). One possible reason to the discrepancy regarding reiteration frequency of *HINTW* in both orders can be different levels of female-specific selection for *HINTW* gene copies in the concerned species. Observations in Galliformes are reported in birds such as chicken, turkey and ducks starting from the embryo stage. These species are domesticated for importance in poultry farming activities such as egg-laying, where female-driven selection of individuals in a population is preferred. In comparison, songbirds have not reported to experience any such strong female-specific selection. Combined with degradation events on the female-specific W- chromosome, the decrease in the reported copy number might be lower in Passerines than in Galliformes.

Future studies focusing on expanding the knowledge regarding copy number variation of *HINTW* gene might help us understand this hypothesis slightly better.

HINTW gene copies reduce with age in Great reed warblers

Evidence provided by Hori., *et al.* 2016 supports the theory that the copies of *HINTW* gene decrease in number when compared between embryos and young individuals in Galliformes such as chicken and duck. This trait was observed in different breeds of the same species with varying copy number of the gene being reported. The study conducted involved a different mode of testing than the protocol used in my case. The quantification for copy number frequency for *HINTW* gene in my experiment involved comparison of ΔCq values in qPCR analysis in reference to single copy *HINTZ* gene. The estimates obtained through two methods used for analysis might result in discrepancies in results for the absolute copy number in the studied species. The observed difference in the estimated copies can possibly result from the difference in age groups of the study samples where the aforementioned study on Galliformes *HINTW* was reported for embryonic stage of the birds.

It was found that in four of the five samples tested for the two age groups studied, there was a considerable decrease in the copy number of the gene (~10-15 copies in younger, ~5-7 copies in older). The copy number decrease might not compare to the drastic decrease reported between embryo and matured stages of the species in Galliformes. It is however suggestive of the theory that cessation of recombination on the W chromosome results in continued loss of copies of *HINTW* gene through the lifetime of individual.

The opposition to the theory might arise from the fact that the sample size of test individuals analyzed to posit the assumption is quite low to be deemed significant. The non-significance of the data confirmed by the statistical analysis of the data is suggestive of the same. Another fact that might affect the observation is absence of documentation of copy number in different embryo stages and even nestlings. Existing possibility of point mutations in region of interest might cause certain fragments to elude amplification and subsequent analysis. Despite the arguments that can be presented, the studies can stand true if future studies expand the analysis on a larger sample size of test individuals. It also remains to be seen as to how the copy number would vary if observed between individuals at a higher age difference. The reported results might help to provide a case for further studies to be undertaken in understanding the copy number variation in *HINTW* and its implications on the characterization of female-specific traits in an individual, especially prompting the need to study the effect of parasitic infection.

Conclusion

The molecular mechanism which influences sex determination and gonadal differentiation in birds is still a subject of interest for many researchers (Estermann., *et al.* 2020; Ayers., *et al.* 2013; Graves, 2014). Identifying essential genes involved in the process of sex-determination and understanding their functioning and molecular mechanism is core intrigue. *HINTW* gene was looked to as a potential candidate due to similarities with testis-specific genes human Y-chromosome. While studies related to *HINTW* are well-documented in Galliformes such as chicken and ducks, a consensus for studies in other species was lacking. During the course of the project, primers designed were tested and found to work in gene copies in *Acrocephalus* species and were reported to work in other species of superfamily Sylvioidea. Evidence focusing on exon- III region of *HINTW* gene depicting diversification between *HINTZ* and *HINTW* copies in the two taxonomic orders of birds is presented. Copy number for species from genus *Acrocephalus* is reportedly conserved and was found to be identical in closely related species from superfamily Sylvioidea. Reduction in copy number of *HINTW* with age in great reed warbler individuals is seen. This puts emphasis on considering the age of the individual sampled in order to estimate the gene copy number in *HINTW*. The small sample size of the data calls for a future study encompassing individuals from related species to further understanding of the copy number variation in reproductive fitness of the species. Moreover, the effect of malarial infection on the reproductive success needs to be analyzed in a larger pool of samples to eventually determine its effect in the survival of the population.

