A peer-reviewed version of this preprint was published in PeerJ on 24 February 2015.

Noble JA, Duru KC, Guindo A, Yi L, Imumorin IG, Diallo DA, Thomas BN. 2015. Interethnic diversity of the CD209 (rs4804803) gene promoter polymorphism in African but not American sickle cell disease. PeerJ 3:e799 https://doi.org/10.7717/peerj.799

Interethnic diversity of the CD209 (rs4804803) gene promoter polymorphisms in African but not American sickle cell disease

Elucidating the genomic diversity of CD209 gene promoter polymorphisms could assist in clarifying disease pathophysiology as well as contribution to co-morbidities. CD209 gene promoter polymorphisms have been shown to be associated with susceptibility to infection. We hypothesize that CD209 mutant variants occur at a higher frequency among Africans and in sickle cell disease. We analyzed the frequency of the CD209 gene (rs4804803) in healthy control and sickle cell disease (SCD) populations and determined association with disease. We obtained genomic DNA from 145 SCD and 244 control Africans (from Mali), 331 SCD and 379 control African Americans and 159 Caucasians. Comparative analysis among and between groups was carried out by polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP). Per ethnic diversification, we found significant disparity in genotypic (23.4% versus 16.9% versus 3.2%) and allelic frequencies (36.1% versus 25.1% versus 11.6%) of the mutant variant of the CD209 (snp 309A/G) gene promoter between Africans, African Americans and Caucasians respectively. Surprisingly, there was a wide disparity in the genotypic and allelic frequencies among African SCD versus healthy controls (10.4% versus 23.4% (genotypes) and 25.2% versus 36.1% (alleles), which is completely absent among African Americans. Comparing SCD groups, there was no difference between Africans and Americans, implying a lack of association between CD209 polymorphisms and sickle cell disease in either population. The higher frequency of CD209 mutant variants in the non-SCD group reveals an impaired capacity to mount an immune response to infectious diseases. We conclude that CD209 polymorphism play a major role in susceptibility to infectious pathogens and could potentially delineate susceptibility to and severity of co-morbidities.

Interethnic diversity of the CD209 (rs4804803) gene promoter polymorphisms in African but not American sickle cell disease

Jenelle A. Noble[1], Kimberley C. Duru[1], Aldiouma Guindo[2], Li Yi[3], Dapa A. Diallo[2], Bolaji N. Thomas[1]

[1]Department of Biomedical Sciences, College of Health Sciences and Technology, Rochester Institute of Technology, Rochester NY; [2]Centre de Recherche et de Lutte contre la Drepanocytose, Bamako, Mali; [3]School of Statistics, Shanxi University of Finance and Economics, Shanxi, China

Jenelle Noble: jxn6742@rit.edu

Kimberley Duru: kcd5949@rit.edu

Aldiouma Guindo: guindo100@yahoo.fr

Yi Li: liyi3344520@hotmail.com

Dapa Diallo: dadiallo@icermali.org

Bolaji Thomas: bolaji.thomas@rit.edu

16 *Corresponding author:
17 Dr Bolaji N. Thomas
18 Department of Biomedical Sciences
19 College of Health Sciences and Technology
20 Rochester Institute of Technology
21 153 Lomb Memorial Drive
22 Rochester NY 14623
23 Office: (585) 475-6382
24 Fax: (585) 475-5809
25 Email: bolaji.thomas@rit.edu

Abstract

Elucidating the genomic diversity of CD209 gene promoter polymorphisms could assist in clarifying disease pathophysiology as well as contribution to co-morbidities. CD209 gene promoter polymorphisms have been shown to be associated with susceptibility to infection. We hypothesize that CD209 mutant variants occur at a higher frequency among Africans and in sickle cell disease. We analyzed the frequency of the CD209 gene (rs4804803) in healthy control and sickle cell disease (SCD) populations and determined association with disease. We obtained genomic DNA from 145 SCD and 244 control Africans (from Mali), 331 SCD and 379 control African Americans and 159 Caucasians. Comparative analysis among and between groups was carried out by polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP). Per ethnic diversification, we found significant disparity in genotypic (23.4% versus 16.9% versus 3.2%) and allelic frequencies (36.1% versus 25.1% versus 11.6%) of the mutant variant of the CD209 (*snp 309A/G*) gene promoter between Africans, African Americans and Caucasians respectively. Surprisingly, there was a wide disparity in the genotypic and allelic frequencies ***among*** African SCD versus healthy controls (10.4% versus 23.4% (genotypes) and 25.2% versus 36.1% (alleles), which is ***completely absent*** among African Americans. Comparing SCD groups, there was no difference between Africans and Americans, implying a lack of association ***between*** CD209 polymorphisms and sickle cell disease in either population. The higher frequency of CD209 mutant variants in the non-SCD group reveals an

impaired capacity to mount an immune response to infectious diseases. We conclude that CD209 polymorphism play a major role in susceptibility to infectious pathogens and could potentially delineate susceptibility to and severity of co-morbidities.

Introduction

Sickle cell disease (SCD) is an inherited multisystem disorder, characterized by chronic hemolytic anemia, vaso-occlusive crises and several other disease outcomes such as acute chest syndrome, bacteremia, leg ulcers and priapism (Bunn 1997; Benkerrou *et al.,* 2002). SCD has shown marked variability in severity between individuals, with evidence of extensive differences in both clinical and genotypic presentations, with a global distribution, especially in sub-Saharan Africa, Middle East, parts of the Indian subcontinent, and Americans with an African or Caribbean descent (Hassell, 2010; Piel *et al.,* 2013; Bandeira *et al.,* 2014; Saraf *et al.,* 2014; Thakur *et al.,* 2014). SCD occurs in patients that are homozygous for the hemoglobin S gene, produced by a defective β-globin gene on chromosome 11 and has also been defined as resulting from compound heterozygosity for hemoglobin S and another β-globin chain abnormality (typically hemoglobin C or β-thalassemia), with α-thalassemia serving as a modifier of the clinical manifestations (Weatherall, 2010; Saraf *et al.,* 2014). Patients commonly require red cell transfusions to manage complications, with alloimmunization a common occurrence (Charache *et al.,* 1983; Rosse *et al.,* 1990; Tatari-Calderone *et al.,* 2013) leaving such multiply transfused patients at risk for delayed hemolytic transfusion reactions (Piomelli *et al.,* 1985; Petz *et al.,* 1997; Taylor *et al.,* 2008; Yazdanbaksh *et al.,* 2012), development of autoimmune hemolytic anemia.

