Received: 27 July 2017 | Revised: 17 November 2017 | Accepted: 9 December 2017

DOI: 10.1002/ajpa.23378

RESEARCH ARTICLE

WILEY American Journal of PHYSICAL ANTHROPOLOGY

Matriclans shape populations: Insights from the Angolan Namib Desert into the maternal genetic history of southern Africa

Sandra Oliveira[1,2] | Anne-Maria Fehn[1,3,4] | Teresa Aço[5] | Fernanda Lages[6] | Magdalena Gayà-Vidal[1] | Brigitte Pakendorf[7] | Mark Stoneking[8] | Jorge Rocha[1,2,6]

[1]CIBIO/InBIO: Research Center in Biodiversity and Genetic Resources, Vairão 4485-661, Portugal

[2]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Porto 4169-007, Portugal

[3]Department of Linguistic and Cultural Evolution, Max Planck Institute for the Science of Human History, Jena 07745, Germany

[4]Institute for African Studies, Goethe University, Frankfurt 60323, Germany

[5]Centro de Estudos do Deserto (CEDO), Namibe, Angola

[6]ISCED/Huíla—Instituto Superior de Ciências da Educação, Lubango, Angola

[7]Laboratoire Dynamique du Langage, UMR5596, CNRS & Université Lyon, Lyon 69007, France

[8]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany

Correspondence
Jorge Rocha, CIBIO/InBIO, Campus de Vairão, Rua Padre Armando Quintas n 7, 4485-661 Vairão, Portugal.
Email: jrocha@cibio.up.pt

Funding information
FEDER funds through the Operational Programme for Competitiveness Factors—COMPETE, by National Funds through FCT—Foundation for Science and Technology under the PTDC/BIA-EVF/ 2907/2012 and FCOMP-01-0124-FEDER- 028341, and by the Max Planck Society. SO was supported by the FCT grant SFRH/BD/85776/2012, and MG-V by FP7-REGPOT-2011-1-project 286431, UID/BIA/50027/2013 and POCI-01-0145-FEDER-006821. BP acknowledges the LABEX ASLAN (ANR-10-LABX-0081) of Université de Lyon for its financial support within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) of the French government operated by the National Research Agency (ANR).

Abstract

Objectives: Southern Angola is a poorly studied region, inhabited by populations that have been associated with different migratory movements into southern Africa. Apart from Kx'a-speaking San foragers and Bantu-speaking pastoralists, ethnographic and linguistic studies have suggested the existence of an enigmatic array of pre-Bantu communities, like the Kwepe (formerly Khoe-Kwadi speakers), Twa and Kwisi. Here, we evaluate previous peopling hypotheses by assessing the relationships between different southern Angolan populations, based on newly collected linguistic data and complete mtDNA genomes.

Materials and methods: We analyzed 295 complete mtDNA genomes and linguistic data from seven groups from the Namib Desert (Himba, Kuvale, Tjimba, Twa, Kwisi, Kwepe) and Kunene Province (!Xun), placing special emphasis on the evaluation of the genealogical consistency of the matriclanic system that characterizes most of these groups.

Results: We found that the maternal genetic structure of all groups from the Namib Desert was strongly shaped by the consistency of their matriclanic system. The tracking of the maternal heritage enhanced population differentiation by genetic drift and is likely to have caused the divergent mtDNA profiles of the Kwepe, Twa, and Kwisi, who probably formed a single population within the spectrum of Bantu genetic variation. Model-based analyses further suggest that the dominant pastoral groups Kuvale and Himba may be grouped into a Bantu proto-population which also included the ancestors of present-day Tjimba and Herero, as well as the Khoe-Kwadi speaking Damara foragers from Namibia.

Discussion: The view from southwestern Angola offers a new perspective on the populating history of southern Africa and the Bantu expansions by showing that social stratification and different subsistence patterns are not always indicative of remnant groups, but may reflect Bantu-internal variation and ethnogenesis.

KEYWORDS
Bantu, Khoe-Kwadi, matriclans, pastoralism, southern Africa

1 | INTRODUCTION

The high ethnic diversity of southwestern Angola, the importance of its pastoral culture, and the likely confluence of different migratory waves in its peopling provide a unique opportunity to explore the significance of different hypotheses about the population history of southern Africa. At present, it is generally accepted that the oldest population stratum in this vast region is represented by groups speaking languages that make extensive use of click consonants, which were previously lumped into a hypothetical "Khoisan" phylum (Greenberg, 1963), but are now divided into three families: Kx'a, Tuu and Khoe-Kwadi (Güldemann & Fehn, 2014). Although Tuu and Kx'a-speaking peoples were historically hunter-gatherers, Khoe-Kwadi languages are spoken by both foraging and food-producing groups, with the Khoekhoe-speaking Nama representing one of the major pastoralist populations of southern Africa. Based on typological observations, it has been speculated that the Khoe-Kwadi languages might constitute a later arrival in the area, possibly linked to a migration of Later Stone Age pastoralists from East Africa, who moved into regions previously inhabited by Kx'a and Tuu-speaking hunter–gatherers (Barnard, 1992; Güldemann, 2008; Westphal, 1963). Although the presence in southern Africa of lactase persistence and Y-chromosome haplotypes that originated in eastern African pastoralists seems to support this hypothesis (Breton et al., 2014; Coelho, Sequeira, Luiselli, Beleza, & Rocha, 2009; Henn et al., 2008; Macholdt et al., 2014), it is unclear whether these traits were dispersed by a massive immigration of Khoe-Kwadi speakers or were introduced through small scale movements leading to the diffusion of livestock and genetic variants across neighboring resident populations (Sadr, 2015).

More recently, about 1,500 years ago, the human population landscape of southern Africa was further modified by the arrival of Bantu-speaking groups with subsistence economies that presently range from almost exclusive pastoralism to mixed farming systems (Russell, Silva, & Steele, 2014). Although the emergence of new combinations of genes, languages and modes of subsistence is an expected outcome of the confluence of different population strata, the prevailing views about the peopling of southern Africa favor the idea that the technological advantages and social dominance of the Bantu considerably restricted the direction and range of genetic and cultural exchange (Cashdan, 1986). Consequently, a strong connection between foraging, low social status, the "Khoisan" languages and phenotypes including small stature and light skin was established, leaving anthropologists puzzled with foraging peoples physically more similar to other non-"Khoisan" African populations (Barnard, 1992; Cashdan, 1986). In this context, the origin of the dark-skinned foragers speaking Khoe-Kwadi languages, such as the Khwe from the Okavango region, or the Damara from Namibia, is often considered enigmatic and has been linked to a hypothetical stratum of pre-Bantu non-"Khoisan" peoples (Barnard, 1992; Blench, 2006; Cashdan, 1986). Intriguingly, the possibility of a historical link to the Bantu has only rarely been considered (Cashdan, 1986; Westphal, 1963).

Located at the southwestern edge of the Bantu expansion and at the northwestern fringe of an area traditionally inhabited by Kx'a-speaking hunter–gatherers, the Angolan Namib Desert forms a contact zone that mirrors the high variability currently observed in the wider region of southern Africa (Figure 1). The dominant populations are the Himba and Kuvale, two matrilineal pastoralist populations commonly considered to be part of the broad Herero ethno-linguistic division that arrived in the area during the Bantu expansions, but whose relationships to one another and to other southwestern African Bantu speakers are not clear (Barbieri et al., 2014b; Coelho et al., 2009; Gibson, 1977; Westphal, 1963).

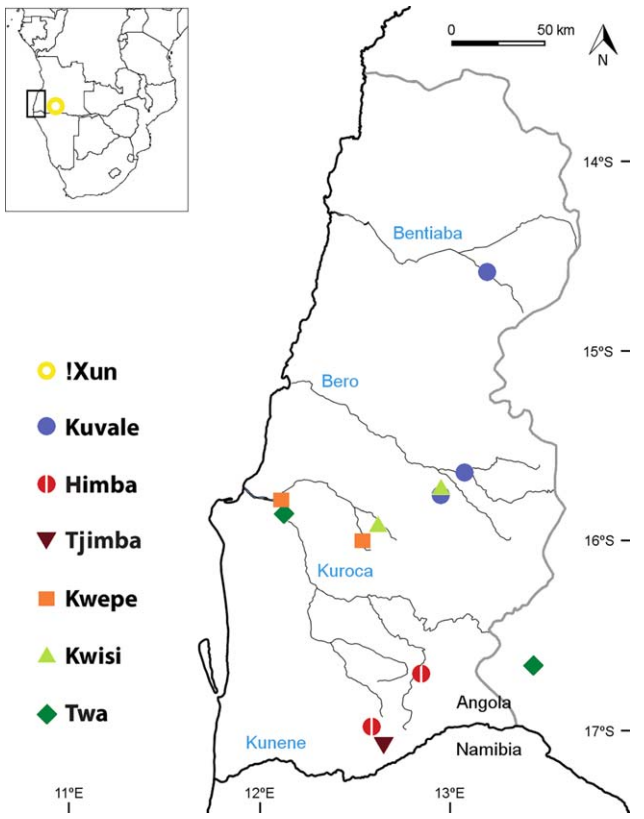

FIGURE 1 Map of sampling locations. Each location is colored by the corresponding population. The Angolan Namib province is delimited by a gray contour, country borders are show in black and the names of the main intermittent rivers are indicated in blue. The area covered by the map is highlighted in the inset

In the orbit of these two groups gravitate several small-scale communities, including the Tjimba, the Kwepe, the Kwisi, and the Twa, who share a matriclanic social organization with their Bantu neighbors, but whose origins remain unknown. Due to their patron–client relationship with the Himba and Kuvale, they are perhaps best described as peripatetic peoples (Bollig, 2004), a category that encompasses small-scale, low-status, endogamous communities that are primarily non-food producing and provide specialized goods and services (e.g., as blacksmiths, healers, sorcerers) to their dominant neighbors. However, it is uncertain whether the peripatetic peoples from the Angolan Namib represent distinct populations, each with a separate demographic history, or are simply outcast groups that are genetically related to, but socially marginalized by the pastoral Bantu.

The early references to human biological variation in the area noted the lack of apparent physical distinctiveness between

peripatetic and Bantu populations, but these observations did not rely on objective, statistically validated anthropometrical measurements (Almeida, 1965; Estermann, 1976). Therefore, previous hypotheses about the origins of the peoples of the Angolan Namib were mostly formulated in the context of ethnographical and linguistic work favoring the idea that the peripatetic communities might be associated with very different migratory movements (Almeida, 1965; Estermann, 1976).

The Kwisi and Twa, who speak the Bantu language Kuvale, are the most marginalized communities of the Angolan Namib. Since they describe themselves as the native peoples of the region and are locally perceived as having a distinct ethnic history from their neighbors, they were suggested to represent two related branches of a group of remnant pre-Bantu hunter-gatherers to which the Damara were also ascribed (Almeida, 1965; Blench, 2006; Cashdan, 1986; Estermann, 1976). This link to an old hunter–gatherer tradition was additionally thought to be supported by the Twa ethnonym, which supposedly goes back to a proto-Bantu word used to designate resident foragers from different parts of Africa (Estermann, 1962; Vansina, 1990). In this context, the present use of the Kuvale language by the Kwisi and Twa was attributed to a hypothetical (and undocumented) shift from an unknown language that would have been lost after contact with the Bantu (Estermann, 1976; Westphal, 1963), similar to what has been claimed for the Pygmies of West and Central Africa (Bahuchet, 2012).

The Kwepe are small stock breeders who until recently spoke Kwadi, a language that has been replaced by Kuvale and is now virtually extinct (Almeida, 1965; Westphal, 1963). Their distinct linguistic heritage and the accepted link between Kwadi and the Khoe language family led to the proposal that they might represent a remnant group from the hypothetical Khoe-Kwadi migration introducing pastoralism to southern Africa (Güldemann, 2008). Finally, the Tjimba are often considered Himba who lost their cattle but retained their language and other aspects of their culture (Warmelo, 1951). Still, it has been suggested that some isolated Tjimba communities from Namibia with a documented use of stone tools might be connected to a more ancient hunter-gatherer tradition (MacCalman & Grobbelaar, 1965).

All of these hypotheses entail a set of testable expectations about the genetic, linguistic and cultural relationships of the peoples living in the Angolan Namib. Specifically, from a genetic perspective it is expected that: (a) the Kuvale and the Himba are related to each other as well as to other Herero-speaking peoples of southern Africa; (b) the Twa and the Kwisi are genetically similar to each other, but clearly distinct from their Bantu neighbors; (c) the Kwepe share genetic similarities with Khoe-speaking peoples from other regions of southern Africa; and (d) the Tjimba are either closely related to the Himba or have a very distinct genetic composition presumably related to the Twa and Kwisi. In this scenario, it is also expected that the matrilineal descent-group systems of the peripatetic peoples are relatively recent and were borrowed from their Himba and Kuvale neighbors, considering that matrilineality is known to be a distinctive feature of Bantu societies in southwestern Africa (Bollig, 2006; Estermann, 1952; Gibson, 1956).

