# Retroposon Insertions Provide Insights into Deep Lagomorph Evolution

Jan Ole Kriegs,<sup>†</sup><sup>‡,1</sup> Anja Zemann,<sup>‡,1</sup> Gennady Churakov,<sup>1</sup> Andreas Matzke,<sup>1</sup> Martina Ohme,<sup>2</sup> Hans Zischler,<sup>3</sup> Jürgen Brosius,<sup>1</sup> Ute Kryger,<sup>4</sup> and Jürgen Schmitz<sup>\*,1</sup>

<sup>1</sup>Institute of Experimental Pathology (ZMBE), University of Münster, Münster, Germany

<sup>2</sup>Primate Genetics, German Primate Center, Göttingen, Germany

<sup>3</sup>Institute of Anthropology, Johannes Gutenberg-Universität Mainz, Mainz, Germany

<sup>4</sup>Department of Zoology & Entomology, University of Pretoria, Pretoria, South Africa

†Present address: LWL-Museum für Naturkunde, Westfälisches Landesmuseum mit Planetarium, Münster, Germany.

‡These authors contributed equally to this work.

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

Associate editor: Norihiro Okada

## Abstract

The homogenous mammalian order Lagomorpha comprises about 80 species in two families, Ochotonidae (pikas) and Leporidae (rabbits and hares). However, the phylogenetic relationships among leporids are controversial. Molecular data, particularly from mitochondrial sequences, give highly homoplasious signals. To resolve the controversy between mitochondrial and nuclear data, we analyzed genomic orthologous retroposon insertion sites, a virtually homoplasy-free marker system. From a differential screen of rabbit genomic data for intronic retroposon insertions of CSINE elements, we polymerase chain reaction–amplified and sequenced 11 retroposons in eight representative lagomorphs. We found three retroposons shared among all lagomorphs but absent in outgroups, four confirmed the monophyly of leporids, and three significantly supported Pronolagus as the sister group to all other leporids. One retroposon supported the monophyly of Lepus. The position of Pronolagus outside of the remaining leporids supports the sequence-based signals of nuclear genes and clearly refutes the misleading signals of mitochondrial genes.

Key words: Lagomorpha, retroposon phylogeny, CSINE, TinT.

About 62-100 million years ago (Ma), lagomorphs (pica, rabbits, and hares) diverged from their shared common ancestor with rodents (Benton and Donoghue 2007). Although hotly debated in the middle of the twentieth century, the clade Glires, comprising lagomorphs and rodents, is clearly supported today by molecular data (Murphy et al. 2001; Douzery and Huchon 2004; Kriegs, Churakov, et al. 2007). Within Glires, lagomorphs represent the well-separated group of Duplicidentata that have a second small upper incisor located behind the primary incisors, in contrast to the single incisors of the Simplicendentata, such as rodents (Rose 2006). Lagomorphs consist of two families, Ochotonidae (pikas) and Leporidae (rabbits, hares) that are distributed over large parts of Africa, Eurasia, and America. Despite intensive investigations, the evolutionary history of lagomorphs, including the European rabbit, an important model for biomedical research, is far from settled. Phylogenetic reconstructions have been made using morphological characters (Dawson 1958; Hibbard 1963; Dawson 1981), molecular sequences of mitochondrial genes (Halanych and Robinson 1999), chromosome rearrangements (Robinson et al. 2002), and a combination of nuclear and mitochondrial data in a so-called supermatrix approach (Matthee et al. 2004). Especially, the comparison of nuclear and mitochondrial phylogenies of leporids led to confounding results. In a mitochondrial data

set of cytochrome *b* and 12S rRNA sequences, Matthee et al. (2004) failed to find any clear phylogenetic signals for grouping the leporids (their fig. 2b). Halanych and Robinson (1997) attributed such inconsistent homoplasious phylogenetic signals to a rapid radiation and mutational saturation of mitochondrial sequences. However, despite the high level of homoplasy in leporid mitochondrial sequences, combining information from mitochondrial genes with that of five nuclear DNA fragments (their fig. 2a) enabled (Matthee et al. 2004) to reconstruct a leporid tree corresponding to the topology derived from their nuclear data.

