The 5q31 region in two African populations as a facet of natural selection by infectious diseases

ARTICLE *in* GENETIKA · FEBRUARY 2013

Impact Factor: 0.37 · DOI: 10.1134/S1022795413020051 · Source: PubMed

6 AUTHORS, INCLUDING:

Hiba S Mohamed
Taibah University
32 PUBLICATIONS **585** CITATIONS
SEE PROFILE

Muntaser Ibrahim
University of Khartoum
119 PUBLICATIONS **3,842** CITATIONS
SEE PROFILE

ГЕНЕТИКА, 2013, том 49, № 2, с. 1–10

ГЕНЕТИКА ЧЕЛОВЕКА

УДК 575.17:599.9

THE 5q31 REGION IN TWO AFRICAN POPULATIONS AS A FACET OF NATURAL SELECTION BY INFECTIOUS DISEASES

A. A. M. Elhassan[1], A. A. Hussein[1, 4], H. S. Mohamed[1], K. Rockett[2, 4], D. Kwiatkowski[2, 3, 4], A. M. Elhassan[1], and M. E. Ibrahim[1, 4]

[1]*Unit of Disease and Diversity, Department of Molecular Biology, Institute of Endemic Diseases, University of Khartoum, Khartoum, Sudan*

e-mail: abjil04@yahoo.com; mibrahim@iend.org

[2]*Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK*

[3]*Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK*

[4]*The MalariaGEN Consortium*

Received May 18, 2011; in final form, March 02, 2012

Cases of extreme natural selection could lead either to rapid fixation or extinction of alleles depending on the population structure and size. It may also manifest in excess of heterozygosity and the locus concerned will be displaying such drastic features of allele change. We suspect the 5q31 in chromosome 5 to mirror situation of such extreme natural selection particularly that the region encompasses genes of type 2 cytokine known to associate with a number of infectious and non-infectious diseases.

We typed two sets of single nucleotide polymorphisms (SNPS) in two populations: an initial limited set of only 4 SNPS within the genes of *IL-4*, *IL-13*, *IL-5* and *IL-9* in 108 unrelated individuals and a replicating set of 14 SNP in 924 individuals from the same populations with disregard to relatedness. The results suggest the 5q31 area to be under intense selective pressure as indicated by marked heterozygosity independent of Linkage Disequilibrium (LD); difference in heterozygosity, allele, and haplotype frequencies between generations and departure from Hardy–Weinberg expectations (DHWE). The study area is endemic for several infectious diseases including malaria and visceral leishmaniasis (VL). Malaria caused by *Plasmodium falciparum*, however, occurs mostly with mild clinical symptoms in all ages, which makes it unlikely to account for these indices. The strong selection signals seems to emanate from recent outbreaks of VL which affected both populations to varying extent.

DOI: 10.7868/S0016675813020057

Advancements in genome sciences are making possible not only the study of single variants associated with susceptibility to disease and responses to the environment but also the complex interaction of genes and their products underlying the evolution of such traits [1]. Infectious diseases are notorious for exerting profound evolutionary pressure on our genomes [2–5], resulting in specific signatures of natural selection i.e. directional selection and balancing selection that could be detected in natural populations through manifestations of different patterns of DNA polymorphism, thus allowing for inferences on past histories of selection and population structure [6, 7].

The populations inhabiting the endemic diseases area of the Rahad River in eastern Sudan are example of such populations because of their past and ongoing disease burden. We have been studying villages in that area over the past 15 years for the impact of infectious diseases namely malaria and visceral leishmaniasis (VL). Malaria has a limited disease burden, possibly due to the prevalence of the sickle-cell gene in the population [7]. On the other hand, VL or kala-azar resulted in high mortalities prior to the intervention of the leishmaniasis research group starting from the mid 1990s. The outbreak of VL in the eighties and nineties in southern Sudan resulted in probably some of the most severe mortality indexes in modern history with hundreds of thousands reported to have succumbed to the diseases [8, 9].

We have reported earlier interethnic differences to the disease among game wardens [10] and populations of endemic areas in eastern Sudan [11]. We have also identified in a separate study candidate genes associated with susceptibility to VL in a group with higher diseases incidence which is the Massalit tribe of Darfur who have recently migrated to an endemic area in eastern Sudan [12, 13].These genes include *SLC11A1* (formerly NRAMP) and *IL-4* which is encoded in the 5q31 region. The 5q31 region contains a myriad of genes that modulate atopic responses, including *IL-3*, *IL-4*, *IL-13*, *IL-5*, *CD14*, and granulocyte macrophage-colony-stimulating factor [14] has particular interest for association studies because these cytokine family genes are candidates also for a number of infec-

tious and non-infectious diseases [3, 15–25]. The region has been studied for susceptibility to malaria, severity of the disease and the level of parasitemia [15–18, 26–28] and is of interest to the science community because of the potential for trade-off mechanisms between infection and chronic diseases in these loci.

The population structure, stratification and ethnicity are thought to bear great implication for case control association studies particularly within the context of the African continent [29]. The population size and SNP density has also much to reveal particularly in situations where the population under study is of limited size or unique genetic structure. The trait of interest could be affecting a small number of individuals or differentially distributed between populations similar to what is seen in malaria and VL in the two ethnic groups of the endemic area of the Rahad River [7]. The 5q31 region is of particular relevance in displaying possible dynamics of natural selection in loci under burden of infectious diseases. We investigate such dynamics based on a limited set of polymorphisms and sample size and compare the outcome with a replicating set of polymorphisms in a larger set of sample from the same populations.

MATERIALS AND METHODS

Ethics Statement: The study has been approved by the Ethics Review Committee of the Institute of Endemic Diseases, University of Khartoum Samples were obtained, following written informed consent, according to the guidelines for genetic research on human subjects (www.iend.org).

