The Rural-Urban Divide: Insights from Immuno-Genetic Profiles and Implications for Health

Reem Salah¹, Safa A.E Aboswar¹, Azza Saeed¹, Ayman A Hussein¹, Ibrahim M Elhassan¹, Kirk A Rockett², Dominic P Kwiatkowski², Muntaser Ibrahim¹, Ahmed M Elhassan¹

1 University of Khartoum
2 University of Oxford

Funding: MI has received a grant as part of the activities of the BioMalPar European Network of Excellence supported by a European grant (LSHP-CT-2004-503578) from the Priority 1 ‘Life Sciences, Genomics and Biotechnology for Health’ in the 6th Framework Programme. The funders did not play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. https://european-union.europa.eu/live-work-study/funding-grants-subsidies_en

Potential competing interests: No potential competing interests to declare.

Abstract

Population disparities in health and disease have been observed and amply documented. While often attributable to genetic underpinnings, such disparities extend beyond population genetic predisposition to include environmental and geographic determinants, most pronouncedly the division between rural and urban lifestyles. Under such influences, genes and gene products may become affected by epigenetic factors, microbial modifiers including infections, and the body microbiome that ultimately shapes the outcome of a complex milieu of protein networks.

Retrospective, demographic, genotype, and expression data from two rural populations in eastern Sudan were analysed for genotype, allele frequency distribution, Hardy-Weinberg equilibrium, and expression profiles using an array panel of Th1, Th2, and Th3 genes in a subset of the rural population sample against matched urban controls.

Differences between urban and rural samples were observed in the departure from HWE, with an excess of heterozygosity in the rural sample. In the Th1, Th2, and Th3 array, cytokines were consistently overexpressed in the rural cohort compared to the urban cohort and were replicated in 7 selected genes that are associated with chronic diseases amongst urban dwellers in contrast to rural village inhabitants. IgE levels, as a feature of parasitic infections, are another difference to include in that dichotomy.

Gene expression appears to be more exposed to the overall outcome of genetic variations, including the interaction with environmental influences within and outside the body. Here, it may be gathered from the contrast in the expression patterns between the rural and urban samples. The presence of signals of natural selection in genes that are key to certain biological functions, such as CD40L and FasL, and the sharp contrast between urban and rural populations in gene variants distribution and expression patterns, may provide important clues towards understanding the disparity between human communities in non-communicable diseases of lifestyle as well as some of the emerging infectious diseases.
Reem Hamad¹, Nahid A Eid¹, Safa A.E Aboswar¹, Azza Saeed¹, Ayman A Hussein¹, Ibrahim M Elhassan¹, Kirk A Rockett², Dominic P Kwiatkowski²†, Muntaser E Ibrahim¹*, and Ahmed M Elhassan¹†

¹ Department of Molecular Biology, Institute of Endemic Diseases, Medical Campus, Qasser Street, University of Khartoum, Khartoum, Sudan.
² Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Headington, Oxford.

* Correspondence: mibrahim@iend.com (MI)
† Deceased

Short title: The rural urban immunogenetic dictum

Keywords: Rural, Urban, immune profiles, CD40L, FasL. Transition.

Introduction

Rural-urban divides are among the earliest social structures in human history, with the term “civilization” often denoting the shift from rural-based economies and habits into cities and urbanized lifestyles. Since the last glacial maxima, the process has been ensuing progressively, reaching its zenith in the past few centuries with the industrial revolution. Such a fundamental shift has its impact on human biological and psychological well-being and disease transmission patterns. It has been speculated that a major trade-off took place whereby infectious diseases like Malaria and Helminthes, i.e., communicable diseases prevalent in rural communities, have been replaced by chronic and non-communicable diseases in urban environments, with some far-reaching implications.

During its biological history, the human species was under continuous challenge by infectious agents. It is the host-parasite interactions that shaped our evolutionary history, particularly diseases like malaria, which remains one of the major causes of mortality among children worldwide and thus among the strongest known forces for evolutionary selection in recent history [1]. Analysis of malaria candidate genes from malaria endemic areas has indicated the potential impact of the population structure on the outcome of association in susceptibility to malaria and that the structure is apparently a function of both the unified ethnicity and relatedness within the population [2]. The expression of these genes, however, is not necessarily a direct reflection of such variation and could be subject to the complex interaction of the protein networks of the body and that with the environment.

