Author's personal copy

Mitochondrion 13 (2013) 630–636

Contents lists available at ScienceDirect

Mitochondrion

journal homepage: www.elsevier.com/locate/mito

Plasmodium falciparum mitochondrial genetic diversity exhibits isolation-by-distance patterns supporting a sub-Saharan African origin ☆

Kazuyuki Tanabe [a,b,*,1], Thibaut Jombart [c], Shun Horibe [d], Nirianne M.Q. Palacpac [b], Hajime Honma [a,e], Shin-Ichiro Tachibana [a,f], Masatoshi Nakamura [g], Toshihiro Horii [b], Hirohisa Kishino [d], Toshihiro Mita [e,h,**]

[a] *Laboratory of Malariology, Research Institute for Microbial Diseases, Osaka University, Osaka 565-0871, Japan*
[b] *Department of Molecular Protozoology, Research Institute for Microbial Diseases, Osaka University, Osaka 565-0871, Japan*
[c] *MRC Centre for Outbreak Analysis and Modelling, Department of Infectious Disease Epidemiology, School of Public Health, Imperial College, London W2 1PG, UK*
[d] *Graduate School of Agriculture and Life Sciences, University of Tokyo, Tokyo 113-8657, Japan*
[e] *Department of International Affairs and Tropical Medicine, Tokyo Women's Medical University, Tokyo, Japan*
[f] *Department of Biology and Geosciences, Graduate School of Science, Osaka City University, Osaka 558-8585, Japan*
[g] *Department of Tropical Medicine and Parasitology, Dokkyo Medical University, Tochigi 321-0293, Japan*
[h] *Department of Molecular and Cellular Parasitology, Juntendo University School of Medicine, Tokyo 113-8421, Japan*

ARTICLE INFO

Article history:
Received 29 April 2013
Received in revised form 13 July 2013
Accepted 26 August 2013
Available online 1 September 2013

Keywords:
Mitochondrion
Mitochondrial genome
Malaria
Plasmodium falciparum
Genetic diversity
Polymorphism

ABSTRACT

The geographical distribution of single nucleotide polymorphism (SNP) in the mitochondrial genome of the human malaria parasite *Plasmodium falciparum* was investigated. We identified 88 SNPs in 516 isolates from seven parasite populations in Africa, Southeast Asia and Oceania. Analysis of the SNPs postulated a sub-Saharan African origin and recovered a strong negative correlation between within-population SNP diversity and geographic distance from the putative African origin over Southeast Asia and Oceania. These results are consistent with those previously obtained for nuclear genome-encoded housekeeping genes, indicating that the pattern of inheritance does not substantially affect the geographical distribution of SNPs.

© 2013 Elsevier B.V. and Mitochondria Research Society. All rights reserved.

1. Introduction

Malaria, a mosquito-borne infectious disease caused by protozoa of the genus *Plasmodium*, continues to impose a serious public health concern throughout the tropics with 219 million estimated cases and 660,000 deaths in 2010 (WHO, 2012), mostly due to infections of the most virulent species, *Plasmodium falciparum*. Elucidating and understanding the genetic diversity of this parasite is of great importance to gain insights into its pathogenesis as this has compromised malaria control strategies by facilitating drug resistance and immune evasion.

Malaria parasite populations in geographically diverse regions exist in highly variable epidemiological settings with respect to intensity of transmission (Hay et al., 2009; Kelly-Hope and McKenzie, 2009), malaria interventions, evolutionary history and population structures (Anderson et al., 2000; Joy et al., 2003). We have recently shown that within-population diversity of *P. falciparum* housekeeping genes was primarily determined by geographic distance from a postulated origin in central sub-Saharan Africa over Africa, Asia and Oceania (Tanabe et al., 2010). In a subsequent study, we also recovered significant negative (or inverse) correlation between within-population diversity of *P. falciparum* vaccine candidate antigen genes and geographic distance (Tanabe et al., 2013). These results suggest that regional variation in transmission intensity, malaria interventions, and host immune level might not be tightly associated with the geographic distribution of within-population parasite

Abbreviations: mt, mitochondria; SNP, single nucleotide polymorphism; θ_S, the standardized number of polymorphic sites per site; θ_π, the average number of pair-wise nucleotide differences; dS, the mean number of synonymous substitutions per synonymous site; dN, the mean number of non-synonymous substitutions per non-synonymous site; F_{ST}, the Wright's fixation index of inter-population variance in allele frequencies; MAF, minor allele frequency; *serca*, sarcoplasmic/endoplasmic reticulum Ca^{2+}-transporting ATPase gene; *adsl*, adenylosuccinate lyase gene; ATQ, atovaquone; *cox1*, cytochrome *c* oxidase subunit I gene; *cox3*, cytochrome *c* oxidase subunit III gene; *cob*, cytochrome *b* gene.

☆ We dedicate this paper to the memory of the late Kazuyuki Tanabe.

* Corresponding author.
** Correspondence to: T. Mita, Department of Molecular and Cellular Parasitology, Juntendo University School of Medicine, Tokyo 113-8421, Japan. Tel.: +81 3 5802 1042; fax: +81 3 5800 0476.
E-mail addresses: t.jombart@imperial.ac.uk (T. Jombart), shunhoribe@gmail.com (S. Horibe), nirian@biken.osaka-u.ac.jp (N.M.Q. Palacpac), homma@research.twmu.ac.jp (H. Honma), sitachi@sci.osaka-cu.ac.jp (S.-I. Tachibana), mnakamura8823@gmail.com (M. Nakamura), horii@biken.osaka-u.ac.jp (T. Horii), kishino@lbm.ab.a.u-tokyo.ac.jp (H. Kishino), tmita@juntendo.ac.jp (T. Mita).

[1] Deceased, August 12th 2013.

1567-7249/$ – see front matter © 2013 Elsevier B.V. and Mitochondria Research Society. All rights reserved.
http://dx.doi.org/10.1016/j.mito.2013.08.008

Author's personal copy

genetic diversity. Rather, the smooth patterns of reduction in parasite genetic diversity along with geographic distance from the African origin mirror the ones previously documented in modern humans (Handley et al., 2007; Liu et al., 2006; Prugnolle et al., 2005; Ramachandran et al., 2005). We then inferred that *P. falciparum* "piggybacked" with modern humans in a central sub-Saharan area and colonized the world alongside modern human migrations out of Africa (Tanabe et al., 2010).

