Syst. Biol. 68(3):482–493, 2019

DOI:10.1093/sysbio/syy076
Advance Access publication November 15, 2018

True Homoplasy of Retrotransposon Insertions in Primates

LILIYA DORONINA[1,*], OLGA REISING[1], HIRAM CLAWSON[2], DAVID A. RAY[3], AND JÜRGEN SCHMITZ[1]

[1]*Institute of Experimental Pathology (ZMBE), University of Münster, Von-Esmarch-Str. 56, D-48149 Münster, Germany;* [2]*Department of Biomolecular Engineering, University of California, 1156 High Street, Santa Cruz, CA, USA; and* [3]*Department of Biological Sciences, Texas Tech University, 2901 Main Street, Lubbock, TX, USA*

[*]*Correspondence to be sent to: Institute of Experimental Pathology (ZMBE), University of Münster, Von-Esmarch-Str. 56, D-48149 Münster, Germany; E-mail: doronina@uni-muenster.de.*

Received 17 August 2018; reviews returned 05 November 2018; accepted 13 November 2018
Associate Editor: Edward Susko

Abstract.—How reliable *are* the presence/absence insertion patterns of the supposedly homoplasy-free retrotransposons, which were randomly inserted in the *quasi* infinite genomic space? To systematically examine this question in an up-to-date, multigenome comparison, we screened millions of primate transposed *Alu* SINE elements for incidences of homoplasious precise insertions and deletions. In genome-wide analyses, we identified and manually verified nine cases of precise parallel *Alu* insertions of apparently identical elements at orthologous positions in two ape lineages and twelve incidences of precise deletions of previously established SINEs. Correspondingly, eight precise parallel insertions and no exact deletions were detected in a comparison of lemuriform primate and human insertions spanning the range of primate diversity. With an overall frequency of homoplasious *Alu* insertions of only 0.01% (for human–chimpanzee–rhesus macaque) and 0.02–0.04% (for human–bushbaby–lemurs) and precise *Alu* deletions of 0.001–0.002% (for human–chimpanzee–rhesus macaque), real homoplasy is not considered to be a quantitatively relevant source of evolutionary noise. Thus, presence/absence patterns of *Alu* retrotransposons and, presumably, all LINE1-mobilized elements represent indeed the virtually homoplasy-free markers they are considered to be. Therefore, ancestral incomplete lineage sorting and hybridization remain the only serious sources of conflicting presence/absence patterns of retrotransposon insertions, and as such are detectable and quantifiable. [Homoplasy; precise deletions; precise parallel insertions; primates; retrotransposons.]

Homoplasy is the independent occurrence of identical characteristics or traits in two or more species that were absent in their common ancestor (convergence, parallelism) or deletion of an inherited trait in one of the descendent lineages (reversal). Homoplasies are present in many taxa and at different organizational levels (e.g., molecular, epigenetic, developmental, and phenotypic). During the last decade, the phylogenetic relationships among many taxa were reliably established, piquing an interest in homoplastic traits and their underlying mechanisms of emergence. Most often, homoplasies, especially those leading to phenotypical changes, were exposed coincidentally rather than via systematic searches (Wake et al. 2011). By contrast, the majority of homoplasies occurring at the molecular level do not show phenotypic expression or selective relevance, and therefore, occur more often and provide a good substance for quantitative estimations.

Retrotransposons are mobile genetic elements that propagate in genomes by "copy-and-paste" processes via an RNA intermediate. They are abundant in all eukaryotes and occupy around 42% of the human (*Homo sapiens*) genome (Lander et al. 2001). After random insertion of a retrotransposon in a germline genomic locus and fixation in the prevalent population, a retrotransposon is inherited to all descendants, and its presence-state at the orthologous genomic position of modern lineages is commonly assumed to indicate their common ancestry. Almost three decades ago, Ryan and Dugaiczyk (1989) proposed the use of presence/absence patterns of retrotransposons (namely, *Alu* elements) for phylogenetic reconstructions of primates. The pioneering retrophylogenetic studies in fishes (Murata et al. 1993, 1996; Takahashi et al. 1998), cetariodactyls (Shimamura et al. 1997; Nikaido et al. 1999), and primates (Hamdi et al. 1999; Schmitz et al. 2001), as well as *Alu*-based population genetic studies in human (Perna et al. 1992; Stoneking et al. 1997; Roy-Engel et al. 2001) inspired early interest in retrotransposons as clade-specific markers. Currently, informative data for retrotransposon presence/absence patterns in phylo- and population genetics are accumulating as large-scale by-products of the pioneering Earth BioGenome Project that aims to cover all of eukaryotic biodiversity (Lewin et al. 2018). Automated multigenome extraction and analysis tools make it possible to filter and visualize their unique, assorting signals (Noll et al. 2015). Once glorified as the perfect characters for phylogenetic analyses, it is also recognized that they are not immune to problems associated with incomplete lineage sorting (ILS; also known as hemiplasy [Avise and Robinson 2008]) and even, albeit rarely, to homoplasy (Hillis 1999). To counter such problems in quantitative phylogenetic analyses, statistical tests were developed to evaluate the reliability of presence/absence markers (Waddell et al. 2001; Kuritzin et al. 2016).

ILS, as a phenomenon of evolution itself, affects all phylogenetic marker systems equally. Its presence and influence are most significant during periods of rapid speciation, represented by short internodes in phylogenetic trees, when the times between speciation events are too short for polymorphic markers to become fixed. To resolve the phylogenetic relationships in such rapidly radiating clades it is especially crucial to use a homoplasy-free marker system to distinguish phylogenetically informative signals from

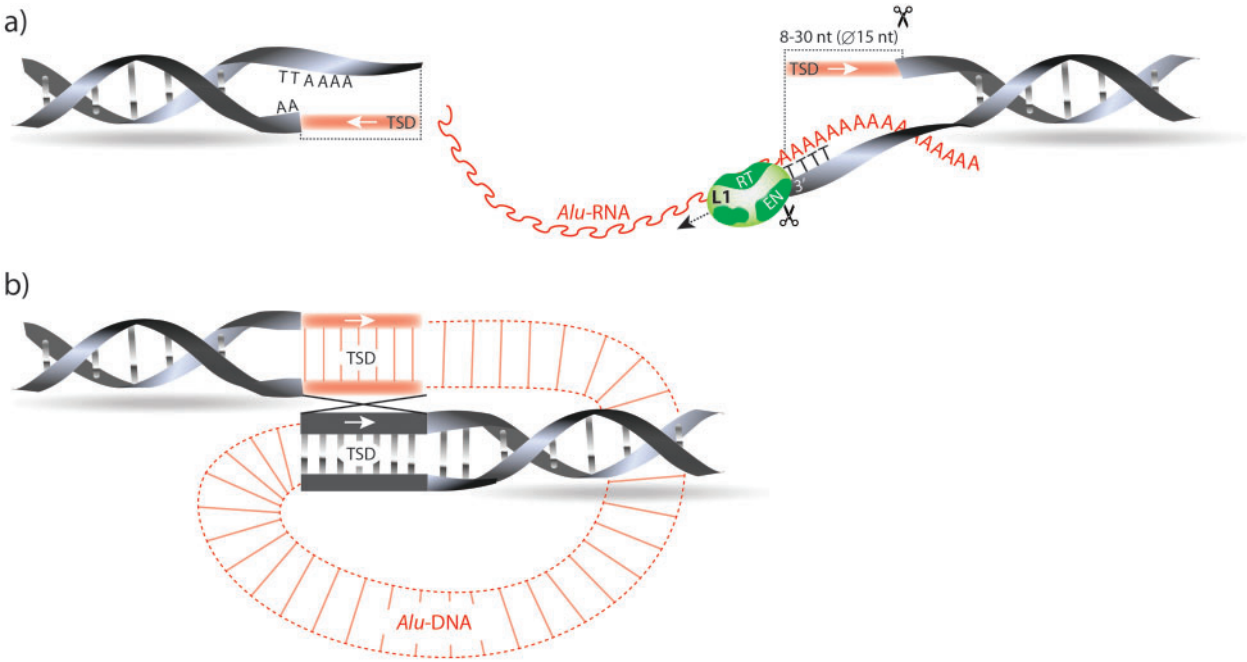

FIGURE 1. The two sources of retrotransposon presence/absence homoplasy. a) Target-primed reverse transcription of LINE1 (L1)-mobilized insertions exhibit a high level of uniqueness, but do not completely exclude the possibility of parallel insertions into orthologous genomic loci in different lineages. The endonuclease activity of the reverse transcriptase introduces shifted nicks in the target DNA. A new 3′-OH end is then used as a primer for the reverse transcriptase activity, and an RNA (here *Alu* RNA) as the template for reverse transcription. On both sites of the inserted retrotransposon, the single strand gaps are filled by the host repair system (red rectangles) to form complete TSDs flanking the element. RT = reverse transcriptase; EN = endonuclease. When a precise parallel insertion into an orthologous locus involved the same element type, we diagnosed real homoplasy. b) Precise retrotransposon deletion via illegitimate recombination is permitted by the TSDs. Deletion of the entire retrotransposed element and one of the two TSDs (red part) leaves behind no trace of the previously inserted element and can be diagnosed as real homoplasy. Arrows indicate the original orientation of the TSDs. The crossed black lines indicate the recombination areas.

