

The Population Genetics of the α -2 Globin Locus of Orangutans (*Pongo pygmaeus*)

Michael E. Steiper,¹ Nathan D. Wolfe,² William B. Karesh,³ Annelisa M. Kilbourn,^{3*} Edwin J. Bosi,⁴ Maryellen Ruvolo⁵

¹ Department of Anthropology, Hunter College—City University of New York, 695 Park Avenue, New York, NY 10021, USA

² Epidemiology and Molecular Microbiology and Immunology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD 21205, USA

³ Field Veterinary Program, Wildlife Conservation Society, The Bronx, NY 10460, USA

⁴ Sabah Wildlife Department, Sepilok Orangutan Rehabilitation Center, Sandakan, Sabah, Malaysia

⁵ Department of Anthropology, Harvard University, Cambridge, MA 02138, USA

Received: 29 June 2004 / Accepted: 4 October 2004; [Reviewing Editor: Dr. Deborah Charlesworth]

Abstract. In this study, the molecular population genetics of the orangutan's α -2 globin (*HBA2*) gene were investigated in order to test for the action of natural selection. Haplotypes from 28 orangutan chromosomes were collected from a 1.46-kilobase region of the α -2 globin locus. While many aspects of the data were consistent with neutrality, the observed heterogeneous distribution of polymorphisms was inconsistent with neutral expectations. Furthermore, a single amino acid variant, found in both the Bornean and the Sumatran orangutan subspecies, was associated with different alternative synonymous variants in each subspecies, suggesting that the allele may have spread separately through the two subspecies after two distinct origination events. This variant is not in Hardy–Weinberg equilibrium (HWE). These observations are consistent with neutral models that incorporate population structure and models that invoke selection. The orangutan *Plasmodium* parasite is a plausible selective agent that may underlie the variation at α -2 globin in orangutans.

Key words: Primates — Hominoids — Malaria — Population genetics — Natural selection — Alpha-globin

Introduction

Malaria is a parasitic human disease caused by four different species of *Plasmodium*: *P. falciparum*, *P. vivax*, *P. malariae*, and *P. ovale*. The disease has long been considered a recent selective force in humans, related to the invention of agriculture (Livingstone 1958), which occurred roughly 12,000 years ago (Smith 1995). Initially, phylogenetic evidence suggested a recent lateral transfer of *P. falciparum* from birds (Waters et al. 1991, 1993), but subsequent evidence has revealed that *Plasmodium* has a more ancient relationship with the human lineage (Ayala et al. 1999; Escalante and Ayala 1994, 1995; Escalante et al. 1995; 1997, 1998; Perkins and Schall 2002; Qari et al. 1996). Many studies bolster malaria as a recently expanded disease in humans, including a recent genetic 'Eve' of the major human parasite, *P. falciparum* (Ayala et al. 1999; Rich and Ayala 2000; Rich et al. 1998; Volkman et al. 2001; see Mu et al. [2002] for an older date), a recent population

*Deceased

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession numbers AY372078–AY372110.

Correspondence to: Michael E. Steiper; email: msteiper@hunter.cuny.edu

expansion of *Anopheles* mosquito species (Donnelly et al. 2001), and a recent age for malaria resistance alleles at G6PD (Tishkoff et al. 2001; Saunders et al. 2002; Verrelli et al. 2002) and β -globin (Currat et al. 2002). These observations suggest that malaria's prevalence extends into the past, but likely in a restricted fashion prior to the recent virulent explosion of *Plasmodium* in humans (Ayala et al. 1999; Coluzzi 1999; Rich and Ayala 2000; Rich et al. 1998).

Studies of living apes provide key information for reconstructing the lives of early hominids. In humans, malaria is responsible for high rates of mortality, but few studies have investigated *Plasmodium's* impact on wild primates. At least 20 other primates also harbor their own *Plasmodium* species, including apes (Coatney et al. 1971). One well-studied case is in wild and semicaptive orangutans (*Pongo pygmaeus*), where *Plasmodium* parasites are present at high frequencies (Wolfe et al. 2002). Individuals can become ill from the parasite (Wolfe 1999), and antimalarial medicines ameliorate their symptoms (Wolfe 1999). Skinner and Hopwood (2004) link diseases including malaria to enamel hypoplasia, an indicator of stress, in large-bodied hominoids. In macaques (Genus *Macaca*), experimental infections with non-native *Plasmodium* species are often fatal in novel hosts (Fooden 1994). These studies suggest that similar to humans, malaria exerts a selective pressure on nonhuman primates, predicting that primates may also be adapting to their *Plasmodium* parasites.

In this study, we investigated the population genetics of the orangutan α -2 globin locus as a candidate gene under malarial selection. This locus was chosen for two reasons. First, molecular phylogenies show that the orangutan parasites, *P. pitheci* and *P. silvaticum*, belong to the same evolutionary radiation as *P. vivax* (Wolfe 1999), and generally, these parasites are only distantly related to *P. falciparum* and the other human parasites (Perkins and Schall 2002). Human α -thalassemias (deletions of α -globin genes) interact with *Plasmodium* (Flint et al. 1986), and there is evidence for some form of relationship between α -thalassemias and *P. vivax* (Allen et al. 1997; Williams et al. 1996; Zimmerman et al. 1999). Second, orangutan α -globin proteins have two variants that are found in both the Bornean and Sumatran subspecies (Barnicot and Jolly 1966; Bruce and Ayala 1979; de Boer and Meera Khan 1982; Sullivan and Nute 1968), suggesting balancing selection or allelic replacement may play a role.

