Tracking Ancient Polyploids: A Retroposon Insertion Reveals an Extinct Diploid Ancestor in the Polyploid Origin of Belladonna

Yao-wu Yuan,*† Zhi-yun Zhang,* Zhi-duan Chen,* and Richard G. Olmstead†

*State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, People's Republic of China and †Department of Biology, University of Washington

Polyploidy is a prominent process in plant evolution and adaptation, but molecular phylogenetic studies of polyploids based on DNA sequences have often been confounded by their complex gene and genome histories. We report here a retroposon insertion in the nuclear gene granule-bound starch synthase I (GBSSI or "waxy") that clearly reveals the ancient hybrid history of the medically important polyploid species belladonna (*Atropa belladonna*) and resolves the controversy over the taxonomic group to which it belongs, the tribe Hyoscyameae (Solanaceae). Our inferences based on the pattern of presence or absence of the retroposon insertion are corroborated by phylogenetic analyses of the GBSSI gene sequences. This case may suggest that retroposons are promising molecular markers to study polyploid evolution.

Polyploidy is a prominent process in plant evolution, where 50% or more of flowering plants and 95% of ferns and fern allies are polyploids (Goldblatt 1980; Grant 1981; Masterson 1994), including many crop plants of worldwide importance (e.g., rice, wheat, cotton, and soybean). More recent genomic studies even suggest that probably all angiosperms have had at least one polyploidization event somewhere in their evolutionary history (Vision et al. 2000; Bowers et al. 2003). Tracing the evolutionary history of polyploids is essential for crop genetic engineering and understanding the extant plant diversity in general. However, reticulate evolution and the complex gene or genome histories in polyploids often confound phylogenetic inferences based on DNA sequences (Wendel 2000) and subsequently restrict other evolutionary studies of polyploids in the light of phylogeny.

In the last several years, a new source of phylogenetic characters, retroposable elements, especially SINE (short interspersed element) families, have been employed as a powerful tool to complement DNA sequence data for addressing phylogenetic questions in a number of animal groups (Murata et al. 1993; Shimamura et al. 1997; Nikaido et al. 1999; Terai et al. 2003; Sasaki et al. 2004; Nikaido et al. 2006). Retroposons move within the genome via a copy-and-paste process: the parent locus is transcribed into RNA and then the genetic information is reverse transcribed from RNA back into chromosomal DNA at a new locus (Rogers 1983; Okada 1991). Because of this replicative mechanism, the integration of a retroposon at a new locus is considered to be an irreversible event, and its target site is chosen almost at random (Okada 1991). These characteristics of retroposons make them excellent tools for the determination of phylogenetic relationships, with the probability of homoplasy very small (Okada 1991; Nikaido et al. 1999). However, in contrast to the extensive data on retroposons available for animals, only a few retroposons have been well characterized in plants. For this reason, the "SINE method" (Cook and Tristem 1997; Shedlock and Okada 2000; Okada et al. 2004; Shedlock et al. 2004) has not stimulated much attention among plant systematists and evolutionists. The

extraordinary potential of using this type of molecular marker to trace the evolutionary history of polyploids has been largely unexplored.

The delimitation of tribe Hyoscyameae (Solanaceae) has been debated widely (Bentham 1876; Tetenyi 1987; D'Arcy 1991; Hoare and Knapp 1997; Olmstead et al. 1999; Hunziker 2001). The focus of these arguments is the affinity of *Atropa* to the traditionally recognized Hyoscyameae, which includes 6 genera (*Anisodus*, *Atropanthe*, *Hyoscyamus*, *Physochlaina*, *Przewalskia*, and *Scopolia*). *Atropa* contains 3 species (Hoare and Knapp 1997; Hunziker 2001), including belladonna (*Atropa belladonna*), source of the drug atropine. Traditional classifications that treated *Atropa* and the tribe Hyoscyameae separately did so mainly on the basis that the plants in *Atropa* have fleshy, indehiscent fruits, whereas the other 6 genera have unusual circumscissile capsules. Previous molecular studies based on chloroplast DNA (cpDNA) (Olmstead and Sweere 1994; Olmstead et al. 1999) showed that *Atropa* was grouped with the traditionally recognized Hyoscyameae to form a well-supported monophyletic group. Consequently, *Atropa* was included in the tribe Hyoscyameae in their provisional phylogenetic classification (Olmstead et al. 1999). Chromosome numbers have been counted extensively for this group in the past. Numbers for the traditionally recognized Hyoscyameae, with their corresponding references, have been very nicely summarized in a recent paper (Tu et al. 2005). These 6 genera are predominantly tetraploids, with base chromosome number $X = 12$, which has been used as a synapomorphy to define a major clade within Solanaceae (Olmstead and Sweere 1994). *Anisodus*, *Atropanthe*, and *Scopola* all have $2n = 48$, *Physochlaina* has $2n = 42$, *Przewalskia* has $2n = 44$, whereas *Hyoscyamus* possesses much variations in both base chromosome number and ploidy level. There are 8 counts for *A. belladonna*, with corresponding references, in the Missouri Botanical Garden TROPICS database (http://mobot.mobot.org/W3T/Search/ipcn.html). Seven of them reported this species as hexaploids, with $2n = 72$, the other one reported $2n = 60$.

