Molecular Evolution of the Avian *CHD1* Genes on the Z and W Sex Chromosomes

Anna-Karin Fridolfsson and Hans Ellegren

Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, SE-752-36 Uppsala, Sweden

Manuscript received December 8, 1999
Accepted for publication April 14, 2000

ABSTRACT

Genes shared between the nonrecombining parts of the two types of sex chromosomes offer a potential means to study the molecular evolution of the same gene exposed to different genomic environments. We have analyzed the molecular evolution of the coding sequence of the first pair of genes found to be shared by the avian Z (present in both sexes) and W (female-specific) sex chromosomes, *CHD1Z* and *CHD1W*. We show here that these two genes evolve independently but are highly conserved at nucleotide as well as amino acid levels, thus not indicating a female-specific role of the *CHD1W* gene. From comparisons of sequence data from three avian lineages, the frequency of nonsynonymous substitutions (K_a) was found to be higher for *CHD1W* (1.55 per 100 sites) than for *CHD1Z* (0.81), while the opposite was found for synonymous substitutions (K_s, 13.5 *vs.* 22.7). We argue that the lower effective population size and the absence of recombination on the W chromosome will generally imply that nonsynonymous substitutions accumulate faster on this chromosome than on the Z chromosome. The same should be true for the Y chromosome relative to the X chromosome in XY systems. Our data are compatible with a male-biased mutation rate, manifested by the faster rate of neutral evolution (synonymous substitutions) on the Z chromosome than on the female-specific W chromosome.

THE underlying factors affecting the molecular evolution of sex-linked genes differ in some important ways as compared to those governing the evolution of autosomal genes. First, the effective population size of sex-linked genes is always smaller than that of autosomal genes, implying different fixation probabilities of a given selection coefficient (Charlesworth *et al.* 1987; Li 1997). Second, while autosomal genes spend an equal amount of time in the male as in the female germline, sex-linked genes show a bias with respect to their transmission through the two sexes. The mammalian X chromosome, for example, is two-thirds of the time in the female germline. Moreover, genes from the nonrecombining part of one of the sex chromosomes are exclusively transmitted by a single sex. This means that if the patterns of mutation or selection differ between sexes, sex-linked genes will evolve in a "sex-biased" fashion (Miyata *et al.* 1987). For instance, several lines of evidence from several organisms indicate that the mutation rate of males is higher than that of females, a situation commonly attributed to the many more mitotic germline cell divisions in spermatogenesis than in oogenesis (Miyata *et al.* 1987; Shimmin *et al.* 1993; Ellegren and Fridolfsson 1997; Hurst and Ellegren 1998). Moreover, the degree of methylation of CpG sites, which increases the mutability of such sites, may differ between sexes (Driscoll and Migeon 1990). Third, while there as yet is no evidence for chromosome-specific mutation rates of vertebrate autosomes, a lowered mutation rate has been suggested for the mammalian X chromosome, which could be adaptive by reducing the effect of slightly deleterious mutations being exposed in hemizygote males (McVean and Hurst 1997). Fourth, possible dosage and dominance effects might act differentially on genes on sex chromosomes (Charlesworth *et al.* 1987) and their recombination rates may also differ. Thus, a number of sex- or chromosome-specific factors may be manifested in the molecular evolution of sex-linked genes.

Since the selection pressure on individual genes varies enormously, empirically addressing the effects of sex- and chromosome-specific factors in molecular evolutionary processes ideally requires analyses of the same gene exposed to different genomic environments. This is obviously not possible for single-copy genes and most multigene families are either autosomal or sex-linked, at least with respect to expressed gene copies. However, a very special class of genes offers a possibility to study these factors, namely, genes shared between the nonrecombining parts of the two types of sex chromosomes. In principle, sex chromosomes are thought to evolve from an ancestral pair of autosomes, where, following the arrest of recombination, one of the chromosomes gradually becomes degraded and devoid of most genes (Charlesworth 1996; Rice 1996). Since degradation will in most cases not be complete, a few genes will remain on the smaller sex chromosome (*e.g.*, the mam-

Corresponding author: Hans Ellegren, Department of Evolutionary Biology, Evolutionary Biology Ctr., Uppsala University, Norbyvägen 18D, SE-752-36 Uppsala, Sweden.
E-mail: hans.ellegren@ebc.uu.se

malian Y chromosome) and will thus be present in a copy both on this and on the larger nondegraded sex chromosome (*e.g.*, the X chromosome). As shown for mammals, some of these genes are associated with male-specific or male-enhancing functions and may actually become silenced or deleted from the X chromosome (Graves 1995). Others, however, will be expressed from both sex chromosomes. Only a limited number of such genes have yet been identified (Lahn and Page 1997).

In birds, the female is the heterogametic sex and she has one Z and one W chromosome, whereas the male has two Z chromosomes. Physically, the W chromosome resembles the mammalian Y chromosome in several respects; it is small, gene-poor, and mainly heterochromatic (Stefos and Arrighi 1971). Studies of genes shared by the Z and W chromosomes would be important for reasons discussed above, and the avian sex chromosome system makes it possible to distinguish between some sex- and chromosome-specific factors that confound analyses in mammals (*cf.* Crow 1997; Ellegren and Fridolfsson 1997; Lessels 1997). Moreover, such studies are also motivated by the fact that the role of the W chromosome in avian sex determination is still unclear (Ellegren 2000). The critical issue is whether it is the W chromosome that is required for female development or if it is the number of Z chromosomes that regulates male development, *i.e.*, a dominance (as in mammals) or a balance (as in Drosophila and *Caenorhabditis elegans*) mode of genic sex determination. If it is the latter, and circumstantial evidence lends some support to this idea (Crew 1954; Halverson and Dvorak 1993; Raymond *et al.* 1999; Smith *et al.* 1999), the question is what selective constraints act on W-linked genes and why they have been retained on this chromosome.

Two avian genes have recently been shown to exist in a copy on both the Z and the W chromosome, the *CHD1Z*/*CHD1W* gene pair (Ellegren 1996; Griffiths *et al.* 1996; Griffiths and Korn 1997; Fridolfsson *et al.* 1998) and the ATP5A1Z/ATP5A1W pair (Dvorak *et al.* 1992; Fridolfsson *et al.* 1998; Carmichael *et al.* 2000). The avian *CHD1* genes belong to a family of genes composed of a chromatin organization modifier (*chromo*) domain, a SNF2-related *helicase*/ATPase domain, and a *DNA-binding* domain and the protein has been named CHD to denote these domains. Functional studies in model organisms have indicated that *CHD1* alters the chromatin structure and thereby facilitates gene expression (Stokes and Perry 1995; Stokes *et al.* 1996). It is not yet known if avian *CHD1Z* and *CHD1W* are functionally differentiated.

