Published in final edited form as:

Am J Primatol. 2013 February ; 75(2): 135–144. doi:10.1002/ajp.22091.

Single Nucleotide Polymorphisms Reveal Patterns of Allele Sharing across the Species Boundary between Rhesus (*Macaca mulatta*) and Cynomolgus (*M. fascicularis*) Macaques

Jessica A. Satkoski Trask^{1,2}, Wendy T. Garnica^{1,2}, David Glenn Smith^{1,2}, Paul Houghton³, Nicholas Lerche², and Sree Kanthaswamy^{2,4}

¹Department of Anthropology, 330 Young Hall, One Shields Avenue, University of California, Davis, CA, 95616, USA

²California National Primate Research Center, One Shields Avenue, University of California, Davis, CA, 95616, USA

³Primate Products, Inc., PO Box 1588, Immokalee, FL, 34143

⁴Department of Environmental Toxicology, One Shields Avenue, University of California, Davis, CA, 95616, USA

Abstract

Both phenotypic and genetic evidence for asymmetric hybridization between rhesus (*Macaca mulatta*) and cynomolgus (*Macaca fascicularis*) macaques has been observed in the region of Indochina where both species are sympatric. The large-scale sharing of MHC class II alleles between the two species in this region supports the hypothesis that genes, and especially genes involved in immune response, are being transferred across the species boundary. This differential introgression has important implications for the incorporation of cynomolgus macaques of unknown geographic origin in biomedical research protocols. Our study found that for 2808 SNP markers, the MAF and observed heterozygosity calculated from a sample of Vietnamese cynomolgus macaques was significantly different from those calculated from samples of both Chinese rhesus and Indonesian cynomolgus macaques. SNP alleles from Chinese rhesus macaques were overrepresented in a sample of Vietnamese cynomolgus macaques relative to their Indonesian conspecifics and located in genes functionally related to the primary immune system. These results suggest that Indochinese cynomolgus macaques represent a genetically and immunologically distinct entity from Indonesian cynomolgus macaques.

Keywords

SNPs; macaques; hybridization; introgression; ancestral character states

Introduction

Four species comprise the fascicularis group of macaques: the cynomolgus macaque, *M. fascicularis*; the rhesus macaque, *M. mulatta*; the Formosan macaque, *M. cyclopis* and the Japanese macaque, *M. fuscata*. The most recent common ancestor of this group of macaque species is thought to have originated on the islands of Indonesia approximately 900 kya before spreading throughout East, South and Southeast Asia [Abegg and Thierry 2002;

Delson 1980]. Both rhesus and cynomolgus macaques are believed to have diverged from a common ancestor in Indochina after that ancestor reached the mainland. The completion of the rhesus macaque genome sequence in 2007 [Gibbs and Consortium 2007] led to an explosion of marker discovery, the development of bioinformatics tools for genetic investigation of non-human primates and the evaluation of human disease models. While the rhesus macaque is the most common primate model for human disease research and provides the primary animal model for HIV research [Haigwood 2009], cynomolgus macaques are used to study infectious diseases, such as tuberculosis and malaria, transplantation methods [Doxiadis et al. 2006] and environmental and reproductive toxicology. However, the increasing popularity of the cynomolgus macaque in biomedical research is confronted with a lack of genetic and genomic data.

Rhesus macaques incorporated in biomedical research protocols in the United States are typically either bred domestically from the descendants of animals imported from India prior to an exportation embargo in April, 1978 [Hendrickx and Hendrickson 1988; Kerber and Herbert 1979] or imported directly from breeding facilities in China [VandeBerg 1995; Williams-Blangero 1993], where animals from different regions of China are housed together [Satkoski et al. 2008a]. Although cynomolgus macaques do not occur naturally in China, 59% of them imported into the US during FY2009 were bred at facilities in that country and are essentially of unknown or undocumented geographic origin. In contrast, 19% of the cynomolgus macaques imported to the US during that year originated on the island of Mauritius and 13% were imported directly from either Vietnam or Indonesia [Creager et al. 2011]. Increasing numbers of cynomolgus macaques are being exported from Cambodia both to China and the US, but many cynomolgus macaques exported to the US from Chinese breeding farms are of mixed ancestry because cynomolgus macaques from different origins are not bred separately at these farms.

The ranges of rhesus and cynomolgus macaques overlap in an area of approximately 2000 km stretching from southeastern Bangladesh to central Vietnam [Fooden 1964]. Interbreeding between the two species has been postulated based on behavioral, physiological and genetic evidence [Fooden 2006; Hamada et al. 2006; Malaivijitnond et al. 2008; Tosi et al. 2002]. The inter-species transfer of genes is hypothesized to be primarily unidirectional from rhesus macaques to cynomolgus macaques [Bonhomme et al. 2009], with a genome-wide comparison suggesting that approximately 30% of the cynomolgus macaque genome is of rhesus macaque origin, exclusive of genes identical by descent [Yan et al. 2011]. Recent genetic investigation of the major histocompatibility complex (MHC) of cynomolgus macaques suggests that hybridization has led to large-scale introgression of genes from rhesus macaques to Indochinese cynomolgus macaques both within and beyond the known hybrid zone. Sano [2006] identified four MHC class II DBP1 alleles in both rhesus and cynomolgus macaques, but in all but one case, limited to cynomolgus macaques of Indochinese (Vietnamese) origin. Ling [2011] described a similar pattern among DPB1, DOB and DQB1 alleles – seven alleles were identical between the two species, and were limited to, or of higher frequency in, Vietnamese cynomolgus macaques, compared to cynomolgus macaque samples from other regions. Of the 122 MHC class II alleles discovered in 24 cynomolgus macaques by Creager et al. [2011], 12 alleles, all in animals of Vietnamese origin, were identical to alleles previously identified in rhesus macaques. Given the role that MHC gene products play in immune response to pathogens, the combination of cynomolgus macaque and rhesus macaque alleles in the MHC of Indochinese animals has a potential to produce an immunologic profile that is distinct from both progenitor populations.

