Plasmodium vivax genome evolution for erythrocyte invasion in Duffy-negative infections from East Africa

Anthony Ford[1,2], Cheikh Cambel Dieng[3], Beka Raya Abagero[3], Guiyun Yan[4], Delenesaw Yewhalaw[5], Daniel A. Janies[1,2*], Eugenia Lo[3*]

[1] Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, USA

[2] Center for Computational Intelligence to Predict Health and Environmental Risks, University of North Carolina at Charlotte, USA

[3] Department of Microbiology and Immunology, College of Medicine, Drexel University, USA

[4] School of Public Health, University of California at Irvine, USA

[5] Tropical Infectious Disease Research Center, Jimma University, Ethiopia

***Correspondence:** Daniel A. Janies: djanies@charlotte.edu; Eugenia Lo: el855@drexel.edu

Running Title: *Plasmodium vivax* genome evolution for erythrocyte invasion

Abstract

Compared to other regions around the world, *Plasmodium vivax* (*Pv*) malaria is historically less common in Africa due to the lack of Duffy antigen receptor for chemokines (DARC) expression on host erythrocytes that *Pv* Duffy Binding Protein (DBP1) interacts and utilizes for erythrocyte invasion. However, increased reports of *Pv* across Africa lead to the hypothesis that the parasites have evolved alternate invasion mechanisms that are DARC-independent but rely on other ligand-receptor interactions. In this study, we analyzed single nucleotide polymorphisms (SNPs) based on whole genome sequence data of 110 *Pv* samples isolated from 38 Duffy-negative and 72 Duffy-positive individuals in Ethiopia. SNPs detected in *Pv* from Duffy positives (477,561 SNPs) were more than two-fold higher than Duffy negatives (197,461 SNPs). Chromosomes 9, 10, and 12 were shown to be most variable in both Duffy groups. Among 43 erythrocyte binding gene candidates, tryptophan rich antigen 3 and 34 (TRAg3 and TRAg34) and members of the merozoite surface protein (MSP) 3 family had the highest variation. In Duffy positive *Pv*, significant signal of positive selection was observed in genes related to drug resistance (chloroquine resistance associated protein CG1 and the 26S proteasome regulatory subunit RPN2) and erythrocyte binding (membrane associated erythrocyte binding-like protein MAEBL); whereas in Duffy negative *Pv*, positive selection was observed in the plastid replication repair enzyme and the AP-5 complex subunit beta 1 genes associated with parasite mobility, transport and DNA repair. Several amino acid substitutions were found in *PvDBP*1, *PvEBP/DBP*2, and *PvRBP*2b of Duffy positive but not Duffy negative samples, with almost half of the mutations located in the binding regions. No genetic differentiation was detected

between Duffy positive and Duffy negative *Pv* from Ethiopia. These isolates were genetically similar to other East Africa *Pv*, but markedly different from those from Southeast Asia and South America. The highly conserved *Pv* genomes in Duffy negatives suggest stringent selective pressures and restricted binding interactions of *Pv* with Duffy negative erythrocytes. These findings shed light on alternate parasite invasion mechanisms in Duffy negative Africans.

Keywords: *Plasmodium vivax*, Duffy negative, genomic variation, erythrocyte invasion

Introduction

Plasmodium vivax (*Pv*) and *P. falciparum* (*Pf)* are two major parasite species accounting for the majority of malaria cases globally. While *Pf* is the most lethal, *Pv* poses a significant health burden due to its widespread distribution across Southeast Asia, South America, the Middle East, and Africa. In Africa, *Pv* cases have been reported in Ethiopia, Botswana, Cameroon, Angola and Equatorial Guinea, Senegal, Sudan, and beyond.[1-4] Despite an estimated 2 billion people at risk of the infection, *Pv* malaria is often overlooked as a neglected tropical disease due to its lower mortality rate compared to *Pf*.[5] African human populations were thought to be protected from *Pv* because people of African descent often lack the expression of Duffy Antigen Receptor for Chemokines (DARC), a protein on the surface of host erythrocytes that is central to *Pv* invasion.[6] However, several recent studies reported *Pv* infection in Duffy-negative individuals broadly across Africa including areas where Duffy negatives are predominant. It is hypothesized that *Pv* has undergone evolutionary events that allow erythrocyte invasion independent of DARC. Though *P. vivax* infections in Duffy-negative individuals may be less severe, successful adaptation and transmission of the parasites in Duffy-negative human populations require reassessment of malaria control and treatment strategies in Africa and elsewhere.[7,8]

High levels of polymorphism have been observed in the *Pv* genome, likely as a result of gene flow due to human migration, intense parasite transmission, and/or variations in host susceptibility.[9] In *Pf,* merozoites use several erythrocyte receptors and ligands to facilitate erythrocyte invasion. In *Pv,* merozoites primarily use Duffy binding protein

(PvDBP1) that interacts with DARC for erythrocyte invasion. Our previous study showed that *PvDBP1* mutations unique to *Pv* in Duffy-negative Ethiopians did not lead to binding of Duffy-negative erythrocytes *in vitro*.[7,10] Yet, other ligands such as *PvRBP1a*, *PvRBP1b*, *PvRBP2a* and *PvRBP2*-P2 in the reticulocyte binding protein (*PvRBP*) gene family were known to play a role in invasion, with *PvRBP2b* binding to transferrin receptor 1 (TfR1) prior to reticulocyte invasion.[11-16] The Long Homology Repeat A (LHR-A) subdomain of the complement receptor 1 (CR1) was recently shown to bind to region II of *PvEBP/DBP2*.[17] Furthermore, highly polymorphic genes in the *PvMSP* and *PvTRAg* multigene families have been shown to associate with erythrocyte invasion and host immunity evasion.[18-21] These different potential invasion mechanisms independent of Duffy antigen may account for widespread distribution of *Pv* in Africa.

Recent analysis of 730 global *Pv* isolates showed lower genetic polymorphisms in African *Pv* compared to isolates from Southeast Asia and South America, pointing to the notion that *Pv* could be originated from Southeast Asia and genetic diversity decreases and linkage disequilibrium increases as isolates spread out to other geographical regions.[22-24] Among the African isolates, those from Eritrea and Sudan clustered together and were mostly distinct from the Ethiopian *Pv*.[23] At the genome level, we previously observed high genetic variations on chromosomes 9 and 10 in *Pv* from Duffy positive Ethiopians.[9] Among 43 erythrocyte-binding gene candidates, the *MSP3* multigene family (*MSP3.5*, *MSP3.8*, *MSP3.9*) on chromosome 10 was most variable in the Duffy positives. By contrast, *PvDBP1* and *PvEBP/DBP2*, both located on chromosome six, as well as the *TRAg* gene family exhibited relatively low variation. It is

yet unclear the level of polymorphisms of the genomes, and particularly genes related to host-parasite interactions in *Pv* from Duffy negative Africans. Because of the highly polymorphic nature of *Pv*, genetic differences between Duffy negative and Duffy positive individuals could be a function of evolved mutations emerged from parasite's survival, local selection, human migrations, intense transmission, and/or host susceptibility. In this study we compare the extent of genomic differences between *Pv* in Duffy-positive and Duffy-negative Ethiopians, determine genetic variants in 43 erythrocyte binding genes that play a role in host-parasite interactions, and infer the evolutionary patterns of *Pv* from Duffy-negative Ethiopians relative to other geographical regions.[9] Findings shed light on alternate or evolved invasion mechanisms *Pv* uses, as well as clarify the origins of African *Pv* leading to improved methods of malaria control and treatment.

Materials and Methods

Ethics statement

Scientific and ethical clearance was obtained from the Institutional Scientific and Ethical Review Boards of Jimma University, Ethiopia and Drexel University, USA. Written informed consent/assent for study participation was obtained from all consenting heads of households, parents/guardians (for minors under 18 years old), and each individual who was willing to participate in the study.

Sample Processing and Whole Genome Sequencing

A total of 120 *Pv*-confirmed whole blood samples were included, of which 40 were from Duffy negative and 80 from Duffy-positive patients. We used a

Lymphoprep™/Plasmodpur™-based protocol to deplete white blood cells and enrich red blood cell (RBC) pellets. Genomic DNA was extracted from approximately 1 mL of RBC pellets using Quick-DNA™ Miniprep Kit (Zymo Research), following the manufacturer's protocol. DNA quality and concentration were assessed using NanoDrop™ spectrophotometer. DNA library was prepared using the NEBNext® Ultra™ II DNA Library Prep Kit (New England BioLabs) following the manufacturer's protocol for fragmentation, end repair, ligation, and PCR enrichment. Library quality and quantity were evaluated using Qubit fluorometry and Agilent TapeStation. Sequencing was carried out on the Illumina NovaSeq™ 6000 platform, targeting a sequencing depth of 30X per sample to ensure comprehensive coverage of the genome. Quality control metrics included average Phred quality scores above 30 (Q30) across all bases and all samples that met the minimum threshold for both read depth and quality, ensuring reliable downstream analysis. Only samples with monoclonal infections, as confirmed by microsatellite genotyping were selected for sequencing. After sequencing, 110 of the 120 total samples, with 38 Duffy negative and 72 Duffy positive, were proceeded for further analysis. Illumina sequencing reads were trimmed using Trimmomatic.

