

OPEN

The Early Peopling of the Philippines based on mtDNA

Miguel Arenas^{1,2,3*}, Amaya Gorostiza⁴, Juan Miguel Baquero^{3,5}, Elena Campoy⁵, Catarina Branco^{1,2}, Héctor Rangel-Villalobos⁶ & Antonio González-Martín⁵

Despite the efforts made to reconstruct the history of modern humans, there are still poorly explored regions that are key for understanding the phylogeography of our species. One of them is the Philippines, which is crucial to unravel the colonization of Southeast Asia and Oceania but where little is known about when and how the first humans arrived. In order to shed light into this settlement, we collected samples from 157 individuals of the Philippines with the four grandparents belonging to the same region and mitochondrial variants older than 20,000 years. Next, we analyzed the hypervariable I mtDNA region by approximate Bayesian computation based on extensive spatially explicit computer simulations to select among several migration routes towards the Philippines and to estimate population genetic parameters of this colonization. We found that the colonization of the Philippines occurred more than 60,000 years ago, with long-distance dispersal and from both north and south migration routes. Our results also suggest an environmental scenario especially optimal for humans, with large carrying capacity and population growth, in comparison to other regions of Asia. In all, our study suggests a rapid expansion of modern humans towards the Philippines that could be associated with the establishment of maritime technologies and favorable environmental conditions.

Traditionally it is considered that the first humans who colonized Oceania departed from Southeast Asia (SEA) and reached the austral continent through Sumatra, Java and Timor¹. According to this approach, the Philippine archipelago, constituted by more than 7,000 islands and located in a more northern region, would not be involved in the ancestral human migrations that colonized Oceania². However, there are some studies that contradict this approach. The fossil record confirms that our modern humans were in the Philippines at least 40,000–50,000 years ago^{3–5}, the genus *Homo* possibly 66,700 years ago^{6,7}. Another finding is the presence in the archipelago of *Negritos* groups related to the first migrations of *Homo sapiens* outside Africa^{4,8,9}. Indeed, several genetic studies detected an archaic substrate in current populations of the Philippines^{4,5,10–14}. Hence, the human presence in the Philippines during the first stages of expansion of anatomically modern *Homo sapiens* is very likely. The recent finding of *Homo luzonensis*⁷ in the Philippines does not provide new information on the origin and expansion process of our species although it does demonstrate the importance of the archipelago to understand the past of our lineage. Consequently, the Philippines probably presented a relevant role in the out of Africa expansion of modern humans to colonize the SEA and Oceania.

Despite the relative certainty concerning the timing of the entry into the Philippines, only some studies investigated the population genetics and evolutionary processes occurring in the first expansion of modern humans towards this region. Alves, *et al.*¹⁵ and Pugach, *et al.*¹⁶ found that Paleolithic populations in Eurasia and Oceania, respectively, could have expanded with long-distance dispersal (LDD) events, rather than through a progressive expansion along the landscape. However, little is known about LDD and the migration routes used to reach the Philippines and about the Palaeolithic population dynamics in this region. The Philippines is an especially interesting region to test these population processes because its colonization required an expansion through large stretches of sea (bottlenecks), which could be done using boats or rafts as proposed for other regions of the world^{17,18}.

In the present work, we investigated the timing of colonization, migration patterns and the population dynamics that occurred in the Philippine archipelago at the first stages of human expansion. Mitochondrial DNA

¹Department of Biochemistry, Genetics and Immunology, University of Vigo, Vigo, Spain. ²Biomedical Research Center (CINBIO), University of Vigo, Vigo, Spain. ³Centre for Molecular Biology "Severo Ochoa", Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain. ⁴Forensic Genetics and Identification Laboratory, GENOMICA S.A.U. Pharma Mar Group, Madrid, Spain. ⁵Department of Biodiversity, Ecology and Evolution, University Complutense of Madrid, Madrid, Spain. ⁶Institute of Research in Molecular Genetics, University of Guadalajara (CUCienega-UdeG), Ocotlán, Jalisco, Mexico. *email: marenas@uvigo.es