those arising from ancestral ILS. Another potential source of evolutionary noise leading to the occurrence of discordant signals might be ancestral hybridization (e.g., Kuritzin et al. 2016). Studies suggest that the presence/absence patterns of retrotransposed elements contain negligible homoplasious noise (Van de Lagemaat et al. 2005; Ray et al. 2006) and are therefore, projected to have the exceptional power to detect and investigate phylogenomic zones of ILS and/or ancestral hybridization (Doronina et al. 2015, 2017; Suh et al. 2015; Kuritzin et al. 2016). However, it should be noted that cases of homoplasy in retrotransposon presence/absence patterns have been reported. For example, Slattery et al. (2000) and Pecon-Slattery et al. (2004) found homoplasious, parallel insertions in felids (retrotransposons present in several closely related *Felis* species and independently inserted in bobcat [*Lynx rufus*]). Doronina et al. (2015) described three cases of precise parallel insertions in arctoid species. Therefore, the question arises: in quantitative analyses of retrotransposon presence/absence patterns in large-scale genome data sets, to what extent is the retrotransposon marker system homoplasy-free? In other words, how often do retrotransposons independently insert into exactly the same orthologous locus in different lineages, falsely implying common ancestry, and how often do precise excisions of previously inserted retrotransposons occur, leading to the false exclusion of a lineage from a monophyletic group?

Short interspersed elements (SINEs) represent the most commonly used transposed element-based phylogenetic marker system applied to elucidate intraordinal lineage affiliations, whereby shared genomic insertions are diagnostic for shared ancestry. SINEs are derived from tRNA, 7SL RNA, or 5S RNA, small, nonprotein coding, housekeeping sequences. They are nonautonomous and coretropose via autonomous long interspersed elements (LINEs) that supply the protein-coding equipment for reverse transcription and transinsertion. Among many other nonautonomous elements, LINE1-mobilized SINEs are integrated in a process known as LINE1 target-primed reverse transcription (TPRT; Fig. 1a). Integration takes place at kinkable DNA sites that include a TT/AAAA consensus recognition pattern, with excision occurring between the TT and AAAA (Jurka 1997). Similar to many restriction enzymes, a pair of shifted nicks is placed in the double helix at an internick interval of 8–30 nt (Luan et al. 1993), producing approximately 15-nt-long, target site duplications (TSDs). TSDs represent a hallmark of retrotransposition and are important indicators of orthology among different lineages.

To date, only two attempts were published to estimate the frequency of homoplasy for transposable element

(TE) insertions. In their review, Ray et al. (2006) evaluated the level of precise parallel insertions of TEs in primates and found a rate of 0.0005–0.005 events/insertion. However, it should be noted that in this study mostly the precise parallel insertions of obviously different families/subfamilies of elements were taken into consideration with a view to distinguish them from the apparently confounding signals arising from ILS. However, strictly speaking they do not represent homoplasy. Van de Lagemaat et al. (2005) focused on precise deletions and provided the first and currently, to our knowledge, the only systematic screening of genome sequence information for homoplasy. They also proposed a mechanism of illegitimate recombination between 10 nt and 20 nt (not less than 10 nt) of TSDs (Fig. 1b) leading to exact deletions without visible traces of the preceding integration in a human–chimpanzee (*Pan troglodytes*)-rhesus macaque (*Macaca mulatta*) comparison of genomes and genomic trace data. It should be mentioned, that initially perfect TSDs accumulate mutations over time that might soon impair their ability to recombine. In a partial draft genomic comparison of human–chimpanzee–rhesus macaque, Van de Lagemaat et al. (2005) estimated that in 0.5–1% of retrotransposon insertion loci, the elements had been precisely deleted. They rightly considered that the phylogenetic relationships among these three species are well established and that ILS did not produce interfering phylogenetic signals in this group. However, at that time with only a restricted number of primate genomes having been sequenced, they could not test for incidences of precise parallel insertions in their proposed precise deletion cases.

By contrast, we present a systematic screening for retrotransposon homoplasies, including both parallel insertions and precise deletions. To detect, classify, and quantify cases of homoplasies of retrotransposed elements we analyzed genomes and multiway genome alignments of relatively closely as well as distantly related primate lineages in which ILS/ancestral hybridization was not considered to have occurred and so was unlikely to influence the results. To screen orthologous transposed elements and their flanking TSDs for homoplasious signals and compute their frequencies we applied the genome presence/absence compiler (GPAC, Noll et al. 2015) using RepeatMasker coordinates of millions of *Alu* elements and their orthologous representations in multiway alignments of primate lineages.

Materials and Methods

Human–Chimpanzee–Rhesus Macaque Comparison

The human–chimpanzee–rhesus macaque (HCR) group provides a good model for investigating the frequency of retrotransposon homoplasies because the highly significant evidence establishing the phylogenetic relationships among these lineages, with rhesus macaque as first divergent group (e.g., Perelman et al. 2011), is not disputed. We then focused our investigation on LINE1-dependent monomeric and dimeric SINE elements. Primate-specific SINEs evolved from monomers (FAM, FLAM, FRAM) to the highly efficiently propagating dimeric *Alu* elements (Kriegs et al. 2007; Kramerov and Vassetzky 2011). The oldest group of *Alu* dimers, *Alu*J elements were active in early primates until the Tarsiiformes divergence, intermediate *Alu*S elements were mainly active in the common Simiiformes ancestor, and the youngest group *Alu*Y mobilized in the Catarrhini ancestor; while some *Alu*Y subfamilies are still active in modern species (Churakov et al. 2010; Konkel et al. 2015). We extracted the element coordinates for all families of di- and monomeric *Alu*s from the RepeatMasker reports for human and chimpanzee (http://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/ and http://hgdownload.soe.ucsc.edu/goldenPath/panTro5/bigZips/, respectively; Fig. 2). The rhesus macaque genome was used as a reference in our homoplasy screenings (loci representing the presence state in the rhesus macaque genome were searched); therefore, we extracted all families of *Alu*s except rhesus macaque-specific *Alu*YRs (Han et al. 2007) from the RepeatMasker report (http://hgdownload.soe.ucsc.edu/goldenPath/rheMac8/bigZips/). Then we used the GPAC (Noll et al. 2015; http://retrogenomics.uni-muenster.de/tools/gpac) to visualize presence/absence patterns of *Alu* elements in genomes for all possible relationships among HCR that contradict the well-established phylogenetic relationships of primates (e.g., Perelman et al. 2011) (HCR +−+ and −++ vs. ++−; where + and − are the presence and absence states, respectively, of an element). To calculate the frequency of homoplasies (see below), we also performed GPAC screenings for all other possible HCR presence/absence patterns. We embedded the public primate 30-way multigenome alignment with human as the reference species (Santa Cruz UCSC Genome Browser; http://hgdownload.soe.ucsc.edu/goldenPath/hg38/multiz30way/) in GPAC for homoplasy screening HCR (+−+), as well as for HCR (+++), (++−), (+−−) screenings. For HCR (−++) and (−+−) screenings, we further embedded the 30-way-compatible 2-way genome alignments of chimpanzee–human (http://hgdownload.soe.ucsc.edu/goldenPath/panTro5/vsHg38/) and of chimpanzee–rhesus macaque (http://hgdownload.soe.ucsc.edu/goldenPath/panTro5/vsRheMac8/) into GPAC. For HCR (−−+) screening, we embedded 2-way genome alignments of rhesus macaque–human (http://hgdownload.soe.ucsc.edu/goldenPath/rheMac8/vsHg38/) and of rhesus macaque–chimpanzee (http://hgdownload.soe.ucsc.edu/goldenPath/rheMac8/vsPanTro5/). Using GPAC screening, we collected all clear *perfect* presence/absence patterns (+ or − in GPAC; see Noll et al. 2015 for criteria) for HCR, and to avoid misalignments extracted only *Alu* elements from the genome of human, chimpanzee, and rhesus macaque

