Recommendations for the genetic management of nonhuman primate colonies at the National Primate Research Centers

NHP Genetics and Genomics Working Group

Introduction

Regardless of the species, proper genetic management is an important component of any successful breeding program. Nonhuman primate genetic management has unique challenges due to their outbred nature, genetic substructure among natural populations, longer generation times, and complex social dynamics among other factors. These issues are relevant to varying degrees across species and the approaches and methodologies for genetic management discussed here are broadly applicable. Our prime focus here, however, will be on rhesus macaques both as an exemplar and due to their significant and widespread use.

There are several captive rhesus monkey colonies throughout the United States. A number of these are maintained at either National Primate Research Centers (NPRCs) or the Caribbean Primate Research Center. These various colonies contain between a few hundred and several thousand animals. Most of these rhesus macaque colonies were initially established during the 1960s by importation of animals from India. Following the export ban on rhesus monkeys by the Indian government in 1978, the addition of new animals to increase or maintain genetic diversity or to add specific desirable genotypes to existing colonies has been restricted to acquisition of animals from other colonies or private vendors. In response to the Indian export ban and faced with a sudden increase in demand for rhesus monkeys for AIDS research, importation of rhesus monkeys from China began in the late 1980s and separate colonies of Chinese rhesus monkeys now exist in several primate centers.

Two major aims of any breeding strategy for captive Indian- or Chinese-ancestry rhesus macaques should be to maintain a genetic composition consistent with their geographic ancestry (i.e. avoid cross-breeding) and to maintain or maximize genetic diversity within these populations. The genetic management of animals in the various rhesus colonies has never been coordinated across primate centers and consequently it is likely that breeding systems and strategies differ. Furthermore, some colony breeding strategies may have to accommodate the production of animals with specific genotypes to meet research demands. This may occasionally and inadvertently result in the removal of genetically valuable animals from the breeding pool and should not be undertaken without consideration of the short-term and long-term consequences..

This document presents a set of recommendations that the Nonhuman Primate Genetics and Genomics Working Group has developed to aid colony managers in making breeding decisions that complement and enhance other existing data-driven husbandry practices and goals. These recommendations are based primarily on the development and use of multigenerational pedigree information for each colony, which facilitates the calculation of genetic relationships between animals. While we realize that future advances in genetic technology and the genetic information available for rhesus macaques may facilitate the use of whole genome sequence data for the characterization and management of research colonies, whole genome data is not available at present, and the computational tools and platforms currently available not sufficiently well established for such data to be immediately and readily useful for the genetic management of breeding colonies. Consequently, the application of large-scale whole genome sequencing data is only discussed briefly.

Lastly we wish to point out that the metrics and approaches we describe here are not meant to be used as a tool for comparisons of genetic health between colonies and primate centers, but for monitoring progress within colonies. Using these guidelines should enable colony managers to ensure that they are meeting their goals of maintaining genetic diversity over the span of many generations. We also acknowledge that in some cases these recommendations will serve as a starting point for discussion and/or planning, and that general breeding strategies may need to be adapted to specific local circumstances.

Defining Genetic Diversity

The amount of genetic variation that is present in a population is generally referred to as the genetic diversity of that population. A high degree of genetic diversity means that a population contains members with many different genotypes, making that population more resistant and adaptable to environmental changes. Higher genetic diversity also potentially provides a wider range of phenotypes for researchers to study, a benefit of using non-human primates as research animals.

Genetic diversity is determined by the number of alleles that are present in a population at a given genetic locus and/or the degree of heterozygosity at specific nucleotides within individual members. A higher level of heterozygosity generally results in increased phenotypic variability in a population which leads to greater population-level responsivity to changing selection pressures. A higher level of heterozygosity (diversity) will also produce a wider variety of phenotypes including those associated with disease-related traits, thus providing a wider range of biomedical research opportunities.

