Supporting Online Material for

Mobile DNA in Old World Monkeys: A Glimpse Through the Rhesus Macaque Genome

Kyudong Han, Miriam K. Konkel, Jinchuan Xing, Hui Wang, Jungnam Lee, Thomas J. Meyer, Charles T. Huang, Erin Sandifer, Kristi Hebert, Erin W. Barnes, Robert Hubley, Webb Miller, Arian F. A. Smit, Brygg Ullmer, Mark A. Batzer*

*To whom correspondence should be addressed. E-mail: mbatzer@lsu.edu

This PDF file includes:

> Materials and Methods
> SOM Text
> Figs. S1 and S2
> Tables S1 to S7
> References

Other Supporting Online Material for this manuscript includes the following:
(available at www.sciencemag.org/cgi/content/full/316/5822/238/DC1)

> Source Code

Materials and Methods

Computational Analysis

Putative rhesus-specific insertions – or more precisely, Old World Monkey

(OWM)-specific insertions present in the rhesus genome but not in the human and

chimpanzee genomes – were computationally retrieved from the multi-Z 3-way

alignment of the human, chimpanzee, and rhesus macaque genomes (Cornell version

2006/06) using in-house scripts (see supplemental file source_code.doc) and a

customized variant of BLAT (BLAST like alignment tool, *S1*). Given more than 400,000

indels to analyze, several large cluster computers were used to conduct extraction,

filtering, and analysis, altogether processing roughly 5 million files per complete analysis

(two full analyses were completed). As the rhesus macaque genome is a draft genome

and portions of the current multi-Z 3-way alignment are not well aligned, we most likely

undercounted OWM-specific elements. Therefore, our count represents a conservative

estimate of retrotransposon insertions in the OWM lineage.

The extraction process for retrieving and filtering indel data for the three species

is illustrated in Fig. S1. Briefly, the headers from aligned sequences containing

information regarding each chromosome, including both their location and orientation,

were extracted from the multi-Z 3-way alignment. Indels were identified through an *in*

silico comparison of consecutive entries in the 3-way alignment. If the end of one

sequence did not terminate with the beginning of the next, and the orientation and

1

chromosome did not change, the indel was retrieved using our modified BLAT installation. Additionally, 500 bp of flanking sequence on both sites of all indels were extracted and saved in separate files. Next, all indels and flanking sequences were analyzed with a locally installed version of RepeatMasker using the repbase library of repetitive elements (http://girinst.org, version 11.05). After parallel execution of RepeatMasker on the 400,000 candidate indels, we filtered the output files for a variety of criteria (target element type, length, number and position of repetitive elements within indels, etc.). Through this approach, all OWM-specific *Alu* and L1 elements were extracted. Roughly five million files were generated for the genome analysis and were sorted by chromosomal location and repetitive element content using in-house Tcl, Python, and shell scripts (see supplemental file source_code.doc). For brevity, we will describe sequences present in the rhesus genome but not in the human and chimpanzee genome as "OWM-specific insertions" in the following text, although some of these gaps may be deletions in the common ancestor of humans and chimpanzees.

Subfamily identification, copy number and age estimates

For ERVs, a neighbor joining tree (*S2*) of the full-length Pol proteins of OWM retroviruses was constructed using ClustalX (*S3*) with 1000 replicates bootstrap analysis. Non-macaque-specific ERV proteins were extracted from the transposable element protein database curated at and available from http://www.repeatmasker.org. This database contains a subset of proteins less than 90% identical to each other (*i.e.* closely related proteins are represented by one representative sequence).

To examine the L1 subfamily structure in the OWM lineage, we aligned 871 bp

sequences from 1620 OWM-specific L1 elements, a region that corresponds to the 3' end

of ORF2 and the entire 3' UTR (excluding the $G_4TG_6AG_6AG_3$ repeat unit due to length

variations among elements), using ClustalW, as embedded in BioEdit v.7.0 (*S4*). The

MegAlign program (available in the DNAStar software package) was used to generate L1

subfamily consensus (or ancestral) sequences on the basis of putative diagnostic

substitutions. The nomenclature of OWM-specific L1 subfamilies follows the method

employed with chimpanzee-specific L1 subfamilies (*S5*). Briefly, each L1 subfamily

name is identified by CER (which stands for Cercopithecidae, indicating the origin of the

consensus sequence) and an Arabic number indicating its lineage (*e.g.* L1CER-1),

followed by a capital letter classifying the subfamilies within the lineage (*e.g.* L1CER-

2A). Lowercase letters are also added for classified subfamilies outside of the sequential

lineage (*e.g.* L1CER-2Ea). The relationship among all subfamilies was constructed using

a median-joining network (*S6*). However, due to an 18-bp duplication in the 3' UTR of

some subfamily consensus sequences, separate networks of the consensus sequences with

and without the duplications (dashed line and red arrow in Fig. 2, respectively) were

reconstructed independently. The two networks were then combined along with the older

L1PA6, the L1PA5, and Hominoid-specific L1 subfamily consensus sequences (L1PA4

to L1Hs and L1Pt). The copy number of L1 subfamilies was estimated by multiplying the

percentage of elements in each subfamily from the initial analysis with the estimated total

number of OWM-specific L1s from the computational analysis. The age of L1

subfamilies was estimated with NETWORK 4.200 using the time estimates module (*S7*),

3

on the basis of the divergence among all copies of each subfamily. A neutral nucleotide substitution rate of 0.0015/site/million years (*S8*) was used for the age estimate.

For *Alu* elements, multiple sequence alignments were constructed with ClustalW using 5500 *Alu*Y elements randomly selected from the rhesus genome. The consensus sequence of each subfamily was determined using a majority rules approach available in BioEdit 7.0 (*S4*). Due to the faster mutation rate of CpG dinucleotide sequence (*S9*), only non-CpG substitutions were considered as diagnostic substitutions. If a group of elements shared more than three non-CpG substitutions (as compared to the *Alu*Y consensus sequence) and the group had more than ten members, the group was considered to be an individual subfamily. During the reconstruction process, the highly variable CpG dinucleotide sites posed a problem for the majority rule, and the program failed to infer the consensus sequence. To resolve this issue, we manually examined the dinucleotide sites exhibiting high diversity. If any dinucleotide sites were comprised primarily of CpA and TpG dinucleotides, which were assumed to be the decayed highly mutable CpG sites, CG was assigned to the consensus sequence for that site. The four *Alu* lineages in the rhesus genome were named *Alu*YRa to *Alu*YRd ("R" standing for rhesus, indicating the origin of the consensus sequence), followed by a number and a letter indicating the subfamily within a certain lineage (*e.g. Alu*YRa1 represents the first *Alu* subfamily belonging to lineage a). The copy number of *Alu* subfamilies was estimated by searching the rhesus macaque genome with subfamily consensus sequences with BLAT. The age of each subfamily was estimated as previously described (*S9*); age estimates on the basis of CpG and non-CpG mutations were averaged to obtain the final results. Due to the conservative criteria we used for mapping *Alu* elements to subfamilies (only one mutation

4

allowed at the diagnostic sites), our copy number estimate is likely to be an under-

estimate.

Analysis of full-length and nearly full-length L1 elements

To identify retrotransposition competent OWM-specific L1s with intact coding

sequences for the enzymatic machinery used for retrotransposition (ORF1 and ORF2), we

extracted all OWM-specific L1PA5 insertions that were >5.5kb and started between

position 1 and 40 of the consensus sequence on the basis of their RepeatMasker

annotations. This approach has previously been shown to be effective in identifying full-

length L1 elements (*S10*). With these criteria, a total of 3248 L1PA5 elements were

extracted. Among them, 978 elements were excluded from further analyses due to

missing data (*i.e.* Ns) in their sequences. The remaining 2270 L1 elements were analyzed

with L1Xplorer (http://l1xplorer.molgen.mpg.de/) and/or by manual inspection after

alignment to determine whether ORF1 and ORF2 were intact.

Primer design

On the basis of the diverged primate species used for this study, primers were

designed either using Primer3 (http://cbr-rbc.nrc-cnrc.gc.ca/cgi-bin/primer3_www.cgi) or

manually in conserved genomic regions. For both scenarios, areas of repetitive sequences

were skipped for primer design. This information was obtained by RepeatMasker analysis

of flanking sequences for all three species across each locus. For the manual primer

design the sequences of all three species were aligned with ClustalW (*S11*).

Regardless of the primer design-process all primers were analyzed with BLAT or BLAST (*S12*) to ensure that primers resided in unique sequence. Also, a virtual PCR (http://genome.ucsc.edu) was performed for all primer pairs to further confirm that only one PCR product was likely generated, and to ascertain the size of the PCR products for both filled and empty sites. Sequences of the primers (including the size of PCR products) are shown in Tables S6 and S7.

PCR amplification and sequencing

PCR amplification (if not otherwise indicated) was performed in 25 μl reactions containing 10 to 25 ng of template DNA; 5-10 pmol of each oligonucleotide primer (Sigma-Genosys, St. Louis, MO); 1.5mM $MgCl_2$ (Applied Biosystems, Foster city, CA); 10X PCR buffer (50mM KCl; 10mM Tris–HCl, pH 8.4, Applied Biosystems, Foster city, CA); 0.2 mM dNTPs (Invitrogen, Carlsbad, CA); and 2 U Taq DNA polymerase (Invitrogen, Carlsbad, CA) on an iCycler (Bio-rad, Hercules, CA). Each PCR reaction was performed under the following conditions: initial denaturation step at 94 °C for 120 s, followed by 32 cycles of denaturation at 94 °C for 15 s, 45 s at annealing temperature (specific for each locus), and extension at 72 °C for 30 s, followed by a final extension step at 72 °C for 3 min. For analysis, 15 μl of each PCR product was fractionated on a 2% agarose gel (Bio-rad, Hercules, CA) containing 0.1 μg/ml ethidium bromide (Amresco, Solon, OH) in a horizontal gel chamber for 45 min at 180 V, and directly visualized using UV-fluorescence (Bio-rad, Hercules, CA). Some unexpected PCR products were purified from the gels using the Wizard® gel purification kit (Promega, Madison, WI) and sequenced directly using the respective primers. DNA sequencing was

performed using chain termination sequencing on an Applied Biosystems ABI3130*XL*

(Foster city, CA) automated DNA sequencer.

Supporting online text

Horizontal transmission of ERVs in the OWMs

The OWM-specific horizontal invasions are supported by the fact that the human and rhesus macaque genome did not contain closely related ERVs at orthologous sites. Nor are any of the families similar enough to each other so that they could represent a single invasion followed by divergence into two groups during its at most 25 million-year propagation in the rhesus genome. Finally, the consensus sequences for all eight contain intact *env* gene coding regions that are essential for horizontal transfer but are generally lost in long-term vertically transmitted ERVs.

Three of the five horizontally transmitted ERV class I retroviruses closely resemble gamma retroviruses in other species. MacERV1 is 98% identical to PtERV1, which has invaded the gorilla, chimpanzee, baboon and macaque germlines (*S13-S15*). Another, MacERV2, has a consensus sequence that is <80% similar to that of MacERV1. This divergence level and the paucity of sequence intermediates between MacERV1 and MacERV2 suggest separate invasion scenarios for these closely related retroviruses. A third class I element, MacERV3, is very closely related to the Baboon Endogenous Virus (BaEV) and has been detected previously in macaque (*S16*). The remaining three invading ERVs are class II elements. MacERV4 is very closely related to SRV1a, a Simian Retrovirus (SRV). The wider phylogenetic range of exogenous SRVs suggests the relatively recent endogenization of an SRV1a-like virus in the macaque genome. The remaining two class II invaders are somewhat related to the HERVK group of ERVs. Although they are close relatives, they again are far too distant to challenge the notion of two separate germline infections.

Evolutionary history of L1s in OWMs (Fig. 2)

The OWM-specific L1 lineage rooted with the L1PA6 consensus sequence and propagated as a single lineage from 33 to 25 million years ago (mya) (until subfamily L1CER-1D), up to the divergence of a new L1 lineage (L1CER-2). The L1CER-2 lineage contained eight sequential L1 subfamilies (L1CER-2A to L1CER-2G), but all became extinct (or silenced) within ~3 million years (myrs). The youngest subfamily, L1CER-2G, is estimated to be ~20 myrs old. After the divergence of the L1CER-2 lineage, the L1CER-1 lineage later separated into two groups of subfamilies after L1CER-1F. Both groups evolved in a sequential order but their relationship with L1CER-1F is less clear. Group I subfamilies (L1CER-3A to -3H) are characterized by an 18-bp duplication in their 3'UTR region and the subfamilies within the group were active between ~22 (L1CER-3A) and ~10 mya (L1CER-3H). Group II (L1CER-4A to -4G) seems to have started their amplification more recently and seven of them have estimated ages of less than 10 myrs (Table S2). The network results produced two alternative pathways for the origin of the group II subfamilies. The first pathway, when networks were separately constructed, identified the subfamily L1CER-1F as the common ancestor of group II subfamilies (dashed line in Fig. 2). The second pathway suggested that group II subfamilies were derived from a younger subfamily (L1CER-3F) of group I because the L1CER-3F and L1CER-4A consensus sequences are identical except for the 18-bp duplication (red arrow in Fig. 2).