Like other eukaryotic organisms, *P. falciparum* has mitochondria essential for a range of cellular processes and cellular signaling. The parasite mitochondria showed remarkable changes in both morphology and biochemical activities during parasite development (Krungkrai, 2004; Vaidya and Mather, 2009). In erythrocytic stages, although the organelle has underdeveloped structure and relatively limited function, enzymes for oxidative phosphorylation were expressed, suggesting heterogeneity of metabolic status at the blood stage (Daily et al., 2007). At the mosquito stages of the parasite, the importance of TCA cycle members has recently been demonstrated (Hino et al., 2012) consistent with the organelle being metabolically more active and developed showing typical cristate morphology.

Additionally the mitochondrial (mt) genome of *P. falciparum*, as well as other *Plasmodium* species, has several unique genomic features (Vaidya and Mather, 2009): (i) the genome is present in the form of a circular and/or tandemly repeated linear 6 kb element, the smallest mt genome size known so far (Preiser et al., 1996); (ii) the 6 kb element encodes only three protein-coding genes (*cox1*, *cox3* and *cob*: cytochrome *c* oxidase subunit I, cytochrome *c* oxidase subunit III and cytochrome *b* genes, respectively) along with highly fragmented large subunit (LSU) and small subunit (SSU) ribosomal RNA (rRNA) genes (Feagin et al., 1997, 2012; Hikosaka et al., 2011); and (iii) curiously, no transfer RNA genes exist.

These unique genomic features may impact strongly the genetic diversity observed in parasite populations. Moreover, similar to a great majority of sexual organisms, the parasite mt genome is uniparentally inherited and does not undergo recombination (Creasey et al., 1993; Joy et al., 2003; Vaidya et al., 1993). Thus, it is likely possible that the geographic distribution of within-population genetic diversity of the mt genome is different from that of the nuclear genome, which is inherited biparentally and undergoes recombination. For example, once a deleterious mutation is newly generated in the mt genome, all single nucleotide polymorphisms (SNPs) previously present in the genome must disappear, due to the absence of recombination in the mt genome. In contrast, in the nuclear genome only SNPs linked to a newly generated deleterious mutation in a chromosomal recombination segment would be lost by recombination and genetic hitchhiking with SNPs outside the segment surviving. This may suggest that geographical distribution of SNPs in the mt genome could be inherently unique. However, investigations on polymorphisms of the *P. falciparum* mt genome are still limited (Conway et al., 2000; Joy et al., 2003). There is also a clinical and epidemiological reason for studying polymorphisms of the mt genome relating to its biological role. Atovaquone (ATQ), an anti-malarial drug, is known to target cytochrome *b*, leading to the inhibition of pyrimidine biosynthesis and eventually parasite death (Korsinczky et al., 2000; Srivastava et al., 1999). The prevalence of ATQ-resistant parasites in targeted populations needs to be assessed before and after implementation of the drug usage for effective treatment and malaria control.

In this study, we investigated the geographical distribution of SNPs in the *P. falciparum* mt genome with the same parasite populations from Africa, Southeast Asia and Oceania that were previously used for the analysis of nuclear genome-encoded housekeeping genes (Tanabe et al., 2010). Using the same sample set allows us to address whether within-population mt SNP diversity is driven by a negative correlation with the geographic distance from a sub-Saharan African origin, as in nuclear genome-encoded housekeeping genes (Tanabe et al., 2010), or whether specific genetic features such as uniparental inheritance and the absence of recombination in the mt genome determine the geographic distribution of mt SNPs.

2. Materials and methods

2.1. Parasite isolates

P. falciparum isolates were collected from seven countries: Ghana, Tanzania, Thailand, Philippines, Papua New Guinea (PNG), the Solomon Islands, and Vanuatu. Details of the parasite isolates have been described previously: in Ghana, samples were collected in three villages near Winneba, coastal Ghana in 2004 (Tanabe et al., 2010). In Tanzania, samples were collected in the Rufiji River Delta in 1993, 1998 and 2003 (Tanabe et al., 2007). In Thailand, samples were collected in Mae Sod, Tak Province in 1995 (Sakihama et al., 1999). In the Philippines, samples were collected in Puerto Princesa, Palawan island in 1997 (Sakihama et al., 2007). In Papua New Guinea (PNG), samples were collected from five villages in Wewak, East Sepik Province in 2001 and 2002 (Tanabe et al., 2010). In the Solomon Islands, samples were collected in three local areas in northeastern Guadalcanal island in 1995–1996 (Sakihama et al., 2006). In Vanuatu, samples were collected from four islands in 1996–1998 (Sakihama et al., 2001). These seven parasite populations were also used in our previous polymorphism study for two nuclear housekeeping genes, the sarcoplasmic/endoplasmic reticulum Ca^{2+}-transporting ATPase gene (*serca*) and the adenylosuccinate lyase gene (*adsl*) (Tanabe et al., 2010). Ethical approval was obtained from relevant ethical committees.

2.2. Sequencing

Near full-length sequences (5883–5885 bp) of the *P. falciparum* mitochondrial genome were amplified by PCR using Takara LA Taq (Takara Bio, Japan). Procedures and conditions for PCR amplification were described (Tanabe et al., 2011). Four sets of primers were used in two successive reactions. The 5′ 3-kb region was first amplified using primers mtF0 and mtR3076, followed by a second PCR using primers mtF1 and mtR3076; the 3′ 3-kb region was first amplified using primers mtF2992 and mtR1, followed by a second PCR using primers mtF3019 and mtR2 (Supplementary Table 1). The PCR products were purified using QIAquick (PCR Purification Kit, QIAGEN). DNA sequencing was performed directly from two independent PCR products, using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA) and an ABI 3130 Genetic Analyzer (Applied Biosystems). Sequencing primers were designed to cover target regions in both directions (Supplementary Table 1). Mixed genotype infections, judged from overlapping peaks at given positions in an electropherogram, were excluded from further analysis. We obtained 516 sequences from the seven parasite populations. The nucleotide sequences were deposited in the DDBJ/EMBL/GenBank database (accession numbers: AB570434–AB570542, AB570544-AB570765, AB570767-AB570951).

