Genome-wide analysis of the human head louse (*Pediculus humanus capitis*) reveals geographically structured genetic populations

Niyomi House[1], Aida Miro-Herrans[1], Julie Allen[2], Muhammad Ashfaq[3], Bret Boyd[4], Henk Braig[5], Oleg Mediannikov[6], Kosta Mumcuoglu[7], Alejandra Perotti[8], Ramírez-Rozzi Fernando[9], Didier Raoult[6], Lajos Rozsa[10], Jan Štefka[11], Ariel Toloza[12], Natacha Wirdemark[13], and David Reed[1]

[1]University of Florida
[2]University of Nevada Reno
[3]University of Guelph
[4]Virginia Commonwealth University
[5]Bangor University
[6]Aix-Marseille Universite
[7]Hebrew University Hadassah Medical School
[8]University of Reading
[9]Centre National de la Recherche Scientifique
[10]Centre for Ecological Research
[11]University of South Bohemia in Ceske Budejovice
[12]Centro de Investigaciones de Plagas e Insecticidas
[13] LusFri Inspagat AB

April 4, 2023

Abstract

The human head louse (*Pediculus humanus capitis*) is an obligate ectoparasite of humans and has the potential to uncover aspects of human history that cannot be directly inferred from genetic data derived from humans. Previous studies have shown that global louse populations exhibit restricted patterns of genetic variation. However, these studies were restricted both genetically and lacked a global sampling. With the aim of capturing the genetic diversity of head louse populations from around the world, we generated whole genome sequences of human head lice from 43 countries, spanning five continents and Oceania, to determine if louse nuclear diversity mirrors its mitochondrial haplotypes or if population genetic structure, genetic diversity, and population connectivity are associated with geographical regions or host behavior. Here we show that there are five nuclear genetic clusters that are associated with large geographical regions, either at continental or intercontinental levels. High genetic variation was found between African and non-African individuals and the highest genetic diversity was found in samples from sub-Saharan Africa, similar to that of humans. Unlike the mitochondrial clades examined in previous studies, nuclear genetic clusters of lice examined here are highly structured based on geography (continentally and major regions within continents). Results from our genome analyses revealed that host-mediated global dispersal as the likely primary process in shaping diversity and maintaining genetic population boundaries within the nuclear genome of the human head louse.

Genome-wide analysis of the human head louse (*Pediculus humanus capitis*) reveals geographically structured genetic populations

Running title: Population genomics of human head lice

Niyomi House[1,2,4 *], Aida T. Miro-Herrans[2,3], Julie M. Allen[4], Muhammad Ashfaq[5], Bret M. Boyd[6], Henk Braig[7,7a], Oleg Mediannikov[8], Kosta Y. Mumcuoglu[9], Alejandra Perotti[10], Ramírez-Rozzi V. Fernando[11], Dider Raoult[8], Lajos Rozsa[12,13], Jan Stefka[14], Ariel C. Toloza[15], Natacha V. Wirdemark[16] and David L. Reed[2]

[1] Department of Biology, University of Florida, Gainesville, Florida, USA

[2] Florida Museum of Natural History, University of Florida, Gainesville, Florida, USA

[3] George Smathers Libraries, University of Florida, Gainesville, Florida, USA

[4] Department of Biology, University of Nevada, Reno, Nevada, USA

[5] Centre for Biodiversity Genomics and Department of Integrative Biology, University of Guelph, Ontario, Canada

[6] Center for Biological Data Science, Virginia Commonwealth University, Richmond, Virginia, USA

[7] School of Natural Sciences, Bangor University, Wales, UK

[7a] Institute and Museum of Natural Sciences, National University of San Juan, San Juan, Argentina

[8] Aix Marseille Univ, IRD, AP-HM, MEPHI, Marseille, France.

[9] Department of Microbiology and Molecular Genetics, The Hebrew University-Hadassah Medical School, Jerusalem, Israel

[10] Ecology and Evolutionary Biology Section, School of Biological Science, University of Reading, UK

[11]UMR 7206 Ecoanthropologie, MNHN, CNRS, UP

[12] Institute of Evolution, ELKH Centre for Ecological Research, Hungary

[13] Centre for Eco-Epidemiology, National Laboratory for Health Security, Hungary

[14] Faculty of Science, University of South Bohemia in Ceske Budejovice, CZ and Institute of Parasitology, Biology Centre CAS, CZ

[15] Centro de Investigaciones de Plagas e Insecticidas (CONICET-UNIDEF), Villa Martelli, Buenos Aires, Argentina

[16] LusFri Inspagat AB, Stockholm, Sweden

[*] Address correspondence to Niyomi House, niyomiw88@gmail.com

Abstract

The human head louse (*Pediculus humanus capitis*) is an obligate ectoparasite of humans and has the potential to uncover aspects of human history that cannot be directly inferred from genetic data derived from humans. Previous studies have shown that global louse populations exhibit restricted patterns of genetic variation. However, these studies were restricted both genetically and lacked a global sampling. With the aim of capturing the genetic diversity of head louse populations from around the world, we generated whole genome sequences of human head lice from 43 countries, spanning five continents and Oceania, to determine if louse nuclear diversity mirrors its mitochondrial haplotypes or if population genetic structure, genetic diversity, and population connectivity are associated with geographical regions or host behavior. Here we show that there are five nuclear genetic clusters that are associated with large geographical regions, either at continental or intercontinental levels. High genetic variation was found between African and non-African individuals and the highest genetic diversity was found in samples from sub-Saharan Africa, similar to that of humans. Unlike the mitochondrial clades examined in previous studies, nuclear genetic clusters of lice examined here are highly structured based on geography (continentally and major regions within continents).

Results from our genome analyses revealed that host-mediated global dispersal as the likely primary process in shaping diversity and maintaining genetic population boundaries within the nuclear genome of the human head louse.

Keywords

host-mediated dispersal, nuclear, genome, population structure, genetic diversity

Introduction

The study of organisms that are obligately or facultatively associated with humans can provide insight into the history of its host. For example, human head lice have been used as markers of the past to uncover prehistoric contact between extinct and anatomically modern humans (AMH) (Reed et al., 2004), the timing of the split between human and chimpanzees (5-7 Ma) (Reed et al., 2007), and the origin of clothing use by AMH (Kittler et al., 2003; Toups et al., 2011). The human head louse *Pediculus humanus capitis* De Geer, 1778 (Phthiraptera: Anoplura) is a permanent, blood-feeding ectoparasite of humans completing their entire life cycle on their human host. Lice cannot live off their host for more than 36 hours (Burkhart & Burkhart, 2007), making their life cycle closely tied to their hosts for thousands of generations. Head lice are primarily found among school-aged children and dispersal (transmission) from one head to another occurs mainly when the host individuals have close physical contact (Mumcuoglu et al., 2021). Ancient human populations also carried lice, and this is demonstrated by the finding of nits still attached to ancient human hair as well as trapped in the combs used (Amanzougaghene et al., 2016; Arriaza et al., 2014; Pedersen et al., 2022). Therefore, host migration and contact should define the population structure of lice. Historically, this appears to be true, that the evolutionary history of human lice appears to have been shaped by human migration (Ewing, 1926; Pedersen et al., 2022) that facilitated host-switching (Reed et al., 2007). However, the increase in ease of global migration and travel could break down geographical barriers and erase historical population structure. The question remains, is louse population structure defined by geographic regions and historic migration or recent host movement.

Anatomically modern humans migrated out of Africa 60-70 thousand years ago (Carto et al., 2009; Stringer, 2000; Walter et al., 2000). A growing number of studies now support two main dispersal events out of Africa: a Southern route to Southeast Asia and a Northern route to Eurasia (Reyes-Centeno et al., 2014; Tassi et al., 2015). As AMHs dispersed throughout Eurasia they interbred with other species of archaic humans including Neanderthals and Denisovans who migrated out of Africa much earlier (Reich et al., 2011; Sankararaman et al., 2014; Villanea & Schraiber, 2019). While genetic and archeological evidence has increased our understanding of human evolution, a complete picture remains elusive because behavior and dispersal are hard to discern from fossilized remains.

As speciation happens in the host lineage, host-specific parasites like lice often speciate in tandem as a result of imposed isolation (Reed et al., 2007). Therefore, the processes shaping human evolution may leave a signal in the genome of human lice. To date, genetic diversity and population structure of head lice has mostly been examined in terms of mitochondrial clade diversity and distribution (Al-Shahrani et al., 2017; Amanzougaghene et al., 2016; Ashfaq et al., 2015; Boumbanda Koyo et al., 2019). Mitochondrial genes; partial cytochrome b and cytochrome oxidase I of lice showed six genetically distinct, geographically unique clades (clades A-F). Clade A is distributed world-wide, while the other clades are more geographically restricted. Clade B is found in Europe, the Americas, and Australia. Clade C is found in Africa and southern Asia. Clades D and E are found only in Africa, while Clade F has been found on individuals in Argentina and Mexico (Amanzougaghene et al., 2019). Of these clades, only A and D include both head and body lice, while the rest of the clades have consisted only of head lice.

