# Nuclear genetic diversity of head lice sheds light on human dispersal around the world

Marina S. Ascunce[1,2,*], Ariel C. Toloza[3], Angélica González-Oliver[4], and David L. Reed[5]

[1]Emerging Pathogens Institute, Department of Plant Pathology, University of Florida, Gainesville, Florida, United States of America

[2]USDA-ARS Center for Medical, Agricultural, and Veterinary Entomology, Gainesville, Florida, United States of America

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

[4]Departamento de Biología Celular, Facultad de Ciencias, Universidad Nacional Autónoma de México, Ciudad de México, México

[5]Florida Museum of Natural History, University of Florida, Gainesville, Florida, United States of America

[*]Corresponding author E-mail: marina.ascunce@usda.gov (MSA)

# Abstract

The human louse, *Pediculus humanus*, is an obligate blood-sucking ectoparasite that has coevolved with humans for millennia. Because of the intimate relationship between this parasite and the human host, the study of human lice has the potential to shed light on aspects of the human evolution that are obscured or difficult to interpret using other biological evidence. In this study, we analyzed the genetic variation in 274 human lice from 25 geographic sites around the world by using nuclear microsatellite loci and female-inherited mitochondrial DNA sequences. Nuclear genetic diversity analysis revealed the presence of two distinct genetic clusters I and II, which are subdivided into subclusters: Ia-Ib and IIa-IIb, respectively. Among these samples, we observed the presence of the two most common louse mitochondrial haplogroups: A and B. Both mitochondrial haplotypes were found in nuclear Clusters I and II. Evidence of nuclear admixture was uncommon (33 lice) and was predominately found in the New World potentially mirroring the pattern of very recent host admixture between Native Americans and Europeans. These findings were supported by novel DIYABC-simulations that were built using both host and parasite data to define parameters and models. In addition to providing new evolutionary knowledge about this human parasite, our study could guide the development of new analyses in other host-parasite systems.

# Introduction

The human louse is a single insect species composed of two distinct morphological, behavioral, and ecological types: the human head louse (*Pediculus humanus capitis* De Geer), and the clothing or body louse (*Pediculus humanus humanus* L.) [1]. Human lice (Phthiraptera: Anoplura) are blood-sucking, wingless, host-specific ectoparasites of humans that are both obligate (cannot live off the host) and permanent (complete their life cycle on a single host). Thus, human lice are inextricably tied to their host in ecological and evolutionary time. The coevolution of these parasites and their primate hosts over the last 25 million years (MY) is very well documented [2,3]. Cophylogenetic analyses revealed that human and chimpanzee lice in the genus *Pediculus* diverged ca. 5.6 MY ago (MYA), which is contemporaneous with the known divergence of chimps and humans [4]. Human head lice have been found in archeological human remains, including a head louse nit obtained from human remains in northeastern Brazil dated 10,000 years old [5] and a nit from Israel dated as 9,000 years old [6] (Fig 1). In host-parasite systems, these long and intimate associations lead to reciprocal adaptations in the hosts and parasites (coevolution), which allows one to identify events of the same age in the host and parasite phylogenies, and provide an internal time calibration for comparative studies of rates of evolution in the two groups [7]. Thus, in cases where parasites coevolve with their hosts, parasite phylogenies will, to some extent, reflect host evolutionary history [8–10]. Parasites have some biological features that differ from their hosts, such as faster evolutionary rates and larger population sizes in parasites, which allows them to maintain greater levels of genetic diversity than their hosts. These attributes of human parasites have provided insights into aspects of human evolution that were unclear from studies of direct human evidence, such as fossil or molecular data [10–13]. Evolutionary studies of human lice have already provided insights about the ecology of early humans, such as when humans started

56 using clothes [14,15] and how humans travelled around the globe [16–18]. Recently, human DNA
57 was extracted from nits found on hairs from ancient mummies in South America revealing
58 ancestral migration routes [19]. Moreover, almost 20 years ago, Reed et al. [2] described highly
59 divergent human louse mitochondrial clades and hypothesized that each louse mitochondrial
60 lineage evolved on different hominin hosts. The authors suggested that as anatomically modern
61 humans (AMH) migrated out of Africa and interacted with other hominis, those different lineages
62 of lice were transferred among hominins through direct physical contact [2]. This hypothesis of
63 direct contact between modern and archaic hominins predicted using human louse data was later
64 confirmed through host genomic studies comparing ancient DNA genomes from hominid remains
65 with AMH genomes [20].

66 **Fig 1. Worldwide geographic distribution of collected modern lice (green dots).**
67 Archeological findings of human lice are shown with the figure of human lice on the map and
68 the corresponding estimated dates from: [3,5,6,22,100]. In addition, the map reflects the
69 proposed distribution of hominids: The approximate distribution and locations with dates of
70 known Neanderthal (Blue), Denisovan (Black), and Anatomical Modern Humans (Orange) fossil
71 remains based on: [101−116].

72 The genetic diversity of human lice has been widely studied based on the mitochondrial DNA.
73 However, we know little about the nuclear genetic diversity of this human parasite, particularly at
74 a worldwide scale. The first mitochondrial studies revealed the presence of several deeply
75 divergent mtDNA clades or haplogroups named A, B and C [2,3,14,17,21]. Haplogroup A is the
76 most common of the human louse haplogroups, has a worldwide distribution and has been found
77 in lice from human archeological remains in Israel from 2,000 years ago [22]. Haplogroup B has
78 been found in the Americas, Europe, Australia, North and South Africa [17,21,23,24], as well in

79 ancient human remains in Israel [22] and Peru [3]. Haplogroup C has a more geographically
80 restricted distribution and is only present in lice found in Nepal, Ethiopia, Republic of Congo,
81 Pakistan and Thailand [14,22,25]. Finally, recent studies have suggested the presence of three
82 more clades: D, E and F [24,26,27]. Clade D is sister to clade A found in Ethiopia, Zimbabwe and
83 Republic Democratic of Congo [24,26,28]. Clade E is sister to clade C and was detected in head
84 lice from Senegal and Mali [24]. Clade F appears to be specific to South America, as it has not
85 been reported elsewhere [27]. Head lice can be found in all the mitochondrial clades (A-F),
86 whereas clothing lice are only found in clades A and D [28]. Because of the haploid nature and
87 maternal inheritance of the mitochondrial DNA, events of hybridization cannot be detected with
88 this marker, thus new nuclear markers were developed [16] to provide a comprehensive
89 understanding of human louse genetic diversity and to assess admixture among human lice.
90 In our previous study using the same microsatellite markers with fewer samples [16], we
91 found that nuclear genetic diversity in human lice revealed continental differences and suggested
92 that the genetic pattern found in the Americas could be mirroring the (human host) colonization
93 of the New World represented by two major human settlements of the New World: the first
94 peopling of America and the European colonization after Columbus. In this new study we
95 expanded the sampling of lice to include more localities to deepen our understanding of human
96 louse nuclear genetic diversity while also assessing its correlation with mitochondrial diversity.
97 For this study, we examined 15 microsatellite loci and a mitochondrial gene (COX1) for nearly
98 300 individual head lice collected from 25 sites distributed throughout 10 main geographic
99 regions (Table 1, S1 Table, Fig 1). Furthermore, we included an new analysis by using novel
100 DIYABC-simulations that were built using both host and parasite data to define parameters and
101 models. In addition to providing new evolutionary knowledge about this human parasite, our

102 study could guide the development of new analyses in other host-parasite systems.

103 **Table 1. Microsatellite polymorphisms among human louse populations.** Groups reflects
104 their geographic origin and STRUCTURE clustering: Africa-Israel-Spain (AIS), South and
105 Southeast Asia (SAs), East Asia (EAs), Europe (E), North America (NA), Mesoamerica (MA),
106 Central America (CA), and South America (SAm).

| Geographic Region | Group | Country | Site | N | *Na* | $H_O$ | $H_E$ | $F_{IS}$ | *P-value* |
|---|---|---|---|---|---|---|---|---|---|
| Africa | AIS | Rwanda | Rw | 1 | 1.533 | n/a | n/a | n/a | - |
| West Asia | AIS | Israel | Isr | 2 | 1.667 | n/a | n/a | n/a | - |
| Central Asia | SAs | Nepal | Np | 1 | 1.667 | n/a | n/a | n/a | - |
| South and Southeast Asia | SAs | Thailand | Th | 1 | 1.333 | n/a | n/a | n/a | - |
| | SAs | Cambodia | Cam | 22 | 6.8 | 0.483 | 0.703 | 0.319 | *0.000* |
| | SAs | Laos | Ls | 4 | 3.2 | n/a | n/a | n/a | - |
| East Asia | EAs | Mongolia | Mn | 10 | 5 | 0.524 | 0.706 | 0.267 | *0.005* |
| Europe | E | Turkey | Tur | 12 | 3 | 0.232 | 0.408 | 0.407 | *0.000* |
| | E | Croatia | Croa | 6 | 1.133 | n/a | n/a | n/a | - |
| | E | Norway | Nw | 1 | 1.067 | n/a | n/a | n/a | - |
| | E | Netherlands | Neth | 64 | 2.267 | 0.032 | 0.248 | 0.863 | *0.000* |
| | E | France | Fr | 3 | 1.133 | n/a | n/a | n/a | - |
| | E | United Kingdom | UK | 27 | 2.533 | 0.145 | 0.329 | 0.550 | *0.000* |
| | AIS | Spain | Spain | 3 | 2.067 | n/a | n/a | n/a | - |
| North America | NA | USA | Oce | 25 | 2.8 | 0.083 | 0.352 | 0.793 | *0.000* |
| | NA | USA | SF | 1 | 1 | n/a | n/a | n/a | - |
| | NA | USA | Martin | 1 | 1 | n/a | n/a | n/a | - |
| | NA | USA | WPB | 1 | 1.467 | n/a | n/a | n/a | - |
| | NA | USA | Nash | 18 | 2.467 | 0.037 | 0.311 | 0.864 | *0.000* |
| | NA | USA | MC | 22 | 3.867 | 0.173 | 0.452 | 0.587 | *0.000* |
| Meso America | MA | Mexico | MX-1 | 19 | 3.13 | 0.208 | 0.396 | 0.481 | *0.000* |
| | MA | Mexico | DF | 1 | 1 | n/a | n/a | n/a | - |
| Central America | CA | Honduras | Hon | 20 | 4.2 | 0.365 | 0.567 | 0.312 | *0.000* |
| South America | SAm | Peru | Peru | 1 | 1.267 | n/a | n/a | n/a | - |
| | SAm | Argentina | Arg | 9 | 3.2 | 0.349 | 0.482 | 0.231 | *0.012* |
| Total/Average | | | 25 | 274 | 12 | 0.257 | 0.402 | 0.469 | *0.000* |

107
108 *N* = sampling size; *Na* = Average number of alleles; $H_O$ = observed heterozygosity; $H_E$ =
109 expected heterozygosity; $F_{IS}$ = index of inbreeding of individuals resulting from the non-random
110 union of gametes within a subpopulation, positive value indicates heterozygote deficit; *P*
111 proportion of permutations leading to a $F_{IS}$ value larger or equal to the observed one. n/a
112 indicates values not estimated due to sample size smaller than 9 individuals.

113 # MATERIALS AND METHODS

114 ## Ethics statement

115 The Institutional Review Board of the University of Florida exempted the study from review
116 (Exemption of Protocol #2009-U-0422) and waived the need for written informed consent of the
117 participants. This exemption is issued based on the United States Department of Health and
118 Human Services (HHS) regulations. Specifically, HHS regulation 45 CFR part 46 applies to
119 research activities involving human subjects. Because louse removal was voluntary and no
120 information was recorded that would allow patients to be identified directly or through identifiers
121 linked to them, the University of Florida waived the need for written informed consent from the
122 participants.

123 ## Sampling

124 Sampling details are provided in Fig 1, Table 1 and S1 Table. Briefly, 274 human head lice
125 were collected from different individuals at 25 localities throughout the world covering 10 major
126 geographic regions.

127 ## DNA extraction

128 Lice were cut in half, placed into 1.5 ml Eppendorf tubes containing a cell lysis and
129 proteinase K solution, and ground using a pestle. DNA was extracted from each individual louse
130 using the Puregene Core Kit A (QIAGEN, Valencia, California) following manufacturer
131 protocols with modifications. We used a NanoDrop 1000 spectrophotometer (Thermo Scientific)
132 to determine the amount of DNA in each sample and then diluted them from their original
133 concentrations to approximately 5-10 ng/µl.

134 ## Multiple co-amplification: Multiplexes

135 A total of 15 microsatellite loci were amplified using four multiplexes as described in [16].
136 We used fluorescently labeled universal primers: M13 (5'-CACGACGTTGTAAAACGAC-3')
137 and CAG (5'-CAGTCGGGCGTCATCA-3'), each labeled with a unique fluorescent tag (e.g.
138 FAM, VIC, NED, PET; Applied Biosystems) to co-amplify multiple loci. Locus-specific primers
139 were modified by adding the matching 5′ universal primer sequence tails. The complete set of 15
140 loci was amplified in four separate 15 μL multiplex-PCR reactions containing 7.5 μL of 2X Master
141 Mix (Type-It Microsatellite PCR kit, Qiagen, Venlo, Netherlands), 0.015-0.15 μL of unlabelled
142 forward primer with tail (10 μM), 0.15-0.8 of unlabelled reverse primer (10 μM), 0.5 μL labelled
143 universal tail primer (10 μM), 1-2 μl of total genomic DNA (10-20 ng) and $sdH_20$ added up to final
144 volume. Two thermal cycling profiles were used, both beginning with initial denaturation at 95°C
145 (5 min) and ending with a final extension of 72°C (40 min). The v1 (touch-down PCR) protocol
146 consisted of 10 cycles of 94°C (30 sec), 60°C/55°C (-0.5°C/cycle) for 45 sec, and 72°C (45 sec)
147 followed by 25 cycles of 94°C (30 sec), 55°C (45 sec), and 72°C (1 min). The v2 protocol
148 consisted of 35 cycles of 94°C (30 sec), 52°C (45 sec), and 72°C (45 sec). PCR products were
149 electrophoresed on 1.5% and 4% agarose gels stained with ethidium bromide and visualized under
150 ultraviolet light. Dilutions in $dH_20$ from amplicons (1:100) were run on an ABI 3730xl 96-
151 capillary sequencer using GeneScan 600 LIZ as an internal size standard (Applied Biosystems) at
152 the University of Florida biotechnology facility (ICBR). Microsatellite genotypes were scored
153 using the GeneMarker version 1.60 software (SoftGenetics, LLC).