Clarifying the phylogenetic affinity and evolutionary history of this medically important polyploid species is essential for potential genetic improvement of belladonna cultivars by conventional or molecular assisted plant breeding. Here we report a retroposon insertion in the nuclear gene granule-bound starch synthase I (GBSSI or "waxy") that reveals the ancient hybrid history of belladonna

Key words: retroposon insertion, SINEs, Hyoscyameae, *Atropa belladonna*, polyploid, phylogeny.
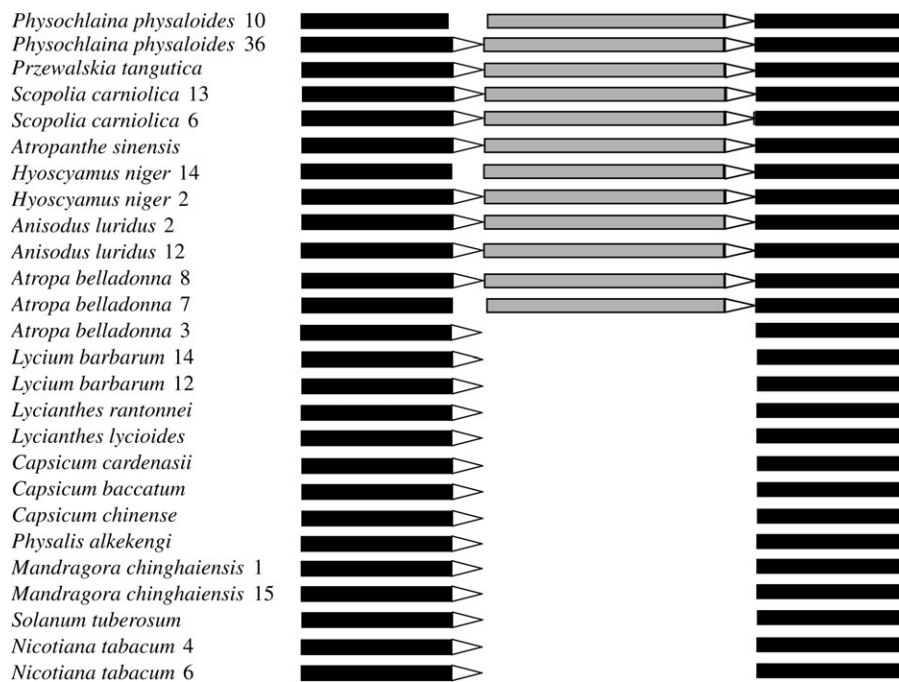
E-mail: zhangzy@ibcas.ac.cn.

Fig. 1.—Schematic representation of a portion of the alignment that shows the retroposon insertion. Black boxes represent flanking sequences; gray boxes, inserted retroposon sequences; and white arrows, target-site duplications (GGTCCTGAG). Numbers following species names designate clones. This retroposon is found in all the traditionally recognized Hyoscyameae taxa and 2 of the 3 copies from *Atropa belladonna*. The third copy, *Atropa. belladonna* 3, does not possess this insertion. Note that 3 gene copies (*Physochlaina physaloides* 10, *Hyoscyamus niger* 14, and *Atropa belladonna* 7) have lost one of target-site duplications.

(*A. belladonna*) and defines the monophyly of the tribe Hyoscyameae to which it belongs.

