Retroposon Insertions and the Chronology of Avian Sex Chromosome Evolution

Alexander Suh,*[,1] Jan Ole Kriegs,[1,2] Jürgen Brosius,[1] and Jürgen Schmitz[1]

[1]Institute of Experimental Pathology (ZMBE), University of Münster, Münster, Germany

[2]LWL-Museum für Naturkunde, Westfälisches Landesmuseum mit Planetarium, Münster, Germany

***Corresponding author:** E-mail: asuh@uni-muenster.de.

Associate editor: Yoko Satta

Abstract

The vast majority of extant birds possess highly differentiated Z and W sex chromosomes. Nucleotide sequence data from gametologs (homologs on opposite sex chromosomes) suggest that this divergence occurred throughout early bird evolution via stepwise cessation of recombination between identical sex chromosomal regions. Here, we investigated avian sex chromosome differentiation from a novel perspective, using retroposon insertions and random insertions/deletions for the reconstruction of gametologous gene trees. Our data confirm that the *CHD1Z*/*CHD1W* genes differentiated in the ancestor of the neognaths, whereas the *NIPBLZ*/*NIPBLW* genes diverged in the neoavian ancestor and independently within Galloanserae. The divergence of the *ATP5A1Z*/*ATP5A1W* genes in galloanserans occurred independently in the chicken, the screamer, and the ancestor of duck-related birds. In Neoaves, this gene pair differentiated in each of the six sampled representatives, respectively. Additionally, three of our investigated loci can be utilized as universal, easy-to-use independent tools for molecular sexing of Neoaves or Neognathae.

Key words: birds, retroposon phylogeny, CR1, gametolog, sex chromosome, molecular sexing.

Unlike mammals with male XY heterogamety, birds possess a sex chromosomal system with ZW heterogamety in females. Comparisons of such contrasting features have helped to understand not only the evolution of female heterogamety (Fridolfsson et al. 1998) but also the differentiation of sex chromosomes in general (Bergero and Charlesworth 2008). The avian Z and W sex chromosomes evolved from a pair of autosomes via regional suppression of interchromosomal recombination, independently from those in mammals (Fridolfsson et al. 1998; Bellot et al. 2010). The neognaths, representing most extant birds, exhibit highly differentiated sex chromosomes (Fridolfsson et al. 1998; Shetty et al. 1999) with a largely degenerated W chromosome. The paleognaths, the early diverged sister group of neognaths, feature a more plesiomorphic autosome-like situation, for the Z and W chromosomes of ratites (ostriches and related flightless birds) recombine along most of their length (Ogawa et al. 1998; Shetty et al. 1999). Tinamous, a paleognath taxon branching off within the paraphyletic ratites (Harshman et al. 2008), show signs of intermediate Z–W differentiation (Pigozzi and Solari 1999; Tsuda et al. 2007) parallel to that of neognaths (Mank and Ellegren 2007). Irrespective of the degree of W chromosomal degradation, some genes have been retained that are shared by Z and W chromosomes but are presently divergent because of the lack of genetic interchange. Such gene pairs on opposite sex chromosomes, termed "gametologs" (García-Moreno and Mindell 2000), have been assigned to three evolutionary strata on the Z chromosome of chicken by Nam and Ellegren (2008) based on divergence times of gametologous nucleotide sequences. In contrast to some mammals, where interchromosomal gene conversion (homogenization of diverged sequences) obscures the chronology of sex chromosome evolution (e.g., Pecon Slattery et al. 2000), the difference in GC content (on third codon positions) between avian Z and W gametologs suggests that this phenomenon has not played an important role in the evolution of avian sex chromosomes (Nam and Ellegren 2008).

Despite the importance of understanding the sex chromosome evolution through the divergences of gametologous gene pairs, data have been limited to nucleotide sequence analyses (e.g., Fridolfsson et al. 1998; García-Moreno and Mindell 2000; Ellegren and Carmichael 2001; Nam and Ellegren 2008), which are prone to homoplasious results as known, for example, from reconstructions of avian species phylogeny using nuclear versus mitochondrial DNA (e.g., Hackett et al. 2008 vs. Pratt et al. 2009). To examine avian sex chromosome evolution from an independent and novel perspective, we analyzed presence/absence patterns of retroposon insertions and random insertions/deletions (indels) to reconstruct gene trees of gametologs. Retroposons, mobile elements that propagate and integrate almost randomly in genomes via RNA intermediates, leave behind complex, virtually homoplasy-free phylogenetic signals of common ancestry (Shedlock et al. 2004; Ray et al. 2006). Thus, a resultant maximum parsimony reconstruction is effectively a maximum likelihood estimation (Steel and Penny 2000). In birds, this approach has been successfully used to unambiguously reconstruct the species phylogenies of penguins (Watanabe et al. 2006) and gamebirds (Kaiser et al. 2007; Kriegs et al. 2007).

