Triosephosphate isomerase gene promoter variation: -5G/A and -8G/A polymorphisms in clinical malaria groups in two African populations

Mónica Guerra[a], Patricia Machado[a], Licínio Manco[b], Natércia Fernandes[c], Juliana Miranda[d], Ana Paula Arez[a]

[a]Centro de Malária e outras Doenças Tropicais, Unidade de Parasitologia Médica, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Portugal

[b]Centro de Investigação em Antropologia e Saúde (CIAS), Universidade de Coimbra, Portugal

[c]Faculdade de Medicina da Universidade Eduardo Mondlane, Maputo, Mozambique

[d]Hospital Pediátrico David Bernardino, Luanda, Angola

Corresponding author

AP Arez. Centro de Malária e outras Doenças Tropicais, Unidade de Parasitologia Médica, Instituto de Higiene e Medicina Tropical, Rua da Junqueira, 100, 1349-008 Lisboa, Portugal

E-mail: aparez@ihmt.unl.pt; Tel.: +351 213652657; Fax: +351 213652699

ABSTRACT

TPI1 promoter polymorphisms occur in high prevalence in individuals from African origin. Malaria-patients from Angola and Mozambique were screened for the *TPI1* gene promoter variants rs1800200A>G, (-5G>A), rs1800201G>A, (-8G>A), rs1800202T>G, (-24T>G), and for the intron 5 polymorphism rs2071069G>A, (2262G>A). -5G>A and -8G>A variants occur in 47% and 53% in Angola and Mozambique, respectively while -24T>G was monomorphic for the wild-type T allele. Six haplotypes were identified and -8A occurred in 45% of the individuals, especially associated with the GAG haplotype and more frequent in non-severe malaria groups, although not significantly. The arising and dispersion of -5G>A and -8G>A polymorphisms is controversial. Their age was estimated by analyses of two microsatellite loci, CD4 and ATN1, adjacent to *TPI1* gene. The -5G>A is older than -8G>A, with an average estimate of approximately 35,000 years. The -8A variant arose in two different backgrounds, suggesting independent mutational events. The first, on the -5G background, may have occurred in East Africa around 20,800 years ago; the second, on the -5A background, may have occurred in West Africa some 7,500 years ago. These estimates are within the period of spread of agriculture and the malaria mosquito vector in Africa, which could has been a possible reason for the selection of -8A polymorphism in malaria endemic countries.

Keywords: human malaria, selection signatures, triosephosphate isomerase-deficiency, *TPI1* gene promoter variants

Highlights

Four SNPs variants of *TPI1* were analyzed in malaria patients.

Two STRs were analyzed for evaluated haplotype diversity and antiquity of *TPI1* promoter variants.

TPI1 -8A allele was more frequent in non-severe malaria groups.

The age estimate for -8 variant are within the period of origin the malaria mosquito vector in Africa.

TPI1 polymorphic variants could have been due to a selective advantage against malaria.

1. INTRODUCTION

The enzyme triosephosphate isomerase (TPI1; EC 5.3.1.1) is a housekeeping glycolytic enzyme required for cell maintenance. It catalyzes the isomerization of glyceraldehyde-3-phosphate (G3P) and dihydroxyacetone phosphate (DHAP) and plays an essential role in glycolysis, gluconeogenesis, fatty acid synthesis and pentose-phosphate pathways (Maquat et al. 1985).

TPI1 deficiency (MIM #615512) (http://www.omim.org/) is a rare autosomal recessive multisystem disorder, characterized by a decrease in enzyme activity in all tissues. In erythrocytes, that consume glucose to produce ATP, the conversion of DHAP to G3P is assured exclusively by TPI1 and this reduced activity is accompanied by an increase of DHAP (Schneider et al. 1965).

Heterozygous individuals with one of the promoter polymorphisms in the *TPI1* gene (MIM #190450, 12p13.31, GenBank NM_000365.5) are asymptomatic with intermediate enzyme activity (about 50% of normal TPI1 activity), but homozygous and compound heterozygous patients suffer from a strong reduction of enzyme activity (approximately 2–20% of normal), which manifests not only by congenital hemolytic anemia, but also by an increased susceptibility to infections, cardiomyopathy and progressive neuromuscular impairment (Schneider et al. 1965; Orosz et al. 2009). TPI1 deficiency is described as the most severe glycolytic enzyme abnormality and, almost all cases have ended in death in the fetal period or before age of five (Orosz et al. 2009). There is no effective therapy available.