To date, the remote geographical location and high-mobility status of the peripatetic peoples of the Angolan Namib have made it difficult to evaluate these predictions. Recently, in the course of field research being conducted in the area, we have located and contacted several communities belonging to the Twa, Kwisi, Kwepe, and Tjimba ethnic groups who live in close proximity to the Kuvale and Himba populations. Here, we report for the first time a multidisciplinary assessment of the relationships between these populations based on newly collected linguistic data and 295 complete mtDNA genomes. Our results suggest that the maternal genetic structure of the different ethnic groups dwelling in the Namib Desert is largely derived from endogamous Bantu peoples and was strongly shaped by their matriclanic social organization, with contributions of non-Bantu populations being mostly restricted to "Khoisan" lineages. In this context, we propose that the social stratification and different subsistence patterns found in the area are not indicative of remnant groups, but reflect Bantu-internal variation and ethnogenesis.

2 | MATERIALS AND METHODS

2.1 | Samples

We analyzed 295 whole mitochondrial genomes from six populations living in the Namib Desert (77 Himba; 85 Kuvale; 37 Kwepe; 24 Kwisi; 18 Twa; 15 Tjimba) and from 39 Kx'a-speaking !Xun hunter–gatherers from the Kunene Province (Figure 1; Supporting Information Table S1). At all sampling locations, the purpose of the study was explained with the aid of bilingual native speakers. For each participant, we collected a saliva sample and information about language, matriclan and place of birth, up to the grandparental generation. With the exception of the !Xun, who do not have a clanic system, all sampled individuals identified as members of one out of 13 distinctive matriclans. Additional genealogical information, including relatedness with other donors, was also recorded. Given the intrinsic social structure of these highly endogamous groups, we only avoided including siblings and mother–offspring pairs in the final dataset (see Pinto et al., 2016 for details). The linguistic analyses were based on lexical data collected from individuals belonging to each sampled group, including two elder community members of the Kwepe community, who still remember Kwadi (see Pinto et al., 2016). As previously described in Pinto et al. (2016), the saliva samples, as well as the linguistic and the personal information, were collected with the donors' written informed consent in the framework of a collaboration between the Portuguese-Angolan TwinLab established between CIBIO/InBIO and ISCED/Huíla Angola, with the ethical clearance of ISCED and the CIBIO/InBIO-University of Porto boards, and the support and permission of the Provincial Governments of Namibe and Kunene.

2.2 | MtDNA sequencing

Multiplexed sequencing libraries were produced from genomic DNA and enriched for mtDNA sequences following Meyer and Kircher (2010) and Maricic, Whitten, and Pääbo (2010) with some

modifications as detailed in Barbieri et al. (2012). The sequencing was performed on the Illumina Miseq platform with paired-end runs of 214 or 314 cycles. Base calling was performed with Bustard, adapters trimmed with leeHom (Renaud, Stenzel, & Kelso, 2014) and reads demultiplexed using deML (Renaud, Stenzel, Maricic, Wiebe, & Kelso, 2015). The reads were aligned against the human reference genome 19 using a customized version of BWA v0.5.10-evan (https://bitbucket.org/ustenzel/network-aware-bwa; Li & Durbin, 2009). Reads that aligned to the mitochondrial genome and known nuclear insertions of mitochondrial DNA (numts) (Li, Schroeder, Ko, & Stoneking, 2012) were re-aligned to the mtDNA revised Cambridge Reference Sequence (Andrews et al., 1999) using BowTie2 (Langmead & Salzberg, 2012), and the consensus sequences were called using an in-house script for detecting mtDNA heteroplasmies (Li & Stoneking, 2012). The resulting mitochondrial genomes have a mean coverage of 400×. Missing nucleotides were replaced with the nucleotide that was present in all otherwise identical haplotypes of the dataset. With this imputation approach the missing data of the whole dataset (1,057 missing nucleotides distributed across 10 samples) was reduced to three missing sites in a single sample. The Haplogrep webtool and Phylotree Build 16 were used to assign the haplogroup of each sample (Kloss-Brandstätter et al., 2011; Oven & Kayser, 2008). Sequence alignments were performed with MUSCLE v.3.8 (Edgar, 2004). The two poly-C regions (np 303–315, 16,183–16,194) were removed in all further analyses.

Sequences are available from GenBank with accession numbers MF381287–MF381581.

2.3 | Genetic data analysis

Analyses of molecular variance (AMOVA), pairwise Φst values and genetic diversity indices were computed in Arlequin v3.5.2.2 (Excoffier & Lischer, 2010). Non-metric multidimensional scaling (MDS) and *k*-means analyses based on pairwise Φst distance matrices were carried out in R, using the functions "isoMDS" from the package MASS (Venables & Ripley, 2002) and "kmeans" with several random starts (Hartigan & Wong, 1979), respectively. An additional matrix describing the relationships between populations solely on the basis of matriclan frequencies was generated in Arlequin v3.5.2.2 using a FST-like distance and coding each clan as a sequence of sites that differed from all the other clans in a single site. The correlation between genetic and clanic distances was assessed by performing a Mantel test with 1,000 permutations of matrix elements to determine significance.

Neighbor-Joining trees were generated using the R function "nj" from the package "ape" (Paradis, Claude, & Strimmer, 2004). Bootstrap analyses were performed with 100 replicates, using the R function "boot.phylo" from the package "ape" (Efron, Halloran, & Holmes, 1996; Felsenstein, 1985). After each resampling step, the function "stat.phist" from the package "strataG" v0.9.2 was used to calculate a Φst distance matrix, and a new Neighbor-Joining tree was generated. We report the percentage of bipartitions that match the original neighbor-joining tree.

Median-joining networks (Bandelt, Forster, & Rohl, 1999) were computed with Network 5.0 (www.fluxus-engineering.com) and customized in Network Publisher v2.1.1.2. The time to the most recent common ancestor (TMRCA) of sub-haplogroups was estimated with Network from the *rho* statistic (Forster, Harding, Torroni, & Bandelt, 1996), using a mutation rate of 1.665×10^{-8} substitutions per nucleotide per year (Soares et al., 2009), and a generation time of 28 years (Fenner 2005). The root defining the ancestral haplotype in each sub-haplogroup was identified by using the full mtDNA network. In addition, alternative TMRCA estimates were obtained with BEAST v1.8 (Drummond, Suchard, Xie, & Rambaut, 2012), using the same mutation rate, and a piecewise-linear *Coalescent: Bayesian Skyline* tree model (Drummond, Rambaut, Shapiro, & Pybus, 2005). For each sub-haplogroup, we ran a chain of 10 million steps with parameters logged every 1000 steps. The first one million steps were discarded as burn-in.

For population-based and sequence-based comparisons, we compiled a dataset comprising approximately 2,500 previously-published whole mitochondrial genomes from different regions of Africa (Supporting Information Table S2).

We additionally used an Approximate Bayesian Computation (ABC) approach (Beaumont, Zhang, & Balding, 2002) to test three alternative isolation-with-migration models for the relationships between the Herero, Himba, and Damara and the neighboring Kuvale, Nyaneka-Nkhumbi and !Xun populations. After model choice, ABC was again used to estimate the parameters of the most favored scenario. For each model, two million datasets of complete mtDNA genomes were simulated with wide uniform prior distributions of demographic parameters (population sizes, asymmetric migration rates and population splitting times). Due to the high complexity of the tested models, involving as many as 29 parameters (see Results), we opted to use a fixed mutation rate of 1.665×10^{-8} substitutions/nucleotide/year (equal to that used in BEAST analysis), which has been carefully estimated using a comprehensive phylogeny of complete mtDNA genomes (Soares et al., 2009). Although this simplification may be too restrictive, studies on pseudo-observed genome-wide data (Li & Jakobsson, 2012) suggest that the impact of using fixed mutation rates on ABC procedures is most relevant for deviations from true values that are larger than the current variation in available estimates of the overall mtDNA mutation rate (Fu et al., 2013, 2014; Posth et al., 2016; Rieux et al., 2014; Soares et al., 2009). Simulations were performed with fastsimcoal v2.5.2.1.1 (Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013) and summary statistics computed with Arlequin v3.5.2.2 (Excoffier & Lischer, 2010), both within the framework of ABCtoolbox (Wegmann, Leuenberger, Neuenschwander, & Excoffier, 2010). The summary statistics used for comparing the observed and simulated data were the number of haplotypes (k), sequence diversity (H), number of segregating sites (S), number of private segregating sites (prS), Tajima's D (D) and mean number of pairwise differences (MPD), all computed within populations. In addition, population pairwise Φst and pairwise MPD was computed between pairs of populations. All summary statistics were standardized.

The ABC estimations were performed with a general linear model regression adjustment (Leuenberger & Wegmann, 2010; Wegmann et al., 2010) applied to the 10,000 retained simulations (0.5%) closest to the observed data. Model selection was based on posterior probabilities (PPs) estimated using the marginal density of each model relative

to the density of all models. The power to correctly select a given model was assessed by using 1,000 pseudo-observed datasets taken from that model and calculating the number of times it had the highest PP when compared with alternative models (Veeramah et al., 2012). To reduce the effects of including summary statistics that are redundant or do not capture the main features of the data, we additionally performed model selection using a subset of summary statistics that were only moderately correlated (Pearson's $r^2 < 0.8$) and exhibited the highest power to discriminate between models, as proposed by de Filippo et al. (2016) (Supporting Information Table S3).

To estimate parameters from the most supported model, we transformed summary statistics from simulated and observed data into partial least squares (PLS) using the R scripts provided in ABCtoolbox (Wegmann & Excoffier, 2010; Wegmann, Leuenberger, & Excoffier, 2009). By using root mean square error (RMSE) plots we found that an optimal set of five PLS components provided the largest amount of information about the model parameters (Wegmann et al., 2009). The ABC estimation was then performed as described above. To assess the agreement between the observed data and the retained data generated under the chosen model, we checked the *p*-value provided by ABCtoolbox, which corresponds to the proportion of retained simulations with a smaller or equal likelihood than the observed data. In addition, we plotted density distributions of pairs of PLS components from the retained simulations together with transformed observed statistics to visually inspect how well the simulated data explored the space surrounding the observed data.

To check for bias in the posterior distributions of individual parameters, we randomly selected 1,000 pseudo-observed datasets and assessed the uniformity of the posterior quantiles of the true parameter values using a Kolmogorov-Smirnov test with Bonferroni correction (Wegmann & Excoffier, 2010; Wegmann et al., 2009). To determine the power of parameter estimation, we computed the coefficient of variation R^2 by regressing the PLS components against each model parameter (Neuenschwander et al., 2008). To evaluate the accuracy of the mode as a point estimate, we calculated the $RMSE_{mode}$ for each parameter based on 1,000 pseudo-observed datasets (Wegmann & Excoffier, 2010).

2.4 | Linguistic data analysis

We collected data from Kuvale and Himba, the two major Bantu languages of southwestern Angola, as well as comparative samples from Kwadi (as remembered by two Kwepe elders) and the !Xun variety of the Kunene Province. The linguistic data analyzed in this work are based on a 600-item wordlist that is a subset of the Summer Institute of Linguistics Comparative African Wordlist (Snider & Roberts, 2006). Data were elicited through interviews with mother-tongue speakers who had grown up in their respective communities. Portuguese was used to communicate with the speakers, either directly or through a translator. All elicitation sessions were recorded on tape and subsequently transcribed with the help of a local assistant fluent in both Kuvale and Himba. We obtained at least one 600-item wordlist from each ethnic community, which was then cross-checked and supplemented by an additional 200-item list of core vocabulary (Supporting Information Table S4). For comparative purposes, we added material from Namibian Herero (Möhlig & Kavari, 2008), and several varieties belonging to the Nyaneka-Nkhumbi cluster (Humbe, Muhila, Nyaneka, Ngambwe, Handa; Jordan, 2015; unpublished data from Jordan & Manuel). We then compiled a summary wordlist for each ethnic group, allowing for synonyms but excluding obvious borrowings from Portuguese. In a subsequent step, we excluded meanings displaying a high amount of missing data, as well as those with a high degree of variation suggesting uncertainty by the speakers. Based on these considerations, we drew up a final list of 273 meanings, which include the Swadesh 200 (Swadesh, 1952) and Leipzig-Jakarta wordlists (Haspelmath & Tadmore, 2009), minus function words, personal pronouns, and question words. We then performed an analysis of regular sound correspondences in our dataset of southwestern Bantu languages that led to the establishment of 693 cognate sets (available upon request), which were used in all subsequent analyses.