## 154 Mitochondrial PCR amplification and sequencing

155 A polymerase chain reaction was performed to amplify the mitochondrial gene cytochrome *c*
156 oxidase 1 (COX1) using the primers H7005 and L6625 [9] as described in [2]. PCRs consisted of
157 25 μL total volume including 10 μL of MasterMix (5 PRIME), 1 μL of each primer, 2-4 μL of total

158 genomic DNA, and water. The thermal cycling profile began with an initial denaturation at 94°C
159 (10 min) followed by 10 cycles of 94°C (1 min), 48°C (1 min), and 65°C (2 min) (decreased by
160 0.5°C per cycle). This was followed by 35 cycles of 94°C (1 min), 52°C (1 min), and 65°C (2
161 min) and then a final extension of 65°C (10 min). Amplified fragments were purified using
162 ExoSAP-IT (USB Corporation, Cleveland, Ohio). Sequencing was performed at the University of
163 Florida DNA Sequencing Core Laboratory (ICBR, Gainesville, Florida) using standard fluorescent
164 cycle-sequencing PCR reactions (ABI Prism Big Dye terminator chemistry, Applied Biosystems).

165 ***Sequence editing, alignment, and haplotype reconstruction***

166 The forward and reverse sequences were edited and aligned using Sequencher 4.5 (Gene Codes
167 Corporation, Ann Arbor, Michigan) with base calling confirmed by eye. Consensus sequences
168 were generated for each sample using both forward and reverse sequences. These new louse COX1
169 sequences were aligned along with previous sequences from Ascunce et al. (2013) [17]. A final
170 alignment of 379 bp was used to construct a matrix of pairwise differences using uncorrected *p*-
171 distances (proportion of nucleotide sites at which two sequences being compared are different)
172 among COX1 unique haplotypes was calculated with MEGA 4 [29]. The genetic relationships
173 were estimated by constructing neighbor-joining (NJ) trees [30] using PAUP* 4.0b10 [31].
174 Bootstrapping was performed using 100 pseudo-replications of the data set.

## 175 Analyses of genetic diversity and population structure among 176 collection sites

177 We used the software Arlequin 3.5.1.2 [32] to estimate population genetic diversity values.
178 First, for all sites regardless of the number of lice per site, we determined the number of alleles per
179 locus. Then, for those 11 sites with sample sizes exceeding 8 lice per site, we also estimated
180 observed and expected heterozygosity ($H_O$ and $H_E$), and the mean $F_{IS}$ over loci per population with

181 significance tested using 1,023 random permutations. For those same 11 sites with sample sizes
182 exceeding 8 lice per site, population pairwise $F_{ST}$ values were estimated using Arlequin 3.5.22
183 [33]. Significance values for pairwise $F_{ST}$ values were estimated using the permutation procedure
184 with 1,000 permutations.

185 To assess the distribution of the human louse genetic diversity among world populations
186 through clustering of similar genotypes, we used three different methods as they have different
187 assumptions. First, we used STRUCTURE [34], a model-based Bayesian clustering method, that
188 assumes Hardy–Weinberg and linkage equilibrium among loci. Second, we also used two
189 multivariate methods, which do not carry these assumptions and could visualize potential cryptic
190 structure, including PCoA (Principle Coordinates Analysis) in GenAlEx 6.5 [35,36] and
191 Discriminant Analysis of Principal Components (DAPC) [37].

## 192 Bayesian clustering analyses

193 The Bayesian clustering approach implemented in STRUCTURE 2.3.4 [34] uses the individual
194 multi-locus genotypic data to evaluate models assuming different numbers of genetic clusters ($K$)
195 based on the posterior probabilities given the data and model. Based on the individual multilocus
196 genotypes and the allele frequencies estimated for the reconstructed clusters, each individual's
197 genome was probabilistically partitioned into membership fractions ($q$) (ancestry) in each cluster.
198 Admixed ancestry is modeled by assuming that an individual $i$ has inherited some fraction ($q$) of
199 its genome from ancestors in population $k$ [34]. Hybrids can then be inferred when observing an
200 individual being intermediate of some degree between two or more clusters in terms of the $q$-value.
201 For example, individuals with $q = 0.5$ would be likely first-generation hybrids. We identify
202 possible hybrids based on $q$-values threshold of 0.2 as suggested in VäHä and Primmer (2005)
203 [38], where threshold $q$-value of 0.2 means that individuals with $q$-value between 0 and < 0.20 or >

204 0.8 and 1 are classified as non-admixed and any individual with *q*-value between 0.2 and 0.8 are
205 classified as a hybrids. All simulations used 50,000 Markov chain Monte Carlo (MCMC)
206 generations in the burn-in phase and 100,000 generations in the data collection phase. Ten
207 independent runs were performed using default parameters for each *K*, to ensure equilibration
208 during burn-in and consistency in estimation of the posterior probabilities. Selection of the number
209 of distinct clusters was based on the evaluation of the $\Delta K$ statistic [39] using the online tool
210 STRUCTURE HARVESTER [40]. The ten STRUCTURE runs at each *K* produced nearly
211 identical individual membership coefficients. The run with the highest likelihood of the data given
212 the parameter values for the predominant clustering pattern (i.e. the mode) at each *K* was used for
213 plotting with DISTRUCT [41].

## 214 Multivariate analysis using Principal Coordinates Analysis (PCoA)

215 This multivariate analysis was conducted using pairwise genetic difference between
216 individuals calculated in GenAlEx 6.5 [35,36] to validate and further define genetic clusters found
217 using STRUCTURE.

## 218 Multivariate analysis using Discriminant Analysis of Principal Components 219 (DAPC) [37]

220 This analysis was conducted using the '*adegenet'* package 2.1.3 [42], in the R software
221 environment version 3.6.3 [43]. For these analyses we only included 175 lice from which we were
222 able to obtain both microsatellite genotype and mitochondrial haplogroup sequence. An additional
223 DAPC included the same louse samples, however we considered four groups depending on nuclear
224 cluster and haplogroup membership.

225 ## Allelic richness within each main nuclear STRUCTURE genetic
226 cluster

227 We estimated the allelic richness withing each main nuclear STRUCTURE genetic cluster
228 using the rarefaction method implemented in the computer program ADZE (Allelic Diversity
229 Analyzer) version 1.0 [44]. Because allelic richness measurements considered sample size, the
230 rarefaction method allows the estimation of allelic richness for different random subsamples of
231 size "g" from the populations [45,46].

232 ## Analyses based on STRUCTURE genetic clusters and mitochondrial
233 clades

234 To gain a better understanding of the geographic distribution of the nuclear and mitochondrial
235 diversity, we conducted a set of Kruskal–Wallis nonparametric analysis of variance tests [47]. We
236 included the same 175 lice from which we were able to obtain both microsatellite genotype and
237 mitochondrial haplogroup sequence data and the same eight geographic regions as described above.
238 Calculated medians were compared and separated at $P < 0.05$ [48]. We compared the Observed
239 vs. Expected Frequency $X^2$ test to evaluate the fit of the data (frequencies) to any arbitrary set of
240 expected frequencies among the eight geographic regions. All statistical tests were performed with
241 $\alpha$=0.05 for significance of statistical tests.

242 Analysis of molecular variance (AMOVA) considering the number of different microsatellite
243 alleles ($F_{ST}$ based) was used to investigate the partitioning of genetic variation between the two
244 nuclear clusters: cI and cII using Arlequin 3.5 [32]. Statistical significance of differentiation at
245 each level was assessed by means of 1,023 permutations.

246 The historical migration patterns among the genetic clusters identified using the methods
247 describe above were estimated in Migrate-n 3.3.0 using the maximum likelihood approach [49,50].
248 Migrate employed estimates of gene genealogies to sample areas of the coalescence space with
249 highest likelihood. The random number of seed and starting values of Θ and $4N_em$ (where $N_em$ is
250 the indirect measure of gene flow) were based on $F_{ST}$ as starting parameters and we used a
251 Brownian mutation model. The data was analyzed using 10 short ($10^4$ MCMC steps) and 3 long
252 chains ($10^5$ MCMC steps), together with adaptive heating to assure convergence.

253 **Comparison of demographic models**

254 We further inferred the demographic history of human lice, based on microsatellite data, using
255 the approximate Bayesian computation (ABC) approach [51], implemented in the program
256 DIYABC, v2.0 [52]. We tested different evolutionary scenarios in which each louse genetic
257 cluster evolves on different human host populations and diverge in tandem with their hosts. A
258 total of seven different evolutionary scenarios were considered with population size estimates and
259 divergence times based on inferences done using data on human evolution and louse life history
260 traits references (Table 2, Fig 2). The first three scenarios are directly testing Reed et al. (2004)
261 [2] hypothesis of ancient hybridization between modern and archaic humans. The scenarios 1 and
262 2 were trying to unravel the retention of ancient polymorphisms on the hybrid lice, where the main
263 nuclear genetic clusters were derived from lice that parasitized the heads of Neandertals (scenario
264 1), or the heads of anatomical modern humans (AMH) (scenario 2). In scenario 3, the model tested
265 hybridization events between lice parasitizing Neanderthals and AMH during periods of
266 contemporaneity between the two hosts. The last four scenarios (scenario 4, 5, 6 and 7) do not
267 require ancient hominids, and are based solely on hybridization events within AMH during recent
268 time periods. In scenario 4, we tested the hypothesis that the hybrid lice were the product of

269 hybridization between lice from the Native Americans and the first Europeans during the European
270 colonization of the Americas 500 years ago. In scenario 5, we tested the hybridization occurring
271 in the last 100 to 40 years around the World Wars I and II. Scenario 6 reflects the beginning of
272 the globalization from 1980's to ten years ago, while scenario 7 refers to very recent events in the
273 last 10 years.

274 **Table 2. Prior distributions of historical and demographic parameters for the demographic scenarios.** Between [ ] are the host
275 values obtained from anthropological references. For estimates of times, those values for louse represents louse generations (g), and
276 the values between [] for hominids are in years (y). Louse values are estimate considering an infection of 20 lice in 30% of hominin
277 population.

| Details/Historic Events | Parameter | Distribution | Minimum-Maximum | References |
|---|---|---|---|---|
| *Effective population* | | | | |
| Louse Ancestral Population | $N_A$ | Uniform | 1,800-36,000 | |
| [Ancestral hominids] | | | [300-6,000] | |
| Louse cluster I | $N_1$ | Uniform | 1,800-36,000 | |
| [AMH] | | | [300-6,000] | |
| Louse cluster II | $N_2$ | Uniform | 1,800-36,000 | |
| [Neanderthals] | | | [300-6,000] | [113] |
| Louse Hybrids | $N_3$ | Uniform | 270-5,400 | 15% hybrids from current study |
| [Hybrids between AMH and Neanderthals] | | | [?] | |
| *Time Events** | | | | |
| Time of split clusters I and II | $t_S$ | Uniform | 1,200,000 - 16,000,000g | |
| [Split between AMH and Neanderthals] | | | [600,000-800,000y] | [113] |
| T1 ancestral polymorphism from cII | T1 | Uniform | 2,400,000 - 2,800,000g | [114] |
| [Presence of Neanderthals in Mongolia] | | | [120,000 - 140,000y] | [116] |
| T2 ancestral polymorphism from cI | T2 | Uniform | 1,000,000 - 1,600,000g | |
| [Presence of AMH in Middle East] | | | [50,000-80,000y] | [105] |
| Time of ancestral admixture | $t_{aA}$ | Uniform | 1,000,000 - 1,600,000g | |
| [AMH-Neanderthals in sympatry] | | | [50,000-80,000y] | [20] |
| *Times of Recent Admixture* | | | | |
| Time of admixture E1 | $T_{aE1}$ | Uniform | 2,000 – 20,000g | |
| [European colonization] | | | [100 – 1,000y] | |
| Time of admixture E2 | $T_{aE2}$ | Uniform | 800 – 2000g | |
| [Migration after World Wars] | | | [40 – 100y] | |
| Time of admixture E1 | $T_{aG}$ | Uniform | 200 – 800g | |
| [Initial Globalization] | | | [10 – 40y] | |
| Time of admixture E1 | $T_{aC}$ | Uniform | 20 – 200g | |
| [Current time -t sampling] | | | [1 – 10y] | |

| | | | |
|---|---|---|---|
| *Admixture rate* | | | |
| Rate of admixture | ar | Uniform | 0.001-0.999 |
| *Mutation rate* | | | |
| Tetranucleotide microsatellite mutation rate | Um1 | Loguniform | $10^{-9}$–$10^{-3}$ |
| Trinucleotide microsatellite mutation rate | Um2 | Loguniform | $10^{-7}$–$10^{-3}$ |
| Dinucleotide microsatellite mutation rate | Um3 | Loguniform | $10^{-6}$–$10^{-2}$ |

*Time parameters for lice are expressed as the number of generations of lice back in time assuming a generation time of 18 days [15]. The values for hominids refer to years

281 **Fig 2. Evolutionary scenarios.** A total of seven different scenarios were compared using the
282 approach from the DIY-ABC program.

283 For each scenario, we considered three levels of parasitism: from a low level of parasitism
284 where each head human would have 10 lice and only 10% of the heads would be infected (S2
285 Table); an intermediate level, where each head would have 20 lice and only 30% of the population
286 is infested (Table 2); and a high level of lice, where there are 50 lice per head in 50% of the
287 population (S3 Table) [53]. The generation time of *P. humanus* was considered as 18 days [15].

288 All microsatellites used the default generalized stepwise mutation model and were assigned a
289 log uniform distribution with mean mutation rates between $10^{-9}$–$10^{-2}$. The standard mutation rate
290 for microsatellites is considered to be between $10^{-6}$–$10^{-2}$ per locus per generation [54]. However,
291 due to the possibility of different repeat lengths having different mutation rates [55] for
292 microsatellites, tetra-, tri- and di- nucleotide repeats were modeled separately. Thus, preliminary
293 runs were carried out in order to optimize mutation rates to account for other factors [56] following
294 Bertorelle et al. [57]. A range in allele size of 40 contiguous allelic states was used, as
295 recommended by the program developers (A. Estoup, personal communication).

