American Journal of Primatology 73:1031–1040 (2011)

RESEARCH ARTICLE

Genotyping Single Nucleotide Polymorphisms (SNPs) Across Species in Old World Monkeys

RIPAN S. MALHI[1,2*], JESSICA SATKOSKI TRASK[3], MILENA SHATTUCK[1], JESSE JOHNSON[4], DEBRAPRIYO CHAKRABORTY[5], SREE KANTHASWAMY[6,7], UMA RAMAKRISHNAN[5], AND DAVID GLENN SMITH[3,7]
[1]*Department of Anthropology, University of Illinois Urbana-Champaign, Illinois*
[2]*Department of Animal Biology and Institute for Genomic Biology, University of Illinois Urbana-Champaign, Illinois*
[3]*Department of Anthropology, University of California, Davis, California*
[4]*School of Integrative Biology, University of Illinois Urbana-Champaign, Illinois*
[5]*National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, India*
[6]*Department of Environmental Toxicology, University of California, Davis, California*
[7]*California National Primate Research Center, University of California, Davis, California*

The development of DNA markers is becoming increasingly useful in the field of primatology for studies on paternity, population history, and biomedical research. In this study, we determine the efficacy of using cross-species amplification to identify single nucleotide polymorphisms (SNPs) in closely related species. The DNA of 93 individuals representing seven Old World Monkey species was analyzed to identify SNPs using cross-species amplification and genotyping. The loci genotyped were 653 SNPs identified and validated in rhesus macaques. Of the 653 loci analyzed, 27% were estimated to be polymorphic in the samples studied. SNPs identified at the same locus among different species (coincident SNPs) were found in six of the seven species studied with longtail macaques exhibiting the highest number of coincident SNPs (84). The distribution of coincident SNPs among species is not biased based on proximity to genes in the samples studied. In addition, the frequency of coincident SNPs is not consistent with expectations based on their phylogenetic relationships. This study demonstrates that cross-species amplification and genotyping using the Illumina Golden Gate Array is a useful method to identify a large number of SNPs in closely related species, although issues with ascertainment bias may limit the type of studies where this method can be applied. Am. J. Primatol. 73:1031–1040, 2011.

Key words: DNA; coincident SNP; cross-species amplification

INTRODUCTION

The development of DNA markers has become increasingly important within the field of primatology. Reliable indicators of genetic variation allow for more efficient genetic management of captive primate colonies and within wild primate populations, DNA markers can be used to infer population history and paternity [Kayser et al., 1996; Morin & Smith, 1995; Morin et al., 1997]. Of particular interest is the study of primate species employed as subjects in biomedical research, such as macaques, baboons, and vervet monkeys. The development of thousands of DNA markers in these biomedical model species can be used to identify the heritable component of diseases using whole genome association studies [Hirschorn & Daly, 2005]. Therefore, the establishment of a reliable and efficient method for identifying DNA markers is critical for future primatological studies.

One such method is cross-species amplification, which uses known polymorphisms identified in one species to find potential polymorphic sites in closely related species. Given the close genetic relationship among many anthropoid species, cross-species amplification provides a method for the rapid development of DNA markers. This approach has been applied extensively to short tandem repeats (STRs). After the identification of thousands of STRs in the human genome, researchers began to use heterologous PCR primers to estimate population genetic parameters and characterize genetic structure in nonhuman primates. However, while STRs are particularly useful as markers of genetic variation due to their high heterozygosity, their analysis can be very time consuming and subjective due to the presence of stutter bands and allelic dropout.

Contract grant sponsor: National Institutes of Health; Contract grant number: RR05090.

*Correspondence to: Ripan S. Malhi, 607 South Mathews Avenue, Urbana, IL 61801. E-mail: malhi@illinois.edu

Received 15 August 2010; revised 2 May 2011; revision accepted 2 May 2011

DOI 10.1002/ajp.20969
Published online in Wiley Online Library (wileyonlinelibrary.com).

Additionally, in the absence of a standardized allelic ladder, alleles at a locus may run at different sizes in different laboratories due to differences in laboratory environments (e.g. equipment/instrumentation, temperature, polymer, etc.), making it difficult to compare data across laboratories.

As an alternative, basing studies of genetic structure on single nucleotide polymorphisms (SNPs) instead of STRs avoid the problems of stutter bands and cross-laboratory consistency in genotyping. SNPs are also much more abundant in genomes, exhibiting closer linkage to sites of interest, and their analysis is amenable to automation and high throughput analysis with a high degree of accuracy in genotype calls. With advances in DNA sequencing technology (e.g. massive parallel sequencing-by-synthesis), SNP marker development and genetic analysis is relatively cost effective and can be applied cross all primate species, both captive and wild [Kanthaswamy et al., 2009; Williams et al., 2010]. For the foregoing reasons, the Genetics and Genomic Working Group of the National Primate Research Centers (NPRCs), under the sponsorship of the National Center for Research Resources of the National Institutes of Health (NIH), has developed panels of SNPs to replace STR panels currently in use at the NPRCs for genotyping rhesus macaques [Kanthaswamy et al., 2009]. However, while cross-species amplification is known to work in STRs, it is not yet known how effective this approach will be with SNPs. It is of interest to know whether these SNPs established in rhesus macaques are also present in, and hence informative for, other primate species of interest (coincident SNPs).