While the biomedical use of animals separated by biogeographic barriers may increase the proportion of phenotypic differences (presumably due to genetic differentiation) among

them [Smith et al. 2006], hybridization between rhesus and cynomolgus macaques in the zone of sympatry has the potential to produce a genetically distinct population, not through differentiation but through admixture and recombination. Thus, the combination of Indochinese cynomolgus macaques with animals from other regions may also increase the phenotypic variance of traits that are the focus of research protocols, obscure positive results through spurious association and require a larger number of research subjects to resolve phenotypic differences that are due to experimental treatment effects.

In this study, we compare genome-wide SNP genotypes from Vietnamese cynomolgus macaques with those of Chinese-origin rhesus macaques and cynomolgus macaques collected from Indonesia. Assuming that Chinese-origin rhesus macaques and Indonesian cynomolgus macaques are appropriate proxies for their respective ancestors and are, therefore, representative of the ancestral parental sources of Indochinese cynomolgus macaques, we employ bioinformatics methods to identify regions of the rhesus macaque genome that are significantly overrepresented into the genome of Indochinese cynomolgus macaques when compared to a null model of admixture. We postulate that while not necessarily a phylogenetically distinct group, Indochinese cynomolgus macaques are genetically distinct from other cynomolgus macaque populations and that this fact should be considered when using cynomolgus macaques of Indochinese or unknown origin in biomedical research.

Methods

The research described herein was approved by the Institutional Animal Care and Use Committee of the University of California-Davis (#16549) and adheres to the American Society of Primatologists (ASP) Principles for the Ethical Treatment of Non-Human Primates and all legal requirements for vertebrate animal research in the United States. We extracted DNA from 100 rhesus macaques obtained from a breeding farm in Suzhou, eastern China and 14 cynomolgus macaques of Indonesian origin that are housed at the California National Primate Research Center (CNPRC). DNA was also extracted from 19 Vietnamese and seven southern Sumatran cynomolgus macaques obtained from Primate Products, Inc. (Immokalee, FL). DNA was extracted from all samples using a guanadinium thiocyanate protocol without a phenol/chloroform step (Froenicke & Sullivan, pers. comm.), with the exception of those from Sumatra that were stored in PaxGene Blood DNA Tubes (Qiagen, Inc., Valencia, CA) and extracted using the manufacturer's recommended protocol. After extraction, all DNA samples were normalized to a concentration of 60 ng/μl.

All samples (100 Chinese rhesus macaques, 19 cynomolgus macaques, and 21 Indonesian cynomolgus macaques) were submitted to the DNA Technologies Core at the University of California-Davis Genome Center for SNP genotyping on the Illumina GoldenGate™ platform. The 3072 SNPs genotyped are spaced approximately every 750 kilobases (Kb) along the 20 autosomes and every 1.9 megabases (Mb) on the X chromosome. No markers were located on the Y chromosome due to the lack of a rhesus Y reference sequence. The locations and properties of these SNPs are described by Satkoski Trask et al. [2011] and validation methods are described in Malhi et al. [2007] and Satkoski et al. [2008b]. The genotypes were edited to include only markers that were polymorphic in at least one of the three samples (Chinese rhesus macaques, Vietnamese cynomolgus macaques or Indonesian cynomolgus macaques), resulting in a total of 2808 genotypes for each individual.

To confirm the provenience of the parental populations, we used the SNP genotypes for each population as input for the software ADMIXTURE 1.22 in cross-validation mode [Alexander et al. 2009] assuming from 1 to 10 populations. To create the input file for ADMIXTURE, the Indonesian cynomolgus macaque and Chinese rhesus macaque samples

were edited to remove any SNPs that were monomorphic in one or both parental populations or had a minor allele frequency of less than 0.01. When the resulting edited data sets were merged, 1747 genotypes remained. The trimmed parental genotypes were then analyzed along with the corresponding genotypes in the Vietnamese cynomolgus macaque sample.

All subsequent analyses were run on the full 2808 SNP data set. Minor allele frequencies (MAF) and observed heterozygosities were calculated with PLINK 1.07 [Purcell et al. 2007, <http://pngu.mgh.harvard.edu/purcell/plink/>]. Comparisons of MAF and observed heterozygosity were performed with R version 2.13.1 [Team 2011]. A Bartlett test of homogeneity of variance [Snedecor and Cochran 1989] was used to compare sample variance of MAF and observed heterozygosity between pairs of the three samples. Two paired t-tests, one assuming unequal variance and one assuming equal variance [Archie 1985], were used to compare MAF and observed heterozygosities between samples. Hybrid index [Buerkle 2005] and interspecific heterozygosity were calculated with the INTROGRESS package for R [Gompert and Buerkle 2010]. The former is an estimate of the proportion of the individual's genome inherited from each of the two presumed parental species (Chinese rhesus and Indonesian cynomolgus macaques), calculated via a maximum likelihood method. The latter is the proportion of an individual's genotypes comprised of heterozygotes representing one allele from each parental population. In addition to hybrid index and interspecific heterozygosity, INTROGRESS uses the genomic clines method to estimate the change in marker genotype frequencies across the entire genome. The INTROGRESS algorithm performs permutations to generate a neutral expectation of genotype frequency for statistical significance testing based on a null model of introgression. For this analysis, SNP alleles were designated as co-dominant. INTROGRESS outputs a *p* value based on the log likelihood ratio test statistic for each marker. The markers that significantly exceed neutral expectation are likely to be linked to genes contributing to reproductive isolation or hybrid vigor. To account for multiple testing we used the Bonferroni correction and considered SNPs significantly introgressed only when $p < 0.05/2808 = 1.8 \times 10^{-5}$. The identities of the chromosome and nucleotide position of each such SNP were used to generate a custom track for the UCSC genome browser. To detect these SNPs with the greatest potential impact on phenotype, we used the UCSC table browser and the rhesus macaque genome in combination with the GALAXY web tool [Blankenberg et al. 2010; Fujita et al. 2010; Goecks et al. 2010] to identify those located within rhesus EST or ENSEMBL genes.

