12-1-2006

SINEs of a nearly perfect character

David A. Ray
West Virginia University

Jinchuan Xing
Louisiana State University

Abdel Halim Salem
Suez Canal University

Mark A. Batzer
Louisiana State University

Recommended Citation

Ray, D., Xing, J., Salem, A., & Batzer, M. (2006). SINEs of a nearly perfect character. *Systematic Biology, 55* (6), 928-935. https://doi.org/10.1080/10635150600865419

Syst. Biol. 55(6):928–935, 2006

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150600865419

SINEs of a *Nearly* Perfect Character

DAVID A. RAY,[1] JINCHUAN XING,[2] ABDEL-HALIM SALEM,[3] AND MARK A. BATZER[2]

[1]*Department of Biology, West Virginia University, PO Box 6057, Morgantown, West Virginia 26506, USA*
[2]*Department of Biological Sciences, Biological Computation and Visualization Center, Center for Bio-Modular Multiscale Systems, Louisiana State University, 202 Life Sciences Building, Baton Rouge, Louisiana 70803, USA; E-mail: mbatzer@lsu.edu (M.A.B.)*
[3]*Department of Anatomy, Faculty of Medicine, Suez Canal University, Ismailia, Egypt*

Abstract.—Mobile elements have been recognized as powerful tools for phylogenetic and population-level analyses. However, issues regarding potential sources of homoplasy and other misleading events have been raised. We have collected available data for all phylogenetic and population level studies of primates utilizing *Alu* insertion data and examined them for potentially homoplasious and other misleading events. Very low levels of each potential confounding factor in a phylogenetic or population analysis (i.e., lineage sorting, parallel insertions, and precise excision) were found. Although taxa known to be subject to high levels of these types of events may indeed be subject to problems when using SINE analysis, we propose that most taxa will respond as the order Primates has—by the resolution of several long-standing problems observed using sequence-based methods. [*Alu*; mobile element; phylogenetics; retrotransposon; SINE.]

Although implemented previously in smaller studies (Minghetti and Dugaiczyk, 1993; Murata et al., 1993, 1996, 1998; Ryan and Dugaiczyk, 1989), the SINE method of phylogenetic reconstruction began receiving increased attention when Norihiro Okada and colleagues published their investigation into the relationship between whales and artiodactyls (Nikaido et al., 1999; Shimamura et al., 1997). The potential for a close relationship among artiodactyls and cetaceans had been proposed and investigated in earlier studies (Flower, 1883; Gatesy et al., 1999; Sarich, 1985) but had never been so clearly and simply elucidated. These authors pronounced that the SINE method represented a revolution in phylogenetic inference because of its apparent freedom from the problems associated with homoplasy.

SINE insertions in a genome offer two important advantages over other markers used for systematic and population genetic studies. First, the presence of an element in an individual is presumed to represent identity by descent (Batzer and Deininger, 2002). Polymorphic mobile element insertions will thus reflect relationships more accurately than many other markers (i.e., sequence data, restriction fragment length polymorphisms [RFLPs], and microsatellites) that may only reflect identity by state (i.e., homoplasy) (Batzer et al., 1994). A second advantage of these genetic markers is that the ancestral state of an insertion polymorphism is known to be the absence of the element at a particular genomic location (Batzer et al., 1994; Perna et al., 1992). Precise knowledge of the ancestral state of a genomic polymorphism allows us to draw trees of population relationships without making unnecessary assumptions (Batzer et al., 1994; Perna et al., 1992).

Hillis (1999), however, offered several cautious notes about hailing SINEs as the answer to all of our phylogenetic woes. He correctly observed that almost every time a new methodology is developed, it is heralded, at least in the short term, as superior to all methods that have come before. In particular, Hillis worried that aspects of SINE analysis that may contribute to homoplasy had not been sufficiently investigated. Analysis of phylogenies using the SINE method may be adversely affected by various events that could distort the true evolutionary history of species. This has been a problem for every method yet devised to examine evolutionary relationships and SINEs are indeed not immune. They are, however, thought to be less susceptible to these problems.

There are four potential sources for confusion in SINE analysis of phylogenetics (Fig. 1): lineage sorting, parallel insertions (including precise- and near-parallel insertions), precise excision, and paralogous insertions. Lineage sorting is caused by the presence of a polymorphic insertion in a common ancestor that alternatively becomes fixed or extinct in the genomes of daughter species. Parallel insertions include the insertion of distinct elements at or near the same location (within the polymerase chain reaction [PCR] amplicon) in the genomes of different taxa under study. Precise SINE excision was until very recently not thought to be an issue as there was no known mechanism for these sorts of events to occur. Paralogous insertions include duplicated regions of the genome at which an insertion may have occurred at one of the duplicates but not the other.

Shedlock and Okada produced two reviews of the issues associated with the SINE method (Shedlock and Okada, 2000; Shedlock et al., 2004), focusing on the problems of lineage sorting. In particular, they utilized the relatively large data sets that have been collected over the years on cichlid phylogeny and population biology. Since then, however, there have been numerous published studies using SINEs as phylogenetic and population genetic characters in a wider variety of taxa (Bamshad et al., 2003; Batzer and Deininger, 2002; Carroll et al., 2001; Churakov et al., 2005; Cotrim et al., 2004; Kawai et al., 2002; Nasidze et al., 2001; Nishihara et al., 2002; Ray et al., 2005; Roos et al., 2004; Roy-Engel et al., 2002; Salem et al., 2003; Sasaki et al., 2004; Schmitz et al., 2001, 2005; Schmitz and Zischler, 2003; Singer et al., 2003; Takahashi et al., 2001b; Terai et al., 2003, 2004; Watkins et al., 2003; Xing et al., 2005; Zampicinini et al., 2004). These studies have proven very successful and have served to confirm many of the positive