One potential problem with using cross-species amplification to identify SNPs is a bias toward finding coincident SNPs that are influenced by nearby genes. A SNP is a polymorphism that occurs *within* a species. Here, we define a coincident SNP as a polymorphism that occurs at the same locus in multiple species. Coincident SNPs are either due to identity by descent (IBD) or identity by state (IBS). For coincident SNPs that result from IBD, balancing selection near genes may maintain polymorphism in different species. Likewise, coincident SNPs due to independent mutations would result in IBS and might be expected to occur near genes as Hodgkinson et al. [2009] suggest that cryptic variation in mutation rate resulted in coincident SNPs near CpG dinucleotides and CpG islands are often found at the 5′ end of genes [Cross & Bird, 1995]. Thus, there are two reasons to expect that identification of coincident SNPs will be biased toward loci that are near to genes.

Malhi et al. [2007] used massive parallel pyrosequencing to develop approximately 23,000 candidate SNPs in *Macaca mulatta* (rhesus macaque), 85% of which are estimated to be polymorphic [Satkoski et al., 2008]. In this study, we used the SNP genotyping assays developed for *Macaca mulatta* on individuals from seven other old world monkey species: *Macaca fuscata* (Japanese macaque), *Macaca fascicularis* (longtail macaque), *Macaca radiata* (bonnet macaque), *Macaca nemestrina* (pigtail macaque), *Macaca sylvaunus* (Barbary macaque), *Papio anubis* (olive baboon), and *Chlorocebus sabaeus* (vervet monkey) to identify SNPs in these species. We investigate if SNP loci identified in *M. mulatta* are also polymorphic in other species of Old World Monkeys. The results of this study suggest that cross-species amplification is a useful technique for developing SNP DNA markers among closely related species and the frequency of coincident SNPs is not biased based on the proximity to genes.

METHODS

Samples and Assays

DNA from 93 Old World Monkeys was genotyped for 768 candidate SNPs (constructed in two 384 SNP panels) for rhesus macaques. Of the 768 candidate SNPs screened in this study, 653 were confirmed to be polymorphic in rhesus macaques in parallel studies [Kanthaswamy et al., 2010; Trask et al., submitted]. Studies used to confirm candidate SNPs in rhesus macaques relied on genotyping individuals on Illumina Golden Gate assays as described in Satkoski et al. [2008]. The Old World Monkeys screened in this study consist of Japanese macaques ($N = 10$) from the Oregon NPRC (ONPRC), longtail macaques ($N = 29$) from the California NPRC (CNPRC)/COVANCE Research Products/New Iberia Research Center, bonnet macaques ($N = 6$) from zoos in Central and Southern India, pigtail macaques ($N = 14$) from the CNPRC, Barbary macaques ($N = 8$) from La Foret des Singes, olive baboons ($N = 17$) from the University of Oklahoma Health Sciences Center (OHSC), and vervet monkeys ($N = 9$) from the University of Miami Division of Veterinary Resources. This research complies with the protocols of Illinois Institutional Animal Care and Use Committee and adheres to the requirements of the American Society of Primatologits and of the United States.

DNA Extraction, Genotyping, and Validation

DNA was extracted from whole blood using the Qiagen QIAamp DNA Mini Kit (Qiagen Inc., Valencia, CA). Following DNA extraction, the bonnet macaque DNA (whose concentration was far too low for SNP analysis) was whole genome amplified using the Qiagen Repli-G Kit (Qiagen Inc., Valencia, CA). There is no suggestion that the whole genome amplification of the bonnet macaque DNA resulted in a bias of loci genotyped [Giardina et al., 2009; Hansen et al., 2007]. The DNA samples were screened with the Illumina Golden Gate Assay using the BeadXpress Reader at the U.C. Davis Genome

Center. The Golden Gate Assay provides locus specificity by annealing oligonucleotides, one allele-specific and one locus-specific, upstream and downstream to the SNP, respectively. Each oligonucleotide is replicated about 30 times on each array. The high degree of replication allows for robust measurement at each SNP. In addition, three samples (one representative each from the sample of bonnet, Japanese and pigtail macaques) were used to generate multiple independent genotypes as a control.

The SNPs were analyzed on the BeadStudio software using a stringent analysis protocol designed to minimize the inclusion of type I errors. A cluster file was generated for the data from *Macaca mulatta* samples alone. Then data from the seven species were combined with the data from *M. mulatta* for a combined analysis. However, to maintain rigor in the analysis, the cluster file generated for *M. mulatta* alone was used in the combined analysis. Individual genotypes within a species plus the overall pattern of a plot for a locus were examined visually. Plots displaying unusual cluster patterns (different from the expected pattern) were rejected and excluded from further analysis (see Fig. 1 for examples of expected patterns at a locus). For each locus, if all individuals of a species were homozygotes, that locus was excluded for that species to preclude errors caused by allelic dropout. In addition, if an individual genotype did not cluster within the 95% confidence intervals calculated in the BeadStudio software analysis, it was excluded from the analysis for that particular locus (see Fig. 1 for an example).