Infectious pathogens are a threat to those individuals with SCD, particularly children, that are prone to frequent and severe attacks (Overturf, 1999; Halasa *et al.,* 2007; Szczepanek *et al.,* 2013). For children in endemic countries, with very high circulating immune complexes due to constantt exposure to multiple pathogenic stimuli, the added burden of these co-morbidities can severely impact immune response and survival (Thomas *et al.,* 2012). Recent reports showing high mortality rates post-vaccination in transgenic animals demonstrates that a dysregulated immune response might be responsible for such mortality and could be a major drawback to the current push to vaccinate (McCavit *et al.,* 2011; Szczepanek *et al.,* 2013). Infact, other reports have shown that there is an over-stimulation of pro-inflammatory cytokines in sickle cell disease patients, which might be be related to vaso-occlusion (Makis *et al.,* 2000; Pathare *et al.,* 2004; Steinberg 2006; Conran *et al.,* 2009; Qari *et al.,* 2012; Bandeira *et al.,* 2014). In fact, this hyperstimulation has been associated with sickle cell haplotype in Brazil, and as such, the obvious consequence of worsening immune response to secondary infectious pathogens or co-morbidities of infection.

Recently published data have shown that there are wide differences in infection rates and multiplicity of infection between children who are carriers of the sickle cell trait (hemoglobin AS) and those patients that possess the normal hemoglobin (HbAA) gene. In addition, extensive

differences in genomic diversity of endothelial nitric oxide synthase (eNOS) genes, that had been reported to bear clinical significance on sickle cell pathogenesis, has been reported between Africans and African Americans (Thomas *et al.,* 2013). These polymorphisms have been shown to be potential modifiers of clinical disease, with significant differences reported between Indian and African sickle cell disease patients (Nishank *et al.,* 2013; Thakur *et al.,* 2014), and these differences could be potentially linked to disease haplotype. These interethnic differences can be attributed to the introduction of single nucleotide polymorphisms over a very long period, which can ultimately influence gene expression, protein structure and potentially function. Therefore, single nucleotide polymorphisms located in certain promoter regions can affect transcription thereby altering variability in the immune response, and contributing to disease susceptibility or host resistance (Sakuntabhai *et al.,* 2005). Despite the fact that African Americans can trace their ancestry to sub-Saharan Africa, recombination and genetic diversity in the African American gene pool has facilitated the introduction of single nucleotide polymorphisms leading to differing immune response to infectious pathogens. In addition, they are exposed to different groups of infectious agents compared to their African counterparts, which in turn directs immune system development, alongside circulating antibodies. These phenomena would undergo a similar diversification in the sickle cell population as well.

One of the most common immunogenetic markers, usually evaluated for immune system response and susceptibility to infectious pathogens is dendritic cell-specific ICAM-3 grabbing non-integrin (DC-SIGN) encoded by CD209. It assists in the migration dendritic cells on endothelium as well as enabling the activation of signal transduction pathways (Rappociolo *et al.,* 2006; Dettogni *et al.,* 2013). They are targets for pathogens, seeking to impair the immune response in early infection, and are known to recognize diverse pathogens, with reports showing association between CD209 gene polymorphisms and infectious agents (Mummidi *et al.,* 2001; Martin *et al.,* 2004). The guanine (G) to adenine (A) transition within the gene promoter (SNP -336 A/G; rs4804803) polymorphism has shown the most significance, demonstrating association with susceptibility to HIV, tuberculosis, leishmaniasis and dengue (Tailleux *et al.,* 2003; Tassaneetrithep *et al.,* 2003; Van Kooyk *et al.,* 2003; Martin *et al.,* 2004; Sakuntabhai *et al.,* 2005; Barreiro *et al.,* 2006). Sickle cell disease presents with variability in clinical severity, alongside genetic diversity and selection pressure imposed on patients by infectious diseases, leading to single nucleotide polymorphisms that can exacerbate or ameliorate disease outcome, especially among Africans, exposed to multiple infectious assaults and co-morbidities (Thomas *et al.,* 2012a, 2012b). We have shown that there is an extensive diversity in the ethnogenomic distribution of endothelial nitric oxide synthase (eNOS) polymorphisms (Thomas *et al.,* 2013). Despite reports to the contrary, we have also demonstrated that endothelin-1 polymorphisms rather than eNOS are the most important in African SCD (Thakur *et al.,*

2014). Therefore, since infections are common occurrences in SCD, there is a need to characterize the genomic diversity as well as haplotype frequency of immunogenetic markers and extrapolate their potential role in susceptibility to infectious diseases. This could clarify disease pathophysiology as well as their contribution to co-morbidities. To this end, what is the genotypic and allelic frequency of CD209 gene promoter polymorphisms (SNP -336 A/G; rs4804803) in control groups (African versus African American versus Caucasians) and between sickle cell disease populations (African versus African American), and does this polymorphism ameliorate or exacerbate disease pathophysiology? We will conduct our analyses using a polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) assay.

Materials and Methods

Subjects

This study encompasses sickle cell disease patients (cases) and control groups (Africans versus African Americans), as well as diverse ethnic groups (Africans, African Americans and Caucasians). The African portion was conducted at the Centre de Recherche et de Lutte contre la Drepanocytose (CRLD), a sickle cell disease treatment and referral center in Mali. Approval was received from the national ethical review board, and a written consent obtained before study was initiated. Inclusion criteria include diagnosis with sickle cell disease and presentation during crisis or during regular follow-up. Sickle cell disease demographic data has been

described previously (Thakur *et al.,* 2014). Healthy population controls comprised of family members or those recruited by word of mouth, able to provide informed consent and without a diagnosis of sickle cell disease. In the United States, control groups are African American and Caucasian self-identified individuals, recruited from Shreveport, Louisiana. African American sickle cell disease patients were recruited as part of the National Institute of Health-funded Cooperative Study of Sickle Cell Disease (CSSCD).

Samples and Genomic DNA Extraction

Discarded EDTA-anticoagulated blood samples, from 376 subjects (145 sickle cell disease patients and 231 controls) were spotted onto filter papers (GE Healthcare Sciences, Piscataway NJ) and genomic DNA samples extracted from the dried, spotted samples with the Qiagen Blood Mini Kit (Qiagen Inc., Valencia, CA), with some changes to the manufacturer's instruction (Thakur *et al.,* 2014). Final elution volume was 100 μl and DNA samples were stored at −20 °C until further analysis. Genomic DNA samples from African American sickle cell disease patients as well as African American and Caucasian controls were gratefully provided (Betty Pace, Georgia Regents University and Joann Moulds, Grifols USA respectively).

Genotyping for CD209 single nucleotide polymorphism

To genotype for the single nucleotide polymorphisms of the CD209 gene promoter, we utilized a previously published primer and PCR assay (Dettogni *et al.,* 2013), with a slight modification to the protocol. The

primer sequences are 5′- GGATGGTCTGGGGTTGACAG-3 (forward reaction) and 5′- ACTGGGGGTGCTACCTGGC-3′ (reverse reaction). 1 μl of genomic DNA served as template for PCR amplification, with conditions optimized to 25μl final volume and amplified using the Lucigen EconoTaq Plus Green 2X Master Mix PCR system (Lucigen Corporation, Middleton WI), as described previously (Thomas *et al.,* 2012), and PCR cycling parameters as published (Sakuntabhai et al., 2005). Amplified PCR products (5 μl) was examined on a 2% (w/v) agarose gel and photographed. Positive amplification yielded products of 150 bp, with size estimated with a TriDye 100 bp DNA ladder (New England Biolabs, Boston MA).