Although identifying the distribution of mitochondrial clades through only fragments of genes has helped elucidate the possibility of a demographic history of human head lice, these patterns are only vaguely indicative of just the maternal louse lineages, and longer term co-speciating and coevolutionary events. An understanding of genome-wide patterns of genetic diversity and population structure across global head louse populations are needed to characterize the complete picture of louse evolutionary history. For example,

considering that lice are horizontally transmitted, it is unclear if the nuclear genetic structure of the lice mirrors that of its host genetic population structure, or if human migrations and more recently the ease of global travel has erased these patterns by transporting lice around the world. The question remains whether lice are more defined by geography or host movement. Given the long history of human migration across the globe, and the co-demography observed in lice (Reed et al., 2007), it is useful to evaluate worldwide louse populations as geographical units (i.e., countries or continents) rather than genetic clades. Investigating population structure across geographically separated populations using neutral nuclear loci can show interactions between genetic drift and gene flow (Slatkin, 1987) that may not be captured by uniparentally inherited markers (i.e., mitochondrial DNA) (Ebert & Fields, 2020).

In attempts to understand the nuclear genetic diversity of human lice, a previous study used microsatellite markers from 93 specimens in 11 countries and found no relationship between the mitochondrial clades and the nuclear population clusters (Ascunce et al., 2013). Additionally, these microsatellite markers were able to detect geographically structured genetic population clusters indicating low levels of gene flow between global louse populations (Ascunce et al., 2013). Further investigations with a much broader sampling regime including genome-wide representation are needed to have the resolution to identify the distribution of genetic variants as well as the genetic population structure of the global human head louse populations.

Using the human body louse, *Pediculus humanus humanus* genome as a reference, we analyzed whole nuclear genomes of the human head louse*P. humanus capitis* from across the world to overcome the current lack of geographic and genome-wide representation. The main goals of this study were to: 1) investigate nuclear genetic diversity across the globe ; 2) examine the geographic distribution of genetic variants worldwide ; 3) identify the population structure based on nuclear markers; and 4) investigate possible modes of gene flow and historical population connectivity between global louse populations. Our main questions are, does the population structure of human lice match that of its host migration patterns or does it reflect the same evolutionary history shown through its mitochondrial genome?

Materials and Methods

Ethics Statement

The Institutional Review Board of the University of Florida exempted the study from review (Exemption of Protocol #2009-U-0422) and waived the need for written informed consent of the participants. This exemption is issued based on the United States Department of Health and Human Services (HHS) regulation 45 CFR part 46, because louse removal was voluntary and no identifiable information was recorded and so the need for written informed consent from the participants was waived.

Sample Collection

Head louse specimens were collected through collaborations including both scientific researchers and the non-scientific community (head louse removing clinics). In this study, we included a total of 377 specimens collected worldwide and DNA extractions that were previously collected and stored in Reed lab, Florida Museum of Natural History at University of Florida, USA from Ascunce et al. 2013 publication. The samples that were included in this study are summarized in Table 1. A special focus was made in the classification and selection of individual lice collected from different human heads, to maximize the genetic sampling within an area. All specimens were stored in 95% ethanol at -80^{o}C prior to DNA extractions.

Table 1: Geographic localities and the number of head louse samples included in this

study

Continent	Region	Country	Number of Head lice
Asia	West Asia	Turkey	16
		Israel	9
	South Asia	Pakistan	12

Continent	Region	Country	Number of Head lice
		India	3
		Nepal	6
	Southeast Asia	Cambodia	19
		Philippines	9
		Laos	4
		Thailand	3
		China	4
	East Asia	Mongolia	7
Africa	Sub-Saharan Africa	Ethiopia	11
		Burundi	1
		Democratic Republic of Congo	12
		Senegal	16
		Guinea	4
		Rwanda	1
		Gabon	6
		Cameroon	3
	North Africa	Egypt	3
		Algeria	8
Europe	Europe	United Kingdom	23
		Czech Republic	3
		Norway	3
		Netherlands	9
		France	5
		Sweden	10
		Spain	10
		Hungary	3
		Greece	4
		Italy	14
North America	North America	United States of America	48
		Mexico	15
	Central America	Guatemala	12
		Honduras	13
		Panama	4
South America	South America	Colombia	5
		Ecuador	5
		Argentina	18
		Venezuela	1
		Chile	2
	Oceania	Papua New Guinea	10
		Cook Islands	3
	Total		377

DNA Extractions

We extracted louse DNA using the MasterPure^{TM} Complete DNA and RNA Extraction Kit (Lucigen, Middleton, WI, USA). First, each louse specimen was prepared for DNA extraction by making a small incision on their abdomen under a dissecting microscope. Each specimen was then incubated overnight (<24hrs) at 52^0C in a solution of Tissue and Cell Lysis Solution mixed with 1ul of 20ug/ul Proteinase K. Occasional vortexing of the tubes facilitated cell lysis by allowing the solution to move into the abdominal incision. After the incubation period, the exoskeletons of the lice were removed from the tubes and stored in 70% ethanol

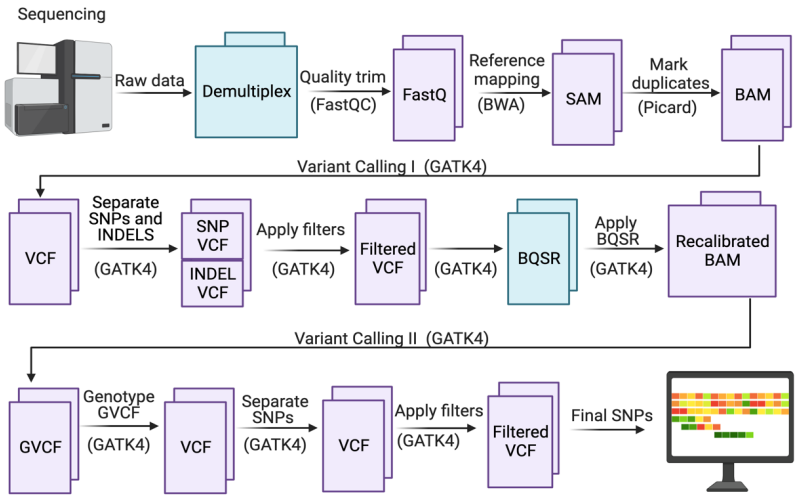
for future morphological analyses. We followed the manufacturer's protocol to purify the remaining digestion buffer. Using QUBIT dsDNA HS Assay Kit (Thermo Fisher Scientific), we quantified DNA concentrations. As a quality control step, to confirm the extraction of louse DNA, we ran a polymerase chain reaction (PCR) using published louse cytochrome B primers (Raoult et al., 2008) and PCR products were run on a 2% agarose gel. Samples that failed to amplify (resulting in no visual bands on the gel) were removed from any further analyses.

Library Construction and Sequencing

For whole genome resequencing, libraries were built using unique in-line indexed adapters designed by Rohland & Reich, 2012. First, 100 ng of DNA per sample were sheared to 400bp on the Covaris E220 focused-ultrasonicator using recommended factory settings for 400bp. Library construction included blunt-end repair, adapter ligation, and adapter fill-in that followed the protocol by Rohland & Reich, 2012. The individual samples were then PCR amplified with Pre-hyb primers following Rohland et al., 2015 and PCR products were purified using bead purification and pooled (Rohland & Reich, 2012). Sample pools were reamplified with indexed Illumina-compatible primers and they were size selected for fragments between 400-1000 bp and sequenced at 2x150 on the Illumina NovaSeq 6000 at the Interdisciplinary Center for Biotechnology Research (ICBR), University of Florida.

Bioinformatic Pipeline

To analyze the raw whole genome sequencing reads, we used a custom made, bioinformatic pipeline for our study organism following the GATK best practices method (Van der Auwera et al., 2013) (Figure 1). Because the whole genome analysis of each louse was computationally expensive and time consuming, we employed a GPU-accelerated software system utilizing NVIDIA RTX6000 GPU nodes on University of Florida's high performance computing cluster, HiPerGator, to run the software suite, Clara Parabricks (https://github.com/clara-parabricks) and significantly accelerated the processing time. The first 3 steps of the pipeline: reference mapping, coordinate sorting and mark duplicates were performed in Parabricks under BWA-MEM (H. Li, 2013) and GATK4 (McKenna et al., 2010; see Supplementary Materials for parameter details) .



Post-process Filtering

The resulting SNP dataset consisted of 19,256,706 SNPs. To ensure the quality of our dataset, we applied

additional filters to the VCF (Variant Calling Format) file. First, we removed all the singleton SNPs from our dataset (as these carry no useful information to address our main questions). Singleton detection was carried in vcftools under –singletons and the variants were extracted from the dataset. We then analyzed the dataset in vcftools (–freq2, –depth, –site-mean-depth, –site-quality, –missing-individuals, –missing-site, –het) to identify additional quality filters. Based on the results, we applied the following filters using vcftools; –minQ 20 (minimum quality score of 20), –min-meanDP 10 (minimum read depth), –max-meanDP 50 (maximum read depth), –maf <0.05 (minor allele frequency cutoff at 0.05). We then used PLINK v2.0 (Chang et al., 2015) to apply "missing data" filters; –mind 0.2 (filter out individuals that have >20% missing data), –geno 0.2 (filter out variants that have >20% missing data). The resulting dataset consisted of 575,096 biallelic SNPs. This SNP set was tested for Linkage Disequilibrium (LD) to remove SNPs that were in complete or in very high LD. We used PLINK –indep-pairwise with a window of 50kbp, sliding by 10kbp with a $r^2 < 0.7$. This resulted in a final set of 366,078 SNPs for all downstream analyses.