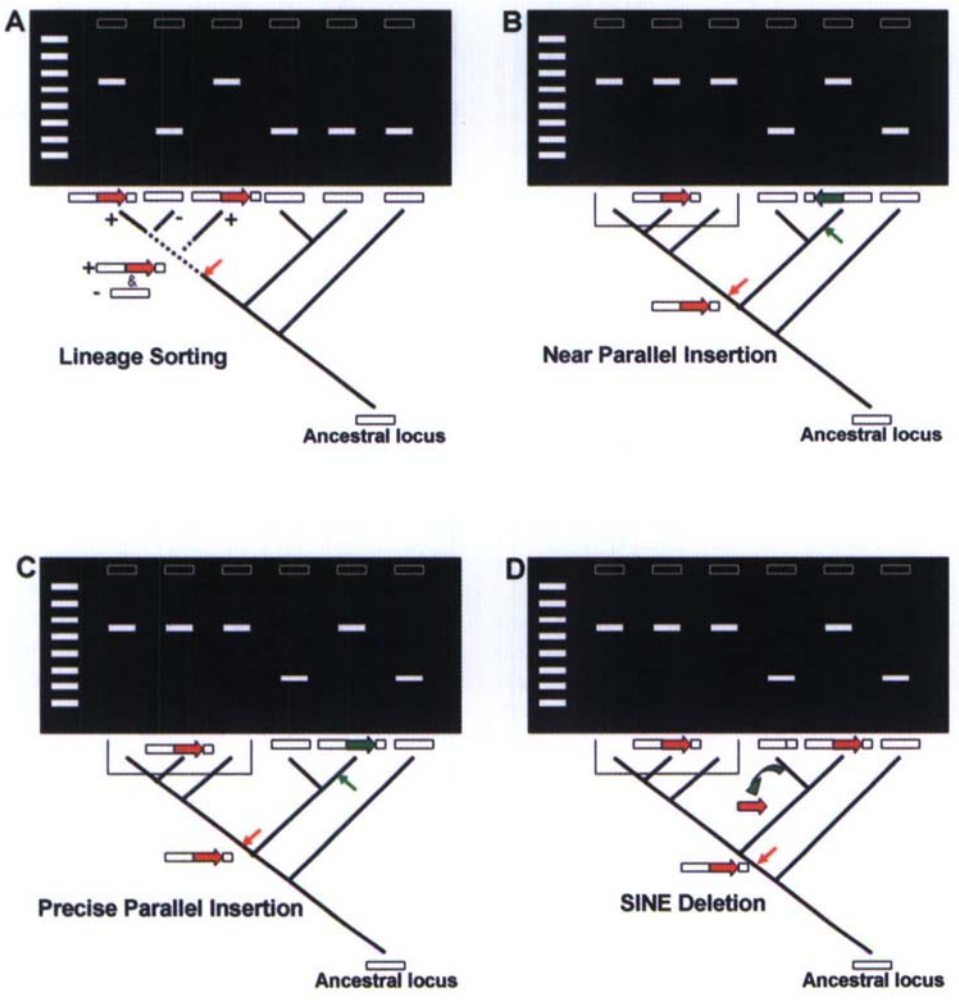

FIGURE 1. Schematic of the types of potentially misleading events encountered during a SINE-based analysis of phylogeny. (A) Lineage sorting due to retention of an ancestral polymorphism and subsequent random fixation or loss in daughter lineages; (B) a potentially confusing pattern due to a near-parallel insertion; (C) insertion homoplasy due to a precise-parallel insertion; and (D) mobile element excision. Straight arrows along branches indicate mobile element insertion events and the arrow head indicates that orientation of the inserted element. The curved arrow indicates the precise removal of an element from the genome.

aspects of SINE analysis of phylogenies and population dynamics.

The most comprehensive work has been performed in primate taxa using *Alu* elements. In fact, nearly the entire order has been investigated using the SINE method (see Table 1). One of the reasons that *Alu* elements have been so successfully used in investigations of primate phylogenetics is the existence of the human genome draft sequence. This has been an invaluable tool for determining the flanking sequences of *Alu* insertions during the process of primer design (Ray et al., 2005; Salem et al., 2003; Schmitz et al., 2001; Xing et al., 2005). In addition, the presence of the draft sequence has also allowed us to ascertain ~90% of the *Alu* elements in the human genome. These elements have been divided into subfamilies, each of which began expanding in the primate genome at different times. For example, the Ye family of *Alu* elements first arose in the common ancestor of all hominoid primates (Salem et al., 2005), whereas the Yb8 subfamily is restricted to gorillas, chimpanzees, and humans (Han et al., 2005).

In the current work, we have taken the presence/absence data sets currently available for analyses of primate phylogenetic and population genetic analysis (as well as a few other appropriate sources) to determine the levels of potentially homoplasious events and additional misleading patterns on the study of primate relationships. Although other taxa that have experienced rapid divergences may in fact be subject to more substantial problems resulting from lineage sorting in particular (Takahashi et al., 2001a, 2001b), we suspect that most phylogenetic groups will respond to SINE analysis much the way primates have—by allowing for the resolution of several previously unresolved relationships. A combination of careful data analysis and adequate sampling of informative loci and taxa should minimize most of the issues raised by the specter of homoplasy.

Lineage Sorting

Lineage sorting is a form of homoplasy introduced by the random fixation or extinction of alternative alleles in distinct lineages. The potential for lineage sorting is an issue whenever researchers deal with species groups that may have had a large ancestral effective population size and/or experienced rapid speciation (Nei, 1987; Pamilo and Nei, 1988; Takahata, 1989). A classic example from the application of SINEs to phylogeny is found in cichlids (Takahashi et al., 2001b) in which 14 of 38 loci appeared to be subject to lineage sorting including an event thought to have occurred as long as 14 million years ago. However, as we will discuss, this high level of lineage sorting is probably not typical of SINE studies, as shown in the studies in primates and other organisms.

