Population Genetic Analysis of M/LWS Opsin Polymorphism in a Social Group of New World Monkey though Estimating Population Mutation Rate

（集団突然変異率の推定による野生新世界ザル M/LWS オプシン多型の集団遺伝学的解析）

Master Thesis
Department of Integrated Biosciences
Graduate School of Frontier Sciences
University of Tokyo

Toko Tsutsui
January 31th 2006

Contents

Abstract

The color vision of New World monkeys is highly polymorphic. The polymorphism has persisted for at least 20 million years and the long duration of polymorphism has often been explained by some form of strong balancing selection. However, an alternative hypothesis has not been tested that the polymorphism is selectively neutral. Neutral theory that was put forward by Motoo Kimura proposes that most polymorphisms at the molecular level are selectively neutral (Kimura, 1968), so that their frequency dynamics in a population is incidentally determined by random genetic drift and mutation rate. Under this theory, the polymorphism of color vision in New World monkeys could be kept for such a long duration merely by a balance between mutation rate and random genetic drift if the effective population size is large enough.

In order to test this possibility, I examined the neutrality of the allelic polymorphism of the middle-to-long-wavelength-sensitive (M/LWS) opsin gene using population genetic parameters for a wild population of New World monkeys, white-faced capuchins (*Cebus capucinus*), sampled at Santa Rosa national park, Costa Rica. For the M/LWS opsin gene and neutral references (i.e., a pseudogene and introns of several genes characterized for capuchins), I measured nucleotide diversity (π; proportion of nucleotide difference between two sequences averaged for all pairs) and nucleotide

polymorphism (S/A_n; where S is proportion of segregating sites (= polymorphic sites among samples), A_n is $\sum_{i=1}^{n-1}\frac{1}{i}$, and n is the number of samples). From π and S/A_n values of neutral references, I estimated the population mutation rate (θ) of the capuchin population by coalescent computer simulation. On the basis of the θ value, expected ranges of π and S/A_n for neutral genes were derived by, again, coalescent simulation. Then π and S/A_n values of M/LWS opsin gene were compared to the theoretical distribution of the values under the neutral model.

I showed that π and S/A_n values of the M/LWS opsin gene were significantly deviated from the theoretical distribution of the parameters. Furthermore, I showed that Tajima's D ($\pi - S/A_n$ divided by its estimated standard error) of the M/LWS opsin gene was significantly larger than 0 whereas those of neutral references were indistinguishable from 0. These results indicate that polymorphism of M/LWS opsin genes cannot be explained either by random genetic drift of neutral mutations or by population historic events such as population reduction or fusion of highly differentiated populations, but is explained only by balancing selection. This is the first study to demonstrate that polymorphism in color vision of New World monkeys is not neutral variation in a wild population.

1. Introduction

1.1 The color vision of New World monkeys

The color vision of New World monkeys (Platyrrhine primates) is unique among all animals in that it is highly polymorphic (Jacobs, 1998). This phenotypic variation results from allelic polymorphism of the single locus middle-to-long-wavelength-sensitive (M/LWS) cone visual pigment (opsin) gene residing on the X chromosome (Mollon et al., 1984; Kawamura et al., 2001). There are typically three M/LWS opsin alleles found in the New World monkeys. Absorption spectra of the three M/LWS opsin alleles are determined by three amino acid sites, 180, 227 and 285 (Yokoyama and Radlwimmer, 1999; Hiramatsu et al., 2004). The combinational differences of three alleles determine their phenotypic variation. Male has dichromatic color vision, while female has either dichromatic or trichromatic color vision.

It has been considered that the tri-allelic system arose only once in the New World monkey lineage, and that the amino acid differences at the functionally critical sites among alleles persisted in New World monkeys for about 20 million years (Hunt et al., 1998). This extremely long duration of polymorphism has often been explained by a form of strong natural selection, that is, a balancing selection.

However, the balancing selection hypothesis has not verified, and the mechanisms maintaining the genetic variability in natural populations

have remained unclear and been a subject of great interest for many behavioral ecologists, geneticists, and evolutionary biologists. Many scientists have proposed possible utilities of the color vision polymorphism in New World monkeys, but these are often contradicted each other. Dichromats can penetrate some camouflages that defeat trichromats (Morgan et al., 1992), making dichromats possibly more adaptive in some foraging tasks because their major source of nutrition is figs which are cryptically colored with green-yellow. Color vision of catarrhines (human, apes and Old World monkeys) is considered to be adaptive to detect young leaves against a background of mature foliage (Dominy and Lucas, 2001). Advantage of trichromacy on detecting ripe fruts, a long-standing hypothesis, has been questioned recently, since many fruits cab be discriminated from mature leaves by yellow-blue color channel equipped in both dichromats and trichromats and also be their luminance. The M/LWS opsin alleles appear to exist at about equal frequencies among different species of squirrel monkeys under different dietary demands (Cropp et al., 2002).Then, color vision variation may not be much relevant to foraging efficiency, and the selective force remains unsettled and controversial

1.2. Neutral theory

Motoo Kimura argued that most polymorphisms observed at the molecular level are selectively neutral, so that their frequency dynamics in a

population is determined by random genetic drift and mutation rate (Kimura 1968). This is called the neutral theory, where some mutations spread to the population and eventually fixed, whereas others are lost by a simple stochastic process. Formulation of any natural selection requires a rejection of the "null hypothesis" of this neutral evolution by some statistical test using popul ation genetic parameters.

