International Journal of Infectious Diseases 133 (2023) 1–4

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

Short Communication

First genome-wide association study for lymphatic filariasis in a West African population points to a human leukocyte antigen-mediated disease pathophysiology

Sandeep Grover[1,#], Vera S. Opoku[2,#], Linda B. Debrah[2,3], Carlo Maj[1], Jubin Osei-Mensah[2,4], Derrick A. Mensah[2], Achim Hoerauf[5,6], Alexander Y. Debrah[2,7], Johannes Schumacher[1,#,*], Kenneth Pfarr[5,6,#]

[1] *Center for Human Genetics, Philipps University of Marburg, Marburg, Germany*
[2] *Kumasi Center for Collaborative Research in Tropical Medicine (KCCR), Kumasi, Ghana*
[3] *Department of Clinical Microbiology, School of Medicine and Dentistry, KNUST, Kumasi, Ghana*
[4] *School of Veterinary Medicine, KNUST, Kumasi, Ghana*
[5] *Institute of Medical Microbiology, Immunology and Parasitology (IMMIP), Bonn, Germany*
[6] *German Center for Infection Research (DZIF), Partner Site Bonn-Cologne, Bonn, Germany*
[7] *Department of Medical Diagnostics, School of Medicine and Dentistry, KNUST, Kumasi, Ghana*

ARTICLE INFO

Article history:
Received 22 February 2023
Revised 10 April 2023
Accepted 22 April 2023

Keywords:
Lymphatic filariasis
Wuchereria bancrofti
Genome-wide association study (GWAS)
West African population
Lymphedema
Hydrocele

ABSTRACT

Objectives: Lymphatic filariasis (LF) represents a parasitic disease caused by filarial nematodes. Although some infected individuals present an asymptomatic course, others suffer severe chronic lymphatic pathology, including lymphedema, hydrocele, and elephantiasis. Several studies have shown that host genetic factors influence LF susceptibility and chronic pathology. The current study aimed to conduct the first genome-wide association study to systematically determine LF susceptibility.
Methods: We analyzed genome-wide single-nucleotide polymorphism data from 1459 LF cases and 1492 asymptomatic controls of West African (Ghanaian) descent.
Results: We identified two independent genome-wide significant associated genetic variants near the genes *HLA-DQB2* (rs7742085) and *HLA-DQA1* (rs4959107) contributing to LF and/or lymphedema susceptibility ($P < 5.0 \times 10^{-8}$, odds ratios [ORs] >1.30). We also observed suggestive evidence of LF associations ($P < 1.0 \times 10^{-6}$) at two non-HLA loci, near the genes *ZFHX4-AS1* (rs79562145) and *CHP2* (rs12933387). In contrast, we could not replicate any previously reported LF associations drawn from candidate gene association studies. On the polygenic level, we show that our genome-wide association study data explain 24-42% of LF heritability, depending on an assumed population prevalence of 0.5-5.0%.
Conclusion: Our findings point to an involvement of HLA-mediated immune mechanisms in LF pathophysiology.

Introduction

Lymphatic filariasis (LF) is a debilitating infectious disease caused by nematode parasites transmitted through mosquitoes. Three filarial nematode species are disease-causing, namely Wucheria bancrofti, Brugeria timori, and Brugeria malayi [1–3]. The mosquito deposits the infective larval stage of the parasite on human skin, where they migrate to the lymphatic system, resulting in significant lymphatic damage. The disease is endemic, with an estimated 59 million infections, in tropical and subtropical regions, including the sub-Saharan African region [4]. The parasitic infections are a major health concern and affected individuals respond differently. Although the majority show an asymptomatic course, a proportion of the infected present with LF characterized by lymphedema (LE), hydrocele (HYD), and elephantiasis. The current treatment is limited to antifilarial drugs and management of morbidity [5]. LF tends to aggregate in families, independent of

* Corresponding author.
E-mail address: johannes.schumacher@uni-marburg.de (J. Schumacher).
# Contributed equally.

https://doi.org/10.1016/j.ijid.2023.04.408

household and environment, suggesting a role of host genetic factors in response to infection [6].

To date, all genetic studies on LF have used a candidate gene approach based on prespecified hypotheses in a small number of cases and controls. Supplementary Table 1 lists all studies that were performed on at least 100 patients with LF. Although some associations were found in these studies, it is unclear to what extent they represent true-positive findings because of the small numbers of included cases. The objectives of the current study was to overcome the limitations of previous studies. First, we conducted a genome-wide association study (GWAS) of susceptibility to LF in a sizable Ghanaian population comprising 1459 LF cases and 1492 controls. Second, we aimed to estimate the single-nucleotide polymorphism (SNP)-based genetic heritability of LF. Third, we aimed to replicate the previously reported candidate gene-based LF associations.

