IBIS The International Journal of Avian Science BOU
Ibis (2010), **152**, 223–225

Commentary

When non-coding is non-neutral: the role of CHD1 gene polymorphism in sexing, in phylogenetics and as a correlate of fitness in birds

LAURYN BENEDICT,[1] RAURI C. K. BOWIE,[1]* JEROME FUCHS[1,2] & MATTHEW MACMANES[1]
[1]*Museum of Vertebrate Zoology and Department of Integrative Biology, 3101 Valley Life Sciences Building, University of California, Berkeley, CA 94720, USA*
[2]*DST/NRF Centre of Excellence at the Percy FitzPatrick Institute, University of Cape Town, Rondebosch 7701, Cape Town, South Africa*

Ornithologists routinely use DNA data for all kinds of studies, from measuring population sex ratios to determining phylogenetic relationships between avian orders. Although the size of an avian genome is over 1 billion base pairs (Hillier *et al.* 2004), not all of this genetic information is presently useful to researchers (Ellegren 2005). Natural selection can create both divergence and convergence at the level of DNA, causing genetic markers that are under selection to evolve differently than the rest of the genome. Thus, regions of non-coding DNA (introns) have increasingly become popular markers, mostly because they are thought to evolve under a model of neutral evolution (but see Bejerano *et al.* 2004). An intron within the Chromo-Helicase-DNA Binding (CHD1) gene on the avian Z-chromosome is a specific example of a widely used, non-coding and presumably neutral marker. Researchers studying population parameters frequently use CHD1 as a sex-specific marker with analogues on the Z- and W-chromosomes (Griffiths *et al.* 1998). Phylogeneticists often sequence the CHD1 intron as a marker for use in studies determining population and species relationships (Garcia-Moreno & Mindell 2000, Peters *et al.* 2005).

In this issue of *Ibis*, Schroeder *et al.* (2010) demonstrate that in a population of Black-tailed Godwits *Limosa l. limosa*, CHD1-Z is polymorphic, but surprisingly, particular alleles correlate with fitness, thus violating the basic assumption of neutrality. The authors studied a population of Black-tailed Godwits with a two-allele (Z, Z*) length polymorphism in CHD1 on the Z-chromosome together with the single allele on the W-chromosome. Heterozygotic males (the homogametic sex) with the less common Z* allele had more attractive plumage characteristics and had mates that laid larger eggs (Schroeder *et al.* 2009, 2010). Females (the heterogametic sex) with the Z* allele were larger and initiated clutches earlier in the season than females with the Z allele (Schroeder *et al.* 2010). Schroeder *et al.* (2010) do not suggest that the CHD1 gene region has any direct effects on fitness, but rather that this gene region is linked by physical proximity on the chromosome to genes that affect plumage coloration and body mass. A previous study (Lee *et al.* 2002) also found fitness correlates of a rare CHD1–Z allele in Common Moorhens *Gallinula chloropus*, but in this species heterozygosity in the CHD1 intron region was correlated with higher mortality. Unlike in the Moorhen system, CHD1 heterozygosity in Black-tailed Godwits is correlated with traits that have positive fitness consequences. These results underline that non-coding DNA regions may evolve as the products of natural selection, thus violating one of the fundamental assumptions of both population genetic and phylogenetic analyses. Such an outcome is, obviously, a well-discussed possibility but has not often been demonstrated for a single avian allele, particularly one at a locus which is so widely used for a variety of research questions (Chapman *et al.* 2009). Interestingly, the Z-chromosome is important in the speciation process in birds (Sæther *et al.* 2007), with some loci potentially involved in the formation and/or maintenance of reproductive isolation (e.g. Haldane 1922, Carling & Brumfield 2008). Hence, the fact that some loci are under strong selection on this chromosome is not surprising.

What do the results of Schroeder *et al.* (2010) mean for other researchers? Most obviously, they illustrate the known caution that sexing birds with the CHD1 locus should be done carefully, always checking for the presence of more than two alleles in the population (Dawson *et al.* 2001). Their results also suggest that researchers using CHD1 to determine relationships between populations or species should do so with caution. One strategy to minimize problems associated with selection at one locus is to use multiple unlinked markers. Introns are now probably the most widely used markers in avian phylogenetics, with more than 250 loci available for use to help resolve the avian tree of life (Backström *et al.* 2008, Hackett *et al.* 2008, Kimball *et al.* 2009).