Mohrenweiser and Fielek (1982) analyzed the incidence of heterozygous TPI1 deficiency in an interethnic population and reported allele frequencies of 0.024 in African-Americans, ten times the frequency of 0.0024 observed in individuals of non-African origin. Watanabe et al. (1996), analyzing the same samples, identified three different polymorphisms within or in close proximity to *cis*-active regulatory elements in *TPI1* gene promoter: rs1800200A>G SNP (-5G>A), rs1800201G>A SNP (-8G>A) within the cap proximal element (CPE), and rs1800202T>G SNP (-24T>G) within the TATA box. The -5G>A substitution had no effect on TPI1 enzyme activity, in contrast to the -8G>A and -24T>G substitutions, associated with a progressive reduction of enzyme activity (Schneider et al. 1998).

Variant *TPI1* promoter alleles were found to be geographically widespread. However, while the -5G minor allele has global geographical distribution and has attained high frequency in African, Caribbean and Oriental populations, the -8A minor allele had been found in

geographically dispersed populations, being more frequent in individuals of African origin (Humphries et al. 1999; Manco et al. 2009).

Humphries et al. (1999) reported that -5G and associated haplotypes were found to be in linkage disequilibrium (LD) with an intragenic polymorphism at nucleotide 2262 in intron 5 (Arya et al. 1997), which could be indicative of a common ancestral origin.

Schneider et al. (1998) suggests that the high incidence of these *TPI1* promoter polymorphisms in African-origin populations, associated to a reduced enzyme activity, may have resulted from some selective advantage for survival similarly to glucose-6-phosphate dehydrogenase (G6PD) deficient variants (Luzzatto, 2006) and pyruvate kinase (PK) deficient variants (Machado et al. 2010, 2012) and hemoglobin polymorphisms (HbS) (Fleming et al. 1979). These are reminiscent examples of genetic diversity in African populations associated to protection of human host against malaria infection and resulting from the selective pressure that malaria has imposed to human populations.

Following this rationale, we investigated *TPI1* promoter variants -5G>A, -8G>A, -24T>G and rs2071069G>A SNP in intron 5 (2262G>A), in infected malaria individuals from Angola and Mozambique, grouped according to the severity of malaria disease outcome. In addition, to explore their natural history we evaluated haplotype diversity and estimated their relative antiquity using microsatellites close to the *TPI1* gene.

2. MATERIALS AND METHODS

2.1 SAMPLING

Peripheral whole blood samples were collected from unrelated individuals (mainly children under or equal to 15 years-old; only 7 adults with more or equal than 16 years-old in the asymptomatic group from Maputo, Mozambique) from two malaria endemic sub-Saharan African areas, Angola (N=124 from Luanda, Machado et al. 2010) and Mozambique (N=250 from Maputo, Machado et al. 2010, 2012 and N=13 from Manhiça district, Marques et al. 2005).

These were all *P. falciparum* infected individuals with different malaria outcomes, defined as follows: (i) Severe malaria (SM): slide positive for blood-stage asexual *P. falciparum* at any parasite density, fever (axillary temperature ≥37,5ºC), hemoglobin level of Hb<50 g/l and/or other symptoms, such as coma, prostration or convulsions; (ii) Uncomplicated malaria (UM): slide positive for blood-stage asexual *P. falciparum* at any parasite density, fever (axillary temperature ≥37,5ºC) and hemoglobin level of Hb>50 g/l; and (iii) Asymptomatic infection

(AI): slide positive for blood-stage asexual *P. falciparum* at any parasite density in the absence of fever or other symptoms of clinical illness.

The 124 individuals from Angola belong to the following malaria outcome groups: 44 SM, 43 UM and 37 AI (Machado et al. 2010). Same criteria were used to group the 263 isolates from Mozambique: 93 SM, 150 UM and 20 AI (Marques et al. 2005; Machado et al. 2010, 2012).

Blood was collected after the clinician examination and before the administration of any antimalarial drug and/or blood transfusion (individuals who had received a blood transfusion in the last six months were excluded from the study).

2.2 ETHICS STATEMENT

As reported before in Marques et al. (2005) and Machado et al. (2010, 2012), the human isolates collection was approved by Ministry of Public Health of Angola and Mozambique and by the local Ethical Committees at the institutions involved in those studies. Informed consent was obtained from all individuals.

2.3 DNA AMPLIFICATION AND GENOTYPING OF THE *TPI1* GENE PROMOTER REGION

The promoter region of *TPI1* gene was amplified following the PCR conditions and primers described by Schneider et al. (1998). The PCR product (258bp) was digested with the restriction enzymes *TseI*, *MscI* and *SfcI* from New England Biolabs (Baverly, MA) to examine -5G>A, -8G>A and -24T>G polymorphisms, respectively (Schneider et al. 1998) (Fig. 1). DNA fragments were resolved by electrophoresis in 2% agarose gel stained with ethidium bromide under UV transillumination.

