Molecular Ecology (2009) **18**, 2457–2475 doi: 10.1111/j.1365-294X.2009.04212.x

Divergence population genetic analysis of hybridization between rhesus and cynomolgus macaques

LAURIE S. STEVISON*,† and MICHAEL H. KOHN*
Department of Ecology & Evolutionary Biology, Rice University, MS170, PO Box 1892, Houston, TX 77251-1892, USA, †Biology Department, Duke University, French Family Science Center 4244, Box 90338, Durham, NC 27708, USA

Abstract

The geographic ranges of rhesus (*Macaca mulatta*) and cynomolgus (*M. fascicularis*) macaques adjoin in Indochina where they appear to hybridize. We used published and newly generated DNA sequences from 19 loci spanning ~20 kb to test whether introgression has occurred between these macaque species. We studied introgression at the level of nuclear DNA and distinguished between incomplete lineage sorting of ancestral polymorphisms or interspecific gene flow. We implemented a divergence population genetics approach by fitting our data to an isolation model implemented in the software IMa. The model that posits no gene flow from the rhesus into the cynomolgus macaque was rejected ($P = 1.99 \times 10^{-8}$). Gene flow in this direction was estimated as $2Nm$~1.2, while gene flow in the reverse direction was nonsignificantly different from zero ($P = 0.16$). The divergence time between species was estimated as ~1.3 million years. Balancing selection, a special case of incomplete sorting, was taken into consideration, as well as potential crossbreeding in captivity. Parameter estimates varied between analyses of subsets of data, although we still rejected isolation models. Geographic sampling of the data, where samples of cynomolgus macaques derived from Indochina were excluded, revealed a lost signature of gene flow, indicating that interspecific gene flow is restricted to mainland Indochina. Our results, in conjunction with those by others, justify future detailed analyses into the genetics of reproductive barriers and reticulate evolution in these two genome-enabled primates. Future studies of the natural hybridization between rhesus and cynomolgus macaques would expand the repertoire of systems available for speciation studies in primates.

Keywords: hybridization, speciation, macaques, gene flow, primates

Received 30 April 2008; revised 21 March 2009; accepted 25 March 2009

Introduction

Research examining introgression between species has investigated the competing pressures gene flow and divergence have on the genome. The combination of various advancements in population genetics methods, computational methods, sequencing methods, and growing sequence data collected for many species (Excoffier & Heckel 2006; Hey 2006), has lead to a great increase in the past decades of careful analyses of gene flow between species and how the interaction between introgression and selection shape various genomes. The initial evidence of introgression was strongly focused in plants, and studies in animals have been growing in the last decade. The study of hybridization in mammals with a published genome sequence has been limited to the mouse (*Mus* spp.). Numerous studies now have utilized the mouse to describe the patterns of population divergence in this genus, and to infer some of the processes that might be driving speciation. While many more details will emerge from studies of hybridization in rodents, it might be of interest also to establish other mammalian systems amenable to the study of natural hybridization.

Studies on hybridization in primates have significant impacts in our understanding of how *Homo sapiens* have

Correspondence: Michael H. Kohn, Fax: 713-348-5232; E-mail: hmkohn@rice.edu

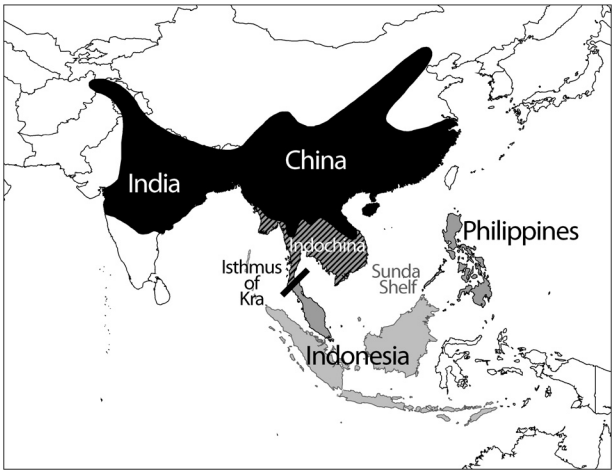

Fig. 1 The geographic ranges of the rhesus and the cynomolgus macaques and the putative area where hybridization between the two species occurs. The geographic distribution of the rhesus macaque and the cynomolgus macaque are shown in black and in grey, respectively (redrawn from Fooden 1980). The putative extent of introgression is highlighted by the hatched region (redrawn from Tosi & Coke 2007).

evolved. Considering the controversy surrounding the possibility of reticulate evolution in the genus *Homo*, it might be timely to initiate the study of natural hybridization in a genome-enabled primate system that, at least partially, can be studied in the laboratory (Patterson *et al*. 2006, 2008; Wakeley 2008). The release of the genome sequence of the macaque (Gibbs *et al*. 2007) and the availability of captive and semi-wild macaque populations, in conjunction with earlier suggestions that hybridization in macaques (and the primate lineage overall) might have been a significant factor that has shaped the evolution of primate genomes (Arnold & Meyer 2006), indicate that a genome-enabled primate system amenable for the study of hybridization might now be available. The potential held by the study of a genome-enabled system (i.e. a system for which a genome sequence has been released and the genomic toolkit is under intensive development) refers to the ability to screen for, and to detect, the differences in the rates of introgression for individual genes or groups of genes that might be particularly important in speciation.

Here we characterize a putative case of natural hybridization between two closely related primate species within the genus *Macaca* that would be well suited for such studies. The rhesus macaque (*Macaca mulatta*) and the cynomolgus macaque (*M. fascicularis*), also known as the crab-eating or long-tailed macaque, have a parapatric geographic distribution with adjoining ranges in Indochina (Fig. 1). The two species belong to the *fascicularis* species group of macaques that contains, in addition to the two aforementioned species, *M. fuscata* and *M. cyclopis* (Fooden 1980). Rhesus and cynomolgus macaques have a proposed divergence time of 0.9–2.5 million years ago, with the minimum estimate based on nuclear loci and the maximum estimate based on mitochondrial sequences (Morales & Melnick 1998; Gibbs *et al*. 2007; Osada *et al*. 2008). A recent analysis based on microsatellite data indicated a divergence time of only 43 500 years, but owing to homoplasy in the microsatellite data this estimate is expected to be downwardly biased (Bonhomme *et al*. 2008). The wide range of divergence time estimates signals a need for additional analyses based on larger data sets of nuclear DNA sequences. Estimates obtained from such data sets should be less affected by homoplasy. Moreover, the long lineage sorting times of nuclear DNA sequences, and the variances thereof, hold highly useful information about the history of speciation of closely related taxa (Hey 2006).

The interpretation of the species status of the rhesus and cynomolgus macaques is complicated by a number of facts. Based on morphological characteristics, Fooden (1964) noted that museum samples presumed to be cynomolgus macaques from Indochina should be classified as potential hybrids between the rhesus and cynomolgus macaques. Weiss *et al*. (1973) noted on protein electrophoretic homoplasies between the two species when cynomolgus macaques from Thailand were compared to rhesus samples. Thailand is located proximal to the southern range limit of the rhesus macaque and to the northern range limit of the cynomolgus macaque (Fig. 1). A number of genetic studies have reported on the sharing of genetic polymorphisms on the Y-chromosome, in some MHC genes, and some nuclear genes of the rhesus and cynomolgus macaques (Morales & Melnick 1998; Tosi *et al*. 2000, 2002, 2003b; Magness *et al*. 2005; Ferguson *et al*. 2007; Blancher *et al*. 2008; Bonhomme *et al*. 2008; Osada *et al*. 2008). The observation of shared variation of Y-chromosome alleles while mitochondrial DNA haplotype trees were monophyletic was considered consistent with recent hybridization between the two species (Tosi *et al*. 2000, 2002). The authors considered it unlikely that the Y-chromosomal variants were ancestral polymorphisms retained by both species, because the philopatric nature of female macaques increases the effective population size of the mitochondrial genome, which otherwise should equal that of the Y-chromosome (Melnick & Hoelzer 1992). Also, the high reproductive success of males relative to females reduces the effective population size of the Y-chromosome further relative to the mitochondrial genome (Hoelzer *et al*. 1998). Owing to the more rapid lineage sorting of ancestral genetic polymorphisms of the mtDNA when compared to the Y-chromosome the opposite would be expected if lineage-sorting times were the sole explanation for the sharing of alleles (Melnick & Hoelzer 1992; Hoelzer *et al*. 1998). Overall, these observations are in conflict with simple models of strictly allopatric speciation

with no gene flow during speciation (parapatric or sympatric speciation) or thereafter (e.g. hybridization upon secondary contact or bouts of reticulate evolution).

Indeed, the present-day parapatric distributions of the two species would make gene flow between the rhesus and cynomolgus macaque a possibility (Fig. 1). Such gene flow could be a result of secondary contact of two previously allopatric species. For example, cynomolgus macaques could have been isolated on the Sunda shelf south of Indochina, and during the lowering of sea levels during the Pleistocene could have recolonized mainland Indochina where cynomolgus macaques now are parapatric with the rhesus macaque (Abegg & Thierry 2002). However, another dispersal scenario for the cynomolgus macaque is under discussion that places the ancestral population of cynomolgus macaques in mainland Indochina whose southern populations, now found in insular Indonesia and the Philippines, have been isolated by Pleistocene glacial phenomena and resulting sea level changes (cf. Blancher *et al*. 2008, and see references therein for a recent discussion). Pleistocene glacial phenomena and sea level changes further maintained genetic subdivisions among cynomolgus macaque populations from the Philippines, other insular populations, and the mainland population thought to be involved in hybridization (Tosi & Coke 2007). Overall, in conjunction with biogeographical evidence, an estimate of the divergence time separating the rhesus and cynomolgus macaques could help clarify whether the two species currently are in secondary contact or have been parapatric in the distant past.

Divergence population genetic analyses of shared nuclear genetic polymorphisms should provide additional support for the proposed hybridization between the rhesus and cynomolgus macaque, and provide an age estimate for the split between the two species while accounting for ancestral genetic variation. Shared nuclear genetic polymorphisms need to be considered carefully because they may mimic gene flow. In addition, balancing selection, as reported for numerous genes involved in immune response, could prolong the time needed for species to become reciprocally monophyletic, thereby falsely implicating gene flow. Further, samples of cynomolgus macaques potentially could be analysed such that it is possible to distinguish secondary contact from more ancient bouts of hybridization. If samples from the proposed hybrid zone would be removed from the analysis, then we might expect the signal for interspecific gene flow to disappear if secondary contact limited the spread of alleles into the cynomolgus macaques' range. Another interesting facet of the system that merits further examination is the apparent unidirectional flow of genes from the rhesus into the cynomolgus genome, and how this may be facilitated by the social structure of macaque societies, and the morphological and behavioural differences between rhesus and cynomolgus macaques. Finally, to be able to promote the hybridization between the rhesus and cynomolgus macaque as a novel genome-enabled model system for evolutionary genetic studies of speciation in primates, we consider the possibility that the proposed hybridization is merely a result of interbreeding in captivity. We recognize, however, that the combination of hybridization and interbreeding in captivity could lead to an exaggeration by the inclusion of captive specimens of cynomolgus macaques that could have crossbred with rhesus macaques.