Methods

Study participants

A total of 1459 cases with LF (961 females, 498 males) and 1492 asymptomatic controls (843 females, 649 males) who have been infected by nematode parasites were enrolled in the study. All participants were of Ghanaian ancestry. The cases included 1166 patients with LE (961 females, 205 males), 278 male patients with HYD, and 15 male patients with both LE and HYD. The recruitment scheme and clinical characteristics of all participants are provided in detail in the Supplementary Methods.

Genotyping, quality control, and imputation

The genotyping was performed using Illumina HumanOmniExpress BeadChip. Standardized quality control (QC) measures were applied to the genotyped data [7]. Multidimensional scaling components were used to identify ancestral outliers (Supplementary Figure 1). All post-QC genotypes were imputed using the TransOmics in Precision Medicine (TOPMed) Imputation server (https://imputation.biodatacatalyst.nhlbi.nih.gov/start.html#). The details on QC and imputation are provided in the Supplementary Methods.

Association analysis and heritability

The GWAS of imputed dosages with LF as the outcome was conducted using an additive logistic regression model, adjusting for sex and multidimensional scaling components 1-10. The details on the association testing are provided in the Supplementary Methods. In addition, we used the restricted maximum likelihood approach implemented in the genome-wide complex trait analysis (https://yanglab.westlake.edu.cn/software/gcta/#Overview) to estimate the LF heritability in unrelated individuals, assuming a disease prevalence of 0.5, 1.0, 2.0, and 5.0% [4].

Functional annotation

Functional annotation was carried out using functional mapping and annotation functions, namely SNP2GENE and GENE2FUNC (https://fuma.ctglab.nl/). We used the genotype tissue expression (GTEx) database (https://gtexportal.org/home) to correlate the associated LF genetic variants with gene expression data. Finally, we assessed the pleiotropic effect of the most significantly associated locus and performed a phenome-wide association study (PheWAS) using the GWAS atlas (https://atlas.ctglab.nl). Details on all functional analyses are presented in the Supplementary Methods.

Results

Association analysis and heritability

The imputation resulted in a final dataset of 5,592,510 autosomal SNPs. The QQ plot showed no inflation ($\lambda = 1.03$) (Supplementary Figure 2). We identified one genome-wide significant associated risk variant (Supplementary Figure 3). SNP rs7742085 on chromosome 6p21 near the gene *HLA-DQB2* was LF-associated, with $P = 1.42 \times 10^{-9}$ (odds ratio [OR] = 1.43; confidence interval [CI] 1.27-1.60) (Table 1). The regional association plot for rs7742085 showed no linkage disequilibrium ($r^2 > 0.8$) with any nearby variants (Figure 1). Moreover, we found suggestive LF associations ($P < 1 \times 10^{-6}$) at two independent loci near the genes *ZFHX4-AS1* (rs79562145) and *CHP2* (rs12933387) (Table 1). Heritability, estimated from individual-level GWAS data, suggested that common genetic variants account for 24-42% of the LF heritability (Supplementary Table 1).

We next performed the subtype analyses using LE and HYD as the outcomes, which are presented in Supplementary Figure 4 and Supplementary Table 2. SNP rs7742085 on chromosome 6p21 was also genome wide significantly LE-associated ($P = 1.91 \times 10^{-8}$, OR = 1.44 [CI 1.27-1.64]). However, a second variant at the same locus, rs4959107, showed an even stronger LE association ($P = 1.10 \times 10^{-8}$, OR = 1.46 [CI 1.28-1.66]) (Supplementary Figure 5). This variant is located near the gene *HLA-DQA1* and 154 kb distant from rs7742085. Because both variants show no linkage disequilibrium in our study population ($r^2 = 0.070$), the association signals seem to be independent. Accordingly, rs7742085 showed an LE association of $P = 2.50 \times 10^{-5}$ after conditioning on rs4959107 (and *vice versa*, rs4959107 showed a *P*-value of 1.36×10^{-5} after conditioning on rs7742085) (data not shown). In contrast, the subtype analysis using HYD as the outcome revealed no genome-wide significant association signal, which is most probably due to the smaller number of HYD cases than LE cases (278 vs 1166 individuals). However, whereas HYD seems to contribute to the association signal of rs7742085 ($P = 0.143$, OR = 1.22 [CI 0.94-1.58]), the association signal of rs4959107 appears to be HYD-independent or LE-specific ($P = 0.899$, OR = 1.02 [CI 0.80-1.29]). Supplementary Table 3 provides an overview of both genome-wide significant associated

Table 1
Suggestive and genome wide significantly associated SNPs contributing to LF susceptibility in the Ghanaian population.