The studies of mobile element insertions in primates include several large-scale studies of *Alu* insertion patterns from the hominid (Salem et al., 2003), cercopithecid (Xing et al., 2005), platyrrhine (Ray et al., 2005; Singer et al., 2003), tarsier (Schmitz et al., 2001), and strepsirrhine (Roos et al., 2004) lineages. Of the 131 insertions characterized in the hominid lineage, only one locus could be interpreted as being the product of lineage sorting. Xing et al.'s study of 285 *Alu* insertion loci in Cercopithecidae recovered four putative lineage sorting events. There were no examples found in examinations of 190 and 74 New World monkey insertions, 118 *Alu* insertions used to determine the affiliations of tarsiers, and 61 insertions from an investigation into strepsirrhine (prosimian) phylogeny.

Several other large-scale analyses not directly aimed at phylogenetic analysis can also help us assess the extent of SINE insertion homoplasy due to lineage sorting. Although these studies were aimed at characterizing specific subfamilies of *Alu* elements, the work did involve amplification of loci isolated from the human genome in various nonhuman primates. For example, one study involved the *Alu* Yd3 subfamily (Xing et al., 2003). These authors examined 133 loci in lineages ranging from human to owl monkey and found no evidence of lineage sorting. Two additional similar studies of the *Alu* Ya (Otieno et al., 2004) and Yb (Carter et al., 2004) lineages analyzed another 2672 individual loci and again found no evidence of lineage sorting.

These projects, in conjunction with several others summarized in Table 1, comprise a total of over 11,000 individual *Alu* loci. Of those, only seven suspected cases of lineage sorting were discovered. Thus, it appears that the overall frequency of lineage sorting events in the primate order is small, ~0.0006 events/insertion. We understand that this collection of studies is biased toward the human lineage, and that this may skew that data. We would argue, however, that the relatively rapid divergence between *Homo*, *Pan*, and *Gorilla* (1 to 3 million years between the *Gorilla* divergence and the divergence of *Homo* and *Pan*) suggested by several studies (Goodman et al., 1998; O'Huigin et al., 2002; Satta et al., 2000; Sibley and Ahlquist, 1984) would promote the occurrence of lineage sorting events and the observation that so few were recovered suggests a low rate of occurrence. Thus, as long as sufficient sampling is present at each node and the species being studied do not belong to a group that is prone to lineage sorting, these types of events should not be a serious problem. In fact, the paucity of these events in some taxa and their higher frequency in others can be an indication of speciation dynamics of a lineage as it evolved; see Shedlock et al. (2004) for a recent discussion.

Parallel Insertions

The insertion of SINEs into regions occupied by similar elements in other taxa is another source of potential confusion. In fact, given the observation that some mobile element families appear to accumulate in particular regions of the genome (Greally, 2002; Jurka et al., 2005) as well as exhibit a preference for certain target sequences (Gentles et al., 2005; Jurka, 1997), these sorts of events are not unexpected. Other factors that can influence their rate of occurrence would include the relative divergence time between taxa and relative rates of retrotransposition (Hillis, 1999). Including a wide range of taxa in any application of SINE data to phylogenetic analysis would also be expected to increase the likelihood of observing these events. Parallel insertions of SINE characters can be separated into two types: near-parallel insertions and precise-parallel insertions.

Near-Parallel Insertions

The first and most common type have been termed near-parallel insertions. In these cases, a secondary insertion has occurred near the insertion originally being studied (usually within a 200- to 600-bp amplicon). Given this definition, these sorts of events are technically not instances of homoplasy. However, preliminary analysis using agarose gel electrophoresis of loci at which these events have occurred can be interpreted as homoplasy if more detailed sequence analysis of the loci is not applied. Because the vast majority of the loci in a SINE analysis will not contradict the final version of the tree, the patterns present on a developing cladogram often begin to become clear relatively early in an analysis. Thus, these events are usually easy to detect and sequencing of the locus will resolve the issue. Therefore, if an anomalous pattern is apparent when compared to the overall tree, it becomes clear that this locus should be investigated more closely.

One clear example comes from the work on platyrrhine primate phylogeny (Ray et al., 2005). Of the 190 loci investigated, 11 contained multiple insertions in the amplified regions. Thus, the vast majority of the loci were clearly consistent with the final tree. At one locus, the original *Alu* insertion under investigation was ascertained from the genome of *Saguinus labiatus*, a New World primate. Primers were designed to amplify a ~300-bp empty site (i.e., without an *Alu* insertion) based on comparisons to the orthologous locus in *Homo sapiens*. The filled site (containing an *Alu* insertion) in taxa sharing the insertion through identity-by-descent should have been ~500 to 600 bp depending on the length of

TABLE 1. Instances of misleading and homoplasy inducing events observed during studies of primate phylogeny and population biology using SINEs.

Reference	Number of insertions examined	Lineage sorting suspected	Near-parallel *Alu* insertions	Other near-parallel insertions	Precise-parallel insertions	Targeted taxa or *Alu* subfamily	Number of taxa compared
Ray et al., 2005	190	0	11	0	0	Platyrrhini	15
Xing et al., 2005	285	4	14	0	1?	Cercopithecidae	20
Schmitz et al., 2001 and personal communication	118	0	6	0	0	*Tarsius*	9
Singer et al., 2003	6	0	0	0	0	Platyrrhini	12
Salem et al., 2003	131	1	0	0	0	Hominidae	8
Roos et al., 2004	61	0	0	0	0	Strepsirrhini	22+
Carter et al., 2004	1202	0	1	0	0	Yb[a]	7
Otieno et al., 2004	1470	0	1	1 (ERV[b])	0	Ya[a]	7
Xing et al., 2003	133	0	1	0	0	Yd[a]	9
Han et al., 2005	12	0	0	0	0	Yb[a]	9
Salem et al., 2005	120	0	3	0	0	Ye[a]	13
Hedges et al., 2004	123	2	0	0	0	*Homo-Pan*	5
Roy-Engel et al., 2002	139	0	3	0	0	*Homo*	7
Conley et al., 2005	2	0	0	0	1 (SVA[c])	*Homo*	1
van de Lagemaat et al., 2005	7010	0	1	0	3	*Homo-Pan*	3
Total	11002	7	41	1	5		

[a] Each subfamily indicated here (Ya, Yb, etc.) represents a distinct group of insertions in primate genomes characterized by their own set of diagnostic mutations.
[b] ERV = endogenous retrovirus.
[c] SVA = a composite repetitive element named after its main components, SINE, VNTR, and Alu.

the poly-A tail of the *Alu* element. Examination of the raw data from the agarose gel electrophoresis suggested that members of four genera share the *Alu* insertion—*Macaca, Chlorocebus, Callithrix, Saguinus*. A fifth taxon, *Aotus*, exhibits an anomalous pattern and the remaining taxa in the panel exhibit either the expected empty site or no amplification at all.