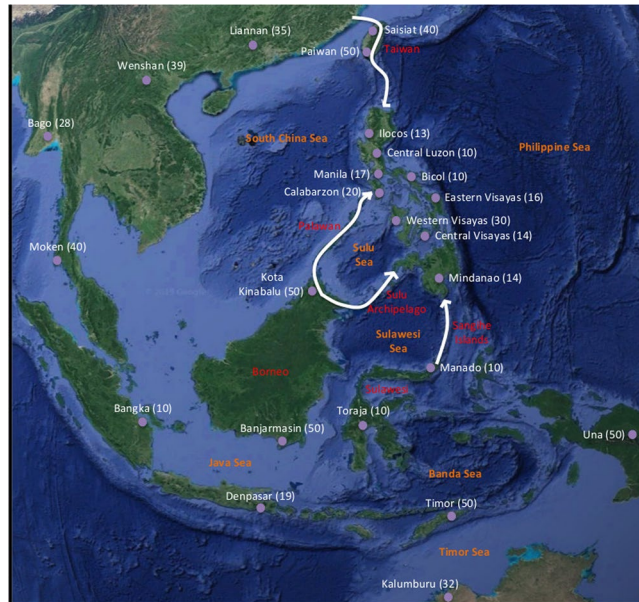


Figure 1. Studied landscape, samples and colonization routes. The figure shows the sampling locations applied in this study, including sample size (number of individuals) in parenthesis. The figure also shows four possible routes of early colonization of the Philippine archipelago: (1) northern route, from Taiwan through the Luzon Strait; (2) southwest route, from Borneo through Palawan; (3) central south route, from east of Borneo crossing the Sulu Archipelago towards western Mindanao; (4) southeast route, from Sulawesi through southern Mindanao through the Sangihe Islands. Combining these four routes we proposed seven possible colonization scenarios that are explored in this study: (i) migration allowing LDD events (*LDD*); (ii) All corridors (*AllCorr*), where the four routes are used; (iii) North corridor (*North*), colonization exclusively from the north; (iv) All south corridors (*AllSouth*); (v) Southwest corridor (*SW*); (vi) South Central corridor (*SC*); (vii) Southeast corridor (*SE*). The satellite image was obtained from the Google Maps application [Google Maps attribution: *Imagery (2019) Data SIO, NOAA, U.S. Navy, NGA, GEBCO, Landsat/Copernicus, Map data (2019) Google*; <https://www.google.co.uk/maps/@10.0038529,108.7048144,3313765m/data=!3m1!1e3>].

(mtDNA) has been shown as a very informative genetic marker to study the genetic diversity and genetic relationships among populations of the Philippines^{19–23}. In agreement with these previous works and in order to address the previously presented questions, we collected sequences from the D-Loop region of mtDNA of 157 individuals selected from native volunteer donors from different geographic regions of the archipelago. All of them carried mitochondrial variants older than 20,000 years. It is important to consider that mtDNA is inherited to the maternal line (but see²⁴), hence here we mainly make focus on the female evolutionary history of the archipelago, which is unequivocally related to the global history²⁵. Next we applied approximate Bayesian computation (ABC) based on extensive spatially explicit computer simulations to identify the scenario of migration towards the archipelago that fits better with the data. We also estimated a variety of population genetic parameters informative about this settlement such as the time of the onset of the expansion, carrying capacity (available resources), demographics (including ancestral population size and population growth rate), mutation rate, migration rate, and proportion of migration events with LDD.

Materials and Methods

Sample collection and ethical statements. A total of 157 samples were collected covering most of the geographical area of the archipelago (Figs. 1 and S1; Table S1). The samples were obtained from Filipinos currently living in Spain but with known geographic origin in the Philippines. Each participant was informed about the study and signed an informed consent document for study participation according to the Helsinki protocol for the use of these samples. Indeed, the protocol was approved by the bioethics Committee of the Complutense University of Madrid. Next, the participants completed a survey about their origins and predecessors (including ethno-linguistic groups; Table S2). The samples were only collected from participants presenting the four grandparents in the region and accounting for the geographical diversity of the Philippines by donors from the three major regions of the archipelago: Luzon ($n = 84$), Visayas ($n = 59$) and Mindanao ($n = 14$). Moreover, given that the first two regions include a large number of islands, intraregional samples were collected. In particular, Luzon was divided into five subregions –Bicol ($n = 10$), Calabarzon ($n = 20$), Ilocos ($n = 13$), Manila ($n = 17$) and a central region ($n = 24$)– and, Visayas was divided into three subregions –West ($n = 30$), Center ($n = 15$) and East ($n = 14$)–.

MtDNA amplification, sequencing and incorporation of additional sequences. Samples consisted of oral swabs (two per individual) that were kept dry in paper envelopes at 4 °C. Next, DNA was extracted by standard phenol-chloroform method including a final purification with AMICON filters following the

instructions of the manufacturer (Merck). The mtDNA D-loop (control region) was sequenced for all 157 individuals using the BigDye Terminator kit v1.1 (Applied Biosystems), in two or three fragments depending on the samples between positions 16,000 and 600²⁶. Amplified fragments were identified previously to the sequencing reaction in 4% agarose gel. Finally, the fragments were sequenced on Applied Biosystems ABI PRISM 3130 genetic analyzer. The genetic sequences were deposited at GenBank with accession numbers MH910797–MH910953.

The identification of spatial genetic variation is more accurate by considering a well-distributed sampling over the landscape²⁷. Because of that, we included genetic samples from neighboring regions of the Philippines. Unfortunately, most of studies involving humans from those regions only shared sequences of the hypervariable I mtDNA region (HVR-I) and therefore posterior analyses were based on this region. We obtained from the bibliography genetic sequences from regions surrounding the Philippines (with a relatively even geographic distribution, Figs. 1 and S1; Table S1) and where each sampling location included genetic data from at least 10 individuals. We only considered sequences belonging to haplogroups older than 20,000 years⁵ to avoid lineages with origin posterior to the Philippines settlement²⁸. The haplogroup assignment was performed with *HAPLOFIND*²⁹ and *MitoTool*³⁰ (Table S2). All the sequences were aligned with *MAFFT*³¹. The final multiple sequence alignment (MSA) included a total of 720 individuals collected from 25 different locations, from which 157 individuals of 9 populations of the Philippines were chosen and sequenced for this study. Details about sample size (number of individuals), geographic location and source (present study or other studies) are shown in Table S1 and depicted on the map in Figs. 1 and S1. Despite the limitations of our data due to the need of incorporating genetic information from different geographic areas, we found that our final dataset was sufficient to yield accurate model selection and estimation of population genetics parameters (see Results).

Testing migration routes to the Philippines. Migration routes to the Philippines could affect the genetic diversity and population structure observed in this archipelago. Therefore, first, we investigated the fitting of diverse scenarios of migration towards the Philippines with the real data using ABC (details shown in following subsections).

