Published in final edited form as:
*Infect Genet Evol*. 2007 March ; 7(2): 180–187.

# A comparative study of the genetic diversity of the 42 kDa fragment of the merozoite surface protein 1 in *Plasmodium falciparum* and *P. vivax*

**M. Andreina Pacheco**[1], **Amanda C. Poe**[2], **William E. Collins**[2], **Altaf A. Lal**[2,†], **Kazuyuki Tanabe**[3], **Simon K. Kariuki**[4], **Venkatachalam Udhayakumar**[2], and **Ananias A. Escalante.**[1,*]

1 *School of Life Sciences, Arizona State University, Tempe, Arizona*

2 *Malaria Branch, Division of Parasitic Diseases, Centers for Disease Control and Prevention, Chamblee, Georgia*

3 *International Research Center of Infectious Diseases, Research Institute for Microbial Diseases, Osaka University, Osaka, Japan*

4 *Center for Vector Biology and Control Research, Kenya Medical Research Institute, Kisumu, Kenya*

## Abstract

We investigated the genetic diversity of the 42 kDa fragment of the merozoite surface protein 1 (MSP-1) antigen in *Plasmodium falciparum* and *P. vivax*, as well as in non-human primate malarial parasites. This fragment undergoes a proteolytic cleavage generating two fragments of 19 kDa (MSP-$1_{19}$) and 33 kDa (MSP-$1_{33}$) that are critical in erythrocyte invasion. We found that overall the MSP-$1_{33}$ fragment exhibits greater genetic diversity than the MSP-$1_{19}$ regardless of the species. We have found evidence for positive natural selection only in the human malaria parasites by comparing the rate of non-synonymous versus synonymous substitutions. In addition, we found clear differences between the two major human malaria parasites. In the case of *P. falciparum*, positive natural selection is acting on the MSP-$1_{19}$ region while the MSP-$1_{33}$ is neutral or under purifying selection. The opposite pattern was observed in *P. vivax*. Our results suggest different roles of this antigen in the host-parasite immune interaction in each of the major human malarial parasites.

### Keywords

Malaria; Merozoite; MSP-1; MSP-1 19 kDa; Genetic Diversity; Vaccine; *Plasmodium*

## Introduction

The malaria burden is particularly high in sub-Saharan Africa where *Plasmodium falciparum* is predominant. However, malaria “out of Africa” is characterized by the presence of *P. vivax*, the second most important malaria parasite in terms of its morbidity. Although there are clear biological and genetic differences between these two parasites (Coatney *et al*.,

*Corresponding Author: Ananias A. Escalante, PhD School of Life Sciences, Arizona State University P. O. Box 874501, Tempe, Arizona 85287-4501 Phone: 1-480-9653739, Fax: 1-480-965-6899 E-mail: ananias.escalante@asu.edu.
†Present address: Health Attaché and HHS Regional Representative for South Asia, U.S. Embassy, New Delhi, India

1971), they overlap in their geographic distribution and there is increasing evidence for their interaction (Snounou and White, 2004).

Among the antigens currently under consideration in malaria vaccine formulations, one of the most promising candidates is the major merozoite surface protein 1 (MSP-1) (Good *et al.*, 2004). The MSP-1 antigen is expressed as a large protein of 190–200 kDa on the parasite surface (Holder *et al*., 1982). This precursor undergoes two steps of proteolytic cleavage during the merozoite maturation. First, it is cleaved into four major fragments of 83, 30, 38 and 42 kDa (further referred to as MSP-$1_{83}$, MSP-$1_{30}$, MSP-$1_{38}$, and MSP-$1_{42}$ ) then, before erythrocyte invasion the MSP-$1_{42}$ fragment undergoes a second cleavage resulting in the generation of 33 and 19 kDa (MSP-$1_{33}$ and MSP-$1_{19}$) fragments where the latter remain on the merozoite surface during invasion.

## Plasmodium spp

MSP-1 exhibits extensive genetic polymorphism (Tanabe et al., 1987, Putaporntip et al., 2002) that appears to be maintained by positive natural selection in *P. falciparum* (Hughes, 1991; Escalante *et al*., 1998; Conway *et al*., 2000) and *P. vivax* (Putaporntip *et al*., 2006). Similar observations have been made about other malarial vaccine antigens (see Escalante *et al.* 2004) on which the host immune system is considered the driving selective force that allows for the accumulation and frequent switch of suitable mutations in the parasite population. Under this scenario, mutations are maintained longer in the parasite population than expected if genetic drift were the sole process acting on the genetic polymorphism.

The conclusion that positive selection maintains the genetic diversity of genes encoding malarial antigens is supported, among others lines of evidence, by the observation in *P. falciparum* that non-synonymous nucleotide substitutions (those that change the amino acid) are more common than synonymous substitutions (mutations that do not change the amino acid) (Hughes and Hughes 1995, Escalante *et al.* 1998, Escalante *et al.* 2004). Since natural selection acts on phenotypic differences, an excess of non-synonymous substitutions over synonymous is considered evidence that natural selection is favoring the maintenance of genetic polymorphism.

In the case of *Plasmodium* spp. MSP-1, most of the genetic diversity analyses have subdivided the gene into blocks (segments) based on their level of genetic diversity but not using any other biological criteria (Tanabe *et al*., 1987; Putaporntip *et al*., 2002; Putaporntip *et al*., 2006); however, few studies have been done considering the proteolytic fragments as functional units (Escalante *et al*., 1998).

The MSP-$1_{42}$ and MSP-$1_{19}$ fragments have received special attention in *P. falciparum* as part of vaccine formulations given that they are relatively conserved and antibodies against these fragments inhibit the parasite invasion into the red blood cells (Yang *et al*., 1999; Stanisic *et al*., 2004). In addition, the critical role of the MSP-$1_{19}$ fragment in the erythrocyte invasion is conserved even among distantly related species (O' Donnell *et al*., 2001).

An important characteristic of *P. vivax* is that it invades reticulocytes, a process that is mediated by specific proteins such as the reticulocyte binding proteins and Duffy receptor (Gallinski *et al*., 1992; Chitnis and Miller, 1994). However, MSP-1 in *P. vivax* also appears to play an important role in this process (Rodriguez *et al*. 2002; Espinosa *et al*., 2003; Han *et al*., 2004; Sachdeva *et al*., 2004). Indeed, peptides with high specific binding activity (HSBA) to reticulocytes have been found in the $MSP1_{33}$ (Espinosa *et al*., 2003; Rodriguez *et al*., 2002).

