


RESEARCH ARTICLE

Open Access



Genome-wide data from the Bubi of Bioko Island clarifies the Atlantic fringe of the Bantu dispersal

Pere Gelabert¹, Manuel Ferrando-Bernal¹, Toni de-Dios¹, Benedetta Mattorre², Elena Campoy³, Amaya Gorostiza⁴, Etienne Patin^{5,6,7}, Antonio González-Martín^{3*} and Carles Lalueza-Fox^{1*} 

Abstract

Background: Bioko is one of the few islands that exist around Africa, the most genetically diverse continent on the planet. The native Bantu-speaking inhabitants of Bioko, the Bubi, are believed to have colonized the island about 2000 years ago. Here, we sequenced the genome of thirteen Bubi individuals at high coverage and analysed their sequences in comparison to mainland populations from the Gulf of Guinea.

Results: We found that, genetically, the closest mainland population to the Bubi are Bantu-speaking groups from Angola instead the geographically closer groups from Cameroon. The Bubi possess a lower proportion of rainforest hunter-gatherer (RHG) ancestry than most other Bantu-speaking groups. However, their RHG component most likely came from the same source and could have reached them by gene flow from the mainland after island settlement. By studying identity by descent (IBD) genomic blocks and runs of homozygosity (ROHs), we found evidence for a significant level of genetic isolation among the Bubi, isolation that can be attributed to the island effect. Additionally, as this population is known to have one of the highest malaria incidence rates in the world we analysed their genome for malaria-resistant alleles. However, we were unable to detect any specific selective sweeps related to this disease.

Conclusions: By describing their dispersal to the Atlantic islands, the genomic characterization of the Bubi contributes to the understanding of the margins of the massive Bantu migration that shaped all Sub-Saharan African populations.

Keywords: Bubi, Bioko Island, Bantu-speaking groups, Population genetics, Isolation

Background

The Gulf of Guinea, which covers a large portion of the African coast, is characterized by complex and rich ecosystems. Of the few islands located in Atlantic Africa, four of them are found within the Gulf of Guinea (Fig. 1a). Bioko is the largest of these islands, with a total area of 2017 km² (Fig. 1b). It is located 32 km offshore of Cameroon but constitutes in fact, the northernmost part of Equatorial Guinea, a former Spanish colony. The island is volcanic and very mountainous, with an abrupt coastline; despite its small size, its highest peak has an

altitude of 3012 m. The indigenous population of Bioko, the Bubi, speak Bubi, a basal Bantu language [1]. The closest Bantu language on the mainland is most likely Galoa, which is spoken by Bantu-speaking groups of the Ogowe basin in Gabon [2]. The Bubi have a distinct and unique culture among Bantu-speaking people [3], including the belief that different spiritual beings reside in specific geographical locations along the island and the existence of well-defined matrilineal clans [4].

The origin of the Bubi people is controversial. Since the British explorer Richard Francis Burton visited the island (then called Fernando Poo) in 1874 [5], ethnographers generally considered the Bubi to be the original settlers of Bioko. However, it is currently accepted that the Bubi agriculturalists arrived from the mainland in dugout canoes about 2000 years ago during the Late

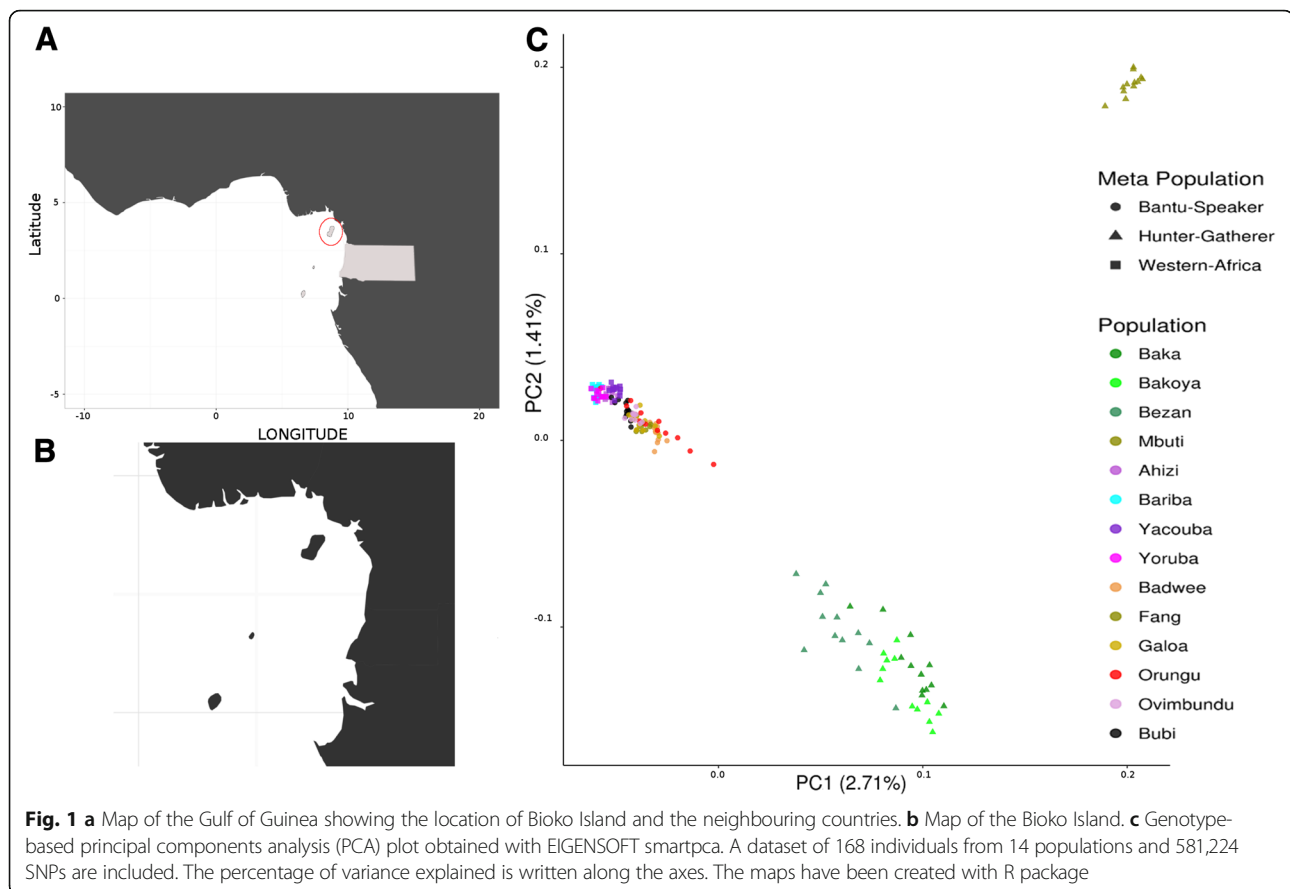
* Correspondence: antonio@ucm.es; carles.lalueza@upf.edu

³Department of Biodiversity, Ecology and Evolution, Complutense University of Madrid, Madrid, Spain

¹Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona, Spain

Full list of author information is available at the end of the article





Neolithic [6–8]. Ever since, the Bubi seem to have been isolated from mainland Bantu-speaking groups [9]. Bubi mythology explains that, upon their arrival to the island, they found other, more robust people living there, a population whom they called Balettérino [1, 9].

In fact, some unsystematic archaeological prospects carried out by Spanish scholars have uncovered pre-Neolithic lithic tools of a typology that has been called banapense, although this lithic typology does not currently have a clear chronological framework [9].

The expansion of the Bantu-speaking farming communities is probably one of the most important human movements that have taken place in recent African history [10]. This movement started approximately 4000 to 5000 years ago [11], likely from a source close to the present-day North Cameroonians [12] or Gabonese/Angolan Bantu-speaking populations [13], depending on which model – “early-split” or “late-split” – is assumed. While the first model suggests that the Bantus made an early separation into western and eastern branches, the second model supports an initial movement south across the rainforest before splitting into two branches, one headed south and the other east. Whatever the route of dispersal, the Bantu-speaking migration triggered an expansion of agriculture and ironwork along with the

spread of Bantu languages (part of the Niger-Congo family) across most of Central, South and East Africa [13–16].