In the case of gametologous gene pairs (fig. 1), we assumed that a retroposon insertion could be either present

Letter

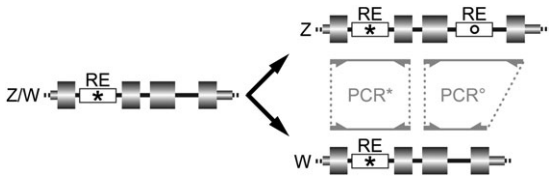

FIG. 1. Evolutionary scenario of a gametologous gene pair indicated by insertions of retroposed elements (RE). Large gray boxes are exons, small gray boxes are untranslated regions, and white boxes are retroposon insertions. Subsequent to frequent recombination (left), cessation of recombination leads to divergence (right) of the two gametologs. Asterisks denote REs that inserted prior to differentiation; REs marked with a circle inserted after cessation of recombination. Short arrows indicate primer positions for simultaneous PCR amplification of both gametologs of a given intron.

in both gametologs of a given genome (Z-/W-presence) and absent in the outgroup's gametolog pair or present in only one of the gametologs and absent in the other (e.g., Z-presence/W-absence). To find such retroposon insertions, for example, LTRs (long terminal repeat elements of endogenous retroviruses) or CR1s (chicken repeat 1 family of long interspersed elements), all 12 gametologous genes known from chicken (Nam and Ellegren 2008), comprising 126 sequenced intron pairs, were computationally screened for cases of either Z-/W-presence or Z-presence/W-absence feasible for polymerase chain reaction (PCR) (for more details, see supplementary fig. S1, Supplementary Material online). We experimentally analyzed and successfully amplified four phylogenetically informative gametologous loci in a taxon sampling that spans the breadth of avian taxa sensu Hackett et al. (2008). All sequences of each locus were aligned and screened for rare genomic changes, such as retroposon insertions (supplementary fig. S2, Supplementary Material online) and random indels (for details and full sequence alignments, see supplementary Methods, Supplementary Material online). Rare genomic changes were interpreted considering maximum parsimony; retroposon-flanking sequences were subjected to maximum likelihood sequence analyses (supplementary fig. S3, Supplementary Material online).