Populations and study area: Um-Salala and Koka, are two villages located on the eastern bank of the Rahad River, a tributary of the Blue Nile; Koka is 35 km north west of Um-Salala. Koka village population is ~1800 individuals, of which 53% are females and the rest are males. The village was established nearly 50 years ago, by members of the Hausa tribe who migrated from northern Nigeria (mainly from the towns of Kanu and Sakatu); the majority of the present inhabitants can therefore be expected to have been born in the village. Um-Salala on the other hand was founded in 1969 by members of the Massalit tribe who migrated from western Sudan (Darfur State, near El-Geneina town). The migration to the Um-Salala village increased dramatically after the drought that hit Darfur in 1984. The village population is ~1450 individuals. Farming is the predominant economic practice and the population of both villages is fairly young consistent with age distribution in a typical African village where the young age groups dominate as follows: In Koka, 0–9 43.8% , 10–19 22.1%, 20–29 14.5%, 30–39 8.2% , 40–49 4.7%, and ≥50 6.7% respectively and in Salala village was respectively, 0–9 36.7, 10–19 23.8%, 20–29 12.2%, 30–39 14.9%, 40–49 5.1%, and ≥50 years 7.3%).

Genotyping. We typed the following SNPs by Amplification Refractory Mutation System (ARMS): *IL-4*: A/G, rs734244 (intronic polymorphism), *IL-5*: A/G, rs1800474 (silent mutation in exon three), *IL-9*: A/C, rs1799962 (intronic polymorphism) and *IL-13*: A/C, rs1881457, (intronic polymorphism), in a total of 109 unrelated individuals (57 from Hausa and 52 from Massalit) in an initial phase. An additional 56 individuals from the same population were subsequently typed for *IL-4* (rs734244) (Table 1).

Another set of genotype data from the same population was generated by the MalariaGen Consortium as part of Consortial Project 2 (http://www.malariagen.net/science/consortialproject2). This dataset comprised genotypes by the IPLEX technology for 82 SNPs from the 5q31 region in 924 individuals, 414 from Hausa and 510 from Massalit. Of those we selected 14 informative SNPs typed within the same genes as those by ARMS and were rendered for analysis.

Amplification Refractory Mutation System: DNA was extracted from buccal washes using an adapted chloroform-based method. Polymerase chain reaction (PCR) was carried out for DNA amplification, using the following primers: **rs734244**, *IL-4FN*, GCAAGCTAAGGAGCTCTGGA, *IL-4FM* GCAAGCTGAGGAGCTCTGGA, *IL-4R* TCTGGCCCATACATTTCTGA; **rs1800474**, *IL-5RN* TTGGCTATCAGCAGAGTTCG, *IL-5RM* TTGGCTATCAGCAGAGTCCG, *IL-5F* TGGGGACGCAGTCTTGTACT; **rs1799962**, *IL-9RN* GAACGCCCAGGTCTGTATTTTC, *IL-9RM* GAACGCCCAGGTCTGTATTGTC, *IL-9F* AGGTGGCAAACTTCATCCTG; **rs1881457**, *IL-13RN* TGACCCCTCTACGGGCCTGTT, *IL-13RM* TGACCCCGCTACGGGCCTGTT, *IL-13F* GATAAGGGGCGTTGACTCAC.

PCR reagents were added into a 200 μl sterile PCR tube. A final total volume of 25 μl containing 0.1–0.5 μg genomic DNA, 200 μM of each of the four dezoxynucleotides, 1.0 μM of each primer, and 0.5 U of *Taq* DNA polymerase in a buffer composed of 10 mM tris pH 8.3, 50 mM KCl, and 1.5 mM $MgCl_2$ (2.5 for *IL-5*) with 0.01% gelatin or 1.0 mM spermidine. The amplifications were performed in 41 cycles in thermal cycler (MJ Research). The ARMS/PCR used here was carried out in two separate tubes, one for the wild type and the other for the mutant. The contents of the two tubes were similar except for the primers. The PCR product was then loaded onto agarose gel and following electrophoresis, visualized under UV light in a Syngene gel-documentation system, the images were saved and printed out.

For purposes of ascertainment of ARMS results for rs1800474 and rs734244 the SNPs were retyped using RFLP for rs1800474 (primers: *IL-5F*: TGGGGACGCAGTCTTGTACT and *IL-5R*: CTGGTGATGTAGTTATCACC) and sequencing for rs734244.

Table 1. Genes and SNP designations (rs) of 4 SNPs typed by ARMS/PCR-RFLP and 14 SNPs by the IPLEX technology in the study population, the type and position of variation in the DNA sequence

SNP	Gene	Variation	Position
4 SNPs typed by RFLP/ARMS			
rs734244	*IL-4*	A/G	Intronic
rs1800474	*IL-5*	T/C	Exonic (syn)
rs1799962	*IL-9*	A/G	Intronic
rs1881457	*IL-13*	T/G	Promoter-1456
14 SNPs typed by IPLEX			
rs2243248	*IL-4*	G/T	Promoter -219
rs2243250	*IL-4*	C/T	Promoter -729
rs2243251	*IL-4*	A/G	Exonic (syn, leu)
rs4986964	*IL-4*	C/T	Exon (Arg/Cys)
rs2243270	*IL-4*	A/G	Intronic
rs2243283	*IL-4*	C/G	Intronic
rs739718	*IL-5*	A/G	NA
rs2069818	*IL-5*	A/C	Exonic (Syn, Thr)
rs2069823	*IL-5*	A/G	Intronic
rs1881457	*IL-13*	A/C	Promoter -1456
rs2069744	*IL-13*	C/T	Intronic
rs20541	*IL-13*	C/T	Exonic (Arg/Glu)
rs848	*IL-13*	G/T	3UTR
rs2243206	*IL13*	C/T	NA

Data analysis. The 4 SNPs primarily analyzed were selected from the web site: (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi) based on their minor allele frequency and location within the gene. The NCBI web site was also used to identify the suitable enzyme (*Taq*1) for the RFLP of the *IL-5* SNP detection. The program was NEBcutter V2.0 (http://tools.neb.com/NEBcutter2/index.php).