2.4 ANALYSIS OF *TPI1* GENE INTRON 5 POLYMORPHISM

The polymorphism at 2262 nucleotide (Arya et al. 1997) (Fig. 1) was identified by sequencing. DNA was amplified by PCR using the tagged sense primer (5'-TGGCTGGAGAGCTCTTTCTT-3') and the antisense primer (5'-AGCCCACTCCACCTCAGC-3') in a 25μL reaction containing 1x GoTaq® Flexi Buffer (Promega, USA), 3mM $MgCl_2$, 0.025uM of each dNTP, 0.08 uM of each primer, 0.5 U of Taq polymerase GoTaq® Flexi DNA polymerase (Promega, USA) and 100 to 150 ng genomic DNA. PCR was conducted for 94ºC for 3 min by denaturation initially, followed by 35 cycles of 94ºC for 45 s, 62ºC for 45 s, 72ºC for 45 s and a final extension step of 72ºC for

7 min. Double-strand sequencing of PCR products was performed by Macrogen (Europe). Sequences were analyzed with Chromas Lite software version 2.01 to identify specific variation.

2.5 MICROSATELLITES TYPING

Two microsatellite loci were analyzed: a pentanucleotide repeat CD4-(CTTTT)n, located about 79 Kb upstream from the *TPI1* gene within the *CD4* gene (MIM #186940, 12p13.31, GenBank NM_000616.4) (Edwards et al. 1991), and a trinucleotide repeat region ATN1-(CAG)n, located about 70 Kb downstream of the *TPI1* gene within the protein-encoding sequence of the atrophin-1 gene (*ATN-1*) (MIM #607462, 12p13.31, GenBank NM_001007026.1) (Koide et al. 1994) (Fig. 1).

Each locus was amplified separately in a 25µl PCR reaction that contained 1× GoTaq® Flexi Buffer (Promega, USA), 3 mM $MgCl_2$, 0.025 mM of each dNTP, 0.08 µM of each primer [CD4 (sense) 5'-TTGGAGTCGCAAGCTGAACTAGAG-3' and CD4 (antisense) 5'-CCAGGAAGTTGAGGCTGCAGTGAA-3'; or ATN1 (sense) 5'-CCCCTTCCCTCCCTCTACT-3' and ATN1 (antisense) 5'-GAGACATGGCGTAAGGGTGT-3'], 0.5 U of Taq polymerase GoTaq® Flexi DNA polymerase (Promega, USA) and 100 ng genomic DNA. For each locus, one of the primers was fluorescently labelled (FAM or HEX) (Eurofins MWG Operon). Thermal cycling was initially at 95ºC for 15 min, followed by 40 cycles of 94ºC for 45 s, annealing at 63ºC to CD4 and 60ºC to ATN1 for 1min, and 72°C for 1min. After a final extension step of 10 min at 72°C, reactions were stopped at 4°C. Amplified products were separated by capillary electrophoresis in a genetic analyzer ABI3730 (Applied Biosystems, USA) at the DNA Analysis Facility on Science Hill, Yale University (USA). Fragment sizes and genotypes were scored using the software GeneMarker 1.4. (Softgenetics, USA).

2.6 AGE ESTIMATION OF *TPI1* VARIANTS

To estimate the age of the most recent common ancestor of haplotypes from the different *TPI1* variants, we simulated the overtime decay in LD in a population of infinite size and calculated the microsatellite allele frequency distribution in each generation by using the following relation:

$$p_{(g,i)} = p_{(g-1,i)}\,(1-\mu-\theta) + \theta q_i + \left(\frac{\mu}{2}\right)\left[p_{(g-1,i-1)} + p_{(g-1,i+1)}\right] \quad (1)$$

where $p_{(g,i)}$ is the frequency of a marker microsatellite allele with i repeats in generation g of mutated chromosomes, *qi* is the frequency of that allele in the whole population, μ represent the microsatellite mutation rate, and θ the recombination fraction, as described in Seixas et al. (2001). The number of generations leading to a match with the observed frequency of the ancestral microsatellite allele within each protein variant was chosen as the estimate of age of mutation (Seixas et al. 2001). A microsatellite mutation rate (μ) of 0.001 and a correspondence of 1cM = 1Mb for the recombination fraction (θ) or 1 cM defined as θ = 0.01 (Ott 1999; Sudbery, 2002; Ulgen et al., 2005) were considered. Considering the distances of approximately 70Kb and 79kb between CD4 and *TPI1* gene and ATN1 and *TPI1* gene, respectively (UCSC Genome Browser), the recombination fractions are 0.00070 and 0.00079 for the ATN1 and CD4 locus, respectively.

2.7 STATISTICAL ANALYSIS

Allele frequency values, observed heterozygosity (Ho), expected heterozygosity (He), Hardy-Weinberg equilibrium probability value (HWE) and exact *P*-values for linkage disequilibrium, as well as the linkage phase from diploid data established by statistical inference via ELB algorithm, were obtained using the software package ARLEQUIN (Excoffier et al. 2005, version 3.1 accessed in 15-01-2014 at http://cmpg.unibe.ch/software/arlequin3/).

