Extensive Gene Conversion Drives the Concerted Evolution of Paralogous Copies of the *SRY* Gene in European Rabbits

Armando Geraldes,*†,[1,2,3] Teri Rambo,[4] Rod A. Wing,[4] Nuno Ferrand,[1,2] and Michael W. Nachman[3]

[1]CIBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto, Campus Agrário de Vairão, Vairão, Portugal

[2]Departamento de Biologia, Faculdade de Ciências da Universidade do Porto, Porto, Portugal

[3]Department of Ecology and Evolutionary Biology, University of Arizona

[4]Arizona Genomics Institute, Department of Plant Sciences, University of Arizona

†Present address: Department of Botany, University of British Columbia, Vancouver BC, Canada

***Corresponding author:** E-mail: geraldes@interchange.ubc.ca.

Associate editor: Hideki Innan

Abstract

The human Y chromosome consists of ampliconic genes, which are located in palindromes and undergo frequent gene conversion, and single-copy genes including the primary sex-determining locus, *SRY*. Here, we demonstrate that *SRY* is duplicated in a large palindrome in the European rabbit (*Oryctolagus cuniculus*). Furthermore, we show through comparative sequencing that orthologous palindrome arms have diverged 0.40% between rabbit subspecies over at least 2 My, but paralogous palindrome arms have remained nearly identical. This provides clear evidence of gene conversion on the rabbit Y chromosome. Together with previous observations in humans, these results suggest that gene conversion is a general feature of the evolution of the mammalian Y chromosome.

Key words: Y chromosome, *SRY*, gene conversion, *Oryctolagus cuniculus*.

The completion of the sequence of the euchromatic portion of the male-specific region of the human Y chromosome (MSY) yielded many novel insights into the structure and evolutionary dynamics of this chromosome (Rozen et al. 2003; Skaletsky et al. 2003). Skaletsky et al. (2003) suggested that the human MSY is composed of three sequence classes (ampliconic, X-degenerate, and X-transposed) and that these three classes have distinct properties. According to this view, ampliconic genes occur in multiple copies, are expressed mostly in the testes, and are located in palindrome arms that undergo extensive gene conversion. X-degenerate genes are typically single copy and are generally widely expressed. X-transposed genes derive from a recent transposition from the X to the Y in the human lineage. Other authors have emphasized that some Y-linked genes do not fit easily into these categories, suggesting that Y-linked genes form an evolutionary continuum of degradation from a proto-XY ancestor (Graves 2006).

The ampliconic genes on the MSY are noteworthy because they have been shown to undergo frequent gene conversion in humans (Rozen et al. 2003). Rozen et al. (2003) suggested that such gene conversion might prevent the decay of these genes, although they noted that this would require a directional bias favoring the restoration of the ancestral state.

The human *SRY* (sex-determining region on the Y) is unusual in that it is the only X-degenerate gene with expression mostly restricted to the testis (primordial gonad), a pattern of expression that is otherwise associated with ampliconic genes (Skaletsky et al. 2003). SRY triggers the development of the undifferentiated gonadal primordia into testis, and its presence or absence determines whether an individual is born as a male or a female (Gubbay et al. 1990; Sinclair et al. 1990; Koopman et al. 1991). Although in humans *SRY* is a single-copy gene, it is present in multiple copies in several rodents (Bianchi et al. 1993; Nagamine 1994; Lundrigan and Tucker 1997; Bullejos et al. 1999; Turner et al. 2007) and lagomorphs (Geraldes and Ferrand 2006; Putze et al. 2007). In the only case where *SRY* gene conversion has been explored, the nucleotide differences among copies suggest that extensive gene conversion is not occurring (Turner et al. 2007).

