Sex Differences in Structure and Expression of the Sex Chromosome Genes CHD1Z and CHD1W in Zebra Finches

Robert J. Agate,¹ Meeryo Choe, and Arthur P. Arnold

Department of Physiological Science, and Laboratory of Neuroendocrinology of the Brain Research Institute, University of California, Los Angeles

Genes on the sex chromosomes are unique because of their sex-specific inheritance. One question is whether homologous gene pairs on the sex chromosomes, which have diverged in their sequence, have acquired different functions. We have analyzed the first homologous pair of genes (CHD1Z and CHD1W) discovered on the avian Z and W sex chromosomes of the zebra finch (Taeniopygia guttata) to examine whether functional differences may have evolved. Sequence analysis revealed that the two genes maintained a high degree of similarity especially within the C, H, and D domains, but outside of these regions larger differences were observed. Expression studies showed that CHD1W was unique to females and has the potential to produce a protein that CHD1Z does not. CHD1Z mRNA was expressed at a higher level in the male brain than in the female brain at various post-hatch ages. Reporter constructs containing the 5′ flanking regions of each gene showed they had the ability to drive reporter expression in primary cell cultures. The 5′ flanking region sequence of CHD1Z and CHD1W exhibited little homology, and differences in putative promoter elements were apparent. These differences between CHD1Z and CHD1W suggest that the two proteins may have diverged in their function.

Introduction

The sex chromosomes in mammals are thought to have evolved from an ancestral pair of autosomes following a dominant male-determining mutation on the proto-Y chromosome which led eventually to an inhibition of recombination between the proto-Y and proto-X (Graves and Shetty 2001). This lack of X-Y recombination over most of the Y chromosome led to an accumulation of mutations and loss of most Y genes which formerly were alleles of X genes (Charlesworth 1991; Lahn and Page 1997; Skaletsky et al. 2003). A significant number of genes encoded on the human Y chromosome, for example, have been classified as “X-degenerate” because they, unlike the vast majority of proto-Y genes, have survived this degradation process (Skaletsky et al. 2003). One explanation for this unique conservation is that these genes perform a critical male-specific function (Lahn and Page 1997), as exemplified by the Y-linked testis-determining gene SRY, which has no functionally similar X-linked homolog. On the other hand, most of these Y genes have a highly homologous partner gene on the X sex chromosome (forming X-Y gene pairs), and appear to be expressed in many tissues; they therefore may play fundamental housekeeping functions that are important in cells in both sexes (Skaletsky et al. 2003). These Y genes may have survived because their loss would leave the heterogametic sex with only a single copy, which would be insufficient for survival. This view requires functional similarity of these X-Y gene pairs, which are often quite similar in nucleotide sequence (Lahn and Page 1997; Graves, Distecho, and Toder 1998). Indeed the human RPS4X and RPS4Y genes have been shown to be functionally interchangeable (Watanabe et al. 1993).

Nevertheless, most X-Y gene pairs have not been closely studied, and some evidence suggests differences may have evolved. For example, the summed expression levels of some X-Y gene pairs in mouse brain are different in males and females and could contribute to sex differences in brain function (Xu, Burgoyne, and Arnold 2002). Moreover, several Y sex chromosome genes, even those with closely related X homologs, have transcripts that are expressed uniquely in male reproductive tissues and therefore may have male-specific functions (Lahn and Page 1997; Skaletsky et al. 2003). Further information is needed to establish similarity or differences in functional roles of X-Y genes.

The W and Z sex chromosomes in birds (females ZW and males ZZ) are analogous to mammalian sex chromosomes in that they differ greatly in gene content and contain highly homologous Z-W genes (Fridolfsson et al. 1998). However, the sex chromosomes of birds contain a complement of genes different from those on the mammalian sex chromosomes, and the heterogametic sex is the female (ZW). Thus, in birds, females contain genes not found in males (i.e., W chromosome genes). An interesting question is whether differences in sex chromosome genes between males and females result in differences in development.

The sex chromosomes of zebra finches (Taeniopygia guttata) are postulated to encode genes that cause sex-specific patterns of brain development (Arnold 1996). Several telencephalic nuclei in this species are highly sexually dimorphic. Males sing a courtship song that females cannot sing, and the neural circuit controlling song is much larger in males than in females (Nottebohm and Arnold 1976). The signals that cause sexually dimorphic development of these telencephalic brain regions are not known. Although sexually dimorphic actions of sex steroid hormones appear to play a role (Gurney and Konishi 1980; Holloway and Clayton 2001), previous studies strongly suggest that sex chromosome genes act in a sex-specific
cell-autonomous manner in the brain to cause sex differences in brain development (Wade and Arnold 1996; Agate et al. 2003). For example, W chromosome genes expressed only in female brain cells may act to prevent masculine development of the song system, or Z genes expressed higher in males may promote masculine brain development.

We have examined the Z-W gene pair CHD1Z and CHD1W (Chromodomain-Helicase-DNA binding protein) (Ellegren 1996; Griffiths, Daan, and Dijkstra 1996) in the zebra finch to compare the structure and brain expression of the two forms. Like X-degenerate genes encoded by the human Y chromosome (Skaletsky et al. 2003), CHD1W may represent a “Z-degenerate” gene that is expressed in many cell types and may serve important housekeeping functions, and which was present on the avian sex chromosomes early in their evolution (as evidenced by the extensive sex chromosome linkage of CHD1 in many avian species (Ellegren 1996; Griffiths, Daan, and Dijkstra 1996; Griffiths et al. 1998)). Our analysis bears on the issue of whether CHD1W has survived on the W sex chromosome because CHD1 is required in two doses as suggested above, or because it may have evolved a critical female-specific function.