Acknowledgement

I am grateful to my supervisor, Professor Bengt Hansson, for his expertise and guidance in completion of my research, I would like to thank faculty members as well as the PhD students at Department of Biology, Lund University, Sweden for their continued guidance and help during my time in the laboratory. I would like to extend my gratefulness towards ASMI for allowing me to share my research insights to a broader audience. Lastly, I would like to thank all the scholars and researchers for their literature. Without their efforts, it would not have been possible to complete my report.

Bibliography

1. Szathmáry E. "Toward major evolutionary transitions theory 2.0". P*roceedings of the National Academy of Sciences* 112.33 (2015): 10104-10111.

2. Gray J C and Goddard M R. "Gene-flow between niches facilitates local adaptation in sexual populations". *Ecology Letters* 15.9 (2012): 955-962.

3. van der Heide E M M., *et al*. "Sexual dimorphism in livestock species selected for economically important traits". *Journal of Animal Science* 94.9 (2016): 3684-3692.

4. Tran HT., *et al*. "An estrogen sensor for poultry sex sorting1". *Journal of Animal Science* 88.4 (2010): 1358-1364.

5. Olanrewaju H A., *et al*. "Effect of light intensity adjusted for species- specific spectral sensitivity on blood physiological variables of male broiler chickens". *Poultry Science* 98.3 (2019): 1090-1095.

6. Warner D A and Shine R. "The adaptive significance of temperature-dependent sex determination in a reptile". *Nature* 451.7178 (2008): 566-568.

7. Ospina-Álvarez N and Piferrer F. "Temperature-dependent sex determination in fish revisited: Prevalence, a single sex ratio response pattern, and possible effects of climate change". *PLoS ONE* 3.7 (2008): e2837.

8. Gilbert SF. Developmental Biology, ed. Sunderland (MA): Sinauer (2000).

9. Ghiselin M T. "The evolution of hermaphroditism among animals". *The Quarterly Review of Biology* 44.2 (1969): 189-208.

10. Edenbrow M and Croft D P. "Kin and familiarity influence association preferences and aggression in the mangrove killifish Kryptolebias marmoratus". *Journal of Fish Biology* 80.3 (2012): 503-518.

11. Bachtrog D. "Y-chromosome evolution: Emerging insights into processes of Y-chromosome degeneration". In Nature Reviews Genetics 14.2 (2013): 113-124.

12. Charlesworth B and Leon JA. "The Relation of Reproductive Effort to Age". *The American Naturalist* 110.973 (1976): 449-459.

13. Warren WC., *et al*. "Genome analysis of the platypus reveals unique signatures of evolution". *Nature* 453.7192 (2008): 175-183.

14. Gruetzner F., *et al*. "How did the platypus get its sex chromosome chain? A comparison of meiotic multiples and sex chromosomes in plants and animals". *Chromosoma* 115.2 (2006): 75- 88.

15. Deakin JE., *et al*. "The evolution of marsupial and monotreme chromosomes". *Cytogenetic and Genome Research* 137.2-4 (2012): 113-129.

16. Chue J and Smith CA. "Sex determination and sexual differentiation in the avian model". *In FEBS Journal* 278.7 (2011): 1027-1034.

17. Cui Z., *et al*. "High-density linkage mapping aided by transcriptomics documents ZW sex determination system in the Chinese mitten crab Eriocheir sinensis". *Heredity* 115.3 (2015): 206-215.

18. Jiang X H and Qiu G F. "Female-only sex-linked amplified fragment length polymorphism markers support ZW/ZZ sex determination in the giant freshwater prawn Macrobrachium rosenbergii". *Animal Genetics* 44.6 (2013): 782-785.

19. Parnes S., *et al*. "Sex determination in crayfish: are intersex Cherax quadricarinatus (Decapoda, Parastacidae) genetically females?". *Genetical Research* 82.2 (2003): 107-116.

20. Ezaz T., *et al*. "The ZW sex microchromosomes of an Australian dragon lizard share no homology with those of other reptiles or birds". *Chromosome Research : An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology* 17.8 (2009): 965-973.

