Supplementary Document

Application of kernel-ABC for three populations

We applied the kernel-ABC method to infer the demography of macaques using the IGS data of Indonesian-Malaysian *M. fascicularis* (2*N* = 30), Philippine *M. fascicularis* (2*N* = 18), and Burmese *M. mulatta* (2*N* = 10), which was previously analyzed using a different method (NAKAGOME *et al.* 2012; OSADA *et al.* 2010). The sequence data were summarized into SFS in each population (SFS$_{IM}$, SFS$_P$, and SFS$_R$,) and three dimensional tensors of HFS (3D-HFS) (Figure S1). The demographic model is shown in Figure S2. In order to investigate the trend of population size change in two *M. fascicularis* populations, we set no assumption for the population size changes and used the same prior density of *LN*(100000, 100000^2) for N_R, N_{C-anc}, N_{IM}, and N_P, respectively. The conditions of kernel-ABC and coalescent simulations were the same as those described in Materials and Methods. The estimated parameters are shown in Table S1, supporting the population expansion in the Indonesian-Malaysian *M. fascicularis* and population contraction in the Philippine *M. fascicularis*.

Validation of kernel–ABC in two population models

We generated samples that have the same number of loci (26 loci) and similar parameter values to our estimated parameter values in a constant population size model, Model 1, and Model 2 with 25% size. For each model, we generated 10,000 simulated datasets using the parameter values shown in Table S2, and calculated the average of summary statistics (2D-SFS and 2D-HFS) across all datasets. Next, we chose a typical dataset as an observation, which has the smallest sum of squared deviations from the average. Posterior means of demographic parameters were estimated using the kernel-ABC method with 20,000 samples, and the process was repeated for 100 times.

References

Nakagome, S., K. Fukumizu and S. Mano, 2012 Kernel approximate bayesian computation for population genetic inferences. ArXiv e-prints**:** arXiv:1205.3246.

Oada, N., Y. Uno, K. Mineta, Y. Kameoka, I. Takahashi *et al.*, 2010 Ancient genome-wide admixture extends beyond the current hybrid zone between *Macaca fascicularis* and *M. mulatta*. Mol. Ecol. **19:** 2884-2895.
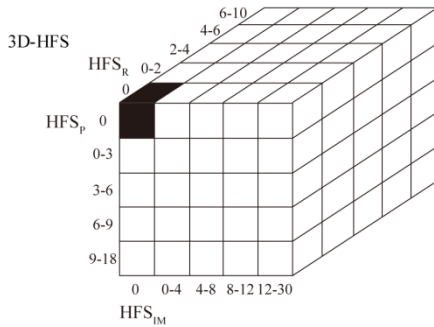
(a)

2D-SFS — SFS_IM / SFS_P

2D-SFS (SFS_P)	SFS_IM: 0	0-4	4-8	8-12	12-30
0	■	159	19	7	5
0-3	17	12	1	1	3
3-6	8	9	5	1	1
6-9	5	2	7	1	2
9-18	4	5	5	9	38

2D-HFS — HFS_IM / HFS_P

2D-HFS (HFS_P)	HFS_IM: 0	0-4	4-8	8-12	12-30
0	■	196	11	5	0
0-3	41	16	6	2	0
3-6	11	9	2	1	3
6-9	1	8	2	3	2
9-18	1	1	2	2	3

(b)

