J Mol Evol (2015) 81:18–20
DOI 10.1007/s00239-015-9692-x

LETTER TO THE EDITOR

# The Specific Requirements for CR1 Retrotransposition Explain the Scarcity of Retrogenes in Birds

**Alexander Suh**[1]

Received: 19 July 2015 / Accepted: 20 July 2015 / Published online: 30 July 2015


**Abstract** Chicken repeat 1 (CR1) retroposons are the most abundant superfamily of transposable elements in the genomes of birds, crocodilians, and turtles. However, CR1 mobilization remains poorly understood. In this article, I document that the diverse CR1 lineages of land vertebrates share a highly conserved hairpin structure and an octamer microsatellite motif at their very 3′ ends. Together with the presence of these same motifs in the tails of CR1-mobilized short interspersed elements, this suggests that the minimum requirement for CR1 transcript recognition and retrotransposition is a complex >50-nt structure. Such a highly specific recognition sequence readily explains why CR1-dominated genomes generally contain very few retrogenes. Conversely, the mammalian richness in retrogenes results from CR1 extinction in their early evolution and subsequent establishment of L1 dominance.

**Keywords** Chicken repeat 1 · Long interspersed element · Retrotransposition · Microsatellite · Retrogene · Land vertebrates

Chicken repeat 1 (CR1) long interspersed elements (LINEs) are present in virtually all amniotes. They dominate the landscape of transposable elements (TEs) in the genomes of all major lineages except mammals (Shedlock et al. 2007; Suh et al. 2015a) and comprise a large diversity that existed since the days of the amniote ancestor (Suh et al. 2015a). Notably, CR1 elements are the only TEs that were active throughout avian evolution and have thus been widely used as phylogenetic markers [e.g., (Baker et al. 2014; Haddrath and Baker 2012; Kaiser et al. 2007; Kriegs et al. 2007; Liu et al. 2012; Suh et al. 2011a, b, 2012, 2015a)]. A few of these presence/absence analyses also revealed that, contrary to earlier assumptions (Hillier et al. 2004; Kapitonov and Jurka 2003), CR1 integration leads to target site duplication (Silva and Burch 1989; Suh et al. 2012, 2015a, b; Wicker et al. 2005) of usually 4 nt (Suh et al. 2012) and up to 9 nt (Suh et al. 2015b). However, despite the expectation that CR1 elements employ target primed reverse transcription (TPRT) as is the case in other non-long terminal repeat (non-LTR) retroelements (Luan et al. 1993), the minimum prerequisites for CR1 transcript recognition and TPRT initiation have not been investigated across the diversity of CR1 from land vertebrates.

In this article, I analyzed the 3′ untranslated regions (UTRs) of 119 CR1 subfamilies from amniote and amphibian genomes that were recently classified into seven ancient groups with differential retention across amniotes (Fig. 1a) (Suh et al. 2015a). Note that only CR1 elements from these organisms, but not other vertebrates, can be confidently aligned with each other on the nucleotide level (Suh et al. 2015a). All amniote CR1 subfamilies exhibit an 8-nt microsatellite motif at their very 3′ ends, which is relatively conserved within each of the seven CR1 groups (Fig. 1b). The only deviation is the avian subfamily CR1-X3_Pass with a 5-nt microsatellite motif that appears to be a shortened version of the 5′-ATTCTRTG-3′ motif of other CR1 in birds (Fig. 1b, Table S1) (Hillier et al. 2004). The presence of a similar octamer motif in the amphibian outgroup suggests that an 8-nt microsatellite (5′-

**Electronic supplementary material** The online version of this article (doi:10.1007/s00239-015-9692-x) contains supplementary material, which is available to authorized users.