In the rapidly escalating urban centres, socioeconomic differences result in diverse environmental acquaintances which largely determine disease patterns. Understanding the factors associated with altering environments and the link to the change of the immune system would be important for both communicable and non-communicable disease prevention strategies. Although risk factors associated with the rural to urban transition, particularly relative to inflammatory diseases,
have been studied comprehensively\textsuperscript{[3][4], little is known about changes that take place in the immune system as a function of the rural–urban gradient.

During routine longitudinal health surveillances of rural populations in eastern Sudan\textsuperscript{[5]}, we have noticed a striking dearth of chronic illnesses like diabetes, hypertension, cancer, and asthma, while these illnesses are registering a steady increase in urban populations nationwide. Taken together, this shift forms what is known as the health features of the transition, the term coined in reference to the changes from a rural, mostly subsistent economy into an urbanized, lifestyle-based economy (reference: article on the sociobiological). The association of the transition with diseases like asthma has been amply described; however, no proper studies on the underlying genomic and molecular basis of chronic illnesses in relation to the transition have been carried out.

Despite the bountiful studies that tackled the genetic elements of gene expression, few have addressed the geographical location impact at the transcriptional level. Idaghdour and his colleagues studied the Amazighs of Morocco -a relatively genetically homogenous population- who live in three distinct geographical localities; they concluded that geographical area differences might have a dominant effect on gene expression profiles, up to one third of the leukocyte transcriptome\textsuperscript{[5][6]. Investigating gene expression patterns is one approach toward scrutinising differences in immune responses between urban and rural populations, as variability in gene expression is a result of environmental factors in addition to genetic ones\textsuperscript{[7]. Moreover, differential gene expression can be a key mechanism in disease manifestation.\textsuperscript{[8]}. To date, a lack of data has hampered the identification of functional genomic features that foster these differences, which might have led to a better understanding as well as the discovery of specific roles of genes in immunity. Here, we investigate whether contrasting cultural and geographical locations in urban and rural areas in Sudan that are endemic to malaria and leishmaniasis may have had an impact on the mRNA expression of selected immune genes among the population and, consequently, their health and disease profiles.

Materials and Methods

Ethics Statement

Analysis based on archived data and research projects approved by the Ethical Committee of the Institute of Endemic Diseases, University of Khartoum, was carried out. Samples were originally taken with written informed consent from all individuals.

The current analysis reports on retrospective population records and archived samples encompassing mainly populations of villages on the Rahad River in Southeast Gedaref State, Sudan. These populations have been the subject of longitudinal research surveys between the years 1994 and 2014. Archived data, collected for various purposes and several research projects at the time and approved by the Ethical Committee of the Institute of Endemic Diseases, University of Khartoum, were re-analyzed to meet the specific questions posed by the study. Samples were originally taken with open written informed consent from all individuals. Expression analysis was based on samples collected cross-
sectionally in April 2007. The anonymized control samples in the expression analysis were obtained from the biobank of the Institute of Endemic Diseases following proper procedures.

The authors had no access to information that could identify individual participants during or after data collection, except the principal investigator.

Study population and sample size

Genotype and phenotype data were analysed retrospectively from two rural village populations of the Hausa and Massalit ethnicities. These populations reside in the malaria-endemic area of the Rahad River in Eastern Sudan and have been under routine surveillance longitudinally over the past decades for different infectious diseases, including malaria and leishmaniasis [2].

A total of 168 SNPs were genotyped in 414 individuals from the Hausa population and 510 from the Masalit population using the Sequenom® iPLEX gold assay. The SNPs were originally chosen based on previously published reports of malaria candidate-gene associations, in addition to SNPs that had shown early promise for associations in a Genome Wide Study undertaken in the Gambia by the MalariaGEN consortium. No urban controls were genotyped in the MalariaGEN project, but controls consisting of 60 supposedly immunologically naïve samples for infectious diseases were selected based on an extended urban lifestyle and habitation; of those, 20 samples had their RNA extracted and subsequently analyzed for gene expression as below.

Expression analysis

To estimate the gene expression profile, we used the RT2 Profiler PCR Array for Th1/Th2/Th3 panel 96, a set of primers that includes 84 cytokine genes involved in the immune response. To analyse the expression profile of the samples from rural and urban areas, we used an online program on the website http://www.sabiosciences.com/rt_pcr_product/HTML/PAHS-011A.html of the company. Then, 7 genes were selected for further expression analysis in the rural and urban areas using RT-PCR based on their putative association with chronic diseases, including cancer, and their signalling pathways. For single gene expression, the following formula was used:

\[ \text{Ct GOI} - \text{Ct norm} = ^\text{C} \]

- **Ct GOI** is genome of interest (Specific gene)
- **Ct norm** is normal gene (House Keeping Gene)

Immunoglobulin E (IgE) Level Estimation
Stored blood samples collected in previous cross-sectional surveys in the years 2000, 2001, 2005, and 2007 from two village populations were analysed, and 50-60 subjects for IgE level using ELISA.