We sampled *A. belladonna* and representatives of all 6 genera of the traditionally recognized Hyoscyameae, as well as several outgroup species in the family Solanaceae (supplementary table 1, Supplementary Material online). The region from exon 3 to exon 10 of the GBSSI gene was amplified, cloned, and sequenced (see Materials and Methods in Supplementary Material online). The total aligned length of these GBSSI sequences is 2,253 bp. A large insertion in intron 3, with the length approximately 258 bp, was found in all the traditionally recognized Hyoscyameae and 2 of the 3 copies from *A. belladonna* but not in other taxa of the family Solanaceae. It is flanked by a short direct repeat of sequence "GGTCCTGAG" (fig. 1, for details see supplementary fig. S1, Supplementary Material online), which is a hallmark of transposition and retroposition (Li 1997). A Blast search of this inserted element in GenBank found no similar sequences, and further mask against Repbase Update (Jurka et al. 2005) using CENSOR (Jurka et al. 2005) also found no similar repeated sequences, which implies that this might be a previously unidentified transposable element. The fact that no terminal inverted-repeat sequences were found in this element suggests a retroposon instead of a MITE (miniature inverted-repeat transposable element) or any other kinds of DNA transposons. Given that the size of this element is fairly small, it is most likely a SINE. Most SINEs characterized in animals and all SINEs characterized in plants so far are derived from tRNAs and thus have a tRNA-related region at the 5′ end (Okada et al. 2004). Unfortunately, a careful step-by-step characterization of both the inserted sequence and its reverse compliment sequence, following the protocol of Okada et al. (2004), failed to identify a tRNA-related region. SINEs with 5′ end tRNA-related region truncated seem to be fairly common in plants (e.g., TSb in Solanaceae, Yoshioka et al. 1993; RathE1 and RathE2 in Brassicaceae, Lenoir et al. 2005). Although it is mostly likely to be a 5′ end–truncated SINE, we cannot rule out the possibility that it is a processed retropseudogene or truncated long interspersed element.

Phylogenetic reconstructions were conducted using the Maximum Parsimony (MP) and maximum likelihood (ML) optimality criteria as implemented in PAUP*4.0b10 (Swofford 2002). MP analysis resulted in 8 equally most parsimonious trees. ML analysis generated a single tree ($-\ln = 9{,}167.0206$) (fig. 2), which is topologically identical to 1 of the 8 MP trees. With *Nicotiana* as outgroup, the 6 tribes (classification see supplementary table 1, Supplementary Material online) in subfamily Solanoideae were divided into 2 main clades. One was composed of the tribe Hyoscyameae and Lycieae, the other comprised the tribe Solaneae, Capsiceae, Physaleae, and Mandragoreae. The relationships among the tribes in the subfamily Solanoideae are consistent with the phylogeny inferred from cpDNA data (Olmstead et al. 1999). *Atropa* and the other 6 traditionally recognized genera of Hyoscyameae form a clade, with 100% bootstrap support in both MP and ML analyses, that is sister group to the tribe Lycieae. However, the 3 GBSSI copies of *A. belladonna* are not monophyletic; 2 of them group with GBSSI copies of the other 6 genera to form a clade that is defined by the retroposon insertion
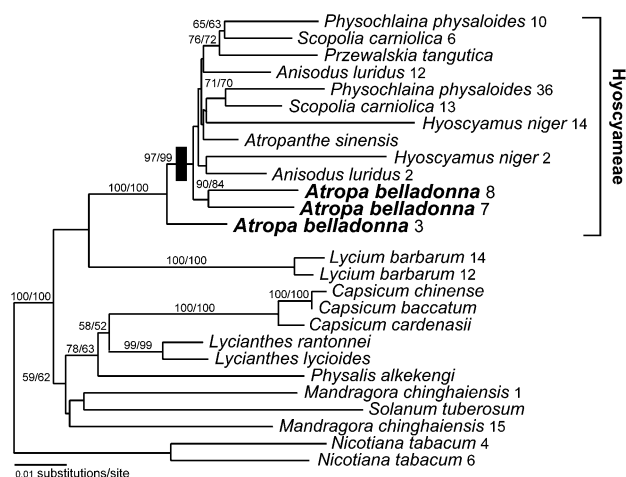
FIG. 2.—Phylogram of the single ML tree resulted from GBSSI data set. This tree was identical to one of the 8 MP trees. The bootstrap values of both ML and MP analyses supporting the corresponding branches are shown when greater than 50% (ML/MP). The shaded box represents the retroposon insertion in intron 3. Of the 26 GBSSI sequences analyzed, 21 were generated in this study and are deposited in GenBank (accession numbers DQ069251–DQ069271). The other 5 sequences were retrieved from GenBank (*Solanum tuberosum* X58453 [van der Leij et al. 1991]; *Capsicum baccatum* var. *baccatum* AF397120, *Capsicum cardenasii* AF397116, *Lycianthes lycioides* AF397111, and *Lycianthes rantonnei* AF397112 [Walsh and Hoot 2001]). For more taxa information, see supplementary table 1 in Supplementary Material online.