SFS_IM	0-4	4-8	8-12	12-16	16-30
	185	37	19	10	43

SFS_P	0-3	3-6	6-9	9-12	12-18
	34	24	17	28	37

SFS_R	0-2	2-4	4-6	6-8	8-10
	102	19	9	10	29

3D-HFS

HFS_R = 6-10

HFS_P \ HFS_IM	0	0-4	4-8	8-12	12-30
0	1	2	0	2	0
0-3	1	0	0	0	0
3-6	0	0	0	0	0
6-9	0	0	0	0	0
9-18	0	0	0	0	0

HFS_R = 4-6

HFS_P \ HFS_IM	0	0-4	4-8	8-12	12-30
0	1	2	0	0	0
0-3	0	0	1	1	0
3-6	0	1	0	0	1
6-9	0	0	0	0	0
9-18	0	0	0	0	0

HFS_R = 2-4

HFS_P \ HFS_IM	0	0-4	4-8	8-12	12-30
0	9	1	1	0	0
0-3	1	1	0	2	0
3-6	0	1	0	0	0
6-9	0	0	0	0	0
9-18	0	0	0	0	0

HFS_R = 0-2

HFS_P \ HFS_IM	0	0-4	4-8	8-12	12-30
0	69	7	2	0	0
0-3	1	0	2	0	0
3-6	1	0	0	1	0
6-9	0	1	0	0	2
9-18	0	0	1	0	0

HFS_R = 0

| HFS_P \ HFS_IM	0	0-4	4-8	8-12	12-30
0	■	182	9	3	0
0-3	37	16	2	0	1
3-6	9	7	2	0	1
6-9	2	7	2	3	0
9-18	1	1	2	1	3

Figure S1 Observed summary statistics for sequence data from (a) two populations (*M. fascicularis*) and (b) three populations (*M. mulatta* and *M. fascicularis*). (a) The data are summarized into 2D-SFS (5 × 5 bins excluding one bin with the frequency "0") and 2D-HFS (5 × 5 bins excluding one bin with the frequency "0"). The frequencies of alleles or haplotypes are binned with "4" in the Indonesian-Malaysian population and "3" in the Philippine population. The 5th bins in SFS_IM and HFS_IM, or SFS_P and HFS_P are "12–30," or "9–18," respectively (b) The data from Burmese *M. mulatta* are added into those from the two populations. SFS_R and HFS_R are binned with "2," and the 5th bins are "8–10" and "6–10." 3D-HFS is represented as each section separated by the bin of HFS_R.
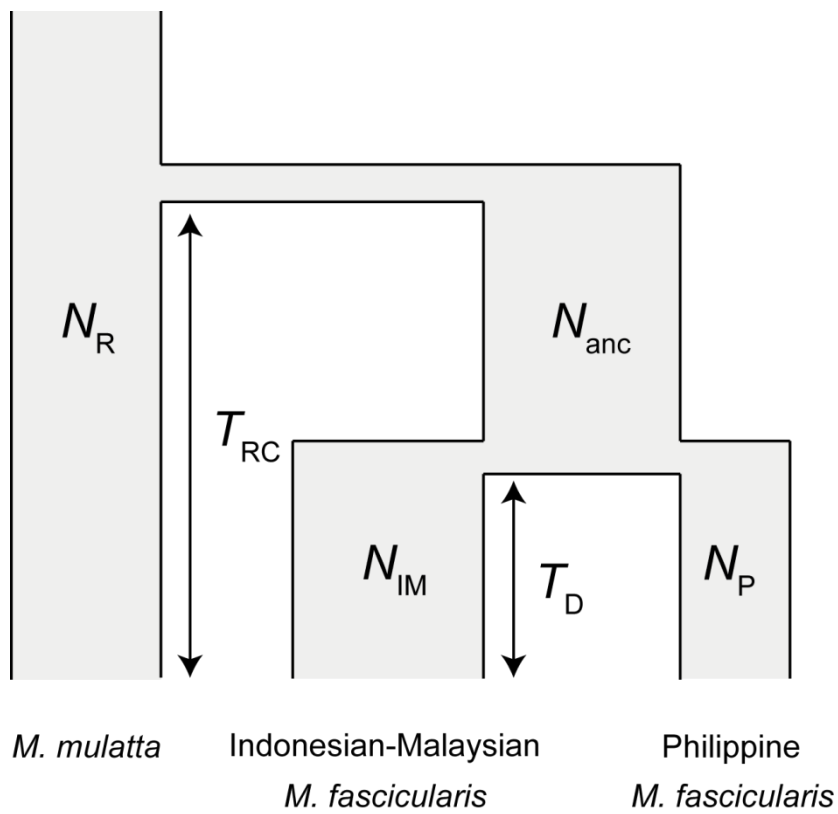
Figure S2 A demographic model of macaques. Two species have split at time T_{RC} and the two local populations of M. fascicularis have split at time T_D.