During their extensive geographical migration, the Bantus encountered and admixed with local rainforest hunter-gatherer (RHG) tribes. Several Bantu-speaking groups have been studied from a genetic point of view over the years, especially using mitochondrial DNA (mtDNA) and Y chromosome markers [12, 17–20]. These studies have detected a substantial fraction of Pygmy mtDNA lineages within the Bantu speakers, but rarely the opposite [17, 18]. For example, traces (around 1%) of RHG Y chromosomes have been found in Bantu-speaking groups from Gabon and Cameroon [12] and signals of hunter-gatherer Khoisan Y chromosomes in Bantu-speaking groups from Mozambique [21].

Despite these efforts, the genetic patterns of variation within the Bantus remained largely unexplored on a genomic scale until fairly recently [14, 15, 22]. The analysis of a large and geographically diverse dataset of 35 Bantu groups has recently confirmed the existence of this RHG ancestry as well as the acquisition of adaptive alleles from these local populations, especially at the human-leukocyte antigen (HLA) loci. By measuring the length of the introgressed genetic fragments, the

admixture between western Bantu-speaking populations and RHG was estimated to occur about 800 years ago, mostly after Bantu-speaking populations began moving throughout Sub-Saharan Africa [13]. According to this study, while the most likely parental source population of Bantu ancestry in both eastern and southern Bantus was located in northern Angola, Bakoya of Gabon and Congo were the best parental source for RHG ancestry. Recent retrieval of ancient genomes from different African localities, notably in the south and the east, could help elucidate these past admixture events [23, 24]. So far, however, no ancient genomes have been retrieved from the Gulf of Guinea.

There are hypothesis that could potentially be explored with genome-wide data from the Bubi. First, their relatively long period of isolation on the island is a potential way to test the age and extent of the Bantu admixture with RHG tribes, an event that supposedly took place among the coastal Bantu groups after isolation of Bubi ancestors on the island. Second, studying a tribe from one of the few islands around Africa could provide information about potential effects of endogamy and isolation that are less likely to occur in mainland tribes. Finally, genome-wide data from the Bubi can offer clues regarding the adaptation of this group to local conditions. For example, it is known that the population of Equatorial Guinea has one of the highest levels of malaria infection in the world [25] and ranks 13th in the list of malaria prevalence countries, representing the second highest cause of death in the country [26]. Despite prevention efforts carried out since 2004, severe malaria prevalence in Bioko children remains high [27]. Thus, screening potential resistant variants can help us further understand the selective pressures faced by the Bubi during the last few hundred years.

We show in this work that the Bantu population of Bioko Island mirrors the genetic makeup of the extant, coastal Bantu-speaking groups, also clarifying the dynamics of this human expansion into the Atlantic islands of Africa.

Results

We sequenced 13 Bubi genomes obtaining a depth of coverage up to 21x–32x (Additional file 1: Table S5). All mtDNA haplogroups present in the Bubi are subclades of the L haplogroup (L1b, L2b, L3e, L3f, and L3e) and are common in other populations from the Gulf of Guinea. All male individuals of this dataset belong to subclades of the E1b1a1 haplogroup, the predominant Y-chromosome lineage in Western, Central and Southern Africa (Additional file 1: Table S6).

On a genome-wide scale, the first two components of the principal component analysis (PCA) separate the RHG from the Bantu speakers and the Western Africa

non-Bantu populations (Fig. 1c). When PC3 and PC4 are plotted, Bantu-speaking and Western African populations cluster separately (Additional file 2: Figure S1). Three Bubi individuals that we named Bubi-subset1 (BBS014, BBS018, BBS023) fall within the Western Africa cluster showing that some individuals share a larger proportion of Western-Africa ancestry than others, while the rest (named Bubi-subset2) cluster within the Bantu diversity (Additional file 2: Figure S1). To examine if this clustering reveals the presence of population substructure, we have used the USCS liftover [28] to convert the chimpanzee reference sequence (*Pan troglodytes* 3.0 assembly, GCA_000001515.5) to human coordinates in 546,558 single nucleotide polymorphisms (SNPs) of our dataset. We have subsequently computed f_4 statistics in the form (Bubi-subset1, Bubi-subset2; *X*, chimpanzee), to examine the homogeneity of the Bubi population (Additional file 2: Figure S2, Additional file 1: Table S7). None of the tested populations showed elevated ($|z| > 3$) values of Z-score; therefore we have treated the Bubi as a single population in subsequent analyses.

Moreover, ADMIXTURE analysis ($K = 4$) (Fig. 2 and Additional file 2: Figure S3) shows that the Bubi contain the same components as the other Bantu-speaking populations of the dataset but lower levels of the RHG component (in grey) than most of the mainland Bantu-speaking populations (with those from Angola being the exception). Some of the remaining ADMIXTURE plots ($K = 2–15$) (Additional file 2: Figure S4) also indicate potential substructuring among the Bubi; however, the sample size is too small to test if this is correlated with geography.

To determine which mainland populations share a higher level of genetic ancestry with the Bubi, we tested all neighbouring populations with the outgroup f_3 statistic (considering San as outgroup). Our results indicate that non-Bantu Western African populations such as Yoruba, Bariba, Fon and Azhizi, as well as Bantu-speaking groups from Angola such as Kongo, Ovimbundu and Kimbundu show more genetic affinities with the Bubi, apparently because they all show lower hunter-gatherer ancestry as compared to other groups (Fig. 3 and Additional file 1: Table S8). To explore the possible source of genetic admixture in the Bubi, we have also calculated the f_4 statistic for the combinations (*Test*, San; Bubi, Mbuti), (*Test*, San; Bubi, Baka), (*Test*, San; Bubi, Yoruba), (*Test*, San; Bubi, Fang) (Additional file 2: Figure S5A–D). This selection represents the four major genomic components in the Gulf of Guinea. We have found that the Bubi show the highest levels of shared ancestry with the Western African populations, which do not overlap with other populations when the comparison is established with RHG components. Angolan Bantu-speaking populations such as Ovimbundu,

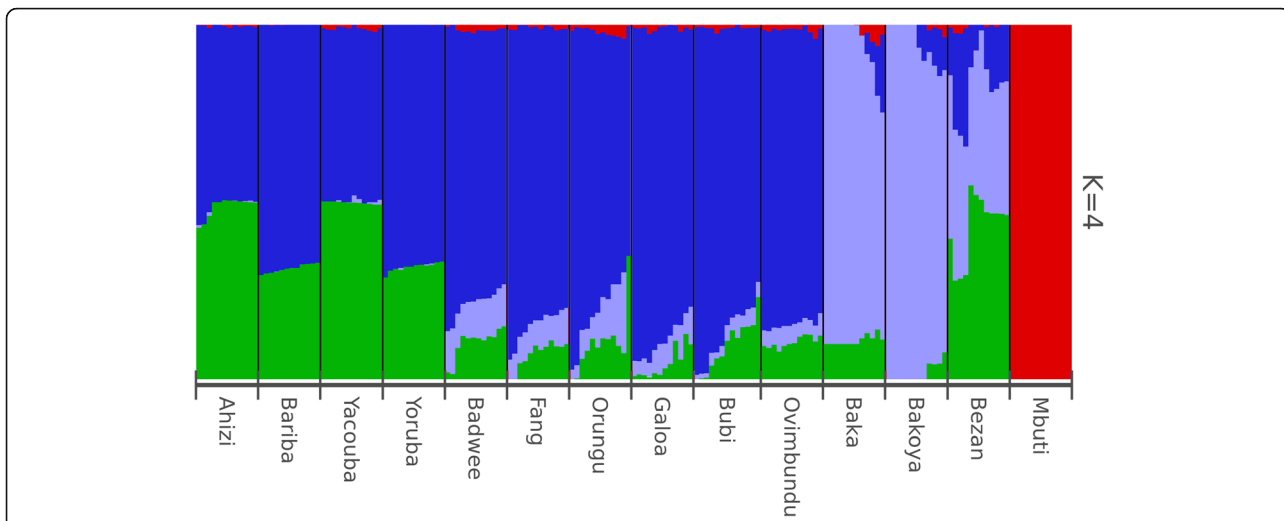


Fig. 2 Graphical representation of shared genetic components performed using the ADMIXTURE (K = 4) software: 169 individuals from 13 populations and 456,095 SNPs have been plotted. Although the cross-validation errors show the lowest value at K = 2, we have chosen K = 4 because it is the first plot where the crucial RHG, Bantu and Western African components are clearly identified