In most apes, there are two α -globin gene copies, α -1 and α -2 globin, encoding the same protein (Bailey et al. 1997; Lauer et al. 1980; Liebhaber et al. 1981; Marks et al. 1986) (Fig. 1), though gene number polymorphism exists (Takenaka et al. 1993; Zimmer et al. 1980). The substitutions responsible for the two

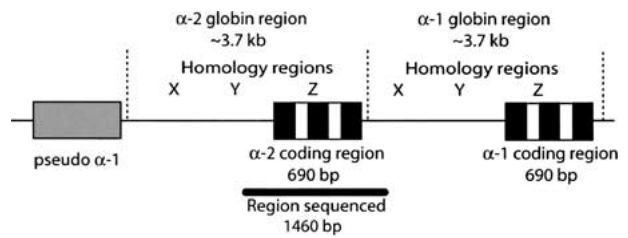


Fig. 1. Genetic map of the α -globin region of hominoids (not to scale).

alleles in orangutans are present at both α -globin loci, α -1 and α -2. In this study α -2 was examined because it is expressed at 2.6 times the rate of α -1 globin in humans (Albitar et al. 1992), implying that the majority of α -globin is most likely translated from this gene in orangutans. Each α -globin gene is approximately 700 base pairs long, with three exons encoding 141 amino acids. The orangutan α -globin variants differ at codon 57. One allele (here termed *WT*) has a glycine residue, a tiny, borderline uncharged amino acid, and the other allele (here termed *MUT*) has an aspartic acid residue, a small, negatively charged amino acid. Structurally, this codon is not in the heme-binding site, it is part of the E helix, one of the hemoglobin molecule's α -helices. This residue's only contact is with site 52, another member of the E helix (Sack et al. 1978). There is some evidence that the different hemoglobins resulting from *MUT* and *WT* alleles of orangutans are functionally different and, therefore, distinguishable by natural selection. Hemoglobins from homozygous orangutans have significantly different MCHC values (the average hemoglobin concentration per volume of packed red cells) and near-significant MCH values (the amount of hemoglobin in a red blood cell) (Huser 1970). However, there are no differences in the oxygenation properties of the different orangutan hemoglobins (Sullivan and Nute 1968) and the same codon 57 replacement (glycine \rightarrow aspartic acid) does not result in any phenotypic differences in humans (Bunn and Forget 1986; Hardison et al. 1998a, b; Huisman et al. 1996; Lehmann and Kynoch 1976), though these studies have examined only heterozygotes.

To test for the action of natural selection at the α -2 globin locus, we collected 1.46 kilobases of α -2 globin DNA sequences from orangutans of both the Bornean and Sumatran subspecies. These data provided information on patterns of intraspecific polymorphism, interspecific divergence, haplotype structure, linkage disequilibrium, and recombination at the α -2 globin locus. These observed patterns allowed testing against the expectations of neutral evolution. The presence of two variants in both subspecies suggested two main selective hypotheses, balancing selection and allelic replacement, which were forwarded previously based on protein elec-

trophoretic evidence (de Boer and Meera Khan 1982; Sullivan and Nute 1968). Though these methods enable tests for the action of selection at this locus, they cannot prove that *Plasmodium* is the selective pressure responsible for the pattern of genetic polymorphism at orangutan α -2 globin, only that it is a plausible pressure.

Materials and Methods

Samples

Samples were obtained from field sites, zoos, and research facilities for both Sumatran and Bornean orangutans (see Supplementary Material for the list of 24 orangutans examined at the DNA sequence level).

PCR

A 1.46-kilobase portion of the α -2 globin region was amplified using seminested PCR. For both PCR rounds, an initial denaturation step (2 min 30 s at 96°C) was followed by 40 cycles of denaturation (30 s at 96°C), annealing (30 s at 59.3°C in round 1 and 58°C in round 2), and extension (60 s at 70°C) with the Epicentre FailSafe PCR System (Madison, WI). Primers were designed using the program Oligo (distributed at <http://www.oligo.net/>) with the human α -globin region as a guide (No. J00153.1). The forward primer used in round 1 was 5' α @ 6065 (5'-AGT-GAC-TAT-CGC-CAG-AGG-GAA-AG-3') (numbers correspond to human sequence accession no. J00153.1) and the reverse was 3' α @ 7885 (5'-AAG-GGG-TGG-GAA-TGA-GAG-AAA-TGT-T-3'). Genomic DNA from the orangutan samples was used in the initial round. In round 2 a different forward primer, 5' α @ 6438 (5'-GGG-ATG-GGC-GGG-AGT-GGA-GT-3'), located ~400 base pairs downstream from the original forward primer, was used with 3' α @ 7885, in a PCR containing a diluted aliquot of the PCR from round one as template DNA. PCRs were analyzed via gel electrophoresis.

Restriction Digest

A larger set of Bornean orangutans ($n = 54$) was screened for *MUT* and *WT* alleles at α -2 globin via restriction digest (list of individuals available upon request). A small region encompassing the *MUT/WT* nonsynonymous site was amplified from the second round of nested PCR using the following primers: Geno6866-F (5'-TCC-TGG-CCC-CGG-ACC-CAA-AC-3') and Geno7130-R (5'-CCG-CCG-CTC-ACC-TTG-AAG-TTG-3'). These PCR products were exposed to the restriction enzyme *AvaII* (NEB), which recognizes 5'-GGWCC-3', found in *MUT* alleles.

Cloning and Sequencing

PCR fragments were cloned using the Zero Blunt TOPO PCR Cloning Kit for Sequencing (Invitrogen Corp., Carlsbad, CA) according to the manufacturer's directions. DNA from individual colonies was amplified using PCR and sequenced by the Dana-Farber/Harvard Cancer Center High-Throughput DNA Sequencing Facility. Primers M13R, M13F, 7098P (5'-GGT-GGA-CCC-GGT-CAA-CTT-C-3'), and 7168R (5'-AAC-CCG-CGG-GAT-GCT-CTG-3') were used for PCR and sequencing. Within Sumatran orangutans, both haplotypes were determined for four individuals, totaling eight haplotypes. Within Bornean orangutans, a single random haplotype was collected from 15 orangutans and

