

An Examination of the *HBB* Gene in Various African Populations

**Charles Washington III
University of Pennsylvania
Department of Anthropology**

May 29, 2013

Abstract

Malarial resistance serves as a prime example of how infectious disease and parasite loads have affected human evolutionary biology. Claiming over 1 million deaths worldwide every year, malaria's wide and severe effects on the human condition are undeniably strong. However, studies on populations existing in malaria endemic areas have revealed numerous biological adaptations that confer resistance against the disease. Major forms of genetic resistance to malaria are seen through variants involved in sickle cell disease, Duffy blood antigens, glyophorins, blood type groupings, and glucose-6-dehydrogenase deficiency.

Of these the best studied are the genetic variants in the *HBB* gene, which is responsible for the construction of hemoglobin in erythrocytes. Variants of the *HBB* gene include HbS, HbC, HbE, and variants involved in beta-thalassemias, which are present at varying frequencies in different areas of the world in response to the pressure of malarial disease. African populations in particular are of great interest due to their greater genetic diversity and the high rates of malarial diseases reported in African countries. As such, sequencing the genotypes of multiple African populations would give a greater understanding to the frequencies of the two major *HBB* variants present in Africa, HbS and HbC.

Furthermore, given the strong selective pressure an infectious disease such as malaria would place on populations in malaria-endemic regions, it is possible that multiple variants may have evolved in these populations in attempt to combat the disease. By sequencing the *HBB* gene and its corresponding upstream region in 260 individuals from 15 different African populations, the frequencies of the HbS and HbC SNPs have been determined.

Background Information

The World Malaria Report has shown that there were about 219 million episodes of malaria and an estimated 660,000 deaths in the year 2011 among African populations, which bore over 90% of all reported malarial deaths (WHO, 2012). Malaria caused by *P. falciparum* results in the highest morbidity and mortality accounting for almost over 1 million deaths every year, largely affecting children under the age of five. It has also been observed that many red cell disorders are more prominent in the tropical regions where malaria is endemic (Haldane et al., 1949). The observation that erythrocytes with sickle cell trait hemoglobin are higher in regions with greater prevalence of malaria has been repeatedly confirmed. It is also undisputable that the variant that causes the disease offers partial protection against severe malaria caused by *P. falciparum* (Aidoo et al., 2002). Further research conducted by Allison (1954) confirmed that due to natural selection acting on the *HBB* gene, which codes for the production of beta hemoglobin in erythrocytes, the frequency of the sickle cell variant is significantly correlated with the geographic distribution of cases of malaria.

The *HBB* (hemoglobin, beta) gene located on chromosome 11 contains 3 exons that provide the instructions for making a 146 amino acid long protein called beta globin which, along with alpha globin, makes up the hemoglobin molecule. The hemoglobin protein (molecular weight of about 64,500 g/mol) is an iron-containing molecule found in the red blood cells of all vertebrates (Perutz, 1978). Hemoglobin carries oxygen from the lungs and delivers it to the cells throughout the body. This molecule consists of two alpha (α) and two beta subunits (β), which are frequently referred to as alpha and beta hemoglobin, respectively. Each of these individual subunits carry an iron-containing molecule called a heme group which is responsible for the oxygen binding and delivery.

More than 250 mutations in the *HBB* gene have been found to cause beta thalassemia while sickle cell anemia, a common form of sickle cell disease, is caused from the presence of a mutated form of hemoglobin. The most common form of this disease, HbS, is an autosomal recessive disorder that causes significant morbidity and mortality particularly in those of African and Mediterranean ancestry. Carriers of the sickle cell trait (who are heterozygous with one normal allele and one HbS allele) have some resistance to the fatal malaria caused by *P. falciparum*. The sickle cell trait is caused by a single amino acid mutation (valine substituted in place of glutamic acid at the 6th position) in the beta chain of the *HBB* gene. While the inheritance of the sickle cell trait from both parents leads to the disease, individuals carrying only one copy of the allele are said to have some form of protection against malaria.