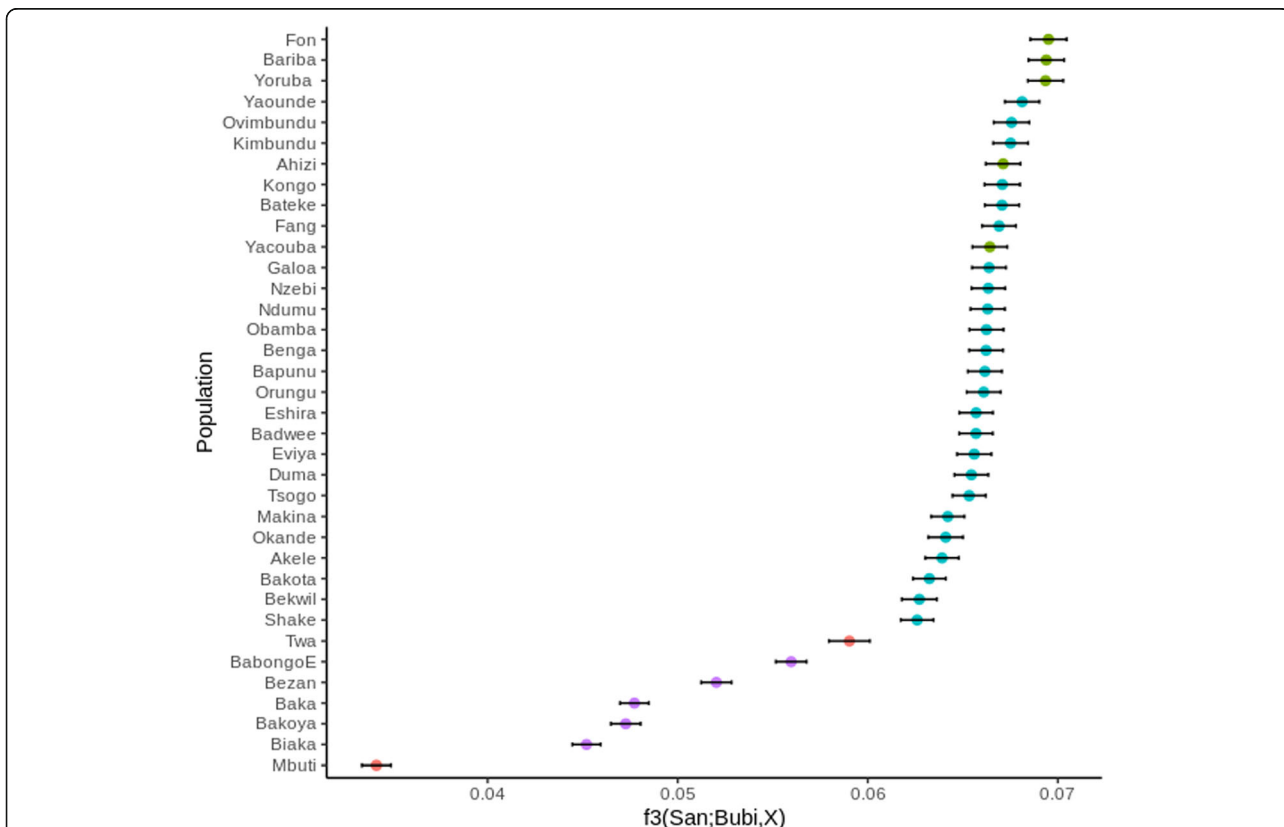


Fig. 3 Statistic f_3 values obtained with popstats when taking San as outgroup. The statistic and one standard error deviation are presented for each combination test. Blue dots represent Bantu-speaking populations, Green dots represent Western-African populations, orange dots Represent Eastern Hunter-Gatherer populations and purple dots indicate Western Hunter-gatherer populations

Kongo and Kimbundu also show elevated levels of shared genetic ancestry (Additional file 2: Figure S5). Bubi people seem to have very low levels of genetic admixture from the RHG populations; however, this small signal is absent in Western Africans.

Furthermore, owing to the colonial history of Bioko we have explored the possibility of some Iberian contribution to the Bubi ancestry by calculating the f_3 statistic in the form (Iberia, X ; Mbuti). The Bubi are placed within the range of Western African and other Bantu-speaking populations (Additional file 2: Figure S6); therefore, no Iberian genetic affinities can be discerned within the current dataset.

Pairwise F_{st} is a statistic used to measure population differentiation based on allelic frequencies. We used this test to quantify the level of genetic differentiation between all combinations of populations in our dataset. Low levels of F_{st} statistic indicate that the tested populations share a large proportion of the genotypes, while high levels are indicators of genetic differentiation. Among all African samples, RHG tribes appear to have

the highest levels of genetic differentiation, both among themselves and with the agriculturalist groups (Fig. 4). The lowest values of genome-wide F_{st} for the Bubi are again with Bantu-speaking Angolan groups (Kimbundu, $F_{st} = 0.0045$; Ovimbundu, $F_{st} = 0.0045$), and also with a Bantu-speaking group from Cameroon (Yaounde, $F_{st} = 0.0045$). On the other hand, we found that the Bubi displayed the highest F_{st} values when compared with RHG populations (Additional file 1: Table S9).

FineSTRUCTURE uses the coancestry matrix obtained from ChromoPainter to classify samples based on haplotype diversity. Using this approach, the matrix and the resulting dendrogram confirm that the closest populations to the Bubi are the Ovimbundu (Fig. 5 and Additional file 2: Figure S7), but also some Gabonese and Cameroonian Bantu-speaking populations. Interestingly, the Bubi are divided into two different clusters (Additional file 1: Table S7), one including only individuals from Bioko Island and the other shared with other Bantu-speaking groups. This, again, suggests some level of substructure in the population. Interestingly, the RHG

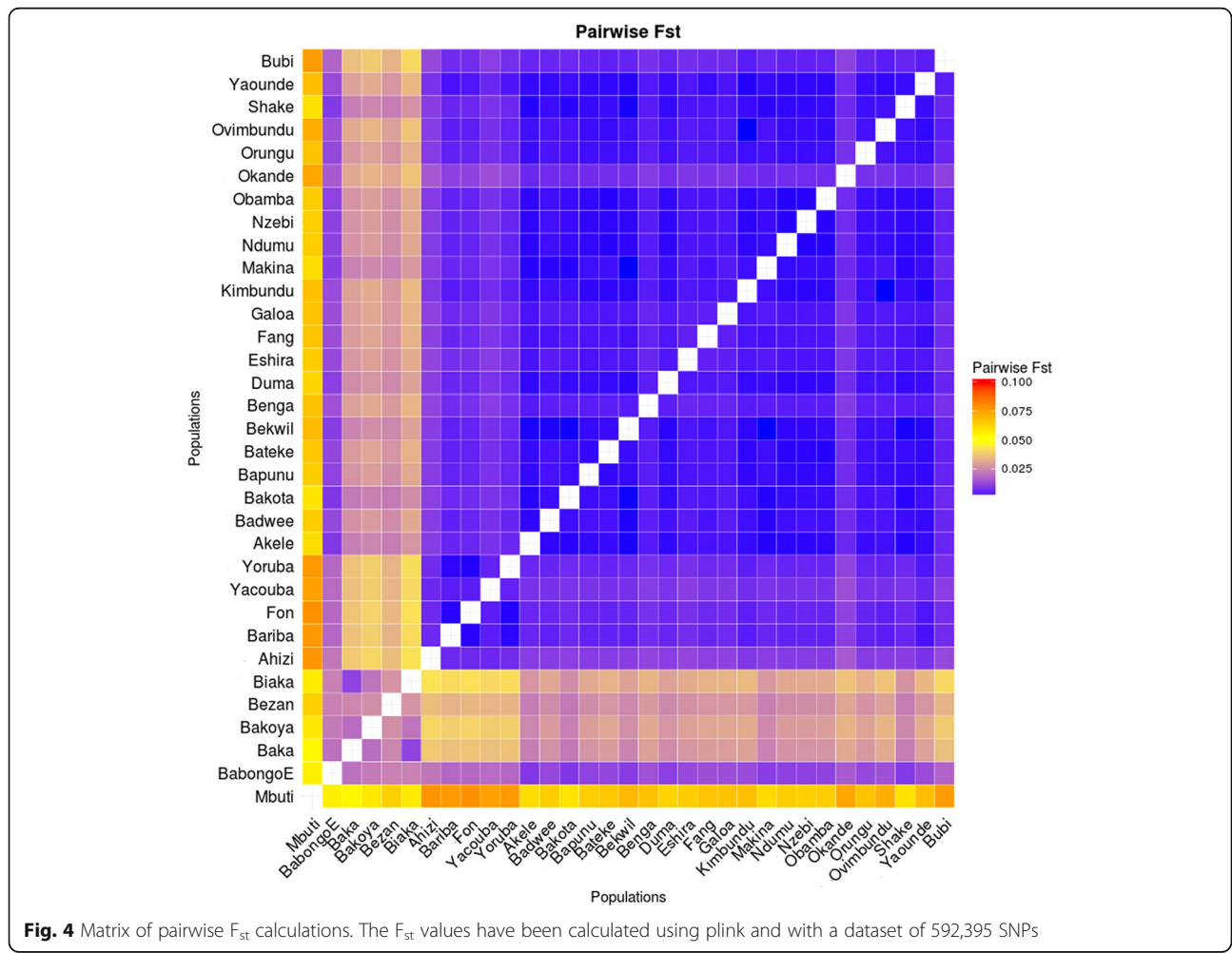


Fig. 4 Matrix of pairwise F_{st} calculations. The F_{st} values have been calculated using plink and with a dataset of 592,395 SNPs