This pattern suggests an insertion shared by two catarrhine (*Macaca* and *Chlorocebus*) and two platyrrhine (*Callithrix* and *Saguinus*) primates and should therefore raise eyebrows. It has been well established by morphological and previous DNA sequence–based studies that there is a clear division between Old and New World taxa. Thus, familiarity with basic primate phylogeny or with the patterns already apparent on the developing SINE-derived tree would suggest that this locus needs to be investigated further via sequence analysis of the locus for all taxa. In fact, sequence analysis subsequently revealed that four independent *Alu* insertion events had occurred within the ancestral locus.

Despite the possible confusion generated by loci such as this one, the problems arising from near-parallel insertions should not be a major issue for the observant researcher. The final results of our data compilation indicate that near-parallel insertion events occur at only ~0.0004 events/insertion when considering the 11,000+ loci collected. All of these events are easily resolved by automated DNA sequence analysis, thus they do not contribute negatively to the phylogeny reconstruction. These observations suggest that a basic knowledge of the taxa involved and a sense of the overall picture generated by the majority of loci in a phylogenetic study will lead to a "red flag" at any problematic locus. In other words, as is most often the case, the anomalous amplification pattern is so different from the overall picture that further investigation is the natural next step.

In cases where near-parallel insertions are discovered, how should each insertion be handled with regard to analysis? Because the presence of each *Alu* element represents a unique event in the evolution of the genomes under study, it is clear that each insertion should be treated as an independent bit of information. Thus, we would even argue that when near-parallel insertions are discovered, it is often a fortuitous event because one now has two or more potentially informative insertions found within a single amplicon.

Precise-Parallel Insertions

Confusion may also be introduced into a phylogenetic analysis by the very rare precise-parallel insertions. These events may in fact mimic the earlier case of lineage sorting. Such an event occurs when a second element has inserted into exactly the same target site in a separate taxon, producing a duplication of the accessed target sequence. This event makes it appear that an insertion is shared when in actuality it represents two independent insertion events. There are only a few cases where such events can be clearly delineated in all of the published mobile element literature. In the first instance, an SVA element and an *Alu* element inserted into the same target site within exon 9 of the human BTK gene (Conley et al., 2005). Cantrell et al. (2001) found two loci within a Sigmodontine retrotransposon that were targets for precise-parallel insertions. The third example comes from an analysis of the Felid Y-chromosome (Slattery et al., 2000). In rats, an independent precise-parallel insertion of ID elements has been identified (Rothenburg et al., 2002). *Mus musculus* and *M. pahari* also share parallel insertions of B1 elements into the same target site (Kass et al., 2000). However, sequence analysis of these loci easily shows that these loci have been subjected to multiple

parallel insertions and thus they do not contribute erroneous information for phylogeny reconstruction. Finally, a report by van de Lagemaat et al. (2005) detailing a mechanism for the precise deletion of mobile elements reported three potential instances of precise parallel insertion when comparing humans and chimpanzees.

Our examination of the current primate data revealed only five precise-parallel *Alu* insertions. The results indicate that of the 11,000+ loci examined, precise parallel insertions are exceedingly rare—occurring at only about 0.0005 events/insertion of the examined loci. The issue, therefore, is not to determine if these types of loci are a difficulty to overcome in phylogenetic analysis but rather to distinguish them from lineage sorting and excision events. In both cases, expanding the number of taxa and/or sequencing of the loci provide resolution of the issue. For example, in many cases, the secondary mobile element insertion belongs either to a different subfamily of the same mobile element family as the original insertion or may belong to a completely different family of repetitive elements.

In fact, sequence characteristics exist that distinguish different types of mobile elements in a majority of the instances listed above. Thus, in cases like these, the homoplasy is only apparent. The presence of differing mobile element families or distinguishing features such as characteristic truncations, deletions, or subfamily diagnostic sites make recognizing parallel insertions relatively simple. This is well illustrated by one large-scale study in which over 4800 mobile element loci were examined to elucidate mammalian phylogeny (Bashir et al., 2005). Nearly 2500 of these loci were *Alu* insertions. The authors suggested that 23 instances of insertion homoplasy (precise-parallel insertions) existed among these *Alu* insertion loci. We have manually reanalyzed a portion of Bashir et al.'s raw data and found that the true number is actually 13 precise-parallel insertion events—0.005 events/insertion. The discrepancy between the two sets of results was due to several loci being counted multiple times and a few instances of near-parallel insertions being characterized as precise. Regardless of this problem, the authors were correct in recognizing that many of the precise-parallel insertions discovered were distinguishable based on the *Alu* insertions belonging to distinct subfamilies.