CHD was originally discovered in the mouse (CHD1) and was named according to the three conserved domains it encodes: Chromo, Helicase, and DNA binding-domains (Delmas, Stokes, and Perry 1993). These domains are homologous to those found in other proteins that have important roles in chromatin remodeling. For example, the chromodomain is similar to that in heterochromatin protein 1 and polycomb, which functionally repress gene expression (Eissenberg et al. 1990; Messmer, Franke, and Paro 1992; Lorentz et al. 1994). The helicase domain is homologous to that found in SNF2/SWI2 and brahma, which are critical for gene activation (Laurent, Treich, and Carlson 1993; Chiba et al. 1994; Peterson and Workman 2000). The DNA-binding domain contains motifs characteristic of the proteins H1 and HMG 1/Y that act to repress and facilitate expression, respectively (Zlatanova 1990; Liu et al. 2001). Because all three domains are associated with proteins involved in the reorganization of chromatin structure, CHD1 probably functions similarly.

Such a role for CHD1 has been supported through studies involving its homolog in different species. For example, in a CHD1-null yeast strain, a search for complementing genes identified the SNF2/SWI2 gene (Tran et al. 2000). Further analysis demonstrated that CHD1 acts similarly to, but in a manner distinct from, SNF2/SWI2 in modifying chromatin structure (Tran et al. 2000). In the mouse, CHD1 was shown to be a component of bulk chromatin (Stokes and Perry 1995), supporting a role in chromatin organization.

The CHD proteins appear to function as facilitators or inhibitors of gene expression, perhaps depending on species and tissue. For instance, the yeast CHD1-null strain is less sensitive to the toxic effects of 6-azauracil, which was interpreted to mean that CHD1 acts as a general inhibitor of transcription (Woodage et al. 1997). However, in Drosophila, CHD1 is localized to the highly transcriptionally active puff regions of polytene chromosomes, an indication of a facilitative role in gene expression (Stokes, Tartof, and Perry 1996). In mammals, CHD1 co-localizes and interacts with the protein SSRP1, an HMG domain containing protein, which positively regulates the expression of the epsilon-globin gene, also suggesting a facilitative role for CHD1 (Dyer, Hayes, and Baron 1998).

We isolated and compared cDNAs encoding the zebra finch CHD1W and CHD1Z proteins. CHD1W contained two alternate first exons and an alternately spliced transcript within the predicted protein-coding region, neither of which was observed for CHD1Z. The predicted translated proteins, although conserved in their C-H-D coded regions, have diverged significantly in functionally associated regions outside of these domains. Both genes were differentially expressed in the brain and were much more abundant during early development than at other times. In addition, sequence information from several kilobases of the 5′ flanking region suggest differences in regulation.

**Materials and Methods**

**Isolation of CHD1Z and CHD1W cDNAs**

Several fragments of the CHD1Z and CHD1W cDNAs were isolated using a combination of library screening and polymerase chain reaction (PCR) amplification. A CHD1 PCR product (Griffiths and Tiwari 1995) was used to probe a zebra finch ovarian cDNA library to obtain an incomplete CHD1W fragment, chd-ov1, which was used in the screening of a zebra finch telencephalic library to isolate several more incomplete fragments (chd-t32, chd-t4, chd-t31, chd-t30). 5′ and 3′ rapid amplification of cDNA ends (RACE) were then used to isolate missing end fragments (3′ CHD1Z chd-racete3; 3′ CHD1W chd-racev3; 5′ CHD1Z chd-racete35; 5′ CHD1W chd-racete3 and chd-race7) using the protocol of Frohman (1990) or SMART RACE cDNA (Invitrogen).

Overlapping fragments formed the putative full mRNAs for both genes (CHD1W: chd-racete3 [371 bp]/chd-race7 [585 bp], chd-ov1 [1,570 bp], chd-ov2 [2,240 bp], chd-ov3 [2,844 bp], chd-ov8 [813 bp]/chd-ov9 [915 bp]; CHD1Z: chd-t4 [1,280 bp], chd-t32 [2,496 bp], chd-t33 [2,993 bp], chd-racete35 [346 bp], chd-racete36 [924 bp]). CHD1Z and CHD1W products were distinguished through an established sex-specific PCR protocol (Griffiths and Tiwari 1995). Sequence differences between the CHD1Z and CHD1W cDNAs allowed us to assign fragments to the Z or W form. The entire length of the CHD1Z and CHD1W open reading frames (ORFs) was amplified via reverse-transcriptase (RT)-PCR (using primers specific to each form, see Reverse Transcriptranscriptase Polymerase Chain Reactions, below; see also table 1) and restriction mapped to confirm that both encode single transcribed products. All isolated products were sequenced in both directions with the Big Dye terminator sequencing reagent (Applied Biosystems, Foster City, Calif.). All sequence analyses were performed using Vector NTI software (InforMax Inc, Bethesda, Md.). GenBank cDNA accession numbers are as follows: CHD1W transcript A: AY217129, CHD1W transcript B: AY217130, CHD1Z: AY217131.
Southern Blot Analysis

Genomic DNA from both adult male and female zebra finches was digested with each of several enzymes (BglII, EcoRI, EcoRV, HindIII), 10 µl resolved by agarose gel electrophoresis and transferred to a nylon membrane (Hybond-N+, Amersham Biosciences Corp., Piscataway, N.J.) and UV cross-linked. The random-primed CHD1W fragment chd-ov3 was hybridized to the membrane overnight at 70°C then rinsed in 2x SSC/1% sodium dodecyl sulfate (SDS), washed twice for 30 min. each in 0.1x SSC/1% SDS at 50°C, and then exposed to autoradiographic film.

