Rooting a Phylogeny with Homologous Genes on Opposite Sex Chromosomes (Gametologs): A Case Study Using Avian CHD

Jaime García-Moreno¹ and David P. Mindell

Museum of Zoology and Department of Biology, University of Michigan

We describe a previously unrecognized form of gene homology using the term “gametology,” which we define as homology arising through lack of recombination and subsequent differentiation of sex chromosomes. We demonstrate use of gametologous genes to root each other in phylogenetic analyses of sex-specific avian Chromo-helicase-DNA binding gene (CHD) sequences. Phylogenetic analyses of a set of neognath bird sequences yield monophyletic groups for CHD-W and CHD-Z gametologs, as well as congruent relationships between these two clades and between them and current views of avian taxonomy. Phylogenetic analyses including paleognath bird CHD sequences and rooting with crocodilian CHD sequences, suggest an early divergence for paleognath CHD within the avian CHD clade. Based on our CHD analyses calibrated with avian fossil dates, we estimate the divergence between CHD-W and CHD-Z at 123 MYA, suggesting an early differentiation of sex chromosomes that predates most extant avian orders. In agreement with the notion of male-driven evolution, we find a faster rate of change in male-linked CHD-Z sequences.

Introduction

“Homology” denotes a relationship among traits due to common ancestry and can be assessed at many levels of organization. Molecular evolutionists recognize multiple kinds of homologous sequences, distinguished primarily by the processes involved in their origins and differentiation. Orthologous genes in different taxa arise due to lineage splitting, as for cytochrome c in humans and in chimpanzees. Paralogous genes arise due to gene duplication, as for alpha-globin and beta-globin in humans. Xenologous genes or sequences arise due to lateral transfer, often via retroviruses, between different taxa.

When homoplasious similarity among traits (due to convergence, parallel evolution, or reversals) is abundant relative to homologous similarity, phylogenetic analyses can be misled (Cavender 1978; Felsenstein 1978; Hendy and Penny 1989). Even when sequence evolution is clocklike, tree-building inconsistency can occur due to positional rate heterogeneity and taxon differences in nucleotide/amino acid composition and distribution of variable sites (see Steel, Huson, and Lockhart 2000), particularly when outgroup taxa are very distant related to the ingroup taxa. One potential solution for outgroup rooting is to use paralogous genes, with one serving as the phylogenetic root for the other. This can be done when the gene duplication occurs after the divergence between the closest organismal outgroup and the ingroup taxa, but before ingroup diversification. This results in the duplicated forms of the gene being more closely related to one another than is to the single gene of the outgroup. This innovation was applied first in phylogenetic analyses for the tree of life, where no suitable outgroup organisms exist (Gogarten et al. 1989; Iwabe et al. 1989; Doolittle and Brown 1994), and more recently in analyses of chaetognaths (Telford and Holland 1997) and angiosperms (Donoghue and Mathews 1998). A general difficulty with this approach is our limited knowledge of paralogous genes that fit the evolutionary pattern described.

Here, we develop a similar, but different, approach to rooting phylogenetic analyses. Instead of using paralogous genes for reciprocal rooting, we use homologous genes on opposite sex chromosomes. Related genes located on opposite sex chromosomes present a unique and previously unnamed form of homology. We recognize “gametologous” genes as arising via nonrecombination and differentiation of sex chromosomes. Barriers to recombination between entire or portions of opposite sex chromosomes facilitate differentiation for gametologs, in a manner similar to lineage splitting and gene duplication facilitating differentiation for orthologs and paralogs. We use “gamete” as the word root because sex chromosomes are distributed differently in sex-specific gametes.

Recognition of gametologous relationships for characters is warranted in theory, as they fit the criterion of sharing common ancestry which lies at the heart of homology definition (Van Valen 1982; Roth 1988; Mindell 1991; Hillis 1994), and they arise by an evolutionary process that is different from those underlying the three kinds of homology relationship described previously. Recognition and use of particular gametologs in phylogenetic analyses can be supported by evidence indicating that (1) they are mutually distinguishable and linked to opposite sex chromosomes, (2) they share common, most recent decent from a homologous autosome pair, and (3) they yield phylogenetic hypotheses that are congruent with each other and with hypotheses based on independent data sets. Ideally, gametologs should evolve independently under similar functional constraints, and their age relative to subsequent divergences within gametolog clades should not yield juxtaposition of very long and short branches. Here, we ex-

¹ Present address: Department of Biology, University of Konstanz, Konstanz, Germany.

Key words: gametolog, homology, phylogeny rooting, avian phylogeny, sex chromosome evolution, Chromo-helicase-DNA binding gene.

Address for correspondence and reprints: Jaime García-Moreno, Department of Biology, University of Konstanz, D-78457 Konstanz, Germany. E-mail: jaime.garcia-moreno@uni-konstanz.de.