These genes were used as search terms in the Gene Ontology Website (AmiGO, <http://www.geneontology.org>) to identify the ontology search term associated with each identified gene or EST. Because term associations are not available for the rhesus or cynomolgus macaque, human reference data and term associations were used as an alternative. All term associations were counted with the GO Classification Counter Tool [Hu et al. 2008] and divided into the three main GO classes.

Results

When ADMIXTURE was run assuming a K value of two, all Indonesian cynomolgus macaques and Chinese rhesus macaques produced Q estimates for their respective groups that equaled or exceeded 0.99. Although K=2 did not necessarily produce the absolute lowest cross validation error or log-likelihood, the analysis did not reach the recommended termination criterion of less than 10^{-4} increase in log-likelihood between iterations even at K=10, an improbably high value.

The sample variances of observed heterozygosity of the Vietnamese and Indonesian cynomolgus macaques were equal and both differed significantly from that of the Chinese

rhesus macaque sample. Both paired t-tests (with the assumptions of equal and unequal variance) for the comparison of observed heterozygosities between the Chinese rhesus macaque and Vietnamese cynomolgus macaque samples and between the Vietnamese and Indonesian cynomolgus macaque samples were statistically significant at, or below, the 2.2×10^{-15} level of probability, with the observed heterozygosity in Vietnamese cynomolgus macaques being statistically significantly smaller than that of Chinese rhesus macaques and statistically significantly greater than that of Indonesian cynomolgus macaques.

MAF values are compared between each pair of samples in Figure 2. The variance in MAF for Vietnamese cynomolgus macaques was statistically significantly different from those of both the Chinese rhesus macaques and the Indonesian cynomolgus macaques. The MAF for Chinese rhesus macaques was statistically significantly higher than that for Vietnamese cynomolgus macaques, which in turn, was statistically significantly higher than those for Indonesian cynomolgus macaques (in both comparisons, $p < 2.2 \times 10^{-16}$). The linear relationship between MAF and observed heterozygosity in the three populations was calculated with a second-order polynomial and is shown with a solid line in Figure 3, while the relationship between MAF and expected heterozygosity is illustrated with a dashed line.

Individual hybrid index and interspecific heterozygosity for the Vietnamese cynomolgus macaques are shown in Figure 4. While the majority of the 19 individuals exhibited a hybrid index between 0.60 and 1 (approximately 60-100% of the genome inherited from cynomolgus macaques), three individuals had a hybrid index between 0 and 0.25 (0-25% of the genome inherited from cynomolgus macaques). These three values suggest that rather than being first generation hybrids; these three sampled Vietnamese cynomolgus macaques are the product of several generations of interspecific breeding involving backcrossing of first generation inter-specific hybrids to rhesus macaques. This result is supported by the results of the ADMIXTURE analysis; when degree of admixture in the Vietnamese cynomolgus macaque population was estimated using 1747 markers, the correspondence between Q value and hybrid index was very high, producing an R-squared value of 0.97, as shown in Figure 5.

Equivalent genetic contribution from both parental populations should produce hybrids with a greater interspecific heterozygosity than either parent population [Gompert and Buerkle 2010]. In this case, the three Vietnamese cynomolgus macaques more genetically similar to Chinese rhesus macaques have the largest interspecific heterozygosity, with the values declining as the hybrid index increases. This result is concordant with the estimates of MAF and observed heterozygosity, both of which indicate that the genetic variation in both Indonesian and Indochinese cynomolgus macaques is a subset of the alleles found in Chinese rhesus macaques.

Of the 2808 markers tested, 216 (8%) showed a statistically significant difference in genotype frequencies between the two parent populations (Chinese rhesus and Indonesian cynomolgus macaques) and Vietnamese cynomolgus macaques, relative to the null expectation, suggesting they are overrepresented in the Vietnamese cynomolgus macaques relative to the Indonesian conspecifics. These markers were located on all 20 autosomes and the X chromosome. Figure 6 shows the number of statistically significant SNPs per Mb of chromosome length ($p < 1.8 \times 10^{-5}$), which was normally distributed (Shapiro-Wilk normality test, $P = 0.8684$). As illustrated in Figure 6, no chromosome exceeded the upper 95% CI, but chromosome 8 fell below the lower CI, indicating that markers are under-represented on this chromosome.