Beyond the traditional sickle cell mutation (HbS) other mutations have been documented in the *HBB* gene such as *HbC* and *HbE*, which arose in Africa and Southeast Asia respectively (Clark and Higgins, 2006). Malaria's severe physiological effect on the survivability of human populations has caused strong evolutionary forces to act on affecting genes such as *HBB*, making them great targets for analysis of selective forces. Proximity to malaria-endemic locales plays an important role in the geographic distribution of several erythrocyte disorders like thalassemia, G6PD deficiency, sickle cell, and mutations in the *Duffy* gene. Malaria is significantly prevalent in many parts of the world in which mutations causing such disorders are still favored despite their physiological consequences, thus serving as a testament to the value of malarial resistance. For geographical regions with a low presence of malaria the frequency of these mutation alleles are low compared to malaria-endemic regions which vary from region to region but can potentially have carrier frequencies as high as 60% for HbS and 50% for HbC (Weatherall and Clegg, 2001), (WHO, 2006). Due to the strong selective force of malaria many typically

deleterious mutations survive in the affected populations, which may result phenotypically as sickle cell and thalassemia depending on the mutation's location (gene proper and gene promoter, respectively); thus, the strong selective pressure of malaria would make study of an associated gene such as *HBB* particularly informative (Ko et al., 2012).

Materials and Methods

Populations and DNA Samples

DNA samples were obtained from 260 unrelated individuals arising from 15 different African populations. These populations include Fulani, Lemande, Mada, Bakola pygmies from Cameroon; Yoruba from Nigeria; Iraqw, Sandawe, Datog and Hadza from Tanzania; Borana, Boni and Sengwer from Kenya, Chad and Sudan (Table 1). Samples were placed within the following categories denoted by their country and population of origins as well as a specific sample number: CAFU, CALM, CAMD, CAPL, CHBU, KEBN, KEBR, KELO, KESN, NGYR, SDDBA, SDHD, TZDT, TZHZ, TZIQ, and TZSW.

Ethics Statement

The purpose of the study was well explained to all the participants before any consent forms were obtained. Written informed consent forms were obtained from all individuals for blood collection and all subjects were kept anonymous. Research and ethics consent were obtained from Commission for science and technology and the National Institute for Medical Research in Dares Salaam in Tanzania; Kenya Medical Research Institute in Nairobi, Kenya; the Ministry of Health and National Committee of Ethics, Cameroon. Samples were collected and analyzed after obtaining approval from the Institutional Review Board of University of Pennsylvania. White blood cells were isolated from whole blood collected from each individual using a

modified salting-out procedure. DNA was then extracted from these samples in the laboratory using a Gentra Purgen DNA extraction kit.

Pre and Post Sequencing Technique

The *HBB* gene (~ 1.6 kbp) encompassing exons 1, 2 and 3 was amplified using PCR. An additional upstream region (~ 1kbp) from exon 1 was also amplified using specific primers (primer list in appendix). In the PCR mixture each reaction contains 17.8uL of ddH₂O, 2.5uL 1x PCR buffer, 1 uL 2.0mM of Mgso₄, 0.2uL of Taq polymerase High Fidelity (Invitrogen), 0.5uL 0.2 mM of dNTP (Promega), 0.5ul of 0.3uM of each primer and 2ul (50 ng) of genomic DNA (final volume 25 µl). PCR cycles consisted of 1 cycle of pre-incubation (94°C for 1 minute), 35 cycles of amplification (95 °C for 30 seconds, 58 °C for 30 seconds, and 68 °C for 2 minutes and 30 seconds), and 1 cycle of extension (72 °C for 5 minutes). PCR products were examined by electrophoresis on 1% agarose gels in 1X TBE buffer pH 8.3. PCR products were then cleaned using Shrimp Alkaline Phosphatase and Exonuclease 1 procedure (United States Biochemicals). Samples were then cleaned by an isopropanol procedure using 40uL 75% isopropanol and then 75uL 70% isopropanol. Samples were re-suspended with 10uL of ABI HiDi Formamide and then sequenced by ABI 96 capillary 3730 XL sequencer (Applied Biosystems). The contigs were assembled and the SNPs were identified using the Sequencer 4.10 program (Gene Codes Corporation). Coverage for the sequences was low; 2.5X average coverage for the gene exon and introns and 1.5X average coverage for the upstream region. The Reference gene was obtained from Genebank.