21. Blackmon H., *et al*. "Sex Determination, Sex Chromosomes, and Karyotype Evolution in Insects". *The Journal of Heredity* 108.1 (2021): 78-93.

22. Ellegren H. "Evolution of the avian sex chromosomes and their role in sex determination". *Trends in Ecology and Evolution* 15.5 (2017): 188-192.

23. Skaletsky H., *et al*. "The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes". *Nature* 423.6942 (2003): 825-837.

24. Estermann M A., *et al*. "Insights into Gonadal Sex Differentiation Provided by Single-Cell Transcriptomics in the Chicken Embryo". *Cell Reports* 31.1 (2020): 107491.

25. Ayers K L., *et al*. "The molecular genetics of avian sex determination and its manipulation". *Genesis* 51.5 (2013): 325-336.

26. Graves J A M. "Avian sex, sex chromosomes, and dosage compensation in the age of genomics". *Chromosome Research* 22.1 (2014): 45-57.

27. Delbridge M L., *et al*. "TSPY, the candidate gonadoblastoma gene on the human Y chromosome, has a widely expressed homologue on the X-implications for Y chromosome evolution". *Chromosome Research* 12 (2004): 345-356.

28. Stiglec R., *et al*. "A new look at the evolution of avian sex chromosomes". *Cytogenetic and Genome Research* 117.1-4 (2007): 103-109.

29. Ellegren H. "Sex-chromosome evolution: Recent progress and the influence of male and female heterogamety". In Nature Reviews Genetics 12.3 (2011): 157-166.

30. Brockdorff N and Turner BM. "Dosage compensation in mammals". *Cold Spring Harbor Perspectives in Biology* 7.3 (2015).

31. DiNapoli L and Capel B. "SRY and the Standoff in Sex Determination". *Molecular Endocrinology* 22.1 (2021): 1-9.

32. Tomaszkiewicz M., *et al*. "Y and W Chromosome Assemblies: Approaches and Discoveries". In Trends in Genetics 33.4 (2017): 266-282.

33. Ioannidis J., *et al*. "Primary sex determination in birds depends on DMRT1 dosage, but gonadal sex does not determine adult secondary sex characteristics". 118.10 (2021): e2020909118.

34. Omotehara T., *et al*. "Spatiotemporal expression patterns of doublesex and mab-3 related transcription factor 1 in the chicken developing gonads and Müllerian ducts". *Poultry Science* 93.4 (2014): 953-958.

35. Charlesworth B., *et al*. "The evolution of restricted recombination and the accumulation of repeated DNA sequences". *Genetics* 112.4 (1986): 947-962.

36. White M A., *et al*. "Purifying selection maintains dosage-sensitive genes during degeneration of the threespine stickleback Y chromosome". *Molecular Biology and Evolution* 32.8 (2015): 1981-1995.

37. Brashear W A., *et al*. "Evolutionary conservation of Y Chromosome ampliconic gene families despite extensive structural variation". *Genome Research* 28.12 (2021): 1841-1851.

38. Sun C., *et al*. "Role and function of the Hintw in early sex differentiation in chicken (Gallus gallus) embryo". *Animal Biotechnology* (2021): 1-11.

39. Séraphin B. "The HIT protein family: a new family of proteins present in prokaryotes, yeast and mammals". *DNA Sequence: The Journal of DNA Sequencing and Mapping* 3.3 (1992): 177-179.

40. Brenner C., *et al*. "Crystal structures of HINT demonstrate that histidine triad proteins are GalT-related nucleotide-binding proteins NIH Public Access Author Manuscript". In Nat Struct Biol 4.3 (1997).

41. Hori T., *et al*. "Wpkci, Encoding an Altered Form of PKCI, Is Conserved Widely on the Avian W Chromosome and Expressed in Early Female Embryos: Implication of Its Role in Female Sex Determination". In Molecular Biology of the Cell 11.10 (2000): 3645-3660.

42. Smith CA., *et al*. "DMRT1 is upregulated in the gonads during female-to-male sex reversal in ZW chicken embryos". *Biology of Reproduction* 68.2 (2003): 560-570.