Genomic locus	SNP	Chr.	Position (bp, hg19)	Function	Nearest gene	Effect allele	Other allele	Effect allele frequency	Odds ratio (95% confidence interval)	Regression coefficient (β)	Standard error of β	*P*-value
1	rs9269041	6	32438243	ncRNA_intronic	*HLA-DRB9*	G	A	0.267	0.73 (0.65-0.82)	-0.313	0.060	1.99E-07
1	rs9274724	6	32637531	Intergenic	*HLA-DQB1*	A	G	0.432	0.77 (0.70-0.86)	-0.259	0.053	9.89E-07
1	**rs7742085**	**6**	**32746178**	**Intergenic**	***HLA-DQB2***	**G**	**T**	**0.268**	**1.43 (1.27-1.60)**	**0.357**	**0.059**	**1.42E-09**
2	rs79562145	8	77576245	ncRNA_intronic	*ZFHX4-AS1*	A	G	0.205	1.39 (1.22-1.58)	0.331	0.066	4.59E-07
3	rs12933387	16	23751216	Intergenic	*CHP2*	G	A	0.319	0.74 (0.66-0.83)	-0.301	0.057	1.71E-07

bp, base pairs; Chr., chromosomes; HLA, human leukocyte antigen; ncRNA, non-coding RNA; SNP, single-nucleotide polymorphism.
Bold = genomewide significantly associated at conventional *P* <5E-08; All other rows = suggestive associated at *P* <1E-6.