The r^2 values of linkage disequilibrium were calculated by software package PLINK (Purcell et al. 2007, accessed in 20-09-2014 at http://pngu.mgh.harvard.edu/purcell/plink/).

For sample differentiation between *TPI1* polymorphisms and clinical malaria groups, Person's Chi-square and Fisher exact probability tests were used and calculated with the on-line software SISA (Simple Interactive Statistical Analysis) (Uitenbroek, 1997), admitting a significance level of *P*=0.05. The genotypic frequencies were estimated by direct gene counting.

3. RESULTS

3.1 *TPI1* VARIANTS

Frequencies of *TPI1* promoter alleles (-5G/A and -8G/A) and 2262G/A nucleotide in different malaria outcome groups from Angola and Mozambique are shown in Table 1. The variant -24T>G is monomorphic for the wild-type T allele in both studied populations and the 2262 is the most polymorphic locus in both populations, showing similar frequencies of each

allele (G, 54% and A, 46%). For the -5 locus, the A allele was the most frequent occurring in 74 and 70% of the Angolan and Mozambican individuals, respectively. For the -8 polymorphism, the allele G presented high frequency in both populations (Angola: 92%; Mozambique: 93%).

Genotype distribution showed no deviations from the Hardy-Weinberg equilibrium (P>0.05) for all polymorphisms.

Seven genotypes for -5 and -8 loci were identified by direct individual counting (Table 2). Six of them were present in both populations: -5AA/-8GG, -5AG/-8GG, -5AG/-8GA, -5GG/-8GG, -5AA/-8GA and -5GG/-8GA, while the -5GG/-8AA homozygous genotype was only observed in Mozambique (1%, UM and SM groups) (Fig. 2). About 45% of the Angolan and Mozambican individuals show the -8A variant. In Angola 8%, 19% and 18% individuals from the AI, UM and SM groups, respectively, presented the –8A allele in both homozygous and heterozygous genotypes and in Mozambique it was present in 20%, 14% and 11% of the individuals in AI, UM and SM groups, respectively. However no statistical significant differences were found when sample differentiation between the three malaria outcome groups was tested for individual polymorphisms -5G>A or -8G>A (exact P values ranged between 0.11 and 1.00).

Six haplotypes were observed, the -5G/-8G/2262A (GGA, H5) haplotype being exclusive to Mozambique in the UM (2%) and SM (2%) groups (Table 3). Frequencies of haplotypes between malaria outcome groups were not significantly different (P>0.05) but the GAG haplotype (H4) was the most frequent within the UM groups (Angola: 8.1%; Mozambique: 7.3%) when compared to the SM groups (Angola: 6.8%; Mozambique: 5.9%).

Linkage disequilibrium between the three bi-allelic polymorphic loci was significant in all malaria groups from Mozambique. In Angola, this did not happen in the severe malaria group for the loci combinations -8/-5 and -8/2262 (Table 4). In concordance, high r^2 values of LD were observed for the loci combination -5/-2262 in all malaria groups, both considering the two populations separately or together (Table 4).

3.2 MICROSATELLITES

Ten and 12 CD4 alleles were identified in Angola (ranging in size from 137-182bp) and Mozambique (137-192bp), respectively; regarding ATN1, 14 alleles (197-242bp) were identified in Angola and Mozambique populations (Fig. 3). Isolates where amplification

failed were excluded from the analysis. The frequency distribution is similar in the two populations for both loci.

Allelic variation of CD4 and ATN1 microsatellites was estimated in Angola and Mozambique populations within *TPI1* gene variants (Fig. 4). Both microsatellites exhibited high allelic variation in association to -5A and -5G alleles. The -8A allele showed lower variability, being associated to six of the 10 and seven of the 12 CD4 alleles and to six and four of the 14 ATN1 alleles in Angola and Mozambique, respectively.

3.3 AGE OF *TPI1* PROMOTER VARIANTS

Mutations age for the single nucleotide variants -5G>A and -8G>A was estimated based on the expected STR mutation rate and decay of LD due to recombination between a STR polymorphic marker and the single nucleotide locus over the generations. Haplotype frequencies for the two microsatellites CD4 and ATN1 located respectively 70 Kb and 79 Kb downstream and upstream the *TPI1* gene were considered. For the -5 locus the -5G was considered as the ancestral allele because it was the most common allele found in non-human primates (chimpanzee) (http://www.ensembl.org) (Humphries et al. 1999). Estimates of mutation age for the *TPI1* promoter variants, -5G>A and -8G>A, based on CD4 and ATN1 microsatellites are depicted in Table 5. The -8G>A mutation age was estimated on two possible backgrounds, -5G and -5A. Results based on CD4 molecular marker show more ancient ages than those calculated from ATN1 locus.