The case of color vision polymorphism is not the exception. The allelic variation of the M/LWS opsin gene in New World monkeys may not necessarily be an outcome of adaptive natural selection. To carry out a population genetic study for color vision polymorphism in a wild population, it is necessary to collect DNA samples from majority of group members who have been individually identified. To conduct such a field study, it is prerequisite to identify all individuals in the population. Our research group has collaborated with a field Primatologist, Dr. Linda M. Fedigan at University of Calgary, who has carried out a long-term (over 20 years) and continuous observation of social behaviors of a species of New World monkey, white-faced capuchins(*Cebus capucinus*) at Santa Rosa national park in Costa Rica. This collaboration has enabled our research group to access wild populations of capuchins and determine color vision type for each individual by PCR-based allelic typing method (Hiramatsu et al., 2005).

In this study, for testing neutrality of M/LWS polymorphism, I estimate population mutation rate θ by computer simulation using several

pseudogenes and intron regions as neutral references.

1.3. Basic knowledge of population genetics

Let me introduce some essential concepts of population genetics necessary to understand my study. The basic knowledge of the theoretical principles especially under the neutral situation will be explained in the following.

Population mutation rate (θ)

A central theme in population genetics is to understand how a population evolves under a given set of conditions. DNA sequence data represents the highest level of genetic resolution and allows us to develop powerful statistical approaches. One of the most powerful parameter in population genetics is the population mutation rate, $4N_e\mu$, usually symbolized as θ, where μ is the neutral mutation rate per nucleotide site per generation and N_e is the effective population size that is the theoretical population size where individuals can breed under random mating. Since individuals which do not participate in reproduction are not counted in, N_e is usually smaller than the actual population size. Under infinite site model, every new mutation occurs at the site where previous mutations have not occurred, and creates a novel allele that has not already existed in the population.

Coalescent theory

Suppose a sort of phylogenetic tree for population samples of DNA sequences. For any samples, a tree-like relationship exists among the samples which successively traces back to a common ancestral sequence of the sample sequences. This tree is called a genealogical tree or a genealogy. In a genealogy, the process of the tracing back to ancestors is called coalescence. The essence of coalescent theory is to start with a sample, and trace backward in time to identify events that occurred in the past since the most recent common ancestor (MRCA) of the sample. Mutations are assumed to occur at random along the sequence. The location of mutation sites is assumed to have no effect on fitness, in other words, mutations are selectively neutral. In a constant-size population, the probability of coalescence at the immediate previous generation between two sequences in the current generation is $1/(2N_e)$. Thus, the mean time of $2N_e$ generations is expected separating the two sequences, and the mean number of mutations per site accumulated in the two sequences is expected to be $4N_e\mu$.

Two approaches estimating of population mutation rate (θ)

Under the stipulation that the sample be from a population that is in equilibrium between mutation and random genetic drift, and also that the polymorphisms be selectively neutral, θ can be related to the following two

estimators.

① $E(\pi)=\theta$

② $E(S)=\theta A_n$,

① θ from nucleotide diversity (π)

A quantity used to assess polymorphisms at the DNA level is the nucleotide diversity, typically denoted π, which is the average proportion of nucleotide differences between all possible pairs of sequences in the sample. π is obtained by summing the proportion of mismatches across the sample and dividing by the total number of pairwise comparisons. That is,

$$\pi=\frac{\sum \pi_{ij}}{{}_nC_2},$$

where π_{ij} represent number of different nucleotide between sequence i and j, divided by the length of the sequence, i.e. the proportion of differences.

The above formulation of π is the measurement of π from actual sequence data. In theory, π is also the expected number of differences per site between randomly chosen sequences from a population. This expected difference is $4N_e\mu$ ($=\theta$) as explained in page 7. Therefore, θ can be estimated by the measure of π.

② θ from proportion of segregating site (S)

Another estimator is the proportion of segregating sites. A segregating site is the site that is polymorphic in the sample. Consider i alleles present at some time in a population size *Ne*. The probability that there is a coalescence in the immediately preceding generation is $\frac{1}{2Ne}{}_i C_2 = \frac{i(i-1)}{4Ne}$. Therefore, for the mean of the coalescence time is expected to be $\frac{4\,Ne}{i(i-1)}$. Thus, the total time in terms of generations encompassed by all the branches of the genealogy is given as $4Ne\sum_{i=1}^{n-1}\frac{1}{i}$. Then the total number of mutations per site in all branches is $4Ne\mu\sum_{i=1}^{n-1}\frac{1}{i} = \theta\sum_{i=1}^{n-1}\frac{1}{i} \equiv \theta A_n$.

In the infinite-sites mutation model, each mutation occurs at different site. Then the total number of mutations per site in all branches equals the expected proportion of segregating sites in the samples That is, $S = \theta A_n$. Hence θ can be estimated as $\theta = S / A_n$ and in this context the measured value S / A_n is expressed as nucleotide polymorphism.

There is therefore a theoretical relation between the two measures nucleotide diversity π and nucleotide polymorphism S / A_n though θ under a simple assumption that the alleles are invisible to natural selection, i.e., $\pi = S / A_n$.

Tajima's D statistic

Tajima (Tajima, 1989) pointed out that the difference between π and

S/A_n could be used as a test statistic for selective neutrality. When $\pi - S/A_n$ is expressed as d, the value of d divided by its estimated standard error is the so-called Tajima's D. Under the null hypothesis of equilibrium between mutation and random genetic drift and selective neutrality, the mean and variance of Tajima's D are approximately 0 and 1, respectively.