In this study, we present a detailed analysis of the molecular evolution of avian *CHD1Z* and *CHD1W* genes. Based on sequence data from three avian species, we show that the two genes are highly conserved both with respect to *CHD1* genes in other organisms and to each other. However, the two genes appear to evolve independently, without signs of genetic exchange through recombination. *CHD1Z* has a lower frequency of nonsynonymous (K_a) but a higher frequency of synonymous (K_s) substitutions compared with *CHD1W*. We attribute these differences to the respective characteristics of effective population size, recombination, and sex-specific mutation rates associated with the two types of sex chromosomes.

MATERIALS AND METHODS

PCR and cloning: mRNA was prepared from 25 μl of fresh whole blood from one male and one female of Tengmalm's owl (*Aegolius funerus*) and of cockatiel (*Lutino cockatiel*), with a Quick Prep Micro mRNA purification kit (Pharmacia Biotech, Piscataway, NJ). The Access reverse transcriptase PCR (RT-PCR) system (Promega, Madison, WI) was used with 1/500 of each mRNA preparation, together with the primer combinations described below, to amplify overlapping fragments of the *CHD1* genes. Obtaining *CHD1Z* was straightforward since amplification of male mRNA yields only this gene, even when using primer sequences conserved between *CHD1Z* and *CHD1W*. The following five pairs were used to amplify *CHD1Z*: 1090F (Ellegren 1996) and 2128R (Fridolfsson *et al.* 1998), 1628F (Fridolfsson *et al.* 1998) and 2469R (Ellegren 1996), 2421F (Ellegren 1996) and 3112R (Ellegren and Fridolfsson 1997), 2895F (5′-CGGCTAGTCACAAAAGGATC-3′) and 3681R (Ellegren and Fridolfsson 1997), and finally P3 (Griffiths and Tiwari 1995) and 4104R (Ellegren 1996).

Specific amplification of *CHD1W* in female birds is complicated by the fact that *CHD1Z* and *CHD1W* are very similar and are both expressed in females. We used a combined strategy of W-specific primers (underlined below), on the basis of sequence information from chicken *CHD1W*, and single-strand conformation polymorphism (SSCP) analysis to identify *CHD1W* products in amplifications of female DNA. The primers used were: 1275F (Ellegren 1996) and 1869R (5′-CATCCATTCATGAGTTCTTAT-3′), 1628F and 2469R, 2421F and 3112R, 2987F (Ellegren and Fridolfsson 1997) and 3829R (5′-GCCAACTCTTCTTCGTGAGAA-3′), and 3468F (Ellegren 1996) and 4105R. RT-PCR conditions were 48° for 45 min, then an initial denaturation step of 94° for 2 min followed by a 10-cycle touchdown profile consisting of 94° for 30 sec, 60–50° (lowering the temperature by 1° per cycle) for 30 sec, and 68° for 1 min. Then 30 cycles of the same profile were run at a constant annealing temperature of 50°, and a final extension step of 68° for 10 min was added after the last cycle.

Amplification products were separated by agarose gel electrophoresis (1.5% agarose, Sea Kem) and fragments of the expected size were excised and purified (Qiaex II gel extraction kit, QIAGEN, Hilden, Germany) and ligated into pGEM-T vector (pGEM-T easy vector systems, Promega). For *CHD1W*, 10 clones of each ligation were reamplified with the same primers and were analyzed with SSCP, together with clones known to contain the Z copy of the fragment (*i.e.*, amplified from males). Clones containing *CHD1W* could thereby be identified on the basis of the contrasting SSCP patterns of *CHD1Z* and *CHD1W* sequences. Clones were sequenced with vector primers using BigDye terminator cycle sequencing chemistry (Perkin Elmer, Norwalk, CT), followed by analysis on an ABI377 automated sequencing instrument (Perkin Elmer, Foster City, CA). The fact that we used overlap-

ping fragments allowed us to ensure that correct clones had always been identified.

Genes were named with a prefix denoting the Latin name of the species of origin (chicken, Gg; Tengmalm's owl, Af; cockatiel, Lc). For use in analyses we obtained from GenBank chicken *CHD1Z* (AF004397), mouse (Mm, L10410), human (Hs, AF006513), *Drosophila melanogaster* (Dm, X99021), *Saccharomyces cerevisiae* (Sc, L10718), and *Arabidopsis thaliana* (At, AC007209) *CHD1* gene sequences. Sequences obtained in this study have been deposited in GenBank under accession nos. AF181824–AF181828.

Sequence analysis: Contigs of the coding sequence of *CHD1Z* and *CHD1W* from each species were constructed using Sequencher 3.0 (Gene Codes, Ann Arbor, MI). Avian sequences were aligned with Sequence Navigator (Applied Biosystems, Foster City, CA) and MEGA (Kumar *et al.* 1993) was used for translation and analyses of amino acid (aa) distances and base composition. Phylogenetic analyses were done by maximum parsimony (MP) and maximum likelihood (ML) as implemented in PAUP* 4.0b2A (Swofford 1998). PHYLIP version 3.5c (Felsenstein 1991) was used for UPGMA clustering of sequences based on synonymous substitutions. The frequency of synonymous and nonsynonymous substitutions and their standard errors were calculated by combining the information from twofold and fourfold degenerate sites and using the Kimura two-parameter model to correct for multiple hits (Li 1993; Pamilo and Bianchi 1993). Patterns of variation in the K_a/K_s ratio across genes were calculated by dividing the gene into 18 nonoverlapping sections, each containing 51 codons. Spearman rank correlation was used to test if patterns were significantly repeatable. To test for positive selection in individual *CHD1W* lineages or among sites, the program codeml included in PAML (Yang 1999) was used. Analysis of CpG sites followed the method described by Smith and Hurst (1999), where sites on both strands were included in the analysis. In several analyses, we present means of K_a and K_s in comparisons of different *CHD1Z* and *CHD1W* sequences. However, since only three avian species were studied, it should be noted that the three possible comparisons (chicken *vs.* Tengmalm's owl, chicken *vs.* cockatiel, and Tengmalm's owl *vs.* cockatiel) do not represent independent observations.

