Adaptive Molecular Evolution of HINTW, a Female-Specific Gene in Birds

Helene Ceplitis and Hans Ellegren
Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

It is well established that many genes on the male-specific Y chromosome of organisms such as mammals are involved in male reproduction and may evolve rapidly because of positive selection on male reproductive traits. In contrast, very little is known about the function and evolution of W-linked genes restricted to the female genome of organisms with female heterogamety. For birds (males ZZ, females ZW), only one W-linked gene (HINTW) is sufficiently different from its Z-linked homolog to indicate a female-specific function. Here, we report that HINTW shows evidence of adaptive molecular evolution, implying strong positive selection for new functional properties in female birds. Moreover, because HINTW is expressed in the gonads of female birds just before sexual differentiation and is thus a candidate for sex determination, it suggests adaptive evolution related to female development. This provides the first example of Darwinian evolution of a gene restricted to the female genome of any organism. Given that HINTW exists in multiple copies on W, similar to some testis-specific genes amplified on mammalian Y, avian HINTW may thus potentially represent a female parallel to the organization and evolution of Y chromosome genes involved in male reproduction and development.

Introduction

While no genes are specific to the female genome of organisms with male heterogamety, such genes occur in organisms with reversed sex chromosome organization such as birds, butterflies, and some reptiles (males ZZ, females ZW). For the latter group of organisms, the W chromosome constitutes an interesting female parallel to the male-specific Y chromosome of, for instance, mammals. As nonrecombining and sex-limited chromosomes, with the exception of pseudoautosomal regions, W and Y may have been subject to similar evolutionary forces during the process of sex chromosome evolution. However, there are also some important differences when it comes to evolutionary aspects. Notably, many genes on the Y chromosome are involved in male reproduction and may therefore evolve rapidly because of sexual selection on male reproductive traits (Wyckoff, Wang, and Wu 2000). It is not obvious how such predictions would apply to genes on the female-specific W chromosome, except perhaps in avian species in which sexual selection operates on both males and females (Jones and Hunter 1999). In general, little is known about the evolutionary forces behind W-linked genes and there are, to our knowledge, no examples of positive selection operating on genes limited to the female genome.

For birds, it is unclear whether genes on the W chromosome are necessary for female development (a dominant role, cf. the mammalian Y chromosome) or whether the number of Z chromosomes regulates sex differentiation (a dosage effect, cf. Drosophila) (Clinton 1998; Clinton and Haines 1999; Ellegren 2001). Only a few genes have been identified on the avian W chromosome, and most of them have very similar homologs (gametologs) on Z, providing no indication of evolution of female-specific functions (Griffiths and Korn 1997; Carmichael et al. 2000; Ellegren 2000; Fridolfsson and Ellegren 2000; Itho et al. 2001). However, the amino acid sequence of one W-linked gene, HINTW, is distinctly different from that of its Z-linked gametolog, HINTZ (Hori et al. 2000; O’Neill et al. 2000; Pace and Brenner 2003; see Notes added in proofs concerning the nomenclature of these genes). The gene encoding a histidine triad nucleotide binding protein (HINT) is likely to have been present as a single-copy gene also on the ancestral protosex chromosomes of birds. After the differentiation of the avian sex chromosomes into Z and W, initiated some 100 to 170 MYA (dating based on the degree of divergence between gametologous gene pairs on Z and W [Lawson-Handley et al. 2004]), HINT diverged into two independently evolving copies, HINTW and HINTZ (Hori et al. 2000) (see fig. 1). Avian HINTZ is highly conserved; for instance, the chicken HINTZ amino acid sequence shows an overall 87% identity to human HINT and even higher similarity in functional domains (Hori et al. 2000). In contrast, and despite more recent common ancestry, chicken HINTZ shows only 65% amino acid identity to its W-linked gametolog HINTW (Hori et al. 2000). Moreover, the latter lacks a histidine triad (HIT; HisφHisφHisφ where φ is a hydrophobic amino acid) found in HINT and other members of the HIT superfamily of proteins (Séraphin 1992; Brenner et al. 1999; Brenner 2002) but contains a unique Leu-rich and Arg-rich region of unknown function (Hori et al. 2000; O’Neill et al. 2000).