The major discrepancies of D from 0 occur when

① The frequencies of polymorphic variants are too nearly equal. This pattern increases the proportion of pairwise differences over its neutral expectation, hence $\pi - S/A_n$ is positive. The finding typically suggests either some type of balancing selection, in which heterozygous genotypes are favored, or some type of diversifying selection, in which genotypes carrying the less common alleles are favored. Positive D value can occur also by population historic effects such as reduction of population size (which preferentially eliminates rare alleles, then decreasing S more severely than π) or recent mixture of populations highly differentiated each other.

② The frequencies of the polymorphic variants are too unequal, with an excess of the most common type and a deficiency of the less common types. This pattern results in a decrease in the proportion of pairwise differences, so $\pi - S/A_n$ is negative. The finding typically suggests that many of mutations in the region are slightly deleterious and tend to be eliminated by selection (i.e., purifying selection), thus many mutations cannot

increase frequency, affecting increase of π more severely than S. Negative D value can also occur by population historic effects such as expansion of population (which preferentially increases rare alleles and thus S than π).

In both ① and ②, population historic effects are genome-wide and affect neutral references and the gene of interest in the same way. Thus, differences in Tajima's D between neutral references and the target gene can constitute an evidence that population historic events can be ruled out for interpreting the D values.

2. Materials and Methods

2.1. Subjects and DNA sequencing

Genomic DNA was obtained from feces of White-faced capuchin monkeys (*Cebus capucinus*) using QIAamp DNA Stool Mini kit (Qiagen, Crawley,UK). The group of White-faced capuchin monkeys used in this study is called LV group that live in the Santa Rosa National Park located in the tropical dry forest of Costa Rica. The group consists of 32 monkeys which are individually identified

DNA analysis

M/LWS Opsin gene

The region spanning introns 2 to 3 including entire exon 3 of M/LWS opsin gene (later called exon 3) and that spanning introns 4 to 5 including entire region of exon 5 (later called exon 5) were amplified using the following PCR primers, Exon3F (GGAAAACAGGGGTCAGTGGGGAAGC), Exon3R(ATGAGGAGGGCGGGAGACAGAGACA), Exon5F (CCCTCCAGCCACGGCTCTCGCCT), and Exon 5R (CTAGAGCGTCGCAACGCCAGGAGACCC). The PCR reactions (50 μl) contained 1.5 units of high fidelity Pyrobest polymerase (Takara, Tokyo, Japan) with 1xPyrobest buffer, 0.2mM each of dNTPs, 0.6 μM each of the forward and reverse primers, and 5 μl of the DNA extract from the feces. I used pure water as the template for negative control in every reaction. All

procedures were conducted in an isolated clean room which is only for this purpose to avoid contaminations. After initial denaturing at 98℃ for 5 minutes, the PCR profile consisted of 40 cycles at 98℃ for 1 minute, 67-67.5℃ for 5 seconds, 72℃ for 3 minutes for exon 3. For exon 5, the same cycles was applied as exon 3, but annealing temperature was changed to 60-67.5℃. PCR products were purified from other DNA fragments in agarose gel electrophoresis by Ultra Clean (MO bio, CA, USA).

All PCR products were cloned into an EcoRV-digested pBluescript vector and then transformed into competent cells; *Escherichia coli*. Sequencing analysis was performed on ABI3100 (Applied Biosystems, Warrington, UK) in both strands.

Neutral genes

Since there are not many available data of pseudogenes for New World monkeys, I used as neutral reference the sequence of eta-globin pseudogene and those of three previously characterized introns; SWS opsin (intron 4), Von Willebrand factor, and beta 2-microglobin. In all genes forward and reverse primers were positioned at 500-bp intervals. PCR-amplified genes were directly sequenced for double strands. Primer sequences: eta-glogin pseudogene, eta-F (GAGGTGCATTTCACTGCTGAT) and eta-R (TTTTATCATGCAGGACCTCCC); SWS opsin gene intron 4, sws-F (CCTCTGAATTTCCAGTCCTTGCATT) and sws-R

(TAACCTTGGCAGACAAGAAGTATGG); Von Willebrand factor intron, von-F (TCCATTGTCATTGAGACGGTGCAGG) and von-R (CAGGAAGGTGCTGAGCACATCCGT); beta 2-microglobin intron, beta2-F (CCCAAGACAGTAAAGTGGGGTAAGT) and beta2-R (CCATGTACTAACAAATGTCTAAAAC). For the pseudogene and all introns, the reaction condition of PCR was the same as M/LWS opsin gene, but annealing temperature was changed.

2.2. Coalescent simulations

Coalescent simulations were performed using the "sarg" softawer (Nordborg and Innan, 2003). This software simulates patterns of nucleotide polymorphism under the infinite site model with recombination. The population size was assumed to be constant. To simulate a pattern of polymorphism when all mutations are neutral, two parameters are required: population mutation rate θ, and population recombination rate ρ. Recombination was assumed to occur just like mutation. The effective population size (N_e) is explained from $\theta=4N_e\mu$ and $\rho=4N_er$.

Rejection sampling algorithm to estimate θ

A rejection sampling method was used to estimate θ from four reference regions (eta-globin pseudogene, SWS opsin intron 4, Von Willebrand factor, and beta 2-microglobin) which are supposed to be neutral.

The procedure is as follows.

1. Simulate a random value (θ') from the prior distribution of the π or S/A_n values of the four neutral references. A uniform distribution is used. $\rho=\theta$ is assumed first, and this assumption is relaxed later.
2. Simulate polymorphism in the four regions conditional on θ, and calculate the average π over the four regions, which is denoted by π_ave'.
3. Accept θ if $|\pi\_ave' - \pi\_ave\_observation| < 0.01$, otherwise go to step 1.