The program Genetix 404 (http://www.genetix.univ-montp2.fr/genetix/genetix.htm) was used to calculate allele frequencies and heterozygosity. Arlequin 3.1 (http://cmpg.unibe.ch/software/arlequin/) was used to perform the neutrality tests, haplotype frequencies, and to measure population differences. LD blocks were obtained using HaploView program. Selection coefficient was calculated manually using the formula $p' = \{(1 + 2s)p^2 + (0.5)(1 + s)2pq\}/(1 + 2ps)\text{-}p'$, where p' is the allele frequency of the major allele in the 2nd generation, p is the frequency of the major allele in the first generation, q is the frequency of the minor allele in the first generation and s is the selection coefficient. To ensure independence of genotypes, a sample of unrelated individuals from these village, we selected the parents in each household. The range of age of sampled individuals allowed dividing them into two age groups of less and more than thirty years of age, and referred to as generations/age groups; as the generation time in these communities could range between 15–25 years, these age groups may correspond to one or two generations, not considering the generational overlap. Microsoft Excel was used to draw the figures. In order to verify this pattern identified in small sample set, we looked at heterozygosity values of extra 5 SNPs in *IL-4* in larger sample set of 321 individuals from Hausa and 510 individuals from Massalit over 3 generations/Age groups <15 years, 15–30, >30.

Chi-squared, *t* Test, *F* statistics and their *P* values were calculated using the Online Statistic Calculator Vassarstat (http://faculty.vassar.edu) and the Program R.

RESULTS

Allele and genotype frequency differences as depicted by ARMS assays: values for departure from Hardy–Weinberg expectations (DHWE) are shown in table 2 as well as values of *F* statistics for the collective sample in Hausa and Massalit at each locus. The *P* values for DHWE is calculated for the age groups/generations when sample was divided into those above or less than 30-years old typed by ARMS. The *F* statistics values are consistent with excess of heterozygosity and the considerable departure from HWE in all SNPs in the studied population except for *IL-9* in the less than 30-years old groups in both Hausa and Massalit. This is apparently due to the sample bias of sampling par-

Table 2. *P* values of the departure from HWE in Hausa and Massalit in the ≥30 under <30 years age groups

Population	*IL-4*	*IL-5*	*IL-9*	*IL-13*
Hausa (>30, $n = 31$)	0.0004	0.0016	0.2467	0.0743
F	−0.75	−0.56	−0.316	−0.39
Hausa (<30, $n = 28$)	0.0042	0.0121	0.1797	0.032
Massalit (>30, $n = 20$)	0.00155	0.0155	0.0599	0.012
F	−0.65	−0.57	−0.405	−0.529
Massalit (<30, $n = 32$)	0.0232	0.0024	0.1086	0.0108

Note: *F* statistic is calculated for the collective village sample at each locus.

ents, which comes as a cost of seeking genotype independence.

The mean heterozygosity in Hausa and Massalit were 0.657 and 0.687 respectively, there was no significant difference in allele frequency between the two populations (fig. 1) shows the genotype frequency for the *IL-4* rs734244 SNP as compared to other populations using data from the HapMap. There was marked difference from values reported in non-African populations while they more closely resembled the Yoruba (YRI).

After stratification of the populations into two age groups/generations, those less than 30-years of age as compared to thirty and above, the allele frequencies were found to be different, but not overall significant significant.

The greatest heterozygosity difference in the age-groups was found for rs1800474 in *IL-5* (Hausa: <30-years 0.73 to ≥30-years 0.59 and Massalit: <30-years 0.69 to ≥30-years 0.58). The mean heterozygosity in Hausa for those over and below 30 years, were 0.58 and 0.71 respectively ($P = 0.118$), and in Massalit were 0.69 and 0.70 ($P = 0.410$). Heterozygosity decreased with age in all loci in Hausa but in Massalit there was a decrease in *IL-9*, increase in *IL-4* and *IL-5* frequencies, while *IL-13* remained unaltered. The decrease in Hausa in *IL-5* was significant, (0.83 for the overall sample and 0.54 in the under 30yrs with *P* value = 0.02) and the increase in Massalit in *IL-5* was also significant (0.64 in the older and 0.84 in the younger age group, $P = 0.026$) (Fig. *a*, *b*).

T test was calculated for observed and expected heterozygous in Hausa and Massalit. In Hausa *T* test was 3.55 and *P* value was 0.038 and it was 7.94 and *P* value was 0.0042 in Massalit.

Temporal Change in allele frequency, heterozygosity and selection measure in the IL-4 locus. The temporal variation (inter generation or between age-groups) in allele frequency was calculated and taken as a measure of natural selection at the given locus. This also permitted having an impression of the trajectory of alleles over time. The *IL-4* locus was singled out because of its remarkable increased heterozygosity and reported association with VL (Mohamed et al. [13]). Further 56 individuals were genotyped for the single nucleotide polymorphism (rs734244), and analyzed along a larger sample of 5 other SNPS across the *IL-4* (rs2243248, rs2243250, rs2243251, rs2243270, rs2243283) from the MalariaGEN consortium genotype data resource (Fig. 3).

Change in allele frequency in both Hausa and Massalit shows that following an initial divergence in the age groups between 30 and 45 and >45 in 164 individuals (coinciding with the peak of VL outbreak for the susceptible age groups), a restoration of equilibrium was reached accompanied by decline in heterozygosity in both populations in the younger age groups: In order to verify this pattern identified in small sample set, we looked at heterozygosity values of extra 5 SNPs in *IL-4* in larger sample set of 321 individuals from Hausa and 510 individuals from Massalit over 3 generations/Age groups, and the outcome was consistent with a general decline of heterozygosity over time, indicating that older generations had been subjected to intense selective pressure favoring heterozygote forms. Figure 3 shows this pattern in Hausa, where the selection coefficient (*S*) was calculated for these SNPs based on an additive model and the results ranged for an *S* of −0.79 to −0.97 (shown on top of each line) indicating the presence of a strong selection signal along these loci. We excluded the recessive and dominant models and opted for an additive model based on the fact that the mode of selection of the minor allele is still unknown to us.