296 Summary statistics calculated for microsatellite loci were as follows: Mean number of alleles,
297 Mean genetic diversity, and Mean size variance, and Mean $F_{ST}$. The selection of summary
298 statistics was empirical based on preliminary tests where these summary statistics were used and
299 able to recover the models and parameters following Bertorelle et al. [57]. Through these
300 preliminary tests the above summary statistics demonstrated their suitability for our data.

301 We performed about 714,000 simulations using prior parameters for each scenario, thereby
302 producing a reference table comprising 5,000,000 simulated data sets. First, we pre-evaluated the
303 fit of observed values to prior distributions of scenarios, by conducting a Principal Components

304 Analysis (PCA) using data sets simulated with the prior distributions of parameters and the
305 observed data as well as data sets from the posterior predictive distribution to check the suitability
306 of the model (scenarios and prior parameters). The posterior probability of each scenario was
307 assessed by performing direct approach and a polychotomous weighted logistic regression that
308 estimates the difference between simulated (from the reference table) and observed data sets
309 [58,59] after linear discriminant analysis on summary statistics [52]. Only the posterior
310 probabilities for the 1% (50,000) of simulated data sets closest to the observed data were retained.
311 Confidence in choice of a scenario was estimated as the proportion of times the selected scenario
312 did not possess the highest posterior probability compared to competing scenarios for the 500
313 simulated data sets (type I error). Confidence in scenario choice was further tested including the
314 estimation of the prior error rate, which is calculated as the probability of choosing an incorrect
315 model when drawing model index and parameter values into priors (the probability of selecting
316 the chosen scenario when it is not correct, a type-II error). After scenario choice, we proceeded to
317 parameter inference estimated from the modes and 95% confidence intervals (CI) of their posterior
318 distributions.

319 # RESULTS

320 ## Nuclear Genetic Diversity

321 The 274 human louse specimens comprising 25 geographic sites through 10 worldwide
322 geographic regions revealed high microsatellite diversity with overall five to 20 alleles per locus
323 and an average of 12 alleles per locus (Table 1). In the 11 sites with more than 8 lice analyzed,
324 $H_O$ values ranged from 0.032 to 0.524 (average of 0.257) and $H_E$ values ranged from 0.248 to
325 0.706 (average of 0.402), reflecting a general patter of deficit of heterozygotes (Table 1).
326 Furthermore, in all 11 louse populations, genotype proportions deviated significantly from Hardy-

327 Weinberg expectations with values of $F_{IS}$ significantly positive and an overall $F_{IS}$ value of 0.469
328 ($P < 0.00001$) (Table 1). Overall, there were high values of genetic differentiation among the 11
329 sites, with $F_{ST}$ values of up to 0.663 between the sites: Mongolia (Mn) and Netherlands (Neth) (S4
330 Table). The lowest $F_{ST}$ of 0.053 was found between sites in New York (Oce) and Tennessee
331 (Nash).

## 332 Mitochondrial diversity

333 We also obtained the COX1 sequence data from 175 head lice spanning 18 sampling locations
334 throughout nine geographic regions (Table 3). One hundred four (104) lice belonged to
335 haplogroup A and 71 to haplogroup B (Table 3). As expected based on previous studies [17,21,23],
336 haplogroup A shows a worldwide distribution including Africa, Asia, Europe and the Americas.
337 For our sampling, haplogroup B has a narrower distribution and it is found only in lice from Europe,
338 North and Central America (Table 3).

339 **Table 3. Nuclear and mitochondrial louse genetic diversity.** Geographic distribution of 175
340 human lice including combined information of their mitochondrial DNA haplogroups based the
341 COX1 gene and their assignment to each louse major nuclear genetic clusters described in this
342 study.

| Geographic Region | Group | Country | Site | mtA+clusterI | mtB+clusterI | mtA+clusterII | mtB+clusterII |
|---|---|---|---|---|---|---|---|
| Africa | AIS | Rwuanda | Rw | 1 | | | |
| South Asia | SAs | Nepal | Np | 1 | | | |
| Southeast Asia | SAs | Cambodia | Cam | 21 | | | |
| | | Laos | Ls | 1 | | | |
| East Asia | EAs | Mongolia | Mn | 6 | | | |
| Europe | E | Turkey | Tur | | | | 7 |
| | E | Croatia | Croa | | | | 5 |
| | E | Norway | Nw | | | 1 | |
| | E | Netherlands | Neth | | | 49 | 5 |
| | E | France | Fr | | | 2 | |
| | E | U. Kingdom | UK | | | | 26 |
| North America | NA | USA | Oce | | | 12 | 4 |
| | NA | USA | SF | | | 1 | |
| | NA | USA | Nash | | | | 3 |
| | NA | USA | MC | | | 1 | 9 |
| Mesoamerica | MA | Mexico | MX | | | 6 | |
| Central America | CA | Honduras | Hon | | 12 | | |
| South America | SAm | Argentina | Arg | | | 2 | |
| Total | | | | 30 | 12 | 74 | 59 |

343

344 ## Nuclear genetic clusters at the worldwide scale

345 All clustering methods: STRUCTURE, PCoA and DAPC were concordant in identifying the
346 distinct nuclear genetic clusters: Cluster I (cI) with subclusters Ia and Ib and Cluster II (cII)
347 including subclusters IIa and IIb (Fig 1; S1 Fig). It is striking, the difference in their geographic
348 distribution, where cI defined by head lice from Africa, Israel, Asia and the Americas compass a
349 more worldwide distribution, while cII composed of head lice mainly Europe and the Americas
350 reflects a more restricted geographic distribution. While most of the genetic diversity of human
351 louse populations from around the world seems to be highly structured geographically,
352 approximately 12% of the lice we sampled (33 of 274 lice) appear to be hybrids (lice with *q*-values
353 less than 0.8), with 75% (25 lice) of the hybrids found in the New World.

354 In the STRUCTURE analysis, after a signal for $K = 2$ ($\Delta K = 858.853$, S2 Fig), increasing $K$
355 values in the STRUCTURE plots revealed geographically localized structure. At $K = 3$, a new
356 cluster formed within Europe from the original cII and it is composed of lice primarily from the
357 Netherlands forming the subcluster cIIb (light blue color). When $K$ was increased to 4 a population
358 from Honduras and other parts of the Americas emerged from cluster I as the Ib subcluster, with
359 the other samples remaining in subcluster Ia. This Honduras (Ib subcluster) might be indicative
360 of Native American louse lineage.

361 The Principal Coordinate Analysis (PCoA) supported the two main nuclear genetic clusters
362 and subclusters identified by STRUCTURE, with putative hybrid lice not assigned to either cluster,
363 rather were placed between them (Fig 3). Finally, the eigenvalues obtained in the DAPC analysis
364 showed that the genetic structure was captured mainly by the first principal component
365 representing cI (to the right) and cII (to the left) (S1 Fig).

366 **Fig 3. Human louse worldwide nuclear genetic diversity**. A) Genetic clusters inferred from
367 STRUCTURE simulations at $K = 2$: Cluster I (cI, Orange) and Cluster II (cII, Blue), $K = 3$ cI and
368 cIIa-cIIb, and $K = 4$ cIa-cIb and cIIa-cIIb. In the bar plot, each louse is represented by a single
369 vertical line and the length of each color segment represents the proportion of membership ($Q$) to
370 each cluster. B) Distributions of points in the first two dimensions resulting from Principal
371 Coordinates Analysis (PCoA) conducted using pairwise genetic distance comparisons. Each site
372 was color-coded to match STRUCTURE results (e.g. sites in Asia were color-coded in orange,
373 and sites in Europe were color-coded in blue) codes correspond to geographic origin and
374 abbreviations as defined in Table 1.

375 Additionally, we conducted further analysis with all the dataset and excluding hybrids as
376 identified through STRUCTURE $q$-value. Thus, when using the reduced dataset only non-
377 admixed individuals with shared ancestry $q$-values larger than 0.8 as suggested by [38], the
378 STRUCTURE analysis reflected similar results identifying cI and cII, with their corresponding
379 subclusters at $K = 4$ (S3 Fig).

## 380 Worldwide cluster I

381 All lice that belong to the cI ($q$-values larger than 0.5) were included in a second STRUCTURE
382 simulations to assess the substructure within this cluster. At $K = 2$, this figure strikingly showed
383 the presence of two clusters: one composed of lice from the Old World (Africa, Asia and Europe),
384 and the other of lice from the New World (Americas) (S4 Fig). The further split of the genetic
385 clusters during the increase of defined $K$ in the STRUCTURE analysis showed at $K = 3$ the
386 emergence of an Asian cluster except for most of the lice from Mongolia that remains within the
387 Old Word cluster, however at $K = 4$, a subset of those lice from Mongolia formed a unique cluster
388 (S4 Fig).

## Geographically restricted cluster II

Considering our current sampling, cII has narrower geographic distribution than the cI and it is limited mostly to Europe and the Americas (Fig 3). This cluster showed a substructure at $K = 2$ where most of the lice from the Netherlands plus few others formed a distinct cluster (light blue in plot) (S5 Fig). Both these individual cluster analyses for cI and cII highlight the geographic structure observed among louse populations.

## Allelic richness analysis per main genetic cluster

We estimated the allelic richness per main non-admixed clusters I and II (without hybrids) using the rarefaction method implemented in the computer program ADZE (Allelic Diversity Analyzer) Version 1.0 [44]. Because allelic richness measurements considered sample size, the rarefaction method allows the estimation of allelic richness for different random subsamples of size "g" from the populations [45,46]. In addition, we conducted the same analysis for each cluster separately using the sub-clusters defined by STRUCTURE.

Allelic richness per non-admixed cluster was higher in cluster I than II (S6 Fig). This pattern reflects very different evolutionary histories for the two clusters. While high allelic richness is consistent with a large effective population size, it could also be concordant with greater genetic structure. It is noteworthy also that the clusters I and II differ in sample size (80 lice in the cI versus 174 in cII).

## Statistical analysis of genetic clusters and mitochondrial clades

The combined analysis of the nuclear cluster membership frequency and the mitochondrial haplogroup were plotted using a Correspondence Analysis (CA) and showed the presence of four well defined groups (S7 Fig). The dimensions 1 and 2 represented the largest amount of inertia or

411 variance representing a measure of deviation of independence. Each one of the dimensions possess
412 an eigenvalue that represents its relative importance and how much of the variance is explained.
413 In our study, one dimension represents 58.7 % of the inertia and two dimensions represents 99.6%
414 of the inertia. Of this, dimension 1 differentiated both nuclear genetic clusters: cI and cII; while
415 dimension 2 differentiated the mitochondrial haplogroups. One group corresponds to lice of
416 nuclear cluster II-haplogroup A and included lice from Europe, North America, Mesoamerica and
417 South America. The second group is also nuclear cluster II-haplogroup B from the same regions,
418 except South America. The third group is composed of individuals of cluster I-haplogroup A:
419 South Asia, Africa-Israel-Spain, some Europe, while the last group belonged to cluster I with
420 haplogroup B and were found in Central America (S7 Fig). Furthermore, multiple comparisons
421 (non-parametric Kruskal–Wallis test) indicated that the distribution of the individuals assigned to
422 the nuclear genetic clusters I and II were not distributed randomly among the eight geographic
423 regions (H = 103.55, $P < 0.0001$). The cluster membership median values were significantly
424 different among: South Asia (SAs), East Asia (EA), Africa-Israel-Spain (AIS), Central America
425 (CA), and Europe (E) ($P < 0.05$). On the other hand, cluster membership medians values for North
426 America (NA), South America (Sam), and Mesoamerica (MA) were not significantly different.
427 Both nuclear genetic assignment and haplogroup frequencies differed significantly among the
428 eight geographic regions (nuclear cluster assignment $X^2 = 175$, $P < 0.0001$; haplogroup frequencies
429 $X^2 = 42$, $P < 0.0001$).

## 430 AMOVA by nuclear clusters and mitochondrial clades

431 Hierarchical analyses of molecular variance (AMOVA) employed to explore the distribution
432 of nuclear genetic variation considering the two nuclear genetic clusters: cI and cII revealed
433 significant genetic differentiation between them ($\Phi_{ST} = 0.33$; $P < 0.00$). When examining only the

434 non-admixed individuals (those with at least 0.8 of coefficient of ancestry values) the
435 differentiation between cI and cII increased ($\Phi_{ST}$ = 0.4; $P$ < 0.00).

## 436 Migration patterns considering nuclear genetic clusters and 437 mitochondrial clades

438 The historical migration pattern was estimated in Migrate-n 3.3.0 using maximum likelihood
439 inference and the Brownian motion mutation model [49,50]. Given that previous studies have
440 suggested a different evolutionary history between the mitochondrial haplogroups A and B, we
441 evaluated a full migration model with four populations defined by both the nuclear clusters inferred
442 from STRUCTURE and the mitochondrial haplogroups: cI – mtDNA A; cI – mtDNA B; cII –
443 mtDNA A and cII – mtDNA B. Migration rates between these four populations varied. In general,
444 migration rates within nuclear clusters were high, contrasting with low migration rates within
445 mitochondrial haplogroups (S8 Fig). For example, we observed the highest migration rates within
446 cluster II (cluster II – mtDNA A and cluster II – mtDNA B) and some of the lowest migration rates
447 between cluster I – mtDNA A and cluster II – mtDNA A. Because STRUCTURE analysis
448 indicated the effect of geography on the genetic structure of human lice, further studies including
449 more samples from each sampling site are needed to further understand migration patterns among
450 geographic human louse populations.