Mobile Element Excision

The potential for the precise removal of mobile elements from the genome has long been contemplated. However, it was previously thought that no mechanism existed to remove these elements in such a way that a site identical to the pre-integration sequence was obtained. Two recent manuscripts have provided evidence that mobile element excision may play a larger role than expected in the evolution of some genomes. In plants, Lenoir et al. (2005) described a rapid turnover of SINEs in Arabidopsis. This suggests a limitation to the SINE method for phylogenetic analysis in these organisms. However, it appeared that the excision of particular elements typically proceeded over several steps and often provided a signature of the removal that could be detected using sequence analysis or high-resolution electrophoresis. A potentially more significant problem was introduced by van de Lagamaat et al. (2005). They proposed a theoretical model for the precise excision of *Alu* elements from human and chimpanzee genomes. Their model suggests that the target site duplications (TSDs) generated upon the insertion of the *Alu* elements act as locations for illegitimate recombination. This development may seem to be a setback for proponents of the idea that SINEs are essentially homoplasy-free markers. However, results of the human-chimpanzee comparisons reveal an extremely low rate of complete precise mobile element excisions, about 0.5% of length polymorphisms (van de Lagemaat et al., 2005). In addition, the model described depends on recombination between the direct repeats that flank *Alu* insertions. After short evolutionary times, the condition described will likely no longer exist. This may help to explain the relative rarity of the observation of precise excisions and its minor impact on phylogenetic studies using SINEs.

One of the major issues involved with these types of events is distinguishing precise deletion events from other occurrences that can mimic them: for example, lineage sorting and precise-parallel insertions. Several examples exist in the data sets from Old World primates (Xing et al., 2005) and human-chimpanzee-gorilla comparisons (Hedges et al., 2004; Salem et al., 2003). Although lineage sorting can be invoked to explain anomalous SINE patterns in taxa that either recently shared a common ancestor (Salem et al., 2003) or shared a common ancestor that was thought to have undergone a rapid speciation in the past (Takahashi et al., 2001b), it is unlikely that the polymorphism would have been retained through several successive speciation events over evolutionarily long periods of time. The most consistent interpretation in such events might be that the mobile element has been precisely deleted in some taxa. When making decisions as to the most likely scenario (deletion or lineage sorting), information on relative divergence times between taxa will have to be employed, but we may never know the correct mechanism. However, even if all of the examples currently reported as lineage sorting events are actually instances of precise excision, the total number of events would still remain low enough that their impact on phylogeny reconstruction would be minimal and, unlike most other genetic systems, it would also be well defined/quantified in a SINE-based analysis.

Paralogous Loci

One final occurrence is important enough to be included in any discussion of potentially misleading events in a SINE-based analysis. Paralogous loci include duplicated regions of the genome at which an insertion may have occurred at one of the duplicates but not the other. Given the relatively high occurrence of segmental

duplications in primate and other genomes (influenced in part by the presence of the SINEs themselves (Bailey et al., 2003)), one might expect this to be an issue in SINE analysis. Ruling out potential paralogs in a phylogenetic analysis requires careful comparison. One potential indicator of paralog amplification is the consistent amplification of two bands in all samples from a single species. A single individual with a filled and empty site could be interpreted as a heterozygote, but if all individuals show the same pattern, it should be seen as a potential case of paralogous loci. If there are functional versus nonfunctional open reading frames (ORFs) in the suspect sequences, the paralogs can also be relatively easily detected. To date, only two problematic paralogs have been identified (Luis et al., 2003). Unfortunately, this aspect of SINE analysis has not been adequately addressed and may be a fertile area for future investigations.

Dealing with Potentially Misleading Events

As described above, it is clear that in most cases the vast majority of the SINE data collected for phylogeny reconstruction will be internally consistent. Thus, instead of focusing on whether or not SINE analysis is effectively free of homoplasy, the real task is to determine how best to identify and deal with potentially confounding events. Identification of the events is the most problematic issue. In small studies (<300 loci examined), identifying potential homoplasy is relatively easy; however, at larger scales the number of dates to be examined becomes problematic. For example, several authors have developed purely computational approaches to identify informative insertion patterns for phylogeny construction (Schwartz et al., 2003; Thomas et al., 2003). We believe this is an excellent idea given the large amount of sequence data currently and soon to be available. However, using these approaches it is important to incorporate methods to identify potentially misleading events. In one example, a study incorporated the data available from the NISC Comparative Vertebrate Sequencing program to elucidate mammalian phylogeny (Bashir et al., 2005). Although the final topology is undoubtedly correct, several problems in the data analysis need to be addressed in future studies using this type of approach. For example, in some nonprimate comparisons, the authors did not realize that they compared DNA transposons with RNA transposons at (nearly) the same loci as shown in their figure 3. Furthermore, as discussed above, several loci were counted multiple times, potentially inflating the support for the affected nodes. Such incomplete evaluations make it very difficult to convince skeptics of the advantages of presence/absence data. It also serves to emphasize the need for careful interpretation of the data and for sequencing of SINE insertion loci in a phylogenetic analysis.

Once potentially misleading events have been identified, the most important next step is to obtain enough loci to support each node so that these events can be either ruled out or recognized for what they are—the occasional rare confounding factor. Waddel et al. (2001) proposed a likelihood method for evaluating support for nodes defined by SINE data. Their results suggest that the minimum number of non-contradictory loci required for a significant level of support ($P = 0.037$) at any node is three. Even if there is an anomalous locus that is not easily explained by near-parallel insertion or precise-parallel insertion of a different mobile element family, additional loci at the node can produce significant levels of support. For example, if one locus appears to have been subjected to a precise deletion, four noncontradictory loci will still produce a P-value of 0.045. Therefore, although distinguishing between the three potentially homoplasy-inducing events (e.g., lineage sorting, precise mobile element deletion, and precise-parallel insertion) that can most closely resemble one another is one of the more difficult problems at present, their overall effects are minimal as long as taxonomic sampling is adequate and careful consideration of the developing data is used.

Despite the possibility of several confounding events that may occur to disrupt the interpretation of a SINE-derived cladogram, it remains clear that SINE-based markers are some of the most powerful phylogenetic tools available. In the data we examined, none of the events that would be considered troublesome occurred at a high enough rate to be considered a significant problem. Near-parallel insertion events, which are the most easily resolved by sequencing efforts, were the most common but only noted in 41 of 11,000+ loci. The other potentially confounding events are more difficult to resolve but are also less common and therefore less likely to be encountered.