This investigation aims to compare the genetic diversity of the MSP-$1_{42}$ in *Plasmodium* spp. focusing on *P. falciparum* and *P. vivax.* We have analyzed 120 sequences of the MSP-$1_{42}$ of

*P. falciparum* and 75 sequences of the homologous region in *P. vivax*, and we have explored the genetic diversity of the MSP-$1_{33}$ and MSP-$1_{19}$ fragments. In the case of *P. vivax* we further explored its genetic diversity by comparing it with the homologous regions in primate malarial parasites that are closely related to *P. vivax* (Escalante *et al.*, 2005). Although we find evidence that positive natural selection is acting on the observed polymorphism in MSP-$1_{42}$, it operates differently in each of the two major human malarial parasites. We conclude that inferences made about *P. falciparum* MSP-1 cannot simply be "translated" into *P. vivax*.

## Materials and Methods

The gene encoding the 42kDa fragment of MSP-1 or MSP-$1_{42}$ was amplified by polymerase chain reaction (PCR). The primers forward- GAA TGA TAT TCC TAA GAA GTT AGA GG and reverse- GAT AGA TTA TTT AAT AAG AAA AAA GAA CTT GGC CAA GAC AAA ATG C were used to amplify the partial *P. falciparum* 3' sequences. The PCR conditions for *P. falciparum* were: a partial denaturation at 94 °C for 1 minute and 30 cycles with 1 minute at 94 °C, 1 minute at °50 and 3 minutes extension at °72. A final extension of 3 minutes was added in the last cycle. The primers forward-GAC CAA GTA ACA ACG GGA G and reverse-CAA AGA GTG GCT CAG AAC C were used for *P. vivax*, *P. cynomolgi*, *P. inui*, and *P. knowlesi*. In the case of *P. fragile* we used the forward primer GAC CAA GTA ACA ACG GG. The PCR conditions for *P. vivax* and non human primate malarias were: a partial denaturation at 94 °C for 3 minute and 35 cycles with 1 minute at 94 °C, 45' at 50-58 °C and 2 minutes extension at 72 °C, a final extension of 10 minutes was added in the last cycle.

The amplified product was purified, cloned using the pGEM-TEasy Vector System I from Promega (USA), and sequenced. Both strands were sequenced from at least two clones. The alignment was performed using ClustalW version 1.7 with manual editing using the alignment reported by Miller *et al.* (1993) in the case of *P. falciparum* and those reported by Putaporntip *et al*. (2002 (2006) in the case of *P. vivax* and related species.

In the case of *P. falciparum*, we sequenced the MSP-1 42 kDa in 34 isolates from Asembo Bay, western Kenya in this investigation. In addition, a total of 20 isolates (5 from India, 9 from Venezuela, and 6 from Thailand) were sequenced for the 3' end. We used in our investigation prior published sequences (Chang *et al.*, 1988; Qari *et al.*, 1998; Jangwutiwes *et al.*, 1992; Jangwutiwes *et al.*, 1993; Tanabe *et al.*, 2004) and unpublished sequences under the accession numbers U20726-U20733 and U20653-U20656. A total of 120 MSP-$1_{42}$ sequences were considered in our analyses. In addition, we included 55 sequences of the MSP-$1_{19}$ reported in the literature (Kaneko *et al*., 1997; Kumar *et al*., 2005) and unpublished sequences under the accession numbers AF29507 to AF29537 in order to obtain a complete picture of the MSP-$1_{19}$ alleles that have been reported.

In the case of *P. vivax*, we report 5 sequences from laboratory isolates (Rio Meta, Sumatra I, Indonesia I, Mauritania I, and Vietnam II) and used the sequences reported in the literature (Putaporntip *et al*., 2000; Putaporntip *et al*., 2002) for a total of 75 sequences. In addition, we included 10 sequences from different isolates of *P. cynomolgi* (the sequence AY869723 from the GenBank together with new sequences from the strains B strain, Berok, Cambodian, Ceylonensis, Gombok, Mulligan, PT1, PT2, and RO), 15 sequences from isolates of *P. inui* (Celebes I and II, Hackeri, Hawking, Leaf Monkey I and II, Leucosphyrus, Mulligan, N-34, OS, Perak, Perlis, Philippine, Taiwan I and II), a sequence of *P. knowlesi* (Hackery strain), *P. hylobati* (parasite from gibbons), and *P. fragile* (Nilgiri strain). Information about the biology of these species and the origin of the isolates can be found elsewhere (Coatney *et al*. 1971). All the primate malaria strains were provided by the Centers for Disease Control and Prevention. The sequences reported in this study are deposited in the GenBank with the accession numbers DQ907617-DQ907702.

### Statistical analysis

We estimate genetic polymorphism by using the parameter π, which estimates the average number of substitutions between any two sequences. The average number of synonymous (Ds) and nonsynonymous substitutions (Dn) between a pair of sequences was investigated to explore the effect of natural selection. The average numbers of synonymous and nonsynonymous substitutions are estimated using two methods: Nei and Gojobori's method (1986) with the Jukes and Cantor correction, and the Li's method (1993) as implemented in the MEGA program (Kumar *et al.*, 2001). We estimated the difference between Ds and Dn, its standard deviation was calculated using bootstrap with 1000 pseudo-replications for Ds and Dn, as well as a two tail Z test on the difference between Ds and Dn (Nei and Kumar 2000). The null hypothesis is that Ds = Dn; thus we assumed as null hypothesis that the observed polymorphism was neutral.

The Tajima's D statistic and F* from Fu and Li were estimated for testing the hypothesis that the allele frequency spectrum is compatible with the neutral model (Tajima, 1989; Fu and Li, 1993). Under the neutral model, Tajima's D and F* are approximately equal to zero, thus any deviation from zero would indicate a departure from neutrality in the allele frequency spectrum.

Evidence for recombination was assessed by using the Rm parameter that estimates the minimum number of recombination events in the history of the sample. Rm is obtained using the four-gamete test (Hudson and Kaplan 1985) and, as the name of the parameter indicates, it is a conservative estimate of the number of recombination events.