In most captive populations, the goal of management is generally to maintain a high level of genetic diversity and this can be accomplished with careful breeding strategies. In contrast, genetic diversity can be reduced by various practices. An unintentional consequence of selecting breeding animals based on a specific trait will be reduced variability at genetic loci related to that trait, and at other linked loci. Breeding strategies may also inadvertently choose closely related animals as breeders, or deliberately make that choice when the aim is to produce populations with similar genetic backgrounds. In such instances, mating of closely related individuals over generations will result in inbred populations. Animals from such populations can sometimes be desirable for researchers because inbreeding reduces the variation among individuals in their responses to treatments that may otherwise be caused by their genotypic differences. However, the concomitant decrease in genetic diversity may be associated with lower individual fitness that may manifest itself in impairment of reproductive traits or a diminished ability to maintain immune responses. In addition, reduction of variation reduces options or opportunities for future research studies of different phenotypes.

In captive rhesus monkey colonies, genetic diversity can be increased by importation of breeding animals which, depending on the specific genetic make-up, may include animals from the wild or animals from other captive populations. Given the ban on new importation from India, we note that any general loss of genetic variability from Indian-ancestry rhesus colonies will be difficult to restore. While genetic diversity could be increased by crossbreeding members of different subspecies i.e. crossing Indian with Chinese macaques, this is generally not considered desirable by the biomedical community and is not advocated here. There will also be a very small increase in genetic diversity that occurs naturally by *de novo* mutations, but this is true of natural populations as well and will be a negligible contribution in comparison with even the most basic inbreeding avoidance.

Choosing Metrics for Colony Genetic Management

Several metrics may be recommended to aid in making decisions concerning colony management goals and actions, e.g., formation of new breeding groups, selection of animals for sale, or for assignment to research protocols, etc. These metrics can be factored into a ***genetic value analysis***, which ranks animals according to their genetic value to the breeding colony using measures of: 1) mean kinship, a measure of how many relatives an animal has within the current population; 2) z-score, a parameter which summarizes how the mean kinship of each animal compares to the overall kinship in the population, and 3) genome uniqueness, the probability that an animal possesses genetic alleles that have a low frequency in the population, making such animals potentially genetically valuable. Once these metrics have been calculated for each animal in the colony, a ranking scheme may be developed according to the priorities specific to each research center, i.e., primarily by mean kinship and z-score, by genome uniqueness, or by some combination of these metrics. Once animals have been ranked, a threshold value can be chosen above which animals are recommended for retention in the breeding colony, and below which animals may be removed from the breeding colony and assigned to another purpose.

One example of how a genetic value analysis may be used is in selecting animals for new breeding groups. From the perspective of genetic management, the goal of animal assignment to a new breeding group is to preferentially select animals with the highest genetic value and the lowest relatedness to other potential group members. Selecting animals according to these criteria will prevent the production of offspring between closely related individuals in the short-term, and maximize genetic diversity in the population over the long-term. A genetic value analysis of all animals under consideration for a new breeding group will produce a list of animals ranked by genetic value (from most to least valuable), and this list should include a threshold genetic value below which animals will not be considered for a new group. All animals above this threshold may then be assessed for pair-wise relatedness, and animals should be removed from further consideration in the new breeding group if they have relatedness to any other breeding-age adult ≥0.03 (i.e., kinship coefficient ≥0.015). Note that some breeding age animals with close relatedness may be retained within the new group depending on whether the natural transgenerational social group structure is allowed at the facility, e.g., for macaques, whether female matrilines are maintained within breeding groups or if daughters are separated from their natal groups. If female matrilines are allowed, relatedness between males and females, and relatedness between males should be avoided, while the potential genetic costs of relatedness between breeding age females may be outweighed by the potential benefits afforded by improved social stability.