HINTW/HINTZ represents the only known gene pair in an organism with female heterogamety where the W-linked copy is clearly differentiated from its Z-linked counterpart. This may be because of relaxation of selective constraints or to the acquisition of new and female-specific function(s) through adaptive evolution. To address this question, we target in this study the molecular evolution of avian HINTW. We demonstrate that HINTW shows evidence of positive selection, providing the first example of Darwinian evolution of a gene restricted to the female genome.

Materials and Methods

Sequence Data

Sequences were from GenBank with the following accession numbers: AB026677 (chicken HINTW),
AB026675 (chicken HINTZ), AB033881 (quail HINTW),
AB033882 (quail HINTZ), U27143 (human HINT), and
CA844757 (fugu HINT).

Phylogenetic Analysis

Maximum parsimony was used to construct phylo-
genetic trees in PAUP* version 4.0b10 (Swofford 2000).
A tree of cDNA sequences of HINTZ and HINTW from
both chicken and quail and of HINT from human was
constructed using a 399-bp alignment with HINT from
fugu as an outgroup (fig. 1). Levels of support were
obtained from 1,000 bootstrap replicates. For subsequent
molecular evolutionary analyses, an unrooted tree (without
fugu) was also constructed (fig. 2). Alignments of cDNA
sequences were done using ClustalW (available from
http://pbil.ibcp.fr [Thompson, Higgins, and Gibson 1994])
and when necessary edited by eye.

Analysis of Adaptive Evolution

The ratio of the number of nonsynonymous substi-
tutions per nonsynonymous site to the number of
synonymous substitutions per synonymous site ($K_a/K_s$)
was computed using the CODEML program in the PAML
software package (Yang 1997). To test if lineages in
a phylogeny (the unrooted tree described above) have
different $K_a/K_s$ ratios (Yang 1998), we compared a model
that assumes the same $K_a/K_s$ ratio for all branches (one-
ratio model) with a model that assumes an independent
$K_a/K_s$ ratio for each branch (free-ratio model). To test
whether a free-ratio model fitted the data significantly
better than a one-ratio model, we used the likelihood ratio
test statistic $\text{LRT} = -2[\ln L_0 - \ln L_1]$, where $L_0$ is the
likelihood under the null hypothesis (i.e., the one-ratio
model) and $L_1$ is the likelihood under the alternative
hypothesis (the free-ratio model). When the models
compared are nested, like the one-ratio and free-ratio
models, the LRT-value will be $\chi^2$ distributed (Goldman
and Yang 1994). The number of degrees of freedom will
depend on how many parameters are being estimated
under the different models.

$K_a/K_s$ ratios were also estimated from pairwise
comparisons between sequences (table 1). We used PAML
to calculate the likelihood values under a model in which
$K_a/K_s$ was fixed at 1 ($L_0$) and under a model in which the
$K_a/K_s$ ratio were estimated from the data ($L_1$). The
likelihood ratio test statistics ($\text{LRT} = -2[\ln L_0 - \ln L_1]$)
was then compared with the $\chi^2$ distribution with one
degree of freedom to test for significant deviations from
neutrality (table 1). Because frameshift mutations occur in
HINT genes, the whole gene sequence could not be used in
every alignment. Specifically, the reading frame of avian
HINTW and HINTZ differs from human HINT.
These regions were thus excluded from the
corresponding alignments and analyses.