In the case of *P. vivax* and related non-human primate malarial parasites, the gene genealogy of the MSP-$1_{42}$ alleles was determined by using the Neighbor Joining (Saitou and Nei, 1987) method with the Tamura-Nei model. The reliability of the nodes in the NJ tree was assessed by the bootstrap method with 1,000 pseudo-replications. The genealogy was estimated using the MEGA program (Kumar *et al.*, 2001). The assumption of neutrality was also tested in *P. vivax* MSP-1 by using the McDonald and Kreitman test (McDonald and Kreitman 1991), which compares the intra-and interspecific number of synonymous and nonsynonymous sites; significance was assessed by using a Fisher's exact test for the 2x2 contingency table as implemented in the programs DNAsp version 4.0 (Rosas *et al.* 2003 ). In this analysis we compare *P. vivax* with *P. cynomolgi* and *P. inui* (see below).

## Results

Table 1 shows the genetic diversity found in the MSP-$1_{42}$ fragments in *P. falciparum* and *P. vivax*. Overall, the genetic diversity of *P. falciparum* is twice that observed in *P. vivax* (π of 0.05042 vs. 0.02184). Analysis of the genetic diversity of the MSP-$1_{33}$ and MSP-$1_{19}$ fragments confirmed previous observations that the MSP-$1_{19}$ fragment is more conserved than the MSP-$1_{33}$ fragment (Table 1) in both human malarial parasites. *P. vivax* MSP-$1_{19}$ has only one polymorphic site while in *P. falciparum* the substitutions are concentrated in five residues within the epidermal growth factor like domains (EGF). In an extended alignment that included all the MSP-$1_{19}$ sequences reported in the literature at the time of this study (n=175); we found 11 alleles reported based in these five residues, among them, there are four common alleles that have a worldwide distribution: E-KNG-L (n=54), E-TSR-L (n=41), Q-KNG-F (n=20), Q-KNG-L (n=33). It is worth noting that some alleles, although reported in low frequency, have been found in two continents; such are the cases of E-KNG-F (n=8 reported in India and Kenya), E-KSR-L (n=4 reported in Kenya, South Africa, and Vanuatu), and Q-TSR-L (n=3 reported in India and Papua New Guinea). The allele E-TSG-L (n=9) has been reported three times in India (including this study) and is the one observed in *P. reichenowi*, the most closely related species to *P. falciparum* found in chimpanzees (Coatney *et al*. 1971).

We found two recombination-convergent events using the Rm method (Hudson and Kaplan, 1985); these events are illustrated using the relative positions of the residues in the allele E-KNG-L, specifically between the position held by the amino acids E and K (separated by 138 bp) and between the positions filled by amino acids K and G (separated by 30 bp). Recombination events have been previously reported in MSP-$1_{19}$ (Qari *et al*., 1998).

In order to explore the role of natural selection we further analyzed the genetic polymorphism in the MSP-$1_{42}$ as a unit by estimating the number of synonymous (Ds) and non-synonymous (Dn) substitutions per site estimated by the Nei and Gojobori method with the Jukes and Cantor correction. When this comparison is made, both parasites exhibit opposite patterns: MSP-$1_{42}$ in *P. falciparum* shows more synonymous than non-synonymous substitutions while the homologous region in *P. vivax* shows more non-synonymous than synonymous substitutions. In both cases the differences are significant with a Z test (Kumar and Nei, 2000) (Table 1). The Li's method gives identical results. We explore departure from neutrality by using the Tajima's D test (Tajima, 1989) and F* test (Fu and Li, 1993). These tests should to be used with caution since they aim to detect departures from a neutral panmictic population, an assumption that is violated by these geographically and temporally spaced samples. Nevertheless, we used them to explore the distribution of haplotypes in our samples as was used previously to compare *P. vivax* and *P. knowlesi* (Putaporntip *et al.*, 2006). These tests could not detect departure from neutrality in *P. falciparum*, although they did so in *P. vivax* when the complete MSP-$1_{42}$ was considered as a unit.

We explored the diversity in the MSP-$1_{33}$ and MSP-$1_{19}$ fragments separately by comparing the number of synonymous and non-synonymous substitutions in each species. In the case of the MSP-$1_{33}$ of *P. falciparum* there are more synonymous than non-synonymous substitutions ($P<0.05$) (Table 1), while the contrary was observed in the MSP-$1_{19}$ where there are more non-synonymous than synonymous substitutions ($P<0.05$). These results suggest that while the MSP-$1_{19}$ is under positive selection in *P. falciparum*, the MSP-$1_{33}$ is under purifying selection; that is, natural selection favors the maintenance of amino acid polymorphism in the MSP-$1_{19}$ while it holds back the rate of amino-acid polymorphism in the MSP-$1_{33}$. Differences between the MSP-$1_{33}$ and MSP-$1_{19}$ were also observed by using the Tajima's D and F* tests (Table 1): there is not a departure from neutrality in the MSP-$1_{33}$ while the MSP-$1_{19}$ polymorphism rejects the expectation under the neutral model. Although the significance level by the Tajima's D test is weak for MSP-$1_{19}$ ($0.05<P<0.1$), there is almost no synonymous variation, substantiating a departure from the neutrality in this region. It is important to notice that the Tajima's D and F* tests have a negative value indicating that there is an excess of low frequency variants in the sample (Table 1).

In the case of *P. vivax* the pattern is the opposite. There are more non-synonymous than synonymous substitutions in the MSP-$1_{33}$ while there is almost no variation in the MSP-$1_{19}$ (Table 1). The polymorphism in the *P. vivax* MSP-$1_{33}$ is not evenly distributed. Indeed, there is a region of 105 bp out of 848 bp in MSP-$1_{33}$ (35 amino acids) where a clear excess of nonsynonymous versus synonymous substitution is observed driving the overall MSP-$1_{33}$ results. In addition, there is a departure from neutrality in the MSP-$1_{33}$ when the Tajima's D and F* tests are applied. However, contrasting with *P. falciparum*, the value of the test is positive as the result of an excess of variants in intermediate frequencies.

We further explore the hypothesis that positive selection is acting on the *P. vivax* MSP-$1_{33}$ fragment by comparing it with its closely related non-human primate malarial parasites (Escalante *et al*., 2005). The genealogy of the MSP-$1_{42}$ fragments from the species reported in this study is depicted in figure 1. *P. cynomolgi* appears as sister taxa of *P. vivax*; however, this clade does not have strong support. *P. cynomolgi* strains are subdivided into two clear clades; no evidence for allele families could be observed with this fragment. *P. inui* and *P.*

*hylobati* are closely related as previously reported (Escalante *et al*., 2005b). The close relationship of these two species was further supported by the presence of a repetitive sequence in the MSP-$1_{33}$ fragment. Specifically, a motif with the residues NE*QEEI* is inserted in some of the *P. inui* isolates while *P. hylobati* has the residues NE*QEEI*KIR*QEEI*. We also found an insertion in *P. knowlesi* that emerged as a duplication of the motif *INNCQIEK* conserved in *P. inui* and *P. vivax* (figure 2). Given the lack of resolution of the phylogeny using this region, we used both *P. cynomolgi* and *P. inui* for comparison with *P. vivax*.

