Published in final edited form as:
Chromosome Res. 2010 July ; 18(5): 543–553. doi:10.1007/s10577-010-9134-y.

A W-linked palindrome and gene conversion in New World sparrows and blackbirds

Jamie K. Davis,
Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322, USA

Pamela J. Thomas,
Genome Technology Branch and NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

NISC Comparative Sequencing Program, and
Genome Technology Branch and NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA

James W. Thomas
Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322, USA

James W. Thomas: jthomas@genetics.emory.edu

Abstract

A hallmark feature of the male-specific region of the human Y chromosome is the presence of large and near-identical palindromes. These palindromes are maintained in a state of near identity via gene conversion between the arms of the palindrome, and both neutral and selection-based theories have been proposed to explain their enrichment on the human Y and X chromosomes. While those proposed theories would be applicable to sex chromosomes in other species, it has not been established whether near-identical palindromes are a common feature of sex chromosomes in a broader range of taxa, including other tetrapods. Here, we report the genomic sequencing and features of a 279-kb region of the non-recombining portion of the W chromosome spanning the *CHD1W* locus in a New World sparrow, the white-throated sparrow (*Zonotrichia albicollis*), and the corresponding region on the Z chromosome. As has been observed for other Y and W chromosomes, we detected a high repetitive element content (51%) and low gene content on the white-throated sparrow W chromosome. In addition, we identified a 22-kb near-identical (>99%) palindrome on the W chromosome that flanks the 5′ end of the *CHD1W* gene. Signatures of gene conversion were readily detected between the arms of this palindrome, as was the presence of this palindrome in other New World sparrows and blackbirds. Near-identical palindromes are therefore present on the avian W chromosome and may persist due to the same forces proposed for the enrichment of these elements on the human sex chromosomes.

Keywords

palindrome; gene conversion; sex chromosome

Introduction

The evolution and structure of male-specific Y and female-specific W sex chromosomes have been of considerable interest and studied in a broad range of taxa (for example, see

Correspondence to: James W. Thomas, jthomas@genetics.emory.edu.

Ohno 1967; Charlesworth et al. 2005; Marshall Graves 2008; Wilson and Makova 2009). Compared to the rest of the genome, the Y and W chromosomes are unique in that they include a sex-specific non-recombining chromosomal segment that displays signatures of genetic degeneration, including high repetitive element content and low gene density. Amongst tetrapod Y and W chromosomes, the human Y chromosome is the best characterized. Like other Y (W) chromosomes, the euchromatic male-specific region of the human Y chromosome (MSY) encodes very few functional genes and is enriched in repetitive elements (Skaletsky et al. 2003).

A particularly striking feature of the human (and chimpanzee) MSY is the prevalence of large, near-identical palindromes which comprise at least 25% of the MSY in those species (Skaletsky et al. 2003; Hughes et al. 2010). While near-identical palindromes (also referred to here as inverted repeats) are found throughout the human genome, these structures are enriched on the MSY and the X chromosome (Warburton et al. 2004). The over-representation of near-identical palindromes on the X and Y chromosomes strongly suggests that these sequence features functionally contribute to and/or are neutral by-products of aspects of genome biology restricted to the sex chromosomes. For example, in the eutherian male germ line, sex chromosomes are subject to meiotic sex chromosome inactivation (MSCI), and it has been proposed that the secondary structures (cruciforms) formed by the intrastrand pairing of near-identical palindrome arms might enable some X- and Y-linked genes to escape MSCI (Skaletsky et al. 2003; Warburton et al. 2004). Gene conversion between the palindrome arms has been shown to be prevalent on the ape Y (Rozen et al. 2003; Hughes et al. 2010) and X chromosomes (Warburton et al. 2004; Bagnall et al. 2005) and can maintain the palindrome arms in a state of near identity for extended periods of time (Caceres et al. 2007). Gene conversion between palindrome arms on the Y chromosome has been proposed to be a mechanism by which the Y chromosome can compensate for its lack of normal homologous recombination and stave off the genetic degradation that is inevitable on a non-recombining chromosome (Skaletsky et al. 2003). An alternative model is that the high frequency of intra-palindrome arm gene conversion on both the X and Y chromosomes is simply a neutral by-product of their position in the genome. That is, due to the limited homologous pairing of the X and Y chromosomes during male meiosis, the inverted repeats on the sex chromosomes may be in a more optimal environment for intrastrand pairing, and therefore gene conversion, than if they were located on an autosome (Rossiter et al. 1994). Although sex-linked inverted repeats may have been maintained in the genome due to important biological functions, it is also true that crossovers resulting from non-allelic homologous recombination between palindrome arms can cause pathogenic inversions (Lakich et al. 1993) or isodicentric/acentric chromosomal abnormalities (Lange et al. 2009).

Another important model for understanding the evolution of sex chromosomes in tetrapods is the avian Z/W system. In birds, unlike mammals, females are the heterogametic sex. Thus, studies of the avian female-specific W have the potential to provide valuable points of comparison for evaluating the evolutionary history and genomic features associated with the eutherian Y chromosome. For example, reconstruction of the history of the eutherian and avian sex chromosomes have revealed evolutionary strata that are consistent with XY and ZW recombination having independently ceased in a step-wise process over the course of the past ~150 million years (Lahn and Page 1999; Ellegren and Carmichael 2001; Skaletsky et al. 2003; Marshall Graves 2008; Nam and Ellegren 2008). As a consequence of genetic degradation caused by the lack of recombination in the female-specific segment of the W chromosome, like eutherian Y chromosomes, most avian W chromosomes are physically smaller than the Z chromosome (reviewed in Mank and Ellegren 2007). Another established characteristic of avian W chromosomes is the presence of rapidly evolving repetitive elements (Itoh et al. 2008 and references therein). Unfortunately, only a limited amount of contiguous genomic sequence from avian W chromosomes has been reported (Hillier et al.

2004; Itoh et al. 2008, 2009), and as a result, little else is known about the sequence features and content of the avian W chromosome. Here, we report the targeted physical mapping and sequencing of segments of the Z and W chromosomes containing the *CHD1Z* and *CHD1W* gametologs (located within the oldest avian Z-W evolutionary strata) in a New World sparrow, the white-throated sparrow (*Zonotrichia albicollis*), and the genomic properties associated with this ancient non-recombining segment of the avian W chromosome.

