MULTI-COPY GENE FAMILY EVOLUTION ON THE AVIAN W CHROMOSOME

Thea F. Rogers[1*], Tommaso Pizzari[2] & Alison E. Wright[1*]

1 Department of Animal and Plant Sciences, University of Sheffield, United Kingdom

2 Edward Grey Institute, Department of Zoology, University of Oxford, United Kingdom

* Corresponding authors: tfrogers1@sheffield.ac.uk, a.e.wright@sheffield.ac.uk

Running title: Evolution of gene copy number on the W chromosome

Keywords: sex chromosomes, HINTW, Y chromosome, copy number evolution

ABSTRACT

The sex chromosomes often follow unusual evolutionary trajectories. In particular, the sex-limited Y and W chromosomes frequently exhibit a small but unusual gene content in numerous species, where many genes have undergone massive gene amplification. The reasons for this remain elusive with a number of recent studies implicating meiotic drive, sperm competition, genetic drift and gene conversion in the expansion of gene families. However, our understanding is primarily based on Y chromosome studies as few studies have systematically tested for copy number variation on W chromosomes. Here, we conduct a comprehensive investigation into the abundance, variability, and evolution of ampliconic genes on the avian W. First, we quantified gene copy number and variability across the duck W chromosome. We find a limited number of gene families as well as conservation in W-linked gene copy number across duck breeds, indicating that gene amplification may not be such a general feature of sex chromosome evolution as Y studies would initially suggest. Next, we investigate the evolution of HINTW, a prominent ampliconic gene family hypothesized to play a role in female reproduction and oogenesis. In particular, we investigate the factors driving the expansion of HINTW using contrasts between modern chicken and duck breeds selected for different female-specific selection regimes and their wild ancestors. Although we find the potential for selection related to fecundity in explaining small-scale gene amplification of HINTW in the chicken, purifying selection seems to be the dominant mode of evolution in the duck. Together, this challenges the assumption that HINTW is key for female fecundity across the avian phylogeny.

INTRODUCTION

Sex chromosomes are subject to unique evolutionary pressures due to their sex-limited inheritance and exhibit many unusual characteristics compared to the rest of the genome (Furman et al., 2020). They evolve when an autosome acquires a sex determining locus followed by halting of recombination between the sex chromosome pairs (Bergero & Charlesworth, 2009; Charlesworth, 1991). This recombination suppression triggers a cascade of neutral and adaptive processes that cause the once identical chromosomes to diverge from each other, often leading to the evolution of heteromorphic sex chromosomes (Bachtrog, 2013). These effects are most pronounced for the sex-limited Y and W chromosomes, which experience a reduction in the efficacy of selection, often resulting in rapid decay of gene content and activity due to processes such as Muller’s ratchet, the Hill-Robertson effect and genetic hitchhiking (Bachtrog, 2008; Bachtrog & Charlesworth, 2002; Charlesworth, 1978; Charlesworth & Charlesworth, 2000; Rice, 1996). In addition, because the Y and W chromosomes are haploid and only present in one sex, their effective population size is a fraction of that of the autosomes (Bachtrog & Charlesworth, 2002; Haddrill et al., 2007), making them more susceptible to genetic drift. Indeed, many Y chromosomes often consist of very few functional genes (Mank, 2012), however, intriguingly many of these genes have undergone massive gene amplification and persist as members of multi-copy gene families. For instance, the human Y chromosome harbours nine multi-copy ampliconic gene families which constitute the majority of protein-coding genes present on the Y (Skaletsky et al., 2003). Why these ampliconic gene families have evolved on heteromorphic sex chromosomes is an open question and their phenotypic consequences remain debated. It is also becoming increasingly apparent that copy number of these gene families can vary substantially, not only across closely related species but also individuals of the same species (Brashear et al., 2018; Lucotte et al., 2018; Poznik et al., 2016; Vegesna et al., 2019; Vegesna et al., 2020; Ye et al., 2018). Understanding the factors driving this variability can provide insight into the adaptability and functional importance of sex chromosomes more broadly.

It is widely assumed that the expansion of multi-copy ampliconic gene families is an adaptive response to lack of recombination between the sex chromosomes, where non-allelic homologous gene conversion between copies can escape Muller’s ratchet and the accumulation of deleterious mutations (Betrán et al., 2012; Charlesworth & Charlesworth, 2000; Connallon & Clark, 2010). Indeed, gene conversion appears to be a common feature of amplicons on both the Y and W chromosome across multiple species (Backström et al., 2005; Davis et al., 2010; Geraldes et al., 2010; Rozen et al., 2003; Skov et al., 2017). Furthermore, many Y amplicons are expressed exclusively within the testes (Mueller et al., 2008; Skaletsky et al., 2003; Vegesna et al., 2020) and often implicated in spermatogenesis and male fertility in humans (Kuroda-Kawaguchi et al., 2001; Lahn & Page, 1997; Vogt et al., 1996), leading to the hypothesis that selection on male fertility, often as a consequence of sperm competition, drives the expansion of multi-copy gene families. While there appears to be a positive relationship between copy number and expression level across some gene families (Vegesna et al., 2019), as well as with sperm mobility in humans (Yan et al., 2017), comparative approaches across species have failed to detect a significant correlation between copy number and intensity of sperm competition (Vegesna et al., 2020), although this may be due to the small number of species examined to date. Intriguingly, in several species there has been rapid co-amplification of genes on both sex chromosomes, suggestive of genomic conflict during gametogenesis to bias the transmission of the X versus Y (Bachtrog et al., 2019; Hughes et al. 2020; Soh et al., 2014). Detailed molecular analysis of the Sly and Slx gene families in the mouse provides strong support for antagonistic interactions and segregation distortion as a major force in driving gene amplification (Cocquet et al., 2012; Larson et al., 2018). Similarly, meiotic drive has been implicated in the evolution of gene families on the *Drosophila* Y chromosome (Bachtrog et al., 2019). Finally, many amplicons appear to be evolving under relaxed purifying selection, consistent with the reduced efficacy of selection on the non-recombining Y (Ghenu et al., 2016; Vegesna et al., 2020). Thus, while a myriad of forces have been implicated in the amplification of gene families on the Y and W chromosomes, the relative importance of each remains unclear.

To date, our understanding of multi-copy ampliconic gene families is primarily based on Y chromosome studies across mammals and *Drosophila*, and the W chromosome has been largely overlooked. Although the W is in many ways comparable to the Y chromosome, as both are sex-limited and do not recombine, the W is only present in females and the Y is only present in males. Therefore, the W chromosome, unlike the Y chromosome, does not experience sperm competition and might be subject to weaker sexual selection than the Y

(Bachtrog et al., 2011). Additionally, in polygynous mating systems where a small proportion of males in the population mate with multiple females, the effective population size of the Y relative to the autosomes is smaller than that of the W (Mank, 2012; Wright & Mank, 2013) As a result, the W chromosome may be less susceptible to genetic drift than the Y. Therefore, if multi-copy gene families are a consequence of random gene amplification due to genetic drift, they should be more pronounced on the Y chromosome rather than represent a general feature of heteromorphic sex chromosomes. It remains unclear whether W-linked amplicons have followed similar patterns of evolution to ampliconic genes on Y chromosomes, and whether gene amplification always occurs in parallel with sex chromosome degeneration.

A limited number of W-linked multi-copy gene families have been documented in a handful of species, primarily avian (Backström et al., 2005; Davis et al., 2010; Moghadam et al., 2012; Smeds et al., 2015; Zhou et al., 2020). The best studied is HINTW, an ampliconic gene family present on the avian W chromosome that is hypothesized to play a role in female reproduction and oogenesis (Ceplitis & Ellegren, 2004; O'Neill et al., 2000), and was originally proposed as the avian sex determining gene (Moriyama et al., 2006; Pace & Brenner, 2003; Parks et al., 2005). While an initial study of HINTW indicated that large scale amplification of copy number is conserved across avian non-ratites (Hori et al., 2000), a recent study suggested that HINTW is single-copy in the Pekin duck (Li et al., 2021). To date, there has been no comprehensive investigation into the abundance, variability, and evolution of multi-copy ampliconic gene families on the W chromosome both across and within species.