Even with the replication and stringent protocols used in the analysis, genotyping errors can occur due to artifacts and gene duplication. To increase our confidence in the data we performed two additional analyses. First, we trimmed the data to conform to Hardy Weinberg equilibrium expectations. A locus where the observed genotype differed from the expected by more than 0.025 was excluded from analysis for that species. This threshold was chosen so that the outlier SNPs, but not those moderately influenced by selection, would be excluded. Second, we performed an additional validation by dideoxy DNA sequencing (Sanger) a SNP at position 8,405,370 on Chromosome 5 in Japanese macaques. Flanking regions of the SNP locus were identified in rhesus macaque by searching for the ''MamuSNP 454 sequence identifier'' at the website www.mamusnp.ucdavis.edu [Malhi et al., 2007]. The sequence was compared with the rhesus macaque draft sequence at the UCSC genome browser using BLAT and a target sequence of 400 base pairs was identified for primer design.

Analysis

Because the number of coincident SNPs identified is influenced by the sample size of each species investigated, the number of coincident SNPs identified was adjusted for variation in sample size by

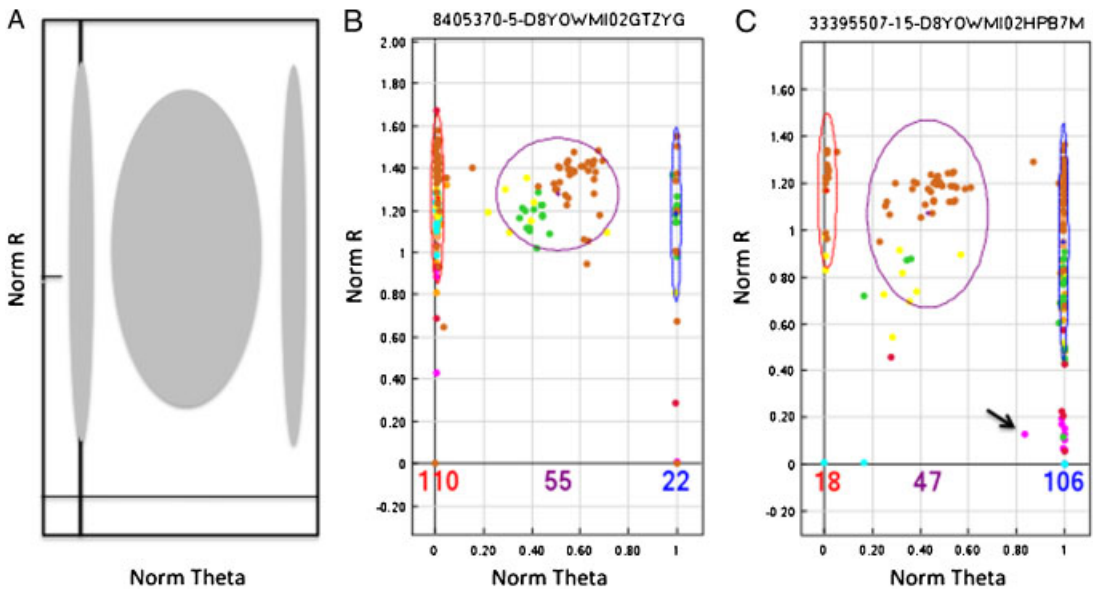


Fig. 1. (**A**) Illustration of expected distribution plot of homozygotes and heterozygotes in SNP analysis on the BeadStudio software. Homozygotes are expected in the gray ovals on the left and right side. Heterozygotes are expected in the gray oval in the middle of the homozygotes. (**B**) Example of BeadStudio SNP plot where individuals of Japanese and longtail macaques were identified as being heterozygous for a coincident SNP. Circles represent center of genotype clusters used to calculate 95% confidence intervals. Marker color identifies species (brown = rhesus macaque, green = longtail macaque, yellow = Japanese macaque, red = bonnet macaque, orange = pigtailed macaque, blue = baboon, pink = vervet). (**C**) Example of Bead Studio SNP plot where Japanese, longtail and bonnet macaques were identified as being heterozygous, however a single vervet monkey (identified by an arrow) fell outside of the 95% confidence interval and was therefore excluded from the analysis as a heterozygote. [Color figures can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

adapting the Ewens–Watterson estimator for segregating sites [Watterson, 1975]. The number of coincident SNPs was adjusted with the formula

$$\theta = \frac{K}{\sum_{i=1}^{n-1} \frac{1}{i}},$$

where K is the number of coincident SNPs and n is the number of samples of a given species. In addition, minor allele frequency, expected and observed heterozygosity, and all plots were calculated and created on MICROSOFT EXCEL.

To estimate whether or not a SNP was located in or near a gene, we used the ''MamuSNP 454 sequence identifier'' to obtain the pyrosequence read at the MamuSNP website. We then used the BLAT program to identify the location of the pyrosequence read in the rhesus macaque genome with the RefSeq and ENSEMBL gene prediction activated. If the pyrosequence read matched the rhesus draft sequence or any mammalian genome within 2 kb distance from a gene via RefSeq, it was considered to be influenced by a gene. However, if the read matched genomic regions beyond 2 kb limit it was considered outside the influence of a gene. The same method was used on all 653 SNPs assayed in this study. A chi-square analysis comparing number of SNPs influenced by genes from the 653 SNPs in rhesus macaque and the list of coincident SNPs was performed in SPSS.