Population-specific Summary Statistics

To better understand the genetic diversity and demographics of the worldwide head louse populations, we first calculated summary statistics. We grouped our individuals by continent and calculated the nucleotide diversity (pi) and Tajima's D values from SNPs across the entire nuclear genome using the sliding window approach using the R package *PopGenome* (Pfeifer et al., 2014). We then calculated SNP heterozygosity for each country (observed;H_o and expected heterozygosity;H_e), to identify populations and loci that significantly deviated from Hardy Weinberg Equilibrium (HWE). By grouping individuals into their respective countries, we considered a spatial based population substructure to account for the Wahlund effect (Wahlund,1928) that would otherwise have an impact on the countrywide HWE calculations. All locus-by-locus HWE analyses were done using Arlequin 3.5 (Excoffier & Lischer, 2010) with 1,000,000 steps in Markov chain after 100,000 steps discarded as burnin. We also calculated statistical significance of the deviations at p <0.05 and proportion of loci that significantly deviated from HWE for each population. In addition to SNP heterozygosity, we also calculated autosomal heterozygosity for each country which indicates the proportion of heterozygous sites in each country. SNP heterozygosity and autosomal heterozygosity differ in that both polymorphic and monomorphic sites are included in autosomal heterozygosity calculations. Polymorphic sites are those which at least two individuals in each population had in common. We eliminated the monomorphic sites for our HWE analyses because they provided no useful information about the evolutionary forces acting upon the population.

To examine the distribution of genetic variability within and among groups and populations, we ran an Analysis of Molecular Variance (AMOVA) by using Arlequin 3.5 (Excoffier & Lischer, 2010). Hierarchical groups were defined as follows: samples were first grouped into their respective populations (by countries) and then they were further grouped into the five geographic continents (North America, South America, Asia, Europe, Africa, and Oceania. The significance of the variance components was tested using 1000 permutations. Lastly, we calculated the pairwise genetic differentiation (F_{ST}) between all pairs within and between countries and continents. The analysis was run in Arlequin, and significant genetic differentiation was determined after 1000 permutations.

Population Structure Analyses

We employed several methods to investigate head louse genetic population structure, we first conducted a Principal Component Analysis (PCA) using the R package *SNPRelate* (Zheng et al., 2012). Secondly, a Discriminant Analysis of Principal Components (DAPC) was performed using the R package *Adegenet* (Jombart, 2008). Unlike PCA, which summarizes the overall variability among individuals, DAPC maximizes the between group variation while minimizing within group variation (Jombart et al., 2010). In the first DAPC analysis, we used an a priori population assignment by continents and used these as the population identifiers. In the second analysis, we used the find.cluster() method without prior population assignment. Both DAPC analyses were run with 500 Principal Components (PC) and retained the first 5 discriminant functions. Optimum.a.score() was used to assess the optimum number of PCs to retain after 100 simulations (Figure S2). We tested K values from 1 to 50 and selected the optimum number of K based on the lowest

Bayesian Information Criterion (BIC) for each value of K (Figure S1). Results for both PCA and DAPC were visualized in a scatterplot in R.

We used a model-based Bayesian clustering analysis, fastSTRUCTURE v 1.0 (Raj et al., 2014) to infer ancestral population structure. We ran 10 iterations each for K=1 through K=10 with a 10-fold cross-validation and selected the optimum K value through the "choose K" method. All the input files were generated using PLINK v 2.0 (Chang et al., 2015) and the results were visualized using CLUMPAK (Kopelman et al., 2015) and DISTRUCT (Rosenberg, 2004). We also used the program ADMIXTURE to implement a maximum-likelihood approach to infer population structure. ADMIXTURE was run using the cross-validation flag (–cv) to select the best K with the lowest cv error. Additional ADMIXTURE analyses were run on each major genetic cluster that resulted from the main population structure analyses through fastSTRUCTURE that included all samples. For the population sub-structure analysis, individuals were selected if their genetic ancestry was greater than 0.8 in the resulting q matrix, for a given cluster.

Gene flow and phylogeny

To detect gene flow and possible migration events between louse populations, we used TREEMIX (Pickrell & Pritchard, 2012) to build a population tree and add on likely migration events. TREEMIX can model gene flow between populations and its direction by comparing the covariance modeled by the bifurcating tree to the observed covariance between populations. We ran the TREEMIX analysis with m=1 to m=14 migration events with four replicates for each migration edge (m) with varying SNP window sizes. We assessed the variance explained by the models, standard error and likelihood scores using the Evanno method (Evanno et al., 2005) implemented in the *optM* R package (Fitak, 2021). Delta "m" was calculated for each run and the optimum number of migration edges that fit the data were chosen accordingly. A summary of the percentage of variance explained by each model, significance of the migration events, and the associated likelihood scores can be found in Figure S3.

Finally, to evaluate the evolutionary relationships among the lice from the countries in our dataset, we constructed a neighbor-joining tree that uses Euclidean genetic distances for each population using the R package *dartR*(Gruber et al., 2018). To root the phylogenetic tree with an outgroup and evaluate ancestral populations, we identified SNPs from 16 chimpanzee lice (*Pediculus schaeffi*) using the body louse genome as the reference (Kirkness et al., 2010) and following the same bioinformatic pipeline as outlined in Figure 1. The resulting dataset underwent the same post-process filtering steps as the head lice (see methods above). The final filtered SNPs from both species were intersected using bcftools -isec to identify variants that were common to both species. The final dataset for the phylogenetic analysis including the outgroup contained 889 variants. The resulting neighbor-joining tree using these 889 SNPs was imported into Figtree for visualization.

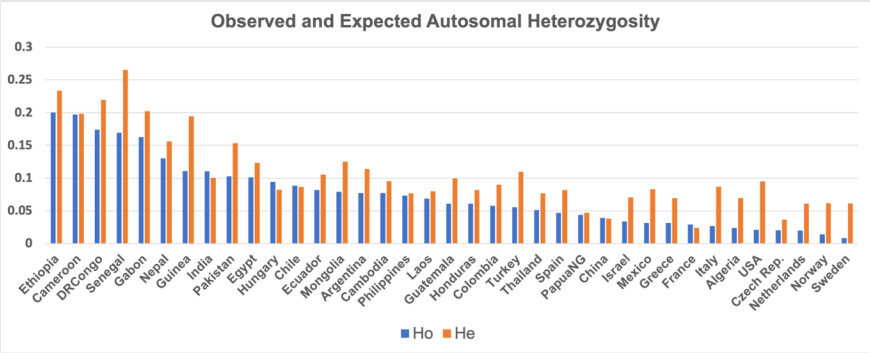
Results

Population specific summary statistics

The nucleotide diversity (pi) of lice from each continent revealed African lice had the highest diversity (0.318) compared to all other non-African lice. Asian lice had the second highest diversity (0.175) followed by South America (0.125), North America (0.110), Europe (0.099) and Oceania (0.073). Hardy Weinberg Equilibrium (HWE) tests using SNP heterozygosity revealed a pattern where most of the European and North American countries had a higher percentage of loci that were significantly out of HWE compared to Asian and African countries (Table 2). Each country had a different number of polymorphic sites. Noticeably, African countries had the highest number of polymorphic sites. Several countries had all polymorphic loci in HWE, which was unexpected. Each of these were instances where the lice used are suspected of coming either from a single human host or from a relatively few. Based on the Hardy-Weinberg Equilibrium (HWE) calculations, whether a specific locus was in/out of HWE seems to be a population-specific phenomenon rather than a locus-specific event. The fraction of loci that were out of HWE was different in each population. Summarized HWE results can be found in Table 2.

Table 2: Population summary statistics by country. Columns include major geographic regions, country, number of individuals, total number of polymorphic SNPs in each population, total SNPs that are significantly out of HWE, mean observed SNP heterozygosity (H_o), mean expected SNP heterozygosity (H_e) and percentage of SNPs that deviates from HWE. SNP heterozygosity mean values were calculated per country. Countries in each region are arranged by percentage out of HWE from high to low where it was calculated as, number of loci with$p$ <0.05/total polymorphic loci*100.

