Research Article

Palenque de San Basilio in Colombia: genetic data supports an oral history of a paternal ancestry in Congo

Naser Ansari-Pour,[1*] Yves Moñino,[2] Constanza Duque,[3,4] Natalia Gallego,[3,5] Gabriel Bedoya,[3] Mark G Thomas[6] and Neil Bradman[7]

[1] Faculty of New Sciences and Technology, University of Tehran, Tehran, Iran
[2]LLACAN, CNRS, Villejuif, Paris, France
[3] Universidad de Antioquia UdeA, Calle 70 No 52-21 Medellín Colombia
[4] Universidad Cooperativa de Colombia, Sede Medellín, Colombia
[5] Institución Universitaria Colegio Mayor de Antioquia, Medellín, Colombia
[6]Department of Genetics, Evolution and Environment, University College London, London, UK
[7]Henry Stewart Group, 29/30 Little Russell Street, London, UK

Corresponding Author:

Dr Naser Ansari-Pour

E-mail: n.ansaripour@ut.ac.ir

Abstract

The Palenque, a black community in rural Colombia, have an oral history of fugitive African slaves founding a free village near Cartagena in the 17th Century. Recently linguists have identified some 200 words in regular use that originate in a Kikongo language, with Yombe, mainly spoken in the Congo region, being the most likely source. The non-recombining portion of the Y chromosome (NRY) and mitochondrial DNA were analysed to establish whether there was greater similarity between present day members of the Palenque and Yombe than between the Palenque and 42 other African groups (for all individuals, n=2,799) from which forced slaves might have been taken. NRY data are consistent with the linguistic evidence that Yombe is the most likely group from which the original male settlers of Palenque came. Mitochondrial DNA data suggested substantial maternal sub-Saharan African ancestry and a strong founder effect but did not associate Palenque with any particular African group. In addition, based on cultural data including inhabitants' claims of linguistic differences, it has been hypothesized that the two districts of the village (Abajo and Arriba) have different origins, with Arriba founded by men originating in Congo and Abajo by those born in Colombia. Although significant genetic structuring distinguished the two from each other, no supporting evidence for this hypothesis was found.

Introduction

In many locations throughout the Caribbean and Latin America during the Atlantic slave trade, runaway slaves in pursuit of freedom established fortified villages. In Colombia, these walled towns (known as palenques) were famed for their resistance to the Spanish military conquest. This reputation is evident from colonial records which tell how inhabitants successfully repulsed attacks by the authorities [1]. Despite their resistance, Palenque de San Basilio is the only palenque to have survived to the present day [2].

Palenque de San Basilio (Palenque for short) is located some 70 km south east of the regional capital of Bolivar, Cartagena, in North West Colombia (10.1°N, 75.2°W) (see Fig. 1). The residents comprise a community of about 3,500 individuals divided between two major districts, Arriba and Abajo, although the reason for the division is not established [2, 3]. They have remained largely isolated from the prevailing Hispanic culture, living by subsistence farming together with cattle husbandry [4]. Their oral history is one of descent from a group of male slaves who escaped captivity early in the 17th century from nearby Cartagena (then a major centre of the Latin American slave trade [5]).

Interestingly, Palenque is the only Colombian black community that speaks a creole Spanish known as Palenquero [6]. Linguistic analysis of this creole led to the suggestion that the language of the founding group originated in the area of present day Congo and/or northern Angola [4, 7]. More recently, detailed lexical research has established that Palenquero contains more than 200 words of African origin [8, 9] and that Kikongo is the only demonstrable donor of the vocabulary [10]. The Kikongo group of languages encompasses several extant tongues which are spoken by approximately one million people in Republic of the Congo [11]. Although the recorded vocabulary in Palenquero does not suggest a particular origin among them, the ritual vocabulary [10] and oral history [12] suggest that Yombe is the most probable source. Today Yombe is spoken by the Yombe people, an ethnic group living mainly in Pointe-Noire (Republic of the Congo). Furthermore, many members of the Palenque community have claimed that a) Arriba residents are more traditional and have better conserved Palenquero than their Abajo counterparts; and b) the founding men of Arriba were born in Congo while Abajo would have been populated by Maroons born in Colombia (Yves Moñino (YM), field work in Palenque and [3]).

To clarify these questions concerning Palenque history, we undertook a genetic analysis of individuals from both Arriba and Abajo. DNA analysis has proved useful in revealing origins of ethnic communities (for example see [13-15]). Sex specific genetic systems (the non-recombining portion of the Y chromosome (NRY) and mitochondrial DNA (mtDNA)) have been analysed to reveal connections between geographically separated diaspora communities sharing a common identity [15-17] and to evaluate support for alternative oral histories [18]. Recently, the geographic distribution of NRY haplotypes and time to the most recent common ancestor (TMRCA) of paternal haplogroups were interpreted as suggesting a late, exclusively eastern, expansion of the Bantu speaking peoples (EBSP) [19].