43. Rogers T F., *et al*. "Multi-copy gene family evolution on the avian w chromosome". In Journal of Heredity, 112.3 (2021): 250-259.

44. Backström N., *et al*. "Gene conversion drives the evolution of HINTW, an ampliconic gene on the female-specific avian W chromosome". *Molecular Biology and Evolution* 22.10 (2005): 1992-1999.

45. McCarroll S A and Altshuler DM. "Copy-number variation and association studies of human disease". *Nature Genetics* 39.7S (2007): S37-S42.

46. Sharp A J., *et al*. "Segmental Duplications and Copy- Number Variation in the Human Genome". *American Journal of Human Genetics* 77.1 (2005): 78-88.

47. Ceplitis H and Ellegren H. "Adaptive Molecular Evolution of HINTW, a Female-Specific Gene in Birds". *Molecular Biology and Evolution* 21.2 (2004): 249-254.

48. Brenner S E. "Errors in genome annotation". *Trends in Genetics* 15.4 (1999): 132-133.

49. Brenner C. "Hint, Fhit and GalT: Function, Structure, Evolution and Mechanism of Three Branches of the Histidine Triad Superfamily of Nucleotide Hydrolases and Transferases". In Biochemistry 41.29 (2002).

50. Handley LJ L., *et al*. "Evolutionary Strata on the Chicken Z Chromosome: Implications for Sex Chromosome Evolution". *Genetics* 167.1 (2004): 367-376.

51. Sánchez L and Chaouiya C. "Logical modelling uncovers developmental constraints for primary sex determination of chicken gonads". *Journal of the Royal Society Interface* 15.142 (2018).

52. Rogers L J., *et al*. "Advantages of having a lateralized brain". *Proceedings of the Royal Society B: Biological Sciences* 271 (2004): S420-S422.

53. Zhang Z., *et al*. "Whole-genome resequencing reveals signatures of selection and timing of duck domestication". *GigaScience* 7.4 (2018).

54. Fregin S., *et al*. "New insights into family relationships within the avian superfamily Sylvioidea (Passeriformes) based on seven molecular markers". *BMC Evolutionary Biology* 12.1 (2012): 1-12.

55. Sigeman H., *et al*. "Avian neo-sex chromosomes reveal dynamics of recombination suppression and W degeneration". *Molecular Biology and Evolution* 38.12 (2021): 5275-5291.

56. Letunic I and Bork, P. (n.d.). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation.

57. Schoch CL., *et al*. "NCBI Taxonomy: A comprehensive update on curation, resources and tools'. In Database (2020).

58. Nei M and Gojobori T. "Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions". *Molecular Biology And Evolution* 3.5 (1986): 418-426.

59. Felsenstein J. "CONFIDENCE LIMITS ON PHYLOGENIES: AN APPROACH USING THE BOOTSTRAP. Evolution". *International Journal of Organic Evolution* 39.4 (1985): 783-791.

60. Tamura K., *et al*. "MEGA11: Molecular Evolutionary Genetics Analysis Version 11". *Molecular Biology and Evolution* 38.7 (2021): 3022-3027.

61. Edgar R C. "MUSCLE: multiple sequence alignment with high accuracy and high throughput'. *Nucleic Acids Research* 32.5 (2004): 1792-1797.

62. Saitou N and Nei M. "The neighbor-joining method: a new method for reconstructing phylogenetic trees". *Molecular Biology and Evolution* 4.4 (1987).

63. Smith C A., *et al*. "Genetic evidence against a role for W-linked histidine triad nucleotide binding protein (HINTW) in avian sex determination". *The International Journal of Developmental Biology* 53.1 (2009): 59-67.

64. Warren W C., *et al*. "Genome analysis of the platypus reveals unique signatures of evolution". *Nature* 453.7192 (2008): 175-183.

65. O'Neill M., *et al*. "ASW: a gene with conserved avian W-linkage and female specific expression in chick embryonic gonad". *Development Genes and Evolution* 210.5 (2000): 243-249.