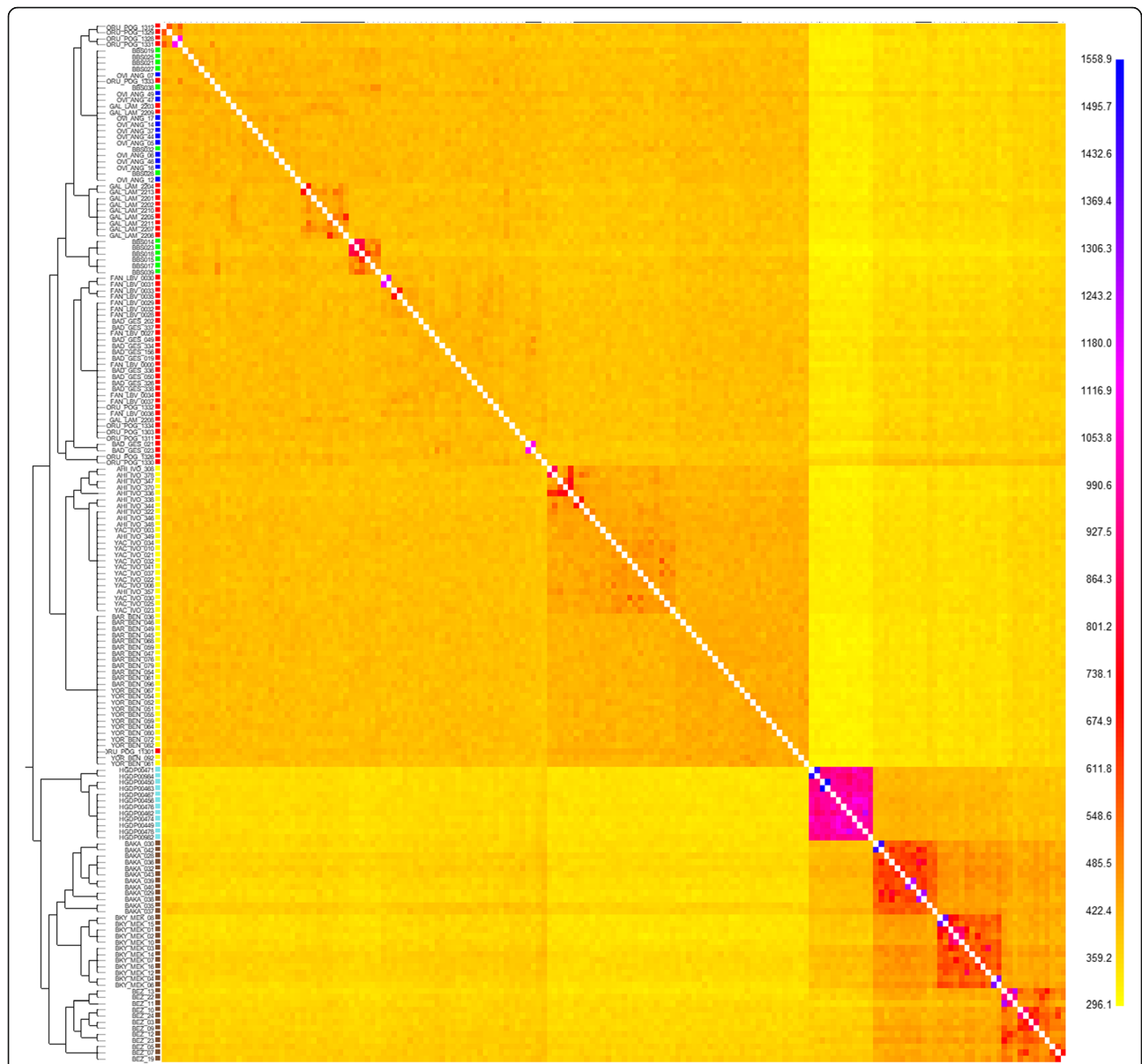


Fig. 5 Matrix of shared counts of haplotypes obtained using fineSTRUCTURE. The dataset includes 491,203 SNPs and 169 individuals. The structure of the matrix has been adapted using the fineSTRUCTURE dendrogram. Green label represents the Bubi individuals, red label is used to represent the Gabon-Cameroon Bantu-speaking individuals, Blue label represents the Angolan Bantu-speaking individuals, Yellow label represents the Western-African individuals, Light blue label represents the Eastern Hunter-Gatherer individuals and Brown label represents the and Western Hunter-gatherer populations

show a higher level of haplotype differentiation compared to other populations. To additionally test for the presence of Iberian ancestry we have repeated the fineSTRUCTURE analysis including 12 Iberian individuals (Additional file 2: Fig. S8), following the same methodology previously described and 244,897 phased SNPs.

To explore possible admixture events that could have led to the origins of the Bubi, we also performed GLOBETROTTER analysis (based on ChromoPainter); however, no admixing events could be identified. We also used fineSTRUCTURE to plot a PCA based on

haplotype differentiation. This method clusters the populations into the same groups as those used when considering genotypes. In this analysis, the Bubi present a clear intermediate position between Western African and Bantu-speaking populations (Additional file 2: Figure S9). For computational convenience, we performed these analyses with a restricted dataset.

Furthermore, we also estimated identity by descent (IBD) tracks to help elucidate the recent co-ancestry links between the Bubi and mainland populations. We found the Bubi to share the highest number of IBD

blocks longer than 2 cM (indicative of genealogical connections occurring during the last 2500 years [29]) with populations from Angola (Ovimbundu, Kongo), as well as some from Gabon (e.g., Obamba, Duma, Bateke and Bapunu) (Fig. 6 and Additional file 1: Table S10). Additionally, we estimated the average of runs of homozygosity (ROHs) in each population, as well as the fraction of the genome in homozygosity as signals of endogamy. We found that the Bubi are the second most endogamous Bantu-speaking group (Additional file 1: Table S11) after the Bekwil population, although the ROHs of RHG tribes were longer on average than those observed among the agriculturalist groups (Additional file 2: Figure S10, Additional file 1: Table S12).