Multiple Copies of HINTW

HINTW is a multicopy gene present in at least 40
highly similar copies on chicken W and is repeated on the
W also in other bird species (Hori et al. 2000; O’Neill et al.
HINTW. reading frame shifts between avian rich and Arg-rich region unique to copies of chicken similarity between copies from each species (the four homogenization could occur in parallel as well. The or unequal crossing-over. Potentially, amplification and homogenization by mechanisms such as gene conversion amplified in different lineages or that it is subject to indication that the gene is frequently and independently revealed that all (studied) copies within species are more similar to each other than they are to any copy in closely related species (our unpublished observations). This would indicate that the gene is frequently and independently amplified in different lineages or that it is subject to homogenization by mechanisms such as gene conversion or unequal crossing-over. Potentially, amplification and homogenization could occur in parallel as well. The clustering of individual copies within species in phylogenetic analysis and the very high degree of sequence similarity between copies from each species (the four copies of chicken HINTW available in GenBank only show two segregating sites over 387 bp) mean that the estimation of substitution rates between fairly divergent species, such as chicken and quail, is rather insensitive to the particular copies used. The analyses presented herein are therefore based on just one copy of HINTW per species.

Results
Nonneutral Evolution of HINT on Avian Sex Chromosomes

Natural selection can be inferred from DNA sequence data using the ratio of the nonsynonymous to synonymous substitution rates (Kd/Ks). This ratio is assumed to be equal to 1 under neutral evolution, whereas a ratio above 1 or below 1 indicates adaptive (Darwinian) or negative (purifying) selection, respectively. For a phylogeny based on avian HINTW and HINTZ and human HINT sequences (fig. 2a), we found the estimated Kd/Ks ratios along individual branches to be more compatible with a free-ratio model where Kd/Ks varies between branches than with a one-ratio model where Kd/Ks is the same for all branches (the Kd/Ks ratio from the one-ratio model was estimated to 0.3460; LRT = 33.670, df = 6, P < 0.01). This shows that avian HINTW and HINTZ and human HINT have not been subject to similar selection pressures.

An analysis of Kd/Ks along individual branches (fig. 2a and table 1) revealed high Kd/Ks ratios (1.31 to 1.96) for HINTW branches, indicating positive selection and adaptive molecular evolution towards new function in a female-specific gene. In contrast, HINTZ has been subject to strong purifying selection during avian evolution (Kd/Ks = 0 to 0.13).

Adaptive Evolution of HINTW

To study further the selective pressures that have been operating on HINTW, we made pairwise comparisons of chicken and quail HINTW genes (table 1). This permits the use of the whole gene sequence in the analysis. The alignment, including both HINTW and HINTZ (fig. 3), contains frame shift mutations, meaning that part of the alignment cannot be used for the estimation of Kd/Ks ratios (see Materials and Methods). For the pairwise comparison Kd/Ks for chicken and quail HINTW is high (2.06) although not significantly different from neutral expectations (LRT = 2.38, df = 1, P = 0.123).

Although our data indicate, albeit not with statistical support, that the avian HINTW gene has been subject to positive selection, it is possible that the gene also contains conserved amino acid positions, subject to purifying selection, that reduce the overall Kd/Ks ratio (Endo, Ikeo, and Gojobori 1996; Nielsen and Yang 1998; Silberg and Liberles 2002). As a matter of fact, an analysis of the amino acid identity between chicken HINTW and HINTZ has revealed two regions with strong sequence similarity, including an α-helix and a C-terminal loop (Hori et al. 2000) (fig. 3). Three-dimensional analysis suggests that these regions are crucial for dimer formation of HINT proteins (Lima et al. 1996), and their conservation in HINTW could indicate that this protein is similarly involved in dimerization. We therefore constructed a phylogenetic tree based on an alignment without the

