A Universal Method for the Study of CR1 Retroposons in Nonmodel Bird Genomes

Alexander Suh,*[,1] Jan Ole Kriegs,[1,2] Stephen Donnellan,[3] 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

[3]Evolutionary Biology Unit, South Australian Museum, Adelaide, Australia

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

Associate editor: Claudia Kappen

Abstract

Presence/absence patterns of retroposon insertions at orthologous genomic loci constitute straightforward markers for phylogenetic or population genetic studies. In birds, the convenient identification and utility of these markers has so far been mainly restricted to the lineages leading to model birds (i.e., chicken and zebra finch). We present an easy-to-use, rapid, and cost-effective method for the experimental isolation of chicken repeat 1 (CR1) insertions from virtually any bird genome and potentially nonavian genomes. The application of our method to the little grebe genome yielded insertions belonging to new CR1 subfamilies that are scattered all across the phylogenetic tree of avian CR1s. Furthermore, presence/absence analysis of these insertions provides the first retroposon evidence grouping flamingos + grebes as Mirandornithes and several markers for all subsequent branching events within grebes (Podicipediformes). Five markers appear to be species-specific insertions, including the hitherto first evidence in birds for biallelic CR1 insertions that could be useful in future population genetic studies.

Key words: birds, CR1, genomic library, retroposon extraction, inverse PCR, biallelic insertion.

Insertions of retroposons, repetitive genomic elements that propagate via an RNA intermediate and integrate in genomes after reverse transcription, have proven to be powerful tools both for the reconstruction of gene trees (Suh, Kriegs, et al. 2011) as well as species trees (Suh, Paus, et al. 2011) of early bird evolution. In such a complex marker system, the number of possible character states (i.e., insertion sites in the genome) is extremely large and thus events leading to homoplasy (such as exact excisions or parallel insertions) are extremely rare (Shedlock et al. 2004; Ray et al. 2006; Han et al. 2011). Consequently, the identification of three or more congruent retroposon markers for a phylogenetic branch is statistically significant (likelihood ratio test after Waddell et al. 2001) and such a maximum parsimony tree topology converges to a maximum likelihood estimation (Steel and Penny 2000).

Despite the advantages of retroposed element (RE) insertions for phylogenetic reconstructions, as well as population genetics (reviewed by Ray 2007), large-scale phylogenetic studies among birds (Kaiser et al. 2007; Kriegs et al. 2007; Suh, Paus, et al. 2011) are so far limited to the lineages of model birds where complete genome sequences are available (Hillier et al. 2004; Warren et al. 2010), namely chicken and zebra finch. A few studies (Watanabe et al. 2006; Treplin and Tiedemann 2007) have investigated retroposons in nonmodel birds via methods such as hybridization-based library screening, but each yielded only a few phylogenetic markers. An alternative and efficient procedure based on magnetic bead capture was reported by St John and Quinn (2008a) but was not tested on the suitability for finding retroposon markers.

To overcome the lack of retroposon information in nonmodel birds and test if there are young retroposons that could be used for avian population genetics, we have established a rapid and easy-to-use polymerase chain reaction (PCR)-based protocol (fig. 1 and supplementary methods [Supplementary Material online]; adapted from a general procedure described by Wang and Kirkness 2005) that yields a genomic retroposon-enriched library of the most abundant (Hillier et al. 2004; Warren et al. 2010) group of avian retroposons, namely the CR1 family of long interspersed elements.

To examine the applicability of this method, we conducted a case study using the little grebe. From the CR1-enriched library established via the procedure outlined in figure 1, we sequenced 180 randomly selected clones of CR1-containing genomic fragments, masked all repetitive regions using CENSOR (http://www.girinst.org/censor/index.php) or RepeatMasker (http://www.repeatmasker.org) and subsequently BLAT screened (Kent 2002) the sequences for similarity to single-copy sequences within the chicken and zebra finch genomes (for more details, see fig. 2 and the corresponding figure legend). We aligned these loci and applied standard criteria for selecting marker candidates (see supplementary methods, Supplementary Material online and Suh, Paus, et al. 2011), identifying 38 retroposon candidate loci (see supplementary table S1, Supplementary Material online). Two of these loci (each exhibiting an orthologous CR1 insertion shared among little grebe and zebra finch) were not further analyzed, because early bird phylogeny of the lineage leading to the zebra finch has already been studied extensively by Suh, Paus, et al. (2011). For the