2.3. Statistical analysis

DNA sequences were aligned as described in Tanabe et al. (2011), and insertions/deletions (indel), were excluded. The aligned sequences (5883 bp) were subjected to statistical analysis and the following statistical indices were calculated: number of haplotypes, number of polymorphic sites, number of singleton polymorphic sites, the average number of pair-wise nucleotide differences per site (θ_π), and the standardized number of polymorphic sites per site (θ_S), i.e., Watterson's estimator. All statistics were calculated by DnaSP version 5.00 (Librado and Rozas, 2009). The mean numbers of synonymous substitutions per synonymous site (dS) and non-synonymous substitutions per non-synonymous site (dN) were estimated using MEGA 5.0 (Tamura et al., 2011). When dN was significantly higher than dS, diversifying

Author's personal copy

selection was inferred and if an excess of dS over dN was significant, purifying selection was inferred.

Allele frequencies in SNPs were computed using Arlequin v3.1 (Excoffier et al., 2005) and singleton SNPs as well as SNPs with minor allele frequency (MAF) of <5% were identified. SNPs were also classified into either SNPs unique to a geographic area or those shared by multiple populations. Differences in SNP frequencies among various groups were assessed by the Student's *t*-test.

Genetic differentiation among populations was estimated by Wright's fixation index of inter-population variance in allele frequencies, termed F_{ST}. Since admixture of structured parasite populations results in artifactually elevated genetic diversity (Tanabe et al., 2010), we first tested the level of local population variation. Among the five countries where samples were collected from several local populations (Ghana, Philippines, PNG, Solomon Islands and Vanuatu), we detected significant population sub-structuring only within PNG. We thus conducted statistical analysis for individual local subpopulations in PNG and then calculated the average of statistical indices across subpopulations. Overall F_{ST} and pairwise F_{ST} values were estimated, using Arlequin v3.1. Statistical significance of F_{ST} values higher than 0 was assessed from 16,000 random-permutations.

2.4. Inference for geographic origin of P. falciparum and correlation between within-population genetic diversity and geographic distance

All geographic distances were computed via graph theory (Manica et al., 2005; Prugnolle et al., 2005) using shortest distances along landmasses within a spherical referential of 40,962 vertices, and adopting the same assumptions of friction cost and a colonization route out of Africa as we previously described (Tanabe et al., 2010). The best supported origin was inferred by searching for the shortest routes to all analyzed populations from 312 hypothetical origins on land arranged on a regular grid. For the isolation by distance analysis, a matrix of pairwise physical distances was computed for all populations using the distance through landmasses described above. Geographic distance was measured as the logged travel cost over friction routes along landmasses (Tanabe et al., 2010). Correlation between within-population nucleotide diversity and geographic distance was tested by standard regression analysis. The coefficient of determination (R^2) was calculated to quantify the strength of this association.

2.5. Construction of haplotype network

The 516 mt genome sequences were used to construct a haplotype network using the median joining method implemented in the NETWORK software version 4.6.1.1 (Bandelt et al., 1999). Outgroup probability or root probability of a given haplotype was generated using TCS 1.21 (Clement et al., 2000) to estimate the relative age of each haplotype based on the frequencies and the number of connections to other haplotypes within the network.

3. Results

3.1. Polymorphism in the P. falciparum mt genome

We identified 88 SNPs in 516 *P. falciparum* isolates from seven parasite populations in Africa (Ghana and Tanzania), SEA (Thailand and Philippines) and Oceania (PNG, Solomon Islands and Vanuatu): 57 in *cox1*, *cox3* and *cob*, 25 in RNA gene fragments and 6 in the intergenic regions (Table 1). Of the 88 SNPs, 75 (85%) were newly identified here. 13 of 32 previously reported SNPs (40%) described from 100 worldwide samples were represented (Joy et al., 2003). Nineteen of the previously reported SNPs were not detected in our samples, suggesting the probable presence of more abundant SNPs in natural parasite populations. Nucleotide diversity was low: 0.214% and 0.027% for θ_S (standardized number of polymorphic sites per site) and θ_π (average pairwise nucleotide differences per site), respectively (Table 1), indicating high sequence similarity among isolates. These values are comparable to those found in the parasites' two nuclear housekeeping genes, *serca* + *adsl* (0.181% and 0.035% for θ_S and θ_π, respectively) (Table 1). In the mt genome, θ_S was 7.9-fold higher than θ_π. An excess of θ_S relative to θ_π indicates the abundance of rare alleles. Of the seven parasite populations studied here, two African populations (Tanzania and Ghana) showed the highest nucleotide diversity (θ_S) (Supplementary Table 2).

A signature of purifying selection was detected for *cox1* and *cob* where a significant excess of synonymous substitutions (dS) over non-synonymous substitutions (dN) was observed (Table 1). For *cox3*, dN was seemingly higher than dS but the difference was not significant. Thus, the three protein coding genes in the mt genome did not show evidence for strong diversifying selection, confirming previous observations by Joy et al. (2003).

In the present sample set, we did not observe a mutation at amino acid position 268 in *cob* at which a change from Tyr to either Ser, Cys or Asn was correlated to atovaquone resistance following treatment failure (Fivelman et al., 2002; Korsinczky et al., 2000; Musset et al., 2007; Srivastava et al., 1999).

3.2. Geographic origin of the mt genome and negative correlation between within-population diversity and geographical distance from a postulated origin

We searched for a geographical origin of the *P. falciparum* mt genome using within-population nucleotide diversity (θ_S) from the seven parasite populations. As previously observed in nuclear genome-encoded housekeeping genes, results suggest an origin in sub-Saharan Africa (Fig. 1). A clear pattern of negative correlation between within-population nucleotide diversity (θ_S) and geographical distance from a postulated origin was obtained ($R^2 = 0.96$, $P < 0.0001$ by Pearson correlation test) (Fig. 1, inset). This indicates that 96% of within-population nucleotide diversity (θ_S) of the parasite mt genome can be ascribed to geographic distance from the eastern sub-Saharan African origin over

Table 1
Diversity profiles of the *P. falciparum* mitochondrial genome.