Materials and methods

Isolation and sequencing of white-throated sparrow CHD1Z/W BAC clones

Overgo-hybridization probes designed from sequences conserved between the published zebra finch *CHD1Z* and *CHD1W* loci (Agate et al. 2004) using the Uprobe custom overgo-probe design tool (Sullivan et al. 2008) were used to screen a female white-throated sparrow BAC library (CHORI-264, http://bacpac.chori.org/library.php?id=469) as described in Thomas (2008). Probe-content and restriction-enzyme fingerprint maps (Thomas et al. 2002) were constructed from the positive clones, and a polymerase chain reaction (PCR) assay for sexing birds based on a length polymorphism between the *CHD1Z* and *CHD1W* gametologs (Griffiths et al. 1998) was used to assign the BAC clones to either the Z or W chromosome. The combined probe-content and restriction-enzyme fingerprint maps in conjunction with comparative mapping data of BAC-end sequences (generated at the British Columbia Cancer Agency Genome Sciences Centre, Vancouver, Canada) were then used to select overlapping pairs of Z and W chromosome BAC clones for sequencing. The individual BACs were Sanger shotgun sequenced, assembled into an ordered set of contigs (Blakesley et al. 2004), and used to generate multi-BAC assemblies representing each chromosome (GenBank Ac# DP001175 and DP001176).

Genomic sequence annotation

The intron–exon structures were annotated based on interspecies genomic-cDNA alignments generated with Spidey (Wheelan et al. 2001) where the genomic sequence was white-throated sparrow BAC sequence assemblies, and the cDNAs were primarily zebra finch mRNAs or predicted mRNAs. Common repetitive elements within the BAC sequence assemblies were first identified using RepeatMasker (library version 20090604) (Smit et al. 1996–2004) using the species "*Z. albicollis*" option. RetroTector (Sperber et al. 2009) was then used to identify potential endogenous retroviral elements in both the Z and W sequence assemblies. RepeatMasker was then run again on Z and W assemblies using the sequences identified by Retro-Tector as a custom repeat library. The first pass of RepeatMasking, not including the custom repeat library, estimated the common repetitive element content of the Z and W sequences as 8.7% and 16.4%, respectively.

PCR and sequencing of the palindrome arm-spacer junction fragments in other songbirds

To identify potential repetitive regions not detected by RepeatMasker or RetroTector, the inverted repeats and flanking intervals were compared to the zebra finch genome assembly (taeGut1; Warren et al. 2010) using MEGABLAST (-t 16, -N 2, -W 11, -e 1e-10; Zhang et al. 2000). PCR primers were then designed from the non-repetitive sequence to amplify the palindrome arm-internal spacer junctions. The sequences of the PCR primers used were as follows: Left.1f: 5′-GGC TAC CAA GAA GTG TTA TTT CCT GTC CTC ATC-3′, Left.1r: 5′-ACC CTG TAG CCT AAT TCA CAG ACA GAT CTT GG-3′, Left.2f: 5′-AGC AAA GGG AAC AAA GGG TAT TAA GTT TAG GC-3′, Left.2r: 5′-AGA TGG TTA AGT AAA GCC AGC TCT TTG TCA GC-3′, Right.1f: 5′-CAA CAC AAT TAC AGA ATG CTG GAA ACA GGA AG -3′, Right.1r: 5′-CTT CAT GAC AGA CAG TTA TAC AAG CCA TGC TG-3′, Right.2f: 5′-AGA CAG CTC TGA GCA ACA CAA TTA CAG AAT GC-3′, and Right.2r: 5′-CCA GAA TGA AGT AAC AGA CAC TAA ATG CCA AGC-3′.

Note that no PCR primers that met these criteria could be designed to amplify the junctions of the palindrome arms with the external flanking sequence. The PCR primers were then used to amplify female DNA from a locally sampled dark-eyed junco (*Junco hyemalis*), Harris's sparrow (*Zonotrichia querula*, UWBM #69719), and Bullock's oriole (*Icterus bullockii*, UWBM #55978) in a 50-μl reaction volume including ~120 ng of genomic DNA, 2.5 mM $MgCl_2$, 400 μM of each dNTP, 0.2 μM of each PCR primer, 2.5 units of Takara LA Taq, and LA PCR buffer II (Takara Bio Inc.) and the following cycling conditions: 1 min at 94°C, and 30 cycles of 30 s at 94°C, 5 min at 68°C, and a single 10 min extension at 72°C. The PCR amplicons were then directly sequenced (Sanger sequencing) using internal primers (GenBank Ac# GU985567-GU985569).

Genomic sequence alignments

Pairwise alignments of the BAC assemblies and dotplots were generated with PipMaker (Schwartz et al. 2000). ClustalX (Jeanmougin et al. 1998) was used to generate a multiple sequence alignment of the inverted repeats and internal flanking segments. Divergence estimates were calculated using MEGA (Kumar et al. 2004). Note the divergence estimates of the white-throated sparrow W chromosome palindrome arms were based only on sites with a Phred score ≥40.

Results

General genomic features and gene content of the Z and W BAC assemblies

To sample the genomic landscape of the Z and W chromosomes in the white-throated sparrow, we used a BAC-based targeted mapping and sequencing approach to sequence the homologous regions on the sex chromosomes containing the *CHD1Z/W* gametologs. The sequence of a pair of overlapping BAC clones from the Z chromosome resulted in a final assembly of 289 kb that included just three sequencing gaps. Similarly, the BAC-based assembly from the W chromosome was 279 kb and included five sequencing gaps. The GC content for the assembled Z and W chromosome sequences were 40.1% and 43.7%, respectively. Because no systematic study has been conducted to identify common repetitive elements in the white-throated sparrow genome, we used an iterative approach to identify common repetitive elements (see "Materials and Methods" for details) and estimated the repetitive element content of the Z chromosome sequence assembly as 25.7% and 51.1% for the W chromosome sequence.

In terms of gene content, the *CHD1Z* and *CHD1W* gametologs were fully contained within the respective Z and W chromosome sequence assemblies. The predicted proteins encoded by the white-throated sparrow *CHD1Z* and *CHD1W* genes were 91.7% identical, which is similar to what has been reported previously for the identity (91%) between these two proteins in the zebra finch (Agate et al. 2004). In addition, as has been reported for other interspecies comparisons of these genes in other birds (Fridolfsson and Ellegren 2000), the white-throated sparrow and zebra finch CHD1Z and CHD1W orthologous proteins were more identical to one another (98.1% and 97.1%, respectively), than the intraspecies gametologs.