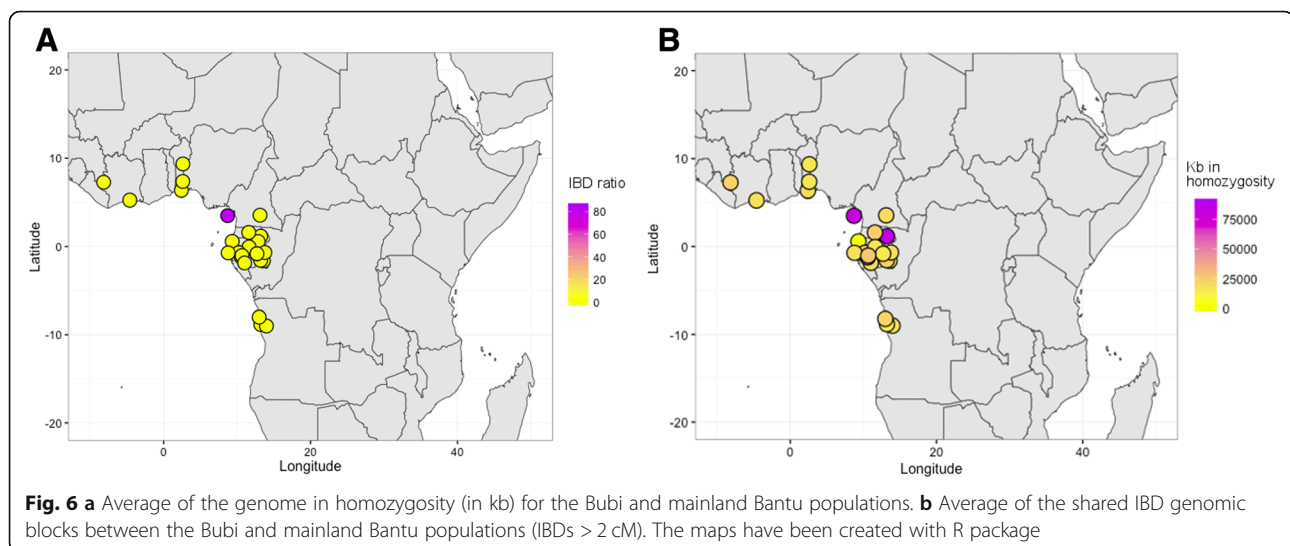
We found that all Bubi individuals possessed the malaria-resistant allele of the *ACKRI* [30] and *CD36* genes [31], in addition to certain variations in other genes such as *G6PD* [32], *ATP2B4* [33], *GRK5* [34], and *IL-10* [35]. Moreover, resistant variants are absent in *ABO*, *HBB* [36] and *TIRAP* [37] (Additional file 1: Table S13) [38]. However, considering these mutations are observed at low frequency among other African groups, it is likely they were simply not present in the ancestors of the Bubi (Additional file 1: Table S14). We compared the allelic frequencies of malaria SNPs in the Bubi with those from neighbouring populations of the Gulf of Guinea – such as Esan, Gambian, Mende, and Yoruba – for which genome-wide sequence data was available. We found statistically significant (Fisher's exact test) differences in some alleles, but nothing that indicated a unique trend in the Bubi (Additional file 1: Table S15). We subsequently conducted a genome-wide F_{st} scan between the Bubi and Yoruba, using all the variable positions with $MAF > 0.05$ and missing genotypes < 0.05 , plotting the mean F_{st} values in 0.5 Mb windows. We set a threshold of significance in 0.25 [39]. We have failed

to detect any signal of a selective event, including those regions related to immunity against malaria (Additional file 2: Figure S11).

Discussion

Our genomic study of the population of Bioko Island confirms that the Bantu-speaking migration that shaped most of the present-day human diversity in Sub-Saharan Africa [40] also had a significant impact on African islands of the Gulf of Guinea. The general components of ancestry found in the Bubi are not different from those found in mainland Bantu-speaking groups, although in the case of the Bubi, the RHG ancestry is lower than the amount detected in most Western Bantu-speaking groups. Moreover, we did not detect a significant difference in the origin of the Bubi RHG genetic signal to the one observed in other Bantu populations. One potential explanation could be that an admixture event between the ancestors of the Bubi and the RHG tribes started about 2000 years ago and was brought to the island upon settlement, but continued to increase thereafter in most mainland Bantu-speaking groups. It is worth noting that the time of admixture can be underestimated when using methods based on linkage-disequilibrium decay if continuous admixture events actually occurred [41]. Therefore, the current 800 years estimate [13] could in fact be the end of a long period of gene flow between mainland Bantu-speakers and RHG. This scenario could help explain the clustering of the Bubi with Western African groups in some analyses (the latter groups also show residual or no traces of RHG ancestry).

Due to a certain degree of heterogeneity detected within the Bubi that was evident from PCA, ADMIXTURE and fineSTRUCTURE analyses, the possibility that different populations from Bioko could harbour



slightly different genetic histories existed. Notably, some Bubi show almost no signs of RHG; interestingly, one of these individuals is from Bariobé, a relatively isolated province in the mountainous interior of the island. An alternative possibility could be that the small fraction of RHG ancestry was acquired by gene flow from coastal regions after the ancestors of the Bubi settled in Bioko. For example, the presence of both Fang and Benga people in Bioko has been described in historical times, partly related to the slave trade. In fact, although the slave trade was not so important in Bioko, it was very active in other coastal centres of the Gulf of Guinea, especially in some of the minor islands such as Corisco and Annobón [42]. Nonetheless, due to the cultural particularities of the Bubi and the clear genetic signals of endogamy and isolation, it seems unlikely they would assimilate a significant number of foreign people. In addition, no signals of potential Iberian admixture have been detected among the Bubi.

Within the Bantu-speakers, the Bubi are more closely related to Angolan than to the geographically closer Cameroon groups (this is supported for instance by fineSTRUCTURE or f_3 statistics). Based on the evidence that Bantu expansion likely moved from Angola northwards [13], it is possible that Bantu-speaking groups from Cameroon experienced subsequent admixture events with neighbouring RHG populations.

The Bubi particularities are mirrored by their geographically induced genetic isolation as well as their linguistic differences with neighbouring, mainland populations. At the linguistic level, the Bubi language is basal to most Bantu languages [40, 43] and clusters together with northwest Bantu speakers. This correlates with archaeological findings from the region dated from 5000 to 2500 years ago and associated to the spread of Bantu languages [40]. This decoupling between language and genetics could be explained if the former was acquired by or imposed onto the Bubi mainland ancestors. Accordingly, there are some historical accounts that consider the Bubi to be an enslaved tribe that escaped to Bioko [44].

The Bubi seem to have experienced a certain history of isolation that left a mark in their genomes. Out of all the Bantu-speaking groups, for instance, we found that the Bubi have some of the highest levels of IBD tracks shared among members of the same population, a signal of low diversity that is compatible with endogamy. In fact, in the ROH analysis, the Bubi rank as the second most endogamous Bantu-speaking group, only after the Bekwil. Nevertheless, the fact that the Bubi do not show a large genetic differentiation from potential source populations along the coast also indicates that drift did not have time to operate at large scale and that colonization of the island did not occur a long time ago.

The Bubi, like other groups from the Gulf of Guinea, display a high frequency of some mutations associated with protection against malaria. Other mutations, however, are absent or segregating. The underlying mutation for the Duffy-negative phenotype (at the *ACKR1* gene) that is known to protect against *Plasmodium vivax* and *P. knowlesi*, seems to be fixed, or at least is present at extremely high frequencies, in the Bubi population. In fact, this is a common trait in all Western Africa. At the beginning of the twenty-first century, malaria was responsible for a child mortality rate of 152 per every 1000 births in Bioko island, a figure that is only beginning to decrease thanks to recent malaria control projects [27]. However, in a genome-wide scan performed against the Yoruba, we were unable to identify genomic regions in the Bubi that appear to be shaped by natural selection, even despite their insular conditions.

However, due to the limited sampling size and restricted distribution within Bioko, our study has to be considered as a preliminary assessment of the current Bubi genetic diversity. Despite evidences that our sampling size can effectively estimate parameters of genetic diversity from a larger population (see Methods last section), additional sampling with a broader geographical distribution should be undertaken in the future.

Conclusions

In addition to the general population affinities of the Bubi, we have unveiled genetic evidences of a certain degree of isolation, which can be related to the insular conditions; this trait is quite unique in most of African mainland populations. Our study of the genomic composition of the Bubi not only adds further information to the current genetic diversity within Africa and its Atlantic islands, but also points to the importance of the genome-wide analyses in unravelling population affinities, selective pressures and past migrations that can be correlated with linguistic and archaeological information. We conclude that the origins of the Bantu expansion still needs further research and that future retrieval of ancient genomes from Central and Western Africa could shed light on the cradle of the Bantu migrations.

Methods

Samples

All thirteen individuals analysed in this study are members of the Cultural Bubi Association of Fuenlabrada, Madrid (Spain). We obtained informed consent from all subjects. We discarded 25 of the interviewed individuals because of admixed ancestry; many of them had a recent Fang ancestor from the mainland. Even though most of the individuals were not born in Bioko, we verified that the selected individuals had all grandparents born in the island; many of the volunteers' direct ancestors come

from Malabo, Bariobé and Baney, which are located in the Northeast region of Bioko (Additional file 1: Table S1).

Extraction, sequencing and mapping

We isolated DNA from cotton swabs using all the available material and an organic-based DNA extraction method adapted to Amicon® Ultra 0.5-mL columns [45]. After extraction, we concentrated the DNA by centrifugation up to 50 μ L and subjected samples to a quality control. To ensure there was a proper DNA concentration, 1 μ L of sample was loaded in a 1% agarose gel and stained with ethidium bromide. Only a single band was observed. The samples were quantified with BioTek's Epoch and yielded values, on average, of 68.88 ng/ μ L.