	Nucleotide length (bp)	No. of SNPs	$\theta_S \pm$ SD	$\theta_\pi \pm$ SD	dS ± SE	dN ± SE	P value
Whole region (mt)	5883	88	0.00214 ± 0.00046	0.00027 ± 0.00001	NA	NA	NA
Three protein genes							
cox1	1434	22	0.00225 ± 0.00063	0.00012 ± 0.00002	0.00037 ± 0.00014	0.00004 ± 0.00002	dS > dN: 0.011
cox3	753	20	0.00391 ± 0.00112	0.00115 ± 0.00008	0.00098 ± 0.00057	0.00120 ± 0.00081	dN > dS: 0.416
cob	1131	15	0.00181 ± 0.00058	0.00012 ± 0.00002	0.00028 ± 0.00011	0.00008 ± 0.00004	dS > dN: 0.045
cox1 + *cox3* + *cob*	3318	57	0.00248 ± 0.00056	0.00035 ± 0.00002	0.00048 ± 0.00016	0.00031 ± 0.00018	dS > dN: 0.247
[a]*serca* + *adsl*	5043	62	0.00181 ± 0.00023	0.00035 ± 0.00002	0.001107 ± 0.00049	0.00016 ± 0.00006	dS > dN: 0.034

The numbers of samples analyzed were 516 and 453 for mitochondrial genome and *serca* + *adsl*, respectively.
Abbreviations: *cox1*, cytochrome *c* oxidase subunit 1 gene; *cox3*, cytochrome *c* oxidase subunit 3 gene; *cob*, cytochrome *b* gene.
[a] For comparison these nuclear genome-encoded housekeeping genes are included in the analysis.

Author's personal copy

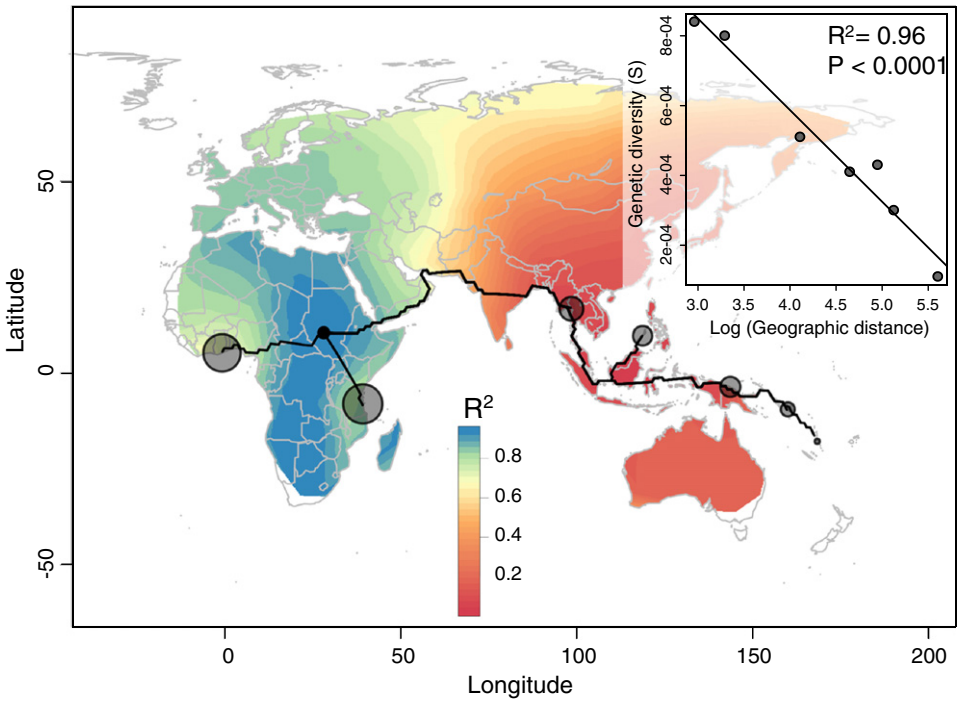

Fig. 1. African origin for *P. falciparum* and out of Africa migration over Southeast Asia and Oceania postulated from within-population diversity of the parasite mitochondrial genome. Gray circles are the geographic regions of seven parasite populations: Ghana, Tanzania, Thailand, Philippines, Papua New Guinea, Solomon Islands, and Vanuatu. The size of the circles is proportional to within-population diversity (θ_S) of the parasite mitochondrial genome. The background color represents the strength of the correlation between geographic distance from various origins and θ_S, with dark blue representing more likely origins. A black dot in Africa represents the best supported origin. The postulated colonization routes from the origin through landmasses to all parasite populations analyzed are shown on the map. The inset shows the correlation between geographic distance and within-population genetic diversity for the best supported origin.

SEA and Oceania. A significant negative correlation was, however, not detected when nucleotide diversity was expressed by θ_π.

3.3. Geographical distribution of SNPs

Almost all SNPs detected in the mt genome were those with minor allele frequency (MAF) <5% (84/88 = 95%), and mostly (63%) singletons (Table 2). The number of SNPs unique to either Africa or SEA + Oceania was very high: 95% (=41/43) and 96% (=45/47), respectively. Consistently, very limited numbers of SNPs (2/47 = 4%) were shared between Africa and non-Africa. We also saw in *serca* + *adsl* a high frequency of SNPs with MAF <5% (58/62 = 94%), although the number of SNPs shared between Africa and non-Africa in the two nuclear genes (40% = 8/20) was significantly higher than that in the mt genome (P < 0.001).

To estimate geographic differentiation of parasite populations, the apportionment of SNPs among populations was analyzed using Arlequin. Estimated F_{ST} value among all seven populations was 0.23, indicating that the majority (77%) of genetic variation was apportioned within populations (Supplementary Fig. 1). F_{ST} values between three major geographic regions/continents (Africa, Southeast Asia and Oceania) were generally higher than that among populations within these continents (Supplementary Fig. 1 & Supplementary Table 3). Moreover, the observed F_{ST} values for overall populations, between the continents and among populations within the continents, appeared to be higher in the mt genome than for *serca* + *adsl* (Supplementary Fig. 1), except between Ghana and Tanzania, where population differentiation was not significant

Table 2
Geographical distribution of SNPs in the *P. falciparum* mitochondrial genome.