Restriction Fragment Length Polymorphism Assay

We utilized the *MscI* (New England Biolabs, Boston MA) restriction endonuclease for restriction fragment length polymorphism analysis of CD209 (DC-SIGN 336A/G) variants. 10 μl of PCR product was mixed with 0.5μl of enzyme (5000U/ml), 5μl of 1X CutSmart buffer and incubated at 37°C for 1 hour. Digested products were analyzed on an ethidium bromide-stained agarose gel, and band analysis carried out with a Doc-It LS Image Analysis Software (UVP Life Sciences, Upland CA). Restriction analysis was conducted by two investigators anonymously and 50% of amplified products subjected to repeat digestion (3rd investigator), with 100% concordance. Wild type variants (-336A/A) were undigested (150 bp) while mutant variants (-336G/G) produced bands of 131 and 19 bp.

Statistical analysis

Genotypic and allelic frequencies were determined with a simple PERL

script, as described previously (Thakur *et al.,* 2014). Differences in genotype and allele frequencies between populations were assessed by chi-square tests, while differences between sickle cell disease and controls were assessed by odds ratio. Tests for deviation from Hardy-Weinberg equilibrium (HWE) were performed, with SNP's rejected based on the recommended threshold of $p<0.001$ in control individuals.

Results

We found a wide disparity in the genetic diversity of the promoter variant of CD209 (DC-SIGN1-336A/G; rs4804803) gene polymorphisms in different populations. Genotypic frequencies of 23.4%, 16.9% and 3.2% were observed for the mutant variant between Africans, African Americans and Caucasians respectively (Table 1). Similar findings were made for the allelic frequencies (36.1%, 25.1% and 11.6% respectively), with a significant

difference in both genotypic and allelic frequencies (P<0.05) of CD209 gene promoter variants between all population groups. Surprisingly, the mutant variant (GG) is almost absent among Caucasians (3.2%). The genotypic and allelic frequencies of the mutant variant (snp-336GG) had the highest frequency among Africans (23.4% and 36.1% respectively). The wild type and heterozygote variants (AA and AG), that are necessary to facilitate dendritic cell activation and function during immune response, occurred at higher frequencies among African Americans (83.1%) and Caucasians (~97%), and an unprecedented low frequency among Africans (26%) (Fig 1).

We also examined the diversity of CD209 (snp 336A/G) gene promoter polymorphisms between sickle cell disease and healthy control groups in Africa and United States. There was a an extensive and significant disparity in the genotypic (Fig 2a, Table 2) frequency of the CD209 mutant variant (snp 336G/G) between sickle cell disease and control populations in Africa (P=0.002). Surprisingly, this was not the case between sickle cell disease and control populations recruited from the United States (Fig 2b) (P=0.54). In addition, the mutant variant has a higher frequency among healthy control groups than sickle cell patients (23.4% versus 10.4% respectively) in Africa, but no difference in the United States (16.9% versus 15.1% for controls and cases respectively). Similar observation was made for the allelic frequencies between controls and cases in Africa and United States (Table 3).

252 Since sickle cell disease has been known to display disease severity
253 between population groups, we evaluated the diversity of CD209 (snp
254 336A/G) gene promoter polymorphisms between sickle cell groups
255 recruited from Africa and United States. Surprisingly, there was no
256 difference either in genotypic (P=0.19) or allelic frequencies (P=0.72) of
257 mutant variants (snp 336G/G) between sickle cell disease groups (Table 4).
258 The similarities in the genotypic and allelic frequencies (10.4% versus
259 15.1% and 25.2% versus 28.1% for genotypes and alleles respectively) of
260 mutant variants were statistically insignificant.

261 **Discussion**

262 Sickle cell disease is the most commonly inherited hemoglobinopathy with
263 a worldwide distribution. It is a major disease represented in populations of
264 sub-Saharan Africa, the Middle East and several parts of India, and remains

a significant health burden borne by the African American population in the United States, and several Caribbean island nations, whose populations are dominated by ethnicities of African origin. It has recently been classified as a disease that would create a global challenge to the population of three major countries, therefore requiring a need to clarify, elucidate and decipher the various parameters contributing to its severity and diverse clinical pathophysiology among and between populations. To our understanding, this is the first report to elucidate the genomic diversity of CD209 promoter gene (snp-336A/G) polymorphisms in sickle cell disease, with the potential to clarify its role or otherwise in susceptibility to infectious pathogens. We chose three definitively classified populations, and as such permits conclusive inferences based on our finding. The African samples are from Mali facilitating analysis from a homogeneous population in comparison to the heterogeneous nature of the African American group.

Our observation that the CD209 promoter gene wild-type allele (snp-336A/A) occurred at a lower frequency among Africans compared to African Americans and Caucasians is significant, though not unexpected considering the degree of genetic admixture in the African American population. This is similar to our previous finding while examining the genomic diversity of endothelial nitric oxide synthase genes in differing populations (Thomas *et al.,* 2013). Though both populations share a common ancestry, it is expected that the several hundred years of sexual

recombination and both the uncomfortable and under-reported legacies of slavery would affect the genetic contribution of African genes into the African American genome. The wild type variant is necessary for dendritic cell activation and initiation of adaptive immune response. Therefore the reduced frequency of this allele among Africans potentially is a major contributing factor to their susceptibility to infectious pathogens. Unfortunately, sub-Saharan Africa is blessed with geographic and weather pattern that sustains the endemicity of many pathogens, especially neglected tropical diseases, and could potentially explain the often-encountered cases of disease co-morbidities with multiple infectious agents in a single host. In addition, this could be an evolutionary disadvantage in the African continent enhancing susceptibility and infectivity, thereby underscoring the preponderance of infections. The possibility that these infectious agents may have imposed a selection pressure on dendritic cells, that are imperative to initiate and exert immune pressure, is of potential significance and deserves further analysis.

In addition, this observation in Africans is enhanced by the reverse observation in the Caucasian population of the United States. The wild type variants (AA, AG) allele is ~97% among Caucasians and 83% among African Americans, with the mutant variant almost absent in both groups (3.2% among Caucasians and 16% among African Americans). This low genotypic frequency of the mutant variant is similar to results from previous reports, which showed 0%, 3% and 5% in the Taiwan, general