Genomic DNA libraries were prepared using TruSeq DNA PCR-Free Library Preparation Kit (in accordance with the general settings of the preparation guide). The procedure produced a PCR-free library with 350 bp average insert size that requires 20 ng/ μ L (in 50 μ L samples). DNA samples were randomly fragmented by Covaris system and sequenced in HiSeqX10 (Illumina) with hiSeq2x150bp settings plus 65 bp paired-end adapters at Macrogen (South Korea).

We evaluated the paired-end sequenced reads with FASTQC to check their quality. The sequencing adapters were then removed using Adapter removal [46], reads shorter than 30 bp were removed, and the reads were mapped against the Human reference genome [National Center for Biotechnology Information (NCBI) 37, hg19] using Burrows-Wheeler Aligner (BWA) with default parameters [47]. Duplicated reads were removed using Picard Tools MarkDuplicates version 2.8.3 and low quality mapping reads (< 30) were removed with SAMtools version 1.6.23 [48].

Genotyping and quality control tests

Unique aligned reads were processed with Base Quality Score Recalibration (BQSR) implemented in the GATK version 3.7 software [49]. Even if the plots did not show signals of systematic errors, we applied recalibration to all filtered reads. We used GATK HaplotypeCaller in GVCF mode for scalable variant calling (using the GRCh37 as a reference sequence). Individual variant calls were merged in a single VCF file using GATK genotypeGVCFs tool, and the variants were filtered using Variant Quality Score Recalibration (VQSR) with a filter level of 99%. We used QD, MQ ReadPosRankSum, FS, and SOR annotations in this step. We excluded any variant with less than 70% of the main depth coverage or more than 200%. We also removed those variants exhibiting qualities below 30. We removed those called variants with a minimum allele frequency below 0.05 and a Hardy-Weinberg disequilibrium p -value below $1e-6$.

Population genetics dataset

We merged our filtered variants with 690,739 SNPs from 1235 genotyped individuals belonging to 35 Western Africa populations. This dataset includes: Bantu-speaking populations, hunter-gatherers and Western African groups [13], using Plink 1.9 [50] (Additional file 1: Table S2). We excluded triallelic sites, A/T and C/G mutations and all sites with a minor allele frequency (MAF) below 0.05. We subsequently removed positions with > 10% missing data and those individuals with > 5% missing values. To ensure that genotypes were properly called after merging the dataset, the Yoruba SNP genotypes were compared against the 1000 Genomes Yoruba population. However, subsequent analyses were performed only with the Yoruba genotypes from Western Africa dataset [51]. Positions that exhibited > 0.2 values of pairwise F_{st} between both samples were also removed. Based on the colonial history of Bioko, we have assessed the presence of potential genetic admixture of the Bubi with Spanish individuals, adding Iberian samples from 1000 Genomes [52] to the SNP dataset. After this procedure, we again removed positions with MAF below 0.05, missing data above 0.1 and Hardy-Weinberg disequilibrium p -values below $1e-10$.

For most of the analyses, we have extracted a sub-dataset with representative populations from Western and Central Africa. This reduced dataset includes 14 populations and 169 individuals (Additional file 1: Table S3). Some of the population genomics analysis require an unrelated outgroup to the tested populations. We have merged our genotypes with data of eleven San individuals [53] from the Human Origins array [54], followed with the same merging procedure previously detailed. The resulting African dataset –including the Bubi– comprises 130,647 SNPs present in 1259 individuals.

Mitochondrial (mt) DNA and Y-chromosome analysis

Reads were mapped against the Revised Cambridge Reference Sequence (rCRS) of the human mtDNA [55]. After calling variants with GATK version 3.7 [49] as has been previously described, the mtDNA haplogroups were predicted using Haplogrep version 2 [56]. Y chromosome haplogroups were predicted by classifying the observed mutations according to the PhyloTree database [57].

Population genomic analyses

To situate the Bubi within the present diversity of the Gulf of Guinea and Western Africa, a principal components analysis (PCA) with the reduced dataset was generated using EIGENSOFT software [58]. Results were plotted using R package ggplot2 [59, 60].

ADMIXTURE plots were generated to estimate the proportions of K ancestral components on each individual genome [61] of the reduced dataset. As the analysis assumes linkage disequilibrium (LD), we pruned the dataset. We used Plink 1.9 to remove SNPs with an $LD > r^2 = 0.5$ in windows of 50 SNPs. ADMIXTURE analyses were performed with K from 2 to 15 and were repeated five times. The ADMIXTURE iterations were consolidated using CLUMPP with the large K greedy algorithm [62] and the results were plotted using R package *pophelper* [63].

Outgroup f_3 statistic is a useful test to determine the closest population to a target one using one outgroup population and measuring the amount of shared genetic drift with a test population. San were selected as outgroup, as they represent the most distant African population with genome-wide data, Bubi population was compared to all other populations in the dataset. The f_3 (San; Bubi, *Test*) statistic was calculated with *popstats* [64] and the results were again plotted using R. f statistics can also be implemented in order to determine which populations exhibited the highest genetic drift with the Bubi people, to do so, we used the *popstats* software to compute the f_4 statistic (*Test*, San; Bubi, Mbuti), (*Test*, San; Bubi, Baka), (*Test*, San; Bubi, Yoruba), (*Test*, San; Bubi, Fang). These combinations allow us to dissect the genetic admixture of the tested populations with the Bubi in relation to all the representative source of genetic ancestry in Western Africa: Eastern RHG, Western RHG, Western-African populations and Bantu-speaking populations.

The fixation index (F_{st}) is a measure of population differentiation. We calculated the mean pairwise F_{st} values between all the populations present in the global dataset. All autosomal SNPs were included in this analysis using the approach of Cockerham and Weir integrated in Plink 1.9 [65]

The reduced dataset was phased with SHAPEIT2 [66], using 500 states, 50 MCMC main steps, 10 burn-in and 10 pruning steps; recombination maps were interpolated from the HapMap phase 2 genetic maps. After excluding all positions with at least one missing site, we ended up with a dataset of 491,203 variable positions with no missing data.

We used CHROMOPAINTER to build a coancestry matrix based on haplotype data from the phased-reduced dataset. This software estimates the admixture proportions in recipient chromosomes by painting the proportion of each genetic component from the donor populations. We ran CHROMOPAINTER with linked data, estimating n and M parameters through an observation run with no prefixed parameters and including 30 randomly selected samples and three randomly selected chromosomes; fineSTRUCTURE analysis was

performed with the counts obtained in CHROMOPAINTER and ran with 1000,000 Markov chain Monte Carlo (MCMC) iterations and output printed every 10,000 iterations. The best tree was calculated with 10,000 state attempts. We also generated a haplotype-based PCA with fineSTRUCTURE.

To identify any admixture events between Bubi ancestors and other populations during the last 4500 years, we used the GLOBETROTTER [41] software on the basis of the defined clusters from fineSTRUCTURE (Additional file 1: Table S4).

Identity by descent (IBD) analysis

Identity by descent (IBD) blocks are defined as identical chromosome fragments present in multiple individuals that have been inherited from the same ancestral chromosome [67]. We have used RefinedIBD software [68] setting “ibdcm” = 0.5, “ibdtrim” = 62, “ibdwindow” = 2478, and “overlap” = 413; the rest of the parameters were assigned by default. All IBD blocks longer than five centimorgans (cM) were kept and the statistical threshold marked by LOD (the base 10 log of the likelihood ratio of the IBD segments, which is a figure that will depend of the size of the database and the genetic diversity within it) was assigned by default (> 3). The number of SNPs used here was 685,382. We then filtered the IBD segments to keep only those that were shared by any Bubi and another individual of the dataset (including the IBD fragments shared by two Bubi individuals). To reduce the impact that the population size could have on the global counts of IBD blocks per population, we corrected the value of the shared IBD fragments (IBD_n) by the population size (t). In order to obtain the average of the IBD blocks shared by any Bubi with any other individual or population, we divided each number obtained in the previous step by the number of Bubi individuals, $13:\text{ratioBubi_pop} = (\text{IBD}_n/t)/13$.

Runs of homozygosity (ROHs) analysis

ROHs (> 1000 kb) were estimated with Plink software. First, we calculated the average (in kilobases) of the genome that it is in homozygosity for each population. Second, we calculated the average of the number of genomic fragments that are in homozygosity for each population.

Malaria resistance

Relevant mutations associated with malaria resistance in 10 different genes (Additional file 1: Table S11) – as found in genome-wide association studies (GWAS) and other previous studies [31, 69] – were genotyped in Bubi and the 1000 Genomes African populations. Fisher’s exact test was used to determine the statistical significance of the observed differences ($p < 0.001$).