	Africa	Southeast Asia	Oceania	Southeast Asia + Oceania	Overall
Mitochondrial genome					
No. of SNPs	43	25	26	47	88
No. of singletons	34	13	12	22	55
No. of SNPs with MAF < 5%[a]	42	22	22	42	84
No. of geographically unique SNPs	41	21	21	45	–
SNPs shared with Africa	–	1	2	2	–
Nuclear housekeeping genes:serca + adsl					
No. of SNPs	50	16	13	20	62
No. of singletons	32	5	1	6	36
No. of SNPs with MAF < 5%[a]	44	12	9	17	58
No. of geographically unique SNPs	42	4	4	12	–
SNPs shared with Africa	–	8	5	8	–

Colored data highlights the preponderance of minor allele frequency and singletons in mt genome vis-a-vis to housekeeping genes.
[a]MAF < 5%: minor allele frequency of less than 5%.

Author's personal copy

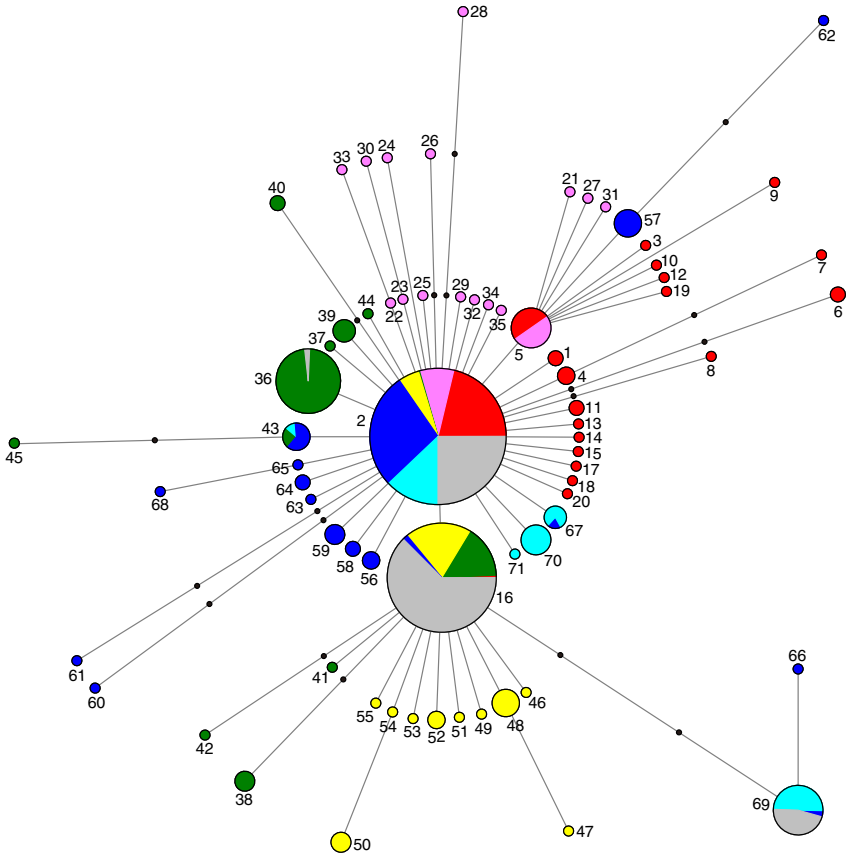

Fig. 2. Haplotype network tree showing genetic relationships among 516 *P. falciparum* mitochondrial genome sequences from seven countries in Africa, Southeast Asia and Oceania. The tree shows 71 haplotypes (nodules) based on 88 SNPs in the genome sequences analyzed (5883 base pairs). Each line represents one mutational step and black dots are hypothetical missing intermediates. The size of the circles is proportional to haplotype frequency. Node colors indicate the geographic origins of the isolates: red, Tanzania; pink, Ghana; green, Thailand; yellow, Philippines; dark blue, Papua New Guinea; light blue, Solomon Islands; gray, Vanuatu. Numbers refer to haplotype ID numbers in Supplementary Fig. 2.

for both mitochondrial and *serca* + *adsl* genes. Hence, there appears to be relatively higher population differentiation in the mt genome compared to *serca* + *adsl*.

3.4. Geographic distribution of mt genome haplotypes

Haplotype network analysis was used to assess the geographic distribution of mt genome haplotypes. Among 71 haplotypes identified here (Supplementary Fig. 2), the majority of haplotypes (64 haplotypes) were those of singletons (n = 49) or of low frequency, and were unique to each of the seven countries except for hap-2 (n = 188) and hap-16 (n = 122) (Fig. 2). The haplotype network had a star-like structure with hap-2 positioned at the center and distributed in all seven populations. Moreover, haplotypes with a single mutation that can be traced back to hap-2 are distributed in all seven countries. This haplotype had the highest outgroup probability (0.1871) (Supplementary Table 4), suggesting that hap-2 is an ancestral haplotype. Hap-16 had the second highest outgroup probability (0.1416) and is distributed in five countries in Southeast Asia and Oceania. Importantly, no haplotype that contained only African isolates radiate off from hap-16 or other haplotypes that contain only isolates from Southeast Asia and/or Oceania. Some degree of geographical clustering suggests that from the most likely ancestral hap-2, this haplotype spread from Africa to Southeast Asia and Oceania; from whence other Southeast Asian and Oceania haplotypes originated. A separate migration from Africa is associated with hap-5, which has a single-step connection to PNG, suggesting multiple unrelated founding events in PNG as reported previously (Joy et al., 2003).

4. Discussion

The analysis of mt genomes is useful in evolutionary and phylogeographic studies but its uniparental inheritance may affect geographic distribution of within-population genetic diversity in *P. falciparum*. In this context, we assessed whether within-population genetic diversity in the mt genome would reflect the observed geographic patterns of isolation by distance seen in two nuclear genes using the same parasite populations (Tanabe et al., 2010). The results obtained suggested a sub-Saharan African origin of the *P. falciparum* mt genome and recovered a strong negative association between within-population mt SNP diversity (θ_S) and geographic distance from the African origin to Southeast Asia and Oceania. We note that more extensive geographic sampling would be useful in future studies, particularly expanding the populations from Africa and acquiring isolates from other regions. Our results, nevertheless, agree well with our analysis of two nuclear housekeeping genes, *serca* + *adsl* (Tanabe et al., 2010). Thus, the present study strongly suggests that the pattern of inheritance, irrespective of uniparental (mt genome) or biparental (nuclear genome), does not appear to substantially affect the geographic distribution of within-population genetic diversity of *P. falciparum*.