An additional 177 samples of whole genome sequencing (WGS) data were retrieved from MalariaGen and select studies.[25,26] These data represent *Pv* isolates from Southeast Asia (Thailand, Cambodia), South America (Peru, Brazil, Guyana), and other parts of Africa (Uganda, Sudan, Madagascar and Eritrea). All FASTQ files were mapped to the P01 reference genome, which was derived from a *Pv* clinical patient from the Papua region of Indonesia, using BWA-mem v2 with default parameters.[27-30]

Compared to the earlier Salvadoran-1 *Pv* assembly, the P01 assembly offers improved completeness in the subtelomeric regions of the *Pv* genome [cite]. Post-mapped reads were subjected to quality control checks of the alignment metrics including removal of duplicates and assessment of mapping quality using Samtools v1.10 and Picard.[31,32] Only reads that were mapped to the reference genome were included in further analyses. All genome data were deposited in NCBI SRA database (PRJNA1255830).

Variant Discovery, Nucleotide Diversity, and Codon Prevalence

Variant call format (VCF) files were generated using GATK's HaplotypeCaller procedure with the best practices workflow.[33] To ensure high-confidence variant calls, insertions and deletions were excluded from further analysis and SNPs were retained only if the following quality thresholds were met: QD (quality by depth) less than 2.0, QUAL (read quality) less than 30, SOR (strand odds ratio) greater than 3·0, and MQ (map quality) less than 40. Additionally, the latest version of gene annotation files were retrieved from PlasmoDB and functional annotations were applied to the remaining high quality variants using snpEff v5.1.[29,] Nonsynonymous and synonymous variants were compared between the two Duffy groups across the 14 chromosomes in the nuclear genome. Coding sequences for 43 erythrocyte binding genes, which include genes of the *PvTRAg*, *PvMSP*, *PvRBP*, *PvGAMA*, and *PvDBP* multigene families, were extracted from the assembled genomes.[9] Nucleotide diversity was calculated for each target gene and compared between Duffy-positive and Duffy-negative samples excluding gaps from pairwise comparison. Because the number of Duffy-positive samples exceeded the Duffy-negative samples (72 to 38), we employed a bootstrap resampling approach to

control for differences in sample size and to assess the robustness of the nucleotide diversity calculations. For each gene, a fixed number of samples (n=38, equal to the number of Duffy-negative samples) were randomly sampled with replacement from the Duffy-positive population. This procedure was repeated 500 times per gene and for each bootstrap replication, we calculated the nucleotide diversity. A summary of the bootstrap results is included in Supplementary Table 2. All resampling, computation and data aggregation were implemented in a custom Python script using pandas and NumPy. Amino acid variation for *PvDBP1*, *PvEBP/DBP2*, and *PvRBP2b* was calculated using an in-house python script, with the relative frequency of amino acid substitutions recorded at each codon site.

Test for signal of positive selection

We detected loci under recent positive selection using the integrated haplotype score (iHS) implemented in the R-based package *rehh*.[35,36] For iHS analysis, the reference and non-reference alleles were used in place of the ancestral and derived alleles. We used a two-sided p-value for significance testing under the null hypothesis of no evidence of signals of recent selection. SNPs with iHS scores that had a *p*-value of -$\log_{10}$(*p*-value) greater than four were considered significant and therefore identified as undergoing potential positive selection.

Population structure and differentiation

Population genetic cluster analysis was assessed using the Admixture software, which estimates individual ancestry proportions using an accelerated maximum likelihood

estimation by means of a novel quasi-Newton acceleration method. The analysis used a combination of the 110 Duffy-positive and Duffy-negative Ethiopian genomes alongside 177 publicly available genomes representing diverse global regions from South America and Asia.[37].[37] Clustering was performed for 1 to 6 clusters (*K*) and cross-validation scores were calculated for each run, with the optimal *K* being the smallest cross-validation error.

Results

SNP discovery and nucleotide diversity

In subsequent analyses, we examine genetic diversity in 72 Duffy-positive samples and 38 Duffy-negative samples relative to the P01 reference genome. A total of 477,561 high quality SNPs were detected in *Pv* from Duffy-positive individuals, more than two-fold higher than the 197,461 SNPs observed in Duffy negatives. Among all observed SNPs, 134,266 were shared between the two Duffy groups, 343,295 were unique to Duffy-positive and 63,195 were unique to Duffy-negative *Pv*. Across the genomes, the highest genetic variations were observed in chromosomes 9, 10, and 12 of *Pv* in Duffy-positive individuals (Figure 1). Similar pattern was observed in the Duffy negatives, despite a lower number of nonsynonymous to synonymous SNPs.

For the 43 erythrocyte binding genes, a total of 13,527 SNPs including 8,018 synonymous and 5,509 nonsynonymous mutations were detected in Duffy positive samples, with an average nucleotide diversity of $4{\cdot}39 \times 10^{-3}$ among Duffy positive *Pv*. The lowest genetic polymorphisms were observed in *PvEBPDBP*2, *PvTRAg*2B, *PvTRAg*7,

and *PvRON*4, which account for 0·3-0·5% of all SNPs in these genes. By contrast, markedly fewer SNPs (3,168 synonymous and 2,877 nonsynonymous SNPs) were detected in Duffy-negative samples, with an average nucleotide diversity of $0{\cdot}98 \times 10^{-3}$. The lowest polymorphisms were observed in *PvTRAg*21, *PvEBPDBP*2, *PvRBP*1a and *PvRON*4 which accounted for 0·2-0·4% of all SNPs in these genes. In both Duffy negative and positive *Pv*, the highest polymorphisms were observed in *PvTRAg34* and members of the *PvMSP3* multigene family including *PvMSP*3G, *PvMSP*3.11, *PvMSP*3.10, and *PvMSP*3.8, despite Duffy-positive showing higher polymorphisms than Duffy-negative *Pv* samples (Figure 2). These four genes accounted for 46% of the 13,527 SNPs in Duffy positive and 43% of the 6,212 SNPs in Duffy negative *Pv*.

Amino acid substitutions and gene differentiation

In *PvDBP*1, a total of 35 codon mutations were detected in Duffy-positive *Pv*, each ranging from 12-51% of the isolates. Of these 35 mutations, 15 were located in the binding region. Over 50% of Duffy positive samples were observed with mutation at four codon positions including G175E, S263R, K402S, and I458K, all were within the binding region (Figure 3A). In contrast, less than 5% of Duffy negative samples had these mutations in the same position. In Duffy-negative *Pv*, 20 codon mutations were detected and they all accounted for only 5-10% of the isolates. In *PvEBP/DBP*2, 14 codon mutations were detected in Duffy positive *Pv,* ranging from 10-43% of the isolates (Figure 3B). Eight of these mutations occurred in the binding region with almost 50% of the samples having mutations 306E and 421T. However, in Duffy negative *Pv*, only four mutations were detected, each account for 5-7% of the isolates. The two mutations

306E and 421T that occurred in the binding region were found in less than 5% of the samples. A similar pattern was observed in *PvRBP*2b, of which 37 codon mutations were observed in Duffy-positive *Pv,* ranging from 1-46% of the isolates, but only 16 mutations were detected in 5-8% of Duffy-negative *Pv* (Figure 3C). Among the 37 mutations, 23 occurred in the binding region with up to 50% of Duffy positive *Pv* having these mutations (e.g., 217R and 558D). Eight mutations were detected in the *PvRBP2b* binding region for Duffy negative *Pv*, but only less than 5% of the samples had these mutations. No mutations were detected in binding region of *PvEBP/DBP*2 nor *PvRBP*2b that were markedly different between the Duffy-positive and Duffy-negative samples. For other erythrocyte gene including *PvMSP1*, *PvRBP1a*, and *PvRBP2a*, four amino acid mutations were detected in >50% of Duffy positive but <10% of the Duffy-negative samples (Supplementary Figure 1A-E). They included E1692K in *PvMSP1*, P552Q in *PvRBP1a*, and E438G and M511K in *PvRBP2a*. These mutations were located at the binding sites of the protein ligand. Overall, *PvDBP*1, *PvEBP/DBP2*, and *PvRBP*2b were genetically more conserved in Duffy-negative than Duffy-positive *Pv*, with *PvEBP/DBP*2 being the most conserved among the three genes.

