Genotyping of Cynomolgus and Rhesus Macaques Used in Research

Jean Dubach, Loyola University Medical Center, Maywood, IL
Donald Waller, PreLabs, LLC, Oak Park, IL

INTRODUCTION

DNA analysis of targeted genetic loci has revealed genetic variation and similarities that underlie observed idiosyncratic and generalized trends of responses to xenobiotics and lend efficiency to the pursuit of therapies for intractable human diseases. The non-human primate model has become important for evaluating the safety and efficacy of many therapeutic biologics due to its genetic similarity with humans. Improved predictability and reproducibility of research data in non-human primate studies should be possible if the genetic variability of sourced non-human primates were better defined.

The cynomolgus macaque (*Macaca fascicularis*) and the rhesus macaque (*Macaca mulatta*) are the most commonly used non-human primates in biomedical research. They have been used for many years without a careful evaluation of their genetic diversity. Cynomolgus monkeys used in research may come from several geographical sources, that include Vietnam, Cambodia, Thailand, Malaysia, Indonesia, Mauritius, Philippines, Japan, Burma, China, Borneo, and Java. China, however, does not have an indigenous population of cynomolgus monkeys. Consequently, animals sourced as Chinese originate from breeding programs of animals derived from other geographic areas. Rhesus monkeys originate from both India and China.

There is a growing body of knowledge suggesting that genetic background can directly influence study parameters that may be important when selecting a population of animals for a particular research project. Behavior, disease resistance and pathology have been shown to be influenced by genetic background (Bethea, et. al. 2005; Shibata et. al. 1997, Drevon-Gaillot, et. al. 2006, Vidal, et al 2010).

Genetic uniformity can insure a more consistent response to a test agent and reduce the occurrence of outlier responses. Several considerations direct us to better understand the genetic homogeneity of the animals used in a study. Group sizes in a typical non-human primate study are relatively small and when combined with unknown genetic heterogeneity, produce results with a large variance around test parameters. Previous genetic studies of the cynomolgus macaque clearly indicate population differences depending upon their geographical source (Bonhomme et al. (2005, 2008). Physical/morphological characteristics of individuals differ geographically providing additional evidence of their overall genetic diversity. There is a large contact zone (Striped area) between Rhesus (Black area) and Cynomolgus macaques (Grey area) (Figure 1). Hybridization is possible for the two species in the contact zone. *Macaca mulatta* alleles have been introgressing into the northern *M. fascicularis* stocks over time. This may be correlated with the higher level of aggression exhibited by Indochinese *M. fascicularis* relative to their insular conspecifics (Stevison

and Kohn, 2009). These studies have measured significant divergence in the populations from the mainland and island populations. Therefore, cynomolgus monkeys from isolated colonies should have less genetic variance between individual animals whereas animals from geographically less isolated areas could be more genetically diverse. Non-human primates for research are sourced from both types of genetic backgrounds. It is clear that the genetic profiles of the cynomolgus monkey can vary greatly depending on location of the source population.

Although end users usually request research animals sourced from a single geographical area, this selection may be based upon past experience, price and availability with minimal regard for genetic background. Currently no information is provided with the animals regarding their genetic background or pedigree.

We present here results from our genetic analysis of several groups of cynomolgus and rhesus macaques from various sources to characterize the genetic diversity among individual sources and between groups of animals. These data were generated to determine if the animals currently in the supply chain follow the genetic patterns of animals as defined by previous studies.
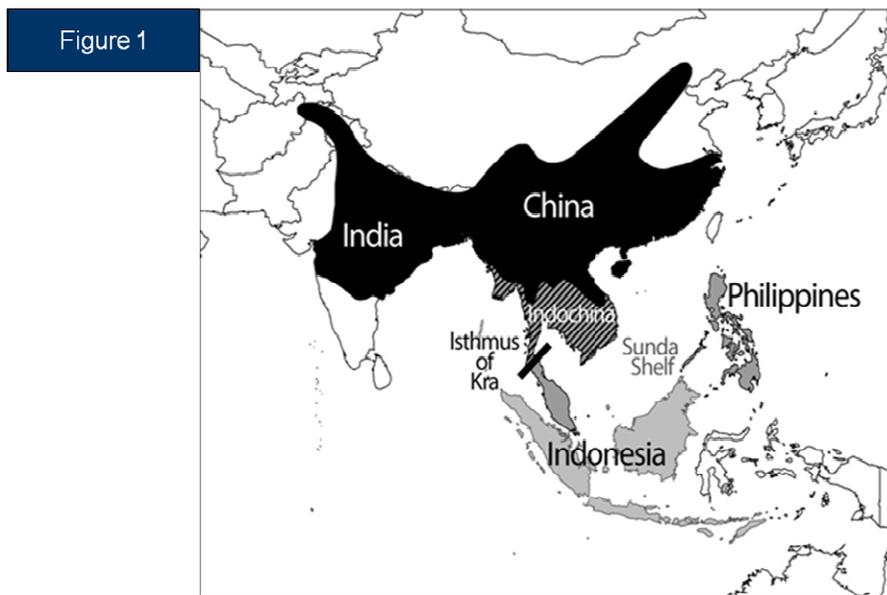
Figure 1