both haplotypes were sequenced from 5 orangutans. This totaled 25 Bornean haplotypes. The observed frequencies of the *MUT* and *WT* alleles in the Bornean orangutan sample did not closely match the allele frequencies derived from the restriction analysis, which sampled a larger pool of orangutans (Fisher's exact test, $p = 0.11$). To correct the *WT* and *MUT* frequencies in the DNA sequence data set, only one haplotype was included from each of the five Bornean individuals from which two haplotypes were collected. This resulted in a data set of 20 Bornean haplotypes. Culling the data set in this manner had no significant impact on the analyses. To determine each haplotype, between 4 and 15 colonies were sequenced from at least two different PCR and cloning reactions. This strategy was carried out both to identify colonies with inserts having undergone PCR-mediated recombination (Bradley and Hillis 1997; Cronn et al. 2002; Judo et al. 1998; Pääbo et al. 1990) and to detect PCR errors so they could be removed from the data set. Because the primers anchored in the downstream region were specific to the α -2 globin gene, no sites specific to the downstream region of the orangutan α -1 gene were detected in the clones sequenced. Several clones of PCR products from each orangutan sample were compared to one another. In addition, these clones were compared with sequences from other samples in the study, as well as an α -2 globin sequence that was obtained from a genomic library and not subject to PCR recombination (accession no. M12158).

Population Genetic Analyses

Population genetic statistics were calculated with the program DnaSP, versions 3.53 and 3.99 (Rozas and Rozas 1999; Rozas et al. 2003). Probability values for Tajima's D (1989) and Fu and Li's (1993) tests were generated from 5000 coalescent simulations conditional on the $\theta_{W(\text{Gene})}$ and recombination rates estimated from the data set. McDonald's heterogeneity tests (Runs tests) (1996, 1998) were carried out with the program DNA Slider (distributed by J. McDonald at <http://udel.edu/~mcdonald>). Probability values were estimated as suggested by McDonald (1998). For each test, a range of recombination parameters was used (0–128). From these simulations, the recombination parameter with the highest probability value was chosen, and additional simulations were run over a smaller interval of recombination values (e.g., 12–20), resulting in a conservative test for rate heterogeneity. The HKA test (Hudson et al. 1987) was carried out with the programs HKA (distributed by J. Hey at <http://lifesci.rutgers.edu/~heylab>) and DnaSP. In this test the α -2 globin data set was compared to 14 orangutan haplotypes from the Xq13.3 region (Kaessmann et al. 2001), which is a noncoding, presumably neutral 10-kb region. This test was corrected for the difference in N_e between the autosomally linked α -2 globin gene and the X-linked Xq13.3 region. The Haplotype test (Hudson et al. 1994) was implemented with a version of Hudson's psubs program used by Parsch et al. (2001) (10,000 random samples were generated). A molecular clock approach for estimating the date of the ancestor of the α -2 globin alleles was conducted with MEGA v. 2.1 (Kumar et al. 2001) using calibration dates of 12 million years for the human–orangutan split and 17 million years for the gibbon–orangutan split (Young and MacLatchy 2004). Where required, human and gibbon sequences were used as outgroups (human accession no. J00153.1, gibbon accession no. M94634).

Results

Fifty-four Bornean orangutans were screened for the α -2 globin *MUT* and *WT* alleles by restriction digest.

Table 1. Test for fit to Hardy–Weinberg expectation

Genotype	Region	Observed	Expected ^a	(O–E) ² /E
<i>WT</i> / <i>WT</i>	Borneo	31	26.04	0.94
	Sabah	29	25.52	0.47
<i>WT</i> / <i>MUT</i>	Borneo	13	22.92	4.29
	Sabah	12	18.96	2.55
<i>MUT</i> / <i>MUT</i>	Borneo	10	4.88	4.40
	Sabah	7	3.52	3.44
χ^2 (<i>p</i>) ^b	Borneo			10.11 (0.002**)
	Sabah			6.47 (0.011*)

^aThe number of expected individuals based on gene frequencies found in the text.

^bProbability values are two-tailed with 1 df.

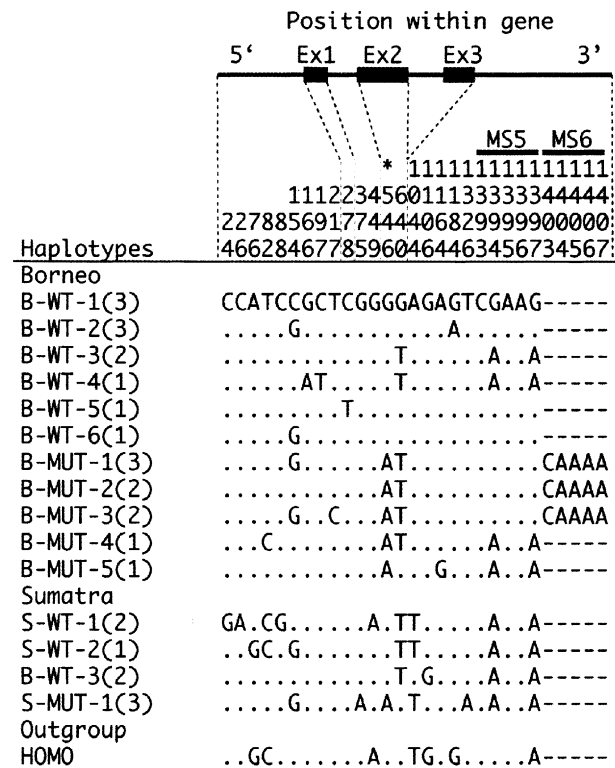


Fig. 2. Twenty-eight orangutan α -2 globin haplotypes and one outgroup (human). Each haplotype name is followed by the number of times observed in parentheses. Two repeats from the 3' microsatellite region are also included which harbor two motif point mutations and one indel mutation. There are three states in this microsatellite: five repeats, six repeats, and five repeats with an altered motif (GG at sites 1394 and 1397), referred to as 5, 6, and GG5, respectively, in the text. Although there are two differently sized variants in the population, only mutations in the motif were utilized because it was unclear how to code for this label and potentially homoplastic mutation. The structure of the gene is given above the data matrix. The nonsynonymous variant at position 546 is noted with an asterisk. Characters matching the state in the first haplotype are indicated by a "." notation. The initiation codon AUG is at site 117.