In order to investigate whether *SRY* in the European rabbit (*Oryctolagus cuniculus*) is located in a palindrome and thus potentially subject to gene conversion, we sequenced 92,657 bp of a bacterial artificial chromosome (BAC) clone previously shown to contain *SRY*. This clone was mapped to Yp12 by fluorescent in situ hybridization (Hayes et al. 2002) and then sequenced and assembled following standard protocols (The Rice Chromosome 10 Sequencing Consortium 2003) (GenBank accession numbers for all sequences in this manuscript: HM230423–HM230433). In this clone, we detected 23,086 bp of an inverted repeat (Y Palindrome Arms 1 and 2—YP1A1 and YP1A2, fig. 1*a*) separated by 18,911 bp of unique sequence (Y Palindrome 1 Spacer—YP1S, fig. 1*a*). The *SRY* open reading frame was located in the arms of a palindrome, and transcription was oriented outwards (fig. 1a). This organization is similar to the palindromic organization of ampliconic genes on the human Y chromosome. We found 11 single nucleotide polymorphisms (SNPs) between palindrome arms (99.94% homology)

Letter

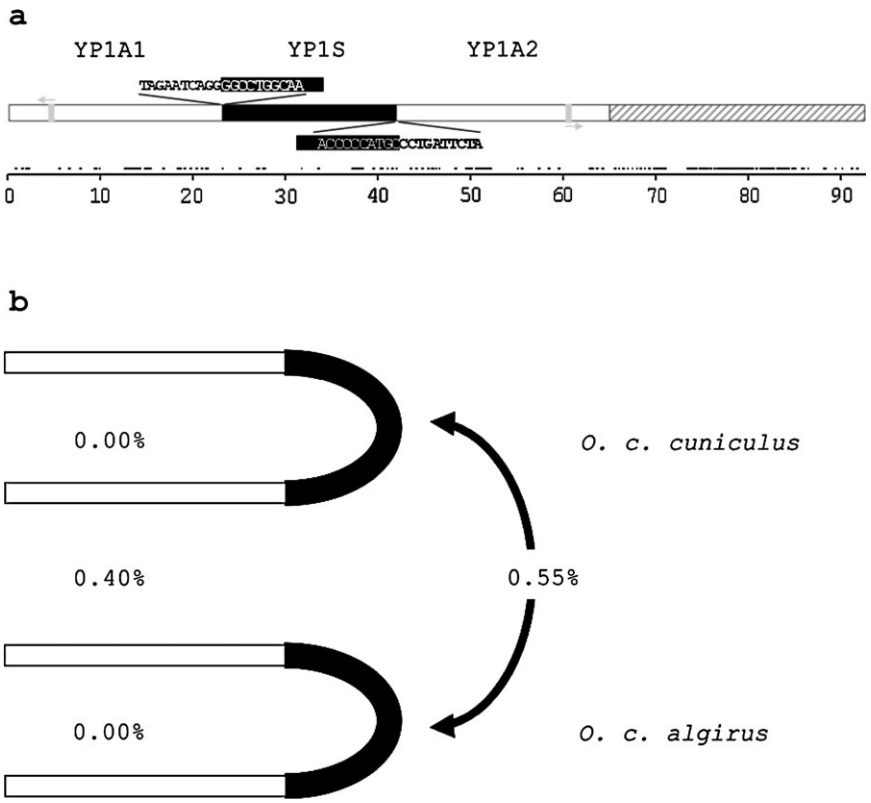

Fig. 1. *SRY* palindromic structure in rabbits and levels of divergence between orthologous and paralogous regions. (*a*) Shown are the results of the sequencing and assembly of a MSY BAC clone from a domestic rabbit genomic library. The two arms of the palindrome are shown as white boxes. YP1A1 extends from the beginning of the BAC clone to 23,086 bp, and YP1A2 begins at 41,999 bp and extends at least until 65,085 bp. The hatched box represents 27,572 bp of sequence that is single copy in the sequenced clone but that possibly is still part of YP1A2. The black box indicates the 18,911 bp YP1S between the two palindrome arms. Ten base pairs of sequence at the boundary between YP1A1 and YP1S and between YP1S and YP1A2 are shown, as well as 10 bp of the YP1S sequence in the same regions. Note that these YP1A1 and YP1A2 sequences are complementary. The *SRY* gene is shown as a gray box, and the direction of transcription of each copy is indicated by an arrow. Above the scale (in kilobases), each line indicates an interspersed repeat identified with Repeat Masker. (*b*) Divergence between orthologous sequences of YP1S (in black) of *Oryctolagus cuniculus cuniculus* and *O. c. algirus* in the right. Divergence between paralogous YP1A1 and YP1A2 in *O. c. cuniculus* at the top, and in *O. c. algirus* at the bottom (all in white). Divergence between orthologous palindrome arms at the center.