To explore these questions, we worked within the comparatively novel evolutionary framework of 'divergence population genetics' (Nielsen & Wakeley 2001; Hey & Nielsen 2004, 2007). We fitted autosomal, mitochondrial, X- and Y-linked molecular sequence data collected by sequencing and compiled from GenBank to an isolation model (Hey & Nielsen 2007). Our main objective was to assess the validity of the null model of strict allopatric speciation (allopatric divergence followed by no gene flow) of rhesus and cynomolgus macaques. In addition, because data from the Y-chromosome indicated an asymmetrical pattern of gene flow from the rhesus into the cynomolgus macaque, our second objective was to statistically evaluate the significance of this asymmetrical pattern of gene flow by using autosomal nuclear makers. Our third objective was to further delineate the geographic extent of this hybridization, i.e. whether analyses including cynomolgus macaque samples from outside the proposed zone of hybridization (Fig. 1) would still reveal evidence for gene flow between rhesus and cynomolgus macaque. Finally, we assess whether the signal of interspecific gene flow remains when only wild-derived macaques were included in the analysis.

Materials and methods

Collection of DNA sequence data

Genetic information on the rhesus and the cynomolgus macaques was collected using two approaches. First, published sequence data for 14 loci from GenBank (V157, 12/2006) were chosen for this study. Details on the genes are listed in Table 1. We used the criterion of a minimum of eight haplotypes per species (Felsenstein 2006), with the exception of the Y-chromosomal locus, for which only five haplotypes were available for the rhesus macaque.

Second, we generated our own sequences for the intron-containing regions of one X-chromosomal and four autosomal loci. For sequencing, 15 rhesus and 14 cynomolgus macaque samples were obtained from the New England Primate Research Center (NEPRC; cynomolgus samples received as whole blood preserved in EDTA buffer; rhesus samples received as concentrated DNA), the Oregon National Primate Research Center (ONPRC; concentrated DNA),

Table 1 Genes included in this analysis and their annotations. Compilation of 19 genes analysed in this study, 14 of which were taken from GenBank and five of which were generated during this study. For each gene, the chromosomal location (Chr), sample type (either wild caught with known geographic origin or from primate research centres (PRC) with some samples of known geographic origin), the IMa tests where each gene was included, and accession numbers are indicated.

Gene symbol	Gene name	Chr	Sample type	IMa analyses	Accession numbers
Published sequences used in this study					
3kgg	3K gamma-globin 1 gene, exon 3 and partial cds	14	PRC	I	DQ515100–20, DQ515122–29
abo	*abo* histo-blood group A transferase, exon 7 and partial cds	15	PRC	I	AF052078–86, XM_001118424, AF071830, AF094693&695, AF100981&984
c4	C4 gene, exon 9 and partial cds; and endogenous retrovirus ERV-K, partial sequence	4	Wild	I,II,IV	AY224332–47,47–50
ccl2	Chemokine ligand receptor 2, proximal promoter region and partial cds	16	PRC	I	DQ515029–58
ccr5	Chemokine receptor 5, complete cds	2	PRC	I	AF005660,62, AF161950–54, 58–75, AF291669, DQ902484–543, NM_001042773, U73739, U77672, U96762
dpb1	Major histocompatibility complex, class 2, DP beta-1, partial cds	4	PRC	I	AB235856–87, 89–96, AF024562–65, AM086061–69, AM086165, D13335–36, D16608–10, Z32402–10,12–15
irbp	Interstitial retinol-binding protein, intron 3 and partial cds	9	Wild	I,II,III,IV	AY224277–87,92–95
mhca	Major histocompatibility class I heavy chain antigen, complete cds	4	PRC	I	AB154760–73, AF157394,396–401, AJ539307–8, AJ542567–80, AJ551315–21, NM_001048246, AY707076–77, AY958087–121, DQ979878, EF028175, U50837
mhcg	Major histocompatibility class I antigen G, exon 3	4	PRC	I	L41257–64,U33295,301–306,312
mtdna	12S rRNA, partial; tRNA-Val, complete; 16S rRNA gene, partial	mt	Wild	I,II,III,IV	AF424949–70, EU399479–EU399507
nbpf	Neuroblastoma breakpoint family, member 1, partial cds	1	PRC	I	AY894616–35
nramp1	Natural resistance-associated macrophage protein 1, introns 4 and 5	12	PRC	I,III	AF352993–3027,40–49,52–55,74–77
nry	Testis-specific protein Y, partial cds and sex-determining region Y, partial cds	Y	Wild	I,II,III,IV	AF284297–300,302–304,309–312, AF425282–88,91–95
r1c2	*r1c2* gamma globin 2 gene, promoter region and partial cds	14	PRC	I	DQ515168–97
Sequences generated during this study					
afp	Alpha fetoprotein, intron 1 and partial cds	5	PRC	I,II,III	FJ846491–FJ846518
apoe	Apolipoprotein E, intron 3 and partial cds	19	PRC	I,II,III	FJ846519–FJ846540
b2m	Beta-2 microglobulin, intron 2 and partial cds	7	PRC	I,II	FJ846541–FJ846566
slc25a5	Solute carrier protein family 25 member A5, exon 2–3	X	PRC	I,II,III	FJ846567–FJ846591
ttr	Transthyretin, intron 1 and partial cds	18	PRC	I,II,III	FJ846592–FJ846621

and the New Iberia Research Center (NIRC; whole blood preserved in EDTA buffer). In Table S1, Supporting information, we indicate the known or deduced (see below) geographic origin of samples as described in Stevison & Kohn (2008).

DNA extraction of whole blood samples was performed using QIAamp DNA Blood Midi Kit (QIAGEN). Samples were handled under Rice University Institutional Biosafety protocols. Primers for polymerase chain reaction (PCR) were designed based on the rhesus macaque genome sequence (Gibbs *et al*. 2007) as accessed through the Ensembl genome browser, release 39 (Hubbard *et al*. 2007). The intragenic regions were chosen based on genetic variation in both the rhesus and cynomolgus macaques exhibited by

the genes in an analysis of expressed sequence tagged sites (ESTs) (Magness *et al*. 2005). The loci chosen were two serum/plasma proteins (*afp* and *ttr*), a microglobulin protein (*b2m*), an apolipoprotein (*apoe*), and a membrane protein of the mitochondrion (*slc25a5*).

Primers (provided in Table S2, Supporting information) were placed within the coding regions flanking the introns approximately 1-kb pairs in length. Each locus was amplified from genomic DNA template in 10 μL reaction volumes using a touch-down thermal cycling profile consisting of an initial denaturation (94 °C for 2 min) followed by 15 cycles of a 94 °C denaturation for 30 s, a 65 °C annealing for 30 s reduced by 1 °C/cycle, and a 72 °C elongation for 1 min, followed by 25 cycles at an annealing temperature of 50 °C, and a final elongation at 72 °C for 4 min. PCR products were cleaned using ExoSAP-IT (USB Corporation). An M13-labelled primer was added to each forward and reverse primer and these universal primers were used for sequencing in both the forward and reverse directions (Schuelke 2000). Sequencing and clean-up reactions were performed at SeqWright, Inc. (Houston, TX) on ABI PRISM 3730*xl* DNA sequencers (Applied Biosystems). DNA sequence chromatograms were assembled and edited in Lasergene SeqMan 7.0 (DNASTAR, Inc.), exported to BioEdit 7.0 (Hall 1999) and aligned using Clustal_X 1.8 (Higgins & Sharp 1988) (incorporated into the BioEdit software). For sequences with mixed bases, haplotypes were reconstructed using PHASE 2.1 (Stephens *et al*. 2001; Stephens & Donnelly 2003). Sequences were deposited in GenBank (Accessions numbers are provided in Table 1).

Analysis of DNA sequence data

Genetic variation and differentiation. Genetic variation within the rhesus macaques and within the cynomolgus macaques was measured based on the number of segregating sites θ_S (S) and expected site heterozygosity θ_π (π) (Hartl & Clark 2007) (Table 2). Genetic differentiation between the rhesus and cynomolgus macaques was expressed as F_{ST} and D_{XY} (Hartl & Clark 2007) (Table 3). All above measures were calculated in Arlequin 3.11 (Excoffier *et al*. 2005). To document the extent of shared variation, gene genealogies were estimated for each locus based on pairwise sequence comparisons (gaps were excluded) using the neighbour-joining method as implemented in MEGA 3.1 (Kumar *et al*. 2004). Support for each node in each haplotype tree was determined by 1000 bootstrap replicates (Table 3).

Selective neutrality. Selection violates the assumptions made by the tests implemented in the Isolation Migration-analytical (IMa) software (Hey & Nielsen 2007). To test for selection, Tajima's *D* (Tajima 1989) was calculated for each species separately and significance was assessed using Arlequin. The Ewens–Watterson–Slatkin exact method (Ewens 1972; Watterson 1978; Slatkin 1994) was used to calculate the expected homozygosity (*F*) in each species. Expected and observed *F* values were calculated in Arlequin and the difference of these values along with a *P* value were calculated according to Slatkin (1994) (Table 2). Significance was assessed using 1000 simulated samples. Finally, we noted whether the annotated or known function of genes, with particular attention paid to their possible role in immunity (cf. Table 2), could make these the target of balancing selection.

Divergence population genetics. The 'isolation with migration' model implemented by IMa attempts to fit the data to a null model where an ancestral population bifurcates into two strictly allopatric populations 1 and 2. It implements a Markov chain Monte Carlo (MCMC) simulation to estimate the joint posterior probability of six demographic parameters: population size of species 1 and 2 (N_{e1}, N_{e2}) and of the ancestral species (N_{eA}), migration (m_1, m_2), and divergence time (t). In our study, the subscript 1 refers to the rhesus macaque and the subscript 2 refers to the cynomolgus macaque, and the m_1 parameter refers to migration from the cynomolgus macaque into the rhesus macaque, whereas the m_2 parameter refers to migration from the rhesus macaque into the cynomolgus macaque. The model of molecular evolution employed was the infinite sites model (Kimura 1969), except for mitochondrial DNA sequences for which the Hasegawa–Kishino–Yano (HKY; Hasegawa *et al*. 1985) model was used. Inheritance scalars of 1 for the nuclear loci, 0.75 for the X-linked locus, and 0.25 for the Y-chromosomal and mitochondrial sequences were used. Input files were executed in IMa using 10 Metropolis-coupled chains with a 300 000-step burn-in followed by 50 million iterations of the Markov chain. Autocorrelation values, effective sample sizes (ESS), and inspection of parameter trend plots by eye indicated adequate convergence of the Markov chain (not shown).