(fig. 2). The third copy does not possess this retroposon insertion and stands out as sister group of all the other Hyoscyameae copies.

The fact that this one copy does not have the retroposon insertion (fig. 1, *A. belladonna* 3), similar to those from other Solanaceae, implies that this copy has an origin outside the clade defined by the presence of this insertion, or alternatively, the retroposon insertion was deleted in this copy subsequent to the diversification of the group. The latter hypothesis is very unlikely because retroposon insertion events have been argued to be irreversible (Okada 1991; Nikaido et al. 1999) due to their intrinsic replicative mechanism. Even when a rare deletion of a retroposon does occurs, it leaves behind a molecular signature in the flanking sequences (Edwards and Gibbs 1992; Shedlock et al. 2004). There is no such signature in the flanking sequence of this particular gene copy (fig.1 and supplementary fig. S1, Supplementary Material online). Furthermore, the phylogenetic analyses based on GBSSI gene sequences support the former hypothesis. The 2 copies of *A. belladonna* that possess the insertion group with copies from the other Hyoscyameae in a strongly supported clade. The third copy occupies the position sister to this clade (fig. 2). Based on this clear and direct evidence from the retroposon insertion, and on evidence from phylogenetic analyses, we conclude that *A. belladonna* has an allopolyploid origin. Given the fact that samples for the other genera of Hyoscyameae are tetraploids and *A. belladonna* is a hexaploid, it appears that *A. belladonna* was involved in an ancient natural hybridization between a tetraploid species on the main stem of Hyoscyameae evolution and a now-extinct diploid lineage sister to the tetraploid lineage.

The phylogenetic affinity of *Atropa* to the traditionally recognized Hyoscyameae has been controversial for over a century (Bentham 1876; Tetenyi 1987; D'Arcy 1991; Hoare and Knapp 1997; Olmstead et al. 1999; Hunziker 2001). Our phylogenetic analyses showed that *Atropa* groups with the traditionally recognized Hyoscyameae in a strongly supported clade (fig. 2), with *Lycium* as its sister group, which is consistent with previous molecular studies (Olmstead and Sweere 1994; Olmstead et al. 1999). This result is not inconsistent with traditional classifications but highlights the problem of nonphylogenetic classification. *Atropa*, along with most Lycieae and subfamily Solanoideae to which they belong, are characterized by fleshy fruits. The capsular fruits of the other Hyoscyameae are a derived trait (Knapp 2002), thus indicative of their monophyly but not contradicting monophyly of a more inclusive group including *Atropa*. The weak molecular support for monophyly of the other 6 genera also argues for including *Atropa* in the Hyoscyameae.

We infer that this retroposon is an insertion constrained in the Hyoscyameae by aligning the Hyoscyameae GBSSI sequences with those from outgroup genera, which represent 4 tribes in the subfamily Solanoideae and a more distantly related subfamily, Nicotianoideae (supplementary table 1, Supplementary Material online). It is absent from GBSSI of all outgroup taxa. Furthermore, it was not found in any other taxa of Solanaceae that have been investigated using GBSSI gene sequences (Peralta and Spooner 2001; Walsh and Hoot 2001). This fact suggests that the retroposon is a synapomorphy for Hyoscyameae and thus corroborates our phylogenetic analyses and the previous cpDNA studies (Olmstead and Sweere 1994; Olmstead et al. 1999) that Hyoscyameae, comprising 7 genera, *Anisodus*, *Atropa*, *Atropanthe*, *Hyoscyamus*, *Physochlaina*, *Przewalskia*, and *Scopolia*, is a monophyletic group.