and 5 indel polymorphisms located in regions of repetitive sequence. We scanned this clone for interspersed repeats with Repeat Masker and found that repeats make up 37.8% of the total sequence. In the human MSY, repeats make up a similar fraction of ampliconic regions (35.5%), a value much lower than the value for the entire euchromatic MSY (47.4%) or the X-degenerate regions (56.8%) (Skaletsky et al. 2003). After masking all interspersed repeats, we used GeneScan to identify other genes in this BAC clone. Two putative genes were identified, but these were discarded because they had significant Blast hits to retroviral elements present throughout the rabbit genome. Our results demonstrate that there are at least two copies of *SRY* in the rabbit genome and that these copies are located in the arms of an inverted repeat.

Next, we sought to determine the age of the *SRY* duplication by asking whether the duplication boundaries are shared among closely related species. We designed primers (all primers used in this manuscript are available upon request) to polymerase chain reaction (PCR) amplify the two boundaries between palindrome arms and the internal single-copy spacer in the rabbit subspecies (*O. c. algirus* and *O. c. cuniculus*), one hare (*Lepus* spp.), and one cottontail (*Sylvilagus* spp.). Amplification of the boundary between YP1A1 and YP1S was successful in all samples, but the boundary between YP1S and YP1A2 was only successfully amplified in *O. c. algirus* and *O. c. cuniculus*. The two rabbit subspecies are distributed parapatrically in Iberia and diverged approximately 2 Ma (Bijuduval et al. 1991; Branco et al. 2000; Geraldes et al. 2006; Carneiro et al. 2009), whereas *Oryctolagus* diverged from *Lepus* and *Sylvilagus* approximately 11.8 Ma (Matthee et al. 2004). This indicates that the *SRY* gene is located in the same palindromic structure in both rabbit subspecies and that the duplication occurred at least 2 Ma. It is also possible that the palindrome predates the divergence between *Oryctolagus* and *Lepus*. Although we could only amplify one of the boundaries in *Lepus*, previous work showed that *SRY* is not a single-copy gene in this genus (Putze et al. 2007). The finding that in rabbits *SRY* has the characteristics of an ampliconic gene, although in humans it is a single-copy X-degenerate gene, reinforces the idea that these three classes of genes are somewhat fluid and that ampliconic genes can arise from X-degenerate genes or from genes acquired from other genomic regions (Graves 2006).

Table 1. Nucleotide Divergence between Orthologous MSY Sequences in Two Rabbit Subspecies and Hares.

Region	Length (bp)	Dxy[a] (%)		Divergence Time[b] (Ma)
		Oryctolagus cuniculus algirus to *O. c. cuniculus*	*O. cuniculus* to *Lepus granatensis*	*O. c. algirus* to *O. c. cuniculus*
YP1A	1,239	0.40	5.51	0.86
YP1S	1,831	0.55	3.03	2.14

[a] Average pairwise distance.
[b] Divergence time between rabbit subspecies was calculated assuming a molecular clock and divergence time between rabbits and hares of 11.8 Ma (Matthee et al. 2004).