Haplotypes of Hausa and Massalit. Difference in haplotype frequencies in the two populations was calculated based on the three loci for *IL-4*, *IL-5* and *IL-13* since *IL-9* was beyond the accepted range of linkage. The results in Table 3 show differences in the haplotype frequencies in the two populations. The detrimental locus seems to be the *IL-5*, in agreement with the allele frequency and heterozygosity data above.

Linkage disequilibrium was calculated using the Genetix 404 program for *IL-4*, *IL-5* and *IL-13*. The *D'* values were 0.6228, 0.6415 and 0.7453 respectively in Hausa, and were 0.4859, 0.6141 and 0.4764 in Massalit. The largest difference in LD was between *IL-5* and *IL-13*, 0.7453 in Hausa and 0.4764 in Massalit

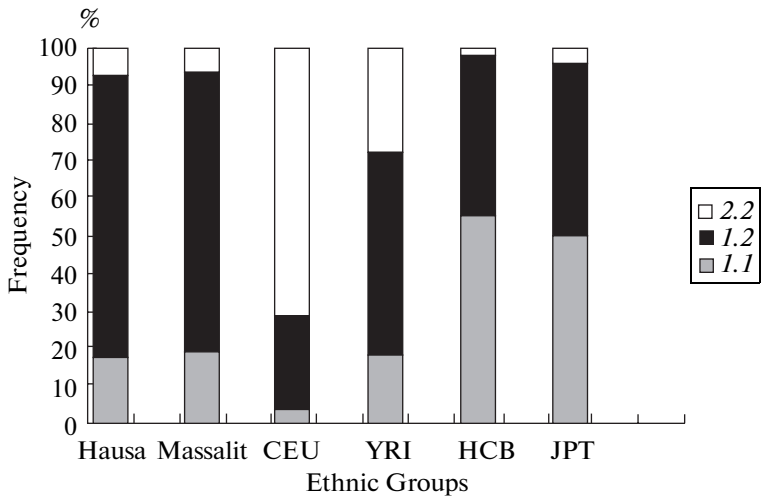

Fig. 1. Genotype frequencies in different Ethnic Groups of the *IL-4* rs734244. CEU: Europeans, YRI: Yoruba, HBC: Chinese, JPT: Japanese. *2.2*: homozygous minor allele, *1.2*: heterozygous, *1.1*: homozygous major allele.

(Table 4). These results were also compared to LD patterns based on 14 SNPs in the 5q31 in a larger MalariaGen data set (414 Hausa and 510 Massalit, Fig. 4,*a* and *b*). The results were comparable to values of LD based on the limited number of SNPs tested in our analysis above, although Massalit showed relatively higher LD and longer blocks than the Hausa. None of the SNPs used in the primary analysis however was among the MalariaGen SNP panel.

DISCUSSION

Our primary observation of differences in diseases profile between the two populations of the Rahad River in susceptibility to visceral leishmaniasis and malaria was not reflected in the allele frequencies of the limited set of the 5q31 SNPs investigated initially or in heterozygosity values. Nearly all SNPs compared were in excess of heterozygosity. Excess of heterozygosity which indicates some degree of out-breeding is supported by the results of the *F* statistic. However this occurred despite the fact that >90% of the marriages in Koka was within village in the past decades [7]. In these circumstances, one has to guard against the locus specific effect, given that the history of each could depict different evolutionary background. We speculate that the relatively large effective size (N_e) of the village founders may have influenced the impact of random mating over generations. Because of this, the potential of sampling effect (Drift) does not seem to affect the allele frequency drastically; mtDNA and Y chromosome data indicate that Hausa effective population size is large enough [21], to avoid an effect of endogamy on HWE in the few generations since the foundation of the village.

In comparison with European and even HapMap data of African populations, heterozygosity in Hausa

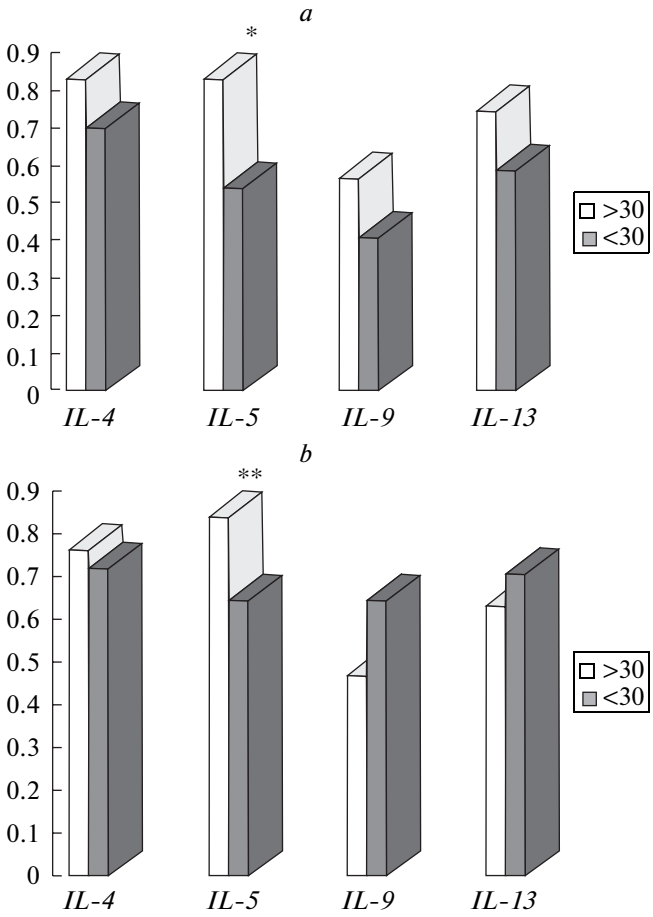

Fig. 2. Heterozygosity in two age groups (<30 years or ≥30 of age) using four SNPs in the genes of *IL-4*, *IL-5*, *IL-13* and *IL-9* in Hausa (*a*) and Massalit (*b*). Significant differences * in Hausa *P* value = 0.02 and ** in Massalit *P* value = = 0.026. See Materials and Methods for SNP designations.