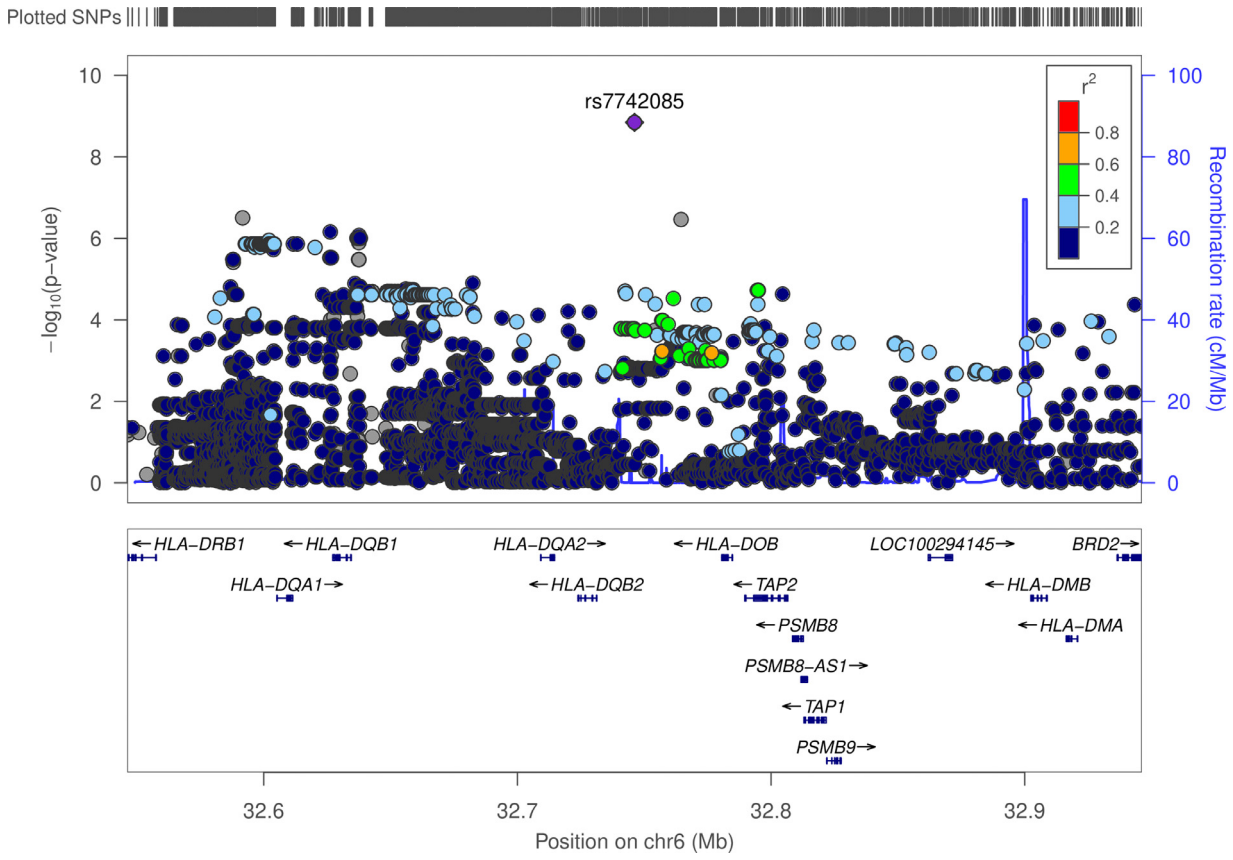

Figure 1. Regional association plot for LF at the HLA region in the Ghanaian population.
The plot shows a 400 kb window centered on the genomewide significantly associated LF SNP. The upper panel shows the significance of SNP associations with LF. The vertical lines represent the local recombination rate. The lower panel lists all genes in the region.
Chr., chromosomes; HLA, human leukocyte antigen; LF, lymphatic filariasis; SNP, single-nucleotide polymorphism.

SNPs within the human leukocyte antigen (HLA) region in LF, LE, and HYD.

Identification and replication of previously reported loci

A systematic literature search yielded six publications reporting a significant association of 11 genetic polymorphisms with LF infection (Supplementary Table 4). Among these, three were detected in our GWAS and none of them were LF-associated (Supplementary Table 5).

Functional annotation

The MAGMA gene-based association analysis implicated two genes ($P < 2.65 \times 10^{-6}$) as disease-associated: *HLA-DRA* and *BTNL2* using LF and LE as the outcomes (Supplementary Table 6). The MAGMA gene set analysis further implicated 30 genes involved in pre-mRNA binding in the pathogenesis of LE ($P_{\text{Bonferroni}} = 0.035$) (Supplementary Table 7). The differentially expressed gene set (DEG) analysis using GTEx did not show dysregulated gene sets across any tissue subtype. However, the enrichment analysis for Kyoto encyclopedia of genes and genomes (KEGG) terms showed antigen processing and presentation as the most significant pathways for both LF and LE (Supplementary Figure 6). In contrast, the expression quantitative trait locus (eQTL) analysis using GTEx data did not yield any association of the genome-wide significantly associated LF or LE SNPs (rs7742085, rs4959107) with gene expression. This is also the case for the suggestive associated LF risk SNPs (rs79562145, rs12933387), which showed no eQTL effects in the GTEx data. Furthermore, the PheWAS identified gastrointestinal (celiac disease and primary sclerosing cholangitis), endocrine traits (type 1 diabetes), and connective tissue-related traits (rheumatoid arthritis) as the most significantly associated traits at the *HLA-DQB2/DQA1* locus (Supplementary Table 8, Supplementary Figure 7).

Discussion

We observe two independent genome-wide significant associations with susceptibility to LF and/or LE in the major histocompatibility complex class II region, which are located near the genes *HLA-DQB2* (rs7742085) and *HLA-DQA1* (rs4959107). Both genes encode for proteins that play an important role in presenting peptides derived from extracellular antigens to T-helper cells (CD4+ T cells), which, in turn, leads to the synthesis of specific antibodies by B cells [8]. According to its prominent role in immune-mediated processes, it is well known that genetic variants within the HLA region, including *HLA-DQB2* and *HLA-DQA1*, play a major role in predisposition to various infectious diseases [9]. Accordingly, previous studies have speculated on the involvement of HLA-mediated immunity in LF development [10,11].

Although our GWAS revealed the first genome-wide significant LF/LE associations, we were not able to study the associated region at a higher resolution. Accordingly, we could not impute specific HLA alleles due to the absence of the *HLA-DQB2/DQA1* locus in any publicly available reference panels. Thus, HLA long-read sequencing data in the Ghanaian population are now needed to identify the

HLA alleles that represent the true LF/LE risk variants within this region [12].

In addition, to the best of our knowledge, our GWAS provides the first genetic heritability estimate. Our data show that common genetic variants account for 24-42% of LF heritability. These data are higher than the heritability estimates based on circulating filarial antigens (CFAs). The detection of CFAs is considered as the gold standard for LF diagnosis, and the heritability estimates for *Wucheria bancrofti* CFAs was previously calculated to be 23% in the Republic of Congo [13].