The discovery that *SRY* is located in a palindrome provided the opportunity to ask whether there is evidence for gene conversion between palindrome arms in rabbits, as seen in the human MSY (Rozen et al. 2003). If there is ongoing gene conversion between palindrome arms, divergence between paralogous arms of palindromes should be reduced relative to divergence between orthologous arms. Thus, we predict that YP1A1 and YP1A2 within each rabbit subspecies should be more similar to each other than either is to its ortholog in the other subspecies. To test this hypothesis, we PCR amplified and sequenced ~1.2 kb from the palindrome arms of both rabbit subspecies. We were able to amplify fragments corresponding to each arm of the palindrome by placing one PCR primer within the YP1S region and one primer within the palindrome arm for each fragment. We found that paralogous copies (YP1A1 and YP1A2) within each rabbit subspecies were identical, yet divergence between orthologous copies was 0.40% (fig. 1*b* and table 1), providing clear evidence for gene conversion. Furthermore, previous work provided some indirect evidence for gene conversion at *SRY*: a population survey of the distribution of a 7 bp insertion in the 3′ region of *SRY* in rabbit subspecies in Iberia (Geraldes and Ferrand 2006) showed that although this insertion was absent in *O. c. cuniculus*, in *O. c. algirus* some individuals had the insertion in the two *SRY* copies, some had the insertion in just one *SRY* copy, and some, just like *O. c. cuniculus*, completely lacked the insertion. The authors suggested that either 1) an identical insertion arose independently in each *SRY* copy or 2) the insertion arose in just one of the *SRY* copies and was then converted in some but not all individuals onto the other *SRY* copy.

We also sequenced ~1.8 kb in the YP1S region and found that divergence between rabbit subspecies was 0.55% (fig. 1*b* and table 1). This value is slightly higher than the divergence between orthologous palindrome arms, though both estimates are based on only a few substitutions and are therefore approximate. This is in contrast to the divergence between *Oryctolagus* and *Lepus*, where divergence at YP1S (3.03%) is lower than divergence at YP1A (5.51%). Assuming a divergence time between *Oryctolagus* and *Lepus* of 11.8 Ma (Matthee et al. 2004), we estimate that rabbit subspecies diverged 0.87 Ma based on palindrome arms and 2.14 Ma based on the palindrome spacer. Although approximate, the value of 2.14 Ma is similar to previous reports based on mitochondrial DNA (Bijuduval et al. 1991; Branco et al. 2000), X-linked (Geraldes et al. 2006), and autosomal (Carneiro et al. 2009) data.

Here, we provide the first evidence of gene conversion within MSY palindromes outside of primates, suggesting that this might be a general feature of mammalian Y chromosomes. This, coupled with previous evidence for gene conversion on the Z chromosome of birds (Backstrom et al. 2005), suggests that gene conversion might be a common feature of sex chromosomes that lack homologous recombination. Rozen et al. (2003) suggested that gene conversion of Y ampliconic genes might prevent gene decay by preferentially restoring ancestral alleles. Consistent with this hypothesis, they observed lower levels of sequence divergence between humans and chimpanzees for palindromes than for single-copy regions. In contrast, we observed slightly higher levels of divergence between *Oryctolagus* and *Lepus* in the palindrome than in the single-copy spacer (table 1), providing no evidence that gene conversion is reducing the rate of evolution in these taxa. Although multiple copies of *SRY* have been described in other mammalian species, the genomic context was not characterized and therefore it was not possible to determine if they were located in a palindrome or not. It remains to be seen whether *SRY* is an ampliconic gene in other species of mammals and whether primates are the exception to the rule by having just one copy of *SRY*.

Acknowledgments

We thank Claire Rogel-Gaillard and Helene Hayes for providing the *SRY*-containing clone (828D7), members of the Nachman Lab and CIBIO who provided invaluable suggestions during the course of this project, and two anonymous reviewers for their insightful comments. This work was supported by Fundacao para a Ciencia e a Tecnologia (PTDC/BIA-BDE/72304/2006 project and SFRH/BD/4621/2001 doctoral fellowship to A.G.); National Science Foundation and National Institute of Health (to M.W.N.).

References

Backstrom N, Ceplitis H, Berlin S, Ellegren H. 2005. Gene conversion drives the evolution of HINTW, an ampliconic gene on the female-specific avian W chromosome. *Mol Biol Evol.* 22: 1992–1999.