For computational purposes, we coded languages for presence (1) or absence (0) of a particular lexical root. As our data from Himba and Tjimba displayed a high degree of linguistic homogeneity, they were combined and treated under the label "Himba". Based on our coded dataset, we generated a matrix of linguistic distances (one minus the percentage of cognate sharing) and computed a Neighbor-Joining tree with 1,000 bootstrap replicates using SplitsTree v4.14.2 (Huson & Bryant, 2006). Linguistic distances were compared with genetic distances with a Mantel test, as described earlier.

We further used a Bayesian phylogenetic approach as implemented in the BEAST2 framework (Bouckaert et al., 2014) and tested three models included in the Babel package (Bouckaert, 2016): (a) Continuous Time Markov Chain (CTMC); (cf. Greenhill & Gray, 2009); (b) Covarion (Atkinson, Meade, Venditti, Greenhill, & Pagel, 2008; Penny, McComish, Charleston, & Hendy, 2001); (c) Dollo (Nicholls & Gray, 2006). We ran an analysis for each model, with a chain length of 10,000,000, sampling every 1,000 steps. The first 100,000 steps were discarded as burn-in.

To evaluate the performance of these models with our dataset, we used the Tracer software (Rambaut & Drummond, 2007) to compare the Akaike Information Criteria through Markov chain Monte Carlo (AICM) of each analysis, where lower AICM values indicate better model fit (Baele, Li, Drummond, Suchard, & Lemey, 2013). We found that the model displaying the best fit for our data was Covarion (AICM = 7,116), outranking both CTMC (AICM = 7,146), and Dollo (AICM = 7,682).

The output of the analysis was visualized in DensiTree (Bouckaert, 2010) in order to display reticulations and conflicting signals.

3 | RESULTS

3.1 | Genetic and matriclanic diversity in the Angolan Namib

By performing an AMOVA, we found that 25.2% of the total genetic variation in our sample is due to differences between populations. This

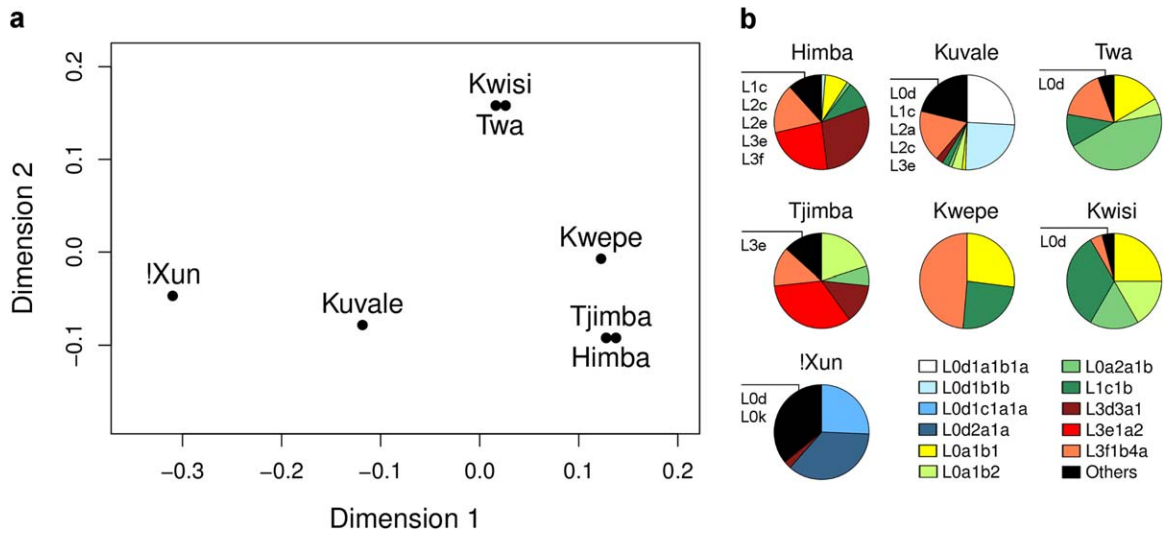

FIGURE 2 MDS analysis and haplogroup variation in southwestern Angola. (a) MDS plot based on Φst genetic distances. The pairs Kwisi-Twa and Tjimba-Himba are not significantly different, with *p*-values 0.11 and 0.16, respectively. Stress value: 0.006. (b) Frequencies of the most common sub-haplogroups (≥20% in at least one population) are shown for each population. The remaining sub-haplogroups are pooled under the category "Others" (black), with the major haplogroup assignments within this category listed for each population. Note that major haplogroups that are represented in the plots by a specific sub-haplogroup might appear again in the category "Others" to indicate other low frequency sub-haplogroups

level of genetic differentiation is 20.2% even when the !Xun are removed and is higher than previously observed among Bantu (5.5%; Barbieri et al., 2014b) and "Khoisan" populations (16.6%; Barbieri et al., 2014a). The levels of intra-population diversity are highest in the Kuvale and Himba (mean value of haplotype diversity, 0.95) and lowest in the Kwepe (0.67), who display only five different haplotypes (Supporting Information Table S1).

A non-metric MDS plot based on pairwise Φst distances reveals three main vertices of divergence (Figure 2a): (a) the !Xun from Kunene Province, who have high frequencies (97%; Supporting Information Table S5) of haplogroups L0d and L0k that typically predominate in most "Khoisan" populations from southern Africa (Barbieri et al., 2014a); (b) the Tjimba and Himba, whose close genetic relationship supports the view that the two groups are merely distinguished by their socio-economic status (Vashro & Cashdan, 2015; Warmelo, 1951); (c) the Kwisi and Twa, whose genetic proximity is consistent with previous claims that these communities represent northern and southern branches of the same ethnic group respectively (Estermann, 1976).

The differences in the mtDNA composition of the Namib peoples are mainly due to the uneven distribution of nine common sub-haplogroups that collectively account for 90% of their observed variation, each with a small number of haplotypes rarely exhibiting more than five pairwise differences (Figures 2b and 3; Supporting Information Table S5): L0a1b1, L0a1b2, L0a2a1b, and L1c1b are very common in the Kwisi and Twa; L3e1a2 and L3d3a1a predominate in the Himba and Tjimba, while L0d1a1b1a and L0d1b1b are very frequent in the Kuvale, placing them closer to the !Xun (Figure 2a). With the exception of the Kwisi, L3f1b4a is found at relatively high frequencies in most groups.

An assessment of lineage sharing among different populations shows that the most common sub-haplogroups among the Himba/Tjimba and Kuvale (Supporting Information Figure S1) are rarely found in other groups, except for one single L3f1b4a haplotype that is very frequent in the Kwepe but is likely to have originated in the Himba, who display a higher L3f1b4a diversity (Supporting Information Figure S1). Conversely, haplotypes belonging to sub-haplogroups that are frequent and diverse in the Kwisi, the Twa or the Kwepe can be found at moderate frequencies in the Himba and Kuvale (Supporting Information Figure S1), suggesting that gene flow occurs preferentially from these peripatetic communities into the dominant groups.

The nine most common sub-haplogroups are associated with all 13 matriclans identified during our survey, with the number of clans in each sub-haplogroup varying from one to five (Figure 3). The occurrence of several clans in the same sub-haplogroup has several potential explanations, including adoption, patrilineal transmission, or chance. However, this pattern can also be explained by a well-documented Herero custom of splitting the same line of descent into different clans, forming clan-groups with a claimed common ancestor designated as phratries (Gibson, 1956; Vivelo, 1977). Interestingly, we found that three pairs of clans that were reported to us as sharing a distant ancestor were also associated with the same sub-haplogroup: Mukwalukune/Mukwatjiti (L0d1a1b1a); Mukwanambula/Mukwangombe (L0d1b1b and also L0a2a1b) and Mukwandjata/Mukwambua (L3f1b4a).

Although most clans are distributed across multiple populations (Figure 4a), we found several cases where the same clan is associated with different sub-haplogroups in different populations (Supporting Information Figure S2b, e, g, h, and k), suggesting that clan sharing is not always due to migration. All these cases involve at least one common sub-haplogroup from a dominant population (Kuvale or Himba),

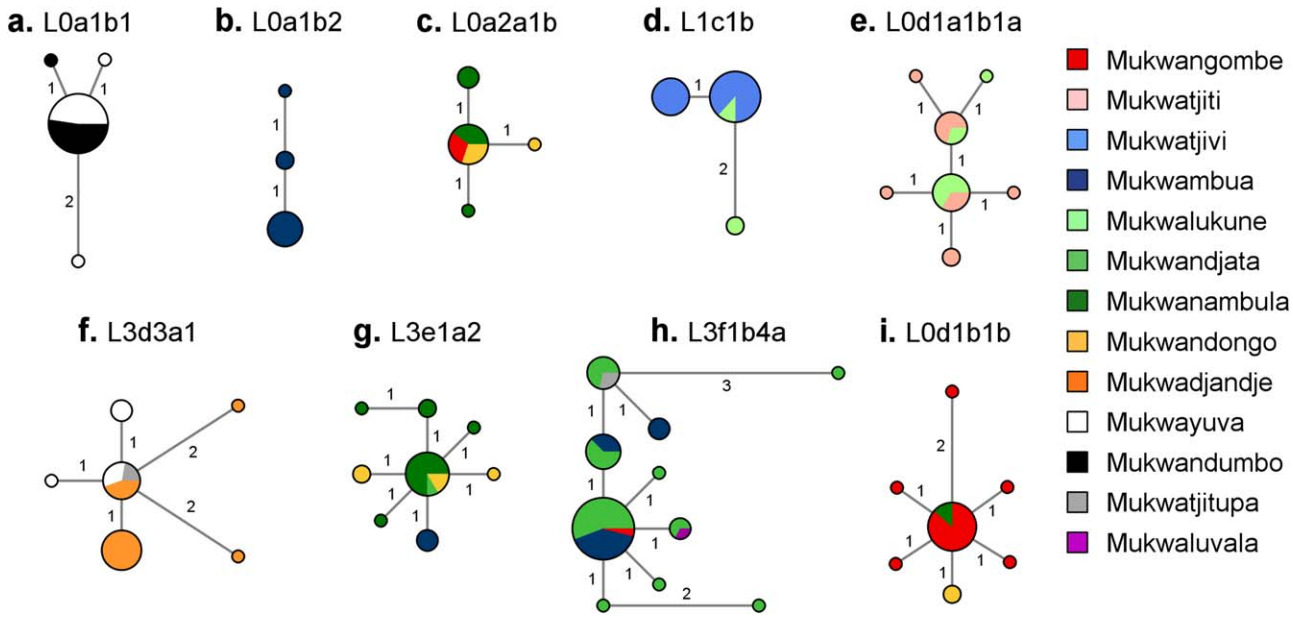

FIGURE 3 Median-joining networks showing haplotype variation within the most common sub-haplogroups of the Angolan Namib. Circles represent mtDNA haplotypes, with size proportional to frequency and color corresponding to clan affiliation. The number of mutational steps is indicated for each branch. Indels were not included

and one common sub-haplogroup from the Twa, Kwisi, and Kwepe peripatetics.