## 451 Demographic history

452 We inferred relationships between sets of individuals based on the STRUCTURE genetic
453 clusters using approximate Bayesian computation (ABC) incorporated in the software DIYABC,
454 v2.0 [52]. Individuals were assigned to the genotype pool based on their $q$ values, where
455 individuals with $q$-value between 0 and < 0.20 or > 0.8 and 1 are classified as non-admixed and

456 any individual with *q*-value between 0.2 and 0.8 are classified as hybrids. We tested seven
457 different evolutionary scenarios in which each louse genetic cluster evolves on different human
458 host populations and diverges in tandem with their hosts. The scenarios reflect different
459 hypothesizes that followed human evolutionary demographic events from the admixture between
460 Neanderthals and AMH to current human movement around the globe during globalization (Table
461 2, Fig 2). We repeated the same 7 models using different levels of louse parasitism. Comparison
462 of the seven scenarios showed that the most strongly supported scenario was scenario 7 suggesting
463 a very recent admixture between the two nuclear clusters with a logistic posterior probability of
464 ca. 0.90 (Fig 4, Table 4). The three most strongly supported scenarios 5, 6 and 7, are topologically
465 similar each involving admixture between the two nuclear clusters at different time paralleling
466 different human host colonization and migration events from Europe into the New World.
467 **Fig 4. Posterior probabilities for each of the seven demographic scenarios.** Graphs of
468 posterior probabilities based on DIY-ABC analysis comparing the seven demographic scenarios
469 when considering intermedia louse infestation level where 30% of the hominin population have
470 lice with 20 lice per head. Each scenario was color-coded as follow Scenario 1 – green, Scenario
471 2 – red, Scenario 3 – light blue, Scenario 4 – pink, Scenario 5 – yellow, Scenario 6 – black,
472 Scenario 7 – gray.

473 **Table 4. Posterior probabilities for the seven demographic scenarios shown in Fig 2.** Probabilities obtained through the
474 logarithmic method (with 95% confidence intervals in brackets) for the seven scenarios inferred from DIY-ABC analysis. For each
475 scenario three parasitism level were considered where 10% indicates that 10% of the hominin population have lice with 10 lice per
476 head, 30% and 50% indicate that 30% and 50% of the hominin population have 20 and 50 lice per head, respectively. For each
477 comparison, the selected scenario (bold entry in shaded cell) was the one with the highest probability value. Confidence in scenario
478 choice, shown under probability values, was evaluated by estimating the type I error rate from 1000 randomly constructed simulated
479 data sets, for scenarios 5 and 7, respectively.

| Scenarios | Parasitism Level | | |
|---|---|---|---|
| | 10% | 30% | 50% |
| 1: Retain Polymorphisms from cI | 0.0000 [0.0000,0.0000] | 0.0000 [0.0000,0.0715] | 0.0000 [0.0000,0.1130] |
| 2: Retain Polymorphisms from cII | 0.0000 [0.0000,0.0000] | 0.0000 [0.0000,0.0715] | 0.0000 [0.0000,0.1130] |
| 3: Ancestral admixture cI and cII | 0.0000 [0.0000,0.0000] | 0.0000 [0.0000,0.0715] | 0.0000 [0.0000,0.1130] |
| 4: European colonization of the New World | 0.0000 [0.0000,0.0000] | 0.0001 [0.0000,0.0716] | 0.0560 [0.0000,0.1522] |
| 5: European migration to the New World after WWs | 0.0000 [0.0000,0.0000] | 0.0071 [0.0000,0.0757] | **0.5204 [0.4477,0.5930]** |
| 6: Globalization | 0.0026 [0.0011,0.0041] | 0.3792 [0.2622,0.4963] | 0.2754 [0.2042,0.3467] |
| 7: Current time - sampling | **0.9974 [0.9959,0.9989]** | **0.6136 [0.4959,0.7312]** | 0.1482 [0.0095,0.2869] |
| | 0.335 | 0.317 | 0.372 |

480

481 # DISCUSSION

482 In this study, we found that the nuclear microsatellite data from lice were grouped in two
483 distinct genetic clusters: cI and cII. Cluster I had a worldwide distribution whereas cII was
484 restricted to Europe and the Americas (Fig 5). Remarkably, the two nuclear clusters revealed a
485 significant genetic differentiation between them ($\Phi_{ST} = 0.33$; $P < 0.00$) despite having geographic
486 overlap in either Europe or the Americas. In the Structure analysis at $K$ =4 each of the two Clusters
487 further revealed the subclusters: Ia-Ib and IIa-IIb, which were sorted based on their geographical
488 origin (i.e., continents). These results are consistent with our previous work using fewer samples
489 where we showed that human louse populations are genetically structured based on geography
490 [16]. In that previous study, we assessed louse microsatellite genetic diversity of a small sampling
491 from four main geographic regions (North America, Central America, Asia, and Europe) that
492 showed that the head louse populations were strongly structured geographically, where one cluster
493 included head lice from North America and Europe, a second cluster encompassed only head lice
494 from Central America, with a third Asian cluster [16]. This current study with a larger dataset
495 highlighted a similar pattern showing a close relationship between Central America (cIa) and Asia
496 (cIb) subclusters suggesting that the Central America cluster is probably from Native American
497 origin since East Asia has been suggested as a potential source population for the first people of
498 the Americas [60–62] (Fig 5). Thus, microsatellite data support the idea that current populations
499 of human lice in the Americas retain human louse genetic diversity brought by the first people
500 [16]. This work adds to our previous studies and shows that the study of the genetic diversity of
501 human lice provides insights to better understand patterns of human migrations worldwide.

502 **Fig 5. Proposed global co-migration human lice and humans.** Top: Map depicting the
503 geographic distribution of the nuclear genetic diversity among the human louse populations

504 included in this study. Colored circles on map indicate collecting sites, with the color of each
505 circle corresponding to the majority nuclear genetic cluster to which sampled individuals were
506 assigned. Sites with admixed lice are indicated with patterned circles including colors of the two
507 major genetic clusters at that site. Thick grey arrows indicate proposed migrations of anatomically
508 modern humans out of Africa into Europe, Asia and the Americas, as well as the more recent
509 European colonization of the New World. Colored arrows represent hypothetical human louse co-
510 migrations. The bottom panel is the STRUCTURE plot from Fig 3A corresponding to the
511 assignment of 274 lice from 25 geographical sites at $K = 4$ (Table 1).

512 Other human parasites have also shed light into different aspects of human evolution [11–13].
513 For example, the evolutionary history of the bacterial parasite, *Helicobacter pylori*, which infects
514 the stomach of most humans, reflected major events in human migrations such as the out of Africa
515 expansion [63,64], the colonization of Austronesian speaking region [65], and the settlement of
516 the Americas by Asians across the Bering Strait [66]. However, microorganisms such as bacteria
517 and viruses often reproduce clonally, increasing the probability of observing a selective sweep
518 such that the parasite phylogeny is likely to reflect the selection history of some genes rather than
519 of the host tree. On the other hand, macroparasites (such as parasitic arthropods) often reproduce
520 sexually and may track host evolutionary history more accurately because selective sweeps affect
521 only closely linked loci in their genomes [67]. In cases where obligate parasites coevolve with
522 their hosts, parasite phylogenies will, to some extent, reflect host evolutionary history [8,9]. For
523 this reason, the use of macroparasites such as pinworms [68] and lice [2,3] has become increasingly
524 common in the study of human evolution. In fact, the human louse, *Pediculus humanus*, is one of
525 the most well studied human parasites [1–6,14–19,21–28].

526 A key feature of lice is that human lice can disperse from a host to another at any motile stage
527 (nymphs and adults), however adults are most likely to transfer from one head to another [69].
528 Lice can be transferred via direct body contact, for example when people fight, live in group and
529 their heads are in contact [70]. Additionally, lice could be transferred via indirect contact, like
530 when an infected person shares their hat and combs with other person, there could be a possibility
531 that live lice stayed on the personal staff and might be transferred to the other person head [71].
532 In addition, because parasites have larger population sizes and faster rates of molecular evolution
533 compared to their hosts, population genetic patterns of parasites can reveal life history traits of
534 their host such as host dispersal, interspecies interactions and population structure that might not
535 be evident from host data alone or not yet discernible in the host's genome [9,72,73]. Differences
536 between parasite and host intrinsic life-history traits may lead to incongruities in patterns of
537 population structure between hosts and their parasites [74–78]. Several types of parasites are
538 known to demonstrate this phenomenon, including flies from bats [78], helminths from pikas [79],
539 and feather lice from Galápagos hawks [80]. For example, the comparative evolutionary histories
540 between Arctic Galliformes and their parasitic lice suggest consistent movement of lice among
541 different host species and populations in Alaska [81]. While the bird hosts present some level of
542 population structure and/or isolation by distance, no structure was detected within either genus of
543 louse on these hosts [81]. These incongruent patterns in host and parasite genetic structure suggest
544 there is still contact or at least distributional overlaps in host populations which are not perceptible
545 from the host data alone. In our current study, we show the presence of two louse distinct nuclear
546 genetic clusters: I and II. These could be the result of the retention of ancient polymorphisms due
547 to incomplete lineage sorting (ILS) or admixture. To test if louse current genetic diversity is the
548 result of ILS or admixture we used the approximate Bayesian computation (ABC) method as

549 implemented in the DIY-ABC package [52]. We performed ABC analyses for seven scenarios
550 and for three different level of parasitism (Table 2, 4, Figs 2, 4, S9 Fig). These analyses suggested
551 a scenario of recent admixture between nuclear Clusters I and as observed in the STRUCTURE
552 analysis, where most of the hybrids are being observed in the New World. While most of the
553 genetic diversity in human louse population from around the world seems to be highly
554 geographically structured, there are a total of 33 hybrids out of 274 lice analyzed reaching 12% of
555 our current sampling. However, 25 of those hybrids (75%) are found in the New World, which
556 may be mirroring human colonization history and reflecting a zone of secondary contact. All the
557 three scenarios considering recent admixture events from European migration after World Wars
558 (scenario 5), initial globalization (scenario 6) and current times at sampling (scenario 7) were the
559 most supported (Table 2, 4, Fig 4, S9 Fig).

560 A new and puzzling insight from this study was the finding of such low prevalence of hybrid
561 lice and that these hybrids are geographically concentrated in the New World. This pattern could
562 be reflecting the human colonization history from Europe to the New World as mentioned above.
563 However, if these colonization events have been occurring for many louse generations, why are
564 there so few hybrids? A tentative explanation could be associated to sampling issues as a large
565 number of samples came from the Americas (43%) (Table 1). Thus, future studies on genetic
566 diversity of human lice should consider larger sampling if possible. Despite that it is beyond the
567 scope of the present paper to investigate the reason behind the lack of hybrids, we hypothesize that
568 this pattern could be indicating some sort of genetic isolation between cI and cII such that the
569 hybrids are not able to persist for many generations within the human louse populations. Thus, it
570 is worthwhile to enumerate the potential causes for future researcher.

571 One potential mechanism of reproductive isolation may involve the effect of endosymbiont on
572 host biology. Endosymbionts could also alter host biology in diverse ways, including the induction
573 of reproductive manipulations, such as feminization, parthenogenesis, male killing and sperm–egg
574 incompatibility, which could lead to reproductive incompatibility [82]. *Wolbachia* is a common
575 intracellular bacterial symbiont of arthropods and filarial nematodes and it has been shown to
576 infect human lice [83–85]. When populations of insect host that contain different strains of
577 *Wolbachia* interbreed, or when one population has no *Wolbachia* at all and the other population
578 does have *Wolbachia*, the interbreeding matings ("hybrids") result in reduced fitness or the
579 complete absence of progeny. If this process were to occur in human lice, it could explain the low
580 number of hybrids; however, because *Wolbachia* endosymbiont is transmitted together with the
581 mitochondrial genome through the egg cytoplasm, endosymbionts and mtDNA should reflect a
582 similar history. In our study, mitochondrial lineages were grouped in two mitochondrial
583 haplogroups. We found individuals of both the A and B haplogroups in both nuclear clusters (cI
584 and cII). This is consistent with our initial findings [16] in that there is no one-to-one correlation
585 between mitochondrial haplogroups (A and B) and nuclear genetic clusters.
586 Insect species with nutritionally incomplete diets (e.g., phloem or blood), such as human lice
587 often harbor mutualistic bacteria that synthesize missing nutrients [86–88]. As *Wolbachia*, these
588 endosymbionts are intracellular, and are transmitted vertically from mother to offspring. The
589 human lice endosymbiont is *Candidatus* Riesia pediculicola [89], and based on genomic studies,
590 this endosymbiont carries a small plasmid needed for vitamin B5 synthesis [90] suggesting that
591 this bacterial symbiont supplies the lice with vitamin B-complex, which the louse cannot produce
592 by itself [91,92]. In Boyd et al. [90] they compared the genomic sequence of *Candidatus* Riesia
593 pediculicola found in lice from Cambodia (cI-mtDNA A) and the Netherlands (cII-mtDNA A)

594 with the lab strain, and they found no indels between these two genomes. However, when they
595 included the genome obtained from louse from Honduras (cI-mtDNA B), the authors found
596 indels [90], reflecting that correlation between mtDNA clades and the endosymbiont. Here
597 again, since there is no correlation between the mitochondrial DNA and the microsatellites,
598 *Candidatus* Riesia pediculicola cannot help to understand the low prevalence of louse hybrids.
599 An alternative explanation could be related to an unusual paternal genome elimination (PGE)
600 that has been reported in other members of Psocodea and has been suggested as the mode of sex
601 determination for the entire lineage that includes Liposcelididae and Phthiraptera [93]. Earlier
602 studies have shown that some, but not all, male of the clothing lice *P. humanus* only transmitted
603 their maternal copy of microsatellite markers [94]. Particularly, the allele transmission in
604 females followed Mendelian expectations but it was non-Mendelian in males: heterozygous male
605 parents systematically passed on one of their two alleles to their offspring, the one from the
606 mother [94]. This mode of inheritance is consistent with paternal genome elimination (PGE), a
607 type of pseudohaplodiploid reproduction found across several arthropod orders in which males
608 do not transmit paternally inherited alleles to their offspring [95]. A more recent study [96]
609 assessed this phenomenon in both clothing and head lice. They found genome-wide male
610 transmission ratio distortion in both ecotypes of *P. humanus*: males exclusively (or, in some
611 cases, preferentially) transmit only one of their alleles to their offspring, in agreement with
612 previous findings from McMeniman and Barker (2006) (92). In our study, out of the 33 hybrids,
613 19 were females, 10 males and 4 were juveniles. Under the PGE scenario, all those 10 male
614 hybrids will only pass their maternally inherited alleles to the offspring, this will lessen
615 admixture in the populations. However, as human louse sex ratios are female-biased, PGE

616 would have a lesser impact. Future studies including more louse samples and genomic tools may
617 help to understand the low number of hybrid lice.
618 Another cause of genetic isolation may be related to epigenetic effects. In the broad sense,
619 epigenetics refers to the differential expression of genes and phenotypes that cannot be explained
620 by changes in the DNA sequence. Epigenetic mechanisms like DNA methylation and histone
621 modification are well known to influence gene regulation, phenotypic plasticity, development,
622 and the preservation of genome integrity. However, a growing number of studies are showing
623 direct and indirect contributions of epigenetics to evolutionary processes since some epigenetic
624 states are transmitted intergenerationally and affect the phenotype of offspring [97]. One of the
625 modes for transgenerational epigenetic inheritance is through "genomic imprinting" for which a
626 gene is distinguished as patrigene or matrigene, depending on its paternal or maternal inheritance
627 [98]. This mode is often referred as parent-specific gene expression (PSGE) and has been well
628 study among social insects, where it is shown that epigenetic changes are functionally
629 responsible for caste formation [99]. Among lice epigenetics has not been studied and opens an
630 exciting area of new research. There could also be a relationship between the mechanism
631 mentioned in the paragraph above related to paternal genome elimination (PGE) and parent-
632 specific gene expression (PSGE) that could be interfering with the viability and/or fecundity of
633 hybrid lice, a question that warrants epigenomic studies in human lice.