Considering the geography of the Philippines, its neighboring regions and suggestions from previous studies^{14,32,33}, we designed a total of 7 scenarios based on migration patterns and migration routes to the archipelago: (i) migration presenting LDD events (*LDD*, as suggested for Eurasia¹⁵; details about the LDD model are shown in the Supplementary Material) and, 6 scenarios of migration through different routes and without LDD proposed in the “Beyer’s Wave Migration Theory”³² and other studies^{14,33}: (ii) a north route and three south routes (*AllCorr*), (iii) only a north route (*North*), (iv) only the three south routes (*AllSouth*), (v) only the southwest route (*SW*), (vi) only the south-central route (*SC*) and, (vii) only the southeast route (*SE*) (Figs. 1 and S2). Next, we evaluated non-nested scenarios by fitting with real data to identify the most likely scenario. In particular, in order to evaluate the influence of LDD, we evaluated: *LDD vs AllCorr* and, *LDD vs AllSouth vs North*. We also performed comparisons between other non-nested scenarios without LDD to explore the fitting of particular migration routes: *AllSouth vs North* (to explore migration from the north respect to migration from the south) and, *SW vs SC vs SE* (to explore the most likely migration route/s from the south).

Spatially explicit computer simulations. We performed spatially explicit computer simulations with the evolutionary framework *SPLATCHE3*³⁴. Information about the applied evolutionary models and landscape is presented in the Supplementary Material. The computer simulations were based on prior distributions for population genetic parameters following previous studies (Table S3). We assumed a range expansion with origin in Bangladesh at a time (T_{ANC}) between 60,000 and 70,000 years ago³⁵, with an ancestral population size (N_{ANC}) between 5,000 and 50,000 individuals³⁶. At each generation, individuals could migrate to adjacent demes at a migration rate (MIG_R) between 0.2 and 0.3, following previous studies on Eurasia^{15,37}. The population density for each deme was determined by the population growth rate (N_{GR}) that varied between 0.4 and 1.0^{15,37} and the carrying capacity (K) that varied between 1,000 and 3,000^{15,37}. For the LDD scenario, we applied a LDD proportion (LDD_P) between 0.0001 (0.01%) and 0.05 (5%)^{15,27}, we assumed migration distances sampled from a gamma distribution with parameters estimated from human data³⁸ and we considered a maximum dispersal distance of 20 demes (500 km) to avoid unrealistically long LDD movements. Based on the previously simulated demographic history, we simulated genetic data for 720 individuals from 25 populations (Table S1 and Fig. 1). DNA sequence evolution was simulated according to a prior distribution for the mutation rate (MUT_R) based on estimates of this rate in mtDNA presented in previous studies³⁹ (Table S3). All these parameters were sampled from uniform prior distributions (Table S3) and were estimated with ABC under the best fitting evolutionary scenario.

Approximate Bayesian computation for selection among evolutionary scenarios and parameters estimation. The real data was analyzed with the ABC approach⁴⁰, which basically applies extensive computer simulations to estimate posterior probabilities of alternative scenarios and parameters by a statistical fitting between simulated and real data. Next, we describe the steps of our ABC method.

Computer simulations. For each of the seven evolutionary scenarios previously described, we performed a total of 100,000 spatially explicit computer simulations according to the prior distributions presented in Table S3. For each scenario, we also performed 100 additional and independent simulations (hereafter, *test datasets*) to evaluate the power of selection among evolutionary scenarios and parameters estimation (details shown later).

Summary statistics. Summary statistics (SS) from real and simulated datasets were obtained with *ARLEQUIN ver.3.5*⁴¹. To perform this task, the 25 populations were classified into five geographic groups (Fig. S1) that allowed the estimation of diverse group-based SS. Group 1 included samples from the north of the continent; group 2 included samples from the south of the continent; group 3 included samples from the Philippines; groups 4 and 5

included the remaining samples separated according to the Wallace line (Fig. S1). We considered that this general grouping was proper since the ABC performance reduces quickly as the number of statistics grows⁴⁰. We applied a total of 18 SS with biological meaning: (i) genetic diversity in every group and in all groups together (pairwise differences, π ; including standard deviation) and, (ii) genetic differentiation (F_{ST}) between every two groups and all groups together. We found that this set of SS was informative for the ABC inferences (see validations in Results section). We did not apply partial least square (PLS) components to reduce the number of SS because our selected SS were informative enough and because PLS components can generate SS without biological meaning, being difficult to interpret and leading to biases when comparing components that are calculated separately⁴².

Selection among evolutionary scenarios. As indicated above, we generated 100,000 simulated datasets for each of the seven migration scenarios. In addition, we simulated 100 *test datasets* for each scenario to evaluate the accuracy of the estimation method. We applied three ABC approaches to perform the selection of the best-fitting evolutionary scenario with the real data: The rejection approach developed by Pritchard *et al.*⁴³ (*Pr*) and, the rejection approach implemented in the *abc* package of *R*⁴⁴ (*Rrej*) and the neuralnet (*Rnn*; nonlinear heteroscedastic regression⁴⁵) approach also implemented in the *abc* package of *R*. Following the authors recommendation we retained 1% of simulations with SS closer to the SS of the real data⁴⁴.

For the evaluation of the selection among alternative scenarios, we applied the 100 *test datasets* (considering them as true data) and also for *Rnn* the leave-one-out cross-validation method implemented in the *abc* library of *R* (*cv4postpr* function) that accounts for non-linearity adjustment based on 100 permutations and again with a tolerance of 1%⁴⁴. The neuralnet method required more computational costs but, according to Csilléry *et al.*⁴⁴, it could provide more accurate inferences.