The irreversible nature of retroposon insertion events suggests their 2 major advantages over other molecular markers (Shedlock et al. 2004): 1) extremely low probability of homoplasy and 2) straightforward identification of character polarity and thus no ambiguities in outgroup selection and tree rooting. Retroposons as phylogenetic markers may avoid many shortcomings of sequence data (Cook and Tristem 1997), including long-branch attraction, heterogeneous substitution rate, and dependence on methodology of phylogenetic analyses. One of the big drawbacks of this type of maker is the large effort:signal ratio for a typical systematic project (Shedlock et al. 2004). The other major limit is that they are not suitable for addressing deep phylogeny questions (Shedlock et al. 2004) because these elements will become unrecognizable after a long period of time as random mutations accumulate. In addition, both historical and recent incomplete lineage sorting could cause incongruence among different retroposon loci (Terai et al. 2003; Nikaido et al. 2006). For plants, in particular, extensive hybridization, introgression, and polyploidization could further confound phylogenetic reconstructions. These complications have been demonstrated in a few pioneer applications of retroposons as phylogenetic markers in "model system" plants (Mochizuki et al. 1992; Deragon et al. 1994; Jing et al. 2005). However, complications caused by incomplete lineage sorting,

hybridization, introgression, and polyploidization also apply to all other kinds of markers. The interaction between these taxon-intrinsic complications and shortcomings of common molecular markers such as DNA sequences makes plant phylogenetic reconstruction at species level remarkably difficult. Comprehensive analyses of multiple retroposon insertions (e.g., >50 loci) could shed a light on this problem by avoiding most drawbacks of typical molecular markers. The wealth of retroposons in eukaryotic genomes is exemplified by the human genome (~3 million retroposons) (Lander et al. 2001), and because polyploidy may lead to activation of transposons and retroposons (Comai 2000), they may be very abundant in polyploids. The fact that a retroposon insertion defines a monophyletic Hyoscyameae and discloses the allopolyploid origin of belladonna highlights the potential of retroposons as promising markers for studying plant systematics, especially, the origin and evolution of polyploids. Although the implications of this study are limited by the use of only a single locus, we wish to stimulate more applications of retroposons from multiple loci in plant systematics.

Supplementary Material

Materials and Methods (Materials_Methods.pdf), supplementary figure S1 (supplemental_figure_1.pdf), and supplementary table 1 (supplemental_table_1.pdf) are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

Literature Cited

Bentham G. 1876. Solanaceae. In: Bentham G, Hooker JD, editors. Genera plantarum. Vol. 2. London: L. Reeve & Co. p. 882–913.

Bowers JE, Chapman BA, Rong JK, Paterson AH. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature. 422:433–438.

Comai L. 2000. Genetic and epigenetic interactions in allopolyploid plants. Plant Mol Biol. 43:387–399.

Cook JM, Tristem M. 1997. 'SINEs of the times'—transposable elements as clade markers for their hosts. Trends Ecol Evol. 12:295–297.

D'Arcy WG. 1991. The Solanaceae since 1976, with a review of its biogeography. In: Hawkes JG, Lester RN, Nee M, Estrada RN, editors. Solanaceae III: taxonomy, chemistry, evolution. Kew (UK): Royal Botanic Gardens. p. 75–138.

Deragon JM, Landry BS, Pelissier T, Tutosis S, Tourmente S, Picard G. 1994. An analysis of retroposition in plants based on a family of SINEs from Brassica-Napus. J Mol Evol. 39:378–386.

Edwards MC, Gibbs RA. 1992. A human dimorphism resulting from loss of an Alu. Genomics. 14:590–597.

Goldblatt P. 1980. Polyploidy in angiosperms: monocotyledons. In: Lewis WH, editor. Polyploidy: biological relevance. New York: Plenum Press. p. 219–239.

Grant V. 1981. Plant speciation. 2nd ed. New York: Columbia University Press.

Hoare AL, Knapp S. 1997. A phylogenetic conspectus of the tribe Hyoscyameae (Solanaceae). Bull Nat Hist Mus Lond (Bot). 27:11–29.

Hunziker AT. 2001. Genera Solanacearum: the genera of Solanaceae illustrated, arranged according to a new system. Konigstein (Germany): Koeltz Scientific Books.

Jing RC, Knox MR, Lee JM, Vershinin AV, Ambrose M, Ellis THN, Flavell AJ. 2005. Insertional polymorphism and antiquity of PDR1 retrotransposon insertions in *Pisum* species. Genetics. 171:741–752.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 110:462–467.

Knapp S. 2002. Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae. J Exp Bot. 53: 2001–2022.

Lander ES, Linton LM, Birren B, et al. (242 co-authors). 2001. Initial sequencing and analysis of the human genome. Nature. 409:860–921.

Lenoir A, Pelissier T, Bousquet-Antonelli C, Deragon JM. 2005. Comparative evolution history of SINEs in *Arabidopsis thaliana* and *Brassica oleracea*: evidence for a high rate of SINE loss. Cytogenet Genome Res. 110:441–447.