Prior to analysis with IMa, sequences were processed to meet restrictions/assumptions. First, gaps in the alignments were removed (Hey & Nielsen 2004). Second, intragenic recombination for each gene and species was tested using the four-gamete test (Hudson & Kaplan 1985; Hudson 1985). For nuclear loci with evidence of intragenic recombination between alleles, the locus was divided into nonrecombining segments using the algorithm of Hudson & Kaplan (1985). We estimated genealogies for those segments of nuclear alleles that did not exhibit signs of recombination. We removed sites failing the four-gamete test in order for the IMa analysis to proceed because recombination can be ruled out in the mtDNA and Y-chromosomal sequences. To reduce potential bias in our choice of nonrecombining segments, we calculated nucleotide site polymorphism [the ratio of the number of segregating sites, S, to the total

Table 2. Analysis of genetic variation. For each gene, we report the aligned length and total number of polymorphisms (*S*). For each species, we indicated the number of haplotypes (2N), genetic diversity (π, θ_S), and the results of the tests of selective neutrality: Tajima's *D* and Ewens–Watterson–Slatkin (EWS). Statistical significance is indicated by *(< 0.05), **(< 0.01), or ***(< 0.001), all in bold.

							EWS		
Gene	Length (bp)	S (total)	Species	2*N*	π	θ_S	Tajma's *D*	$F_{obs\text{-}exp}$	Immunity-related
3kgg	640	16	*M. mulatta*	34	0.002	0.003	–0.587	0.060	Yes
			M. fascicularis	24	0.004	0.005	–0.125	–0.554	
abo	459	19	*M. mulatta*	14	0.003	0.003	0.184	**0.283***	
			M. fascicularis	16	0.014	0.010	1.289	**0.194****	
c4	3326	47	*M.mulatta*	8	0.003	0.003	1.485	–0.515	Yes
			M. fascicularis	20	0.002	0.003	–0.687	–0.560	
ccl2	529	8	*M. mulatta*	36	0.003	0.003	0.379	–0.069	Yes
			M. fascicularis	24	0.002	0.003	–0.538	–0.074	
ccr5	1019	42	*M. mulatta*	166	0.002	0.006	**–2.022****	–0.837	Yes
			M .fascicularis	14	0.003	0.003	–0.410	0.044	
dpb1	257	98	*M. mulatta*	34	0.071	0.065	0.367	**0.095*****	Yes
			M. fascicularis	110	0.103	0.072	1.420	**0.046*****	
irbp	1615	34	*M. mulatta*	8	0.003	0.002	1.250	–0.844	
			M. fascicularis	22	0.004	0.005	–0.766	–0.362	
mhca	1019	292	*M.mulatta*	68	0.048	0.038	0.867	**0.060*****	Yes
			M. fascicularis	102	0.053	0.049	0.229	**0.034*****	
mhcg	236	24	*M.mulatta*	16	0.005	0.004	–0.740	0.057	Yes
			M. fascicularis	16	0.003	0.003	0.513	**0.214****	
mtdna	1478	141	*M.mulatta*	20	0.012	0.011	0.471	–0.165	
			M. fascicularis	31	0.015	0.015	–0.086	–0.787	
nbpf	1169	240	*M. mulatta*	20	0.050	0.050	0.000	**0.155*****	
			M. fascicularis	20	0.079	0.048	2.640	**0.156*****	
nramp1	904	11	*M. mulatta*	98	0.001	0.002	–0.823	–0.137	Yes
			M. fascicularis	8	0.002	0.001	1.474	0.281	
nry	3018	12	*M. mulatta*	5	0.000	0.000	1.459	0.206	
			M. fascicularis	18	0.002	0.001	2.221	0.235	
r1c2	416	9	*M. mulatta*	36	0.001	0.002	–0.810	–0.022	Yes
			M. fascicularis	24	0.005	0.005	0.282	–0.323	
afp	743	10	*M.mulatta*	20	0.016	0.019	1.078	0.326	
			M. fascicularis	26	0.030	0.027	–0.736	–0.613	
apoe	455	13	*M. mulatta*	20	0.001	0.001	1.399	**0.538***	
			M. fascicularis	22	0.003	0.004	–1.214	–0.480	
b2m	835	6	*M.mulatta*	26	0.002	0.001	1.924	–0.079	Yes
			M. fascicularis	26	0.001	0.001	0.600	0.280	
slc25a5	724	2	*M.mulatta*	18	0.000	0.001	–1.508	–0.445	
			M. fascicularis	9	0.001	0.001	0.986	0.315	
ttr	820	17	*M. mulatta*	26	0.001	0.001	–0.600	–0.191	
			M. fascicularis	28	0.003	0.005	–1.354	–0.838	

number of sites, L (Hartl & Clark 2007, p. 176)] for the whole gene and for each of the nonrecombining segments. We selected the nonrecombining segment of each locus with the nucleotide site polymorphism value closest to the whole locus alignment. Third, loci expected to be in linkage disequilibrium were concatenated into a single locus to meet the assumption of free recombination between loci. For example, the locus *nry* is a product of concatenation of two loci, *sry* and *tspy* on the Y-chromosome, and the *NRAMP1* locus is a product of the concatenation of the introns of the *NRAMP1* gene. Finally, IMa assumes selective neutrality. As an auxiliary aspect of our study, we liberally interpreted the results of neutrality tests, and for one of our analyses, we excluded genes for which there was any evidence for selection, however tentative.

Four separate IMa analyses were conducted that differed from one another with respect to the question addressed and the genes and macaque samples included. In Table 1, we indicated for each gene whether these were included in any of the analyses I–IV.

I. The full data set consisted of 673 rhesus and 560 cynomolgus macaque haplotypes and 19 genes.
II. A subset including only cynomolgus macaque samples of known geographic origin (with a subsequent exclusion of mainland Indochinese cynomolgus samples from the shaded area in Fig. 1) was analysed to determine if introgression extends beyond the Isthmus of Kra (Fig. 1). This analysis depended on the availability of information on the geographic origin of wild–derived macaque samples, and the inference of geographic origin of macaque samples obtained from primate research centres based on mitochondrial DNA analysis (cf. Table S1) (Stevison & Kohn 2008). This analysis considered 151 rhesus and 127 (202 prior to removal of Indochina haplotypes) cynomolgus macaque haplotypes and nine genes.
III. A subset of genes excluding those that might violate selective neutrality, consisting of 215 rhesus, and 164 cynomolgus macaque haplotypes and eight genes.
IV. A subset including only wild-caught macaque samples was included in the analysis to explore whether the level of estimated interspecific gene flow differed in the wild relative to captive stocks. This analysis consisted of 26 rhesus and 77 cynomolgus macaque haplotypes and four genes.

The values resulting from the IMa test were converted to values of effective population size (N_e), population migration rate ($2N_e$m), and divergence time, t (in years). These conversions were made using the geometric mean per gene mutation rate per year estimated for each locus (3.5×10^{-7}) using *Homo sapiens* as an outgroup (Hernandez *et al*. 2007) (Table 4). The average per site mutation rate was calculated as 1.1×10^{-9}, which is close to the per-site genome-wide mutation rate recently reported using the macaque genome, 5.9×10^{-9}. We used 25 million years as the divergence time between human and macaque, and a generation time of 6.5 years for macaques (Gibbs *et al*. 2007; Hernandez *et al*. 2007).

Log-likelihood ratio (2LLR) tests were performed to test for the statistical significance of various nested models implemented in the IMa software. In essence, these models describe cases where maximum-likelihood estimates of migration parameters, population size parameters, and combinations thereof, are fixed at the value of zero, are set as equal to other parameters, or are allowed to vary. The 2LLR values should be approximately chi-squared distributed, and thus, the associated probability values were calculated based on this distribution as recommended (Hey & Nielsen 2007). However, for nested models where one of the migration parameters was set at zero, the 2LLR was approximately 50% chi-square distributed with 1 degree of freedom (Hey & Nielsen 2007). This approximation should be sufficient for the data sets with 6–25 loci (I–III). However, for small data sets, such as those used for analysis IV, the test is liberal, and thus, P values for these analyses tend to erroneously reject the null model in favour of erroneously accepting the alternative hypothesis (see Results and Discussion).

Table 3 Analysis of genetic differentiation. Here we report the results of the species differentiation as measured by F_{ST}, species divergence as measured by D_{XY} and the results of our phylogenetic analysis with the topology for each species. Statistical significance is indicated by *(< 0.05), **(< 0.01), or ***(< 0.001).

Gene	F_{ST}	D_{xy}	Species	Gene tree topology
3kgg	0.44***	0.006***	*M. mulatta*	Monophyletic-69†
			M. fascicularis	Paraphyletic
abo	0.45***	0.016***	*M. mulatta*	Paraphyletic
			M. fascicularis	Paraphyletic
c4	0.10	0.003*	*M. mulatta*	Paraphyletic
			M. fascicularis	Paraphyletic
ccl2	0.22***	0.003***	*M.mulatta*	Paraphyletic
			M. fascicularis	Paraphyletic
ccr5	0.23***	0.003***	*M. mulatta*	Paraphyletic
			M. fascicularis	Paraphyletic
dpb1	0.03***	0.092	*M. mulatta*	Paraphyletic
			M. fascicularis	Paraphyletic
irbp	0.24***	0.005**	*M. mulatta*	Monophyletic-54
			M. fascicularis	Paraphyletic
mhca	0.05***	0.053***	*M. mulatta*	Paraphyletic
			M. fascicularis	Paraphyletic
mhcg	0.12***	0.026***	*M. mulatta*	Monophyletic-62
			M. fascicularis	Paraphyletic
mtdna	0.63***	0.04***	*M. mulatta*	Monophyletic-99
			M. fascicularis	Monophyletic-98
nbpf	0.16***	0.081***	*M.mulatta*	Paraphyletic
			M. fascicularis	Paraphyletic
nramp1	0.35***	0.002***	*M. mulatta*	Paraphyletic
			M. fascicularis	Paraphyletic
nry	0.36*	0.003	*M. mulatta*	Monophyletic-98
			M. fascicularis	Paraphyletic
r1c2	0.67***	0.008***	*M. mulatta*	Monophyletic-96
			M. fascicularis	Paraphyletic
afp	0.35***	0.004***	*M. mulatta*	Paraphyletic
			M. fascicularis	Paraphyletic
apoe	0.25***	0.004***	*M. mulatta*	Paraphyletic
			M. fascicularis	Paraphyletic
b2m	0.59***	0.004***	*M.mulatta*	Paraphyletic
			M. fascicularis	Monophyletic-51
slc25a5	0.15	0.001	*M.mulatta*	Paraphyletic
			M. fascicularis	Paraphyletic
ttr	0.08***	0.002	*M.mulatta*	Monophyletic-27
			M. fascicularis	Paraphyletic

†for monophyletic topologies, bootstrap support is shown as assessed using 1000 replicates.