The geographic origins of African diasporas, in particular those created by the Atlantic slave trade, have been investigated using NRY and mtDNA. In studies of the populations of Cape Verde Islands [20, 21] and Sao Tome Island [22, 23], sex specific genetic systems were used to elucidate both maternal and paternal origins. In the case of the Palenque, analysis of HLA autosomal markers and antigens [24, 25], and recently NRY variation [26] has suggested a greater proportion of recent African descent (RAD) than other Colombian groups.

Although culturally and geographically isolated for most of its existence, during the past few decades, the Palenque people have experienced more contact with those from outside their group [10]. Therefore, in recent times, an increased level of gene flow may have occurred. Given the substantial geographic structuring of NRY and mtDNA haplotypes at the continental level, and assuming that the founding group was of RAD, genetic analysis can provide evidence of geographic ancestry and potential gene flow from non-RAD groups. If the male founders were of RAD and there has been little gene flow from Europeans and Amerindians, it can then be expected that NRY haplotypes will match those common in sub-Saharan Africa and will have low diversity respectively.

Palenque oral tradition provides a testable hypothesis for NRY but not mtDNA variation. However, from colonial records, it appears that in the second half of the 18th century 178 black families occupied Palenque [5]. Therefore, it can be hypothesised that the majority of the females at that time had RAD. If there has been little female gene flow since that time, then the expectation is that mtDNA haplotypes will match those commonly seen in sub-Saharan Africa.

Sub-Saharan Africa is known for its relatively high human genetic diversity [27, 28], and geographic structuring of mtDNA haplotypes has been recognized [29]. Furthermore, the

considerable increase in NRY polymorphic sites identified in recent years [30, 31] has revealed geographic structuring of NRY haplotypes [19, 32]. These findings have made it possible, in some cases, to reveal recent shared paternal descent of men with a RAD born outside sub-Saharan Africa with men still living there [21, 23, 26].

To explore these questions about Palenque history based on anthropological and linguistic studies we analysed NRY and mtDNA in the Palenque and 42 sub-Saharan groups. We address the following three questions: a) Is there greater genetic similarity between the inhabitants of Palenque and Yombe speakers than between Palenque and non-Yombe African groups?; b) Is there a significant difference between the sex-specific genetic systems profiles of present day residents of Abajo and Arriba?; and c) Are the NRY and mtDNA of the Arriba inhabitants more similar to those of Republic of the Congo than are the NRY and mtDNA of Abajo residents? Genetic data analysed in this paper support the prior hypotheses that (a) Palenque have a paternal line founding origin in the Yombe and (b) there is significant difference in NRY distribution between Abajo and Arriba, but not mtDNA. There is limited NRY but not mtDNA support for an affirmative answer to (c).

Materials and Methods

Sample collection

In Palenque, buccal swabs were collected from males over eighteen years old currently living in, or born in, the community. Donors were initially selected randomly but after questioning, only one of each set of donors having a common paternal grandfather was included in the study. Samples were collected from a very substantial proportion of individuals satisfying the above criterion (estimated at >90%, n = 153: Abajo area n = 88, Arriba area n = 52, others n = 13). Samples from eight groups in the Republic of the Congo (n = 591) were collected at local gatherings in different areas of Brazzaville, Pointe Noire and in the villages of Kakamoeka and Lovoulou, 90km and 70km inland from Pointe Noire respectively.

Ethnographic data were gathered from each Palenque individual, adopting the procedure reported in Ansari-Pour et al. [19]. Buccal swabs previously collected from 34 sub-Saharan groups in West, Central West and South-East Africa, representing other potential source populations for the Atlantic slave trade, were also analysed in this study (see Supplementary Table S1; samples from all population groups other than Palenque were included in Ansari-Pour et al. [19]). DNA from all Congolese and Palenque samples was

extracted using the Gentra protein precipitation method (Gentra Systems, Minneapolis) while the standard phenol-chloroform method was used for all other samples

DNA typing

The battery of Y-chromosome presumed unique event polymorphisms (UEPs), consisting of single nucleotide polymorphisms (SNPs) and insertion/deletion polymorphisms, as well as a set of short tandem repeats (STRs) were typed in the Palenque samples as described by Ansari-Pour et al. [19]. Briefly: (a) sixteen UEPs (see Figure 2) were used to classify NRY into haplogroups, applying the nomenclature of the Y Chromosome Consortium [31] with the 'capital letter- mutation' system, and within each haplogroup, (b) six STRs (DYS19, DYS388, DYS390, DYS391, DYS392 and DYS393) were used to define haplotypes. Equivalent Y chromosome data for all 42 sub-Saharan African population samples, including the eight Congolese groups (see Table S1) were taken from Ansari-Pour et al. [19].

The mtDNA HVR-1 region of all Congolese groups and Palenque was sequenced as described by Veeramah et al. [18]. For all samples, HVR-1 variable site only (VSO) haplotypes were determined by comparing sequences of nucleotide range 16020-16400 with the revised Cambridge Reference Sequence [33]. Haplotypes were defined by substitutions, insertions and deletions, and their corresponding nucleotide positions. Tentative mtDNA haplogroup assignment, based on HVR-1 sequences, were inferred according to the scheme of Salas et al. [34], although it should be noted that inferred haplogroups frequencies were not used in our statistical analyses and are only presented for reference. To extend the mtDNA dataset, HVR-1 haplotypes were also determined for 30 out of 34 non-Congolese sub-Saharan population samples considered in the NRY analyses (i.e. all groups except Sena, Tumbuka, Bantu speakers from Pretoria and Yao; unpublished data except for the Nigerian groups [35]). To facilitate comparison of all population samples, the range of the HVR-1 region considered was reduced to 16023-16380.