Researchers interested in the phylogeny and phylogeography of birds have moved away from the use of single locus datasets (mostly mitochondrial) toward the combination of several mitochondrial and nuclear (both on autosomes and on sex chromosomes) loci in an effort to estimate more accurately the demographic parameters underpinning species histories (e.g. effective population size, migration rates among populations, time of

*Corresponding author.
Email: bowie@berkeley.edu

population divergence; see Jennings & Edwards 2005, Bowie *et al.* 2009). The CHD1-Z intron is a common nuclear marker used for such studies (Peters *et al.* 2005) and should be carefully evaluated for evidence of selection. Although requiring population-level sampling, one way to test for selection is to make use of several different neutrality tests such as Tajima's D (Tajima 1989), Fu's Fs (Fu 1996) and Hudson–Kreitman–Aguadé (Hudson *et al.* 1987), carefully evaluating their underlying assumptions in light of the specific research question being addressed (reviewed in Nielsen 2005, Nielsen & Beaumont 2009). All of these tests can be applied to intron data and are readily implemented in widely used software (reviewed by Excoffier & Heckel 2006).

The finding that the CHD1 intron is affected by selection may be most problematic for studies of avian relationships above the level of species. Selection *per se* may not be a problem for phylogeneticists if it is occurring in a consistent manner across the species studied. Problems will start to appear, however, when selection is limited to a specific set of branches within the tree. In this case, one may observe a slowdown or an acceleration of the substitution rate for this branch and eventually all of its descendants. Such biases may result in spurious phylogenetic trees with very different branch-lengths at the tips and should be carefully evaluated. Several tests exist to detect selection in a phylogenetic context, but such tests only apply to protein-coding genes and hence are of no use when the sequence data are from introns (McDonald & Kreitman 1991, Yang 2006).

Studies such as that of Schroeder *et al.* have important implications not only for researchers using CHD1 for phylogenetics, but also for phylogeneticists in general, as selection is not likely to be restricted to the CHD1 intron. However, we maintain that introns should continue to be used in phylogenetic studies, as data from a genomic dataset of mammals (Pollard *et al.* 2010) suggest that only a small fraction of sites in intron sequences (2.2%) are experiencing selective constraints; it would be surprising if the same were not true for bird genomes. Furthermore, new methods have been developed recently to test for the existence of constrained sites in non-coding sequences (e.g. Kim & Pritchard 2007, Pollard *et al.* 2010). Unfortunately, these methods mostly apply to genomic datasets and require an already established phylogeny. As such, until more specific methods are developed for non-coding sequence data, perhaps the simplest way to detect selection is to perform a molecular clock test to assess the constancy of rates of molecular evolution across the phylogenetic tree of interest. We expect that selection will rarely be detected, but when it is, intron loci will provide a valuable tool in helping to isolate candidate loci of major phenotypic effect.

Researchers seeking to understand the evolution of phenotypic traits have made great strides in identifying regions of the genome associated with fitness (Stinchcombe & Hoekstra 2008). We now know that specific genes control phenotypic traits related to reproductive success or longevity. Well-studied systems include MHC genes that affect mating preferences (Penn & Potts 1999, Zelano & Edwards 2002), and coat or plumage colour genes that affect survival in different environments or play a central role in mate preference/species recognition (Mullen *et al.* 2009, Uy *et al.* 2009, Rosenblum *et al.* 2010). The vast majority of phenotypic traits, however, have genetic control systems that remain a mystery. Methods for isolating these genes are generally time-intensive, but if well-known non-coding sequences are linked to genes of major effect, then researchers can target a potentially very narrow area of the genome when searching for the genes underpinning a particular fitness trait (Edwards & Dillon 2004). For example, the apparent role of the Z-chromosome in speciation is due to the presence of both species-recognition genes and species-specific plumage coloration genes on this chromosome (Sæther *et al.* 2007). Schroeder *et al.* also suggest that regions of the Z-chromosome control plumage coloration and possibly patterning, providing an excellent target for identifying genes with a direct effect on phenotype.