The Z chromosome sequence also included the entire *RGMB* locus and a portion of the *RASA1* gene, and the order and orientation of these three linked Z chromosome genes in the white-throated sparrow were the same as that observed in another songbird, the zebra finch (GenBank Ac# AC188310 (Itoh et al. 2008; Warren et al. 2010). Alignment of the white-throated sparrow chromosome Z and W assemblies did not reveal any evidence for the presence of gametologs of either *RGMB* or *RASA1* within the sequenced interval on the W chromosome (Fig. 1). The absence of those two genes in the W chromosome assembly could

be due to the loss of these genes resulting from genetic deterioration, which is the expected fate of most genes on a non-recombining chromosome. However, it is also possible that the absence of *RGMB* and *RASA1* on the W chromosome assembly is simply the result of the fact that the W chromosome sequence assembly did not extend far enough to include those loci, or that those genes are present on the W chromosome but no longer linked to *CHD1W*. To determine if any other genes were present in the W chromosome sequence assembly, we used GENSCAN (Burge and Karlin 1997) to generate a list of predicted genes. Though GENSCAN produced a number of predicted gene models, the best BLASTP (Altschul et al. 1997) matches in the GenBank nr database to the protein products of the predicted GENSCAN genes were limited to homologs of *CHD1W* and proteins encoded by common repetitive elements. Thus, there was no evidence for the presence of genes other than *CHD1W* in the white-throated sparrow W chromosome sequence assembly.

Palindromes on the W chromosome

Near-identical palindromes are one of the hallmark features of the human Y chromosome (Skaletsky et al. 2003) and are also present on the chimpanzee Y chromosome (Hughes et al. 2010). To determine if the sequenced region of the white-throated sparrow W chromosome contained similar structural elements, we aligned the W chromosome sequence assembly to itself (Fig. 2a). After excluding examples where the presence of the inverted repeats were primarily due to an alignment between a pair of common repetitive elements, one pair of 22-kb inverted repeats was detected (Fig. 2a). These inverted repeats are separated by an 8-kb spacer region that contains the 5′ end of the *CHD1W* gene, including exons 1 and 2, and the immediate upstream sequence (Fig. 2b). The arms of this palindrome therefore include a portion of *CHD1W* intron 2 and a large block of flanking sequence just 5′ of the *CHD1W*. Note that no analogous inverted repeats were present within the Z chromosome assembly (see Fig. 1).

Gene conversion between the palindrome arms in the white-throated sparrow

Sequence alignment of the palindrome arms to one another identified just ten single-nucleotide differences and a 5-bp indel in 21745 aligned sites. The near identity of inverted repeats (Kimura 2-parameter distance of 0.00046±0.00015) could be the result of a very recent duplication, or alternatively could reflect homogenization of the inverted repeats by the process of gene conversion, which has been observed on the human and great ape Y chromosomes (Rozen et al. 2003). To distinguish between those different evolutionary scenarios, we first looked for signatures of gene conversion based on the pattern of divergence between the inverted repeats. Because gene conversion is likely to be the least effective at the edges of the duplications, we evaluated the spatial distribution of the sequence differences between the inverted repeats. Consistent with homogenization of the inverted repeats by past gene conversion events, there was on average just one substitution per 10,000 bp within internal segments of the inverted repeats, whereas the lone indel and most of the single-nucleotide differences (eight out of ten) clustered within 23 bp of the internal edge of the palindrome arms. In addition, we tested for the presence of gene conversion tracts using GENECONV (Sawyer 1989). This method identified two potential gene conversion tracts (p value ≤0.01) totaling 15,994 bp. Thus, signatures of gene conversion were detected between the arms of the white-throated sparrow W-linked palindrome.

The inverted repeats are present in other New World sparrows and blackbirds

Given the evidence for gene conversion between the palindrome arms in the white-throated sparrow, we cannot accurately estimate how old the inverted repeats are based on sequence divergence alone. However, because gene conversion acts to retard the tendency for duplicated sequences to diverge from each other as a linear function of time, we can infer

that the inverted repeats must be older than the observed divergence between them would suggest. In particular, assuming a mutation rate previously determined for another songbird, the zebra finch (Balakrishnan and Edwards 2009), the observed divergence between the inverted repeats would suggest they arose just ~158,000±61,000 years ago. Thus, even though the estimated time since the duplication that gave rise to the white-throated sparrow inverted repeats is well below the minimum estimated time of 750,000 years since the divergence of the white-throated sparrow lineage from other birds in the *Zonotrichia* genus (Zink et al. 1991), we hypothesized that the inverted repeats might have originated prior to the radiation of this genus. As such, the W chromosome inverted repeats might be present in other *Zonotrichia* sparrows, or perhaps even more distantly related birds. To test this hypothesis, we developed PCR primers designed to amplify a portion of the internal edges of the palindrome arms and the adjacent spacer region and tested these on a female Harris's sparrow (*Z. querula*), a dark-eyed junco (*J. hyemalis*), and a Bullock's oriole (*I. bullockii*). In all three species, we were able to amplify and sequence both palindrome arm-spacer junctions. Specifically, we generated between 1,053 and 2,399 bp of sequence flanking each of the palindrome arm-internal spacer junctions, which included a minimum of 683 bp of palindromic sequence and 380-bp of the internal spacer. These results therefore support the hypothesis that the inverted repeats are present in other species and that the inverted repeats originated prior to the most recent common ancestor of the four members of the Emberizinae subfamily of New World songbirds sampled here.

Additional signatures of gene conversion between palindrome arms

Since the inverted repeats were inferred to have arisen at some point prior to the most recent common ancestor of the Emberizinae species in our sample, in the absence of gene conversion, we would expect that at a minimum, the intraspecies divergence between the palindrome arms would be comparable to the maximum interspecies divergence amongst the four sampled birds. However, as expected from the inference of gene conversion based on the white-throated sparrow sequence, the observed intraspecies divergence between the palindrome arms in the three additionally sample species was between 0.011 and 0.013 substitutions per site, which is lower than the ~0.03 substitutions per site observed between Bullock's oriole and the other three species (Table 1). In addition, gene conversion would also be expected to result in the palindrome arms from one species being more similar to one another than to their orthologous sequences in the other species. To visualize the relationship among the palindrome arms, we constructed a phylogenetic network (Fig. 3). While the Bullock's oriole palindrome arms were more similar to one another than the orthologous palindrome arms, the palindrome arms from the other three species clustered with their orthologs, not their intraspecies paralogs. Note, however, that the phylogenetic network was based on <700 bp of sequence that we had sampled in all species from the internal edge of both palindrome arms, and thus was heavily biased toward the segment of the inverted repeats that was by far the most divergent in the white-throated sparrow. This sampling bias is likely also the basis for the much higher inter-arm divergence observed in Harris's sparrow, the dark-eyed junco, and Bullock's oriole of ~0.012 versus the 0.00046 substitutions per site observed across the entire length of the palindrome arms in the white-throated sparrow.