Table 2 shows the basic statistics for the MSP-$1_{42}$ in these two non-human primate malarial parasites. As in the cases of the human parasites, the MSP-$1_{33}$ fragment is more diverse than the MSP-$1_{19}$. However, in the case of the non human primate malarias, there is no excess of non-synonymous substitutions over synonymous substitutions in the MSP-$1_{42}$ as a unit or considering the MSP-$1_{33}$ and MSP-$1_{19}$ fragments separated. Thus, by comparing the rate of non-synonymous versus synonymous substitutions we could not detect evidence for positive selection acting on *P. cynomolgi* or *P. inui* MSP-$1_{42}$. An identical pattern can be observed in *P. knowlesi* when the two complete MSP-$1_{42,}$ the one reported in this investigation and the one available in the literature (Putaporntip *et al*., 2006) are compared, specifically Ds = 0.04275 and Dn = 0.00240 for MSP-$1_{42.}$

We then analyzed the genetic diversity of *P. vivax* MSP-$1_{42}$ by using the McDonald and Kreitman test (McDonald and Kreitman, 1991) and compared it with both *P. cynomolgi* and *P. inui* samples. In the case of the complete 42Kda, there was an overall excess of non-synonymous over synonymous in the *P. vivax* polymorphism when compared with *P. cynomolgi* ($p < 0.05$ using a Fisher's exact test). Similar results were found with *P. vivax* and *P. inui* ($p < 0.001$ using a Fisher's exact test). In both cases, the significance of the MK test was explained by an excess of amino acid replacements in the polymorphism of the *P. vivax* MSP-$1_{33}$. It is worth noting that no departure from neutrality was found when only MSP$1_{19}$ was considered. It is also important to emphasize that no departure from neutrality was observed when *P. cynomolgi* and *P. inui* were compared considering the MSP-$1_{42}$ as a unit, or separating it into the MSP-$1_{33}$ and MSP-$1_{19}$ fragments.

## Discussion

The available data, mostly derived from *P. falciparum*, indicate the importance of the antibody response against block 2 (located in the 83 kDa or MSP-$1_{83}$) and the MSP-$1_{42}$ fragments in developing protective immunity. In this study, we have described the selective forces operating on the polymorphism observed in the MSP-$1_{42}$ fragment in the two major human malaria parasites. We have shown how the MSP-$1_{33}$ and MSP-$1_{19}$ fragments are under different selective pressures in each of the major human malarial parasites by using the rate of non-synonymous versus synonymous substitutions.

In the case of *P. falciparum*, the polymorphism in MSP-$1_{33}$ appears to be neutral or under purifying selection while the polymorphism in MSP-$1_{19}$ is under positive selection. In this case, our results are consistent with immunologic evidence suggesting that the MSP-$1_{19}$ but not MSP-$1_{33}$ elicits a protective immune response, though the latter being highly immunogenic (Ahlborg *et al.*, 2002). Positive selection has been previously proposed as an important mechanism in maintaining the *P. falciparum* MSP-1 polymorphism in the form of balancing selection (Hughes, 1991; Conway *et al*., 2000); that is, natural selection maintains genetic polymorphism for a longer time than expected under a scenario where only genetic drift is acting. A polymorphism under balancing selection is expected to have an excess of alleles in intermediate frequencies, a pattern that translates into positives Tajima's D and F* tests. In the case of MSP-$1_{19}$, however, there is an excess of alleles in low frequency as evidenced by significant and negative values of the Tajima's D and F* tests, not consistent with balancing

selection. This could be the result of several factors. First, we found four alleles that are particularly common while several others are found in low frequency in our sample; low frequency alleles that are found even in different continents suggest an artifact due to a poor sampling effort. Indeed, lack of appropriate sampling could generate negative Tajima's D tests as a result of several sub-populations being analyzed together (Hammers *et al*., 2003). A second alternative is that a limited number of alleles are increasing in frequency, a scenario expected under a population expansion which coincides with the results reported for mitochondrial data (Joy *et al*., 2003).

Nevertheless, if the population demographic history and inappropriate sampling were the only factors leading to this result (significant and negative Tajimas's D and F* tests), then the MSP-$1_{33}$ should have shown a similar trend. The Tajimas's D and F* tests for MSP-$1_{33}$ are not only non significant but also have an opposite sign. Interestingly, the MSP-$1_{33}$ also shows more synonymous than non-synonymous substitutions. Therefore, we propose that the negative Tajimas's D and F* tests, together with the excess of non-synonymous over synonymous substitutions in MSP-$1_{19}$, are the result of directional selection, that is, there are few MSP-$1_{19}$ alleles increasing in frequency because they are positively selected.

Although the immune response against *P. falciparum* MSP-$1_{19}$ is still under intense investigation, there is evidence suggesting that fine specificity rather than prevalence could be an important factor in the observed immune reactivity (Okech *et al*., 2004). Indeed, only partial cross-reactivity has been found in holoendemic areas among the most common MSP-$1_{19}$ alleles (Udhayakumar *et al*., 1995; Shi *et al*., 1996; John *et al*., 2004). It has been also shown that immunity against MSP-$1_{19}$ in *P. falciparum* has a short lifespan to the extent that its elicited antibody responses allow detecting differences in local transmission (Drakeley *et al*., 2005). Therefore, the pattern in the genetic polymorphism of MSP-$1_{19}$ could be the result of differences of the most common alleles in their specificity and/or life spans of their elicited immune responses when compared with the less frequent MSP-$1_{19}$ alleles, differences that give them a selective advantage favoring their transmission.

Our hypothesis that directional selection is operating on MSP-$1_{19}$ does not contradict previous claims for balancing selection since they are well supported by the extensive divergence observed in MSP-$1_{83}$, MSP-$1_{30}$, and MSP-$1_{38}$ fragments allowing the identification of two very distinctive allele families (Tanabe *et al*., 1987) that have been found to be an ancient polymorphism (Hughes, 1991; Polley *et al*., 2005) as well as evidence derived from population base studies of the MSP-$1_{83}$ (Conway *et al*., 2000; Takala *et al*., 2006). Indeed such divergent allele families are not observed when only the MSP-$1_{19}$ is considered.