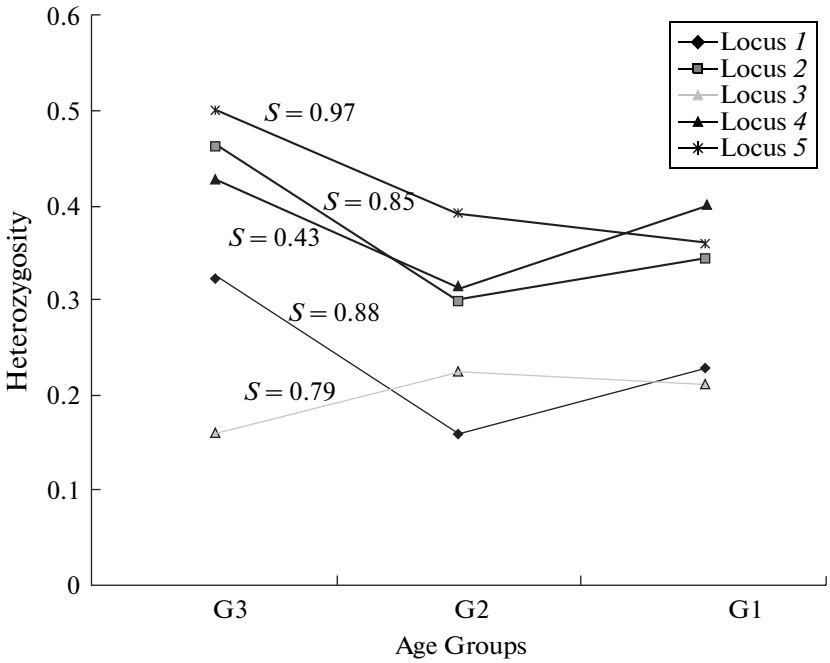

Fig. 3. Heterozygosity in five SNPs in *IL-4* in Hausa ($N = 321$) within three age groups. G1: 1–15-years old, G2: 16–30-years old, G3: over 30-years old. Selection coefficient (S) in the same loci indicated on top of each line. S is calculated based on maximum intergenerational differences.

and Massalit was generally far much higher for the same SNPs. This is consistent with findings that Heterozygosity in Africans is significantly greater than Asian and European populations [30]. A study on the *IL-4*, *IL-13* loci in American, Japanese, African and European populations employing an informative set of 31 SNPS [31]. Have shown that haplotype and LD patterns are significantly different between populations and between close loci with the *IL-4* showing signatures of natural selection that are different from the *IL-13* and in agreement with the proposition that genes that are selectively significant require to be decoupled for optimal function something reflected in LD patterns [32].

We believe that the extreme heterozygosity values and the DHWE are partly due to a sampling bias where we have selected parents as unrelated individuals for purposes of independence of genotypes in this study. This may have inflated the already present measures of selection (heterozygosity and DHWE). Inclusion of genotypes of younger age groups as shown in the *IL-4* locus have manifested in a steady trend towards equilibrium, but not enough to diminish the high already present signal of natural selection. This signal seems to be authentic, as it was replicated in the MalariaGen larger set of data and manifested as differences in heterozygosity between age groups/generations, trends towards increased heterozygosity with age and a significantly high selective coefficient based on an additive model.

Table 3. Frequencies of the three-SNPs (*IL-4*, *IL-5* and *IL-13* respectively) haplotypes in Hausa and Massalit

Haplotype	Hausa	Massalit
AAA	0.241	0.175
ABA	0.045	0.355
BAA	0.128	0.122
BAB	0.000	0.271
BBA	0.297	0.000
AAB	0.287	0.074

Differences in allele frequencies and heterozygosity values between age groups (when existing), are the ultimate means to elucidate the impact of natural selection, particularly in its extreme forms. In contrast to indirect methods based on violating a null hypothesis of neutrality, this approach could be applied to a limited number of SNPs/loci as shown in this study. Such differences were also pronounced in the *IL-5* locus in which the heterozygosity increased in Massalit but decreased in Hausa within few generations. It has been shown that for loci under natural selection heterozygosity could reach extremely high values as in the case of the lactase gene [33]. Interestingly this particular *IL-5* SNPs has shown to be monomorphic in all populations in the HapMap and heterozygosity data was therefore unavailable.

The conditions for a correlation between genomic heterozygosity and fitness in populations departing