Table 4 Isolation model fitting. IMa estimates of demographic parameters: theta, population size, migration estimates ($2N_1m_1$ is migration of alleles from the cynomolgus monkey into the rhesus monkey and $2N_2m_2$ is migration from the rhesus monkey into the cynomolgus monkey), and divergence time, t (in years), using different subsets of the data in Table 1.

n*	μ†	θ_1‡	N_{e1}	θ_2	N_{e2}	θ_A	N_{eA}	$2N_1m_1$§	$2N_2m_2$	t (years)
Full data set (I)										
19	3.49E-07	1.025¶	113 000	1.7058	188 000	1.351	149 000	0.120	1.214	1 295 000
		0.75–1.53**	82 000–169 000	1.24–2.743	137 000–303 000	0.478–21.303	53 000–2 351 000	0.019–0.477	0.37–2.012	977 000–8 430 000
Exclusion and inclusion (in parentheses) of samples from putative hybrid zone (II)										
9	1.00E-06	0.914 (0.86††)	35 000 (33 000)	2.03 (2.5)	78 000 (97 000)	1.55 (1.8)	60 000 (69 000)	0.102 (0.078)	0.023 (0.48)	598 000 (445 000)
		0.53–1.68	20 500–65 000	1.27–3.95	49 000–152 000	0.15–19.71	5900–758 000	0.02–0.74	0.01–0.96	376 000–2 880 000
Exclusion of loci violating neutrality (III)										
8	9.48E-07	1.18	287 000	3.03	735 000	0.088	21 000	0.052	0.493	785 000
		0.67–2.4	164 000–571 000	1.73–5.8	420 000–1 400 000	0.088–30.08	21 000–7 300 000	0.005–1.21	0.051–5.45	400 000–2 850 000
Exclusion of captive macaques (IV)										
4	2.16E-06	4.16	74 000	8.1	144 000	8.82	157 000	0.002	0.249	798 000
		1.59–9.02	28 000–161 000	4.13–14.29	74 000–254 000	1.25–31.93	22 000–569 000	0.002–2.79	0.04–3.1	274 000–1 362 000

*number of genes included in the isolation model.
†an estimate of the mutation rate per year *per gene* as estimated from the data using human as an outgroup.
‡population 1 is defined as the rhesus macaque, and therefore, pop 2 is defined as the cynomolgus macaque.
§values plotted in Fig. 2.
¶high point of posterior distribution (mode), values rounded to nearest thousand individuals and years.
**95% credible interval for the estimated value, values rounded to nearest thousand individuals and years.
††estimates when same data set is used, and samples from Indochina are included.

Results

Genetic variation and differentiation

The sequencing of five genes yielded 110 haplotypes of the rhesus macaque and 111 haplotypes of the cynomolgus macaque (Table 2). In addition, the compilation of 14 genes from GenBank yielded 563 haplotypes of the rhesus macaque and 449 haplotypes of the cynomolgus macaque. In total, this analysis of 19.7 kb of DNA sequence included 673 rhesus haplotypes and 560 cynomolgus haplotypes defined by 1028 segregating sites (Table 2).

Levels of genetic polymorphisms as measured by π in the rhesus and cynomolgus macaques ranged from 0 to 0.103. Genetic polymorphism measured as θ ranged from 0 to 0.072 (Table 2). Average π in the rhesus macaque was 0.012 (SD 0.021) and in the cynomolgus macaque 0.017 (SD 0.029). Average θ in the rhesus macaque was 0.011 (SD 0.019) and in the cynomolgus macaque 0.014 (SD 0.02).

Values of Tajima's *D* computed for each locus and species revealed significant deviations from neutrality for the *ccr5* locus; however, this deviation was consistent with a selective sweep rather than balancing selection (Table 2). The Ewens–Watterson–Slatkin homozygosity test identified six loci (*nbpf, dpb1, mhca, mhcg, abo,* and *apoe*) with lower-than- expected homozygosity ($P < 0.05$), which is consistent with balancing selection (Table 2). Finally, inspection of the annotated functions of genes identified 10 loci that are known to mediate some kind of immune responses (*3kgg*, *b2m*, *c4*, *ccl2*, *ccr5*, *dpb1*, *mhca*, *mhcg*, *nramp1*, *r1c2*) and may therefore be suspected or known (*abo*, *ccr5*, *mhca*, *mhcg*) (Garrigan & Hedrick 2003) to be under balancing selection (Chen *et al*. 1997; Marcon *et al*. 1997; Doxiadis *et al*. 1998, 2006; Deinard *et al*. 2002; Weiler *et al*. 2006). Overall, selection, although notoriously difficult to establish conclusively, might have been a factor in the evolution of as many as 11 loci (Table 2). These loci were removed during one of four of our analyses (analysis III, see below) to examine whether inclusion of these loci resulted in erroneous conclusions.

Genetic differentiation between the rhesus and cynomolgus macaques as measured by F_{ST} ranged between 0.05 and 0.67, with 17 of the 19 loci having significant differentiation (Table 3). Differentiation as measured by D_{XY} between the two species was significant for 15 of the 19 loci (Table 3). Despite these considerable levels of genetic differentiation, 174 shared genetic polymorphisms were observed in the data. In contrast, only 15 fixed polymorphisms were observed. Analysis of gene genealogies also revealed shared nuclear and Y-chromosomal polymorphisms, while the mtDNA haplotype tree was reciprocally monophyletic (Table 3) (Melnick & Hoelzer 1992; Morales & Melnick 1998; Tosi *et al*. 2002; Blancher *et al*. 2008; Bonhomme *et al*. 2008; Stevison & Kohn 2008). Overall, each species was found to be paraphyletic at 11 nuclear genes (*abo, nbpf, ccl2, c4, mhca, apoe, slc25a5, nramp1, dpb1, ccr5, afp*) (Table 3). Only the *b2m* gene showed the cynomolgus macaque as monophyletic and the rhesus macaque as paraphyletic. Six loci (*mhcg, 3kgg, r1c2, nry, ttr*, and *irbp*) showed the rhesus macaque as monophyletic and the cynomolgus macaque as paraphyletic. Included in this was the *nry* gene, which had been reported previously as having a genealogy where cynomolgus macaques were reported as more closely related to rhesus macaques than to members of their own species (Tosi *et al*. 2000, 2002, 2003a).

Divergence population genetics

Data were analysed in four separate configurations I–IV (see methods). The results of isolation model fitting from each analysis are summarized in Table 4, with the posterior distribution of the migration estimates from each analysis depicted graphically in Fig. 2. Below, we present the results obtained from each analysis of the four configurations of samples and genes with respect to parameter estimates and significance (Tables 4 and 5, respectively).

We performed log-likelihood ratio tests on 16 different nested models implemented in IMa. For each configuration of the data, we were interested in five nested models crucial to evaluate the main hypotheses of this study: (i) the model where the two migration parameters are set as equal while all other model parameters are free to vary ($\theta_1\theta_2\theta_A m_1 = m_2$), (ii and iiii) models where each migration parameter is set to zero while the other model parameters are free to vary ($\theta_1\theta_2\theta_A m_1 m_2 = 0$ & ($\theta_1\theta_2\theta_A m_1 = 0 m_2$), (iv) the model where the two species population sizes are set as equal while all other model parameters are free to vary ($\theta_1 = \theta_2\theta_A m_1 m_2$), and (v) the model where the two species' population sizes and the population of their ancestor are set as equal while the other model parameters are free to vary ($\theta_1 = \theta_2 = \theta_A m_1 m_2$). These five select models are summarized in Table 5 for each configuration of the data used to conduct analyses I–IV. The results reported below from IMa are for the full model, even when nested models could not be rejected.

(I) Full data set. The extant population sizes of the rhesus macaque (China and India combined) and of the cynomolgus macaque were estimated as $N_{e1} = 113\,000$ and $N_{e2} = 188\,000$, respectively (Table 4). The effective ancestral population was estimated as $N_{eA} = 149\,000$. Gene flow from the rhesus macaque to the cynomolgus macaque had a broad non-zero distribution with peak at $2N_2m_2 = 1.214$; the reverse direction was non-zero and one order of magnitude lower ($2N_1m_1 = 0.120$, Fig. 2A). The divergence time between the two species was estimated as 1.3 million years. The 95% credible intervals surrounding estimations were large (Table 4), but the probability distributions had well-defined narrow peaks.

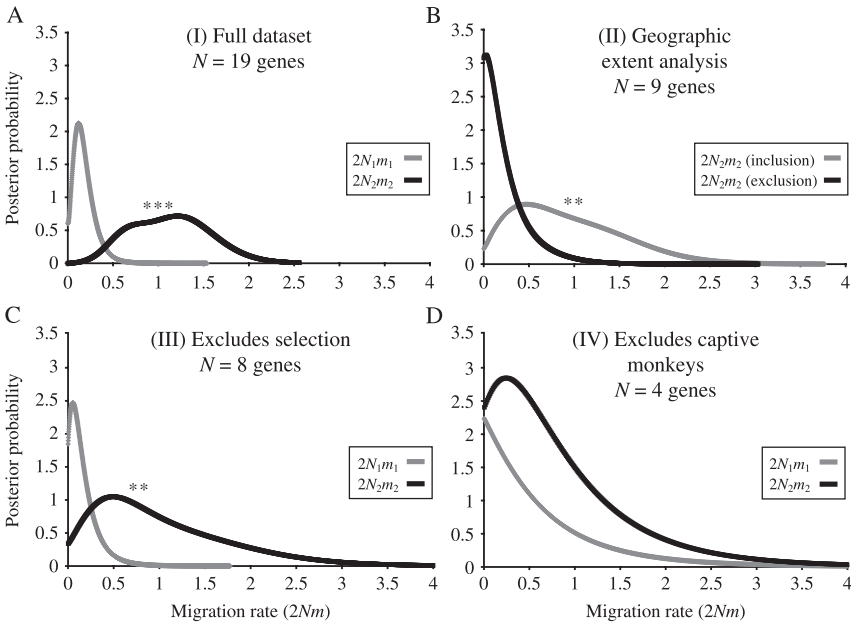

Fig. 2 Results depicting maximum-likelihood estimations of population migration parameters of analysis I–IV. Shown are results for 2Nm for all 19 genes analysed in this study (A), when only samples of known geographic origin excluding and including cynomolgus samples from the putative area of hybridization were used (B), a conservative subset of loci least likely to be influenced by balancing selection (C), and a subset of loci with samples of wild origin only (D). Statistical significance as assessed by log-likelihood ratio tests is indicated by *(< 0.05), **(< 0.01), or ***(< 0.001).

Table 5 Nested model tests. Results of log-likelihood ratio tests of select nested models. Statistical significance is indicated by *(< 0.05), **(< 0.01), or ***(< 0.001), all in bold.