In the case of *P. vivax*, however, the MSP-$1_{33}$ and MSP-$1_{19}$ fragments appear to be under different selective pressures than the ones just described in the homologous region in *P. falciparum*. We observed an excess of non-synonymous over synonymous substitutions in the MSP-$1_{33}$ and not in the MSP-$1_{19}$ ; in addition, we found that the Tajimas's D and F* tests are significant and positive for MSP-$1_{33}$, which is expected under the scenario of balancing selection although it could be the result of population structure, a clear possibility given the origin of the sample analyzed. Nevertheless, when we studied the genetic variation in the MSP-$1_{33}$ and MSP-$1_{19}$ by using the McDonald and Kreitmant test against *P. cynomolgi* and *P. inui* we found an excess of non-synonymous substitutions in the *P. vivax* MSP-$1_{33}$ no matter which species we used to compare it with, suggesting that positive natural selection is operating in this fragment.

Our results support previous observations that *P. vivax* MSP-$1_{33}$ could play an important role in reticulocyte invasion (Espinosa *et al*., 2003; Rodriguez *et al*., 2002). However, the polymorphism in the *P. vivax* MSP-$1_{33}$ appears more complicated; indeed, there is a 105 bp

fragment with high polymorphism located between regions where peptides with high specific binding activity (HSBA) to reticulocytes have been found (Espinosa *et al*., 2003; Rodriguez *et al.*, 2002). These regions with HSBA are not only highly conserved among *P. vivax* isolates (n= 75) but also show more synonymous than non-synonymous substitutions when compared with *P. cynomolgi* (peptides 1735, 1738 and 1747 *sensu* Rodriguez et al 2002 have Ks of 0.30, 0.31 and 0.22 versus Kn of 0.16, 0.025, and 0.11 respectively) and a similar pattern is observed when compared with *P. inui* (peptides 1735, 1738 and 1747 *sensu* Rodriguez *et al.* 2002 have Ks of 0.29, 0.37, and 0.041 versus Kn of 0.17, 0.10, and 0.16 respectively) . This overall pattern indicates that these HSBA regions are under selective constraints to accumulate amino acid replacements; as a result, they could be a valuable target for a vaccine against *P. vivax* as has been suggested previously (Espinosa *et al.* 2003).

There is no information regarding the immunologic role played by the variation observed in *P. vivax* MSP-$1_{33}$. Elucidating whether it hampers effective natural immune responses against these conserved regions with HSBA to reticulocytes or whether it plays any other role requires further investigation. Nevertheless, it seems clear from this comparative analyses that we cannot simply extrapolate information derived from *P. falciparum* into *P. vivax* in the case of MSP-$1_{42}$.

In summary, we have investigated the genetic diversity of the sequence encoding the MSP-$1_{42}$ in the two major human malarial parasites. We found evidence supporting positive natural selection as an important factor in the maintenance and generation of the observed polymorphism. However, we describe how natural selection is acting differently in the MSP-$1_{33}$ and MSP-$1_{19}$ fragments of the MSP-$1_{42}$ in each of the two human malarial parasites. That is, our results suggest that these fragments, MSP-$1_{33}$ and MSP-$1_{19}$, could play different roles in each of the two human malarial parasites.

**Acknowledgements**

This research is supported by grants from the National Institutes of Health (R01 GM60740) to A.A.E. and from the Japanese Ministry of Education, Culture, Sports, Science and Technology (14021125) to K.T. We thank Andrea McCollum and two anonymous reviewers for their commentaries that improved this manuscript. The findings and conclusions in this report are those of the author(s) and do not necessarily represent the views of the Centers for Disease Control and Prevention.

## References

Ahlborg N, Ling IT, Howard W, Holder AA, Riley EM. Protective immune responses to the 42-hilodalton (kDa) region of *Plasmodium yoelii* merozoite surface protein 1 are induced by the C-terminal 19-kDa region but not by large adjacent 33-kDa region. Infect Immun 2002;70:820–825. [PubMed: 11796616]

Certa U, Rotmann D, Matile H, Reber-Liske R. A naturally occurring gene encoding the major surface antigen precursor p190 of *Plasmodium falciparum* lacks tripeptide repeats. EMBO J 1987;6:4137–4142. [PubMed: 3327688]

Chang SP, Kramer KJ, Yamaga KM, Kato A, Case SE, Siddiqui WA. *Plasmodium falciparum*: Gene structure and hydropathy profile of the major merozoite surface antigen (gp195) of the Uganda-Palo Alto isolate. Exp Parasitol 1988;67:1–11. [PubMed: 3049134]

Coatney, GR.; Collins, WE.; Warren, M.; Contacos, PG. The Primate Malarias. U.S. Government Printing Office; 1971.

Conway DJ, Cavanagh DR, Tanabe K, Roper C, Mikes ZS, Sakihama N, Bojang KA, Oduola AM, Kremsner PG, Arnot DE, Greenwood BM, McBride JS. A principal target of human immunity to malaria identified by molecular population genetic and immunological analyses. Nat Med 2000;6:689–692. [PubMed: 10835687]

Drakeley CJ, Corran PH, Coleman PG, Tongren JE, McDonald SL, Carneiro I, Malima R, Lusingu J, Manjurano A, Nkya WM, Lemnge MM, Cox J, Reyburn H, Riley EM. Estimating medium- and long-

term trends in malaria transmission by using serological markers of malaria exposure. Proc Natl Acad Sci U S A 2005;102:5108–5113. [PubMed: 15792998]

Escalante AA, Lal AA, Ayala FJ. Genetic Polymorphism and Natural Selection in the Malaria Parasite *Plasmodium falciparum*. Genetics 1998;149:189–202. [PubMed: 9584096]

Escalante AA, Cornejo OE, Rojas A, Udhayakumar V, Lal AA. Assessing natural selection in malarial parasites. Trends Parasitol 2004;20:388–395. [PubMed: 15246323]

Escalante AA, Cornejo OE, Freeland DE, Poe AC, Durrego E, Collins WE, Lal AA. A monkey's tale: The origin of *Plasmodium vivax* as a human malaria parasite. Proc Nat Acad Sci USA 2005;102:1980–1985. [PubMed: 15684081]