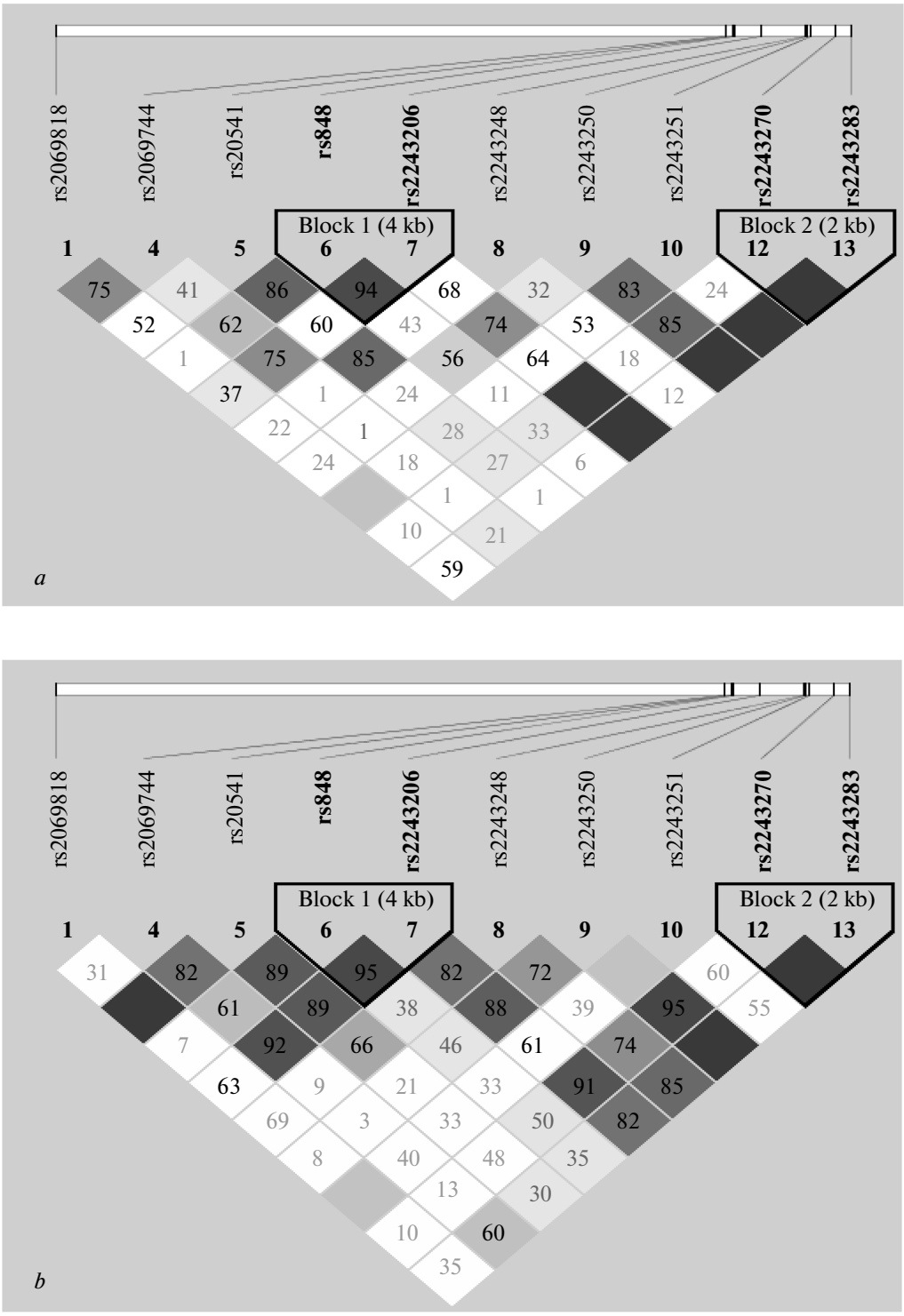

Fig. 4. LD blocks based on 13 SNPs in the 5q31 region in Hausa (*a*) and Massalit (*b*).

from Hardy–Weinberg Equilibrium, to arise are discussed by Deng and Fu [34]. In the absence of solid data on natural selection in human populations, there is a dire need to model the dynamics of such heterozygosity under situations where heterozygosity report with values higher than expected from HWE.

Mortality records are sparse in both villages, but what was available from oral autopsies indicate that

Table 4. Measures of natural selection and populations' structure in Hausa and Massalit, θ values based on the 4 SNPs typed by ARMS/RFLP and Tajema's *D* based on 4 and 14 SNPs across the 5q31.

Populations	θ (S)	θ (hom)	θ (phi)	θ (K)	Tajima *D*	
					4 SNPs	14 SNPS
Hausa	0.375	4.208	0.904	9.505	2.1	5.98
Massalit	0.766	2.013	1.816	4.136	2.6	5.09

the sampled individuals are mostly from families that had suffered from the impact of VL and an ensuing high death rate in these villages. The Leishmaniasis Research Group (LRG) began its intervention in the Rahad area in 1990. Prior to that, children and young adults were at higher risk of mortality because of the lack of treatment.

A selection signal in the *IL-5* region could well be associated with VL mortality. This is suggested by haplotype differences between Hausa and Massalit, since LD patterns excludes a SNP distal to the *IL-5*, and Massalit are the population that suffered most from VL. Mohamed et al. [12] have already implicated the 5q31 region in susceptibility to VL, particularly *IL-4* and *IL-9*. *D*' values show lack of LD consistent not only with African demographic history of lower LD [35–37], but also of a region under natural selection with several genes operating which requires decoupling and independence of gene function as suggested for the β globin locus [32].

Signals of natural selection are often confused with those of population substructure and drift [6]. However, under substructure, allele frequencies should remain constant over generations in the absence of drift or migration. The intergeneration difference and the fluctuation in heterozygosity and the alleles trajectory of the *IL-4* SNP encountered in this study is more in line with natural selection, to which we ascribe to fatalities among children due to VL a common diseases in the Rahad River with a well documented death toll [10, 38]. The estimated time for effective intervention to have started in the Rahad region is around 18 years ago when the LRG identified the spot and started treating patients. The time frame makes the age group in the range of 15–30 to fall within the vulnerable age for susceptibility to the disease. Figure 3 (*a* and *b*) interestingly show a pattern of initial heterozygosity in both population, that widens in time, but then falls to near equilibrium in the younger generations possibly following LRG intervention. The older age groups which include people born in Western Sudan Darfur State, as in the case of Salala village are believed to have been exposed to cutaneous leishmaniasis in their native home, and thus became relatively immune as they convert in their leishmanin [37]. The fact that we have originally sampled from the >30 age group may explain the values of heterozygosity and the DHWE obtained. The elevated heterozygosity remains to be independent from sampling as it declined in both populations over time, thus reflecting also the impact of random mating in the absence of selection. *IL-4* (rs734244) minor allele which seems to be negatively selected has been restored in both populations, even exceeding the frequency of the major allele which conforms to the genotypes ratio in the Yoruba as reported in the HapMap. The selection signatures along this gene, extends to other SNPs with strong values of (*S*) making a strong case for an extreme case of natural selection.