The absence of a one-to-one correspondence between matriclans and sub-haplogroups decreases the association between the distribution of matriclans and the genetic differentiation among populations (Figure 4b): in some cases, sub-haplogroups that are associated with the same matriclan predominate in populations that are genetically very divergent; in others, sub-haplogroups that are shared across genetically similar populations are associated with different matriclans. Consequently, distance matrices between

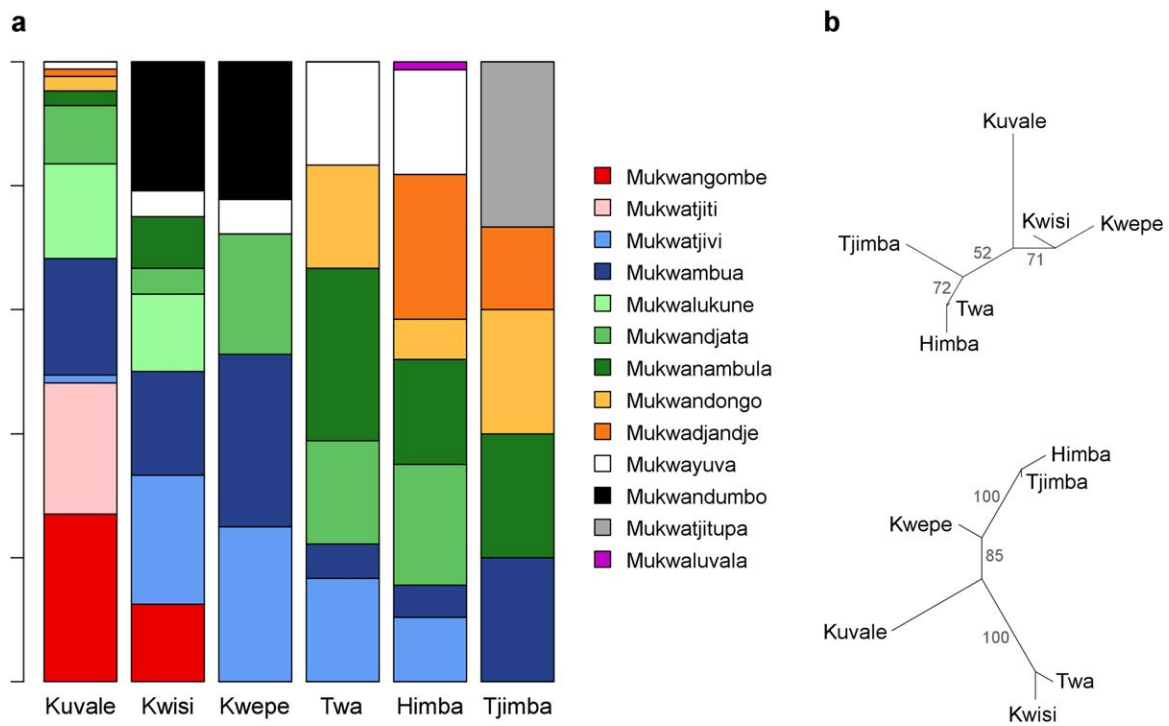

FIGURE 4 Relationship between genetic and clanic distances in populations of the Angolan Namib. (a) Matriclan distribution within each population. (b) Neighbor-joining tree based on clan distances (top) and Φst genetic distances (bottom). Bootstrap values (%) are shown in gray

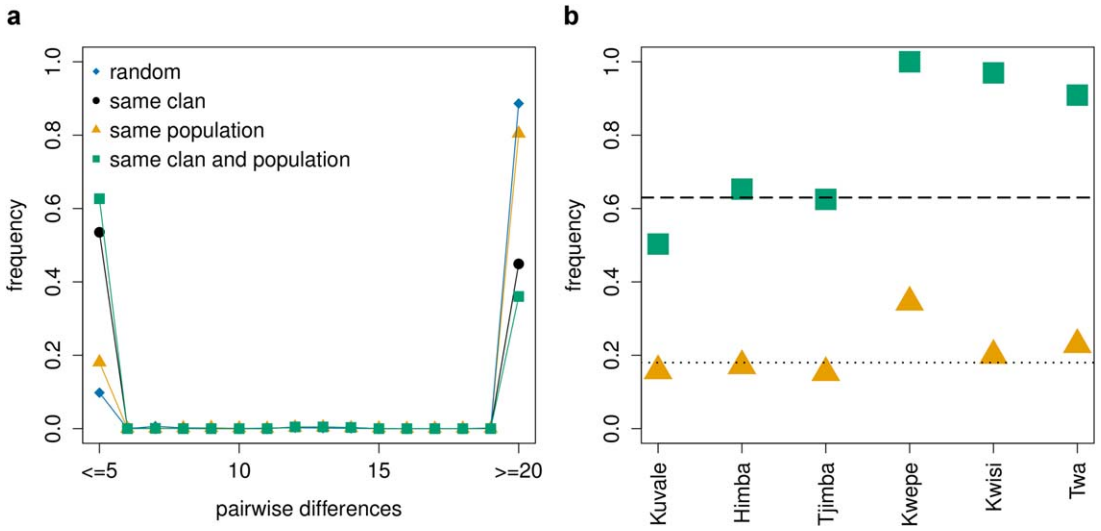

FIGURE 5 Genealogical consistency of matriclans. (a) Distribution of pairwise differences obtained by randomly drawing pairs of sequences from: (i) the whole Angolan Namib pool, (ii) the same population, (iii) the same clan, and (iv) the same clan and population. (b) Sequence similarity in Angolan Namib populations computed for pairs of sequences randomly drawn from each population (orange triangles) and from individuals belonging to the same clan in each population (green squares). The dotted and dashed lines show the average sequence similarity computed within populations, regardless of the clan, and within clans, respectively. Sequence similarity was measured by the frequency of sequence pairs with ≤ 5 differences

populations based on matriclans and mtDNA are clearly uncorrelated (Mantel test $p = 0.57$; Figure 4b).

In spite of these exceptions, we found that as much as 51% of the total mtDNA variation reflects differences between matriclans, a highly significant value (AMOVA; $p < 0.00001$) that is more than two times greater than the 20.2% proportion calculated among ethnic groups, indicating that there are remarkable differences in the mtDNA sequence profiles of individual matriclans (Supporting Information Table S6).

Moreover, as shown in Figure 5a, the distributions of pairwise differences clearly indicate that mtDNA sequences drawn from the same clan have a significantly higher average probability of being closely related ($\leq$5 pairwise differences) than two sequences randomly sampled from the whole Namib pool (0.53 vs. 0.10; $p < 0.001$, Fisher exact test), or from the same population (0.53 vs. 0.18; $p < 0.001$), indicating that individuals from the same clan are more likely to share a sub-haplogroup. This association becomes even stronger when mtDNA sequences are sampled from the same clan and the same population (0.53 vs. 0.63; $p < 0.001$).

The probability of sampling related sequences within clans is significantly elevated in all populations (Figure 5b; $p < 0.001$ in all comparisons), and is remarkably high in the Kwepe, Twa, and Kwisi, who display greater levels of within-clan sequence similarity than the Tjimba, Himba, and Kuvale.

To rule out the possibility that close kin relationships could have inflated the likelihood that individuals from the same clan have exactly the same haplotype, we restricted the analysis to closely related but non-identical haplotypes. We randomized one million times the matriclan labels on observed matriclan/haplotype pairs and then calculated the probability of finding within the same matriclan two haplotypes with one to five differences (Supporting Information Figure S3). As shown in Supporting Information Table S7, in most groups the observed value for this probability is too high to be obtained by chance, indicating that similar (but non-identical) haplotypes have a high probability of sharing clans by inheritance. The only non-significant values were found among the Kwepe and the Tjimba, whose low levels of within haplogroup diversity reduce the power of the test (Figure 2b; Supporting Information Figure S1). Note that by using this approach we made the conservative assumption that all individuals within the same matriclan/haplotype pair share a common ancestor, which drastically reduces the number of independent matriclan assignments that are needed to randomly match the observed data (Supporting Information Figure S3).

Supporting Information Table S8 presents the estimates of the TMRCAs of the nine predominant sub-haplogroups obtained with the *rho* statistic (Forster et al., 1996) and with BEAST (Drummond et al., 2012). Due to the association between clans and sub-haplogroups exhibited by most populations, these TMRCAs can be used as proxies for the coalescent ages of the oldest clans in each clan-group. However, it is not possible to provide separate estimates for matriclans associated with the same sub-haplogroup, since these clans often share the TMRCA of the whole sub-haplogroup and represent different samples from the same genealogy (Figure 3). The TMRCA estimates calculated with *rho* and BEAST range from ~560 to ~3,140 years (average ~1,800 years) and from 1,080 to 3,390 years (average ~2,025 years), respectively, with large uncertainty. Although BEAST estimates tend to be higher than *rho*, the values calculated by the two methods are highly correlated (r^2=0.93).

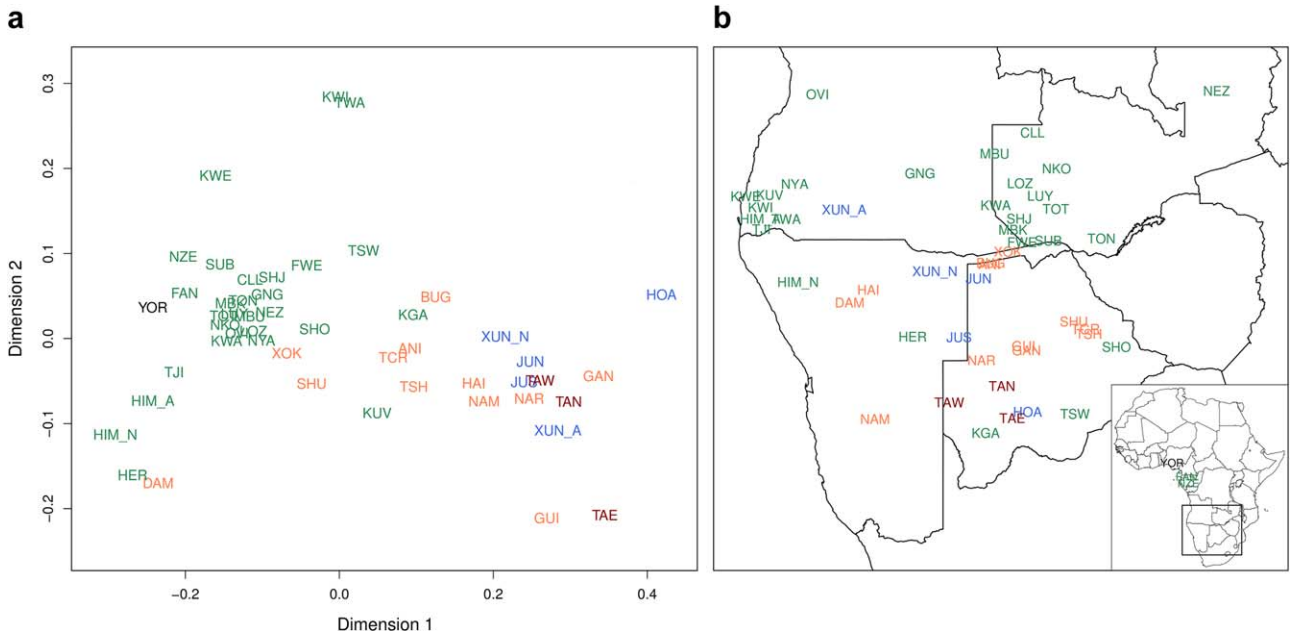

FIGURE 6 MDS analysis in the wider region of southern Africa. Colors correspond to language families: Niger-Congo non-Bantu (black), Niger-Congo Bantu (green), Kx'a (blue), Tuu (dark red), Khoe-Kwadi (orange). The code used for each population can be found in Supporting Information Table S2. (a) MDS plot based on Φst genetic distances. Stress value: 9.4. (b) Geographic origin of the populations included in the MDS analysis

3.2 | Relationships with other populations

When the genetic profiles of the populations from Namib are compared with an extended mitochondrial genome-dataset including other groups from Angola (Nyaneka-Nkhumbi, Ovimbundu, Ganguela) and the wider region of southern Africa (Figure 6), the Kwisi and the Twa remain outliers, while the Tjimba and Himba fall close to the Herero, Himba and Damara from Namibia (see also Barbieri et al., 2014a, 2014b; Soodyall & Jenkins, 1993). The Kuvale, in contrast, are more similar to other populations with high levels of maternal Bantu-"Khoisan" admixture, including the Tshwa, Shua, TcireTcire, and ||Ani. The Kwepe are not close to any Khoe-speaking group, even though they spoke the related Kwadi language until recently (Almeida, 1965; Pinto et al., 2016). The !Xun from the Angolan Kunene Province are related to Kx'a- and Tuu-speaking groups from Namibia and Botswana.

These patterns are confirmed and complemented by the clustering results obtained with the k-means algorithm (Supporting Information Figure S4). With the exception of the Kuvale, all the populations from Namib are initially lumped into a cluster encompassing most Bantu-speaking peoples ($k = 2$ in red). Further partitions: (a) isolate a homogeneous group of Bantu-speaking populations that forms a central core in the MDS plot ($k = 4$ in green); (b) separate the Twa and Kwisi from the other clusters ($k = 6$ in yellow); and (c) group the Angolan Himba with the Herero, Himba and Damara from Namibia ($k = 7$ in orange). An outstanding feature of the k-means partitions is the wide dispersal across different clusters of the Khoe-Kwadi-speaking populations represented in our dataset. Some groups from the Central Kalahari (G|ui, G||ana, and Naro) and Namibia (Nama and Hai||om) cluster together with Kx'a- and Tuu-speaking "Khoisan" peoples ($k = 2$–7). Groups from the eastern Kalahari (Tshwa, TcireTcire) and Okavango (||Ani and Buga) form a cluster with high levels of maternal Bantu/"Khoisan" admixture together with the Bantu-speaking Kuvale, Tswana and Kgalagadi ($k = 3$–7). Finally, the Damara, the ||Xokhoe and the Kwepe (formerly speaking Kwadi), in spite of their high levels of genetic differentiation, are grouped together with Bantu-speaking populations that have low amounts of "Khoisan" admixture ($k = 2$–7).