Results

Helminthic infections are also prevalent, but their incidence in these rural communities is not properly determined, except for a limited survey in school children (157 in Um Salala and 44 in Koka), where in Koka only 7 cases (5 cases of Taenia, 1 case of H. nana, and one of Enterobiasis) were found, while in Salala there were 39 cases (1 case of Ascaris, 8 cases of H. nana, 24 cases of Taenia, and 6 cases of Schistosoma mansoni).

Immunoglobulin E (IgE) Level

The normal value range of IgE is 0-380 IU/ml. Both villages show yearly fluctuation in the average IgE level, which varied between the two villages, with the Koka village consistently having higher readings than Salala. This variability was not significant in the cross-sectional surveys of years 2000 and 2001. However, in the years 2005 and 2007, the readings in Koka were much higher than the normal values, being 710 and 723 IU/ml in the two years, respectively, while in Salala the level was 387 and 374 IU/ml, a significant difference with \( P \leq 0.00081 \). (Fig. 1).

![Fig. 1. Average IgE level in Koka and Salala in years 2000, 2001, 2005, and 2007](image)

Genotype Analysis

Malaria-Gen genotype data, originally generated for association studies on genetic susceptibility to malaria, were used here to capture signals of natural selection by infectious diseases in rural populations and as a likely explanation for the differences in expression patterns.

Departure from HWE was observed in 14 SNPs in both cases and controls Table 1& Table 2, indicating either the impact
of natural selection or the effects of population substructuring. A control sample from the genome database, which is predominantly urban, was analysed, and the results showed that only 4 of those SNPs are in departure from HWE.

### Table 1. Hardy-Weinberg Equilibrium and significant association P value between urban and rural (Massalit)

<table>
<thead>
<tr>
<th>SNP</th>
<th>HWE_P_Case</th>
<th>HWE_P_Control</th>
<th>Urban sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOC44rs2706381</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>LOC441108.rs7704457</td>
<td>0.00284</td>
<td>0.6426</td>
<td>NS</td>
</tr>
<tr>
<td>IL3.rs35415145</td>
<td>NS</td>
<td>NS</td>
<td>0.014851485</td>
</tr>
<tr>
<td>IL5.rs2069818</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>IL22 rs2227485</td>
<td>0.83832</td>
<td>0.020623</td>
<td>NS</td>
</tr>
<tr>
<td>IL22 rs2227478</td>
<td>0.82505</td>
<td>0.024136</td>
<td>NS</td>
</tr>
<tr>
<td>IL1A rs17411697</td>
<td>0.56827</td>
<td>0.01637</td>
<td>NS</td>
</tr>
<tr>
<td>CTL4 rs2242665</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>IL22 rs2227491</td>
<td>0.72206</td>
<td>0.005468</td>
<td>NS</td>
</tr>
<tr>
<td>IL22 rs1012356</td>
<td>0.13729</td>
<td>0.039258</td>
<td>NS</td>
</tr>
<tr>
<td>RAD rs4621555</td>
<td>0.00347</td>
<td>0.66165</td>
<td>NS</td>
</tr>
<tr>
<td>IRF1.rs2070729</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>IL13.rs848</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>CD40LG rs3092945</td>
<td>0.03279</td>
<td>7.15E-11</td>
<td>NS</td>
</tr>
<tr>
<td>G6PD-376.rs1050829</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>IL17RE.rs708567</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>IL4.rs2243283</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>IL4 rs2243270</td>
<td>0.28867</td>
<td>0.02693</td>
<td>NS</td>
</tr>
<tr>
<td>IL20RA rs1555498</td>
<td>0.92658</td>
<td>0.04570</td>
<td>NS</td>
</tr>
<tr>
<td>ABO rs8176719</td>
<td>0.52372</td>
<td>0.00126</td>
<td>NS</td>
</tr>
<tr>
<td>P4HA2 rs3805685</td>
<td>0.14016</td>
<td>0.00159</td>
<td>0.044554455</td>
</tr>
</tbody>
</table>