Here, we conduct a comparative analysis of copy number variation of W-linked genes across chicken and duck breeds. Multi-copy gene families are notoriously challenging to study due to their highly repetitive nature (Tomaszkiewicz et al., 2016). This problem is confounded on the sex-limited Y and W chromosomes where amplicons are often located in repeat-rich regions that are poorly annotated in reference genomes. We employ NanoString technology, which is based on fluorescent probes, to provide high-throughput fine-scale estimates of gene copy number and variability (Ahn et al., 2016; Cui et al., 2014). First, we quantify the frequency and variability of multi-copy gene families on the W across duck breeds, and find a limited number of amplicons on the duck W as well as conservation in copy number of W-linked genes. Next, we investigate the role of selection for fecundity in driving the amplification of HINTW using contrasts between chicken and duck breeds selected for either egg laying, male meat production or male plumage. We find that although large scale amplification of HINTW is ancestral to land and waterfowl species, smaller scale gene duplications have occurred independently across chicken breeds. Our results support a potential role of female-specific selection in driving amplification of the HINTW gene family in the chicken but not the duck, challenging the assumption that HINTW is key for female fecundity across the avian phylogeny.

MATERIALS & METHODS

Samples and DNA extraction

Our workflow is summarised in Figure S1. We obtained tissue samples from Khaki Campbell, Indian Runner, Aylesbury and Cayuga duck breeds and their modern ancestor, the Mallard duck (*Anas platyrhynchos*) (Zhang et al., 2018), and. In addition, we sampled the White Leghorn, Black Minorca, Oxford Old English and Black Sumatra chicken breeds and their main modern ancestor, the Red Junglefowl (*Gallus gallus*) (Frisby et al., 1979; Fumihito et al., 1996).

Samples were collected in accordance with national and ethical guidelines. Specifically, we obtained feathers from White Leghorn and Black Minorca. We also obtained 50 microlitres of Red Junglefowl blood in 1ml of absolute ethanol from a captive population at Oxford University (PPL P50402706). We obtained fertilised eggs from the following duck breeds; Mallard, Khaki Campbell, Cayuga, Aylesbury, Indian Runner, and the following chicken breeds; Oxford Old English and Black Sumatra. All eggs were kept under standard incubation conditions at The University of Sheffield. Samples were collected according to national and ethical guidelines and the liver was dissected at embryonic day 19 and 24 in chicken and duck breeds respectively, then stored in 95% ethanol.

DNA was extracted from feather and embryonic liver samples using DNeasy blood and tissue kit (QIAGEN) using standard protocols. DNA was extracted from blood samples using the ammonium acetate precipitation method. In total, DNA was obtained for three female and two male samples from each of the domesticated breeds, and two female and two male samples from each of the modern ancestor breeds. Embryonic birds were sexed visually and feather and blood samples were sexed using published sexing primers (Fridolfsson & Ellegren, 1999).

The majority of modern chicken breeds originated at the start of the 20th century (Rubin et al., 2010). Most modern chicken breeds are descended from the Red Junglefowl (Frisby et al., 1979; Fumihito et al., 1996) with some genes introgressed from the Grey Junglefowl and possibly other Junglefowl species (Eriksson et al., 2008). The Black Minorca and White Leghorn are layer breeds, which have been selected for female reproductive traits (e.g. fecundity), and the Oxford Old English and Black Sumatra chickens have been selected for male traits such as plumage for ornamentation purposes and aggression for cockfighting. The Oxford Old English and Black Sumatra lay fewer eggs than the two layer breeds and experience numerous female fecundity problems (Ekarius, 2007; Lewis, 2010). Importantly, the chicken breeds used in this study have independent origins (Moghadam et al., 2012) and so we can treat them as independent replicates of increased or relaxed female-specific selection. Most modern duck breeds are descended from the Mallard duck (Zhang et al., 2018). The Indian Runner and Khaki Campbell duck breeds have been subject to strong female-specific selection for egg laying, and the Aylesbury and the Cayuga for meat production (Ashton et al., 1999). Selection for meat- and egg-purpose breeds occurred at the early stages of duck domestication (Zhang et al., 2018) and so it is unclear whether the two layer breeds in our study can be considered independent replicates of increased female-specific selection.

Identification of W-linked genes

Previously, we identified 26 W-linked genes in the duck reference genome (Wright et al., 2014) using a combination of phylogenetic analyses and PCR validation in females. Some of these W genes share the same Z-linked ortholog, indicating they are either paralogs of a multi-copy gene family or fragments of the same gene, which have been assembled into separate genic sequences in the reference genome. Genome assemblies of sex chromosomes can be unreliable due to their repetitive nature and low sequencing coverage (Tomaszkiewicz et al., 2017) and so the latter scenario is plausible. To distinguish between these two scenarios, we aligned W-linked coding sequences with their Z-linked ortholog using PRANK (Löytynoja,

2014) and calculated pairwise distances. For the majority of cases, W-linked sequences shared no sequence similarity with each other, indicating they are fragments of the same gene that have been incorrectly assembled and annotated into separate genes. For subsequent analyses, we averaged data across fragments for these genes. Our results are quantitatively identical whether fragments are analysed separately or combined (see Supplementary Tables). The exception was KCMF1 in which the two annotated W sequences in the reference align and have a low pairwise distance, where the proportion of nucleotide differences was 0.091, suggesting these are paralogs of the same multigene family.

However, HINTW is not annotated in the duck reference genome and a previous study only identified a short fragment of sequence (Hori et al., 2000). Therefore, we sequenced a 702 bp fragment of HINTW in the Mallard using Sanger sequencing at the Core Genomic Facility, University of Sheffield with primers designed for the black oystercatcher (*Haematopus bachmani*) (Guzzetti et al., 2008). Primers are listed in Table S1.

For each PCR reaction the following volumes and concentrations of reagents were used: 4 ul multiplex PCR Master Mix (QIAGEN), 2 ul forward primer, 2 ul reverse primer (initial conc of each 0.2 uM) and 1 ul DNA (initial conc 15 ng/ul). In addition to this, 1 ul of nuclease free H_2O was added to reach a total volume of 10 ul per reaction. The PCR conditions were: initial denaturing stage of 95°C for 15 minutes, then 35 cycles of the following three steps; 94°C for 30 seconds, an annealing step at 57°C for 90 seconds, and an extension at 72°C for 90 seconds. This was then followed by a final extension at 72°C for 10 minutes.

Identification of autosomal invariant genes

The NanoString pipeline relies on the identification of invariant genes, autosomal single copy genes in that do not vary in copy number, as internal controls. We identified invariant genes in the duck and chicken separately using a genomic coverage approach. SOLiD DNA-seq data from nine chicken breeds were obtained from Rubin et al. (2010) and reads were aligned to the chicken reference genome (Gallus_Gallus-5.0, Zerbino et al., 2018) using SHRiMP v. 2.2.2 (Rumble et al., 2009). Mapped reads with a quality score of 10 or above were retained using SAMtools v. 1.8 (Li et al., 2009). Illumina DNA-seq reads from seven duck breeds (Zhang et al. 2018) were aligned to the duck reference genome (BGI_duck_1.0, Zerbino et al., 2018) using

BWA v. 0.5.7 (Li & Durbin, 2009) with the 'mem' algorithm. Read depth for each gene was calculated for both the chicken and the duck using the depth function in SAMtools. For each species, we conducted pairwise regressions of read depth per gene across every breed. We ranked residuals and identified genes in the lowest 35% quantile across all pairwise comparisons, indicative of limited or no copy number variation. We then used SNP data to test for nucleotide polymorphism across these genes, and we only called SNPs if the minor allele was present in one than one read. We chose genes with an absence of nucleotide polymorphism, and therefore an absence of multiple copies, as our invariant genes.