Northern Blot Analysis

cDNA probes used in northern blot analyses were designed to distinguish between the two CHD1 forms (CHD1Z: chd-racet35 and CHD1W: chd329, see figures 1 and 2 of the Supplementary Material online for probe sequences). Telencephalic mRNA (3 µg polyA+) from three pooled adult males and three pooled adult females were separated on a 1.25% agarose gel, transferred to a Hybond-N+ nylon membrane, UV cross-linked, then probed with 32P end-labeled and hybridized with the oligo-specific ULTRAhyb reagent. Blots were exposed and expression was quantified relative to GAPDH using a PhosphorImager (Amersham Biosciences Corp., Piscataway, N.J.). Group differences were analyzed by one-way or two-way ANOVAs followed by Tukey-Kramer posthoc tests using NCSS software (www.ncss.com). Alpha level was set at P < 0.05.

Reverse-Transcriptase Polymerase Chain Reaction

Amplification of the complete ORF was performed using a nested PCR protocol with pairs of primers specific to CHD1Z or CHD1W (table 1). Total RNA from adult male and female telencephalon was reverse transcribed (RT primer = chd5850r, table 1) with the Superscript II reverse-transcriptase enzyme (Invitrogen Corp., Carlsbad, Calif.). The initial PCR was performed with primers CHD1Zf/CHD1Z5660r or CHD1W45f/CHD1W5645r and Advantage 2 taq DNA polymerase (BD Biosciences Clontech, Palo Alto, Calif.) in a RoboCycler Gradient 96 (Stratagene Corp., La Jolla, Calif.; 1 cycle, 95°C, 5 min.; 20 cycles, 95°C, 40 s, 56°C, 40 s, 72°C, 5 min; and 1 cycle, 72°C, 5 min). The nested PCR was performed with

Table 1
Primer and Oligonucleotides Used in Different Procedures as Described in Materials and Methods

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>chdwp239-2</td>
<td>TAGAAAGTGCGACGACTAGC</td>
</tr>
<tr>
<td>chdz334r</td>
<td>ATCATACTCTCTGTCCTTCC</td>
</tr>
<tr>
<td>chdz-warF</td>
<td>TCTGCACATCTAGAGAAGA</td>
</tr>
<tr>
<td>chdz-varR</td>
<td>ACATTGCGAGTCTTCTACTT</td>
</tr>
<tr>
<td>chdz5750f</td>
<td>TTTTGGCTGAGCTGTGCTT</td>
</tr>
<tr>
<td>chdz6606r</td>
<td>GTTGGAGTTGGGTCTGTTAAGAAG</td>
</tr>
<tr>
<td>chdz-60f</td>
<td>CTTGATCTTTTCAAGTGCTTAGAACAT</td>
</tr>
<tr>
<td>chdz5650r</td>
<td>AAGCCTAAAGACGACATACAGGTC</td>
</tr>
<tr>
<td>chdz-45f</td>
<td>GCAATTTGGAAAAATCTGCGAGA</td>
</tr>
<tr>
<td>chdz-645r</td>
<td>ACTGTTGCTAACATTTCCAGTCCA</td>
</tr>
<tr>
<td>chdz-70f</td>
<td>GAACTGTTGCTACCTGGAAGAA</td>
</tr>
<tr>
<td>chdz-5580r</td>
<td>GTATATGCTCAAAGGATGTCAGAG</td>
</tr>
<tr>
<td>oligo 1</td>
<td>CATGAGCTGACGACCCGAGATCATCATCGAG</td>
</tr>
<tr>
<td>oligo 2</td>
<td>TTTTGGCTTATTGAGGTTATCTTATTACATCAGTCAAGG</td>
</tr>
<tr>
<td>oligo 3</td>
<td>ATGTGCTAGCAAGTTTCTT</td>
</tr>
<tr>
<td>oligo 4</td>
<td>TTTTCTTCATACCTTGTGGAAGGAGT</td>
</tr>
<tr>
<td>oligo 5</td>
<td>ATCTTTTCCAGCGTCCATA</td>
</tr>
<tr>
<td>oligo 6</td>
<td>GAAGGAGTGATCTTGAGTATTTCCAGTGGTCT TG</td>
</tr>
<tr>
<td>gchdz-f2 (MluI)</td>
<td>TTAGCGCGT-TGGGCGTACACCCCACTTCAGGCT</td>
</tr>
<tr>
<td>gchdz-v1 (HindIII)</td>
<td>TTAAGCTT-GACTCAACGTTCTAGGAGATCAGGATCCATTTTT</td>
</tr>
<tr>
<td>gchdz-f2 (Xhol)</td>
<td>TCTCTAGA-CGCCGTAACCCCAATCCCAGATG</td>
</tr>
<tr>
<td>gchdz-r12 (HindIII)</td>
<td>TTAAGCTT-GTTCGCTTAGTTCGTCGTCGAG</td>
</tr>
</tbody>
</table>

Note.—The last four primers had restriction sites (indicated) incorporated into the 5’ end and represent the nucleotides before the hyphen in each sequence.
Amplification of the alternately spliced region of CHD1W and the corresponding region of CHD1Z began with the reverse transcription (reagents and conditions as above) of total RNA isolated from adult male telencephalon for CHD1Z (RT primer = CHD1Z834r) and adult female telencephalon for CHD1W (RT primer = CHD1Wp239-2). PCR was performed using Titanium Taq DNA polymerase and reagents (BD Biosciences Clontech) and primers (CHD1Zw-varF/CHD1Zw-varR; table 1) that recognized CHD1Z and CHD1W equally (1 cycle, 95°C, 5 min; 40 cycles, 95°C, 0 s, 57°C, 0 s, 72°C, 1 min; and 1 cycle, 72°C, 5 min).

Amplification of several male and female adult tissues (gonads, telencephalon, liver, and muscle) for CHD1Z-specific and CHD1W-specific expression was performed with primers (CHD1Zw5750f/CHD1Zw5850r) that amplify a similar region of the two genes that were distinguishable by size (CHD1Z: 103 bp; CHD1W: 82 bp). Total RNA (2 μg) from each tissue was reverse-transcribed with oligo (dT) as primer and Invitrogen’s Superscript II reverse transcriptase and buffers; it was then PCR amplified (1 cycle, 95°C, 5 min; 35 cycles, 95°C, 40 s, 54°C, 30 s, 72°C, 40 s; and 1 cycle, 72°C, 5 min).