To look for additional signatures of gene conversion between the palindrome arms, we reconstructed the most parsimonious origin for each of the observed intra- and interspecies differences identified in an alignment of the palindrome arm sequences (Fig. 4a). In particular, we assumed that the inverted repeats were present in the common ancestor of all four birds; the intra- or interspecies differences between the palindrome arms arose either from a single mutation on one palindrome arm or a single mutation in one arm of the palindrome followed by gene conversion of the mutation to the other arm; and the

occurrence of two independent mutations at the equivalent position in both arms of a repeat or in the orthologous position in different species was rare enough to be discounted as a potential basis for the sequence differences we observed in our data set. This logic was then applied in the context of the species phylogeny and used to assign a most parsimonious time point and mechanism for each of the observed differences. Of the 28 observed sequence differences, 15 were classified as a point mutation followed by gene conversion between the palindrome arms (Fig. 4a, b). Note that as would be expected based on the relative evolutionary distances between the four sampled species (Table 1), the greatest proportion of the sequence differences was predicted to have occurred on the branches separating the Bullock's oriole from the most recent common ancestor of the white-throated sparrow, Harris's sparrow, and dark-eyed junco (Fig. 4b).

Discussion

Studies of the avian W chromosome offer a novel point of comparison and perspective on the features of the eutherian MSY. Though a definitive history of the avian sex chromosomes has yet to be fully established (Matsubara et al. 2006; Ezaz et al. 2009), molecular studies have suggested the oldest Z-W evolutionary strata originated ~132–150 million years ago (Nam and Ellegren 2008) and therefore, like the MSY, has chromosomal segments that have been subjected to the suite of forces that are expected to lead to genetic degeneration over extended periods of time (Charlesworth and Charlesworth 2000). As a result, only a handful of genes have been identified on the avian W chromosome, though some, like *CHD1W*, have persisted in the absence of recombination for over 100 million years (Nam and Ellegren 2008). In this study, we used targeted mapping and sequencing of *CHD1Z* and *CHD1W* gametologs and flanking regions as a means to empirically survey the genomic features within the oldest known evolutionary strata on the avian W chromosome.

The chicken genome was the first avian genome to be sequenced, and while a female (ZW) was used as the whole-genome shotgun sequencing template, very little, ~850 kb, of the female-specific W sequence was included in the assembly (Hillier et al. 2004). The assembled chicken W chromosome sequence, similar to the non-recombining Y (W) segments reported in a broad range of species (Skaletsky et al. 2003; Kondo et al. 2006; Yu et al. 2007; Bachtrog et al. 2008), was high in repetitive element content, ~55.3% compared to the genome average of 12.7%. Similarly, we observed that the 279-kb W chromosome sequence of the white-throated sparrow, including the *CHD1W* locus and flanking regions, was high in repetitive content, 50.9%, and a had low gene density (i.e., one gene in 279 kb), which has also been observed on other Y (W) non-recombining chromosomal segments (Skaletsky et al. 2003; Kondo et al. 2006; Yu et al. 2007; Vibranovski et al. 2008). Another established feature of avian W chromosomes is the presence of repetitive elements that are particularly enriched on the W chromosome (Itoh et al. 2008 and references therein). However, sequence similarity searches of published W-enriched repetitive elements detected in another songbird, the zebra finch (Itoh et al. 2008), did not detect those elements in the white-throated sparrow Z and W assemblies. This result is consistent with the observation that the W chromosome repeats tend to be evolving rapidly and are not conserved between avian lineages (Itoh et al. 2008 and references therein). In addition, comparison of the W chromosome assembly to itself did not reveal any obvious short common repeated elements, though it is worth noting that we did detect an array of related 432- and 189-bp direct repeats in the Z chromosome assembly (data not shown, but the location of the direct repeats is annotated in GenBank Ac# DP001176).

The most striking findings from analyses of the white-throated sparrow Z and W chromosome sequence assemblies were the presence of a 22-kb near-identical palindrome and signatures of gene conversion between the palindrome arms. Large, near-identical

palindromes and gene conversion between the palindrome arms are hallmark features of the human and chimpanzee MSY (Skaletsky et al. 2003; Hughes et al. 2010). Near-identical palindromes are also over-represented in the human genome on both the X and Y chromosomes (Warburton et al. 2004). As far as we are aware, no genome-wide scans of near-identical palindromes in birds have been performed either on the chicken or zebra finch genome assemblies (Hillier et al. 2004; Warren et al. 2010). Thus, it is possible that the discovery of a near-identical palindrome on the white-throated sparrow W chromosome is not necessarily a reflection of a similar bias for these elements on the sex chromosomes, but may simply be a chance encounter. It should be noted, however, that while the palindrome we observed on the W chromosome was quite small compared to those present on the human and chimpanzee Y chromosomes (Skaletsky et al. 2003; Hughes et al. 2010), the neutral and selection-based theories as to why near-identical palindromes are over-represented on the human X and Y chromosomes (Rossiter et al. 1994; Skaletsky et al. 2003; Warburton et al. 2004) could be applied equally well to the W chromosome. That is, the female-specific W chromosome does not pair in female meiosis, which in turn may lead to an increased rate of intrastrand pairing and gene conversion. Gene conversion between the palindrome arms could be one mechanism by which genetic degeneration is retarded on the non-recombining segment of the W chromosome, which in this case might influence the evolution not of a protein-coding region of *CHD1W*, but potentially its upstream *cis*-regulatory elements. Note that signatures of gene conversion have also been detected on the avian W chromosome between amplified copies of the *HINTW* genes (Backstrom et al. 2005), though it appears those amplicons are in tandem array versus an inverted repeat orientation (Hori et al. 2000). The potential for the palindrome described here to mediate the formation of a cruciform could be important for the normal expression of *CHD1W*, which for example, in the zebra finch is relatively widespread (Agate et al. 2004), or during the process of MSCI in the female germ line (Schoenmakers et al. 2009). Additional large-scale sequencing in this or other birds, specifically BAC-based sequencing where large near-identical repeats are not collapsed in the assembly process, will be needed to determine the prevalence of near-identical palindromes across the avian genome and if those structures are enriched on the W, and perhaps Z, chromosome.