The 216 statistically significant markers were used to create a custom SNP track, and through the Galaxy interface, this track was loaded into the UCSC Main Table Browser. The

custom track was compared with the following tracks to calculate intersection: rhesus macaque EST (1 intersection), rhesus macaque mRNA (0 intersections), rhesus macaque RefSeq genes (0 intersections) and rhesus macaque ENSEMBL genes (3 intersections). Ten intersections with mRNA from organisms other than rhesus macaques were observed, but few of these mRNAs have been associated with known genes, making their identification of limited use. Both the rhesus macaque EST and ENSEMBL tracks identified intersection with the *ELAVL1* gene [(embryonic lethal, abnormal vision, *Drosophila*)-like 1] on rhesus chromosome 19, while the ENSEMBL track identified intersection with the genes *PIK3CG* (phosphoinositide-3-kinase, catalytic, gamma polypeptide) on rhesus macaque chromosome 3 and *GPR78* (G protein-coupled receptor 78) on rhesus macaque chromosome 5.

Because the Gene Ontology website does not contain functional information for the rhesus macaque, human associations were used instead. The AmiGO tool on the Gene Ontology website identifies all the known functional roles of each gene (called term associations) and classifies the child terms into three broader categories: “biological processes”, or multi-step molecular interactions, “cellular components”, or a product that forms part of a larger object within the cell, and “molecular function”, or catalytic or binding activities that occur at the molecular level. AmiGO identified 13 term associations for *ELAVL1*, 15 term associations for *PIK3CG* and 6 term associations for *GPR78*. The distribution of term associations across the three AmiGO functional categories is shown in Table 1. When the term associations were considered individually, *GPR78* did not share any associations with either of the other genes. However, *PIK3CG* and *ELAVL1* shared two term associations: presence in the cytosol (GO: 0005829) and nucleotide binding (GO: 0000166).

Discussion

The polymorphisms used in this study were discovered through the sequencing of rhesus macaques, and not through the sequencing of cynomolgus macaques [Malhi et al. 2007; Satkoski et al. 2008b], and as such, the issue of ascertainment bias must be addressed. Although cynomolgus macaques were not sequenced, successful SNP genotyping requires an exact match between the target sequence and the genotyping probe – differences in DNA sequence between the target animal and the probe will produce a null result and signals resulting from duplication of the SNP site will not meet quality screening criteria [Kanthaswamy et al. 2012]. Thus, identity of the SNP position as well as the immediately surrounding nucleotides, although not tested explicitly, is very likely to be extremely similar, if not exactly the same.

Although ascertainment bias produced by genotyping genetic markers discovered in one species to another closely related species can influence population-based studies, the allele frequency differences between rhesus and cynomolgus macaques not only meets the assumptions of the INTROGRESS software, but is actually advantageous. This is because the INTROGRESS package uses alleles specific to parental populations in combination with differences in the frequency of shared alleles to calculate the likelihood of relative ancestry in hybrid individuals [Gompert and Buerkle 2010]. The greater the number of population-specific alleles and the greater the allele frequency difference between the parental populations, the greater the accuracy with which hybrid origin and introgression can be assessed. The only disadvantage is that because these markers were discovered in rhesus macaques, private alleles are represented only in the Chinese rhesus population – there are no alleles private to Indonesian cynomolgus macaques [Kanthaswamy et al. 2010]. Thus, we can only assess unidirectional introgression, that is, from rhesus macaques to cynomolgus macaques. To identify introgression in the other direction with this method would require a comparable set of polymorphisms private to cynomolgus macaques. This necessitates

additional sequencing in regional populations of cynomolgus macaques, an effort that is currently underway [Satkoski Trask et al., in prep.].

These results support previous research suggesting that interbreeding between rhesus and cynomolgus macaques in the Indochinese zone of sympatry has led to substantial gene flow between the two species. If a polymorphism is found in both rhesus macaques and all or most populations of cynomolgus macaques, it is likely that the mutation occurred in the common ancestor of both and predates the divergence of the two species. However, if a polymorphism is found in Indochinese cynomolgus macaques and Chinese rhesus macaques to the exclusion of other cynomolgus macaque populations, the most likely hypothesis is that the allele moved across the species boundary. Although some of these shared polymorphisms are undoubtedly the result of convergent evolution and homoplasy occurring post-speciation, the probability of homoplasy is very low and probably accounts for only a small number of the shared SNPs.

The common ancestor of the *fascicularis* species group (containing *M. cyclopis* and *M. fuscata* along with *M. mulatta* and *M. fascicularis*) may have been present in South East Asia as early as 2.5 million years ago [Delson 1980; Fooden 1976], with the divergence of rhesus macaques from a fascicularis-like ancestor on the Asian mainland occurring about 1 mya [Morales and Melnick 1998; Tosi et al. 2003]. The distributions of mitochondrial and Y-chromosome markers in cynomolgus and rhesus macaques support contemporary introgression [Tosi et al. 2002; Tosi et al. 2003]. Given the widely accepted Indonesian origin and northward dispersal route hypothesized by Delson [1980] and Fooden [2006], in the absence of introgression, the Indonesian sample, being of greater antiquity, should be more variable than the Vietnamese sample. As illustrated in Figures 1 and 2, this is not the case, suggesting that introgressive hybridization has occurred in the Vietnamese cynomolgus macaques. Although an Indonesian origin for cynomolgus macaques is supported by the bulk of morphological and genetic data [Fooden 1997; Tosi et al. 2002], Melnick and Kidd [1985] suggested that greater genetic similarity between Chinese rhesus macaques and cynomolgus macaques in Thailand results from the divergence of rhesus and cynomolgus macaques within Thailand during a glacial maximum, followed by dispersal of cynomolgus macaques to the south. Under this alternative hypothesis, allele sharing between Chinese rhesus macaques and Indochinese, but not Indonesian, cynomolgus macaques would reflect shared derived traits rather than introgression, and the greater variability of Indochinese cynomolgus macaques would be due to clinal variation, rather than hybridization alone.