Statistical analysis

Pairwise genetic differences between population samples were assessed using the exact test of population differentiation (ETPD) [36] which is analogous to Fisher's exact test extended to an m x n matrix, where m is the number of groups and n is the number of distinct haplotypes. Gene diversity and its standard error were estimated using the unbiased formula

of Nei [37]. Genetic distances calculated were: F_{ST} [38] based on UEP haplogroups, STR haplotypes (respecting their classification within haplogroups, i.e. UEP+STR haplotypes), and mtDNA HVR-1 haplotypes and imputed haplogroups, R_{ST} [39] based on six STRs on the NRY, and Kimura's two-parameter model with gamma value of 0.47 [40] for mtDNA HVR-1 sequences. It should be noted that F_{ST} is used here as in Thomas et al. [17] as a convenient statistic summarizing multidimensional differences in allele frequencies. No further assumptions regarding the underlying population genetic model were applied in its interpretation, other than a monotonic relationship between F_{ST} and genetic differences.

Analyses based on a selection of UEPs may suffer from biases in their ascertainment. However, given the geographic structuring of the NRY variation, the choice of UEPs is appropriate for the comparisons of sub-Saharan and RAD individuals. To test if genetic distances differed significantly from zero, haplogroups/haplotypes were permuted among samples; 1,000 permutations were performed to generate a null distribution of pairwise genetic distances.

All of the above analyses were performed using Arlequin software version 3.0 [41]. Principal Component Analysis (PCA) and the nonparametric 'Sign Test' were performed using the 'R' statistical programming language (www.R-project.org) [42], using 'princomp' and 'binom.test' functions respectively. PCA plots were used to visualize relationships among population samples based on NRY haplogroup frequencies.

Results

Frequencies of NRY haplogroups and NRY based genetic distances

The frequencies of all observed NRY haplogroups in the Palenque and the 42 sub-Saharan groups analysed in this study are included in Table 1. The phylogenetic relationships of the haplogroups can be seen in Figure 2. Thirteen NRY haplogroups were observed, of which ten were present in the Palenque. The modal haplogroup in the Palenque was E-U175 (27%), but the two districts of the village had different modal haplogroups (details below). Notably, there were only three haplogroups present in sub-Saharan African groups not observed in the Palenque dataset: DE-YAP which is found at very low frequency in Nigeria [43]; A-M13 which forms a very basal clade in the NRY phylogeny and has a wide distribution at low frequency in Africa [44-46]; and E-U181 which has been proposed as a signature of an exclusively eastern expansion of the Bantu speaking peoples [19]. P-92R7 and R1a1, both widely considered to be 'non-African origin haplogroups' [47], were observed at

18% and 2.7% respectively in Palenque while observed as a singleton or at low frequencies and completely absent in sub-Saharan African groups respectively. STR haplotypes within each haplogroup were then analysed (see Supplementary Table 2). Of note, the two most common STR haplotypes within P-92R7 in Palenque were haplotype 14-12-24-11-13-13 (N=5) and its one-step neighbour (14-12-24-12-13-13) (N=5). Both were absent from the sub-Saharan African dataset. The former has been designated the Atlantic Modal Haplotype (AMH) due to its high frequency in Western European populations [48, 49].

Gene diversity in the Palenque based on all haplogroups and E-sY81 component haplogroups was 0.830 ± 0.013 and 0.638 ± 0.035 respectively, while the equivalent statistics in the sub-Saharan African dataset were 0.753 ± 0.007 and 0.679 ± 0.008 respectively (for gene diversity in each individual group see Supplementary Table S3).

The genetic distinctiveness of Palenque compared with each of the sub-Saharan African groups was apparent as assessed by ETPD ($P<0.001$). Also, all F_{ST} values between Palenque and the sub-Saharan African groups were significant as assessed by random permutation (see Methods) ($P<0.00001$) with only two below 0.05 (Chewa, an East African group from Malawi ($F_{ST}=0.027$) and Yombe ($F_{ST}=0.035$)) (see Supplementary Table S4). F_{ST} between the Chewa and Yombe was not significant. This pattern was also consistently observed based on R_{ST} (see Supplementary Table S5). Comparison of haplogroup profiles in Palenque, Chewa and Yombe, revealed twelve NRY haplogroups present in at least one of the groups. Six were observed in all three groups (see Fig 3). Of the remaining six, four were observed in Palenque and Yombe, one in Palenque and Chewa, and one was observed only in the Palenque (see Fig. 3). Most notably, all the haplogroups observed in the Yombe were also observed in the Palenque, while the proposed signature haplogroup of the eastern EBSP (E1b1a8a1a; E-U181) [19] was absent in the Palenque and the Yombe.