Another intriguing parallel with respect to human sex-linked palindromes is the potential for deleterious chromosomal rearrangements to be caused by non-allelic recombination between the arms of the W chromosome palindrome. In particular, intrachromatid non-allelic recombination resulting in a crossover between the palindrome arms would lead to an inversion that reverses the orientation of the first two *CHD1W* exons, thereby disrupting the physical structure of this gene. Barring the presence of alternative promoters and/or 5′ exons for *CHD1W* that are not within the spacer region between the palindrome arms, such a physical disruption of the gene structure is predicted to result in a null allele akin to the pathogenic factor VIII inversion mutations in observed in humans (Lakich et al. 1993). It is also possible that sister chromatid non-allelic recombination-mediated crossovers between the palindrome arms could lead to isodicentric and acentric W chromosomes as has been observed on the human Y (Lange et al. 2009). While we have no data to support that either type of chromosomal rearrangement is in fact mediated by the avian palindrome that flanks the 5′ end of the *CHD1W* locus, we can predict that, like some human X and Y palindromes, there is a potential deleterious cost associated with maintaining this W-linked palindrome in the genome. It is therefore notable that we detected this palindrome in other members of the Emberizinae subfamily of songbirds. Though the relatively rapid radiation of this clade makes inferences difficult as to the extent to which this palindrome is a conserved feature across these species (Yuri and Mindell 2002), based on our sampled set of birds, we can at a minimum predict it is likely to be present in New World sparrows (n=157 species) and New World blackbirds (n=97 species).

In conclusion, we identified a W chromosome-linked near-identical palindrome in New World sparrows and blackbirds, observed signatures of gene conversion between the palindrome arms similar to that previously described for human and ape Y chromosomes, and demonstrated the utility of the recently constructed white-throated sparrow BAC library for targeted mapping and sequencing of the avian W chromosome.

Acknowledgments

The authors thank Judith E. Mank for helpful comments and discussions, Mario Cáceres for comments, Cheryl T. Strauss for technical writing edits, Donna L. Maney for the dark-eyed junco DNA, Greg K. Tharp for computer support, the BC Cancer Agency Genome Sciences Centre, Vancouver, Canada, for generation of the BAC-end sequences, and members of the NIH Sequencing Center including E.D. Green, R. Blakesley, G. Bouffard, and J. McDowell. DNA samples (excluding the dark-eyed junco) were provided by The Burke Museum of Natural History and Culture. J.K.D. and J.W.T. were supported by a grant from the National Institutes of Health (1R21MH082046), and the NIH Intramural Sequencing Center was supported in part by the Intramural Research Program of the National Human Genome Research Institute of the National Institutes of Health.

Abbreviations

BAC	Bacterial artificial chromosome
CHD1W	Chromo-helicase DNA-binding protein, W chromosome
CHD1Z	Chromo-helicase DNA-binding protein, Z chromosome
CHORI	Children's Hospital Oakland Research Institute
dNTP	*deoxy-nucleotide triphosphate*
HINTW	Histidine triad nucleotide binding protein W
kb	Kilobase pairs
MSCI	Meiotic sex chromosome inactivation
MSY	Male-specific region of the Y chromosome
PCR	Polymerase chain reaction
RASA1	RAS p21 protein activator (GTPase activating protein) 1
RGMB	RGM domain family, member B
UWBM	University of Washington Burke Museum

References

Agate RJ, Choe M, Arnold AP. Sex differences in structure and expression of the sex chromosome genes CHD1Z and CHD1W in zebra finches. Mol Biol Evol 2004;21:384–396. [PubMed: 14660691]

Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402. [PubMed: 9254694]

Bachtrog D, Hom E, Wong KM, Maside X, De Jong P. Genomic degradation of a young Y chromosome in *Drosophila miranda*. Genome Biol 2008;9:R30. [PubMed: 18269752]

Backstrom N, Ceplitis H, Berlin S, Ellegren H. Gene conversion drives the evolution of HINTW, an ampliconic gene on the female-specific avian W chromosome. Mol Biol Evol 2005;22:1992–1999. [PubMed: 15972846]

Bagnall RD, Ayres KL, Green PM, Giannelli F. Gene conversion and evolution of Xq28 duplicons involved in recurring inversions causing severe hemophilia A. Genome Res 2005;15:214–223. [PubMed: 15687285]

Balakrishnan CN, Edwards SV. Nucleotide variation, linkage disequilibrium and founder-facilitated speciation in wild populations of the zebra finch (*Taeniopygia guttata*). Genetics 2009;181:645–660. [PubMed: 19047416]

Blakesley RW, Hansen NF, Mullikin JC, et al. An intermediate grade of finished genomic sequence suitable for comparative analyses. Genome Res 2004;14:2235–2244. [PubMed: 15479945]

Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. J Mol Biol 1997;268:78–94. [PubMed: 9149143]

Caceres M, Sullivan RT, Thomas JW. A recurrent inversion on the eutherian X chromosome. Proc Natl Acad Sci USA 2007;104:18571–18576. [PubMed: 18003915]

Charlesworth B, Charlesworth D. The degeneration of Y chromosomes. Philos Trans R Soc Lond B Biol Sci 2000;355:1563–1572. [PubMed: 11127901]

Charlesworth D, Charlesworth B, Marais G. Steps in the evolution of heteromorphic sex chromosomes. Heredity 2005;95:118–128. [PubMed: 15931241]

Ellegren H, Carmichael A. Multiple and independent cessation of recombination between avian sex chromosomes. Genetics 2001;158:325–331. [PubMed: 11333240]

Ezaz T, Moritz B, Waters P, Marshall Graves JA, Georges A, Sarre SD. The ZW sex microchromosomes of an Australian dragon lizard share no homology with those of other reptiles or birds. Chromosome Res 2009;17:965–973. [PubMed: 19967443]

Fridolfsson AK, Ellegren H. Molecular evolution of the avian CHD1 genes on the Z and W sex chromosomes. Genetics 2000;155:1903–1912. [PubMed: 10924484]

Griffiths R, Double MC, Orr K, Dawson RJ. A DNA test to sex most birds. Mol Ecol 1998;7:1071–1075. [PubMed: 9711866]