In conclusion, basic cautions including careful analysis and interpretation of the data can limit the potential impact of misleading events. In addition, as with any study of phylogeny, a basic knowledge of the taxa under investigation is a must in order to avoid being confused by the rare confounding locus. Complete taxon sampling and collection of sufficient numbers of informative insertion events are also basic requirements for any SINE based phylogenetic reconstruction project. We believe that most researchers follow these guidelines as a normal course of events. It is our hope that the information presented here will encourage researchers to consider the utility of the SINE method of phylogeny reconstruction for their organisms.

ACKNOWLEDGMENTS

We would like to thank L. van de Lagemaat and D. L. Mager for their communication clarifying the details of their mobile element deletion data. We would also like to thank A. Bashir and V. Bafna for providing us with their raw data. This research was supported by National Science Foundation BCS-0218338 (M.A.B.) and EPS-0346411 (M.A.B.); Louisiana Board of Regents Millennium Trust Health Excellence Fund HEF (2000-05)-05 (M.A.B.), (2000-05)-01 (M.A.B.), and (2001-06)-02 (M.A.B.); National Institutes of Health R01 GM59290 (M.A.B.); and the State of Louisiana Board of Regents Support Fund (M.A.B.). Funding was also provided by the Eberly College of Arts and Sciences at West Virginia University (D.A.R) and by the West Virginia University Research Corporation (D.A.R).

REFERENCES

Bailey, J. A., G. Liu, and E. E. Eichler. 2003. An Alu transposition model for the origin and expansion of human segmental duplications. Am. J. Hum. Genet. 73:823–834.

Bamshad, M. J., S. Wooding, W. S. Watkins, C. T. Ostler, M. A. Batzer, and L. B. Jorde. 2003. Human population genetic structure and inference of group membership. A. J. Hum. Genet. 72:578–589.

Bashir, A., C. Ye, A. L. Price, and V. Bafna. 2005. Orthologous repeats and mammalian phylogenetic inference. Genome Res. 15:998–1006.

Batzer, M. A., and P. L. Deininger. 2002. *Alu* repeats and human genomic diversity. Nat. Rev. Genet. 3:370–379.

Batzer, M. A., M. Stoneking, M. Alegria-Hartman, H. Bazan, D. H. Kass, T. H. Shaikh, G. E. Novick, P. A. Ioannou, W. D. Scheer, R. J. Herrera, et al. 1994. African origin of human-specific polymorphic *Alu* insertions. Proc. Natl. Acad. Sci. USA 91:12288–12292.

Cantrell, M. A., B. J. Filanoski, A. R. Ingermann, K. Olsson, N. DiLuglio, Z. Lister, and H. A. Wichman. 2001. An ancient retrovirus-like element contains hot spots for SINE insertion. Genetics 158:769–777.

Carroll, M. L., A. M. Roy-Engel, S. V. Nguyen, A. H. Salem, E. Vogel, B. Vincent, J. Myers, Z. Ahmad, L. Nguyen, M. Sammarco, W. S. Watkins, J. Henke, W. Makalowski, L. B. Jorde, P. L. Deininger, and M. A. Batzer. 2001. Large-scale analysis of the *Alu* Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. J. Mol. Biol. 311:17–40.

Carter, A. B., A. H. Salem, D. J. Hedges, C. N. Keegan, B. Kimball, J. A. Walker, W. S. Watkins, L. B. Jorde, and M. A. Batzer. 2004. Genome-wide analysis of the human Alu Yb-lineage. Hum. Genomics 1:167–178.

Churakov, G., A. F. A. Smit, J. Brosius, and J. Schmitz. 2005. A novel abundant family of retroposed elements (DAS-SINEs) in the nine-banded Armadillo (Dasypus novemcinctus). Mol. Biol. Evol. 22:886–893.

Conley, M. E., J. D. Partain, S. M. Norland, S. A. Shurtleff, and H. H. Kazazian, Jr. 2005. Two independent retrotransposon insertions at the same site within the coding region of BTK. Hum. Mutat. 25:324–325.

Cotrim, N. H., M. T. Auricchio, J. P. Vicente, P. A. Otto, and R. C. Mingroni-Netto. 2004. Polymorphic *Alu* insertions in six Brazilian African-derived populations. Am. J. Hum. Genet. 16:264–277.

Flower, W. H. 1883. On the arrangement of the orders and families of existing Mammalia. Proc. Zool. Soc. Lond. 1883:178–186.

Gatesy, J., M. Milinkovitch, V. Waddell, and M. Stanhope. 1999. Stability of cladistic relationships between Cetacea and higher-level artiodactyl taxa. Syst. Biol. 48:6–20.

Gentles, A. J., O. Kohany, and J. Jurka. 2005. Evolutionary diversity and potential recombinogenic role of integration targets of non-LTR retrotransposons. Mol. Biol. Evol. 22:1983–1991.

Goodman, M., C. A. Porter, J. Czelusniak, S. L. Page, H. Schneider, J. Shoshani, G. Gunnell, and C. P. Groves. 1998. Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. Mol. Phylogenet. Evol. 9:585–598.

Greally, J. M. 2002. Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. Proc. Nati. Acad. Sci. USA 99:327–332.

Han, K., J. Xing, H. Wang, D. J. Hedges, R. K. Garber, R. Cordaux, and M. A. Batzer. 2005. Under the genomic radar: The stealth model of Alu amplification. Genome Res. 15:655–664.

Hedges, D. J., P. A. Callinan, R. Cordaux, J. Xing, E. Barnes, and M. A. Batzer. 2004. Differential *Alu* mobilization and polymorphism among the human and chimpanzee lineages. Genome Res. 14:1068–1075.

Hillis, D. M. 1999. SINEs of the perfect character. Proc. Nat. Acad. Sci. USA 96:9979–9981.

Jurka, J. 1997. Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. Proc. Nat. Acad. Sci. USA 94:1872–1877.

Jurka, J., O. Kohany, A. Pavlicek, V. V. Kapitonov, and M. V. Jurka. 2005. Clustering, duplication and chromosomal distribution of mouse SINE retrotransposons. Cytogenet. Genome Res. 110:117–123.