In conclusion, two independent sets of SNPs and sample sizes resulted in similar outcomes for measures of genetic structure of the 5q31 region in the two populations. Although natural selection is an obvious candidate for the allele change in this study, the true culprit underlying its dynamics is not entirely clear. The effect of the new environment of the Rahad River on the two populations, with its impact of endemic diseases may be the driving force for such an impact. Among the diseases affecting the studied populations, visceral leishmaniasis seems as the strongest candidate for acting as a potent selective force.

ACKNOWLEDGMENTS

The authors wish to acknowledge the cooperation of the populations of the villages of Koka and Salala. This work is part of the activities of the EviMalar European Network of Excellence supported by a 7th Framework Programme. We would like to acknowledge also support from MalariaGen consortium in particular Anna Jeffreys, Kate Rowlands, Rachel Craik, Angela Green and Christina Hubbert for processing and genotyping the DNA samples.

REFERENCES

1. Lipsitch, M.A., Sousa, O., Historical intensity of natural selection for resistance to tuberculosis, *Genetics*, 2002, vol. 161, pp. 1599–1607.
2. Hill, A.V., The immunogenetics of human infectious diseases, Annu. Rev., *Immunol.*, 1998, vol. 16, pp. 593–617.
3. Kwiatkowski, D., Genetic susceptibility to malaria getting complex, *Curr. Opin. Genet. Dev.*, 2000, vol. 10, pp. 320–324.
4. Lengeling, A., Pfeffer, K., Balling, R., The battle of two genomes: genetics of bacterial host/pathogen interac-

tions in mice, *Mamm. Genome*, 2001, vol. 12(4), pp. 261–271.

5. Wertheim, K., Kutkowska-Kaźmierczak, A., Bal, J., Genetically determined human susceptibility to selected infectious diseases, *Med. Wieku Rozwoj*, 2008, vol. 12(3), pp. 738–747.
6. Harris, E.E., Meyer, D., The molecular signature of selection underlying human adaptations, *Am. J. Phys. Anthropol.*, 2006, vol. 43, pp. 89–130.
7. Salih, N.A., Almugtaba, I., Elzein, M.E., et al., Loss of balancing selection in the *S*, globin locus in a nonpanmictic African population, *BMC Med. Genetics*, 2010, vol. 3, pp. 11–21.
8. Perea, W.A., Moren, A., Ancelle, T., Sondorp, E., Epidemic visceral leishmaniasis in southern Sudan, *Lancet*, 1989, pp. 1222–1223.
9. Seaman, J., Mercer, A.J., Sondorp, E., The epidemic of visceral leishmaniasis in western Upper Nile, southern Sudan: course and impact from 1984 to 1994, *Int. J. Epidemiol.*, vol. 25, pp. 862–871.
10. Ibrahim, M.E., Lambson, B., Yousif, A.O., et al., Kala-azar in a high transmission focus: an ethnic and geographic dimension, *Am. J. Trop. Med. Hyg.*, 1999, vol. 61(6), pp. 941–944.
11. El Hassan, A.M., Ziljstra, E.E., Ismael, A., Ghallb, H.W., Recent observations on the epidemiology of kala-azar in the eastern and central states of the Sudan, *Trop. Geogr. Med.*, 1995, vol. 47, pp. 151–156.
12. Mohamed, H., Ibrahim, M.E., Miller, E.N., et al., SLC11A1 (formerly NRAMP1) and susceptibility to visceral leishmaniasis in the Sudan, *Europ. J. Human Genetics*, 2004, vol. 12, pp. 63–74.
13. Mohamed, H.S., Ibrahim, M.E., Miller, E.N., et al., Susceptibility to visceral leishmaniasis in the Sudan: Linkage and association with *IL-4* and *IFNGR1. Genes and Immunity*, 2003, vol. 4, pp. 351–355.
14. Frazer, K.A., Ueda, Y., Zhu, Y., et al., Computational and biological analysis of 680 kb of DNA sequence from the human 5q31 cytokine gene cluster region, *Genome Res.*, 1997, vol. 7, pp. 495–512.
15. Rihet, P., Traore, Y., Abel, L., et al., Malaria in humans: Plasmodium falciparum blood infection levels are linked to chromosome 5q31-q33, *Am. J. Hum. Genet.*, 1998, vol. 63, pp. 498–505.
16. Ohashi, J., Naka, I., Patarapotikul, J., et al., A single-nucleotide substitution from C to T at position –1055 in the *IL-13* promoter is associated with protection from severe malaria in Thailand, Genes Immunol., 2003, vol. 4(7), pp. 528–531.
17. Flori, L., Kumulungui, B., Aucan, C., et al., Linkage and association between plasmodium falciparum blood infection levels and chromosome 5q31-q33., Genes Immunol., 2003, vol. 4(4), pp. 265–268.
18. Naka, I., Nishida, N., Patarapotikul, J., et al., Identification of a haplotype block in the 5q31 cytokine gene cluster associated with the susceptibility to severe malaria, *Malar. J.*, 2009, vol. 8, pp. 232–237.
19. Wills-Karp, M., The gene encoding interleukin-13: a susceptibility locus for asthma and related traits, *Respir. Res.*, 2000, vol. 1(1), pp. 19–23.
20. Marsh, D., Neely, J.D., Breazeale, D.R., Linkage analysis of *IL-4* and other chromosome 5q31.1 markers and total serum immunoglobulin E concentrations, *Science*, 1994, vol. 264, pp. 1152–1156.
21. Bleecker, E.R., Amelung, P.J., Levitt, R.C., et al., Evidence for linkage of total serum IgE and bronchial hyperresponsiveness to chromosome 5q: A major regulatory locus important in asthma, *Clin. & Exp. Allergy*, 1995, vol. 25, pp. 84–88.
22. Postma, D.S., Bleecker, E.R., Amelung, P.J., et al., Genetic susceptibility to asthma: bronchial hyperresponsiveness coinherited with a major gene for atopy, *N. Engl. J. Med.*, 1995, vol. 333, pp. 894–900.
23. Jallow, M., Teo, Y.Y., Small, K.S., et al., Genome-wide and fine-resolution association analysis of malaria in West Africa, *Nat. Genet.*, 2009, vol. 41(6), pp. 657–565.
24. Zaharieva, I., Georgieva, L., Nikolov, I., et al., Association study in the 5q31-32 linkage region for schizophrenia using pooled DNA genotyping, *BMC Psychiatry*, 2008, pp. 8; doi:10.1186/1471-244X-8-11.
25. Chu, X., Dong, C., Lei, R., at al., Polymorphisms in the interleukin 3 gene show strong association with susceptibility to Graves' disease in Chinese population, *Â Genes Immunol.*, 2009, vol. 10, pp. 260–266.
26. Rihet, P., Traore, Y., Abel, L., et al., Malaria in humans: Plasmodium falciparum blood infection levels are linked to chromosome 5q31-q33, *Am. J. Hum. Genet.*, 1998, vol. 63, pp. 498–505.
27. Milet, J., Nuel, G., Watier, L., et al., Genome wide linkage study, using a 250K SNP map, of Plasmodium falciparum infection and mild malaria attack in a Senegalese population, *PLoS One*, 2010, vol. 5(7), e11616.
28. Hassan, H.Y., Underhill, P.A., Cavalli-Sforza, L.L., Ibrahim, M.E., Y-chromosome variation among Sudanese: Restricted gene flow, concordance with language, geography, and history, *Am. J. Phys. Anthropol.*, 2008, vol. 137(3), pp. 316–323.
29. Bowcock, A.M., Ruiz-Linares, A., Tomfohrde, J., et al., High resolution of human evolutionary trees with polymorphic microsatellites, *Nature*, 1994, vol. 368, pp. 455–457.
30. Sakagami, T., Witherspoon, D.J., Nakajima, T., et al., Local adaptation and population differentiation at the interleukin 13 and interleukin 4 loci, *Genes Immunol.*, 2004, vol. 5, pp. 389–397.
31. Wood, E.T., Stover, D.A., Slatkin, M., et al., The beta-globin recombinational hotspot reduces the effects of strong selection around HbC, a recently arisen mutation providing resistance to malaria, *Am. J. Hum. Genet.*, 2005, vol. 77, pp. 637–642.
32. Hollox, E.J., Poulter, M., Zvarik, M., et al., Lactase haplotypes diversity in the Old World, *Amer. J. Human Genetics*, 2001, vol. 68, pp. 160–172.
33. Deng, H., Fu, Y., Conditions for Positive and Negative Correlations Between Fitness and *Heterozygosity* in Equilibrium Populations, *Genetics*, 1998, vol. 148, pp. 1333–1340.
34. Tishkoff, S.A., Williams, S.M., Genetic analysis of African populations: Human evolution and complex disease, *Nat. Rev. Genet.*, 2002, vol. 3(8), pp. 611–621.
35. Zhu, Z., Yan, D., Cooper, R.S., Linkage disequilibrium and haplotype diversity in the genes of the renin–angiotensin system: Findings from the family blood pres-