Brazilian and Sao Paulo populations respectively (Kashima *et al.,* 2009; Wang *et al.,* 2011; Dettogni *et al.,* 2013). In fact, in a study conducted among three groups of healthy control populations of Thailand, a similar scenario was observed, with a genotypic frequency of 5%, 1% and 3% (Sakuntabhai *et al.,* 2005). This observation potentially confirms our hypothesis that this immunogenetic marker has undergone evolutionary changes over time, conferring a selective advantage on populations outside of Africa (Miller *et al.,* 1994; Gibbons, 2001; Simmer *et al.,* 2001; Thomas *et al.,* 2005). In otherwords, populations with the wild type variant are able to fight infections, hence the reduced prevalence of infectious agents, while the reverse is the case in Africa. The ancestral-susceptibility model, which states that disease susceptibility alleles are ancestral while derived variants are protective, has been proposed and validated (Di Rienzo and Hudson, 2005; Biswas and Akey 2006). It further emphasizes that ancestral alleles were adapted to historical environmental conditions, becoming maladaptive based on changes in human lifestyle and dispersal into new environmental niches (Biswas and Akey 2006). In fact, extensive reports of geographically restricted selection have been found in genome-wide studies of humans and human diseases (Carlson *et al.,* 2005; Weir *et al.,* 2005; Voight *et al.,* 2006; Nakajima *et al.,* 2004; Zhou *et al.,* 2004; Sakagami *et al.,* 2004; Di Rienzo and Hudson, 2005; Young *et al.,* 2005). It seems clear therefore that local adaptation in extant populations is a major contributor to this observation (Fullerton *et al.,* 2002; Rockman *et al.,* 2004;

Thompson *et al.,* 2004), and is a confirmation of the out-of-Africa hypothesis (Biswas and Akey 2006; Thomas *et al.,* 2013).

Additionally, contrary to other reports, we conclude that the sickle cell gene potentially confer a protective mechanism against common infectious co-morbidities in Africa, based on our present observation. The higher frequency of CD209 mutants in the non-SCD group reveals an impaired capacity to mount an immune response to infectious diseases, potentially a contributor to the dominance of co-morbidities in this population. The red cell abnormality, which causes sickle cell disease, is probably protective in the present case, compared to normal individuals. We conclude that CD209 polymorphism play a major role in susceptibility to infectious pathogens among Africans and could potentially delineate severity of SCD. The implications of this finding for co-morbidities or as modifiers of SCD pathophysiology, and its significance in African Americans with SCD deserves extensive and detailed elucidation. The next step would be to determine if this protection is due to disease haplotypes and evaluate immunoassays for immunoglobulin E and eosinophilia as markers of common helminthic infections between both disease and control groups. Our endpoint would be to decipher the synergistic or pathogenic advantage of the sickle cell gene in disparate disease and population groups.

Acknowledgements

We are grateful to Joann Moulds, Grifols USA for the African American and Caucasian control DNA samples. Special gratitude to Betty Pace, Georgia Regents University, who graciously provided the African American sickle cell DNA samples, collected as part of the Cooperative Study of Sickle Cell Disease. Special thanks to Iuri Drummond Louro, Universidad Federal do Espirito Santo, Brazil for technical assistance. College of Health Sciences and Technology, Rochester Institute of Technology provided the funding for this project. Additional support by RIT Louis Stokes Alliance for Minority Participation (LSAMP) to JAN is also acknowledged.

List of Abbreviations

SCD: sickle cell disease; OR: odds ratio; HbAA: hemoglobin AA; HbSS: hemoglobin SS; PCR-RFLP: polymerase chain reaction-restriction fragment length polymorphism; ACS: acute chest syndrome.

Competing interests

The authors declare that they have no competing interests.

Author Contributions

BNT conceived and designed the experiment, and optimized protocols; AG and DAD carried out sample collection and sickle cell genotyping; BNT, KCD and JAN carried out DNA extraction, genotyping and restriction digestion; BNT drafted the manuscript; BNT and YL carried out the statistical analyses. All authors read and approved the final version of the manuscript. There are no conflicts of interest.

Author Details

[1]Department of Biomedical Sciences, College of Health Sciences and Technology, Rochester Institute of Technology, Rochester, NY

[2] Centre de Recherche et de Lutte contre la Drepanocytose (CRLD), Bamako, Mali

[3]School of Statistics, Shanxi University of Finance and Economics, Shanxi, China

References

1. Adamkiewicz TV, Sarnaik S, Buchanan GR, Iyer RV, Miller ST, Pegelow CH, Rogers ZR, Vichinsky E, Elliott J, Facklam RR, O'Brien KL, Schwartz B, Van Beneden CA, Cannon MJ, Eckman JR, Keyserling H, Sullivan K, Wong WY, Wang WC. 2003. Invasive pneumococcal infections in children with sickle cell disease in the era of penicillin prophylaxis, antibiotic resistance, and 23-valent pneumococcal polysaccharide vaccination. *Journal of Pediatrics* **143**:438–44
2. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. 2002. Interrogating a high density SNP map for signatures of natural selection. *Genome Research*, 12:1805–1814
3. Bandeira IC, Rocha LB, Barbosa MC, Elias DB, Querioz JA, Freitas MV, Gonçalves RP. 2014. Chronic inflammatory state in sickle cell anemia patients is associated with HBB(*)S haplotype. *Cytokine* **65**:217-21
4. Barreiro LB, Neyrolles O, Babb CL et al. 2006. Promoter variation in the DC-SIGN-encoding gene CD209 is associated with tuberculosis. *PLoS Medicine* **3**:e20
5. Benkerrou M, Delarche C, Brahimi L, Fay M, Vilmer E, Elion J, Gougerot-Pocidalo MA, Elbim C. 2002. Hydroxyurea corrects the dysregulated L-selectin expression and increased H2O2 production of polymorphonuclear neutrophils from patients with sickle cell anemia. *Blood* **99**:2297–303
6. Bunn HF. 1997. Pathogenesis and treatment of sickle cell disease. *New England Journal of Medicine* **337**:762–9

7. Carlson CS, Thomas DJ, Eberle MA, Swanson JE, Livingston RJ, Rieder MJ, Nickerson DA. 2005. Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Research*, 15:1553–1565

8. Charache S, Bleecker ER, Bross DS. 1983. Effects of blood transfusion on exercise capacity in patients with sickle-cell anemia. *American Journal of Medicine* **74**: 757–764

9. Conran N, Franco-Penteado CF, Costa FF. 2009. Newer aspects of the pathophysiology of sickle cell disease vaso-occlusion. *Hemoglobin* **33**:1–16

10. Dettogni RS, Sa RT, Tovar TT, Louro ID. 2013. Polymorphic genetic variation in immune system genes: a study of two populations of Espirito Santo, Brazil. *Molecular Biology Reports* **40**:4843-4849

11. Di Rienzo A, Hudson RR. 2005. An evolutionary framework for common diseases: the ancestral-susceptibility model. Trends in Genetics, 21:596–601

12. Di Rienzo A, Hudson RR. 2006. Genomic insights into positive selection. Trends in Genetics 22:437-446

13. Fullerton SM, Bartoszewicz A, Ybazeta G, Horikawa Y, Bell GI, Kidd KK, Cox NJ, Hudson RR, Di Rienzo A. 2002. Geographic and haplotype structure of candidate type 2 diabetes susceptibility variants at the calpain-10 locus. *American Journal of Human Genetics*, 70:1096–1106