Model	log(*P*)*	d.f.†	2LLR‡	*P* value§
Full data set (I)				
$\theta_1\theta_2\theta_A m_1$¶ = m_2**	0.2194	1	1.6974	0.19
$\theta_1\theta_2\theta_A m_1 m_2 = 0$††	–14.0137	1	30.1637	**1.99E-08*** **
$\theta_1\theta_2\theta_A m_1 = 0$ m_2††	0.5821	1	0.9721	0.16
$\theta_1 = \theta_2\theta_A m_1 m_2$	–1.212	1	4.5603	**3.3E-02***
$\theta_1 = \theta_2 = \theta_A m_1 m_2$	–1.4118	2	4.9599	8.40E-02
Exclusion and inclusion (in parentheses) of samples from putative hybrid zone (II)				
$\theta_1\theta_2\theta_A m_1 = m_2$	0.2488 (–0.9615)	1	1.437 (4.3321)	0.23 (**3.7E-02***)
$\theta_1\theta_2\theta_A m_1 m_2 = 0$	0.7344 (–2.4337)	1	0.4659 (7.2766)	0.25 (**3.5E-03****)
$\theta_1\theta_2\theta_A m_1 = 0$ m_2	–0.2195 (–1.4035)	1	2.3737 (5.2161)	0.06 (**1.1E-02***)
$\theta_1 = \theta_2\theta_A m_1 m_2$	–2.5991 (–3.4635)	1	7.1329 (9.3361)	**7.6E-03** (2.3E-03****)
$\theta_1 = \theta_2 = \theta_A m_1 m_2$	–3.4501 (–7.5425)	2	8.8348 (17.4941)	**1.2E-02* (1.6E-04***)**
Exclusion of loci violating neutrality (III)				
$\theta_1\theta_2\theta_A m_1 = m_2$	0.2959	1	3.9044	**4.8E-02***
$\theta_1\theta_2\theta_A m_1 m_2 = 0$	–1.4505	1	7.3972	**3.3E-03****
$\theta_1\theta_2\theta_A m_1 = 0$ m_2	2.2462	1	0.0039	0.48
$\theta_1 = \theta_2\theta_A m_1 m_2$	1.031	1	2.4342	0.12
$\theta_1 = \theta_2 = \theta_A m_1 m_2$	–6.0817	2	16.6596	**2.41E-04*****
Exclusion of captive macaques (IV)				
$\theta_1\theta_2\theta_A m_1 = m_2$	–3.5286	1	0.2413	0.62
$\theta_1\theta_2\theta_A m_1 m_2 = 0$	–3.7862	1	0.7565	0.19
$\theta_1\theta_2\theta_A m_1 = 0$ m_2	–3.4082	1	0.0005	0.49
$\theta_1 = \theta_2\theta_A m_1 m_2$	–4.7219	1	2.628	0.11
$\theta_1 = \theta_2 = \theta_A m_1 m_2$	–4.7933	2	2.7707	0.25

*log probability value result from IMa.
†degrees of freedom for the model.
‡log-likelihood ratio result from IMa; approximates a chi-square distribution.
§probability of model as assessed using a chi-square distribution.
¶m_1 is the migration value from cynomolgus macaque into rhesus macaque.
**m_2 is the migration value from rhesus macaque into cynomolgus macaque.
††for models where migration estimates are set to zero, 2LLR value approximates 1/2chi-square(1) distribution.

Log-likelihood ratio tests rejected significance of two of the select models and were unable to reject three others (Table 5). We were unable to reject the nested model where the two migration parameters are equal ($P = 0.19$). However, we were able to reject the model where the migration parameter representing gene flow from the rhesus macaque to the cynomolgus macaque, m_2, was set to zero ($P = 1.99 \times 10^{-8}$, Table 5), but we were unable to reject the model where the migration parameter representing gene flow from the cynomolgus macaque to the rhesus macaque, m_1, was set to zero (i.e. no gene flow) ($P = 0.16$). Thus, this analysis of the combined data set indicated that gene flow might be asymmetrical, from the rhesus to the cynomolgus macaque. The second model that was rejected assumes equal population sizes for the two species and their ancestor ($P = 0.033$). We were unable to reject the nested model where the two species' population sizes are equal ($P = 0.084$).

(II) Geographic extent of hybrid zone. Hybridization between rhesus and cynomolgus macaques has been postulated to be restricted to the mainland Indochinese populations of cynomolgus macaques found north of the Isthmus of Kra (Fooden 1964; Tosi *et al*. 2002) (cf. Fig. 1). Thus, we would expect that the inclusion of samples derived from the putative hybrid zone in Indochina would increase estimates of migration (in parentheses in Tables 4 and 5). This analysis included only samples of known or inferred geographic origin (cf. Table 1 and Table S1). We observed that prior to the inclusion of the Indochinese samples, migration estimates were $2N_1m_1 = 0.102$ and $2N_2m_2 = 0.023$. However, the inclusion of samples derived from the proposed area of hybridization elevated $2N_2m_2$ to a value of 0.48, the estimation for the reverse gene flow remained low ($2N_1m_1 = 0.078$).

Exclusion of samples of the putative area of hybridization altered log-likelihood ratio tests in two important ways. First, in the absence of samples derived from the putative area of hybridization, we failed to reject the model where the migration parameter describing gene flow from the rhesus macaque to the cynomolgus macaque is set to zero ($m_2 = 0$, $P = 0.25$). In addition, we rejected models where population sizes are set as equal both for all three populations and for just the two species' population sizes ($P = 7.6 \times 10^{-3}$). Inclusion of samples from the putative hybrid zone enabled us to reject all models (Table 5), thereby supporting the hypothesis that gene flow between the two species is non-zero in the putative area of hybridization, but gene flow beyond the putative zone of hybridization in mainland Indochina is zero.

This analysis yielded an additionally unexpected, yet very important observation. Because this analysis included loci with evidence of balancing selection and individuals from captivity, this analysis is additionally able to exclude the interaction of these factors as contributing to the significantly non-zero estimates of gene flow in this study. As it is unlikely that either balancing selection or hybridization in captivity should affect some individuals differently than others, either of these factors or the interaction of these factors *could not have been responsible* for the significantly non-zero estimate of gene flow, or the exclusion of some haplotypes would not have revealed a zero estimate of gene flow.

(III) Influence of selection. Based on our results of significance testing for selection and potential involvement in immunity (Table 2), we removed the loci *3kgg*, *abo*, *b2m*, *c4*, *ccl2*, *ccr5*, *dpb1*, *mhca*, *mhcg*, *nbpf*, and *r1c2* to examine whether selection might have erroneously lead to the rejection of the null model of no gene flow between the two species ($m_1 = m_2 = 0$). Gene flow from the rhesus macaque to the cynomolgus macaque had a broad non-zero distribution with peak at $2N_2m_2 = 0.493$; the reverse direction was one order of magnitude lower ($2N_1m_1 = 0.052$). The 95% credible intervals surrounding estimations were large (Table 4). Differences of results when compared to analysis I may indicate a role of balancing selection and/or differences in the power of analysis due to the lower number of loci included. However, all estimations had 95% credible intervals that overlapped between analyses I and III (Table 4), indicating that the inclusion of genes that may be prone to selection has not led us to infer interspecific gene flow during analysis I. However, the reduction of the maximum-likelihood estimate for $2N_2m_2$ from 1.214 in analysis I to 0.493 in analysis III indicates that the inclusion of immunity-related genes, which may be under balancing selection, might have inflated the estimated interspecific gene flow.

Despite the reduced size of the data set, we were able to reject a model where the migration parameter representing gene flow from the rhesus macaque to the cynomolgus macaque is set to zero ($m_2 = 0$) ($P = 3.3 \times 10^{-3}$) and not the model assuming gene flow from the cynomolgus macaque to the rhesus macaque ($P = 0.48$). Qualitatively, however, balancing selection does not explain the significant pattern of asymmetrical gene flow based on the results of the log-likelihood ratio tests of nested models. This analysis yielded additional power to reject the model where $2N_1m_1 = 2N_2m_2$ ($P = 0.048$). However, this test was not significant after correction for multiple tests.

(IV) Influence of captive macaques. We conducted an analysis of only wild-caught samples to assess the potential effect of crossbreeding between species in captive stocks or stocks used to found these. Such crossbreeding could inflate our estimates for interspecific gene flow. Analysis of a subset of four loci (*c4*, *irbp*, *mtDNA* and *nry*) avoiding this potential source for error yielded estimates of gene flow from the rhesus macaque to the cynomolgus macaque with a peak at $2N_2m_2 = 0.249$; the reverse direction was two orders of

magnitude lower ($2N_1m_1 = 0.002$, Fig. 2C). These estimates reduced relative to analysis I were in the expected direction when captive breeding may have included hybridization events. Although, the credible intervals between analyses I and IV overlapped, indicating that the overall conclusions of the study are not erroneous owing to interbreeding in captive stocks. We found evidence indicative of lower levels of gene flow in the wild compared to gene flow in some of the founding troops used to establish captive macaque colonies. However, the LLR tests did not reject any of the nested models (Table 5), but the power underlying the tests based on this small data set was expected to be low (less than 10% the size of data set I). Another concern with this reduced data set was the high parameter estimate for the mutation rate (Table 4). Although convergence of the MCMC takes longer with larger data sets, smaller data sets often have the opposite problem of not having enough information to converge on a particular value. Inspection of MCMC convergence indicated that this analysis did not converge well for the m_2 parameter, which could also account for the reduced power to significantly reject the nested model. However, as noted above, analysis II was able to account for captivity independently of this analysis and exclude it as a potential source of gene flow in this study.

Discussion

The use of analytical approaches rooted in the contemporary framework of divergence population genetics has highlighted the importance of considering the complex interplay of demographic parameters when testing a null model of allopatric speciation with no subsequent gene flow against more complex models. Complex models include parapatric or sympatric speciation, or conceivably, scenarios that include recent or ongoing hybridization or bouts of reticulate evolution in the past. Further, various multiple-locus approaches, surveying a broad range of genes and gene functions, allow researchers to examine how selection against hybridization results in lowered introgression in some parts of the genome relative to neutral loci (Machado *et al*. 2002; Hey & Nielsen 2004; Borge *et al*. 2005; Geraldes *et al*. 2006; Kronforst *et al*. 2006; Lecis *et al*. 2006; Munoz-Fuentes *et al*. 2007). Research in hybridizing species, focusing on sets of loci that display restricted gene flow across species boundaries, can identify genes important in the evolution and maintenance of reproductive barriers. However, this potential of studies of natural hybrid zones to reveal the influence of selection on the introgression rates of genes has not been realized in primates yet. Our study now further supports that there is a primate system available for such studies. In the following sections, we present the evidence for hybridization and genetic introgression between the rhesus and cynomolgus macaques in an area of mainland Indochina, and we discuss future avenues of investigation.

Hybridization between the rhesus and cynomolgus macaques

The concomitant estimation of numerous demographic parameters is important when attempting to reconstruct speciation histories. Here we provided significant evidence for the introgression of nuclear genes between rhesus and cynomolgus macaques, and that gene flow between the two macaque species is highest in mainland Indochina. Gene flow appears to be asymmetric, with higher rates of introgression from the rhesus macaque to the cynomolgus macaque. Gene flow in the reverse direction is not significantly different from zero in two of the four analyses. The only instance where a model assuming genetic isolation between the two species was supported was when the samples derived from the putative area of hybridization in mainland Indochina (Fig. 1) were removed from the analysis. In all other data sets, gene flow from the rhesus macaque to the cynomolgus macaque (m_2) was non-zero, but during analysis IV, this could not be statistically supported due to low power and lack of convergence of Markov chains (see below).