While both θ_S and θ_π recovered negative correlation between geographic distance and within-population diversity in the two

Author's personal copy

nuclear housekeeping genes (Tanabe et al., 2010), we also note that a negative correlation was not recovered when nucleotide diversity was represented by θ_π. We consider that the absence of recombination events together with strong functional and structural constraints in the mt genome probably contributes to this difference. In the two nuclear housekeeping genes, the ratio of θ_π to θ_S was moderately constant in the seven parasite populations studied (Tanabe et al., 2010). In contrast, in the mt genome the ratio varied considerably among populations and was particularly low in the two African populations (Supplementary Table 2). In Africa, virtually all mt SNPs (42/43) were those with MAF <5%, (Table 2) rendering θ_π values very low and comparable with those in non-African populations (Supplementary Table 2). The frequency of SNPs with MAF <5% in the two housekeeping genes in Africa was relatively low (44/50) compared with that of the mt genome (Table 2). Our study demonstrated a high mt sequence similarity (99.8%) among isolates, suggesting strong functional and/or structural constraints in the mt genome. In the 6 kb *P. falciparum* mt genome, the three protein coding genes and rRNA gene fragments are densely packed with few intergenic sequence regions (Feagin et al., 1997), where sequences are also highly conserved. A recent study also demonstrated that almost all the intergenic regions are transcribed and some of the transcriptional regions contain rRNA gene fragments (Feagin et al., 2012). In such conserved regions and protein coding regions, most new mutations would likely cause a deleterious effect on the function and/or structure of the mt genome and as a consequence SNP(s) already present in the parasite mt genome would disappear through genetic hitchhiking because the mt genome does not undergo recombination. This would result to a high frequency of SNPs with MAF <5% and low θ_π values. In *P. falciparum*, a chromosomal recombination segment in the nuclear genome is fairly short, particularly in African parasite populations (Conway et al., 1999; Neafsey et al., 2008; Tanabe et al., 2007), likely making the frequency of SNPs with MAF <5% relatively low and θ_π values high. These factors may likely account for the absence of a negative correlation between within-population genetic diversity, expressed as θ_π, and geographical distance in mt the genome, in contrast with the parasite nuclear genome that undergoes recombination where genetic hitchhiking is restricted to a recombination segment. Further studies utilizing more African populations could confirm this hypothesis.

The high frequency of geographically unique SNPs observed in the mt genome clearly contributes to a high geographic differentiation among all parasite populations studied here, with the exception of Ghana and Tanzania. Gene flow between the two African populations may have prevented the formation of a population structure. The high frequency of geographically unique SNPs also contributes to a high frequency of geographically unique mt genome halotypes, so that we can infer the demographic history of the *P. falciparum* mt genome through haplotype network analysis. We confirmed a star-like network, which suggests a recent expansion of *P. falciparum* mt genome (Joy et al., 2003). The network suggested an ancestral haplotype that originated in Africa, although the precise origin in Africa remains to be elucidated. The postulated ancestral haplotype with the highest outgroup probability, hap-2, was the same as hap-1 in Fig. 1 of Joy et al. (2003). An African origin supports the one inferred from geographical distribution of SNP diversity in this and previous study (Tanabe et al., 2010). We consider that a limited number of mt haplotypes that expanded and spread out of Africa suggest severe bottlenecks along with migration from Africa to Southeast Asia and Oceania.

In the present data set, we did not see a mutation at codon 268 of cytochrome *b* gene. The relevance of codon 268 polymorphisms in vivo as potential resistance marker has been reported (Schwöbel et al., 2003). However, this absence of polymorphism is not unexpected because the samples used here were collected before the usage of ATQ. A previous large scale survey in Senegal, Ivory Coast, Central African Republic and Madagascar (Ekala et al., 2007), as well as other surveys in Tanzania (Schönfeld et al., 2007) and Ghana (Muehlen et al., 2004), also revealed the absence of this mutation. Moreover, the drug, commonly introduced in combination with proguanil (distributed under the trade name of Malarone) is currently limited to travelers/tourists from industrialized countries owing to its high cost. We also note that the low frequency of cytochrome *b* polymorphism is consistent with strong conservation of mt genes linked with functions/structure crucial for parasite survival. Even if the mutation is generated spontaneously, its deleterious effect on the protein and/or rRNA and the associated fitness cost may lead to a quick elimination from the parasite population (Peters et al., 2002).

5. Conclusion

Analysis of within-population SNP diversity in the mt genome of 516 isolates from seven parasite populations in Africa, Southeast Asia and Oceania postulated a sub-Saharan African origin and recovered a strong negative correlation between within-population SNP diversity and geographical distance from the postulated origin over Africa, Southeast Asia and Oceania. These results are in good agreement with those obtained for two nuclear housekeeping genes, *serca* and *adsl*, suggesting that the pattern of inheritance and the lack of recombination do not substantially affect the geographical distribution of mt SNPs. This study extends our argument (Tanabe et al., 2010, 2013) that within-population genetic diversity of *P. falciparum* is largely determined by ancient, repeated bottlenecks during which pathogen populations progressively lost diversity as they spread out of Africa rather than recent regional variation in transmission intensity, malaria interventions, levels of host immunity, or modes of mt or nuclear genome inheritance. We observed notably high frequency of geographically unique SNPs in the mt genome, compared with nuclear genome-encoded housekeeping genes. This can probably be ascribed to the lack of recombination in the mt genome. Additionally, the present haplotype network analysis supported an African origin of *P. falciparum*. Finally, we did not detect a mutation at 268 of cytochrome *b* in the present samples, suggesting that the mutation can only survive under the pressure of ATQ.

Acknowledgments

We thank Anna Färnert, Anders Björkman, Kenji Hirayama, Hiroshi Ohmae, and Akira Kaneko for providing parasite isolates, Jun Ohashi for useful suggestions and Naoko Sakihama for technical assistance. This work was supported by Grant-in-Aids for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan (18073013, 23659211 and 23590498); from the Japan Society for Promotion of Sciences (18GS03140013); and the Foundation of Strategic Research Projects in Private Universities (S0991013). TJ is funded by the NIGMS MIDAS initiative and the Medical Research Council.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.mito.2013.08.008.