In Situ Hybridization

In situ hybridization was carried out as previously described (Jacobs, Arnold, and Campagnoni 1999). Brain tissue from adult male and female zebra finches (two of each) was collected after deeply anesthetizing the birds with Equithesin and perfusing them intracardially with PBS (100 mM NaPO4 pH 7.2, 0.15 M NaCl) followed by 4% paraformaldehyde. Tissue was postfixied for 1 day, cryoprotected using 20% sucrose in PBS overnight, then sectioned coronally at 18 μm and hybridized with 32P-UTP labeled RNA probes. Sense (chd-t32) and anti-sense (chd-t32 and CHD1W35) probes were labeled in an in vitro transcription. CHD1W35 is a 346-bp fragment generated by PCR and recognizes 5′CHD1W where CHD1Z and CHD1W have diverged in sequence. A sense probe was not generated for CHD1W35 because it is expressed only in females. Male tissue was used as a negative control.

Genomic Sequence

Five prime flanking regions of CHD1Z and CHD1W were isolated by using the Genome Walker kit (BD Biosciences Clontech). Libraries were generated according to the kit protocol and using NcoI, BamHI, BglIII, Stul,
PvuII, XhoI, XbaI, and HindIII. For CHD1Z, the first product (634 bp) was obtained from the BglIII library; the second (1,383 bp), from the StuI library; and a third (2,278 bp), from the HindIII library. The CHD1Z genomic fragments were concatenated to give 4.092 kb of sequence information upstream of the putative transcription start site (TSS).

For CHD1W, the 5’ RACE results indicated that there were two different exons at the 5’ end, resulting in two different TSSs. We used primers to amplify the genomic sequence from the more 5’ exon to the shared downstream exon. This product yielded a 1,811-bp genomic fragment that contained both exons. Sequence from this product was then used to isolate a more 5’ product (2,882 bp) from the BamHI Genome Walker library. The overall size of the concatenated CHD1W genomic fragment was 4,535 kb upstream of a similar point to the TSS described for CHD1Z. The identification of avian transcription factor binding sites within each sequence was performed with Tsitescan (Institute for Transcriptional Informatics, Pittsburgh, Pa.). GenBank genomic accession numbers: CHD1W 5’ flanking genomic sequence: AY217132; CHD1Z 5’ flanking genomic sequence: AY217133.

CpG Island Region Analysis

CpG island regions were analyzed for CHD1Z and CHD1W and for X-Y homologous gene pairs from humans by aligning (using Vector NTI software) several kilobases of genomic DNA containing the islands. All CpG dinucleotides within this alignment were then highlighted, and the region where the highest concentration of CpG dinucleotides plus 100 to 200 bp on either end was used in the comparison. The genomic sequences were identified by using cDNA sequences for each gene (accession numbers are given in the Supplementary Material online) to search the human genome (www.ensemble.org) with BLAST software (Altschul et al. 1990). Several kilobases of genomic sequence containing a CpG island upstream of the 5’ end of this alignment were identified and used for the gene pair analysis.

Transfections

Using PCR we constructed a single genomic fragment representing the 5’ flanking region for each gene (CHD1Z primers: gCHD1Z-f2, gCHD1Z-r2; CHD1W primers: gCHD1W-f2, gCHD1W-r1; table 1). The products (CHD1Z 4.092 kb and CHD1W 4.535 kb) were directionally subcloned into the pGL2 basic vector containing a luciferase reporter (Promega Corp., Madison, Wis.) and designated gCHD1Z-pgl2 and gCHD1W-pgl2. Cell cultures were prepared from P1 male and female zebra finch telencephalon (three males, three females) and fibroblasts from body tissues (two males, three females) according to methods outlined previously (Schlinger et al. 1994). Cells from each telencephalon and body were distributed into 12 wells of a 24-well plate (500 µl/well) and grown until confluent.

Transient transfections were carried out with the LipofectAMINE 2000 reagent used as described by the manufacturer (Invitrogen). For the telencephalic tissue, three different reporter constructs were used in the transfection: a pgl2 vector (Promega), a gCHD1Z-pgl2 construct, and a gCHD1W-pgl2 construct (four wells each, 600 ng/well) (2 µl lipofectAMINE per well, no antibiotics). Transfection was allowed to proceed for 40 h. Media was removed and cells were washed with PBS (no Ca2+ or Mg2+). Cells in each well were incubated with 50 µl of 1X luciferase lysis buffer (Luciferase assay and detection system, Pharmingen, San Diego, Calif.) at room temperature for 20 min; 40 µl of the cellular lysate was used to measure luciferase activity with a Monolight 2010 luminometer (Analytical Luminescence Laboratory, San Diego, Calif.). For the cultures from body tissue, the same protocol as above was followed. The expression of each construct (gCHD1Z-pgl2 and gCHD1W-pgl2) in the cultures was compared with the expression of the vector (pgl2) alone and plotted as the fold difference.

Results

Southern blot analysis of genomic DNA from zebra finch males and females probed with chd-ov3 yielded several bands shared between males and females, as well as a number of female-specific bands (fig. 1A). The greater intensity of some bands in males, as well as the presence of female-specific bands, suggested that the CHD1 genes we report on here are linked to the Z and W chromosomes in zebra finches, as has been shown by FISH in chickens (Griffiths and Korn 1997) and suggested by Southern and PCR analysis in several avian species (Ellegren 1996; Griffiths, Daan, and Dijkstra 1996; Griffiths et al. 1998).