Hillier LW, Miller W, Birney E, et al. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 2004;432:695–716. [PubMed: 15592404]

Hori T, Asakawa S, Itoh Y, Shimizu N, Mizuno S. 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. Mol Biol Cell 2000;11:3645–3660. [PubMed: 11029061]

Hughes JF, Skaletsky H, Pyntikova T, et al. Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. Nature 2010;463:536–539. [PubMed: 20072128]

Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 2006;23:254–267. [PubMed: 16221896]

Itoh Y, Kampf K, Arnold AP. Molecular cloning of zebra finch W chromosome repetitive sequences: evolution of the avian W chromosome. Chromosoma 2008;117:111–121. [PubMed: 17972090]

Itoh Y, Kampf K, Arnold AP. Disruption of FEM1C-W gene in zebra finch: evolutionary insights on avian ZW genes. Chromosoma 2009;118:323–334. [PubMed: 19139913]

Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ. Multiple sequence alignment with Clustal X. Trends Biochem Sci 1998;23:403–405. [PubMed: 9810230]

Kondo M, Hornung U, Nanda I, et al. Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of medaka. Genome Res 2006;16:815–826. [PubMed: 16751340]

Kumar S, Tamura K, Nei M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 2004;5:150–163. [PubMed: 15260895]

Lahn BT, Page DC. Four evolutionary strata on the human X chromosome. Science 1999;286:964–967. [PubMed: 10542153]

Lakich D, Kazazian HH Jr, Antonarakis SE, Gitschier J. Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. Nat Genet 1993;5:236–241. [PubMed: 8275087]

Lange J, Skaletsky H, Van Daalen SK, et al. Isodicentric Y chromosomes and sex disorders as byproducts of homologous recombination that maintains palindromes. Cell 2009;138:855–869. [PubMed: 19737515]

Mank JE, Ellegren H. Parallel divergence and degradation of the avian W sex chromosome. Trends Ecol Evol 2007;22:389–391. [PubMed: 17573147]

Marshall Graves JA. Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. Annu Rev Genet 2008;42:565–586. [PubMed: 18983263]

Matsubara K, Tarui H, Toriba M, et al. Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise differentiation of snake sex chromosomes. Proc Natl Acad Sci USA 2006;103:18190–18195. [PubMed: 17110446]

Nam K, Ellegren H. The chicken (*Gallus gallus*) Z chromosome contains at least three nonlinear evolutionary strata. Genetics 2008;180:1131–1136. [PubMed: 18791248]

Ohno, S. Sex chromosomes and sex-linked genes. Springer-Verlag; New York: 1967.

Rossiter JP, Young M, Kimberland ML, et al. Factor VIII gene inversions causing severe hemophilia A originate almost exclusively in male germ cells. Hum Mol Genet 1994;3:1035–1039. [PubMed: 7981669]

Rozen S, Skaletsky H, Marszalek JD, et al. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature 2003;423:873–876. [PubMed: 12815433]

Sawyer S. Statistical tests for detecting gene conversion. Mol Biol Evol 1989;6:526–538. [PubMed: 2677599]

Schoenmakers S, Wassenaar E, Hoogerbrugge JW, Laven JS, Grootegoed JA, Baarends WM. Female meiotic sex chromosome inactivation in chicken. PLoS Genet 2009;5:e1000466. [PubMed: 19461881]

Schwartz S, Zhang Z, Frazer KA, et al. PipMaker—a web server for aligning two genomic DNA sequences. Genome Res 2000;10:577–586. [PubMed: 10779500]

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, et al. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 2003;423:825–837. [PubMed: 12815422]

Sperber G, Lovgren A, Eriksson NE, Benachenhou F, Blomberg J. RetroTector online, a rational tool for analysis of retroviral elements in small and medium size vertebrate genomic sequences. BMC Bioinform 2009;10(6):S4.

Sullivan RT, Morehouse CB, Thomas JW. Uprobe 2008: an online resource for universal overgo hybridization-based probe retrieval and design. Nucleic Acids Res 2008;36:W149–W153. [PubMed: 18515352]

Thomas, JW. Comparative physical mapping: universal overgo hybridization probe design and BAC library hybridization. In: Murphy, WJ., editor. Methods in molecular biology: phylogenomics. Humana Press; Totowa: 2008. p. 119-132.

Thomas JW, Prasad AB, Summers TJ, et al. Parallel construction of orthologous sequence-ready clone contig maps in multiple species. Genome Res 2002;12:1277–1285. [PubMed: 12176935]

Vibranovski MD, Koerich LB, Carvalho AB. Two new Y-linked genes in *Drosophila melanogaster*. Genetics 2008;179:2325–2357. [PubMed: 18660539]

Warburton PE, Giordano J, Cheung F, Gelfand Y, Benson G. Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. Genome Res 2004;14:1861–1869. [PubMed: 15466286]

Warren WC, Clayton DF, Ellegren H, et al. The genome of a songbird. Nature 2010;464:757–762. [PubMed: 20360741]

Wheelan SJ, Church DM, Ostell JM. Spidey: a tool for mRNA-to-genomic alignments. Genome Res 2001;11:1952–1957. [PubMed: 11691860]

Wilson MA, Makova KD. Genomic analyses of sex chromosome evolution. Annu Rev Genomics Hum Genet 2009;10:333–354. [PubMed: 19630566]

Yu Q, Hou S, Hobza R, et al. Chromosomal location and gene paucity of the male specific region on papaya Y chromosome. Mol Genet Genomics 2007;278:177–185. [PubMed: 17520292]

Yuri T, Mindell DP. Molecular phylogenetic analysis of Fringillidae, “New World nine-primaried oscines” (Aves: Passeriformes). Mol Phylogenet Evol 2002;23:229–243. [PubMed: 12069553]

Zhang Z, Schwartz S, Wagner L, Miller W. A greedy algorithm for aligning DNA sequences. J Comput Biol 2000;7:203–214. [PubMed: 10890397]

Zink RM, Dittmann DL, Rootes WL. Mitochrondrial DNA variation and the phylogeny of *Zonotrichia*. Auk 1991;108:578–584.