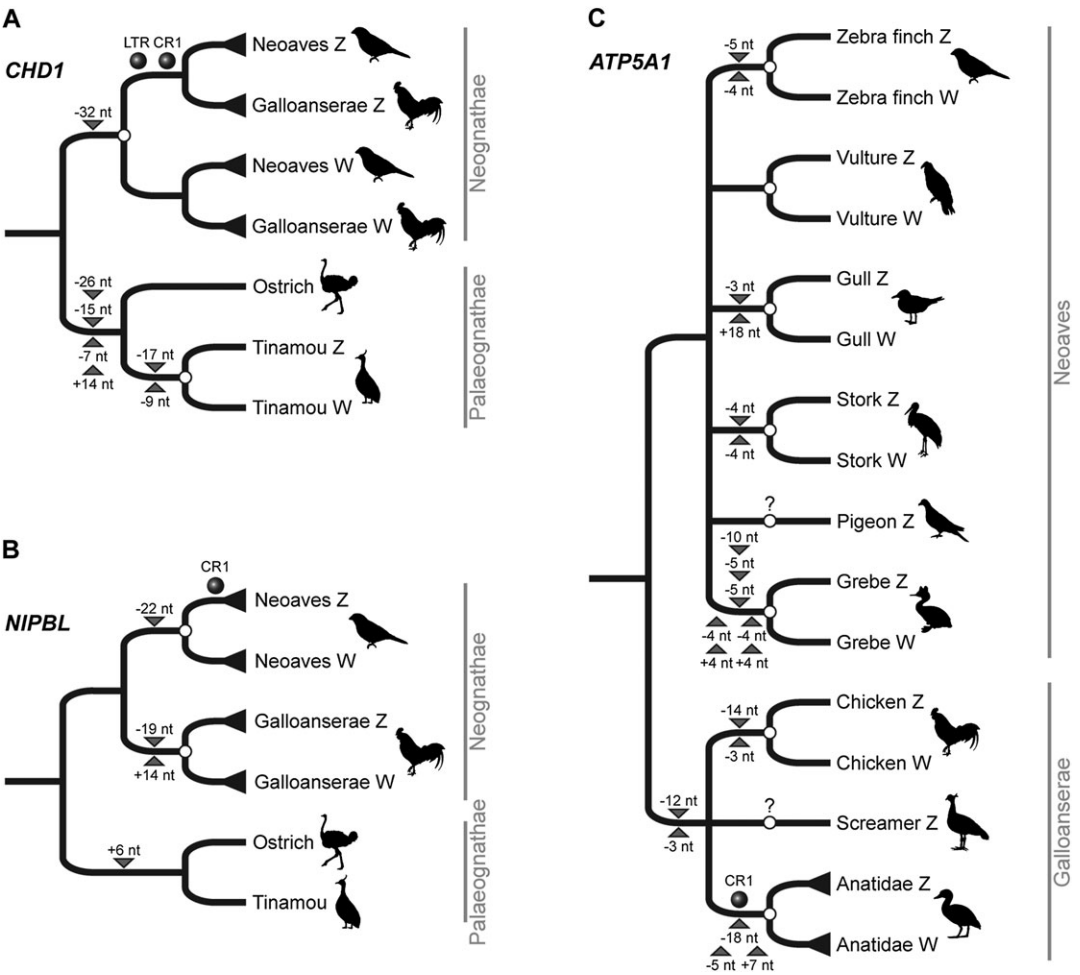

FIG. 2. Rare genomic changes and the simplified gene trees of (*A*) *CHD1* introns 9 plus 16, (*B*) *NIPBL* intron 16, and (*C*) *ATP5A1* intron 3. Tree topologies (branches not to scale) correspond to the maximum parsimony–based interpretation of rare genomic changes and maximum likelihood sequence analyses (for branch lengths, see supplementary fig. S3, Supplementary Material online). White circles pinpoint cessation of gametolog recombination. Gray balls indicate retroposon insertions of CR1 elements (CR1-Y2_Aves in (*A*), CR1-J2_Pass in (*B*), and CR1-X2 in (*C*) and LTRs (TguLTR5e in (*A*)). Random indels are depicted by triangles with the numbers of inserted/deleted nucleotides (nt). The respective outgroup taxa for each gene tree are not shown.