✉ Alexander Suh
alexander.suh@ebc.uu.se

[1] Department of Evolutionary Biology, Evolutionary Biology Centre (EBC), Uppsala University, Norbyvägen 18D, 752 36 Uppsala, Sweden

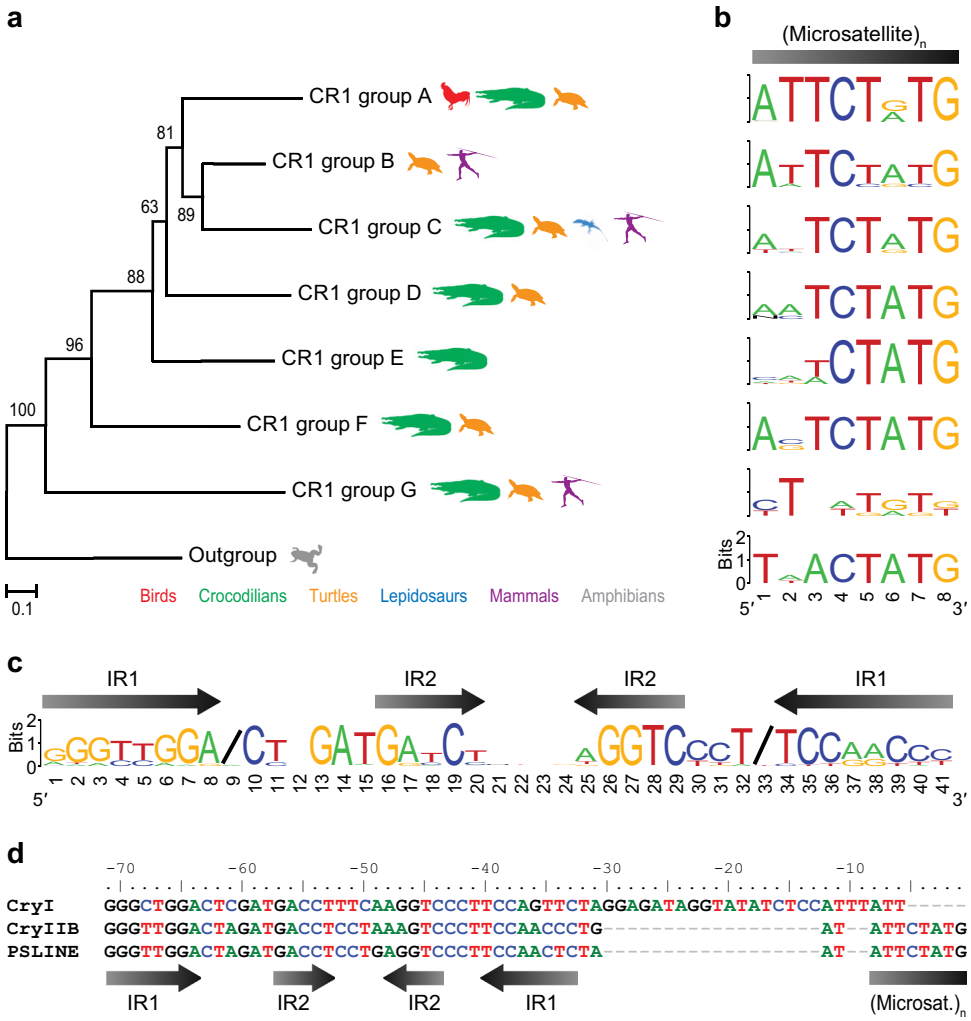


**Fig. 1** The 3′ ends of CR1 retroposons are highly conserved across land vertebrates. **a** Simplified phylogram of the seven ancient amniote CR1 groups and amphibian CR1 (Suh et al. 2015a). Maximum likelihood bootstrap values are shown in percent. **b** 8-nt microsatellite motifs derived from each of the CR1 groups of panel **a**. **c** Hairpin structure with inverted repeat motifs (IR1 and IR2) derived from all 119 CR1 subfamilies. Forward slashes denote alignment gaps. **d** CryI and CryIIB are CR1-mobilized SINEs from turtles (Sasaki et al. 2004), and exhibit tails with hairpin and microsatellite structures highly similar to those of CR1 (cf. panels **b** and **c**). Nucleotide residues are numbered backward from the reverse transcription start site. Sequence logos of panels **b** and **c** were generated in WebLogo (Crooks et al. 2004) using the data shown in Table S1, with blank nucleotide positions indicating poorly conserved residues

NNNCTATG-3′) is common to nearly all CR1 subfamilies in land vertebrates.