RESULTS

Independent evolution of *CHD1W* and *CHD1Z*: Based on overlapping fragments amplified by RT-PCR of mRNA prepared from blood, we sequenced 2754 bp of the coding region of the *CHD1Z* and *CHD1W* genes from two divergent bird species, Tengmalm's owl and cockatiel. This continuous region covers most of the three functional domains of the CHD protein, *i.e.*, the chromo domain, the helicase domain, and the DNA-binding domain. The obtained sequences could be aligned with chicken *CHD1Z* (Griffiths and Korn 1997) and *CHD1W* (Ellegren 1996) without gaps.

As a starting point for further analysis, we first asked whether *CHD1Z* and *CHD1W* genes are evolving independently. Phylogenetic analysis with both MP and ML, using mouse and human *CHD1* as outgroups, clustered the three *CHD1Z* and the three *CHD1W* genes separately (Figure 1). The ML tree has a stronger bootstrap support (84/100) than the MP tree (66/100), which is not unexpected given that maximum-likelihood analysis is less sensitive to long-branch attraction (Huelsenbeck 1997; the outgroups are only distantly related to birds and the branch leading to Gg*CHD1Z* is the longest within the avian *CHD1* tree; moreover, the rate of evolution varies among the branches; see below). The best alternative MP tree (two steps longer) places the root on the Gg*CHD1Z* branch, and the best alternative ML tree ($\Delta \ln L = -5.34$) places the root on the Ag*CHD1Z*/Lc*CHD1Z* branch; both alternatives are indeed unlikely. These results indicate that the *CHD1Z* and *CHD1W* genes of the three avian lineages under study have evolved without signs of genetic exchange (*e.g.*, through recombination) between the Z and the W chromosomes. Importantly, the respective molecular evolution of *CHD1Z* and *CHD1W* should therefore reflect the intrinsic and different evolutionary forces operating on the two sex chromosomes.

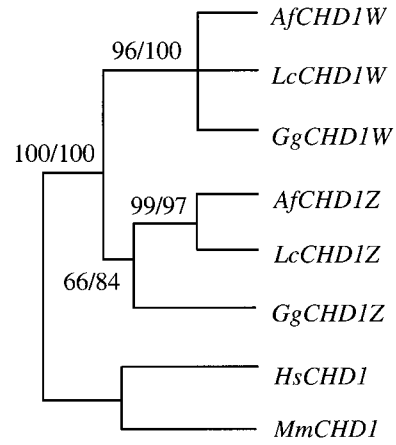

Figure 1.—Phylogenetic tree showing the relationship between six avian *CHD1* gene sequences, using human and mouse *CHD1* as outgroups. Numbers indicate bootstrap support for branches using maximum-parsimony and maximum-likelihood analysis (1000 replicates in each case).

High degree of amino acid conservation in *CHD1* genes: The frequency of aa replacements between different gene copies was derived from alignments of avian *CHD1Z* and *CHD1W* sequences and of *CHD1* from mouse, Drosophila, yeast, and Arabidopsis (Table 1). While avian sequences could be aligned to the mouse sequence without gaps, gaps had to be introduced relative to the more distantly related species. Overall levels of conservation were very high, with, for instance, about five replacements per 100 sites between avian and mouse genes.

Comparisons of avian *CHD1Z* and *CHD1W* aa sequences revealed that the two proteins are very similar to each other (mean = 3.2 ± 0.6 aa replacements per 100 sites, range 2.5–3.9; Figure 2), suggesting shared functional properties. Within the respective class of genes, *CHD1Z* proteins (mean = 1.2 ± 0.1, range 1.2–1.3) are more slowly evolving than *CHD1W* proteins (3.4 ± 0.5, range 2.8–3.7).

TABLE 1

Number of amino acid replacements per site between CHD1 genes

	MmCHD1	DmCHD1	AtCHD1	ScCHD1
All domains[a]				
GgCHD1W	0.06	0.42	0.56	0.60
GgCHD1Z	0.04	0.42	0.56	0.60
Chromo domain[b]				
GgCHD1W	0.09	0.50	0.58	0.66
GgCHD1Z	0.04	0.51	0.58	0.67
Helicase domain[c]				
GgCHD1W	0.03	0.27	0.40	0.47
GgCHD1Z	0.02	0.26	0.39	0.46
Intervening region between the H and D domains[d]				
GgCHD1W	0.09	0.54	0.44	0.70
GgCHD1Z	0.05	0.53	0.45	0.70
DNA-binding domain[e]				
GgCHD1W	0.09	0.56	0.77	0.71
GgCHD1Z	0.06	0.58	0.77	0.71

[a] 981 aa.
[b] 162 aa.
[c] 461 aa.
[d] 140 aa.
[e] 218 aa.

K_a and K_a/K_s ratios of avian *CHD1Z* and *CHD1W* genes: In accordance with the aa data, K_a was lower for *CHD1Z* (mean = 0.81 ± 0.08 nonsynonymous nucleotide substitutions per 100 sites) than for *CHD1W* (1.55 ± 0.30; Table 2), which in turn was only marginally less than that for *CHD1Z vs. CHD1W* (1.85 ± 0.31). However, since the overall mutation rate may differ between the Z and W chromosomes (Ellegren and Fridolfsson 1997), a more appropriate measure of the evolutionary forces operating on *CHD1Z* and *CHD1W* should be their