Although traditionally the goodness of fit analysis is only applied to the preferred scenario, we performed goodness of fit analyses (based on principal component analysis (PCA) and a comparison among the null distributions from the selected SS⁴⁴) to all the evolutionary scenarios designed in the study. In addition, for these analyses of goodness of fit we included the position of the real dataset on the simulated parameters landscape for every comparison of scenarios, providing a preliminary view of the fitting of every scenario with the real data. Finally, we selected the scenario that better fits with the real data applying the ABC approaches previously described (*Pr*, *Rrej* and *Rnn*).

Parameters estimation. We performed the parameters estimation under the evolutionary scenario with LDD, which was the scenario that fitted best with the real data (see Results). The parameters estimation was performed with the multiple regression adjustment implemented in the program *ABCtoolBox*⁴⁶. Following the software documentation, we retained the best 5,000 simulations (with SS closer to the SS of the real data) and also because they provided robust parameters estimation (see Results). The robustness of the parameter estimation was assessed with the 100 *test datasets* simulated under the *LDD* scenario, considering them as true data. For each *test dataset* and parameter we computed the distance between the true value and the estimated value (median, mean and mode).

Results

We first describe the selection between alternative evolutionary scenarios of migration to the Philippines. After that, we present the parameters estimation under the best fitting evolutionary scenario.

Migration to the Philippines occurred with long-distance dispersal and through all north and south migration routes. The power to select among the designed scenarios of migration to the Philippines varied depending on the scenarios that are evaluated. We obtained high posterior probabilities to identify the correct scenario under the different ABC methods when comparing (i) *LDD vs AllCorr* (probabilities around 1, Table S4A), (ii) *LDD vs AllSouth vs North* (probabilities generally above 0.8, Table S4B) and, (iii) *AllSouth vs North* (probabilities around 1, Table S4C). However, we obtained a lower accuracy when comparing the three scenarios with only migration routes from the south, *SW vs SC vs SE* (probabilities above 0.4, Table S4D). This result was expected because both *SW* and *SC* routes come from the same island (Borneo). In order to further explore this finding, we also evaluated the scenarios *SW vs SE* and there we found probabilities generally above 0.9 (Table S4E). Note that we performed comparisons among non-nested scenarios. If nested scenarios are analyzed together (i.e., comparing all the scenarios) the performance of the method will be poor because nested scenarios under certain evolutionary circumstances can lead to similar genetic signatures (i.e., *AllCorr vs AllSouth* could be similar in cases where under *AllCorr* the migration rate from the north is very small by chance sampling from the prior distribution).

Real data clearly fitted better with the scenario of *LDD*, with posterior probabilities above 0.9 under any ABC method (Table 1). Indeed, the analyses of goodness of fit also supported this finding (Fig. S4).

Even considering that *LDD* was the best fitting evolutionary scenario, we explored the fitting of the different migration routes (without *LDD*) with the real data. We found that the migration to the Philippines through the south routes was generally more likely than the migration through the north route (Taiwan), but the north route could also be used (Table 1). When comparing the routes from the south, we found that all the routes were likely to be used, although the routes from Borneo (*SW* and *SC*) presented the highest probabilities. Note that the selection between *SW* and *SC* routes presented a poor performance (Table S4D), and therefore we cannot distinguish between *SW* and *SC* when fitting with the real data. The results slightly varied among ABC approaches.

The settlement of the Philippines was rapid and caused by a rapidly increasing population. The estimation of the population genetic parameters was performed under the best-fitting evolutionary scenario (*LDD*). Concerning the validation of the parameters estimation, the method presented a performance

Evaluated evolutionary scenarios	Posterior probability		
	Pritchard's method	R rejection method	R neuralnet method
LDD/AllCorr	0.94 /0.06	0.96 /0.04	1.00 /0.00
LDD/North/AllSouth	0.93 /0.05/0.02	0.95 /1.30E-03/0.05	1.00 /0.00/0.00
North/AllSouth	0.46/0.54	0.32/0.68	0.29/0.71
SW/SC/SE	0.35/0.33/0.31	0.39/0.38/0.23	0.54/0.24/0.22
SW/SE	0.52/0.48	0.59/0.41	0.69/0.31

Table 1. Fitting of the studied evolutionary scenarios with the real data. Fitting of the studied evolutionary scenarios with real data based on the same comparisons (between non-nested scenarios) presented in the evaluation of the model selection (Table S4). The scenarios of migration to the Philippines are the following: LDD (LDD), all corridors without LDD (AllCorr), all the south corridors without LDD (AllSouth), only the north corridor without LDD (North), only the southwest corridor without LDD (SW), only the south-central corridor without LDD (SC), only the southeast corridor without LDD (SE). The table shows the probabilities based on the three applied ABC methods (see Materials and Methods) when fitting evolutionary scenario with the real data. Posterior probabilities above 0.9 (very high fitting) are shown in bold, note that the best fitting evolutionary scenario is LDD. The information of this table is graphically presented in the Figs. S3A–E.