Figure from: Stevison, L.S. and Kohn, M.H. 2009
Molecular Ecology 18:2457-2475

METHODS

Blood samples and DNA Extraction:
The Cynomolgus samples were from animals attributed to sources in Indochina, Vietnam, Mauritius, Indonesia and the Philippines. The Rhesus samples were from animals attributed to sources in India and China. Samples from groups of animals placed in studies were also obtained (GRP1 – Mf, GRP2 – Mf, GRP 3-Mf and GRP1 – Mm, GRP2 – Mm). Blood samples were preserved with EDTA and DNA was extracted using the QIAamp Blood Kit (Qiagen, Courtaboeuf) and protocol.

Microsatellite analysis: A total of 336 Cynomolgus and Rhesus macaques were genotyped for 12 human microsatellite loci following conditions in Bonhomme et al. (2005, 2008) and include D1S548, D2S1326, D3S1768, D4S2365, D5S820, D7S2204, D8S1106, D10S1432, D14S306, D16S402, D17S791, and D18S536. PCR reactions used 1 X PCR buffer; 1.5 mM $MgCl_2$; 0.2 mM dNTPs; 4 pmol of each primer; 0.5 U GoTaq Flexi DNA polymerase (Promega, Corp.) and 40 to 90 ng genomic DNA in a total volume of 12.5 µl. PCR products were separated using the Beckman/Coulter CEQ8000 capillary electrophoresis system. The DNA Analysis System Software, version 4.3.9 (Beckman Coulter, Inc.) was used to visualize and size all fragments. Forward primers were fluorescently labeled with WellRed™ Beckman/Coulter dyes.

Genotype scoring errors were monitored by re-amplification of approximately 20 individuals at all loci and allele binning by plotting fragment size distribution per locus for each genotype by hand (Amos et al. 2007). Possible scoring errors and null alleles were also checked using Microchecker V.2.2.3 (van Oosterhout et al. 2004).

Statistical analysis: MStools: Microsatellite Toolkit v 3.1 for PC Microsoft Excel (Park 2001) was used to calculate allele frequencies observed and expected heterozygosities and to convert data files for other applications. Observed and expected heterozygosities over all loci per population were also calculated using GENEPOP v 3.3 (Raymond and Rousset 1995). GENEPOP was used to test for linkage disequilibrium between all pairs of loci and significance was determined using Fisher's Exact test with 1000 permutations. FSTAT v 2.9.3 (Goudet 1995) was used to calculate the average number of alleles per locus and FIS per population (Weir and Cockerham 1984). Genetic distance among populations, measured as pairwise FST, was also calculated in FSTAT and GENEPOP.

Structure analysis: The Bayesian clustering method, Structure, version 2.1 (Pritchard et al. 2000, Pritchard and Wen 2004) was used to infer the number of populations based on genetic variation, without information regarding origin, and assign individuals to a population based on their composite genotype. Genotype data from all individuals was analyzed with a burn-in of 200,000 iterations, followed by 500,000 iterations. The number of populations varied from K =1 to K = 10 and there were 10 repetitions at each value of K.

RESULTS

Observed levels of genetic variation in populations of Cynomolgus and Rhesus Macaques based on 12 microsatellite loci.