Quantification of gene copy number using NanoString

Copy number was quantified using the NanoString nCounter platform at the NERC Biomolecular Analysis facility (NBAF), University of Liverpool. NanoString nCounter technology uses fluorescent probes to estimate fine scale variation in gene copy number across samples (Ahn et al., 2016; Cui et al., 2014). Probes were designed for W-linked genes and invariant genes in the Red Jungle Fowl and Mallard duck separately in accordance with NanoString protocol (Table S2). Specifically, two or three probes were designed for HINTW in the chicken and 26 W-linked genes in the duck. One or two probes were designed for each invariant gene.

We implemented a number of controls to ensure copy number was quantified for only W-linked and not their Z-linked orthologs. Genome assemblies of sex chromosomes are often unreliable due to their repetitive nature and low sequencing coverage (Tomaszkiewicz et al., 2017) and therefore accurately identifying W-specific regions can be problematic. Furthermore, given that the Z and W chromosome evolved from the same pair of autosomes, certain regions of W-linked genes have high sequence similarity to their Z-linked gametolog (Wright et al., 2012). First, we designed probes to W-linked exons with low sequence similarity to Z-linked orthologs. Second, we included male samples in the CNV CodeSet analysis, making it possible to identify and exclude probes that bind to the Z chromosome.

The NanoString nCounter assay was performed according to standard protocol. Briefly, at least 300ng of DNA per sample was fragmented via AluI digestion and then hybridized to the custom CNV CodeSet. Samples included three females and two males from each of the selectively bred breeds, and two female and two male samples from each of the modern ancestor breeds. Samples were distributed randomly over the CNV CodeSets to avoid batch effects. The nCounter Digital Analyzer was used to count and quantify signals of reporter probes. Data analysis was performed using the nSolver Analysis Software.

We implemented a number of sanity checks as recommended by NanoString. First, we removed probes with count data above background noise in males and therefore affinity to the Z chromosome (Table S2). Background noise was calculated for each sample according to NanoString protocol as the average plus two standard deviations of the count number in the negative controls. We also removed one probe with count data below background noise in females, indicating low binding affinity. Second, as multiple probes were designed per W-linked gene, we calculated the coefficient of variation for copy number across probes. A high coefficient of variation is indicative of a probe that is not binding as predicted. As recommended by NanoString, we removed two probes from two different genes where the sum of the coefficient of variation across samples was >= 100 (Table S2). We averaged count data across all remaining probes of each gene in every individual.

Quantification of gene copy number from SNP data

We used polymorphism estimates from publicly available DNA-seq data to independently verify the results obtained from the NanoString nCounter assay in the Mallard duck. Given that we expect many gene copies to share identical sequences due to gene conversion (Backström et al., 2005), we can only use SNPs to estimate a minimum copy number.

Illumina data from nine unsexed Mallard ducks (Zhang et al., 2018) were quality trimmed to a minimum of 34 bp using Trimmomatic v. 0.36 (Bolger et al., 2014). Data were then aligned to the duck reference genome (BGI_duck_1.0, Zerbino et al., 2018), with the 702 bp sequenced fragment of HINTW added, using BWA v. 0.7.17 (Li & Durbin, 2009) with the 'aln' algorithm. Alignments were filtered for uniquely mapped reads by keeping only lines of the BAM files that matched the flag 'XT:A:U'. We used read coverage to sex individuals, where Z-linked genes should show half the number of reads in females relative to males. Read depth per gene was calculated using the depth function in SAMtools. To control for differences in overall sequencing depth between individuals we divided read depth on the Z chromosome by average autosomal read depth in each sample. Six females were identified and used in subsequent analyses.

BCFtools v. 1.9 (Narasimhan et al., 2016) was used to call SNPs at sites with a mapping quality > 20. In order to classify a SNP that indicated copy number variation, both the major and minor allele had to be supported by at least four reads and be present in more than half the individuals. Minor allele read depth was also required to be supported by at least 10% the number of reads that supported the major allele.

RESULTS

Copy number of genes on the Mallard W chromosome

We surveyed copy number of 26 genes on the Mallard duck W chromosome using count data obtained from NanoString nCounter. First, count data for W genes were normalised to invariant genes, autosomal genes present in a single copy, following NanoString protocol to account for any differences across samples in genomic DNA input arising from pipetting error or inaccuracies in DNA quantitation. Specifically, in each individual separately, we calculated average counts across all 10 invariant genes and bootstrapped with 1000 replicates to obtain the 95% confidence intervals. We divided the confidence intervals by two to account for comparisons between autosomal genes, which are present in two copies, and W-linked genes, which are present at a minimum of one copy. We then divided count number for each W gene by invariant count values to obtain estimates of W copy number in each individual and 95% confidence intervals.

In the Mallard duck, most W genes are present in a single copy. We found that HINTW is ampliconic, present in approximately 18 copies. This is in contrast to recent work suggesting that HINTW is single-copy in the Pekin duck (Li et al., 2021; Xu & Zhou, 2020). Furthermore, we found that KCMF1W is a multi-copy gene family present in 2 to 3 copies (Tables 1 & S3).

We independently verified copy number estimates using publicly available sequence data from Mallard individuals and nucleotide polymorphism analyses. No SNPs were found in any of the W genes with the exception of KCMF1W (ENSAPLG00000003106), where a single SNP was identified. This supports our finding that the majority of W-linked genes are present in a single copy in the Mallard. Although we verified that HINTW is ampliconic using NanoString data, we did not identify any nucleotide polymorphism across copies. This instead may indicate the occurrence of gene conversion across HINTW in the duck, which acts to homogenise gene sequence among variants, and is consistent with previous results in galliform birds (Backström et al., 2005).

Copy number variation across duck breeds

We used the same approach to estimate copy number of W-linked genes across the four duck breeds, with the exception of HINTW which we discuss separately below. Copy number was broadly conserved, as the majority of genes are present in a single copy across all breeds (Tables 1 & S3), with the exception of KCMF1W. This multi-copy gene family varies from 2 to 3 copies in some breeds to 3 to 4 copies in others, suggesting there may have been lineage-specific duplications in certain breeds (Tables S3 & S4).

In order to verify these results using a separate approach, we next estimated copy number in each breed relative to the Mallard duck. For each W-linked gene, normalised count data in each individual were divided by the average normalised count data for the Mallard to estimate relative copy number. We found that every W gene had a copy number ranging from 0.88 to 1.21 relative to the Mallard in all individuals, supporting our finding that there is limited copy number variation across duck breeds.

Finally, we estimated variation in copy numbers by calculating the coefficient of variance of raw count data across all individuals and breeds for each W-linked gene. Coefficient estimates ranged from 0.078 to 0.112 across individuals (Tables S5 & S6), and importantly no value exceeded the maximum coefficient of variation for invariant genes (mean COV = 0.131, max COV = 0.416), indicating limited variation in W-linked copy number. We repeated the analysis across breeds using average copy number in each breed and found a similar pattern, whereby coefficients of variation ranged from 0.043 to 0.106. No W gene exhibited higher variation across breeds than that observed across invariant genes (mean COV = 0.111, max COV = 0.356).

Copy number variation of ampliconic HINTW across duck and chicken breeds

Next, using contrasts between modern chicken and duck breeds selected for different female-specific selection regimes and their wild ancestors, we investigated the factors driving the expansion of HINTW. First, we estimated the size of the ampliconic HINTW gene family across duck breeds and found limited differences, where the number of copies ranged from 15 to 18 across individuals (Figure 1A, Tables S3 & S4). In addition, the coefficient of variance of HINTW count data across individuals (mean COV = 0.080) and breeds (mean COV = 0.043) was not higher than variation across invariant genes (Tables S5 & S6). Importantly, there is no significant difference in average copy number between breeds (ANOVA; p = 0.312). This suggests that copy number of HINTW is broadly conserved across duck individuals and breeds (Table S7), consistent with our predictions for purifying selection.