Parameter	Mode	Mean	Median	95% HPDI
Time of the onset of the expansion (T_{ANC}) ^a	<i>60,151</i>	63,364	62,713	59,975–68,310
Population size at the onset of the expansion (N_{ANC})	<i>49,999</i>	46,316	47,060	25,760–49,886
Population growth rate (N_{GR})	0.822	<i>0.817</i>	0.825	0.591–0.998
Migration rate (MIG_R)	0.285	<i>0.272</i>	0.280	0.212–0.299
Carrying capacity (K)	<i>3,000</i>	2,578	2,709	1,159–3,000
Mutation rate (MUT_R)	<i>1.31E⁻⁰⁶</i>	1.50E ⁻⁰⁶	1.41E ⁻⁰⁶	2.49E ⁻⁰⁸ –3.08E ⁻⁰⁶
LDD proportion (LDD_P)	0.040	<i>0.036</i>	0.038	0.011–0.049

Table 2. Population genetics parameters estimated with ABC under the best-fitting evolutionary scenario. For each parameter, we present the mode, mean and median of the estimated posterior distribution. In italic we highlight the most accurate measure (median, mean or mode) according to Table S5. 95% HPDI refers to the 95% highest posterior density interval. ^aTime is shown in years and was estimated assuming a generation time of 25 years.

with overall small estimation errors that always fell within the 50% HPDI with respect to the true value (Table S5). The estimated population genetic parameters for the real data are shown in Table 2. The precision was low for the estimation of some parameters, in particular for the time and population size at the onset of the expansion and the carrying capacity. This was expected because the estimation of these parameters requires considerable genetic signatures of ancestral processes, also found in other studies (e.g.^{15,47}). However, we accounted for this uncertainty instead of fixing parameter values, since previous studies often disagree concerning these estimates⁴⁸. We considered acceptable the precision of the estimation of the remaining parameters, presenting in general more narrow posterior distributions (Table 2).

Our estimated date for the onset of the expansion of Anatomically Modern Humans (AMH) from Bangladesh (≈ 60 kya) fitted with previous studies⁴⁹. The estimated ancestral size (above 30,000 individuals and probably near 50,000) was larger than we expected. However, some previous works (e.g.⁵⁰), also suggested large ancestral population sizes (near 40,000 individuals) in the region. Indeed, the estimated carrying capacity ($\approx 3,000$) and population growth rate per generation (≈ 0.82) were higher than expected, although they fell within the range estimated in other studies (i.e., in Europe and north of Africa the estimated carrying capacity varied between 500 and near 7,000⁵¹ and the estimated population growth rate in Eurasia varied between 0.4 and 0.9^{15,51}). The estimated migration rate (≈ 0.27) was high but fell within estimates in Eurasia¹⁵. The proportion of LDD events (≈ 0.04) was in agreement with previous findings¹⁵. Although the estimated mutation rate ($\approx 1.3E^{-06}$) fitted with other studies⁵², this parameter usually varies among studies, especially when they are obtained from different genetic markers (e.g.^{39,52}).

Discussion

Despite the efforts made to understand human phylogeography, there are still relevant regions of the world where the population genetics processes and migration patterns occurring in first colonizations have not been resolved. An especially interesting example is the Philippine archipelago. There is a clear evidence of an early settlement of the Philippines^{6,8,10,11}, however little is known about the population genetics, migration routes and migration patterns that occurred in this settlement. Concerning its geography, the Philippines is a particularly relevant region to test the influence of the geography on human migration and to understand the causes of the currently observed population structure and genetic diversity.

Addressing the early settlement of the Philippines by the study of current genetic variants is complex given that its current composition is the result of a superposition of migratory waves distant in time¹⁰. Despite this complexity, here we investigated the fitting of diverse evolutionary scenarios of migration with real genetic data from

a large number of individuals from the Philippines and neighboring regions in order to provide clues about the population genetics of the settlement and migration routes that could have occurred in this archipelago. Note that the Philippine archipelago, given its geographic location, could be colonized by two major routes: from the north (through Taiwan) and from the south (which could contemplate different routes from Borneo or from northern Sulawesi). Indeed, diverse combinations of potential migration routes could have occurred. Moreover, this migration could also present LDD events. Here, we developed and applied an ABC method based on extensive spatially explicit computer simulations to determine whether the migration to the Philippines occurred with LDD or by progressive (gradual) migration from neighboring regions through particular migration routes. Finally, we estimated demographic and evolutionary parameters under the best fitting evolutionary scenario.

We found that the evolutionary scenario fitting best with the real data was the one that presents migration with LDD. This finding supports a previous study on the Palaeolithic expansion on Eurasia¹⁵. Indeed, this result could be expected because the colonization of the archipelago implies displacements from island to island, which can lead to LDD events. The finding is also in agreement with some previous studies on the arrival of modern humans to Australia, which occurred crossing water barriers with watercrafts such as boats and rafts^{17,18}. Next, we explored the fitting of particular migration routes with the real data to identify possible migratory pathways towards the Philippines. Our results showed that migration routes from both the north and the south were used to colonize the Philippines, although the probability was slightly higher for the colonization from the south. Moreover, within migration from the south, all the proposed routes were likely to occur. Altogether, we found that the most likely migratory scenario involving the early colonization of the Philippine archipelago by modern humans implied LDD events and migration through routes from both the north and the south. In theory, the southern route could be more used given the low sea level during the Last Glacial Maximum (LGM) period that led to the emergence of the continental mass of Sunda, connecting Indochina with the islands at the west of the Wallace line⁵³. A northern route favored by the LGM seems more difficult since this would require the use of advanced navigation techniques to connect distant islands. However, for many authors navigation technologies were already developed by *Homo erectus*⁵⁴ and with respect to our species, the presence of archaic fossil remains confirm that our ancestors must have navigated, more or less frequently, more than 60,000 years ago^{1,17,18,55}. In this concern, our findings present genetic signatures of migration through LDD, which could fit with the use of techniques of navigation in the expansions of modern humans.