Furthermore, we analyzed the 5' UTR of the L1CER-3 and L1CER-4 subfamilies; the results indicate that their 5' UTRs are very similar to each other, suggesting that they possibly utilize similar host factors for their retrotransposition. If the L1CER-4

9

subfamilies are indeed derived from L1CER-1F, as the first pathway suggested, these two

lineages (L1CER-3 and L1CER-4) would have been in competition with each other for

~20 myrs. However, if the L1CER-4 subfamilies are derived from younger L1CER-3

subfamilies as the second pathway suggested, these two lineages only overlapped for a

short period of time and long-term competition would not have occurred. We therefore

favor the second hypothesis for the origin of L1CER-4 lineages: that the L1CER-4A

subfamily is derived from the L1CER-3F subfamily.

Retrotransposon-mediated DNA sequence transduction

The three L1 elements with 5' transduced exon sequence were identified during

the investigation of L1 retrotransposon elements in the rhesus macaque genome. In each

case, the transduced sequence contained the spliced form of several exons. The locations

of the transduced sequences (original locus and place of insertion), and the genes

involved in the transduction process are shown in Table S3. Further detailed analyses are

necessary to investigate these loci, in particular locus 145670, regarding internalization of

spliced exons into the coding sequence of other genes.

Evolutionary analysis of **Alu** subfamilies

In order to determine the evolutionary origin of each *Alu* subfamily, we

genotyped a series of *Alu* elements on the primate phylogenetic DNA panel (Table S5).

The genotyping results showed that the oldest subfamily, *Alu*YRa1, started its

amplification not long after the divergence of OWMs and Hominoids and some of its

members are present in all OWM species included in the analysis. Some *Alu*YRa2

elements, the second oldest subfamily, are present in all species of subfamily Cercopithecinae analyzed, suggesting an origin before the divergence of the Cercopithecinae species. The *Alu*YRb1 subfamily originated before the divergence of tribe Papionini (represented in our panel by macaques and baboon). The remaining 11 subfamilies (YRa3, YRa4, YRb2-4, YRc1-2, and YRd1-4) are restricted to genus *Macaca*.
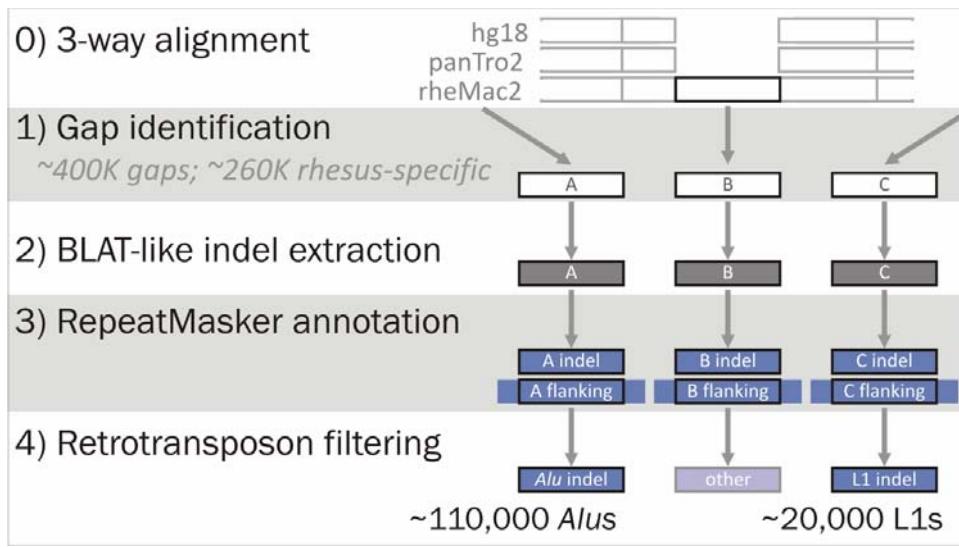
Figure S1. Overview of the computational workflow of OWM-specific retrotransposon identification. Our computational analysis followed a four-stage process. Beginning with a Cornell 3-way alignment of the human, chimpanzee, and rhesus macaque genomes (0), we computationally identified ~400k gaps between stretches of aligned sequence (1). For each of these coordinates, we extracted the corresponding DNA sequences from the three source genomes (2). RepeatMasker was run on each of these candidates -- both on the gap itself, and the gap + surrounding flanking sequence (3). Finally (4), we extracted species-specific mobile elements.
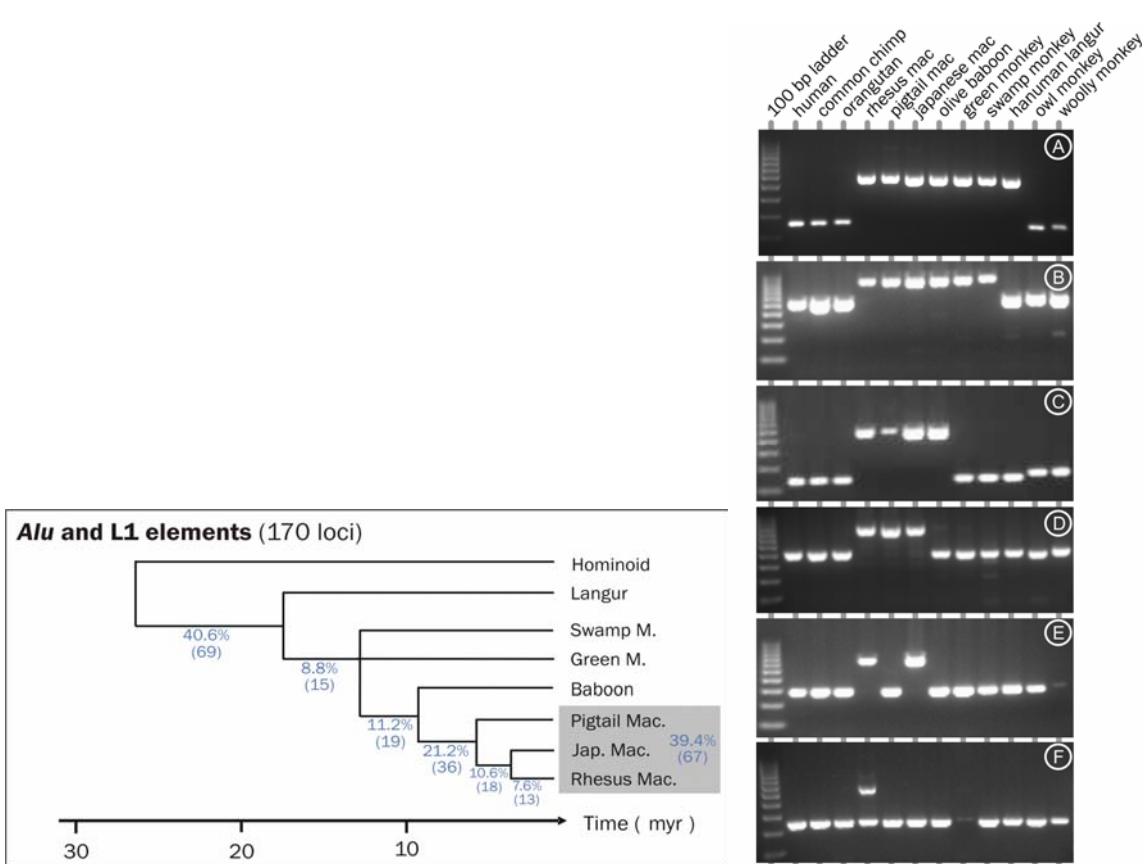
Figure S2: A. Phylogenetic distribution of OWM-specific *Alu* and L1 insertion polymorphisms. The OWM-specific *Alu* and L1 insertion polymorphisms (170 loci) were plotted on a primate evolutionary cladogram with Hominoids as an outgroup. NWMs are not included in the figure, as they also represent a distant taxa. The number of insertions specific to each OWM clade is shown below each node along with the percentage distribution of inserts. **B. Gel chromatographs of OWM-specific *Alu* and L1 insertions.** Six examples of gel chromatographs used to determine the phylogenetic origin of individual insertions (here *Alu* elements) are shown. Upper DNA fragments indicate "filled" sites (insertion present) and lower fragments indicate "empty" sites (no insertion). The DNA template used in each reaction and the locus designation are shown on the top of the panel. (A) insertion present in all OWMs; (B) insertion restricted to subfamily Cercopithecinae; (C) insertion restricted to tribe Papionini; (D) insertion restricted to the genus *Macaca*; (E) insertion shared between *Macaca mulatta* and *Macaca fuscata*; (F) insertion restricted to *Macaca mulatta*.

Table S1. Endogenous retroviruses that have invaded OWM genomes

Name	Class	Closest related retrovirus [a]	Subfamilies [b]	Copy No. [c]	Divergence [d]
MacERV1	I (γ)	PtERV1 (98%)	10	473 (259)	0-3%
MacERV2	I (γ)	MacERV1 /PtERV1 (93%)	6	528 (276)	0-3%
MacERV3	I (γ)	BaEV (97%)	6	395 (261)	0-3%
MacERV5	I	HERV-Fc2 (73%)	3	177 (157)	4-7%
MacERV6	I	MacERV5/HERV-Fc2 (67%)	7	296 (256)	5-8%
MacERV4	II (β)	Simian Retrovirus 1a (94%)	5	329 (165)	0-2%
MacERVK1	II	Mys-ERV_RN (68%)	6	356 (335)	3-4%
MacERVK2	II	MacERVK1 (69%)	6	203 (129)	1-3%

[a] The nearest known neighbor is given with the percent identity of the Pol protein that contains, among others, the reverse transcriptase.
[b] Number of subfamilies for which a consensus is deposited in the repeat database created on the basis of the LTR sequences.
[c] Estimated total number of recognizable remainders of the elements in the rhesus genome. In parentheses is the number of copies only represented by a solitary LTR.
[d] The range in substitution level of the copies compared to the appropriate consensus sequence.

Table S2. Classification, copy numbers, and age estimates for OWM-specific L1 elements

Subfamily Classification	Proportion of OWM-specific L1 elements (%)[†]	Number of OWM-specific L1 elements[‡]	Age ±SD (Myrs)
L1CER-1A	10.9	2076	33.1 ± 0.6
L1CER-1B	9.5	1806	27.7 ± 0.5
L1CER-1C	2.8	540	27.0 ± 0.8
L1CER-1D	2.1	399	25.8 ± 0.9
L1CER-2Aa	4.3	809	22.9 ± 0.7
L1CER-2Ba	0.8	152	23.6 ± 1.3
L1CER-2C	1.4	258	22.2 ± 1.0
L1CER-2D	1.1	211	21.3 ± 1.0
L1CER-2E	8.8	1677	21.2 ± 0.5
L1CER-2Ea	1.1	211	22.2 ± 1.1
L1CER-2F	0.9	176	21.2 ± 1.1
L1CER-2G	0.8	152	20.9 ± 1.2
L1CER-1E	3.8	727	26.3 ± 0.7
L1CER-1F	0.9	164	21.2 ± 1.2
L1CER-3A*	2.7	516	21.9 ± 0.8
L1CER-3B*	1.0	188	18.3 ± 1.0
L1CER-3Ba*	2.3	446	19.5 ± 0.7
L1CER-3C*	1.0	199	18.2 ± 1.0
L1CER-3D*	1.8	340	18.6 ± 0.9
L1CER-3Da*	1.7	328	16.1 ± 0.7
L1CER-3E*	2.3	434	19.3 ± 0.8
L1CER-3F*	7.9	1501	14.0 ± 0.4
L1CER-3G*	6.4	1208	11.1 ± 0.4
L1CER-3H*	7.8	1490	9.8 ± 0.4
L1CER-4A	1.1	211	14.4 ± 0.8
L1CER-4Ba	1.0	188	10.1 ± 0.8
L1CER-4C	0.6	117	7.0 ± 0.9
L1CER-4D	2.3	434	6.4 ± 0.4
L1CER-4Da	2.2	422	6.2 ± 0.7
L1CER-4E	6.0	1149	6.3 ± 0.4
L1CER-4F	1.5	281	4.9 ± 0.5
L1CER-4G	1.0	188	4.6 ± 0.5