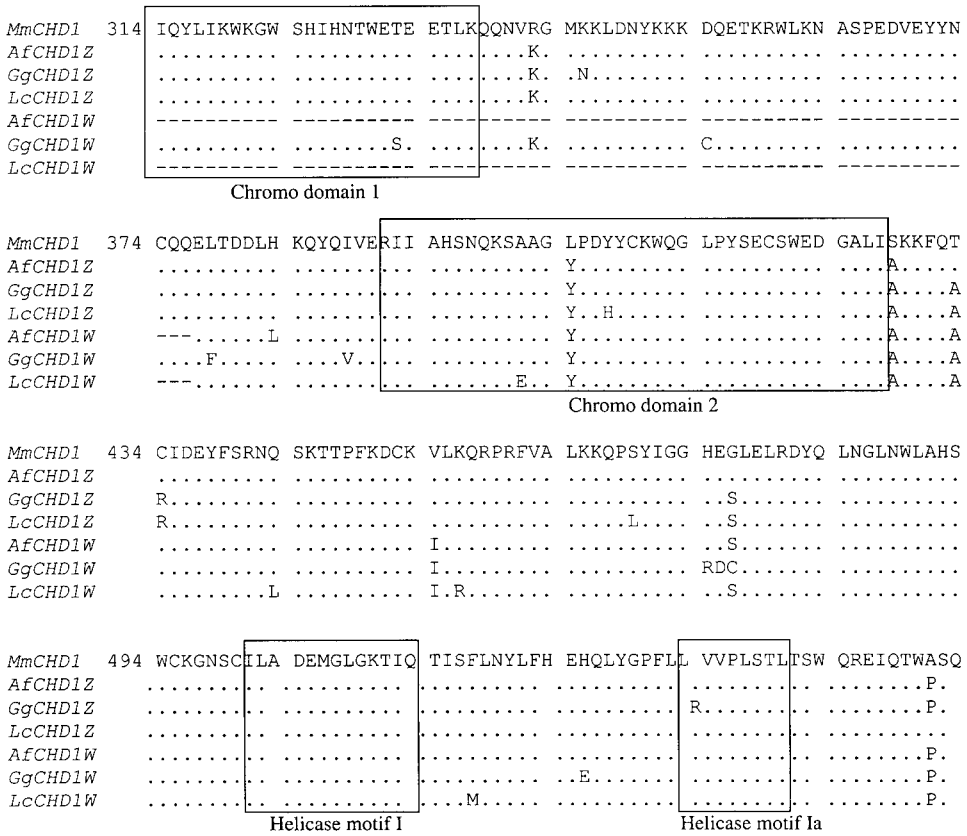

Figure 2.—Amino acid alignment of avian *CHD1* genes, with mouse *CHD1* as master sequence. Identical positions are denoted by dots and positions for which data are lacking are denoted by dashes. There are no gaps. Positions are numbered according to the complete aa sequence of mouse. Known functional domains or motifs are boxed.

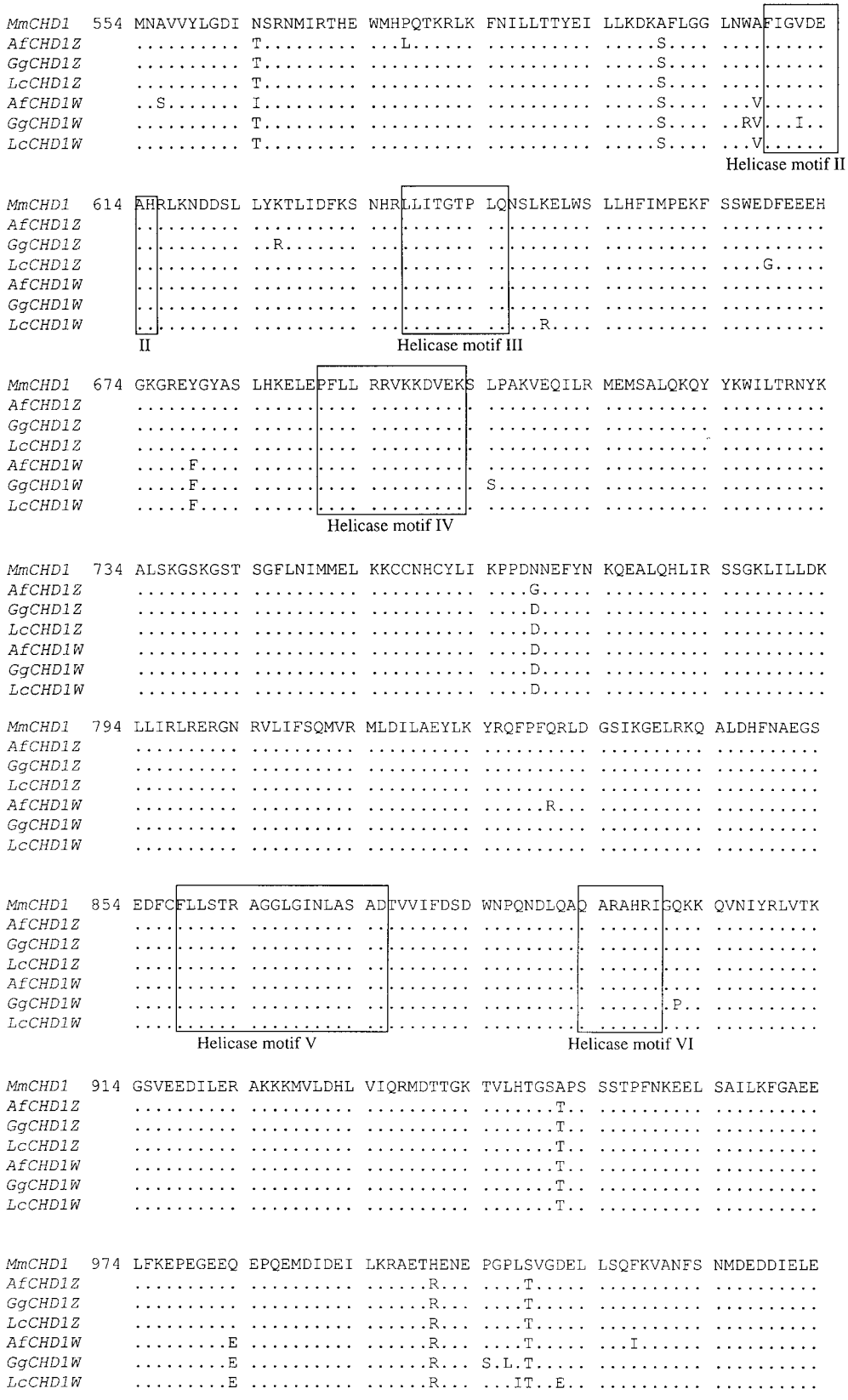

Figure 2.—*Continued.*

K_a/K_s ratios. Mean K_a/K_s for *CHD1Z* (0.037 ± 0.01) was considerably lower than for *CHD1W* (0.11 ± 0.01).