In this sample, the observed frequencies of the *WT* and *MUT* alleles were 0.69 and 0.31, respectively. Based on these gene frequencies, the Hardy–Weinberg expectations were tested against the observed number of orangutans of each genotype. This test

showed a significant difference from the equilibrium hypothesis when all Bornean orangutans were included and when restricted only to orangutans from the Sabah subpopulation (Table 1).

In total, DNA sequences from 28 orangutan haplotypes were determined for a 1.46-kilobase region encompassing the α -2 gene (Fig. 2). The haplotypes were partitioned into four groups (all orangutans, Sumatran orangutans, Bornean orangutans, and Sabah orangutans) and summary statistics were calculated (Table 2). No deviations from the neutral model were found using Tajima's (1989) test statistic of neutral evolution (*D*) or similar tests (Fu and Li 1993). The HKA test (Hudson et al. 1987), which detects deviations in polymorphism and divergence between two loci, was used to compare α -2 globin and the Xq13.3 region. Balancing selection at α -2 globin would increase levels of intraspecific diversity relative to Xq13.3, while allelic replacement would lower the level of diversity at θ -2 globin relative to Xq13.3. No significant disparity in the levels of polymorphism and divergence was found between these loci (all orangutans, $p = 0.66$; Bornean, $p = 0.86$; Sumatran, $p = 0.54$). The allelic replacement hypothesis was tested by ascertaining whether a subset of alleles had a disproportionately large or small proportion of the overall variation present in the data set. The likeliest case of allelic replacement in this data set was in Sumatra, where there was a lack of variation within their *MUT* alleles. A pattern consistent with neutral evolution was found by the haplotype test, showing that the *MUT* alleles in Sumatran orangutans did not undergo a selective sweep.

Deviations from neutrality can cause heterogeneity in the distribution of polymorphic sites among a group of DNA sequences (McDonald 1996, 1998). For example, due to hitchhiking, balancing selection can skew a region of DNA toward an excess of polymorphic sites, while a selective sweep can skew a region toward fewer polymorphic sites. Evidence of heterogeneity in the distribution of polymorphic and fixed differences in the (α -2 globin data set was

Table 2. Summary statistics

Group	N_{chroms}	S	θ_w	π	$R_{\text{(between sites)}}$	R_m sites
All	28	21	0.00378	0.00362	0.0157	(82;154) (154;546) (546;640) (640;1044)
Sumatra	8	13	0.00351	0.00413	0.0157	(82;154)
Borneo	20	12	0.00237	0.00247	0.0176	(154;546) (546;640)
Sabah	13	9	0.00203	0.00242	0.0131	(154;546) (546;640)

Note. N_{chroms} —number of chromosomes; S —number of segregating sites; θ_w —Watterson's θ ; π —nucleotide diversity; R —recombination parameter (Hudson 1987). θ_w , π , and R calculated per site. R_m sites—sites comprising the minimum number of recombination events (Hudson and Kaplan 1985).

Table 3. McDonald's (1996, 1998) tests for site heterogeneity

Ingroup	Outgroup	p value (R) ^a	Test ^b
All	Homo	0.040* (5)	K-S
	Gibbon	0.015* (6)	K-S
Sumatra	Homo	0.038* (12)	Avg. G
	Gibbon	0.027* (18)	K-S
Borneo	Homo	0.019* (10)	K-S
	Gibbon	0.019* (10)	K-S
	Homo	0.057 (8)	Avg. G
	Gibbon	0.026* (60)	Runs
		0.019* (10)	Var.
		0.049* (11)	Avg. G

^aProbability values calculated as described in Materials and Methods. R —recombination parameter resulting in the maximum p value.

^bK-S—Kolmogorov–Smirnov test statistic; Avg. G—mean sliding G test statistic; Var.—interval length variance; Runs—runs statistic.

* $p < 0.05$.

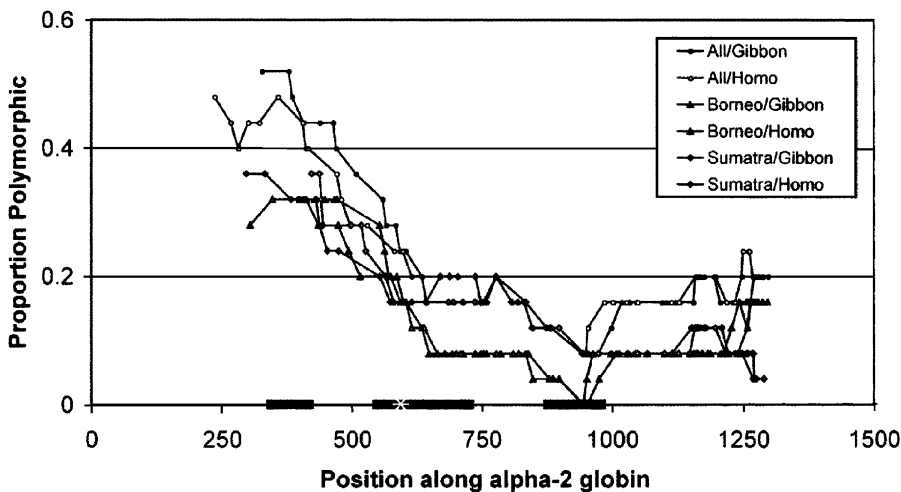


Fig. 3. Proportion of polymorphic sites, relative to all polymorphic and fixed sites. Y-axis—proportion of polymorphic sites; X-axis—nucleotide position, with black boxes indicating exons and an asterisk marking the approximate location of position 546. Sliding windows contain 25 total sites. Each line represents the data from a particular ingroup and outgroup; legend shown in the inset.

uncovered using multiple tests (Table 3). Significantly nonrandom distributions of fixed and polymorphic sites were found at the species and subspecies levels, using both human and gibbon as outgroups. No significant deviations were found in the Sabah subpopulation, possibly because the small number of polymorphisms affected the test's power. The significant deviations were recovered mainly with the Kolmogorov–Smirnov and mean sliding G statistics. These two test statistics are best at detecting a graded pattern where the ratio of polymorphism to divergence is different at the two ends of a gene