Our results were consistent with those of previous studies, extending these to the nuclear genome, and putting these in perspective with other demographic parameters. Our estimation of gene flow from the rhesus into the cynomolgus macaque with a broad peak at $2N_2m_2 = 1.214$ rejected the nested model where $2N_2m_2 = 0$ (Fig. 2A, $P = 1.99 \times 10^{-8}$). The broad peak reflecting a large variance of Nm on a gene-by-gene basis indicated variations in the probability of introgression of genes, which is subject to stochastic factors as well as purifying (not positive or balancing) selection against genes that have a detrimental effect on the fitness of hybrids. With respect to the latter assumption, it is interesting to view our results in comparison to the results obtained from the IM analysis of microsatellite data by Bonhomme *et al*. (2008). Population migration estimates from the cynomolgus macaque to the rhesus macaque for microsatellites were only 0.0615, which is compatible with our estimate based on genes, where we observed a 95% credible interval of $Nm = 0.019$–0.477. In contrast to the results obtained from microsatellites where Nm from the rhesus to the cynomolgus macaque was estimated as 10.59, we observed a much smaller credible interval for Nm of only 0.37–2.012, which corresponds to an 80–96.5% reduced probability of successful introgression compared to presumably selectively neutral and nonfunctional microsatellites.

The introgression of the rhesus macaque into the cynomolgus macaque was not limited to the Y-chromosome, as exclusion of the nonrecombining Y-chromosome yielded results that did not differ (as judged by overlap of 95% credible intervals) from the whole data set ($2N_2m_2 = 0.95$; data not shown). Nevertheless, during a gene-by-gene analysis, the Y-chromosome displayed migration rates

that were at the high end of the 95% credible interval ($2N_2m_2 = 2.13$; data not shown).

The isolation model implemented in IMa assumes no population structure. The effects of violations to this assumption have not been explored in detail so far. However, violation of this assumption does not seem to preclude the power of the approach to accept or reject the null model, as has been shown, for example, during an analysis of chimpanzee subspecies and bonobo. IMa test accepted the null model of genetic isolation between chimp and bonobo despite the fact that the chimpanzee subspecies included in the analyses were genetically subdivided (Won & Hey 2005). Conversely, the null model was rejected in other cases where there was population subdivision. For example, the null model was rejected for three pairs of *Heliconius* butterfly species where two of the three species displayed within-species subdivision (*H. melpomene* and *H. cydno*) while the third species (*H. pachinus*) did not (Kronforst *et al*. 2006). In sum, systematic bias of IMa results due to the inclusion of subdivided populations has yet to be reported. We also note that our result from analysis II should not be affected by population subdivision in the cynomolgus macaque in that we erroneously infer gene flow from rhesus to cynomolgus macaques.

The observation of hybridization between the two macaques is in agreement with earlier studies (Fooden 1964; Tosi *et al*. 2002, 2003b). Many researchers have argued for a broad role of reticulate evolution and/or recent hybridization in the primates (Arnold & Meyer 2006), or have documented such cases (Fooden 1964; Bernstein 1966; Blouch & Groves 1990; Watanabe *et al*. 1991; Bynum *et al*. 1997; Tosi *et al*. 2002, 2003b; Schillaci *et al*. 2005). Our main goal was to establish hybridization for our system in particular, because the rhesus macaque and cynomolgus macaque are the most widely held species in primate research centres and genomic resources and toolkits for both are rapidly growing (Abegg & Thierry 2002; Magness *et al*. 2005; Gibbs *et al*. 2007; Hernandez *et al*. 2007; Osada *et al*. 2008). It has become evident that there are numerous pitfalls when arguing for hybridization between recently split pairs of species, foremost owing to the sharing of ancestral polymorphisms. Thus, our objective was to rule out this specific source for error that would erroneously lead us to conclude that hybridization between the rhesus and cynomolgus macaque occurs. As is discussed below, we also attempted to rule out other issues that would erroneously lead us to conclude that the two species hybridize. However, by establishing this case more conclusively, this study lends support to previous reports on the prevalence of between-species hybridization in the macaque genus, and could render the rhesus–cynomolgus macaque species pair into an important study system to help understand reticulate evolution in the primate branch as a whole.

Demographic parameters

Some of our estimates for demographic parameters from the analysis of the full data set should be of interest, in particular for the less-well studied cynomolgus macaque. We estimated a population size for the cynomolgus macaque of $N_{e2} = 188\,000$ (95% CI: 137 000–303 000; consult Table 4 for additional CI values), which exceeded the estimate obtained for the rhesus macaque ($N_{e1} = 113\,000$). The cynomolgus macaque also harbours more genetic variation than the rhesus macaque ($\theta = 1.24$–2.74 vs. 0.75–1.53) (see also Smith *et al*. 2007). The effective ancestral population size was estimated as $N_{eA} = 149\,000$, with a large 95% credible interval (53 000–2.3 million). The large variance in the ancestral population size is consistent with the recent divergence between species, i.e. a prominent stochastic effect of between-locus variations in lineage sorting of ancestral polymorphisms. Moreover, this large variance is consistent with a model of divergence with gene flow, which would be in violation with a strict allopatric speciation model (Osada *et al*. 2008). This variance, however, is also consistent with a variety of other scenarios. For example, genetic subdivision of the ancestral population, or varying rates of gene flow between the recently diverged species over time, both might result in a larger-than-expected variance of the ancestral N_e. Of course, the result might also reflect limitations of the data itself, resulting in a failure to converge to a fine estimate for ancestral population size. Interestingly, if only the rhesus macaque is studied, then the ancestral population is estimated as 73 070 (Hernandez *et al*. 2007). While not outside of our 95% credible interval, this lower-bound estimate can be explained by the asymmetry of introgression, where the rhesus macaque hardly receives (or retains) any alleles from the cynomolgus macaque. Thus, estimates of ancestral population size based on rhesus macaque alone do not appear to be inflated to the same extent when compared to estimates including both the rhesus macaque and the cynomolgus macaque, because the latter carries alleles of the rhesus macaque (cf. also Osada *et al*. 2008).

The divergence time between the two species was estimated as 1.3 million years (95% credible interval of 977 000–8.43 million years; Table 4). This estimate was smaller but consistent (as judged by the 95% CI) with previous estimates of divergence between these two species inferred from mtDNA as 1.83 to 2.5 million years (Morales & Melnick 1998; Hernandez *et al*. 2007). Recent simulations on the basis of the inferred demographic history for Indian and Chinese rhesus macaques placed their ancestral population at 1.94 million years ago, and this split between rhesus macaques has occurred close in time to the split of rhesus macaque and cynomolgus macaque (Hernandez *et al*. 2007). As described by Osada *et al*. (2008), a large variance component in the estimation of divergence

once ancestral polymorphism was considered indicated that the lower bound of such estimates are more reflective of the actual time since speciation.

Asymmetry of introgression

We observed asymmetrical gene flow from the rhesus macaque genome into the genome of the cynomolgus macaque. Postzygotic barriers tend to have an asymmetrical pattern which result in asymmetric genetic exchange by limiting gene flow in one direction (Coyne & Orr 1989; Tiffin *et al*. 2001; Takami *et al*. 2007; Lowry *et al*. 2008). Thus, our observations predict an asymmetric pattern of postzygotic isolation in these species based on the asymmetrical introgression. It appears to be difficult to conclusively reject the presence of any gene flow from the cynomolgus monkey into the rhesus monkey genome. However, we found good evidence for this interesting biological facet of the system — introgression is biased strongly in the direction of males of the larger rhesus macaque into the females of the smaller cynomolgus macaque (Fooden 1964; Tosi *et al*. 2002). Two scenarios could facilitate this. Either the smaller cynomolgus female prefers the larger rhesus males, or, the larger rhesus males can coerce the smaller cynomolgus females easier than their conspecifics. Body size appears be an important factor to asymmetry of introgression in macaques, as deduced from the hybrid zone between *Macaca tonkeana* and *M. maura* where gene flow from the larger (*M. tonkeana*) to the smaller species (*M. maura*) occurs (Evans *et al*. 2001). It is unclear, however, what other factors may contribute to asymmetrical gene flow observed in other biological systems and whether those factors are operating in this species pair.

Studies on captive colonies reported on more aggressive cynomolgus male macaques isolated from the region of putative introgression compared to their conspecifics from outside of the region of introgression (Brent & Veira 2002). In the presence of rhesus macaque males, female preference may have shifted for higher male aggression, even for the conspecific male cynomolgus macaques (Bernstein & Gordon 1979). Observations made on mixed groups of captive macaques revealed that, when joining a group, heterospecific males encounter less aggression by resident males compared to conspecific males. If this were true also in the wild, then this higher aggression between cynomolgus males could facilitate access for rhesus males, which encounter less aggression by other male troop members, to cynomolgus females.

Our study should help define specific research directions that address this aspect of the biology of the study system. First, field studies would be necessary to specifically investigate how rhesus males gain access to the cynomolgus females, and whether this is facilitated by body size, behaviour, the social structure of macaque societies, or combinations of these factors. Such studies could also be done in captive research colonies. In addition, hybrids resulting from the mating between cynomolgus males and rhesus females may have reduced viability or fertility compared to mating between rhesus males and cynomolgus females. Such asymmetries in fitness reduction have been observed in the mouse hybrid zone in Europe (Oka *et al*. 2004; Storchova *et al*. 2004). Careful monitoring of the genetics and reproductive performance of captive cynomolgus macaques that are identified to carry genetic introgressions of the rhesus macaque could enable research in this area.

Geographic location of hybrid zone

Distinguishing between scenarios of divergence with gene flow and secondary contact are an important step in characterizing any hybrid system; however, it comes with some difficulty. One approach is to determine the timing of introgression to rule out a divergence with gene flow model when introgression is determined to have occurred long after speciation time. However, for nonmodel systems, this can be difficult due to lack of data. Secondary contact also predicts that alleles may not have spread across an entire species range and isolation-by-distance approaches are often useful for distinguishing between these two scenarios as well. Here, we take a similar approach by hypothesizing that any signal of gene flow will be eliminated with the exclusion of particular samples from the zone of contact.

Hybridization has been postulated to be restricted to populations of cynomolgus macaques in mainland Indochina north of the Isthmus of Kra. These populations of the cynomolgus macaque are parapatric with respect to the Chinese populations of the rhesus macaque (Fooden 1964; Tosi *et al*. 2002) (Fig. 1). We estimated gene flow prior to and after the inclusion of cynomolgus samples that are derived from the proposed area of hybridization in northern mainland Indochina (Fig. 1), resulting in an increase of the population migration parameter $2N_2m_2$ from 0.023 to 0.48, respectively. LLR tests of nested models were able to reject zero migration for the latter analysis only, confirming previous work that has pinpointed the area of hybridization to mainland Indochina north to the Isthmus of Kra (Tosi *et al*. 2002). It is important to note that we were also able to reject the model where gene flow from the cynomolgus macaque to the rhesus macaque (i.e. N_1m_1, Tables 4 and 5) was zero for the analysis of gene flow where cynomolgus macaque samples from the putative area of hybridization were excluded. This is an important result not shown by others or in any of our other analyses I, II and IV (but see Hamada *et al*. 2006). The significance of this result will require further investigation because it is consistent with gene flow also in the direction of the cynomolgus macaque to the rhesus

macaque as long as only samples of known (or inferred) origin are considered. In conjunction with our results obtained from the analysis of captive macaques (see below), this result may point to a more complex history of speciation and hybridization. For example, we may need to postulate that during the early stages of divergence of the two species, hybridization was symmetrical, but after a period of allopatric isolation and upon secondary contact, in the wild, the rhesus macaque has evolved barriers to gene flow from the cynomolgus macaque.