Selective forces upon replacement substitutions can obviously be different for different parts of a gene, leading to variation in the pattern of K_a/K_s across genes (Alvarez-Valin *et al.* 1998). Repeatability of K_a/K_s patterns in comparisons of independent pairs of gene lineages is an indication of nonrandom substitution rates

```
MmCHD1 1074 PERNSKNWEE IIPEEQRRRL EEEERQKELE EIYMLPRMRN CAKQISFNGS EGRRSRSRRY
AfCHD1Z     .....R.... ....V....I .......... .......... .......... ..........
GgCHD1Z     .....R.... ....S....I .......... .......... .......... ..........
LcCHD1Z     ....CR.... ....V....I .......... .......... .......... ..........
AfCHD1W     ..Q.F.K... ....V....I .......... .......... .......... ..K.....K.
GgCHD1W     ..Q.LR.... ....V.W..I .......... .......... .........N ...C......
LcCHD1W     ..Q.LR.... ....V....I .......... K......... .......... .R.H......

MmCHD1 1094 SGSDSDSISE RKRPKKRGRP RTIPRENIKG FSDAEIRRFI KSYKKFGGPL ERLDAIARDA
AfCHD1Z     .......... .......... .......... .......... .......... .....V....
GgCHD1Z     ........T. .......... .......... .......... .......... .....V....
LcCHD1Z     .......... .......... .......... .......... .......... .....V....
AfCHD1W     .......... .......... .......... .......... .......... ..........
GgCHD1W     .......... .......... .......... .......... .......... ..........
LcCHD1W     .......... ...R...... .......... .......... .......... ..........
                       DNA-binding motif

MmCHD1 1154 ELVDKSETDL RRLGELVHNG CVKALKDSSS GTERAGGRLG KVKGPTFRIS GVQVNAKLVI
AfCHD1Z     .......... .......... .I........ .Q........ .......... ..........
GgCHD1Z     .......... .......... .I.....N.. .Q........ .......... ..........
LcCHD1Z     .......I.. .......... .I........ .Q........ .......... ..........
AfCHD1W     .......... .......... .I.....N.F .Q........ .......... ..........
GgCHD1W     .......... .......... .I...N.NDF .QG.T...F. .........A ..........
LcCHD1W     .......... .......... .I...N.N.F VQ...A.... .......... ..........

MmCHD1 1214 AHEDELIPLH KSIPSDPEER KQYTIPCHTK AAHFDIDWGK EDDSNLLIGI YEYGYGSWEM
AfCHD1Z     S..E..A... .......... .R.V...... .......... .......... ..........
GgCHD1Z     S..E..A... .......... .R.V...... .......... .......V.. ..........
LcCHD1Z     S..E..A... .......... .R.V...... .......... .......... ..........
AfCHD1W     S..E..A... ........D. .R.V...... .......... .......... ..........
GgCHD1W     S..E..A... .......... .R.V..Y... .......... .......... ..........
LcCHD1W     S..E..A... .......... .R.V...... ........C. .......... ..........

MmCHD1 1274 IKMDPDLSLT HKILPDDPDK K
AfCHD1Z     .......... Q......... .
GgCHD1Z     .......... Q......... .
LcCHD1Z     .......... Q......... .
AfCHD1W     .......... Q......... .
GgCHD1W     .......... Q......... .
LcCHD1W     .......... Q......... .
```

Figure 2.—*Continued.*

(Smith and Hurst 1998) and is suggestive of different gene copies sharing functional properties. The patterns of K_a/K_s variation across avian *CHD1* genes were roughly similar in the three possible comparisons of *CHD1Z* and *CHD1W* genes (Figure 3). For instance, K_a/K_s was particularly low in the 3′ end of the helicase domain of both *CHD1Z* and *CHD1W*. Repeatability was statistically significant for Tengmalm's owl *vs.* cockatiel ($R_s = 0.60$, $P = 0.013$), but not so in the two other comparisons.

Although the fact that K_a/K_s never exceeded 0.35 suggests an absence of positive selection, a higher K_a/K_s ratio in *CHD1W* than in *CHD1Z* genes might be indicative of adaptive changes in individual lineages or in parts of the *CHD1W* gene. To investigate this further, we used a likelihood-ratio test implemented in PAML (Yang and Nielsen 1998; Yang 1999). However, this failed to reject a null hypothesis of equal K_a/K_s ratios in individual lineages, tested in all possible topologies of *CHD1W* trees [$2\Delta l = 0.92$, d.f. = 2, not significant (NS)]. Similarly, a likelihood-ratio test failed to reject a

TABLE 2

Frequency of nonsynonymous substitutions between avian CHD1 genes

	AfCHD1Z	*GgCHD1Z*	*LcCHD1Z*	*AfCHD1W*	*GgCHD1W*	*LcCHD1W*
AfCHD1Z		0.81	0.73	1.45	2.07	1.67
GgCHD1Z	0.15		0.88	1.61	2.30	1.82
LcCHD1Z	0.14	0.15		1.58	2.32	1.84
AfCHD1W	0.19	0.20	0.20		1.68	1.21
GgCHD1W	0.23	0.25	0.25	0.21		1.77
LcCHD1W	0.21	0.22	0.22	0.17	0.22	

Estimated number and standard error of nucleotide substitutions per 100 nonsynonymous (K_a) sites between CHD1W and CHD1Z genes. K_a values are above the diagonal; standard errors are below the diagonal.

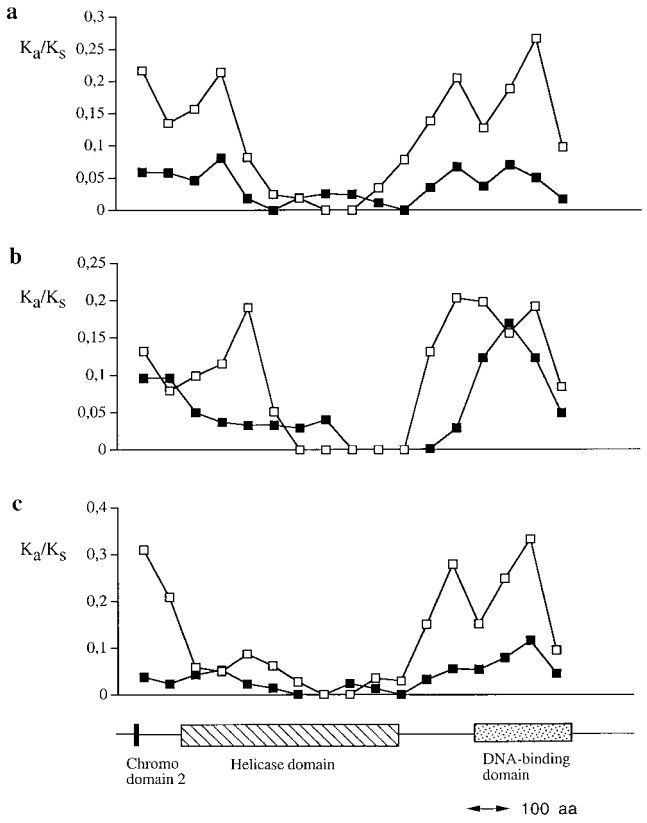

Figure 3.—Variation in K_a/K_s across *CHD1Z* (solid squares) and *CHD1W* (open squares) genes. Each data point represents 102 codons, with an overlapping window of 51 codons. (a) Tengmalm's owl *vs.* chicken; (b) Tengmalm's owl *vs.* cockatiel; and (c) chicken *vs.* cockatiel.

null hypothesis of equal K_a/K_s ratios among sites ($2\Delta l = -27$, d.f. = 2, NS).