[†] On the basis of 1620 Old World monkey-specific L1 elements.
[‡] Estimated from 19,000 Old World monkey-specific L1PA5 elements.
* 18-bp duplication in the 3' UTR, grey shaded

Table S3. Genomic locations of L1 element 5'transduction sequences

ID	locus TD^	appr°. size	original locus	TD^ gene	into gene
101344	rhesus.chr13:38138336-38138577	242 bp	rhesus.chr1:46259387-46262337	MPL	--------------
145670	rhesus.chr2:80335717-80336029	311 bp	rhesus.chr12:35496393-35534319	SLC25A12	ERC2
210588	rhesus.chr11:80418102-80418690	589 bp	rhesus.chr2:189379179-189384466	TFRC*	--------------

^ TD stands for TransDuction
° approximate size of transduced sequence (locus TD)
* provisional

Table S4. Classification, copy number, age, and insertion polymorphism levels for OWM-specific *Alu* subfamilies

Subfamily Classification	Estimated Copy Number	Age (Myrs)
YRa1	~30.2% of OWM-specific *Alu*Ys	9.5
YRa2	~25.1% of OWM-specific *Alu*Ys	8.6
YRa3	~30	3.5
YRa4	~130	1.6
YRb1	~60	3.2
YRb2	~165	2.0
YRb3	~184	1.2
YRb4	~110	1.0
YRc1	~120	2.1
YRc2	~70	1.0
YRd1	~14	1.9
YRd2	~30	1.7
YRd3	~150	1.7
YRd4	~33	1.2

Table S5. Primate DNA samples examined in this study

Species Names	Common Names	Origin	ID
Allenopithecus nigroviridis	Allen's Swamp Monkey	IPBIR [a]	PR00198
Aotus trivirgatus	Three-striped Owl Monkey	ATCC [b]	CRL1556
Chlorocebus aethiops	African Green Monkey	ATCC	CCL70
Homo sapiens	Human	ATCC	CCL2
Lagothrix lagotricha	Wolley Monkey	Coriell [c]	AG05356
Macaca nemestrina	Pigtailed Macaque	Coriell	NG08452
Macaca fuscata	Japanese Macaque	IPBIR	PR00782
Macaca mulatta	Rhesus Macaque	Coriell	NG07109
Papio Anubis	Olive Baboon	SFBR [d]	1X2816
Pan troglodytes	Common Chimpanzee	IPBIR	NS06006
Pongo pygmaeus	Borneo Orangutan	Coriell	AG05252A
Semnopithecus entellus	Hanuman Langur	IPBIR	PR00739

[a] Integrated Primate Biomaterials and Information Resource, http://www.ipbir.org
[b] From cell lines provided by the American Type Culture Collection, P.O. Box 1549, Manassas, VA 20108.
[c] Coriell Institute for Medical Research, 403 Haddon Avenue, Camden, NJ 08103.
[d] Southwest Foundation for Biomedical Research, http://www.sfbr.org

Table S6. Old World monkey-specific L1 insertion amplification conditions

Locus name	Locus(rheMac2)	Primer-f	Primer-r	Filled	Empty	A.T.
B2-967	chr2:136159913-136161030	CAAAATCTGAGCCATCTCTGC	GATGGGAAAACCACTGCCTA	2169	981	60
B2-1037	chr2:127816061-127818303	TGCAGTGTGGTGAGACTGGT	CCACTTCCTCCTTCCATCAA	2556	293	60
F11-28	chr11:96903083-96909225	AGGCTGGTTCGCTCTTATTGT	ATACTGGGGCTTCGGAGACTAT	7031	830	55
F11-28_internal	chr11:96903083-96909225	AGACTGGAGCTGTTCCTATTCG	ATACTGGGGCTTCGGAGACTAT	762	x	60
B2-1199F	chr2:115854975-115861032	TGTCTACTGAGATTCAAACAAAGG	TTGCAAATAGGAGGCAAAGTC	6937	458	60
B2-1199F_internal	chr2:115854975-115861032	TGTCTACTGAGATTCAAACAAAGG	TTTTCCAGGTGTTGTGTGTCTC	626	x	60
F9-35	chr9:111006368-111012532	TCTCAGTGTGACCAAAAGAGGA	TTTCTTTTGCCCTAAACAGCA	6694	499	60
F9-35_internal	chr9:111006368-111012532	CCCAGTGAGATGAACCCAGT	TTTCTTTTGCCCTAAACAGCA	415	x	60
F3-52	chr3:159930481-159936588	AATGACCTGGGGAAGAGGAG	CACTGCTTTTTAGCTGTTGCAT	6700	521	60
F3-52_internal	chr3:159930481-159936588	CAGTATTGGGGTGGGAGTTG	CACTGCTTTTTAGCTGTTGCAT	376	x	60
B2-20	chr2:57404374-57406948	TGGCTCCAGTTTGTACTCCA	GCCTAGGTCCTGTTGGGAAT	4496	1926	60
B2-20_internal	chr2:57404374-57406948	AGCATGAAGGGCTGTTGAAT	GCCTAGGTCCTGTTGGGAAT	1496	x	60
F1-21	chr1:150652233-150658375	TGATCCTCAAGGAGTCCACA	CAAACTCCTTTTCATTGCTCTTC	7275	1105	60
F1-21_internal	chr1:150652233-150658375	TGATCCTCAAGGAGTCCACA	CCTGGCTAGGAAAAGGGATT	1080	x	60
B2-952	chr2:140217914-140219791	TGCCAATCAGAAGAATCACTG	GGCTTTGATATTGGGATTGAA	2389	460	60
B2-952_internal	chr2:140217914-140219791	TGCCAATCAGAAGAATCACTG	GAAACTCATTGGTAGCTTGATGG	595	x	60
B2-532	chr2:9756586-9758694	CCATGTGCTTGTCAGATCCA	TGTGGACAAAAATGGCAACT	2568	435	60
B2-532_internal	chr2:9756586-9758694	CCATGTGCTTGTCAGATCCA	GAAACTCATTGGTAGCTTGATGG	487	x	60
B2-307	chr2:85512999-85518304	GGGAATCAGACAGACATGGA	GGGCCTTGACAtTTCTTCAG	5821	500	60
B2-307_internal	chr2:85512999-85518304	GGGAATCAGACAGACATGGA	GCTTCACCTCTCGCTGGTC	460	x	60
F8-01	chr8:40279440-40285577	AAGACTGTGAAAGATTGCTTAGTACC	AGGAAATGGTTGTGGTCAGAGT	6400	228	60
F8-01_internal	chr8:40279440-40285577	AAGACTGTGAAAGATTGCTTAGTACC	GGCCAGGTGTGGGATATGAT	544	x	60
F18-15	chr18:47115670-47121793	ATGGCCACTTCAATAACCTGAC	AGGCTTAAGGTTGAAGAGTAGGA	6905	760	60
F18-15_internal	chr18:47115670-47121793	ATGGCCACTTCAATAACCTGAC	GTGTCTCAGTTCCCCTGGTTAG	789	x	60
F6-96	chr6:28885043-28891192	AGTAATATTTAACCAGAAGAAGGCATA	CCAACTTTAAACATTTATGGGATT	7047	875	60
F6-96_internal	chr6:28885043-28891192	AGTAATATTTAACCAGAAGAAGGCATA	CGGGTAGATCACAGACTGCT	971	x	60
F1-22	chr1:149820881-149827027	AATCTTTCCAAATTGACCCTCT	TAAACAAACCTGGATTCTAGCC	6414	237	60

F1-22_internal	chr1:149820881-149827027	AATCTTTCCAAATTGACCCTCT	CCTGGCTAGGAAAAGGGATTC	318	x	60
F5-17	chr5:74171085-74177228	AAGCCAGGCAGTGGACTATTTA	GAGATCAGTGGATGTTGAAGTGA	6754	582	60
F5-17_internal	chr5:74171085-74177228	AAGCCAGGCAGTGGACTATTTA	CCAGGTGTTGTGTGTCTCAGTT	549	x	60
F8-52	chr8:115837150-115843276	GTCCTGATCTTATCCCCCAAAT	GTGTGTGTGTCCATAAAATCAAA	6667	829	60
F8-52_internal	chr8:115837150-115843276	GTCCTGATCTTATCCCCCAAAT	GCCTTGCAGGTAGATCACAGA	705	x	60
F13m-21	chr13:82065455-82071576	ACACCTATGTGGGACACCAAGT	CAGGGTTTTATTCCAACTTTGAT	6532	391	60
F13m-21_internal	chr13:82065455-82071576	ACACCTATGTGGGACACCAAGT	TCCAGGTGTTGTGTGTCTCAGT	557	x	60
RSF20-190	chr20:16609048-16615027	TCTAGCTTTGATGGTTGCCATT	CCTACTTCAAAACCCCGTTGTA	6877	895	60
RSF20-90_internal	chr20:16609048-16615027	TCTAGCTTTGATGGTTGCCATT	gtgggatatgatctcctggtgt	559	x	60
F12-30	chr12:52434457-52440595	AGGCACAGAATGCATTTCATAAT	TCATAAGATTTTAGGCCAGGACA	6979	982	60
F12-30_internal	chr12:52434457-52440595	TTTTCCAGGTGTTGTGTGTCTC	TCATAAGATTTTAGGCCAGGACA	877	x	60
F7-12	chr7:82410190-82416340	AGACtCAAACAAGATAAGGTGGT	CAATCTAGTTGCTGGAATGTCAAC	6762	554	60
F7-12_internal	chr7:82410190-82416340	AGACtCAAACAAGATAAGGTGGT	TTTTCCAGGTGTTGTGTGTCTC	496	x	60
F17-18	chr17:85155657-85161788	TGAGGGTTTCATATTAGAAATGACC	TGACAATTTAGAAGTTAATGCTCCT	6401	235	60
F17-18_internal	chr17:85155657-85161788	TGAGGGTTTCATATTAGAAATGACC	CCCAGCAGCTTTGTTTACCTAC	501	x	60
B2-166	chr2:167384222-167385359	AGTGGGGTCTTAGTTCCTCTCA	TTCTCTAAACCCAGCCATCATT	1421	267	55
B2-166_internal	chr2:167384222-167385359	CATGCCTATGTCCTGAATGGTA	TTCTCTAAACCCAGCCATCATT	178	x	60
F12-47	chr13:29871791-29877879	TTTGCCCTGATTTCTTTTTCTTA	GCAAATGACACAAACAGAACAA	7072	962	60
F12-47_internal	chr13:29871791-29877879	TTTGCCCTGATTTCTTTTTCTTA	TGAACCCAGTACCTCAGTTGAA	893	x	60
F17-22	chr17:11080232-11086407	AAACAAGCTCATTTCACAAGCA	TCCACTTGCCACCTGTGA	6831	628	60
F17-22_internal	chr17:11080232-11086407	AAACAAGCTCATTTCACAAGCA	TTTTCCAGGTGTTGTGTGTCTC	457	x	60
F11-19	chr11:71039770-71045953	GCTGATTAGATAGGCAAAATTACAA	CACAGCAAAACATATGTCACTCAA	7398	1211	60
F11-19_internal	chr11:71039770-71045953	GCTGATTAGATAGGCAAAATTACAA	TTTTCCAGGTGTTGTGTGTCTC	669	x	60
F4-26	chr4:96051807-96057947	ATGTATCTGGATCTCCAATGTTTT	GGGGTAATTGAGGACCAAAAA	6574	359	60
F4-26_internal	chr4:96051807-96057947	ATGTATCTGGATCTCCAATGTTTT	CCCAGCAGCTTTGTTTACCTAC	597	x	60
F5-64	chr5:127087452-127093594	CAGGGTCTAAGCAGCTAGAGTT	GAGGTCTTATGTAGAAAAGGAATTTG	6664	494	60
F5-64_internal	chr5:127087452-127093594	CAGGGTCTAAGCAGCTAGAGTT	GTGGGATATGATCTCCTGGTGT	567	x	60
RS20-179	chr20:75703098-75708402	GGAATGCCTAAATGCAAGAAAG	AAGGCTCCAAACGTAGTATCCA	5653	219	60
RS20-179_internal	chr20:75703098-75708402	GGAATGCCTAAATGCAAGAAAG	gtcaaagtcattctccgtccag	433	x	60