Li WH. 1997. Molecular evolution. Sunderland (MA): Sinauer.

Masterson J. 1994. Stomatal size in fossil plants—evidence for polyploidy in majority of Angiosperms. Science. 264:421–424.

Mochizuki K, Umeda M, Ohtsubo H, Ohtsubo E. 1992. Characterization of a plant SINE, p-SINE1, in rice genomes. Jpn J Genet. 67:155–166.

Murata S, Takasaki N, Saitoh M, Okada N. 1993. Determination of the phylogenetic-relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. Proc Natl Acad Sci USA. 90:6995–6999.

Nikaido M, Rooney AP, Okada N. 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: Hippopotamuses are the closest extant relatives of whales. Proc Natl Acad Sci USA. 96: 10261–10266.

Nikaido M, Hamilton H, Makino H, Sasaki T, Takahashi K, Goto M, Kanda N, Pastene LA, Okada N. 2006. Baleen whale phylogeny and a past extensive radiation event revealed by SINE insertion analysis. Mol Biol Evol. 23:866–873.

Okada N. 1991. SINEs: short interspersed repeated elements of the eukaryotic genome. Trends Ecol Evol. 6:358–361.

Okada N, Shedlock AM, Nikaido M. 2004. Retroposon mapping in molecular systematics. In: Miller WJ, Capy P, editors. Mobile genetic elements: protocols and genomic applications. Totowa (NJ): Humana Press. p. 189–226. (Methods in molecular biology series).

Olmstead RG, Sweere JA. 1994. Combining data in phylogenetic systematics—an empirical-approach using 3 molecular-data sets in the Solanaceae. Syst Biol. 43:467–481.

Olmstead RG, Sweere JA, Spangler RE, Bohs L, Palmer JD. 1999. Phylogeny and provisional classification of the Solanaceae based on chloroplast DNA. In: Nee M, Symon DE, Lester RN, Jessop JP, editors. Solanaceae IV: advances in biology and utilization. Kew (UK): Royal Botanic Gardens. p. 111–137.

Peralta IE, Spooner DM. 2001. Granule-bound starch synthase (GBSSI) gene phylogeny of wild tomatoes (Solanum L. section Lycopersicon Mill. Wettst. subsection Lycopersicon). Am J Bot. 88:1888–1902.

Rogers J. 1983. Retroposons defined. Nature. 301:460.

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.

Shedlock AM, Okada N. 2000. SINE insertions: powerful tools for molecular systematics. Bioessays. 22:148–160.

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

Shimamura M, Yasue H, Ohshima K, Abe H, Kato H, Kishiro T, Goto M, Munechika I, Okada N. 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature. 388:666–670.

Swofford DL. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4b10. Sunderland (MA): Sinauer Associates, Inc.

Terai Y, Takahashi K, Nishida M, Sato T, Okada N. 2003. Using SINEs to probe ancient explosive speciation: "hidden" radiation of African cichlids? Mol Biol Evol. 20:924–930.

Tetenyi P. 1987. A chemotaxonomic classification of the Solanaceae. Ann Mo Bot Gard. 74:600–608.

Tu TY, Sun H, Gu ZJ, Yue JP. 2005. Cytological studies on the Sino-Himalayan endemic Anisodus and four related genera from the tribe Hyoscyameae (Solanaceae) and their systematic and evolutionary implications. Bot J Linn Soc. 147:457–468.

van der Leij FR, Visser RGF, Ponstein AS, Jacobsen E, Feenstra WJ. 1991. Sequence of the structure gene for granule-bound starch synthase of potato (Solanum tuberosum L.) and evidence for a single point deletion in the amf allele. Mol Gen Genet. 228:240–248.

Vision TJ, Brown DG, Tanksley SD. 2000. The origins of genomic duplications in Arabidopsis. Science. 290:2114–2117.

Walsh BM, Hoot SB. 2001. Phylogenetic relationships of capsicum (Solanaceae) using DNA sequences from two noncoding regions: the chloroplast atpB-rbcL spacer region and nuclear waxy introns. Int J Plant Sci. 162:1409–1418.

Wendel JF. 2000. Genome evolution in polyploids. Plant Mol Biol. 42:225–249.

Yoshioka YY, Matsumoto S, Kojima S, Ohshima K, Okada N, Machida Y. 1993. Molecular characterization of a short interspersed repetitive element from tobacco that exhibits sequence homology to specific tRNAs. Proc Natl Acad Sci USA. 90: 6562–6566.