Kass, D. H., M. E. Raynor, and T. M. Williams. 2000. Evolutionary history of B1 retroposons in the genus *Mus*. J. Mol. Evol. 51:256–264.

Kawai, K., M. Nikaido, M. Harada, S. Matsumura, L. K. Lin, Y. Wu, M. Hasegawa, and N. Okada. 2002. Intra- and interfamily relationships of Vespertilionidae inferred by various molecular markers including SINE insertion data. J. Mol. Evol. 55:284–301.

Lenoir, A., T. Pelissier, C. Bousquet-Antonelli, and J. M. Deragon. 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.

Luis, J. R., M. C. Terreros, L. Martinez, D. Rojas, and R. J. Herrera. 2003. Two problematic human polymorphic Alu insertions. Electrophoresis 24:2290–2294.

Minghetti, P. P., and A. Dugaiczyk. 1993. The emergence of new DNA repeats and the divergence of primates. Proc. Nat. Acad. Sci. USA 90:1872–1876.

Murata, S., N. Takasaki, T. Okazaki, T. Kobayashi, K. Numachi, K.-H. Chang, and N. Okada. 1998. Molecular evidence from short interspersed elements (SINEs) that *Onchorhynchus masou* (cherry salmon) is monophyletic. Can. J. Fish. Aqua. Sci. 55:1864–1870.

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

Murata, S., N. Takasaki, M. Saitoh, H. Tachida, and N. Okada. 1996. Details of retropositional genome dynamics that provide a rationale for a generic division: The distinct branching of all the pacific salmon and trout (*Oncorhynchus*) from the Atlantic salmon and trout (*Salmo*). Genetics 142:915–926.

Nasidze, I., G. M. Risch, M. Robichaux, S. T. Sherry, M. A. Batzer, and M. Stoneking. 2001. *Alu* insertion polymorphisms and the genetic structure of human populations from the Caucasus. Eur. J. Hum. Genet. 9:267–272.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

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

Nishihara, H., Y. Terai, and N. Okada. 2002. Characterization of novel Alu- and tRNA-related SINEs from the tree shrew and evolutionary implications of their origins. Mol. Biol. Evol. 19:1964–1972.

O'Huigin, C., Y. Satta, N. Takahata, and J. Klein. 2002. Contribution of homoplasy and of ancestral polymorphism to the evolution of genes in anthropoid primates. Mol. Biol. Evol. 19:1501–1513.

Otieno, A. C., A. B. Carter, D. J. Hedges, J. A. Walker, D. A. Ray, R. K. Garber, B. A. Anders, N. Stoilova, M. E. Laborde, J. D. Fowlkes, C. H. Huang, B. Perodeau, and M. A. Batzer. 2004. Analysis of the human Alu Ya-lineage. J. Mol. Biol. 342:109–118.

Pamilo, P., and M. Nei. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5:568–583.

Perna, N. T., M. A. Batzer, P. L. Deininger, and M. Stoneking. 1992. *Alu* insertion polymorphism: A new type of marker for human population studies. Hum. Biol. 64:641–648.

Ray, D. A., J. Xing, D. J. Hedges, M. A. Hall, M. E. Laborde, B. A. Anders, B. R. White, N. Stoilova, J. D. Fowlkes, K. E. Landry, L. G. Chemnick, O. A. Ryder, and M. A. Batzer. 2005. *Alu* insertion loci and platyrrhine primate phylogeny. Mol. Phylogenet. Evol. 35:117–126.

Roos, C., J. Schmitz, and H. Zischler. 2004. Primate jumping genes elucidate strepsirrhine phylogeny. Proc. Natl. Acad. Sci. USA 101:10650–10654.

Rothenburg, S., M. Eiben, F. Koch-Nolte, and F. Haag. 2002. Independent integration of rodent identifier (ID) elements into orthologous sites of some RT6 alleles of *Rattus norvegicus* and *Rattus rattus*. J. Mol. Evol. 55:251–259.

Roy-Engel, A. M., M. L. Carroll, M. El-Sawy, A. H. Salem, R. K. Garber, S. V. Nguyen, P. L. Deininger, and M. A. Batzer. 2002. Non-traditional *Alu* evolution and primate genomic diversity. J. Mol. Biol. 316:1033–1040.

Ryan, S. C., and A. Dugaiczyk. 1989. Newly arisen DNA repeats in primate phylogeny. Proc. Nat. Acad. Sci. USA 86:9360–9364.

Salem, A. H., D. A. Ray, D. J. Hedges, J. Jurka, and M. A. Batzer. 2005. Analysis of the human *Alu* Ye lineage. BMC Evol. Biol. 5:18.

Salem, A. H., D. A. Ray, J. Xing, P. A. Callinan, J. S. Myers, D. J. Hedges, R. K. Garber, D. J. Witherspoon, L. B. Jorde, and M. A. Batzer. 2003. *Alu* elements and hominid phylogenetics. Proc. Nat. Acad. Sci. USA 100:12787–12791.

Sarich, V. M. 1985. Rodent macromolecular systematics. Pages 423–452 *in* Evolutionary relationships among rodents: A multidisciplinary analysis (W. P. Luckett, and J.-L. Hartenberger, eds.). Springer-Verlag, Berlin.

Sasaki, T., K. Takahashi, M. Nikaido, S. Miura, Y. Yasukawa, and N. Okada. 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.

Satta, Y., J. Klein, and N. Takahata. 2000. DNA archives and our nearest relative: The trichotomy problem revisited. Mol. Phylogenet. Evol. 14:259–275.

Schmitz, J., M. Ohme, and H. Zischler. 2001. SINE insertions in cladistic analyses and the phylogenetic affiliations of *Tarsius bancanus* to other primates. Genetics 157:777–784.

Schmitz, J., C. Roos, and H. Zischler. 2005. Primate phylogeny: Molecular evidence from retroposons. Cytogenet. Genome Res. 108:26–37.