In contrast, we found notable variation in the size of the HINTW gene family across chicken breeds and individuals, ranging from 7 to 17 copies. The coefficient of variance for the chicken was 0.213 across individuals and 0.221 across breeds, both of which are higher than mean variation exhibited across invariant genes (mean COV = 0.151, max COV = 0.244 across individuals and mean COV = 0.116, max COV = 0.166 across breeds). Importantly, we found that the average size of HINTW gene family varied significantly between breeds (ANOVA; p = 0.001). Interestingly, all breeds have higher copy number of HINTW than the Red Junglefowl, and this was significant for three breeds (Figure 1B), indicating that the early domestication of chicken breeds may have been associated with a period of female-specific selection, presumably for egg laying. We find a general trend that breeds which have been selected for egg production via artificial female-specific selection (Kerje et al., 2003), had on average higher number of copies relative to breeds which have been bred for male fighting and plumage and subject to relaxed female-specific selection (Ekarius, 2007; Lewis, 2010) (Figure 1B). However, this relationship was only significant for the Black Minorca and not the White Leghorn (Table S8).

DISCUSSION

The sex-limited Y and W chromosomes exhibit a small but unusual gene content in many species compared to the rest of the genome. One striking feature is the existence of ampliconic gene families, arising from massive gene amplification of distinct classes of genes. Our understanding of how and why these ampliconic regions have evolved is primarily based on detailed Y chromosome studies across mammals and *Drosophila*, which have implicated a multitude of factors in the expansion of gene families, including meiotic drive, sperm competition, genetic drift and gene conversion (Bachtrog et al., 2019; Cocquet et al., 2012; Ellis et al., 2011; Ghenu et al., 2016; Good, 2012; Larson et al., 2018; Skaletsky et al., 2003; Soh et al., 2014; Vegesna et al., 2020). However, the evolution of multi-copy gene families on the W chromosome has been largely overlooked, with the exception of a handful of studies (Backström et al., 2005; Davis et al., 2010; Hori et al., 2000; Moghadam et al., 2012; Zhou et al., 2020). As a result, it remains unclear whether ampliconic genes are a fundamental feature of heteromorphic sex chromosome evolution or a peculiar quirk of Y chromosomes. Here, we conduct a comparative analysis to examine the abundance, variability, and evolution of ampliconic gene families on the avian W chromosome both across and within two avian species.

Our results show little evidence for gene amplification on the duck W chromosome. Of the 26 W-linked genes we studied, only two are present in multiple copies. One of these is HINTW, a large well-known ampliconic gene family, that has previously been characterized across a wide range of avian species (Backström et al., 2005; Hori et al., 2000). The fact that HINTW is ampliconic in the Mallard and four duck breeds is in contrast to recent work in the Pekin duck (Li et al., 2021; Xu & Zhou, 2020). Moreover, our finding that the W chromosome in the Mallard and domesticated duck breeds is generally depauperate in multi-copy gene families is consistent with a growing body of avian literature, including studies in the chicken (Moghadam et al., 2012), flycatcher (Smeds et al., 2015), sparrow (Davis et al., 2010), songbirds (Xu et al., 2019) and Pekin duck (Li et al., 2021). Outside of birds, to our knowledge, there is only one report of a W-linked ampliconic gene family in the willow *Salix purpurea* (Zhou et al., 2020), though few W chromosomes have been studied in sufficient detail. This deficit of gene families on the W is in stark contrast to the Y chromosome in mammals and *Drosophila*, where there has been massive amplification of gene sets.

This emerging pattern is consistent with theoretical predictions for how we expect the W to evolve differently to the Y due to their contrasting inheritance patterns (Bachtrog et al., 2014;

Mank, 2012). First, as the W chromosome is maternally inherited it is not subject to sperm competition, a factor which has been hypothesised, with mixed empirical support, to drive the expansion of Y-linked gene families (Hughes et al., 2010; Vegesna et al., 2020). It should be noted that the lack of support Vegesna et al. (2020) find for this hypothesis could be due to the small number of species examined in their study. Second, genetic drift is predicted to be weaker on the W in comparison to the Y chromosome. In polygynous mating systems, where a small proportion of males in the population mate with several females, the effective population size of the Y relative to the autosomes is smaller than that of the W (Mank, 2012; Wright & Mank, 2013). Relaxed purifying selection has been invoked to explain amplification of certain gene families on the primate and human Y chromosome, and the large variability in copy number across individuals and populations (Ghenu et al., 2016; Vegesna et al., 2020; Ye et al., 2018). Under drift, we expect variance in copy number to be approximately proportional to gene family size, where larger gene families will have a greater chance of gene duplication. Interestingly, we do not observe this pattern on the duck W chromosome where variability in the size of the HINTW gene family, present in ~18 copies, was similar to KFMC1, present in ~2 copies, across individuals and breeds. This is consistent with previous work showing evidence for purifying selection on the Mallard W (Wright et al., 2014).

Lastly, Y and W chromosomes are exposed to different types of gametogenesis, where the W is subject to oogenesis and the Y to spermatogenesis. Importantly, these contrasting environments likely lead to differences in the potential for antagonistic coevolution between the sex chromosomes. Antagonistic coevolution is predicted to drive the co-amplification of X and Y-linked genes (Bachtrog, 2020), but should be weaker during oogenesis than spermatogenesis. This is because the window for intragenomic conflict between chromosomes is restricted to the first meiotic division during oogenesis as only a single oocyte is produced containing either the Z or W (Bellott et al., 2017). Therefore, antagonistic coevolution between the Z and W will be limited to the first meiotic division. In contrast, competition between the X and Y can occur during meiosis I and II of spermatogenesis as both of these cell divisions produce viable gametes. As a result, we expect meiotic drive to play a less prominent role in the evolution of the W compared to the Y, and might explain why meiotic drive has been heavily implicated in the amplification of gene families on the mouse and *Drosophila* Y chromosomes (Bachtrog et al., 2019; Cocquet et al., 2012; Ellis et al., 2011; Good, 2012; Larson et al., 2018; Soh et al., 2014).

In addition, expression of the sex chromosomes is repressed during the post meiotic stages of spermatogenesis, leading to intragenomic conflict between X- and Y-linked genes over the transcriptional machinery and selection for gene amplification to maintain gene expression (Moretti et al., 2020). In contrast, no corresponding mechanism of sex chromosome repression in oogenesis has been reported thus far, and so we expect less co-amplification due to antagonistic coevolution in ZW systems. In support of these predictions, there is no evidence for co-amplification of HINTZ or KCMF1 on the avian Z chromosome (Bellott et al., 2010), indicating that antagonistic coevolution is unlikely to be a major factor influencing gene amplification on the W. Together, our results indicate that large scale expansions of gene families does not always occur in parallel with sex chromosome degeneration and so may not be such a general feature of sex chromosome evolution as Y studies would initially suggest.

Finally, as the W chromosome is maternally inherited it is not subject to sperm competition, a factor which has been hypothesised, with mixed empirical support, to drive the expansion of Y-linked gene families (Hughes et al., 2010; Vegesna et al., 2020). However, in theory, sex-specific selection for increased expression of genes associated with fecundity could drive amplification of gene families on the W chromosome, analogous to the hypothesised role of sperm competition on the Y chromosome (Hughes et al., 2005). In order to examine the factors driving the evolution of multi-copy gene families, we contrasted copy number of HINTW across breeds of the duck and chicken. Specifically, we chose breeds that have been subject to stronger or relaxed female-specific selection. In theory, sex-specific selection for increased expression of genes associated with fecundity could drive amplification of gene families. This seems particularly relevant for HINTW, which is expressed in the developing ovaries (O’Neill et al., 2000) and hypothesized to play a role in female reproduction (Ceplitis & Ellegren, 2004; O’Neill et al., 2000). Furthermore, increased copy number of Y-linked genes has been shown to result in greater gene expression level across primates, although this pattern is not universal across all gene families (Vegesna et al., 2019; Yan et al., 2017). However, in general there is uncertainty over whether the W chromosome is subject to female-specific selection, and is enriched for female reproductive functions (Moghadam et al., 2012), or subject to purifying selection for dosage effects (Bellott et al., 2017; Smeds et al., 2015; Xu et al., 2019; Xu & Zhou, 2020).