While Vietnamese cynomolgus macaques are not as variable or heterozygous as Chinese rhesus macaques, they have both significantly higher MAF and observed heterozygosity than Indonesian cynomolgus macaques; the southward gene flow appears to have differentiated the Vietnamese cynomolgus macaques from their Indonesian conspecifics. The greater frequency of minor alleles suggests that alleles that are rare or absent in the Indonesian sample are being supplemented in the Indochinese cynomolgus macaque population through interbreeding with rhesus macaques, a source of alleles not available to other, insular cynomolgus macaque populations. In the Chinese rhesus macaque sample, MAF is strongly predictive of observed heterozygosity ($R^2=0.95$), compared to the sample of Indonesian cynomolgus macaques ($R^2=0.76$). However, this relationship between MAF and observed heterozygosity has decayed even further in Vietnamese cynomolgus macaques, largely due to markers of high MAF but low heterozygosity.

The sample of Vietnamese cynomolgus macaques has a much higher average minor allele frequency than their Indonesian conspecifics, but lower than expected heterozygosity (Figure 3). This departure suggests a force of evolution at work, possibly selection against heterozygote genotypes. The common expectation for cases of low heterozygote fitness is

the maintenance of the novel allele at a low frequency and its eventual extinction [Barton and Pialek 1997]. This is clearly not the case in the Vietnamese cynomolgus macaques. As shown in the lower half of Figure 4, the majority of Vietnamese cynomolgus macaques have hybrid index values of 70% or greater, with interspecific heterozygosity ranging from 5-20%. The high hybrid index values combined with the low interspecific heterozygosity suggest that these individuals are actually the product of at least one generation of backcrossing – the mating of hybrid individuals with a parent population, in this case, cynomolgus macaques. Were backcrossing (rather than hybrid x hybrid mating) the norm, alleles of rhesus macaque origin would be embedded in a largely cynomolgus macaque genomic background. The fitness conferred by a particular introgressed allele is not limited to the effect produced by the genotype, but is also dependent on the action of the allele in a recombinant genomic background [Barton 2001]. It could be hypothesized that neither homozygote genotype is subject to negative selection in these backcrossed individuals, but that the fitness of heterozygotes condition is diminished.

While cynomolgus macaques in Indochina are generally more genetically variable than insular cynomolgus macaques, these statistics describe the entire Vietnamese cynomolgus macaque sample; when individuals are considered, the proportion of probable rhesus ancestry varies considerably, with the INTROGRESS analysis estimating that the majority of the sampled Vietnamese cynomolgus macaques have between 0 and 25% rhesus ancestry. However, the presence of a few individuals with a large proportion of rhesus ancestry and the lack of 50% hybrids in the sample suggest that 1) hybrids are fertile and present in the breeding populations of both species, 2) the majority of hybrid individuals are the result of several interspecies recombination events and 3) hybridization is not a recent phenomenon. Fooden [1964] suggested that the proportion of the rhesus macaque genome transferred to Indochinese cynomolgus macaques may vary geographically, with cynomolgus macaques in the north carrying more rhesus macaque alleles than animals closer to peninsular Malaysia. More detailed genetic sampling within the Indochinese ranges of cynomolgus and rhesus macaques would be required to test this hypothesis, with Indochinese rhesus macaques providing a more accurate model for the parent population of hybrids than the Chinese rhesus used in this study. Finally, although the reduced interspecific heterozygosity of the more cynomolgus-like individuals would seem to support unidirectional, rhesus to cynomolgus, interbreeding, the lack of alleles specific to cynomolgus macaques in the present analysis prevents the rigorous testing of this hypothesis.

Of the 2808 markers included in the INTROGRESS analysis, only 216, or approximately 8%, were statistically significantly over-represented in Vietnamese cynomolgus macaques relative to their Indonesian conspecifics. In addition, rhesus macaque alleles from markers located on chromosome 8 are significantly *under*-represented in the cynomolgus macaque genome. Functional interpretation of the genes located on rhesus chromosome 8 is generally lacking, making this result difficult to interpret at the present time. One hypothesis for the conserved nature of the chromosome is presented by Burrows et al. [2010] who identified a sequence orthologous to the human DUB/USP17 subfamily of enzymes near a copy-number variant beta-defensin gene on chromosome 8 in rhesus macaques. The authors state that while the copy number is highly unstable, the individual genes are conserved at a sequence level, probably to preserve the important functions of both genes in the primary immune system. This conservation is one potential explanation for why Vietnamese cynomolgus macaques have resisted introgression on chromosome 8, relative to other chromosomes.

When the AmiGo functional categories of the three genes were considered, *ELAVL1* and *PIK3CG* were similar in that they were approximately equally divided between “biological process” and “molecular function”. They also had cytosol and nucleotide binding activity in common. Both genes code for products that appear to be generally active in immune

function. ELAVL1 is especially nonspecific in its action, with approximately 15% of the human transcriptome targeted, including (but not limited to) proteins for cell apoptosis and immune response [Uren et al. 2011]. The nucleotide binding functionality of the PIK3CG protein operates within the TLR signaling pathway [Pennings et al. 2008], which regulates the primary immune response [Alderem and Ulevitch 2000]. While most *GPR78* AmiGo term associations were classified as “molecular function”, this gene product is also generally active in immune response, including protecting cells from damage and apoptosis [Ramsay et al. 2005].