Population	Sample size	Observed Heterozygosity[†]	Expected Heterozygosity	Average # alleles per locus	# private alleles
Indochina – Mf[‡]	43	0.75	0.82	11.4	2
Vietnam - Mf	41	0.77	0.81	10.9	4
Mauritius - Mf	32	0.59	0.66	4.8	0
Indonesia - Mf	46	0.76	0.81	10.3	4
Philippines - Mf	10	0.72	0.73	5.6	2
GRP1 - Mf	4	0.56	0.63	3.2	0
GRP2 - Mf	16	0.73	0.82	9.8	1
GRP3 - Mf	10	0.76	0.78	7.6	0
India – Mm[‡]	43	0.69	0.74	8.6	4
China - Mm	46	0.72	0.81	11.4	13
GRP1 - Mm	16	0.74	0.80	7.9	0
GRP2 - Mm	29	0.78	0.83	10.5	7

Total   336

‡ Mf is *Macaca fascicularis*; Mm is *Macaca mulatta*
†Observed heterozygosity is the average number of heterozygous genotypes/ locus/ population
Source populations are from Indochina (IC), Vietnam (VT), Mauritius (MR), Indonesia (IN), Philippines (PH), India (IRH), and China (CRH), or one of five research groups.
GRP represents one of five research groups.

Estimates of genetic uniformity for each population, F_{IS} values.

Indochina Mf	0.075
Vietnam Mf	0.075
Mauritius Mf	0.126
Indonesia - Mf	0.062
Philippines - Mf	0.018
GRP1 - Mf	0.129
GRP 2 - Mf	0.118
GRP 3 - Mf	0.031
India Mm	0.065
China Mm	0.102
GRP1 Mm	0.085
GRP2 Mm	0.066

Estimates of departure from Hardy-Weinberg equilibrium (Fis) showing a deficit of heterozygotes in sourced and research populations of Macaques.

F_{IS} can vary from 1 to –1.

The larger the value, the greater the deviation from expected equilibrium.

A positive value indicates a loss of heterozygosity possibly due to inbreeding.

	Macaca fascicularis					Macaca. mulatta		Research Populations				
	IC	VT	MR	IN	PH	IRH	CRH	GRP 1 Mf	GRP 2 Mf	GRP 3 Mf	GRP 1 Mm	GRP 2 Mm
IC	-	NS	***	***	***	***	***	**	NS	NS	*	***
VT	0.0025	-	***	***	***	***	***	**	NS	NS	***	***
MR	0.092	0.104	-	***	***	***	***	NS	***	***	***	***
IN	0.028	0.032	0.093	-	***	***	***	***	***	***	***	***
PH	0.043	0.043	0.159	0.066	-	***	***	NS	***	**	***	***
IRH	0.086	0.084	0.185	0.099	0.129	-	***	***	***	***	***	***
CRH	0.026	0.027	0.109	0.044	0.058	0.072	-	***	**	***	NS	NS

Genetic differentiation (Fst) values are below the diagonal and significance levels are above the diagonal among pairs of sourced Macaque populations.

Levels of significance between sourced populations/groups of research Macaques are based on 66000 permutations in FSTAT [1] and adjusted for multiple comparisons using the standard Bonferroni correction. (* = p<0.05; ** = p<0.01; *** = p<0.005).

F$_{ST}$ can vary from 0 to 1.
A zero value indicates there is no difference among populations.
A positive value is an estimate of genetic distance between a pair of populations.
A small negative value indicates that pair of populations is more similar than expected by chance.

Structure assignment of Cynomolgus and Rhesus macaques sourced from seven geographic locations.

Each vertical bar is a single individual macaque. Colors indicate the mixture of genotypes that assign the individual to a group or location.