There may also be more than one possible combination of males and females that may be candidates for a new breeding group. For the best possible outcome, all possible combinations that fulfill the conditions outlined above should be derived from the starting set of available animals. Having more than one possible set of animals for a new group is useful because there are usually additional considerations beyond genetic value that will also determine the selection of animals for a new breeding group, such as behavioral compatibility and social dominance rank. For example, one proposed combination of individuals may be ideal in terms of genetic management for the new breeding group but may not be behaviorally compatible, while a second possible combination of animals may avoid this problem.

Metrics for Reporting Colony Genetic Health

Many of the same metrics used for actively managing a breeding colony can be adapted for an overall assessment of colony genetic health. It is worth noting here, however, that many of these metrics are dependent on colony size, number of founders, and historical breeding choices. While comparing how these metrics change for a given colony over time can be useful in determining the overall genetic trajectory for that colony, it is often not meaningful to compare colonies against one another.

A grand mean kinship score summarizes the overall kinship measure within the colony. Measures of the distribution of mean kinship values are also important to give an accurate description of the structure within the population. While the mean value provides a measure of the location of the center of a distribution, statistical values such as skewness and kurtosis provide a quantified description of the shape of the distribution around the average value. Skewness describes the degree to which extreme values exist (a long "tail" to the distribution) while kurtosis describes the peakedness (flat vs. spikey), in the distribution. The overall mean of the distribution of kinship values as well as the shape and presence of extreme values (average, kurtosis, and skewness values, respectively) would definitively describe the genetic health of a population.

Similar values can be generated using the genomic uniqueness measure as well, answering questions such as: what is the overall or most common genome uniqueness score (mean), to what degree do extreme values exist (skewness) and how uniformly are the scores distributed (kurtosis).

One scale-free population metric that may be useful for reporting purposes is the proportion of the genetic diversity that was originally present in the population founders (any animal without ancestors in the population under study) that remains in the current population, as defined by Lacy (1989). This assessment of genetic diversity incorporates information on ***founder genome equivalents***, i.e. the expected number of founders that would be required to provide the level of genetic diversity observed in the living population if the founders were all equally represented and had lost no alleles. Thus, an estimate of founder genome equivalents incorporates information on the proportion of the extant gene pool descended from each founder (i.e., individual founder contribution), as well as the expected proportion of each founder's alleles that are retained in the living population. Maximizing founder genome equivalents maximizes gene diversity as defined above.

An important point to consider is that metrics to determine overall population genetic health will often be influenced by the populations or subsets thereof that are included (e.g. all adults vs. all animals above a certain age vs. all breeders, etc.). Because one primary goal of these recommendations is to standardize breeding colony monitoring and to facilitate comparisons of genetic parameters across centers, it is essential that metrics which describe the genetic status of whole populations be based on similar groups of animals. It can be useful to report these population genetic metrics on the population that includes all animals that have reached puberty and could join breeding groups immediately if they were selected. Alternatively, one could evaluate the population of animals in active breeding, or animals born during a given time period (e.g. a particular breeding season or year). These population subsets each have different implications, and will produce different results for genetics metrics. As a consequence it is important to report what population is under study along with the metrics.

While these metrics are valuable tools, there are several caveats that should be kept in mind regarding their use:

1) A critical limitation is that there is no consensus on what constitutes “ideal” values for these metrics. It is clear, however, that lower mean kinship is better.
2) These metrics are pedigree-based. They do not rely on genetic markers except as a means for validating the pedigree itself. An incorrect pedigree can bias these metrics in either direction. Notably, an absence of a link, for instance an unknown sire or dam, will artificially lower mean kinship and make an animal appear genetically more valuable.
3) There is also considerable uncertainty about how to deal with new animals that enter the colony and are used as breeders. To some extent, these metrics are most effective, and most critical, in multi-generational closed populations.
4) In the future, we anticipate an increased reliance on marker-based methodologies (segregating sites, nucleotide diversity, measures of linkage disequilibrium, etc.). We expand briefly on these opportunities elsewhere in this document.