While a greater density of SNPs could potentially identify more significantly over-represented alleles, the three genes detected by this analysis all have roles in the immune system. This shared property supports the hypothesis that the incorporation of gene variants from Chinese rhesus macaques into Indochinese populations of cynomolgus macaques has caused immune differentiation and potentially allowed rapid adaptation to pathogens not encountered by insular cynomolgus macaque populations. Wheatley [1980] suggested that the malarial plasmodium *P. knowlesi*, which is fatal to rhesus macaques but not to cynomolgus macaques, shapes the ranges of the two species, but adaptation of resistance to *P. knowlesi* would favor the introgression of cynomolgus genes into rhesus macaques, an hypothesis whose test would require use of SNPs identified in cynomolgus macaques rather than the rhesus macaque SNPs employed in the present study. Nevertheless, the three Vietnamese cynomolgus macaques that exhibited 0-25% cynomolgus macaque ancestry suggest the possibility of bidirectional introgression. Moreover, malaria is not the only pathogen to which cynomolgus and rhesus macaques display differential susceptibility. For example, both species also exhibit differential susceptibility to *Mycobacterium tuberculosis* and Simian Immunodeficiency Virus. The identification of additional shared alleles that reflect adaptations to these and other infectious pathogens and a further investigation of the genes identified in this study will elucidate the immune differences between regional populations of rhesus and cynomolgus macaques.

Phenotypic and genetic evidence for hybridization has been observed in the 2000 km boundary between the ranges of rhesus and cynomolgus macaques. MHC class II alleles shared between rhesus macaques and Indochinese cynomolgus macaques support the hypothesis that not only are genes being transferred across the species boundary, but that the specific genes being transferred could produce a unique immunological phenotype in Indochinese cynomolgus macaques. Interbreeding between rhesus macaques and cynomolgus macaques in the Indochinese zone of sympatry has produced a population that appears to be genetically intermediate between, but significantly differentiated from, both parent populations, and highly variable in individual hybridity. These results bring into question the general suitability of Indochinese cynomolgus macaques as a biomedical model.

Our study found that the MAF and observed heterozygosity for 2808 SNP markers calculated from a sample of Vietnamese cynomolgus macaques was not only significantly different from the values calculated from a sample of Chinese rhesus macaque but also from Indonesian cynomolgus macaques. The program INTROGRESS identified 216 SNPs at which alleles from rhesus macaques were significantly over-represented in the Vietnamese cynomolgus sample compared to the Indonesian cynomolgus sample. Three of these markers were located within genes: *ELAVL1*, *PIK3CG* and *GPR78*. Although each of the genes' products plays multiple roles in the cell, each influences immune function. These results demonstrate that even when the scope of genomic inquiry is expanded beyond the MHC, the genes crossing the cynomolgus/rhesus species boundary is still primarily related to immune function. This supports the hypothesis that introgression represents a source of genetic variation in the immune system that provides Indochinese cynomolgus macaques

with resistance to an as yet unidentified pathogen or group of pathogens. Given that the majority of cynomolgus macaques imported for biomedical research are either collected in or near the zone of sympatry or are exported from China with unknown (but probably predominantly Indochinese) geographic origin, the possibility that they represent a distinct, extremely variable genetic entity has important implications for their use as subjects in research protocols.

Acknowledgments

This work has been supported by NIH AARA supplement RR018144-07 to SK and NL and NIH grants RR05090-20 and RR025871-03 to DGS. The authors would also like to thank the staff of the Molecular Anthropology Lab, University of California – Davis as well as the staff of Primate Products, Inc. for assistance in preparing samples for SNP genotyping and Dr. Charles Nicolet and Vanessa Rashbrook for their invaluable assistance with SNP detection and validation studies.

References

- Abegg C, Thierry B. Macaque evolution and dispersal in insular South-East Asia. *Biological Journal of the Linnean Society*. 2002; 75(4):555–576.
- Alderem A, Ulevitch RJ. Toll-like receptors in the induction of the innate immune response. *Nature*. 2000; 406:782–787. [PubMed: 10963608]
- Alexander DH, Novembre FJ, Lange K. Fast model-based estimation of ancestry is unrelated individuals. *Genome Research*. 2009; 19:1655–1664. [PubMed: 19648217]
- Archie J. Statistical Analysis of Heterozygosity Data: Independent Sample Comparisons. *Evolution*. 1985; 39(3):623–637.
- Barton NH. The role of hybridization in evolution. *Molecular Ecology*. 2001; 10(3):551–568. [PubMed: 11298968]
- Barton NH, Pialek J. The spread of an advantageous allele across a barrier: the effects of random drift and selection against heterozygotes. *Genetics*. 1997; 145(2):493–504. [PubMed: 9071602]
- Blankenberg D, Von Kuster G, Corador N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J. Galaxy: a web-based genome analysis tool for experimentalists. *Current Protocols in Molecular Biology*. 2010:1–21.
- Bonhomme M, Cuartero S, Blancher A, Crouau-Roy B. Assessing natural introgression in 2 biomedical model species, the rhesus macaque (*Macaca mulatta*) and the long-tailed macaque (*Macaca fascicularis*). *Journal of Heredity*. 2009; 100(2):158–169. [PubMed: 18974398]
- Buerkle C. Maximum likelihood estimation of a hybrid index based on molecular markers. *Molecular Ecology*. 2005:684–1371. Notes 5(79dcb01b-f1aa-d2a2-435b-b0e0758b1a0d).
- Burrows JF, Scott CJ, Johnston JA. The DUB/USP17 deubiquitinating enzymes: A gene family within a tandemly repeated sequence, is also embedded within the copy number variable Beta-defensin cluster. *BMC Genomics*. 2010; 11:250. [PubMed: 20403174]
- Creager HM, Becker EA, Sandman KK, Karl JA, Lank S, Bimber BN, Wiseman RW, Hughes AL, O'Connor SL, O'Connor DH. Characterization of full-length MHC class II sequences in Indonesian and Vietnamese cynomolgus macaques. *Immunogenetics*. 2011
- Delson, E. Fossil macaques, phyletic relationships and a scenario of deployment. In: Lindburg, DG., editor. *The Macaques: Studies in Ecology, Behavior and Evolution*. New York: Van Nostrand-Reinhold; 1980. p. 10-30.
- Doxiadis GGM, Rouweler AJM, de Groot NG, Louwerse A, Otting N, Verschoor EJ, Bontrop RE. Extensive sharing of MHC class II alleles between rhesus and cynomolgus macaques. *Immunogenetics*. 2006; 58(4):259–268. [PubMed: 16470376]
- Fooden J. Rhesus and crab-eating macaques: intergradation in Thailand. *Science*. 1964; 143(3604): 363–364. [PubMed: 17821058]
- Fooden J. Provisional classification and key to living species of macaques (Primates: *Macaca*). *Folia Primatologica*. 1976; 25:225–236.
- Fooden J. Tail length variation in *Macaca fascicularis* and *M. Mulatta*. *Primates*. 1997; 38(3):221–232.