Although RT-PCR amplification of the complete ORF demonstrated that cloned fragments were transcribed as a single mRNA (see diagrams in fig. 2B and C), CHD1W but not CHD1Z showed alternately transcribed products. The transcripts contained ORFs encoding...
FIG. 3.—Comparison of the CHD1Z and CHD1W predicted protein products. C, H, and D domains are underlined. Bolded regions represent differences in CHD1Z and CHD1W amino acid coding. The boxed region represents the alternately spliced 34 amino acids that are absent in some CHD1W transcripts.

Structure and Expression of CHD1 Genes in Zebra Finches 389
a predicted protein of 1,806 amino acids for CHD1Z and proteins of 1,786 and 1,752 amino acids for CHD1W (fig. 3). The overall identity between the CHD1Z and CHD1W was 90% at the nucleic acid level and 91% at the amino acid level. Each cDNA encoded the characteristic C-H-D domains (fig. 3; also see figs. 1 and 2 in the Supplementary Material online), which were more highly conserved (98% aa identity) than other regions (aa identity 77% in 5' region, and 85% 3' region). An overall comparison of CHD1Z and CHD1W to the GenBank protein database showed that these genes were most similar to the chicken CHD1Z (95% for TgCHD1Z and 92% for TgCHD1W) and the mammalian CHD1 (86% to human CHD1 for TgCHD1Z, and 83% to human CHD1 for TgCHD1W) (Delmas, Stokes, and Perry 1993; Woodage et al. 1997).

The expression of CHD1Z and CHD1W was determined in various tissues and in brain at different ages. Both genes appeared to be expressed in a wide variety of tissues and cell types when examined by RT-PCR (fig. 4A). In situ hybridization in the brain using a CHD1Z antisense probe that hybridizes to both Z and W forms also showed a wide distribution of expression. However, expression appeared to be greater in the cerebellum, ectlstriatum, paleostriatum, and layers of the optic tectum, suggesting regional differences in regulation (see fig. 3B in the Supplementary Material online). No differences in the distribution of CHD1Z and CHD1W expression were detected between males and females at this level. However, CHD1W expression was limited to females, as expected (see fig. 3A in the Supplementary Material online). Northern blot analysis confirmed the female-specific expression of CHD1W in adult brain (fig. 1C), and also indicated that the expression of CHD1Z was higher in the adult male brain than in female brain (fig. 1B). Additionally, the alternately spliced CHD1W product was detected through RT-PCR and found to be unique to females (fig. 4B).

Expression of the two mRNAs was quantified at P2, P6, P10, and P14 by Northern blot analysis. Both mRNAs were expressed at significantly higher levels at P2 and P6 than at P10 and P14 (P14 not shown) (fig. 5A). In addition, CHD1Z expression was significantly higher in males than in females (fig. 5A, ANOVA, P < 0.000001). Two CHD1W transcripts were detected at approximately 7.5 kb and 8.5 kb, but only the 7.5-kb transcript of CHD1Z was observed in males (fig. 1B and D). This difference may, however, have resulted because of the different regions recognized by the CHD1Z (5' end) and CHD1W (3' end) probes (see figs. 1 and 2 of the Supplementary Material online).

In females, the expression of CHD1Z and CHD1W mRNAs each differed as a function of age. CHD1W declined over P2, P6, and P10, whereas the expression of CHD1Z mRNA increased from P2 to P6 and then declined (fig. 5A and B). To compare expression of Z and W forms in females, the values were expressed as a percentage of the maximum value for each probe; this served to eliminate artificial differences in expression values from differences in probe characteristics. In a two-way ANOVA with factors of age (P2, P6, or P10) and linkage (Z vs. W), there was a significant effect of age (P < 0.000001, and no significant interaction between linkage and day (P = 0.07). A similar ANOVA was run to compare expression of CHD1W with summed expression of both CHD1Z and CHD1W (the latter determined using oligonucleotide probes that recognize the two forms equally; fig. 5C). In this case the two factors were age and CHD1 type (CHD1W vs. total). There was a highly significant effect of age (P < 0.000001), and a significant interaction of the two factors (P < 0.002). The interaction means that the developmental change in expression of CHD1W differed significantly from the summed expression of both CHD1 mRNAs. A similar comparison of expression of CHD1Z and total CHD1 expression showed no significant interaction.

Genomic sequence containing the 5' flanking region of the two genes was compared to determine whether differences in sequence suggested differences in regulation. For CHD1Z, the 4,092-bp sequence (fig. 6A) upstream of the putative transcription start site was compared to 4,535 bp of CHD1W sequence upstream of a similar point (fig. 6B). The two alternate TSSs discovered for CHD1W in the 5' RACE were identified in the genomic sequence (fig. 6B, exons 1a and 1b). Both CHD1Z and CHD1W 5' flanking regions contained a local, high concentration of CpG dinucleotides (fig. 6, highlighted), which met the criteria for CpG islands (length > 200 bp; G + C > 50%; and observed/expected CpGs ≥ 0.6) (Gardiner-Garden and Frommer 1987). The CpG island of CHD1Z had a length of approximately 1,500 bp, a GC content of 71%, and an obs/exp CpG ratio of 0.96; for CHD1W, the length was also approximately 1,500 bp with a 63% GC content and an obs/exp CpG ratio of 0.68.
Within the CpG islands the number of CpG dinucleotides for CHD1Z (208) was nearly double that for CHD1W (105). The CpG island of CHD1W, but not CHD1Z, overlapped the 5' UTR (untranslated region).

Cultured brain cells and body fibroblasts were transfected to determine the promoting ability of the two 5' genomic sequences, measured relative to that of the vector alone and plotted as the fold difference. The CHD1Z construct drove expression over 100-fold relative to the vector in both brain and body tissues (fig. 7A). The CHD1W construct also increased expression, but less in brain tissue than in body tissue (fig. 7B). The promoting abilities of the CHD1Z and CHD1W constructs were not directly compared because the constructs differed in size and may not have encoded the complete promoter regions.