We find that HINTW copy number across duck breeds and individuals is remarkably conserved, in contrast to ampliconic gene families of equivalent size on the mammalian and *Drosophila* Y chromosomes (Bachtrog, 2013). We were unable to identify any sequence polymorphism across copies of HINTW, indicative of persistent gene conversion. While gene conversion is unlikely to explain the origin of multi-copy gene families, because it acts at a scale of a few hundreds of bases as opposed to a much larger scale of whole gene duplicates (Chen et al., 2007; Connallon & Clark, 2010; Marais et al., 2010), it has been proposed to select for the maintenance of ampliconic gene families and has been shown to operate across HINTW copies in a number of avian species (Backström et al., 2005). However, it is worth noting that the duck HINTW fragment in our study was only 702 bp, lowering the probability of finding a SNP in this gene and increasing our chances of inferring the action of gene conversion. Together, our results are inconsistent with the role of female-specific selection in driving the evolution of HINTW copy number in the duck. Instead, the conservation in copy number we observe across breeds suggests that HINTW copy number is under strong purifying selection. This is consistent with a number of recent studies showing that the avian W chromosome evolves predominantly under purifying selection to maintain ancestral gene dosage (Bellott et al. 2017; Bellott & Page, 2021; Smeds et al., 2015; Wright et al., 2014).

In contrast, in the chicken, we find notable variation in HINTW copy number across breeds. Breeds subject to female-specific selection tend to exhibit a greater number of HINTW copies This is consistent with the prediction that the chicken HINTW plays a role in female fecundity (Ceplitis & Ellegren, 2004; O'Neill et al., 2000). However, there is considerable variation in this trend, potentially indicating that female-limited selection is not the dominant force driving the evolution of HINTW.

The discrepancy between levels of variation in the size of the HINTW gene family in the chicken and duck is intriguing, particularly as large-scale gene amplification likely occurred in the ancestor of non-ratite birds (Hori et al., 2000). While evidence from the chicken indicates that HINTW plays a role in oogenesis (Ceplitis & Ellegren, 2004; O'Neill et al., 2000), evidence for functionality of HINTW in the duck is lacking. In fact, HINTW in the duck has been shown to lack the C-terminal 14 residues (Hori et al., 2000). HINTW forms a heterodimer with, and inhibits HINTZ in the chicken (Hori et al., 2000), and it is possible that the deletion in the duck has altered its ancestral functionality. Alternatively, it is possible that HINTW may have evolved differential gene expression across duck breeds without a corresponding increase in copy number. Consistent with this explanation, many W-linked genes have evolved increased expression in the chicken embryonic gonad in response to female-specific selection relative to the modern ancestor Red Junglefowl in the absence of copy number variation (Moghadam et al., 2012). It is also possible that the chicken has been subjected to stronger or more consistent sex-specific selection regimes than the duck, although evidence for this is currently lacking. Similarly, it is possible that the timing of domestication differs between the duck and chicken breeds in our study, or that there are differences in the extent of interbreeding. Although the exact breed history of chicken and ducks is obscure, evidence indicates that duck breeds selected for egg laying and meat production form two monophyletic groups that split early in duck domestication approximately 2200 years ago (Zhang et al., 2018). Therefore, we think that the lack of inter-breed copy number variation in the duck is unlikely to be a consequence of more recent origin or greater levels of interbreeding, although we cannot rule out this possibility.

In addition, we find that gene amplification has proceeded independently on the chicken or duck W chromosome (Van Tuinen & Hedges, 2001). When we contrast copy number estimates from previous work for the chicken (Moghadam et al., 2012) with our study, we find that W genes tend to duplicate independently, albeit at low copy number, in each species separately (Table 1). This suggests that the W is not an inert genetic wasteland but seems to evolve dynamically even after recombination was halted between the sex chromosomes.

Lastly, it is worth discussing the difficulties and limitations associated with studying copy number variation in ampliconic gene families. First, while our NanoString probe-based approach offers high-throughput fine-scale estimates of gene copy number and variability, we were not able to distinguish between functional and non-functional gene copies. This is particularly relevant for our conclusions surrounding the evolution of HINTW in the duck. Furthermore, it is not possible to detect gene copies with sequences that are substantially divergent from the probe sequences used. However, gene conversion should homogenise the sequence of gene copies, limiting the potential for this to confound our results. Finally, there is evidence that certain ampliconic genes on the Y are lineage-specific, for instance Sly and Slx are specific to the mouse lineage (Moretti et al., 2020). The list of W-linked genes we included in our analyses is not exhaustive (Wright et al., 2014) due to the challenges of sequencing sex chromosomes. Expanding the scope of this work to test whether lineage-specific loci are more likely to undergo massive scale amplification would be an interesting future avenue.

CONCLUDING REMARKS

Massive gene amplification is a characteristic feature of Y chromosome evolution. However, until now, it has remained unclear whether gene duplication is as prevalent on the W chromosome. We reveal that on the duck W chromosome, only two out of 26 W-linked genes show evidence of gene duplication. We hypothesise that this may be because genetic drift is reduced on the W relative to Y chromosomes, and we find limited variation of within-species gene copy number consistent with purifying selection. Contrary to this, we find some evidence that expansion of the HINTW gene family has evolved in response to female-specific selection for egg laying in the chicken but not the duck, calling into question the broad functionality of this prominent gene family. Taken together, our results suggest that in terms of gene duplication, the W chromosome follows a different evolutionary trajectory to that of the Y.

FUNDING

This work was supported by a Natural Environment Research Council (NERC) Independent Research Fellowship to A.E.W. (NE/N013948/1), a NERC ACCE PhD studentship to T.F.R., and an EECG Research Award from the American Genetics Association to A.E.W.

ACKNOWLEDGEMENTS

We thank the following for providing avian samples; Edward Boothman (White Leghorn), Richard Windsor (Black Minorca), David Hackett (Oxford Old English), Charles Hardcastle (Mallard), Graham Mortimer (Khaki Campbell, Indian Runner), and Richard Hedges and Paul

Hayes (Cayuga, Aylesbury, Black Sumatra). We thank Nicola Hemmings for helpful discussions on staging and incubating avian eggs, Geoff Parker for advice on chicken breeds and chicken breeders, The British Waterfowl Association and Rare Poultry Club for helping finding breeders, Judith Mank, Daniela Palmer and Peter Price for productive discussions. The authors acknowledge the use of the Sheffield HPC Cluster, and associated support services, in the completion of this work. Data generation was carried out by the Centre for Genomic Research, which is based at the University of Liverpool.

DATA AVAILABILITY

The raw data underlying these analyses is available in Dryad, DOI: https://doi.org/10.5061/dryad.18931zcw9.

REFERENCES

Ahn, S., Hong, M., Van Vrancken, M., Lyou, Y. J., Kim, S. T., Park, S. H., … Kim, K. M. (2016). A nCounter CNV assay to detect HER2 amplification: a correlation study with immunohistochemistry and in situ hybridization in advanced gastric cancer. *Molecular Diagnosis and Therapy*, *20*(4), 375–383.

Ashton, C., Ashton, M., & Donner, C. (1999). *British waterfowl standards*. British Waterfowl Association.

Bachtrog, D. (2008). The temporal dynamics of processes underlying Y chromosome degeneration. *Genetics*, *179*(3), 1513–1525.

Bachtrog, D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nature Reviews Genetics*, *14*(2), 113–124.