Table 1
Estimates of Kd/Ks Ratios from Pairwise Sequence Comparisons

<table>
<thead>
<tr>
<th>Pairwise Comparison</th>
<th>Length (bp)</th>
<th>Kd/Ks</th>
<th>LRT-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chicken HINTZ/quail HINTZ</td>
<td>375</td>
<td>0.063</td>
<td>17.56***</td>
</tr>
<tr>
<td>Chicken HINTZ/human HINT</td>
<td>366</td>
<td>0.079</td>
<td>79.96***</td>
</tr>
<tr>
<td>Quail HINTW/human HINT</td>
<td>366</td>
<td>0.082</td>
<td>79.80***</td>
</tr>
<tr>
<td>Chicken HINTW/quail HINTW</td>
<td>387</td>
<td>2.061</td>
<td>2.38 (ns)</td>
</tr>
<tr>
<td>Chicken HINTW/quail HINTW*</td>
<td>300</td>
<td>3.748</td>
<td>4.23*</td>
</tr>
</tbody>
</table>

Note.—* p < 0.05, *** p < 0.001; ns, not significant.

*a Excluding the α-helix and the C-terminal loop.

Fig. 3.—A schematic figure of the avian HINTW gene showing the regions encoding the α-helix and the C-terminal loop (solid boxes), the Leu-rich and Arg-rich region unique to HINTW (vertical lines), and the location of the HIT motif of HINTZ (horizontal lines). Regions 1 and 3 indicate reading frame shifts between avian HINTW and other HINT genes. Region 2 indicates a reading frame shift between human HINT and avian HINTZ/HINTW.

2000). Generally, the analysis of substitution rates in multicopy genes requires careful handling of orthology and paralogy. However, avian HINTW represents somewhat of a special situation. A detailed analysis of the evolution of multiple copies of HINTW in Galliformes has revealed that all (studied) copies within species are more similar to each other than they are to any copy in closely related species (our unpublished observations). This would indicate that the gene is frequently and independently amplified in different lineages or that it is subject to homogenization by mechanisms such as gene conversion or unequal crossing-over. Potentially, amplification and homogenization could occur in parallel as well. The clustering of individual copies within species in phylogenetic analysis and the very high degree of sequence similarity between copies from each species (the four copies of chicken HINTW available in GenBank only show two segregating sites over 387 bp) mean that the estimation of substitution rates between fairly divergent species, such as chicken and quail, is rather insensitive to the particular copies used. The analyses presented herein are therefore based on just one copy of HINTW per species.
regions encoding the α-helix and the C-terminal loop, and repeated Ks/Ka analyses (fig. 2b). Estimates of Ks/Ka ratios were now very high along the HINTW branches (1.51 to 19.3). Note that a very high Ks/Ka ratio (48.98) was found in the internal branch leading to galliform HINTW sequences, indicating that positive selection in HINTW is not only a recent phenomenon during avian evolution. A pairwise comparison of chicken and quail HINTW without the sequence encoding the α-helix and C-terminal loop gives Ks/Ka = 3.74 (table 1), in this case providing statistical support for deviation from neutrality (LRT = 4.23, df = 1, P = 0.04). We therefore conclude that while parts of avian HINTW are likely to have been constrained by negative selection, other parts show evidence of adaptive molecular evolution.

Discussion

HINTW is expressed in both gonads and the developing urogenital tract in the female chicken embryo before the onset of gonadal differentiation (Hori et al. 2000; O’Neill et al. 2000), making HINTW a candidate gene for avian sex determination. HINTZ is also expressed during day 3 to day 6 of embryonic development of both sexes, although at lower levels than HINTW (Hori et al. 2000). One model for the action of HINTW in the differentiation of the female gonad postulates that HINTW forms a heterodimer with HINTZ (HINTW:HINTZ), thereby blocking the formation, and the function, of a HINTZ homodimer (HINTZ:HINTZ) (Hori et al. 2000). Testes development may in this way be inhibited. Another possibility is that HINTW, either as a monomer, a homodimer, or a heterodimer with HINTZ, interacts with a target protein leading to the differentiation of ovaries (Hori et al. 2000). These hypotheses are in many respects similar to the proposed interaction of SRY with SOX3 and SOX9 in the mammalian sex determination pathway (Graves 1998). Our observation of adaptive evolution in parts of the HINTW gene but selective constraints in regions potentially involved in dimer formation is compatible with, but does not prove, these hypotheses and could suggest that the HINTW protein has derived new function(s) related to female development. Acquisition of new function(s) is further supported by the presence of a Leu-rich and Arg-rich region unique to HINTW and the absence of the histidine triad typical of other HINT proteins, including avian HINTZ.