(McDonald 1998). This pattern was confirmed by a plot of the polymorphism to divergence ratio, which showed that the first half of α -2 globin had more polymorphism than the second half (Fig. 3). This was consistent with either a force having increased polymorphism in the 5' region, such as balancing selection, or a force that reduced variation in the 3' region, such as purifying selection. While these tests would not be significant after a Bonferroni correction, McDonald (1998) suggested that each test of heterogeneity could be considered independent, though p values between 0.05 and 0.017 should not be taken as

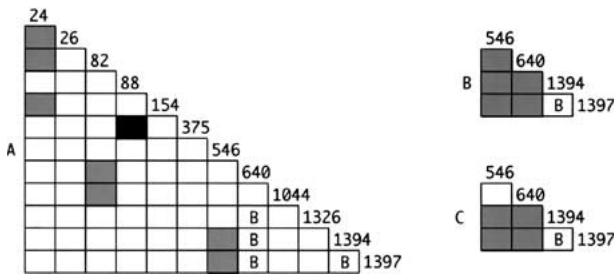


Fig. 4. Sites determined to be in significant linkage disequilibrium via Fisher's exact test. Gray denotes p values < 0.01 ; black denotes $p < 0.001$; B denotes significance after the Bonferroni correction. **A** Graph for the entire data set. **B** Graph for Bornean orangutans. **C** Graph for Sabah orangutans. No sites were found to be in significant linkage disequilibrium in the Sumatran data set.

definitive rejections of the equilibrium model. In the tests of α -2 globin, nearly all the p values were within this interval. There were two reasons why these tests were likely ascertaining a significant, though weak pattern. First, statistical significance was achieved by a nonrandom set of heterogeneity tests, those that specifically perform best at detecting a graded pattern. If these were spurious rejections, there would not have been a pattern in the heterogeneity tests. Second, the finding of a graded pattern was confirmed with an HKA test: when splitting the α -2 globin locus into two data sets of equal length, the halves had significantly different levels of polymorphism and divergence (all orangutans, $p = 0.048^*$; Bornean, $p = 0.089$; Sumatran, $p = 0.044^*$).

Within the Bornean subspecies a noteworthy pattern of linkage disequilibrium and recombination was found. Specifically, the lone nonsynonymous variant in the data set (position 546) and a synonymous variant (position 640) were in linkage disequilibrium, but also had experienced recombination (Table 2 and Fig. 4). Both of these variants were in linkage disequilibrium with a microsatellite in the 3' region. In Borneo, the *MUT* haplotypes mainly consisted of an A at position 546 most often associated with a T at site 640 and a six-repeat motif (A-T-6) (Fig. 2). The Bornean *WT* alleles had a G at position 546, which was most often associated with a G at position 640 and a GG5 microsatellite allele (G-G-GG5) (some G-T-5 haplotypes were also present). In Sumatra, the opposite states were associated at positions 546 and 640, and all Sumatran haplotypes had a five-repeat microsatellite allele (A-G-5 and G-T-5). One Bornean *MUT* haplotype (B-MUT-5) had the Sumatran A-G-5 haplotype but lacked Sumatran-typical variants at positions 1044 and 1106, suggesting that this haplotype was the result of a recombination or gene conversion event. A second Bornean haplotype (B-MUT-4) had the typical Borneo *MUT* combination at sites 546 and 640 (A-T), but with a five-repeat allele, possibly a homoplastic mutation.

The association of the nonsynonymous mutation (site 546) with different sites in the two orangutan subspecies suggested that this mutation might have occurred as independent point mutations in each orangutan subspecies. However, alternative explanations exist for the presence of this mutation in both Borneo and Sumatra, such as recombination or an intra- or interlocus gene conversion event. One possibility is that recombination or an intralocus gene conversion event placed the mutation onto two different haplotypic backgrounds in the last common ancestor of Bornean and Sumatran orangutans. Then, in each descendant subspecies, the opposite haplotypes became fixed while the other was lost. However, this model does not easily explain the presence of the Sumatran-typical substitutions that occur on both *MUT* and *WT* alleles (positions 1044 and 1106) (Fig. 2). Under this model, these subspecies-typical mutations must have spread to both alleles through gene conversion or recombination. A second hypothesis is an interlocus gene conversion event between the α -1 and the α -2 globin genes, as the point mutation at position 546 occurs at both loci in orangutans. If the mutation originated in the last common ancestor the orangutan subspecies at α -1 globin, two separate interlocus gene conversion events in each descendant subspecies could have independently transferred the G \rightarrow A mutation from α -1 to α -2 globin. This second scenario more parsimoniously explains the presence of the Sumatran-typical substitutions common to both *WT* and *MUT* alleles.

Discussion

The estimates for variation at orangutan α -2 globin are similar to those from another nuclear region sequenced in orangutans, Xq13.3 ($\pi = 2.65 \times 10^{-3}$ and $\theta_w = 2.39 \times 10^{-3}$) (Kaessmann et al. 2001). α -2 Globin is slightly more diverse than Xq13.3, which is expected due to its larger effective population size as an autosomal locus. Compared to 16 genes examined in humans (Przeworski et al. 2000), orangutan α -2 globin diversity is approximately 3.5 times greater than the diversity found at 5 autosomal loci and 4.75 times higher than at 11 X-linked loci. Though fewer Sumatran individuals were sampled, this subspecies has higher nucleotide diversity than Bornean orangutans. Higher polymorphism in Sumatra is found in most other orangutan population genetic data sets (Kaessmann et al. 2001; Noda et al. 2001; Muir et al. 2000; Warren et al. 2001; Zhang et al. 2001; Zhi et al. 1996) and may reflect past geological events and a higher historical population size in Sumatra (Muir et al. 2000). The α -2 globin data support this interpretation.