Concerning the demographic and environmental conditions under which the colonization of the archipelago was carried out, we found that the expansion from Bangladesh occurred at 60,000 years ago or earlier (HPDI around 60,000–68,000), a date consistent with findings from other disciplines regarding the human presence in the SEA and Australia (e.g.^{1,56}) and previous studies proposing the presence of our species in Australia at 65,000 years ago⁵⁷. Next, the best fitting model (LDD) considers a rapid colonization of the Philippines (around 200 generations after the onset of the expansion, this is 5,000 years later) leading to an interval of 55,000–63,000 years ago. These dates agree with other studies (e.g.^{4,5}) and are slightly earlier than the dating of the remains of Callao Cave in Luzon (67,000 years old) although such remains have been considered as another species (*Homo luzonensis*)^{6,7}. In any case it should be noted that for some authors these remains are fragmentary and taxonomically ambiguous⁵⁸.

Our results described a rapid Palaeolithic expansion towards the Philippine archipelago. If that dispersal pattern is maintained for the posterior Neolithic expansion from current China⁵⁹, which could be likely considering the technological and cultural innovations developed by Neolithic populations^{59,60} but still to be demonstrated in the study area, our results would support the variant of the *Out of Taiwan* theory (e.g.^{22,59}), known as the *Express Train to Polynesia*, that defends a rapid migration of Polynesian ancestors from the SEA towards the Southeast Pacific (e.g.⁵).

The estimated effective population size at the onset of the expansion varied between 25,000 and 50,000 individuals, with a higher probability for a large size. This large population size agrees with previous studies for the SEA^{5,50,61}. Indeed, our estimates are based on mtDNA and thus they may represent only the female component of the population. For some authors the female demographic component can be about eight times higher than the male component^{62–64}, although this was found for post-Neolithic periods characterized by patrilineal patterns. Studies focused on periods before the Neolithic, as in the present study, found a balanced sex-bias (e.g.⁶⁵), considering cultural, historical and geographical singularities of the study regions. We believe that this large population could favor the rapid spread of modern humans throughout the archipelago, in agreement with Jinam, *et al.*⁵. The estimated population growth rate and carrying capacity, directly related to *Annual Net Primary Production* (ANPP), fit with previously found population demographic patterns⁶⁶, especially in hunter-gatherer societies⁶⁷. The average value obtained for the population growth rate was 0.82 per generation (but ranging between 0.59 and 1.0). This result agrees with previous estimates that presented values between 0.3 and 0.9 (e.g.^{15,51,68}). On the other hand, the estimated carrying capacity (around 3,000) was higher than previous estimates in Eurasia (around 1,000 but with an upper estimate around 2,000)¹⁵ but lower than estimates for Neolithic populations (above 5,000)⁵¹. This suggests that Palaeolithic populations of the SEA could be favoured by an environment rich in natural resources, facilitating demographic growth. Actually, the paleoclimatic reconstructions of the SEA describe an ecological environment very similar to current sub-Saharan Africa and the presence of forests⁶⁹, which could be beneficial for the living-style of Palaeolithic humans. Note that the SEA is one of the most important geographical spaces in our evolutionary history from where movements to Europe, America and Australia started⁶⁸.

A possible consequence of the rapid demographic growth could be a rapid expansion throughout the islands of the Philippine archipelago. This affirmation is supported by the estimated high migration rates (around 27%) that reflect important human contingents movement, higher than those (23–25%) estimated for other groups of hunter-gatherers in Europe and north of Africa⁵¹. In addition, the estimated proportion of individuals migrating with LDD events was approximately 4%, similar to that obtained for Palaeolithic populations in Eurasia¹⁵, which again supports the rapid expansion of hunter-gatherers along Asia and expanding to Indonesia and Philippines.

The Philippines constitutes an important geographic region in the expansion of modern humans despite it was frequently ignored in human evolutionary analyses of the SEA and Oceania. Here we explored the expansion of modern humans towards the Philippines and we found that this expansion was rapid, presenting LDD events, and following diverse migration routes that geography and existing technology could allow. We believe that the rapid expansion an explosive population growth could be favored by optimal environmental conditions present in the region at the time of the expansion.

Data availability

The genetic sequences were deposited at GenBank with accession numbers MH910797–MH910953.

Received: 21 March 2019; Accepted: 3 March 2020;