Cynomolgus
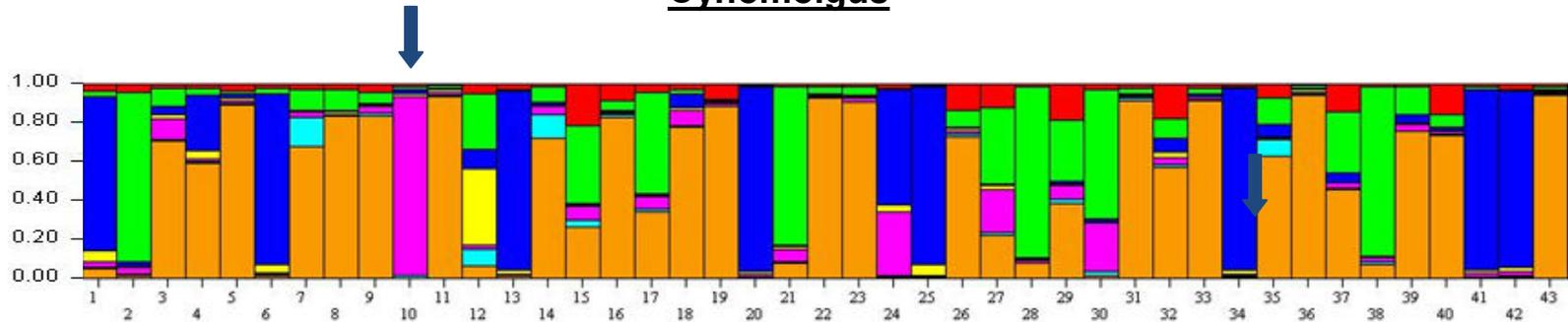
Indochina - Mf
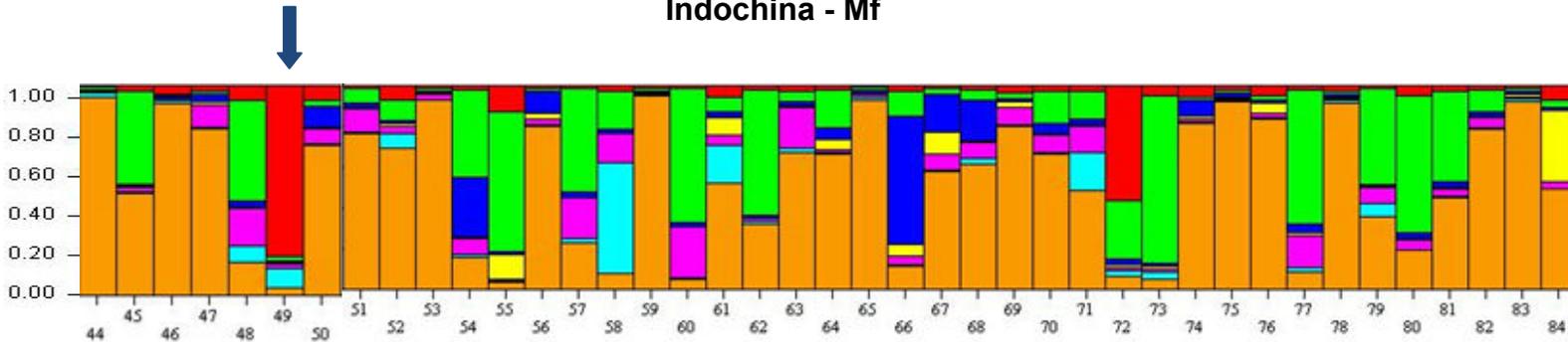
Vietnam - Mf
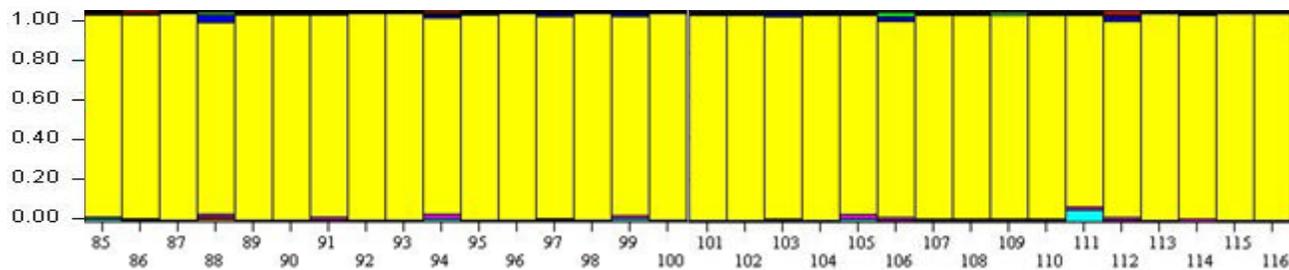
Mauritius - Mf
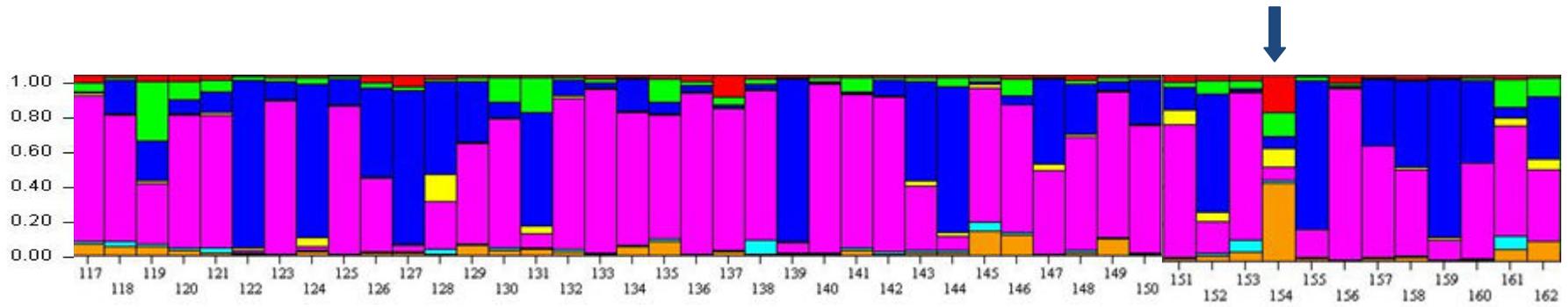
Indonesia
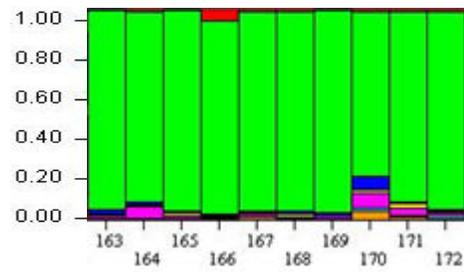
Philippines