Collecting Records for Breeding Management

Genetic management of breeding populations ultimately relies on the ability to identify an individual and its relationships with the rest of its breeding group. This, in turn, facilitates the assembly of pedigrees which is an essential prerequisite for the calculation of metrics that can be used for breeding decisions. This necessitates the ability to reliably identify the parents of each animal born into the colony. In smaller colonies, or where there are subgroups with particular breeding aims, this might be accomplished by caging males and females together in pairs, but in larger colonies it is much more likely that females are housed with several males making a determination of parenthood a much more challenging proposition. It is noteworthy to keep in mind that even in cases where breeding pairs are housed together, but especially so in large breeding groups, a substantial number of dams are misidentified (some estimates range as high as 5% of infants) because of recording and handling errors or spontaneous adoption of infants by other females within a social group.

Unequivocally identifying parents in rhesus macaques breeding groups is now straightforward and simple. Genetic tools, including microsatellites (STRs), provide reliable parentage assignments and this can now also be accomplished by using SNP (single nucleotide polymorphism) genotyping methods. The Genetics and Genomics Working Group has developed a panel of 96 SNPs that can be used for paternity determination and there are service laboratories that will generate SNP profiles for a relatively modest fee. SNP testing allows for greater platform flexibility and inter-center reliability compared to microsatellites. It also standardizes the tools and resulting information allowing for greater comparison and integration of results across breeding colonies. There are software programs available such as the parentage tool on the NHP Resource Portal that facilitate analysis of parentage based on SNP profiles.

As stated previously, pedigree information is essential to develop mating strategies that ensure that the representation of original colony founders across generations is equalized as much as possible, and that inbreeding is avoided. Input information for pedigree analyses, regardless of the specific file type required by the analytical software in use, can be as simple as: focal_animal, focal_sire, focal_dam, and sex, for each animal in the colony. Additional fields may be included in order to split analyses among species, populations, locations, time blocks, etc. Finally, for demographic analyses of fertility and mortality, fields for birth- or entry-date and death- or exit-date are often also included.

On the analytical side, there are a number of computer programs that can be used to aid in generating and interpreting genetic health information on colonies. PMX

(www.vortex9.org/pm2000.html) includes this type of monitoring and support for proper group assignment. Programs like Pedigree/Draw (www.pedigree-draw.com) and others allow for the visualization of pedigree structure; and programs like MateRx (www.vortex9.org/materx.html) and Cervus (www.fieldgenetics.com) provide very sophisticated analyses of the genetic health of a population. More recently a software package has been developed called Pedscope (http://pedscope.co.uk) that can be used to calculate many of the metrics for breeding management decisions that are described in this document.

Tools and Implementation

These metrics are calculated from pedigree information that connects all animals of interest. That set of animals to be included in the analysis of these metrics is already the result of some assessment and some decisions (see above).

Several computational tools exist for the calculation of these measures:
1) Pedscope will calculate both kinship and genome uniqueness
2) PedSys will calculate kinship coefficients
3) Custom designed programs have also been developed which calculate kinship coefficients

Using Genomics Information for Breeding Decisions

The genetic markers used for colony management can be any type of DNA sequence polymorphism. Traditionally, primate colonies have used microsatellite polymorphisms to perform paternity and maternity testing, and in some cases to quantify genetic diversity. The basic significance of any of these genetic markers for colony management is that they can be used to reliably compare genetic composition across a set of animals. Any reliable and polymorphic genetic markers can be used for pedigree analysis, to assess the genetic ancestry of animals, quantify the average heterozygosity across different sets of animals, or for other purposes.

Recent advances in genomic methods and in the genetic information available about rhesus macaques now make it practical to use single nucleotide polymorphisms (SNPs) for many of the analyses previously done using microsatellites. Individual SNPs provide less information than an individual microsatellite marker, but the genotyping is more reliable, faster, more comparable across laboratories, and the testing can be done at large scale in a cost effective manner. In addition, as we learn more about specific SNPs in the rhesus genome, it will become possible to refine our genetic tests to accomplish various goals, such as better ancestry testing, evaluation of functional polymorphisms that may affect phenotypes of interest, and other possibilities.