This process is continued until 10,000 accepted values of θ are accumulated, which represent a sample of the posterior distribution of θ conditional on the observation, π_ave_observation.

Note that instead of π, S/A_n can be used as a summary statistic to represent the amount of variation.

Testing neutrality in the M/LWS opsin gene in terms of the level of polymorphism

As balancing selection increases the level of polymorphism in the surrounding region of the selection target site, I tested if such an elevation of the level of polymorphism is observed in the opsin gene region. In other words, I examine if the observed level of polymorphism in the opsin gene is significantly higher than other regions, whish are supposed to be neutral. The four neutral regions (eta-globin pseudogene, SWS opsin intron 4, Von Willebrand factor, and beta 2-microglobin) were used as reference regions. To

obtain the null distribution of the level of polymorphism (measure by π) in the opsin gene, I used neutral coalescent simulations conditional on the estimate of θ from the four reference regions (see above). From the simulation, the null distribution of π for the opsin gene assuming there is no selection is obtained, and the observed π in the opsin gene is compared with this null distribution.

3. Results

Numbers and distribution of opsin alleles

I sequenced M/LWS opsin gene from six females and eleven males, totally 17 individuals, in this group, which corresponded to total 23 X chromosomes. The sequences were aligned by the genetic software, Genetyx (SDC, Tokyo, Japan). I found only one indel (insertion or deletion) site in intron 4. I aligned 929 bp for exon 3 (Figure 1), and 850 bp for exon 5 (Figure 2), and total 1779 bp were analyzed in the present study. Total 59 polymorphic sites (segregating sites) were found in this 1779-bp region. The allele frequencies of P530, P545, and P560 (in terms of spectral type of opsin alleles; numbers represent wavelength of maximal absorbance) were 0.13, 0.26, 0.61 respectively (Table 1). Five out of six females were heterozygous (and thus have trichromatic color vision).

Variation within each allele of M/LWS gene

Since accumulation of nucleotide changes are expected to occur at more rapid rate in noncoding sequences than that in coding sequences (Gojobori et al., 1982; Li et al., 1984), polymorphism was expected in M/LWS intron region within each allelic type. However, there was no polymorphism within any allele, and the polymorphisms were found only between different allelic types.

Nucleotide diversity (π) and nucleotide polymorphism (S/A_n) in neutral genes

Eta-globin is the best studied pseudogene for New World monkeys (Koop et al., 1986; Fitch et al., 1988; Bailey et al., 1991). There was only two segregating site in eta-globin pseudogene in 32 samples. The S/A_n was 0.0009932 and π was 0.001270 in eta-globin.

There was only single segregating sites in SWS intron 4 (50 samples), Von Willebrand factor intron (36 samples) and beta2-microglobin intron (28 samples). The nucleotide diversity of SWS intron4, Beta-globin intron, Von Willebrand factor intron were 0.000785, 0.000846, and 0.000406 respectively. Taken together, the average nucleotide diversity in four distinct neutral genes was 0.000827 (Table 2). I further calculated S/A_n (θ estimated from segregating sites) of individual neutral genes (Table 2), and the average was 0.000609.

Nucleotide diversity (π) and nucleotide polymorphism S/A_n in M/LWS gene

I next examined the nucleotide diversity of two distinct region in M/LWS; exon 3 region containing introns 2 and 3 and exon3, and exon 5 region including introns 4 and 5 and exon 5. The π of exon3, exon5, and total of both was 0.017797, 0.13741, and 0.015059, respectively (Table2). Consistent values were obtained for the cases using only coding regions; π of the cording regions for exon 3 and exon 5 was 0.015671 and 0.014575,

respectively (Table2). Thus, the $\pi$of M/LWS is about 20 times as large as those of the neutral genes which I examined in this study.

Figure 3 shows the windowed average nucleotide diversity in the sequenced M/LWS opsin gene region using sliding windows size of 100 bp. The sliding windows analysis clearly indicated that remarkably large nucleotide diversity (no lower than 0.005) was maintained throughout the M/LWS gene region sequenced in comparison with average diversity of the neutral genes (0.000827).

I further calculated S/A_n. The S/A_n of exon3, exon 5, and total 1779 bp of M/LWS opsin gene was 0.010814, 0.008925, and 0.008985, respectively. These S/A_n values in M/LWS were also much larger than (more than ten times as large as) the average value of multiple neutral genes.

Tajima's D test

Tajima's D test explained evolutional process of nucleotide based on the differences between π and S/A_n. The D values of neutral genes were -2.0043, -1.3388, -1.5894, -1.4699, respectively (Table 2) and the average D value of all neutral genes were -1.6139. There was no significant difference betweens π and S/A_n. Under the neutral condition, the D value is expected to be zero. Therefore, my data in neutral genes were consistent with neutral expectation. On the other hand, all D values in M/LWS were clearly

positive and the average of M/LWS was 2.5931 (Table 2). In the subsequent coalescent simulation (see next section), the 95% confidence interval for the distribution of D in neutral genes were {-1.805, 2.029}. Thus, the D value of M/LWS opsin gene was significantly larger than zero ($p<0.05$).