Additional DNA Sequencing

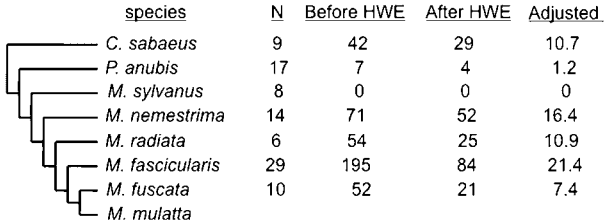
Because some results of the SNP analysis were not consistent with our expectations, additional loci were Sanger sequenced in order to verify the results found in the Golden Gate Assay. Six unlinked genomic regions thought to be neutral (distant from genes) were sequenced in 10 rhesus macaques, 10 longtail macaques, and 10 Japanese macaques. The six regions sequenced for all three species are found on chromosomes 4 (478 bp), 6 (511 bp), 9 (526 bp), 14 (477 bp), 20 (585 bp), and X (325 bp) and some of the data for these loci in rhesus and longtail macaques were previously reported in Satkoski et al. [in press]. The genomic regions were sequenced at least twice using forward and reverse sequencing primers and were aligned using SEQUENCHER. Heterozygotes were detected using SEQUENCHER and by visual inspection of double peaks in electropherograms. DNA sequences are available on GENBANK. Primers are available from the authors upon request.

RESULTS

Of the 653 rhesus macaque SNPs assayed in individual samples of Old World Monkeys, 421 met the criteria and expected pattern described above (See Fig. 1 for examples) and, of these, 215 conformed to the Hardy–Weinberg Equilibrium (HWE) test described above (Fig. 2, Table I). Of the 215 SNPs, 137 occur in one of the seven species, 30 occur in two of the seven species, and 6 occur in three of the seven species (Table I) accounting for 173 (27%) of the 653 assayed. A single representative from the Japanese, bonnet and pigtail macaque samples that were genotyped multiple times for quality control provided identical genotypes with each iteration of genotyping, confirming the accuracy of genotyping.

Following the stringent genotype calling protocol and HWE test, 21 (7.4 adjusted for sample size) rhesus macaque SNPs were coincident in Japanese macaques, 84 (21.4 adjusted for sample size) in longtail macaques, 25 (10.9 adjusted for sample size) in bonnet macaques, 52 (16.4 adjusted for sample size) in pigtail macaques, 0 in Barbary macaques, 4 (1.2 adjusted for sample size) in olive baboons and 29 (10.7 adjusted for sample size) in vervet monkeys (Fig. 2). A single SNP, at position 8,405,370 on chromosome 5, was confirmed by Sanger sequencing in Japanese macaques. All genotype calls for this SNP generated by the Illumina Golden Gate Assay matched genotype calls inferred from the Sanger sequencing for these samples, confirming the accuracy of the instruments used in this study.

Adjustments for sample size reflected the over-representation of coincident SNPs in the longtail macaques (which adjusted less drastically downward) and the under-representation of coincident SNPs in bonnet macaques (which adjusted more drastically downward), suggesting that our method of adjusting numbers of coincident SNPs for variation in sample size was efficacious. Unexpectedly, the number of coincident SNPs, both with and without adjustment for sample size, does not correlate with the phylogenetic distance expected between rhesus macaques and each of the OWM species studied [Li et al., 2009, Fig. 2]. For example, the species most (Japanese macaques) and least (vervet monkeys) closely related to rhesus macaques share, respectively, 21 (7.4 adjusted) and 29 (10.7 adjusted) coincident SNPs with rhesus macaques after trimming. This reflects both an unexpectedly low coincidence for Japanese macaques and an unexpectedly high coincidence for vervet monkeys. To confirm this



species	N	Before HWE	After HWE	Adjusted
C. sabaeus	9	42	29	10.7
P. anubis	17	7	4	1.2
M. sylvanus	8	0	0	0
M. nemestrima	14	71	52	16.4
M. radiata	6	54	25	10.9
M. fascicularis	29	195	84	21.4
M. fuscata	10	52	21	7.4
M. mulatta				

Fig. 2. Phylogeny and table of coincident SNPs. N = sample size. ''Before HWE'' are the number of SNPs before Hardy–Weinberg Equilibrium trimming and ''After HWE'' are the number after trimming. ''Adjusted'' are the number of SNPs ''After HWE'' adjusted for sample size.