To obviate the problem of homoplasy in the deep splits of lagomorph phylogeny, we examined an entirely different set of characters. We conducted a genome-wide screening for phylogenetically informative rare genomic changes (RGCs) in rabbits (*Oryctolagus cuniculus*). RGCs, especially retroposed elements, leave behind distinctive unambiguous traces of relatedness recognizable as orthologous genomic insertions (Shedlock and Okada 2000). Such shared insertions were previously used to resolve, for example, the extremely intricate deep nodes in mammalian phylogeny (Kriegs et al. 2006; Nishihara et al. 2006; Churakov et al. 2009; Nishihara et al. 2009) and the highly complex comparable to lagomorphs rodent phylogeny (Farwick et al. 2006; Churakov et al. 2010). Random indels are also useful markers of phylogenomic relatedness, but, due to the lack

© The Author 2010. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org





Fig. 1. TinT analysis for different CSINEs in pika and rabbit. The lower parts of (A) (*Ochotona princeps*) and (B) (*Oryctolagus cuniculus*) represent the data matrices used for deriving the TinT profiles shown in the upper parts of the figures. The identifiers across the top of the matrix represent elements fragmented by those elements listed along the side. The copy numbers of CSINEs and the sums ( $\Sigma$ ) of nested insertions are indicated. The TinT profiles present the elements sorted by their specific chronological activities, with older elements to the left and younger ones to the right along the relative time lines shown below. Ovals represent the 75% probability of the activity distribution, vertical lines the 95%, and the horizontal lines the 99% probability. OP denotes potential *O. princeps*-specific variations. 2/3A and B are element variations. Note that the current CSINE classification should be updated, as CSINE2B\_OP and CSINE2C\_OP consensus sequences are probably also distributed in *O. cuniculus*. CSINE3\_OP and CSINE3A arose independently in pica and rabbit (see also supplementary fig. 1, Supplementary Material Online).

of complexity of their insertions (e.g., no flanking direct repeats), they are not as comparably informative and are much more exposed to homoplasy.

We focused our retroposon search on lagomorph-specific Short INterspersed Elements (SINEs), so called CSINEs that are subdivided by diagnostic mutations into three subtypes: CSINE1, CSINE2, and CSINE3 (Krane et al. 1991). We first used the Transpositions in Transpositions (TinT) tool (Kriegs, Matzke, et al. 2007; http://www.bioinformatics.unimuenster.de/tools/tint) to determine the correct chronological activity periods of CSINE subtypes for pica (*Ochotona princeps*) and rabbit (*O. cuniculus*). For full distribution patterns, we used the preliminary genomic assemblies available for pica at http://www.broadinstitute.org/ftp/pub/assemblies /mammals/pika/OchPri2.0 and for rabbit at http://www .broadinstitute.org/ftp/pub/assemblies/mammals/rabbit



Fig. 2. Phylogenetic tree of lagomorphs reconstructed from CSINE insertion patterns. Green dots depict phylogenetically informative CSINE2 insertions, purple dots represent CSINE1, and gray dots represent CSINE3 events. Triangles depict the presence of random indels with the numbers of inserted/deleted nucleotides. A corresponding presence/absence matrix and all alignments are presented as supplementary table S1 and supplementary material S1, Supplementary Material online. The images show Ochotona princeps, Oryctologus cuniculus, and Lepus capensis.

/oryCun2. These TinT results showed that the CSINE2 elements were the oldest, followed by the CSINE1, and then the CSINE3 elements (fig. 1). Thus, for our first genome-wide RepeatMasker screening (http://www .repeatmasker.org/RMDownload.html) of rabbit trace sequences (ftp://ftp.ncbi.nih.gov/pub/TraceDB/oryctolagus\_ cuniculus), we searched for CSINE2 elements, and then for CSINE1 and CSINE3 elements in a second screening. Extracted sequences were blasted against 185,000 downloaded human introns (http://genome.ucsc.edu/cgi-bin/ hgTables) to identify those intronic CSINE elements with conserved exonic/intronic flanks for primer design. This strategy enabled the most successful polymerase chain reaction (PCR) amplification and sequencing in all taxa analyzed [regions <800 nucleotide (nt)]. PCR amplification, cloning, and sequencing were performed using standard protocols (Farwick et al. 2006). We experimentally analyzed 74 potential phylogenetically informative loci; ten of them were amplifiable in a representative set of species and contained 11 phylogenetically informative retroposon insertions (one of the loci contains two diagnostic retroposon insertions) and additionally three diagnostic random indels greater than 3 nt (supplementary table S1, Supplementary Material online). The species sampling and all diagnostic insertions are shown in figure 2.