As expected, the -5A polymorphism is older than -8A polymorphism, with an average estimate of approximately 35,000 years in both populations, assuming a mean value of 28 years per generation (Fenner, 2005). When the -8A variant is considered to be associated to the -5G background, it shows an average age of 10,976 (7,868–14,784) years in Angola population and 20,804 (16,828–26,880) years in Mozambique, whereas when associated to the -5A background, an age of 7,504 (2,716–14,784) years in Angola and only 1 generation (28 years) in Mozambique population was observed.

4. DISCUSSION

The clinical importance of the *TPI1* promoter polymorphisms regarding TPI1 deficiency remains unclear. Schneider et al. (1998) suggest that the association between the high prevalence of polymorphisms and moderate reduction of enzyme activity may result from a selective advantage for survival. On the other hand, Humphries et al. (1999), based on the

high allelic variation at CD4 locus among the *TPI1* promoter variants and the loss of LD between CD4 and *TPI1* promoter haplotypes, suggested that the haplotype diversification precedes the separation between African and non-African populations that occurred about 100,000 years ago (Flint et al. 1993).

Following the hypothesis of Schneider et al. (1998) we may propose that *TPI1* polymorphisms could have appeared as a selective advantage for survival against malaria, as reported to other erythrocyte enzymatic deficiencies.

The lack of a carboxylic acid cycle makes the malaria parasite exclusively dependent on glycolysis to gets ATP, mainly from its host cell glycolytic pathway. Subbayya et al. (1997) described that glucose consumption is 50 to 100 times higher in malaria-parasitized erythrocytes than in healthy erythrocytes. Consequently, with lower erythrocyte TPI1 enzyme activity the erythrocyte cycle of *Plasmodium* could be impaired by ATP deprivation. Ritter et al. (1993) observed the presence of autoantibodies against TPI1 (IgM) in *P. falciparum* infected individuals, meaning that TPI1 activity is self-limited in direct association with prolonged hemolytic anemia condition. These findings suggest that total or partial inhibition of TPI1 enzyme may affect the malaria parasite proliferation.

There are other evidences of a possible effect of TPI1 deficiency on the malaria parasite. TPI1 deficiency is associated with about 30-40 times higher cellular accumulation of DHAP (Orosz et al. 1996) that degrades spontaneously to methylglyoxal (MG), a precursor of potentially lethal advanced glycation endproducts (AGEs) that are detoxified by the glyoxalase system. The high glycolytic fluxes observed in parasitized erythrocytes are responsible for the inhibition of the detoxification system of MG, which would result in the accumulation of AGEs (Ahmed et al. 2003). The accumulation of MG is toxic to the parasite, as well as to humans in high concentrations, and the loss of enzymatic activity of TPI1, promoting the MG formation, may indicate a possible pathway of limiting parasite proliferation. In addition, Pavlovic-Djuranovic et al. (2006) tested a permeability of an aquaglyceroporin of *P. falciparum* (PfAQP), a bi-functional channel permeable for water and solutes, to MG and in fact, the incubation of MG in cultures of *P. falciparum* shows that the proliferation of malaria parasites was inhibited. Due to the major public health problem that malaria continues to be, to find potential new targets for new antimalarial therapeutics, such as the inhibition of metabolic pathways or enzymes necessary to parasite survival, is an urgent need.

Individuals from Angola and Mozambique present 47% and 53% of genotypes with *TPI1* promoter variants, respectively, which are consistent with those previously described by Humphries et al. (1999) and Manco et al. (2009) for the African population.

The -24 locus was monomorphic for the wild-type allele, which can be explained by the strong reduction of enzyme activity associated to this mutation and consequently to severe cases of TPI1 deficiency (Schneider et al. 1998). Functional studies of erythrocyte TPI1 enzyme activity showed that -5G>A substitution had no effect on TPI1 enzyme activity, in contrast to the -8A polymorphism, which is associated with a moderate reduction of enzyme activity (Schneider et al. 1998). Both polymorphisms, -8A and -24G are localized in conserved promoter regions CPE and TATA and most probably can impair the binding of transcription factors on these regions, affecting gene transcription (Humphries et al. 1999b).

The -5 locus showed higher genetic variation comparing with the -8 locus (He= 0.41 and 0.14, respectively), that confirm the results from Humphries et al. (1999) of an elevated prevalence for -5 *TPI1* polymorphism in African and Oriental populations. On the other hand, this author reported a low genetic variation for the -8 *TPI1* variant in individuals of African origin but we found this polymorphism in 45% of the individuals genotyped.

Regarding malaria outcome groups, -8A *TPI1* variant was more frequent in the UM group (19%) from Angola and AI group (20%) from Mozambique. Also GAG haplotype was more frequent in these groups although haplotype frequencies among -5, -8 and 2262 *TPI1* variants did not significantly differ. Schneider et al. (1998) observed that most of the Afro-American subjects with lower TPI1 enzyme activity values carried the -5G-8A haplotype and in this study, although not significantly, it was more frequent in non-severe malaria cases. These results could suggest some effect and association with lower TPI1 activity and protection against malaria severity.