Letter

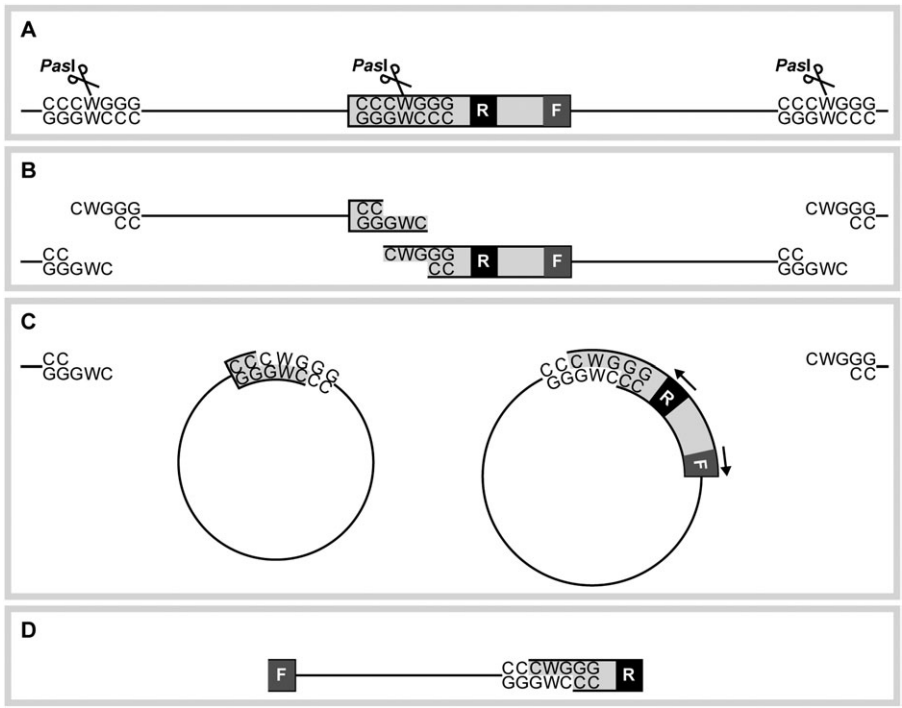

Fig. 1. Schematic construction of a PCR-based CR1-enriched library. The protocol (see supplementary methods, Supplementary Material online) was adapted from a population genetic study of short interspersed elements of dogs (Wang and Kirkness 2005). After preparation (*A*) of genomic DNA, DNA is digested (*B*) using the restriction enzyme *Pas*I, followed by thermal inactivation. Genomic fragments are diluted in water and circularized (*C*) using T4 DNA ligase. After DNA purification, inverse PCR (*D*) using outward-facing, CR1-specific primers yields CR1-containing fragments including one flank of each CR1 insertion. The resultant genomic library is then cloned and sequenced. The light gray box is a CR1 element, black lines are nonrepetitive sequences, dark gray boxes are binding sites for the F (forward) primer and black boxes are binding sites for the R (reverse) primer. Note that due to the ambiguous fourth position of the *Pas*I restriction site (W = A or T), only half of the sticky-ended genomic fragments are expected to be compatible for self-ligation in the circularization step (*C*).

remaining 36 candidate loci, PCR primers were generated (as described in fig. 2) and tested, leading to 15 markers (see supplementary tables S2 and S3 and supplementary material for full alignments, Supplementary Material online) that could be sequenced in different representatives of the grebe lineages (Podicipediformes), their potential sister group candidates and out group representatives (see supplementary methods, Supplementary Material online).

The enrichment method presented here yielded CR1 retroposons that inserted during many different parts of grebe evolution after the neoavian radiation (fig. 3). Two of our markers are shared among the three grebes and the flamingo but absent in the remaining sampled Neoaves (fig. 3, branch A, two REs, $P = 0.1111$, [2 0 0]). This corroborates the Mirandornithes hypothesis (Sangster 2005) established by sequence analyses (e.g., van Tuinen et al. 2001; Ericson et al. 2006; Hackett et al. 2008) and a few morphological characters (Mayr 2004). Considering this body of evidence, the traditional morphology-based grouping of grebes and loons (e.g., Mayr and Clarke 2003; Livezey and Zusi 2007) can be rejected. Furthermore, three markers are exclusively present in the sampled grebes; together with one marker (B-4) found via screening of GenBank (http://www.ncbi.nlm.nih.gov/genbank/) sequences, this is statistically significant evidence sensu Waddell et al. (2001) for the monophyly of Podicipediformes (fig. 3, branch B, four REs, $P = 0.0123$, [4 0 0]). Five markers unite the little grebe and the Australasian grebe to the exclusion of the crested grebe (fig. 3, branch C, five REs, $P = 0.0041$, [5 0 0]) in congruence with grebe taxonomy (Fjeldså 2004). Five retroposon insertions are present solely in the little grebe (fig. 3, branch D) and are, as the Australasian grebe (where these insertions are absent) belongs to the same genus *Tachybaptus*, either species- or population-specific RE insertions. We note that one of these markers appears to be dimorphic (i.e., the presence allele is not yet fixed within the little grebe population; supplementary fig. S1, Supplementary Material online). This is, to our knowledge, the first report of a biallelic RE insertion in birds and promises that CR1 retroposons might be useful for future population genetic studies.