Espinosa AM, Sierra AY, Barrero CA, Cepeda LA, Cantor EM, Lombo TB, Guzman F, Avila SJ, Patarroyo MA. Expression, polymorphism analysis, reticulocyte binding and serological reactivity of two *Plasmodium vivax* MSP-1 protein recombinant fragments. Vaccine 2003;21:1033–1043. [PubMed: 12559776]

Fu YX, Li WH. Statistical tests of neutrality of mutations. Genetics 1993;133:693–709. [PubMed: 8454210]

Good MF, Stanisic D, Xu H, Elliott S, Wykes M. The immunological challenge to developing a vaccine to the blood stages of malaria parasites. Immunol Rev 2004;201:254–267. [PubMed: 15361246]

Han HJ, Park SG, Kim SH, Hwang SY, Han J, Traicoff J, Kho WG, Chung JY. Epidermal growth factor-like motifs 1 and 2 of *Plasmodium vivax* merozoite surface protein 1 are critical domains in erythrocyte invasion. Biochem Biophys Res Commun 2004;320:563–570. [PubMed: 15219866]

Holder AA, Freeman RR. Biosynthesis and processing of a *Plasmodium falciparum* schizont antigen recognized by immune serum and a monoclonal antibody. J Exp Med 1982;156:1528–1538. [PubMed: 6752328]

Hudson RR, Kaplan NL. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 1985;111:147–164. [PubMed: 4029609]

Hughes AL. Positive selection and intrallelic recombination at the merozoite surface antigen-1 (MSA-1) locus of *Plasmodium falciparum*. Mol Biol Evol 1992;9:381–393. [PubMed: 1584009]

Hughes MK, Hughes AL. Natural selection on *Plasmodium* surface proteins. Mol Biochem Parasitol 1995;71:99–113. [PubMed: 7630387]

Jongwutiwes S, Tanabe K, Nakazawa S, Yanagi T, Kanbara H. Sequence variation in the tripeptide repeats and T cell epitopes in P190 (MSA-1) of *Plasmodium falciparum* from field isolates. Mol Biochem Parasitol 1992;51:81–89. [PubMed: 1373473]

Jongwutiwes S, Tanabe K, Kanbara H. Sequence conservation in the C-terminal part of the precursor to the major merozoite surface proteins (MSP-1) of *Plasmodium falciparum* from field isolates. Mol Biochem Parasitol 1993;59:95–100. [PubMed: 8515786]

Joy DA, Feng X, Mu J, Furuya T, Chotivanich K, Krettli AU, Ho M, Wang A, White NJ, Suh E, Beerli P, Su XZ. Early origin and recent expansion of *Plasmodium falciparum*. Science 2003;300:318–321. [PubMed: 12690197]

Kumar S, Tamura K, Jakobsen IB, Nei M. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 2001;17:1244–1245. [PubMed: 11751241]

Kumar SV, Ranjan S, Saxena V, Rajesh V, Roy SK, Kochar D, Ranjan A, Das A. *Plasmodium falciparum*: Genetic diversity of C-terminal region of MSP-1 in isolates from Indian sub-continent. Exp Parasitol 2005;110:384–388. [PubMed: 15907839]

Li WH. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J Mol Evol 1993;36:96–99. [PubMed: 8433381]

Miller LH, Roberts T, Shahabuddin M, McCutchan T. Analysis of sequence diversity in the *Plasmodium falciparum* merozoite surface protein-1 (MSP-1). Mol Biochem Parasitol 1993;59:1–14. [PubMed: 8515771]

Nei, M. Molecular Evolutionary Genetics. Columbia University Press; 1987.

Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 1986;3:418–426. [PubMed: 3444411]

Nei, M.; Kumar, S. Molecular Evolution and Phylogenetics. Oxford University Press; 2000.

O'Donnell RA, Saul A, Cowman AF, Crabb BS. Functional conservation of the malaria vaccine antigen MSP-$1_{19}$ across distantly related Plasmodium species. Nat Med 2000;6:91–95. [PubMed: 10613831]

Okech BA, Corran PH, Todd J, Joynson-Hicks A, Uthaipibull C, Egwang TG, Holder AA, Riley EM. Fine specificity of serum antibodies to *Plasmodium falciparum* merozoite surface protein, PfMSP-1 (19), predicts protection from malaria infection and high-density parasitemia. Infect Immun 2004;72:1557–1567. [PubMed: 14977962]

Pan W, Tolle R, Bujard H. A direct and rapid sequencing strategy for the *Plasmodium falciparum* antigen gene gp190/MSA1. Mol Biochem Parasitol 1995;73:241–244. [PubMed: 8577332]

Polley SD, Weedall GD, Thomas AW, Golightly LM, Conway DJ. Orthologous gene sequences of merozoite surface protein 1 (MSP1) from *Plasmodium reichenowi* and P. gallinaceum confirm an ancient divergence of *P. falciparum* alleles. Mol Biochem Parasitol 2005;142:25–31. [PubMed: 15907558]

Putaporntip C, Jongwutiwes S, Seethamchai S, Kanbara H, Tanabe K. Intragenic recombination in the 3' portion of the merozoite surface protein 1 gene of *Plasmodium vivax*. Mol Biochem Parasitol 2000;109:111–119. [PubMed: 10960170]

Putaporntip C, Jongwutiwes S, Sakihama N, Ferreira MU, Kho WG, Kaneko A, Kanbara H, Hattori T, Tanabe K. Mosaic organization and heterogeneity in frequency of allelic recombination of the *Plasmodium vivax* merozoite surface protein-1 locus. Proc Natl Acad Sci U S A 2002;99:16348–16353. [PubMed: 12466500]

Putaporntip C, Jongwutiwes S, Iwasaki T, Kanbara H, Hughes AL. Ancient common ancestry of the merozoite surface protein 1 of *Plasmodium vivax* as inferred from its homologue in *Plasmodium knowlesi*. Mol Biochem Parasitol 2006;146:105–108. [PubMed: 16337018]

Qari SH, Shi YP, Goldman IF, Nahlen BL, Tibayrenc M, Lal AA. Predicted and observed alleles of *Plasmodium falciparum* merozoite surface protein –1 (MSP-1), a potential malaria vaccine antigen. Mol Biochem Parasitol 1998;92:241–252. [PubMed: 9657329]

Rodríguez LE, Urquiza M, Ocampo M, Curtidor H, Suarez J, Garcia J, Vera R, Puentes A, Lopez R, Pinto M, Rivera Z, Patarroyo ME. *Plasmodium vivax* MSP-1 peptides have high specific binding activity to human reticulocytes. Vaccine 2002;20:1331–1339. [PubMed: 11818151]