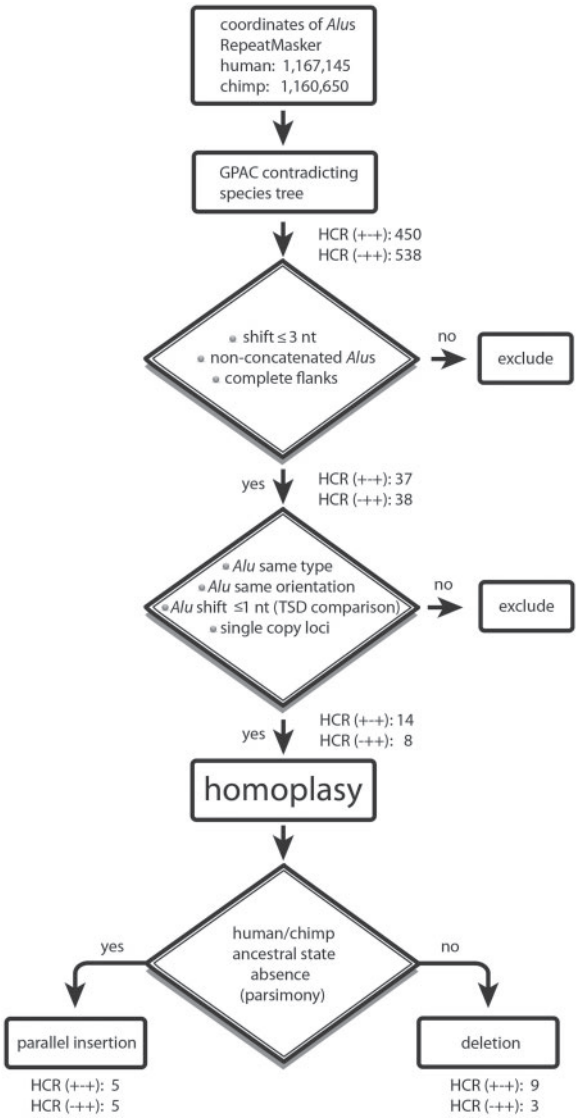

FIGURE 2. A flowchart for detecting homoplasious *Alu* SINEs in human and chimpanzee. Starting from the RepeatMasker genome coordinates of *Alu*s, we extracted *Alu* presence/absence loci that contradict the phylogenetic species tree of primates by using the graphical GPAC and PhyDE visualization tools. To filter out noise, we manually excluded all loci with ambiguous patterns as described in detail in the Materials and Methods section. Finally, using parsimony criteria, homoplasy was classified as precise parallel insertions or precise deletions. The numbers given are representative of the HCR screening. One of five loci from the HCR (+−+) parallel insertions category was classified in Table 1 as a locus with a complex evolutionary scenario (see Materials and Methods section and Table 1).

that were longer than 50 nt. All loci that met the condition of HCR (+−+) or (−+ +) were selected for exhaustive manual inspection and primate multispecies alignment reconstruction (corresponding to Haplorrhini species in Fig. 3) in the phylogenetic data editor (PhyDE, http://www.phyde.de/). For each remaining category that was not diagnostic for homoplasy, we randomly selected a sample of 100 loci and reconstructed primate multispecies alignments corresponding to Figure 3 to manually verify the frequency of reliable presence/absence patterns. For HCR (+++) loci, we checked that the *Alu* insertions were present in all Catarrhini species and perhaps in other primates as well; for HCR (++−) loci, we checked that the *Alu* insertions were present in human, chimpanzee, and possibly in other Hominoidea species, but absent in non-Hominoidea; for HCR (+−−) loci, we checked that the *Alu* insertions were absent in all primates except human; for HCR (−+−) loci, we checked that the *Alu* insertions were absent in all nonchimpanzee primates; and for HCR (−−+) loci, we checked that the *Alu* insertions were absent in all non-Old World monkeys. Then, to reduce potential GPAC errors, we derived a basic percentage of correctly identified loci to the total number of primary GPAC *perfect* cases and applied this percentage of correctly identified loci to the total number of GPAC *perfect* case loci from each category as a normalization. Loci present in more than one genomic copy in key species (human, chimpanzee, and rhesus macaque) were excluded from analyses. Loci with multiple concatenated elements (risk of ectopic *Alu*–*Alu* recombination) were also excluded. However, we retained loci in which younger *Alu*s were inserted into more ancient insertions, provided that we could clearly determine the orthology of the corresponding *Alu* insertions and as long as the nested insertions did not interfere with the original TSDs of the diagnostic element.

Alu presence/absence patterns analyzed in orthologous loci of multiple primate species were identified as homoplasious cases if the same type of *Alu* element (*Alu*J, *Alu*S, or *Alu*Y; and no diagnostic indels in consensus *Alu* subfamily sequences >3 nt) had a presence state in the same orientation, with identical TSDs (shift ⩽1 nt) in some of the investigated species. *Alu*s inserted in orthologous loci with TSD shifts of 2–3 nt were identified as *nearly precise* insertion cases. *Alu*s inserted in orthologous loci with TSD shifts >3 nt were rejected as nonhomoplasious, noninformative, independent insertions. The cases of *Alu* deletions in which parts of *Alu*s or TSDs were present, giving a trace of their ancestral presence-state, were rejected as imprecise nonhomoplasious deletions.

To determine whether the analyzed homoplasious presence/absence patterns represented cases of precise parallel insertions in human and rhesus macaque HCR (+−+) or in chimpanzee and rhesus macaque HCR (−++) or cases of precise deletions in human or chimpanzee, HCR (−++) and (+−+), respectively, we complemented the alignments by additional primate species (corresponding to Supplementary Tables S1–S3 and Supplementary Material Files S1–S3 available on Dryad at http://dx.doi.org/10.5061/dryad.532c28n) using BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and BLAT (https://genome.ucsc.edu/cgi-bin/hgBlat) and searched for the most parsimonious explanations of *Alu* presence/absence patterns. We derived the ancestral state of the element for the clade HC using presence/absence data from all available genomes

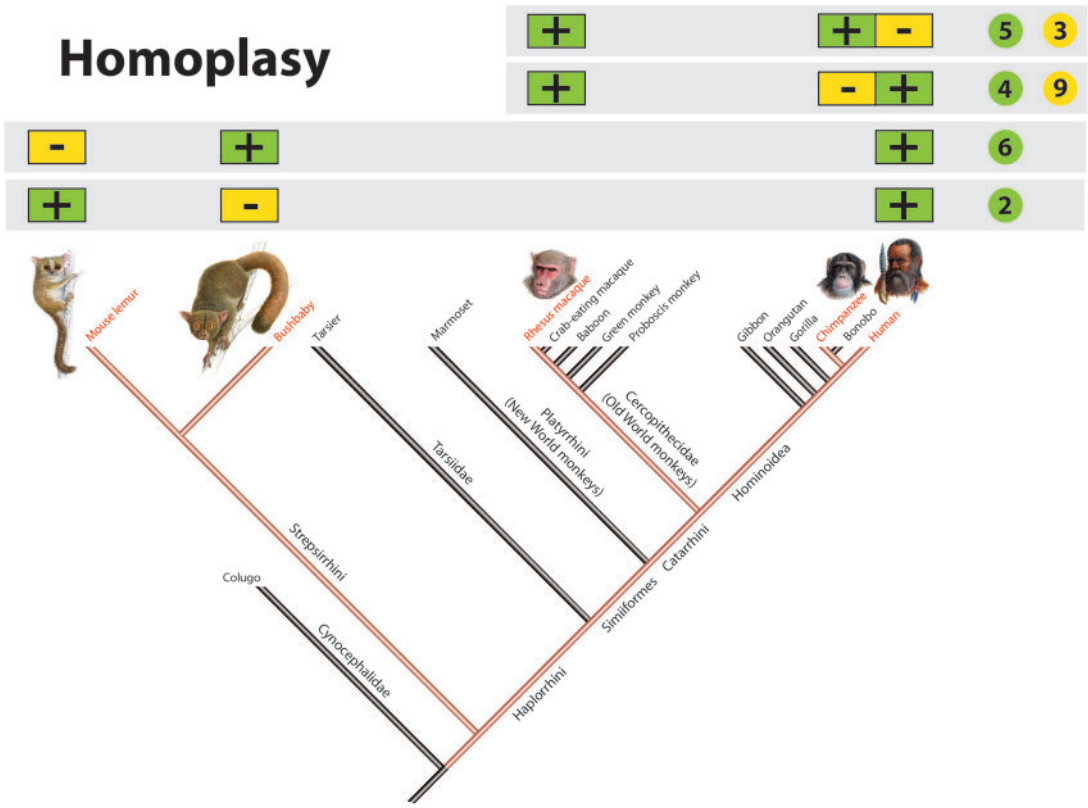

FIGURE 3. Homoplasious *Alu* retrotransposons in Primates. Phylogenetic screening for homoplasy in GPAC-derived *Alu* transposon presence/absence patterns. Red branches and labels indicate the targeted lineages. Sequences of all other species were added from BLAST and BLAT. The colugo (*Galeopterus variegatus*) was taken as the nonprimate outgroup. Green circles indicate precise parallel insertions, yellow circles, precise deletions. Boxes with pluses and minuses indicate presence/absence patterns identified in GPAC. Numbers indicate the cases of homoplasies derived from *perfect* GPAC patterns established on the lineages leading to apes or Strepsirrhini. Cases of uncharacterized homoplasy are not included in this figure.