The phylogeographic analysis of the mtDNA lineages from the Namib populations provides additional information about their relationships with groups from adjacent areas (Supporting Information Figure S5). Sub-haplogroups L1c1b and L0a1b2, are remarkable for their molecular divergence and geographical confinement to southwestern Angola (Figure S5c and i). Other major sub-haplogroups have molecularly close neighbors in several Bantu-speaking populations from southern Africa (L0a2a1b and L3f1b4a; Supporting Information Figure S5d and l) or are related to sequences that are mostly shared by Bantu and Khoe-Kwadi groups from the area (L0a1b1, L3d3a1, and L3e1a2; Supporting Information Figure S5b, j, and k). None of the L0d lineages common in the Kuvale (L0d1a1b1a and L0d1b1b) were found in the !Xun from Angola, the nearest "Khoisan" group from the Namib Desert. Instead, their L0d1a1b1a haplotypes, also observed in the Himba from Namibia, are close to lineages that were found in the Khoe-Kwadi-speaking Shua from Botswana, while the L0d1b1b haplotypes are remotely related to sequences observed in the Damara from Namibia and the Luyana from Zambia (Supporting Information Figure S5e and f). The most common sub-haplogroups in the !Xun (L0d1c1a1a, 26% and L0d2a1a, 36%) have unique haplotype matches with !Xun and Ju|'hoan from northern Namibia and display sequences that are closely related to "Khoisan" groups from southern Africa (Supporting Information Figure S5g and h). Taken together, these results indicate that, with two

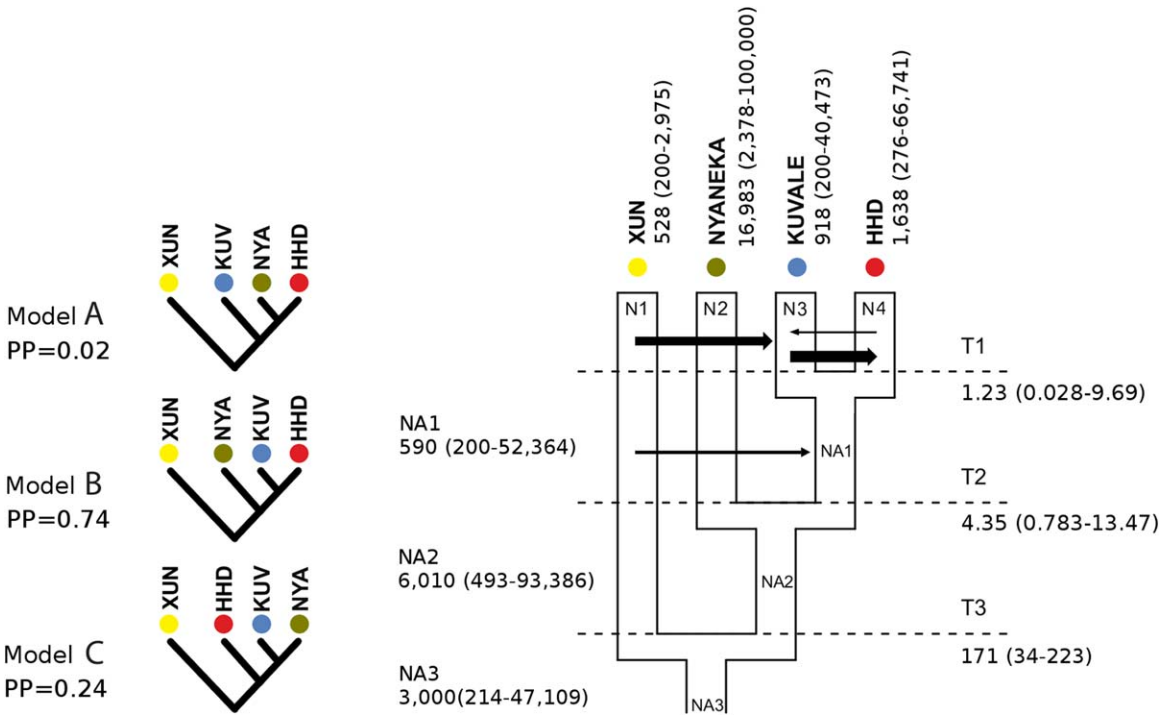

FIGURE 7 Demographic models tested by ABC. The three tested models are shown on the left with their respective PPs. Migration ratios above 0.0001 or effective migration (Nm) above 2 are represented in the plot by arrows with width proportional to Nm. NA1–NA3: ancestral effective population sizes; N1–N4: current effective population sizes; T1–T3: divergence times in kya

exceptions (L1c1b and L0a1b2), most sequences from southwestern Angola are nested in the phylogeographic pattern that emerged from the contact of previously identified population strata from southern Africa.

3.3 | Testing relationships of Kuvale and Herero/Himba/Damara

As previously noted (Barbieri et al., 2014b), the close genetic proximity of the Himba and Herero pastoralists to the Damara, who speak the same Khoe language as the Nama and have a peripatetic lifestyle, stands in stark contrast to their genetic distinctiveness from the linguistically and culturally similar Kuvale. Based on resampling tests, Barbieri et al. (2014b) suggested that the sharing of a common ancestry by the Herero, Himba, and Kuvale was not compatible with a scenario of shared ancestry between the Herero, Himba and Damara. Here, we address the relationships of the closely related Herero, Himba and Damara with other groups. We lumped the Herero, Himba and Damara (all clustered by *k*-means at $k = 7$; Supporting Information Figure S4) into a single metapopulation (HHD) and tested three evolutionary scenarios relating this metapopulation with the Kuvale and two neighboring populations (Nyaneka-Nkhumbi and !Xun), using an ABC approach (Beaumont et al., 2002). The !Xun-speaking "Khoisan" from Angola were always used as an outgroup and we assumed that their split predated all other events (Figure 7). The Nyaneka-Nkhumbi provide a southwestern Bantu-speaking reference population located to the northeast of the Namib Desert (Figure 6b). In the first scenario, an early divergence of the Kuvale is followed by a more recent split between the Nyaneka-Nkhumbi and the HHD metapopulation (Figure 7, Model A). The second scenario postulates a recent common origin of the Kuvale and HHD (Figure 7, Model B). The third scenario assumes that the most recent common origin is between the Kuvale and the Nyaneka-Nkhumbi (Figure 7, Model C). For the time of the oldest divergence event (T3; Supporting Information Table S9), we used a flat prior ranging from 14 to 224 kya that contains and extends beyond the range of estimates reported for the split between Khoisan and other groups (Gronau, Hubisz, Gulko, Danko, & Siepel, 2011; Schlebusch et al., 2012; Veeramah et al., 2012). Irrespective of the model, the prior ranges of the two subsequent divergent events (T1 and T2) were allowed to reach 14 kya (the minimum value of T3), which predates the beginning of the Bantu expansion 4–5 kya by a large margin (Rocha and Fehn, 2016) (Supporting Information Table S9). Since the prior ranges of T1 and T2 overlap, the prior of T1 is not flat. To obtain a flat prior, we followed the approach of Veeramah et al. (2012) and additionally sampled a scale factor from a uniform prior 0–1 (T1_sc) that expresses the time of the most recent population split as a fraction of the time of the second divergence event (Supporting Information Table S9). Asymmetric migration was allowed between all pairs of populations, and migration rates were sampled from wide log-uniform priors between 1×10^{-10} and 0.2 representing residual and very high migration, respectively (Supporting Information Table S9). The power to predict the correct model was 0.47, 0.48, and 0.44, in simulated models A, B, and C, respectively. These values are significantly different from the expected 0.33 if there was no discriminatory power ($p < 0.001$, binomial test).

Model B, assuming a recent common origin of the Kuvale and HHD, was the most supported scenario, with a PP of 0.74 (Figure 7,

Model B). By iteratively excluding summary statistics that were highly correlated (Pearson's $r^2 > 0.8$), starting with those which had less power to discriminate between models (de Filippo et al., 2016), we found that model B was still the most supported model. The model also yielded a relatively high *p*-value (0.4) suggesting a good fit between the simulated and the observed data. The PLS components of the observed summary statistics were clearly within the density distribution of the PLS components of the retained simulations, further indicating that the data simulated under Model B adequately surround the observed data (Supporting Information Figure S6).

We additionally used the ABC framework to estimate the demographic parameters of the best supported scenario based on 2 million simulations (Figure 7, Model B). The posterior distributions of all parameters are shown in Supporting Information Figure S7 and the corresponding point estimates are listed in Table S9. The point estimate for the time of split of the !Xun (T3; ~170 kya; 95% CI: 34–223 kya) is consistent with previous calculations of the divergence time of "Khoisan" peoples from other sub-Saharan African populations (Behar et al., 2008; Schlebusch et al., 2012; Veeramah et al., 2012), but has a wide credible interval and the corresponding posterior distribution does not depart substantially from the prior (Supporting Information Figure S7). A similar problem arises with the estimate for the age of divergence of the Nyaneka-Nkhumbi (T2; ~4.35 kya; 95% CI: 0.783–13.47 kya), which falls within the time frame of the Bantu expansions but predates the 1–2 kya date for the divergence of Angolan southwestern Bantu-speaking groups that can be inferred from the linguistic and archeological data presented by Grollemund et al. (2015; see their Figure 1). The posterior distribution for the time of split between the Kuvale and HHD (T1) is more clearly peaked around 1.231 kya; (95% CI: 0.028–9.692 kya; Supporting Information Figure S7), but this parameter does not have a uniform posterior quantile distribution (Supporting Information Table S9) and is likely to be biased upwardly, as suggested by the concentration of true values in the left half of the posterior distribution (Supporting Information Figure S8). In contrast, the scaled Kuvale-HHD split time (T1_sc) is unbiased (Supporting Information Table S9 and Supporting Information Figure S8) and well peaked around 0.15 (Supporting Information Figure S7), in spite of a large credible interval (95% CI: 0–0.9), suggesting that the Kuvale and the HHD metagroup may have shared a substantial part of their demographic history after splitting from other southwestern Bantu populations.

Our estimates of Ne show that the Nyaneka-Nkhumbi have the largest effective population size (~17,000; 95% CI: 2,378–100,000), followed by the HHD, the Kuvale and the !Xun, with estimates of ~1,600 (95% CI: 276–66,741), ~900 (95% CI: 200–40,073) and ~500, respectively (95% CI: 200–2,975) (Figure 7; Supporting Information Table S9). The point estimates of ancestral effective population sizes (NA1 and NA2) suggest that the Nyaneka-Nkhumbi experienced a ~3-fold growth after their split (Ne A2 = 6,000 to Ne Nyaneka-Nkhumbi = 17,000), while the size of the ancestors of the Kuvale and HHD underwent a ~10-fold reduction (Ne A2 = 6,000 to Ne A1 = 600) (Figure 7; Supporting Information Table S9).

Many migration estimates have flat posterior distributions, suggesting that our data cannot infer adequately all patterns of gene flow among the studied populations (Supporting Information Figure S7). Moreover, the migration rates M(A1>!Xun), M(!Xun > HHD), and M (HHD>!Xun) were biased, exhibiting distributions of posterior quantiles that significantly deviate from uniformity (Supporting Information Table S9 and Supporting Information Figure S8). However, we obtained reasonably peaked, high migration estimates from the !Xun into the Kuvale ($m = 0.021$; 95% CI: 5.6×10^{-10}–0.20; Nm = 18.9 migrants/generation) that are in accordance with the high number of L0d haplotypes found in the Kuvale (Figures 2 and 7; Supporting Information Figure S8 and Supporting Information Table S9). Still, this result should be interpreted with caution since most L0d lineages that were found in the Kuvale belong to two sub-haplogroups that are likely to be derived from only two ancestral women (Supporting Information Table S5 and Supporting Information Figure S1), and probably were not transferred by the continuous gene flow process simulated in our ABC analysis. Finally we additionally estimated high migration rates from the !Xun into the common ancestor of the Kuvale and HHD ($m = 0.010$; 95% CI: 4.9×10^{-10}–0.19; Nm = 5.7), from the Kuvale into HHD ($m = 0.015$; 95% CI: 5.2×10^{-10}–0.20; Nm = 24.5), and to a lesser extent from the HHD into the Kuvale ($m = 0.005$; 95% CI: 3.1×10^{-10}–0.12; Nm = 4.2) (Figure 7; Supporting Information Figure S8 and Supporting Information Table S9).

3.4 | Linguistic analyses

The high amount of genetic divergence among the Namib peoples (Figure 2a) contrasts with the relative linguistic homogeneity of the area, where all groups presently speak either Himba or Kuvale. Although the classification of Himba as a variety of the Herero language is fairly straightforward and widely accepted, the position of Kuvale is less clear (Maho, 2009; Vansina, 2004; Westphal, 1963). Moreover, the Bantu languages spoken by the Kwisi and Twa have long been the subject of speculation (Westphal, 1963). To evaluate the relationships between the Himba and Kuvale languages that are currently spoken in the Angolan Namib, as well as their links to Namibian Herero and to Nyaneka-Nkhumbi southwestern Bantu varieties, we first undertook a lexicostatistical analysis and calculated a language distance matrix based on 693 cognate sets. In the Neighbor-Joining tree based on the language distance matrix, Kuvale forms its own cluster, separated from Himba and Herero on one side and various dialects of Nyaneka-Nkhumbi on the other (Figure 8a). Kuvale as spoken amongst the Kwisi and Kwepe is fully within the range of the cluster. The variety spoken by the Twa seems to have been influenced by Himba and lies in-between the Kuvale and the Himba/Herero clusters. Furthermore, based on a careful comparison of our Bantu wordlists with lexical data from Kwadi (Westphal, 1963, supplemented by our own field notes), Khoe (Vossen, 1997), and !Xun (König & Heine, 2008) we note that no linguistic variety spoken in Namib displays any lexical peculiarities that could be linked to influence from a non-Bantu substrate. As a result of the nesting of linguistic varieties spoken by the peripatetic Kwisi, Twa and Kwepe within the range of Kuvale and Herero/Himba, distance matrices based on linguistic and genetic distances between the Namib groups are uncorrelated (Mantel test, $p = 0.18$).