## 634 Final remarks

635 This work is part of the growing number of studies looking into human parasites as probes
636 for human migration around the world. We focused on the human louse, which is one of the
637 oldest known human ectoparasites as indicated by archeological remains, the worldwide co-
638 distribution of both humans and lice, and coevolutionary studies. Here, we used microsatellites

639 to assess human louse genetic diversity worldwide, which revealed the presence of two deep
640 divergent nuclear clusters, with very limited and recent admixture, mostly in the New World.
641 We suggest that this pattern is reflecting human migrations during the population of the New
642 World and the most recent European migration. However, we hypothesize that there is an
643 undetermined mechanism that is preventing admixture between these two divergent nuclear
644 clusters, probably epigenetic. We also provide a new approach to evaluate demographic
645 scenarios through DIY-ABC models by using data from the human host, which could guide the
646 development of new analyses in other host-parasite systems. Due to the use of microsatellites
647 that are known to be fast-evolving markers, our analyses are more suited to recent events. Other
648 slower evolving markers could provide insights into more ancient events. Further studies in
649 human lice including more samples, whole genomic and epigenomic approaches could provide
650 new knowledge about louse evolution as well as their human host.

651 **Acknowledgments**

652 We would like to thank former University of Florida undergraduate students: Gebreyes Kassu
653 and Jackie Fane for their help in the lab and former UF postdoc Marie de Gracia Coquerel,
654 currently at the Donald Danforth Plant Science Center for comments in early version of the
655 manuscript. We greatly appreciate the contribution of Katie Shepherd (Founder & CEO of The
656 Shepherd Institute for Lice Solutions), Marieta A. H. Braks (Laboratory for Zoonoses and
657 Environmental, National Institute for Public Health and the Environment, Netherlands) and
658 collectors worldwide for providing lice used in this study. We would like to thank Aida Miró-
659 Herrans and Niyomi House (University of Florida, USA) and Jan Štefka (Institute of Parasitology,
660 Biology Centre CAS, Czech Republic) for comments on earlier version on this manuscript.

661 **References**

662 1. Light JE, Toups MA, Reed DL. What's in a name: The taxonomic status of human head and
663 body lice. Mol Phylogenet Evol. 2008;47: 1203–1216. doi:10.1016/j.ympev.2008.03.014

664 2. Reed DL, Smith VS, Hammond SL, Rogers AR. Genetic analysis of lice supports direct
665 contact between modern and archaic humans. PLoS Biol. 2004;2: e340.
666 doi:10.1371/journal.pbio.0020340

667 3. Raoult D, Reed DL, Dittmar K, Kirchman JJ, Rolain J, Guillen S, et al. Molecular
668 Identification of Lice from Pre-Columbian Mummies. J Infect Dis. 2008;197: 535–543.
669 doi:10.1086/526520

670 4. Reed DL, Light JE, Allen JM, Kirchman JJ. Pair of lice lost or parasites regained: the
671 evolutionary history of anthropoid primate lice. BMC Biol. 2007;5: 7. doi:10.1186/1741-
672 7007-5-7

673 5. Araújo A, Ferreira LF, Guidon N, Maues da Serra Freire N, Reinhard KJ, Dittmar K. Ten
674 Thousand Years of Head Lice Infection. Parasitol Today. 2000;16: 269.
675 doi:10.1016/S0169-4758(00)01694-X

676 6. Mumcuoglu KY. Human lice: Pediculus and Pthirus. Springer, Berlin, Germany; 2008.

677 7. Page RDM, Holmes EC. Molecular evolution: a phylogenetic approach. Oxford ; Malden,
678 MA: Blackwell Science; 1998.

679 8. Hafner MS, Nadler SA. Phylogenetic trees support the coevolution of parasites and their
680 hosts. Nature. 1988;332: 258–259. doi:10.1038/332258a0

681 9. Hafner MS, Sudman PD, Villablanca FX, Spradling TA, Demastes JW, Nadler SA.
682 Disparate rates of molecular evolution in cospeciating hosts and parasites. Science.
683 1994;265: 1087–1090.

684 10. Perry GH. Parasites and human evolution. Evol Anthropol Issues News Rev. 2015;23: 218–
685 228. doi:10.1002/evan.21427

686 11. Nozais J-P. The origin and dispersion of human parasitic diseases in the old world (Africa,
687 Europe and Madagascar). Mem Inst Oswaldo Cruz. 2003;98 Suppl 1: 13–19.

688 12. Reed DL, Toups MA, Light JE, Allen JM, Flanagin S. Lice and other parasites as markers
689 of primate evolutionary history. Primate Parasite Ecology: The Dynamics and Study of
690 Host-Parasite Relationships. Cambridge University Press; 2009. p. 531.

691 13. Dittmar K, Araújo A, Reinhard KJ. The Study of Parasites Through Time:
692 Archaeoparasitology and Paleoparasitology. A companion to paleopathology. Chichester,
693 West Sussex; Malden, MA: Wiley-Blackwell; 2012.

694 14. Kittler R, Kayser M, Stoneking M. Molecular evolution of Pediculus humanus and the
695 origin of clothing. Curr Biol CB. 2003;13: 1414–1417.

696 15. Toups MA, Kitchen A, Light JE, Reed DL. Origin of clothing lice indicates early clothing
697 use by anatomically modern humans in Africa. Mol Biol Evol. 2011;28: 29–32.
698 doi:10.1093/molbev/msq234

699 16. Ascunce MS, Toups MA, Kassu G, Fane J, Scholl K, Reed DL. Nuclear Genetic Diversity
700 in Human Lice (*Pediculus humanus*) Reveals Continental Differences and High
701 Inbreeding among Worldwide Populations. PLOS ONE. 2013;8: e57619.
702 doi:10.1371/journal.pone.0057619

703 17. Ascunce MS, Fane J, Kassu G, Toloza AC, Picollo MI, González-Oliver A, et al.
704 Mitochondrial diversity in human head louse populations across the Americas. Am J Phys
705 Anthropol. 2013;152: 118–129. doi:10.1002/ajpa.22336

706 18. Reed DL, Allen JM, Toups MA, Boyd BM, Ascunce MS. The study of primate evolution
707 from a lousy perspective. In: Krasnov BR, Littlewood DTJ, Morand S, editors. Parasite
708 Diversity and Diversification: Evolutionary Ecology Meets Phylogenetics. Cambridge:
709 Cambridge University Press; 2015. pp. 202–214. doi:10.1017/CBO9781139794749.014

710 19. Pedersen MW, Antunes C, De Cahsan B, Moreno-Mayar JV, Sikora M, Vinner L, et al.
711 Ancient human genomes and environmental DNA from the cement attaching 2,000 year-
712 old head lice nits. Mol Biol Evol. 2022; msab351. doi:10.1093/molbev/msab351

713 20. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A Draft Sequence
714 of the Neandertal Genome. Science. 2010;328: 710–722. doi:10.1126/science.1188021

715 21. Light JE, Allen JM, Long LM, Carter TE, Barrow L, Suren G, et al. Geographic
716 Distributions and Origins of Human Head Lice (Pediculus humanus capitis) Based on
717 Mitochondrial Data. J Parasitol. 2008;94: 1275–1281. doi:10.1645/GE-1618.1

718 22. Amanzougaghene N, Mumcuoglu KY, Fenollar F, Alfi S, Yesilyurt G, Raoult D, et al. High
719 Ancient Genetic Diversity of Human Lice, Pediculus humanus, from Israel Reveals New
720 Insights into the Origin of Clade B Lice. PLOS ONE. 2016;11: e0164659.
721 doi:10.1371/journal.pone.0164659

722 23. Boutellis A, Abi-Rached L, Raoult D. The origin and distribution of human lice in the
723 world. Infect Genet Evol. 2014;23: 209–217. doi:10.1016/j.meegid.2014.01.017

724 24. Ashfaq M, Prosser S, Nasir S, Masood M, Ratnasingham S, Hebert PDN. High diversity
725 and rapid diversification in the head louse, Pediculus humanus (Pediculidae: Phthiraptera).
726 Sci Rep. 2015;5. doi:10.1038/srep14188

727 25. Sunantaraporn S, Sanprasert V, Pengsakul T, Phumee A, Boonserm R, Tawatsin A, et al.
728 Molecular survey of the head louse Pediculus humanus capitis in Thailand and its
729 potential role for transmitting Acinetobacter spp. Parasit Vectors. 2015;8: 127.
730 doi:10.1186/s13071-015-0742-4

731 26. Drali R, Davoust B, Shako J-C, Raoult D, Diatta G. A New Clade of African Body and
732 Head Lice Infected by Bartonella quintana and Yersinia pestis—Democratic Republic of
733 the Congo. Am J Trop Med Hyg. 2015;93: 990–993. doi:10.4269/ajtmh.14-0686

734 27. Amanzougaghene N, Fenollar F, Davoust B, Djossou F, Ashfaq M, Bitam I, et al.
735 Mitochondrial diversity and phylogeographic analysis of Pediculus humanus reveals a new
736 Amazonian clade "F." Infect Genet Evol. 2019;70: 1–8. doi:10.1016/j.meegid.2019.02.006

737 28. Amanzougaghene N, Fenollar F, Raoult D, Mediannikov O. Where Are We With Human
738 Lice? A Review of the Current State of Knowledge. Front Cell Infect Microbiol. 2020;9:
739 474. doi:10.3389/fcimb.2019.00474

740 29. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis
741 (MEGA) software version 4.0. Mol Biol Evol. 2007;24: 1596–1599.
742 doi:10.1093/molbev/msm092

743 30. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing
744 phylogenetic trees. Mol Biol Evol. 1987;4: 406–425.

745 31. Swofford DL. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sunderland,
746 Massachusetts: Sinauer Associates; 2003.

747 32. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform
748 population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010;10: 564–
749 567. doi:10.1111/j.1755-0998.2010.02847.x

750 33. Excoffier L, Laval G, Schneider S. Arlequin (ver. 3.0): An integrated software package for
751 population genetics data analysis. Evol Bioinforma Online. 2005;1: 47–50.

752 34. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus
753 genotype data. Genetics. 2000;155: 945–959.

754 35. Peakall R, Smouse PE. genalex 6: genetic analysis in Excel. Population genetic software for
755 teaching and research. Mol Ecol Notes. 2006;6: 288–295. doi:10.1111/j.1471-
756 8286.2005.01155.x

757 36. Peakall R, Smouse P. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software
758 for teaching and research - an update. Bioinforma Oxf Engl. 2012 [cited 3 Aug 2012].
759 doi:10.1093/bioinformatics/bts460

760 37. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new
761 method for the analysis of genetically structured populations. BMC Genet. 2010;11: 94.
762 doi:10.1186/1471-2156-11-94

763 38. VäHä J-P, Primmer CR. Efficiency of model-based Bayesian methods for detecting hybrid
764 individuals under different hybridization scenarios and with different numbers of loci:
765 GENETIC DETECTION OF HYBRIDS. Mol Ecol. 2005;15: 63–72. doi:10.1111/j.1365-
766 294X.2005.02773.x

767 39. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the
768 software STRUCTURE: a simulation study. Mol Ecol. 2005;14: 2611–2620.
769 doi:10.1111/j.1365-294X.2005.02553.x

770 40. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: a website and program for
771 visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet
772 Resour. 2012;4: 359–361. doi:10.1007/s12686-011-9548-7

773 41. Rosenberg NA. distruct: a program for the graphical display of population structure. Mol
774 Ecol Notes. 2004;4: 137–138. doi:10.1046/j.1471-8286.2003.00566.x

775 42. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers.
776 Bioinformatics. 2008;24: 1403–1405. doi:10.1093/bioinformatics/btn129

777 43. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation
778 For Statistical Computing, Vienna, Austria. 2015.

779 44. Szpiech ZA, Jakobsson M, Rosenberg NA. ADZE: a rarefaction approach for counting
780 alleles private to combinations of populations. Bioinformatics. 2008;24: 2498–2504.
781 doi:10.1093/bioinformatics/btn478

782 45. Hurlbert SH. The Nonconcept of Species Diversity: A Critique and Alternative Parameters.
783 Ecology. 1971;52: 577–586. doi:10.2307/1934145

784 46. Petit RJ, El Mousadik A, Pons O. Identifying Populations for Conservation on the Basis of
785 Genetic Markers. Conserv Biol. 2008;12: 844–855. doi:10.1111/j.1523-
786 1739.1998.96489.x

787 47. Kruskal WH, Wallis WA. Use of Ranks in One-Criterion Variance Analysis. J Am Stat
788 Assoc. 1952;47: 583. doi:10.2307/2280779

789 48. StatSoft, Inc. STATISTICA (Data Analysis Software System). Tulsa, OK, USA; 2001.