The hybridization between the rhesus and cynomolgus macaque likely is due to secondary contact. We infer this from the dramatic drop-off of the signal for gene flow after removal of samples derived from the proposed area of hybridization in mainland Indochina (Fig. 1). This result could reflect genetic bottlenecks that populations south of the Isthmus of Kra have undergone. However, such a scenario would require us to assume that predominantly rhesus macaque alleles were lost due to genetic drift. This indeed could be the case as long as it is assumed that the frequency of introgressed rhesus alleles follows an isolation-by-distance pattern resulting in low frequencies at the southern Indochinese range of the cynomolgus macaque. Such rare alleles likely would be lost during subsequent drift.

Our data are consistent with a mixed-model scenario of relatively recent secondary contact combined with an isolation-by-distance model and genetic drift. First, the dispersal scenario depicting cynomolgus macaques spreading southward through mainland Indochina could have brought the rhesus and cynomolgus macaques into contact during earlier periods of their evolution. Thus, ancient bouts of reticulate evolution cannot be excluded based on our data. However, the subsequent dispersal of the cynomolgus macaque across the Sunda shelf would favour the formation of an isolation-by-distance pattern enhanced by genetic bottlenecks, promoting the loss of rhesus alleles acquired during earlier bouts of hybridization. Clearly, there is evidence of recent hybridization when considering the results of the analysis of microsatellite data, which estimated divergence times as low as 43 500 years (Bonhomme *et al*. 2008). The use of this approach in application to future studies of hybrid zones should allow researchers an alternative method to distinguishing between the possible scenarios of speciation between incipient species.

Effects of balancing selection

Two alternative phenomena could wrongfully lead to the inference of hybridization between rhesus and cynomolgus macaque, or, at least affect the ability to accurately estimate demographic parameters. In particular, balancing selection on ancestral polymorphic loci or the accidental crossbreeding of rhesus with cynomolgus macaques in stocks leading up to the samples obtained from primate research centres would mimic hybridization.

First, balancing selection acts to maintain diversity of alleles in a population, and such maintained polymorphisms often persist for long after speciation (Garrigan & Hedrick 2003). For example, a recent analysis of MHC alleles (under balancing selection in other species) showed extensive sharing of alleles between the rhesus and cynomolgus macaque (Doxiadis *et al*. 2006). Our gene-by-gene analyses of the two loci *mhcg* and *abo*, both functioning in immune response, revealed symmetrical migration patterns in our analyses consistent with trans-species polymorphism (data not shown).

Our statistical tests revealed balancing selection acting on six loci (*nbpf*, *dpb1*, *mhca*, *mhcg*, *abo*, and *apoe*) and gene annotations indicated the possibility that an additional six loci, although not detected statistically (*3kgg*, *b2m*, *c4*, *ccl2*, *ccr5*, and *r1c2*), may be under balancing selection. To obtain IM estimates for parameters and a conservative result regarding the possibility that balancing selection accounts for hybridization estimates, we excluded a total of 11 loci during analysis III. The potential for balancing selection as inferred from gene annotations was perceived as the main concern (Table 2). Differences of the results obtained in analysis III when compared to analysis I may indicate a role of balancing selection and/or differences in the power of analyses due to the lower number of loci included. We observed a relative (nonsignificant) reduction of the best estimate for migration rates when loci potentially under balancing selection were excluded, indicating that while balancing selection may have upwardly biased our migration estimates obtained from the full data set, it is unlikely that balancing selection led us to erroneously conclude that there is gene flow between the rhesus and cynomolgus macaque. Moreover, the recovery of the originally observed asymmetry in migration in analysis I argues against a dominant role for balancing selection, because otherwise we would expect more symmetrical rates. Finally, we observed a relative (nonsignificant) reduction of divergence time when loci were removed from the analysis, indicating a number of loci with long coalescent times appear to be present in the full data set (I). It is interesting to note that should these genes be experiencing ongoing balancing selection, the IMa program seems to be robust to their use in this study for the purposes of detecting gene flow.

Hybridization in captivity

It is not completely unlikely that the rhesus and cynomolgus macaques may have interbred sometime along the lineages leading to some of the captive stocks held at primate research centres today. Analysis of a subset of loci minimizing this potential source for error yielded an estimate for gene flow from rhesus into cynomolgus populations with a

broad distribution with peak at $2N_2m_2 = 0.249$; the reverse direction was two orders of magnitude lower ($2N_1m_1 = 0.002$, Fig. 2D; Table 4). Migration rates were (nonsignificantly) reduced relative to analysis (I), which does not exclude a role for crossbreeding in captivity, although a loss of statistical power could explain this result as well (Fig. 2D; Table 4).

Our study was able to reject the possibility that hybridization in captivity accounts for *all* of the shared variation between species, although this rejection was not through the analysis we designed specifically for this purpose, but in analysis II designed for an alternative purpose. A recent study assessed introgression between the rhesus and cynomolgus macaque using individuals recently derived from captivity, which were mostly derived from the F_1 generation (Bonhomme *et al*. 2008). This study was able to detect high migration values, although they did not perform LLR tests on their data sets, and thus, the possibility that captive hybridization occurs has not been rejected (Bonhomme *et al*. 2008). As noted in the introduction, this hybridization might *exaggerate* the rates of hybridization that are inferred when including samples of captive macaques, which would account for the high migration estimates when only captive monkeys were used. However, this putative captive hybridization does not affect the conclusions of this or previous studies of hybridization between the rhesus and cynomolgus macaque in the wild (Fooden 1964; Tosi *et al*. 2002). Interestingly, the asymmetry of introgression might be more pronounced in the wild than is apparent from the data (e.g. analysis I–III). While we were unable to attach significance to values and credible intervals overlapped, we observed that introgression rates of the cynomolgus macaque into the rhesus macaque were lowest when we excluded samples obtained from captive populations ($Nm = 0.002$ vs. 0.052–1.02).

Conclusion

By applying a multilocus approach, we confirmed the previously proposed natural hybridization between the rhesus and cynomolgus macaques in mainland Indochina. One of the main advancements over most previous studies is that we explicitly accounted for the possibility that shared genetic variants between the two species are due to their recent ancestry, rather than hybridization. Moreover, we were able to substantiate the asymmetry of this gene flow, with the predominant direction from the rhesus macaque into the cynomolgus macaque. Finally, we were able to obtain estimates for the recent divergence time and ancestral population sizes of the two species. During the analyses of subsets of genes and samples, we were able to exclude the possibility that genes under balancing selection and crossbreeding in captivity might have led us to falsely conclude that hybridization between the rhesus macaque and the cynomolgus macaque occurs in the wild. A third analysis of a subset of genes and samples enabled us to confirm the location of the hybrid zone in mainland Indochina. The latter analysis also allowed us to exclude the interaction of balancing selection and crossbreeding in captivity jointly affecting our signal of gene flow.

Our analyses endorse the collection of genetic polymorphism data for a larger set of genes and wild-derived macaques from the area where hybridization occurs, and from outside this area. In particular, while we were able to deduce that the inclusion of genes under balancing selection, and the inclusion of macaques of unknown geographic origin kept in captivity, both appear to inflate estimates of the rates of introgression. Larger, more systematically collected data sets would further enable the detection of variation among genes in rates of interspecific gene flow. For example, it would be interesting to examine if genes on the X chromosome display reduced rates of introgression in the macaques as has been demonstrated in mice (Geraldes *et al*. 2006).

The natural hybridization between the rhesus and cynomolgus macaques could provide an additional model system to study the genetics of reproductive isolation and the genetic underpinnings of diversification in primates. Both species could be considered genome-enabled, and thus, genomic tools can be used to interrogate large portions of the macaque genome for clines in allele frequency across the hybrid zone in Indochina, for the detection of fixed genetic differences between the species and, potentially, for physiological or behavioural studies on late-generation hybrids that are likely to occur in captive colonies.

Acknowledgements

The Bob E. & Lore Merten Watt Fellowship as well as the Texas Rodeo and Livestock Fellowship provided support for L.S.S. We thank Marcus Kronforst, Dave Queller, Kenneth Whitney, and Naoki Osada for constructive suggestions. We would also like to thank Greg Miller at the New England Primate Research Center and Thomas J. Rowell and Dana Hasselschwert at the New Iberia Research Center for providing primate blood and/or DNA samples. We thank anonymous reviewers and the editors for suggestions that improved our manuscript.

References

Abegg C, Thierry B (2002) Macaque evolution and dispersal in insular south-east Asia. *Biological Journal of the Linnean Society*, **75**, 555–576.

Arnold ML, Meyer A (2006) Natural hybridization in primates: one evolutionary mechanism. *Zoology*, **109**, 261–276.

Bernstein IS (1966) Naturally occurring primate hybrid. *Science*, **154**, 1559–1560.

Bernstein IS, Gordon TP (1979) Interspecific and intraspecific sexual-behavior in two species of macaque — possible behavioral barrier to gene flow. *Behavioural Processes*, **4**, 265–272.

Blancher A, Bonhomme M, Crouau-Roy B *et al*. (2008) Mitochondrial DNA sequence phylogeny of 4 populations of the widely distributed cynomolgus macaque (*Macaca fascicularis fascicularis*). *Journal of Heredity*, **99**, 254–264.

Blouch RA, Groves CP (1990) Naturally-occurring suid hybrid in Java. *Zeitschrift Fur Saugetierkunde-International Journal of Mammalian Biology*, **55**, 270–275.

Bonhomme M, Cuartero S, Blancher A, Crouau-Roy B (2008) Assessing natural introgression in two biomedical model species, the rhesus macaque (*Macaca mulatta*) and the long-tailed macaque (*Macaca fascicularis*). *Journal of Heredity*, **100**, 158–169.

Borge T, Webster MT, Andersson G, Saetre GP (2005) Contrasting patterns of polymorphism and divergence on the Z chromosome and autosomes in two *Ficedula* flycatcher species. *Genetics*, **171**, 1861–1873.

Brent L, Veira Y (2002) Social behavior of captive Indochinese and insular long-tailed macaques (*Macaca fascicularis*) following transfer to a new facility. *International Journal of Primatology*, **23**, 147–159.

Bynum EL, Bynum DZ, Supriatna J (1997) Confirmation and location of the hybrid zone between wild populations of *Macaca tonkeana* and *Macaca hecki* in Central Sulawesi, Indonesia. *American Journal of Primatology*, **43**, 181–209.

Chen Z, Zhou P, Ho DD, Landau NR, Marx PA (1997) Genetically divergent strains of simian immunodeficiency virus use CCR5 as a coreceptor for entry. *Journal of Virology*, **71**, 2705–2714.

Coyne JA, Orr HA (1989) Patterns of speciation in *Drosophila*. *Evolution*, **43**, 362–381.

Deinard AS, Lerche NW, Smith DG (2002) Polymorphism in the rhesus macaque (*Macaca mulatta*) NRAMP1 gene: lack of an allelic association to tuberculosis susceptibility. *Journal of Medical Primatology*, **31**, 8–16.