In the absence of pedigree information, a set of genetic markers can be used to estimate the kinship among pairs of animals. In other words, once a set of animals is genotyped for a given set of markers, one can calculate the predicted kinship among all pairs of individuals, thus identifying pairs that are closely related and pairs that are more distantly related. This would in principle be useful in making decisions about colony breeding groups and the selection of specific individuals for mating, as well as facilitating the process of identifying well-represented animals that may be released for research assignment.

Genetic markers can also be used to determine the level of heterozygosity within a population. There are a number of software applications that will estimate the expected and actual degree of heterozygosity among a set of animals. Actual heterozygosity can

deviate from expected heterozygosity under various conditions, but this comparison (actual vs. expected) is generally used as an indicator of either inbreeding or unrecognized population substructure (i.e. that the set of animals that is considered one fairly homogeneous population actually consists of two sets of animals with different genetic compositions). These analyses can be useful for evaluating breeding programs.

Next Generation Sequencing

Characterizing the whole genome or whole exome DNA sequence of individuals has become much more practical as next generation sequencing technologies allow for much higher throughput at much lower costs. As NGS becomes more affordable and pervasive it will gradually assume many of the roles of existing marker-based technologies as well as incorporate new techniques and advantages. Firstly, large-scale sequencing will likely lead to an increased awareness of functional variation. Current colony genetic management practices related to, for instance, MHC haplotypes will broadly be expanded into other important functional loci. The specific loci under study and how they will be incorporated into colony management and animal allocation decisions will likely be driven by administrative choices such as whether to breed animals with targeted genotypes (e.g. specific MHC haplotypes) or to prioritize specific research directions, but in general the appreciation of functional genetic variation in management practices will become higher profile.

Secondly, NGS will replace or supplement marker-based technologies. Existing SNP or microsatellite-based approaches take advantage of a relatively small group of the most informative genetic markers to determine ancestry, kinship relationships, heterozygosity, etc. NGS will incorporate much larger numbers (four to five orders of magnitude greater) of variants though many of these will be individually less informative. The net result will be a greater statistical power and broader confidence but with greater computational and bioinformatics needs. Lastly, NGS will also allow for the development of novel metrics based on its larger scope (as compared to targeted markers). This will likely include metrics making use of linkage disequilibrium blocks, pieces of DNA inherited together from a single ancestral chromosome, to quantitate actual genome uniqueness or founder representation. These tools have been developed somewhat for recombinant mouse lines, but their full implementation and implications for traditionally outbred populations like NHPs will need to be further explored.

Moving Forward: An Integrated Approach to Genetics at the Primate Research Centers

At present the rhesus colonies at the primate research centers across the United States function essentially as subpopulations with migration. It is well established in population genetics literature that over time subpopulations can experience genetic drift or differential selective pressures that ultimately may result in genetic divergence among the groups. Indeed, it is this effect which led to the stratification of the Indian and Chinese populations of rhesus macaques naturally. The goals and methods described here are a set of recommendations concerning how colony managers can best interpret and manage the genetic diversity that exists within their individual colonies. Ultimately, however, closed populations will eventually drift apart.

Ideally, from a genetics perspective, the entirety of the US academic rhesus population would be managed as a single panmictic (randomly mating) population. This would maximize genetic diversity and minimize drift and divergence among sites. There are, of course, practical limitations to our ability to realize this; transport of animals

between centers will always be difficult to accomplish on a large scale. Colony managers already appreciate the desire to bring in “outside blood”, but that must be balanced with practical husbandry issues (including cost, quarantine/viral status, behavioral integration, etc.). While moving the US primate center rhesus colonies to become a single unified national population may be unlikely, there are specific steps that can be taken now to fairly easily minimize the stratification between them.