Analysis of positive selection

The iHS estimates for *Pv* in Duffy-positive Ethiopians showed significant signal of positive selection in six genes or gene regions across the 14 chromosomes ($p<0{\cdot}001$; Figure 4A). Two of the genes were related to antimalarial drug resistance, including the chloroquine resistance associated protein 1 Cg1 (PVP01_0109400) located on chromosome one and the 26S proteasome regulatory subunit RPN2, of which mutation

was shown to confer resistance to artemisinin in *Pf*. In *PvCg*1, 72% of the samples had the mutant allele A (G to A). In *PvRPN*2, 69% of the samples had the mutation. These SNPs and flanking regions had an overall sequencing depth greater than 10. PVP01_0533300 (*PvSTP1*) on chromosome five, a gene which shows structural and sequence similarities with surface associated interspersed genes (*surf* genes) and is expressed on the surface of both infected erythrocytes and merozoites, also showed signs of recent positive selection (p<0·001).[38] In Duffy positive *Pv*, most of the SNP variants on chromosome nine that show signals of positive selection occurred mostly in the intergenic regions upstream and downstream of hypothetical proteins, with the exception of the membrane associated erythrocyte binding like protein *MAEBL* (PVP01_0948400). Contrary to Duffy-positive *Pv*, signals of positive selection were observed in fewer genomic locations in Duffy-negative samples (Figure 4B). Significant signals were found in one unknown conserved protein on chromosome 13 (PvP01_1322000, p< 0·001), a plastid replication-repair enzyme on chromosome 13 (PVP01_1337500, p< 0·001), and the AP-5 complex subunit beta one (PVP01_1438600, p< 0.001) on chromosome 14.

Population structure and differentiation

The most probable number of genetic clusters identified by *K*-fold cross validation scores in the admixture analysis were three, four and five (Figure 5). In *K*=3 (error = 0.363), *Pv* in Duffy-positive Ethiopians clustered together with those in Duffy-negatives. These isolates shared the same cluster and were genetically similar with the Sudan and Uganda *Pv*. Though most Southeast Asian *Pv* were clustered together, a small

proportion of the isolates shared the same cluster with the African ones. The African and Southeast Asian *Pv* were genetically distinct from South American *Pv*. Similar clustering patterns were observed in *K*=4 (error = 0.354) and *K*=5 clusters (error = 0.356), of which samples were grouped into distinct geographical clusters by continent, consistent with previous studies.[24] More specifically, four major pools were identified, one in Southeast Asia, one in South America, and two in Africa (Figure 5). *Pv* in Southeast Asia were consistently distinct from South American *Pv*, whereas the East African *Pv* were further divided into two separate clusters, of which some *Pv* isolates in Ethiopia were different from those in Eritrea, Uganda, Sudan, and Madagascar. Nevertheless, *Pv* in Duffy negative Ethiopians were not clearly distinct from those in Duffy positives (Figure 5).

Discussion

Our study revealed distinct patterns of genetic variation, selection and population structure of *Pv* in Duffy positive and Duffy negative populations, providing new insights into the evolutionary dynamics of *Pv* in Africa. We identified a substantial number of high-quality SNPs, with a marked difference in prevalence between Duffy-positive and Duffy-negative individuals. Duffy-positive *Pv* exhibited greater genomic variability, with a higher number of variants, suggesting a more diverse parasite population in host reservoirs. Similar patterns of genetic diversity have also been observed in other parts of East Africa based on microsatellites, supporting the notion of genetically diverse *Pv* populations in these regions.[39,40] The high polymorphisms in Duffy-positive individuals, especially on chromosomes 9, 10, and 12, were in part due to the highly variable

PvMSP3 multigene family. *PvMSP3* expresses on the surface of merozoites and plays a role in evading the host's immune response and subsequent invasion of erythrocytes.[41] Such polymorphisms may reflect a complex interplay between host immunity and parasite adaptation. Higher genetic diversity in Duffy-positive hosts could be driven by strong immune selection, frequent recombination due to intense transmission, and gene flow across regions. These factors may promote genomic changes in the parasites. In contrast, Duffy negative individuals account for only about one-third of the general populations in Ethiopia and other East African countries and present a smaller host reservoir for *Pv*.[4] This may impose bottlenecks, limiting parasite evolution and transmission and resulting in low genome diversity.

Among the 43 erythrocyte binding genes, *PvDBP1*, *PvEBP/DBP2*, *PvRBP2b,* and some members of the *PvTRAg* family were among some of the most conserved in both Duffy groups. *PvDBP*1, *PvEBP/DBP*2, and *PvRBP*2b play significant roles in erythrocyte invasion, with region II of *PvEBP/DBP2* binds to the LHR-A subdomain of CR1 and residues 161-1454 of *PvRBP2b* to the transferrin receptor 1 (TfR1 or CD71) on the surface of immature erythrocytes upon entry.[11,16,17,42] Nevertheless, recent studies indicated that binding inhibitors blocking the TfR1 receptor showed no inhibition of *Pv* invasion in the Cambodian *Pv* isolates, suggesting that the functional significance of the PvRBP2b/TfR1 pathway may be varied by parasite strains.[43] The conservation of these genes suggests their importance in maintaining parasite functions for successful invasion of Duffy-negative erythrocytes. It has been hypothesized that unique mutations in *PvDBP*1 may offer the parasites invasion capability to Duffy negative

erythrocytes. While our earlier study revealed a few unique *PvDBP*1 mutations in two Duffy negative *Pv* samples, broader samples in this study did not indicate any unique *PvDBP*1 mutations between the two Duffy groups.[10] The low variability among codons in the key binding domain of *PvDBP1*, *PvEBP/DBP2* and *PvRBP2b* in Duffy-negative *Pv* may reflect more stringent selective pressures and restricted binding interactions with Duffy negative than Duffy positive erythrocytes. Indeed, recent studies showed that DARC is transiently expressed on erythroid precursor cells within the bone marrow of Duffy-negative individuals. This would imply that although Duffy negative erythrocytes circulating in the peripheral system may lack DARC expression, invasion may occur readily in bone marrow and/or other organs where DARC expressed in young reticulocytes during erythropoiesis.[44,45] PvDBP1-DARC may still be the primarily pathway for Duffy-negative erythrocyte invasion.

A number of *Pv* ligands and receptors have been recently shown to involve in erythrocyte invasion and it is possible that *Pv* from different geographical populations utilize different pathways for invasion. Our recent transcriptomic analyses showed that *PvDBP1* and *PvEBP/DBP2* were highly expressed in the Cambodian but not Brazilian or Ethiopian *Pv*, whereas *PvRBP2a* and *PvRBP2b* were highly expressed in Ethiopian and Cambodian isolates compared to Brazilian ones.[46] Genes in the *TRAg* (TRAg38, TRAg6, TRAg7, and TRAg14) and *RBP* multigene families (RBP2a, RBP1b, and RBP2b) are generally conserved in Duffy positive and Duffy negative *Pv* compared to the *MSP* gene family. *Pv*-infected individuals from Northern India have shown strong humoral immune responses against conserved regions of *PvTRAg*.[47,48] *PvTRAg*34 is

one of four *PvTRAg* genes that bind to erythrocytes and shares the same erythrocyte receptors with *PvTRAg*33.5.[49] Selection related to immune evasion could explain the higher genetic variation observed in *PvTRAg34* but not in other *PvTRAg* genes.

Selective pressures differ significantly between Duffy-positive and Duffy-negative hosts, shaping *Pv* genome diversity. In Duffy-negative *Pv*, three genes were detected with significant positive selection. The adaptor protein 5 (AP-5) complex subunit beta 1 (PVP01_1438600) plays a critical role in intracellular membrane trafficking by recognizing cargo and facilitating the formation of post-Golgi vesicles.[50] AP complexes consist of two large subunits for membrane binding and interaction with cargo adaptors, one medium subunit for recognizing transmembrane cargo motifs, and one small subunit for stability and binding dileucine-based motifs.[51] While AP-1 and AP-2 interact with clathrin to mediate trafficking between the trans-Golgi network (TGN) and endosomes or plasma membrane, the roles of AP-4 and AP-5 are less defined.[52] Another gene that encodes the plastid replication-repair enzyme on chromosome 13 was also shown with positive selection in Duffy-negative *Pv*. These genes may offer selection advantages for the parasites to facilitate the movement of cargo such as proteins and nutrients between organelles, which is essential for parasite survival and adaptation, as well as maintain cellular integrity and functionality while invading Duffy-negative red blood cells.[50,53] In Duffy-positive *Pv*, *PvCG*1 and *PvRPN*2 showed evidence of positive selection. Both genes are associated with parasite's drug resistance, with *PvCG*1 conferring resistance to chloroquine and *PvRPN*2 to artemisinin based on previous studies in *Pf*.[54,55] In *Pf,* mutation E738K in *PfRPN*2 was reported

with increased parasite survival when exposed to dihydroartemisinin.[55] Another gene *PvMAEBL*, which was detected under positive selection, is a minor erythrocyte binding protein expressed on the surface of the merozoites during the sporozoite phase.[56] This gene may offer selection advantages for sporozoite invasion of the salivary glands of the *Anopheles* mosquitoes.[56,57]