of other Hominoidea species basal to the HC clade (corresponding to those in Fig. 3). If all basal species of Hominoidea (gibbon (*Nomascus leucogenys*), orangutan (*Pongo pygmaeus*), gorilla (*Gorilla gorilla*)) exhibited the absence-state of the *Alu* insertion, we assumed the ancestral state for the HC clade to be absence and interpreted HCR (+−+) or (−++) patterns as precise parallel insertions in human or chimpanzee and the rhesus macaque lineage. If all basal species of Hominoidea exhibited the presence-state of the *Alu* insertion, we assumed that the human–chimpanzee ancestor also exhibited the presence-state, and therefore, interpreted the patterns HCR (+−+) and (−++) as precise deletions in chimpanzee or human, respectively. In cases where we determined that a precise parallel insertion had occurred, we checked to make sure that the *Alu* insertion occurred only once in the Cercopithecoidea group (e.g., in the ancestor of rhesus macaque and crab-eating macaque [*Macaca fascicularis*], or in the ancestor of rhesus macaque and baboon [*Papio anubis*], etc.) and had a clear absence-state in New World monkeys. For loci that we determined to contain a precise deletion of an *Alu* element in human or chimpanzee, we checked to make sure that the ancestral state of the entire Cercopithecoidea group was presence. If an *Alu* presence/absence pattern in primates could not be explained by a single event of precise parallel insertion or precise deletion, we registered this locus as a homoplasious locus with a *complex evolutionary scenario*. Figure 2 shows a schematic representation of the applied screening strategies. The presence/absence table and all alignments of conflicting patterns HCR (+−+) and (−++) are presented in Supplementary Table S1 and Supplementary Material File S1, respectively (available on Dryad).

We calculated the frequency of homoplasious parallel insertions as a ratio of the found cases to the total number of *Alu*s inserted in the two investigated lineages (human and rhesus macaque lineages for HCR +−+; chimpanzee and rhesus macaque lineages for HCR −++). To estimate the frequency of precise deletions, we computed the ratio of precise deletion cases to the number of *Alu*s inserted in the common ancestor of HCR.

Reanalyzing the van de Lagemaat et al. Loci

To reassess the results of Van de Lagemaat et al. (2005) in light of the currently available genome data, we reconstructed alignments of the 36 described

homoplasious *Alu* loci that were presented as precise deletions. We used the current versions of human (GRCh38/hg38, December 2013), chimpanzee (Pan_tro 3.0/panTro5, May 2016), and rhesus macaque (BCM Mmul_8.0.1/rheMac8, November 2015) genomes and aligned and visualized them in PhyDE. In instances where they contained complex multi-copy loci, we extracted and aligned all copies. We completed alignments by adding sequences of the orthologous loci in additional great apes, lesser apes, Old World monkeys, New World monkeys, and prosimian species to newly reconstruct the evolutionary history of each proposed diagnostic *Alu* deletion. A presence/absence table and alignments of loci described as homoplasy in Van de Lagemaat et al. (2005) are shown in Supplementary Table S2 and Supplementary Material File S2, respectively (available on Dryad).

Human–Bushbaby–Lemur Comparison

To investigate the frequency of retrotransposon homoplasy in distant primate lineages we also screened the human–bushbaby (*Otolemur garnettii*)-lemur (HBL) group, in which the sister group relationships of bushbaby and lemurs comprising the clade Stepsirrhini are supported by significant evidence (e.g., Roos et al. 2004; Perelman et al. 2011). For the Strepsirrhini-human comparison we restricted our search for homoplasy to *Alu* families that were active during the early Strepsirrhini-Haplorrhini diversification (*Alu*Js, and *Alu* monomers), extracting coordinates of elements >50 nt from RepeatMasker reports of human, bushbaby, and mouse lemur (*Microcebus murinus*) genomes (http://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/, http://hgdownload.soe.ucsc.edu/goldenPath/otoGar3/bigZips/, and ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/165/445/GCF_000165445.2_Mmur_3., respectively). Similar to the exclusion of rhesus macaque-specific *Alu*s (see above), mouse lemur-specific *Alu*J_Mim elements were omitted from further analyses.

Similar to previous screens, we used the 30-way alignment described above to search in GPAC for possible conflicting presence/absence patterns of *Alu*s in HBL relationships (++−) and (+−+). Only *perfect* cases of presence or absence states were selected; for lemurs, at least one of four species (mouse lemur, Coquerel's sifaka [*Propithecus coquereli*], black lemur [*Eulemur macaco*], and Sclater's lemur [*Eulemur flaviofrons*]) was indicative for the presence or absence state. Manual analyses of extracted loci were performed as described above for the HCR group. Furthermore, we generated two additional 3-way genome alignments, one with bushbaby as reference species (http://hgdownload.soe.ucsc.edu/goldenPath/otoGar3/multiz3way/ and one with mouse lemur as reference species (http://hgdownload.soe.ucsc.edu/goldenPath/micMur3/multiz3way/), and embedded them in GPAC. Using the multiway alignments, we also performed GPAC screening and verification for all possible nonhomoplasious HBL presence/absence patterns as described above for the HCR comparison.

To determine whether HBL (++−) and (+−+) patterns represented cases of precise parallel insertions in human and bushbaby, HBL (++−), or in human and lemur, HBL (+−+) or the case of a precise deletion in bushbaby or lemurs (HBL +−+ and ++−, respectively), we determined their ancestral state in Haplorrhini using the additional primate species shown in Figure 3 and searched for the most parsimonious explanation of corresponding *Alu* presence/absence patterns in primates. If the ancestral state of Haplorrhini was absence, we interpreted HBL (++−) or (+−+) patterns to be precise parallel insertions in the lineage leading to human (Simiiformes) and bushbaby or in Simiiformes and lemurs, respectively. If the ancestral state of Haplorrhini was presence, then we interpreted HBL (++−) or (+−+) patterns to be a precise deletion of the *Alu* in the lemur or bushbaby lineages, respectively, or to be parallel insertions in HB or in HL lineages. The presence/absence table and alignments of conflicting HBL (++−) and (+−+) patterns are presented in Supplementary Table S3 and Supplementary Material File S3, respectively (available on Dryad). We calculated the frequency of homoplasy in HBL as described above for HCR.

Results

Human–Chimpanzee–Rhesus Macaque Comparison

We screened a total of 3,410,175 *Alu*-like presence/absence patterns among the three lineages, human (1,167,145 loci), chimpanzee (1,160,650 loci), and rhesus macaque (1,082,380 loci). Selecting the setting *Display Perfect* in the results part of GPAC yielded the following phylogenetic signals that contradicted the established HCR (++−) relationships: 450 cases for HCR (+−+) and 538 cases for HCR (−++) (Fig. 2). We then manually inspected all loci and added data from blast screens of related species and consensus sequences of diagnostic *Alu* elements to identify *Alu* types to help confirm our analysis. Of these, we identified 22 cases of true homoplasy: nine precise parallel insertions, 12 precise deletions (Fig. 3), and one locus with a complex evolutionary scenario that we cannot clearly attribute to either parallel insertion or deletion. Furthermore, we found one case of nearly precise insertions of the same *Alu*. We also found three cases of precise and one of nearly precise insertions of *Alu* elements belonging to different subfamilies with >3-nt diagnostic indels, none of which were considered to be homoplasious (Table 1, Supplementary Table S1, Supplementary Material File S1, available on Dryad). Eleven of our detected homoplasious loci overlap with cases described in Van de Lagemaat et al. (2005), but two of the parallel insertions were wrongly assigned as precise deletions (Supplementary Table S2 available on Dryad). However, our stringent GPAC screening conditions did not enable us to retrieve the six remaining homoplasious loci that

TABLE 1. HCR conflicting presence/absence patterns

Diagnostic loci	HCR (+−+)	HCR (−++)
Precise insertions/same *Alu*s (homoplasy)	**4**	**5**
Precise deletions (homoplasy)	**9**	**3**
Precise ins/del/same *Alu*s (homoplasy)	**1**	**0**
Nearly precise insertions (2–3 nt shift)/same *Alu*s	1	0
Precise insertions/different *Alu*s	3	0
Nearly precise insertions (2–3 nt shift)/different *Alu*s	1	0

Notes: Cases of true homoplasy are bold.
Ins = insertion; del = deletion.