The greater number of CpG dinucleotides in the 5' flanking region of CHD1Z suggested that similar differences in CpG islands might be found in X-Y gene pairs. An analysis of the CpG island regions in the 5' flanking regions of paired human X-Y sex chromosome genes confirmed this hypothesis. The number of CpG dinucleotides in X-Y gene pairs was typically greater for the X gene than for its Y homolog (table 2), suggesting that genes on the minor sex chromosome—W in birds or Y in mammals—generally contain fewer CpG islands.

The 5' flanking sequences did not contain the clear elements for binding of the basal transcription machinery that are commonly found in other gene promoters (i.e., TATA and CAAT boxes). However, several putative regulatory elements were identified in each gene (fig. 6). For example, CHD1Z contained a 16-bp poly(dG) sequence, seven putative SP1 binding elements (GGCGGG or CCGCCC), and a palindromic sequence quite similar to those found in the promoter of the chicken β-globin gene ([Lewis et al. 1988]. CHD1W contained three putative BetaE-F3 elements (GGGTGGG) and a single SP1 element (GGCGGG). An alignment between the sequences (4,092 bp [CHD1Z] to 4,535 bp [CHD1W]) oriented them in a manner such that the CpG islands of the two genes overlapped (see fig. 4 of the Supplementary Material online). The overall identity between the sequences was 51% over the aligned region and 56% over the CpG island region. Several tandem blocks of sequence appeared to have maintained a relatively greater degree of conservation. Most of these were found within the CpG island region and either coincided with or were close to the CHD1W 5' exons (underlined in fig. 4 of the Supplementary Material online). These blocks did not appear to code for known avian trans- or cis-acting promoter elements.

Discussion

We analyzed the sex chromosome genes CHD1Z and CHD1W in zebra finches to cast light on the question of whether the two genes might have diverged sufficiently so that CHD1W evolved a female-specific function, or that CHD1Z might have evolved a male-specific function because of its higher dose in males. The avian CHD1 genes are estimated to have begun diverging approximately 123 MYA (Garcia-Moreno and Mindell 2000), giving sex-specific evolutionary pressures a considerable amount of time to act on them. We hypothesized that functional differences in the two genes would be reflected in differences in their structure, their promoter regions, and...
FIG. 6.—Genomic sequences 5' to the TSS of CHD1Z (A) and from a comparable position for CHD1W (B). CpG dinucleotides are highlighted; underlined sequences are exons; boxed sequences are putative SP1 binding sites. The double underlined sequence in (A) is a putative binding site for BGP1, and the double underlined sequences in (B) are putative BetaE-F3 binding sites.
or their expression. Our results show a number of substantial differences between CHD1Z and CHD1W that could imply functional differences, and that therefore are consistent with the idea that CHD1W may have survived on the W sex chromosome in part because it evolved a female-specific function.

We report for the first time the full coding sequence of an avian CHD1W gene, as well as the first comparison of the entire coding regions of CHD1W and CHD1Z. The cDNA sequences of the two genes were conserved (~90% relative to each other) except at the 5′ end, where CHD1W contained two different noncoding exons not observed for CHD1Z, as well as an alternately spliced exon in the ORF that was not produced by CHD1Z. The predicted proteins encoded by these genes were highly similar within the C-H-D domains, indicating that they are functionally conserved. An analysis of the C-H-D domains across several species has been used to conclude that CHD1Z and CHD1W overall are functionally conserved (Fridolfsson and Ellegren 2000). However, this analysis was limited to these domains. The degree of conservation is lower outside the C-H-D domains, where roughly half of the coding sequence is found (97%, C-H-D region; 77%, 5′ region; and 85%, 3′ region). The similarity is even lower in one CHD1W isoform that lacks 34 amino acids at its 5′ end, unlike CHD1Z. Although no functional role has been clearly assigned to regions outside the C-H-D domains, the 5′ region in the mammalian CHD1 has been shown to interact and co-localize with the HMG box containing protein SSRP1 (Kelley, Stokes, and Perry 1999), suggesting that this region may be functionally important. If so, the decreased homology in this region between CHD1Z and CHD1W may impart functional differences.

The expression of CHD1Z and CHD1W was examined in the brain through northern blot analysis and in situ hybridization. In situ hybridization on adult brain tissue of males and females did not reveal differences in the distribution of CHD1Z and CHD1W expression. In addition, both CHD1Z and CHD1W were widely expressed in the brain, with somewhat higher expression in specific regions. The expression of CHD1Z and CHD1W from the northern blots of the early developing telencephalon showed that CHD1Z was highest in the telencephalon of males and females at P2 and P6, but significantly lower by P10 and P14. Moreover, the expression in males was significantly greater than in females. This indicates that dosage compensation mechanisms, if present, do not successfully balance the expression of CHD1Z between males and females. Dosage compensation in birds has been proposed but remains incompletely resolved (McQueen et al. 2001; Teranishi et al. 2001; Ellegren 2002). The lack of CHD1Z normalization between males and females can be explained if CHD1Z and CHD1W are functionally the same, so that the combined expression of CHD1Z and CHD1W in females is equivalent to the expression of two copies of CHD1Z in males. This idea is supported by our northern blot analysis that used several oligonucleotides that recognize CHD1Z and CHD1W equally. That analysis showed that the combined expression of CHD1Z and CHD1W was equal to the expression of two copies of CHD1Z during P2 to P10.

Differences in CHD1Z and CHD1W expression in females between P2 and P14 suggested, however, that they were regulated differently. The expression of CHD1W decreased consistently from P2 to P10, whereas CHD1Z expression increased significantly from P2 to P6 and then declined by P10 and P14. The dissimilar expression profile suggests differences in the regulatory elements controlling these genes. This idea was further supported by the finding that the CHD1W expression profile was significantly different from the total CHD1 expressed during this period, but CHD1Z expression was not.