Doxiadis GG, Otting N, Antunes SG *et al*. (1998) Characterization of the ABO blood group genes in macaques: evidence for convergent evolution. *Tissue Antigens*, **51**, 321–326.

Doxiadis GGM, Rouweler AJM, de Groot NG *et al*. (2006) Extensive sharing of MHC class II alleles between rhesus and cynomolgus macaques. *Immunogenetics*, **58**, 259–268.

Evans BJ, Supriatna J, Melnick DJ (2001) Hybridization and population genetics of two macaque species in Sulawesi, Indonesia. *Evolution*, **55**, 1686–1702.

Ewens WJ (1972) Sampling theory of selectively neutral alleles. *Theoretical Population Biology*, **3**, 87–112.

Excoffier L, Heckel G (2006) Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics*, **7**, 745–758.

Excoffier L, Laval LG, Schneider S (2005) Arlequin version 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47–50.

Felsenstein J (2006) Accuracy of coalescent likelihood estimates: Do we need more sites, more sequences, or more loci? *Molecular Biology and Evolution*, **23**, 691–700.

Ferguson B, Street SL, Wright H *et al*. (2007) Single nucleotide polymorphisms (SNPs) distinguish Indian-origin and Chinese-origin rhesus macaques (*Macaca mulatta*). *BMC Genomics* **8**, 43.

Fooden J (1964) Rhesus and crab-eating macaques: intergradation in Thailand. *Science*, **143**, 363–365.

Fooden J (1980) Classification and distribution of living macaques (*Macaca* Lacepede, 1799). In: *The Macaques: Studies in Ecology, Behavior, and Evolution* (ed. Lindburg DG), pp. 1–9. Van Nostrand Reinhold Co., New York.

Garrigan D, Hedrick PW (2003) Perspective: Detecting adaptive molecular polymorphism: Lessons from the MHC. *Evolution*, **57**, 1707–1722.

Geraldes A, Ferrand N, Nachman MW (2006) Contrasting patterns of introgression at X-linked loci across the hybrid zone between subspecies of the European rabbit (*Oryctolagus cuniculus*). *Genetics*, **173**, 919–933.

Gibbs RA, Rogers J, Katze MG *et al*. (2007) Evolutionary and biomedical insights from the rhesus macaque genome. *Science*, **316**, 222–234.

Hall TA (1999) BioEdit: a user friendly biological sequence alignment editors and analysis program for Windows 95/98/NT. *Nuclear Acids Symposium Series* 41, 95–98.

Hamada Y, Urasopon N, Hadi I, Malaivijitnond S (2006) Body size and proportions and pelage color of free-ranging *Macaca mulatta* from a zone of hybridization in northeastern Thailand. *International Journal of Primatology*, **27**, 497–513.

Hartl DL, Clark AG (2007) *Principles of Population Genetics*, 4th edn. Sinauer Associates, Sunderland, Massachusetts.

Hasegawa M, Kishino H, Yano TA (1985) Dating of the human ape splitting by a molecular clock of mitochondrial-DNA. *Journal of Molecular Evolution*, **22**, 160–174.

Hernandez RD, Hubisz MJ, Wheeler DA *et al*. (2007) Demographic histories and patterns of linkage disequilibrium in Chinese and Indian rhesus macaques. *Science*, **316**, 240–243.

Hey J (2006) Recent advances in assessing gene flow between diverging populations and species. *Current Opinion in Genetics and Development*, **16**, 592–596.

Hey J, Nielsen R (2004) Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–760.

Hey J, Nielsen R (2007) Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences, USA*, **104**, 2785–2790.

Higgins DG, Sharp PM (1988) CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. *Gene*, **73**, 237–244.

Hoelzer GA, Wallman J, Melnick DJ (1998) The effects of social structure, geographical structure, and population size on the evolution of mitochondrial DNA. II. Molecular clocks and the lineage sorting period. *Journal of Molecular Evolution*, **47**, 21–31.

Hubbard TJP, Aken BL, Beal K *et al*. (2007) Ensembl 2007. *Nucleic Acids Research*, **35**, D610–D617.

Hudson RR (1985) The sampling distribution of linkage disequilibrium under an infinite allele model without selection. *Genetics*, **109**, 611–631.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.

Kimura M (1969) Number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics*, **61**, 893–903.

Kronforst MR, Young LG, Blume LM, Gilbert LE (2006) Multilocus analyses of admixture and introgression among hybridizing *Heliconius* butterflies. *Evolution*, **60**, 1254–1268.

Kumar S, Tamura K, Nei M (2004) MEGA 3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics*, **5**, 150–163.

Lecis R, Pierpaoli M, Biro ZS *et al*. (2006) Bayesian analyses of admixture in wild and domestic cats (*Felis silvestris*) using linked microsatellite loci. *Molecular Ecology*, **15**, 119–131.

Lowry DB, Modliszewski JL, Wright KM, Wu CA, Willis JH (2008) The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **363**, 3009–3021.

Machado CA, Kliman RM, Markert JA, Hey J (2002) Inferring the history of speciation from multilocus DNA sequence data: The case of *Drosophila pseudoobscura* and close relatives. *Molecular Biology and Evolution*, **19**, 472–488.

Magness CL, Fellin PC, Thomas MJ *et al*. (2005) Analysis of the *Macaca mulatta* transcriptome and the sequence divergence between *Macaca* and human. *Genome Biology*, **6**, 16.

Marcon L, Choe H, Martin KA *et al*. (1997) Utilization of C-C chemokine receptor 5 by the envelope glycoproteins of a pathogenic simian immunodeficiency virus, SIVmac239. *Journal of Virology*, **71**, 2522–2527.

Melnick DJ, Hoelzer GA (1992) Differences in male and female macaque dispersal lead to contrasting distributions of nuclear and mitochondrial-DNA variation. *International Journal of Primatology*, **13**, 379–393.

Morales JC, Melnick DJ (1998) Phylogenetic relationships of the macaques (Cercopithecidae: *Macaca*), as revealed by high resolution restriction site mapping of mitochondrial ribosomal genes. *Journal of Human Evolution*, **34**, 1–23.

Munoz-Fuentes V, Vila C, Green AJ, Negro JJ, Sorenson MD (2007) Hybridization between white-headed ducks and introduced ruddy ducks in Spain. *Molecular Ecology*, **16**, 629–638.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.

Oka A, Mita A, Sakurai-Yamatani N *et al*. (2004) Hybrid breakdown caused by substitution of the X chromosome between two mouse subspecies. *Genetics*, **166**, 913–924.

Osada N, Hashimoto K, Kameoka Y *et al*. (2008) Large-scale analysis of *Macaca fascicularis* transcripts and inference of genetic divergence between *M. fascicularis and M. mulatta*. *BMC Genomics*, **9**, 90.

Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D (2006) Genetic evidence for complex speciation of humans and chimpanzees. *Nature*, **441**, 1103–1108.

Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D (2008) Complex speciation of humans and chimpanzees — Reply. *Nature*, **452**, E4–E4.

Schillaci MA, Froehlich JW, Supriatna J, Jones-Engel L (2005) The effects of hybridization on growth allometry and craniofacial form in Sulawesi macaques. *Journal of Human Evolution*, **49**, 335–369.

Schuelke M (2000) An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, **18**, 233–234.

Slatkin M (1994) An exact test for neutrality based on the Ewens sampling distribution. *Genetical Research*, **64**, 71–74.

Smith DG, McDonough JW, George DA (2007) Mitochondrial DNA variation within and among regional populations of longtail macaques (*Macaca fascicularis*) in relation to other species of the *fascicularis* group of macaques. *American Journal of Primatology*, **69**, 182–198.

Stephens M, Donnelly P (2003) A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics*, **73**, 1162–1169.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

Stevison LS, Kohn HM (2008) Determining genetic background in captive stocks of cynomolgus macaques (*Macaca fascicularis*). *Journal of Medical Primatology*, **37**, 311–317.

Storchova R, Gregorova S, Buckiova D *et al*. (2004) Genetic analysis of X-linked hybrid sterility in the house mouse. *Mammalian Genome*, **15**, 515–524.

Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Takami Y, Nagata N, Sasabe M, Sota T (2007) Asymmetry in reproductive isolation and its effect on directional mitochondrial introgression in the parapatric ground *beetles Carabus yamato* and *C. albrechti*. *Population Ecology*, **49**, 337–346.

Tiffin P, Olson MS, Moyle LC (2001) Asymmetrical crossing barriers in angiosperms. *Proceedings of the Royal Society B: Biological Sciences*, **268**, 861–867.

Tosi AJ, Coke CS (2007) Comparative phylogenetics offer new insights into the biogeographic history of *Macaca fascicularis* and the origin of the Mauritian macaques. *Molecular Phylogenetics and Evolution*, **42**, 498–504.

Tosi AJ, Disotell TR, Morales JC, Melnick DJ (2003a) Cercopithecine Y-chromosome data provide a test of competing morphological evolutionary hypotheses. *Molecular Phylogenetics and Evolution*, **27**, 510–521.

Tosi AJ, Morales JC, Melnick DJ (2000) Comparison of Y chromosome and mtDNA phylogenies leads to unique inferences of macaque evolutionary history. *Molecular Phylogenetics and Evolution*, **17**, 133–144.

Tosi AJ, Morales JC, Melnick DJ (2002) 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*, **23**, 161–178.

Tosi AJ, Morales JC, Melnick DJ (2003b) Paternal, maternal, and biparental molecular markers provide unique windows onto the evolutionary history of macaque monkeys. *Evolution; International Journal of Organic Evolution*, **57**, 1419–1435.

Wakeley J (2008) Complex speciation of humans and chimpanzees. *Nature*, **452**, E3–E4.

Watanabe K, Lapasere H, Tantu R (1991) External characteristics and associated developmental-changes in two species of Sulawesi macaques, *Macaca tonkeana* and *M. hecki*, with special reference to hybrids and the borderland between the species. *Primates*, **32**, 61–76.

Watterson GA (1978) Homozygosity test of neutrality. *Genetics*, **88**, 405–417.

Weiler A, May GE, Qi Y, Wilson N, Watkins DI (2006) Polymorphisms in eight host genes associated with control of HIV replication do not mediate elite control of viral replication in SIV-infected Indian rhesus macaques. *Immunogenetics*, **58**, 1003–1009.

Weiss ML, Goodman M, Prychodk W, Moore GW, Tanaka T (1973) Analysis of macaque systematics using gene frequency data. *Journal of Human Evolution*, **2**, 213–226.

Won YJ, Hey J (2005) Divergence population genetics of chimpanzees. *Molecular Biology and Evolution*, **22**, 297–307.

This research was completed in fulfilment of master's thesis requirements for Laurie Stevison at Rice University. She is currently continuing her doctoral research at Duke University in the laboratory of Dr Mohamed Noor, researching speciation genetics in *Drosophila*. The research of Michael H. Kohn spans mammalian evolutionary genomics, genetics of adaptation, conservation genetics, and medical genetics, using drosophila, rat and the macaque as model systems.

Supporting information

Additional supporting information may be found in the online version of this article:

Table S1 List of specimens used to sequence additional genes in this study.

Table S2 List of Forward and Reverse primers used to amplify additional genes in this study.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.