20

F15-39	chr15:52288351-52294492	GCGTACTTCCCAATAAGGGACT	TTGAGAACAgAAGGAAAGGAAA	6655	481	60
F15-39 _internal	chr15:52288351-52294492	GCGTACTTCCCAATAAGGGACT	CTCAGATCTCAACCTCCGTGTT	731	x	60
F15-40	chr15:48908166-48914320	CATTATGCTACAGTGTCTTCTTTGG	AAGTCAGAAGAGACAGGCAAGG	6599	405	60
F15-40_internal	chr15:48908166-48914320	CATTATGCTACAGTGTCTTCTTTGG	CCCAGCAGCTTTGTTTACCTAC	734	x	60
B2-216	chr2:174254655-174257304	TTTACCTATCTCAAGATCAGACA	CATTATTTCCTGTGATCCAGAGC	2900	240	60
B2-216_internal	chr2:174254655-174257304	TTTACCTATCTCAAGATCAGACA	ATTGATTTGCGAATGTTGAACC	470	x	60
B2-238	chr2:178237344-178241475	GAATAATCAACCTGGGCATCTT	TGGAGGAAAACATTAGAAATAGCAA	5831	611	60
B2-238_internal	chr2:178237344-178241475	GAATAATCAACCTGGGCATCTT	GTGTGCCTGTTTGCTTAAAGTG	517	x	60
F14-11	chr14:80260668-80266851	AATATGGGAGCAGAAATGGAAA	GACATAATGAGTCTGGACAGCAA	6654	459	60
F14-11_internal	chr14:80260668-80266851	AATATGGGAGCAGAAATGGAAA	TTTTCCAGGTGTTGTGTGTCTC	469	x	60
F8-14	chr8:57889231-57895368	CCCCAAGACCAGATAAACTCAG	TCATGGAATGGATGTGAGTAGG	6608	442	60
F8-14_internal	chr8:57889231-57895368	CCCCAAGACCAGATAAACTCAG	ACGATATCGCAGACTGCTGTG	750	x	60
F8-60	chr8:134181890-134188053	TGGGCTCAAGGAAAGTATTGTT	CAAGAAAAGCTGCTCTGGTTTC	6946	752	60
F8-60_internal	chr8:134181890-134188053	TGGGCTCAAGGAAAGTATTGTT	CCTGGCTAGGAAAAGGGATTC	480	x	60
F17-20	chr17:88273756-88279868	TCCTATTAGATTCCCTCTGGATT	TGGGCTTAAGTCAACAGTGAGA	6886	748	60
F17-20_internal	chr17:88273756-88279868	TCCTATTAGATTCCCTCTGGATT	CTTGCCGGTAGATCACAGACT	517	x	60
F18-22	chr18:8462589-8468729	TGATGATGAAGTGCAAAAGTGA	TTGGTCCCTAATTCATTGGTCT	6590	426	60
F18-22_internal	chr18:8462589-8468729	TGATGATGAAGTGCAAAAGTGA	TTTTCCAGGTGTTGTGTGTCTC	432	x	60
B19-33	chr19:35334012-35339465	GGACATGTCAACCCAGCACT	AGGTTGCTATATGTGGAAATGC	8382	220	60
B19-33_internal	chr19:35334012-35339465	GGACATGTCAACCCAGCACT	TACCCGATTTTCCAGGTGTT	455	x	60
B2-676	chr2:24846994-24852918	TTGAAAAGTATGTCTAGAGCCTTG	GTAAACTTTTTGACTGGAACAACAT	6565	592	55
B2-676_internal	chr2:24846994-24852918	TTGAAAAGTATGTCTAGAGCCTTG	TTTTCCAGGTGTTGTGTGTCTC	749	x	60
RS10-72	chr10:21635932-21637144	GGGAGAGAAGAGGCAAAAGATT	TAGTTTTGCTCCTGGGAGATTG	1492	291	60
F11-5	chr11:41704911-41711036	CAACCTGTGCATTACTGAGTGG	TGGGTTAGAATTTGGACTCATT	7076	900	60
F11-5_internal	chr11:41704911-41711036	CAACCTGTGCATTACTGAGTGG	TTTTCCAGGTGTTGTGTGTCTC	743	x	60
F6-88	chr6:160527776-160533920	ATACATAGCCCAATGCCAATTC	ACCATcCCCATAGTGTCATTTT	7344	1161	60
F6-88_internal	chr6:160527776-160533920	ATACATAGCCCAATGCCAATTC	CTGCTTCAGCTCTCACTGGTC	760	x	60
F8-76	chr8:21031146-21037295	TTTTCCTGATCCATACCCTTTG	GCATCTAATCTGCCTGTTTGG	7208	789	60
F8-76_internal	chr8:21031146-21037295	TTTTCCTGATCCATACCCTTTG	CCAGGTGTTGTGTGTCTCAGTT	799	x	60

F15-31	chr15:60982774-60988919	GCATTATGTTGCCAGTTACTATCC	ATGGTGACTAACCCAAATCCTG	6457	283	60
F15-31_internal	chr15:60982774-60988919	GCATTATGTTGCCAGTTACTATCC	CTTGCCGGTAGATCACAGACT	520	x	60
F5-25	chr5:67090642-67096850	TATGGCATGTTCAGCAGACAG	GGAGTTTCTAGTTTCCCTGTaC	6584	364	60
F5-25_internal	chr5:67090642-67096850	TATGGCATGTTCAGCAGACAG	CCCAGCAGCTTTGTTTACCTAC	713	x	60
F9-17	chr9:73093546-73099685	CATTTACaAGAAGCTGGGCTTT	CAGAAAGGTTTTGCCAAGCTAC	7702	1546	60
F9-17_internal	chr9:73093546-73099685	CATTTACaAGAAGCTGGGCTTT	TTTTCCAGGTGTTGTGTGTCTC	626	x	60
RS20-77	chr20:13880546-13881451	TCAAACAATCTTCTGGCTCAAG	TTATGGGATTCTGCCTGGAAC	1589	649	55/60
B2-231	chr2:177046802-177049141	GTTTGTATTGAGCACCAGTCTTT	GCTTGAGTTTAGGAATCAGACAGA	2476	125	60
F4-61	chr4:115780346-115786520	TTaATAAAGTGGGCTTTGCTGT	TGTCTTTGAAATCACCATTAACA	6573	371	60
F4-61_internal	chr4:115780346-115786520	TTaATAAAGTGGGCTTTGCTGT	TTTTCCAGGTGTTGTGTGTCTC	424	x	60
F6-04	chr6:42158479-42164633	TGTGAAACATTATCTATGCTGTGA	GTACTTGAAAAGGGGAGTCCTG	6597	406	60
F6-04_internal	chr6:42158479-42164633	TGTGAAACATTATCTATGCTGTGA	GGGATATAATCTCCTGGTGTGC	649	x	60
RS20-159	chr20:63254080-63260150	GCTTTTAAGAGCAAtGGtCCAc	CTGAATACATCAATCAGTTCTCTCA	7201	1206	55
RS20-159_internal	chr20:63254080-63260150	GCCGGTAGATCACAGACTGC	CTGAATACATCAATCAGTTCTCTCA	982	x	55
B2-121	chr2:163631736-163636835	TCCACTGCCTCaAATTCTCTC	TAGCCTCTGCAACCCAGAGT	7646	2479	60
B2-121_internal	chr2:163631736-163636835	CTGACCAGCACCGATTGTC	TAGCCTCTGCAACCCAGAGT	465	x	60
B2-641	chr2:21293134-21294138	GGGAAACTATTAAGAAGCTGTTGA	CCTTATACCTAGGTTTGTTAGGTCA	1530	327	60
B2-84	chr2:159578602-159580174	TGAGGCATCAATTAAGCAGAAA	GGGTTCAGTTCCTGGCTGT	1910	229	60
F12-41	chr12:74839861-74846011	CTACTGAAGCTGAAGTGGCAGA	AACAATTCCCACCTCTATGCTA	6764	599	60
F12-41_internal	chr12:74839861-74846011	CTACTGAAGCTGAAGTGGCAGA	CTCAGATCTCACCCTCCGTGT	847	x	60
F11-01	chr11:40238337-40244500	GCCAAGTTATAGAGGCTGCAC	GCATTTCCTTAGcAGGCAAGT	6913	737	60
F11-01_internal	chr11:40238337-40244500	GCCAAGTTATAGAGGCTGCAC	TTTTCCAGGTGTTGTGTGTCTC	479	x	60
B2-690	chr2:26873840-26880072	TGTGGAGTTACATATGCTTCCTTT	TGTTTTGATGAAGGCACCTG	6789	542	60
B2-690_internal	chr2:26873840-26880072	AAATCACCAGTCTTCTGTGTCG	TGTTTTGATGAAGGCACCTG	577	x	60
B2-670	chr2:24030143-24031185	CTGATTGGAGAATTAATGGGAAAC	TCTTCCTTAACATATCCTACCACTG	2291	388	60
B2-670_internal	chr2:24030143-24031185	CTGATTGGAGAATTAATGGGAAAC	CATCACGCTACCTGACTTCAAA	641	x	60
B2-620	chr2:19297004-19303163	GCATCACTTGCTTCTCAGCTT	TCCCCACACTAAGAATATTGGAA	6825	633	60
B2-620_internal	chr2:19297004-19303163	ATCACCGGTCTTCTGTGTCC	TCCCCACACTAAGAATATTGGAA	622	x	60
B2-495	chr2:6985479-6986856	CTGTGACCATAAATGTGTATAGAGC	ATAGGCCACTAAAGGAAAATCTC	2089	693	60

B2-495_internal	chr2:6985479-6986856	CATGCCTATGTCCTGAATGGTA	ATAGGCCACTAAAGGAAAATCTC	304	x	60
B2-455	chr2:915687-917194	TCAAAGCTCTTGGTAAGCAATG	TGTACCACAGGGCACAAAAA	2954	159	60
B2-455_internal	chr2:915687-917194	TCAAAGCTCTTGGTAAGCAATG	CAGATGGCTGTAGATGTGTGGT	268	x	60
B2-295	chr2:91209855-91211129	GGCTGGGTGGTCCATAGATT	GGGTACAAACTGCTTTAACTTGC	1523	227	60
B2-37	chr2:54692352-54698611	GAGAAAATAGGTGTAAAGACACTGC	TGCTCAATGCCTTGGAATAA	8818	2540	55
B2-37internal	chr2:54692352-54698611	GAGAAAATAGGTGTAAAGACACTGC	GGAGTTGGAGACTGGAGCTG	2017	x	60
B2-14	chr2:57224018-57226503	GGGCCATTTTCTCACTGGT	AATGGTGCTTCACCTGTTCC	3503	1376	60
B2-14_internal	chr2:57224018-57226503	ATGTTGAACCAGCCTTGCAT	AATGGTGCTTCACCTGTTCC	410	x	60
B2-672	chr2:24349683-24355895	TGTGGTTCTTGTGGAAGAGG	GCTGACAATGGTCACAGGAA	6880	438	60
B2-672_internal	chr2:24349683-24355895	TGTGGTTCTTGTGGAAGAGG	CGATTGTCCAGCACTCCCTAGT	422	x	60
B2-150	chr2:165010713-165012091	TGGCTTTCAAATTTGCTTGA	GGCACATATAAGGCTAAGTACTGTT	2579	143	60
B2-150_internal	chr2:165010713-165012091	CCATAAAAACCCTAGAGGAAAACC	GGCACATATAAGGCTAAGTACTGTT	200	x	55
B2-498	chr2:7114005-7115311	TTCCTGGAGGTCAGGAGTTTT	CCAAAATATACAGCTCTTTCCATTC	2069	749	60
B2-498_internal	chr2:7114005-7115311	TTCCTGGAGGTCAGGAGTTTT	tagaaagctgaaactggaacct	630	x	60
B2-181	chr2:169285295-169287669	AGAACTGCAAATGAAAGGGGTA	ACAGCGGACAGATGAAAGGTAT	2920	490	60
B2-181internal	chr2:169285295-169287669	AGAACTGCAAATGAAAGGGGTA	ATGAAGGGCTGTTGAATTTTGT	452	x	60
F6-73	chr6:125231337-125237477	CTATTCCCAAAATTCCTGACCA	AATAGACAATTTGTGCCAATGC	6469	295	60
F6-73_internal	chr6:125231337-125237477	CTATTCCCAAAATTCCTGACCA	TTTTCCAGGTGTTGTGTGTCTC	503	x	60
B2-1065	chr2:125427912-125429447	GAAATGTAAGCTCTCAGGTCCAA	AACAGATGCCTCATAGGAGAAG	2200	652	60
B2-1065_internal	chr2:125427912-125429447	GAAATGTAAGCTCTCAGGTCCAA	GGCACATATACACCATGGAATG	345	x	60
B2-959	chr2:139013607-139020063	GGGTCTCAAAGTTACCATAGGC	TGGCTGAGAAGTTGATAAAGTG	7327	880	60
B2-959_internal	chr2:139013607-139020063	GGGTCTCAAAGTTACCATAGGC	TTTTCCAGGTGTTGTGTGTCTC	419	x	60
F5-72	chr5:141809395-141815543	TCTGGGCTCAGGTGAATAATTT	ACTGCTTCCTGGGCATCTT	6747	575	60
F5-72_internal	chr5:141809395-141815543	TCTGGGCTCAGGTGAATAATTT	TTTTCCAGGTGTTGTGTGTCTC	394	x	60
F4-16	chr4:82524744-82530915	CGTCACTTCCCTACAACTGTGT	AAAAGGGAAGACCCAAAATCTC	6879	704	60
F4-16_internal	chr4:82524744-82530915	CGTCACTTCCCTACAACTGTGT	TTTTCCAGGTGTTGTGTGTCTC	439	x	60