The polymorphism and divergence, haplotype patterns, and the deviations from HWE at this locus are consistent with different interpretations. The deviations from HWE in Borneo and Sabah are suggestive of a nonequilibrium pattern of evolution and may reflect selection for *MUT* homozygotes, selection against heterozygotes, inbreeding, and/or the Wahlund effect. Microsatellite analyses of Bornean orangutans also recover departures from HWE that are suggestive of inbreeding (Warren et al. 2000). In the case of α -2 globin, interpretation of the HWE deviation is further complicated by the presence of the alleles at both α -2 and α -1 globin (individuals homozygous at one α -globin gene may be effectively heterozygous at the phenotypic level). This departure may reflect the interaction of these two α -globin genes, as well as a third α -globin gene (Takenaka et al. 1993; Zimmer et al. 1980).

The heterogeneous distribution of polymorphism at the locus is consistent with two interpretations: a force increasing polymorphism in the first half of the gene or a force reducing variation in the second half. An elevation of polymorphism in the 5' region may have been caused by the hitchhiking effect on sites near to the nonsynonymous variant. Because there was no overall increase in intraspecific polymorphism found at α -2 globin relative to Xq13.3, this is not a straightforward case of balancing selection. Instead, these data may represent balancing selection in its early stages, complicated by the presence of multiple α -globin loci. In a case of young balancing selection, the expected buildup of variation or ancient coalescent dates may not have had the time to occur. Dates calculated for the ancestor of the orangutan α -2 globin alleles are similar to those estimated for other genes, supporting this interpretation (σ -2 globin, 2.2 million years ago [MYA; gibbon outgroup] and 1.5 MYA [human outgroup]; Xq13.3, 1.4–3.4 MYA [Kaessmann et al. 2001]; COII, 3.5 MYA [Ruvolo 1994]; ND5, 2.3 MYA [Zhang et al. 2001]; multiple loci, 1.7 MYA [Zhi et al. 1996]). A neutral alternative, incorporating population subdivision, could explain the deviation from HWE and lack of significant differences between α -2 globin and Xq13.3, but it does not easily explain the skew toward polymorphism in the first half of the gene. The separate geographical distribution of the α -2 globin haplotypes, which may be attributable to the spread of a favorable amino acid polymorphism following two gene conversion events, further supports the proposition that this gene may not follow a neutral pattern of evolution. Additional studies of the α -1 globin gene and other neighboring regions would aid in clarifying the evolutionary history of the orangutan α -globin region.

Although there is evidence for functional differences among orangutan hemoglobins (albeit limited),

these differences cannot be related to variation in malarial resistance. It is crucial to conduct further studies of orangutan population genetics including the association of their hemoglobin polymorphisms with *Plasmodium* incidence. If *Plasmodium* is the selective pressure mediating the evolution of orangutan α -globin region, this suggests that humans are not the only primates adversely affected by malaria. Other genetic studies bolster this idea, including work on the hominoid glycophorin genes, which are evolving under positive selection, possibly to evade *Plasmodium* parasites (Wang et al. 2003). If *Plasmodium* infection and disease is not limited to postagricultural *Homo sapiens*, this predicts the presence of ancient *Plasmodium* resistance alleles in the contemporary human gene pool. In addition, other nonhuman primate species that harbor *Plasmodium* may have molecular and behavioral adaptations to fight this parasite.

Acknowledgments. We thank the WCS–Harvard Orangutan Study for the Sabah orangutan DNA samples. For non-Sabah samples we thank Oliver Ryder and Leona Chemnick at the San Diego Zoo's Center for the Reproduction of Endangered Species (CRES)/Frozen Zoo, the Orangutan SSP, Yerkes RPC, and the Topeka Zoo. This work was supported by grants from the National Science Foundation (BCS-0097057), the Lindbergh Foundation, and the Harvard Anthropology Department to M.S. and M.R. Research Centers in Minority Institutions award RR-03037 from the National Center for Research Resources of the National Institutes of Health, which supports the infrastructure of the Anthropological Genetics Laboratory at Hunter College, is also acknowledged. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NCR/NIH. We also thank Andrew Berry, Dan Hartl, and David Pilbeam for comments on an early version of the manuscript and two anonymous reviewers for helpful comments and suggestions.

References

- Albital M, Cash FE, Peschle C, Liebhaber SA (1992) Development switch in the relative expression of the α -1-globin and α -2-globin genes in humans and in transgenic mice. *Blood* 79:2471–2474
- Allen SJ, O'Donnell A, Alexander NDE, Alpers MP, Peto TEA, Clegg JB, Weatherall DJ (1997) α^+ -Thalassemia protects children against disease caused by other infections as well as malaria. *Proc Natl Acad Sci USA* 94:14736–14741
- Ayala FJ, Escalante AA, Rich SM (1999) Evolution of *Plasmodium* and the recent origin of the world populations of *Plasmodium falciparum*. *Parassitologia* 41:55–68
- Bailey AD, Shen CC, Shen C-KJ (1997) Molecular origin of the mosaic sequence arrangements of higher primate α -globin duplication units. *Proc Natl Acad Sci USA* 94:5177–5182
- Barnicot NA, Jolly CJ (1966) Haemoglobin polymorphism in the orangutan and an animal with four major haemoglobins. *Nature* 210:640–642
- Bradley RD, Hillis DM (1997) Recombinant DNA sequences generated by PCR amplification. *Mol Biol Evol* 14:592–593