Schmitz, J., and H. Zischler. 2003. A novel family of tRNA-derived SINEs in the colugo and two new retrotransposable markers separating Dermopterans from Primates. Mol. Phylogenet. Evol. 28:341–349.

Schwartz, S., L. Elnitski, M. Li, M. Weirauch, C. Riemer, A. Smit, E. D. Green, R. C. Hardison, and W. Miller. 2003. MultiPipMaker and supporting tools: Alignments and analysis of multiple genomic DNA sequences. Nucleic Acids Res. 31:3518–3524.

Shedlock, A. M., and N. Okada. 2000. SINE insertions: Powerful tools for molecular systematics. Bioessays 22:148–160.

Shedlock, A. M., K. Takahashi, and N. Okada. 2004. SINEs of speciation: Tracking lineages with retroposons. Trends Ecol. Evol. 19:545–553.

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

Sibley, C. G., and J. E. Ahlquist. 1984. The phylogeny of the hominoid primates, as indicated by DNA-DNA hybridization. J. Mol. Evol. 20:2–15.

Singer, S. S., J. Schmitz, C. Schwiegk, and H. Zischler. 2003. Molecular cladistic markers in New World monkey phylogeny (Platyrrhini, Primates). Mol. Phylogenet. Evol. 26:490–501.

Slattery, J. P., W. J. Murphy, and S. J. O'Brien. 2000. Patterns of diversity among SINE elements isolated from three Y-chromosome genes in carnivores. Mol. Biol. Evol. 17:825–829.

Takahashi, K., M. Nishida, M. Yuma, and N. Okada. 2001a. Retroposition of the AFC family of SINEs (short interspersed repetitive elements) before and during the adaptive radiation of cichlid fishes in Lake Malawi and related inferences about phylogeny. J. Mol. Evol. 53:496–507.

Takahashi, K., Y. Terai, M. Nishida, and N. Okada. 2001b. Phylogenetic relationships and ancient incomplete lineage sorting among cichlid fishes in Lake Tanganyika as revealed by analysis of the insertion of retroposons. Mol. Biol. and Evol. 18:2057–2066.

Takahata, N. 1989. Gene genealogy in three related populations: Consistency probability between gene and population trees. Genetics 122:957–966.

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

Terai, Y., N. Takezaki, W. E. Mayer, H. Tichy, N. Takahata, J. Klein, and N. Okada. 2004. Phylogenetic relationships among East African haplochromine fish as revealed by short interspersed elements (SINEs). J. Mol. Evol. 58:64–78.

Thomas, J. W., J. W. Touchman, R. W. Blakesley, G. G. Bouffard, S. M. Beckstrom-Sternberg, E. H. Margulies, M. Blanchette, A. C. Siepel, P. J. Thomas, J. C. McDowell, B. Maskeri, N. F. Hansen, M. S. Schwartz, R. J. Weber, W. J. Kent, D. Karolchik, T. C. Bruen, R. Bevan, D. J. Cutler, S. Schwartz, L. Elnitski, J. R. Idol, A. B. Prasad, S. Q. Lee-Lin, V. V. Maduro, T. J. Summers, M. E. Portnoy, N. L. Dietrich, N. Akhter, K. Ayele, B. Benjamin, K. Cariaga, C. P. Brinkley, S. Y. Brooks, S. Granite, X. Guan, J. Gupta, P. Haghighi, S. L. Ho, M. C. Huang, E. Karlins, P. L. Laric, R. Legaspi, M. J. Lim, Q. L. Maduro, C. A. Masiello, S. D. Mastrian, J. C. McCloskey, R. Pearson, S. Stantripop, E. E. Tiongson, J. T. Tran, C. Tsurgeon, J. L. Vogt, M. A. Walker, K. D. Wetherby, L. S. Wiggins, A. C. Young, L. H. Zhang, K. Osoegawa, B. Zhu, B. Zhao, C. L. Shu, P. J. De Jong, C. E. Lawrence, A. F. Smit, A. Chakravarti, D. Haussler, P. Green, W. Miller, and E. D. Green. 2003. Comparative analyses of multi-species sequences from targeted genomic regions. Nature 424:788–793.

van de Lagemaat, L. N., L. Gagnier, P. Medstrand, and D. L. Mager. 2005. Genomic deletions and precise removal of transposable elements mediated by short identical DNA segments in primates. Genome Res. 15:1243–1249.

Waddell, P. J., H. Kishino, and R. Ota. 2001. A phylogenetic foundation for comparative mammalian genomics. Genome Inform. Ser. 12:141–154.

Watkins, W. S., A. R. Rogers, C. T. Ostler, S. Wooding, M. J. Bamshad, A. M. Brassington, M. L. Carroll, S. V. Nguyen, J. A. Walker, B. V. Prasad, P. G. Reddy, P. K. Das, M. A. Batzer, and L. B. Jorde. 2003. Genetic variation among world populations: Inferences from 100 *Alu* insertion polymorphisms. Genome Res. 13:1607–1618.

Xing, J., A. H. Salem, D. J. Hedges, G. E. Kilroy, W. S. Watkins, J. E. Schienman, C. B. Stewart, J. Jurka, L. B. Jorde, and M. A. Batzer. 2003. Comprehensive analysis of two *Alu* Yd subfamilies. J. Mol. Evol. 57:S76–S89.

Xing, J., H. Wang, K. Han, D. A. Ray, C. H. Huang, L. G. Chemnick, C. B. Stewart, T. R. Disotell, O. A. Ryder, and M. A. Batzer. 2005. A mobile element based phylogeny of Old World monkeys. Mol. Phylogenet. Evol. 37:872–880.

Zampicinini, G., A. Blinov, P. Cervella, V. Guryev, and G. Sella. 2004. Insertional polymorphism of a non-LTR mobile element (NLRCth1) in European populations of *Chironomus riparius* (Diptera, Chironomidae) as detected by transposon insertion display. Genome 47:1154–1163.

First submitted 4 November 2005; reviews returned 31 December 2005; final acceptance 15 March 2006
Associate Editor: Andrew Shedlock