Bachtrog, D. (2020). The Y chromosome as a battleground for intragenomic conflict. *Trends in Genetics*, *36*(7), 510–522.

Bachtrog, D., & Charlesworth, B. (2002). Reduced adaptation of a non-recombining neo-Y chromosome. *Nature*, *416*(6878), 323–326.

Bachtrog, D., Kirkpatrick, M., Mank, J. E., McDaniel, S. F., Pires, J. C., Rice, W. R., & Valenzuela, N. (2011). Are all sex chromosomes created equal? *Trends in Genetics*, *27*(9), 350–357.

Bachtrog, D., Mahajan, S., & Bracewell, R. (2019). Massive gene amplification on a recently formed *Drosophila* Y chromosome. *Nature Ecology and Evolution*, *3*(11), 1587–1597.

Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T. L., … Vamosi, J. C. (2014). Sex determination: why so many ways of doing it? *PLoS Biology*, *12*(7), 1–13.

Backström, N., Ceplitis, H., Berlin, S., & Ellegren, H. (2005). Gene conversion drives the evolution of HINTW, an ampliconic gene on the female-specific avian W chromosome. *Molecular Biology and Evolution*, *22*(10), 1992–1999.

Bellott, D. W., & Page, D. C. (2021). Dosage-sensitive functions in embryonic development drove the survival of genes on sex-specific chromosomes in snakes, birds, and mammals. *Genome Research*, *31*(2), 198–210.

Bellott, D. W., Skaletsky, H., Cho, T. J., Brown, L., Locke, D., Chen, N., … Page, D. C. (2017). Avian W and mammalian Y chromosomes convergently retained dosage-sensitive regulators. *Nature Genetics*, *49*(3), 387–394.

Bergero, R., & Charlesworth, D. (2009). The evolution of restricted recombination in sex chromosomes. *Trends in Ecology and Evolution*. 24: 94-102

Betrán, E., Demuth, J. P., & Williford, A. (2012). Why chromosome palindromes? *International Journal of Evolutionary Biology*, 1–14.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.

Brashear, W. A., Raudsepp, T., & Murphy, W. J. (2018). Evolutionary conservation of Y chromosome ampliconic gene families despite extensive structural variation. *Genome Research*, *28*(12), 1826–1840.

Ceplitis, H., & Ellegren, H. (2004). Adaptive molecular evolution of HINTW, a female-specific gene in birds. *Molecular Biology and Evolution*, *21*(2), 249–254.

Charlesworth, B. (1978). Model for evolution of Y chromosomes and dosage compensation. *Proceedings of the National Academy of Sciences of the United States of America*, *75*(11), 5618–5622.

Charlesworth, B., & Charlesworth, D. (2000). The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *355*(1403), 1563–

Charlesworth, B. (1991). The evolution of sex chromosomes. *Science*, *251*(4997), 1030–1033.

Chen, J.-M., Cooper, D. N., Chuzhanova, N., Férec, C., & Patrinos, G. P. (2007). Gene conversion: mechanisms, evolution and human disease. *Nature Reviews Genetics*, *8*(10), 762–775.

Cocquet, J., Ellis, P. J. I., Mahadevaiah, S. K., Affara, N. A., Vaiman, D., & Burgoyne, P. S. (2012). A genetic basis for a postmeiotic X versus Y chromosome intragenomic conflict in the Mouse. *PLoS Genetics*, *8*(9), e1002900.

Connallon, T., & Clark, A. G. (2010). Gene duplication, gene conversion and the evolution of the Y chromosome. *Genetics*, *186*(1), 277–286.

Cui, W., Cai, Y., Wang, W., Liu, Z., Wei, P., Bi, R., … Zhou, X. (2014). Frequent copy number variations of PI3K/AKT pathway and aberrant protein expressions of PI3K subunits are associated with inferior survival in diffuse large B cell lymphoma. *Journal of Translational Medicine*, *12*(1), 1–11.

Davis, J. K., Thomas, P. J., & Thomas, J. W. (2010). A W-linked palindrome and gene conversion in New World sparrows and blackbirds. *Chromosome Research*, *18*(5), 543–553.

Ekarius, C. (2007). *Storey's Illustrated Guide to Poultry Breeds*. Marceline, MO: Walsworth Publishing Company.

Ellis, P. J. I., Bacon, J., & Affara, N. A. (2011). Association of Sly with sex-linked gene amplification during mouse evolution: A side effect of genomic conflict in spermatids? *Human Molecular Genetics*, *20*(15), 3010–3021.

Eriksson, J., Larson, G., Gunnarsson, U., Bed'hom, B., Tixier-Boichard, M., Strömstedt, L., … Andersson, L. (2008). Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genetics*, *4*(2).

Fridolfsson, A.-K., & Ellegren, H. (1999). A simple and universal method for molecular sexing of non-ratite birds. *Journal of Avian Biology*, *30*(1), 116.

Frisby, D. P., Weiss, R. A., Roussel, M., & Stehelin, D. (1979). The distribution of endogenous chicken retrovirus sequences in the DNA of galliform birds does not coincide with avian phylogenetic relationships. *Cell*, *17*(3), 623–634.

Fumihito, A., Miyake, T., Takada, M., Shingu, R., Endo, T., Gojobori, T., … Ohno, S. (1996). Monophyletic origin and unique dispersal patterns of domestic fowls. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(13), 6792–6795.

Furman, B. L. S., Metzger, D. C. H., Darolti, I., Wright, A. E., Sandkam, B. A., Almeida, P., … Fraser, B. (2020). Sex chromosome evolution: so many exceptions to the rules. *Genome Biology and Evolution*, *12*(6), 750–763.

Geraldes, A., Rambo, T., Wing, R. A., Ferrand, N., & Nachman, M. W. (2010). Extensive gene conversion drives the concerted evolution of paralogous copies of the SRY gene in European rabbits. *Molecular Biology and Evolution*, *27*(11), 2437–2440.

Ghenu, A. H., Bolker, B. M., Melnick, D. J., & Evans, B. J. (2016). Multicopy gene family evolution on primate Y chromosomes. *BMC Genomics*, *17*(1), 157.

Good, J. M. (2012). The conflict within and the escalating war between the sex chromosomes. *PLoS Genetics*, *8*(9), e1002955.

Guzzetti, B. M., Talbot, S. L., Tessler, D. F., Gill, V. A., & Murphy, E. C. (2008). Secrets in the eyes of Black Oystercatchers: a new sexing technique. *Journal of Field Ornithology*, *79*(2), 215–223.

Haddrill, P. R., Halligan, D. L., Tomaras, D., & Charlesworth, B. (2007). Reduced efficacy of selection in regions of the *Drosophila* genome that lack crossing over. *Genome Biology*, *8*(2), 1–9.

Hori, T., Asakawa, S., Itoh, Y., Shimizu, N., & Mizuno, S. (2000). *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. *Molecular Biology of the Cell*, *11*(10), 3645–3660.

Hughes, J. F., Skaletsky, H., Pyntikova, T., Minx, P. J., Graves, T., Rozen, S., …Page, D. C. (2005). Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. *Nature*, *437*(7055), 100-3.

Hughes, J. F., Skaletsky, H., Pyntikova, T., Graves, T. A., Van Daalen, S. K. M., Minx, P. J., … Page, D. C. (2010). Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature*, *463*(7280), 536–539.

Hughes, J. F., Skaletsky, H,. Pyntikova, T., Koutseva, N., Raudsepp, T., Brown, L. G., ... Page, D. C. (2020). Sequence analysis in *Bos taurus* reveals pervasiveness of X-Y arms races in mammalian lineages. *Genome Research, 30*, 1716–1726.

Kerje, S., Carlborg, Ö., Jacobsson, L., Schütz, K., Hartmann, C., Jensen, P., & Andersson, L. (2003). The twofold difference in adult size between the Red Junglefowl and White Leghorn chickens is largely explained by a limited number of QTLs. *Animal Genetics*, *34*(4), 264–274.