Table S7. Old World monkey-specific *Alu* insertion amplification conditions

Locus Name	Locus(rheMac2)	Primer-f	Primer-r	Filled	Empty	A.T.
0008WA_1	chr20:79917196-79917477	AtGACAGAAAATCAtGGAAGCA	GCTTATCcATCATCCTCATTCTTT	486	176	55
0008WA_2	chrX:136343561-136343842	GTGATCTTTTcCTATGCCCAAA	TTTGAGAAGCAGCtCAGATCC	393	82	55
0008WA_3	chr9:68868923-68869204	TTGAGAGTTTTCCCAAAAGATT	TCAAAAGTTGCAGGTAAGGATTG	584	278	55
0008WA_4	chr4:21355410-21355691	TGCAGGGAAGGCTTTACATATC	AcCTTttGTATCGGCTAATGACTC	382	73	55
0008WA_6	chr4:112161335-112161615	ACGGCAACCACATTTTATGAAC	TGAGAATAATCCAGGAAGgATCA	399	94	55
0008WA_7	chr4:150317515-150317795	ATGCATAGCAAGTTATGCTCCA	AGAGCTATACCCTCCCTGCTTC	484	146	55
0008WA_8	chr20:85894060-85894342	TCAcAGGCcGGGAATTTCT	GGATcATGTGGTGATTCTtTGTT	388	81	55
0008WA_9	chr6:167789848-167790128	AAaACCATTAAATGCAGCcATC	GCCAAAATCATCCTAACgTGAA	568	260	55
0008WA_10	chr5:82304128-82304408	AGTagGTGCTAAGGGGAGAcAG	TGGAAgGGATTTCTGCAACTAT	694	388	55
0008WA_11	chr5:35092333-35092613	TCTGATGAAACCATCCAcATAGA	AATGTTAGgTTCTCAGCAACAGC	847	543	55
0008WA_12	chr2:102533808-102534088	TCAGTCCCAAGCATgTAGTGAG	CTTCCATAAGtgATGCGATTCC	394	88	55
0008WA_13	chr15:84951098-84951378	AAACAGGGTTAAATGGTTTCCA	TAATGGGGACTGTTCaGGTTTC	498	199	55
0008WA_14	chr7:25418682-25418962	CTCTGACAGTGTAAAAGAGATCTTCA	TGATTGGTTTaaTGGTCTTTTGG	400	95	55
0008WA_15	chr6:158845392-158845673	ATGcaCTAGCaTTTCCTCTCAc	ATTGGCTCACCATTATGTTTcC	763	456	55
0008WA_16	chr3:96656155-96656435	AATGGCCAGAGCTACCATAAGA	AACCGTAAACAAATtGCTTCAA	398	90	55
0008WA_17	chr3:89946564-89946844	GCATGTTAGGTATACcAGGGATTT	TCAAGGCCAGATTTATTCACCT	463	162	55
0008WA_18	chr2:127512009-127512289	TCTCTTTTATGAAGATGTAAGAATTGC	AATAGAACTCTCCCCAACAGTTCA	386	75	55
0008WA_19	chr15:72420377-72420657	GCACAAGAAAATCTTCCCAAAC	AATGGCTACcAACCCACATTAC	462	153	55
0008WA_20	chr15:71149518-71149798	AAGGtGaTAACATATTTTGAAAGTCC	TATCTATAGCAATGCacGGTCA	400	94	55
0008WA_21	chr10:46386339-46386619	TTTCCAAAGGAGGGTTAGGAAT	GAAggGTAGGAGTGCATTTGTG	376	73	55
0008WA_22	chr1:197677024-197677304	ACAAGCGAAAATCAGATTGAAAA	AGTAAagTGTTTTCATGAAGTAGGTTG	377	75	55
0008WA_23	chr1:18433836-18434116	AGAAGCAGTTAGTCCAtTGGGTA	TCgTTCGTTCTATTTAGAATGCTG	842	515	55
0008WA_24	chr1:763215-763495	AGTGGGTgTATGAAAACGAACA	GGACACAtCTTTTCACTGTCGTC	397	86	55
0008WA_25	chr9:101435100-101435383	TTTTGATGGAATGCTAtTTTCTCTT	TTGAAAATTTGTaTCTAAAAACCCTTA	384	84	55
0008WA_26	chr5:1935965-1936247	GCACCTTGAGGCACAGAAAATA	GAAATTCCCTGAAGACAcCAAG	378	72	55
0008WA_27	chr3:51392110-51392392	CTGCCCCAcATTTCcTTATCT	TTTCTCCCCaCAgCTTTCC	400	87	55
0008WA_28	chr1:192251843-192252122	TGTTTGATAACCCcTTTTCtCC	AACTTGTTGGTCTTCACCCAAG	599	257	55

(continued on next page)

0008WA_29	chr7:136745558-136745838	AATGTTTGCTGTGGACTGAGTG	TTTGAGAATCTGGAACTCAAAATG	561	254	55
0008WA_30	chr2:156277378-156277658	AGCCtGGGATTCTTCAAATGTA	AGATAGTTTCCATGGGGAGGAA	540	229	55
0008WA_31	chr14:47195875-47196155	GAACATAGTTTTATTAACTGTCcCAAA	AATTGcTTGacTTACAaCAGAGGA	392	84	55
0008WA_32	chr11:3214792-3215072	TGAAAtCCCTGCAATATGAAGAT	CTGCCTGCTAGAcAACTGACTT	654	345	55
0008WB_1	chr20:79232410-79232710	TAAGTGGAAACCTGAAtGTGGT	AGAGGAAAGACAaAGGAAGGTG	400	98	55
0008WB_2	chr4:145086603-145086903	CCTCTcCAACCCAGGATACTTA	TTACCTTCAtTGGTTTCCACATT	835	528	55
0008WB_3	chr16:39074914-39075214	CAGTGTTATCAATTGCCAGGAA	TCAAAGTCACTGTGCAGTTCAA	470	159	55
0008WB_4	chr6:159834124-159834424	TGcAGAGTTTCTCAGATTTTcAG	TACAGGTTtTTAGGCTCCATCC	451	140	55
0008WB_5	chr3:57888859-57889159	cGAGCAGTTGATGTGAgTCTTG	TTGAGAaCACCAACCTGTTCCT	561	248	55
0008WB_6	chr1:190999080-190999380	TAGgGGACTGGCAGTAAGTTTC	AGAAAGGGGGTGcCACTTAG	386	75	55
0008WB_7	chr6:10631550-10631850	GGTCACAAGCTAAGGTGCATAA	GAGGCAAGCAaCCTCTAGGA	379	70	55
0008WB_8	chr17:23150063-23150363	ATTGCAGCCAATTCTTAATGCT	CATCTTTGGGAAtCACCTTAACA	800	492	55
0008WB_9	chr13:123620803-123621103	ATCTGCATTGGAATTTGTGTGA	TGCTAAAGTGTCATATTGTACAGAAAA	382	71	55
0008WB_10	chr11:93895728-93896028	CTTGGGGAAACAAATGTAGAGG	CCTTTCCAGAGTCAtGATTTCC	762	453	55
0008WB_11	chr8:58793188-58793489	TGgCAtTGATCAtAAAGGAATG	AGCCATGATTATTGGTGaCCTT	523	211	55
0008WB_12	chr11:81593893-81594194	CACATTTTCCCAACgTGTCTT	ATGACaTGCCTTTCTATAATTTCTC	399	87	55
0008WB_13	chr4:67743297-67743597	GTAGTCcCGGAAGATGAAGAAC	TTTCTTTCCCATTGATTTCTGC	548	237	55
0008WB_14	chr3:157600750-157601050	GGTTCCTcCTgAATTCAAAACA	CAAGATTGcTCTTGTGAAAGGA	645	334	55
0008WB_15	chr2:37127814-37128114	AATAATGCTCAGGGTGGAAGAG	TCAGATgCATTTCATGAGTTCTG	400	95	55
0008WB_17	chr1:181234973-181235273	ACATTGGAAACATcaAAACTGC	ACCCTttTAAGAACCAGGGTGA	694	383	55
0008WB_18	chr1:175423002-175423302	AACAGGGcaGCATTATTCaAAA	ACTTCTgGtTACTCTGTTCCCTTTA	500	196	55
0008WB_19	chr1:120611588-120611888	CAGAtAgCTTCCTTCAtGGCTTA	CAGGTTTaCAAGGGAAAcGATG	600	289	55
0008WB_20	chr1:37900784-37901084	CAACATaTGGCTATAAGAAGAGCAAA	TTGTTGGAAAAGGAAGTAATGGA	398	87	55
0008WB_21	chrX:35896473-35896773	AGcATATCAGGAACTTCCCATC	TtAATGCCCTCAtTCTCCTTTC	481	173	55
0008WB_22	chr6:134846227-134846559	TGGAGTCaTTTTAGGAcGCATT	ACTCTGAAggGGAgAGAAATGA	695	396	55
0008WB_23	chr6:37082488-37082787	CAAgTTGGATTAAACACCACCA	CAGAAAGACCCCAAGGTCAGTA	777	467	55
0008WB_24	chr19:14872768-14873068	AGGAGGtTCTGAAGATAACaAATGA	TTATGGTGATCcCCCCTTaTG	700	390	55
0008WB_25	chr18:69623903-69624203	AcTTCTTTtGGGAAACGTGTTG	CCTTAGCTcTTAATCATTCgTTTTG	396	94	55
0008WB_26	chr12:5216493-5216793	CTTCTAAcTGGCACgTgAGTTT	AAGTAGCCTGCAAGAAGGAGGT	679	376	55