Similar to the approach taken by Di Giacomo et al. [50], we compared the distribution of NRY variation within E-sY81 (E1b1a; the signature haplogroup of EBSP [19, 35, 51, 52]), a clade which was present in all population samples including Palenque, and observed only in men of RAD. F_{ST} between the Palenque and the other groups, based on the frequencies of the E-sY81 component haplogroups, revealed only two groups with a non-significant F_{ST} (Chewa and Yombe) with the Yombe-Palenque F_{ST} <0.001 (Supplementary Table S6). Based on the same dataset, pairwise differentiation between Palenque and all sub-Saharan African

population samples were also assessed using ETPD. Interestingly, all were significant at the 5% level except the Yombe (P=0.507).

A PCA plot using only E-sY81 component haplogroup frequencies showed Palenque as an outlier. While a mixed collection of Bantu speaking groups are nearer than other Niger-Congo groups to Palenque, strikingly it is the Yombe who are the closest of all (Figure 4).

The distribution of mtDNA variation

The Palenque sample contained 26 mtDNA HVR-1 haplotypes. The modal haplotype was at a frequency of 0.166 and five common haplotypes together accounted for 66.2% of the total. The imputed haplogroups were almost all sub-lineages of L (the sub-Saharan African modal haplogroup) with over 70% within either L1 or L3 (see Supplementary Table S7). Nei's gene diversity, based on mtDNA HVR-1 haplotypes, was 0.903 ± 0.010. Across all sub-Saharan African groups (38 groups), 723 mtDNA HVR-1 haplotypes (427 singletons) were observed. Gene diversity for the combined set was 0.993 ± 0.004 and in individual groups ranged from 0.968 to 0.997 (see Supplementary Table S8) with mean of 0.987 and SD of 0.007.

ETPD based on HVR-1 haplotypes was significant between Palenque and all sub-Saharan groups. This was also the case based on imputed haplogroups (except Sundi (N=25) with borderline P-value of 0.057). F_{ST} and K2P between Palenque and the sub-Saharan African groups were also all significant, and all were in the range of 0.037-0.066 and 0.039-0.126 respectively (see Supplementary Tables S9 and S10 respectively).

Intra-village analysis of Palenque

Summary statistics were calculated for both districts (Abajo and Arriba) in Palenque (see Supplementary Tables S11 and S12 for NRY and mtDNA raw data respectively). The modal NRY haplogroups in Abajo and Arriba were E-U175 (34.1%) and E-U290 (23.1%), respectively. The proportion of the E-sY81 clade in Abajo and Arriba was 55.3% and 51.9% respectively and not significantly different (P=0.727). Gene diversity based on all NRY UEP in Abajo and Arriba was 0.800 ± 0.024 and 0.853 ± 0.019 respectively, and 0.5597 ± 0.0632 and 0.6809 ± 0.0495 respectively when restricting analysis to E-sY81 NRY types. At the UEP+STR level, the modal haplotype was one STR mutation different from the EBSP modal haplotype (i.e. E-sY81-15-12-21-10-11-13) [19] in Abajo (E-sY81-16-12-21-10-11-13; 10.6%) and in Arriba (E-sY81-15-12-21-10-12-13; 19.2%), while the EBSP modal haplotype was at a frequency of 5.9% and 5.8% respectively.

Analysing P-92R7 NRY, the frequency was not significantly different in the two districts (15.3% in Abajo and 19.2% in Arriba, P=0.639). The distribution of constituent haplotypes was also not significantly different as measured by ETPD (P=0.738).

To investigate the hypothesis derived from observed cultural differences and local oral history (YM, personal field notes) that the two village districts can be distinguished from each other, several statistical tests were performed using the NRY and mtDNA data (see Supplementary Table S13). Strikingly, genetic comparisons based on mtDNA were not significant based on both haplotypes and imputed haplogroups, while all comparisons using NRY markers were significant at the 5% level. Notably F_{ST} between the two village districts, calculated using E-sY81 component haplogroups only, was both significant ($P<0.05$) and greater than between either of them and Yombe (Supplementary Table S14). The distinctiveness of Abajo and Arriba NRY was also confirmed by ETPD (UEP, P=0.006; UEP-E-sY81, P=0.012; UEP+STR, P=0.017). We then examined whether the ETPD was significant because of haplogroups introduced into Palenque probably through non-RAD introgression. Y chromosomes were divided into a) those collectively belonging to haplogroups K, P and R (all with inferred origins outside sub-Saharan Africa) and b) Y(xK,P,R). The difference was driven by the African Y(xK,P,R) NRY at both UEP (P=0.008) and UEP+STR (P=0.008) levels and not by the non-African (Y(K,P,R) NRY (P=0.114 and 0.271 at UEP and UEP+STR levels respectively). No significant difference was observed between the two districts based on mtDNA HVR-1 haplotypes and imputed haplogroups as assessed by ETPD (P=0.985 and P=0.77). When applying the same test both districts differed significantly from all sub-Saharan African groups ($P<0.00001$) at the haplotype level. Analysis of imputed haplogroups also showed a consistent pattern. This may be due to the presence of non-African mtDNA haplotypes including those defined by 16290T and 16319A (possibly belonging to the Amerindian A2 haplogroup) and high frequency of founder haplotypes such as that bearing 16294T and 16309G (possibly haplogroup L2a1).