In addition to reconstructing the species tree evolution in the lineage leading to grebes, the RE sequences themselves provide insights into the evolution of avian CR1 retroposons and their temporal impact on the little grebe genome (supplementary fig. S2, Supplementary Material online). The CR1 element (CR1-X3_Pod) of one of the Mirandornithes markers (marker A-1) is closely related to CR1-X3_Pass, a CR1 subtype that was only active prior to and during the neoavian radiation in the lineage leading to the zebra finch (Suh, Paus, et al. 2011). Thus, the master gene of this subtype appears to have had a prolonged

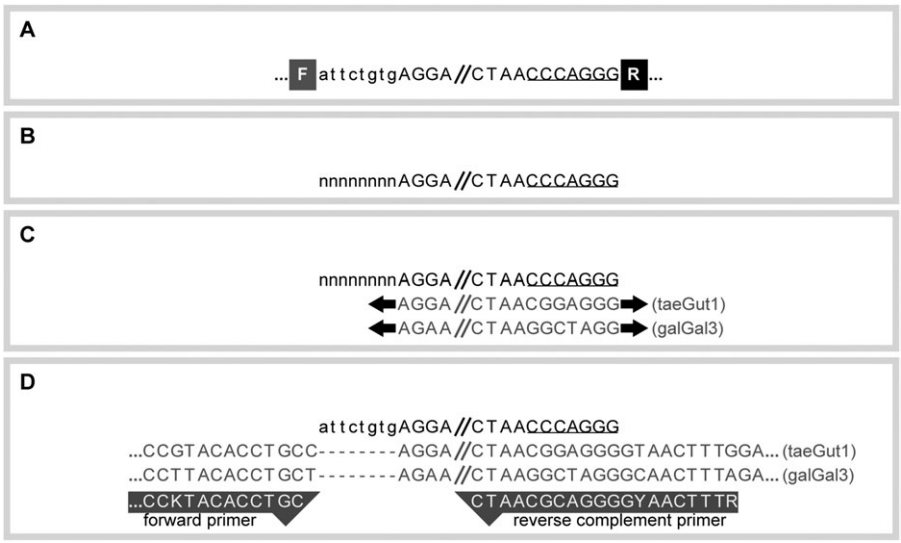

Fig. 2. Schematic identification of a CR1 insertion marker candidate. The construction of a PCR-based CR1-enriched library from a bird genome of interest (fig. 1 and supplementary methods, Supplementary Material online) and subsequent cloning yields sequences (*A*) including a CR1 retroposon 3′ end (lowercase letters) and *Pas*I restriction site (underlined) as well as primer (boxed F and boxed R) and vector sequences. After removal of primer and vector sequences (*B*), the remaining sequence is masked (i.e., nucleotides of CR1 and other retroposed sequences are replaced by the letter "n") using RepeatMasker or CENSOR. If BLAT searches (Kent 2002) using the reference genomes of the chicken (galGal3) and the zebra finch (taeGut1) yield a match score >50, the sequence hit (*C*) is extended (black arrows) by 2,000 bp upstream and downstream, respectively. If a match is found only in one of the two reference genomes, the extended sequence hit from one reference genome is BLAT screened against the other. The resultant sequences and the initial unmasked sequence from the bird genome of interest (after masking all CR1 and other retroposons using lowercase letters) are aligned (*D*) using MAFFT (E-INS-i, version 6, http://mafft.cbrc.jp/alignment/server/index.html; Katoh and Toh 2008). This permits the generation of one primer that specifically binds to a region that is at least partially well conserved between the bird of interest and the two reference genomes (here, the reverse complement primer binding to the right flank). The appropriate second primer is inferred on the basis of strong sequence conservation between the two reference genomes (here, the forward primer binding to the left flank), as there is no prior sequence information on this CR1-flanking sequence from the bird of interest. In rather rare cases, no well-conserved region on that flank is located within a <1,000 bp distance to the primer on the other flank. In some other cases, the above mentioned BLAT searches yield flanking sequences from only one of the two reference genomes, hampering the identification of well-conserved regions. For such marker candidates, two or more alternative primers (here, forward primers) should be generated to maximize the possibility of obtaining positive PCR results. The subsequent presence/absence analysis sampling the bird of interest as well as more or less closely related species is conducted following standard methods and strict criteria for character interpretation (Suh, Paus, et al. 2011).