0008WB_27	chr1:107489502-107489802	CTCTGAGTTGGATGTTGTAAAgAAAT	ATGAGATGAAGAGCCATGGAAG	372	70	55
0008WB_28	chr20:81097387-81098128	CCAGTGGAAAGAAGCCTAAAGA	CACTAACgTATGGTTCTGaGGAA	399	90	55
0008WB_30	chr2:37127835-37128114	AATAATGCTCAGGGTGGAAGA	CAGATgCATTTCATGAGTTCTG	399	94	55
0008WB_31	chr18:2425655-2425934	ATAAACCTGGGgATGTGcAGAG	CCAGAATCATATGGCTGAAGTG	595	288	55
0005wD_1	chr1:107168847-107169238	AAccCTGTTGGCAAAAATAGAA	ATTTTGGCAGTCAcATAATgTTT	552	241	55
0005wD_2	chr12:26438295-26438699	GTAATTAttCTGGCTGCAAATGA	ATTAcAATGTATGGGCCTTTG	479	168	55
0005wD_3	chr1:11780838-11781385	TcTCATTTCCTCcTAGCAGTCC	TGTGCcACTTGGtgTACTATTT	545	235	55
0005wD_4	chr3:105448261-105448781	GGAATATGcAAATGGAATATCTG	GAACATAGGACAGGGAACTTGC	697	390	55
0005wD_5	chr5:104378298-104378686	TAGGCATgGGTGATAAAACAGG	CCTTGAAACTGGGTTGCTAGTG	517	209	55
0005wD_6	chr10:73340050-73340453	CCTTCtCAGAATGtgGACAGAG	ATTTCCaCTTCTCCAaTGGTTC	533	224	55
0005wD_7	chr20:70918204-70918932	AAACATTCcCTGTGATCAAAAA	TTCACCaAAcATGTTAGTATTTCCA	1168	867	55
0005wD_8	chr2:111785672-111785976	ACCATTCAGcTGAGAAGCATTT	TaCATTCCAAACCCATACTGAAA	521	210	55
0005wD_9	chr2:83329622-83329926	GAAGAGGTgTCAcATCACATGG	CTCTTGCCAAAATGCTCTCAC	818	507	55
0005wD_10	chr6:108459725-108460022	GAACTTTATTGgCCCTGTACCC	TGGAAGCCCTAAATCAtCCATA	632	321	55
0005wD_11	chr10:46539667-46539970	CCAGTGtATCCTaGCTGCCTCT	AGCACTCCTcAATGAGAAAACC	536	225	55
0005wD_12	chr4:128867766-128868300	TCAAGTAAACCAAGCAAAAGCA	ACATTCTTTACCCcCgTCTTCT	515	214	55
0005wD_13	chr4:71507571-71507912	TTTTCACATTGACCCTTCAGAC	ACCATCAgTCTGTTGTGAATGC	523	216	55
0005wD_14	chr17:26859097-26859823	TGACTGCCTCAGAAAGGGTATT	TAAAGTGAACACACcCCTGAAC	596	296	55
0005wD_15	chr11:134336653-134337074	TGAGGAaGTTTTTGTcGATTTTT	CAAATTTTGCCAAGtTAaAACAAG	818	507	55
0005wD_16	chr17:74625247-74625550	TTGTgCAGGAATTTTGAACACT	AGCAACCACTGATAACCACTCA	699	392	55
0005wD_17	chr14:104986645-104986948	tTTTATGATGAAGGCTCCTGCT	CCATCTCTGCTTCACcCTATTC	587	280	55
0005wD_18	chr7:97475750-97476052	ATCATAAGGTCCTTGTCTTCCA	TCTACAAcGCaCTGCCAAGAtA	481	172	55
0005wD_19	chr20:72373676-72373977	GGTACTAAGTCAAATGGGGATG	GTGCAGACGCACATCCTCTAA	583	272	55
0005wD_20	chr2:145645983-145646289	AAATGCTCAgTTTTGTGCTTGA	GGAAATGTCATGAACCCTCAAT	910	599	55
0005wD_21	chr18:15121225-15121592	CATGTTCAGGACAAGAGGAACA	ATATGCCACAAGAAGGTTGTGA	542	242	55
0005wD_22	chr15:107277828-107278130	CAAGCCAGTTCCTAAATTCCTG	CTGTTGTGCAAATGGCACTTA	579	270	55
0005wD_23	chr11:113160243-113160804	ATCCCAGGAAATTGAGAAGgTA	AAAATGCAACTGAAAAGCCACT	692	391	55
0005wD_24	ch1:12061174-12061939	CTTCAcTcACCTACACCTGCAC	CTtGAATGGGaTGACTTCTGGT	522	216	55
GROUP6&7_1	chr17:57245605-57245885	GTTTTGCCAAACCTGACAGAAT	CACTCCCAACATCCTCAGaACT	422	116	55

GROUP6&7_2	chr11:156172-156452	GCTTATGGAATGTGGAGTGTCA	TTGTACTTGCTGCTCCTTCTTG	542	231	55
GROUP6&7_3	chr20:66591376- 66591656	TTTAAAGgGTGAATGGAAGCaC	GGCAGAACTAACTTTCCCTTCC	808	497	55
GROUP6&7_4	chr2:56354243-56354524	TGATGCTCTGGTTTTGTAATGG	AAGGATTTAcGTTCTGCAAATGA	560	248	55
GROUP6&7_5	chr14:29564411-29564691	ATATGGaGAAAGATcGGGgTTG	CACAAAGCAAATTCCCATGaC	583	284	55
GROUP6&7_6	chrX:73256176-73256456	TGCTTAATGTTCCCACTTTgTG	CCTGCCCAGAATaCCAGATaAC	554	244	55
GROUP6&7_7	chr9:16694481-16694762	CTGAAACCCATACCATCAGACA	gCTTAGtTTTTCCAAACaCCAGT	524	219	55
GROUP6&7_8	chr7:119923545-119923825	GCATTTCCTAGTTTCCCTTTCA	AGAATCCATGtTTCTTGGGTTG	761	448	55
GROUP6&7_9	chr5:30341453-30341739	TGTTACaGTGAACTAGcGGTCAA	TGGGTTAGTGAACTTGGACTCA	619	291	55
GROUP6&7_10	chr14:62305284-62305565	AGGTTTAGAAGGTGCCACCAA	GCgAACATCCTAACACATCAAA	623	319	55
GROUP6&7_11	chr8:53430137-53430418	CATCTAagTTTTCCATTTCCCTTT	TGTTGGgATAGGATTGTCATTG	818	508	55
GROUP6&7_12	chr13:72924839-72925130	GGATGTCTAAATCTTTTTAGGCaCAG	CTTTGAAAAGACACCTGAAAGG	846	524	55
GROUP6&7_14	chr15:92103752-92104033	TATTGCCCCCAgAtAATCAGTT	TCTGGAGAGCAATTAGGAGCTT	674	362	55
GROUP6&7_15	chr13:88748284-88748564	AAGGAaGAAgAGGGTcATGCTT	TACCTCCGCTTTTCACCTTCTA	600	295	55
GROUP6&7_16	chr12:4175791-4176071	TGTTCTTTTGGaGGTCTCCTTT	TTTACTTAAATTCAGATACCATTTCg	552	218	55
GROUP6&7_17	chr11:106137834-106138115	TGCCTGTATCCATGAgtGTGAT	ACCCAGGATTGTGAACTGATGT	816	504	55
GROUP6&7_18	chr6:111112629-111112909	TGGAGTAcaACCACATGCtAGG	CAGTGaAACAGGTTCTTCTGAAAG	569	267	55
GROUP6&7_19	chr2:31276369-31276650	ACAGTGATATTATgGACACTGACATT	AGGTgAAaGAATCAAGCCCTTT	458	142	55
GROUP6&7_20	chr17:43661067-43661348	GCCAaAAAGAGTGATAACTAAAGACAC	GGAGGGAAATGGAGTTACaGAA	700	388	55
GROUP6&7_21	chr10:65330510-65330792	CCCTGGAACATGTCTTGGTTAT	TCTTGGGGGACATAgATAAaCCT	681	369	55
GROUP6&7_22	chr4:83191-83472	GCCACCTCTgGATTACTGCTAT	CCTCTTTCaGCTCTGGGAATTA	637	327	55
GROUP6&7_23	chr3:145546613-145546890	ATCAGGTATAACATTTCCGTTGATG	AGcTcAATTTCCCTAAGTCCTG	518	210	55
GROUP6&7_24	chr12:98325091-98325372	ATGcTGGGAacCTTCTGAGAGT	AaGCATTTCCACCATTGTCTCT	824	512	55
YRa3_1	chr2:188762670-188763952	CACGCTAAGAAATCAACATTGG	GAAAAGAGCTGATGCATGTTTG	575	263	60
YRa3_2	chr2:101577949-101579231	GTCCTCTTGCAGCTGCTTTC	ACTGACTCATAGGCCCTGACTG	569	275	60
YRa3_3	chr6:146115980-146117262	AATATGCAAGGAGGCAAACTGT	GCAGCAGGGCTTTTCTGAT	983	672	60
YRa3_4	chr5:35880344-35881626	ATAGGGATTTCGTCACTTGGAA	TGAGTTTGGCTTCCTGTTATTG	501	196	60
YRa3_5	chr19:14926875-14928157	TAACAGTTTGAAGTGCGGTGAG	GCTCAGGGACCAATTGTAGAGA	562	257	60
YRa3_6	chr1:171600659-171601941	CCTGAAAGGAAATGATCCCATA	GGTGATGTGAAGGTCCCAAAAT	511	183	60
YRa3_7	chr7:6583650-6584932	ACAATAAATGCTTTAAGTGTGCTTT	TCTGAAGAAAGGTGCTCTGTGA	571	260	60

YRa3_8	chr11:40596760-40598042	TTTAGCCTTTCTTCCCTGCTCT	GATTCGCCATTTCTCTGATACC	504	201	60
YRa3_9	chr10:31755199-31756481	TGCATGTGGAATGAAAATCAAT	TTAGGGTGTCTGACATCAGTGG	845	544	60
YRa3_10	chr1:135190122-135191404	CTCTAACCCAAGGGCTGATATG	GGAGACCCACAGAGAAATGAAC	542	229	60
YRa3_11	chr1:65212356-65213638	TCTTTGTATCTCAGCCCCTAACA	AGTAGTTGTCCAGCCCCACCT	582	272	60
YRa3_12	chr1:48794973-48796255	TTTATTTGCCTCCCTTGAGTGT	AGAAGAAGCCACATATTAGTGTCA	584	282	60
YRa3_13	chr4:42224434-42225716	TGTGCCAGGTGGAATATTTTAAG	ACTTACGTGAACCATAGGGAAA	681	368	60
YRa3_14	chr12:45162702-45163983	GCTACCCATTGGTTAATTTGTGA	TAATGTTTGCATGCCAGTGAAC	551	239	60
YRa3_15	chr14:110368175-110369722	TATTGGTTTCTCATTTTCTCCTCA	TGGATGGACATGAAAAAGAACC	503	202	60
YRa3_16	chr1:3228297-3229587	GTAAACGGGTCTGACAGGACT	TTTTCCTAATGGACCACACAAA	849	538	60
YRa3_17	chr7:92086999-92088277	TCTGAAAGGCAAAATTATGTGTG	GCCTTCATTGACAAAGCCTAAA	466	150	60
YRa3_18	chr6:49066949-49068232	TGTGAGAAGTTACATTTAAAATAGGTG	GAGGCTAGTTCTGGGGCATAAA	686	383	60
YRa3_19	chr3:89872421-89873703	AAAAATCCTTAGCTTCAAACCA	TTACCTATGAAAGTGCCTTGGA	801	491	60
YRa3_20	chr9:81648754-81650036	GGAATGATGCCTTCATGGTATT	TACTTGGCCAGTGGGTTTTCTA	511	185	60
YRa3_21	chr3:142523866-142525148	CACATGTGGGTAAATACTGGAGTT	CTCTTTCGTGGGCATTTGAC	740	433	60
YRa3_22	chr2:45746736-45748017	GTGTGCGTATTTTGTAGCCAGA	TAGCAGTCTGTGAGGTGCTAGG	560	251	60
YRa3_23	chr15:43332017-43333300	AAATTTCCAGGAATGTTCAGGAT	GAAGGGCAAAAGGTTTCAATC	554	245	60
YRa1_1	chr1:70241118-70242453	CTCATTTTTGCTGATCACCTTC	TCACACTGGATCATCAGACTGC	596	315	60
YRa1_2	chr1:69705631-69706963	TCTGAACCTAAAAAGACCCACTG	GATAAAGAGCAGGGGATCACAG	577	252	60
YRa1_3	chr1:70517606-70518936	TCTAATGACCATTTGAGGAGCA	AAACCCTTTTCAATATGGCTGT	582	261	60
YRa1_4	chr11:27266364-27267677	TCTTTCCCTCTACCATCCAGAG	CTCATTTCTTTCCCATTCCACT	532	229	60
YRa1_5	chr11:28072724-28074027	ATTCTACGGAATGCTTCTCTGG	ATTCAGGAGGTGAAGACTGGAA	512	217	60
YRa1_6	chr11:70443967-70445277	TGACACTGCTTTTGAGCTATGC	TATCCTATGCCAAGGTGGAAAT	556	263	60
YRa1_7	chr11:27076794-27078098	ACAAATGCCAACAGTCAAGAGA	CATCAGGGGGAATATAGCTTCA	516	223	60
YRa1_8	chr14:69969711-69971040	TCACAAACATCCACTAAATGGT	TTATAGTCCACTGATTGGGTGAC	599	303	60
YRa1_9	chr11:28003140-28004438	TTGAGATTGAGAATTGGGAGGT	TGGAAAACAGTTCATTTTCCTT	577	277	60
YRa1_10	chr11:70812596-70813912	TCTCATCAGGAAAACAGAACCA	GGCCCAGCCTAAAGTAGTTTCT	596	287	60
YRa1_11	chr20:60659894-60661227	GAATCAATCGCAGGCTGTTTA	GCAATGTTTTGCTTGATCTTCA	598	284	60
YRa1_12	chr20:61023078-61024387	TTCCCAAGTTCTCCACTAATGA	TATCCCAACCCAAACTGGTAAA	591	282	60
YRa1_13	chr17:74985040-74986374	CTTTGGGCAAACAGATCAATG	GCAGTCAGACCAGTATGTCAGG	627	329	60