A higher K_s in *CHD1Z* than in *CHD1W*: The frequency of synonymous substitutions (K_s) was higher for *CHD1Z* (mean = 22.70 ± 6.62) than for *CHD1W* (mean = 13.48 ± 2.06; Table 3), which contrasts to the situation for K_a. This indicates an underlying sex difference in the mutation rate, assuming that synonymous substitutions in *CHD1Z* and *CHD1W* are selectively neutral or are at least under the same constraints. Pairwise comparisons of K_s revealed estimates of the male-to-female mutation rate ratio (α_m) of 2.1 ± 0.3 (Tengmalm's owl *vs.* chicken), 2.1 ± 0.3 (cockatiel *vs.* chicken), and 1.5 ± 0.2 (cockatiel *vs.* Tengmalm's owl). A mean value of $\alpha_m \approx 1.7$ was estimated from the branch lengths of a dendrogram based on K_s distances (*cf.* Shimmin *et al.* 1993). Since the phylogenetic relationship of the *CHD1W* genes was unresolved, this mean value is only an approximation.

Low influence of CpG sites on K_s: The GC content of *CHD1Z* (mean = 40.5 ± 0.1) and *CHD1W* genes (39.1 ± 0.1) was lower than an average of 53.2% estimated from 399 chicken genes (Olivier and Marin 1996), but did not differ between the two types of genes ($\chi^2 = 0.04$, NS). The GC3 content was even lower (*CHD1Z*, 36.0 ± 0.5; *CHD1W*, 33.0 ± 0.4; chicken average, 69.4%; Bernardi *et al.* 1988), but again did not differ between *CHD1Z* and *CHD1W* ($\chi^2 = 0.20$, NS). The observed number of CpG sites was about five times lower than expected based on base composition in both *CHD1Z* (ratio of observed/expected = 0.17) and *CHD1W* (0.20). This underrepresentation is of the same magnitude as the average for genes in the human genome (Schorderet and Gartler 1992).

In separate analyses of *CHD1Z* and *CHD1W*, we counted the number of synonymous and nonsynonymous sites where all three sequences had a CpG dinucleotide. This number was compared to the number of sites where at least one sequence had a TpG dinucleotide while the other/s had a CpG dinucleotide, *i.e.*, possible cases of C–T transitions at methylated CpG sites. Since both the total number of CpG sites (*CHD1Z*, 15; *CHD1W*, 15) and the number of sites with possible C–T transitions (*CHD1Z*, 4; *CHD1W*, 5) were low, and did not differ between genes, we conclude that methylated CpG sites seem not to affect the molecular evolution of *CHD1Z* and *CHD1W* evolution in a contrasting way.

DISCUSSION

Very few aa changes distinguish avian *CHD1Z* and *CHD1W* proteins (eight positions represent fixed differ-

TABLE 3

Frequency of synonymous substitutions between avian CHD1 genes

	AfCHD1Z	GgCHD1Z	LcCHD1Z	AfCHD1W	GgCHD1W	LcCHD1W
AfCHD1Z		26.45	15.06	21.36	25.75	22.82
GgCHD1Z	1.70		26.59	29.30	31.90	29.76
LcCHD1Z	1.20	1.74		23.53	27.31	25.46
AfCHD1W	1.50	1.87	1.57		15.19	11.19
GgCHD1W	1.65	1.93	1.75	1.23		14.07
LcCHD1W	1.54	1.88	1.66	1.07	1.14	

Estimated number and standard error of nucleotide substitutions per 100 synonymous (K_s) sites between CHD1W and CHD1Z genes. K_s values are above the diagonal; standard errors are below the diagonal.

ences, six of which are conservative changes). Similarly, comparisons of eukaryotic *CHD1* genes, including avian *CHD1Z* and *CHD1W*, reveal extensive conservation, particularly in the functional domains. For instance, only one fixed amino acid difference distinguishes birds from mammals over a region of 180 aa residues in the helicase domain (Figure 2). In fact, the helicase domain is highly conserved even between different members of the CHD gene family (Woodage *et al.* 1997), indicating strong functional constraints. The DNA-binding activity of the *CHD1* protein has been located to a domain of 229 aa residues and within this region a sequence of 11 aa is essential for DNA binding by A · T minor-groove interactions (Stokes and Perry 1995). This sequence motif is identical between avian and mouse *CHD1* genes. Overall, this suggests (i) that *CHD1Z* and *CHD1W* share similar functional properties and (ii) that this function should be more or less the same as in other organisms.

Comparative analyses of nonsynonymous substitution rates are preferably made using the K_a/K_s ratio to account for local variation in the mutation rate. In our study, we found K_a/K_s to be higher for *CHD1W* (0.11) than for *CHD1Z* (0.04), which in turn was higher than for *CHD1* in mammals (0.025). Since K_a/K_s for *CHD1W* is $\ll 1$, which is the strict requirement for demonstration of positive selection, we found no overall suggestion that *CHD1W* would be rapidly diverging in an adaptive way. Likelihood-ratio tests similarly failed to detect signs of positive selection in terms of variation in K_a/K_s among *CHD1W* lineages or among *CHD1W* sites. Moreover, the patterns of K_a/K_s variation across the gene were similar between *CHD1Z* and *CHD1W* genes. This, together with the high degree of aa conservation seen between *CHD1Z* and *CHD1W*, strongly argues against a female-specific role of *CHD1W*. In fact, it might be argued that *CHD1Z* and *CHD1W* act in concert and in a sense should be seen as allelic variants of the same functional protein. It should be noted that positive selection has been recognized in male-specific and Y-linked sequences in mammals. For example, the mammalian SRY gene shows a K_a/K_s ratio of 1.3 (Tucker and Lundrigan 1993; Whitfield *et al.* 1993).