Region	Population (Country)	Number of individuals	Total polymorphic SNPs	Total SNPs
Europe	Sweden	10	62,330	41,877
	UK	23	100,938	66,416
	Italy	14	103,001	66,596
	Netherlands	9	68,826	22,808
	Spain	10	100,419	16,432
	France	5	23,974	1,248
	Norway	3	53,975	0
	Greece	4	66,624	0
	Czech Republic	3	31,291	0
	Hungary	3	78,039	0
North America	USA	48	145,883	120,077
	Mexico	15	101,635	54,292
Central America	Guatemala	12	121,137	26,900
	Honduras	13	90,922	13,998
	Panama	4	76,950	0
South America	Argentina	18	166,011	38,168
	Colombia	5	93,602	7,521
	Ecuador	5	115,355	4,794
	Chile	2	75,427	0
West Asia	Turkey	16	153,718	56,627
	Israel	9	79,396	21,006
South Asia	Pakistan	12	181,009	34,422
	Nepal	6	171,734	6,383
	India	3	94,357	0
Southeast Asia	Cambodia	19	137,931	19,388
	Philippines	9	87,813	3,439
	Laos	4	81,663	0
	Thailand	3	71,761	0
Oceania	Papua New Guinea	10	69,088	1,576
	Cook Islands	3	62,413	0
East Asia	Mongolia	7	147,135	14,560
	China	4	38,980	0
North Africa	Algeria	8	85,068	21,205
	Egypt	3	115,388	0
Sub-Saharan Africa	Senegal	16	265,774	65,235
	D.R.Congo	12	221,449	29,496
	Ethiopia	11	239,829	16,785
	Gabon	6	194,737	12,751
	Guinea	4	179,765	0
	Cameroon (lice from Baka Pygmy)	3	182,616	0



Observed autosomal heterozygosity (including both polymorphic and monomorphic sites) varied across all populations. Sub-Saharan African countries had the highest heterozygosity values observed (Figure 2). Asian and South American countries had the second highest heterozygosity estimates, followed by North American and European countries (Figure 2). Across all the countries, the observed heterozygosity was lower than the expected heterozygosity for all populations. Tajima's D values for each continent ranged from -0.0001-0.0003, indicating no demographic changes within populations among continents (Figure S4).

Hierarchical population structure analysis (AMOVA) revealed significant variation partitioned at each population hierarchical level (Table 3). The variation among continents was 25%, among countries within continents was 21%, among individuals within countries was 24% and lastly the individual variation among all individuals was 29%. F_{ST} values between continents were also relatively high. The highest F_{ST} values were found between Africa vs non-African countries and Oceania vs Europe (0.441) (Table 4). Low F_{ST} values between Asia and other continents suggest that it is the least differentiated continent relative to others. We also observed the lowest F_{ST} values between geographically proximal continents (e.g.: between North and South America; between Asia and Oceania). The pairwise-F_{ST}values between countries further explains the overall continental differentiation patterns in Table 4.

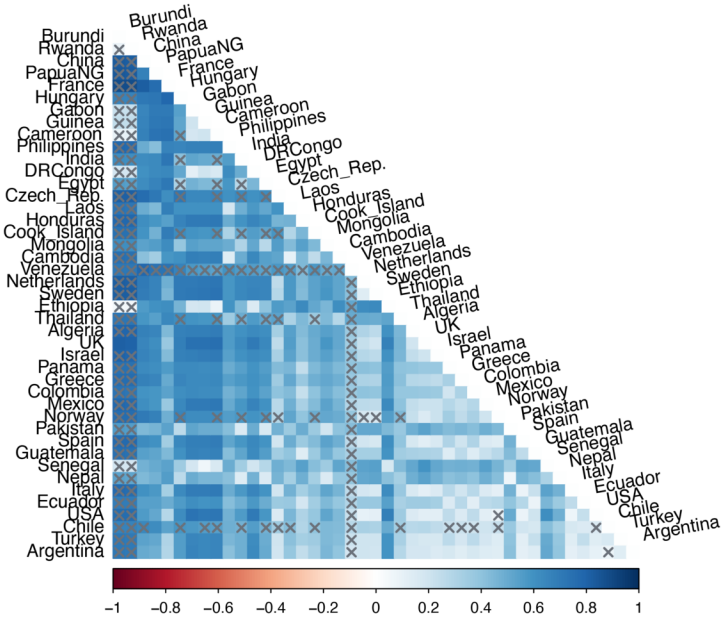
Table 3: Global AMOVA results as a weighted average across all polymorphic loci the SNP dataset. Significant p values ($p<0.00005$) are indicated by an asterisk (*)

Source of variance	Sum of squares	Variance components	Percentage variation (%)
Among continents	4087677.400	5560.54718	24.72825*
Among countries within continents	3672984.746	5373.26137	23.89538*
Among individuals within countries	5393862.265	4986.70023	22.17630*
Within individuals	2431033.500	6566.10712	29.20007*
Total	15585557.911	22489.08438	100

	North America	**Asia**	**Africa**	**Europe**	**South America**
Asia	0.174				
Africa	0.455	0.325			
Europe	0.061	0.202	0.458		
South A.	0.039	0.142	0.376	0.103	
Oceania	0.385	0.149	0.343	0.441	0.364

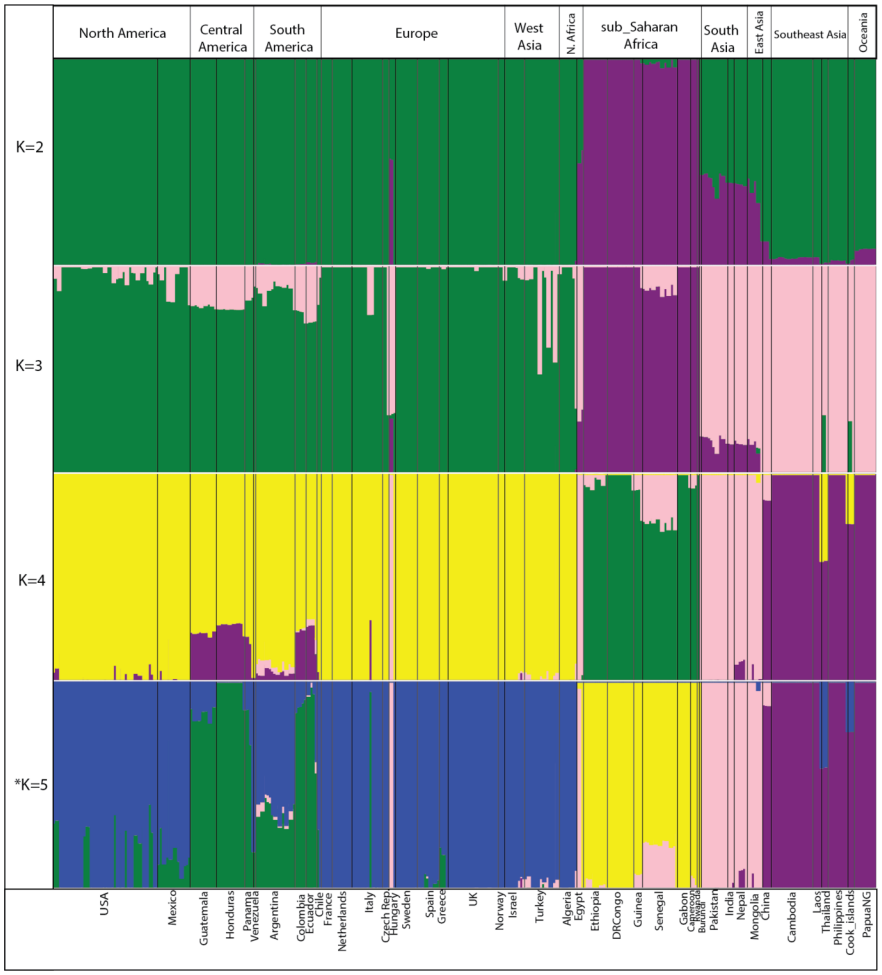
Between countries, some F_{ST} values were as high as 0.83, indicating strong genetic differentiation at the country level (Figure 3). Geographically proximal countries, that is, countries within a continent, and those that are in close proximity to one another, show less differentiation than those farther apart. One important

pattern that emerges as we examine country-level variation is that South Asian countries (India, Pakistan, and Nepal) and Mongolia seem to be genetically similar (low F_{ST} values) compared to African countries. This same pattern does not seem to be apparent when other countries are compared to African countries. Non-significant F_{ST} values ("X" marks on the heatmap; Figure 3) were between populations that had three or fewer samples.