Rhesus

Indian - Mm

Chinese - Mm
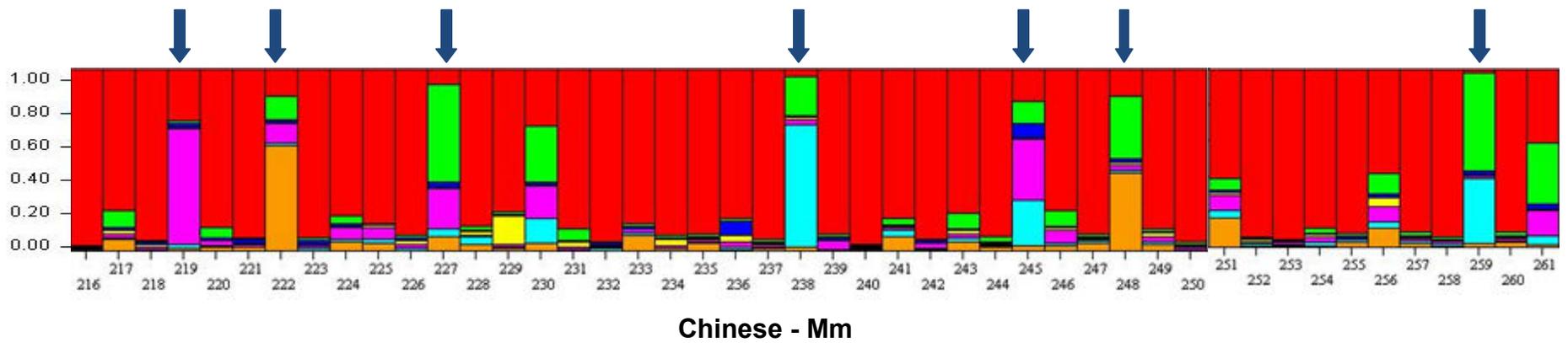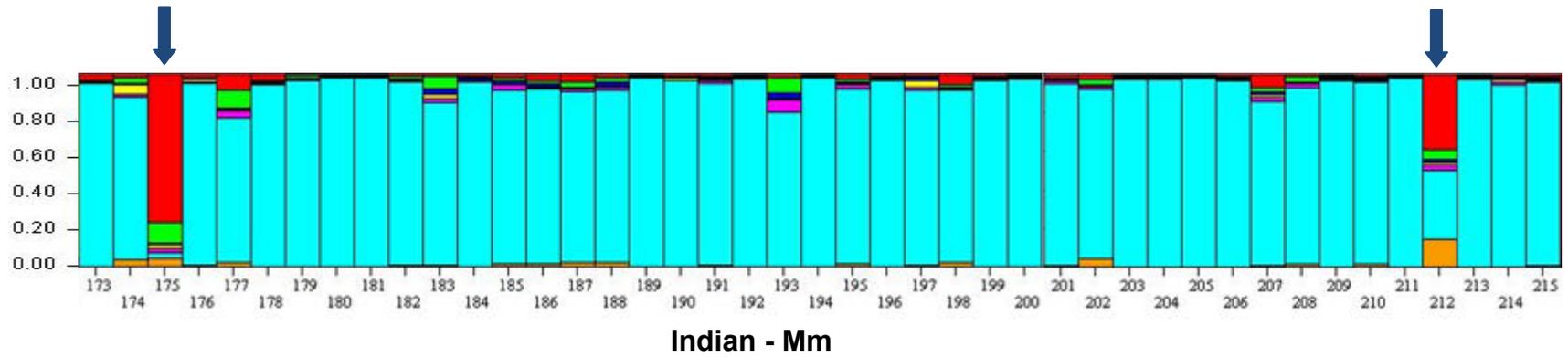
CONCLUSIONS