790 49. Beerli P, Felsenstein J. Maximum likelihood estimation of a migration matrix and effective
791 population sizes in n subpopulations by using a coalescent approach. Proc Natl Acad Sci.
792 2001;98: 4563–4568. doi:10.1073/pnas.081068098

793 50. Beerli P, Palczewski M. Unified Framework to Evaluate Panmixia and Migration Direction
794 Among Multiple Sampling Locations. Genetics. 2010;185: 313–326.
795 doi:10.1534/genetics.109.112532

796 51. Beaumont MA, Zhang W, Balding DJ. Approximate Bayesian computation in population
797 genetics. Genetics. 2002;162: 2025–2035.

798 52. Cornuet J-M, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, et al. DIYABC
799 v2.0: a software to make approximate Bayesian computation inferences about population
800 history using single nucleotide polymorphism, DNA sequence and microsatellite data.
801 Bioinformatics. 2014;30: 1187–1189. doi:10.1093/bioinformatics/btt763

802 53. Toloza AC, Laguna MF, Ortega-Insaurralde I, Vassena C, Risau-Gusman S. Insights About
803 Head Lice Transmission From Field Data and Mathematical Modeling. J Med Entomol.
804 2018;55: 929–937. doi:10.1093/jme/tjy026

805 54. Schlötterer C. Evolutionary dynamics of microsatellite DNA. Chromosoma. 2000;109:
806 365–371. doi:10.1007/s004120000089

807 55. Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TF, Aquadro CF. The
808 mutation rates of di-, tri- and tetranucleotide repeats in Drosophila melanogaster. Mol Biol
809 Evol. 1998;15: 1751–1760.

810 56. Stolle E, Kidner JH, Moritz RFA. Patterns of Evolutionary Conservation of Microsatellites
811 (SSRs) Suggest a Faster Rate of Genome Evolution in Hymenoptera Than in Diptera.
812 Genome Biol Evol. 2013;5: 151–162. doi:10.1093/gbe/evs133

813 57. Bertorelle G, Benazzo A, Mona S. ABC as a flexible framework to estimate demography
814 over space and time: some cons, many pros. Mol Ecol. 2010;19: 2609–2625.
815 doi:10.1111/j.1365-294X.2010.04690.x

816 58. Cornuet J-M, Santos F, Beaumont MA, Robert CP, Marin J-M, Balding DJ, et al. Inferring
817 population history with DIY ABC: a user-friendly approach to approximate Bayesian
818 computation. Bioinformatics. 2008;24: 2713–2719. doi:10.1093/bioinformatics/btn514

819 59. Cornuet J-M, Ravignie V, Estoup A. Inference on population history and model checking
820 using DNA sequence and microsatellite data with the software DIYABC (v1.0). BMC
821 Bioinformatics. 2010;11: 401. doi:10.1186/1471-2105-11-401

822 60. Torroni A, Sukernik RI, Schurr TG, Starikorskaya YB, Cabell MF, Crawford MH, et al.
823 mtDNA variation of aboriginal Siberians reveals distinct genetic affinities with Native
824 Americans. Am J Hum Genet. 1993;53: 591–608.

825 61. Tamm E, Kivisild T, Reidla M, Metspalu M, Glenn Smith D, Mulligan CJ, et al. Beringian
826 standstill and spread of Native American founders. PLoS ONE. 2007;2: e829.

827 62. Raghavan M, Steinrücken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M, et al.
828 Genomic evidence for the Pleistocene and recent population history of Native Americans.
829 Science. 2015;349: aab3884. doi:10.1126/science.aab3884

830 63. Falush D. Traces of human migrations in Helicobacter pylori populations. Science.
831 2003;299: 1582–1585. doi:10.1126/science.1080857

832 64. Moodley Y, Linz B, Bond RP, Nieuwoudt M, Soodyall H, Schlebusch CM, et al. Age of the
833 Association between Helicobacter pylori and Man. Ochman H, editor. PLoS Pathog.
834 2012;8: e1002693. doi:10.1371/journal.ppat.1002693

835 65. Moodley Y, Linz B, Yamaoka Y, Windsor HM, Breurec S, Wu J-Y, et al. The peopling of
836 the Pacific from a bacterial perspective. Science. 2009;323: 527–530.
837 doi:10.1126/science.1166083

838 66. Ghose C, Perez-Perez GI, Dominguez-Bello M-G, Pride DT, Bravi CM, Blaser MJ. East
839 Asian genotypes of Helicobacter pylori strains in Amerindians provide evidence for its
840 ancient human carriage. Proc Natl Acad Sci. 2002;99: 15107–15111.
841 doi:10.1073/pnas.242574599

842 67. Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA, et al. Unifying the
843 Epidemiological and Evolutionary Dynamics of Pathogens. Science. 2004;303: 327–332.
844 doi:10.1126/science.1090727

845 68. Araújo A, Reinhard KJ, Ferreira LF, Gardner SL. Parasites as probes for prehistoric human
846 migrations? Trends Parasitol. 2008;24: 112–115. doi:10.1016/j.pt.2007.11.007

847 69. Takano-Lee M, Edman JD, Mullens BA, Clark JM. Transmission potential of the human
848 head louse, Pediculus capitis (Anoplura: Pediculidae). Int J Dermatol. 2005;44: 811–816.
849 doi:10.1111/j.1365-4632.2005.02418.x

850 70. Rózsa L, Apari P. Why infest the loved ones – inherent human behaviour indicates former
851 mutualism with head lice. Parasitology. 2012;139: 696–700.
852 doi:10.1017/S0031182012000017

853 71. Mumcuoglu KY, Pollack RJ, Reed DL, Barker SC, Gordon S, Toloza AC, et al.
854 International recommendations for an effective control of head louse infestations. Int J
855 Dermatol. 2021;60: 272–280. doi:10.1111/ijd.15096

856 72. Johnson KP, Allen JM, Olds BP, Mugisha L, Reed DL, Paige KN, et al. Rates of genomic
857 divergence in humans, chimpanzees and their lice. Proc R Soc B Biol Sci. 2014;281:
858 20132174. doi:10.1098/rspb.2013.2174

859 73. Ricklefs RE, Outlaw DC. A Molecular Clock for Malaria Parasites. Science. 2010;329:
860 226–229. doi:10.1126/science.1188954

861 74. Olival KJ, Dick CW, Simmons NB, Morales JC, Melnick DJ, Dittmar K, et al. Lack of
862 population genetic structure and host specificity in the bat fly, Cyclopodia horsfieldi,
863 across species of Pteropus bats in Southeast Asia. Parasit Vectors. 2013;6: 231.
864 doi:10.1186/1756-3305-6-231

865 75. van Schaik J, Dekeukeleire D, Kerth G. Host and parasite life history interplay to yield
866 divergent population genetic structures in two ectoparasites living on the same bat species.
867 Mol Ecol. 2015;24: 2324–2335. doi:10.1111/mec.13171

868 76. Mazé-Guilmo E, Blanchet S, McCoy KD, Loot G. Host dispersal as the driver of parasite
869 genetic structure: a paradigm lost? Thrall P, editor. Ecol Lett. 2016;19: 336–347.
870 doi:10.1111/ele.12564

871 77. Talbot B, Vonhof MJ, Broders HG, Fenton B, Keyghobadi N. Comparative analysis of
872 landscape effects on spatial genetic structure of the big brown bat and one of its cimicid
873 ectoparasites. Ecol Evol. 2017;7: 8210–8219. doi:10.1002/ece3.3329

874 78. Speer KA, Luetke E, Bush E, Sheth B, Gerace A, Quicksall Z, et al. A Fly on the Cave
875 Wall: Parasite Genetics Reveal Fine-Scale Dispersal Patterns of Bats. J Parasitol.
876 2019;105: 555. doi:10.1645/19-20

877 79. Galbreath KE, Hoberg EP. Return to Beringia: parasites reveal cryptic biogeographic
878 history of North American pikas. Proc R Soc B Biol Sci. 2012;279: 371–378.
879 doi:10.1098/rspb.2011.0482

880 80. Whiteman NK, Kimball RT, Parker PG. Co-phylogeography and comparative population
881 genetics of the threatened Galápagos hawk and three ectoparasite species: ecology shapes
882 population histories within parasite communities. Mol Ecol. 2007;16: 4759–4773.
883 doi:10.1111/j.1365-294X.2007.03512.x

884 81. Sweet AD, Wilson RE, Sonsthagen SA, Johnson KP. Lousy grouse: Comparing
885 evolutionary patterns in Alaska galliform lice to understand host evolution and host–
886 parasite interactions. Ecol Evol. 2020;10: 8379–8393. doi:10.1002/ece3.6545

887 82. Werren JH, Baldo L, Clark ME. Wolbachia: master manipulators of invertebrate biology.
888 Nat Rev Microbiol. 2008;6: 741–751. doi:10.1038/nrmicro1969

889 83. Perotti MA, Catalá SS, Ormeño A del V, Zelazowska M, Biliński SM, Braig HR. The sex
890 ratio distortion in the human head louse is conserved over time. BMC Genet. 2004;5: 10.
891 doi:10.1186/1471-2156-5-10

892 84. Kyei-Poku GK, Colwell DD, Coghlin P, Benkel B, Floate KD. On the ubiquity and
893 phylogeny of Wolbachia in lice. Mol Ecol. 2005;14: 285–294. doi:10.1111/j.1365-
894 294X.2004.02409.x

895 85. Covacin C, Barker SC. Supergroup F Wolbachia bacteria parasitise lice (Insecta:
896 Phthiraptera). Parasitol Res. 2007;100: 479–485. doi:10.1007/s00436-006-0309-6

897 86. Baumann P. Biology of Bacteriocyte-Associated Endosymbionts of Plant Sap-Sucking
898 Insects. Annu Rev Microbiol. 2005;59: 155–189.
899 doi:10.1146/annurev.micro.59.030804.121041

900 87. López-Sánchez MJ, Neef A, Peretó J, Patiño-Navarrete R, Pignatelli M, Latorre A, et al.
901 Evolutionary Convergence and Nitrogen Metabolism in Blattabacterium strain Bge,
902 Primary Endosymbiont of the Cockroach Blattella germanica. Bordenstein SR, editor.
903 PLoS Genet. 2009;5: e1000721. doi:10.1371/journal.pgen.1000721

904 88. Manzano-Marín A, Oceguera-Figueroa A, Latorre A, Jiménez-García LF, Moya A. Solving
905 a Bloody Mess: B-Vitamin Independent Metabolic Convergence among
906 Gammaproteobacterial Obligate Endosymbionts from Blood-Feeding Arthropods and the
907 Leech *Haementeria officinalis*. Genome Biol Evol. 2015;7: 2871–2884.
908 doi:10.1093/gbe/evv188

909 89. Sasaki-Fukatsu K, Koga R, Nikoh N, Yoshizawa K, Kasai S, Mihara M, et al. Symbiotic
910 Bacteria Associated with Stomach Discs of Human Lice. Appl Environ Microbiol.
911 2006;72: 7349–7352. doi:10.1128/AEM.01429-06

912 90. Boyd BM, Allen JM, Nguyen N-P, Vachaspati P, Quicksall ZS, Warnow T, et al. Primates,
913 Lice and Bacteria: Speciation and Genome Evolution in the Symbionts of Hominid Lice.
914 Mol Biol Evol. 2017;34: 1743–1757. doi:10.1093/molbev/msx117

915 91. Puchta O. Experimentelle Untersuchungen uber die Bedeutung der Symbiose der
916 Kleiderlaus Pediculus Vestimenti Burm. Z Parasitenkd. 1955;17.
917 doi:10.1007/BF00260226

918 92. Perotti M, Kirkness E, Reed D, Braig H. Endosymbionts of lice. Insect Symbiosis. Bourtzis
919 K Mill T Ed. 2009;3.

920 93. Yoshizawa K, Johnson KP. How stable is the "Polyphyly of Lice" hypothesis (Insecta:
921 Psocodea)?: A comparison of phylogenetic signal in multiple genes. Mol Phylogenet Evol.
922 2010;55: 939–951. doi:10.1016/j.ympev.2010.02.026

923 94. McMeniman CJ, Barker SC. Transmission ratio distortion in the human body louse,
924 Pediculus humanus (Insecta: Phthiraptera). Heredity. 2006;96: 63–68.
925 doi:10.1038/sj.hdy.6800760

926 95. Normark BB. The Evolution of Alternative Genetic Systems in Insects. Annu Rev Entomol.
927 2003;48: 397–423. doi:10.1146/annurev.ento.48.091801.112703

928 96. De La Filia AG, Andrewes S, Clark JM, Ross L. The unusual reproductive system of head
929 and body lice ( *Pediculus humanus* ): Paternal genome loss in human lice males. Med Vet
930 Entomol. 2018;32: 226–234. doi:10.1111/mve.12287

931 97. Ashe A, Colot V, Oldroyd BP. How does epigenetics influence the course of evolution?
932 Philos Trans R Soc B Biol Sci. 2021;376: 20200111. doi:10.1098/rstb.2020.0111

933 98. Patten MM, Ross L, Curley JP, Queller DC, Bonduriansky R, Wolf JB. The evolution of
934 genomic imprinting: theories, predictions and empirical tests. Heredity. 2014;113: 119–
935 128. doi:10.1038/hdy.2014.29

936 99. Oldroyd BP, Yagound B. The role of epigenetics, particularly DNA methylation, in the
937 evolution of caste in insect societies. Philos Trans R Soc B Biol Sci. 2021;376: 20200115.
938 doi:10.1098/rstb.2020.0115

939 100. Drali R, Mumcuoglu K, Raoult D. Human Lice in Paleoentomology and
940 Paleomicrobiology. Drancourt M, Raoult D, editors. Microbiol Spectr. 2016;4.
941 doi:10.1128/microbiolspec.PoH-0005-2014

942 101. Finlayson C. Biogeography and evolution of the genus Homo. Trends Ecol Evol. 2005;20:
943 457–463. doi:10.1016/j.tree.2005.05.019

944 102. McDougall I, Brown FH, Fleagle JG. Stratigraphic placement and age of modern humans
945 from Kibish, Ethiopia. Nature. 2005;433: 733–736. doi:10.1038/nature03258