Directly upstream of this microsatellite motif and a short spacer (1–16 nt, Table S1) is a short, inverted repeat region that was previously proposed to be a recognition site for the *cis*-encoded reverse transcriptase (Haas et al. 2001; Kajikawa et al. 1997; Suh et al. 2011b). My motif analysis of this region across all the known CR1 subfamilies of the present dataset further supports this hypothesis. Most of the highly conserved residues belong to inverted repeat motifs (IR1 and IR2) that span 8 and 5 nt, respectively (Fig. 1c), and are interspersed by small insertions or deletions in less conserved parts (Table S1). Consequently, this ~41-nt region forms a stable hairpin in structural predictions of the RNA transcript (Fig. S1) in mfold (Zuker 2003). The degrees of conservation of both this hairpin structure and the aforementioned microsatellite motif across the studied diversity of CR1 lineages strongly imply their involvement in the mobilization of CR1 LINEs.

Co-mobilized short interspersed elements (SINEs) constitute an additional line of evidence for minimum requirements of retrotransposition, given that they 'mimic' LINEs by having similar or LINE-derived tails (Ohshima et al. 1996). For example, *Alu* SINEs are the most abundant TE in the human genome and exhibit A-rich 3′ ends that closely resemble those of L1 LINEs (Deininger and Batzer 2002). CR1-mobilized SINEs have been intensely studied in turtles (Sasaki et al. 2004), and their SINE tails are nearly identical to 3′ ends of CR1 (Fig. 1d). The tails include the hairpin motifs IR1 and IR2, as well as the 8-nt microsatellite motif (or a truncated 3-nt version thereof in CryI) at the very 3′ end (Fig. 1d, cf. Figure 1b, c). This further suggests that the aforementioned motifs are indeed the minimum sequence requirement for CR1 transcript recognition and TPRT initiation.

These observations have far-reaching implications. Retrogenes and retropseudogenes are remarkably scarce in birds (e.g., 51 cases in the chicken genome) compared to >15,000 cases in eutherian mammals (Hillier et al. 2004). One of the

most striking differences between avian and mammalian genomes is the virtual absence of L1 LINEs in birds and instead a dominance of CR1 LINEs (Hillier et al. 2004). Indeed, the similarity between the A-rich 3′ end of L1 and the polyadenylated 3′ ends of host mRNAs frequently leads to diversion of L1 retrotransposition and L1-mediated generation of retrogenes (Deininger and Batzer 2002; Esnault et al. 2000). In the case of CR1, the results presented here strongly suggest that their retrotransposition generally requires a highly specific hairpin and octamer structure at the transcript 3′ end, thereby largely precluding the recognition of host mRNAs by the enzymatic machinery of CR1.

Further details of CR1 mobilization may be unearthed in the near future. Although initial analyses of TE landscapes revealed only inactive CR1 subfamilies [e.g., (Hillier et al. 2004; Wicker et al. 2005)], there is now a growing body of evidence for the recent activity of CR1 elements in some birds, crocodilians, and turtles (St. John and Quinn 2008; Suh et al. 2012, 2015a). The most striking example is the gharial genome with ~2 Mb of identical CR1 copies [mostly CR1-7B_Gav, (Suh et al. 2015a)], which implies the possibility for resurrection of an intact CR1 master gene. The foreseeable study of CR1 activity in vitro promises to reveal the commonalities and differences between CR1 and L1 retrotransposition, and thereby their differential potential of creating retrogenes in avian and mammalian genomes.