Despite the difference in the level of genome diversity, *Pv* in Duffy positive and Duffy negative individuals were not genetically distinct but clustered together, suggesting a shared ancestry and/or recent gene flow between Duffy populations. The clustering patterns observed at *K*=4 and 5, where some of the Ethiopian *Pv* (from both Duffy-positive and Duffy-negative *Pv*) showed genetic similarity with other East African isolates, further suggest recent parasite gene flow via human movement.[23,26] No significant difference was observed between Duffy positive and Duffy negative *Pv* , consistent with previous findings based on microsatellites.[39] Nevertheless, these isolates were distinct from Southeast Asia and South America, which aligns with prior studies showing distinct genetic structure among global *Pv* with three ancestral populations in Asia, one in Africa, and two in South America.[24] A recent study using microhaplotypes showed lower genetic diversity in *Pv* from the Horn of Africa than other geographical regions.[58] Genetic diversity was the highest in the Asian *Pv*, with genetic diversity decreasing as distance from Asia increases.[24,58] The admixture of genetic materials from geographically distinct regions may contribute to the parasite's ability to infect a broader range of hosts, including Duffy-negative individuals, and facilitate its spread across diverse geographical areas. The origins of *Pv* in humans are believed to

date back to the migration of *Pv* from Southeast Asia, with genetic evidence supporting the emergence of distinct clades in different regions. Recent genomic studies provided further insight into the evolution of *Pv*, revealing that *Pv* populations in different parts of Africa are highly interrelated despite varying rates of Duffy-negative individuals across regions.[4,39,58]

While our study offers valuable insights into the genomic features of Duffy-negative *Pv*, there were few limitations. First, Duffy-negative individuals from Ethiopia alone may limit our ability to generalize our findings. Expanded samples from other African countries would provide a more comprehensive view of genetic variations and adaptive strategies employed by *Pv* in different host populations. Second, our focus on SNPs without exploring other types of genetic variations, such as duplications, structural variants, and gene expression may undermine important elements that contribute to *Pv* adaptation in Duffy-negative hosts. Third, while we identified genes under recent positive selection in Duffy-negative Ethiopians, the functional implications of these genetic changes remain unclear. Experimental validation through *in vitro* studies or animal models would offer better understand of their roles in *Pv* infection. Fourth, this study focused solely on the genomic data of *Pv* without host genetic factors that could significantly influence infection dynamics and parasite adaptation. Future research integrating both host and parasite genomic data would provide a more holistic understanding of the host-parasite interactions driving *Pv* evolution.

To conclude, our findings highlight remarkably conserved *Pv* genomes in Duffy-negative individuals. Positive selection detected in AP5 and the plastid replication repair enzyme

on chromosome 13 in Duffy-negative *Pv* suggests that the parasites may have undergoing changes to optimize its survival and transmission potential in populations previously considered to be less susceptible. The lack of genetic structure suggests frequent gene flow between the two Duffy populations coexist in Ethiopia. Our study is the first to characterize genomic diversity of *Pv* in Duffy-negative and Duffy-positive populations. Further genomic surveillance across broad Africa will offer a deeper understanding of the molecular mechanisms underlying *Pv* infections in diverse host populations, which is crucial for developing effective control strategies against this widespread pathogen. Continued research is needed to explore the functional significance of the identified genetic variants, and its relevance to the changing epidemiology of *Pv* in Africa.

Acknowledgements. We thank the field team from Jimma University for their technical assistance; the communities and hospitals for their support and willingness to participate in this research; Drs. Joshua Mell and Azad Ahmed from the Drexel Genomics Core Facility and SKCC Share Genomics Resources for consultation and sequencing services; and undergraduate students at Drexel University for assistance with the experiments.

Financial support. This research was funded by National Institutes of Health R01 AI162947 and R01 AI173171.

Potential conflicts of interest. The authors declare no conflict of interest.

References

1. Albsheer MMA, Pestana K, Ahmed S, Elfaki M, Gamil E, Ahmed SM, et al. Distribution of Duffy Phenotypes among Plasmodium vivax Infections in Sudan. Genes (Basel). 2019;10(6).

2. Twohig KA, Pfeffer DA, Baird JK, Price RN, Zimmerman PA, Hay SI, et al. Growing evidence of Plasmodium vivax across malaria-endemic Africa. PLoS Negl Trop Dis. 2019;13(1):e0007140.

3. Motshoge T, Haiyambo DH, Ayanful-Torgby R, Aleksenko L, Ntebela D, Malleret B, et al. Recent Molecular Assessment of Plasmodium vivax and Plasmodium falciparum Asymptomatic Infections in Botswana. Am J Trop Med Hyg. 2021;104(6):2159-64.

4. Lo E, Russo G, Pestana K, Kepple D, Abagero BR, Dongho GBD, et al. Contrasting epidemiology and genetic variation of Plasmodium vivax infecting Duffy-negative individuals across Africa. Int J Infect Dis. 2021;108:63-71.

5. World Malaria Report 2022 [Available from: https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2022.

6. Miller LH, Mason SJ, Clyde DF, McGinniss MH. The resistance factor to Plasmodium vivax in blacks. The Duffy-blood- group genotype, FyFy. N Engl J Med. 1976;295(6):302-4.

7. Pestana K, Ford A, Rama R, Abagero B, Kepple D, Tomida J, et al. Copy Number Variations of Plasmodium vivax DBP1, EBP/DBP2, and RBP2b in Ethiopians Who Are Duffy Positive and Duffy Negative. J Infect Dis. 2024;230(4):1004-12.

8. Little E, Shenkutie TT, Negash MT, Abagero BR, Abebe A, Popovici J, et al. Prevalence and Characteristics of Plasmodium vivax Gametocytes in Duffy- Positive and Duffy-Negative Populations across Ethiopia. Am J Trop Med Hyg. 2024;110(6):1091-9.

9. Ford A, Kepple D, Abagero BR, Connors J, Pearson R, Auburn S, et al. Whole genome sequencing of Plasmodium vivax isolates reveals frequent sequence and structural polymorphisms in erythrocyte binding genes. PLoS Negl Trop Dis. 2020;14(10):e0008234.

10. Gunalan K, Lo E, Hostetler JB, Yewhalaw D, Mu J, Neafsey DE, et al. Role of Plasmodium vivax Duffy-binding protein 1 in invasion of Duffy- null Africans. Proc Natl Acad Sci U S A. 2016;113(22):6271-6.

11. Franca CT, He WQ, Gruszczyk J, Lim NT, Lin E, Kiniboro B, et al. Plasmodium vivax Reticulocyte Binding Proteins Are Key Targets of Naturally Acquired Immunity in Young Papua New Guinean Children. PLoS Negl Trop Dis. 2016;10(9):e0005014.

12. Gunalan K, Niangaly A, Thera MA, Doumbo OK, Miller LH. Plasmodium vivax Infections of Duffy-Negative Erythrocytes: Historically Undetected or a Recent Adaptation? Trends Parasitol. 2018;34(5):420-9.

13. Han JH, Lee SK, Wang B, Muh F, Nyunt MH, Na S, et al. Identification of a reticulocyte-specific binding domain of Plasmodium vivax reticulocyte-binding protein 1 that is homologous to the PfRh4 erythrocyte-binding domain. Sci Rep. 2016;6:26993.

14. Gupta ED, Anand G, Singh H, Chaddha K, Bharti PK, Singh N, et al. Naturally Acquired Human Antibodies Against Reticulocyte-Binding Domains of Plasmodium vivax

Proteins, PvRBP2c and PvRBP1a, Exhibit Binding- Inhibitory Activity. J Infect Dis. 2017;215(10):1558-68.

15. Chim-Ong A, Surit T, Chainarin S, Roobsoong W, Sattabongkot J, Cui L, Nguitragool W. The Blood Stage Antigen RBP2-P1 of Plasmodium vivax Binds Reticulocytes and Is a Target of Naturally Acquired Immunity. Infect Immun. 2020;88(4).

16. Gruszczyk J, Kanjee U, Chan LJ, Menant S, Malleret B, Lim NTY, et al. Transferrin receptor 1 is a reticulocyte-specific receptor for Plasmodium vivax. Science. 2018;359(6371):48-55.

17. Lee SK, Crosnier C, Valenzuela-Leon PC, Dizon BLP, Atkinson JP, Mu J, et al. Complement receptor 1 is the human erythrocyte receptor for Plasmodium vivax erythrocyte binding protein. Proc Natl Acad Sci U S A. 2024;121(5):e2316304121.

18. Alam MS, Zeeshan M, Mittra P, Choudhary V, Sharma YD. Receptor specific binding regions of Plasmodium vivax tryptophan rich antigens and parasite growth inhibition activity of PvTRAg35.2. Microbes Infect. 2016;18(9):550-8.

19. Alam MS, Choudhary V, Zeeshan M, Tyagi RK, Rathore S, Sharma YD. Interaction of Plasmodium vivax Tryptophan-rich Antigen PvTRAg38 with Band 3 on Human Erythrocyte Surface Facilitates Parasite Growth. J Biol Chem. 2015;290(33):20257-72.

20. Tyagi RK, Sharma YD. Erythrocyte Binding Activity Displayed by a Selective Group of Plasmodium vivax Tryptophan Rich Antigens Is Inhibited by Patients' Antibodies. PLoS One. 2012;7(12):e50754.

21. Rodriguez LE, Urquiza M, Ocampo M, Curtidor H, Suarez J, Garcia J, et al. Plasmodium vivax MSP-1 peptides have high specific binding activity to human reticulocytes. Vaccine. 2002;20(9-10):1331-9.