YRa1_14	chr7:102069717-102071109	AAAGTTCTAAAACTCACACGGTTTC	TGATATATAATTGGGGGAATTGG	603	289	60
YRa1_15	chr16:69222293-69223600	TGTAACTTGTGCCCCAGCA	AACAGTTCCTTCAGTGTGAGCA	602	304	60
YRa1_16	chr18:73181493-73182793	TTTAAAAATATTGAAACAGTCACAAAG	TTGAGACATTACATTCAGAATTATAGC	601	307	60
YRa1_17	chr19:18020983-18022304	GCCCACAGTCTCTATCTGTGAA	AGACCGTTTCCTCTCAGAAGC	617	317	60
YRa1_18	chr12:52866119-52867437	ACAGCTCTAACAAGTTTGGTGTATT	TGCTGATATTTTTCCAGATTCCT	670	389	60
YRa1_19	chr13:42773257-42774580	CTCTGGTCAAAAGGCAGGTAAC	CAACAGAAGGGGCTTAAAACTG	700	383	60
YRa1_20	chr5:29335675-29336991	CCTCAGCTATTCTCCCATTACG	GGAAATTCTTCAATAAAAGATGATGC	693	383	60
YRa1_21	chr5:29381170-29382514	GAGATTTTCGCCTAAGTCTATATGAT	ACTTCCTTGAGAAAGGGAATGC	518	233	60
YRa1_22	chr5:29010106-29011436	GGAATAAAGCAAGTTAATCGGATA	ACATGAAGAGGAGTCACCCAGT	569	286	60
YRa1_23	chr6:86399529-86400845	ACTTCAGCCATATGGATTTTCA	GATCACTTATCAATCCCCCAAA	681	383	60
YRa1_24	chr6:85927012-85928327	AATCGCTATGACTCTATCTGAACCT	ATTTTCTTCCTCCTGTATTCTACCTTC	603	298	60
YRa1_25	chr4:3764976-3766321	GGAAATATGCTGATGTGACTGG	AAGGCTTAGGGAGAAAACTGAGA	635	300	60
YRa1_26	chr4:98154824-98156161	GGCCAGGGTTCAGAATGTACT	TCCTTTGTAATTTGTGGTAATTATGC	597	286	60
YRa1_27	chr6:138307618-138308949	GACAAGGAAATAGTGGCAGCTT	TAGACCAAGCTTTGTGTGAGGA	778	470	60
YRa1_28	chr6:138200267-138201599	GAATCTATCACATGCCCAATGA	AAACAACGGAACACTGCTTCTT	597	284	60
YRa2_1	chr4:36113979-36115297	TACACCTCCCCGTGAGAGAC	CCTCTTTTTCCAACTTCCCTTT	581	274	60
YRa2_2	chr8:38713758-38715075	ATAGCAGCCTAGATCCTGGTGA	CAGGGTATCTTCCTTGATCAGC	502	195	60
YRa2_3	chr16:49122128-49123445	AGCCCTACCTACTGAGCTTGTG	ATCATGTGACCCCCACAAATAC	531	226	60
YRa2_4	chr10:33751858-33753173	ATGGCAGAATTTGACATTACCC	GTGTCCACTTTCTGCATCTTGA	616	313	60
YRa2_5	chr8:38281061-38282381	TTCTCAGAACAACAGCCCTTCT	GCTTCCTTACATTCAGAGTTTGC	688	380	60
YRa2_6	chr3:28917159-28918499	TTGCACATATCTAGGCATACCA	TTTTGTTACCAGAACATTGAACA	505	197	60
YRa2_7	chr3:29348893-29350299	CCTGTTGGAAAATCACAGTCCT	CTGGTCTACTACACAGCGCATC	522	217	60
YRa2_8	chr14:69653927-69655230	GCCTCAGAGTAAAAGTCCCTGA	TGGCCTCTTGGTTTGGTAGATA	594	289	60
YRa2_10	chr17:36993538-36994869	TCCAGATTTTTCAAAAGCCCTA	ACTTGGCTCCAATTCTATCGTC	651	342	60
YRa2_11	chr13:39423014-39424335	ATCCCTTTACTGCCAGACAAAA	GTGTTGAGTTTCCTCAGCTGTTT	563	253	60
YRa2_12	chr7:124348500-124349829	CCAGTGACTGAAATATTATGCACCT	TGGTTTGGTCTTAATGGGATTT	603	293	60
YRa2_13	chr9:34695819-34697139	ACCTGTTTATCGGAAGCAGAAA	GATTTTATCTGCCAGGCTTTTG	667	358	60
YRa2_14	chr5:35081953-35083286	AAATGCTTGTTAAAGGGAAATGT	GCAGGTCATTACCAAACATTGA	612	299	60
YRa2_15	chr13:39947959-39949278	CCGTTTCCTAGAATGGTGTTGT	CATTTGACAACAATCAGCAGGT	514	203	60

YRa2_16	chr6:43060202-43061529	CAGACAAGGTAGCCATATCAACAC	GAGACACATTAACAGGCCACTTT	636	326	60
YRa2_17	chr17:36462825-36464157	CTTCTTCCACAAAAGCAGATCA	AACAACTGGAAAAGATAGGTCAA	654	345	60
YRa2_18	chr13:40590022-40591333	TTGAATTAGAGGGGCCTTAAAA	TGCAAACAATCATCCAGATACC	532	229	60
YRa2_19	chr19:48535506-48536835	AACTGTGAAGGATTCTGTCTTGC	AGCGTCGGAGAGTTCTGTAGAC	650	334	60
YRa2_20	chr9:34965143-34966475	AAAAGCCAGCAAGTGATTTTGT	ACGGTTATCATCACCATCATCA	501	188	60
YRa2_21	chr13:39644952-39646271	GGATCAGGTCTAAGCCTGTCTG	TGTACCAACCCACTCTGACTTG	538	228	60
YRa2_22	chr17:37290381-37291746	CATATTTGGAGAATGTGCTGACA	TCAATTCATGTGCTGTGGAGTT	602	294	60
YRa2_23	chr15:109618600-109619900	CCAAGTAGAGGACGGAAATGAG	AAAAGGGTGATGAGGCTGTATC	623	312	60
YRa2_24	chr2:93902741-93904050	CTGGGGAACTCAGAATTGTTTC	ACATCCTAGGAACCTGCACCT	561	251	60
YRa2_25	chr17:37615468-37616809	CTCTGAGATGTTGGCCTTTCTT	AAATGATAAGAGGCATGGCAAG	578	268	60
YRa2_26	chr9:34889527-34890839	TAGCAAAAGCAGTAGGGAATGG	CACATCCTGTGTGGTTTGAGTT	699	401	60
YRa2_27	chr3:70605477-70606822	TTTCCTGTGCTTAAATGCTGTT	AATGACGACAGACGATACCTCA	689	379	60
YRa2_28	chr18:38916385-38917731	TCTGATTTGGACAAGCCACATA	TTGTTTTCCTGACGTAATGCAG	805	497	60
YRa2_29	chr5:124168843-124170189	TCCTGAAGGAGTCATACCAATGT	ACCATTCATAGTTCAAGCAGGA	691	380	60
YRa2_30	chr18:39581532-39582864	CCCGGTATTCAAACATAGATGG	TAGTGGTGTGTGACTGAAAGCA	645	336	60
YRa2_31	chr16:48528303-48529627	TTAAGGGTACTCCACTCACCAA	ACCCTTCCTTTGATGTGACCTA	662	354	60
YRa2_32	chr17:36459306-36460647	GTTGAGCTGGACAATGAAACTG	TGAAAACCGAAACTGATGCTATT	613	299	60
YRa2_33	chr11:41054554-41055875	CACAATGTCAAATACTTTCAGACAA	AATCGACACGAATATCCCCTTA	663	354	60
YRa2_34	chr13:40294186-40295490	TCAAAGATCCAGATGTTTTACCA	TTCCAAGACAATTTAGAAGCTATTGA	602	293	60
YRb3_1	chrX:111505418-111507086	GCCAGGTTTAGTCTTTTCCTCA	GTTACTCGCTGGGAAAGAAATG	911	615	55
YRb3_2	chr14:23840559-23841727	TCGCATGTAAATGGATGACTTC	TTCATCTGGAATTTATTTTCATAGTCA	675	387	55
YRb3_3	chr19:39181694-39182827	AGCTCCCAGTCAGTCCTCCTA	TTTGCAGTATGCCACGTAGAAC	561	269	55
YRb3_4	chr14:6421562-6422830	TCTCTTCCTCACTGTGCCAAC	AGGAAGTCATCAAAGGTGCCTA	606	309	55
YRb3_5	chr13:78119460-78120728	GGCCAAAAGTTGAATGTCAGAG	TCTTTCTCCACCAATCAATGTG	502	205	55
YRb3_6	chrX:110809290-110810558	ACCCCCATTTTGAAGTAAACCT	TACCACACAATGTGGCAAAGAC	737	441	55
YRb3_7	chr12:34187335-34188603	AGCAGTCTGAGAAGTCGGAACT	ACCTCACTTCCAGTCATTTTCC	507	214	55
YRb3_8	chr6:143102047-143103315	CATGACCTCTTTGCCCTCTTAC	ACCTGGTGGGACTTCATAGAAC	922	631	55
YRb3_9	chr17:84961553-84962821	TGCAGATAACATGGAGGTAATGA	GGAACTGCCCTAAAGATGCTC	500	205	55
YRb3_10	chr3:13777101-13778369	AAACTCTTGATTTGGGGAGTCA	AGGATGCCATTGAGGAAAATTA	853	557	55

YRb3_11	chr17:9457079-9458345	TCAAGCATTGATAAAGGGAAAA	TGAGCTGCTTTCAGATCTGTTT	484	204	55
YRb3_13	chr5:61040255-61041522	TATTGGGCAAATTATCACCACA	AAGTTCCTTCTCAAATGATTCCA	580	288	55
YRb3_14	chr14:43196373-43197640	TCATTTCACCAAGCTGAGAGATT	ACACTCAGCAGGTACAGAAGCA	679	383	55
YRb3_15	chr16:19167343-19168608	TGGGTGTATGCAGAAGTGACTC	AACTCAGCAGTGCCAATTTCTT	529	236	55
YRb3_16	chr14:851324-852591	CAGCTGGTGTCAAGGCTGAT	AGCTGCCCACTATAGCAAGG	518	223	55
YRb3_17	chr3:116008439-116009706	TTGTTTAGTGTAATTGGGTGAGG	ATGCCCAGGACATAAATTCTCA	629	333	55
YRb3_18	chr6:161342812-161344079	TGGATTGCCTAGAGGAAAAGAA	TGATTCTAGGCTACCTGTTACATCA	839	550	55
YRb3_20	chr9:64502363-64503631	ATCGGCAGGTTTTCCTAAATAA	AGCTTTTTGACGTTGCTGGTAT	561	266	55
YRb3_21	chr10:19858586-19859854	CCCAAGTGTACGTGCTTGTGTA	TGGCTTTAGGAGCAAAGTATCC	538	243	55
YRb3_22	chr13:134177774-134179042	ATATGCACTCAGTGAGCACAGG	AAATAAGGGCATGGTCATTCAG	514	222	55
YRb1_1	chr8:115497214-115498482	TCAGAAAAATGTTGCTTCATGTC	ACTCGGAAAAATAAGTGGCAAT	452	157	55
YRb1_2	chr10:61398735-61400003	ACTGTCGCTTCAATACTCAGCA	AGGGAACTTTGTGACAGTTTCTG	546	249	55
YRb1_3	chr3:12746000-12747268	AAGTCATTTCAAAACGGCTGTAA	AGCATCCTCATCCAATAGCAGT	491	194	55
YRb1_4	chr15:14169897-14171336	TGAGTGATCAGGAAATGGAGTG	GGAGAAAGCTTGTTCTGTGTCC	774	483	55
YRb1_5	chr14:34352445-34353712	GACCACCCTTCCATGAAAATAA	GACCTAGGCATTGAGATTTTGC	534	244	55
YRb1_6	chr4:156546811-156548078	AACAGAGCTTAAGAACAGAGTCCAA	AAATTTGAAATGGGGACTTCCT	464	171	55
YRb1_7	chr5:153758989-153760257	AGGTTCTTTGAGCCATTCAGAT	AAAGTGAACAGCTTCCCCTATG	501	213	55
YRb1_8	chr2:83313780-83315049	ACCACTCTTGAAGGCACCTTAG	TGCTAGACTGGGTGTTTCTGAG	675	392	55
YRb1_9	chr15:43018359-43019628	TTTCTGTCTCCACCAAAGAAGG	ACACGTTTCTCTGCATACCTCA	557	263	55
YRb1_10	chr13:122572549-122573824	TTTTGTGTACCAGGATATTTGTGA	GAGTTGAATCAGGACAGGTTGC	538	241	55
YRb1_11	chr10:51817644-51818913	TGCAGGTTCTCTTTATTGGTGA	TTCCTGGGACTTAGCCTATCAA	509	216	55
YRb1_13	chr4:141222592-141223862	TCAAGCCTGTTACTGTCATTTCA	GCCTAATTTAGTAAACGGTTAAAGTGG	594	267	55
YRb1_14	chr4:61898875-61900143	TGCCAAAGTTGAATCTCTACCA	ATGAATGGGGAGAGATGGTTTA	514	217	55
YRb1_15	chr10:65437238-65438508	CCACACTGGTCGTTGAAGAAC	GGTGACCATGGACAACAAACTT	730	443	55
YRb1_17	chr10:35535309-35536576	TCTGCAGACCCAGAAACTAACA	TTTCACTTTTCCCAGTGGAGAG	522	234	55
YRb1_18	chr9:103112684-103113952	ATTGGTCACCTTTGCTGCTACT	CTGTCTCCCTGGTGTTTTACCT	898	603	55
YRb1_19	chr4:149026540-149027808	GACAAAGTCCCAATATGCCTTC	TCCGTTACCTAGGGTCGTTTTA	636	339	55
YRb1_20	chr8:40085648-40086918	GGGAGATGCACAAGAGTTATCC	TTCAGTTTTGGTGATCTTGCAT	920	629	55
YRb1_21	chr3:194953360-194954628	TTGCAAGGACTGGAACATCTTA	GATAACCTGGGCTCTGTTTGTT	459	168	55