We argue that the difference in K_a/K_s between *CHD1Z* and *CHD1W* is associated with differences in effective population size and recombination characteristics of the two types of sex chromosomes. First, selection is more effective in removing slightly deleterious mutations in a population of larger size (Nei 1970; Li 1997). Other factors being equal, this should imply that such mutations are more easily removed from the Z chromosome since its effective population size is three times that of the W chromosome. Second, since most parts of the W chromosome do not recombine and are thus clonally transmitted, slightly deleterious mutations should be expected to accumulate faster than on the Z chromosome (Charlesworth 1996; Rice 1996). The expectation is in both cases a higher K_a/K_s ratio on the W chromosome than on the Z chromosome, as we observe. The same should be true for genes on the Y (analogous to W) and X (analogous to Z) chromosomes of mammals and is indeed supported by available data. The K_a/K_s ratio is higher for Ube1y (0.19) than for Ube1x (0.0; Chang and Li 1995), for ZFY (0.42) than for ZFX (0.13; Pamilo and Bianchi 1993), and for SMCY (0.17) than for SMCX (0.02; Agulnik *et al.* 1997).

In contrast to the situation for K_a, K_s was higher for *CHD1Z* than for *CHD1W*. From a similar observation based on partial sequence data, we recently interpreted this as evidence for a male-biased mutation rate, given that W is exclusively transmitted through the female germline (Ellegren and Fridolfsson 1997). Applying the formula of Miyata *et al.* (1987), present data suggest a male bias in the mutation rate of $\alpha_m \approx 1.7$ in the lineages studied, which is lower than our previous estimate of $\alpha_m \approx 3.9$ derived from the coding regions of *CHD1* genes of two passerine bird species (Ellegren and Fridolfsson 1997). It is not clear if this suggests variation in α_m between avian lineages, since the validity of statistical analyses is uncertain due to the difficulty in estimating confidence intervals of α_m. Importantly, all presently available bird data indicate more mutations among males than females.

Does the excess of male mutations tie in quantitatively with the difference in the number of germline cell divisions between males and females? This question is difficult to address due to the lack of detailed cytological data, although it seems quite clear that spermatogenesis involves more cell generations than oogenesis in birds (Jones and Lin 1993). Moreover, if there is an intrinsic reduction in the Z chromosome mutation rate, as has been suggested for the mammalian X chromosome (McVean and Hurst 1997), comparisons of the rate of neutral evolution on Z and W chromosomes would tend to underestimate α_m. On the other hand, α_m could overestimate the difference in the number of cell divisions in male and female germlines if the per cell generation mutation rate differs between sexes. One such potential factor is the degree of germline methylation, which affects the mutability of CpG sites (Li 1997). For example, methylation has been invoked to explain the male-biased mutation rate at hemophilia A CpG sites, which are more strongly methylated in male than in female germline (Oldenburg *et al.* 1993; Sommer and Ketterling 1996). According to the present data, however, a potential role of methylated CpG sites in explaining the male-biased mutation rate of avian *CHD1* genes could be excluded.

In summary, the genomic location of the *CHD1Z* and *CHD1W* genes on the avian sex chromosomes is likely to have affected the molecular evolution of these two genes in distinct ways. While the two proteins are highly conserved and do not seem functionally differentiated, they differ with respect to frequency of synonymous and nonsynonymous nucleotide substitutions. Since the

respective factors contributing to these differences (effective population size, recombination, and sex-specific mutation rates) should be valid for sex chromosomes in general, we anticipate the observed patterns of molecular evolution to be general characteristics of sex-linked genes.

We thank Tim Hipkiss and Jan Högberg for providing avian blood samples. We also thank Nick Smith and Bengt-Olle Bengtsson for valuable comments on the manuscript. Financial support was obtained from the Swedish Natural Sciences Research Council and Lars Hiertas Minne.

LITERATURE CITED

Agulnik, A. I., C. E. Bishop, J. L. Lerner, S. I. Agulnik and V. V. Solovyev, 1997 Analysis of mutation rates in the SMCY/SMCX genes shows that mammalian evolution is male driven. Mamm. Genome **8:** 134–138.

Alvarez-Valin, F., K. Jabbari and G. Bernardi, 1998 Synonymous and nonsynonymous substitutions in mammalian genes: intragenic correlations. J. Mol. Evol. **46:** 37–44.

Bernardi, G., D. Mouchiroud, C. Gautier and G. Bernardi, 1988 Compositional patterns in vertebrate genomes: conservation and change in evolution. J. Mol. Evol. **28:** 7–18.

Carmichael, A. N., A. K. Fridolfsson, J. Halverson and H. Ellegren, 2000 Male-biased mutation rates revealed from Z- and W-chromosome linked ATP synthase α-subunit (ATP5A1) sequences in birds. J. Mol. Evol. **50:** 443–446.

Chang, B. H., and W.-H. Li, 1995 Estimating the intensity of male-driven evolution in rodents by using X-linked and Y-linked Ube1 genes and pseudogenes. J. Mol. Evol. **40:** 70–77.

Charlesworth, B., 1996 The evolution of chromosomal sex determination and dosage compensation. Curr. Biol. **6:** 149–162.

Charlesworth, B., J. A. Coyne and N. H. Barton, 1987 The relative rates of evolution of sex chromosomes and autosomes. Am. Nat. **130:** 113–146.

Crew, F. A. E., 1954 *Sex Determination*. John Wiley & Sons, New York.

Crow, J. F., 1997 Molecular evolution—who is in the driver's seat? Nat. Genet. **17:** 129–130.

Driscoll, D. J., and B. R. Migeon, 1990 Sex difference in methylation of single-copy genes in human meiotic germ cells: implications for X chromsome inactivation, parental imprinting and origin of CpG mutations. Somat. Cell Mol. Genet. **16:** 267–282.

Dvorak, J., J. L. Halverson, P. Gulick, K. A. Rauen, U. K. Abbott *et al.*, 1992 cDNA cloning of a Z- and W-linked gene in Gallinaceous birds. J. Hered. **83:** 22–25.

Ellegren, H., 1996 First gene on the avian W chromosome provides a tag for universal sexing of non-ratite birds. Proc. R. Soc. Lond. Ser. B **263:** 1635–1641.

Ellegren, H., 2000 Evolution of the avian sex chromosomes and their role in sex determination. Trends Ecol. Evol. **15:** 188–192.

Ellegren, H., and A.-K. Fridolfsson, 1997 Male-driven evolution of DNA sequences in birds. Nat. Genet. **17:** 182–184.