N. Osada *et al.*
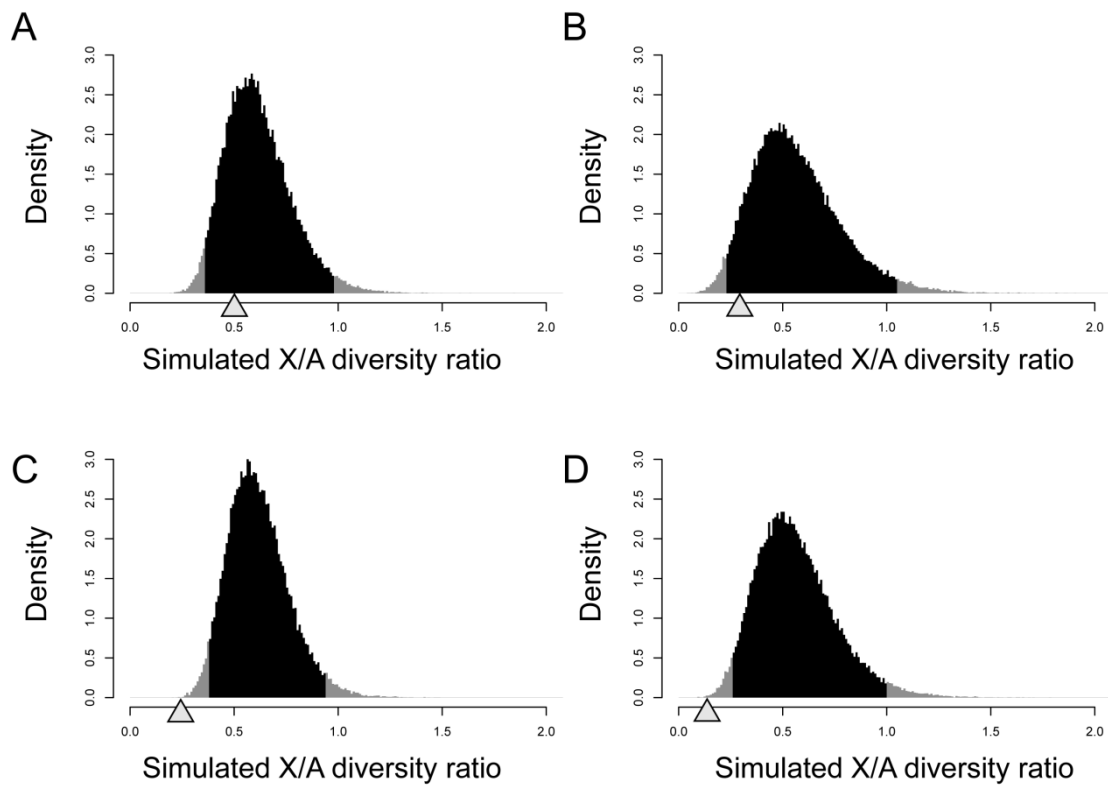
Figure S3 Distribution of X/A diversity ratio in the simulated datasets. The average recombination rate on the X chromosome was assumed to be a half of that on autosomes. The solid bars represent the density of X/A diversity ratio in the simulated datasets. The black bars indicate 95% confidence intervals. The grey triangles point out the observed values. Mean recombination rate on the X chromosome was a half of that on autosomes. Each panel represents A) IGS in the Indonesian-Malaysian population, B) IGS in the Philippine population, C) CDS in the Indonesian-Malaysian population, and D) CDS in the Philippine population.
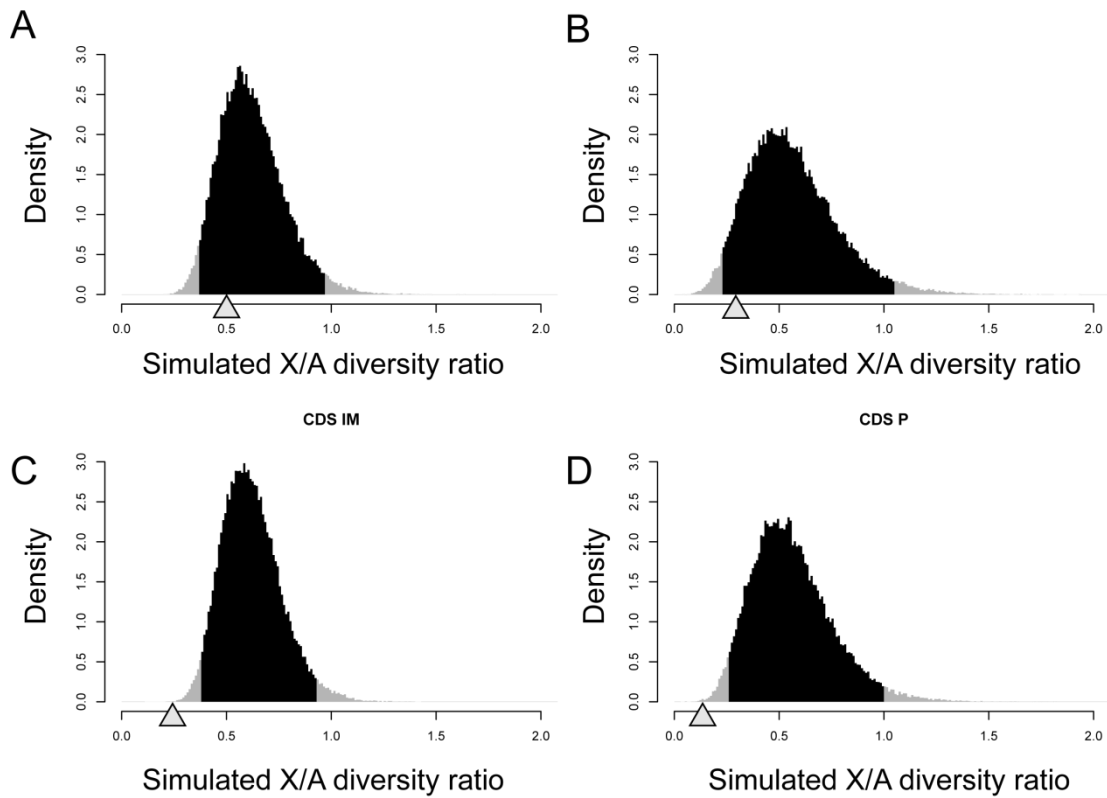
Figure S4 Distribution of X/A diversity ratio in the simulated datasets. *Papio anubis* instead of humans was used for ancestral state inference. The average recombination rate on the X chromosome was assumed to be the same as that on autosomes. The solid bars represent the density of X/A diversity ratio in the simulated datasets. The black bars indicate 95% confidence intervals. The grey triangles point out the observed values. Mean recombination rate on the X chromosome was a half of that on autosomes. Each panel represents A) IGS in the Indonesian-Malaysian population, B) IGS in the Philippine population, C) CDS in the Indonesian-Malaysian population, and D) CDS in the Philippine population.