YRb1_22	chr3:9249375-9250643	GTTACAGAAGACTTGCCCCAAA	CAGGGCATGTAGTAGGTGTTCA	690	393	55
YRb1_23	chr4:128891807-128893078	AGTGCCATCGAAGCTATCCTAA	TGGACTTGGAACTTTTCATTTG	489	189	55
YRb2_1	chr9:56772094-56773362	CTGCTCCCAGAAACAGGTCT	CTAGCTGACCCTGACTCCTCAT	762	465	55
YRb2_2	chr5:129303809-129305077	GCTTCAAACTTTCTGGCTGTG	ACATACACAACCTCTAGGCAGATG	526	235	55
YRb2_4	chr12:77372749-77374017	TTTAGGTGTTTCCCTTCTTTGC	TAGCTCTACCCCCAGCACTACT	706	416	55
YRb2_5	chr11:78407009-78408277	AGCTTCTGGTCTGAGGAGGTTT	TCATGCAAGAAATGGTTGAATC	769	297	55
YRb2_6	chr7:132449355-132450623	GGTGCCTCATATAATGCCTAGC	GTGAGTCCTGGAGTGCCTAAGT	432	150	55
YRb2_8	chr11:5823635-5824903	AAAATACGCACAACTGGGTTTC	TTTCCATATCTGCTTTCCCAGT	616	322	55
YRb2_9	chr8:38740369-38741636	CATCCTGAAATCCATCCACTTT	TCTTACATCAGCAAACCCTCAA	447	153	55
YRb2_10	chr8:28825609-28826877	AAATGTTGGAAAAGTGGAGGAA	AAGGGAACCCCAGAAGAATAAC	540	248	55
YRb2_11	chr7:142150639-142151908	ACCCTTCAGGCTCTATCTGACA	ATGGAGGTTGATCTGAAAATGC	444	159	55
YRb2_12	chr6:285619-286894	AGCTCCCAGTGGAATAAAACCT	GCTTGTGAAAGTGTCTTGTGGA	518	216	55
YRb2_13	chr5:161466164-161467432	AGTCTGGTGATTTTGCGTTTTC	TCCAGTGAATTCCTTTGGATGT	541	246	55
YRb2_15	chr12:96896039-96897307	CAGTCAGGCTCATTGTGTGAGT	GTCCTCGTAGATGGTTTGCTCT	489	199	55
YRb2_16	chr11:57418184-57419452	ACTTTGAACAGGACTCACATGC	ATAGTTCTATCTCCTTTGTCATTGTTT	508	203	55
YRb2_17	chr1:201391519-201392787	GACTATTGGCCTGCTTTTCAGT	ACAGAGTATGTTGCTGGGGAAT	520	223	55
YRb2_18	chr6:86766035-86767303	CCAGACTCCACCAAATCTCATT	AGCATCAAATTGGGAAACAAAG	520	225	55
YRb2_19	chr5:110125640-110126908	TTTCAAAAATGCTGACCTTCCT	GCCATGATGTCTTCAAACACTG	586	293	55
YRb2_20	chr3:117773467-117774738	CCCCAGATTTCACTTGGAATAA	TAACATTGTTCCCCAATCACTG	697	397	55
YRb2_21	chr2:172782270-172783538	TGCATTGATTTATGTCAAAAGG	AGACAGTGGTGACAACATCCAA	513	219	55
YRb2_22	chr15:108656090-108657358	GAGAGTGGTGGAAATTGCTCTT	AGTTCCTCAAACTGCATCACCT	804	522	55
YRb2_23	chr11:88868535-88869803	GCATTCACCCAAAGAGTTAATTG	CCTTCCTCTCTATACTGATTGAAATG	528	210	55
YRb2_24	chr10:60923640-60924908	AATGTTCTGTAAATAGCGCAGTAAT	TGAAGTGTTCTAGGTAAAGAACAAAA	506	218	55
YRb2_25	chr5:177180519-177181786	TTATTACATGCTCTGCCGGAAT	ACTCAGGGCATCAAGTAGAAGG	662	372	55
YRb2_26	chr11:111739204-111740471	GAATTTGGGTCTATGCCATGTT	GTGAGTGTACGGATCAATCGTG	744	457	55
MacYb5_1	chr2:98311406-98312674	GCAAAGGAATGGAATAGATTTTGT	TTTGTGTCTACTCCGAATACATCT	700	403	55
MacYb5_2	chr2:48781055-48782323	AGAATGCTGAATGAAGCAGAGG	TGACATGTTCCCACTTGTGATT	560	263	55
MacYb5_3	chr16:55670874-55672142	TTTGGTTTCATTCAAACTGACCT	AAAGTCTGAACGCAGTGCTCTA	551	254	55
MacYb5_4	chr1:112867300-112868568	TCTTGTGGCGAAATGAAATCTT	GAATGGCTGCTGACAGTGG	576	250	55

MacYb5_5	chr1:65365964-65367237	ACTATGGCAGCTGCAAAAAGAG	AAGAACCTTAAGGAAATGCTGTCT	576	282	55
MacYb5_6	chr9:68802338-68803606	CATTTGGGAATGAAAAGTAGGG	AAAACACATTCCTTCTGATCTTTGA	549	254	55
MacYb5_7	chr8:112974678-112975946	GTTGCCTTTTGTGTCTGAACCT	CCTTGTTTAGGACATGGAATGTT	530	233	55
MacYb5_8	chr5:138827215-138828483	CATGGTCACGTTACTGTCTAGGG	GCACAGTGCTATCCCTTCTGAT	537	240	55
MacYb5_9	chr3:133854596-133855866	AGTGATTCTGCAAAACAATCCA	TTCACAGTAGTGTTTCATGACCTTC	558	261	55
MacYb5_10	chr14:110204130-110205398	TTCAAATAAGACTCCAGCATCAA	GATTGCAGTTCTGTATTTTGAATGA	508	211	55
MacYb5_11	chr14:104673840-104675108	AGTGGCGTAGGTCCTGAGAATA	CGGTGTTACGAAAGTGAATAGG	585	288	55
MacYb5_12	chr14:44785725-44786993	TTCTGCTCAACTGCCAAATAGA	TCAAAAGCAACTGCCTATGTTT	576	274	55
MacYb5_13	chr11:122223550-122224818	GCATGTCCTGGTACATTGAAAA	GTCTCCTGGAGTCGTTTTTGTT	516	220	55
MacYb5_14	chr11:25218655-25219923	AGATCTGGTTCTAATGGGCTGA	CCTCAGGCTAGGCTCAGAGATA	576	288	55
MacYb5_15	chr10:84732600-84733868	ACTGCCACGTTTCAATTACCTG	CCTATAGCAAAGCCATGAAAGG	574	286	55
MacYb5_16	chr17:74299296-74300563	TGTGCAAAAGCACTTCATAGG	TCACTAAATGTCACTCCCTCCA	592	300	55
MacYb5_17	chrX:130339054-130340322	TCAGGTTTCCCACAATTCTACA	ATTTAAGGACCTCTGCCATCCT	550	254	55
MacYb5_18	chr2:117260899-117262167	AGCAATCAGATAGCAGTCATGC	AGTAACCAGTGTTCCTCGCATC	596	299	55
MacYb5_19	chr2:44773928-44775196	TTAGCACTGAGCATGCCTTTTA	TACTTTCGCATATGACCTGCTG	537	225	55
MacYb5_20	chr2:38251204-38252472	TGTCATTTAGAAGGGCTCTAGCA	AAGCAAAAAGTTCCAATCAAATG	529	232	55
MacYb5_21	chr14:64412094-64413362	TCTCCTAGCAGTGCCTCTCAG	GAGTCCCTCTGGATACTTGTGG	595	298	55
MacYb5_22	chr1:162620738-162622006	GCTGTGCAGACATTTGTAGCTC	AGTCAGGGATAGCATTTTCTCG	554	267	55
MacYb5_23	chr1:13312037-13313305	TGCCCAATTGCTTAATTTTGAT	TAGAGTTAGGCAGGGAAAGCTG	648	351	55
MacYb5_24	chrX:37542020-37543288	CGTGATTGGCAAACAAAATCTA	ATGTGATACAGTTGCTCCGTGA	573	278	55
MacYb5_25	chr9:51863356-51864626	CGAGGCCCAGTCCTATAGAGTT	GCAGTTGCTGAAAGTAAGCTGAG	699	390	55
MacYb5_26	chr7:23732205-23733473	AGTGAGTTCCTAAGGCAGTTGG	GATACCTCCTTTCACCGAATCA	523	229	55
MacYb5_27	chr6:88014178-88015446	TAGGATTTTTGCGCTATCTTGA	TGTGGCTCTCAGTCCAAATAAA	682	385	55
MacYb5_28	chr4:162113173-162114441	CTGTGGACTTGCCTAAATTGTG	GTCTGTGCCTCGCTAGTTTTG	572	286	55
MacYb5_29	chr2:36254658-36255926	TGGAGAACACAAGCTCATTCAT	ACTCAGGGACTCAATTCAAAGG	818	521	55

Supplementary References

S1.   W. J. Kent, *Genome Res* **12**, 656 (2002).
S2.   N. Saitou, M. Nei, *Mol Biol Evol* **4**, 406 (1987).
S3.   D. K. Kohli, A. K. Bachhawat, *Nucleic Acids Res* **31**, 3501 (2003).
S4.   T. A. Hall, *Nucleic Acids Symposium Series*, 95 (1999).
S5.   J. Lee *et al.*, *Gene* **390**, 18 (2007).
S6.   H. J. Bandelt, P. Forster, A. Rohl, *Mol Biol Evol* **16**, 37 (1999).
S7.   P. Forster, R. Harding, A. Torroni, H. J. Bandelt, *Am J Hum Genet* **59**, 935 (1996).
S8.   M. M. Miyamoto, J. L. Slightom, M. Goodman, *Science* **238**, 369 (1987).
S9.   J. Xing *et al.*, *J Mol Biol* **344**, 675 (2004).
S10.   S. T. Szak *et al.*, *Genome Biol* **3**, research0052 (2002).
S11.   J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res* **22**, 4673 (1994).
S12.   S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J Mol Biol* **215**, 403 (1990).
S13.   P. Jern, G. O. Sperber, J. Blomberg, *J Virol* **80**, 1367 (2006).
S14.   N. Polavarapu, N. J. Bowen, J. F. McDonald, *Genome Biol* **7**, R51 (2006).
S15.   C. T. Yohn *et al.*, *PLoS Biol* **3**, e110 (2005).
S16.   R. Mang, J. Goudsmit, A. C. van der Kuyl, *J Virol* **73**, 7021 (1999).