Internet references

Smit, AFA.; Hubley, R.; Green, P. RepeatMasker Open-3.0. 1996–2004. http://www.repeatmasker.org

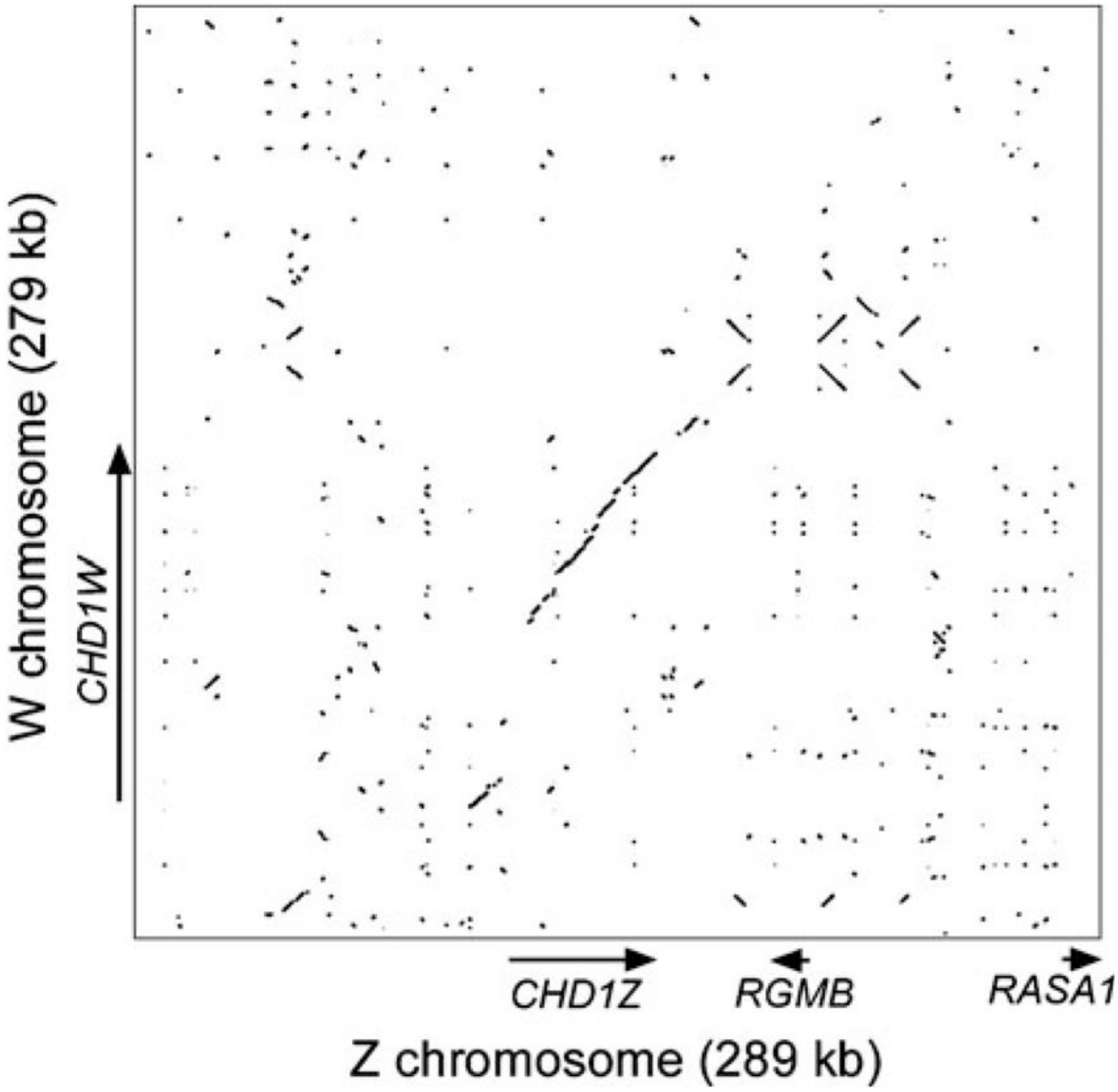

Fig. 1.

Alignment of the white-throated sparrow Z and W chromosome assemblies. The reverse-complement of the Z and W chromosome BAC-based assemblies were aligned with PipMaker allowing matches on both strands and are illustrated in the form of a *dotplot*. The position, orientation, and name of each gene are illustrated by a *labeled arrow*

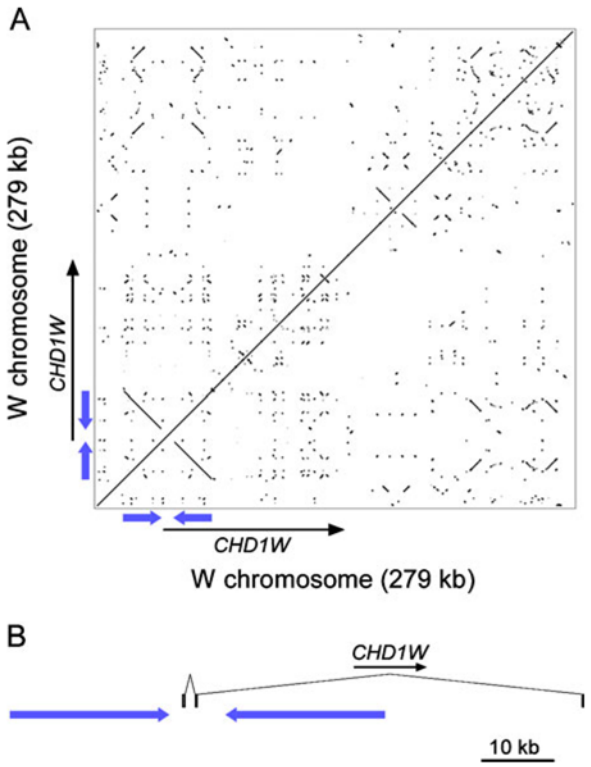

Fig. 2.