As already mentioned before, *TPI1* promoter haplotype diversity in different geographic populations is a controversial issue in the scientific community, namely the discussion about genetic events that could have been the cause for this high diversity. Genetic recombination is unlikely to have contributed significantly to haplotype diversity due the short distance (20bp) on the *TPI1* promoter haplotype framework and Humphries et al. (1999) suggest that -8A polymorphism has arisen independently by two mutation events, which could explain its presence in the two haplotypes (-5G-8A and -5A-8A).

Regarding linkage between the studied loci, a significant LD between -5 and -8 *TPI1* variants was expected due their physical distance of only 3 nucleotides (Fig. 1). However, in the SM group from Angola and in the AI group from Mozambique these loci are in linkage

equilibrium. The lower LD involving the -8 locus in combination with -5 or 2262 loci, also reflected by the r^2 values of LD, could reflect the independent origins of -8G>A in both haplotypes -5G and -5A (see Table 3). LD was significant between *TPI1* loci -5, -8 and 2262 in Mozambique within all malaria groups and in the UM group from Angola population.

The significant LD observed among -5 and 2262 loci in both populations supports the reports of Humphries et al. (1999) on this disequilibrium, whom attributes it to a common ancestral origin, probable in Africa and preceding the modern human dispersion.

Viewing the high frequency of *TPI1* promoter variants in African population, we used LD to measure the association between a single allele at one locus and with multiple loci at various distances to assess the age of each allele by the decay of its association. The positive selection causes an unusually rapid rise in allele frequency, occurring over a short enough time and recombination does not substantially break down the haplotype on which the selected mutation occurs. A signature of positive natural selection is thus an allele having an unusually LD given its population frequency (Sabeti et al. 2002).

Based on these principles, we intended to find any evidence concerning the arising and dispersion of -5G>A and -8G>A polymorphisms during human genetic evolution.

The age estimation for these mutations suggested a much lower age than the dispersion period of modern human populations originating from Africa, rejecting the hypothesis of Humphries et al. (1999). The estimate of about 35,000 years for the -5G>A mutation in both studied African populations would explain the high frequency of this variant in the African subjects. These data are supported by the elevated intra-allelic variation of CD4 and ATN1 loci. Despite this, among non-African populations, the allelic distribution of -5A is very similar to the African population (Humphries et al. 1999; Manco et al. 2009).

The -8A mutation seems to be a much more recent event in African population and considering that the -8A allele had arisen by successive mutations in two different backgrounds (-5G and -5A), we calculated the age of this mutation, on two mutational events. The first mutational event (on the -5G haplotype) may have occurred in Mozambique (East Africa) around 16,828–26,880 years ago. Angola isolates (West Africa) revealed a more recent age estimate for this event, which might have occurred about 7,868–14,784 years ago. Probably due to the phenomenon of migration, it had emerged differently in both populations. The second mutational event (on the -5A haplotype) showed a more recent age, but diverged greatly in the two populations. This mutation would have occurred for the first time in Angola region with an age estimate about 2,716–14,784 years ago; in Mozambique, this mutation appears to be much more recent, associated to only 1 generation (28 years).

Despite the similar allele frequencies of -8A allele in Angola and Mozambique (7.6 and 7.3%, respectively), in Mozambican population the haplotype -5A-8A is linked with only one allele of both microsatellites. This result suggests that not enough time has elapsed yet to accumulate variation or recombination on microsatellites, another sign of a recent age for this second mutational event.

Historical and archaeological evidence suggest an age of 10,000 years as the evolutionary period of the malarial selective impact on human populations. This period is consistent with origin and expansion of agriculture in the Middle East and Africa, leading to the conditions for one of the principal malaria vectors, *Anopheles gambiae*, to spread. Further, the agriculture development allowed an increase of human population density, which also enabled the spread of infectious diseases (Armelagos et al. 1996; Uneke et al. 2009; Tishkoff et al. 2001). Joy et al. 2003 suggest that the parasite migration outside of Africa happened before the human migration out of Africa about 40,000 to 130,000 years ago, during the Pleistocene expansion.

The age calculated for -8A allele on the -5G background (16,828–26,880 years) exceeded the period of 10,000 years, but when associated with the -5A haplotype is much more recent (2,716–14,784 years) and the estimate are within the selective pressure period from malaria. This could be a possible reason for the spread of the -8A polymorphism in malaria endemic countries.