Coalescent simulation

Estimation of θ from π

The simulated distribution of θ that was obtained from observed π vales of neutral genes is shown in Figure4. Random values of θ in the range of 0~0.003 were given to the simulation. Values of θ were partitioned into 10 columns in Figure 4. When the number of neutral source is 1 (m=1), the distribution of accepted θ was broad and has less power for estimation of θ. From m=2 to 4, the distribution seems to be converged and the mean of θ shift into partition 3 or 4. Indeed, my analysis was carried out with four neutral genes, although the distribution of accepted θ may increase with a sharp peak in the range of partitions 3 and 4, i.e. θ values between 0.0006 ~ 0.0012 when the simulated number of neutral genes are 10. However, focusing on the variance of distribution, it is likely that the increasing numbers of neutral genes from 1 to 4 clearly minimized the variance of distribution. Thus my usage of four genes is statistically powerful enough. The simulated distribution indicated that more number of neutral genes clearly increased accuracy of distribution of variable in θ.

Estimating of θ from S/A_n

I further examined the computer simulation using the value of S/A_n (Figure5). Again, using one neutral gene, the distribution of θ was not sharp and the variance was too large. When the numbers increased, the distribution of θ showed steep slope, indicating increasing accuracy of the estimation to the partition 3 where θ value was in the range between 0.0006 and 0.0009. From these two computer simulations, I estimated that real θ value is in the range between 0.0006 and 0.0009.

Testing neutrality in the M/LWS opsin gene

The null distribution of the level of polymorphism (measure by π) in the opsin gene was obtained by using neutral coalescent simulations conditional on the estimate of θ from the four neutral genes (Figure 6). The mean was 0.0009315, and variance was 2.55756×10^{-7}. The 99% confidence interval for the distribution of π in neutral genes was {0.000063, 0.002880 }. Thus 0.015059 of observed π value in M/LWS opsin gene is significantly larger than neutral expectation ($p<0.01$). These results indicate that evolutional process of M/LWS gene is controlled by balancing selection.

4. Discussion

I tested the neutrality of polymorphism in M/LWS opsin gene sampled from a wild population of capuchin monkeys by estimating population mutation rate θ with the coalescent simulation. I rejected the null hypothesis that is the π and S/A_n values of M/LWS opsin gene are consistent with θ value estimated from neutral genes.

Deviation of Tajima's D from 0 occurs not only by the presence of selection but also by effects of population-historic events, such as reduction or expansion of population size and mixture of isolated populations. However, such effect is genome-wide and all genes would show similar tendencies of D values. In my study, only M/LWS opsin gene showed significantly positive D values and all neutral genes showed nearly 0 (even negative) of D values. It is a clear evidence that the polymorphism of M/LWS opsin gene is due to balancing selection.

There is no variation within each allelic type of M/LWS opsin gene even in no-cording region. This suggests that random genetic drift has been operating effectively within allele level. This, in turn, emphasizes the power of balancing selection maintaining the differences between alleles. Consistently, no recombinant allele sequence was found for the M/LWS opsin gene regions encompassing ~10kb including exons 3 and 5. I need a farther study to sequence far downstream or upstream regions of M/LWS to find out recombination between alleles, which would give us deeper insight on the

evolutionary process of M/LWS opsin genes.

Under the equilibrium between mutation and random genetic drift, the expected nucleotide diversity (π) is $4N_e\mu$. Therefore, we can estimate N_e of the capuchin population if we assume that generation time of capuchin is approximately 10 years (Fragaszy and Visalberghi, 1990) and mutation rate per site per year is close to 10^{-9}. The estimation of θ in this study is 0.00075. Then 0.00075= 4X N_e X 10 X 10^{-9}, giving N_e to be 18750, approximately 2X 10^4. This N_e value is fairly large compared to the actual group size of ~30. Similarly, we can estimate of effective population size from S/A_n because under the equilibrium model, π and S/A_n are expected be the same. The large N_e consists of many social groups mutually exchanging genes through migration. This is consisted with the long-term observation of the capuchin groups by our collaborater Dr. Linda M. Fedigan that group members immigrate and emigrate every year (Fedigan and Jack, 2004).

The allele frequencies in most natural populations are affected by mutation, migration, and natural selection. These processes can cause directional changes in allele frequency through time. Random fluctuations in allele frequency can also occur due to chance, because populations are not infinitely large and their sizes are rarely constant. Fifteen out of 32 individuals in this capuchin group disappeared because of their death or moved to other groups. The migration frequency seems to be very high and one should consider the effect of migration on the generation and

maintenance of polymorphism in the future study.

I rejected the null hypothesis that is the π and S/A_n values of M/LWS opsin gene are consistent with θ value estimated from neutral genes. Now we can safely move to the next phase of question: what is the entity of balancing selection in M/LWS opsin gene? Four forms of balancing selection have been hypothesized for color vision polymorphism: overdominant selection, frequency dependent selection, multiple niche polymorphism, and group selections. Combinations of population genetic tools and field behavioral observation would significantly progress our understanding of the evolution of color vision polymorphism and primate color vision.

5. References

Bailey, W.J., Fitch, D.H., Tagle, D.A., Czelusniak, J., Slightom, J.L., Goodman, M., 1991. Molecular evolution of the psi eta-globin gene locus: gibbon phylogeny and the hominoid slowdown. Mol Biol Evol 8, 155-184.

Cropp, S., Boinski, S., Li, W.H., 2002. Allelic variation in the squirrel monkey x-linked color vision gene: biogeographical and behavioral correlates. J Mol Evol 54, 734-745.

Dominy, N.J., Lucas, P.W., 2001. Ecological importance of trichromatic vision to primates. Nature 410, 363-366.