TABLE 2. Reanalyzed conflicting presence/absence patterns described in van de Lagemaat et al. (2005)

Diagnostic loci	Van de Lagemaat et al. (2005)	Reanalysis
Precise insertions/same *Alu*s (homoplasy)	**0**	**4**
Precise deletions (homoplasy)	**36**	**11**
Precise ins/del/same *Alu*s (homoplasy)	**0**	**2**
Nearly precise insertions (2–3 nt shift)/same *Alu*s	0	1
Precise insertions/different *Alu*s	0	4
Imprecise insertions/deletions	0	10
Multi-copied loci	0	4

Notes: Cases of true homoplasy are bold.
Ins = insertion; del = deletion.

they described (see *Reanalyzing the van de Lagemaat Loci* below) due to regional ambiguous alignments.

GPAC screening of nonhomoplasious patterns and subsequent manual verification (see Materials and Methods) yielded the following signals: 544,034 cases for HCR (+++) (insertions occurred in the ancestral primate lineage before the Catarrhini split); 27,327 cases for HCR (++−) (insertions occurred in the ancestral Hominoidea lineage before the human–chimpanzee split); 4215 cases for HCR (+−−) (insertions occurred in the human lineage); 1881 cases for HCR (−+−) (insertions occurred in the chimpanzee lineage); and 33,954 cases for HCR (−−+) (insertions occurred in the Old World monkey lineage leading to rhesus macaque). Thus, in accordance with general knowledge (e.g., Perelman et al. 2011), our analyses revealed high support for the human–chimpanzee sister group relationships (HC:HR:CR 27,327:0:0; KKSC insertion significance test $P < 10^{-298}$; zeros reflect the absence of ILS and ancestral hybridization signals in the HCR group, according to our assumption; Kuritzin et al. 2016). Given the above data, we estimated the frequency of precise parallel insertions to be 0.01% in both human—rhesus macaque and chimpanzee—rhesus macaque pairs (4/(4215 + 33,954 + 4) × 100% and 5/(1881 + 33,954 + 5) × 100%, respectively). We estimated the frequency of precise deletions in human to be 0.001% (3/(544,034 + 3) × 100%) and in chimpanzee to be 0.002% (9/(544,034 + 9) × 100%).

Reanalyzing the van de Lagemaat Loci

Van de Lagemaat et al. (2005) investigated precise deletions of *Alu* elements in human and chimpanzee by assigning their presence in rhesus macaque as the ancestral state. Thus, only *Alu*s with presence/absence patterns HCR (+−+) and (−++) were further investigated. Accordingly, 36 cases of "apparent precise *Alu* deletions" were detected.

We reanalyzed these 36 proposed precise *Alu* deletions after extracting the human and chimpanzee *Alu* elements plus their flanking sequences from the coordinates provided in their publication. Orthologous regions were extracted for a set of genome-sequenced primates using the UCSC Genome Browser. Sequences were realigned and exact insertion/deletion points reconstructed.

Our multigenome/multispecies comparisons of *Alu* presence/absence patterns revealed that only 11 of the 36 loci contained precise deletions of *Alu*s in human or chimpanzee (chimpanzee + bonobo [*Pan paniscus*] ancestor). Of the remaining 25 loci, four contained insertions in the human lineage or in chimpanzee (chimpanzee + bonobo ancestor) and in one of the Old World monkey branches (rhesus macaque/rhesus macaque + crab-eating macaque/rhesus macaque + crab-eating macaque + baboon, etc.), which we interpreted by parsimonious criteria (see Materials and Methods) to have been precise parallel insertions rather than precise deletions in one of the lineages. Two loci reflect more complex evolutionary scenarios, probably including multiple cases of precise *Alu* insertions or deletions or both in various lineages.

The remaining 19 loci did not contain homoplasious events. Four of them contain parallel insertions of different elements. One locus contains nearly precise insertions in human and rhesus macaque + pig-tailed macaque ancestor with a shift of 2 nt. Ten loci contain imprecise insertions or deletions of *Alu*s with a shift of 5–31 nt; moreover, in two of these, the retroelements inserted in the opposite orientation. Four loci were present in several nonattributable copies in the genomes of the model species, and therefore, clear orthology could not be assigned (Table 2, Supplementary Table S2, Supplementary Material File S2 available on Dryad).

Human–Bushbaby–Lemur Comparison

We extracted a total of 826,034 human, bushbaby, and mouse lemur insertions, scanned them in GPAC to determine their presence/absence patterns, and identified the following potentially homoplasious loci: 337 candidates for HBL (++−) and 10 cases for HBL (+−+). As described for the HCR group, we manually inspected all cases. We complemented the alignments with sequences from additional species and relevant RepeatMasker consensus sequences and found 11 cases

TABLE 3. HBL conflicting presence/absence patterns

Diagnostic loci	HBL (++−)	HBL (+−+)
Precise insertions/same REs (homoplasy)	**6**	**2**
Uncharacterized homoplasy	**2**	**1**
Nearly precise insertions (2–3 nt shift)/same REs	5	0
Precise insertions/different REs	22	3
Nearly precise insertions (2–3 nt shift)/different REs	9	1

Notes: Cases of true homoplasy are bold.
RE = retroelement.

of true homoplasy: eight cases of homoplasious precise parallel insertions (Fig. 3) and three that could be equally probable parallel insertions in bushbaby—Haplorrhini (or lemurs—Haplorrhini) or precise deletions. We also found five cases of nearly precise insertions of the same type of *Alu*s, 25 cases of nonhomoplasious precise parallel insertions of different elements, and 10 cases of nearly precise insertions (Table 3, Supplementary Table S3, Supplementary Material File S3 available on Dryad).

Further GPAC screenings with subsequent manual verification revealed the following results: 2283 cases for HBL (+++) (insertions occurred in the common ancestor of primates); 13,118 cases for HBL (−++) (insertions occurred in the common ancestor of Strepsirrhini); 4876 cases for HBL (+−−) (insertions occurred in the common ancestor of Haplorrhini or Simiiformes); 23,859 cases for HBL (−+−) (insertions occurred in the bushbaby lineage); and 139 cases for HBL (−−+) (insertions occurred in the lemur lineage). Thus, our data support the commonly accepted Strepsirrhini monophyly (BL:HB:HL 13,118:0:0; KKSC insertion significance test $P < 10^{-298}$; zeros reflect the absence of ILS and ancestral hybridization signals in the HBL group according to our assumption; Kuritzin et al. 2016). Based on these data, we estimated the frequency of precise parallel insertions to be 0.02% (6/(4876 + 23,859 + 6) × 100%) in human and bushbaby genomes and 0.04% (2/(4876 + 139 + 2) × 100%) in human—lemurs. As we could not clearly classify any locus with a precise deletion, we did not calculate the frequency of precise deletions for the HBL group.

DISCUSSION

The insertion patterns of transposed elements have been successfully applied as clade markers in phylogenetic reconstructions (e.g., Shedlock et al. 2004; Doronina et al. 2015), for population analyses (e.g., Baker et al. 2018), and for tracing back significant genomic changes that are occasionally of adaptive relevance (reviewed in Schrader and Schmitz 2018). A small fraction of retrotransposons, and especially *Alu* elements, provide material for genetic novelties and enrich protein-coding sequences. Theoretically, homoplasious events in such cases might lead to phenotypic effects. However, the vast majority of retrotransposon insertions are evolutionarily neutral; hence, most homoplasies have consequences in phylogenetic reconstruction rather than phenotypic expression.

When a marker system is virtually free of homoplasy, a potential source of conflicting signals in phylogenetic reconstructions is hemiplasy induced by ancestral ILS, which is particularly prevalent in founder populations undergoing rapid radiation. ILS is expressed in polymorphic states of characters that are retained over successive speciation events and subsequently fixed randomly in different lineages. Furthermore, ancestral hybridization may also contribute to the occurrence of conflicting signals. As long as there is no significant presence of homoplasy, ILS, and ancestral hybridization are then very quantifiable.