14. Gibbons A. 2001. Tools show humans reach Asia early. *Science* **293**: 2368–2369

15. Halasa NB, Shankar SM, Talbot TR, Arbogast PG, Mitchel EF, Wang WC, Schaffner W, Craig AS, Griffin MR. 2007. Incidence of invasive pneumococcal disease among individuals with sickle cell disease before and after the introduction of the pneumococcal conjugate vaccine. *Clinical*

433 *Infectious Diseases* **44**:1428–33
434 16. Hassell KL. 2010. Populations estimates of sickle cell disease in the
435 United States. *American Journal of Preventive Medicine* **38**:S512-S521
436 17. Makis AC, Hatzimichael EC, Bourantas KL. 2000. The role of cytokines in
437 sickle cell disease. *Annals of Haematol*ology 79: 407–13
438 18. Martin MP, Lederman MM, Hutcheson HB, Goedert JJ, Nelson GW, van
439 Kooyk Y, Detels R, Buchbinder S, Hoots K, Vlahov D, O'Brien SJ, Carrington
440 M. 2004. Association of DC-SIGN promoter polymorphism with increased
441 risk for parenteral, but not mucosal, acquisition of human
442 immunodeficiency virus type 1 infection. *Journal of Virology* **78**:14053-
443 14056
444 19. McCavit TL, Quinn CT, Techasaensiri C, Rogers ZR. 2011. Increase in
445 invasive *Streptococcus pneumoniae* infections in children with sickle cell
446 disease since pneumococcal conjugate vaccine licensure. *Journal of*
447 *Pediatrics* **158**:505–7
448 20. Miller LH. 1994. Impact of malaria on genetic polymorphisms and
449 genetic diseases in Africans and African-Americans. *Proceedings of the*
450 *National Academy of Science of the USA* **91**: 2415–2419
451 21. Milner PF. 1982. Chronic transfusion regimens in sickle cell disease.
452 *Progress in Clinical and Biological Research* **98**:97–107
453 22. Mummidi S, Catano G, Lam L, Hoefle A, Telles V, Begum K, Jimenez F,
454 Ahuja SS, Ahuja AK. 2001. Extensive repertoire of membrane-bound and
455 soluble dendritic cell-specific ICAM-3-grabbing nonintegrin 1 (DC-SIGN1)
456 and DC-SIGN2 isoforms. Inter-individual variation in expression of DC-SIGN
457 transcripts. *Journal of Biological Chemistry* **276**:196-212
458 23. Nakajima T, Woodling S, Sakagami T, Emi M, Tokunaga K, Tamiya G,
459 Ishigami T, Umemura S, Munkhbat B, Jin F, Guan-Jun J, Hayasaka I, Ishida
460 T, Saitou N, Pavelka K, Lalouel JM, Jorde LB, Inoue I. 2004. Natural

selection and population history in the human angiotensinogen gene (AGT): 736 complete ATG sequences in chromosomes from around the world. *American Journal of Human Genetics*, 74:898–916

24. Nishank SS, Singh MP, Yadav R, Gupta RB, Gadge VS, Gwal A. 2013. Endothelial nitric oxide synthase gene polymorphism is associated with sickle cell disease patients in India. *Journal of Human Genetics* **58**:775-779

25. Overturf GD. 1999. Infections and immunizations of children with sickle cell disease. *Advances in Pediatrics Infectious Diseases* **14**:191–218

26. Pathare A, Al Kindi S, Alnaqdy AA, Daar S, Knox-Macaulay H, Dennison D. 2004. Cytokine profile of sickle cell disease in Oman. *American Journal of Hematology* **77**:323–328

27. Petz LD, Calhoun L, Shulman IA, Johnson C, Herron RM. The sickle cell hemolytic transfusion reaction syndrome. *Transfusion* **3**7:382–392

28. Piel FB, Hay SI, Gupta S, Weatherall DJ, Williams TN. 2013. Global burden of sickle cell anemia in children under five, 2010-2050: modeling based on demographics, excess mortality, and interventions. *PLoS Medicine* **10**:e1001484

29. Piomelli S. 1985. Chronic transfusions in patients with sickle cell disease. Indications and problems. *American Journal of Pediatric Hematology/Oncology* **7**:51–55

30. Qari MH, Dier U, Mousa SA. 2012. Biomarkers of inflammation, growth factor, and coagulation activation in patients with sickle cell disease. *Clinical and Applied Thrombosis/Hemostasis* **18**:195–200

31. Rappocciolo G, Jenkins FJ, Hensler HR et al. 2006. DCSIGN is a receptor for human herpes virus 8 on dendritic cells and macrophages. *Journal of Immunology* **176**:1741–1749

487 32. Rockman MV, Hahn MW, Soranzo N, Loisel DA, Goldstein DB, Wray GA.
488 2004. Positive selection on MMP3 regulation has shaped heart disease risk.
489 *Current Biology*, 14:1531–1539
490 33. Rosse WF, Gallagher D, Kinney TR, et al. 1990. Transfusion and
491 alloimmunization in sickle cell disease, *Blood* **76**:1431–1437
492 34. Sakagami T, Witherspoon DJ, Nakajima T, Jinnai N, Wooding S, Jorde LB,
493 Hasegawa T, Suzuki E, Gejyo F, Inoue I. 2004. Local adaptation and
494 population differentiation at the interleukin 13 and interleukin 4 loci.
495 *Genes and Immunity*, 5:389–397
496 35. Sakuntabhai A, Turbpaiboon C, Casademont I et al. 2005. A variant in
497 the CD209 promoter is associated with severity of dengue disease. *Nature*
498 *Genetics* **37**:507–513
499 36. Sakuntabhai A, Turbpaiboon C, Casadémont I, Chuansumrit A, Lowhnoo
500 T, Kajaste-Rudnitski A, Kalayanarooj SM, Tangnararatchakit K,
501 Tangthawornchaikul N, Vasanawathana S, Chaiyaratana W,
502 Yenchitsomanus PT, Suriyaphol P, Avirutnan P, Chokephaibulkit K, Matsuda
503 F, Yoksan S, Jacob Y, Lathrop GM, Malasit P, Desprès P, Julier C. 2005. A
504 variant in the CD209 promoter is associated with severity of dengue
505 disease. Nature Genetics 37:507-13
506 37. Steinberg MH. 2006. Pathophysiologically based drug treatment of
507 sickle cell disease. *Trends in Pharmacological Sciences* **27**:204–10
508 38. Szczepanek SM, Secor Jr ER, Bracken SJ, Guernsey L, Rafti E, Matson A,
509 Thrall RS, Andemariam B. 2013. Transgenic sickle cell disease mice have
510 high mortality and dysregulated immune responses after vaccination.
511 *Pediatric Research* **74**:141-147
512 39. Tailleux L, Schwartz O, Herrmann JL. et al 2003. DC-SIGN is the major