Fedigan, L.M., Jack, K.M., 2004. The demographic and reproductive context of male peplacements in Cebus capuchinus. Behavior 141, 755-775.

Fitch, D.H., Mainone, C., Slightom, J.L., Goodman, M., 1988. The spider monkey psi eta-globin gene and surrounding sequences: recent or ancient insertions of LINEs and SINEs? Genomics 3, 237-255.

Fragaszy, D.M., Visalberghi, E., 1990. Social processes affecting the appearance of innovative behaviors in capuchin monkeys. Folia Primatol (Basel) 54, 155-165.

Gojobori, T., Li, W.H., Graur, D., 1982. Patterns of nucleotide substitution in pseudogenes and functional genes. J Mol Evol 18, 360-369.

Hiramatsu, C., Radlwimmer, F.B., Yokoyama, S., Kawamura, S., 2004. Mutagenesis and reconstitution of middle-to-long-wave-sensitive

visual pigments of New World monkeys for testing the tuning effect of residues at sites 229 and 233. Vision Res 44, 2225-2231.

Hiramatsu, C., Tsutsui, T., Matsumoto, Y., Aureli, F., Fedigan, L.M., Kawamura, S., 2005. Color-vision polymorphism in wild capuchins (Cebus capucinus) and spider monkeys (Ateles geoffroyi) in Costa Rica. Am J Primatol 67, 447-461.

Hunt, D.M., Dulai, K.S., Cowing, J.A., Julliot, C., Mollon, J.D., Bowmaker, J.K., Li, W.H., Hewett-Emmett, D., 1998. Molecular evolution of trichromacy in primates. Vision Res 38, 3299-3306.

Jacobs, G.H., 1998. A perspective on color vision in platyrrhine monkeys. Vision Res 38, 3307-3313.

Kawamura, S., Hirai, M., Takenaka, O., Radlwimmer, F.B., Yokoyama, S., 2001. Genomic and spectral analyses of long to middle wavelength-sensitive visual pigments of common marmoset (Callithrix jacchus). Gene 269, 45-51.

Kimura, M., 1968. Evolutionaryrate at thr molecular level. Nature 217, 624-626.

Koop, B.F., Goodman, M., Xu, P., Chan, K., Slightom, J.L., 1986. Primate eta-globin DNA sequences and man's place among the great apes. Nature 319, 234-238.

Li, W.H., Wu, C.I., Luo, C.C., 1984. Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its

evolutionary implications. J Mol Evol 21, 58-71.

Mollon, J.D., Bowmaker, J.K., Jacobs, G.H., 1984. Variations of colour vision in a New World primate can be explained by polymorphism of retinal photopigments. Proc R Soc Lond B Biol Sci 222, 373-399.

Morgan, M.J., Adam, A., Mollon, J.D., 1992. Dichromats detect colour-camouflaged objects that are not detected by trichromats. Proc Biol Sci 248, 291-295.

Nordborg, M., Innan, H., 2003. The genealogy of sequences containing multiple sites subject to strong selection in a subdivided population. Genetics 163, 1201-1213.

Tajima, F., 1989. The effect of change in population size on DNA polymorphism. Genetics 123, 597-601.

Yokoyama, S., Radlwimmer, F.B., 1999. The molecular genetics of red and green color vision in mammals. Genetics 153, 919-932.

6. Acknowledgements

I would like to acknowledge the contributions of a number of people who supported this study. First of all, I greatly appreciate Dr. Shoji Kawamura who supported this study by guidance and instruction. I could not accomplish this study without their support. I am also grateful to Dr. Hideki Innan, University of Texas at Huston, who instructed me the computer simulation and genetic statistics, and Chihiro Hiramatsu who instructed me in the method of genetic analysis and determination of allelic types for all monkeys which were used in this study, Amanda Melin who corrected fecal samples at Santa Rosa national park, and other members in this laboratory who encouraged me.

Table 1

Allele frequency of M/LWS opsin gene of capuchin monkeys in LV group of Santa Rosa

Allele type	# of allele	allele frequencies
P530	3	0.13
P545	6	0.26
P560	14	0.61
Total	23	

Table 2
Polymorphisms of Nuclear Genes in Capcin Monkey in LV

Region	No. of samples	No. sites	No. segregating sites	π	Sn/An	Tajima's D
M/LWS opsin gene						
Exon3	23	170	6	15.671	9.563	1.9547
Exon5	23	240	8	14.575	9.031	1.9913
Intron2-Intron3	23	929	37	17.797	10.814	2.4881
Intron4-Intron5	23	850	28	13.741	8.925	2.0423
Total Opsin	23	1779	65	15.059	8.985	2.5931
Neutral genes						
eta-globin	32	500	2	1.2702	0.9932	0.5603
SWS Intron 4	50	500	1	0.7853	0.4465	1.0194
beta2 intron	28	500	1	0.8466	0.5139	1.0324
Von intron	36	500	1	0.4064	0.4823	-0.2319
Average Neutra	36.5	500	1.25	0.8271	0.6090	0.5951