During past usage of retrotransposons as phylogenetic markers, they acquired a gradient of assessments, including "SINEs of the perfect character" (Hillis 1999), "SINEs of a *nearly* perfect character" (Ray et al. 2006), and "Are transposable element insertions homoplasy free?" (Han et al. 2011). In the last study, the authors concluded that "no" they are not. However, it should be mentioned that they examined an inadequate test group, notorious for its impenetrable jungle of ILS-infiltrated short internodes, namely birds. It was shown in neoavian birds that the polymorphic hemiplasious state of characters persists over extremely long evolutionary periods (Suh et al. 2015) and renders this group unsuitable as an ILS-free reference group to access potential homoplasy. Furthermore, the authors examined precise deletions of CR1 elements, which lack the minimal criteria for the described mechanism of illegitimate recombination via 10–20-nt TSDs. CR1 elements are known to have no or mostly very short 4-6-nt TSDs (Ichiyanagi and Okada 2008) that are too short for recombination (Van de Lagemaat et al. 2005).

The question arises: how often *do* homoplasious retrotransposon signals occur or, in particular, how frequent are precise parallel insertions or random exact deletions for such elements that are originally copied and pasted into a quasi-infinite resource of available genomic loci, and which, in contrast to DNA transposons, do not have specific general mechanisms for element cleavage? A retrotransposon presence/absence pattern can be identified as homoplasious only in lineages with well-established phylogenetic relationships. In both of the groups we analyzed, human–chimpanzee–rhesus macaque and human–bushbaby–lemurs, the phylogenetic relationships were firmly established in many previous studies (e.g., Perelman et al. 2011; Herlyn 2016) and reconfirmed by our current data. Furthermore, ILS and ancestral hybridization can largely be ruled out in both groups, because diversification in the HCR group occurred during a period of approximately 25 million years (myr) and in the HBL group during 12 myr (dos Reis et al. 2018), whereas the fixation time for transposed elements in primates is estimated to be approximately 2 myr (Kuritzin et al. 2016).

Thus, all loci with the presence/absence patterns HCR (+−+ and −++) and HBL (++− and +−+) result most probably from independent events and are thus potentially homoplasious.

Precise Parallel Insertions

There are several reports or descriptions of single cases of individual hotspots of parallel insertions of transposed elements (Cantrell et al. 2001; Roy-Engel et al. 2002; Salem et al. 2003; Pecon-Slattery et al. 2004; Doronina et al. 2015); however, not all of them represent true homoplasy because members of distinct element subfamilies in orthologous loci were compared (e.g., Cantrell et al. 2001), different TSDs flanked such insertions (e.g., Roy-Engel et al. 2002), or ancestral ILS rather than homoplasy might have led to the presence/absence insertion pattern (e.g., Salem et al. 2003). On the other hand, systematic screenings for insertion homoplasy are rare, have been inappropriately conducted, or are now outdated considering the rapidly accumulating genomic data. In examining primates, Ray et al. (2006) selected a suitable test group and suitable element types to identify potential homoplasious markers from published presence/absence patterns of *Alu* element insertions, and the inspected lineages were separated by long internodes, minimizing the potential effects of ILS. However, they counted as precise parallel insertions retrotransposons belonging to distinct types. By compiling published "conflict" cases, their strategy did not enable them to investigate real homoplasy (apparently identical elements in orthologous loci).

In our systematic screening of primates, we found nine cases of homoplasious precise parallel insertions of apparently identical *Alu* elements in the human–chimpanzee–rhesus macaque comparison (four cases for HCR +−+, 0.01% and five for HCR −++, 0.01%) and 8 cases in the human–bushbaby–lemur comparison (6 cases for HBL ++−, 0.02%; and 2 for HBL +−+, 0.04%) (Tables 1 and 3, respectively). As the HR and CR parallel insertions occurred after Catarrhini diversification, when *Alu*Y activity was dominant (Churakov et al. 2010), it is not surprising that the vast majority of homoplasious parallel insertions that we found in HCR were *Alu*Y-*Alu*Y integrations. It should be mentioned, however, that we accepted *Alu* elements inserted in HR or CR as apparently identical, even when the RepeatMasker classified them as different subfamilies within *Alu*Y, as long as these subfamilies did not contain diagnostic indels >3 nt. The reason being that some *Alu*Y subfamilies differ only by few point mutations (e.g., the consensi of *Alu*Y and *Alu*YRc0 are distinguished by only two point mutations). *Alu*s that inserted several million years ago in noncoding regions unavoidably accumulate mutations that by chance can lead to misclassification in repeat masking. Thus, to avoid the current bias of TEs being homoplasy-free, we opted for overestimation rather than underestimation.

We found only a few precise insertions of *different* elements in the HCR group. By contrast, in HBL such events occurred much more frequently with a bias to human—bushbaby (22 cases in HB vs. three in HL). That probably corresponds to the high activity of Garn*Alu*s in bushbaby (about 520,000 additional tRNA-derived Garn*Alu*s specific for bushbaby), which also agrees with our observation that 19 of the 22 bushbaby insertions were Garn*Alu*s in the orthologous positions to the human *Alu*J elements.

Precise Deletions

More than two decades ago the reported case of an imprecise *Alu* deletion in human populations proposed that the recent loss of inserted *Alu*s may be a source of population polymorphisms (Edwards and Gibbs 1992). Van de Lagemaat et al. (2005) estimated that 0.5–1% of human–chimpanzee retrotransposon presence/absence patterns represented precise deletions of previously inserted retrotransposons. Reanalysis of these cases revealed that less than half of their results represent true homoplasies and only a third of them represent clear precise deletions. Similar to our assessment, van de Lagemaat et al. also considered retrotransposons of different *Alu*Y subfamilies located in orthologous loci in HR or CR as homoplasy. They found 15 such loci; however, our reanalysis revealed that in eight of them there was a shift of ⩾5 nt between insertions and two of them were in the opposite orientation, which confirms their independent origins. The remaining seven loci contained orthologous insertions, and they suggested gene conversion as the most probable scenario. *Alu* gene conversion is a process by which an originally inserted element from one *Alu* subfamily might subsequently be replaced by an *Alu* belonging to another subfamily, as proposed in some studies (e.g., Roy et al. 2000). It has been shown that such a replacement may occur within an *Alu* element, converting an old *Alu*S element that was inserted before the Catarrhini split (present in Old World monkeys and in nonhuman Hominoidea) to the very young *Alu*Yb8 in the human lineage, leaving the TSDs of the insertion intact (Kass et al. 1995). However, our reanalysis of van de Lagemaat et al.'s potential cases of gene conversion revealed that in all seven there was a clear absence state in all Hominoidea species except human, and in five of the seven the *Alu* element was also absent in some Old World monkeys. Based on the parsimony rule, we propose parallel insertions of two independent elements rather than insertion in the common ancestor with subsequent gene conversion on one of the branches. Thus, we suggest that potential *Alu* gene conversion can be neglected in this study.

Our screenings yielded 12 cases of precise *Alu* deletions in the human–chimpanzee–rhesus macaque comparison (three deletions in human, 0.001% and nine in chimpanzee, 0.002%), suggesting that 0.5% is an excessive overestimation of the frequency of precise deletions and that retrotransposon homoplasy due to

precise deletions of elements is even less problematic for phylogenetic reconstructions than previously proposed.

Interestingly, in the human–bushbaby–lemur comparison no clear cases of precise deletions were detected. There might be two explanations that are not mutually exclusive. One, the time between HBL speciation events was long enough (~12 myr, dos Reis et al. 2018) to accumulate mutations in TSD sequences, making them unsuitable for illegitimate recombination. Two, because there was only a relatively short internode before the first primate diversification (~8 myr, dos Reis et al. 2018), the number of *Alu* insertions was too low for the occurrence of such rare events as precise deletions.

Phylogenetic Relevance of Retrotransposons

Our analyses revealed rates of true homoplasious insertions to be about 0.01% in the HCR comparison and 0.02–0.04% in the HBL comparison, while rates of precise deletions were considerably lower. Thus, primate screening revealed that only a tiny fraction of retrotransposon signals appeared as a result of true homoplasy, indicating that such incidences can be neglected in phylogenetic investigations and that they do not interfere with retrotransposon-based phylogenetic analyses.