Three independent insertions of CSINE2s, which were clearly absent at orthologous positions in human or guinea pig, significantly (Waddell et al. 2001) support the monophyly of the order Lagomorpha. Four independent retroposon insertions ( $3 \times CSINE2s$ ,  $1 \times CSINE1$ ) and one 4-nt deletion that were absent in Ochotonidae, but present in all Leporidae, significantly support the monophyly of this family, confirming most previous investigations (Matthee et al. 2004). We identified three additional independent retroposon insertions ( $2 \times CSINE3s$ ,  $1 \times CSINE1$ ) and one 8-nt deletion that were absent in Pronolagus but present in all other leporids, providing significant evidence that Pronolagus is the sister group to the remaining leporids. In addition, one orthologous CSINE3 element and one 17-nt deletion were found in all Lepus species, but not in other Leporidae, indicating the monophyly of Lepus.

Although the monophyly of lagomorphs was already generally accepted and the distribution of CSINEs in all lagomorphs but absent in all other mammals is a confirmation of this, the three orthologous CSINE2 insertions demonstrate the activity of CSINE2 elements before the first lagomorph splits, approximately 45 Ma (Lopez-Martinez 2008). One of the more controversial divergences in lagomorphs was the earliest branching within Leporidae. Our data provide significant evidence that Pronolagus is the sister group to the remaining leporids. This supports the nuclear gene analyses of Matthee et al. (2004) and clearly refutes the misleading signals from mitochondrial genes, underlining their impracticality and unreliability for phylogenetic reconstructions in this taxonomic group (see also Slimen et al. 2008). The consistent phylogenetic signals of nuclear gene sequences (Matthee et al. 2004) and the retroposon data presented here indicate that previous misinterpretations were more likely due to deficient signals in mitochondrial sequences and not ancestral polymorphism in connection with rapid radiation. There are many other examples of errant mitochondrial phylogenies, one of the most impressive being the "significant support" for primate paraphyly mistakenly suggested by the clustering of colugo (*Cynocephalus variegatus*, Dermoptera) close to anthropoids. This was caused by a significant simultaneous change in the mitochondrial nt composition of these two clades (Schmitz et al. 2002).

Our investigation represents the first successful step in evaluating the deep splits within lagomorph phylogeny with the unambiguous, virtually homoplasy-free SINE marker system. Further investigations will move from the deep to more recent nodes to understand the detailed relationships between rabbits and hares. CSINE1 and CSINE3 elements should be well suited to finally resolve the last uncertainties in this rapidly diverged group. Especially important for further investigations will be to supplement screening for retroposon-free rabbit introns, with high-throughput experimental amplifications of all additional investigated species, to randomly find evidence of speciation events apart from the lineage leading to rabbit. This strategy should enable us to unbiasedly track all speciation events within the Lagomorpha. The resulting completely resolved homoplasy-free reconstruction of relationships within the entire order will be a major step forward in lagomorph research that will open the doors to meaningful management and conservation.

### **Supplementary Material**

Supplementary tables S1 and S2, supplementary figure 1, and supplementary material S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals .org/).

## Acknowledgments

We thank Marsha Bundman for editorial assistance and Per Kryger for organizing the sample transfer. We thank the German Primate Center (DPZ, Göttingen) for support in the initial phase of the project. The lagomorph paintings were provided by Jón Baldur Hlíðberg. Our research was funded by the Deutsche Forschungsgemeinschaft (SCHM1469/3-1).

#### References

- Benton MJ, Donoghue PC. 2007. Paleontological evidence to date the tree of life. *Mol Biol Evol*. 24:26–53.
- Churakov G, Kriegs JO, Baertsch R, Zemann A, Brosius J, Schmitz J. 2009. Mosaic retroposon insertion patterns in placental mammals. *Genome Res.* 19:868–875.
- Churakov G, Sadasivuni MK, Rosenbloom KR, Huchon D, Brosius J, Schmitz J. 2010. Rodent evolution: back to the root. *Mol Biol Evol.* 27:1315–1326.
- Dawson MR. 1958. Later tertiary Leporidae of North America. Univ Kansas Paleontol Contrib Vertebrata. 6:1-75.