22. de Oliveira TC, Rodrigues PT, Menezes MJ, Goncalves-Lopes RM, Bastos MS, Lima NF, et al. Genome-wide diversity and differentiation in New World populations of the human malaria parasite Plasmodium vivax. PLoS Negl Trop Dis. 2017;11(7):e0005824.

23. Benavente ED, Manko E, Phelan J, Campos M, Nolder D, Fernandez D, et al. Distinctive genetic structure and selection patterns in Plasmodium vivax from South Asia and East Africa. Nat Commun. 2021;12(1):3160.

24. Daron J, Boissiere A, Boundenga L, Ngoubangoye B, Houze S, Arnathau C, et al. Population genomic evidence of Plasmodium vivax Southeast Asian origin. Sci Adv. 2021;7(18).

25. Ibrahim A, Manko E, Dombrowski JG, Campos M, Benavente ED, Nolder D, et al. Population-based genomic study of Plasmodium vivax malaria in seven Brazilian states and across South America. Lancet Reg Health Am. 2023;18:100420.

26. Gartner V, Redelings BD, Gaither C, Parr JB, Kalonji A, Phanzu F, et al. Genomic insights into Plasmodium vivax population structure and diversity in central Africa. Malar J. 2024;23(1):27.

27. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60.

28. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589-95.

29. Bahl A, Brunk B, Crabtree J, Fraunholz MJ, Gajria B, Grant GR, et al. PlasmoDB: the Plasmodium genome resource. A database integrating experimental and computational data. Nucleic Acids Res. 2003;31(1):212-5.

30. Auburn S, Bohme U, Steinbiss S, Trimarsanto H, Hostetler J, Sanders M, et al. A new Plasmodium vivax reference sequence with improved assembly of the subtelomeres reveals an abundance of pir genes. Wellcome Open Res. 2016;1:4.

31. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2).

32. Institute B. Picard Toolkit Github Repository: Broad Institute; 2019 [Available from: https://broadinstitute.github.io/picard/.

33. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43:11.0.1-.0.33.

34. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012;6(2):80-92.

35. Gautier M, Vitalis R. rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. Bioinformatics. 2012;28(8):1176-7.

36. Gautier M, Klassmann A, Vitalis R. rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. Mol Ecol Resour. 2017;17(1):78-90.

37. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19(9):1655-64.

38. Winter G, Kawai S, Haeggstrom M, Kaneko O, von Euler A, Kawazu S, et al. SURFIN is a polymorphic antigen expressed on Plasmodium falciparum merozoites and infected erythrocytes. J Exp Med. 2005;201(11):1853-63.

39. Kepple D, Hubbard A, Ali MM, Abargero BR, Lopez K, Pestana K, et al. Plasmodium vivax From Duffy-Negative and Duffy-Positive Individuals Share Similar Gene Pools in East Africa. J Infect Dis. 2021;224(8):1422-31.

40. Hoque MR, Elfaki MMA, Ahmed MA, Lee SK, Muh F, Ali Albsheer MM, et al. Diversity pattern of Duffy binding protein sequence among Duffy-negatives and Duffy-positives in Sudan. Malar J. 2018;17(1):297.

41. Alves KCS, Guimaraes JM, Almeida MEM, Mariuba LAM. Plasmodium falciparum merozoite surface protein 3 as a vaccine candidate: a brief review. Rev Inst Med Trop Sao Paulo. 2022;64:e23.

42. Bozdech Z, Mok S, Hu G, Imwong M, Jaidee A, Russell B, et al. The transcriptome of Plasmodium vivax reveals divergence and diversity of transcriptional regulation in malaria parasites. Proc Natl Acad Sci U S A. 2008;105(42):16290-5.

43. Feufack-Donfack LB, Baldor L, Roesch C, Tat B, Orban A, Seng D, et al. The PvRBP2b-TfR1 interaction is not essential for reticulocytes invasion by Plasmodium vivax isolates from Cambodia. NPJ Vaccines. 2024;9(1):232.

44. Dechavanne C, Dechavanne S, Bosch J, Metral S, Redinger KR, Watson QD, et al. Duffy antigen is expressed during erythropoiesis in Duffy-negative individuals. Cell Host Microbe. 2023;31(12):2093-106.e7.

45. Bouyssou I, El Hoss S, Doderer-Lang C, Schoenhals M, Rasoloharimanana LT, Vigan-Womas I, et al. Unveiling P. vivax invasion pathways in Duffy-negative individuals. Cell Host Microbe. 2023;31(12):2080-92.e5.

46. Kepple D, Ford CT, Williams J, Abagero B, Li S, Popovici J, et al. Comparative transcriptomics reveal differential gene expression among Plasmodium vivax geographical isolates and implications on erythrocyte invasion mechanisms. PLoS Negl Trop Dis. 2024;18(1):e0011926.

47. Zeeshan M, Bora H, Sharma YD. Presence of memory T cells and naturally acquired antibodies in Plasmodium vivax malaria-exposed individuals against a group of tryptophan-rich antigens with conserved sequences. J Infect Dis. 2013;207(1):175-85.

48. Alam MT, Bora H, Singh N, Sharma YD. High immunogenecity and erythrocyte-binding activity in the tryptophan- rich domain (TRD) of the 74-kDa Plasmodium vivax alanine-tryptophan-rich antigen (PvATRAg74). Vaccine. 2008;26(31):3787-94.

49. Zeeshan M, Tyagi RK, Tyagi K, Alam MS, Sharma YD. Host-parasite interaction: selective Pv-fam-a family proteins of Plasmodium vivax bind to a restricted number of human erythrocyte receptors. J Infect Dis. 2015;211(7):1111-20.

50. Hirst J, Barlow LD, Francisco GC, Sahlender DA, Seaman MN, Dacks JB, Robinson MS. The fifth adaptor protein complex. PLoS Biol. 2011;9(10):e1001170.

51. Geier C, Kreysing J, Boettcher H, Pohlmann R, von Figura K. Localization of lysosomal acid phosphatase mRNA in mouse tissues. J Histochem Cytochem. 1992;40(9):1275-82.

52. Park SY, Guo X. Adaptor protein complexes and intracellular transport. Biosci Rep. 2014;34(4).

53. Ferreira JL, Heincke D, Wichers JS, Liffner B, Wilson DW, Gilberger TW. The Dynamic Roles of the Inner Membrane Complex in the Multiple Stages of the Malaria Parasite. Front Cell Infect Microbiol. 2020;10:611801.

54. Su X, Kirkman LA, Fujioka H, Wellems TE. Complex polymorphisms in an approximately 330 kDa protein are linked to chloroquine-resistant P. falciparum in Southeast Asia and Africa. Cell. 1997;91(5):593-603.

55. Goncalves AF, Lima-Pinheiro A, Teixeira M, Cassiano GC, Cravo P, Ferreira PE. Mutation in the 26S proteasome regulatory subunit rpn2 gene in Plasmodium falciparum confers resistance to artemisinin. Front Cell Infect Microbiol. 2024;14:1342856.

56. Preiser P, Renia L, Singh N, Balu B, Jarra W, Voza T, et al. Antibodies against MAEBL ligand domains M1 and M2 inhibit sporozoite development in vitro. Infect Immun. 2004;72(6):3604-8.

57. Singh N, Preiser P, Renia L, Balu B, Barnwell J, Blair P, et al. Conservation and developmental control of alternative splicing in maebl among malaria parasites. J Mol Biol. 2004;343(3):589-99.

58. Siegel SV, Trimarsanto H, Amato R, Murie K, Taylor AR, Sutanto E, et al. Lineage-informative microhaplotypes for recurrence classification and spatio-temporal surveillance of Plasmodium vivax malaria parasites. Nat Commun. 2024;15(1):6757.

Figure Legends

Figure 1. Distribution of synonymous and nonsynonymous SNPS across all 14 chromosomes of the nuclear genome. Consistently higher proportions of synonymous SNPs to nonsynonymous SNPs with more mutations were detected in

Duffy positive than Duffy negative samples. Chromosomes 9, 10, and 12 have the higher portion of mutations across all chromosomes in both Duffy groups, with chromosomes 3 and 6 the least polymorphic.

Figure 2. Distribution of synonymous and nonsynonymous SNPs in 43 erythrocyte binding genes, as well as nucleotide diversity in each gene. Genes in the *PvMSP*3 multigene family had higher levels of SNPs and nucleotide diversity relative to other erythrocyte binding genes. *PvMSP*3G, *PvMSP*3.11, and *PvMSP*3.10 showed the highest levels of polymorphisms in both Duffy groups.