sure program, *Genome Res*., 2003, vol. 13, Issue 2, pp. 173–181.

36. Tarazona-Santos, E., Tishkoff, S.A., Divergent patterns of linkage disequilibrium and haplotype structure across global populations at the interleukin-13 (IL13) locus. *Genes Immunol*., 2001, vol. 6(1), pp. 53–65.
37. Ziljstra, E.E., El-Hassan, A.M., Leishmaniasis in Sudan.Visceral Leishmaniasis, *Trans R. Soc. Trop. Med. Hyg*., 2001, Suppl. 1, pp. 27–58.
38. Ibrahim, M.E., The epidemiology of visceral leishmaniasis in east Africa: Hints and molecular revelations, *Trans R. Soc. Trop. Med. Hyg*., 2002, vol. 96, Suppl. 1, pp. 25–29.

THE 5q31 REGION IN TWO AFRICAN POPULATIONS AS A FACET OF NATURAL SELECTION BY INFECTIOUS DISEASES

© 2013 г. A. M. Elhassan[1], A. A. Hussein[1, 4], H. S. Mohamed[1], K. Rockett[2, 4], D. Kwiatkowski[2, 3, 4], A. M. Elhassan[1], and M. E. Ibrahim[1, 4]

[1]*Unit of Disease and Diversity, Department of Molecular Biology, Institute of Endemic Diseases, University of Khartoum, Khartoum, Sudan*

e-mail: abjil04@yahoo/com; mibrahim@iend.org

[2]*Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK*

[3]*Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK*

[4]*The MalariaGEN Consortium*

Received May 18, 2011; in final form, March 02, 2012

Cases of extreme natural selection could lead either to rapid fixation or extinction of alleles depending on the population structure and size. It may also manifest in excess of heterozygosity and the locus concerned will be displaying such drastic features of allele change. We suspect the 5q31 in chromosome 5 to mirror situation of such extreme natural selection particularly that the region encompasses genes of type 2 cytokine known to associate with a number of infectious and non-infectious diseases. We typed two sets of single nucleotide polymorphisms (SNPS) in two populations: an initial limited set of only 4 SNPS within the genes of *IL-4*, *IL-13*, *IL-5* and *IL-9* in 108 unrelated individuals and a replicating set of 14 SNP in 924 individuals from the same populations with disregard to relatedness. The results suggest the 5q31 area to be under intense selective pressure as indicated by marked heterozygosity independent of Linkage Disequilibrium (LD); difference in heterozygosity, allele, and haplotype frequencies between generations and departure from Hardy Weinberg expectations (DHWE). The study area is endemic for several infectious diseases including malaria and visceral leishmaniasis (VL). Malaria caused by *Plasmodium falciparum*, however, occurs mostly with mild clinical symptoms in all ages, which makes it unlikely to account for these indices. The strong selection signals seems to emanate from recent outbreaks of VL which affected both populations to varying extent.