We used *Alu* elements in primates as a model test system. However, we suggest that our results are representative as well for other groups of species and elements, especially those that make use of the same or similar element mobilization system (LINE1) that produces suitably long TSDs during retrotransposition. Other systems, such as the CR1 element insertions that have no or only short TSDs require further investigations and, most importantly, a suitable pre-selection of a genome-sequenced test group largely free of ILS and ancestral hybridization.

As LINE1- and co-mobilized LINE1-dependent SINEs are the most frequently used presence/absence markers in phylogenetic reconstructions of mammals (e.g., Shimamura et al. 1997; Doronina et al. 2015; Feigin et al. 2018), our results are important to substantiate previous studies and to pre-validate the enormous potential source of future retrophylogenomic data. There are also phylogenetic studies in which, for example, CR1 (e.g., Suh et al. 2011) or LTR (e.g., Hartig et al. 2013) elements were successfully used as phylogenetic markers, both of which lack long enough TSDs as a source for recombination (Wicker et al. 2007; Ichiyanagi and Okada 2008). Moreover, LTR elements themselves provide an especially good substrate for illegitimate recombination involving the long terminal repeats (100–5000 nt) that flank their internal protein-coding part. In human, about 8% of the genome space is occupied by LTR elements, most of them comprising a single LTR unit generated by such an illegitimate recombination (Lander et al. 2001). These solitary LTRs with short, ~5-nt TSDs (Wicker et al. 2007) can be used effectively as phylogenetic markers. Nevertheless, the frequency of homoplasy among such elements with short TSDs might be different than that of *Alu*s, and systematic investigations are needed to be sure.

Using the GPAC tool enabled us to graphically analyze millions of genome-wide distributed insertions on a multispecies level and to automatically pre-filter a substantial but manageable number of *perfect* cases (GPAC setting) with clear orthology for individual manual alignments. We pre-filtered hundreds of such individual loci to identify the few real existing homoplasies that might confound phylogenetic reconstructions. Nevertheless, GPAC depends on basic background data, in this case genome sequence scaffolds compiled in multiway genome alignments. In the future, improved assemblies and improved multiway alignments will further improve the extraction of rare events such as precise parallel insertions and precise element excisions.

Homoplasy is not only an issue of neutral evolutionary processes such as random retrotransposon insertions or deletions but is also emphasized in the diversity of phenotypes. Exciting developmental and evolutionary questions arose with the observation that processes, mechanisms, and phenotypes accumulate in varying magnitudes of independently derived homoplasious characteristics (Wake et al. 2011). A better understanding of the process and frequency of basic molecular homoplasy will help to understand such complex patterns at the organismal level.

Conclusions

Genome-wide comparisons of retrotransposon insertion patterns in phylogenetically well-defined and virtually ILS/hybridization-free multispecies screens provide a reliable source to evaluate the extent and nature of confounding aspects in such a data environment. The screening and analysis strategies presented here in combination with an exhaustive source of available genome information in the selected primate test groups enabled us to find and characterize the few existing cases of real homoplasy present in the *Alu* retroposed element character set, and to identify many previously published but falsely defined cases. The 22 clear cases (nine precise insertions, 12 precise deletions and one locus with a complex evolutionary scenario from ~2,300,000 inspected loci) of homoplasy in retrotransposon presence/absence data of apes demonstrate that the frequency is even much lower than previously estimated, and therefore emphasize the special suitability of this marker system in phylogenetic reconstructions. Nevertheless, retrotransposon insertion homoplasy is a minor but still existing source of evolutionary noise, and especially with growing reports of insertions or deletions accompanied by functional changes, one cannot exclude the low possibility that new characters might repeatedly appear or be deleted again.

We suggest that the present data are not only representative of the broad range of primates and their *Alu*-SINE insertion site uniqueness but, because of shared features (long TSDs, specific target site preference), are also directly comparable to any LINE1-mobilized insertion in any therian taxonomic group. Thus, extensive systematic screenings will set a reference point for future Earth biodiversity genome sequencing projects.

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.532c28n.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft [SCHM1469/10-1].

Acknowledgments

Many thanks go to Gennady Churakov and Michael Haberl for fruitful discussions. We thank Marsha Bundman for editorial assistance and Jon Baldur Hlidberg for the animal paintings.

References

Avise J.C., Robinson T.J. 2008. Hemiplasy: a new term in the lexicon of phylogenetics. Syst. Biol. 57:503–507.

Baker J.N., Walker J.A., Denham M.W., Loupe C.D. 3rd, Batzer M.A. 2018. Recently integrated *Alu* insertions in the squirrel monkey (*Saimiri*) lineage and application for population analyses. Mob. DNA. 9:9.

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

Churakov G., Grundmann N., Kuritzin A., Brosius J., Makalowski W., Schmitz J. 2010. A novel web-based TinT application and the chronology of the Primate *Alu* retroposon activity. BMC Evol. Biol. 10:376.

Doronina L., Churakov G., Shi J., Brosius J., Baertsch R., Clawson H., Schmitz J. 2015. Exploring massive incomplete lineage sorting in arctoids (Laurasiatheria, Carnivora). Mol. Biol. Evol. 32:3194–3204.

Doronina L., Churakov G., Kuritzin A., Shi J., Baertsch R., Clawson H., Schmitz J. 2017. Speciation network in Laurasiatheria: retrophylogenomic signals. Genome Res. 27:997–1003.

dos Reis M.D., Gunnell G.F., Barba-Montoya J., Wilkins A., Yang Z., Yoder A.D. 2018. Using phylogenomic data to explore the effects of relaxed clocks and calibration strategies on divergence time estimation: primates as a test case. Syst. Biol. 67:594–615.

Edwards M.C., Gibbs R.A. 1992. A human dimorphism resulting from loss of an *Alu*. Genomics. 14:590–597.

Feigin C.Y., Newton A.H., Doronina L., Schmitz J., Hipsley C.A., Mitchell K.J., Gower G., Llamas B., Soubrier J., Heider T.N., Menzies B.R., Cooper A., O'Neill R.J., Pask A.J. 2018. Genome of the Tasmanian tiger provides insights into the evolution and demography of an extinct marsupial carnivore. Nat. Ecol. Evol. 2:182–192.

Hamdi H., Nishio H., Zielinski R., Dugaiczyk A. 1999. Origin and phylogenetic distribution of *Alu* DNA repeats: irreversible events in the evolution of primates. J. Mol. Biol. 289:861–871.

Han K., Konkel M.K., Xing J., Wang H., Lee J., Meyer T.J., Huang C.T., Sandifer E., Hebert K., Barnes E.W., Hubley R., Miller W., Smit A.F., Ullmer B., Batzer M.A. 2007. Mobile DNA in old world monkeys: a glimpse through the rhesus macaque genome. Science 316:238–240.

Han K.L., Braun E.L., Kimball R.T., Reddy S., Bowie R.C., Braun M.J., Chojnowski J.L., Hackett S.J., Harshman J., Huddleston C.J., Marks B.D., Miglia K.J., Moore W.S., Sheldon F.H., Steadman D.W., Witt C. C., Yuri T. 2011. Are transposable element insertions homoplasy free?: An examination using the avian tree of life. Syst. Biol. 60:375–386.

Hartig G., Churakov G., Warren W.C., Brosius J., Makalowski W., Schmitz J. 2013. Retrophylogenomics place tarsiers on the evolutionary branch of anthropoids. Sci. Rep. 3:1756.

Herlyn H. 2016. The phylogenetic system of primates—character evolution in the light of a consolidated tree. Org. Divers. Evol. 16:689.

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

Ichiyanagi K., Okada N. 2008. Mobility pathways for vertebrate L1, L2, CR1, and RTE clade retrotransposons. Mol. Biol. Evol. 25:1148–1157.

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

Kass D.H., Batzer M.A., Deininger P.L. 1995. Gene conversion as a secondary mechanism in SINE evolution. Mol. Cell. Biol. 15:19–25.

Konkel M.K., Walker J.A., Hotard A.B., Ranck M.C., Fontenot C.C., Storer J., Stewart C., Marth G.T., Batzer M.A. 2015. Sequence analysis and characterization of active human *Alu* subfamilies based on the 1000 genomes pilot project. Genome Biol. Evol. 7:2608–2622.

Kramerov D.A., Vassetzky N.S. 2011. Origin and evolution of SINEs in eukaryotic genomes. Heredity (Edinb). 107:487–495.

Kriegs J.O., Churakov G., Jurka J., Brosius J., Schmitz J. 2007. Evolutionary history of 7SL RNA-derived SINEs in supraprimates. Trends Genet. 23:158–161.