- Bruce EJ, Ayala FJ (1979) Phylogenetic relationships between man and the apes: Electrophoretic evidence. *Evolution* 33:1040–1056
- Bunn HP, Forget BG (1986) Hemoglobin: Molecular, genetic and clinical aspects. W.B. Saunders, Philadelphia
- Coatney RG, Collins WE, Warren M, Contacos PG (1971) The primate malarial. US Government Printing Office, Washington, DC
- Coluzzi M (1999) The clay feet of the malaria giant and its African roots: hypotheses and inferences about origin, spread and control of *Plasmodium falciparum*. *Parassitologia* 41:277–283
- Cronn R, Cedroni M, Haselkorn T, Grover C, Wendel JF (2002) PCR-mediated recombination in amplification products derived from polyploid cotton. *Theor Appl Genet* 104:482–489
- Curat M, Trabuchet G, Rees D, Perrin P, Harding RM, Clegg JB, Langaney A, Excoffier L (2002) Molecular analysis of the β -globin gene cluster in the Niokholo Mandenka population reveals a recent origin of the β^S Senegal mutation. *Am J Hum Genet* 70:207–223
- de Boer LEM, Meera Khan P (1982) Orangutan haemoglobins: a short review. In: de Boer LEM (ed) *The orangutan: Its biology and conservation*. Dr. W. Junk, The Hague, pp 119–134
- Donnelly MJ, Licht MC, Lehmann T (2001) Evidence for recent population expansion in the evolutionary history of the malaria vectors *Anopheles arabiensis* and *Anopheles gambiae*. *Mol Biol Evol* 18:1353–1364
- Escalante AA, Ayala FJ (1994) Phylogeny of the malarial genus *Plasmodium*, derived from ribosomal-RNA gene-sequences. *Proc Natl Acad Sci USA* 91:11373–11377
- Escalante AA, Ayala FJ (1995) Evolutionary origin of *Plasmodium* and other Apicomplexa based on rRNA genes. *Proc Natl Acad Sci USA* 92:5793–5797
- Escalante AA, Barrio E, Ayala FJ (1995) Evolutionary origin of human and primate malarial—Evidence from the circumsporozoite protein gene. *Mol Biol Evol* 12:616–626
- Escalante AA, Goldman IF, De Rijk P, De Wachter R, Collins WE, Qari SH, Lal AA (1997) Phylogenetic study of the genus *Plasmodium* based on the secondary structure-based alignment of the small subunit ribosomal RNA. *Mol Biochem Parasitol* 90:317–321
- Escalante AA, Freeland DE, Collins WE, Lal AA (1998) The evolution of primate malaria parasites based on the gene encoding cytochrome b from the linear mitochondrial genome. *Proc Natl Acad Sci USA* 95:8124–8129
- Flint J, Hill AVS, Bowden DK, Oppenheimer SJ, Sill PR, Serjeantson SW, Bana-Koiri J, Bhatia K, Alpers MP, Boyce AJ, Weatherall DJ, Clegg JB (1986) High frequencies of α -thalassaemia are the result of natural selection by malaria. *Nature* 321:744–750
- Fooden J (1994) Malaria in macaques. *Int J Primatol* 15:573–596
- Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* 133:693–709
- Hardison R, Riemer C, Chui DH, Huisman TH, Miller W (1998a) Electronic access to sequence alignments, experimental results, and human mutations as an aid to studying globin gene regulation. *Genomics* 47:429–437
- Hardison RC, Chui DH, Riemer CR, Miller W, Carver MF, Molchanova TP, Efremov GD, Huisman TH (1998b) Access to a syllabus of human hemoglobin variants (1996) via the World Wide Web. *Hemoglobin* 22:113–127
- Hudson RR, Kreitman M, Aguadé M (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153–159
- Hudson RR, Bailey K, Skarecky D, Kwiatowski J, Ayala FJ (1994) Evidence for positive selection in the superoxide dismutase (*Sod*) region of *Drosophila melanogaster*. *Genetics* 136:1329–1340
- Huisman THJ, Carver MFH, Efremov GD (1996) A syllabus of human hemoglobin variants. Sick Cell Anemia Foundation, Augusta, GA
- Huser H-J (1970) Atlas of comparative primate hematology. Academic Press, New York
- Judo MS, Wedel AB, Wilson C (1998) Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Res* 26:1819–1825
- Kaessmann H, Wiebe V, Weiss G, Pääbo S (2001) Great ape DNA sequences reveal a reduced diversity and an expansion in humans. *Nat Genet* 27:155–156
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular Evolutionary Genetics Analysis software. *Bioinformatics* 17:1244–1245
- Lauer J, Shen CK, Maniatis T (1980) The chromosomal arrangement of human α -like globin genes: sequence homology and α -globin gene deletions. *Cell* 20:119–130
- Lehmann H, Kynoch PAM (1976) Human haemoglobin variants and their characteristics. North-Holland, Amsterdam
- Liebhaber SA, Goossins M, Kan YW (1981) Homology and concerted evolution at the α -1 and α -2 loci of human α -globin. *Nature* 290:26–29
- Livingstone FB (1958) Anthropological implications of sickle-cell gene distribution in West Africa. *Am Anthropol* 60:533–562
- Marks J, Shaw J-P, Shen C-KJ (1986) The orangutan adult α -globin gene locus: Duplicated functional genes and a newly detected member of the primate α -globin gene family. *Proc Natl Acad Sci USA* 83:1413–1417
- McDonald JH (1996) Detecting non-neutral heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. *Mol Biol Evol* 13:253–260
- McDonald JH (1998) Improved tests for heterogeneity across a region of DNA sequence in the ratio of polymorphism to divergence. *Mol Biol Evol* 15:377–384
- Mu JB, Duan JH, Makova KD, Joy DA, Huynh CQ, Branch OH, Li WH, Su XZ (2002) Chromosome-wide SNPs reveal an ancient origin for *Plasmodium falciparum*. *Nature* 418:323–326
- Muir CC, Galdikas BMF, Beckenbach AT (2000) mtDNA sequence diversity of orangutans from the islands of Borneo and Sumatra. *J Mol Evol* 51:471–480
- Noda R, Kim CG, Takenaka O, Ferrell RE, Tanoue T, Hayasaka I, Ueda S, Ishida T, Saitou N (2001) Mitochondrial 16S rRNA sequence diversity of hominoids. *J Hered* 92:490–496
- Pääbo S, Irwin DM, Wilson AC (1990) DNA damage promotes jumping between templates during enzymatic amplification. *J Biol Chem* 265:4718–4721
- Parsch J, Meiklejohn CD, Hartl DL (2001) Patterns of DNA sequence variation suggest the recent action of positive selection in the janus-ocnus region of *Drosophila simulans*. *Genetics* 159:647–657
- Perkins SL, Schall JJ (2002) A molecular phylogeny of malarial parasites recovered from cytochrome b gene sequences. *J Parasitol* 88:972–978
- Przeworski M, Hudson RR, Di Rienzo A (2000) Adjusting the focus on human variation. *Trends Genet* 16:296–302
- Qari SH, Shi YP, Pieniazek NJ, Collins WE, Lal AA (1996) Phylogenetic relationship among the malaria parasites based on small subunit rRNA gene sequences: monophyletic nature of the human malaria parasite, *Plasmodium falciparum*. *Mol Phylogenet Evol* 6:157–165
- Rich SM, Ayala FJ (2000) Population structure and recent evolution of *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 97:6994–7001
- Rich SM, Light MC, Hudson RR, Ayala FJ (1998) Malaria's Eve: evidence of a recent population bottleneck throughout the