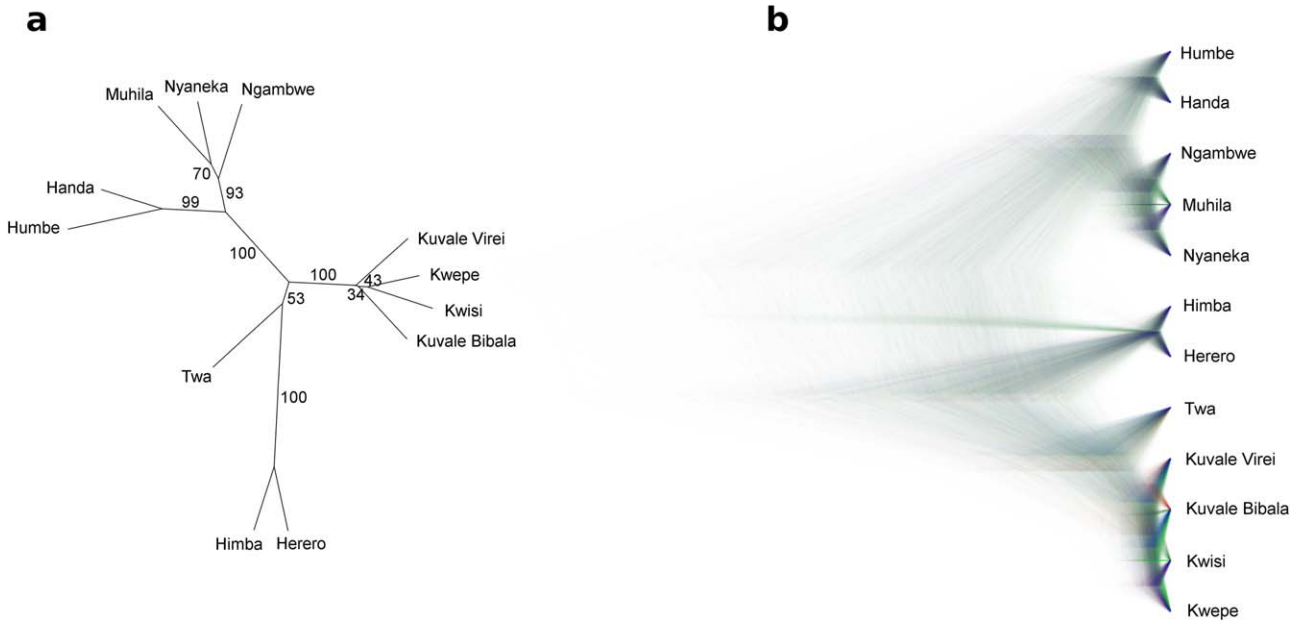

FIGURE 8 Linguistic relationships between Kuvale, Himba, Herero and Nyaneka-Nkhumbi. The Kuvale sample includes varieties spoken by the Kuvale people (Kuvale Virei and Kuvale Bibala), as well as the Kwepe, the Kwisi and the Twa. The Nyaneka-Nkhumbi sample includes varieties spoken by the Handa, Humbe, Ngambwe, Nyaneka and Muhila peoples. (a) Neighbor-joining tree. Numbers indicate bootstrap values (%). (b) Bayesian trees plotted with DensiTree

To gain a better understanding of the historical relations between Nyaneka-Nkhumbi, Herero/Himba and Kuvale, we additionally undertook a Bayesian phylogenetic analysis in BEAST, using the same 693 cognate sets underlying the Neighbor-Joining tree in Figure 8a. We used the software DensiTree to visualize the PP distribution and uncover possible uncertainties in the topology (Figure 8b). While we find some conflicting signal in the internal node heights, the general topology of the tree is well resolved. As in our Neighbor-joining analysis, all language clusters (Herero/Himba, Kuvale, and Nyaneka-Nkhumbi) are unequivocally identified ($p = 1.00$; Figure 8b; Supporting Information Figure S9). The analysis further suggests a more recent common ancestor for Herero and Kuvale ($p = 0.9$) than either language shares with the five varieties of Nyaneka-Nkhumbi we included in our analysis. There is, however, a residual conflicting signal with regards to the position of Herero, which might suggest that the three clusters form a star-like tree. Nevertheless, the result of the linguistic analysis is remarkably congruent with Model B of the ABC analysis, which suggests that Kuvale and Herero/Himba are more closely related than either population is to Nyaneka-Nkhumbi (Figures 7 and 8b). Within Kuvale, we find no well-supported sub-clusters, except for the initial split from Twa ($p = 1.00$), which is grouped with the other varieties but remains an outlier (Figure 8a; Supporting Information Figure S9).

4 | DISCUSSION

In recent years, a growing number of studies on the population history of southern Africa has considerably broadened our knowledge concerning the historical interactions of groups dwelling in and around the Kalahari Basin (Barbieri et al., 2014b; Marks et al., 2015; Pickrell et al., 2012; Schlebusch et al., 2012). Within this geographical area, the focus has largely been on "Khoisan"-speakers and the southeastern Bantu populations whose genetic and cultural make-up are thought to have been shaped by contact with indigenous foragers and herders. In the Southwest, new genetic data have recently become available for populations from Namibia and southern Africa (Montinaro et al., 2017; Uren et al., 2016), while the groups to their north remain the subject of intense speculation, but constitute a noticeable gap in the available literature. Our study presents for the first time full maternal genomes and linguistic data from Angolan populations previously deemed inaccessible or vanished (Almeida, 1965; Estermann, 1976), including Bantu-speaking groups, as well as the formerly Kwadi-speaking Kwepe. We sampled both foraging and pastoral populations, placing special emphasis on the analysis of the coherence of the matriclanic system that characterizes the area and unites populations of different social status and modes of subsistence. In this framework, we are now able to address different historical hypotheses about the present-day diversity found in the Namib Desert both from a local perspective and within the context of the wider region of southern Africa.

4.1 | Genealogical consistency of matriclans

A remarkable feature of the social organization of all the populations from the Angolan Namib and other southwestern Bantu peoples is their matrilineal descent-group system in which individuals are affiliated to the clan of their mother, and members of the same matriclan (sing. *eanda*) consider themselves as distant relatives that descend from an unknown founder woman (Bollig, 2006; Estermann, 1952; Gibson, 1956). Although some populations may have dual descent systems and additionally form patriclans, it is the matrilineal principle that regulates

key aspects of community life, such as cattle inheritance, social obligations, marriage preferences and group membership (Gibson, 1956). However, the consistency of southwestern African matriclans has been difficult to validate with genealogical data, since the relationships between members of the same clan are often considered to be too distant to be traced accurately (Gibson, 1956; Vivelo, 1977). Furthermore, it has been suggested that members of low-status peripatetic communities borrowed the matriclanic system from their dominant neighbors as a means to achieve better integration into the regional network of the southwestern Bantu societies (Bollig, 2004; Estermann, 1976).

In this study, we relied on the maternal inheritance of mtDNA to show for the first time that matriclans are indeed good descriptors of deep genealogical relationships in pastoral and peripatetic Bantu-speakers from southwestern Angola. Several interrelated lines of evidence support this conclusion: (a) a high proportion of the total mtDNA variation is found among matriclans ($\Phi st = 0.51$; $p < 0.00001$); (b) individuals from the same clan have a significantly increased probability of having related mtDNA haplotypes that are likely to belong to the same sub-haplogroup (Figures 3 and 5); iii) the average TMRCAs of major sub-haplogroups (~1,800 or ~2,025 years, Supporting Information Table S8) suggests that the oldest matriclans are not recent and probably date back to the arrival of Bantu-speaking peoples to southern Africa.

In spite of this evidence, we found that several matriclans likely became associated to more than one sub-haplogroup through multiple founders in different populations. Since these cases often involve a sub-haplogroup restricted to the Himba or Kuvale and a sub-haplogroup predominant in the Kwepe, Twa or Kwisi (Supporting Information Figure S2b, e, g, h, and k), it may be argued that these low-status peripatetic communities were clanless and recently borrowed the matriclanic system from their dominant neighbors, as proposed previously in Bollig (2004) and Estermann (1976). However, such an imitation scenario is difficult to reconcile with the antiquity and the genealogical consistency of the matriclanic system observed in all peripatetic populations (Supporting Information Table S8; Figure 5b). Furthermore, our permutation tests indicate that random assignment of clans, as would be expected in a borrowing situation, is very unlikely in these communities (Supporting Information Table S7).

Alternatively, we find it more plausible that the Twa, Kwisi, and Kwepe may have had their own matriclan systems, and merely replaced their pre-existing clan labels with those of their dominant neighbors. This seems to be particularly evident among the Twa, who have a genealogically consistent matriclan system based on a clan inventory similar to the Himba, despite their close genetic relationship with the Kwisi (Figure 4b). The cultural approximation of the Twa to the Himba, which might be driven by geographical proximity (Figure 1), is also reflected in the apparent influence of Himba on the linguistic variety spoken by the Twa (Figure 8), as well as the documented tendency of the Twa to mimic the distinctive attire of the Himba women (Estermann, 1952). More generally, it is likely that clan-switching has facilitated female gene flow from the peripatetics into the dominant communities (Supporting Information Figure S1), thus explaining the reduced levels of sequence similarity observed within Himba and Kuvale clans (Figure 5b; Supporting Information Figure S2 and Supporting Information Table S7).

The matriclanic organization of the Namib peoples seems to have had a strong impact on their current patterns of mtDNA variation. The fact that the percentage of the total genetic diversity that is found between clans ($\Phi st = 0.51$) is much higher than that observed between populations ($\Phi st = 0.20$) suggests that ethnic groups arose from the assemblage of genetically different clans instead of clans being formed just by fissions occurring within groups. Thus, although both clan and group membership are determined by the mother, it is clear that the matrilineal principle is frequently violated during ethnogenesis. This pattern is especially striking among the Kuvale, who are highly endogamous and ethnically Bantu, yet comprise among their founders two descent groups (L0d1a1b1a and L0d1b1b; Figures 2a and 3), including the powerful clan of the cattle (Mukwangombe), that ultimately trace their origin to "Khoisan" populations. This type of population structure closely mirrors patterns of Y-chromosome variation previously reported in traditional patrilineal societies from other regions of the world (Chaix et al., 2004; Chaix et al., 2007; Sanchez-Faddeev et al., 2013).

4.2 | The southwestern African pastoral scene: Herero, Himba, Damara and Kuvale

The Himba and Kuvale from Angola are generally considered to be part of a broad cultural cluster of Bantu-speaking cattle herders that includes adjacent Himba groups from Namibia, as well as Herero populations extending from Namibia to Botswana (Bollig & Gewald, 2009). Besides sharing many aspects of their pastoral culture, these peoples are commonly thought to speak dialects of the same Herero language, which has been grouped with Nyaneka-Nkhumbi and Ovambo into a division of southwestern Bantu referred to as Cimbebasia (Vansina, 2004). However, the internal relations and migration routes of the southwestern Bantu herders, as well as the origins of their pastoral tradition remain poorly understood (Bollig & Gewald, 2009; Gibson, 1977).

Our results, together with previous work, show that the Himba, Tjimba, and Herero share a mtDNA profile that sets them apart from the Kuvale and other Bantu-speaking populations, but is not significantly different from the Damara who speak the same Khoe-Kwadi language as the pastoral Nama (Figures 2b and 6a) (Barbieri et al., 2014b; Coelho et al., 2009).

The most striking aspect of the Kuvale's maternal heritage is the high frequency (~50%) of characteristic "Khoisan" lineages associated with sequence types (L0d1a1b1a and L0d1b1b) that are likely to be derived from only two ancestral women (Figures 2a and 3). In contrast, the Himba, Herero and Damara have much lower frequencies of "Khoisan" mtDNA (10–17%), and share unusually high frequencies of sub-haplogroup L3d3a (38–61%), which is present in several Bantu, Kx'a, and Khoe-Kwadi speaking populations of southwestern Africa (Supporting Information Figure S5j; Barbieri et al., 2014b; Soodyall & Jenkins, 1993).