Felsenstein, J., 1991 PHYLIP: phylogeny inference package. Version 3.4. University of Washington, Seattle.

Fridolfsson, A.-K., H. Cheng, N. G. Copeland, N. A. Jenkins, H. C. Liu *et al.*, 1998 Evolution of the avian sex chromosomes from an ancestral pair of autosomes. Proc. Natl. Acad. Sci. USA **95:** 8147–8152.

Graves, J. A. M., 1995 The evolution of mammalian sex chromosomes and the origin of sex determining genes. Philos. Trans. R. Soc. Lond. Ser. B **350:** 305–312.

Griffiths, R., and R. M. Korn, 1997 A *CHD1* gene is Z chromosome linked in the chicken *Gallus domesticus*. Gene **197:** 225–229.

Griffiths, R., and B. Tiwari, 1995 Sex of the last wild Spix's macaw. Nature **375:** 454.

Griffiths, R., S. Daan and C. Dijkstra, 1996 Sex identification in birds using two CHD genes. Proc. R. Soc. Lond. Ser. B **263:** 1251–1256.

Halverson, J. J., and J. Dvorak, 1993 Genetic control of sex determination in birds and the potential for its manipulation. Poult. Sci. **72:** 890–896.

Huelsenbeck, J. P., 1997 Is the Felsenstein zone a fly trap? Syst. Biol. **46:** 69–74.

Hurst, L. D., and H. Ellegren, 1998 Sex bias in the mutation rate. Trends Genet. **14:** 446–451.

Jones, R. C., and M. Lin, 1993 Spermatogenesis in birds, pp. 233–264 in *Oxford Review of Reproductive Biology*, edited by S. R. Milligan. Oxford University Press, Oxford.

Kumar, S., K. Tamura and M. Nei, 1993 *MEGA: Molecular Evolutionary Genetic Analysis*, version 1.0. Pennsylvania State University, University Park, PA.

Lahn, B. T., and D. C. Page, 1997 Functional coherence of the human Y chromosome. Science **278:** 675–680.

Lessels, K., 1997 More mutations in males. Nature **390:** 236–237.

Li, W.-H., 1993 Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. **36:** 96–99.

Li, W.-H., 1997 *Molecular Evolution*. Sinauer Associates, Sunderland, MA.

McVean, G. T., and L. D. Hurst, 1997 Evidence for a selectively favourable reduction in the mutation rate of the X chromosome. Nature **386:** 388–395.

Miyata, T., H. Hayashida, K. Kuma, K. Mitsuyasu and T. Yasunaga, 1987 Male-driven molecular evolution: a model and nucleotide sequence analysis. Cold Spring Harbor Symp. Quant. Biol. **52:** 863–867.

Nei, M., 1970 Accumulation of nonfunctional genes on a sheltered chromosome. Am. Nat. **104:** 311–322.

Oldenburg, J., R. Scwaab, T. Grimm, K. Zerres, K. Hakenberg *et al.*, 1993 Direct and indirect estimation of the sex ratio of mutation frequencies in hemophilia A. Am. J. Hum. Genet. **53:** 1229–1238.

Olivier, J. L., and A. Marin, 1996 A relationship between GC content and coding sequence length. J. Mol. Evol. **43:** 216–223.

Pamilo, P., and N. O. Bianchi, 1993 Evolution of the Zfx and Zfy genes: rates and interdependence between the genes. Mol. Biol. Evol. **10:** 271–281.

Raymond, C. S., J. R. Kettlewell, B. Hirsch, V. J. Bardwell and D. Zarkower, 1999 Expression of Dmrt1 in the genital ridge of mouse and chicken embryos suggests a role in vertebrate sexual development. Dev. Biol. **215:** 208–220.

Rice, W. R., 1996 Evolution of the Y sex chromosome in animals. BioScience **46:** 331–343.

Schorderet, D. F., and S. M. Gartler, 1992 Analysis of CpG suppression in methylated and nonmethylated species. Proc. Natl. Acad. Sci. USA **89:** 957–961.

Shimmin, L. C., B. H. Chang and W.-H. Li, 1993 Male-driven evolution of DNA sequences. Nature **362:** 745–747.

Smith, N. G. C., and L. D. Hurst, 1998 Molecular evolution of an imprinted gene: repeatability of patterns of evolution within the mammalian insulin-like growth factor type II receptor. Genetics **150:** 823–833.

Smith, N. G. C., and L. D. Hurst, 1999 The causes of synonymous rate variation in rodent genome: can substitution rates be used to estimate the sex bias in mutation rate? Genetics **152:** 661–673.

Smith, C. A., P. J. McClive, P. S. Western, K. J. Reed and A. H. Sinclair, 1999 Conservation of a sex-determining gene. Nature **402:** 601–602.

Sommer, S. S., and R. P. Ketterling, 1996 The factor IX gene as a model for analysis of human germline mutations: an update. Hum. Mol. Genet. **5:** 1505–1514.

Stefos, K., and F. E. Arrighi, 1971 Heterochromatic nature of W chromosome in birds. Exp. Cell Res. **68:** 228–231.

Stokes, D. G., and R. P. Perry, 1995 DNA-binding and chromatin localization properties of *CHD1*. Mol. Cell. Biol. **5:** 2745–2753.

Stokes, D. G., K. D. Tartof and R. P. Perry, 1996 *CHD1* is concentrated in interbands and puffed regions of *Drosophila* polytene chromosomes. Proc. Natl. Acad. Sci. USA **93:** 7137–7142.

Swofford, D. L., 1998 *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, version 4.0b2a. Sinauer Associates, Sunderland, MA.

Tucker, P. K., and B. L. Lundrigan, 1993 Rapid evolution of the sex determining locus in Old World mice and rats. Nature **364:** 715–717.

Whitfield, S. L., R. Lovell-Badge and P. N. Goodfellow, 1993 Rapid sequence evolution of the mammalian sex-determining gene SRY. Nature **364:** 713–715.

Woodage, T., M. A. Basra, A. D. Baxevanis, P. Hieter and F. S. Collins, 1997 Characterization of the CHD family of proteins. Proc. Natl. Acad. Sci. USA **94:** 11472–11477.

Yang, Z., 1999 *Phylogenetic Analysis by Maximum Likelihood (PAML)*, version 2.0. University College, London.

Yang, Z., and R. Nielsen, 1998 Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J. Mol. Evol. **46:** 409–418.

Communicating editor: P. D. Keightley