Studies have now demonstrated that the CHD1 gene is affected by selection in at least two different bird species (Lee *et al.* 2002, Schroeder *et al.* 2010). The fact that CHD1 is so widely used in ornithological studies may be both good and bad news in light of this result. The bad news is that other studies using CHD1 as a presumed neutral marker may have skewed outcomes. This problem can be detected by testing for selection as mentioned above, and can be minimized by using multiple unlinked loci. The good news is that the popularity of CHD1 has led to a more complete understanding of how this intron evolves, and may provide the means to help identify genes that underpin plumage polymorphism in a diversity of birds. We encourage other researchers using introns to be aware of the possible influence of selection and to couple genetic information with fitness data, as is so rarely done.

REFERENCES

Backström, N., Fagerber, S. & Ellegren, H. 2008. Genomics of natural bird populations: a gene-based set of reference markers evenly spread across the avian genome. *Mol. Ecol.* **17**: 964–980.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S. & Haussler, D. 2004. Ultraconserved elements in the human genome. *Science* **304**: 1321–1325.

Bowie, R.C.K., Fjeldså, J. & Kiure, J. 2009. Multilocus molecular DNA variation in Winifred's Warbler *Scepomycter winifredae* suggests cryptic speciation and the existence of a threatened species in the Rubeho-Ukaguru Mountains of Tanzania. *Ibis* **151**: 709–719.

Carling, M.D. & Brumfield, R.T. 2008. Haldane's rule in an avian system: using cline theory and divergence population genetics to test for differential introgression of mitochondrial, autosomal, and sex-linked loci across the *Passerina* bunting hybrid zone. *Evolution* **62**: 2600–2615.

Chapman, J.R., Nakagawa, S., Coltman, D.W., Slate, J. & Sheldon, B.C. 2009. A quantitative review of heterozygosity-fitness correlations in animal populations. *Mol. Ecol.* **18**: 2746–2765.

Dawson, D.A., Darby, S., Hunter, F.M., Krupta, A.P., Jones, I.L. & Burke, T. 2001. A critique of avain CHD-based molecular sexing protocols illustrated by a Z-chromosome polymorphism detected in auklets. *Mol. Ecol. Notes* **1**: 201–204.

Edwards, S.V. & Dillon, M. 2004. Hitchhiking and recombination in birds: evidence from Mhc-linked and unlinked loci in Red-winged Blackbirds (*Agelaius phoeniceus*). *Genet. Res.* **84**: 175–192.

Ellegren, H. 2005. The avian genome uncovered. *Trends Ecol. Evol.* **20**: 180–186.

Excoffier, L. & Heckel, G. 2006. Computer programs for population genetics data analysis: a survival guide. *Nat. Rev. Genet.* **7**: 745–758.

Fu, Y.X. 1996. New statistical tests of neutrality for DNA samples from a population. *Genetics* **143**: 557–570.

Garcia-Moreno, J. & Mindell, D.P. 2000. Rooting a phylogeny with homologous genes on opposite sex chromosomes (gametologs): a case study using avian CHD. *Mol. Biol. Evol.* **17**: 1826–1832.

Griffiths, R., Double, M.C., Orr, K. & Dawson, R.J.G. 1998. A DNA test to sex most birds. *Mol. Ecol.* **7**: 1071–1075.

Hackett, S.J., Kimball, R.T., Reddy, S., Bowie, R.C.K., Braun, E.L., Braun, M.J., Chojnowski, J.L., Cox, W.A., Han, K.-L., Harshman, J., Huddleton, C.J., Marks, B.D., Miglia, K.J., Moore, W.A., Sheldon, F.H., Steadman, D.W., Witt, C.C. & Yuri, T. 2008. A phylogenomic study of birds reveals their evolutionary history. *Science* **320**: 1763–1768.

Haldane, J.B.S. 1922. Sex ratio and unisexual sterility in animal hybrids. *J. Genet.* **12**: 101–109.

Hillier, L.W., *et al.* 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**: 695–716.

Hudson, R.R., Kreitman, M. & Aguadé, M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**: 153–159.

Jennings, W.B. & Edwards, S.V. 2005. Speciational history of Australian Grass Finches (*Poephila*) inferred from 30 gene trees. *Evolution* **59**: 2033–2047.