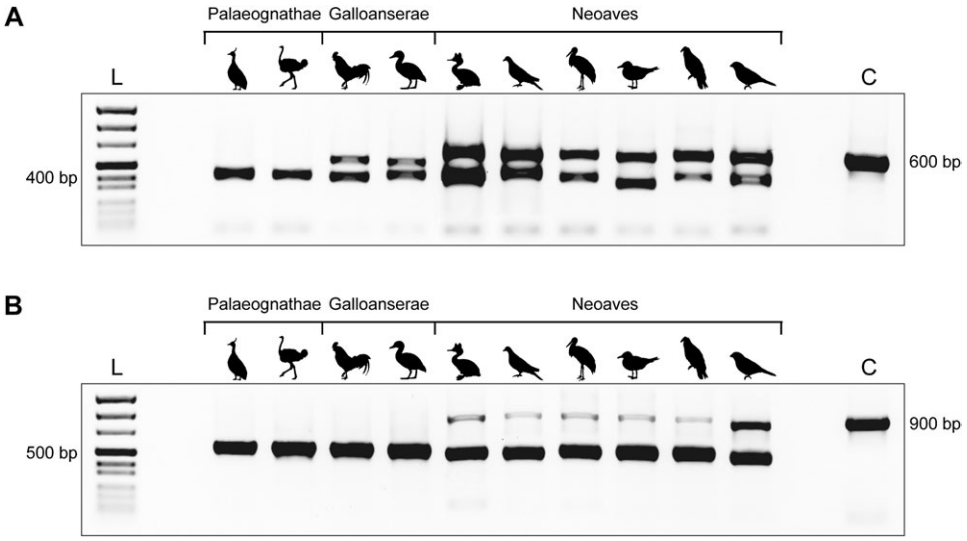

FIG. 3. PCR products of (*A*) *CHD1* intron 16 and (*B*) *NIPBL* intron 16 in female birds of the breadth of avian taxa. Each 1% agarose gel photo includes a pUC8 size marker (**L**) and a male zebra finch Z amplicon for control (*C*). In cases of two visible bands, the lower band represents the W amplicon and the upper band the larger Z amplicon containing a retroposon insertion. In earlier branching species (i.e., those lacking the Z-gametologous retroposon insertion), only one band is visible as the Z and W amplicons show no distinct size differences (for species names, see supplementary Methods, Supplementary Material online).

The gene tree of the chromodomain helicase DNA–binding protein 1 (*CHD1* introns 9 and 16, fig. 2*A*) corroborates an independent differentiation of the gene pair in tinamous and in neognaths as suggested by Tsuda et al. (2007). Several random indels (e.g., two deletions shared by both tinamid gametologs and a 32-nt deletion present in all neognathous Z and W gametologs) support this. Furthermore, an LTR retroposon (TguLTR5e) and a CR1 retroposon (CR1-Y2_Aves) are present in the neognathous *CHD1Z* introns 9 and 16, respectively. As these retroposons are absent in the neognathous W gametolog, they must have inserted subsequent to the divergence of the Z and W gametologs in that lineage (fig. 1).

A different tree topology was obtained for the *Drosophila* Nipped-B homolog (*NIPBL* intron 16, fig. 2*B*), where recombination ceased independently in the neoavian and in the galloanseran lineage. A 22-nt deletion in the Z and W gametologs of Neoaves and a 19-nt deletion plus a 14-nt insertion in both gametologs of Galloanserae corroborate this finding. Subsequent to the divergence of the neoavian *NIPBL* gametologs, a CR1 element (CR1-J2_Pass) inserted into the Z gametolog of the common ancestor of Neoaves because the W gametolog exhibits the ancestral absence state. No phylogenetic resolution was ascertained among the galloanseran gametologs, so it remains unclear whether the differentiation occurred already in the ancestor of Galloanserae.

Even more complex than noted by Ellegren and Carmichael (2001) is the gene tree of the ATP synthase α-subunit isoform 1 (*ATP5A1* intron 3, fig. 2*C*) because recombination seems to have ceased independently in all six neoavian representatives (we did not obtain the pigeon W gametolog, though). More samples are necessary to assess when and how often *ATP5A1* differentiated in Neoaves. Within Galloanserae, the Z and W gametologs differentiated independently in the galliform chicken and in duck-related birds (Anatidae). In the anseriform screamer, we could obtain only the Z-gametologous sequence. As evidenced by a CR1 insertion (CR1-X2) plus three random indels present in *ATP5A1Z* and *ATP5A1W*, genetic interchange persisted for some time between both anatid gametologs, before they diverged in the anatid ancestor (indicated by the sequence analysis, supplementary fig. S3*C*, Supplementary Material online).

This study demonstrates the usefulness of retroposon insertions not only as phylogenetic markers but also as temporal landmarks of gametolog differentiation, permitting an independent evaluation of known gametolog divergence chronologies (Nam and Ellegren 2008). The distribution of a retroposon among extant species enables us to trace its insertion back to the sex chromosomes of a common ancestor. Thus, the time of existence of that ancestral species can be inferred by consideration of dated phylogenies based on autosomal or mitochondrial sequences (e.g., Brown et al. 2008 and Ericson et al. 2006, reviewed by Pereira and Baker 2009 and van Tuinen 2009a, 2009b), circumventing the problems (e.g., stochastic errors, substitution rate heterogeneities, and nucleotide composition biases) inherent in relying only on one limited source of data for molecular dating (i.e., here, the Z–W pairwise distance of the gametologs' nucleotide sequences).