513 *Mycobacterium tuberculosis* receptor on human dendritic cells. *Journal of*
514 *Experimental Medicine* **197**:121–127
515 40. Tassaneetrithep B, Burgess TH, Granelli-Piperno A et al. 2003. DC-SIGN
516 (CD209) mediates dengue virus infection of human dendritic cells. *Journal*
517 *of Experimental Medicine* **197**:823–829
518 41. Tatari-Calderone Z, Tamouza R, Le Bouder GP, Dewan R, Luban NLC,
519 Lasserre J, Maury J, Lionnet F, Krishnamoorthy R, Girot R, Vukmanovic S.
520 2013. The association of CD81 polymorphisms with alloimmunization in
521 sickle cell disease. *Clinical and Developmental Immunology* 2013:1-9
522 42. Taylor JG 6th, Nolan VG, Mendelsohn L, Kato GJ, Gladwin MT, Steinberg
523 MH. Chronic hyper-hemolysis in sickle cell anemia: association of vascular
524 complications and mortality with less frequent vasoocclusive pain. *PLos*
525 *One* **3**:e2095
526 43. Telfer P. 2011. Management of sickle cell disease: out-patient and
527 community aspects. Pediatrics and Child Health 21:357-362
528 44. Thakur TJ, Guindo A, Cullifer LR, Li Y, Imumorin IG, Diallo DA, Thomas
529 BN. 2014. Endothelin-1 but not endothelial nitric oxide synthase gene
530 polymorphism is associated with sickle cell disease in Africa. *Gene*
531 *Regulation and Systems Biol*ogy **8**:119-26
532 45. Thomas BN, Diallo DA, Noumsi GT, Moulds JM. 2012a. Circulating
533 immune complex levels are associated with disease severity and
534 seasonality in children with malaria from Mali. *Biomarker Insights* **7**:81-86
535 46. Thomas BN, Donvito B, Cockburn I, Fandeur T, Rowe JA, Cohen JHM,
536 Moulds JM. 2005. A complement receptor-1 polymorphisms with high
537 frequency in malaria endemic regions of Asia but not Africa. *Genes and*
538 *Immunity* **6**:31-36
539 47. Thomas BN, Petrella CR, Thakur TJ, Crespo SR, Diallo DA. 2012b.
540 Genetic polymorphism of *Plasmodium falciparum* merozoite surface

protein-1 and 2 and diversity of drug resistance genes in blood donors from Bamako, Mali. *Infectious Diseases: Research and Treatment* **6**:49–57

48. Thomas BN, Thakur TJ, Yi L, Guindo A, Diallo DA, Ott JG. 2013. Extensive ethnogenomic diversity of endothelial nitric oxide synthase (eNOS) polymorphisms. *Gene Regulation and Systems Biol*ogy **7**:1–10

49. Thompson EE, Kuttab-Boulos H, Witonskky D, Yang L, Roe BA, Di Rienzo A. 2004. CYP3A variation and the evolution of salt-sensitivity variants. *American Journal of Human Genetics*, 75:1059–1069

50. Van Kooyk Y, Appelmelk B, Geijtenbeek TB. 2003. A fatal attraction: mycobacterium tuberculosis and HIV-1 target DC-SIGN to escape immune surveillance. *Trends in Molecular Medicine* **9**:153–159

51. Voight BF, Kudaravalli S, Wen X, Pritchard JK. 2006. A map of recent positive selection in the human genome. *PLoS Biology*, 4:e72

52. Wang ET, Kodama G, Baldi P, Moyzis RK. 2006. Global landscape of recent inferred darwinian selection for Homo sapiens. *Proceedings of the National Academy of Sciences USA*. 103:135–140

53. Weatherall DJ. 2010. The inherited diseases of hemoglobin are an emerging global health burden. *Blood* **115**:4331–4336

54. Weir BS, Cardon LR, Anderson AD, Nielsen DM, Hill WG. 2005. Measures of human population structure show heterogeneity among genomic regions. *Genome Research*, 15:1468–1476

55. Yazdanbakhsh K, Ware RE, Noizat-Pirenne F. 2012. Red blood cell alloimmunization in sickle cell disease: pathophysiology, risk factors, and transfusion management. *Blood* **120**: 528–537

56. Young JH, hang YP, Kim JD, Chretien JP, Klag MJ, Levine MA, Ruff CB,

566 Wang NY, Chakravarti A. 2005. Differential susceptibility to hypertension is
567 due to selection during the out-of-Africa Expansion. *PloS Genetics*, 1:e82
568 57. Zimmer C. 2001. Genetic trees reveal disease origins. *Science* **292**:
569 1090–1093

570 **Figure Legends**

571 **Fig 1**. Genotypic distribution of CD209 gene promoter polymorphisms (SNP
572 -336 A/G; rs4804803) in Caucasian, African American and African control
573 populations. Wild type variant (snp-336A) showed no digestion (150 bp),

while mutant variant (snp-336G) produced two bands (131 and 19 bp) on digestion (lower band size not shown). Marker: 100 bp ladder, where the 500 bp band stains most intensely (New England Biolabs). Black bars-wild-type homozygotes (AA); blue bars-heterozygotes (AG); red bars-homozygote mutants (GG)

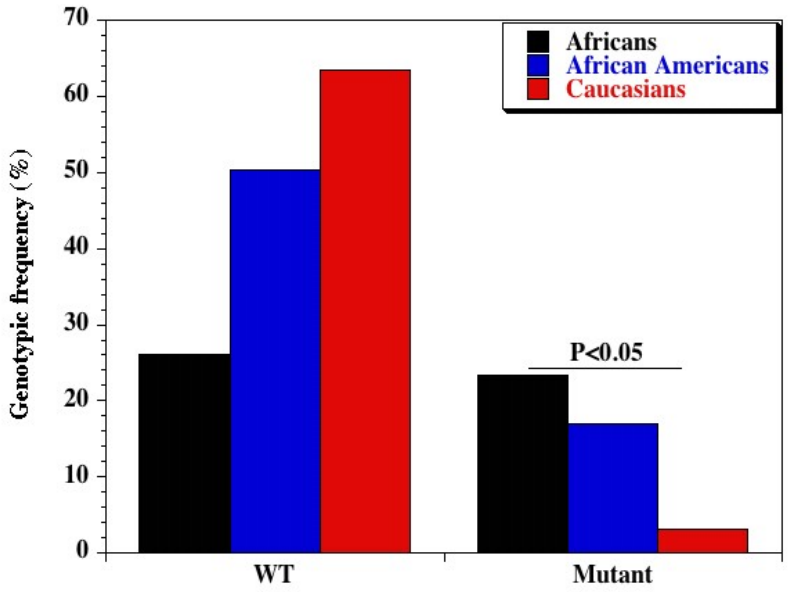

Figure 1

Fig 2: Genotypic frequency of CD209 gene promoter polymorphisms (SNP -336 A/G; rs4804803) in African (Fig 2a) and African American (Fig 2b) sickle cell disease and control groups. Amplified PCR products were digested with *MscI* restriction endonuclease (Fisher Scientific), and

584 expressed on a 2% ethidium bromide-stained agarose gel. Wild type
585 variant (snp-336A) showed no digestion (150 bp), while mutant variant
586 (snp-336G) produced two bands (131 and 19 bp) on digestion (lower band
587 size not shown). Marker: 100 bp ladder, where the 500 bp band stains most
588 intensely (New England Biolabs). Blue bars-sickle cell disease; red bars-
589 control groups