The putative differences in the regulation of CHD1Z and CHD1W are likely to be encoded in the promoter region. We therefore compared the 5′ proximal region of the two genes. The ability of both flanking regions to drive the expression of a luciferase reporter suggested that the constructs encode at least some of the promoter. The several kilobases of 5′ sequence analyzed were not well conserved (51% identity overall). Although the percent homology of non-coding sequence is difficult to interpret, regulatory elements important for controlling gene expression are expected to be conserved and depicted in the alignment. If the two genes are regulated by the same mechanisms, these regulatory elements should be shared. In CHD1Z, we found several putative regulatory elements.
quite similar to those found in the β-globin gene promoter (Lewis et al. 1988). Although CHD1W did not contain these elements, it did have two β-globin enhancer sequences (BetaE-F3) that may serve a regulatory role unique to CHD1W. The two 5′ exons of CHD1W identified by 5′ RACE were not found in the genomic sequence isolated for CHD1Z, and were not products of the CHD1Z 5′ RACE. The putative first exons for each gene were also located in different genomic positions. Together, these findings suggest a substantial divergence of the promoter regions of these genes, which would be expected to alter promoter strength and thus expression. These findings could explain differences in the expression pattern observed for CHD1Z and CHD1W during early development.

Our analysis of the 5′ flanking regions of CHD1Z and CHD1W also revealed that each gene contained a CpG island. CpG islands are commonly found at the 5′ end of many genes and in the vicinity of the transcription start site (Antequera and Bird 1993). This finding supports the view that the sequences we isolated are likely part of the CpG island region only. The two 5′ exons of CHD1W that were not found in the genomic sequence isolated for CHD1Z, and were not products of the CHD1Z 5′ RACE. The putative first exons for each gene were also located in different genomic positions. Together, these findings suggest a substantial divergence of the promoter regions of these genes, which would be expected to alter promoter strength and thus expression. These findings could explain differences in the expression pattern observed for CHD1Z and CHD1W during early development.

Our analysis of the 5′ flanking regions of CHD1Z and CHD1W also revealed that each gene contained a CpG island. CpG islands are commonly found at the 5′ end of many genes and in the vicinity of the transcription start site (Antequera and Bird 1993). This finding supports the view that the sequences we isolated are likely part of the CpG island region only. The two 5′ exons of CHD1W that were not found in the genomic sequence isolated for CHD1Z, and were not products of the CHD1Z 5′ RACE. The putative first exons for each gene were also located in different genomic positions. Together, these findings suggest a substantial divergence of the promoter regions of these genes, which would be expected to alter promoter strength and thus expression. These findings could explain differences in the expression pattern observed for CHD1Z and CHD1W during early development.

### Table 2
Comparison of CpG Island Regions of Sex Chromosome Gene Pairs

<table>
<thead>
<tr>
<th>Gene</th>
<th>X or Z (CpG #)</th>
<th>Y or W (CpG #)</th>
<th>Ratio XZ/YW</th>
<th>% G + C</th>
<th>Region</th>
<th>Aligned Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHD1Z/CHD1W</td>
<td>155</td>
<td>98</td>
<td>1.58</td>
<td>68/64</td>
<td>1.5</td>
<td>56</td>
</tr>
<tr>
<td>SOX3/SRY</td>
<td>144</td>
<td>36</td>
<td>4.00</td>
<td>72/45</td>
<td>1.1</td>
<td>47</td>
</tr>
<tr>
<td>UTX/UTY</td>
<td>181</td>
<td>70</td>
<td>2.59</td>
<td>70/54</td>
<td>1.5</td>
<td>63</td>
</tr>
<tr>
<td>EIF1AX(EIF1AY</td>
<td>105</td>
<td>47</td>
<td>2.23</td>
<td>62/52</td>
<td>1.3</td>
<td>58</td>
</tr>
<tr>
<td>SMCX/SMCY</td>
<td>90</td>
<td>57</td>
<td>1.58</td>
<td>62/46</td>
<td>1.2</td>
<td>48</td>
</tr>
<tr>
<td>RPS4X/RPS4Y</td>
<td>77</td>
<td>52</td>
<td>1.48</td>
<td>56/47</td>
<td>1.3</td>
<td>52</td>
</tr>
<tr>
<td>ZFX/ZFY</td>
<td>197</td>
<td>146</td>
<td>1.35</td>
<td>76/56</td>
<td>1.5</td>
<td>77</td>
</tr>
<tr>
<td>TBL1X/TBL1Y</td>
<td>183</td>
<td>140</td>
<td>1.31</td>
<td>73/67</td>
<td>1.6</td>
<td>74</td>
</tr>
<tr>
<td>PCDH11X/PCDH11Y</td>
<td>102</td>
<td>101</td>
<td>1.01</td>
<td>54/54</td>
<td>1.7</td>
<td>98</td>
</tr>
</tbody>
</table>

Note.—Boldface indicates CHD1Z and CHD1W, the zebra finch sex chromosome genes. All others are from the human X-Y sex chromosomes. All values represent the CpG island region only.