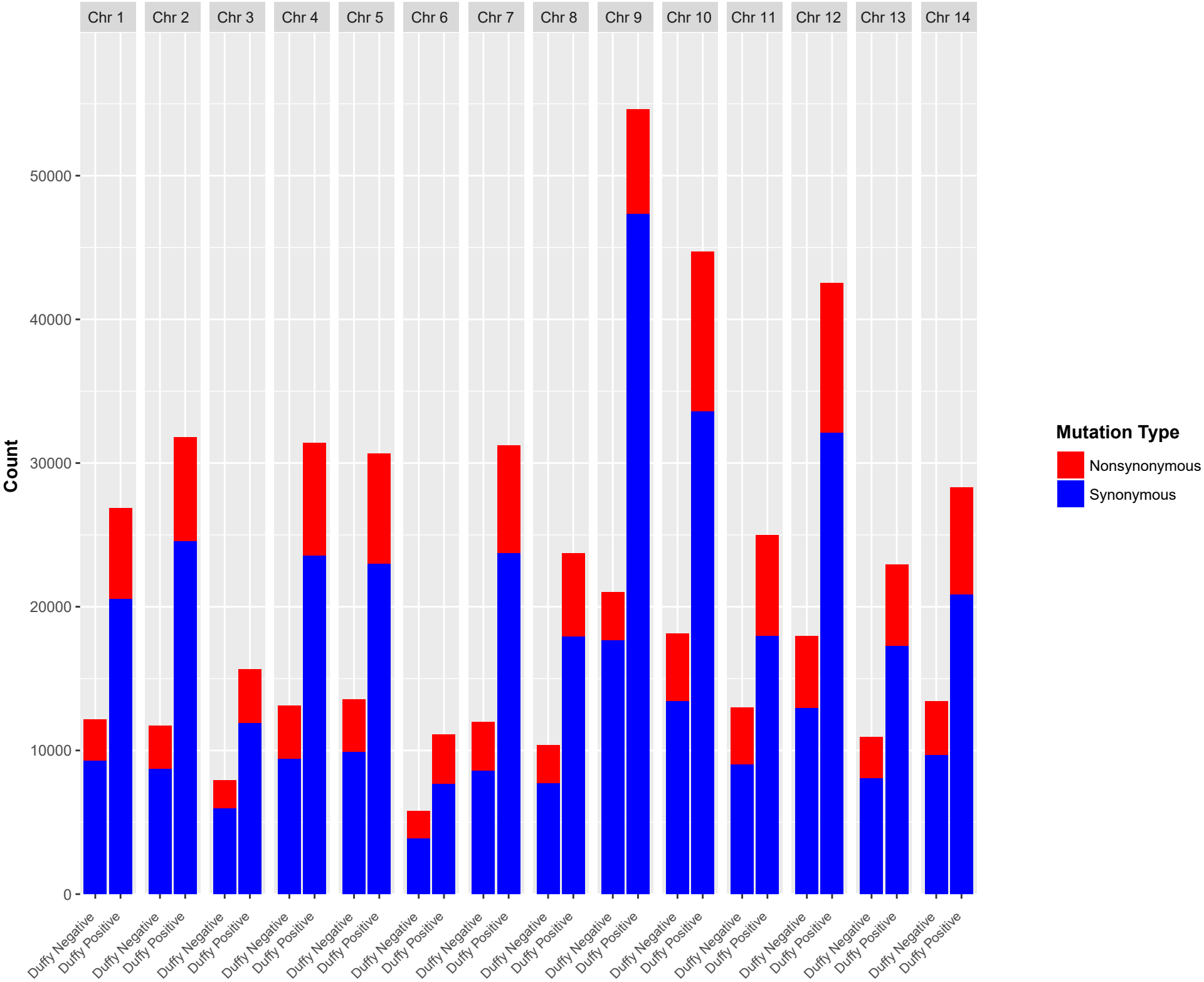
Figure 3. Amino acid variations across three erythrocyte binding genes including *PvDBP*1, *PvEBP/DBP*2, and *PvRBP*2b, all of which are known to involve in *Pv* invasion of Duffy-positive erythrocytes. Asterisks denote codons located in the binding regions. In *PvDBP*1, over 50% of Duffy positive samples were observed with mutation at four codon positions including G175E, S263R, K402S, and I458K (in red), all of which were in the binding region. In contrast, less than 10% of Duffy negative samples had these mutations at the same position. No mutations were detected in binding region of *PvEBP/DBP*2 nor *PvRBP*2b that were markedly different between the Duffy-positive and Duffy-negative samples.

Figure 4. Bar plots based on ADMIXTURE analyses of SNP profiles show clustering patterns of sample data collected from Africa, Southeast Asia and South America. results for *K*=3, 4 and 5 coupled with their respective cross

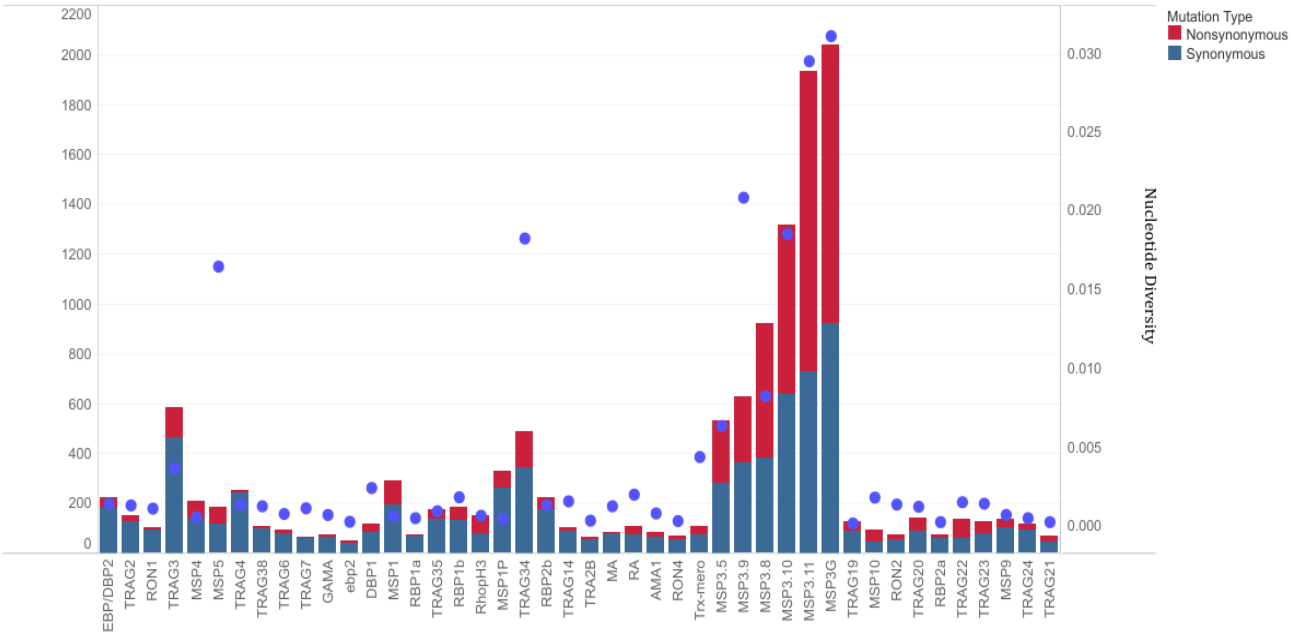
validation errors. Three genetic clusters were consistently observed for each *K*, one for each continental region. At *K*=5, high levels of admixture was seen among the African and South American *Pv*. No clear differentiation was detected between Duffy-positive and Duffy-negative samples.

Figure 5. **Signals of positive selection across the 14 chromosomes in both Duffy positive and negative Ethiopians.** Genes that showed significant signals of positive selection in Duffy positive samples included MAEBL, Cg1, STP1, SPM3 and RPN2, with genes Cg1 and RPN2 involved with drug resistant and MAEBL being a potential erythrocyte binding gene. Genes with significant positive selection in Duffy negative samples include a plastid replication repair enzyme and the AP-5 complex subunit beta one.