The first of these steps is to develop, or at least maintain the ability to develop, a single unified pedigree for rhesus across centers. As animals transfer from one center to another they are often given new identification and are treated as “founder animals” at the new center; any information regarding genetic relationships among individual animals across centers is lost. Maintaining a system-wide integrated pedigree will help determine “identity by descent” or similarities between animals resulting from a shared ancestor. Another approach, large-scale genomics, can measure “identity by state” looking directly at the genetic composition of animals without regard for how shared ancestry led to them. This latter approach requires less *a priori* knowledge. On a smaller scale, this could be accomplished somewhat by the parentage SNP arrays. This would be a major benefit if all the centers used the same panel of markers for parentage testing. Allele frequency differences between centers for that panel can be considered as an initial measure of genetic differentiation.

At present we do not see evidence of substantial population differences among the continental NPRCs. (The CPRC colony does show some differences though it is unclear the degree to which they are meaningful.) It is possible that with the occasional animal transfers between centers that currently occur and proper genetic management within the centers, that significant population differentiation will not occur in a meaningful time frame. Presently, cryptic inter-individual variation within any given center is almost certainly more meaningful than inter-center variation. Nevertheless, as genomic tools progress and our understanding of functional genetics increases the genetic relationships between Primate Research Centers can and should be monitored.

Finally, we see the development and validation of pedigree information for colony animals as information that is valuable both for the genetic management of research colonies and for the use of those animals as research subjects. Pedigree information should be part of an animal’s colony record, like date of birth or sex. Thus, we recommend that the pedigree information concerning research animals be made available to any researcher who is using those colony animals in approved studies, with no restrictions on the use of the pedigree data for research purposes.

Summary

- Aim of breeding program should be preservation of genetic composition reflecting geographic origin and maintenance and maximization of genetic diversity.
- Important to understand the nature of genetic variation, means to preserve or increase it and challenges that certain breeding strategies (e.g. selection for specific genotypes) may pose.
- When selecting animals for breeding (or removal from the breeding population), metrics can be calculated that rank animals according to their genetic value within the population.
- This analysis may include parameters such as mean kinship (how many relatives an animal has in the current population) or genome uniqueness (a measure of how many rare alleles an animal possesses).

- Metrics can also be calculated to provide an assessment of genetic health for the whole colony. This facilitates monitoring breeding colonies over time. However, it is important to recognize that such metrics must be calculated consistently for the same type of population if comparisons are to be made.
- Metrics relying on current technology are mostly based on pedigree information. Consequently, it is essential to establish parentage for all animals. When animals are housed in breeding groups, unequivocal identification of parents can only be made using genetic markers such as microsatellites or single nucleotide polymorphisms.
- Future developments to assess the genetic status of individual animals as well as populations are likely to rely increasingly on genomics information.

NHP Genetics and Genomics Working Group members:

Laura A Cox, Department of Genetics and Southwest National Primate Research Center, Texas Biomedical Research Institute, San Antonio TX

Betsy Ferguson, Oregon National primate Research Center, Beaverton OR

Jim Ha, Washington National Primate Research Center, Seattle WA

Zachary Johnson, Yerkes National Primate Research Center, Atlanta GA

Sree Kanthaswamy, California National Primate Research Center, Davis CA

H. Michael Kubisch, Tulane National Primate Research Center, Covington LA

David O'Connor, Wisconsin National Primate Research Center, Madison WI

Xinxia Peng, Washington National Primate Research Center, Seattle WA

Jeffrey Rogers (Chairman, Genetics and Genomics Working Group), Human Genome Sequencing Center, Baylor College of Medicine and Wisconsin National Primate Research Center, Madison, WI

Angelina V Ruiz Lambides, Caribbean Primate Research Center, San Juan PR

David Glen Smith, California National Primate Research Center, Davis CA

Eric Vallender, New England National Primate Research Center, Southborough, MA

Amanda Vinson, Oregon National Primate Research Center, Beaverton OR

Roger Wiseman, Wisconsin National Primate Research Center, Madison WI

February 2015