590

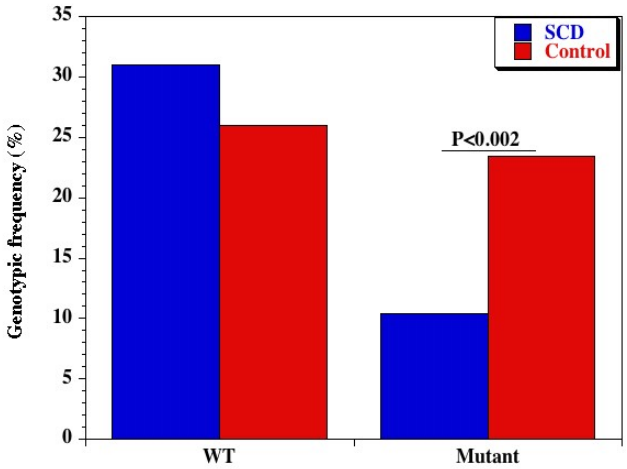

Fig 2a

591

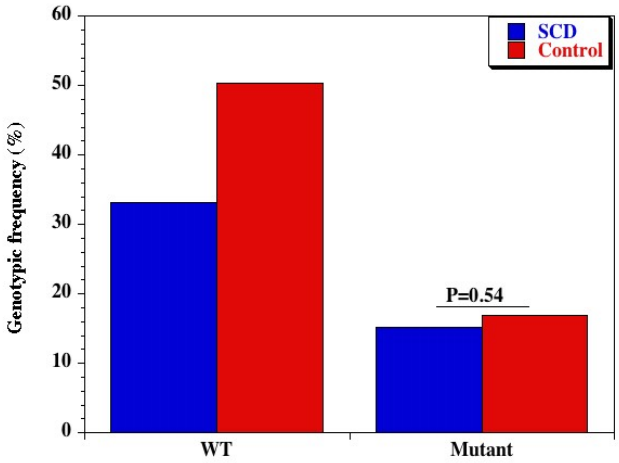

Fig 2b

Table 1(on next page)

Table 1. Genotypic and allelic frequency of CD209 polymorphisms in diverse populations

Table 1. Genotypic and allelic frequency of CD209 polymorphisms in diverse populations

Polymorphism	Genotype	Ethnic groups			Chi square	*P*-value
		African	African American	Caucasian		
		n=244 (%)	n=379 (%)	n=159 (%)		
CD209 (rs4804803)	A/A	60 (26.0)	191 (50.4)	101 (63.5)	59.9243	9.72E-14
	A/G	117 (50.6)	124 (32.7)	53 (33.3)	21.5787	2.06E-05
	G/G	54 (23.4)	64 (16.9)	5 (3.2)	29.1326	4.72E-07
		Allelic diversity				
	Allele	**n=488 (%)**	**n=758 (%)**	**n=318 (%)**	**Chi square**	***P*-value**
CD209 (rs4804803)	A	179 (38.7)	444 (58.6)	229 (72.0)	83.7253	<2.2E-16
	G	167 (36.1)	190 (25.1)	37 (11.6)	83.7253	<2.2E-16

Percentile frequency of the genotypes and alleles at CD209 locus, determined among African, African American and Caucasian ethnic populations. Africans were recruited from Mali while African American and Caucasian populations were recruited from Louisiana. Odds ratio was calculated by Fisher's two-tailed exact test

Table 2(on next page)

Table 2. Genotypic frequency of CD209 polymorphisms between sickle cell and control groups

Table 2. Genotypic frequency of CD209 polymorphisms between sickle cell and control groups

African					
Polymorphism	**Genotype**	**SCD: n=145 (%)**	**Controls: n=231 (%)**	**Odds ratio (95% CI)**	***P*-value**
CD209 (rs4804803)	A/A	45 (31.0)	60 (26.0)	1.28 (0.79-2.08)	0.2907
	A/G	85 (58.6)	117 (50.6)	1.38 (0.89-2.15)	0.1382
	G/G	15 (10.4)	54 (23.4)	0.38 (0.19-0.72)	0.002
African American					
		SCD: n=331 (%)	**Controls: n=379 (%)**	**Odds ratio (95% CI)**	***P*-value**
CD209 (rs4804803)	A/A	110 (33.2)	191 (50.4)	0.49 (0.36-0.67)	4.70E-06
	A/G	171 (51.7)	124 (32.7)	2.20 (1.60-3.01)	4.33E-07
	G/G	50 (15.1)	64 (16.9)	0.88 (0.57-1.33)	0.54

Abbreviations: SCD, sickle cell disease; NS, not significant; CI, confidence interval

Percentile frequency of the genotypes at CD209 locus, determined among African American sickle cell disease patients and control groups. Sickle cell disease populations were recruited from Mali and Georgia. Control populations (individuals without sickle cell disease) were recruited from Mali and Louisiana. Odds ratio was calculated by Fisher's two-tailed exact test

Table 3(on next page)

Table 3. Allelic frequency of CD209 polymorphisms between sickle cell and control groups

Table 3. Allelic frequency of CD209 polymorphisms between sickle cell and control groups

African					
Polymorphism	**Allele**	**SCD: n=290 (%)**	**Controls: n=462 (%)**	**Odds ratio (95% CI)**	***P*-value**
CD209 (rs4804803)	A	133 (45.9)	179 (38.7)	1.70 (1.17-2.47)	0.003432
	G	73 (25.2)	167 (36.1)	0.59 (0.41-0.85)	0.003432
African American					
		SCD: n=662 (%)	**Controls: n=758 (%)**	**Odds ratio (95% CI)**	***P*-value**
CD209 (rs4804803)	A	306 (46.2)	444 (58.6)	0.70 (0.54-0.91)	0.006167
	G	186 (28.1)	190 (25.1)	1.42 (1.10-1.84)	0.006167

Abbreviations: SCD, sickle cell disease; NS, not significant; CI, confidence interval

Percentile frequency of the genotypes at CD209 locus, determined among African American sickle cell disease patients and control groups. Sickle cell disease populations were recruited from Mali (African) and Augusta GA (African American). Control populations (individuals without sickle cell disease) were recruited from Mali and Louisiana. Odds ratio was calculated by Fisher's two-tailed exact test

Table 4(on next page)

Table 4. Genotypic and allelic frequency of CD209 polymorphisms between sickle cell disease groups

Table 4. Genotypic and allelic frequency of CD209 polymorphisms between sickle cell disease groups