Positive selection is not uncommon among genes involved in male reproduction (Yang and Bielawski 2000; Wyckoff, Wang, and Wu 2000; Tsaur, Ting, and Wu 1998; Lee, Ohta, and Vacquier 1995). Sexual selection arising from sperm competition can be invoked as an important driving force in such cases, but this is not applicable in the case of the female-specific HINTW gene. However, a few mammalian genes related to female reproduction have also shown evidence of adaptive evolution (Swanson et al. 2001; Swanson, Nielsen, and Yang 2003; Civetta 2003). These include autosomally encoded genes with a sex-specific pattern of expression. However, avian HINTW is the first gene restricted to the female genome documented with adaptive evolution.

The avian W chromosome shows intriguingly low levels of nucleotide diversity (Berlin and Ellegren 2001; Montell, Fridolfsson, and Ellegren 2001). For instance, a study of 3.4-kb intronic W chromosome sequence of seven different bird species failed to identify a single segregating site (Montell, Fridolfsson, and Ellegren 2001). This limited genetic diversity is unaccounted for (Ellegren 2003), but the demonstration of positive selection in HINTW now offers a clue. As W is a single segregating unit (with the exception of the pseudautosomal region), positive selection on any W-linked gene will introduce selective sweeps, reducing genetic diversity at all loci on the chromosome. Strong directional selection on HINTW may represent a causal link between selection and levels of polymorphism on the avian W chromosome.

It is of interest to note that HINTW is the only gene found in multiple copies on a female-specific W chromosome (Hori et al. 2000; O’Neill et al. 2000). This resembles the situation for several testis-specific genes that are amplified on the mammalian Y chromosome (Lahn, Pearson, and Jegalian 2001). Amplification of gene copy number might be a way of counteracting the decay of genes on the sex-limited chromosome (Vogel and Schmidtko 1998; Dechend et al. 2000; Lahn, Pearson, and Jegalian 2001), especially in cases of adaptively evolving genes, but could also be driven by sexual selection. In any case, and in summary, avian HINTW displays several characteristics that make it represent a potential female parallel to the organization and evolution of Y chromosome genes involved in male reproduction and development.

Note Added in Proofs

The gene described in this study was first reported by O’Neill et al. (2000) under the name ASW (Avian sex-specific W-linked) and by Hori et al. (2000) under the

---

**Table 2**

<table>
<thead>
<tr>
<th>Branch</th>
<th>Kp</th>
<th>Ks</th>
<th>Ks/Kp</th>
<th>N*Ks</th>
<th>S*Ks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Figure 2a</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>0.118</td>
<td>0.978</td>
<td>0.120</td>
<td>22.8</td>
<td>74.6</td>
</tr>
<tr>
<td>b</td>
<td>0.000</td>
<td>0.149</td>
<td>0.000</td>
<td>0</td>
<td>11.4</td>
</tr>
<tr>
<td>c</td>
<td>0.004</td>
<td>0.084</td>
<td>0.051</td>
<td>0.8</td>
<td>6.4</td>
</tr>
<tr>
<td>d</td>
<td>0.008</td>
<td>0.060</td>
<td>0.128</td>
<td>4.5</td>
<td>1.5</td>
</tr>
<tr>
<td>e</td>
<td>0.248</td>
<td>0.190</td>
<td>1.308</td>
<td>48.1</td>
<td>14.5</td>
</tr>
<tr>
<td>f</td>
<td>0.019</td>
<td>0.010</td>
<td>1.968</td>
<td>3.7</td>
<td>0.7</td>
</tr>
<tr>
<td>g</td>
<td>0.060</td>
<td>0.041</td>
<td>1.476</td>
<td>11.7</td>
<td>3.1</td>
</tr>
</tbody>
</table>