TABLE I. List of Coincident SNPs Identified After Trimming

Rhesus macaque chromosome position	Rhesus macaque chromosome	MamuSNP 454 sequence identifier	*Macaca fuscata*	*Macaca fascicularis*	*Macaca radiata*	*Macaca nemestrina*	*Papio anubis*	*Chlorocebus sabaeus*
1225328	1	D8YOWMI02J2VK2				X		
35009123	1	D8YOWMI02GFCBO				X		
81583842	1	D8YOWMI01CX04H	X	X	X			
90794516	1	D8YOWMI01A5XHU		X				
109370014	1	D8YOWMI02G5UK4		X		X		
114884317	1	D8YOWMI01B4Z3C			X			
122140063	1	D8YOWMI02F4IAP		X				
136630342	1	D8YOWMI01BYCMR		X				
177882171	1	D8YOWMI02ICOQZ				X		
189341999	1	D8YOWMI01ALA6J				X		
208621724	1	D8YOWMI02GE084		X				
13837970	2	D8YOWMI02GZD64		X		X		
22248380	2	D8YOWMI01BMXAS		X				
57035184	2	D8YOWMI02HTCYO		X	X			
70607926	2	D8YOWMI01B2H0F	X					
141969056	2	D8YOWMI02FNN17					X	
158529536	2	D8YOWMI01EO0KK		X				
160903336	2	D8YOWMI02HC8IL		X				
162289707	2	D8YOWMI01DW7XC		X				X
180864622	2	D8YOWMI01EC4QW		X				
185888444	2	D8YOWMI01E4LGQ		X				
7026913	3	D8YOWMI02GSRQO		X				
14964730	3	D8YOWMI01BHAP3				X		
118163111	3	D8YOWMI01ALEQO		X				X
128301476	3	D8YOWMI01A94T0	X					
155951478	3	D8YOWMI01CPWNX						X
163059574	3	D8YOWMI02HV31O				X		
167559647	3	D8YOWMI02H5ZKN					X	
171009796	3	D8YOWMI01BCHPT				X		
177722945	3	D8YOWMI02F5WJE		X			X	
8097710	4	D8YOWMI02HP0CJ		X				
14901823	4	D8YOWMI01CYS3U	X	X				
21847630	4	D8YOWMI02GPYVY		X				
53143722	4	D8YOWMI01CVDSB		X		X		
91812668	4	D8YOWMI02I4GKW		X				
119078596	4	D8YOWMI02IS1D8		X				
126818281	4	D8YOWMI01B94SG		X		X		X
152912672	4	D8YOWMI02I24OR						X
163463473	4	D8YOWMI01BB7A7			X			
8405370	5	D8YOWMI02GTZYG	X					
9441528	5	D8YOWMI02FQ3NQ		X				
10991261	5	D8YOWMI01EAFHD		X				
11647796	5	D8YOWMI01DSNBZ				X		
15773318	5	D8YOWMI01BRKYT		X				
18473146	5	D8YOWMI01A437X		X				
23758421	5	D8YOWMI02H1LO8			X			
56662681	5	D8YOWMI01BW2WM		X				
67425781	5	D8YOWMI02J26BD		X				
87729814	5	D8YOWMI02GM8NE		X				X
113434228	5	D8YOWMI02GEC04	X					
122327212	5	D8YOWMI01CPM17				X		
126137157	5	D8YOWMI01CT198		X				
160183031	5	D8YOWMI02IPR19						X
168028900	5	D8YOWMI02JB4HP		X				
181052616	5	D8YOWMI01E56LM				X		
11634432	6	D8YOWMI01CE37I	X	X				
15035209	6	D8YOWMI02IFI5V						X
44209456	6	D8YOWMI02FZ4ZX	X					

TABLE I. Continued

Rhesus macaque chromosome position	Rhesus macaque chromosome	MamuSNP 454 sequence identifier	*Macaca fuscata*	*Macaca fascicularis*	*Macaca radiata*	*Macaca nemestrina*	*Papio anubis*	*Chlorocebus sabaeus*
50319998	6	D8YOWMI01EQYDN	X					X
53623247	6	D8YOWMI02HWJ2W		X				
57767887	6	D8YOWMI02JDW5K		X				
80482477	6	D8YOWMI02F89A5	X					
115816965	6	D8YOWMI02JEDP2		X				
116172769	6	D8YOWMI01COQ28		X				
153671398	6	D8YOWMI02HUEVC		X				X
162595761	6	D8YOWMI01CTS8T				X		
170687211	6	D8YOWMI02JBM8T					X	
31810831	7	D8YOWMI01BXVVK				X		
39655970	7	D8YOWMI02IOTYU		X				
52102599	7	D8YOWMI01CWETE			X			
66188937	7	D8YOWMI02H6Q3M		X				
75404223	7	D8YOWMI01CNX4K		X				
104917188	7	D8YOWMI02ICQT4				X		
112000631	7	D8YOWMI01ACX21		X				
118907410	7	D8YOWMI01C6FJ8			X			
135750880	7	D8YOWMI02G639E						X
146603681	7	D8YOWMI01BZUBP				X		
150166837	7	D8YOWMI01EVYLT			X			
3188570	8	D8YOWMI02H39OG		X				
6614606	8	D8YOWMI02IRK8A						X
9338149	8	D8YOWMI01B67HW				X		
12745272	8	D8YOWMI02F1LPM				X		
17104034	8	D8YOWMI02FOUUY	X					
63900217	8	D8YOWMI01B5196		X				
71193925	8	D8YOWMI02H95BY		X				
89206559	8	D8YOWMI01AIW7Z			X	X		
109676695	8	D8YOWMI02I9BOP		X				
127156610	8	D8YOWMI01EE5YG			X			
9197587	9	D8YOWMI01D9A4J				X		
18315239	9	D8YOWMI02F8M1S	X	X				
32042524	9	D8YOWMI02FT8TY		X				
43994974	9	D8YOWMI01ANQRH				X		
45957464	9	D8YOWMI02F3ZER						X
60129585	9	D8YOWMI01DTEW2		X				
75688312	9	D8YOWMI01E20L8	X					
82509260	9	D8YOWMI02GPQPE		X				
105355955	9	D8YOWMI01AFT2K						X
122398406	9	D8YOWMI01EH54Y			X			X
9692408	10	D8YOWMI01B1YGQ	X	X				
17334887	10	D8YOWMI02I0H6L		X				
33170961	10	D8YOWMI01CKIXV			X			
38086313	10	D8YOWMI02IAWPW		X		X		X
62948254	10	D8YOWMI02IZ1YY				X		
64311100	10	D8YOWMI02GOPYA						X
90335621	10	D8YOWMI01DYCTR		X				
8955710	11	D8YOWMI02HPFQF		X	X			
10920575	11	D8YOWMI01CG4GO				X		
12480611	11	D8YOWMI02JZCOG				X		
15759394	11	D8YOWMI01BBD6B			X			
30082108	11	D8YOWMI02HNMP0				X		X
33074229	11	D8YOWMI02GE7UT		X				
37349535	11	D8YOWMI02J2TFD				X		
52445830	11	D8YOWMI02F5ZX9		X				
59595345	11	D8YOWMI01B4I84		X				
85285618	11	D8YOWMI01B9ZS5		X				
98483405	11	D8YOWMI01DW4J1				X		X