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 2003;19:2496–2497. [PubMed: 14668244]

Sachdeva S, Ahmad G, Malhotra P, Mukherjee P, Chauhan VS. Comparison of immunogenicities of recombinant *Plasmodium vivax* merozoite surface protein 1 19- and 42-kiloDalton fragments expressed in *Escherichia coli*. Infect Immun 2004;72:5775–5782. [PubMed: 15385477]

Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987;4:406–425. [PubMed: 3447015]

Shi YP, Sayed U, Qari SH, Roberts JM, Udhayakumar V, Oloo AJ, Hawley WA, Kaslow DC, Nahlen BL, Lal AA. Natural immune response to the C-terminal 19-kilodalton domain of *Plasmodium falciparum* merozoite surface protein 1. Infect Immun 1996;64:2716–2723. [PubMed: 8698500]

Snounou G, White NJ. The co-existence of *Plasmodium*: sidelights from *falciparum* and *vivax* malaria in Thailand. Trends Parasitol 2004;20:333–339. [PubMed: 15193565]

Stanisic DI, Martin LB, Gatton ML, Good MF. Inhibition of 19-kDa C-terminal region of merozoite surface protein-1-specific antibody responses in neonatal pups by maternally derived 19-kDa C-terminal region of merozoite surface protein-1-specific antibodies but not whole parasite-specific antibodies. J Immunol 2004;172:5570–5581. [PubMed: 15100300]

Takala SL, Escalante AA, Branch OH, Kariuki S, Biswas S, Chaiyaroj SC, Lal AA. Genetic diversity in the Block 2 region of the merozoite surface protein 1 (MSP-1) of *Plasmodium falciparum*: Additional complexity and selection and convergence in fragment size polymorphism. Infect Genet Evol 2006;6:417–424. [PubMed: 16517218]

Tanabe K, Mackay M, Goman M, Scaife JG. Allelic dimorphism in a surface antigen gene of the malaria parasite *Plasmodium falciparum*. J Mol Biol 1987;195:273–287. [PubMed: 3079521]

Tolle R, Bujard H, Cooper JA. *Plasmodium falciparum*: variations within the C-terminal region of merozoite surface antigen-1. Exp Parasitol 1995;81:47–54. [PubMed: 7628566]

Tanabe K, Sakihama N, Nakamura Y, Kaneko O, Kimura M, Ferreira MU, Hirayama K. Selection and drift on polymorphisms within the merozoite surface protein-1 gene of *Plasmodium falciparum*. Gene 2000;241:342–331.

Tanabe K, Sakihama N, Kaneko A. Stable SNPs in malaria antigen genes in isolated populations. Science 2004;303:493. [PubMed: 14739451]

Udhayakumar V, Anyona D, Kariuki S, Shi YP, Bloland PB, Branch OH, Weiss W, Nahlen BL, Kaslow DC, Lal AA. Identification of T and B cell epitopes recognized by humans in the C-terminal 42-kDa domain of the *Plasmodium falciparum* merozoite surface protein (MSP)-1. J Immunol 1995;154:6022–6030. [PubMed: 7538540]

Weber JL, Sim BK, Lyon JA, Wolf R. Merozoite surface protein sequence from the Camp strain of the human malaria parasite *Plasmodium falciparum*. Nucleic Acids Res 1988;16:1206. [PubMed: 3278296]

Yang C, Collins WE, Sullivan JS, Kaslow DC, Xiao L, Lal AA. Partial protection against *Plasmodium vivax* blood-stage infection in Saimiri monkeys by immunization with a recombinant C-terminal fragment of merozoite surface protein 1 in block copolymer adjuvant. Infect Immun 1999;67:342–349. [PubMed: 9864235]