References

Anderson, T.J., Haubold, B., Williams, J.T., Estrada-Franco, J.G., Richardson, L., Mollinedo, R., Bockarie, M., Mokili, J., Mharakurwa, S., French, N., Whitworth, J., Velez, I.D., Brockman, A.H., Nosten, F., Ferreira, M.U., Day, K.P., 2000. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. Mol. Biol. Evol. 17, 1467–1482.

Bandelt, H.J., Forster, P., Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16, 37–48.

Clement, M., Posada, D., Crandall, K.A., 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9, 1657–1660.

Conway, D.J., Roper, C., Oduola, A.M.J., Arnot, D.E., Kremsner, P.G., Grobusch, M.P., Curtis, C.F., Greenwood, B.M., 1999. High recombination rate in natural populations of *Plasmodium falciparum*. Proc. Natl. Acad. Sci. U. S. A. 96, 4506–4511.

Conway, D.J., Fanello, C., Lloyd, J.M., Al-Joubori, B.M., Baloch, A.H., Somanath, S.D., Roper, C., Oduola, A.M., Mulder, B., Povoa, M.M., Singh, B., Thomas, A.W., 2000. Origin of *Plasmodium falciparum* malaria is traced by mitochondrial DNA. Mol. Biochem. Parasitol. 111, 163–171.

Author's personal copy

Creasey, A.M., Ranford-Cartwright, L.C., Moore, D.J., Williamson, D.H., Wilson, R.J., Walliker, D., Carter, R., 1993. Uniparental inheritance of the mitochondrial gene cytochrome b in *Plasmodium falciparum*. Curr. Genet. 23, 360–364.

Daily, J.P., Scanfeld, D., Pochet, N., Le Roch, K., Plouffe, D., Kamal, M., Sarr, O., Mboup, S., Ndir, O., Wypij, D., Levasseur, K., Thomas, E., Tamayo, P., Dong, C., Zhou, Y., Lander, E.S., Ndiaye, D., Wirth, D., Winzeler, E.A., Mesirov, J.P., Regev, A., 2007. Distinct physiological states of *Plasmodium falciparum* in malaria-infected patients. Nature 450, 1091–1095.

Ekala, M.-T., Khim, N., Legrand, E., Randrianarivelojosia, M., Jambou, R., Fandeur, T., Menard, D., Assi, S.-B., Henry, M.-C., Rogier, C., Bouchier, C., Mercereau-Puijalon, O., 2007. Sequence analysis of *Plasmodium falciparum* cytochrome b in multiple geographic sites. Malar. J. 6, 164.

Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin (version 3.0): an integrated soft- ware package for population genetics data analysis. Evol. Bioinforma. Online 1, 47–50.

Feagin, J.E., Mericle, B.L., Werner, E., Morris, M., 1997. Identification of additional rRNA fragments encoded by the *Plasmodium falciparum* 6 kb element. Nucleic Acids Res. 25, 438–446.

Feagin, J.E., Harrell, M.I., Lee, J.C., Coe, K.J., Sands, B.H., Cannone, J.J., Tami, G., Schnare, M.N., Gutell, R.R., 2012. The fragmented mitochondrial ribosomal RNAs of *Plasmodium falciparum*. PLoS One 7, e38320.

Fivelman, Q.L., Butcher, G.A., Adagu, I.S., Warhurst, D.C., Pasvol, G., 2002. Malarone treatment failure and *in vitro* confirmation of resistance of *Plasmodium falciparum* isolate from Lagos, Nigeria. Malar. J. 1, 1.

Handley, L.J.L., Manica, A., Goudet, J., Balloux, F., 2007. Going the distance: human population genetics in a clinal world. Trends Genet. 23, 432–439.

Hay, S.I., Guerra, C.A., Gething, P.W., Patil, A.P., Tatem, A.J., Noor, A.M., Kabaria, C.W., Manh, B.H., Elyazar, I.R., Brooker, S., Smith, D.L., Moyeed, R.A., Snow, R.W., 2009. A world malaria map: *Plasmodium falciparum* endemicity in 2007. PLoS Med. 6, e1000048.

Hikosaka, K., Watanabe, Y., Kobayashi, F., Waki, S., Kita, K., Tanabe, K., 2011. Highly conserved gene arrangement of the mitochondrial genomes of 23 *Plasmodium* species. Parasitol. Int. 60, 175–180.

Hino, A., Hirai, M., Tanaka, T.Q., Watanabe, Y., Matsuoka, H., Kita, K., 2012. Critical roles of the mitochondrial complex II in oocyst formation of rodent malaria parasite *Plasmodium berghei*. J. Biochem. 152, 259–268.

Joy, D.A., Feng, X., Mu, J., Furuya, T., Chotivanich, K., Krettli, A.U., Ho, M., Wang, A., White, N.J., Suh, E., Beerli, P., Su, X.Z., 2003. Early origin and recent expansion of *Plasmodium falciparum*. Science 300, 318–321.

Kelly-Hope, L.A., McKenzie, F.E., 2009. The multiplicity of malaria transmission: a review of entomological inoculation rate measurements and methods across sub-Saharan Africa. Malar. J. 8, 19.

Korsinczky, M., Chen, N., Kotecka, B., Saul, A., Rieckmann, K., Cheng, Q., 2000. Mutations in *Plasmodium falciparum* cytochrome b that are associated with atovaquone resistance are located at a putative drug-binding site. Antimicrob. Agents Chemother. 44, 2100–2108.

Krungkrai, J., 2004. The multiple roles of the mitochondrion of the malarial parasite. Parasitology 129, 511–524.

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452.

Liu, H., Prugnolle, F., Manica, A., Balloux, F., 2006. A geographically explicit genetic model of worldwide human-settlement history. Am. J. Hum. Genet. 79, 230–237.

Manica, A., Prugnolle, F., Balloux, F., 2005. Geography is a better determinant of human genetic differentiation than ethnicity. Hum. Genet. 118, 366–371.

Muehlen, M., Schreiber, J., Ehrhardt, S., Otchwemah, R., Jelinek, T., Bienzle, U., Mockenhaupt, F., 2004. Prevalence of mutations associated with resistance to atovaquone and to the antifolate effect of proguanil in *Plamsodium falciparum* isolates from northern Ghana. Trop. Med. Int. Health 9, 361–363.