In summary, to the best of our knowledge, this study represents the first GWAS on LF using the largest case-control sample analyzed thus far. Our findings point to an involvement of HLA-mediated immune processes in LF pathophysiology. Moreover, our data show that LF has a polygenic risk architecture and provide the first LF heritability estimate.

Declarations of Competing Interest

The authors have no competing interests to declare.

Funding

This study was supported by the German Research Foundation (PF 673/2-0, PF 673/4-1, and PF 673/6-1).

Ethical approval

The study was approved by the ethical committee of the Kwame Nkrumah University of Science and Technology (No. CHRPE/AP/366/17) and the University Hospital of Bonn (No. 018/12). All participants gave their written informed consent for their participation in the study.

Acknowledgements

The authors thank all the study participants, scientists, and clinicians who contributed to this study.

Author contributions

AYD, AH, JS, and KP conceptualized the study. KP and JS obtained funding. AYD, LBD, JOM, VSO, and DAM performed the field work and obtained patient samples. LBD and VSO prepared patient samples for genotyping. SG analyzed the data and drafted the manuscript. JS, CM, LBD, VSO, KP, AH, and AYD edited the manuscript. All authors read and approved the final manuscript.

Data availability

Summary statistics of LF GWAS are available from the corresponding author upon reasonable request.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijid.2023.04.408.

References

[1] Dasgupta A, Bala S, Dutta SN. Lymphatic filariasis in man: demonstration of circulating antigens in Wuchereria bancrofti infection. *Parasite Immunol* 1984;**6**:341–8. doi:10.1111/j.1365-3024.1984.tb00806.x.

[2] David HL, Edeson JFB. Filariasis in Portuguese Timor, with observations on a new microfilaria found in man. *Ann Trop Med Parasitol* 1965;**59**:193–204. doi:10.1080/00034983.1965.11686299.

[3] Soh CT, Lee KT, Im SW, Lee JH. Clinical manifestation of Brugia malayi infection in Korea. *Kisaengchunghak Chapchi* 1966;**4**:1–6. doi:10.3347/kjp.1966.4.2.1.

[4] Local burden of disease 2019 Neglected tropical diseases collaboratorsThe global distribution of lymphatic filariasis, 2000-18: a geospatial analysis. *Lancet Glob Health* 2020;**8**:e1186–e1e94. doi:10.1016/S2214-109X(20)30286-2.

[5] World Health Organization. Ending the neglect to attain the Sustainable Development Goals – A road map for neglected tropical diseases 2021–2030, https://www.who.int/publications/i/item/9789240010352; 2020 [accessed 5 January 2023].

[6] Cuenco KT, Ottesen EA, Williams SA, Nutman TB, Steel C. Heritable factors play a major role in determining host responses to Wuchereria bancrofti infection in an isolated South Pacific island population. *J Infect Dis* 2009;**200**:1271–8. doi:10.1086/605844.

[7] Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. A tutorial on conducting genome-wide association studies: quality control and statistical analysis. *Int J Methods Psychiatr Res* 2018;**27**:e1608. doi:10.1002/mpr.1608.

[8] Kindt T, Goldsby R, Osborne B, Kuby J. *Kuby immunology*. New York: W.H. Freeman; 2007.

[9] Medhasi S, Chantratita N. Human leukocyte antigen (HLA) system: genetics and association with bacterial and viral infections. *J Immunol Res* 2022;**2022**:9710376. doi:10.1155/2022/9710376.

[10] Chan SH, Dissanayake S, Mak JW, Ismail MM, Wee GB, Srinivasan N, et al. HLA and filariasis in Sri Lankans and Indians. *Southeast Asian J Trop Med Public Health* 1984;**15**:281–6.

[11] Yazdanbakhsh M, Sartono E, Kruize YC, Kurniawan A, Partono F, Maizels RM, et al. HLA and elephantiasis in lymphatic filariasis. *Hum Immunol* 1995;**44**:58–61. doi:10.1016/0198-8859(95)00059-d.

[12] Liu C, Yang X, Duffy BF, Hoisington-Lopez J, Crosby M, Porche-Sorbet R, et al. High-resolution HLA typing by long reads from the R10.3 Oxford nanopore flow cells. *Hum Immunol* 2021;**82**:288–95. doi:10.1016/j.humimm.2021.02.005.

[13] Chesnais CB, Sabbagh A, Pion SD, Missamou F, Garcia A, Boussinesq M. Familial aggregation and heritability of Wuchereria bancrofti infection. *J Infect Dis* 2016;**214**:587–94. doi:10.1093/infdis/jiw212.