### Gene X or Z (CpG #) Y or W (CpG #) Ratio XZ/YW % G + C Region Aligned Region

### Table 2
Comparison of CpG Island Regions of Sex Chromosome Gene Pairs

<table>
<thead>
<tr>
<th>Gene</th>
<th>X or Z (CpG #)</th>
<th>Y or W (CpG #)</th>
<th>Ratio XZ/YW</th>
<th>% G + C</th>
<th>Region</th>
<th>Aligned Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHD1Z/CHD1W</td>
<td>155</td>
<td>98</td>
<td>1.58</td>
<td>68/64</td>
<td>1.5</td>
<td>56</td>
</tr>
<tr>
<td>SOX3/SRY</td>
<td>144</td>
<td>36</td>
<td>4.00</td>
<td>72/45</td>
<td>1.1</td>
<td>47</td>
</tr>
<tr>
<td>UTX/UTY</td>
<td>181</td>
<td>70</td>
<td>2.59</td>
<td>70/54</td>
<td>1.5</td>
<td>63</td>
</tr>
<tr>
<td>EIF1AX(EIF1AY</td>
<td>105</td>
<td>47</td>
<td>2.23</td>
<td>62/52</td>
<td>1.3</td>
<td>58</td>
</tr>
<tr>
<td>SMCX/SMCY</td>
<td>90</td>
<td>57</td>
<td>1.58</td>
<td>62/46</td>
<td>1.2</td>
<td>48</td>
</tr>
<tr>
<td>RPS4X/RPS4Y</td>
<td>77</td>
<td>52</td>
<td>1.48</td>
<td>56/47</td>
<td>1.3</td>
<td>52</td>
</tr>
<tr>
<td>ZFX/ZFY</td>
<td>197</td>
<td>146</td>
<td>1.35</td>
<td>76/56</td>
<td>1.5</td>
<td>77</td>
</tr>
<tr>
<td>TBL1X/TBL1Y</td>
<td>183</td>
<td>140</td>
<td>1.31</td>
<td>73/67</td>
<td>1.6</td>
<td>74</td>
</tr>
<tr>
<td>PCDH11X/PCDH11Y</td>
<td>102</td>
<td>101</td>
<td>1.01</td>
<td>54/54</td>
<td>1.7</td>
<td>98</td>
</tr>
</tbody>
</table>

Note.—Boldface indicates CHD1Z and CHD1W, the zebra finch sex chromosome genes. All others are from the human X-Y sex chromosomes. All values represent the CpG island region only.

### Table 2
Comparison of CpG Island Regions of Sex Chromosome Gene Pairs

<table>
<thead>
<tr>
<th>Gene</th>
<th>X or Z (CpG #)</th>
<th>Y or W (CpG #)</th>
<th>Ratio XZ/YW</th>
<th>% G + C</th>
<th>Region</th>
<th>Aligned Region</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHD1Z/CHD1W</td>
<td>155</td>
<td>98</td>
<td>1.58</td>
<td>68/64</td>
<td>1.5</td>
<td>56</td>
</tr>
<tr>
<td>SOX3/SRY</td>
<td>144</td>
<td>36</td>
<td>4.00</td>
<td>72/45</td>
<td>1.1</td>
<td>47</td>
</tr>
<tr>
<td>UTX/UTY</td>
<td>181</td>
<td>70</td>
<td>2.59</td>
<td>70/54</td>
<td>1.5</td>
<td>63</td>
</tr>
<tr>
<td>EIF1AX(EIF1AY</td>
<td>105</td>
<td>47</td>
<td>2.23</td>
<td>62/52</td>
<td>1.3</td>
<td>58</td>
</tr>
<tr>
<td>SMCX/SMCY</td>
<td>90</td>
<td>57</td>
<td>1.58</td>
<td>62/46</td>
<td>1.2</td>
<td>48</td>
</tr>
<tr>
<td>RPS4X/RPS4Y</td>
<td>77</td>
<td>52</td>
<td>1.48</td>
<td>56/47</td>
<td>1.3</td>
<td>52</td>
</tr>
<tr>
<td>ZFX/ZFY</td>
<td>197</td>
<td>146</td>
<td>1.35</td>
<td>76/56</td>
<td>1.5</td>
<td>77</td>
</tr>
<tr>
<td>TBL1X/TBL1Y</td>
<td>183</td>
<td>140</td>
<td>1.31</td>
<td>73/67</td>
<td>1.6</td>
<td>74</td>
</tr>
<tr>
<td>PCDH11X/PCDH11Y</td>
<td>102</td>
<td>101</td>
<td>1.01</td>
<td>54/54</td>
<td>1.7</td>
<td>98</td>
</tr>
</tbody>
</table>

Note.—Boldface indicates CHD1Z and CHD1W, the zebra finch sex chromosome genes. All others are from the human X-Y sex chromosomes. All values represent the CpG island region only.
tissues does not preclude an organ-specific role in sexual differentiation of specific tissues. Finally, if the divergence of CHD1W and CHD1Z in fact causes differences in ZZ vs. ZW cells, the sex differences may or may not be relevant to the large sexual dimorphisms in the telencephalic song nuclei. Rather, they may imply differences in the function of all ZZ and ZW cells and thus contribute to non-specific sex differences in cell function. Because of the putative role of CHD1 in chromatin remodeling, sex differences could influence the transcription of numerous genes.

Supplementary Material

Available online:

CHD1W transcript A: AY217129
CHD1W transcript B: AY217130
CHD1Z: AY217131
CHD1W 5’ flanking genomic sequence: AY217132-CHD1Z 5’ flanking genomic sequence: AY217133

Human sequences for X-Y cDNAs:

EIF1AX: NM_001412; EIF1AY: NM_004681; ZFX: NM_003410; ZFY: NM003414; TBL1X: NM_005647; TBL1Y: NM_033284; UTX: NM_021140; UTY: NM_001007; RPS4X: NM_001008; SMCX: NM_004187; SMCY: NM_004653; SOX3: NM_005634; SRY: NM_003140; PCDH11X: NM_007125; RPS4Y: NM_001008; SMCH: NM_004187; SMCY: NM_004653; SOX3: NM_005634; SRY: NM_003140; PCDH11X: NM_001452; PCDH11Y: NM_032971.

Figures 1, 2, 3, and 4.

Acknowledgments

We are indebted to Anthony Campagnoni for advice and assistance throughout this project, and we also thank Richard Griffiths, Allison Wong, and Valerie Sugiyama for their help. This work was supported by National Institutes of Health grant DC00217.

Literature Cited


Scott Edwards, Associate Editor

Accepted September 29, 2003