Evaluation of the effects of limited sample size

We have used a whole genome F_{st} approach to evaluate the effects of the small sample size used in this work. We have randomly grouped the 186 Yoruba individuals from 1000 Genomes in 14 subsamples of 13–17 individuals and we have estimated the mean pairwise F_{st} values among all population combinations. All autosomal SNPs were included in this analysis using the approach of Cockerham and Weir integrated in Plink 1.9 [65]. No comparison has shown values of mean pairwise F_{st} higher than 0.1, which indicates that the sub samples do not show significant differences in terms of genetic diversity (Additional file 2: Figure S12). This result suggests that the limited Bubi sample size can be used to infer genetic diversity at a higher population level.

Additional files

Additional file 1: Tables. (XLSX 78 kb)

Additional file 2: Supplementary figures. (PDF 1206 kb)

Abbreviations

BQSR: Base Quality Score Recalibration; F_{st} : Fixation index; GWAS: Genome-wide association study; HLA: Human-leukocyte antigen; IBD: Identity by descent; LD: Linkage disequilibrium; MAF: Minor allele frequency; MCMC: Markov chain Monte Carlo; mtDNA: Mitochondrial DNA; NCBI: National Center for Biotechnology Information; PCA: Principal component analysis; RHG: Rainforest hunter-gatherer; ROHs: Runs of homozygosity; SNP: Single nucleotide polymorphism; VQSR: Variant Quality Score Recalibration

Acknowledgements

We are grateful to the Bubi people from the Asocuba Association of Fuenlabrada for their ongoing support along the study and to Justo Bolekia Boloka and Jose F. Gómez for their help.

Funding

This research was supported by a grant from FEDER and Ministry of Economy and Competitiveness (BFU2015–64699-P) of Spain to C.L.-F. CLF is supported by Obra Social "La Caixa" and Secretaria d'Universitats i Recerca Programme del Departament d'Economia i Coneixement de la Generalitat de Catalunya (GRC 2017 SGR 880).

Availability of data and materials

Sequenced reads of the 13 Bubi people analysed in the present study have been submitted to European Nucleotide Archive under the accession numbers: ERR2640217-ERR2640226.

Authors' contributions

A.G.-M. and C.L.-F. conceived and coordinated the study; E.C., A.G. and A.G.-M. collected the samples; A.G. extracted the DNA; E.C., P.G., B.M., T.d.-D., E.P. and M.F.-B. analysed data and interpreted results; A.G.-M., C.L.-F. and P.G. wrote the paper. All authors reviewed the manuscript. All authors have read and approved the manuscript.

Ethics approval and consent to participate

Written informed consent was provided by all participants. The confidentiality and anonymity of all participants has been guaranteed. Experimental protocols and informed consent have been approved by the Clinical Research Ethics Committee from Institut Hospital del Mar d'Investigacions Mèdiques in Barcelona (CEIC-PSMAR)(2018/7845/I).

Consent for publication

The consent for publication has been given by all the participants in the study.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Author details

¹Institute of Evolutionary Biology (CSIC-Universitat Pompeu Fabra), Barcelona, Spain. ²Laboratory of Anthropology, Department of Biology, University of Florence, Florence, Italy. ³Department of Biodiversity, Ecology and Evolution, Complutense University of Madrid, Madrid, Spain. ⁴Forensic Genetics Laboratory, GENOMICA S.A.U., Pharma Mar Group, Madrid, Spain. ⁵Unit of Human Evolutionary Genetics, Department of Genomics & Genetics, Institut Pasteur, Paris, France. ⁶CNRS UMR 2000, Paris, France. ⁷Center of Bioinformatics, Biostatistics and Integrative Biology, Institut Pasteur, Paris, France.

Received: 6 November 2018 Accepted: 13 February 2019

Published online: 06 March 2019

References

- Günter T, Los Bubis de Fernando Póo. Trujillo JR, Rodríguez B, editors. Madrid: SIAL ediciones; 2008. p. 270.
- Bolekia BJ. Lingüística Bantú a través del Bubi. Salamanca: Universidad de Salamanca; 2008. p. 192.
- Jeffreys MDW. Some notes on the Neolithic of West Africa. In: Desmond Clark J, Cole S, editors. Third pan-african congress on prehistory. Livingstone: London Chatto & Windus; 1951. p. 262–73.
- Eteo Soriso JF. Los ritos de paso entre los Bubis. Madrid: SIAL- Casa de África; 2017.
- Burton RF. A visit to Fernando Po Peak, and a Night in the Open. Alp J. 1872;VI(XXXVII).
- Martín del Molino A. Tipología de la cerámica de Fernando Poo. Vol. 1, Estudios del Instituto Claretiano de Africanistas (Separata de la revista "La Guinea Española"). Santa Isabel: Instituto Claretiano de Africanistas; 1960. 1–36 p.
- Martín del Molino A. Secuencia Cultural en el Neolítico de Fernando Póo. Madrid: Inst. Español de Prehistoria-C.S.I.C.; 1965.
- Martín del Molino A. Etapas de la cultura Carboneras de Fernando Poo en el primer milenio de nuestra Era. Madrid: Inst. Español de Prehistoria-C.S.I.C.; 1968.
- Martín del Molino A. Los Bubis. Ritos y Creencias. Madrid: Labrys 54, Ediciones; 1993.
- Diamond J. Guns, germs and steel. Norton W., editor. A short history of everybody for the last 13,000 years. New York: W. W. Norton & Company; 1997. p. 480.
- Nurse D, Philippson G. In: Nurse D, Philippson G, editors. The bantu languages. New York: Routledge, Taylor & Francis Group; 2003.
- Berniell-Lee G, Bosch E, Bertranpetit J, Comas D. Y-chromosome diversity in bantu and pygmy populations from Central Africa. Int Congr Ser. 2006; 1288:234–6.
- Patin E, Lopez M, Grollemund R, Verdu P, Harmant C, Quach H, et al. Dispersals and genetic adaptation of bantu-speaking populations in Africa and North America. Science. 2017;356(6337):543–6. <https://doi.org/10.1126/science.aal1988>.
- Busby GB, Band G, Si Le Q, Jallow M, Bougama E, Mangano VD, et al. Admixture into and within sub-Saharan Africa. eLife 2016;5:15266. doi: <https://doi.org/10.7554/eLife.15266>.
- de Filippo C, Bostoen K, Stoneking M, Pakendorf B. Bringing together linguistic and genetic evidence to test the bantu expansion. Proc R Soc B Biol Sci. 2012;279(1741):3256–63. <https://doi.org/10.1098/rspb.2012.0318>.
- Tishkoff SA, Reed FA, Ranciaro A, Awomoyi AA, Bodo J, Doumbo O, et al. The genetic structure and history of Africans and African Americans. Science. 2009;324:1035–44. <https://doi.org/10.1126/science.1172257>.
- Batini C, Coia V, Battaglia C, Rocha J, Pilkington MM, Spedini G, et al. Phylogeography of the human mitochondrial L1c haplogroup: genetic signatures of the prehistory of Central Africa. Mol Phylogenet Evol. 2007; 43(2):635–44. <https://doi.org/10.1016/j.ympev.2006.09.014>.