Bianchi NO, Bianchi MS, Bailliet G, de la Chapelle A. 1993. Characterization and sequencing of the sex determining region Y gene (Sry) in Akodon (Cricetidae) species with sex reversed females. *Chromosoma* 102:389–395.

Bijuduval C, Ennafaa H, Dennebouy N, Monnerot M, Mignotte F, Soriguer RC, Elgaaied A, Elhili A, Mounolou JC. 1991. Mitochondrial-DNA evolution in lagomorphs—origin of systematic heteroplasmy

and organization of diversity in European rabbits. *J Mol Evol*. 33:92–102.

Branco M, Ferrand N, Monnerot M. 2000. Phylogeography of the European rabbit (Oryctolagus cuniculus) in the Iberian Peninsula inferred from RFLP analysis of the cytochrome b gene. *Heredity* 85(Pt 4):307–317.

Bullejos M, Sanchez A, Burgos M, Jimenez R, Diaz De La Guardia R. 1999. Multiple mono- and polymorphic Y-linked copies of the SRY HMG-box in microtidae. *Cytogenet Cell Genet*. 86: 46–50.

Carneiro M, Ferrand N, Nachman MW. 2009. Recombination and speciation: loci near centromeres are more differentiated than loci near telomeres between subspecies of the European rabbit (Oryctolagus cuniculus). *Genetics* 181:593–606.

Geraldes A, Ferrand N. 2006. A 7-bp insertion in the 3' untranslated region suggests the duplication and concerted evolution of the rabbit SRY gene. *Genet Sel Evol*. 38:313–320.

Geraldes A, Ferrand N, Nachman MW. 2006. Contrasting patterns of introgression at X-linked loci across the hybrid zone between subspecies of the European rabbit (Oryctolagus cuniculus). *Genetics* 173:919–933.

Graves JA. 2006. Sex chromosome specialization and degeneration in mammals. *Cell* 124:901–914.

Gubbay J, Collignon J, Koopman P, Capel B, Economou A, Munsterberg A, Vivian N, Goodfellow P, Lovellbadge R. 1990. A gene-mapping to the sex-determining region of the mouse Y-chromosome is a member of a novel family of embryonically expressed genes. *Nature* 346:245–250.

Hayes H, Rogel-Gaillard C, Zijlstra C, de Haan NA, Urien C, Bourgeaux N, Bertaud M, Bosma AA. 2002. Establishment of an R-banded rabbit karyotype nomenclature by FISH localization of 23 chromosome-specific genes on both G- and R-banded chromosomes. *Cytogenet Genome Res*. 98:199–205.

Koopman P, Gubbay J, Vivian N, Goodfellow P, Lovell-Badge R. 1991. Male development of chromosomally female mice transgenic for Sry. *Nature*. 351:117–121.

Lundrigan BL, Tucker PK. 1997. Evidence for multiple functional copies of the male sex-determining locus, Sry, in African murine rodents. *J Mol Evol*. 45:60–65.

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.

Nagamine CM. 1994. The testis-determining gene, SRY, exists in multiple copies in Old World rodents. *Genet Res*. 64:151–159.

Putze M, Nurnberg S, Fickel J. 2007. Y-chromosomal markers for the European brown hare (Lepus europaeus, Pallas 1778). *Eur J Wildl Res*. 53:257–264.

The Rice Chromosome 10 Sequencing Consortium. 2003. In-depth view of structure, activity, and evolution of rice chromosome 10. *Science* 300:1566–1569.

Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Page DC. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* 423:873–876.

Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischauf AM, Lovell-Badge R, Goodfellow PN. 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* 346:240–244.

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, et al. (40 co-authors). 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423:825–837.

Turner ME, Martin C, Martins AS, Dunmire J, Farkas J, Ely DL, Milsted A. 2007. Genomic and expression analysis of multiple Sry loci from a single Rattus norvegicus Y chromosome. *BMC Genet*. 8:11.