**NS:** not significant

### Table 2. Hardy-Weinberg Equilibrium and significant association P value between Urban and Rural (Hausa).
<table>
<thead>
<tr>
<th>Gene</th>
<th>HWE_P_Case</th>
<th>HWE_P_Control</th>
<th>Urban sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOC44rs2706381</td>
<td>0.02613</td>
<td>0.018582</td>
<td>NS</td>
</tr>
<tr>
<td>LTA rs909253</td>
<td>0.88956</td>
<td>0.0068871</td>
<td>0.04950495</td>
</tr>
<tr>
<td>LTA rs2239704</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>IL5.rs2069818</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>IL22 rs2227485</td>
<td>0.048194</td>
<td>0.45624</td>
<td>NS</td>
</tr>
<tr>
<td>IL22 rs2227478</td>
<td>0.028801</td>
<td>0.88961</td>
<td>NS</td>
</tr>
<tr>
<td>CD36 rs3211938</td>
<td>0.3396</td>
<td>0.04037</td>
<td>0.014851485</td>
</tr>
<tr>
<td>CTL4 rs2242665</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>RTN3 rs542996</td>
<td>0.77931</td>
<td>0.048729</td>
<td>NS</td>
</tr>
<tr>
<td>GNAS rs8386</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>DERL3.rs3177244</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>CFTR rs17140229</td>
<td>NS</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>ADORA2B rs2535611</td>
<td>0.94521</td>
<td>0.041863</td>
<td>NS</td>
</tr>
<tr>
<td>CD40LG rs3092945</td>
<td>6.70E-06</td>
<td>5.31E-09</td>
<td>NS</td>
</tr>
<tr>
<td>CD40LG.rs17424229</td>
<td>0.097609</td>
<td>1.93E-08</td>
<td>NS</td>
</tr>
</tbody>
</table>

Th1/Th2/Th3 Array expression profile

There were clear differences in the pattern of expression between rural and urban areas. Overall, up-regulated genes were dominant in the rural area, while down regulation was the predominant mode in the urban samples. Th2 samples were down-regulated in several rural samples (Fig 2). The numbers of genes down- or upregulated were indicated in each column, which represents a single sample.
There are clear differences in the patterns of expression between rural and urban areas. Overall, up-regulated genes were dominant in the rural area, while down regulation was the predominant mode in the urban samples. Th2 samples were down-regulated in several rural samples.

### Gene expression profile

The analysis of the 10 genes was selected from those expressed in all samples. Seven genes were analysed for both areas. Again, down regulation of the expression of the selected genes was the main feature among the urban ($n = 23$) in comparison to the rural area ($\text{Salala}, n = 21; \text{Koka}, n = 29$).

Unlike the array expression, few rural samples showed a down expression tendency, and three urban samples showed slight overexpression in one gene ($\text{CTLA4}$). (Fig 3)
Fig 3. Shows selected gene expression differences between urban and rural (Koka and Salala). A. Rural expression bar chart. B. Urban expression bar chart. C. Heat map of combined urban and rural expression.

Discussion

In the human host, the outcome of infectious disease is determined by a complex interactive relationship between the host, the parasite, and the environment. Against such a complex multifactorial milieu, the dichotomous health disparities between rural and urban lives are considered. This obvious dichotomy has been validated by various reports and anecdotes across the globe, particularly in developing countries [9][10][11][12][13]. Underlying these disparities are multiple elements involving social and environmental factors that characterize and distinguish urban and rural spaces, and thereby affect health profiles and risk of contracting diseases. Despite observations and ample documentation, few studies have attempted to venture into understanding the genotypic and immunological differences in expression between the two spaces. In one setting, humans are in tandem with their evolutionary history, in terms of nutrition, exposure to pathogens, and psychosocial elements like crowding and nature of the household, etc. On the other hand, people are subject to new nutritional cultures, a pathogens-free environment, and stress where the repercussions on human health are enormous.

Among the interesting observations that may provide some clues to this dichotomous health status are the expression profiles of candidate genes originally tested within the Malaria-Gen consortium, of which some were included in the Th1/Th2/Th3 arrays. Several of these genes are candidates for a disparate burden of chronic diseases like Asthma (IL-4, IL-5, and TLR4) and cancer (CD-40, FASL, CTLA4, and CCR5) and are being investigated against the reported extremely low incidence of asthma and cancer in the area.