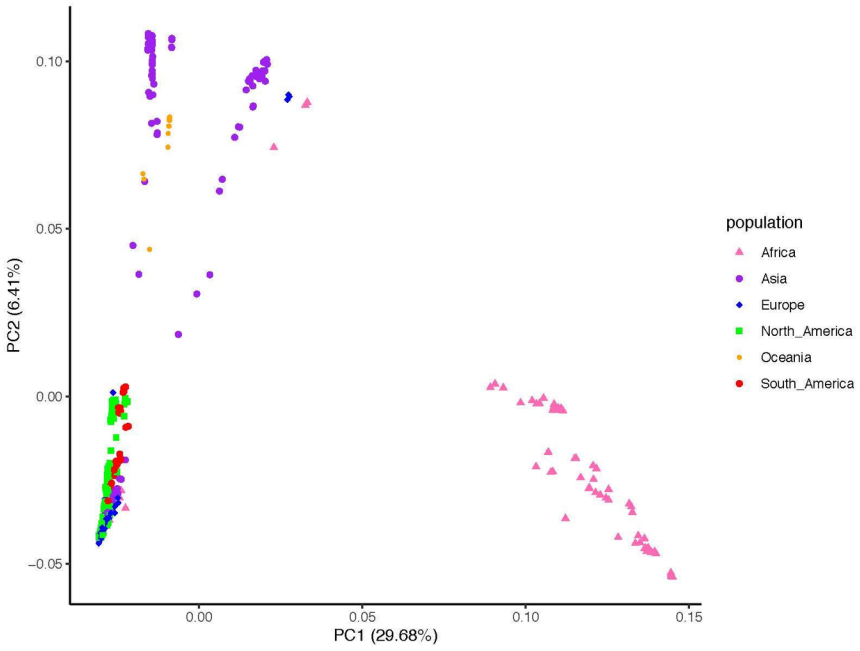
Population Structure

At K=2, fastSTRUCTURE results show a clear separation of lice collected in Africa from the rest of the regions, including a few admixed samples in Asia (Figure 4). As the number of K increases, the genetic clusters start to separate the samples by continents and further subdividing into sub-continents. Out of the 10 K-values tested, the optimum value of K for this dataset was K=5. The five genetic clusters (Figure 4) separate out individuals as follows: Sub-Saharan African (orange cluster), Southeast Asian and Oceanian (dark blue cluster), South Asia (green cluster), Central and South America (maroon cluster), and Europe with the majority of North America (light blue cluster). There are a few exceptions to these general categorizations of individuals. Some that stand out are samples from Hungary, Egypt, Mongolia, Turkey, and Israel. Both fastSTRUCTURE and ADMIXTURE analyses produced identical results. Therefore, we only report fastSTRUCTURE results on the overall genetic population structure.



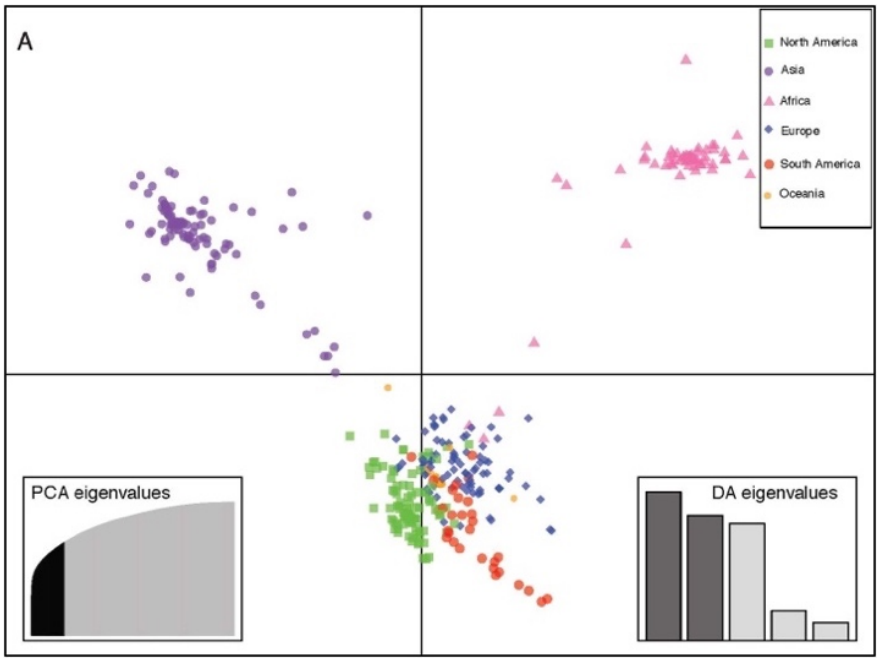
Principal Component Analysis

The principal component analysis showed similar population structure as the fastSTRUCTURE analysis. PC1 and PC2 explain 29.68% and 6.41% of variation, respectively, and separated the sub-Saharan African samples from the rest very clearly (Figure 5). In addition, we observed the same clustering patterns of lice in the fastSTRUCTURE results, including the same outliers (Figure 4). For example, in the PCA, Asia is subdivided into South Asia and Southeast Asia (See table 2 for details on countries that belong to each region), where the Southeast Asian cluster includes individuals from Oceania. Similarly, the main outliers detected in the fastSTRUCTURE analysis were Hungary, Mongolia and Egypt, clustering together with South Asian individuals, and all of the Algerian lice clustering with Europeans. These were also observed in the PCA clusters.



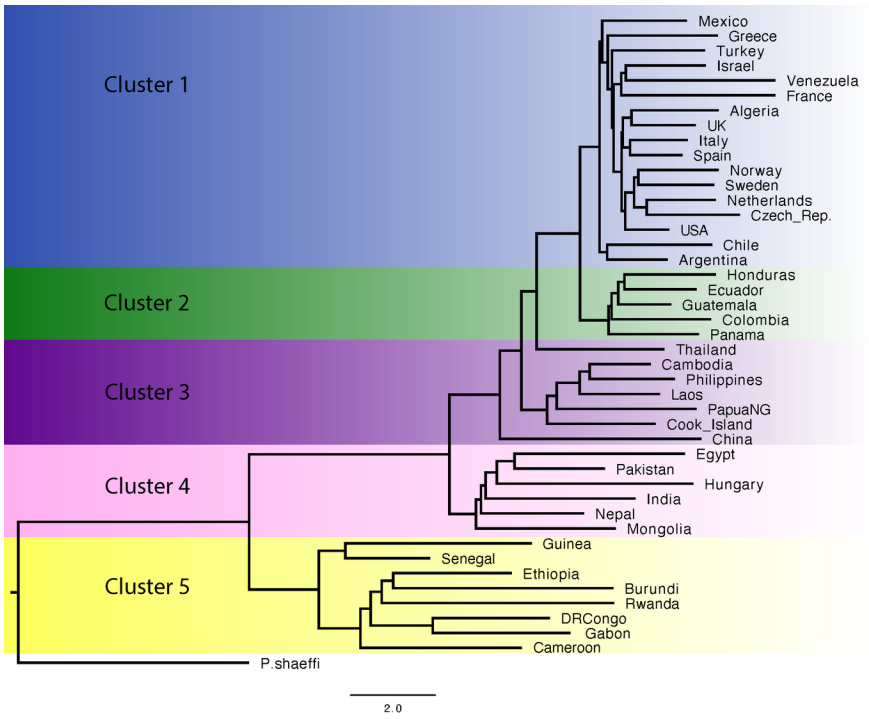
Discriminant Analysis of Principal Components (DAPC)

We ran a DAPC analysis to optimize the variation between populations while reducing within population variation. When grouping individuals by continent, we were able to detect the variation among samples from different continents (Figure 6A). Both the African and Asian groups were separated from the rest of the samples, indicating greater variation among those lice. Even though European, South + Central American and North American samples formed different clusters, the variation among those was small as compared to the African and Asian groups (Figure 6A). We also ran the DAPC analysis without adding any prior information about the sampling localities and tested putative genetic populations K=1 through K=50. We retained K=5 (Figure 6B) as the optimum K after using the find.clusters parameter and calculating the BIC scores for each K tested. This K=5 genetic clustering agrees with the results obtained from both the fastSTRUCTURE and PCA analyses, and the distribution of individuals (Figure 6B) was also similar to that of the PCA and fast STRUCTURE results. In addition, like the PCA results, discriminant component 1 separated the sub-Saharan African samples from the rest. North African (Algeria and Egypt) samples were separated from this main sub-Saharan African cluster

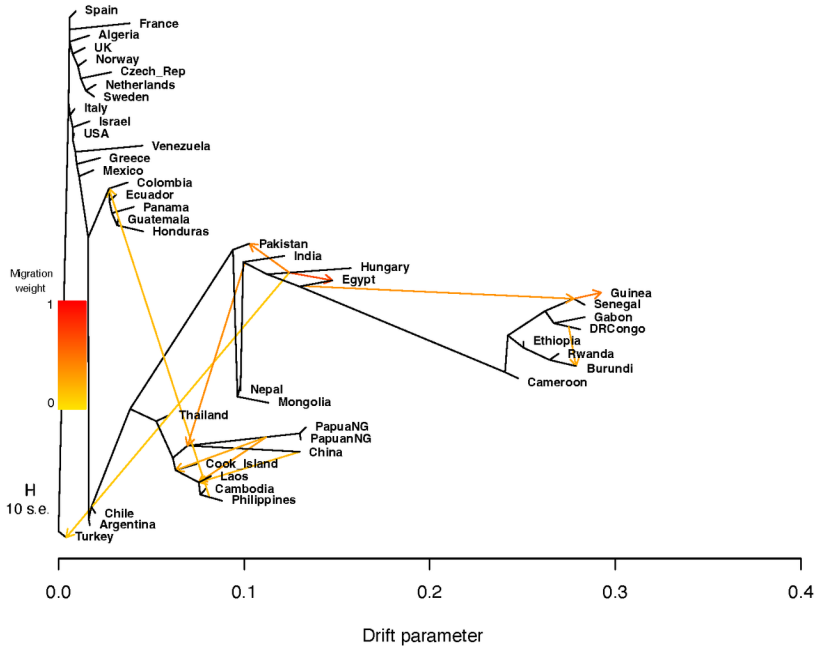


Phylogenetic relationships and Gene Flow

The neighbor-joining SNP phylogeny provided additional support for continental and regional division among the sampled populations (Figure 7). Results show that phylogenetic relationships among worldwide louse populations appear to be correlated with geography. With *Pediculus schaeffi* (chimpanzee louse) as an outgroup, sub-Saharan African populations were sister to all other populations.