Published online: 17 March 2020

References

- Malaspinas, A. S. *et al.* A genomic history of Aboriginal Australia. *Nat.* **538**, 207–214 (2016).
- Nielsen, R. *et al.* Tracing the peopling of the world through genomics. *Nat.* **541**, 302–310 (2017).
- Barker, G. *et al.* The ‘human revolution’ in lowland tropical Southeast Asia: the antiquity and behavior of anatomically modern humans at Niah Cave (Sarawak, Borneo). *J. Hum. Evol.* **52**, 243–261 (2007).
- Jinam, T. A. *et al.* Discerning the Origins of the Negritos, First Sundaland People: Deep Divergence and Archaic Admixture. *Genome Biol. Evol.* **9**, 2013–2022 (2017).
- Jinam, T. A. *et al.* Evolutionary history of continental southeast Asians: “early train” hypothesis based on genetic analysis of mitochondrial and autosomal DNA data. *Mol. Biol. Evol.* **29**, 3513–3527 (2012).
- Mijares, A. S. *et al.* New evidence for a 67,000-year-old human presence at Callao Cave, Luzon, Philippines. *J. Hum. Evol.* **59**, 123–132 (2010).
- Detroit, F. *et al.* A new species of Homo from the Late Pleistocene of the Philippines. *Nat.* **568**, 181–186 (2019).
- Aghakhanian, F. *et al.* Unravelling the genetic history of Negritos and indigenous populations of Southeast Asia. *Genome Biol. Evol.* **7**, 1206–1215 (2015).
- Reich, D. *et al.* Denisova Admixture and the First Modern Human Dispersals into Southeast Asia and Oceania. *Am. J. Hum. Genet.* **89**, 516–528 (2011).
- Lipson, M. *et al.* Reconstructing Austronesian population history in Island Southeast Asia. *Nat. Commun.* **5**, 4689 (2014).
- Morseburg, A. *et al.* Multi-layered population structure in Island Southeast Asians. *Eur. J. Hum. Genet.* **24**, 1605–1611 (2016).
- Consortium, H. P.-A. S. *et al.* Mapping human genetic diversity in Asia. *Sci.* **326**, 1541–1545 (2009).
- Mallick, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nat.* **538**, 201–206 (2016).
- Yew, C. W. *et al.* Genetic relatedness of indigenous ethnic groups in northern Borneo to neighboring populations from Southeast Asia, as inferred from genome-wide SNP data. *Ann. Hum. Genet.* **82**, 216–226 (2018).
- Alves, I. *et al.* Long-Distance Dispersal Shaped Patterns of Human Genetic Diversity in Eurasia. *Mol. Biol. Evol.* **33**, 946–958 (2016).
- Pugach, I. *et al.* The Gateway from Near into Remote Oceania: New Insights from Genome-Wide Data. *Mol. Biol. Evol.* **35**, 871–886 (2018).
- Stringer, C. Palaeoanthropology. Coasting out of Africa. *Nat.* **405**(24–25), 27 (2000).
- Balme, J. Of boats and string: The maritime colonisation of Australia. *Quat. Int.* **285**, 68–75 (2013).
- Gunnarsdottir, E. D., Li, M., Bauchet, M., Finstermeier, K. & Stoneking, M. High-throughput sequencing of complete human mtDNA genomes from the Philippines. *Genome Res.* **21**, 1–11 (2011).
- Loo, J. H. *et al.* Genetic affinities between the Yami tribe people of Orchid Island and the Philippine Islanders of the Batanes archipelago. *BMC Genet.* **12**, 21 (2011).
- Delfin, F. *et al.* Complete mtDNA genomes of Filipino ethnolinguistic groups: a melting pot of recent and ancient lineages in the Asia-Pacific region. *Eur. J. Hum. Genet.* **22**, 228–237 (2014).
- Tabbada, K. A. *et al.* Philippine mitochondrial DNA diversity: a populated viaduct between Taiwan and Indonesia? *Mol. Biol. Evol.* **27**, 21–31 (2010).
- Soares, P. *et al.* Ancient voyaging and Polynesian origins. *Am. J. Hum. Genet.* **88**, 239–247 (2011).
- Luo, S. *et al.* Biparental Inheritance of Mitochondrial DNA in Humans. *Proc. Natl Acad. Sci. USA* **115**, 13039–13044 (2018).
- Jorde, L. B. *et al.* The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am. J. Hum. Genet.* **66**, 979–988 (2000).
- Anderson, S. *et al.* Sequence and organization of the human mitochondrial genome. *Nat.* **290**, 457–465 (1981).
- Mona, S., Ray, N., Arenas, M. & Excoffier, L. Genetic consequences of habitat fragmentation during a range expansion. *Heredity* **112**, 291–299 (2014).
- Behar, D. M. *et al.* A “Copernican” reassessment of the human mitochondrial DNA tree from its root. *Am. J. Hum. Genet.* **90**, 675–684 (2012).
- Vianello, D. *et al.* HAPLOFIND: a new method for high-throughput mtDNA haplogroup assignment. *Hum. Mutat.* **34**, 1189–1194 (2013).
- Fan, L. & Yao, Y. G. MitoTool: a web server for the analysis and retrieval of human mitochondrial DNA sequence variations. *Mitochondrion* **11**, 351–356 (2011).
- Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
- Beyer, H. O. Earliest people of the Philippines. *Manila Bulletin* **27** (1950).
- Pawlik, A. F., Piper, P. J. & Mijares, A. S. B. In *Southern Asia, Australia and the Search for human origins* 135–147 (Cambridge University Press, 2014).
- Currat, M., Arenas, M., Quilodran, C. S., Excoffier, L. & Ray, N. SPLATCHE3: simulation of serial genetic data under spatially explicit evolutionary scenarios including long-distance dispersal. *Bioinformatics* **35**, 4480–4483 (2019).
- Macaulay, V. *et al.* Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Sci.* **308**, 1034–1036 (2005).
- Tenesa, A. *et al.* Recent human effective population size estimated from linkage disequilibrium. *Genome Res.* **17**, 520–526 (2007).
- Arenas, M., Francois, O., Currat, M., Ray, N. & Excoffier, L. Influence of admixture and paleolithic range contractions on current European diversity gradients. *Mol. Biol. Evol.* **30**, 57–61 (2013).
- Ray, N. & Excoffier, L. A first step towards inferring levels of long-distance dispersal during past expansions. *Mol. Ecol. Resour.* **10**, 902–914 (2010).
- Soares, P. *et al.* Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am. J. Hum. Genet.* **84**, 740–759 (2009).
- Beaumont, M. A. Approximate Bayesian Computation in Evolution and Ecology. *Annu. Rev. Ecol. Syst.* **41**, 379–405 (2010).