The study of the *TPI1* promoter variants allele and haplotype frequencies in different malaria outcome groups was not conclusive but we consider that altogether the following gathered results support the hypothesis that the arising and spread of *TPI1* polymorphic variants could have been due to a selective advantage against malaria: (1) the high frequency of *TPI1* promoter polymorphisms observed in the African populations studied, previously associated to a reduced TPI1 activity, (2) higher frequency of the allele -8A and the haplotype GAG (-5/-8/2262) in non-severe malaria groups, (3) the significant LD between *TPI1* promoter region and (4) the age estimate for *TPI1* -8 variant within the period of origin and spread of agriculture and the malaria mosquito vector in Africa. To confirm these results, it would be important now to further investigate not only the *TPI1* promoter haplotypes but also measure the TPI1 activity in the malaria groups to ascertain the correspondence between the genotype and phenotype since the presence of a singular -5 polymorphism (associated to a normal enzyme activity) or an heterozygous -5/-8 (associated to a reduced enzyme activity) may vary between individuals.

ACKNOWLEDGMENTS

We thank all individuals and parents/tutors of children who participated in this study and to all health technicians working at Emergency Services of the Pediatric Hospital David Bernardino (Luanda, Angola), Pediatrics Department of Central Hospital of Maputo, Health Centers of Bagamoyo and Boane (Maputo, Mozambique) and from Manhiça Health Research Center (CISM) for all technical support.

This study was supported by PEst-OE/SAU/LA0018/2011 - Proj. Estratégico LA0018 2011/2012 (http://cmdt.ihmt.unl.pt/index.php/pt/) and PTDC/SAU-MET/110323/2009, ‘‘Fundação para a Ciência e Tecnologia/Ministério da Educação e Ciência’’, FCT/MEC (http://alfa.fct.mctes.pt/index.phtml.pt), Portugal. PM held a FCT grant (SFRH/BD/28236/2006). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions: Conceived and designed the experiments: APA, LM. Performed the experiments: MG, PM. Analyzed the data: MG, LM, APA. Contributed reagents/materials/analysis tools: APA, LM. Wrote the paper: MG, APA. Did the field work: at Mozambique PM, NF, at Angola LM.

Conflict of interest: The authors have declared that no conflict of interest exists.

REFERENCES

Ahmed, N., Battah, S., Karachalias, N., Babaei-Jadidi, R., Horányi, M., Baróti, K., Hollan, S., Thornalley, P.J., 2003. Increased formation of methylglyoxal and protein glycation, oxidation and nitrosation in triosephosphate isomerase deficiency. Biochim. Biophys. Acta. 1639, 121-132.

Armelagos, G.J., Barnes, K.C., Lin, J., 1996. Disease in human evolution: the re-emergence of infectious disease in the third epidemiological transition. AnthroNotes. 18, 1-7.

Arya, R., Lalloz, M.R., Bellingham, A.J., Layton, D.M., 1997. Evidence for founder effect of the Glu104Asp substitution and identification of new mutations in triosephosphate isomerase deficiency. Hum. Mutat. 10, 290-294.

Excoffier, L., Laval, G., Schneider, S., 2007. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol. Bioinform. Online. 1, 47-50.

Fenner, J.N., 2005. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. Am. J. Phys. Anthropol. 128, 415-423.

Fleming, A.F., Storey, J., Molineaux, L., Iroko, E.A., Attai, E,D., 1979. Abnormal haemoglobins in the Sudan savanna of Nigeria. I. Prevalence of haemoglobins and relationships between sickle cell trait, malaria and survival. Ann. Trop. Med. Parasitol. 73, 161-172.

Flint, J., Harding, R.M., Clegg, J.B., Boyce, A.J., 1993. Why are some genetic diseases common? Distinguishing selection from other processes by molecular analysis of globin gene variants. Hum. Genet. 91, 91-117.

Humphries, A., Ationu, A., Lalloz, M.R., Layton DM. 1999a. Ancestral origin of variation in the triosephosphate isomerase gene promoter. Hum. Genet. 104, 486-491.

Humphries, A., Ationu, A., Wild, B., Layton, D.M., 1999b. The consequence of nucleotide substitutions in the triosephosphate isomerase (TPI) gene promoter. Blood. Cells Mol. Dis. 25, 210-217.

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.

Luzzatto, L., 2006. Glucose 6-phosphate dehydrogenase deficiency : from genotype to phenotype. Haematologica. 91, 1303-1306.

Machado, P., Manco, L., Gomes, C., Mendes, C., Fernandes, N., Salomé, G., Sitoe, L., Chibute, S., Langa, J., Ribeiro, L., Miranda, J., Cano, J., Pinto, J., Amorim, A., do Rosário, V. E., Arez, A. P., 2012. Pyruvate kinase deficiency in sub-Saharan Africa: identification of a highly frequent missense mutation (G829A;Glu277Lys) and association with malaria. PloS One. 7, e47071.

Machado, P., Pereira, R., Rocha, A.M., Manco, L., Fernandes, N., Miranda, J., Ribeiro, L., do Rosário, V.E., Amorim, A., Gusmão, L., Arez, A.P., 2010. Malaria: looking for selection signatures in the human PKLR gene region. Br. J. Haematol. 149, 775-784.