Abajo and Arriba in the context of sub-Saharan Africa

F_{ST} between Abajo and Arriba, treated as separate samples, and 42 sub-Saharan African populations were estimated based on all haplogroups and E-sY81 component haplogroups (Supplementary Table S14) but not mtDNA HVR-1 haplotypes, since no significant genetic distance (F_{ST} and K2P) was observed between the two districts.

At the NRY-UEP level, all F_{ST} estimates were significant ($P<0.05$) with the exception of Abajo and Yombe. When considering only E-sY81 component haplogroups, F_{ST} was similarly not significant between Abajo and (a)Yombe and (b) Chewa. In addition, Arriba had a non-significant F_{ST} with Bembe and Yombe. Based on the magnitude of F_{ST}, at the NRY-UEP level, the Congolese groups were split in half; four were closer to Arriba than Abajo, and four were closer to Abajo than Arriba. However, at the E-sY81 level, six of seven (excluding Yombe which had a non-significant F_{ST}) were closer to Arriba. Comparisons with the complete African dataset presented a clearer difference. At both NRY-UEP and E-sY81 levels, Arriba was closer to 33 (Sign Test $P = 0.0001$) and 36 (Sign Test $P <0.0001$) out of 42 sub-Saharan groups, respectively.

Discussion

The evolutionary processes of mutation and genetic drift, including founder-effect, as well as the possible influence of natural selection, make direct inference of population history from genetic data challenging. However, such challenges become more tractable when clear hypotheses can be formulated from existing anthropological, linguistic and ethnographic research. In such circumstances, genetic data can, in some cases, be analysed to test those hypotheses. Even though the NRY and mtDNA are effectively single loci (because they are non-recombining regions) they can be appropriate systems for testing such hypotheses, particularly where the prior hypotheses concern only patrilineal or matrilineal history. In the current study, there were three prior hypotheses which we address in turn in the following sections.

The founding fathers of the Palenque community were primarily Yombe

Sex-specific genetic systems are particularly susceptible to genetic drift [53, 54]. The larger the population size and the fewer the generations since a postulated event, the less the effect of genetic drift should be. Because forced slaves are recorded to be mainly from Niger-Congo speaking groups, we analysed a set of haplogroups (E-sY81) that are collectively in high frequency in Niger-Congo speaking peoples but are at only low frequencies or absent in other groups. This should, at least to some extent, have the added benefit of reducing the effect of any recent contribution from Amerindian and European males. We repeated the

366 analysis including haplogroups within Y(xP,K,R) (see Figure 2), to which the NRY
367 haplotypes of the great majority of residents in sub-Saharan Africa belong.
368 Notably, in the PCA visualization, the Yombe are the closest to the present day
369 Palenque out of all the 42 sub-Saharan groups (see Fig. 4). In addition, Yombe was the only
370 group from the Republic of Congo for which there was not a significant F_{ST} value (Yombe
371 P=0.378, other seven groups $P < 0.001$). We also calculated F_{ST} distances after including ten
372 West African groups from Montano et al. [52] with E-sY81 chromosome set equal to or
373 above the minimum set in this study (Sundi, N(E-sY81)=22). Yombe remained the closest
374 group to Palenque. Analysing the Y(xP,K,R) set of haplotypes produced a similar outcome
375 but with the Chewa marginally closer to the Palenque than were the Yombe (both $F_{ST} <$
376 0.02). Interestingly, in both the E-sY81 and Y(xP,K,R) analyses, Yombe and Chewa had an
377 $F_{ST} <0.001$.
378 Even though there is considerable genetic similarity among the many widely
379 distributed groups having an origin in the rapid EBSP [19], the small genetic distance
380 between Chewa (a group from Malawi) and Palenque is so similar to that between the Yombe
381 and Palenque that it would be surprising, were it not for an oral history of the Chewa that
382 records an origin in the "Luba country of the southern Congo basin" [55]. This description
383 could place their origin only about 400 miles east of the region where Yombe is currently
384 spoken and may even reflect a migration from a more western location, passing through Luba
385 country rather than commencing within it. The date of this migration is uncertain with the
386 earliest record of the group as 'Chévas' only appearing in 1831-2 [55]. Marwick [55] also
387 records that the Chewa have an equally prevalent alternative origin story that places their
388 genesis south west of Lake Malawi. Marwick agrees with Hamilton [56] that the two
389 traditions can be reconciled by the migration from the north being by "chiefly invaders" who
390 gained control over "long-established autochthones". Our results are more consistent with
391 this interpretation of the oral accounts, as the Chewa-Yombe genetic distances were non-
392 significant at the NRY level but highly significant at the mtDNA level ($P < 0.001$).
393 Additional support for the Yombe origin of the Palenque comes from the absence of
394 NRY E-U181 chromosomes in both the Yombe and Palenque and their presence in the
395 Chewa. The presence of the E-U181 – previously reported as characteristic of East African
396 populations – in the Chewa can be explained by post-migration male gene flow following
397 their arrival in Malawi [56]. Nevertheless, given that a prior hypothesis exists – based on
398 linguistic evidence – for a Yombe origin, and that no such evidence has yet been advanced to

support a Chewa origin, it is reasonable to conclude that the genetic analysis of NRY haplotypes supports a Yombe origin.