The identified retroposon insertions, in combination with other rare genomic changes, provide strong evidence that the *CHD1Z*/*CHD1W* genes of the extant neognaths were already differentiated in their common ancestor. As the ancestor of Neognathae lived 119–105 Ma (van Tuinen 2009a), this, together with the estimated Z–W divergence of 132 Ma (Nam and Ellegren 2008), corroborates

the inclusion of this gene pair in the oldest evolutionary stratum 1 of the Z chromosome. Furthermore, retroposon evidence unambiguously indicates that the *NIPBLZ*/*NIPBLW* genes diverged in the neoavian ancestor. This ancestral species lived 105–97.3 Ma (van Tuinen 2009a, 2009b) and thus, this gene pair can be included in stratum 2 of the neoavian Z chromosome. Within Galloanserae, the timing of *NIPBLZ*/*NIPBLW* differentiation could not be elucidated via retroposons or random indels, but Nam and Ellegren (2008) calculated a Z–W divergence of 52 Ma and included this gene pair in stratum 3. The *ATP5A1Z*/*ATP5A1W* genes of chicken also belong to the youngest evolutionary stratum 3 as their Z–W divergence is 53 Ma (Nam and Ellegren 2008). On the anseriform branch, our data suggest that Z–W recombination of this gene pair ceased at least 10 My earlier, namely in the anatid ancestor (who lived 97.9–63.2 Ma, Pereira and Baker 2009; this corresponds to the evolutionary strata 2 and 3).

PCR amplifications of *CHD1* intron 16 in neognaths (Fridolfsson and Ellegren 1999) and *NIPBL* intron 16 in neoavians yielded two distinct amplicons in females (fig. 3) and only the respective larger amplicon in males. The same applies for *CHD1* intron 9 in neognaths (data not shown). Consequently, to the same degree as shared retroposon insertions strongly indicate common ancestry in a gametologous gene tree, retroposon insertions in one of the two gametologs are ideal tools for molecular sexing (Hedges et al. 2003). Due to the distinct size differences in Z and W amplicons (200 bp in *CHD1* intron 16, 400 bp in *NIPBL* intron 16, and 500 bp in *CHD1* intron 9), these markers are not prone to misinterpretation compared with previous markers based on random indels (e.g., 50 bp in the marker of Griffiths et al. (1998)). Thus, these three loci should enable ornithologists to conveniently determine the molecular gender of Neoaves or Neognathae (comprising as much as 95–99% of all bird species) using three universal and independent tests (for a standard operation procedure, see supplementary fig. S1, Supplementary Material online). A patent application was filed which was assigned the number EP 11 152 645.5.

In summary, we have demonstrated that gametologous retroposon insertions are ideal markers, both for a clear-cut examination of the relative chronology of avian sex chromosome differentiation and for molecular gender identification in birds. We are positive that this retroposon-based approach is also applicable to other sex chromosomal systems.

Supplementary Material

Supplementary figures S1, S2, and S3 and Methods are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Elisabeth Suh, Sandra Silinski (Zoo Münster), Holger Schielzeth, Gerald Mayr, the LWL-DNA- und Gewebearchiv, Harald Jockusch, Ommo Hüppop, Anne Hoppmann (Weltvogelpark Walsrode), and Werner Beckmann for providing feather, blood, and tissue samples. Marsha Bundman helped with editing; Yoko Satta and two anonymous reviewers provided valuable comments. This research was funded by the Deutsche Forschungsgemeinschaft (KR3639 to J.O.K. and J.S.).

References

Bellot DW, Skaletsky H, Pyntikova T, et al. (11 co-authors). 2010. Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature* 466:612–616.

Bergero R, Charlesworth D. 2008. The evolution of restricted recombination in sex chromosomes. *Trends Ecol Evol*. 24:94–102.

Brown JW, Rest JS, García-Moreno J, Sorenson MD, Mindell DP. 2008. Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. *BMC Biol*. 6:6.

Ellegren H, Carmichael A. 2001. Multiple and independent cessation of recombination between avian sex chromosomes. *Genetics* 158:325–331.

Ericson PGP, Anderson CL, Britton T, et al. (10 co-authors). 2006. Diversification of Neoaves: integration of molecular sequence data and fossils. *Biol Lett*. 4:543–547.

Fridolfsson AK, Cheng H, Copeland NG, Jenkins NA, Liu HC, Raudsepp T, Woddage T, Chowdhary B, Halverson J, Ellegren H. 1998. Evolution of the avian sex chromosomes from an ancestral pair of autosomes. *Proc Natl Acad Sci U S A*. 95:8147–8151.

Fridolfsson AK, Ellegren H. 1999. A simple and universal method for molecular sexing of non-ratite birds. *J Avian Biol*. 30:116–121.