π and Sn/An were multipllied by 1000

Figure 1 Alignent for exon3

Figure 2 Alignent for exon5

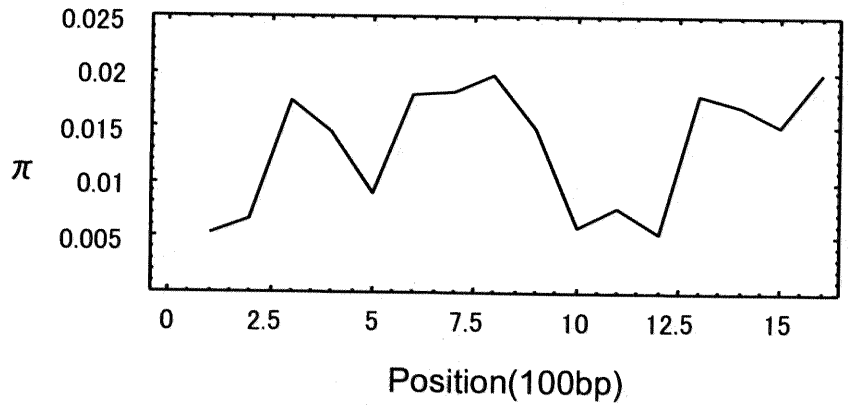

Figure 3 Sliding window analysis of M/LWS gene. Window size is 100 sites .

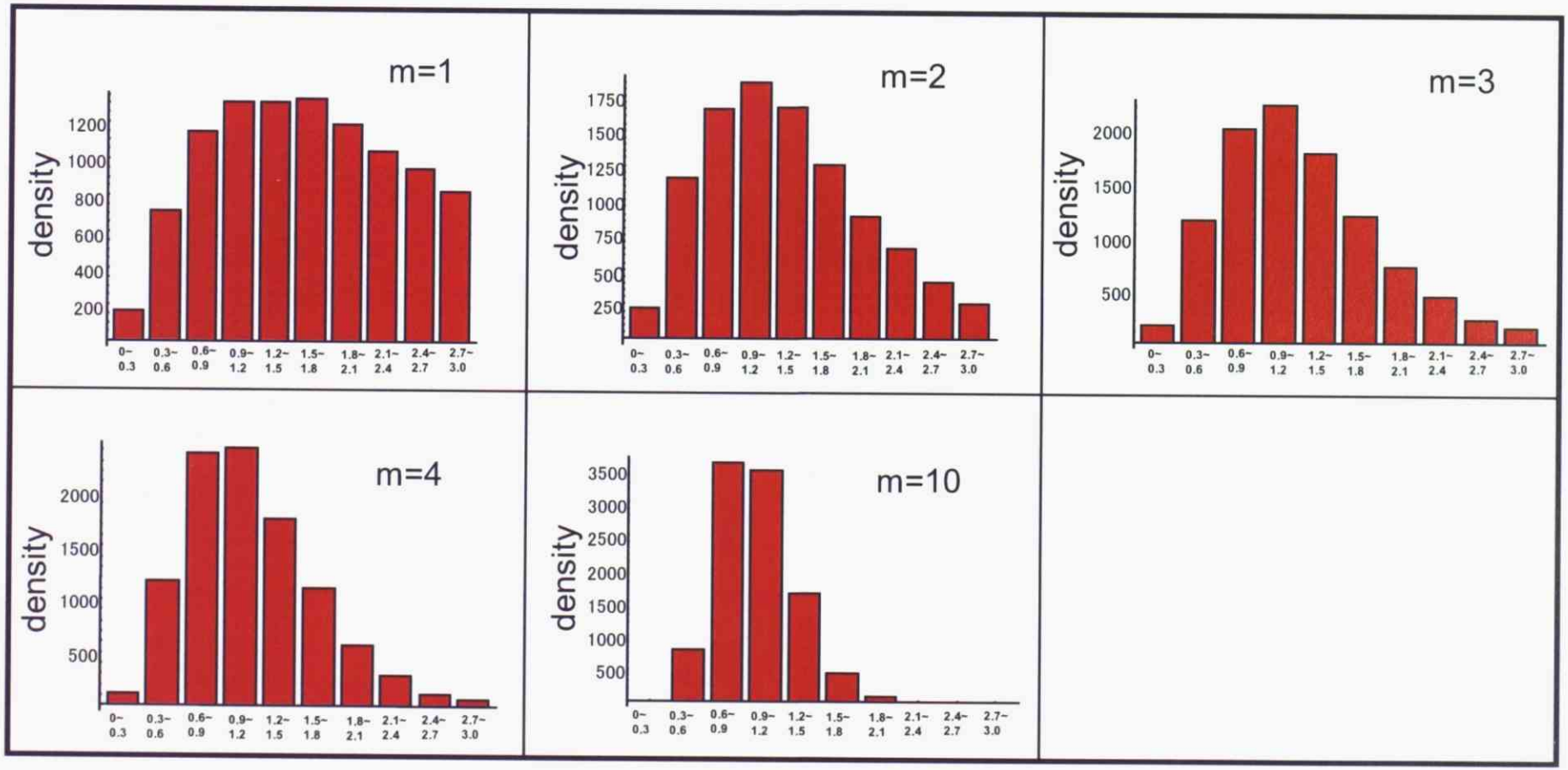

Figure 4. Distribution of θ obtained from coalescent simulation based on nucleotide diversity (π) of varied number (m) of neutral genes.

X-axis:Random variable of θ value from 0 to 0.003 partitioned off into 10 windows. θ value are multiplied by 10^3.