Population relationships were further assessed by modeling migration events with TREEMIX given a number of migration edges from m=1 to m=14. Using the Evanno method we identified that 10 migration events best explained the sample covariance of allele frequencies among louse populations (Figure 8). Gene flow with the highest migration weight across continents was between Hungary and Egypt. The analysis also showed a second gene flow event between Hungary and Turkey; however, the migration weight was lower. Hungary and Egypt still appear to be an outlier (not geographically clustered) in the ML tree generated through TREEMIX. Another intercontinental gene flow event was found between the Philippines and Colombia. All other admixture events were between populations within continents (Figure 8). Overall, the population relationships between the TREEMIX ML tree and SNP phylogeny are in agreement with each other.



Discussion

Diversity and Population Structure

In this study, we focused on worldwide populations of head lice to understand their genetic diversity and distribution, which may help shed light on major events of their human host dispersal throughout the recent past. We investigated the worldwide population structure of human head lice using whole genome nuclear SNPs. Results of multiple clustering analyses agreed with an overall population structure that showed five genetically distinct nuclear clusters. In all our analyses that evaluated population structure, sub-Saharan African individuals separate out from the rest of the individuals when considering only two populations (Africa and the remainder of the world) up to as many as five genetic populations (Figure 4). This African cluster is also the most genetically diverse possessing a greater number of polymorphic sites compared to other non-African countries. Through comparison to an outgroup (the chimpanzee louse) we were able to confirm that this sub-Saharan African population is sister and basal to the rest of the non-African individuals (Figure 7).These results are consistent with Yong et al., 2003 who also found a clear geographic separation between African and non-African lice using partial nuclear genes 18S rRNA and EF1-alpha. Human genetics and history present a similar pattern, with a great deal of genetic diversity being found in sub-Saharan African populations (Campbell & Tishkoff, 2008). Some of these characteristics include highest nucleotide diversity, highest observed heterozygosity, and high percentage variation separating them out from all other populations. In human population genetics, all of these are characteristics of Africa being

the source population to all modern-day humans (J. Z. Li et al., 2008; Xing et al., 2010).

Upon closer investigation of the five nuclear genetic clusters that we uncovered in human head lice, we detected some global patterns arising that are similar to its host. In Asia + Oceania, lice split among two genetic clusters consisting of Southeast Asia and South Asian individuals (Figure 4). This geographical split between South and Southeast Asia is consistent with the southern expansion route proposed for human dispersals into Asia (Macaulay et al., 2005; Reyes-Centeno et al., 2014; Tassi et al., 2015). Southeast Asia further divided into four geographically structured sub-clusters comprised of a Thailand + Laos + Cambodia genetic cluster, and China, Papua New Guinea, and Philippines each forming separate genetic clusters (Figure S7), reflecting the fine scale population structure of humans in that area (Henn et al., 2010). In contrast, the South Asian genetic cluster includes lice from locations that are geographically separated from it (i.e., Mongolia, Hungary, and Egypt). The genetic affinity of the samples from Hungary and Egypt could be the result of sampling from recent immigrants or travelers (in both cases, the samples were obtained from a single individual). However, our TREEMIX analysis showed gene flow between Hungary and Egypt with a relatively high migration weight which was highly significant ($p<10^{-308}$). This gene flow event may explain the similarity between Hungary and Egypt. While the genetic affinity among these groups may be reflective of louse demographic history, further research is needed to better understand these relationships. Furthermore, due to the limited number of sampling in Hungary and Egypt, these samples may not necessarily reflect the genetic structure found across these countries and any interpretations taken from a sample derived from a single host should be done with caution.

The samples from Europe and the Americas showed little differentiation, grouping together in the PCA, DPAC and fastSTRUCTURE analyses. Nuclear diversity, heterozygosity, and F_{ST} values were similarly low between the Europe, North America, and South American samples, which is unexpected given the host population structure and dispersal history. In humans, African populations have the highest genetic diversity, followed by Europeans and Asians, with the lowest genetic diversity in indigenous American populations (Rosenberg et al., 2002). However, one key difference in host and parasite diversity patterns is the sampling strategies. In this study, lice were not collected from isolated ethnic groups or from aboriginal Americans like in human genetic diversity studies. Therefore, genetic similarity between European and American louse populations could be due to more recent gene flow (e.g., during European colonization of the Americas). Alternatively, the low genetic diversity and similarity of European and American (North and South America) lice could be due to selective pressures from insecticide use. The high use of pyrethroid insecticides to control louse infestations in Europe and the Americas (Diamantis et al., 2009) may have reduced the genetic variation among these populations but further investigation is needed, and is currently underway, to test this alternative explanation. In addition, it could also be the case that our lice were sampled from European descendants in cities outside of Europe (i.e in the Americas) . At K=5 clusters, a subset of the continental North American samples, primarily from Central America, separate out from the Europe + Americas cluster. The populations in this cluster also had moderate levels of observed heterozygosity (0.05-0.08), greater than in European and other American populations. It could be that these Central American louse populations may have experienced less insecticide exposure or the genetic variation may reflect earlier louse demographic history that could not be observed in the other American louse populations. For example, our TREEMIX analysis using allele frequency distributions shows a potential ancestral gene flow event between Philippines and the countries in this Central American cluster (Figure 8), suggesting inter-continental mixing at some point in time.

The geographically structured genetic clusters we uncovered here are concordant with previous findings examining microsatellite markers from eight localities around the world (Ascunce et al., 2013). Our current dataset adds information about African populations and how the genetic diversity of human head lice is distributed across the world by ancestral demographic events. Based on our analyses of population substructure, it is evident that the 5 major nuclear clusters that we uncovered are further subdivided into major regions within continents suggesting even more genetically structured louse populations (Figure S5-S9).

Parasite and Host Discordance

Our results found that much of the genetic structure in lice was similar to that of their human hosts, a few notable differences deserve further exploration. The hierarchical structure analysis (AMOVA) found high amounts of variation (>20%) across multiple spatial levels. This differs significantly from human genetic diversity where ~13% of variation is among continents, ~1% is among regions within continents and ~85% is within-population among regions (Jorde et al. 2000). These differences point to lice maintaining higher isolation than their hosts even with the ability of horizontal transfer compared to the strict vertical transfer in the host. Additional research is needed to understand the evolutionary forces acting upon local louse populations, but not on the host, to achieve these levels of louse variation and continental population genetic structure (High F_{ST}values between continents).

The host-mediated geographical structure of nuclear diversity in lice contrasts with the documented distribution of mitochondrial clades in lice. Three deeply divergent mitochondrial clades (A, B and C) that are not geographically well defined have been hypothesized to be associated with lice on now-extinct species of hominids (e.g., Neanderthals) that dispersed onto modern humans during periods of contemporaneity and survived to present day (Reed et al., 2004). In contrast, the more geographically structured nuclear clusters mirror the human host distribution following modern human dispersal out of Africa. This is consistent with Ascunce et al. (2013) where they found no relationship between the mitochondrial clade a given louse was assigned to and the nuclear genetic clusters in which the individual was found.

Mito-nuclear Discordance

Reasons for the mitonuclear discordance include the different inheritance patterns of mitochondrial and nuclear markers. Mitochondrial DNA (mtDNA) markers track the maternal lineage of a species whereas nuclear markers include DNA from both parents and undergo recombination, which adds additional levels of complexity (Rubinoff & Holland, 2005; Zink & Barrowclough, 2008). This discordance raises the question of mtDNA's power to resolve population level relationships. If mtDNA saturation is at play, which can happen due to the rapid evolutionary rate, mtDNA markers could be inaccurately signaling to this deep louse evolutionary history. This seems unlikely, however, because the primary endosymbiont of lice, *Candidatus Riesia pediculicola* , which is maternally inherited, shows similar distribution patterns to louse mitochondrial DNA (Hammoud et al., 2022) .

On the other hand, it might also be possible that the mito-nuclear discordance is resulting from the unusual gene transfer of human head lice where males preferentially eliminate their paternally inherited chromosomes (PGE- paternal genome elimination) (de la Filia et al., 2018). PGE coupled with sex ratio distortion of human head lice (Perotti et al., 2004) could potentially lead to sex-biased gene flow that is reflected through the differences in mitochondrial and nuclear phylogenies.