© 2000 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

1826
amine rooting with Chromo-helicase-DNA binding gene (CHD) gametologs in birds, and we consider the three criteria of potential support mentioned above. Points 1 and 2 have recently been addressed by others (references below), and our new data and analyses focus on point 3. Brief phylogenetic analyses including CHD-Z and CHD-W have been presented by Fridolfsson et al. (1998) and Kahn and Quinn (1999) for one and four bird species, respectively. We seek to extend their analyses with a larger sampling of species providing a more detailed assessment of congruence and to develop and justify the approach more explicitly.

In support of the first criterion, CHD has been shown to exist in two recognizably different forms in most bird species (neognaths; see below). One is found on the Z chromosome and thus is present in both sexes (CHD-Z), and the other is on the W chromosome and present only in females (CHD-W), the heterogametic sex in birds (Woodage et al. 1997; Fridolfsson et al. 1998; Griffiths et al. 1998; Kahn, St. John, and Quinn 1998; Fridolfsson and Ellegren 1999). There is no evidence for recombination between these two genes, and no autosomal copies have been detected (Fridolfsson et al. 1998; Kahn and Quinn 1999). Traditionally, avian taxonomists have recognized two primary clades for extant birds, Paleognathae (ratites and tinamous) and Neognathae (all others). Distinct CHD-Z and CHD-W genes have been found in all neognaths assessed but not in paleognaths. Although subtle sex chromosome differences exist in at least some paleognaths (Ansari, Takagi, and Sasaki 1988), the CHD-Z and CHD-W genes (gametologs) cannot be readily distinguished from each other in paleognaths, and we refer to paleognath CHD sequences as CHD-paleognath.

Materials and Methods
Taxon Sampling

We isolated DNA, using standard methods (Sambrook, Fritsch, and Maniatis 1989), from the following avian species: *Bonasa umbellus* (ruffed grouse), *Phasianus colchicus* (common pheasant), *Acryllium vulturinum* (vulturine guineafowl) (Galliformes); *Aythya americana* (redhead), *Anas platyrhynchos* (mallard) (Anseriformes); *Gampsonyx swainsonii* (mallard); *Accipiter nisus* (red-tailed hawk) (Falconiformes); *Grytvagn oxya* (great skua), *Spheniscus demersus* (penguin) (Sphenisciformes). We also isolated DNA sequences from *Crocodilus porosus* (saltwater crocodile), *Caiman yacare* (Yacare caiman), and *Gavialis gangeticus* (gharial). We also used published sequences for *Gallus gallus* (chicken) and *Taeniopygia guttata* (zebra finch) (Griffiths et al. 1998).

PCR and Sequencing Protocol

We obtained PCR products for CHD following standard protocols using primers P2 and P8, described by Griffiths et al. (1998). The amplification product included approximately equal parts of a complete intron and portions of its two flanking exons. Single products for homogametic male neognaths and for paleognaths of both sexes were cleaned by column purification using QIAquick columns (QIAGEN Inc.) following the supplier’s protocol. Two products for heterogametic female neognaths were resolvable in 3% LMP agarose gels, as the amplified fragment spanned an intron that varied in length between CHD-W and CHD-Z (Griffiths et al. 1998; Miyaki et al. 1998). The two resulting bands were each excised and cleaned using the gel extraction version of the QIAquick kit. Clean products were sequenced using an ABI377 sequencer. Sequences have been deposited in GenBank under accession numbers AF288487–AF288516 and AF006659–AF006662.

Phylogenetic Analyses

Sequences were initially aligned using Clustal X (Thompson et al. 1997) with the following settings: gap opening = 10; gap extension = 0.05; delay divergent sequences = 40; DNA transition weight = 0.5. The resulting alignment was adjusted by eye to minimize mismatches, and gaps were either treated as missing (in analyses including intron sequences) or excluded (in the analysis of exon sequences only). The alignments are available from us on request. We performed a range of phylogenetic analyses using maximum parsimony (MP) and maximum likelihood (ML). For MP, we used equal weights for all characters, with gaps treated as missing. For the ML analyses, we chose the Hasegawa-Kishino-Yano (HKY) model accounting for invariable positions and unequal rates of substitution following a gamma distribution based on model performance comparisons using likelihood ratio tests (Huelsenbeck and Rannala 1997). Alternative tree topologies were compared using the Kishino and Hasegawa (1989) (KH) test. To estimate support for particular nodes within trees, we performed character bootstrap replicates and jackknife replicates with 50% deletion of sequences and random addition of sequences (Felsenstein 1985). We also evaluated relationships for all possible species quartets using the ML criterion and 1,000 puzzling steps (Strimmer and von Haeseler 1996), as implemented in PAUP* (Swofford 1999).