TABLE I. Continued

Rhesus macaque chromosome position	Rhesus macaque chromosome	MamuSNP 454 sequence identifier	*Macaca fuscata*	*Macaca fascicularis*	*Macaca radiata*	*Macaca nemestrina*	*Papio anubis*	*Chlorocebus sabaeus*
100272930	11	D8YOWMI02GOFG9				X		
2820172	12	D8YOWMI01EQOOR			X			
5750047	12	D8YOWMI02JOKZD		X				
30014284	12	D8YOWMI02G7314			X	X		
68241234	12	D8YOWMI01EOBO2						X
96524476	12	D8YOWMI02I34M2				X		
101387294	12	D8YOWMI01A0O5H		X				
104866834	12	D8YOWMI02HAIYA		X				X
8631768	13	D8YOWMI02JFTTP				X		
52692215	13	D8YOWMI01D2AHR		X				
102359734	13	D8YOWMI01A7TMN	X		X			
109279279	13	D8YOWMI01DH96M				X		X
133145357	13	D8YOWMI01A4D38				X		
7607929	14	D8YOWMI02H5WF2			X			
34024212	14	D8YOWMI02G1J53				X		
37521504	14	D8YOWMI01B4JSC	X					
41927931	14	D8YOWMI02F212Z			X			
64204689	14	D8YOWMI02FNMDV						X
71176560	14	D8YOWMI01EQJPW		X				
77510326	14	D8YOWMI01DI1XC				X		
78159529	14	D8YOWMI02FOMJA				X		
85093242	14	D8YOWMI01C1XW1		X		X		
88510771	14	D8YOWMI01AQZO5		X				
90207126	14	D8YOWMI02GBWXW				X		
92048577	14	D8YOWMI02JH24P		X				
120087585	14	D8YOWMI01C20RM				X		
9925136	15	D8YOWMI01CN5Z1			X			X
33395507	15	D8YOWMI02HPB7M	X	X	X			
37206156	15	D8YOWMI01EH01E				X		
44889065	15	D8YOWMI01BZFWE						X
68488712	15	D8YOWMI01EIPZA	X					
73578355	15	D8YOWMI01DSO35		X				
80451870	15	D8YOWMI01DZQJ3		X				
99850716	15	D8YOWMI01BIXMX	X			X		
38951307	16	D8YOWMI01EL12W						X
40639725	16	D8YOWMI02ICB3N			X			
53844615	16	D8YOWMI02GL0VV				X		
5423481	17	D8YOWMI01C43GP		X				
11143522	17	D8YOWMI01AFNLY		X				
17527068	17	D8YOWMI02IP98Z			X			X
37664986	17	D8YOWMI01EHXXT	X					
44520276	17	D8YOWMI02G3LN1		X		X		X
50245329	17	D8YOWMI02IGWMS				X		
66897426	17	D8YOWMI02GX6UA	X		X			
75029428	17	D8YOWMI02H9JNN		X		X		
90265285	17	D8YOWMI02I20II		X	X	X		
18461747	18	D8YOWMI02IB8XJ		X				
44247960	18	D8YOWMI01CMPQJ		X				
47386966	18	D8YOWMI01B3HSS		X				
65033140	18	D8YOWMI01CHYUE		X				
70525231	18	D8YOWMI02GU71J		X		X		
4396820	19	D8YOWMI01CVQRW				X		
7807333	19	D8YOWMI02JIK8C		X				
16063151	19	D8YOWMI02HD6R2		X				
38104151	19	D8YOWMI01A5JSC		X				
9267056	20	D8YOWMI02H8FKM						X
23973946	20	D8YOWMI01AOYH0				X		

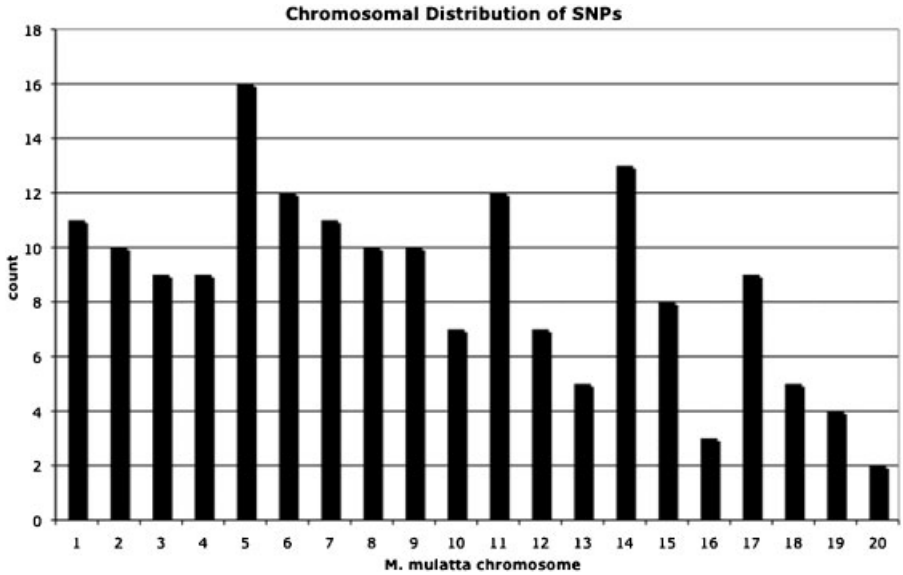


Fig. 3. Distribution of coincident SNPs on rhesus macaque genome.