**Figure 2b**

<table>
<thead>
<tr>
<th>Branch</th>
<th>Kp</th>
<th>Ks</th>
<th>Ks/Kp</th>
<th>N*Ks</th>
<th>S*Ks</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.120</td>
<td>2.050</td>
<td>0.058</td>
<td>16.1</td>
<td>99.0</td>
</tr>
<tr>
<td>b</td>
<td>0.000</td>
<td>0.493</td>
<td>0.000</td>
<td>0</td>
<td>23.8</td>
</tr>
<tr>
<td>c</td>
<td>0.007</td>
<td>0.106</td>
<td>0.065</td>
<td>0.9</td>
<td>5.1</td>
</tr>
<tr>
<td>d</td>
<td>0.010</td>
<td>0.107</td>
<td>0.095</td>
<td>1.4</td>
<td>5.2</td>
</tr>
<tr>
<td>e</td>
<td>0.321</td>
<td>0.007</td>
<td>48.981</td>
<td>43.2</td>
<td>0.3</td>
</tr>
<tr>
<td>f</td>
<td>0.024</td>
<td>0.016</td>
<td>1.513</td>
<td>3.3</td>
<td>0.8</td>
</tr>
<tr>
<td>g</td>
<td>0.049</td>
<td>0.001</td>
<td>0.001</td>
<td>6.6</td>
<td>0</td>
</tr>
</tbody>
</table>

**Note.**—Estimates were obtained from the free-ratio model. Individual branches in the phylogenies were labeled a-g. For data in figure 2a, number of nonsynonymous sites along the gene (N) = 193.7; number of synonymous sites along the gene (S) = 76.3. For data in figure 2b, N = 134.7 and S = 48.3.
name Wpkci (W-linked protein kinase C inhibitor). Both these gene symbols have been used in subsequent literature, with the modification that Wpkci has been referred to as PKCIW, to conform to the nomenclature applied for other gametologous genes shared between avian Z and W chromosomes (e.g., ATP5A1Z/ATP5AIW, CHD1Z/CHD1W, SPINZ/SPINW; see Ellegren 2002). Consistent with this, the Z copy of PKCI has also been referred to as PKCIZ, PKCIZ shows a high degree of homology to postulated mammalian PKCI genes.

Subsequent to the acceptance of this paper it was brought to our attention that the genes originally described as PKCI-1 and PKCI-2 in mammals actually do not encode for protein kinase C inhibitors (Charles Brenner, Dartmouth Medical School, Lebanon). Based on the characteristic histidine triad motif in this group of proteins, the name histidine triad nucleotide-binding protein has been suggested (Brenner et al. 1997), with gene symbol HINT. Although not yet consequently used, HINT has recently been introduced as a gene symbol for the Z-linked copy in birds (e.g., http://www.thearkdb.org/).

Accepting HINT and to follow the abovementioned nomenclature for gametologous genes shared between avian Z and W sex chromosomes, we suggest that the ASW/Wpkci/PKCIW gene should be referred to as HINTW, which is thus used in this paper. It follows that the Z copy should be referred to as HINTZ. This nomenclature is not specific for birds, but is also applied for gametologous genes shared between mammalian X and Y (e.g., SlX1|SlY1, SOX3|SRY, etc.) and plant X related nucleotide-binding proteins. Nat. Struct. Biol. 4:231–238.

Consistent with this, the Z copy of PKCI has also been referred to as PKCIZ, PKCIZ shows a high degree of homology to postulated mammalian PKCI genes.


Scott Edwards, Associate Editor

Accepted August 6, 2003