Results

The results are divided into the 16 distinct populations given earlier. The frequency of the SNPs strongly associated with malarial resistance in Africa, HbS and HbC, and the number of individuals for each population are presented in Table 1. The same data is shown in graphical form in Figure 1.

Table 1 Frequencies of Malaria Associated SNPs in human African populations			
Populations	N	HbS frequency (%)	HbC frequency (%)
CAFU	18	5.56	0
CALM	18	11.11	0
CAMD	18	5.56	0
CAPL	19	0	0
CHBU	11	9.09	0
KEBN	16	1.25	6.25
KEBR	17	29.41	17.65
KELO	17	29.41	0
KESN	19	0	0
NGYR	16	6.25	0
SDBA	8	25	1.25
SDHD	11	27.27	18.18
TZDT	17	23.53	11.76
TZHZ	18	5.56	0
TZIQ	18	5.56	0
TZSW	19	15.79	5.26
Total	260		

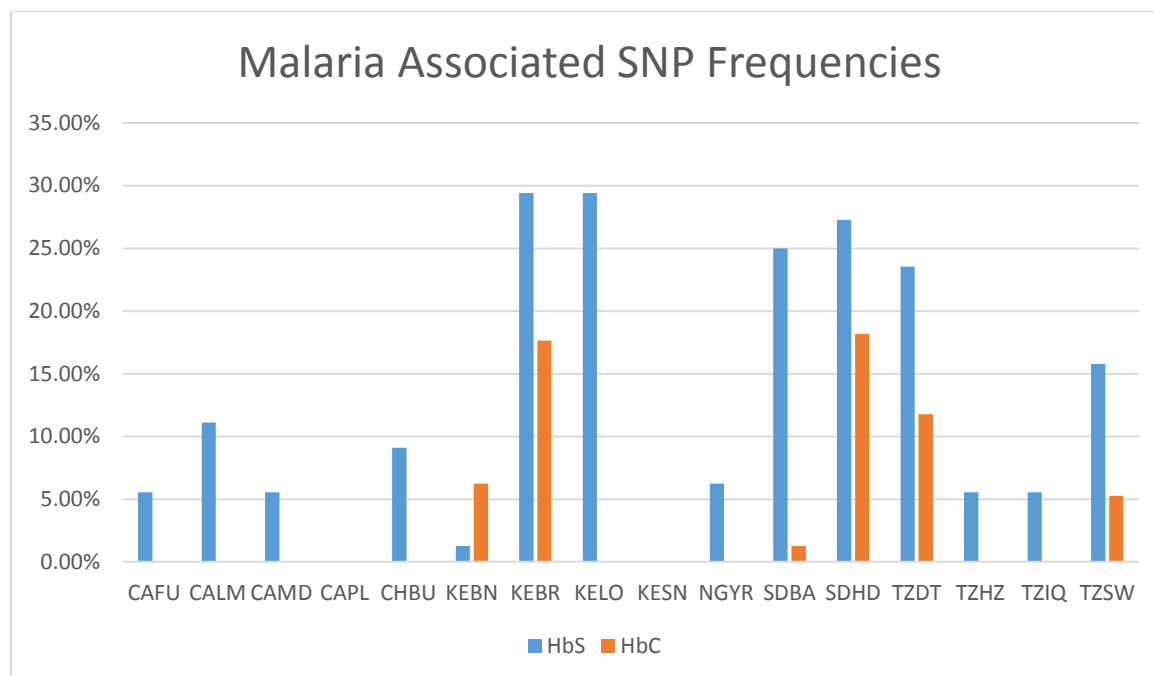


Figure 1. Histogram of HbS and HbC SNP frequencies in human African populations. N = 260.