unexpected pattern, we sequenced six unlinked noncoding genomic regions (a total of approximately 3 kb) in rhesus macaques, Japanese macaques, and longtail macaques. We identified a total of 52 SNPs in rhesus macaques, 31% ($N = 16$) of which were coincident with longtail macaques and only 2% ($N = 1$) of which were coincident with Japanese macaques. This result agrees with the rank order of species with coincident SNPs observed in the genotyping assay.

All coincident SNPs are distributed across all chromosomes when the rhesus macaque genome is used to locate them (Fig. 3). In addition, a majority of the coincident SNPs exhibit an expected heterozygosity (EH) and a minor allele frequency (MAF) of 0.2 and below (Table II) and 0.1 and below (Table III), respectively. Although these values are much lower than those for rhesus macaques, EH and MAF should be considered preliminary estimates until additional study with larger and geographically diverse samples is conducted. Overall, 37% of coincident SNPs were identified in genomic regions influenced by genes compared with 33% of SNPs in the original *M. mulatta* panel ($P = 0.295$).

TABLE II. Fraction of Coincident SNPs With Different Heterozygosity (EH) Estimates

Species	<0.1	<0.2	<0.3	<0.4	<0.5
Macaca fuscata	0.190	0.714	0.000	0.000	0.095
Macaca fascicularis	0.667	0.179	0.024	0.083	0.048
Macaca radiata	0.000	0.962	0.000	0.000	0.038
Macaca nemestrina	0.692	0.231	0.000	0.038	0.038
Papio anubis	0.500	0.250	0.000	0.250	0.000
Chlorocebus sabaeus	0.000	0.964	0.000	0.000	0.036

TABLE III. Fraction of Coincident SNPs With Different Minor Allele Frequency (MAF) Estimates

Species	<0.1	<0.2	<0.3	<0.4	<0.5
Macaca fuscata	0.667	0.238	0.048	0.000	0.048
Macaca fascicularis	0.821	0.071	0.060	0.024	0.024
Macaca radiata	0.962	0.000	0.000	0.000	0.038
Macaca nemestrina	0.846	0.077	0.038	0.000	0.038
Papio anubis	0.750	0.000	0.250	0.000	0.000
Chlorocebus sabaeus	0.857	0.107	0.000	0.036	0.036

DISCUSSION

This study confirms that cross-species amplification and genotyping using the Golden Gate Assay is a useful method for genotyping SNPs in closely related species. The frequency of coincident SNPs near or distant from genes observed in this study are statistically indistinguishable and coincident SNPs likely represent a combination of IBD from shared ancestry and IBS from recurrent mutations. It is difficult to estimate the ratio of coincident SNPs due to IBD vs. IBS in each species, but the phylogenetic relationship of each species of OWM to rhesus macaque can provide some guidance. Species closely related to rhesus macaque (i.e. Japanese macaque, longtail macaque) likely reflect a higher ratio of coincident SNPs due to IBD compared with IBS and species distantly related to rhesus macaque (i.e. baboon, vervet monkeys) likely reflect the opposite pattern with SNPs due to IBS being more common than IBD.

The highest number of coincident SNPs was identified in longtail macaques (84 coincident SNPs), the second most closely related species to rhesus macaques. However, the identification of 29 coincident

SNPs in vervet monkeys, the species studied that is most distantly related to rhesus macaques, was unexpected given the lower numbers of coincident SNPs in most other macaque species studied. Likewise, *M. fuscata*, the species most closely related to rhesus macaques, displayed an unexpectedly low number of coincident SNPs, a result that was confirmed with the additional Sanger sequencing. This suggests that population history and the genetic diversity within the samples screened (as well as sampling effects) contributes to the pattern of coincident SNPs observed in this study. For example, the well-known paucity of genetic heterogeneity in Japanese macaques [Marmi et al., 2004] might reflect a genetic bottleneck during which many less common SNPs coincident with rhesus macaques were lost. The reasons for the unexpectedly high coincidence of SNPs in vervet monkeys are more difficult to explain. Although it is likely that coincident SNPs identified in most macaque species have very similar and homologous binding sites to *M. mulatta*, this pattern may not hold true for distantly related species like baboons and vervet monkeys and it is possible that some of the coincident SNPs are the result of artifact due to dissimilarity from *M. mulatta* in binding sites. This dissimilarity could potentially cause the binding of rhesus macaque-specific oligonucleotides to the binding sites of vervet monkeys to occur incompletely, causing an artifact in the reflectance of the oligonucleotide and subsequently being miscalled as a heterozygote by the genotype calling software. Therefore, this method may only be applicable to species that are very closely related to one another (e.g., within the same genus). Additional research with larger and more geographically diverse samples within each species than those surveyed in this study would provide more accurate estimates of genetic diversity (i.e., heterozygosity and MAF) and confirm the polymorphism in populations of species relatively distantly related to *M. mulatta*. Also, studies designed to estimate which of these SNPs are due to IBD vs. IBS across species will help create panels optimized for different types of genetic analysis such as phylogenetics, population history, genetic management, and kinship.