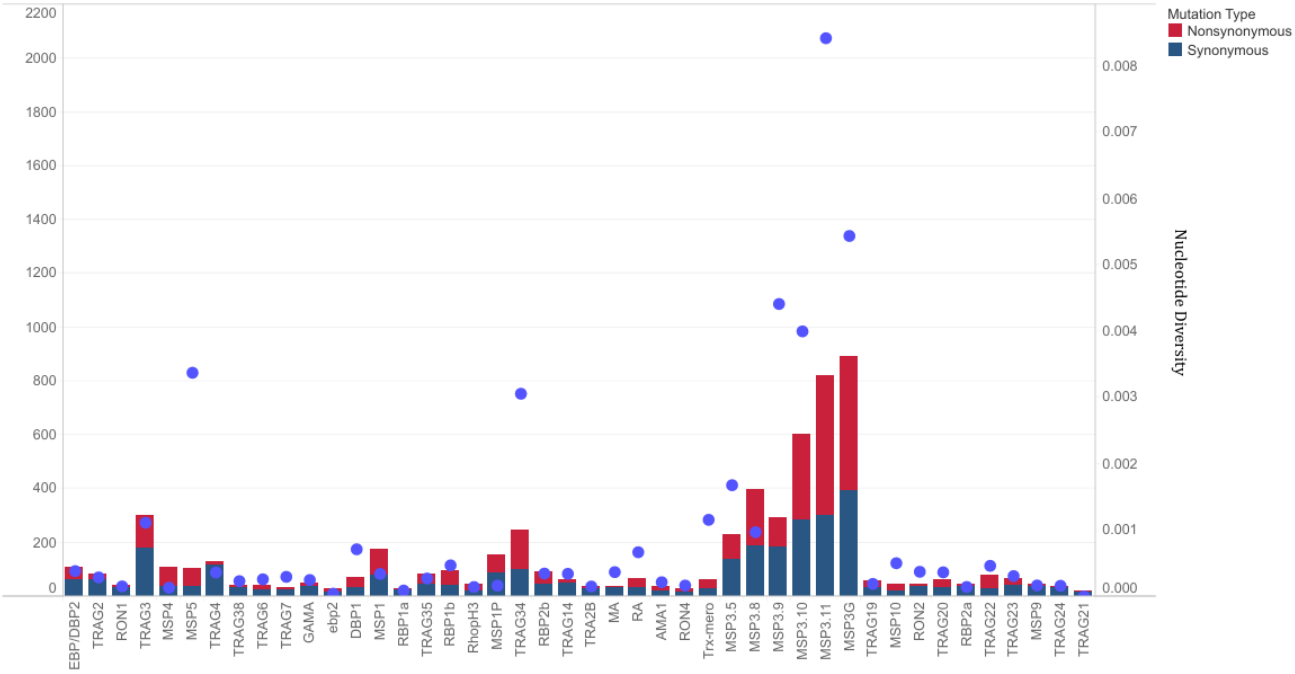
Supplementary Figures 1. **Amino acid variations of erythrocyte binding gene candidates including (A) *PvMSP*1, (B) *PvRBP*1a, (C) *PvRBP*1b, (D) *PvRBP*2a, and (E) *PvTRAg38.*** Asterisks denote codons located in the binding regions. Codons in red indicate mutations in the binding region that were found in over 50% of Duffy positive samples but less than 10% of Duffy negative samples at the same position. They included E1692K in *PvMSP1*, P552Q in *PvRBP1a*, and E438G and M511K in *PvRBP2a*.

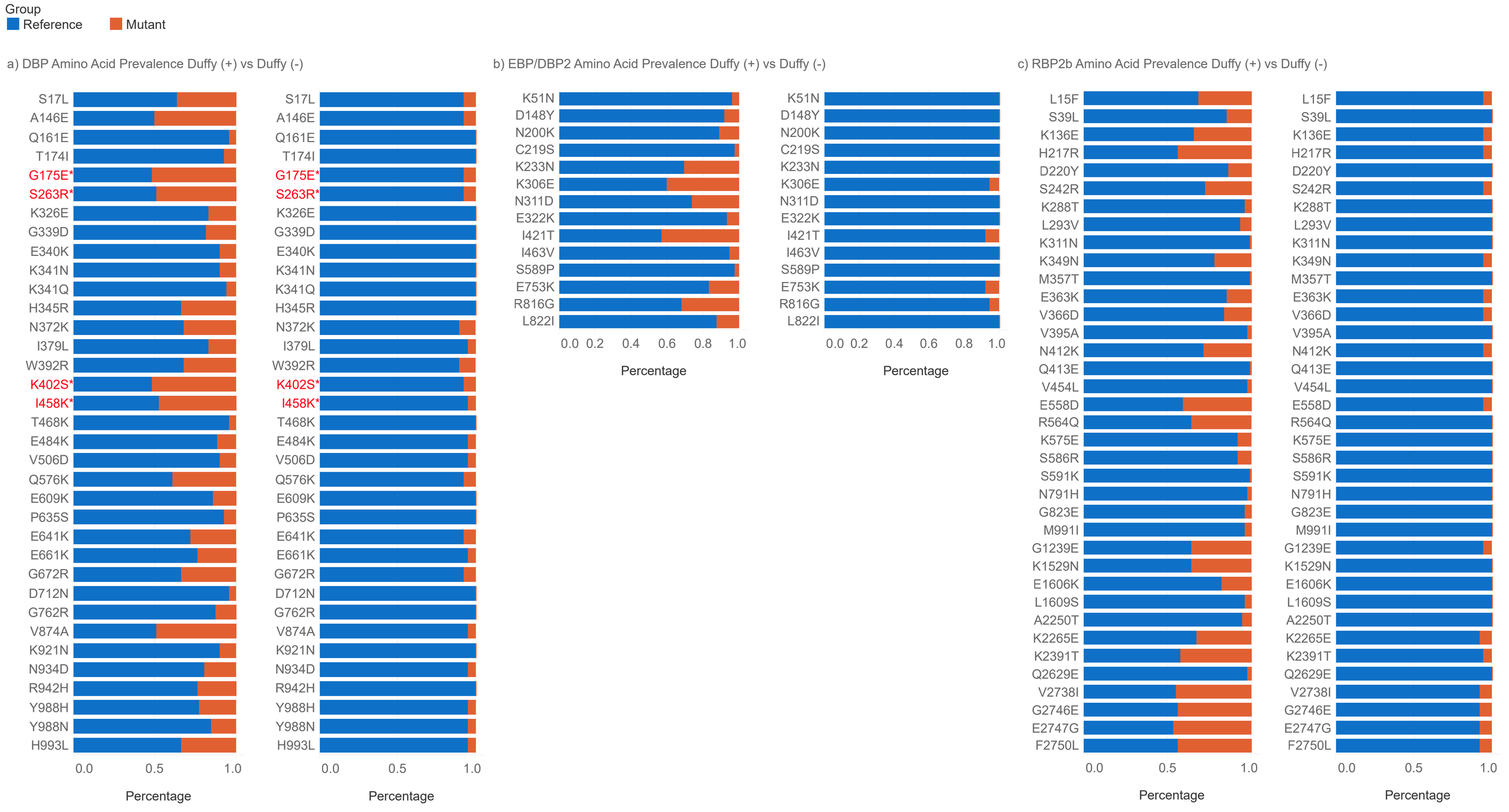

Chr 1
Chr 2
Chr 3
Chr 4
Chr 5
Chr 6
Chr 7
Chr 8
Chr 9
Chr 10
Chr 11
Chr 12
Chr 13
Chr 14
Count
0
10000
20000
30000
40000
50000
Duffy Negative
Duffy Positive
Mutation Type
Nonsynonymous
Synonymous

Duffy Positive Ethiopians Erythrocyte Binding Genes



Duffy Negative Ethiopians Erythrocyte Binding Genes




Group
Reference
Mutant
a) DBP Amino Acid Prevalence Duffy (+) vs Duffy (-)
S17L
A146E
Q161E
T174I
G175E*
S263R*
K326E
G339D
E340K
K341N
K341Q
H345R
N372K
I379L
W392R
K402S*
I458K*
T468K
E484K
V506D
Q576K
E609K
P635S
E641K
E661K
G672R
D712N
G762R
V874A
K921N
N934D
R942H
Y988H
Y988N
H993L
0.0
0.5
1.0
Percentage
b) EBP/DBP2 Amino Acid Prevalence Duffy (+) vs Duffy (-)
K51N
D148Y
N200K
C219S
K233N
K306E
N311D
E322K
I421T
I463V
S589P
E753K
R816G
L822I
0.0
0.2
0.4
0.6
0.8
1.0
Percentage
c) RBP2b Amino Acid Prevalence Duffy (+) vs Duffy (-)
L15F
S39L
K136E
H217R
D220Y
S242R
K288T
L293V
K311N
K349N
M357T
E363K
V366D
V395A
N412K
Q413E
V454L
E558D
R564Q
K575E
S586R
S591K
N791H
G823E
M991I
G1239E
K1529N
E1606K
L1609S
A2250T
K2265E
K2391T
Q2629E
V2738I
G2746E
E2747G
F2750L
0.0
0.5
1.0
Percentage