activity in the ancestor of Mirandornithes in comparison to the zebra finch. Marker A-2 belongs to a chicken CR1-D–related subtype (CR1-D_Pod) that was not only active in the ancestor of Mirandornithes but also before the diversification of Podicipediformes (markers B-2 and B-3). In the podicipediform ancestor, a CR1-Y–related subtype (CR1-Yb_Pod) was active (markers B-1 and B-4). Additionally, our data suggest that two other CR1 subtypes were probably exclusively active in the two youngest branches of the phylogenetic tree, namely CR1-B_Pod (markers C-1 to C-5) in the ancestor of the genus *Tachybaptus* and CR1-Ya_Pod (markers D-1 to D-5) at the species (and population) level of the little grebe.

We are aware that the amount of CR1 retroposon markers and CR1 subtype diversity investigated here is just a fraction of the CR1 landscape of the whole little grebe genome. Nevertheless, the CR1 retroposon enrichment reported here identified five different CR1 subfamilies (four of them not present in zebra finch) that are scattered across the phylogenetic tree of CR1 retroposons. Our presence/absence analysis of these CR1 retroposon insertions suggests that they inserted throughout the last 53 million years (van Tuinen 2009) of grebe evolution since the last common ancestor of the Mirandornithes. Thus, we propose that our protocol is suitable for studying the breadth of avian CR1 diversity without the need for full-genome sequencing and, at the same time, for identifying retroposon insertions located across the avian tree of life. Furthermore, we suggest that this method is suitable for the detection of not only ancient but also very recent activity of CR1 retroposons (at least as recent as the anseriform CR1 activity reported by St John and Quinn 2008b), as we have identified, to our knowledge, the first case of intraspecific CR1 retroposition activity in a bird genome. Such lineage-specific retroposon activity patterns have been previously reported in other animals (e.g., in *Drosophila*, Sánchez-Gracia et al. 2005).

As we have successfully tested this protocol on the genomes of a parrot, a falcon, and an emu (data not shown), we predict that PCR-based CR1 enrichment will be an easy-to-use and cost-effective alternative to hybridization-based techniques, as it is applicable virtually to all birds and yields a wealth of retroposon markers suitable for studies of avian phylogeny, taxonomy, genome