- Fooden J. Comparative review of fascicularis-group species of macaques (primates: *Macaca*). *Fieldiana Zoology*. 2006; 107:1–43.
- Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, et al. The UCSC Genome Browser database: update 2011. *Nucleic Acids Research*. 2010; 39:1–7. [PubMed: 20805246]
- Gibbs R. Consortium RMGSaA. Evolutionary and biomedical insights from the rhesus macaque genome. *Science*. 2007; 316:222–234. [PubMed: 17431167]
- Goecks J, Nekrutenko A, Taylor J, Team TG. Galaxy: a comprehensive approach for supporting accessible, reproducible and transparent computational research in the life sciences. *Genome Biology*. 2010; 11:R86. [PubMed: 20738864]
- Gompert Z, Buerkle CA. INTROGRESS: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*. 2010; 10(2):378–384. [PubMed: 21565033]
- Haigwood NL. Update on animal models for HIV research. *European Journal of Immunology*. 2009; 39:1991–2058. [PubMed: 19672889]
- Hamada Y, Urasopon N, Hadi I, Malaivijitnond S. Body size, proportions and pelage color of free-ranging rhesus macaques (*Macaca mulatta*) from a zone of hybridization in northern Thailand. *International Journal of Primatology*. 2006; 27:497–513.
- Hendrickx, AG.; Hendrickson, R. Breeding rhesus monkeys in outdoor cages. In: Neubert, D.; Hendrickx, AG.; Merker, HJ., editors. *Non-Human Primates: Developmental Biology and Toxicology*. Vienna: Ueberreuter Wissenschaft; 1988. p. 41–52.
- Hu Z-L, Bao J, Reecy JM. CateGORizer: a web-based program to batch analyze gene ontology classification categories. *Online Journal of Bioinformatics*. 2008; 9(2):108–112.
- Kanthaswamy S, Satkoski J, Kou A, Malladi V, Glenn Smith D. Detecting signatures 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*. 2010; 39(4):252–265. [PubMed: 20618591]
- Kanthaswamy S, Satkoski Trask JA, Ross CT, Kou A, Houghton P, Smith DG, Lerche N. A large-scale SNP-based genomic admixture analysis of the captive rhesus macaque colony at the California National Primate Research Center. *American Journal of Primatology*. 2012; 74(8):747–757.
- Kerber WT, Herbert HJ. Establishing a free-ranging colony of rhesus monkeys, part I. *Journal of Medical Primatology*. 1979; 8:129–142. [PubMed: 116000]
- Ling F, Wei L-q, Wang T, Wang H-b, Zhuo M, Du H-l, Wang J-f, Wang X-n. Characterization of the major histocompatibility complex class II DOB, DPB1, and DQB1 alleles in cynomolgus macaques of Vietnamese origin. *Immunogenetics*. 2011; 63(3):155–166. [PubMed: 21132285]
- Malaivijitnond S, Wattana SL, Yuzuru H. The human ABO blood groups of free-ranging long-tailed macaques (*Macaca fascicularis*) and parapatric rhesus macaques (*M. mulatta*) in Thailand. *Journal of Medical Primatology*. 2008; 37(31–37)
- Malhi RS, Sickler B, Lin D, Satkoski J, Tito RY, George D, Kanthaswamy S, Smith DG. *MamuSNP*: a resource for rhesus macaque (*Macaca mulatta*) genomics. *PLoS ONE*. 2007; 2:e438. [PubMed: 17487284]
- Melnick DJ, Kidd KK. Genetic and evolutionary relationships among Asian macaques. *International Journal of Primatology*. 1985; 6(2):123–160.
- Morales JC, Melnick DJ. Phylogenetic relationships of the macaques (Cercopithecidae: *Macaca*) as revealed by high resolution restriction site mapping of mitochondrial ribosomal genes. *Journal of Human Evolution*. 1998; 34(1):1–23. [PubMed: 9467779]
- Pennings JLA, Kimman TG, Janssen R. Identification of a common gene expression response in different lung inflammatory diseases in rodents and macaques. *PLoS ONE*. 2008; 3(7):e2596. [PubMed: 18612392]
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, Bakker PIW, Daly MJ, et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*. 2007; 81:559–575. [PubMed: 17701901]