Is there a significant difference between the sex-specific genetic systems profiles of residents of Abajo and Arriba?

Differences in the paternal demographic histories of the two areas of the village are clearly supported by the presence of different modal haplotypes, a slightly higher haplotype diversity in Arriba compared with Abajo, and a significant ETPD between the two. The summed frequencies of P-92R7 haplotypes and the distribution of STR haplotypes within this haplogroup were, however, similar suggesting a similar extent of non-African genetic introgression into the two districts. These results, in general, contrasted with comparisons using mtDNA where no statistical differences in diversity were observed and ETPD was not significant ($P = 0.985$). The similarity in mtDNA profiles but not NRY supports field observations of YM that patrilocality is practiced in Palenque with men choosing to live close to their fathers and grandfathers, and women marrying men either from their own district or another, with the latter being common.

Are the NRY and mtDNA profiles of the Arriba inhabitants compared with Abajo residents more similar to those of residents of Republic of the Congo?

Since no significant difference was observed between Arriba and Abajo residents in mtDNA haplotype distribution, the answer to the question posed is "no". With respect to paternal ancestry alone, the proposition has some limited support from the results of NRY when restricting analysis to E-sY81 component haplogroups. Here, where F_{ST} was significant, six out of seven of Congo groups had a smaller F_{ST} with the Arriba. More striking is that when compared with all 42 sub-Saharan groups, Arriba had lower genetic distances ($p=0.0001$). Although genetic drift cannot be discounted as the cause, one possible explanation is that practices associated with Africa such as matrilinearity (involving inheritance from a maternal uncle to his nephew, as seen in the Congo) were retained longer in Arriba than Abajo. There might therefore be an association between cultural practice and patterns of genetic diversity but not necessarily a causative relationship.

This study has explored an important aspect of the genetic ancestry of a fugitive African slave community in Colombia and contributed to a fuller understanding of their history. Further analysis of DNA of the Palenque alongside that of Colombian, European and sub-Saharan African groups using genome-wide markers and a more detailed characterisation of NRY and mtDNA should reveal more of the genetic history of the Palenque including contributions made by other communities in Colombia.

Ethics

All samples were collected anonymously with informed consent. This study received ethical approval from the Ministry of Health of the Republic of Congo (741/MSP/DGS/S), the Scientific Committee of the Academic Corporation for the Studies of Tropical Pathologies of the Universidad de Antioquia in Colombia (CPT-8840-03-054), the village council of Palenque de San Basilio and the Joint UCL/UCLH Committees on the Ethics of Human Research Committee A (99/0196).

Data accessibility

Supplementary Tables S1–S14 have been uploaded as part of the electronic supplementary material.

Competing interests

The authors declare no conflict of interest.

Author contributions

N.A., N.B. and Y.M. conceived and designed the study, C.D., N.G., G.B. and N.B. collected DNA samples. N.A. and M.G.T genotyped and sequenced the samples, Y.M. analysed the anthropological data, N.A., M.G.T. and N.B. analysed the genetic data, and N.A., N.B. and M.G.T wrote the paper.

Acknowledgments

We thank all DNA donors and those assisting in sample collection, and David Balding for advice on statistical analysis. N.A was supported by NERC CASE award.

References

[1] Wade P. 1995 The cultural politics of Blackness in Colombia. *Am Ethnol* **22**, 341-357.
[2] Friedemann N,Patiño Rosselli C. 1983 *Lengua y sociedad en El Palenque de San Basilio*. Bogotá, Instituto Caro y Cuervo.
[3] Moñino Y. 2012 Pasado, presente y futuro de la lengua de Palenque. In *Palenque (Colombia): Oralidad, identidad y resistencia* (eds. G. Maglia & A. Schwegler), pp. 221-256. Bogotá, Editorial Pontificia Universidad Javeriana.
[4] Bickerton D,Escalante A. 1970 Palenquero: A Spanish-based creole of northern Colombia. *Lingua* **24**, 254-267.
[5] Del Castillo N. 1984 El lexico negro-africano de San Basilio de Palenque. *Thesaurus* **39**, 80-169.
[6] Schwegler A. 2006 Palenquero. In *Concise encyclopedia of languages of the world* (eds. K. Brown & S. Oligvie). Oxford, UK, Elsevier Ltd.
[7] Granda G. 1971 Sobre la procedencia africana del habla "criolla" de San Basilio de Palenque. *Thesaurus* **26**, 84-94.
[8] Schwegler A. 2000 The African vocabulary of Palenque (Colombia). Part 1: Introduction and corpus of previously undocumented Afro-Palenquerisms. *J Pidgin Creole Lang* **15**, 241-312.
[9] Schwegler A. 2002 El vocabulario africano de Palenque (Colombia). Segunda Parte: compendio de palabras (con etimologías). In *Palenque, Cartagena y Afro-Caribe: historia y lengua* (eds. Y. Moñino & A. Schwegler). Tubingen, Max Niemeyer.
[10] Schwegler A. 2011 Palenque(ro): the search for its African substrate. In *Creoles, their substrates, and language typology* (ed. C. Lefebvre), pp. 225-249. Amsterdam, John Benjamins.
[11] Lewis MP. 2009 *Ethnologue: Languages of the World.* Dallas, Texas, SIL International.
[12] Moñino Y. 2007 Les rôles du substrat dans les créoles et dans les langues secrètes: le cas du palenquero, créole espagnol de Colombie. In *Grammaires créoles et grammaire comparative* (eds. K. Gadelii & A. Zribi-Hertz), pp. 49-72. Saint-Denis, Vincennes University Press.
[13] Helgason A, Sigureth ardottir S, Gulcher JR, Ward R, Stefansson K. 2000 mtDNA and the origin of the Icelanders: deciphering signals of recent population history. *Am J Hum Genet* **66**, 999-1016. (doi:S0002-9297(07)64026-9 [pii]).
[14] Hill EW, Jobling MA, Bradley DG. 2000 Y-chromosome variation and Irish origins. *Nature* **404**, 351-352. (doi:10.1038/35006158).
[15] Thomas MG, Skorecki K, Ben-Ami H, Parfitt T, Bradman N, Goldstein DB. 1998 Origins of Old Testament priests. *Nature* **394**, 138-140. (doi:10.1038/28083).
[16] Thomas MG, Parfitt T, Weiss DA, Skorecki K, Wilson JF, le Roux M, Bradman N, Goldstein DB. 2000 Y chromosomes traveling south: the cohen modal haplotype and the origins of the Lemba--the "Black Jews of Southern Africa". *Am J Hum Genet* **66**, 674-686. (doi:10.1086/302749).