Kuritzin A., Kischka T., Schmitz J., Churakov G. 2016. Incomplete lineage sorting and hybridization statistics for large-scale retroposon insertion data. PLoS Comput. Biol. 12:e1004812.

Lander E.S., Linton L.M., Birren B., Nusbaum C., Zody M.C., Baldwin J., Devon K., Dewar K., Doyle M., Fitzhugh W., Funke R., Gage D., Harris K., Heaford A., Howland J., Kann L., Lehoczky J., Levine R., McEwan P., McKernan K., Meldrim J., Mesirov J.P., Miranda C., Morris W., Naylor J., Raymond C., Rosetti M., Santos R., Sheridan A., Sougnez C., Stange-Thomann Y., Stojanovic N., Subramanian A., Wyman D., Rogers J., Sulston J., Ainscough R., Beck S., Bentley D., Burton J., Clee C., Carter N., Coulson A., Deadman R., Deloukas P., Dunham A., Dunham I., Durbin R., French L., Grafham D., Gregory S., Hubbard T., Humphray S., Hunt A., Jones M., Lloyd C., McMurray A., Matthews L., Mercer S., Milne S., Mullikin J.C., Mungall A., Plumb R., Ross M., Shownkeen R., Sims S., Waterston R.H., Wilson R.K., Hillier L.W., McPherson J.D., Marra M.A., Mardis E.R., Fulton L.A., Chinwalla A.T., Pepin K.H., Gish W.R., Chissoe S.L., Wendl M.C., Delehaunty K.D., Miner T.L., Delehaunty A., Kramer J.B., Cook L.L., Fulton R.S., Johnson D.L., Minx P.J., Clifton S.W., Hawkins T., Branscomb E., Predki P., Richardson P., Wenning S., Slezak T., Doggett N., Cheng J.F., Olsen A., Lucas S., Elkin C., Uberbacher E., Frazier M., Gibbs R.A., Muzny D.M., Scherer S.E., Bouck J.B., Sodergren E.J., Worley K.C., Rives C.M., Gorrell J.H., Metzker M.L., Naylor S.L., Kucherlapati R.S., Nelson D.L., Weinstock G.M., Sakaki Y., Fujiyama A., Hattori M., Yada T., Toyoda A., Itoh T., Kawagoe C., Watanabe H., Totoki Y., Taylor T., Weissenbach J., Heilig R., Saurin W., Artiguenave F., Brottier P., Bruls T., Pelletier E., Robert C., Wincker P., Smith D.R., Doucette-Stamm L., Rubenfield M., Weinstock K., Lee H.M., Dubois J., Rosenthal A.,. Platzer M, Nyakatura G., Taudien S., Rump A., Yang H., Yu J., Wang J., Huang G., Gu J., Hood L., Rowen L., Madan A., Qin S., Davis R.W., Federspiel N.A., Abola A.P., Proctor M.J., Myers R.M., Schmutz J., Dickson M., Grimwood J., Cox D.R., Olson M.V., Kaul R., Raymond C., Shimizu N., Kawasaki K., Minoshima S., Evans G.A., Athanasiou M., Schultz R., Roe B.A., Chen F., Pan H., Ramser J., Lehrach H., Reinhardt R., McCombie W.R., de la Bastide M., Dedhia N., Blocker H., Hornischer K., Nordsiek G., Agarwala R., Aravind L., Bailey J.A., Bateman A., Batzoglou S., Birney E., Bork P., Brown D.G., Burge C.B., Cerutti L., Chen H.C., Church D., Clamp M., Copley R.R., Doerks T., Eddy S.R., Eichler E.E., Furey T.S., Galagan J.,

Gilbert J.G, Harmon C., Hayashizaki Y., Haussler D., Hermjakob H., Hokamp K., Jang W., Johnson L.S., Jones T.A., Kasif S., Kaspryzk A., Kennedy S., Kent W.J., Kitts P., Koonin E.V., Korf I., Kulp D., Lancet D., Lowe T.M., McLysaght A., Mikkelsen T., Moran J.V., Mulder N., Pollara V.J., Ponting C.P., Schuler G., Schultz J., Slater G., Smit A. F., Stupka E., Szustakowki J., Thierry-Mieg D., Thierry-Mieg J., Wagner L., Wallis J., Wheeler R., Williams A., Wolf Y.I., Wolfe K.H., Yang S.P., Yeh R.F., Collins F., Guyer M.S., Peterson J., Felsenfeld A., Wetterstrand K.A., Patrinos A., Morgan M.J., de Jong P., Catanese J.J., Osoegawa K., Shizuya H., Choi S., Chen Y.J., Szustakowki J. 2001. Initial sequencing and analysis of the human genome. Nature. 409:860–921.

Lewin H.A., Robinson G.E., Kress W.J., Baker W.J., Coddington J., Crandall K.A., Durbin R., Edwards S.V., Forest F., Gilbert M.T.P., Goldstein M.M., Grigoriev I.V., Hackett K.J., Haussler D., Jarvis E.D., Johnson W.E., Patrinos A., Richards S., Castilla-Rubio J.C., van Sluys M.A., Soltis P.S., Xu X., Yang H., Zhang G. 2018. Earth BioGenome Project: Sequencing life for the future of life. Proc. Natl. Acad. Sci. USA. 115:4325–4333.

Luan D.D., Korman M.H., Jakubczak J.L., Eickbush T.H. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell. 72:595–605.

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.

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

Nikaido M., Rooney A.P., 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.

Noll A., Grundmann N., Churakov G., Brosius J., Makalowski W., Schmitz J. 2015. GPAC-genome presence/absence compiler: a web application to comparatively visualize multiple genome-level changes. Mol. Biol. Evol. 32:275–286.

Pecon-Slattery J., Pearks Wilkerson A.J., Murphy W.J., O'Brien S.J. 2004. Phylogenetic assessment of introns and SINEs within the Y chromosome using the cat family Felidae as a species tree. Mol. Biol. Evol. 21:2299–2309.

Perelman P., Johnson W.E., Roos C., Seuanez H.N., Horvath J.E., Moreira M.A., Kessing B., Pontius J., Roelke M., Rumpler Y., Schneider M.P., Silva A., O'Brien S.J., Pecon-Slattery J. 2011. A molecular phylogeny of living primates. PLoS Genet. 7:e1001342.

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

Ray D.A., Xing J., Salem A.H., Batzer M.A. 2006. SINEs of a nearly perfect character. Syst. Biol. 55:928–935.

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

Roy A.M., Carroll M.L., Nguyen S.V., Salem A.H., Oldridge M., Wilkie A.O., Batzer M.A., Deininger P.L. 2000. Potential gene conversion and source genes for recently integrated *Alu* elements. Genome Res. 10:1485–1495.

Roy-Engel A.M., Carroll M.L., Vogel E., Garber R.K., Nguyen S.V., Salem A.H., Batzer M.A., Deininger P.L. 2001. *Alu* insertion polymorphisms for the study of human genomic diversity. Genetics. 159:279–290.

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

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

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

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

Schrader L., Schmitz J. 2018. The impact of transposable elements in adaptive evolution. Mol. Ecol. doi:10.1111/mec.14794.

Shedlock A.M., 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.

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

Stoneking M., Fontius J.J., Clifford S.L., Soodyall H., Arcot S.S., Saha N., Jenkins T., Tahir M.A., Deininger P.L., Batzer M.A. 1997. *Alu* insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res. 7:1061–1071.

Suh A., Paus M., Kiefmann M., Churakov G., Franke F.A., Brosius J., Kriegs J.O., Schmitz J. 2011. Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. Nat. Commun. 2:443.

Suh A., Smeds L., Ellegren H. 2015. The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. PLoS Biol. 13:e1002224.

Takahashi K., Terai Y., Nishida M., Okada N. 1998. A novel family of short interspersed repetitive elements (SINEs) from cichlids: the pattern of insertion of SINEs at orthologous loci support the proposed monophyly of four major groups of cichlid fishes in Lake Tanganyika. Mol. Biol. Evol. 15:391–407.

van de Lagemaat L.N., Gagnier L., Medstrand P., Mager D.L. 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., Kishino H., Ota R. 2001. A phylogenetic foundation for comparative mammalian genomics. Genome Inform. 12:141–154.

Wake D.B., Wake M.H., Specht C.D. 2011. Homoplasy: from detecting pattern to determining process and mechanism of evolution. Science. 331:1032–1035.

Wicker T., Sabot F., Hua-Van A., Bennetzen J.L., Capy P., Chalhoub B., Flavell A., Leroy P., Morgante M., Panaud O., Paux E., SanMiguel P., Schulman A. H. 2007. A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. 8:973-982.