Manco, L., Machado, P., Lopes, D., Nogueira, F., Do Rosário, V.E., Alonso, P.L., Varandas, L., Trovoada, Mde. J., Amorim, A., Arez, A.P., 2009. Analysis of TPI gene promoter variation in three sub-Saharan Africa population samples. Am. J. Hum. Biol. 21, 118-120.

Maquat, L.E., Chilcote, R., Ryan, P.M., 1985. Human triosephosphate isomerase cDNA and protein structure. Studies of triosephosphate isomerase deficiency in man. J. Biol. Chem. 260, 3748-3753.

Marques, P.X., Saúte, F., Pinto, V.V., Cardoso, S., Pinto, J., Alonso, P.L., do Rosário, V.E., Arez, A.P., 2005. *Plasmodium* species mixed infections in two areas of Manhiça district, Mozambique. Int. J. Biol. Sci. 1, 96-102.

Mohrenweiser, H.W., Fielek, S., 1982. Elevated frequency of carriers for triosephosphate isomerase deficiency in newborn infants. Pediatr. Res. 16, 960-963.

Orosz, F., Oláh, J., Ovádi, J., 2009. Triosephosphate isomerase deficiency: new insights into an enigmatic disease. Biochim. Biophys. Acta. 1792, 1168-1174.

Orosz, F., Vértessy, B.G., Hollán, S., Horányi, M., Ovádi, J., 1996. Triosephosphate isomerase deficiency: predictions and facts. J. Theor. Biol. 182, 437-447.

Ott J., 1999. Analysis of human genetic linkage, third ed. Johns Hopkins University Press, Baltimore.

Pavlovic-Djuranovic, S., Kun, J.F., Schultz, J.E., Beitz, E., 2006. Dihydroxyacetone and methylglyoxal as permeants of the *Plasmodium* aquaglyceroporin inhibit parasite proliferation. Biochim. Biophys. Acta. 1758, 1012-1017.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559-575.

Ritter, K., Kuhlencord, A., Tomssen, R., Bommer, W., 1993. Prolonged haemolytic anemia in malaria and autoantibodies against triosephosphate isomerase. Lancet, 342, 1333-1334.

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B, Platko, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., Lander, E.S., 2002. Detecting recent positive selection in the human genome from haplotype structure, Nature. 419, 832-837.

Schneider, A., Forman, L., Westwood, B., Yim, C., Lin, J., Singh, S., Beutler, E., 1998. The relationship of the -5, -8, and -24 variant alleles in African Americans to triosephosphate isomerase (TPI) enzyme activity and to TPI deficiency. Blood. 92, 2959-2962.

Schneider, A.S., Valentine, W.N., Hattori, M., Heins, H.L.Jr., 1965. Hereditary hemolytic anemia with triosephosphate isomerase deficiency. N. Engl. J. Med. 272, 229-235.

Seixas, S., Garcia, O., Trovoada, M.J., Santos, M.T., Amorim, A., Rocha, J., 2001. Patterns of haplotype diversity within the serpin gene cluster at 14q32.1: insights into the natural history of the α1-antitrypsin polymorphism. Hum. Genet. 108, 20-30.

Subbayya, I.N., Ray, S.S., Balaram, P., Balaram, H., 1997. Metabolic enzymes as potential drug targets in *Plasmodium falciparum*. Indian J. Med. Res. 106, 79-94.

Sudbery, P., 2002. Human Molecular Genetics, second ed. Pearson Education Limited (Prentice-Hall), London

Tishkoff, S.A., Varkonyi, R., Cahinhinan, N., Abbes, S., Argyropoulos, G., Destro-Bisol, G., Drousiotou, A., Dangerfield, B., Lefranc, G., Loiselet, J., Piro, A., Stoneking, M., Tagarelli, A., Tagarelli, G., Touma, E.H., Williams, S.M., Clark, A.G., 2001. Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. Science. 293, 455-462.

Uitenbroek, D.G., 1997. SISA Binomial. Available online: http://www.quantitativeskills.com/sisa/distributions/binomial.htm, accessed on [January 2014].

Ulgen, A., Li, W., 2005. Comparing single-nucleotide polymorphism marker-based and microsatellite marker-based linkage analyses. BMC Genetics. 6(Suppl 1):S13.

Uneke, C., 2009. Deforestation and Malaria in sub-Saharan Africa: an overview. Int. J. Trop. Med. 6.

UCSC Genome Browser: Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res. 2002 Jun;12(6):996-1006

Watanabe, M., Zingg, B.C., Mohrenweiser, H.W., 1996. Molecular analysis of a series of alleles in humans with reduced activity at the triosephosphate isomerase locus. Am. J. Hum. Genet. 58, 308-316.