Discussion

It is made evident from Table 1 that in the African populations examined HbS is the more frequent malaria associated SNP. All populations except CAPL and KESN reported some frequency for the HbS. Additionally, all populations reported a higher frequency of HbS than HbC except for KEBN. It should be noted that the presence of HbS and HbC is not mutually exclusive. In fact, many individuals are heterozygotes of HbSC, in which they have one allele of both variants, and suffer less severe phenotypic effects from this combination in relation to a homozygous HbS individual (Nagel et al., 2003). It is also particularly interesting that the populations of West Africa show very low frequencies of HbC, which is noted to be most significant in those of Yoruba and West African ancestry (Fairhurst et al., 2003), (Modiano et al., 2003), (May et al., 2007). In contrast, it is actually the other regions of Africa such as Sudan, Kenya, and Tanzania that display the greatest frequencies of the HbC variant.

Future Directions

The first step forward with this project is to increase the coverage at which the samples were sequenced. Increasing the confidence of some of the data, especially in the upstream region, would allow for a more accurate portrayal of potential novel variants in the gene. Another endeavor would be to identify more functional variants on the *HBB* gene and on other erythrocyte-relevant genes among diverse ethnic groups in malaria-endemic regions. Candidate variants can be identified either through population genetics-based analysis or through experimental/clinic measurements of their functional consequences (e.g., resistance to malaria susceptibility or deleterious effects causing various symptoms of hemoglobinopathy) (Rees et al., 2010). The first approach would have some technical issues due to the segregation of these variants being at low frequencies in the sampled populations and thereby garnering low statistical power for detecting their selection signatures. In addition, the evolutionary origin of HbS and other variant alleles has been controversial because it is wide spread and many haplotype structures have been observed across different geographic ranges; as such, both single-origin and multiple-origin hypotheses have gained some supports theoretically or empirically (Flint et al., 1993), (Flint et al., 1998). In general, future studies on the origin of *HBB* gene variants will help to understand how the human genome evolved to fight against infectious diseases (Barreiro and Quintana-Murci, 2009).

Acknowledgements

I would like to thank the University of Pennsylvania Anthropology Department and the Anthropology Undergraduate Research Fellowship for presenting me with the opportunity to conduct such research. I also thank Dr. Sarah Tishkoff and Dr. Theodore Schurr for all their

guidance and assistance in the undertaking of the thesis work. Additionally, for their patience and assistance in my laboratory experiences I want to thank all the members of the Tishkoff lab, particularly Prianka Rajan, MS, Wen-Ya Ko, PhD, Laura Scheinfeldt, PhD, and Bill Beggs, MS.