Rural-urban differences seen in the current data sets were of great interest. Despite the small urban sample size, in the
overall outcome, both communities corresponded to the dichotomous urban-rural divide. How much natural selection is responsible for such differences in expression remains to be answered. Populations of the rural area sampled in this study were the subject of extended investigation over three decades for the transmission of infectious diseases, where the area is endemic for both malaria and visceral leishmaniasis. Conversely, prolonged exposure over time in some cases may result in genetic and epigenetic differences. In these villages, the impact of natural selection in some candidate genes is possibly due to pressure by infectious diseases like VL rather than malaria. Natural selection is best seen in the departure from Hardy-Weinberg expectations, which is shown in a group of loci including CTLA4 and CFTR. Interestingly, it is recorded more frequently in controls than cases, suggesting that these signals might emanate from reasons other, in addition to natural selection, like sampling effects and population structure, both known causes of DHWE to be considered.

Whether helminthic infections could explain some of the health features in these rural communities, including elevated IgE, a Bonafede marker of asthma, remains to be established. Bearing in mind that IgE has been seen to be elevated during malaria infection, differences in IgE levels between the two villages, however, could not be ascribed solely to malaria, as the prevalence and incidence rates of the disease are quite comparable between the villages. Such an observation might insinuate an innate/genetic component in the Hausa community that favours the production of high IgE levels. In Nigeria, a study was done among children where the mean serum level of IgE was significantly raised in children with helminthic or protozoan infection. In contrast, the level of IgE was reduced in children with *Ascaris lumbricoides* compared to *Plasmodium* species. While another study in Ghana found elevated levels of IgE among rural inhabitants, explained by high incidence of helminthic infections.

Differences in malaria transmission between rural and urban centres could be one of the main drivers of these patterns, as several of these markers like IL-4, IL-5, IL-10, and IL-13 are markers of CD4 T-Helper 2 activity. Lower CD40 activity in urban compared to rural populations conforms to the literature of higher susceptibility to chronic diseases among urban populations. On the other hand, the low burden of chronic diseases like asthma, diabetes, and cancer in rural communities could have complex underlying molecular aetiologies beside lifestyle.

However, it might prove challenging to pinpoint or specify actual effector molecule(s) in the trade-off between chronic and infectious diseases, given the intricacies of signalling pathways and epigenetic determinants. One might be tempted, though, to hypothesize that the observed low cancer incidence in these villages might have something to do with the endemicity of the malaria parasite. Salanti et al., (2015), have shown homology between cancer cells and the malaria parasite and hence antibodies against malaria proteins may be potentially protective against cancer. TNF is obviously another potential candidate for these trade-offs. In Ghana, it was found that TNF-a, IL-6, and IL-12 levels in urban dwellers with malaria were significantly higher, while IL-10, CD4, CD3, CD8T-cell levels, and the CD4/CD3 ratio were significantly lower. However, in the current dataset, the urban population displayed an overall low expression of key cytokine genes in both Th1/Th2 classes, which is consistent with lower exposure to the microbial environment, something observed in some other studies. When comparing the expression between three geographically distinct groups in Indonesia, a significant association was found, especially among genes involved in immune function, suggesting a potentially adaptive...
response to local environmental pressures [26].

One limitation of the current study is that it was conducted with relatively small numbers of subjects in each study area. A larger sample size with several points of collection over time may have meant greater statistical power in detecting area differences for some of the genes. Common to all mRNA studies of whole blood, our study also suffers from the fact that mRNA expression might not be directly related to protein expression levels. In addition, the expression of the mRNA is in whole blood and does not reveal any cell-specific profiles, which might be important when considering their function in determining disease profiles.

Nevertheless, results and observations reported here address some of the crucial matters pertaining to public health, especially in developing countries under changing disease patterns accompanying changes of lifestyle. Given that subsistence economies in rural Africa, like that in the Gedaref state, are vanishing modes of life, leaves us with a narrow window of opportunity to foster an in-depth knowledge of the biology of health transition that might inform future health policies and research directions.

Acknowledgments

Authors are indebted to the inhabitants of the villages of Koka and Salala who diligently cooperated in providing biological samples for continued health surveillance in the Rahad River locality, thus making it possible to unlock many of the peculiarities of the diseases endemic to their area. In turn, the manuscript is dedicated to the memories of two giant figures in science, who co-authored this manuscript before passing away: A M Elhassan and D P Kwiatkowski, whose tribute to the well-being of populations of these villages remains some of their everlasting legacy.

References


3. ^a, b^ Fezeu L, Balkau B, Kangne AP, Sobngwi E, Mbanya JC. Metabolic syndrome in a sub-Saharan African setting: central obesity may be the key determinant. Atherosclerosis. 2007;193(1):70-76. doi:10.1016/j.atherosclerosis.2006.08.037