- world populations of *Plasmodium falciparum*. Proc Natl Acad Sci USA 95:4425–4430
- Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15:174–175
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497
- Ruvolo M, Pan D, Zehr S, Goldberg T, Disotell TR, von Dornum M (1994) Gene trees and hominoid phylogeny. Proc Natl Acad Sci USA 91:8900–8904
- Sack JS, Andrews LC, Magnus KA, Hanson JC, Rubin J, Love WE (1978) Location of amino acid residues in human deoxy hemoglobin. Hemoglobin 2:153–169
- Saunders MA, Hammer MF, Nachman MW (2002) Nucleotide variability at G6pd and the signature of malarial selection in humans. Genetics 162:1849–1861
- Skinner MF, Hopwood D (2004) Hypothesis for the causes and periodicity of repetitive linear enamel hypoplasia in large, wild African (*Pan troglodytes* and *Gorilla gorilla*) and Asian (*Pongo pygmaeus*) apes. Am J Phys Anthropol 123:216–235
- Smith BD (1995) Emergence of agriculture. WH Freeman, New York
- Sullivan B, Nute PE (1968) Structural and functional properties of polymorphic hemoglobins from orangutans. Genetics 58:113–124
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595
- Takenaka A, Udono T, Miwa N, Varavudhi P, Takenaka O (1993) High frequency of triplicated α -globin genes in tropical primates, crab-eating macaques (*Macaca fascicularis*), chimpanzees (*Pan troglodytes*), and orangutans (*Pongo pygmaeus*). Primates 34:55–60
- Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G, Drousiotou A, Dangerfield B, Lefranc G, Loiselet J, Piro A, Stoneking M, Tagarelli A, Tagarelli G, Touma EH, Williams SM, Clark AG (2001) Haplotype diversity and linkage disequilibrium at human G6PD: Recent origin of alleles that confer malarial resistance. Science 293:455–462
- Verrelli BC, McDonald JH, Argyropoulos G, Destro-Bisol G, Froment A, Drousiotou A, Lefranc G, Helal AN, Loiselet J, Tishkoff SA (2002) Evidence for balancing selection from nucleotide sequence analyses of human G6PD. Am J Hum Genet 71:1112–1128
- Volkman SK, Barry AE, Lyons EJ, Nielsen KM, Thomas SM, Choi M, Thakore SS, Day KP, Wirth DF, Hartl DL (2001) Recent origin of *Plasmodium falciparum* from a single progenitor. Science 293:482–484
- Wang HY, Tang H, Shen CK, Wu CI (2003) Rapidly evolving genes in human. I. The glycoporphins and their possible role in evading malaria parasites. Mol Biol Evol 20:1795–804
- Warren KS, Nijmian IJ, Lenstra JA, Swan RA, Heriyanto, den Boer M (2000) Microsatellite DNA variation in Bornean orangutans (*Pongo pygmaeus*). J Med Primatol 29:57–62
- Warren KS, Verschoor EJ, Langenhuijzen S, Heriyanto, Swan RA, Vigilant L, Heeney JL (2001) Speciation and intraspecific variation of Bornean orangutans, *Pongo pygmaeus pygmaeus*. Mol Biol Evol 18:472–480
- Waters AP, Higgins DG, McCutchan TF (1991) *Plasmodium falciparum* appears to have arisen as a result of lateral transfer between avian and human hosts. Proc Natl Acad Sci USA 88:3140–3144
- Waters AP, Higgins DG, McCutchan TF (1993) Evolutionary relatedness of some primate models of *Plasmodium*. Mol Biol Evol 10:914–923
- Williams TN, Maitland K, Bennett S, Ganczakowski M, Peto TEA, Newbold CI, Bowden DK, Weatherall DJ, Clegg JB (1996) High incidence of malaria in α -thalassaemic children. Nature 383:522–524
- Wolfe ND (1999) Pathogen evolution and exchange in Bornean orangutans. PhD thesis, Harvard School of Public Health, Cambridge, MA
- Wolfe ND, Karesh WB, Kilbourn AM, Cox-Singh J, Bosi EJ, Rahman HA, Prosser AT, Singh B, Andau M, Spielman A (2002) The impact of ecological conditions on the prevalence of malaria among orangutans. Vector-Borne Zoonot 2:97–103
- Young NM, MacLachy L (2004) The phylogenetic position of *Morotopithecus*. J Hum Evol 46:163–184
- Zhang YW, Ryder OA, Zhang YP (2001) Genetic divergence of orangutan subspecies (*Pongo pygmaeus*). J Mol Evol 52:516–526
- Zhi L, Karesh WB, Janczewski DN, Frazier-Taylor H, Sajuthi D, Gombek F, Andau M, Martenson JS, O'Brien SJ (1996) Genomic differentiation among natural populations of the orangutan (*Pongo pygmaeus*). Curr Biol 6:1326–1336
- Zimmer EA, Martin SL, Beverley SM, Kan YW, Wilson AC (1980) Rapid duplication and loss of genes coding for the α chains of hemoglobin. Proc Natl Acad Sci USA 77:2158–2162
- Zimmerman PA, Wooley I, Masinde GL, Miller SM, McNamara DT, Hazlett F, Mgone CS, Alpers MP, Genton B, Boatman BA, Kazura JW (1999) Emergence of *FY*A^{null}* in a *Plasmodium vivax*-endemic region of Papua New Guinea. Proc Natl Acad Sci USA 96:13973–13977