Kuroda-Kawaguchi, T., Skaletsky, H., Brown, L. G., Minx, P. J., Cordum, H. S., Waterston, R. H., … Page, D. C. (2001). The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nature Genetics*, *29*(3), 279–286.

Lahn, B. T., & Page, D. C. (1997). Functional coherence of the human Y chromosome. *Science*, *278*(5338), 675–680.

Larson, E. L., Kopania, E. E. K., & Good, J. M. (2018). Spermatogenesis and the evolution of mammalian sex chromosomes. *Trends in Genetics*, *34*(9), 722–732.

Lewis, C. (2010). *The illustrated guide to chickens: how to choose them - how to keep them*. Oswestry, UK: Scotprint.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Li, J., Zhang, J., Liu, J., Zhou, Y., Cai, C., Xu, L., … Zhou, Q. (2021). A new duck genome reveals conserved and convergently evolved chromosome architectures of birds and mammals. *GigaScience*, *10*(1), 1-15.

Löytynoja, A. (2014). Phylogeny-aware alignment with PRANK. *Methods in Molecular Biology*, *1079*, 155–170.

Lucotte, E. A., Skov, L., Jensen, J. M., Macià, M. C., Munch, K., & Schierup, M. H. (2018). Dynamic copy number evolution of X-and Y-linked ampliconic genes in human populations. *Genetics*, *209*(3), 907–920.

Mank, J. E. (2012). Small but mighty: The evolutionary dynamics of W and Y sex chromosomes. *Chromosome Research*, *20*(1), 21–33.

Marais, G. A. B., Campos, P. R. A., & Gordo, I. (2010). Can intra-Y gene conversion oppose the degeneration of the human Y chromosome? A simulation study. *Genome Biology and Evolution*, *2*(0), 347–357.

Moghadam, H. K., Pointer, M. A., Wright, A. E., Berlin, S., & Mank, J. E. (2012). W chromosome expression responds to female-specific selection. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(21), 8207–8211.

Moretti, C., Blanco, M., Ialy-Radio, C., Serrentino, M., Gobé, C., Friedman, R., … Cocquet, J.

(2020). Battle of the sex chromosomes: competition between X and Y chromosome-encoded proteins for partner interaction and chromatin occupancy drives multicopy gene expression and evolution in muroid rodents. *Molecular Biology and Evolution*, *37*(12), 3453-3468.

Moriyama, S., Ogihara, J., Kato, J., Hori, T., & Mizuno, S. (2006). PKCI-W forms a heterodimer with PKCI-Z and inhibits the biological activities of PKCI-Z in vitro, supporting the predicted role of PKCI-W in sex determination in birds. *Journal of Biochemistry*, *139*(1), 91–97.

Mueller, J. L., Mahadevaiah, S. K., Park, P. J., Warburton, P. E., Page, D. C., & Turner, J. M. A. (2008). The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nature Genetics*, *40*(6), 794–799.

Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., & Durbin, R. (2016). BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics*, *32*(11), 1749–1751.

O'Neill, M., Binder, M., Smith, C., Andrews, J., Reed, K., Smith, M., … Sinclair, A. (2000). ASW: a gene with conserved avian W-linkage and female specific expression in chick embryonic gonad. *Development Genes and Evolution*, *210*(5), 243–249.

Pace, H. C., & Brenner, C. (2003). Feminizing chicks: a model for avian sex determination based on titration of Hint enzyme activity and the predicted structure of an Asw-Hint heterodimer. *Genome Biology*, *4*(3), R18.

Parks, K. P., Seidle, H., Wright, N., Sperry, J. B., Bieganowski, P., Howitz, K., … Brenner, C. (2005). Altered specificity of Hint-W123Q supports a role for Hint inhibition by ASW in avian sex determination. *Physiological Genomics*, *20*(1), 12–14.

Poznik, G. D., Xue, Y., Mendez, F. L., Willems, T. F., Massaia, A., Wilson Sayres, M. A., … Tyler-Smith, C. (2016). Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nature Genetics*, *48*(6), 593–599.

Rice, W. R. (1996). Evolution of the Y sex chromosome in animals. *BioScience*, *46*(5), 331–

343

Rozen, S., Skaletsky, H., Marszalek, J. D., Minx, P. J., Cordum, H. S., Waterston, R. H., … Page, D. C. (2003). Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature*, *423*(6942), 873–876.

Rubin, C. J., Zody, M. C., Eriksson, J., Meadows, J. R. S., Sherwood, E., Webster, M. T., … Andersson, L. (2010). Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*, *464*(7288), 587–591.

Rumble, S. M., Lacroute, P., Dalca, A. V., Fiume, M., Sidow, A., & Brudno, M. (2009). SHRiMP: accurate mapping of short color-space reads. *PLoS Computational Biology*, *5*(5).

Skaletsky, H., Kuroda-Kawaguchl, T., Minx, P. J., Cordum, H. S., Hlllier, L. D., Brown, L. G., … Page, D. C. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature*, *423*(6942), 825–837.

Skov, L., Schierup, M. H., & Schierup, M. H. (2017). Analysis of 62 hybrid assembled human Y chromosomes exposes rapid structural changes and high rates of gene conversion. *PLoS Genetics*, *13*(8), e1006834.

Smeds, L., Warmuth, V., Bolivar, P., Uebbing, S., Burri, R., Suh, A., … Ellegren, H. (2015). Evolutionary analysis of the female-specific avian W chromosome. *Nature Communications*, *6*(1), 7330.

Soh, Y. Q. S., Alföldi, J., Pyntikova, T., Brown, L. G., Graves, T., Minx, P. J., … Page, D. C. (2014). Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell*, *159*(4), 800–813.

Tomaszkiewicz, M., Medvedev, P., & Makova, K. D. (2017). Y and W chromosome assemblies: approaches and discoveries. *Trends in Genetics*. 33: 266-282

Tomaszkiewicz, M., Rangavittal, S., Cechova, M., Sanchez, R. C., Fescemyer, H. W., Harris, R., … Makova, K. D. (2016). A time- and cost-effective strategy to sequence mammalian Y chromosomes: An application to the de novo assembly of gorilla Y. *Genome Research*, *26*(4), 530–540.

Van Tuinen, M., & Hedges, S. B. (2001). Calibration of avian molecular clocks. *Molecular Biology and Evolution*, *18*(2), 206–213.

Vegesna, R., Tomaszkiewicz, M., Medvedev, P., & Makova, K. D. (2019). Dosage regulation, and variation in gene expression and copy number of human Y chromosome ampliconic genes. *PLoS Genetics*, *15*(9), e1008369.

Vegesna, R., Tomaszkiewicz, M., Ryder, O. A., Campos-Sánchez, R., Medvedev, P., DeGiorgio, M., & Makova, K. D. (2020). Ampliconic genes on the great ape Y chromosomes: Rapid evolution of copy number but conservation of expression levels. *Genome Biology and Evolution*, *12*(6), 842–859.

Vogt, P. H., Edelmann, A., Kirsch, S., Henegariu, O., Hirschmann, P., Kiesewetter, F., … Haidl, G. (1996). Human Y chromosome azoospermia factors (AZF) mapped to different subregions in Yq11. *Human Molecular Genetics*, *5*(7), 933–943.

Wright, A. E., & Mank, J. E. (2013). The scope and strength of sex-specific selection in genome evolution. *Journal of Evolutionary Biology*, *26*(9), 1841–1853.

Wright, A. E., Harrison, P. W., Montgomery, S. H., Pointer, M. A., & Mank, J. E. (2014). Independent stratum formation on the avian sex chromosomes reveals inter-chromosomal gene conversion and predominance of purifying selection on the W chromosome. *Evolution*, *68*(11), 3281–3295.