Musset, L., Le Bras, J., Clain, J., 2007. Parallel evolution of adaptive mutations in *Plasmodium falciparum* mitochondrial DNA during atovaquone–proguanil treatment. Mol. Biol. Evol. 24, 1582–1585.

Neafsey, D.E., Schaffner, S.F., Volkman, S.K., Park, D., Montgomery, P., Milner Jr., D.A., Lukens, A., Rosen, D., Daniels, R., Houde, N., Cortese, J.F., Tyndall, E., Gates, C., Stange-Thomann, N., Sarr, O., Ndiaye, D., Ndir, O., Mboup, S., Ferreira, M.U., Moraes, S.do L., Dash, A.P., Chitnis, C.E., Wiegand, R.C., Hartl, D.L., Birren, B.W., Lander, E.S., Sabeti, P.C., Wirth, D.F., 2008. Genome-wide SNP genotyping highlights the role of natural selection in *Plasmodium falciparum* population divergence. Genome Biol. 9, R171.

Peters, J.M., Chen, N., Gatton, M., Korsinczky, M., Fowler, E.V., Manzetti, S., Saul, A., Cheng, Q., 2002. Mutations in Cytochrome *b* resulting in atovaquone resistance are associated with loss of fitness in *Plasmodium falciparum*. Antimicrob. Agents Chemother. 46, 2435–2441.

Preiser, P.R., Wilson, R.J., Moore, P.W., McCready, S., Hajibagheri, M.A., Blight, K.J., Strath, M., Williamson, D.H., 1996. Recombination associated with replication of malarial mitochondrial DNA. EMBO J. 115, 684–693.

Prugnolle, F., Manica, A., Balloux, F., 2005. Geography predicts neutral genetic diversity of human populations. Curr. Biol. 15, R159–R160.

Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W., Cavalli-Sforza, L.L., 2005. Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. Proc. Natl. Acad. Sci. U. S. A. 102, 15942–15947.

Sakihama, N., Kimura, M., Hirayama, K., Kanda, T., Na-Bangchang, K., Jongwutiwes, S., Conway, D., Tanabe, K., 1999. Allelic recombination and linkage disequilibrium within *Msp-1* of *Plasmodium falciparum*, the malignant human malaria parasite. Gene 230, 47–54.

Sakihama, N., Kaneko, A., Hattori, T., Tanabe, K., 2001. Limited recombination events in merozoite surface protein-1 alleles of *Plasmodium falciparum* on islands. Gene 279, 41–48.

Sakihama, N., Ohmae, H., Bakote'e, B., Kawabata, M., Hirayama, K., Tanabe, K., 2006. Limited allelic diversity of *Plasmodium falciparum* merozoite surface protein 1 gene from populations in the Solomon Islands. Am.J.Trop. Med. Hyg. 74, 31–40.

Sakihama, N., Nakamura, M., Palanca Jr., A.A., Argubano, R.A., Realon, E.P., Larracas, A.L., Espina, R.L., Tanabe, K., 2007. Allelic diversity in the merozoite surface protein 1 gene of *Plasmodium falciparum* on Palawan Island, the Philippines. Parasitol. Int. 56, 185–194.

Schönfeld, M., Miranda, I.B., Schunk, M., Maduhu, I., Maboko, L., Hoelscher, M., Berens-Riha, N., Kitua, A., Löscher, T., 2007. Molecular surveillance of drug-resistance associated mutations of *Plasmodium falciparum* in south-west Tanzania. Malar. J. 6, 2.

Schwöbel, B., Alifrangis, M., Salanti, A., Jelinek, T., 2003. Different mutation patterns of atovaquone resistance to *Plasmodium falciparum in vitro* and *in vivo*: rapid detection of codon 268 polymorphisms in the cytochrome b as potential *in vivo* resistance marker. Malar. J. 2, 5.

Srivastava, I.K., Morrisey, J.M., Darrouzet, E., Daldal, F., Vaidya, A.B., 1999. Resistance mutations reveal the atovaquone-binding domain of cytochrome b in malaria parasites. Mol. Microbiol. 33, 704–711.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28, 2731–2739.

Tanabe, K., Sakihama, N., Rooth, I., Björkman, A., Färnert, A., 2007. High frequency of recombination-driven allelic diversity and temporal variation of *Plasmodium falciparum msp1* in Tanzania. Am.J.Trop. Med. Hyg. 76, 1037–1045.

Tanabe, K., Mita, T., Jombart, T., Eriksson, A., Horibe, S., Palacpac, N., Ranford-Cartwright, L., Sawai, H., Sakihama, N., Ohmae, H., Nakamura, M., Ferreira, M.U., Escalante, A.A., Prugnolle, F., Björkman, A., Färnert, A., Kaneko, A., Horii, T., Manica, A., Kishino, H., Balloux, F., 2010. *Plasmodium falciparum* accompanied the human expansion out of Africa. Curr. Biol. 20, 1283–1289.

Tanabe, K., Zakeri, S., Palacpac, N.M., Afsharpad, M., Randrianarivelojosia, M., Kaneko, A., Marma, A.S., Horii, T., Mita, T., 2011. Spontaneous mutations in the *Plasmodium falciparum* sarco/endoplasmic reticulum Ca^{2+}-ATPase (PfATP6) gene among wide geographical parasite populations unexposed to artemisinin-based combination therapies. Antimicrob. Agents Chemother. 55, 94–100.

Tanabe, K., Mita, M., Palacpac, N.M.Q., Arisue, N., Tougan, T., Kawai, S., Jombart, T., Kobayashi, F., Horii, T., 2013. Within-population genetic diversity of *Plasmodium falciparum* vaccine candidate antigens reveals geographic distance from a Central sub-Saharan African origin. Vaccine 31, 1334–1339.

Vaidya, A.B., Mather, M.W., 2009. Mitochondrial evolution and functions in malaria parasites. Annu. Rev. Microbiol. 63, 249–267.

Vaidya, A.B., Morrrisey, J., Plowe, C.V., Kaslow, D.C., Wellems, T.E., 1993. Unidirectional dominance of cytoplasmic inheritance in two genetic crosses of *Plasmodium falciparum*. Mol. Cell. Biol. 13, 7349–7357.

World Health Organization, 2012. World Malaria Report 2012.