- Ramsay RG, Ciznadija D, Mantamadoitis T, Anderson R, Pearson R. Expression of stress response protein regulated protein-78 mediated by c-Myb. *The International Journal of Biochemistry and Cell Biology*. 2005; 37(6):1254–1268.
- Sano K, Shiina T, Kohara S, Yanagiya K, Hosomichi K, Simuizu S, Anzai T, Watanabe T, Ogasawara K, Torii R, et al. Novel cynomolgus macaque MHC-DPB1 polymorphisms in three South-East Asian populations. *Tissue Antigens*. 2006; 67(4):297–306. [PubMed: 16634866]
- Satkoski J, George D, Smith D, Kanthaswamy S. Genetic characterization of wild and captive rhesus macaques in China*. *Journal of Medical Primatology*. 2008a; 37(2):67–80. [PubMed: 18333917]
- Satkoski J, Malhi RS, Kanthaswamy S, Tito RY, Malladi V, Smith DG. Pyrosequencing as a method for SNP identification in the rhesus macaque (*Macaca mulatta*). *BMC Genomics*. 2008b; 9:256. [PubMed: 18510772]
- Satkoski Trask JA, Garnica WT, Kanthaswamy S, Malhi RS, Smith DG. 4040 SNPs for genomic analysis in the rhesus macaque (*Macaca mulatta*). *Genomics*. 2011; 98(5):352–358. [PubMed: 21907785]
- Smith DG, George D, Kanthaswamy S, McDonough J. Identification of country of origin and admixture between Indian and Chinese rhesus macaques. *International Journal of Primatology*. 2006; 27:881–898.
- Snedecor, GW.; Cochran, WG. *Statistical Methods*. Ames, Iowa: Iowa State University Press; 1989.
- Team RDC. R: a language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2011.
- Tosi AJ, Morales JC, Melnick DJ. Y-chromosome and mitochondrial markers in *Macaca fascicularis* indicate introgression with Indochinese *M. mulatta* and a biogeographic barrier in the Isthmus of Kra. *International Journal of Primatology*. 2002; 23(1):161–178.
- Tosi AJ, Morales JC, Melnick DJ. Paternal, maternal, and biparental molecular markers provide unique windows onto the evolutionary history of macaque monkeys. *Evolution*. 2003; 57(6):1419–1435. [PubMed: 12894949]
- Uren PJ, Burns SC, Ruan J, Singh KK, Smith AD, Penalva LOF. Genomic analyses of the RNA binding protein Hu Antigen R (HuR) identify a complex network of target genes and novel characteristics of its binding sites. *The Journal of Biological Chemistry*. 2011
- VandeBerg, JL. Genetics of nonhuman primates. In: Taylor Bennett, B.; Abee, CR.; Hendrickson, R., editors. *Nonhuman Primates in Biomedical Research*. San Diego: Academic Press; 1995. p. 129-146.
- Wheatley B. Malaria as a possible selective factor in the speciation of macaques. *Journal of Mammology*. 1980; 61:307–311.
- Williams-Blangero S. Research-oriented genetic management of nonhuman primate colonies. *Laboratory Animal Science*. 1993; 42:535–540. [PubMed: 8158976]
- Yan G, Zhang G, Fang X, Zhang Y, Li C, Ling F, Cooper D, Li Q, Li Y, van Gool A, et al. Genome sequencing and comparison of two nonhuman primate animal models, the cynomolgus and Chinese rhesus macaques. *Nature biotechnology*. 2011; 29:1019–1042. d5ad3af3-e68d-f781-64dd-638c728ac299.

Figure 1.
Observed heterozygosity of 2808 SNP markers in three macaque populations.

Figure 2.
Minor allele frequencies of 2808 SNP markers in three macaque populations.

Figure 3.

This scatterplot illustrates the relationship between minor allele frequency and observed heterozygosity in the three regional macaque samples. In each, the dashed line represents the relationship between minor allele frequency and expected heterozygosity, while the solid line is fit to the second-order polynomial relationship between MAF and H_O .

Figure 4.

The upper half of the figure shows the interspecific heterozygosity (proportion of the individual's genome containing genotypes composed of alleles inherited from both parental populations) of the Vietnamese cynomolgus macaque samples relative to the hybrid index (the genetic similarity of the individual to either of the parental populations). The line represents the theoretical maximum value of interspecific heterozygosity for each value of hybrid index. Individuals falling along this line are likely to be F1 or backcross progeny [Gompert and Buerkle 2010]. The lower half of the figure is a histogram illustrating the number of individuals with particular hybrid index values.

Comparison of Hybrid Index and Admixture

Figure 5. Results of ADMIXTURE Q values for each Vietnamese cynomolgus macaque individual regressed on hybrid index, as calculated by INTROGRESS.

Figure 6. Number of statistically significantly introgressed ($p < 1.8 \times 10^{-5}$) rhesus macaque alleles. Upper and lower 95% CI boundaries are marked with dashed lines, while the solid line is the mean across all chromosomes.

Table 1

The number of AmiGo gene function term associations for each of the three identified genes. The percent of total term associations is shown in parentheses.

Gene	Biological Process	Cellular Component	Molecular Function	Total
<i>ELAVL1</i>	5 (38.5%)	3 (23%)	5 (38.5%)	13
<i>PIK3CG</i>	7 (47%)	2 (13%)	6 (40%)	15
<i>GPR78</i>	1 (17%)	2 (33%)	3 (55%)	6