946 103. Grün R, Stringer C, McDermott F, Nathan R, Porat N, Robertson S, et al. U-series and ESR
947 analyses of bones and teeth relating to the human burials from Skhul. J Hum Evol.
948 2005;49: 316–334. doi:10.1016/j.jhevol.2005.04.006

949 104. Krause J, Lalueza-Fox C, Orlando L, Enard W, Green RE, Burbano HA, et al. The Derived
950 FOXP2 Variant of Modern Humans Was Shared with Neandertals. Curr Biol. 2007;17:
951 1908–1912. doi:10.1016/j.cub.2007.10.008

952 105. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A Draft Sequence
953 of the Neandertal Genome. Science. 2010;328: 710–722. doi:10.1126/science.1188021

954 106. Higham T, Compton T, Stringer C, Jacobi R, Shapiro B, Trinkaus E, et al. The earliest
955 evidence for anatomically modern humans in northwestern Europe. Nature. 2011;479:
956 521–524. doi:10.1038/nature10484

957 107. Benazzi S, Douka K, Fornai C, Bauer CC, Kullmer O, Svoboda J, et al. Early dispersal of
958 modern humans in Europe and implications for Neanderthal behaviour. Nature. 2011;479:
959 525–528. doi:10.1038/nature10617

960 108. Arsuaga JL, Martínez I, Arnold LJ, Aranburu A, Gracia-Téllez A, Sharp WD, et al.
961 Neandertal roots: Cranial and chronological evidence from Sima de los Huesos. Science.
962 2014;344: 1358–1363. doi:10.1126/science.1253958

963 109. Devièse T, Karavanić I, Comeskey D, Kubiak C, Korlević P, Hajdinjak M, et al. Direct
964 dating of Neanderthal remains from the site of Vindija Cave and implications for the
965 Middle to Upper Paleolithic transition. Proc Natl Acad Sci. 2017;114: 10606–10611.
966 doi:10.1073/pnas.1709235114

967 110. Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E. Tracing the
968 peopling of the world through genomics. Nature. 2017;541: 302–310.
969 doi:10.1038/nature21347

970 111. Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, et al. A High-Coverage
971 Genome Sequence from an Archaic Denisovan Individual. Science. 2012;338: 222–226.
972 doi:10.1126/science.1224344

973 112. Wall JD, Yang MA, Jay F, Kim SK, Durand EY, Stevison LS, et al. Higher Levels of
974 Neanderthal Ancestry in East Asians than in Europeans. Genetics. 2013;194: 199–209.
975 doi:10.1534/genetics.112.148213

976 113. Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, et al. The complete
977 genome sequence of a Neanderthal from the Altai Mountains. Nature. 2014;505: 43–49.
978 doi:10.1038/nature12886

114. Kuhlwilm M, Gronau I, Hubisz MJ, de Filippo C, Prado-Martinez J, Kircher M, et al. Ancient gene flow from early modern humans into Eastern Neanderthals. Nature. 2016;530: 429–433. doi:10.1038/nature16544

115. Hershkovitz I, Weber GW, Quam R, Duval M, Grün R, Kinsley L, et al. The earliest modern humans outside Africa. Science. 2018;359: 456–459. doi:10.1126/science.aap8369

116. Douka K, Slon V, Jacobs Z, Ramsey CB, Shunkov MV, Derevianko AP, et al. Age estimates for hominin fossils and the onset of the Upper Palaeolithic at Denisova Cave. Nature. 2019;565: 640–644. doi:10.1038/s41586-018-0870-z

# Supporting information

**S1 Table. Detail of the human louse samples used in the current study.** Groups reflects their geographic origin and STRUCTURE clustering: Africa-Israel-Spain (AIS), South and Southeast Asia (SAs), East Asia (EAs), Europe (E), North America (NA), Mesoamerica (MA), Central America (CA), and South America (SAm).

**S2_S3 Tables. Prior distributions of population sizes parameters for the demographic scenarios described in Table 2 and Fig 2.** Between [ ] are the host values obtained from anthropological references. Louse values were estimated using anthropological references under different louse infection rate scenarios. S2 and S3 Tables values reflect an infection rate of 10% of hominin population having lice with 10 lice per head (S2 Table) and 50% of hominin population having lice with 50 lice per head (S3 Table). Times events are the same as in Table 2 of the manuscript.

**S4 Table. Genetic differentiation.** Louse pairwise genetic differentiation based on FST values from 15 microsatellite loci among sites where 8 or more lice were collected. Sample sizes are in parentheses following population name. Values below the diagonal indicate FST values, p-values are shown above the diagonal; * indicates significant at nominal alpha level of 0.05.

**S1 Fig. Discriminant analysis of principal components (DAPC).** Cluster membership of each louse is depicted by a different colour inside their 95% inertia ellipses. DA eigenvalues are shown in the inserted graphs at the bottom right, whereby the numbers of the discriminants plotted against each other are indicated by dark grey and the remaining discriminants retained for the analysis in light grey.

**S2 Fig. Number of clusters.** ΔK statistic of Evanno et al. [39] from STRUCTURE analysis from Fig. 3. K = 2 (ΔK = 858.853).

1012 **S3 Fig. STRUCTURE plot including only non-admixed lice (no hybrids) from both clusters**
1013 **I and II at K = 4.** An ancestry threshold value of q-values as suggested in [38] was used because
1014 it is efficient and accurate at differentiating between non-interbreed and hybrids. Only individuals
1015 with q-value between 0 and < 0.20 or > 0.8 and 1 are classified as non-admixed and were included
1016 in this analysis.
1017 **S4 Fig. Cluster I substructure.** STRUCTURE plot including only non-admixed lice (no hybrids).
1018 **S5 Fig. Cluster II substructure.** STRUCTURE plot including only non-admixed lice (no hybrids).
1019 **S6 Fig. Allelic richness.** Allelic richness per nuclear genetic cluster (cI, orange and cII, blue) using
1020 the rarefaction method implemented in the computer program ADZE (Allelic Diversity Analyzer)
1021 Version 1.0 [44]. Because allelic richness measurements considered sample size, the rarefaction
1022 method allows the estimation of allelic richness for different random subsamples of size "g" from
1023 the populations [45,46].
1024 **S7 Fig. Correspondence analysis (CA).** CA of the nuclear cluster membership frequency and the
1025 mitochondrial haplogroup of head lice.
1026 **S8 Fig. Gene flow.** Full migration model with four units defined by both the nuclear clusters
1027 inferred from STRUCTURE and the mitochondrial haplogroups: unit 1) cI – mtDNA A; unit 2)
1028 cII – mtDNA A, unit 3) cI – mtDNA B; and unit 4) and cII – mtDNA B. Values above and below
1029 the arrows indicate migration rates, arrows indicated directionality.
1030 **S9 Fig. Posterior probabilities for each of the seven demographic scenarios.** Graphs of
1031 posterior probabilities based on DIY-ABC analysis comparing the seven demographic scenarios
1032 when considering low and high louse infestation levels where 10% and 50% indicate that 10% and
1033 50% of the hominin population have 10 and 50 lice per head, respectively. Each scenario was

color-coded as follow Scenario 1 – green, Scenario 2 – red, Scenario 3 – light blue, Scenario 4 – pink, Scenario 5 – yellow, Scenario 6 – black, Scenario 7 – gray.

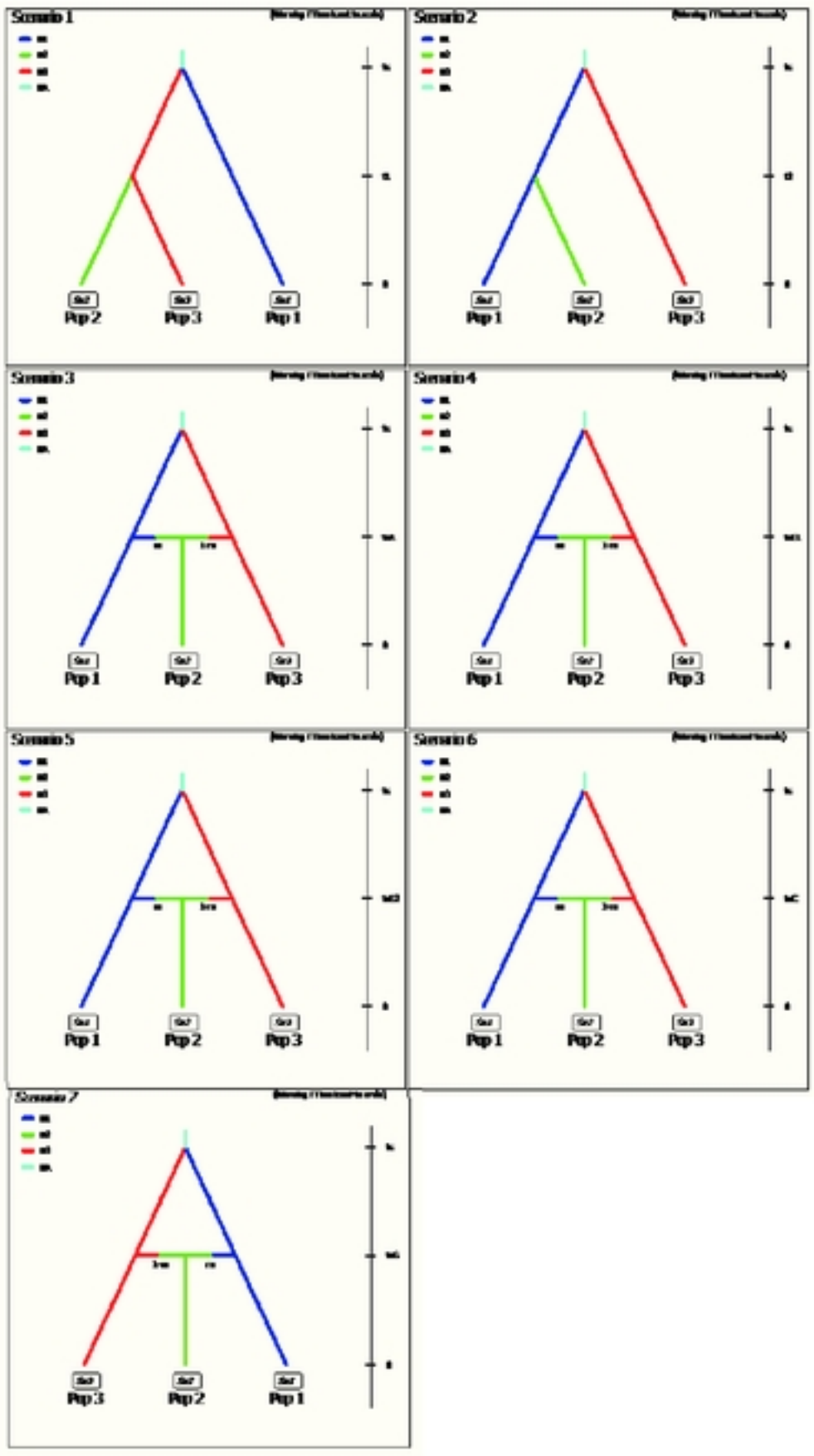

Scenario 1
Pop 2
Pop 3
Pop 1
Scenario 2
Pop 1
Pop 2
Pop 3
Scenario 3
Pop 1
Pop 2
Pop 3
Scenario 4
Pop 1
Pop 2
Pop 3
Scenario 5
Pop 1
Pop 2
Pop 3
Scenario 6
Pop 1
Pop 2
Pop 3
Scenario 7
Pop 3
Pop 2
Pop 1

Figure 2

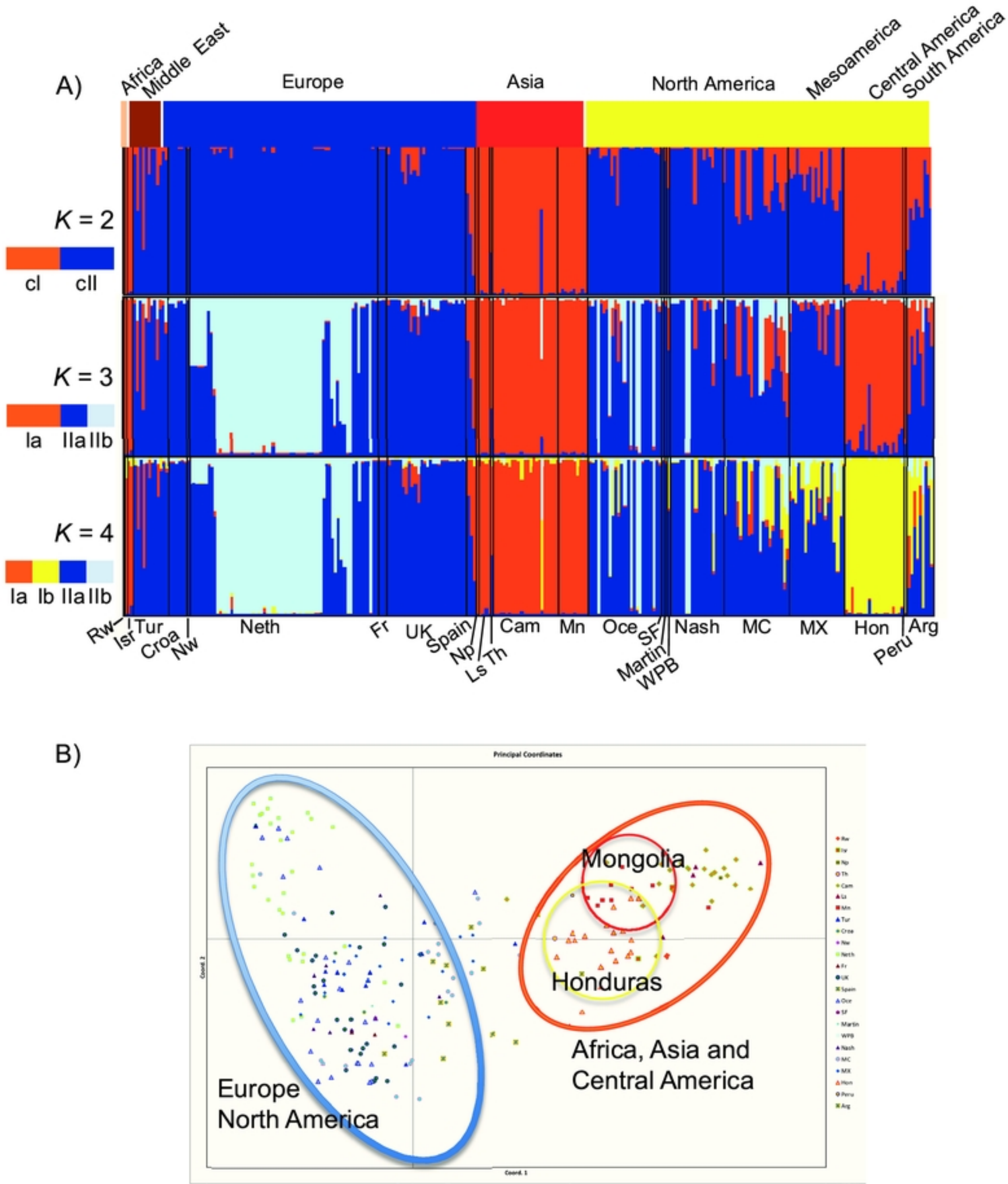

A)
Africa
Middle East
Europe
Asia
North America
Mesoamerica
Central America
South America
K = 2
cI
cII
K = 3
Ia
IIa
IIb
K = 4
Ia
Ib
IIa
IIb
Rw
Isr
Tur
Croa
Nw
Neth
Fr
UK
Spain
Np
Ls
Th
Cam
Mn
Oce
SF
Martin
WPB
Nash
MC
MX
Hon
Peru
Arg
B)
Principal Coordinates
Mongolia
Honduras
Africa, Asia and
Central America
Europe
North America