Kim, S.Y. & Pritchard, J.K. 2007. Adaptive evolution of conserved noncoding elements in mammals. *PLoS Genet.* **3**: 1572–1586.

Kimball, R.T., Braun, E.L., Barker, F.K., Bowie, R.C.K., Braun, M.J., Chojnowski, J.L., Hackett, S.J., Han, K., Harshman, J., Heimer-Torres, V., Holznagel, W., Huddleston, C.J., Marks, B.D., Miglia, K.J., Moore, W.S., Reddy, S., Sheldon, F.H., Smith, J.V., Witt, C.C. & Yuri, T. 2009. A well-tested set of primers to amplify regions spread across the avian genome. *Mol. Phylogenet. Evol.* **50**: 654–660.

Lee, P.L.M., Brain, P.F., Forman, D.W., Bradbury, R.B. & Griffiths, R. 2002. Sex and death: CHD1Z associated with high mortality in Moorhens. *Evolution* **56**: 2548–2553.

McDonald, J.H. & Kreitman, M. 1991. Apaptive protein evolution at the Adh locus in Drosophila. *Nature* **351**: 652–654.

Mullen, L.M., Vignieri, S.N., Gore, J.A. & Hoekstra, H. 2009. Adaptive basis of geographic variation: genetic, phenotypic and environmental differences among beach mouse populations. *Proc. R. Soc. Lond. B* **276**: 3809–3818.

Nielsen, R. 2005. Molecular signatures of natural selection. *Annu. Rev. Genet.* **39**: 197–218.

Nielsen, R. & Beaumont, M.A. 2009. Statistical inferences in phylogeography. *Mol. Ecol.* **18**: 1034–1047.

Penn, D.J. & Potts, W.K. 1999. The evolution of mating preferences and major histocompatibility complex genes. *Am. Nat.* **153**: 145–164.

Peters, J.L., McCracken, K.G., Zhuravlev, Y.N., Lu, Y., Wilson, R.E., Johnson, K.P. & Omland, K.E. 2005. Phylogenetics of wigeons and allies (Anatidae: *Anas*): the importance of sampling multiple loci and multiple individuals. *Mol. Phylogenet. Evol.* **35**: 209–224.

Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R. & Siepel, A. 2010. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**: 110–121.

Rosenblum, E.B., Römpler, H., Schöneberg, T. & Hoekstra, H.E. 2010. Molecular and functional basis of phenotypic convergence in white lizards at white sands. *Proc. Natl Acad. Sci. USA* **107**: 2113–2117.

Sæther, S.A., Sætre, G., Borge, T., Wiley, C., Svedin, N., Andersson, G., Veen, T., Haavie, J., Servedio, M.R., Bureš, S., Král, M., Hjernquist, M.B., Gustafsson, L., Träff, J. & Qvarnström, A. 2007. Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. *Science* **318**: 95–97.

Schroeder, J., Lourenço, P.M., Hooijmeijer, J.C.E.W., Both, C. & Piersma, T. 2009. A possible case of contemporary selection leading to a decrease in sexual plumage dimorphism in a grassland-breeding shorebird. *Behav. Ecol.* **20**: 797–807.

Schroeder, J., Kentie, R., van der Velde, M., Hooijmeijer, J.C.E.W., Both, C., Haddrath, O., Baker, A.J. & Piersma, T. 2010. Linking intronic polymorphism on the CHD1-Z gene with fitness correlates in Black-tailed Godwits *Limosa l. limosa*. *Ibis* **152**: 368–377.

Stinchcombe, J.R. & Hoekstra, H.E. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* **100**: 158–170.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.

Uy, J.A.C., Moyle, R.G., Filardi, C.E. & Cheviron, Z.A. 2009. Difference in plumage color used in species recognition between incipient species is linked to a single amino acid substitution in the Melanocortin-1 Receptor. *Am. Nat.* **174**: 244–254.

Yang, Z. 2006. *Computational Molecular Evoluation*. Oxford: Oxford University Press.

Zelano, B. & Edwards, S.V. 2002. An Mhc component to kin recognition and mate choice in birds: predictions, progress, and prospects. *Am. Nat.* **160**: 225–237.