[17] Thomas MG *et al.* 2002 Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors. *Am J Hum Genet* **70**, 1411-1420. (doi:10.1086/340609).
[18] Veeramah KR, Zeitlyn D, Fanso VG, Mendell NR, Connell BA, Weale ME, Bradman N, Thomas MG. 2008 Sex-Specific Genetic Data Support One of Two Alternative Versions of the Foundation of the Ruling Dynasty of the Nso' in Cameroon. *Curr Anthropol* **49**, 707-714. (doi:10.1086/590119).
[19] Ansari Pour N, Plaster CA, Bradman N. 2013 Evidence from Y-chromosome analysis for a late exclusively eastern expansion of the Bantu-speaking people. *Eur J Hum Genet* **21**, 423-429. (doi:10.1038/ejhg.2012.176).
[20] Brehm A, Pereira L, Bandelt HJ, Prata MJ, Amorim A. 2002 Mitochondrial portrait of the Cabo Verde archipelago: the Senegambian outpost of Atlantic slave trade. *Ann Hum Genet* **66**, 49-60. (doi:doi:10.1017/S0003480001001002).
[21] Goncalves R, Rosa A, Freitas A, Fernandes A, Kivisild T, Villems R, Brehm A. 2003 Y-chromosome lineages in Cabo Verde Islands witness the diverse geographic origin of its first male settlers. *Hum Genet* **113**, 467-472. (doi:10.1007/s00439-003-1007-4).
[22] Trovoada MJ, Pereira L, Gusmao L, Abade A, Amorim A, Prata MJ. 2004 Pattern of mtDNA variation in three populations from Sao Tome e Principe. *Ann Hum Genet* **68**, 40-54.
[23] Goncalves R, Spinola H, Brehm A. 2007 Y-chromosome lineages in Sao Tome e Principe islands: evidence of European influence. *Am J Hum Biol* **19**, 422-428. (doi:10.1002/ajhb.20604).
[24] Jimenez S, Martinez B, Hernandez M. 1996 HLA antigens and gene distribution in San Basilio (SB) an isolated black population of Colombia [Abstract]. *Hum Immunol* **47**, 62.
[25] Arnaiz-Villena A, Reguera R, Parga-Lozano C. 2009 HLA Genes in Afro-American Colombians (San Basilio de Palenque): The First Free Africans in America. *Open Immunol J* **2**, 59-66.
[26] Noguera MC *et al.* 2013 Colombia's racial crucible: Y chromosome evidence from six admixed communities in the Department of Bolivar. *Ann Hum Biol*. (doi:10.3109/03014460.2013.852244).
[27] Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL. 1994 High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**, 455-457. (doi:10.1038/368455a0).
[28] Yu N *et al.* 2002 Larger genetic differences within africans than between Africans and Eurasians. *Genetics* **161**, 269-274.
[29] Rosa A,Brehem A. 2011 African human mtDNA phylogeography at-a-glance. *J Anthropol Sci* **89**, 25-58. (doi:10.4436/jass.89006).
[30] Sims LM, Garvey D, Ballantyne J. 2007 Sub-populations within the major European and African derived haplogroups R1b3 and E3a are differentiated by previously phylogenetically undefined Y-SNPs. *Hum Mutat* **28**, 97. (doi:10.1002/humu.9469).
[31] Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, Hammer MF. 2008 New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* **18**, 830-838. (doi:10.1101/gr.7172008).
[32] de Filippo C *et al.* 2011 Y-chromosomal variation in sub-Saharan Africa: insights into the history of Niger-Congo groups. *Mol Biol Evol* **28**, 1255-1269. (doi:10.1093/molbev/msq312).
[33] Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N. 1999 Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* **23**, 147. (doi:10.1038/13779).