García-Moreno J, Mindell DP. 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 MC, Orr K, Dawson RJG. 1998. A DNA test to sex most birds. *Mol Ecol*. 7:1071–1075.

Hackett SJ, Kimball RT, Reddy S, et al. (18 co-authors). 2008. A phylogenomic study of birds reveals their evolutionary history. *Science* 320:1763–1768.

Harshman J, Braun EL, Braun MJ, et al. (18 co-authors). 2008. Phylogenomic evidence for multiple losses of flight in ratite birds. *Proc Natl Acad Sci U S A*. 105:13462–13467.

Hedges DJ, Walker JA, Callinan PA, Shewale JG, Sinha SK, Batzer MA. 2003. Mobile element-based assay for human gender determination. *Anal Biochem*. 312:77–79.

Kaiser VB, van Tuinen M, Ellegren H. 2007. Insertion events of CR1 retrotransposable elements elucidate the phylogenetic branching order in galliform birds. *Mol Biol Evol*. 24:338–347.

Kriegs JO, Matzke A, Churakov G, Kuritzin A, Mayr G, Brosius J, Schmitz J. 2007. Waves of genomic hitchhikers shed light on the evolution of gamebirds (Aves: Galliformes). *BMC Evol Biol*. 7:190.

Mank JE, Ellegren H. 2007. Parallel divergence and degradation of the avian W sex chromosome. *Trends Ecol Evol*. 22:389–391.

Nam K, Ellegren H. 2008. The chicken (*Gallus gallus*) Z chromosome contains at least three nonlinear evolutionary strata. *Genetics* 180:1131–1136.

Ogawa A, Murata K, Mizuno S. 1998. The location of Z- and W-linked marker genes and sequence on the homomorphic sex chromosomes of the Ostrich and the Emu. *Proc Natl Acad Sci U S A*. 95:4415–4418.

Pecon Slattery J, Sanner-Wachter L, O'Brien SJ. 2000. Novel gene conversion between X-Y homologues located in the non-recombining region of the Y chromosome in Felidae (Mammalia). *Proc Natl Acad Sci U S A*. 97:5307–5312.

Pereira SL, Baker AJ. 2009. Waterfowl and gamefowl (Galloanserae). In: Hedges SB, Kumar S, editors. In: The timetree of life. New York: Oxford University Press. p. 415–418.

Pigozzi MI, Solari AJ. 1999. The ZW pairs of two paleognath birds from two orders show transitional stages of sex chromosome differentiation. *Chromosome Res*. 7:541–551.

Pratt RC, Gibb GC, Morgan-Richards M, Phillips MJ, Hendy MD, Penny D. 2009. Toward resolving deep Neoaves phylogeny: data, signal enhancement, and priors. *Mol Biol Evol*. 26:313–326.

Ray DA, Xing J, Salem AH, Batzer MA. 2006. SINEs of a nearly perfect character. *Syst Biol*. 55:928–935.

Shedlock AM, Takahashi K, Okada N. 2004. SINEs of speciation: tracking lineages with retroposons. *Trends Ecol Evol*. 19:545–553.

Shetty S, Griffin DK, Graves JAM. 1999. Comparative painting reveals strong chromosome homology over 80 million years of bird evolution. *Chromosome Res*. 7:289–295.

Steel M, Penny D. 2000. Parsimony, likelihood, and the role of models in molecular phylogenetics. *Mol Biol Evol*. 17:839–850.

Tsuda Y, Nishida-Umehara C, Ishijima J, Yamada K, Matsuda Y. 2007. Comparison of the Z and W sex chromosomal architectures in elegant crested tinamou (*Eudromia elegans*) and ostrich (*Struthio camelus*) and the process of sex chromosome differentiation in palaeognathous birds. *Chromosoma* 116:159–173.

van Tuinen M. 2009a. Birds (Aves). In: Hedges SB, Kumar S, editors. In: The timetree of life. New York: Oxford University Press. p. 409–411.

van Tuinen M. 2009b. Advanced birds (Neoaves). In: Hedges SB, Kumar S, editors. The timetree of life. New York: Oxford University Press. p. 419–422.

Watanabe M, Nikaido M, Tsuda TT, Inoko H, Mindell DP, Murata K, Okada N. 2006. The rise and fall of the CR1 subfamily in the lineage leading to penguins. *Gene* 365:57–66.