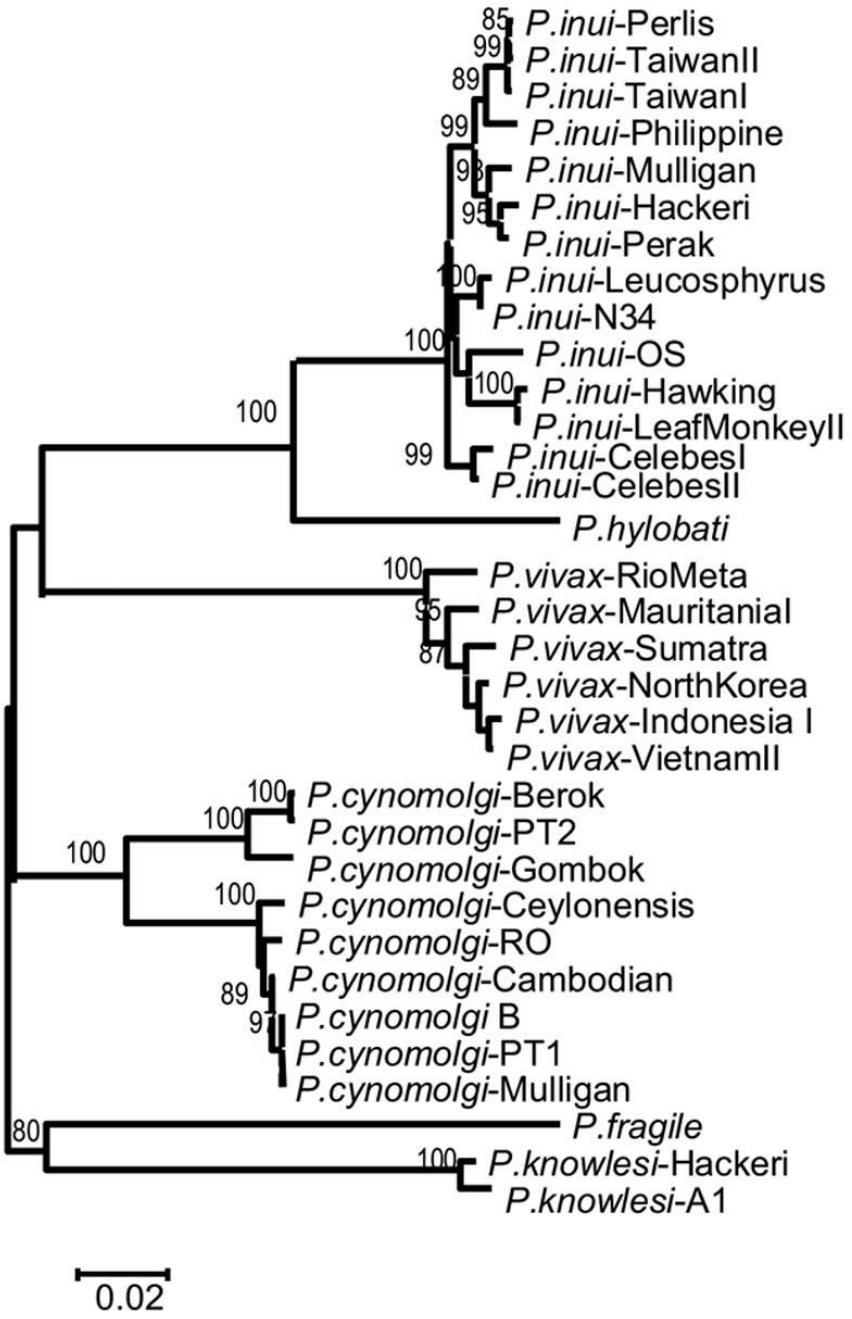


**Figure 1.**
Neighbor-Joining tree of the MSP-$1_{42}$ alleles using Tamura-Neís distance. The numbers on the nodes of the tree are percent of bootstrap values based on 1,000 pseudo-replications. The sequences reported in this study are identified with their species and strain names.

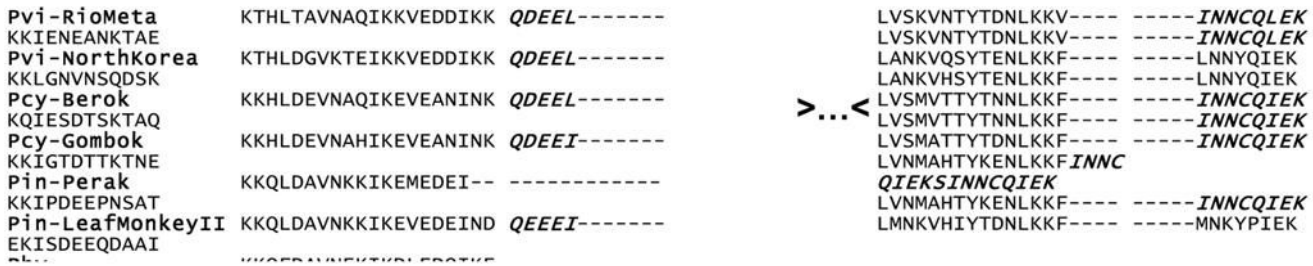


**Figure 2.**

Repetitive sequences observed in the MSP-$1_{42.}$ The observed motifs are in italics. The dots (>**…**<) are indicating a non-repetitive portion of the protein that is not shown. The first three letters in the sequence codes indicate the species: Pvi is *P. vivax,* Pcy is *P. cynomolgi*, Pin is *P. inui*, *Phy* is *P. hylobati*, Pkn is *P. knowlesi*, and Pfr is *P. fragile.*

**Table 1**

Polymorphism found in the MSP-$1_{42}$ in *P. falciparum* and *P. vivax*

| | π | Ds | Dn | Ds-Dn [SD] | Z | Tajima D | F* |
|---|---|---|---|---|---|---|---|
| *P. falciparum* (n= 120) | | | | | | | |
| 42KD | 0.05042 | 0.0821 | 0.0541 | 0.0280 [0.011] | Ds>Dn $_{P < 0.05}$ | −0.11184 n.s. | 0.13353 n.s. |
| 33KD | 0.06551 | 0.1236 | 0.0741 | 0.0494 [0.020] | Ds>Dn $_{P < 0.05}$ | 0.10150 n.s. | 0.86091 n.s. |
| 19KD | 0.00884 | 0.0013 | 0.0107 | −0.009 [0.004] | Ds<Dn $_{P < 0.05}$ | −1.72070 $_{0.10 > P > 0.05}$ | −4.78810 $_{P < 0.05}$ |
| *P. vivax* (n = 75) | | | | | | | |
| 42KD | 0.02184 | 0.0125 | 0.0249 | −0.0123 [0.005] | Ds<Dn | 2.19241 $_{P < 0.05}$ | 2.09599 $_{P < 0.05}$ |
| 33KD | 0.03249 | 0.0162 | 0.0325 | −0.0160 [0.006] | Ds<Dn | 2.31357 $_{P < 0.05}$ | 2.24458 $_{P < 0.05}$ |
| 19KD | 0.0006 | 0.0005 | 0.0006 | 0.0001 [0.000] | Ds=Dn | −1.02018 n.s | −1.02018 n.s |

π, nucleotide diversity; n, number of sequences; Dn is the nucleotide diversity of nonsynonymous mutations per nonsynonymous sites and Ds is the nucleotide diversity of synonymous mutations per synonymous site using the Nei and Gojobori method; Ds-Dn are the difference of Ds and Dn with their standard deviation, SD, estimated by bootstrap with 1000 pseudo replicates; Z is the Z-test (Nei and Kumar 2000); Tajima D and F* are tests for detecting departures from the neutral model.

**Table 2**

Polymorphism found in the MSP-$1_{42}$ in other non-human *Plasmodium* spp.

| | π | Ds | Dn | Ds-Dn [SD] | Z |
|---|---|---|---|---|---|
| *P. cynomolgi* (n= 10) | | | | | |
| 42KD | 0.03805 | 0.0871 | 0.0287 | 0.0585 [0.015] | Ds>Dn $_{P < 0.05}$ |
| 33KD | 0.06551 | 0.1001 | 0.0312 | 0.0687 [0.018] | Ds>Dn $_{P < 0.05}$ |
| 19KD | 0.02502 | 0.0469 | 0.0211 | 0.0257 [0.022] | Ds=Dn |
| *P. inui* (n = 15) | | | | | |
| 42KD | 0.02416 | 0.0284 | 0.0237 | 0.0049 [0.006] | Ds=Dn |
| 33KD | 0.02951 | 0.0358 | 0.0289 | 0.0071 [0.008] | Ds=Dn |
| 19KD | 0.0067 | 0.0051 | 0.0073 | −0.0022 [0.005] | Ds=Dn |

π, nucleotide diversity; n, number of sequences; Dn is the nucleotide diversity of nonsynonymous mutations per nonsynonymous sites and Ds is the nucleotide diversity of synonymous mutations per synonymous site using the Nei and Gojobori method. Ds-Dn are the difference of Ds and Dn with their standard deviation, SD, estimated by bootstrap with 1000 pseudo replicates. Z is the Z-test (Nei and Kumar 2000).