Nevertheless, the different views from both nuclear and mitochondrial DNA of lice suggest a rich evolutionary history of lice that we're only beginning to understand. Despite the differences in inheritance, both mtDNA and nuclear markers in human head lice broadly agree on an out of Africa dispersal, and the differences in mitochondrial and nuclear patterns between human and lice suggest that louse genetics may be capturing human demographic events that are not detectable from human genetics. Recent human dispersal across the world through global ease of travel might be contributing to the continental and regional patterns that we see through the nuclear genome of lice.

It has been suggested that within-species population structure of the parasites have either an equal or higher structure than its host (Criscione et al., 2006). If this was the case for humans and lice, what we can infer from the population structure of lice can potentially be applied towards understanding the history of humans. Therefore, we uncovered some patterns of genetic variation distribution across the world that can only be explained by the recent movements and settlements of its host.

Conclusion

All analyses concluded that sub-Saharan African individuals are ancestral and genetically distinct from all other non-African individuals. Sub-Saharan samples also had the highest heterozygosity mirroring its host.

Unlike the mitochondrial clades, nuclear genetic clusters of lice are highly structured based on geography (continental and major regions within continents). This mito-nuclear incongruence is possibly due to different evolutionary histories of the markers.

Although we do not know the exact timing of this genetic population structure that we observed, it seems possible that these five nuclear genetic clusters are a result of more recent events of human migration (e.g European colonization) and settlement across the world (Ascunce 2013) or it is reflective of major historical events in the human history such as anatomically modern humans moving out of Africa. Further investigations on the timing of these nuclear population splits and coalescent based approaches to investigate the most recent common ancestor are needed to confirm the host influence on shaping the genetic structure of this parasite.

Acknowledgements

We thank Mercedes Domingo and Francesca Reale from Pid free clinic Milano, Italy for sending us louse specimens. We also thank Dr. Marina Ascunce for her DNA extractions on some of the louse specimens that were included in this study. This study was funded by the National Science Foundation P0044766 to DLR and 2206735 and 1925312 to JMA.

Data Accessibility

Raw genomic variants used in this study are publicly available through Dryad (link XXX). Parameter specifications used for processing raw data can be found in the supplementary data file.

Author Contribution

D.L.R, A.M and N.H designed the study. A.M. prepared the genomic libraries. N.H performed the DNA extractions, conducted the analyses, and wrote the original draft. All other authors provided specimens for the study, and everyone contributed to reviewing and editing the final manuscript.

Competing Interests

The authors declare no competing interests.

References

Al-Shahrani, S. A., Alajmi, R. A., Ayaad, T. H., Al-Shahrani, M. A., & Shaurub, E.-S. H. (2017). Genetic diversity of the human head lice,*Pediculus humanus capitis* , among primary school girls in Saudi Arabia, with reference to their prevalence. *Parasitology Research* , *116* (10), Article 10. https://doi.org/10.1007/s00436-017-5570-3

Amanzougaghene, N., Fenollar, F., Davoust, B., Djossou, F., Ashfaq, M., Bitam, I., Raoult, D., & Mediannikov, O. (2019). Mitochondrial diversity and phylogeographic analysis of *Pediculus humanus*reveals a new Amazonian clade "F." *Infection, Genetics and Evolution* , *70* , 1–8.

Amanzougaghene, N., Mumcuoglu, K. Y., Fenollar, F., Alfi, S., Yesilyurt, G., Raoult, D., & Mediannikov, O. (2016). High Ancient Genetic Diversity of Human Lice, *Pediculus humanus* , from Israel Reveals New Insights into the Origin of Clade B Lice. *PLOS ONE* ,*11* (10), Article 10.*https://doi.org/10.1371/journal.pone.0164659*

Arriaza, B., Standen, V. G., Heukelbach, J., Cassman, V., & Olivares, F. (2014). Head combs for delousing in ancient Arican populations: Scratching for the evidence. *Chungara, Revista de Antropología Chilena* , *46* (4), 693–706.

Ascunce, M. S., Toups, M. A., Kassu, G., Fane, J., Scholl, K., & Reed, D. L. (2013). Nuclear genetic diversity in human lice (*Pediculus humanus*) reveals continental differences and high inbreeding among worldwide populations. *PLOS ONE* , *8* (2), 57619. https://doi.org/10.1371/journal.pone.0057619

Ashfaq, M., Prosser, S., Nasir, S., Masood, M., Ratnasingham, S., & Hebert, P. D. N. (2015). High diversity and rapid diversification in the head louse, *Pediculus humanus* (Pediculidae: Phthiraptera.*Scientific Reports* , *5* . https://doi.org/10.1038/srep14188

Boumbanda Koyo, C. S., Amanzougaghene, N., Davoust, B., Tshilolo, L., Lekana-Douki, J. B., Raoult, D., Mediannikov, O., & Fenollar, F. (2019). Genetic diversity of human head lice and molecular detection of associated bacterial pathogens in Democratic Republic of Congo.*Parasites & Vectors* , *12* (1), Article 1. https://doi.org/10.1186/s13071-019-3540-6

Burkhart, C. N., & Burkhart, C. G. (2007). Fomite transmission in head lice. *Journal of the American Academy of Dermatology* ,*56* (6), 1044–1047. https://doi.org/10.1016/j.jaad.2006.10.979

Campbell, M. C., & Tishkoff, S. A. (2008). African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. *Annual Review of Genomics and Human Genetics* , *9* (1), 403–433. https://doi.org/10.1146/annurev.genom.9.081307.164258

Carto, S. L., Weaver, A. J., Hetherington, R., Lam, Y., & Wiebe, E. C. (2009). Out of Africa and into an ice age: On the role of global climate change in the late Pleistocene migration of early modern humans out of Africa. *Journal of Human Evolution* , *56* (2), Article 2. https://doi.org/10.1016/j.jhevol.2008.09.004

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* ,*4* (1), Article 1. https://doi.org/10.1186/s13742-015-0047-8

Criscione, C. D., Cooper, B., & Blouin, M. S. (2006). Parasite genotypes identify source populations of migratory fish more accurately than fish genotypes. *Ecology* , *87* (4), 823–828. https://doi.org/10.1890/0012-9658(2006)87

de la Filia, A. G., Andrewes, S., Clark, J. M., & Ross, L. (2018). The unusual reproductive system of head and body lice *(Pediculus humanus*). *Medical and Veterinary Entomology* , n/a-n/a. https://doi.org/10.1111/mve.12287

Diamantis, S. A., Morrell, D. S., & Burkhart, C. N. (2009). Treatment of head lice. *Dermatologic Therapy* , *22* (4), 273–278. https://doi.org/10.1111/j.1529-8019.2009.01242.x

Ebert, D., & Fields, P. D. (2020). Host-parasite co-evolution and its genomic signature. *Nature Reviews. Genetics* , *21* (12), 754–768. https://doi.org/10.1038/s41576-020-0269-1

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* , *14* (8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Ewing, H. E. (1926). A revision of the American lice of the genus*Pediculus* , together with aconsideration of the significance of their geographical and host distribution. *Proceedings of the United States National Museum* , *68* (2620), Article 2620. https://doi.org/10.5479/si.00963801.68-2620.1

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* , *10* (3), Article 3. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fitak, R. R. (2021). OptM: Estimating the optimal number of migration edges on population trees using Treemix. *Biology Methods and Protocols* , *6* (1), bpab017. https://doi.org/10.1093/biomethods/bpab017

Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). dartr: An r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resources* ,*18* (3), 691–699. https://doi.org/10.1111/1755-0998.12745

Henn, B. M., Gravel, S., Moreno-Estrada, A., Acevedo-Acevedo, S., & Bustamante, C. D. (2010). Fine-scale population structure and the era of next-generation sequencing. *Human Molecular Genetics* ,*19* (R2),

R221–R226.*https://doi.org/10.1093/hmg/ddq403*

Hammoud, A., Louni, M., Missé, D., Cortaredona, S., Fenollar, F., & Mediannikov, O. (2022). Phylogenetic relationship between the endosymbiont "Candidatus Riesia pediculicola" and its human louse host. *Parasites & Vectors* , *15* (1), 73. https://doi.org/10.1186/s13071-022-05203-z

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* , *24* (11), Article 11. https://doi.org/10.1093/bioinformatics/btn129

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics* , *11* (1), Article 1.*https://doi.org/10.1186/1471-2156-11-94*

Jorde, L. B., Watkins, W. S., Bamshad, M. J., Dixon, M. E., Ricker, C. E., Seielstad, M. T., & Batzer, M. A. (2000). The Distribution of Human Genetic Diversity: A Comparison of Mitochondrial, Autosomal, and Y-Chromosome Data. *The American Journal of Human Genetics ,66* (3), 979–988. https://doi.org/10.1086/302825

Kirkness, E. F., Haas, B. J., Sun, W., Braig, H. R., Perotti, M. A., Clark, J. M., Lee, S. H., Robertson, H. M., Kennedy, R. C., Elhaik, E., Gerlach, D., Kriventseva, E. V., Elsik, C. G., Graur, D., Hill, C. A., Veenstra, J. A., Walenz, B., Tubío, J. M. C., Ribeiro, J. M. C., … Pittendrigh, B. R. (2010). Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proceedings of the National Academy of Sciences* , *107* (27), Article 27. https://doi.org/10.1073/pnas.1003379107