The results of this study demonstrate that cross-species amplification and genotyping can be used to identify coincident SNPs in closely related species and extend the benefits of next-generation sequencing technologies to identify polymorphisms. However, the coincident SNPs identified in this study exhibit an ascertainment bias and may be limited in use. Where the presence of ascertainment bias might significantly influence results, as for example in studies of phylogenetic relationships across a wide range of species, this method is not recommended and more expensive methods such as next-generation sequencing with capture methods should be used. Studies that focus on a single, closely related species are unlikely to be affected by this bias, however, and would benefit from the more cost effective method of cross-species amplification described here.

ACKNOWLEDGMENTS

We thank Betsy Ferguson for providing the Japanese macaques used in this study. This study was partially supported by National Institutes of Health grant (No. RR05090 awarded to D.G.S.).

REFERENCES

Cross SH, Bird AP. 1995. CpG islands and genes. Current Opinion Genetics and Development 5:309–314.

Giardina E, Pietrangeli I, Martone C, Zampatti S, Marsala P, Gabriele L, Ricci O, Solla G, Asili P, Arcudi G, Spinella A, Novelli G. 2009. Whole genome amplification and real-time PCR in forensic casework. BMC Genomics 10:159.

Hansen HM, Wiemels JL, Wrensch M, Wiencke JK. 2007. DNA quantification of whole genome amplified samples for genotyping on a multiplexed bead array platform. Cancer Epidemiology Biomarkers and Prevention 16:1686–1690.

Hirschorn JN, Daly MJ. 2005. Genome-wide association studies for common diseases and complex traits. Nature Reviews Genetics 6:95–108.

Hodgkinson A, Ladoukakis E, Eyre-Walker A. 2009. Cryptic variation in the human mutation rate. PLoS Biology 7: e1000027.

Kanthaswamy S, Capitanio JP, Dubay CJ, Ferguson B, Folks T, Ha JC, Hotchkiss CE, Johnson ZP, Katze MG, Kean LS, Kubisch HM, Lank S, Lyons LA, Miller CM, Nylander J, O'Connor DH, Palermo RE, Smith DG, Vallender EJ, Wiseman RW, Rogers J. 2009. Resources for genetic management and genomics research on non-human primates at the National Primate research centers (NPRCs). Journal of Medical Primatology 38:17–23.

Kanthaswamy S, Satkoski J, Kou A, Malladi V, Smith DG. 2010. Detecting signtures of inter-regional and inter-specific hybridization among the Chinese rhesus macaque specific pathogen-free (SPF) population using single nucleotide polymorphic (SNP) markers. Journal of Medical Primatology 39:252–265.

Kayser M, Ritter H, Bercovitch F, Mrug M, Roewer L, Nurnberg P. 1996. Identification of highly polymorphic microsatellites in the rhesus macque *Macaca mulatta* by cross-species amplification. Molecular Ecology 5:157–159.

Li J, Han K, Xing J, Kim HS, Rogers J, Ryder OA, Disotell T, Yue B, Batzer MA. 2009. Phylogeny of the macaques (Cercopithecidae: Macaca) based on Alu elements. Gene 448:242–249.

Malhi RS, Sickler B, Lin D, Satkoski J, Tito RY, George D, Kanthaswamy S, Smith DG. 2007. MamuSNP: a resource for Rhesus Macaque (*Macaca mulatta*) genomics. PLoS One 2:e438.

Marmi J, Bertranpetit J, Terradas J, Takenaka O, Domingo-Roura X. 2004. Radiation and phylogeography in the Japanese macaque, *Macaca fuscata*. Molecular Phylogenetics and Evolution 30:676–685.

Morin PA, Smith DG. 1995. Nonradioactive detection of hypervariable simple sequence repeats in short polyacrylamide gels. Biotechniques 19:223–228.

Morin PA, Kanthaswamy S, Smith DG. 1997. Simple sequence repeat (SSR) polymorphisms for colony management and population genetics in rhesus macaques (*Macaca mulatta*). American Journal of Primatology 42:199–213.

Satkoski JA, Malhi R, Kanthaswamy S, Tito R, Malladi V, Smith D. 2008. Pyrosequencing as a method for SNP

identification in rhesus macaque (*Macaca mulatta*). BMC Genomics 9:256.
Satkoski TJ, Garnica WT, Mahli RS, Kanthaswamy S, Smith DG. In press. High-throughput SNP discovery and the search for candidate genes for long-term SIVmac nonprogression in Chinese rhesus macaques (*Macaca mulatta*). Journal of Medical Primatology.
Trask JA, Malhi RS, Kanthaswamy S, Johnson J, Garnica WT, Malladi VS, Smith DG. 2011. The effect of SNP discovery method and sample size on estimation of population genetic data from Chinese and Indian rhesus macaques (*Macaca mulatta*). Primates 52:129–138.
Watterson GA. 1975. On the number of segregating sites. Theoretical Population Biology 7:256–276.
Williams LM, Ma X, Boyko AR, Bustamante CD, Oleksiak MF. 2010. SNP identification, verification, and utility for population genetics in a non-model genus. BMC Genetics 11:32.