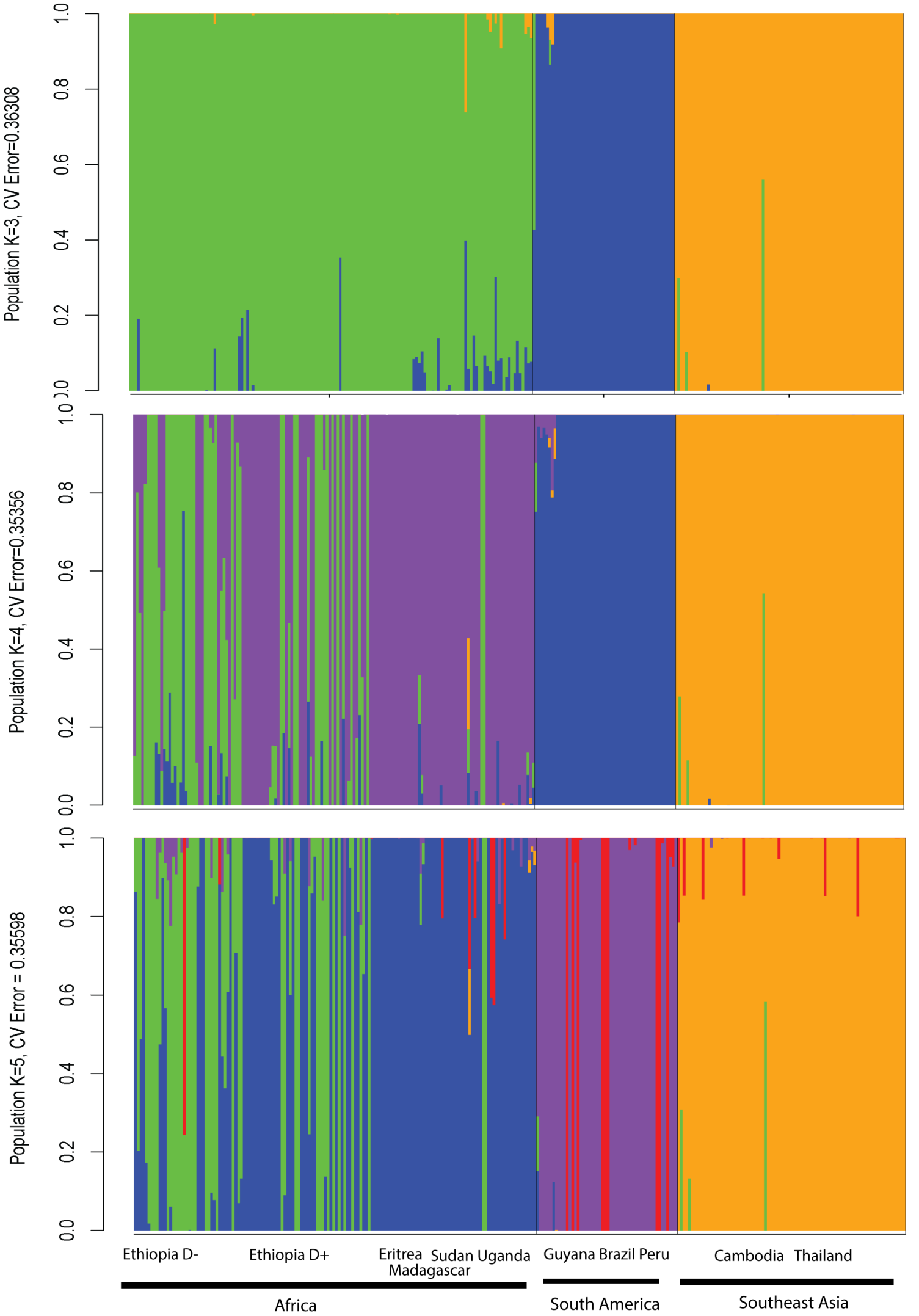
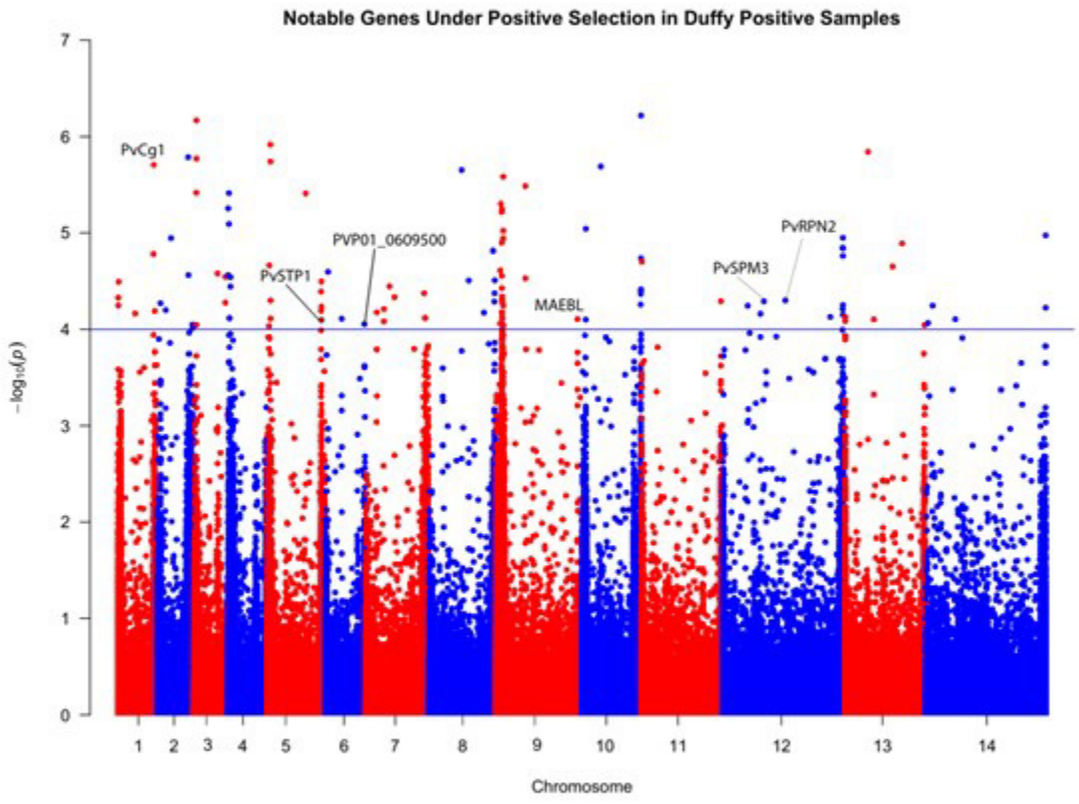
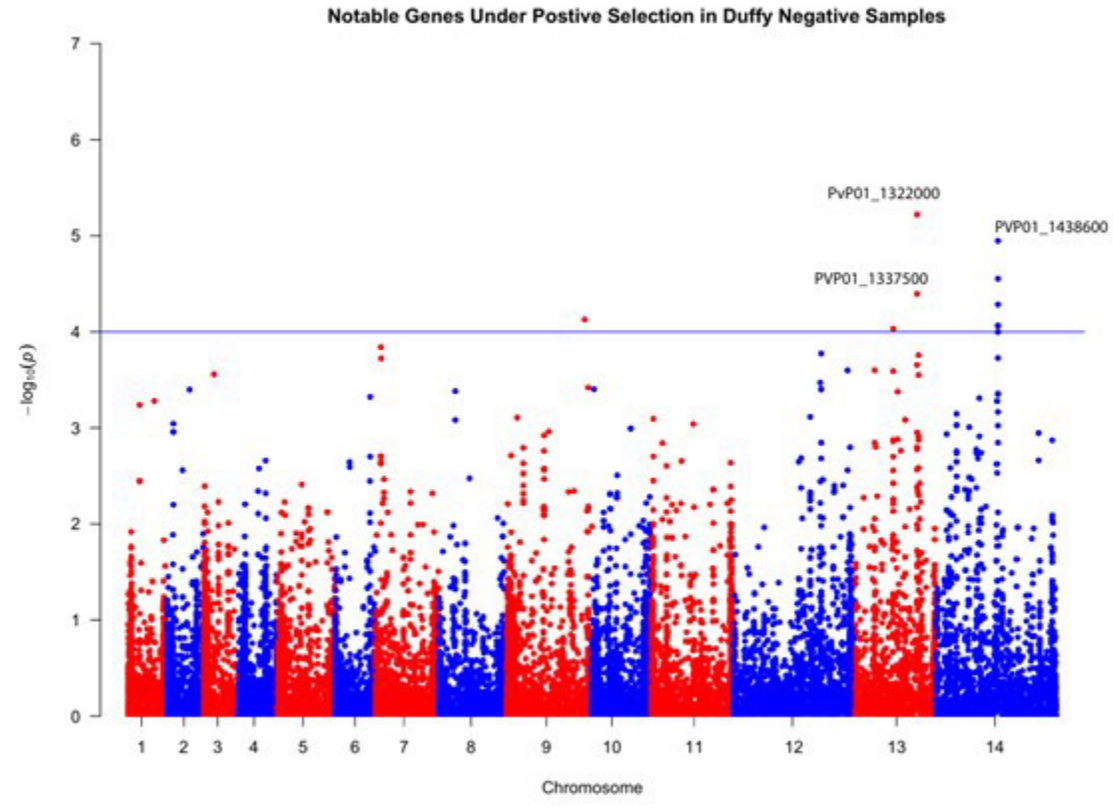

Population K=3, CV Error=0.36308
Population K=4, CV Error=0.35356
Population K=5, CV Error = 0.35598
0.0
0.2
0.4
0.6
0.8
1.0
Ethiopia D-
Ethiopia D+
Eritrea
Madagascar
Sudan
Uganda
Guyana
Brazil
Peru
Cambodia
Thailand
Africa
South America
Southeast Asia

Figure 5

a)


Notable Genes Under Positive Selection in Duffy Positive Samples
PvCg1
PvSTP1
PVP01_0609500
MAEBL
PvSPM3
PvRPN2
$-\log_{10}(p)$
Chromosome

b)


Notable Genes Under Postive Selection in Duffy Negative Samples
PvP01_1322000
PVP01_1438600
PVP01_1337500
$-\log_{10}(p)$
Chromosome