Location of a pair of inverted repeats on the white-throated sparrow W chromosome. **a** A *dotplot* of the white-throated sparrow W chromosome sequence aligned to itself. *Blue arrows* indicate the location and orientation of the inverted repeats. The position and orientation of the *CHD1W* gene is indicated by a *black arrow*. **b** The relative position of the inverted repeats (*blue arrows*) and *CHD1W* exons 1 and 2, which are within the spacer region between the palindrome arms, and *CHD1W* exon 3 are illustrated. *CHD1W* exons are depicted by *small black rectangles*. Note *CHD1W* exons 4–37 are not shown

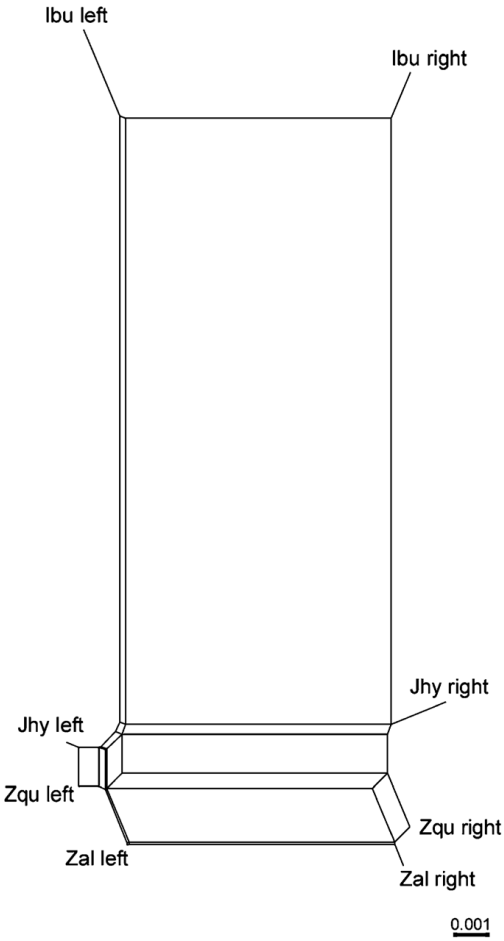

Fig. 3.

Phylogenetic network of the palindrome arms. An alignment of the palindrome arms (*left* and *right*) from the four sampled songbirds was used to construct a phylogenetic network with SplitsTree (Huson and Bryant 2006) using the K2P distance and Neighbor Net drawing option. Ibu=Bullock's oriole, Jhy=dark-eyed junco, Zqu=Harris's sparrow, Zal= white-throated sparrow

A

```
                                           04015273241690
                                 0123457024925006668359258662
                               6789111111122223812222233344456
White-throated sparrow left  -----GGCTGGGATGGCGTTGCCACGGTCTA
White-throated sparrow right TTATTGCATATAAGATCGTTGCCACGGTCTA
Harris's sparrow left        -----GGCTGGGATGGCGTCGCCACGGTCTA
Harris's sparrow right       TTATTGCATATAAGATCGTCGCCACGGTCTA
Dark-eyed junco left         -----GGCTGGGATGGCGTCGCCACGGGCTA
Dark-eyed junco right        TTATTGCATATAAGATCGTCGCCACGGGCTA
Bullock's oriole left        -----GGCCGGGGTGTTTGCTGTGTCAGGGG
Bullock's oriole right       TTGTTACATATAAGGTTTGCTGTGTCAGGGG

                             aaaaacddcdddcdeefffgffffffffhfff
                               b
```

B

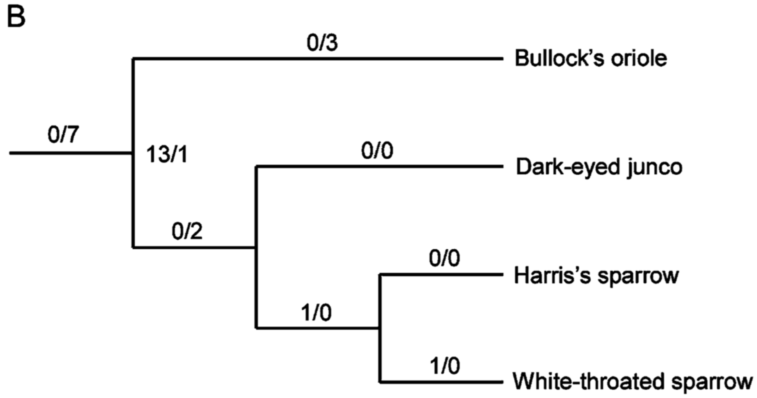

Fig. 4.

Patterns of divergence in the palindrome arms within and between species. A multiple sequence alignment of the palindrome arm sequences was generated with ClustalX and used to identify variable positions in the alignment. **a** The columns from the alignment that included a mismatch or indel are shown on the *right*. The species name and whether the sequence corresponds to the left or right palindrome (relative to GenBank Ac# DP001175) are listed on the *left*. The position of each alignment column relative to the white-throated sparrow right palindrome is listed *above the aligned sequences*. The most parsimonious mechanism and time point for each variable position was inferred and indicated by the *lowercase letter* code below the alignment: *a*, indel and no gene conversion prior to the most recent common ancestor of all four birds (*n*=1); *b*, point mutation and no gene conversion in the terminal branch leading to Bullock's oriole, or in the ancestral branch leading to the dark-eyed junco, Harris's sparrow, and white-throated sparrow (*n*=1); *c*, point mutation and no gene conversion in the terminal branch leading to Bullock's oriole (*n*=3); *d*, point mutation and no gene conversion prior to the most recent common ancestor of all four birds (*n*=6); *e*, point mutation and no gene conversion in the ancestral lineage leading to the dark-eyed junco, Harris's sparrow, and white-throated sparrow (*n*= 2); *f*, point mutation and gene conversion in the terminal branch leading to Bullock's oriole, or in the ancestral branch leading to the dark-eyed junco, Harris's sparrow, and white-throated sparrow (*n*=13); *g*, point mutation and gene conversion in the terminal branch leading to the white-throated sparrow (*n*=1); *h*, point mutation and gene conversion in the ancestral branch leading to Harris's sparrow and the white-throated sparrow (*n*= 1). **b** The inferred time points and type of mutation are illustrated with respect to the species phylogeny as follows. The number of point mutations followed by gene conversion (first number) and the number of point mutations (and indels) and no gene conversion (second number) are listed *above each branch*. The number of events that occurred in either the terminal branch leading to Bullock's oriole or in the ancestral branch leading to the dark-eyed junco, Harris's sparrow, and white-throated sparrow is listed between those two branches

Table 1

Interspecies divergence between the sampled songbirds

Species	1	2	3	4
1. White-throated sparrow	–	–	–	–
2. Harris's sparrow	0.0022±0.0008	–	–	–
3. Dark-eyed junco	0.0094±0.0017	0.0098±0.0018	–	–
4. Bullock's oriole	0.0308±0.0030	0.0304±0.0032	0.0303±0.0036	–

The Kimura 2-parameter distance and standard errors refer to substitutions per site based on interspecies alignments of sequenced segments of the palindrome arms and the adjacent internal spacer region