Previous interpretations of this mtDNA pattern have proposed that L3d3a was a pre-Bantu lineage retained by the Damara that was

subsequently transferred to the Himba and the Herero through admixture, instead of being inherited from a common ancestor by all three populations (Barbieri et al., 2014b). By using ABC analysis to explicitly test alternative evolutionary hypotheses about the relationships between the Kuvale, the Nyaneka-Nkhumbi and a meta-group lumping the Himba, Herero and Damara (HHD), we found that the maternal heritage of the latter group is nested within the southwestern Bantu peoples and shares a recent common ancestor with the Kuvale (Figure 7). In this context, it seems likely that the HHD and Kuvale represent the southern and northern branches, respectively, of a proto-population whose origins may be tentatively placed to the east of their present locations on the basis of the geographic distribution of their most common DNA lineages (Supporting Information Figure S5e, f, j, and k). The separation between the HHD and the Kuvale is paralleled by our linguistic results, which show that the Kuvale language cannot be considered a mere dialect of Herero, as was previously assumed (Estermann, 1981). According to this scenario, it is reasonable to assume that the Damara, like the Tjimba, are a cattleless branch of the Himba/Herero who changed their original Herero language after entering into a subordinate, peripatetic-like relationship with the pastoral Nama. Unlike the Damara, the Kuvale share most aspects of their pastoral culture with the Himba and Herero, in spite of their present genetic divergence (Figure 7).

Recent genome-wide polymorphism data have shown that the Himba, Herero and Damara share a genetic component that is found at lower frequencies in southwestern Bantu populations from the Atlantic coast to the Okavango Delta (Montinaro et al., 2017; Uren et al., 2016). These results, together with our mtDNA and linguistic data, are remarkably consistent with a previously-suggested scenario (Vansina, 2004) in which the Bantu pastoralists from Southwest Africa are an offshoot of the Ovambo and/or Nyaneka-Nkhumbi agropastoralists living around the Kunene river basin, who moved into the dry coastal areas of Namibia and Angola. In this framework, it is likely that the different combinations of genetic, linguistic and cultural profiles currently observed in the Himba, Herero, Damara, and Kuvale result from genetic drift, differential admixture and social stratification, instead of reflecting remote geographic origins or assimilation of pre-Bantu components other than "Khoisan" (cf. Möhlig, 2002; Vedder & Inskeep, 2003)

4.3 | The peoples of the Kuroca River: Kwisi, Twa and Kwepe

Due to their peripatetic way of life, the Kwisi, Twa, and Kwepe are frequently seen as the Angolan representatives of a wider group of populations whose origins are often linked to hypothetical pre-Bantu populations different from the Kx'a and Tuu-speaking foragers (Barnard, 1992; Blench, 2006; Cashdan, 1986; Güldemann, 2008; Westphal, 1963).

Although our results show that the Kwisi and the Twa form a relatively homogeneous group that is remarkably different from all other southern African peoples (Figures 2a and 6a), it is doubtful whether this differentiation could entirely reflect the genetic composition of a pre-Bantu remnant population.

The uniqueness of the two populations can be attributed to their high frequencies of sub-haplogroups L0a1b1 (21%), L0a1b2 (11%), L0a2a1b (31%), and L1c1b (22%), which represent ~85% on average of their mtDNA composition and are collectively much less frequent in the Himba (18%) and Kuvale (8%) (Figure 2b; Supporting Information Table S5). Among these four sub-haplogroups, L0a1b1 and L0a2a1b are most probably of Bantu origin, since their haplotypes are molecularly close to sequences that are observed in several Bantu-speaking populations from Zambia and Botswana (Figure 2b; Supporting Information Figure S5b and d). Haplotypes from sub-haplogroups L0a1b2 and L1c1b are confined to the Angolan Namib and have a less clear origin (Supporting Information Figure S5c and i). Although the long internal branches to their closest sequences suggest ancient isolation (Supporting Information Figure S5c and i), this pattern might also be due to insufficient sampling (Kivisild, 2006), or fragmentation of a large ancestral population (Nielsen & Beaumont, 2009).

Additional evidence for a link between the Kwisi and Twa and other Bantu peoples of the region is provided by the time depth and genealogical consistency of their clan system (see above), which further suggest that they are likely to be part of the constellation of matriclanic peoples that spread across southwestern Africa (Supporting Information Table S8; Figure 5b).

In this context, the genetic uniqueness of the Twa/Kwisi is probably better understood in the frame of a fusion-fission model, where the effects of genetic drift on mtDNA variation are enhanced by the influence of matrilineal kinship on population splitting and ethnogenesis (Fix, 1999; Neel & Salzano, 1967). Moreover, it is likely that this genetic differentiation was maintained and reinforced by the highly hierarchized social setting of pastoral societies, where impoverished cattleless peoples are marginalized by their dominant neighbors (Vansina, 2004).

The relationships between the Twa, Kwisi, and Kwepe have also been a matter of contention (Almeida, 1965; Cashdan, 1986; Estermann, 1976). Recently, based on the fact that the Kwepe formerly spoke Kwadi, and on the conclusion that this language could be grouped with Khoe in a single family, Güldemann (2008) suggested that the Kwepe were part of a putative pre-Bantu Khoe-Kwadi migration introducing pastoralism from eastern to southern Africa.

Our results show that the Kwepe have a very homogeneous mtDNA profile (only five different haplotypes; Supporting Information Table S5) that bears no resemblance to any other Khoe-Kwadi-speaking population and is largely shared with their neighbors from the Angolan Namib (Supporting Information Figure S1). Although the most common haplotype among the Kwepe is an L3f1b4 lineage (49%) with a likely Himba origin (Figure 2b; Supporting Information Figure S1), the other Kwepe haplotypes all belong to sub-haplogroups L0a1b1 (27%) or L1c1b (24%) that are more common and diverse in the Twa and Kwisi (Figure 2b; Supporting Information Figure S1). These observations suggest that the Kwisi, Twa, and Kwepe, who have overlapping residential areas around the Kuroca intermittent river (Figure 1), were originally the same people, and that the Kwepe mtDNA pool was disproportionally impacted by a single woman, or a kin group, migrating out of the Himba. The

genetic similarity of the Kwepe to immediate geographic neighbors displaying Bantu-related mtDNA profiles, rather than to other Khoe-Kwadi-speaking groups, suggests that their former use of Kwadi resulted from language shift after contact with a group of migrants that brought the Kwadi language to the Angolan Namib. So far the only available evidence for a possible genetic contribution of any Khoe-Kwadi migrants to the area is the occurrence in all Namib populations of the lactase persistence −14010*C allele (Pinto et al., 2016), which is found with elevated frequencies in several Khoe-Kwadi-speaking peoples of southern Africa (Macholdt et al., 2014). This evidence suggests that there might have been a measurable genetic impact associated with the original Kwadi-speakers that is not captured in the maternal lineages, and might be revealed by Y-chromosome markers and autosomal genome-wide data (currently under analysis).

In any case, the association of the Kwepe with the Kwadi language and a mtDNA profile that is largely derived from the Bantu, combined with the possibility that the Damara represent a branch of the Herero (see above), has important implications for the understanding of the spread of the Khoe-Kwadi family and pastoralism across southern Africa. When linguistically and geographically diverse populations from the region are compared, the most remarkable characteristic of all Khoe-Kwadi speaking peoples is their lack of a common mtDNA genetic heritage (Figure 6a; Supporting Information Figure S4). This absence of an mtDNA identity is paralleled by recent data on autosomal DNA variation, showing that many Khoe-Kwadi-speaking groups are genetically closer to populations occupying the same broad geographical area than they are to each other (Montinaro et al., 2017; Pickrell et al., 2012; Uren et al., 2016). Taken together, these patterns suggest that the spread of Khoe-Kwadi and its putative pastoral innovations were part of a complex process that cannot be simply modeled by a wave of advance similar to the spread of agriculture in Europe (Pinhasi, Fort, & Ammerman, 2005), nor by a rapid replacement model, analogous to the Bantu expansions (reviewed in Rocha & Fehn, 2016; see also Diamond & Bellwood, 2003). It seems more likely that many southern African groups adopted the Khoe-Kwadi language (and occasionally pastoralism) with only a small genetic contribution of incoming Khoe-Kwadi migrants. Our results now indicate that this type of cultural shift not only affected indigenous "Khoisan" foragers, but also impacted Bantu populations from southwestern Africa, leading to the emergence of new ethnic identities that are commonly perceived as enigmatic.

ACKNOWLEDGMENTS

The authors thank all sample donors for their participation in this study, the governments of Namibe and Kunene Provinces in Angola for supporting our work, João Guerra, Raimundo Dungulo, and Serafim Nemésio for assistance in the preparation of field work, António Mbeape, José Domingos, and Okongo Toko for assistance with sample collection, Roland Schröder and Enrico Macholdt for assistance in the lab, Shengyu Ni for assistance in the processing of sequencing data, Michael Dannemann, Cesare de Filippo, and António M Santos for help in statistical analyses, and Chiara Barbieri for valuable advice. Financial support for this research was provided by FEDER funds through the Operational Program for Competitiveness Factors—COMPETE, by National Funds through FCT—Foundation for Science and Technology under the PTDC/BIA-EVF/2907/2012 and FCOMP-01–0124-FEDER- 028341, and by the Max Planck Society. S.O. was supported by the FCT grant SFRH/BD/85776/2012, and M.G.-V. by FP7-REGPOT-2011-1-project 286431, UID/BIA/50027/2013, and POCI-01–0145-FEDER-006821. B.P. acknowledges the LABEX ASLAN (ANR-10-LABX-0081) of Université de Lyon for its financial support within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) of the French government operated by the National Research Agency (ANR). J.R., B.P., and M.S. planed the study; J.R., A.-M.F., S.O., T.A., and F.L. performed the fieldwork. S.O. generated the laboratory data. S.O., A.-M.F., M.G.-V., M.S., B.P., and J.R. analyzed the data. J.R., S.O., and A.-M.F. wrote the article with input from M. G.-V., B.P., and M.S.. This is scientific paper no. 7 from the Portuguese-Angolan TwinLab established between CIBIO/InBIO and ISCED/Huíla, Lubango. This work is dedicated to the memory of our colleague Samuel Aço.

ORCID

Jorge Rocha http://orcid.org/0000-0001-5460-7615

REFERENCES

Almeida, A. (1965). *Bushmen and other non-Bantu peoples of Angola: Three lectures*. Johannesburg: Witwatersrand University Press.

Andrews, R. M., Kubacka, I., Chinnery, P. F., Lightowlers, R. N., Turnbull, D. M., & Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics*, *23*, 147.

Atkinson, Q. D., Meade, A., Venditti, C., Greenhill, S. J., & Pagel, M. (2008). Languages evolve in punctuational bursts. *Science*, *319*, 588.

Baele, G., Li, W. L. S., Drummond, A. J., Suchard, M. A., & Lemey, P. (2013). Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. *Molecular Biology and Evolution*, *30*, 239–243.

Bahuchet, S. (2012). Changing language, remaining pygmy. *Human Biology*, *84*, 11–43.

Bandelt, H. J., Forster, P., & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*, 37–48.

Barbieri, C., Güldemann, T., Naumann, C., Gerlach, L., Berthold, F., Nakagawa, H., … Pakendorf, B. (2014a). Unraveling the complex maternal history of Southern African Khoisan populations. *American Journal of Physical Anthropology*, *153*, 435–448.

Barbieri, C., Vicente, M., Oliveira, S., Bostoen, K., Rocha, J., Stoneking, M., & Pakendorf, B. (2014b). Migration and interaction in a contact zone: mtDNA variation among Bantu-speakers in Southern Africa. *PLoS ONE*, *9*, e99117.

Barbieri, C., Whitten, M., Beyer, K., Schreiber, H., Li, M., & Pakendorf, B. (2012). Contrasting maternal and paternal histories in the linguistic context of Burkina Faso. *Molecular Biology and Evolution*, *29*, 1213–1223.

Barnard, A. (1992). *Hunters and herders of Southern Africa: A comparative ethnography of the Khoisan peoples*. Cambridge Studies in Social and Cultural Anthropology. Cambridge: Cambridge University Press.

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics*, *162*, 2025–2035.

Behar, D. M., Villems, R., Soodyall, H., Blue-Smith, J., Pereira, L., Metspalu, E., … Rosset, S. (2008). The dawn of human matrilineal diversity. *The American Journal of Human Genetics*, *82*, 1130–1140.

Blench, R. M. (2006). *Archaeology, language,and the African past*. Lanham, MD: Rowman Altamira.

Bollig, M. (2004). Hunters, Foragers, and Singing Smiths: The Metamorphosis of Peripatetic Peoples in Africa. *Customary Strangers. New Perspectives on Peripatetic Peoples in the Middle East, Africa, and Asia* (pp. 195–231).