- Dawson MR. 1981. Evolution of the modern lagomorphs. In: Mayers K, MacInnes CD, editors. Proceedings of the World Lagomorph Conference; 1979. Guelph, Ontario, Canada: University of Guelph Press. p. 1–8.
- Douzery EJ, Huchon D. 2004. Rabbits, if anything, are likely Glires. Mol Phylogenet Evol. 33:922-935.
- Farwick A, Jordan U, Fuellen G, Huchon D, Catzeflis F, Brosius J, Schmitz J. 2006. Automated scanning for phylogenetically informative transposed elements in rodents. *Syst Biol.* 55:936–948.
- Halanych KM, Robinson TJ. 1997. Phylogenetic relationships of cottontails (Sylvilagus, Lagomorpha): congruence of 12S rDNA and cytogenetic data. *Mol Phylogenet Evol.* 7:294–302.
- Halanych KM, Robinson TJ. 1999. Multiple substitutions affect the phylogenetic utility of cytochrome *b* and 12S rDNA data: examining a rapid radiation in leporid (Lagomorpha) evolution. *J Mol Evol.* 48:369–379.
- Hibbard C. 1963. The origin of the P3 pattern of Sylvilagus, Caprolagus, Oryctolagus, and Lepus. J Mammal. 44:1-15.
- Krane DE, Clark AG, Cheng JF, Hardison RC. 1991. Subfamily relationships and clustering of rabbit C repeats. *Mol Biol Evol.* 8:1–30.
- Kriegs JO, Churakov G, Jurka J, Brosius J, Schmitz J. 2007. Evolutionary history of 7SL RNA-derived SINEs in supraprimates. *Trends Genet.* 23:158–161.
- Kriegs JO, Churakov G, Kiefmann M, Jordan U, Brosius J, Schmitz J. 2006. Retroposed elements as archives for the evolutionary history of placental mammals. *PLoS Biol.* 4:e91.
- 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.
- Lopez-Martinez N. 2008. The lagomorph fossil record and the origin of the european rabbit. In: Alves PC, Ferrand N, Hackländer K, editors. Lagomorph biology. Berlin: Springer. p. 27–46.
- Matthee CA, van Vuuren BJ, Bell D, Robinson TJ. 2004. A molecular supermatrix of the rabbits and hares (Leporidae) allows for the identification of five intercontinental exchanges during the Miocene. *Syst Biol.* 53:433–447.
- Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, O'Brien SJ. 2001. Molecular phylogenetics and the origins of placental mammals. *Nature* 409:614–618.
- Nishihara H, Hasegawa M, Okada N. 2006. Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. *Proc Natl Acad Sci U S A*. 103: 9929–9934.
- Nishihara H, Maruyama S, Okada N. 2009. Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. *Proc Natl Acad Sci U S A*. 106:5235–5240.
- Robinson TJ, Yang F, Harrison WR. 2002. Chromosome painting refines the history of genome evolution in hares and rabbits (order Lagomorpha). *Cytogenet Genome Res.* 96:223–227.
- Rose KD. 2006. The beginning of the age of mammals. Baltimore (MD): The Johns Hopkins University Press.
- Schmitz J, Ohme M, Suryobroto B, Zischler H. 2002. The colugo (Cynocephalus variegatus, Dermoptera): the primates' gliding sister? Mol Biol Evol. 19:2308–2312.
- Shedlock AM, Okada N. 2000. SINE insertions: powerful tools for molecular systematics. *Bioessays* 22:148–160.
- Slimen HB, Suchentrunk F, Elgaaied ABA. 2008. On shortcomings of using mtDNA sequence divergence for the systematics of hares (genus Lepus): an example from cape hares. *Mamm Biol.* 73:25-32.
- Waddell PJ, Kishino H, Ota R. 2001. A phylogenetic foundation for comparative mammalian genomics. *Genome Inform.* 12:141–154.