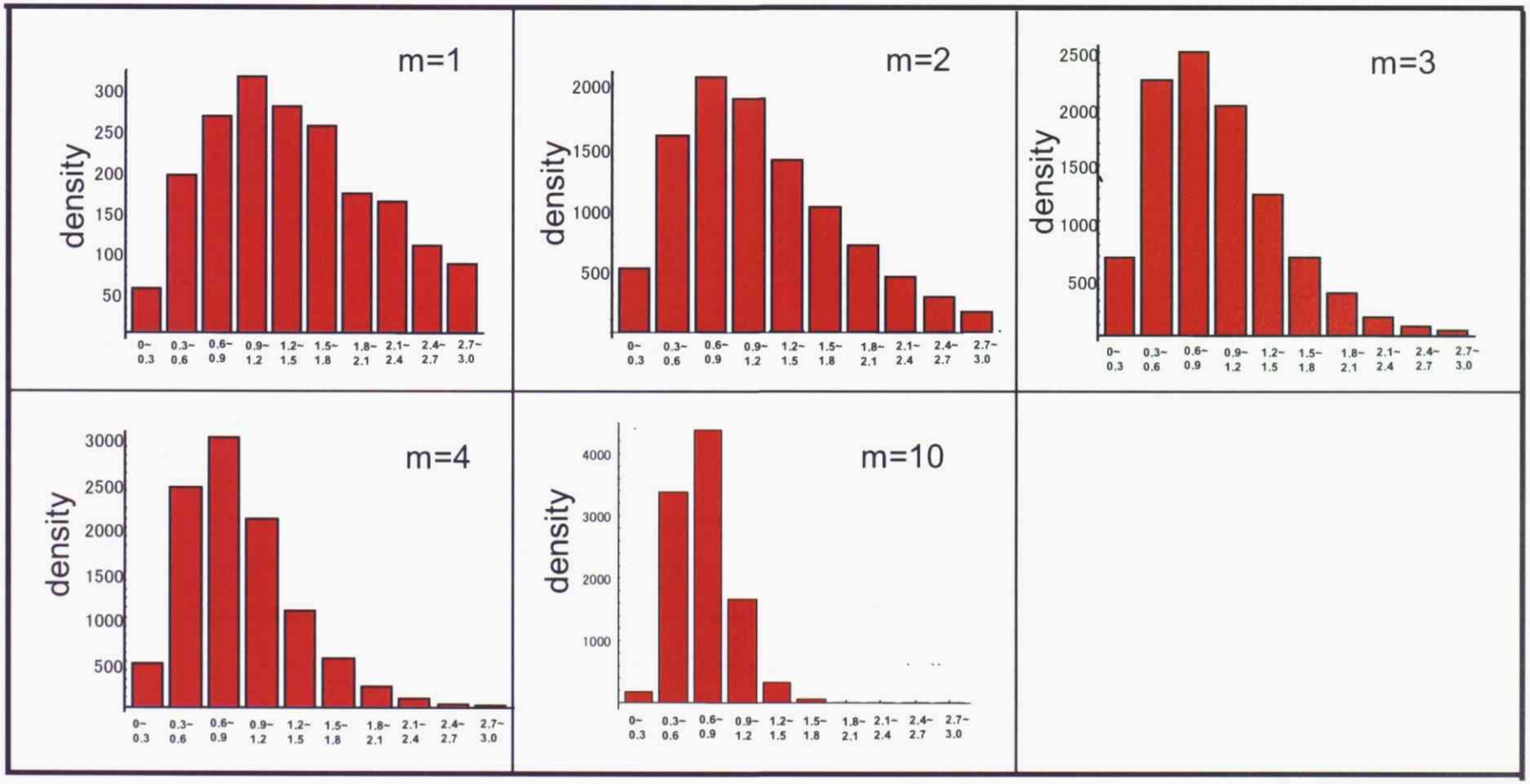

Figure 5. Distribution of θ obtained from coalescent simulation based on nucleotide polymorphism (S/An) of varied number (m) of neutral genes.

X-axis:Random variable of θ value from 0 to 0.003 partitioned off into 10 windows. θ value are multiplied by 10^3.

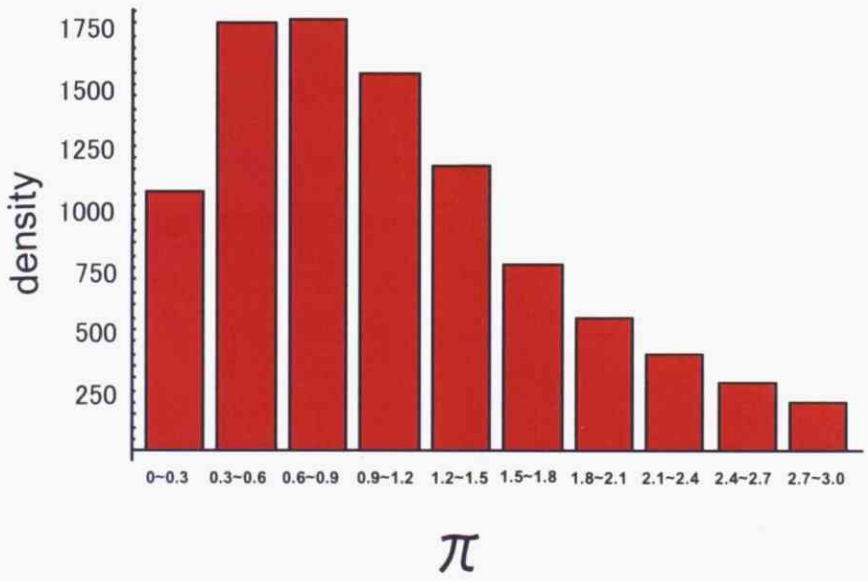

Figure 6. The null distribution of π obtained from neutral coalescent simulations using θ value distribution obtained in Figure 4, m=4.

π values are multiplied by 10^3.