Results and Discussion
Characterization of Avian CHD Genes

We obtained DNA sequences for a fragment of CHD-Z and a fragment of CHD-W for six neognath bird
Phylogenetic Relationships

Phylogenetic analysis of CHD sequences from 19 species of neognath birds yields distinct, monophyletic
groups for CHD-W and CHD-Z sequences (fig. 1a). Inferred relationships for birds within each of these two primary clades are congruent, with each gametolog showing monophyly of Galloanseres (Galliformes [chicken, pheasants] plus Anseriformes [waterfowl]) and Passeriformes (songbirds) and an early separation of Galloanseres from the other neognaths. Phylogenetic analyses excluding the intron (not shown) yields a tree similar to that in figure 1a; the distinction between CHD-W and CHD-Z is maintained, as is the monophyly and early branching of Galloanseres in both clades, despite the small number (168 bases total) of characters considered.

Additional analyses including the three crocodilian and six paleognath bird CHD sequences strongly support two primary monophyletic groups, one comprising the three crocodilian CHD sequences and another, larger, clade including all of the avian CHD sequences (fig. 1b). Within the avian clade, there are three separate monophyletic groups formed by CHD-W, CHD-Z, and the CHD-paleognath sequences. Relatively low bootstrap values for the CHD-Z + CHD-W clade in figure 1b are due to occasional movement of one or another of them to the paleognath CHD clade and may be a function of the small number of sequence characters available. When the tree is rooted with the crocodilian clade, the relationships among and within CHD-W and CHD-Z are essentially the same as those shown in figure 1a, while the paleognath CHD sequences are shown as diverging basally within the avian clade. Analyses using only one or the other gametolog (not shown) gave results fully congruent with each other and with the analyses presented above.

Many of the phylogenetic relationships depicted in the CHD-Z clade, where our taxon sampling was the most extensive, are congruent with current views of avian phylogeny based, variously, on independent morphological (Cracraft 1988), nuclear (e.g., Sibley and Ahlquist 1990; Groth and Barrowclough 1999; Van Tuinen, Sibley, and Hedges 2000), and mitochondrial (Mindell et al. 1997, 1999; Härlid, Janke, and Arnason 1998; Härlid and Arnason 1999) data sets. These include monophyly of Galloanseres, Galliformes, Anseriformes, Piciiformes (woodpeckers and allies), Passeriformes, Accipitriformes (hawks and allies), Picidae (woodpeckers), oscine Passeriformes, and suboscine Passeriformes. Recent analyses of mitochondrial genes indicate a relatively basal position for Passeriformes and a more derived position for paleognaths (Mindell et al. 1997, 1999; Härlid, Janke, and Arnason 1998; Härlid and Arnason 1999); however, recent analyses of nuclear genes support the more traditional view of paleognaths diverging basally and Passeriformes being more derived (Groth and Barrowclough 1999; Van Tuinen, Sibley, and Hedges 2000), and it is not surprising that CHD analyses are congruent with those of other nuclear genes. Relatively sparse taxon sampling and distantly related outgroups remain as problems in many of the analyses cited and may underlie the differences found. Mindell et al. (1999, table 5) did find optimal trees congruent with conventional views (paleognaths diverging basally) based on mitochondrial data when using only 2 reptilian outgroup taxa, rather than 11 outgroup taxa representing reptiles, mammals, and an amphibian, combined with ML analyses accounting for evolutionary rate heterogeneity.

The remaining criterion to be considered in support of using CHD gametologs is that they descend from alternative members of a homologous autosome pair. Our results showing monophyly for CHD-Z and CHD-W sequences and showing those clades to be sister groups do indicate a single common origin for the CHD-Z sequences and for the CHD-W sequences (fig. 1). This is in agreement with analyses by Fridolfsson et al. (1998) and Kahn and Quinn (1999) on smaller sets of taxa.

Others have pointed out that paralogous genes with different functions may evolve under different constraints, yielding differences in rate, nucleotide composition, and distribution of variable sites, and that these can potentially mislead phylogeny reconstruction (Lockhart et al. 1996; Philippe and Laurent 1998). The same potential for problems exists in the use of gametologs. Nevertheless, we note that avian CHD-W and CHD-Z gametologs analyzed here appear to share the same function, differing only in intron size (Fridolfsson et al. 1998; Griffiths et al. 1998).