Wright, A.E., Moghadam, H. K., & Mank, J. E. (2012). Trade-off between selection for dosage compensation and masculinization on the avian Z chromosome. *Genetics*, *192*(4), 1433–1445.

Xu, L., Auer, G., Peona, V., Suh, A., Deng, Y., Feng, S., … Zhou, Q. (2019). Dynamic evolutionary history and gene content of sex chromosomes across diverse songbirds. *Nature Ecology and Evolution*, *3*(5), 834–844.

Xu, L., & Zhou, Q. (2020). The female-specific W chromosomes of birds have conserved gene contents but are not feminized. *Genes*, *11*(10), 1–14.

Yan, Y., Yang, X., Liu, Y., Shen, Y., Tu, W., Dong, Q., … Yang, Y. (2017). Copy number variation of functional RBMY1 is associated with sperm motility: an azoospermia factor-linked candidate for asthenozoospermia. *Human Reproduction*, *32*(7), 1521–1531.

Ye, D., Zaidi, A. A., Tomaszkiewicz, M., Anthony, K., Liebowitz, C., DeGiorgio, M., … Makova, K. D. (2018). High levels of copy number variation of ampliconic genes across major human Y haplogroups. *Genome Biology and Evolution*, *10*(5), 1333–1350.

Zerbino, D. R., Achuthan, P., Akanni, W., Amode, M. R., Barrell, D., Bhai, J., … Flicek, P. (2018). Ensembl 2018. *Nucleic Acids Research*, *46*(D1), D754–D761.

Zhang, Z., Jia, Y., Almeida, P., Mank, J. E., van Tuinen, M., Wang, Q., … Qu, L. (2018). Whole-genome resequencing reveals signatures of selection and timing of duck domestication. *GigaScience*, *7*(4), 1–11.

Zhou, R., Macaya-Sanz, D., Carlson, C. H., Schmutz, J., Jenkins, J. W., Kudrna, D., … Difazio, S. P. (2020). A willow sex chromosome reveals convergent evolution of complex palindromic repeats. *Genome Biology*, *21*(1), 38.

Table 1. Copy number of W-linked genes across duck breeds.

		Average copy number							
Gene name	**Duck Ensembl ID**	**Mallard**	**Caguya**	**Aylesbury**	**Indian Runner**	**Khaki Campbell**	**Δ Copy number**	**Coefficient of variation**	**Stratum^**
HINTW$^+$	NA	18.03	16.35	16.57	17.22	16.83	1.68	0.04	1
CHD1W*	ENSAPLG05191 ENSAPLG02506	0.94	0.97	0.98	0.99	0.98	0.05	0.09	1
KCMF1W	ENSAPLG03026 ENSAPLG03106	2.43	2.59	2.65	2.63	2.63	0.20	0.10	2
RASA1W	ENSAPLG05611 ENSAPLG10611 ENSAPLG10611	0.64	0.69	0.70	0.69	0.70	0.06	0.11	2
ATP5A1W$^+$	ENSAPLG09007	0.82	0.79	0.84	0.85	0.83	0.06	0.07	3
BTF3W	ENSAPLG04652	0.65	0.60	0.64	0.65	0.65	0.05	0.06	3
HNRPKW*$^+$	ENSAPLG10986	0.97	1.00	1.09	1.02	1.05	0.12	0.11	3
MIER3W$^+$	ENSAPLG10850	0.62	0.61	0.65	0.62	0.63	0.04	0.08	3
NIPBLW	ENSAPLG02953 ENSAPLG03022 ENSAPLG05315 ENSAPLG10290 ENSAPLG10560	0.67	0.69	0.71	0.69	0.70	0.04	0.09	3
SMAD2W	ENSAPLG04964	0.69	0.72	0.74	0.71	0.71	0.05	0.09	3
SPIN1W*	ENSAPLG02923	0.63	0.61	0.64	0.66	0.64	0.05	0.08	3
UBAP2W	ENSAPLG16004 ENSAPLG16155	0.61	0.58	0.59	0.60	0.60	0.03	0.06	3
UBE2R2W	ENSAPLG16000	0.76	0.74	0.76	0.78	0.76	0.04	0.07	3
VCPW$^+$	ENSAPLG05806	0.91	0.84	0.90	0.91	0.90	0.07	0.06	3
ZFRW*	ENSAPLG15519	0.67	0.68	0.69	0.70	0.68	0.03	0.08	3
ZSWIM6W	ENSAPLG13555 ENSAPLG14338	0.77	0.80	0.80	0.83	0.80	0.06	0.09	3

* q-PCR analysis showed variation in copy number of ortholog across chicken breeds (Moghadam *et al.* 2012)
+ SNP analysis showed chicken ortholog is multicopy (Moghadam *et al.* 2012)
^ Anseriform strata as defined by Wright *et al.* 2014 Evolution. Strata 1 & 2 are conserved in chicken and duck but Stratum 3 evolved independently.
Note: six zeros have been removed from start of the digits in the Ensembl IDs.

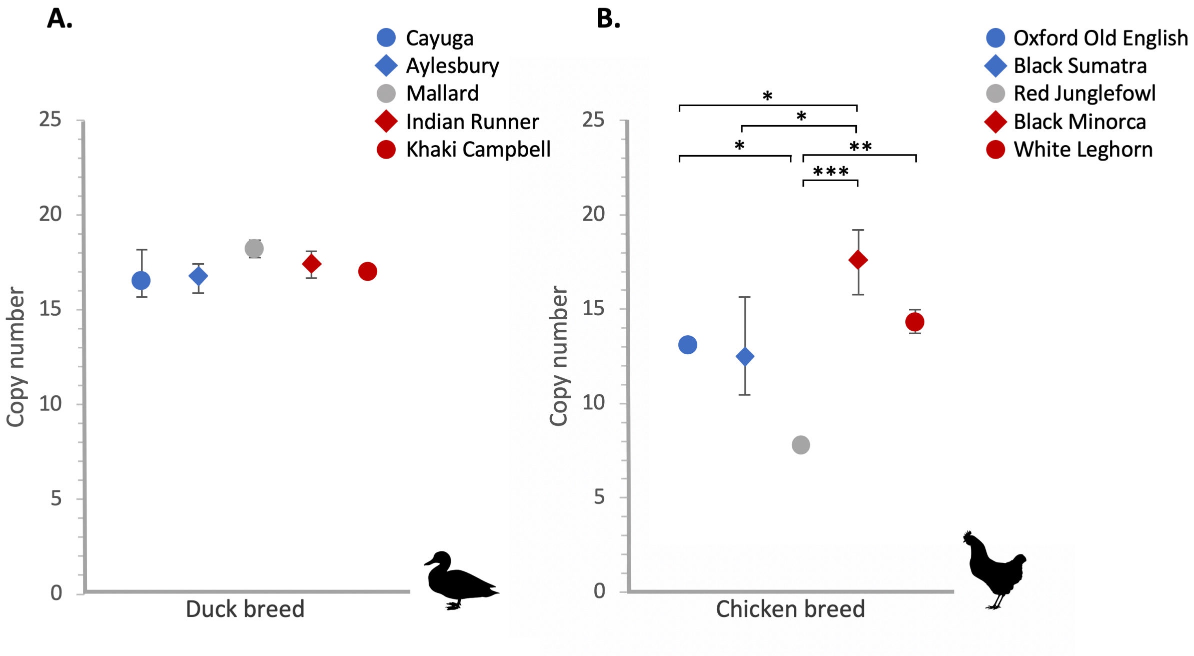

Figure 1. Copy number variation of HINTW across (A) duck and (B) chicken breeds. Copy number was estimated using the NanoString nCounter platform. Each circle or diamond represents the mean HINTW copy number per breed, and bars show the range of HINTW copy number across individuals. Blue markers represent breeds subject to relaxed female-specific selection, red markers represent female-selected breeds, and grey markers denote the modern ancestor for each bird species. Stars indicate pairwise significance values from Tukey multiple comparisons of means where * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$.