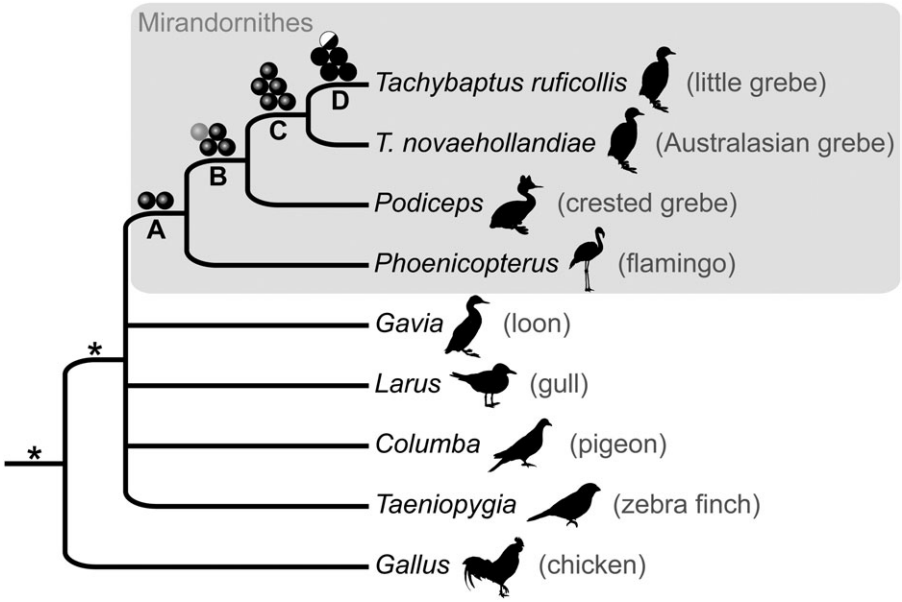

Fig. 3. Retroposon-based phylogenetic tree of Mirandornithes (flamingos and grebes). The tree topology was constructed on the basis of our presence/absence matrix (supplementary table S2, Supplementary Material online) and considering maximum parsimony. Gray balls are phylogenetically informative markers, black circles are autapomorphic retroposon insertions, and the semifilled circle is a dimorphic retroposon insertion (i.e., both the presence and the absence allele appear to be present in the little grebe population; see supplementary figure S1, Supplementary Material online). The light gray marker on branch B was found via screening of GenBank sequences. Branches that previously received strong retroposon support (Suh, Paus, et al. 2011) are highlighted with asterisks.

evolution, and perhaps population genetics. Presumably, the method of the present study will even be suitable for the isolation of CR1s and other retroposons from genomes of nonavian organisms.

Supplementary Material

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

Acknowledgments

We thank Anja Zemann, Astrid Farwick, Carsten Raabe, and Maria Nilsson for methodological discussions. Many thanks go to Werner Beckmann (LWL-DNA- und Gewebearchiv), Frank Grützner, Ommo Hüppop (Vogelwarte Helgoland), Holger Schielzeth, Sandra Silinski (Allwetterzoo Münster), and Leanne Wheaton for providing feather, blood, and tissue samples. Claudia Kappen and four anonymous reviewers provided valuable comments. This research was funded by the Deutsche Forschungsgemeinschaft (KR3639 to J.O.K. and J.S.).

References

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.

Fjeldså J. 2004. The grebes. New York: Oxford University Press.

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.

Han KL, Braun EL, Kimball RT, et al. (17 co-authors). 2011. Are transposable element insertions homoplasy free? An examination using the avian tree of life. *Syst Biol.* 60:375–386.

Hillier LW, Miller W, Birney E, et al. (174 co-authors). 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695–716.

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.

Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform.* 9:286–298.

Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Res.* 12:656–664.

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.

Livezey BC, Zusi RL. 2007. Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. *Zool J Linn Soc.* 149:1–95.

Mayr G. 2004. Morphological evidence for sister group relationship between flamingos (Aves: Phoenicopteridae) and grebes (Podicipedidae). *Zool J Linn Soc.* 140:157–169.

Mayr G, Clarke J. 2003. The deep divergences of neornithine birds: a phylogenetic analysis of morphological characters. *Cladistics* 19:527–553.

Ray DA. 2007. SINEs of progress: mobile element applications to molecular ecology. *Mol Ecol.* 16:19–33.

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

Sánchez-Gracia A, Maside X, Charlesworth B. 2005. High rate of horizontal transfer of transposable elements in Drosophila. *Trends Genet.* 21:200–203.

Sangster G. 2005. A name for the flamingo-grebe clade. *Ibis* 147:612–615.

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

St John J, Quinn TW. 2008a. Rapid capture of DNA targets. *Biotechniques* 44:259–264.

St John J, Quinn TW. 2008b. Identification of novel CR1 subfamilies in an avian order with recently active elements. *Mol Phylogenet Evol.* 49:1008–1014.

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

Suh A, Kriegs JO, Brosius J, Schmitz J. 2011. Retroposon insertions and the chronology of avian sex chromosome evolution. *Mol Biol Evol.* 28:2993–2997.

Suh A, Paus M, Kiefmann M, Churakov G, Franke FA, Brosius J, Kriegs JO, Schmitz J. 2011. Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. *Nat Commun.* 2:443.

Treplin S, Tiedemann R. 2007. Specific chicken repeat 1 (CR1) retrotransposon insertion suggests phylogenetic affinity of rockfowls (genus *Picathartes*) to crows and ravens (Corvidae). *Mol Phylogenet Evol.* 43:328–337.

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

van Tuinen M, Butvill DB, Kirsch JAW, Hedges SB. 2001. Convergence and divergence in the evolution of aquatic birds. *Proc R Soc Lond B Biol Sci.* 268:1345–1350.

Waddell PJ, Kishino H, Ota R. 2001. A phylogenetic foundation for comparative mammalian genomics. *Genome Inform.* 12:141–154.

Wang W, Kirkness EF. 2005. Short interspersed elements (SINEs) are a major source of canine genomic diversity. *Genome Res.* 15:1798–1808.

Warren WC, Clayton DF, Ellegren H, et al. (81 co-authors). 2010. The genome of a songbird. *Nature* 464:757–762.

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.