Kittler, R., Kayser, M., & Stoneking, M. (2003). Molecular Evolution of*Pediculus humanus* and the Origin of Clothing. *Current Biology* , *13* (16), Article 16. https://doi.org/10.1016/S0960-9822(03)00507-4

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources* , *15* (5), Article 5. https://doi.org/10.1111/1755-0998.12387

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv:1303.3997 [q-Bio]* . http://arxiv.org/abs/1303.3997

Li, J. Z., Absher, D. M., Tang, H., Southwick, A. M., Casto, A. M., Ramachandran, S., Cann, H. M., Barsh, G. S., Feldman, M., Cavalli-Sforza, L. L., & Myers, R. M. (2008). Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation.*Science* , *319* (5866), 1100–1104. https://doi.org/10.1126/science.1153717

Macaulay, V., Hill, C., Achilli, A., Rengo, C., Clarke, D., Meehan, W., Blackburn, J., Semino, O., Scozzari, R., Cruciani, F., Taha, A., Shaari, N. K., Raja, J. M., Ismail, P., Zainuddin, Z., Goodwin, W., Bulbeck, D., Bandelt, H.-J., Oppenheimer, S., … Richards, M. (2005). Single, Rapid Coastal Settlement of Asia Revealed by Analysis of Complete Mitochondrial Genomes. *Science* , *308* (5724), 1034–1036. https://doi.org/10.1126/science.1109792

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data.*Genome Research* , *20* (9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Mumcuoglu, K. Y., Pollack, R. J., Reed, D. L., Barker, S. C., Gordon, S., Toloza, A. C., Picollo, M. I., Taylan-Ozkan, A., Chosidow, O., Habedank, B., Ibarra, J., Meinking, T. L., & Vander Stichele, R. H. (2021). International recommendations for an effective control of head louse infestations. *International Journal of Dermatology ,60* (3), 272–280. https://doi.org/10.1111/ijd.15096

Pedersen, M. W., Antunes, C., De Cahsan, B., Moreno-Mayar, J. V., Sikora, M., Vinner, L., Mann, D., Klimov, P. B., Black, S., Michieli, C. T., Braig, H. R., & Perotti, M. A. (2022). Ancient Human Genomes

and Environmental DNA from the Cement Attaching 2,000-Year-Old Head Lice Nits. *Molecular Biology and Evolution* , *39* (2), msab351. https://doi.org/10.1093/molbev/msab351

Perotti, M. A., Catalá, S. S., Ormeño, A. del V., Żelazowska, M., Biliński, S. M., & Braig, H. R. (2004). The sex ratio distortion in the human head louse is conserved over time. *BMC Genetics* , *5* , 10. https://doi.org/10.1186/1471-2156-5-10

Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: An efficient Swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution* , *31* (7), Article 7. https://doi.org/10.1093/molbev/msu136

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLOS Genetics* , *8* (11), Article 11. https://doi.org/10.1371/journal.pgen.1002967

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational inference of population structure in large SNP data sets.*Genetics* , *197* (2), Article 2.

Raoult, D., Reed, D. L., Dittmar, K., Kirchman, J. J., Rolain, J.-M., Guillen, S., & Light, J. E. (2008). Molecular Identification of Lice from Pre-Columbian Mummies. *The Journal of Infectious Diseases* ,*197* (4), Article 4. https://doi.org/10.1086/526520

Reed, D. L., Light, J. E., Allen, J. M., & Kirchman, J. J. (2007). Pair of lice lost or parasites regained: The evolutionary history of anthropoid primate lice. *BMC Biology* , *5* , 7. https://doi.org/10.1186/1741-7007-5-7

Reed, D. L., Smith, V. S., Hammond, S. L., Rogers, A. R., & Clayton, D. H. (2004). Genetic Analysis of Lice Supports Direct Contact between Modern and Archaic Humans. *PLoS Biology* , *2* (11), Article 11. https://doi.org/10.1371/journal.pbio.0020340

Reich, D., Patterson, N., Kircher, M., Delfin, F., Nandineni, M. R., Pugach, I., Ko, A. M.-S., Ko, Y.-C., Jinam, T. A., Phipps, M. E., Saitou, N., Wollstein, A., Kayser, M., Pääbo, S., & Stoneking, M. (2011). Denisova Admixture and the First Modern Human Dispersals into Southeast Asia and Oceania. *The American Journal of Human Genetics* , *89* (4), 516–528. https://doi.org/10.1016/j.ajhg.2011.09.005

Reyes-Centeno, H., Ghirotto, S., Détroit, F., Grimaud-Hervé, D., Barbujani, G., & Harvati, K. (2014). Genomic and cranial phenotype data support multiple modern human dispersals from Africa and a southern route into Asia. *Proceedings of the National Academy of Sciences* ,*111* (20), 7248–7253. https://doi.org/10.1073/pnas.1323666111

Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., & Reich, D. (2015). Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of the Royal Society B: Biological Sciences* , *370* (1660), 20130624. https://doi.org/10.1098/rstb.2013.0624

Rohland, N., & Reich, D. (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research* , *22* (5), Article 5. https://doi.org/10.1101/gr.128124.111

Rosenberg, N. A. (2004). distruct: A program for the graphical display of population structure. *Molecular Ecology Notes* , *4* (1), Article 1. https://doi.org/10.1046/j.1471-8286.2003.00566.x

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., & Feldman, M. W. (2002). Genetic Structure of Human Populations. *Science* , *298* (5602), 2381–2385.*https://doi.org/10.1126/science.1078311*

Rubinoff, D., & Holland, B. S. (2005). Between Two Extremes: Mitochondrial DNA is neither the Panacea nor the Nemesis of Phylogenetic and Taxonomic Inference. *Systematic Biology* , *54* (6), 952–961. https://doi.org/10.1080/10635150500234674

Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., Patterson, N., & Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature* ,*507* (7492), 354–357. https://doi.org/10.1038/nature12961

Slatkin, M. (1987). Gene Flow and the Geographic Structure of Natural Populations. *Science* , *236* (4803), 787–792.

Stringer, C. (2000). Coasting out of Africa. *Nature* ,*405* (6782), 25–27. https://doi.org/10.1038/35011166

Tassi, F., Ghirotto, S., Mezzavilla, M., Vilaça, S. T., De Santi, L., & Barbujani, G. (2015). Early modern human dispersal from Africa: Genomic evidence for multiple waves of migration. *Investigative Genetics* ,*6* (1), 13. https://doi.org/10.1186/s13323-015-0030-2

Toups, M. A., Kitchen, A., Light, J. E., & Reed, D. L. (2011). Origin of Clothing Lice Indicates Early Clothing Use by Anatomically Modern Humans in Africa. *Molecular Biology and Evolution* , *28* (1), Article 1. https://doi.org/10.1093/molbev/msq234

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline.*Current Protocols in Bioinformatics* , *43* (1), 11.10.1-11.10.33. https://doi.org/10.1002/0471250953.bi1110s43

Villanea, F. A., & Schraiber, J. G. (2019). Multiple episodes of interbreeding between Neanderthal and modern humans. *Nature Ecology & Evolution* , *3* (1), Article 1.*https://doi.org/10.1038/s41559-018-0735-8*

Wahlund, S. (1928). Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre aus betrachtet. *Hereditas* , *11* (1), 65-106.

Walter, R. C., Buffler, R. T., Bruggemann, J. H., Guillaume, M. M. M., Berhe, S. M., Negassi, B., Libsekal, Y., Cheng, H., Edwards, R. L., von Cosel, R., Néraudeau, D., & Gagnon, M. (2000). Early human occupation of the Red Sea coast of Eritrea during the last interglacial.*Nature* , *405* (6782), 65–69. https://doi.org/10.1038/35011048

Xing, J., Watkins, W. S., Shlien, A., Walker, E., Huff, C. D., Witherspoon, D. J., Zhang, Y., Simonson, T. S., Weiss, R. B., Schiffman, J. D., Malkin, D., Woodward, S. R., & Jorde, L. B. (2010). Toward a more uniform sampling of human genetic diversity: A survey of worldwide populations by high-density genotyping. *Genomics* , *96* (4), 199–210. https://doi.org/10.1016/j.ygeno.2010.07.004

Yong, Z., Fournier, P.-É., Rydkina, E., & Raoult, D. (2003). The geographical segregation of human lice preceded that of *Pediculus humanus capitis* and *Pediculus humanus humanus* .*Comptes Rendus Biologies* , *326* (6), Article 6. https://doi.org/10.1016/S1631-0691(03)00153-7

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* ,*28* (24), Article 24.*https://doi.org/10.1093/bioinformatics/bts606*

Zink, R. M., & Barrowclough, G. F. (2008). Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology* , *17* (9), 2107–2121. https://doi.org/10.1111/j.1365-294X.2008.03737.x