Measures of genetic divergence (Fst) are highly significant for all population pairs except for the sympatric populations from Indochina and Vietnam. Observed levels of genetic variability were highest in Vietnam and Indochina and lowest in Mauritius and research GRP1, which is also composed of Mauritius macaques. Structure analysis suggests that there are four genetically different groups of Cynomolgus macaques and two different groups of Rhesus macaques. Both Mauritius and Philippine macaques are very uniform within each population, yet there is significant divergence among all groups.

Individuals can be found in some geographic groups that appear to have different origins genetically as indicated by arrows (#10, 49, 154, 175, 212, 219, 222, 227, 238, 245, 248, and 259). The differences among groups may indicate variable responses when evaluated in pharmacological and toxicological testing. Consequently, the origin and mixture of individuals in each study may produce results with large variances.

Genetic testing can measure the genetic variation within individual animals or specimens that form a research group. Individuals that appear to be genetically different from the other members of a group can be identified and their background more fully investigated to determine if they are inappropriate to be included within the small group size often used for toxicology testing. This testing can also be easily applied to past study populations utilizing small samples (1cm^2) of frozen (preferred over formalin-fixed) archived tissues.

REFERENCES

Bethea, C.L., Streicher, J.M., Mirkes, S.J., Sanchez, R.L., Reddy, A.P. Cameron, J.L. (2005) Serotonin-related gene expression in female monkeys with individual sensitivity to stress, Neuroscience 132, 151-166.

Bonhomme, M, A Blancher, B Crouau-Roy (2005) Multiplexed microsatellites for rapid identification and characterization of individuals and populations of Cercopithecidae. American Journal of Primatology 67:385-391.

Bonhomme, M, A Blancher, S Cuartero, L Chikhi, B Crouau-Roy (2008) Origin and number of founders in an introduced insular primate: estimation from nuclear genetic data. Molecular Ecology 17:1009-1019.

Drevon-Gaillot, Elodie., Perron-Lepage, Marie-France., Clément, Christian and Burnett, Roger (2006) A review of background findings in cynomolgus monkeys (*Macaca fascicularis*) from three different geographical origins. Experimental and Toxicologic Pathology 58:77-88.

Goudet, J. (1995) FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). *Available from http://www.unil.ch/izea/softwares/fstat.html*. Updated from Goudet (1995), 2001.

Guo, S. W. and E. A. Thompson (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48:361-372.

Park SDE (2001) Trypanotoleranace in West African cattle and the population genetic effect of selection, PhD Thesis. University of Dublin.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics, 155, 945-959.

Pritchard JK, Wen W (2004) *Documentation for* STRUCTURE *Software* (version 2, p. Structure). Department of Human Genetics University of Chicago, Chicago.

Raymond M, Rousset F (1995) GENEPOP (version 1.2). Population-genetics software for exact tests and ecumenicism. Journal of Heredity, 86, 248-249.

Shibata, R., Maldarelli, F., Siemon, C., Matano, T., Parta, M., Miller, G. (1997), Infection and Pathogenicity of Chimeric Simian-Human Immunodeficiency Viruses in Macaques, *Journal of Infectious Diseases*, 176:362-373

Stevison KS and Kohn MH (2009) Divergence population genetic analysis of hybridization between rhesus and cynomolgus macaques. Molecular Ecology 18:2457-2475.

Van Oosterhout, C., W. F. Hutchinson, D. P. M. Wills, and P. Shipley. "MICROCHECKER: Software for indentifying and correcting genotyping errors in microsatellite data." *Molecular Ecology Notes*, 2004: 4: 535-538.

Vidal, Justin D., Drobatz, Lita S. Holliday, Denise F. Geiger, Lee E. and Thomas, Heath C. (2010) Spontaneous Findings in the Heart of Mauritian-Origin Cynomolgus Macaques (*Macaca Fascicularis*). Toxicologic Pathology 38:297.

Weir, B. S., and C. C. Cockerham. "Estimating F-Statistics for the analysis of population structure." *Evolution*, 1984: 38: 1358-1370.