Figure 3

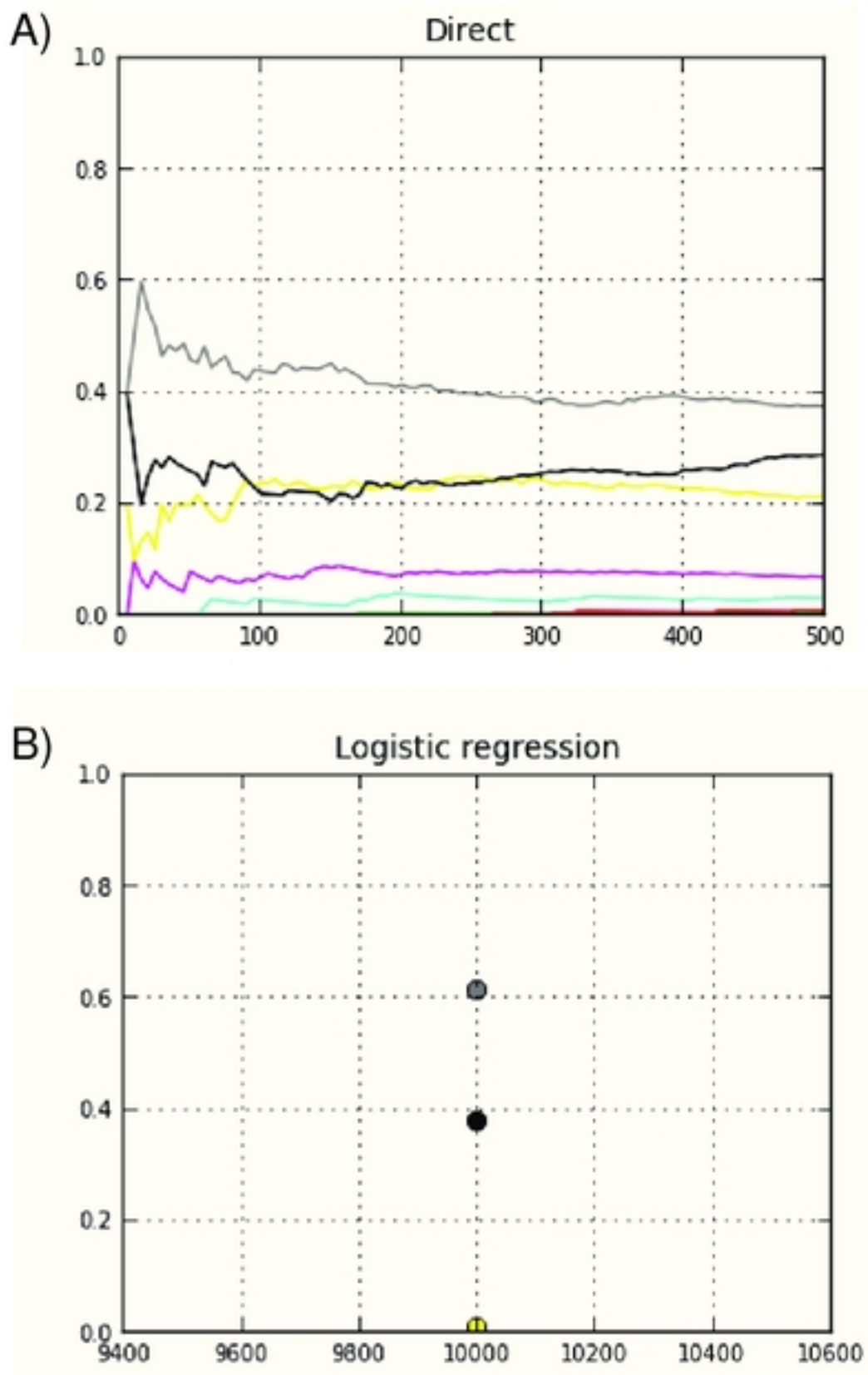

A)
Direct
1.0
0.8
0.6
0.4
0.2
0.0
0
100
200
300
400
500
B)
Logistic regression
1.0
0.8
0.6
0.4
0.2
0.0
9400
9600
9800
10000
10200
10400
10600

Figure 4

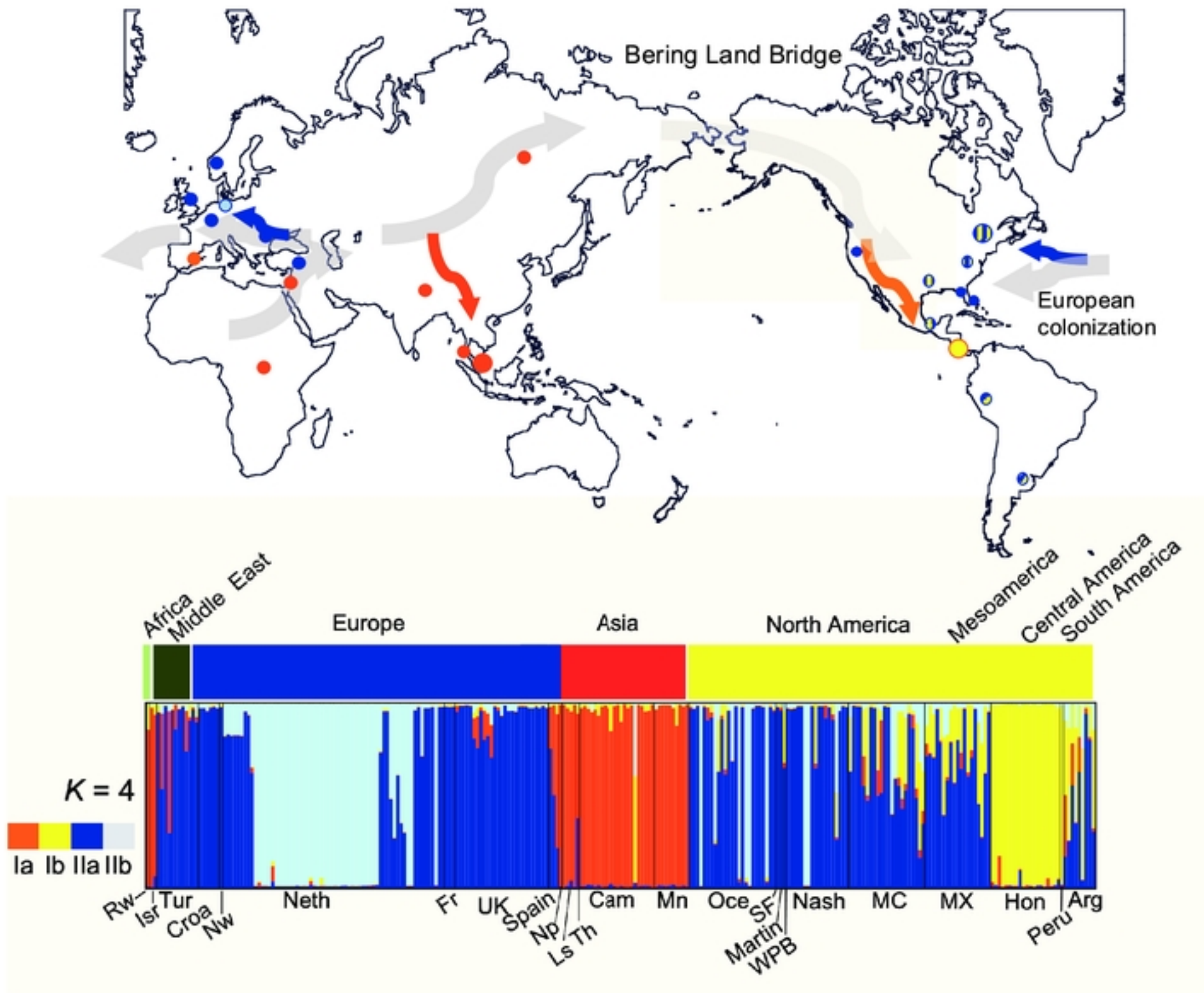
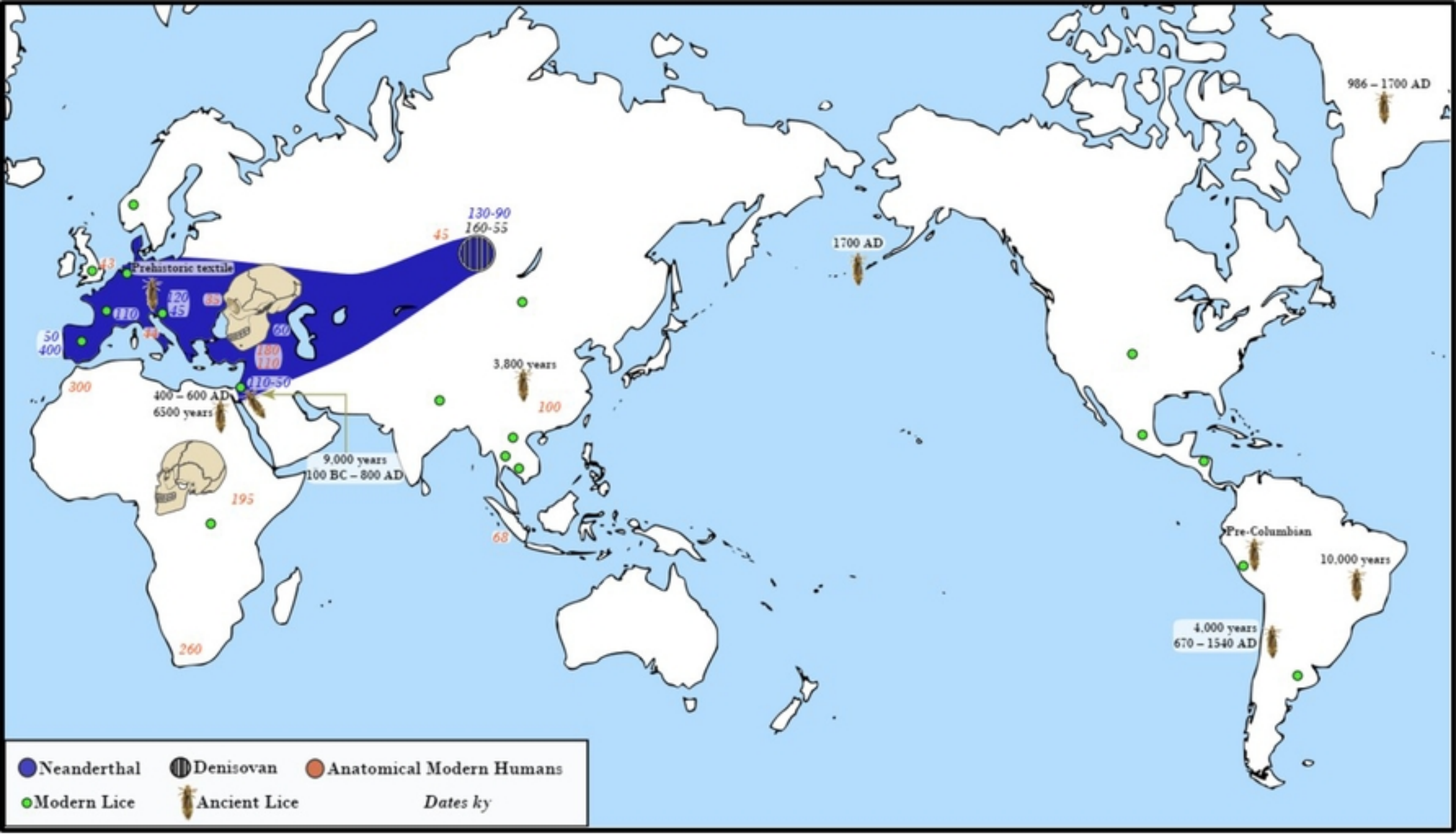

Bering Land Bridge
European colonization
Africa
Middle East
Europe
Asia
North America
Mesoamerica
Central America
South America
K = 4
Ia Ib IIa IIb
Rw
Isr
Tur
Croa
Nw
Neth
Fr
UK
Spain
Np
Ls
Th
Cam
Mn
Oce
SF
Martin
WPB
Nash
MC
MX
Hon
Peru
Arg

Figure 5



986 – 1700 AD
130-90
160-55
45
43
Prehistoric textile
120
45
35
110
44
60
50
400
180
110
110-50
300
400 – 600 AD
6500 years
3,800 years
100
9,000 years
100 BC – 800 AD
195
68
260
1700 AD
Pre-Columbian
10,000 years
4,000 years
670 – 1540 AD
Neanderthal
Denisovan
Anatomical Modern Humans
Modern Lice
Ancient Lice
Dates ky


Figure 1