41. Excoffier, L. & Lischer, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567 (2010).
42. Aeschbacher, S., Beaumont, M. A. & Futschik, A. A novel approach for choosing summary statistics in approximate Bayesian computation. *Genet.* **192**, 1027–1047 (2012).
43. Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A. & Feldman, M. W. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol. Biol. Evol.* **16**, 1791–1798 (1999).
44. Csillery, K., François, O. & Blum, M. G. B. abc: an R package for approximate Bayesian computation (ABC). *Methods Ecol. Evolution* **3**, 475–479 (2012).
45. Blum, M. G. B. & François, O. Non-linear regression models for Approximate Bayesian Computation. *Stat. Comput.* **20**, 63–73 (2010).
46. Wegmann, D., Leuenberger, C., Neuenschwander, S. & Excoffier, L. ABCtoolbox: a versatile toolkit for approximate Bayesian computations. *BMC Bioinforma.* **11**, 116 (2010).
47. Fagundes, N. J. *et al.* Statistical evaluation of alternative models of human evolution. *Proc. Natl Acad. Sci. USA* **104**, 17614–17619 (2007).
48. Benguigui, M. & Arenas, M. Spatial and temporal simulation of human evolution. Methods, frameworks and applications. *Curr. Genomics* **15**, 245–255 (2014).
49. Stanyon, R., Sazzini, M. & Luiselli, D. Timing the first human migration into eastern Asia. *J. Biol.* **8**, 18 (2009).
50. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nat.* **475**, 493–496 (2011).
51. Pimenta, J., Lopes, A. M., Comas, D., Amorim, A. & Arenas, M. Evaluating the Neolithic Expansion at Both Shores of the Mediterranean Sea. *Mol. Biol. Evol.* **34**, 3232–3242 (2017).
52. Madrigal, L. *et al.* High mitochondrial mutation rates estimated from deep-rooting Costa Rican pedigrees. *Am. J. Phys. Anthropol.* **148**, 327–333 (2012).
53. Allen, J., Golson, J. & Jones, R. *Sunda and Sahul: Prehistoric studies in Southeast Asia, Melanesia and Australia.* (Academic Press, 1977).
54. von Koenigswald, G. H. R. Preliminary Report on a Newly-Discovered Stone Age Culture from Northern Luzon, Philippine Islands. *Asian Perspect.* **2**, 69–70 (1958).
55. Norman, K. *et al.* An early colonisation pathway into northwest Australia 70–60,000 years ago. *Quaternary Sci. Rev.* **180**, 229–239 (2018).
56. Davidson, I. The Colonization of Australia and Its Adjacent Islands and the Evolution of Modern Cognition. *Curr. Anthropol.* **51**, S177–S189 (2010).
57. Clarkson, C. *et al.* Human occupation of northern Australia by 65,000 years ago. *Nat.* **547**, 306–310 (2017).
58. Reyes-Centeno, H. *et al.* Genomic and cranial phenotype data support multiple modern human dispersals from Africa and a southern route into Asia. *Proc. Natl Acad. Sci. USA* **111**, 7248–7253 (2014).
59. Diamond, J. & Bellwood, P. Farmers and their languages: the first expansions. *Sci.* **300**, 597–603 (2003).
60. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. Demic expansions and human evolution. *Sci.* **259**, 639–646 (1993).
61. Atkinson, Q. D., Gray, R. D. & Drummond, A. J. mtDNA variation predicts population size in humans and reveals a major Southern Asian chapter in human prehistory. *Mol. Biol. Evol.* **25**, 468–474 (2008).
62. Hamilton, G., Stoneking, M. & Excoffier, L. Molecular analysis reveals tighter social regulation of immigration in patrilineal populations than in matrilineal populations. *Proc. Natl Acad. Sci. USA* **102**, 7476–7480 (2005).
63. Oota, H., Setheetham-Ishida, W., Tiwawech, D., Ishida, T. & Stoneking, M. Human mtDNA and Y-chromosome variation is correlated with matrilineal versus patrilineal residence. *Nat. Genet.* **29**, 20–21 (2001).
64. Seielstad, M. T., Minch, E. & Cavalli-Sforza, L. L. Genetic evidence for a higher female migration rate in humans. *Nat. Genet.* **20**, 278–280 (1998).
65. Wilkins, J. F. & Marlowe, F. W. Sex-biased migration in humans: what should we expect from genetic data? *Bioessays* **28**, 290–300 (2006).
66. Luck, G. W. The relationships between net primary productivity, human population density and species conservation. *J. Biogeogr.* **34**, 201–212 (2007).
67. Binford, L. R. *Constructing Frames of Reference: An Analytical Method for Archaeological Theory Building Using Ethnographic and Environmental Data Sets.* (University of California Press, 2001).
68. Eriksson, A. *et al.* Late Pleistocene climate change and the global expansion of anatomically modern humans. *Proc. Natl Acad. Sci. USA* **109**, 16089–16094 (2012).
69. Finlayson, C. Biogeography and evolution of the genus Homo. *Trends Ecol. Evol.* **20**, 457–463 (2005).

Acknowledgements

This study was supported by the Spanish Ministry of Economy and Competitiveness: Grants HAR2010-21063 and CGL2017-83394-P to AGM and, the “Ramón y Cajal” Grant RYC-2015-18241 to MA. MA was also funded by the Xunta de Galicia “ED431F 2018/08”.

Author contributions

M.A. and A.G.M. conceived and designed the study. A.G., E.C., H.R.V. and A.G.M. collected the data and obtained the sequences. M.A., J.M.B. and C.B. carried out the evolutionary analyses. All authors wrote and approved the final version submitted for publication.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-61793-7>.

Correspondence and requests for materials should be addressed to M.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020