Sex Chromosome Evolution in Neognath Birds

Recognition and analyses of gametologs also provide the opportunity to learn about sex chromosome evolution in neognath birds by estimating the time of divergence between the CHD-Z and CHD-W clades. To obtain this estimate, we used an ML approach based on quartets allowing for evolutionary rate heterogeneity (QDate v1.1; Rambaut and Bromham 1998). A range of sequences and multiple fossil dates are used, rather than reliance on a single calibration rate. We derived the ML model parameters from our optimal tree (fig. 1a) and implemented 100 replicates of the two-rate model, excluding sequences for which rate heterogeneity was detected. The potential for “male-driven evolution” (see below) yielding faster rates in CHD-Z than in female-specific CHD-W is not a biasing factor in these analyses, as two different rates are accommodated in the calculations. We used fossil-based calibration estimates of 40 MYA for the divergence between Phasianidae and Numidae (Benton 1993; Kornegay et al. 1993) and 68 MYA for the divergence between Galliformes and Anseriformes (see Waddell et al. 1999). Eleven quartets passed the rate homogeneity test and were used to estimate the average divergence between CHD-W and CHD-Z at 123 MYA, with a standard deviation of 6 Myr among the 11 quartets (fig. 2). This estimate predates an estimate of 55 Myr of age for the primary radiation of extant avian orders based on fossil evidence (e.g., Feduccia 1996), and is closer to estimates ranging from <90 to >130 Myr based on molecular data (Hedges et al. 1996; Cooper and Penny 1997; Kumar and Hedges 1998; Waddell et al. 1999). If the estimate of 123 MYA
Sex Chromosomes in Paleognath Birds

Gene composition of sex chromosomes in paleognaths appears similar to that of sex chromosomes in other birds (Ogawa, Murata, and Mizuno 1998). Paleognath sex chromosomes, however, are largely homomorphic and show banding patterns similar to each other, in contrast to the more strongly heteromorphic and heterochromatic nature of sex chromosomes in other birds, suggesting recombination in paleognaths (Ansari, Takagi, and Sasaki 1988; Ogawa, Murata, and Mizuno 1998). Even though genetic and sex chromosome differences exist in the paleognaths (Ogawa, Murata, and Mizuno 1998), only one type of CHD sequence is obtained which cannot be readily assigned to CHD-W or CHD-Z (Griffiths et al. 1998; Fridolfsson and Ellegren 1998). Each square represents an estimate based on a quartet including two CHD-W sequences and two CHD-Z sequences. The bars represent the 95% confidence intervals estimated for each quartet.

Conclusions

We have shown that related genes on opposite sex chromosomes (gametologs) can be useful in rooting phylogenetic analyses. Avian gametologs CHD-Z and CHD-W satisfy the three criteria of being mutually distinguishable and linked to opposite sex chromosomes, sharing common decent from alternative members of a homologous autosome pair, and yielding phylogenetic hypotheses that are largely congruent with each other and with hypotheses based on independent data sets. We demonstrated the approach for a set of taxa large enough to assess congruence for multiple avian relationships, and found congruence both between the CHD-W and CHD-Z clades for the same taxa and between our results and traditional views of avian phylogeny based on alternative molecular and morphological data sets. We estimated a divergence time of 123 MYA for the avian gametologs, suggesting a single, early diversification of avian sex chromosomes prior to the radiation of most extant orders.

The general approach we have outlined could be applied to any set of species sharing a unique origination of sex chromosomes with a subsequent lack of recombination. This situation is apparent in multiple vertebrate taxa (reviewed in Solari 1994). Although only environmental sex determination is known for crocodilians, male or female heterogamety may have had three or more separate origins within lizards (Sauria), based on the current distribution of those traits in phylogenetically distant taxa (Bull 1980; Olmo 1986). Similar disparate appearances of male or female heterogamety suggesting independent origins are found within hidden-neck turtles (cryptodires) and snakes. Male heterogamety, considered homologous in placental mammals and marsupials (but see Toder et al. 1997), appears to have arisen independently in monotremes. Although understanding of independent origins of sex determination in tetrapods is limited and relatively few gametologous genes have been identified, the general approach we have presented can be applied where such information is available. In turn, estimation of phylogeny for gametologs as we have done for CHDs can provide insights into the evolution of sex determination and the relative timing of its origins.
Acknowledgments

We thank Lou Densmore and Jon Fjeldså for kindly providing DNA or tissue samples. We also thank Michael Sorenson, Priscilla Tucker, Peter Arctander, Tom Quinn, and two anonymous reviewers for valuable comments on earlier drafts of this paper. Funding was provided by NSF grant DEB 9726427.

LITERATURE CITED


QUINN, AND TWO ANONYMOUS REVIEWERS FOR VALUABLE COMMENTS ON EARLIER DRAFTS OF THIS PAPER. FUNDING WAS PROVIDED BY NSF GRANT DEB 9726427.


OGAWA, A., K. MURATA, AND S. MIZUNO. 1998. The location of Z- and W-linked marker genes and sequence on the ho-


Ross Crozier, reviewing editor

Accepted August 7, 2000