BIBLIOGRAPHY

- Aidoo, M., D. Terlouw, M. Kolczak, P. Mcelroy, F. Terkuile, S. Kariuki, B. Nahlen, A. Lal, and V. Udhayakumar. "Protective Effects of the Sickle Cell Gene against Malaria Morbidity and Mortality." *The Lancet* 359.9314 (2002): 1311-312.
- Allison, A. C. "Protection Afforded by Sickle-cell Trait Against Subtertian Malarial Infection." *Bmj* 1.4857 (1954): 290-94.
- Barreiro, Luis B., and Lluís Quintana-Murci. "From Evolutionary Genetics to Human Immunology: How Selection Shapes Host Defence Genes." *Nature Reviews Genetics* 11.1 (2009): 17-30.
- Clarke, Gwendolyn M., and Trefor N. Higgins. "Laboratory Investigation of Hemoglobinopathies and Thalassemias: Review and Update." *Clinical Chemistry B* 46.8 (2006): 1284-290.
- Fairhurst, R. M., H. Fujioka, K. Hayton, K. F. Collins, and T. E. Wellems. "Aberrant Development of Plasmodium Falciparum in Hemoglobin CC Red Cells: Implications for the Malaria Protective Effect of the Homozygous State." *Blood* 101.8 (2003): 3309-315.
- Flint, Jonathan, Rosalind M. Harding, Anthony J. Boyce, and John B. Clegg. "1 The Population Genetics of the Haemoglobinopathies." *Baillière's Clinical Haematology* 11.1 (1998): 1-51.
- Flint, Jonathan, Rosalind M. Harding, Anthony J. Boyce, and John B. Clegg. "8 The Population Genetics of the Haemoglobinopathies." *Baillière's Clinical Haematology* 6.1 (1993): 215-62.
- Haldane, J.B.S. "Disease and Evolution." *Ricerca Science Supplement* 19 (1949): 3-10. Print.
- Ko, Wen-Ya, Felicia Gomez, and Sarah A. Tishkoff. "Evolution of Human Erythrocyte-specific Genes Involved in Malaria Susceptibility." *Rapidly Evolving Genes and Genetic Systems*. Oxford: Oxford UP, 2012.
- May, J., J. A. Evans, C. Timmann, C. Ehmen, W. Busch, T. Thye, T. Agbenyega, and R. D. Horstmann. "Hemoglobin Variants and Disease Manifestations in Severe Falciparum Malaria." *JAMA: The Journal of the American Medical Association* 297.20 (2007): 2220-226.
- Modiano, David, Gaia Luoni, Bienvenu S. Sirima, Jacques Simpore, Federica Verra, Amadou Konate, Elena Rastrelli, Anna Olivieri, Carlo Calissano, Giacomo M. Paganotti, Leila D'Urbano, Issa Sanou, Alphonse Sawadogo, Guido Modiano, and Mario Coluzzi. "Haemoglobin C Protects Against Clinical Plasmodium Falciparum Malaria." *Nature* 414 (2001): 305-18.

Nagel, R., M. E. Fabry, and M. H. Steinberg. "The Paradox of Hemoglobin SC Disease." *Blood* 17.3 (2003): 167-78.

Perutz, M. F. "Hemoglobin Structure and Respiratory Transport." *Scientific American* 239.6 (1978): 92-125.

Rees, David C., Thomas N. Williams, and Mark T. Gladwin. "Sickle-cell Disease." *The Lancet* 376 (2010): 2018-031.

Weatherall, D. J., and J. B. Clegg. "Inherited Haemoglobin Disorders: An Increasing Global Health Problem." *Bulletin of the World Health Organization* 79.8 (2001): 704-12.

World Health Organization. "Sickle-cell anemia. Report by the Secretariat". 2006
< https://apps.who.int/gb/ebwha/pdf_files/WHA59/A59_9-en.pdf >

World Health Organization. "World Malaria Report 2012 Fact Sheet". 2012.
< http://www.who.int/malaria/publications/world_malaria_report_2012/wmr2012_factsheet.pdf >

APPENDIX

Primers were ordered from Integrated Data Technologies.

Name	Primer
HBB.F.1	5' – GACAGGTACGGCTGTCATCA – 3'
HBB.R.1	5' – GGTTTGAAGCTAGCTCTTCAT – 3'
HBB.F.3	5' – ACATTTGCTTCTGACACAAC – 3'
HBB.R.3	5' – AATCCAGCCTTATCCCAACC – 3'
HBB.NF	5' – ATTCTGGAGACGCAGGAAGA – 3'
HBB.NR1	5' – TATCTCTTGGCCCCATACCA – 3'
HBB.NR2	5' – TAGATGGCTCTGCCCTGACT – 3'
HBB.S	5' – CTGGATTATTCTGAGTCCAAGC – 3'
HBB.int_F	5' – GATCCATCTACATATCCCAA – 3'
HBB.int_R	5' GGCAGAGAGAGTCAGTGCCTA – 3'
HBB.int2_F	5' – GTGTACACATATTGACCAAATCAGGGTA – 3'
HBB.UP.F1	5' – GGAACACTGAGACCCTACGC – 3'
HBB.UP.SEQ.R1	5' – CGATCACGTTGGGAAGCTAT – 3'
HBB.UP.SEQ.R2	5' – CTCTTCCTGCGTCTCCAGAA – 3'