18. Quintana-Murci L, Quach H, Harmant C, Luca F, Massonnet B, Patin E, et al. Maternal traces of deep common ancestry and asymmetric gene flow between pygmy hunter-gatherers and bantu-speaking farmers. *Proc Natl Acad Sci U S A*. 2008;105(5):1596–601. <https://doi.org/10.1073/pnas.0711467105>.
19. Destro-Bisol G, Coia V, Boschi I, Verginelli F, Caglia A, Pascali V, et al. The analysis of variation of mtDNA hypervariable region 1 suggests that eastern and Western pygmies diverged before the bantu expansion. *Am Nat*. 2004;163(2):212–26. <https://doi.org/10.1086/381405>.
20. Belez S, Gusmão L, Amorim A, Carracedo A, Salas A. The genetic legacy of western bantu migrations. *Hum Genet*. 2005;117(4):366–75. <https://doi.org/10.1007/s00439-005-1290-3>.
21. Rowold D, Garcia-Bertrand R, Calderon S, Rivera L, Benedico DP, Alfonso Sanchez MA, et al. At the southeast fringe of the bantu expansion: genetic diversity and phylogenetic relationships to other sub-Saharan tribes. *Meta Gene*. 2014;2:670–85. <https://doi.org/10.1016/j.mgene.2014.08.003>.
22. Li S, Schlebusch C, Jakobsson M. Genetic variation reveals large-scale population expansion and migration during the expansion of bantu-speaking peoples. *Proc R Soc B Biol Sci*. 2014;281(1793):20141448. <https://doi.org/10.1098/rspb.2014.1448>.
23. Skoglund P, Thompson JC, Prendergast ME, Mitnik A, Sirak K, Hajdinjak M, et al. Reconstructing Prehistoric African Population Structure. *Cell*. 2017;171(1):59–71.e21. <https://doi.org/10.1016/j.cell.2017.08.049>.
24. Schlebusch CM, Malmström H, Günther T, Sjödin P, Coutinho A, Edlund H, et al. Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science*. 2017;358(6363):652–5. <https://doi.org/10.1126/science.aao6266>.
25. Cibulskis RE, Aregawi M, Williams R, Otten M, Dye C. Worldwide incidence of malaria in 2009: estimates, time trends, and a critique of methods. *PLoS Med*. 2011;8(12):e1001142. <https://doi.org/10.1371/journal.pmed.1001142>.
26. World Health Organisation. World Malaria Report 2017. 2017.
27. Rehman AM, Mann AG, Schwabe C, Reddy MR, Roncon Gomes I, Slotman MA, et al. Five years of malaria control in the continental region, Equatorial Guinea. *Malar J*. 2013;12(1):154. <https://doi.org/10.1186/1475-2875-12-154>.
28. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res*. 2002;12(6):996–1006. <https://doi.org/10.1101/gr.229102>.
29. Ralph P, Coop G. The Geography of Recent Genetic Ancestry across Europe. *PLoS Biol*. 2013;11(5):e1001555. doi.org/10.1371/journal.pbio.1001555.
30. Tournamille C, Colin Y, Cartron JP, Le Van Kim C. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nat Genet*. 1995;10:224–8. <https://doi.org/10.1038/ng0695-224>.
31. Rockett KA, Clarke GM, Fitzpatrick K, Hubbart C, Jeffreys AE, Rowlands K, et al. Reappraisal of known malaria resistance loci in a large multicenter study. *Nat Genet*. 2014;46(11):1197–204. <https://doi.org/10.1038/ng.3107>.
32. Hirono A, Beutler E. Alternative splicing of human Glucose-6-phosphate dehydrogenase messenger RNA in different tissues. *J Clin Invest*. 1989;83:343–6. <https://doi.org/10.1172/JCI113881>.
33. Timmann C, Thye T, Vens M, Evans J, May J, Ehmen C, et al. Genome-wide association study indicates two novel resistance loci for severe malaria. *Nature*. 2012;489(7416):443–6. <https://doi.org/10.1038/nature11334>.
34. Gupta H, Jain A, Saadi AV, Vasudevan TG, Hande MH, D'Souza SC, et al. Categorical complexities of Plasmodium falciparum malaria in individuals is associated with genetic variations in ADORA2A and GRK5 genes. *Infect Genet Evol*. 2015;34:188–99. <https://doi.org/10.1371/journal.pone.0175702>.
35. Gibson AW, Edberg JC, Wu J, Westendorp RG, Huizinga TW, Kimberly RP. Novel single nucleotide polymorphisms in the distal IL-10 promoter affect IL-10 production and enhance the risk of systemic lupus erythematosus. *J Immunol*. 2001;166(6):3915–22.
36. Gouagna LC, Bancone G, Yao F, Yameogo B, Dabire KR, Costantini C, et al. Genetic variation in human HBB is associated with Plasmodium falciparum transmission. *Nat Genet*. 2010;42(4):328–31. <https://doi.org/10.1038/ng.554>.
37. Khor CC, Chapman SJ, Vannberg FO, Dunne A, Murphy C, Ling EY, et al. A mal functional variant is associated with protection against invasive pneumococcal disease, bacteremia, malaria and tuberculosis. *Nat Genet*. 2007;39(4):523–8. <https://doi.org/10.1038/ng1976>.
38. Manske M, Miotto O, Campino S, Auburn S, Almagro-Garcia J, Maslen G, et al. Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. *Nature*. 2012;487(7407):375–9. <https://doi.org/10.1038/nature11174>.
39. Hartl DL, Clark AG. Principles of population genetics; 1997. p. 546.
40. Currie TE, Meade A, Guillon M, Mace R. Cultural phylogeography of the bantu languages of sub-Saharan Africa. *Proc R Soc London B Biol Sci*. 2013;280:1762. <https://doi.org/10.1098/rspb.2012.0318>.
41. Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D, et al. A genetic atlas of human admixture history. *Science*. 2014;343:747–51. <https://doi.org/10.1126/science.1243518>.
42. Thomas H. In: Paperbacks S& S, editor. The Slave Trade: The story of the Atlantic slave trade: 1440–1870. New York: Simon & Schuster; 1999.
43. Guthrie M. Comparative bantu: an introduction to the comparative linguistics and prehistory of the bantu languages. Ltd GIP, editor. Farnborough: Ltd, Gregg International Publishers; 1971.
44. Aymemí A. Los Búbis en Fernando Poo: colección de los artículos publicados en la revista colonial La Guinea española. Madrid: Ed. G. Sáez; 1942.
45. Sambrook J, Fritsch EF, Maniatis T. Molecular cloning: a laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1989.
46. Lindgreen S. AdapterRemoval: easy cleaning of next generation sequencing reads. *BMC Res Notes*. 2012;5(1):337. <https://doi.org/10.1186/1756-0500-5-337>.
47. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. 2009;25(14):1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
48. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
49. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1298–303. <https://doi.org/10.1101/gr.107524.110>.
50. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81(3):559–75. <https://doi.org/10.1086/519795>.
51. Patin E, Siddle KJ, Laval G, Quach H, Harmant C, Becker N, et al. The impact of agricultural emergence on the genetic history of African rainforest hunter-gatherers and agriculturalists. *Nat Commun*. 2014;5:3163. <https://doi.org/10.1038/ncomms4163>.
52. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68–74. <https://doi.org/10.1038/nature15393>.
53. Lazaridis I, Patterson N, Mitnik A, Renaud G, Mallick S, Kirsanow K, et al. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature*. 2014;513(7518):409–13. <https://doi.org/10.1038/nature13799>.
54. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y. Ancient Admixture in Human History. *Genetics*. 2012;192:1065–93. <https://doi.org/10.1534/genetics.112.145037>.
55. Andrews RM, Kubacka I, Chinnery PF, Lightowler RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet*. 1999;23(2):147. <https://doi.org/10.1038/13779>.
56. Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, Specht G, Bandelt HJ, et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res*. 2016;44(W1):W58–63.
57. van Oven M. PhyloTree build 17: growing the human mitochondrial DNA tree. *Forensic Sci Int Genet Suppl Ser*. 2015;5:e392–4.
58. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006;2(12):2074–93. <https://doi.org/10.1371/journal.pgen.002019>.
59. Ross I, Robert G, Ihaka R, Gentleman R. R: A language for data analysis and graphics. *J Comput Graph Stat*. 1996;5:299–314.
60. Gómez-Rubio V. ggplot2 - Elegant Graphics for Data Analysis (2nd Edition). *J Stat Softw*. 2017;77:2–5.
61. Alexander DH, Novembre J. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*. 2009;19(9):1655–64. <https://doi.org/10.1101/gr.094052.109>.
62. Jakobsson M, Rosenberg NA. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*. 2007;23(14):1801–6. <https://doi.org/10.1093/bioinformatics/btm233>.
63. Francis RM. Pophelper: an R package and web app to analyse and visualize population structure. *Mol Ecol Resour*. 2017;17(1):27–32. <https://doi.org/10.1111/1755-0998.12509>.

64. Skoglund P, Mallick S, Bortolini MC, Chennagiri N, Hünemeier T, Petzl-Erler ML, et al. Genetic evidence for two founding populations of the Americas. *Nature*. 2015;525(7567):104–8. <https://doi.org/10.1038/nature14895>.
65. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution* (N Y). 1984;38(6):1358–70.
66. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods*. 2013;10(1):5–6. <https://doi.org/10.1111/j.1558-5646.1984.tb05657.x>.
67. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet*. 2007;81(5):1084–97. <https://doi.org/10.1086/521987>.
68. Browning BL, Browning SR. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics*. 2013;194(2):459–71. <https://doi.org/10.1534/genetics.113.150029>.
69. Gelabert P, Olalde I, De-Dios T, Civit S, Lalueza-Fox. Malaria was a weak selective force in ancient Europeans. *Sci Rep*. 2017;7(1):1377. <https://doi.org/10.1038/s41598-017-01534-5>.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