Table S1 Estimated demographic parameters under the three population model

parameters	Priors	Posterior estimates	
		Mean	S.D.
N_R	LN(E=100000, V=100000^2)	96,958	3,734
$N_{C\text{-anc}}$	LN(E=100000, V=100000^2)	113,605	37,636
N_{IM}	LN(E=100000, V=100000^2)	335,194	30,004
N_P	LN(E=100000, V=100000^2)	50,160	15,720
T_{RC^*}	LN(E=100000, V=100000^2)	82,522	6,295
T_{D^*}	LN(E=100000, V=100000^2)	31,425	3,799

* Unit is a generation

Table S2 Accuracy of kernel ABC under population divergence models with constant size and expansion and bottleneck

		N_{anc}	N_{IM}	N_P	T_D*	T_C*
		Constant size model				
True values		100,000	100,000	100,000	25,000	–
Priors		LN(E=200000, V=200000²)	LN(E=200000, V=200000²)	LN(E=200000, V=200000²)	LN(E=50000, V=50000²)	–
Posterior estimates	Mean	102,064	104,553	98,028	25,602	–
	S.D.	627	4,684	5,059	315	–
		Model 1				
True values		100,000	1,000,000	50,000	25,000	–
Priors		LN(E=200000, V=200000²)	LN(E=500000, V=500000²)	LN(E=100000, V=100000²)	LN(E=50000, V=50000²)	–
Posterior estimates	Mean	109,705	970,719	52,136	25,524	–
	S.D.	759	31,078	787	275	–
		Model 2 (25%)				
True values		100,000	1,000,000	50,000	25,000	5,000
Priors		LN(E=200000, V=200000²)	LN(E=500000, V=500000²)	LN(E=100000, V=100000²)	LN(E=50000, V=50000²)	LN(E=10000, V=10000²)
Posterior estimates	Mean	94,931	817,480	60,852	29,840	6,972
	S.D.	486	16,354	717	293	176

* Unit is a generation