[34] Salas A, Richards M, Lareu MV, Scozzari R, Coppa A, Torroni A, Macaulay V, Carracedo A. 2004 The African diaspora: mitochondrial DNA and the Atlantic slave trade. *Am J Hum Genet* **74**, 454-465. (doi:10.1086/382194).
[35] Veeramah KR *et al.* 2010 Little genetic differentiation as assessed by uniparental markers in the presence of substantial language variation in peoples of the Cross River region of Nigeria. *BMC Evol Biol* **10**, 92. (doi:10.1186/1471-2148-10-92).
[36] Raymond M,Rousset F. 1995 An Exact Test for Population Differentiation. *Evolution* **49**, 1280-1283.
[37] Nei M. 1987 *Molecular Evolutionary Genetics*. New York, Columbia University Press.
[38] Reynolds J, Weir BS, Cockerham CC. 1983 Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* **105**, 767-779.
[39] Slatkin M. 1995 A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457-462.
[40] Kimura M. 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111-120.
[41] Excoffier L, Laval G, Schneider S. 2005 Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* **1**, 47-50.
[42] Team RDC. 2011 *R: a language and environment for statistical computing*. Vienna, Austria, R Foundation for Statistical Computing.
[43] Weale ME *et al.* 2003 Rare deep-rooting Y chromosome lineages in humans: lessons for phylogeography. *Genetics* **165**, 229-234.
[44] Underhill PA *et al.* 2000 Y chromosome sequence variation and the history of human populations. *Nat Genet* **26**, 358-361. (doi:10.1038/81685).
[45] Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL, Underhill PA. 2002 Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am J Hum Genet* **70**, 265-268. (doi:10.1086/338306).
[46] Knight A, Underhill PA, Mortensen HM, Zhivotovsky LA, Lin AA, Henn BM, Louis D, Ruhlen M, Mountain JL. 2003 African Y chromosome and mtDNA divergence provides insight into the history of click languages. *Curr Biol* **13**, 464-473. (doi:S0960982203001301 [pii]).
[47] Wells RS *et al.* 2001 The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci U S A* **98**, 10244-10249. (doi:10.1073/pnas.171305098).
[48] Moore LT, McEvoy B, Cape E, Simms K, Bradley DG. 2006 A Y-chromosome signature of hegemony in Gaelic Ireland. *Am J Hum Genet* **78**, 334-338. (doi:10.1086/500055).
[49] Wilson JF, Weiss DA, Richards M, Thomas MG, Bradman N, Goldstein DB. 2001 Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc Natl Acad Sci U S A* **98**, 5078-5083. (doi:10.1073/pnas.071036898).
[50] Di Giacomo F *et al.* 2004 Y chromosomal haplogroup J as a signature of the post-neolithic colonization of Europe. *Hum Genet* **115**, 357-371. (doi:10.1007/s00439-004-1168-9).
[51] Berniell-Lee G *et al.* 2009 Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. *Mol Biol Evol* **26**, 1581-1589. (doi:10.1093/molbev/msp069).
[52] Montano V, Ferri G, Marcari V, Batini C, Anyaele O, Destro-Bisol G, Comas D. 2011 The Bantu expansion revisited: a new analysis of Y chromosome variation in Central Western Africa. *Mol Ecol* **20**, 2693-2708. (doi:10.1111/j.1365-294X.2011.05130.x).

[53] Scheinfeldt LB, Soi S, Tishkoff SA. 2010 Working toward a synthesis of archaeological, linguistic, and genetic data for inferring African population history. *Proc Natl Acad Sci U S A* **107**, 8931-8938.
[54] Jobling MA,Tyler-Smith C. 2003 The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* **4**, 598-612.
[55] Marwick MG. 1963 History and tradition in East Central Africa through the eyes of the Northern Rhodesian Cewa. *J Afr Hist* **4**, 375-390.
[56] Hamilton RA. 1955 Oral Tradition: Central Africa. In *History and Archaeology in Africa* (ed. R.A. Hamilton). London, London: School of Oriental and African Studies.

Figure Legends

Figure 1. Geographic location of the village of Palenque de San Basilio in Colombia.

Picture adapted from Google maps (www.maps.google.com).

Figure 2. Phylogenetic relationships of UEP markers used to define NRY haplogroups.

The box identifies the E-sY81 (E1b1a) clade, exclusively observed in population groups with recent African ancestry.

Figure 3. NRY haplogroup profiles in Chewa, Palenque and Yombe.

Note that all haplogroups present in Yombe are observed in Palenque. Haplogroups unobserved in a group are shown as blank. For definition of abbreviations of population names see Table 1.

Figure 4. Visual representation of genetic relationships among all groups using PCA based on NRY UEP within E-sY81 (E1b1a).

SE represents population groups in South-East Africa, namely CH, YA, TU, SE and BN. Percentages in parentheses are the amount of variation explained by each component. For definition of abbreviations of population names see Table 1.