Genotypic frequency					
Polymorphism	**Genotype**	**Mali: n=145 (%)**	**USA: n=331 (%)**	**Odds ratio (95% CI)**	***P*-value**
CD209 (rs4804803)	A/A	45 (31.0)	110 (33.2)	0.90 (0.59-1.40)	0.67
	A/G	85 (58.6)	171 (51.7)	1.32 (0.87-2.00)	0.16
	G/G	15 (10.4)	50 (15.1)	0.65 (0.33-1.23)	0.19
Allelic frequency					
Polymorphism	**Allele**	**Mali: n=290 (%)**	**USA: n=662 (%)**	**Odds ratio (95% CI)**	***P*-value**
CD209 (rs4804803)	A	133 (45.9)	306 (46.2)	1.05 (0.79-1.41)	0.72
	G	73 (25.2)	186 (28.1)	0.95 (0.71-1.27)	0.72

Abbreviations: SCD, sickle cell disease; NS, not significant; CI, confidence interval

Sickle cell disease populations were recruited from Mali (African) and Augusta GA (African American), while control populations, who are individuals without sickle cell disease), were recruited from Mali and Louisiana. A/G denotes the alleles at the locus. Odds ratio was calculated by Fisher's two-tailed exact test

Figure 1

Fig 1. Genotypic distribution of CD209 gene promoter polymorphisms (SNP -336 A/G; rs4804803) in Caucasian, African American and African control populations

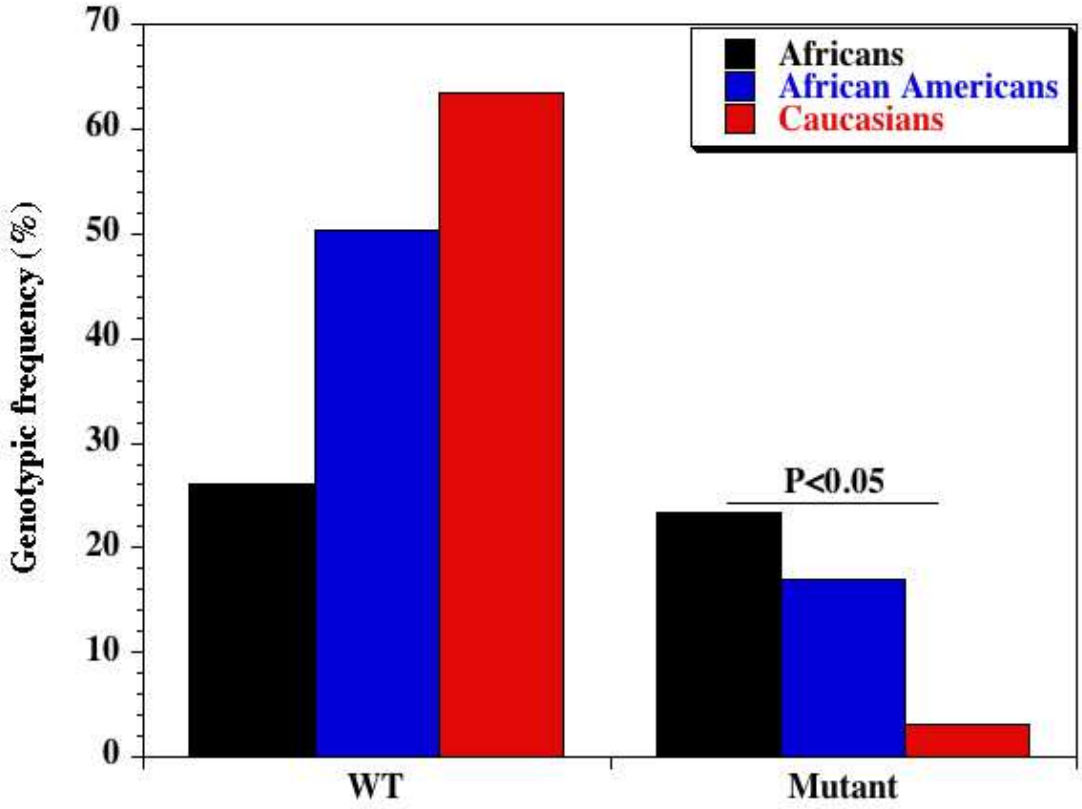

Figure 2

Fig 2a. Genotypic frequency of CD209 gene promoter polymorphisms (SNP -336 A/G; rs4804803) in African sickle cell disease and control groups

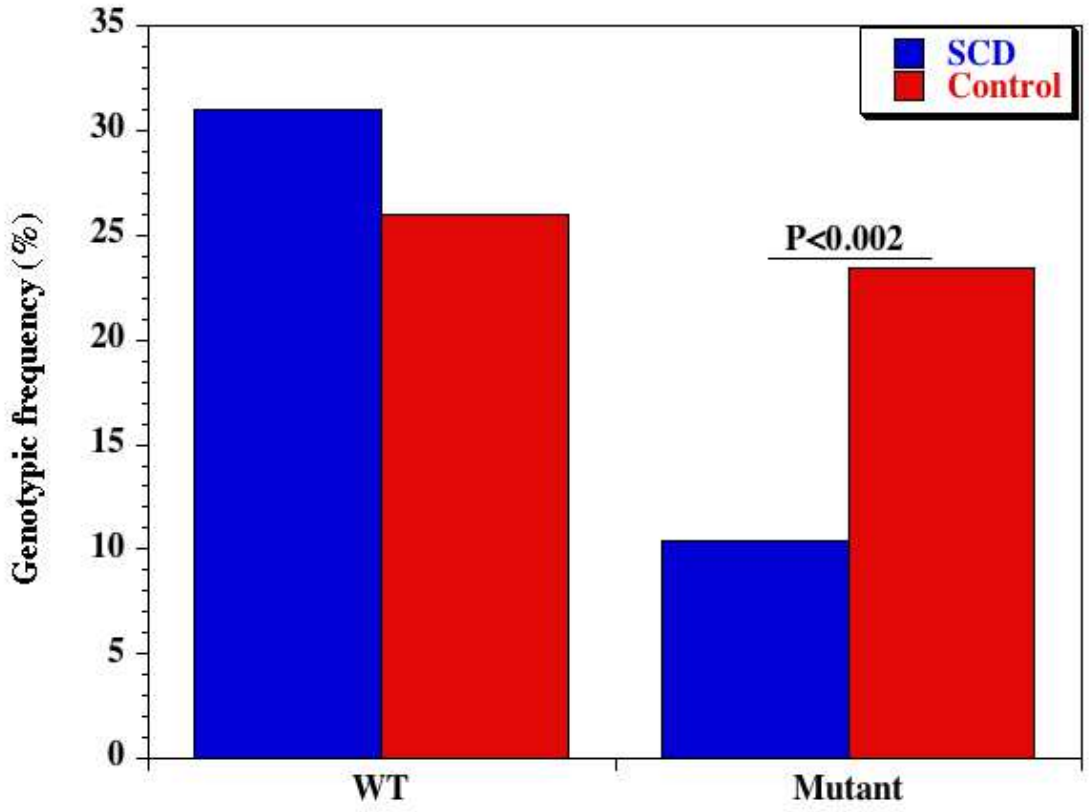

Figure 3

Fig 2b. Genotypic frequency of CD209 gene promoter polymorphisms (SNP -336 A/G; rs4804803) in American sickle cell disease and control groups