**Acknowledgments** The author thanks Aurélie Kapusta and Regine Jahn for critical reading of this manuscript.

## References

Baker AJ, Haddrath O, McPherson JD, Cloutier A (2014) Genomic support for a moa-tinamou clade and adaptive morphological convergence in flightless ratites. Mol Biol Evol 31:1686–1696. doi:10.1093/molbev/msu153

Crooks GE, Hon G, Chandonia J-M, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14:1188–1190. doi:10.1101/gr.849004

Deininger PL, Batzer MA (2002) Mammalian retroelements. Genome Res 12:1455–1465. doi:10.1101/gr.282402

Esnault C, Maestre J, Heidmann T (2000) Human LINE retrotransposons generate processed pseudogenes. Nat Genet 24:363–367

Haas NB, Grabowski JM, North J, Moran JV, Kazazian HH Jr, Burch JBE (2001) Subfamilies of CR1 non-LTR retrotransposons have different 5′UTR sequences but are otherwise conserved. Gene 265:175–183. doi:10.1016/s0378-1119(01)00344-4

Haddrath O, Baker AJ (2012) Multiple nuclear genes and retroposons support vicariance and dispersal of the palaeognaths, and an Early Cretaceous origin of modern birds. Proc R Soc B 279:4617–4625

Hillier LW et al (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695–716. doi:10.1038/nature03154

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. doi:10.1093/molbev/msl164

Kajikawa M, Ohshima K, Okada N (1997) Determination of the entire sequence of turtle CR1: the first open reading frame of the turtle CR1 element encodes a protein with a novel zinc finger motif. Mol Biol Evol 14:1206–1217

Kapitonov VV, Jurka J (2003) The esterase and PHD domains in CR1-like non-LTR retrotransposons. Mol Biol Evol 20:38–46. doi:10.1093/molbev/msg011

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

Liu Z, He L, Yuan H, Yue B, Li J (2012) CR1 retroposons provide a new insight into the phylogeny of *Phasianidae* species (Aves: Galliformes). Gene 502:125–132. doi:10.1016/j.gene.2012.04.068

Luan DD, Korman MH, Jakubczak JL, Eickbush TH (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 72:595–605. doi:10.1016/0092-8674(93)90078-5

Ohshima K, Hamada M, Terai Y, Okada N (1996) The 3′ ends of tRNA-derived short interspersed repetitive elements are derived from the 3′ ends of long interspersed repetitive elements. Mol Cell Biol 16:3756–3764

Sasaki T, Takahashi K, Nikaido M, Miura S, Yasukawa Y, Okada N (2004) First application of the SINE (short interspersed repetitive element) method to infer phylogenetic relationships in reptiles: an example from the turtle superfamily Testudinoidea. Mol Biol Evol 21:705–715. doi:10.1093/molbev/msh069

Shedlock AM et al (2007) Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. Proc Natl Acad Sci USA 104:2767–2772. doi:10.1073/pnas.0606204104

Silva R, Burch JBE (1989) Evidence that chicken CR1 elements represent a novel family of retroposons. Mol Cell Biol 9:3563–3566. doi:10.1128/mcb.9.8.3563

St. John J, Quinn TW (2008) Recent CR1 non-LTR retrotransposon activity in coscoroba reveals an insertion site preference. BMC Genom 9:567

Suh A et al (2015a) Multiple lineages of ancient CR1 retroposons shaped the early genome evolution of amniotes. Genome Biol Evol 7:205–217. doi:10.1093/gbe/evu256

Suh A, Kriegs JO, Brosius J, Schmitz J (2011a) Retroposon insertions and the chronology of avian sex chromosome evolution. Mol Biol Evol 28:2993–2997. doi:10.1093/molbev/msr147

Suh A et al (2011b) Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. Nat Commun 2:443. doi:10.1038/ncomms1448

Suh A, Kriegs JO, Donnellan S, Brosius J, Schmitz J (2012) A universal method for the study of CR1 retroposons in nonmodel bird genomes. Mol Biol Evol 29:2899–2903. doi:10.1093/molbev/mss124

Suh A, Smeds L, Ellegren H (2015b) The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. PLoS Biol. doi:10.1371/journal.pbio.1002224

Wicker T et al (2005) The repetitive landscape of the chicken genome. Genome Res 15:126–136

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406–3415. doi:10.1093/nar/gkg595