Bollig, M. (2006). *Risk management in a hazardous environment* (Vol. *2*). New York: Springer.

Bollig, M., & Gewald, J.-B. (2009). *People, cattle and land: Transformations of a pastoral society. People, cattle and land*. Köln: Rüdiger Köppe.

Bouckaert, R. (2010). DensiTree: Making sense of sets of phylogenetic trees. *Bioinformatics*, *26*, 1372–1373.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., … Drummond, A. J. (2014). BEAST 2: A software platform for bayesian evolutionary analysis. *PLoS Computational Biology*, *10*, e1003537.

Bouckaert, R. (2016). BEAST analysis backing effective linguistics. https://github.com/rbouckaert/Babel. Accessed 18 December 2017.

Breton, G., Schlebusch, C. M., Lombard, M., Sjödin, P., Soodyall, H., & Jakobsson, M. (2014). Lactase persistence alleles reveal partial east African ancestry of southern African Khoe pastoralists. *Current Biology*, *24*, 852–858.

Cashdan, E. (1986). *Hunter-gatherers of the northern Kalahari*. In R. Vossen & K. Keuthmann (Eds.), *Contemporary studies on Khoisan: In honour of Oswin Köhler on the Occasion of his 75th Birthday*. Hamburg: Helmut Buske Verlag.

Chaix, R., Austerlitz, F., Khegay, T., Jacquesson, S., Hammer, M. F., Heyer, E., & Quintana-Murci, L. (2004). The genetic or mythical ancestry of descent groups: Lessons from the Y chromosome. *The American Journal of Human Genetics*, *75*, 1113–1116.

Chaix, R., Quintana-Murci, L., Hegay, T., Hammer, M. F., Mobasher, Z., Austerlitz, F., & Heyer, E. (2007). From social to genetic structures in central Asia. *Current Biology*, *17*, 43–48.

Coelho, M., Sequeira, F., Luiselli, D., Beleza, S., & Rocha, J. (2009). On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC Evolutionary Biology*, *9*, 80.

de Filippo, C., Key, F. M., Ghirotto, S., Benazzo, A., Meneu, J. R., Weihmann, A., … Andrés, A. M. (2016). Recent selection changes in human genes under long-term balancing selection. *Molecular Biology and Evolution*, *33*, 1435–1447.

Diamond, J., & Bellwood, P. (2003). Farmers and their languages: The first expansions. *Science*, *300*, 597–603.

Drummond, A. J., Rambaut, A., Shapiro, B., & Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, *22*, 1185–1192.

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*, 1969–1973.

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*, 1792–1797.

Efron, B., Halloran, E., & Holmes, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proceedings of the National Academy of Sciences of the United States of America*, *93*, 13429–13434.

Estermann, C. (1952). Clans et alliances entre clans dans le sud-ouest de l'Angola. *Anthropos*, *47*, 587–606.

Estermann, C. (1962). Les twa du sud-ouest de l'Angola. *Anthropos*, *57*, 465–474.

Estermann, C. (1976). *The Ethnography of Southwestern Angola: The Non-Bantu Peoples. The Ambo Ethnic Group (translated and* edited *by Gibson GD)* (Vol. *1*). New York: Africana Pub. Co.

Estermann, C. (1981). *The ethnography of southwestern Angola: The Herero people (translated and* edited *by* Gibson GD) (Vol. *3*). New York: Africana Pub. Co.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP Data. *PLoS Genetics*, *9*, e1003905.

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*, 564–567.

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, *39*, 783–791.

Fenner, J. N. (2005). Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *American Journal of Physical Anthropology*, *128*, 415–423.

Fix, A. (1999). *Migration and colonization in human microevolution* (Vol. *24*). Cambridge: Cambridge University Press.

Forster, P., Harding, R., Torroni, A., & Bandelt, H. J. (1996). Origin and evolution of Native American mtDNA variation: A reappraisal. *American Journal of Human Genetics*, *59*, 935–945.

Fu, Q., Li, H., Moorjani, P., Jay, F., Slepchenko, S. M., Bondarev, A. A., … Pääbo, S. (2014). Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature*, *514*, 445–449.

Fu, Q., Mittnik, A., Johnson, P. L. F., Bos, K., Lari, M., Bollongino, R., … Krause, J. (2013). A revised timescale for human evolution based on ancient mitochondrial genomes. *Current Biology*, *23*, 553–559.

Gibson, G. D. (1956). Double descent and its correlates among the Herero of Ngamiland. *American Anthropologist*, *58*, 109–139.

Gibson, G. D. (1977). Foreword. In F. Vivelo (Ed.), *In the Herero of Western Botswana: Aspects of change in a group of Bantu-speaking cattle herders*. St. Paul: West Publishing.

Greenberg, J. (1963). *The languages of Africa*. Bloomington, IN: Indiana University.

Greenhill, S. J., & Gray, R. D. (2009). *Austronesian language phylogenies: Myths and misconceptions about Bayesian computational methods*. In A. Adelaar & A. Pawley (Eds.), *Austronesian historical linguistics and culture history: A festschrift for Robert Blust*. Canberra: Pacific Linguistics.

Grollemund, R., Branford, S., Bostoen, K., Meade, A., Venditti, C., & Pagel, M. (2015). Bantu expansion shows that habitat alters the route and pace of human dispersals. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 13296–13301.

Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G., & Siepel, A. (2011). Bayesian inference of ancient human demography from individual genome sequences. *Nature Genetics*, *43*, 1031–1034.

Güldemann, T. (2008). A linguist's view: Khoe-Kwadi speakers as the earliest food-producers of southern Africa. *Southern African Humanities*, *20*, 93–132.

Güldemann, T., & Fehn, A. (Eds.). (2014). *Beyond "Khoisan": Historical relations in the Kalahari basin*. Amsterdam: John Benjamins Publishing Company.

Hartigan, J. A., & Wong, M. A. (1979). A k-means clustering algorithm. *Applied Statistics*, *28*, 100–108.

Haspelmath, M., & Tadmor, U. (Eds.). (2009). *Loanwords in the world's languages: A comparative handbook*. Berlin & New York: Mouton de Gruyter.

Henn, B. M., Gignoux, C., Lin, A. A., Oefner, P. J., Shen, P., Scozzari, R., … Underhill, P. A. (2008). Y-chromosomal evidence of a pastoralist migration through Tanzania to southern Africa. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 10693–10698.

Huson, D. H., & Bryant D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution 23*, 254–267.

Jordan, L. (2015). A comparison of five speech varieties from southwestern Angola. *SIL: SIL Electronic Survey Report 2015–2017*.

Kivisild, T. (2006). In H. Bandelt, M. Richards, & V. Macaulay (Eds.), *Human mitochondrial DNA and the evolution of Homo sapiens*. Berlin: Springer.

Kloss-Brandstätter, A., Pacher, D., Schönherr, S., Weissensteiner, H., Binna, R., Specht, G., & Kronenberg, F. (2011). HaploGrep: A fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Human Mutation*, *32*, 25–32.

König, C., & Heine, B. (2008). *A concise dictionary of Northwestern !Xun*. Köln: Rüdiger Köppe.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*, 357–359.

Leuenberger, C., & Wegmann, D. (2010). Bayesian computation and model selection without likelihoods. *Genetics*, *184*, 243–252.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*, 1754–1760.

Li, M., Schroeder, R., Ko, A., & Stoneking, M. (2012). Fidelity of capture-enrichment for mtDNA genome sequencing: Influence of NUMTs. *Nucleic Acids Research*, *40*,

Li, M., & Stoneking, M. (2012). A new approach for detecting low-level mutations in next-generation sequence data. *Genome Biology*, *13*, R34.

Li, S., & Jakobsson, M. (2012). Estimating demographic parameters from large-scale population genomic data using Approximate Bayesian Computation. *BMC Genetics*, *13*, 22.

MacCalman, H., & Grobbelaar, B. (1965). *Preliminary report of two stone-working OvaTjimba groups in the northern Kaokoveld of South West Africa*. Windhoek: Staatsmuseum.

Macholdt, E., Lede, V., Barbieri, C., Mpoloka, S. W., Chen, H., Slatkin, M., … Stoneking, M. (2014). Tracing pastoralist migrations to southern Africa with lactase persistence alleles. *Current Biology*, *24*, 875–879.

Maho, J. (2009). NUGL Online: The online version of the New Updated Guthrie List, a referential classification of the Bantu languages. http://goto.glocalnet.net/mahopapers/nuglonline.pdf. Accessed 18 December 2017.

Maricic, T., Whitten, M., & Pääbo, S. (2010). Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PloS One*, *5*, e14004.

Marks, S. J., Montinaro, F., Levy, H., Brisighelli, F., Ferri, G., Bertoncini, S., … Capelli, C. (2015). Static and moving frontiers: The genetic landscape of southern African Bantu-speaking populations. *Molecular Biology and Evolution*, *32*, 29–43.

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, *2010*, pdb.prot5448.

Möhlig, W. (2002). *The language history of Herero as a source of ethnohistorical interpretations*. In M. Bollig & J.-B. Gewald (Eds.), *People, cattle and land: Transformations of a pastoral society in southwestern Africa*. Köln: Rüdiger Köppe, 119–146.

Möhlig, W., & Kavari, J. (2008). *Reference grammar of Herero (Otjiherero) Bantu language of Namibia: With a glossary Otjiherero-English-Otjiherero*. Köln: Rüdiger Köppe.

Montinaro, F., Busby, G. B. J., Gonzalez-Santos, M., Oosthuitzen, O., Oosthuitzen, E., Anagnostou, P., … Capelli, C. (2017). Complex ancient genetic structure and cultural transitions in Southern African populations. *Genetics*, *205*, 303–316.

Neel, J. V., & Salzano, F. M. (1967). Further studies on the Xavante Indians. X. Some hypotheses-generalizations resulting from these studies. *American Journal of Human Genetics*, *19*, 554–574.

Neuenschwander, S., Largiadèr, C. R., Ray, N., Currat, M., Vonlanthen, P., & Excoffier, L. (2008). Colonization history of the Swiss Rhine basin by the bullhead (Cottus gobio): Inference under a Bayesian spatially explicit framework. *Molecular Ecology*, *17*, 757–772.

Nicholls, G. K., & Gray, R. (2006). Quantifying uncertainty in a stochastic Dollo model of vocabulary evolution. In P. Forster & C. Renfrew (Eds.), *Phylogenetic Methods and the Prehistory of Languages*. Cambridge: McDonald Institute for Archaeological Research (pp. 161–172).

Nielsen, R., & Beaumont, M. A. (2009). Statistical inferences in phylogeography. *Molecular Ecology*, *18*, 1034–1047.

Oven, M. V., & Kayser, M. (2008). Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human Mutation*, (September), 386–394.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*, 289–290.

Penny, D., McComish, B. J., Charleston, M. A., & Hendy, M. D. (2001). Mathematical elegance with biogeochemical realism: The covarion model of molecular evolution. *Journal of Molecular Evolution*, *53*, 711–723.

Pickrell, J. K., Patterson, N., Barbieri, C., Berthold, F., Gerlach, L., Güldemann, T., … Pakendorf, B. (2012). The genetic prehistory of southern Africa. *Nature Communications*, *3*, 1143.

Pinhasi, R., Fort, J., & Ammerman, A. J. (2005). Tracing the Origin and Spread of Agriculture in Europe. *PLoS Biology*, *3*, e410.

Pinto, J. C., Oliveira, S., Teixeira, S., Martins, D., Fehn, A. M., Aço, T., … Rocha, J. (2016). Food and pathogen adaptations in the Angolan Namib desert: Tracing the spread of lactase persistence and human African trypanosomiasis resistance into southwestern Africa. *American Journal of Physical Anthropology*, *161*, 436–447.

Posth, C., Renaud, G., Mittnik, A., Drucker, D. G., Rougier, H., Cupillard, C., … Krause, J. (2016). Pleistocene mitochondrial genomes suggest a single major dispersal of non-Africans and a late glacial population turnover in Europe. *Current Biology*, *26*, 827–833.

Rambaut, A., & Drummond, A. (2007). Tracer v1. 4. http://beast.bio.ed.ac.uk/Tracer. Accessed 18 December 2017.

Renaud, G., Stenzel, U., & Kelso, J. (2014). LeeHom: Adaptor trimming and merging for Illumina sequencing reads. *Nucleic Acids Research*, *42*, e141.

Renaud, G., Stenzel, U., Maricic, T., Wiebe, V., & Kelso, J. (2015). DeML: Robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics*, *31*, 770–772.

Rieux, A., Eriksson, A., Li, M., Sobkowiak, B., Weinert, L. A., Warmuth, V., … Balloux, F. (2014). Improved calibration of the human mitochondrial clock using ancient genomes. *Molecular Biology and Evolution*, *31*, 2780–2792.

Rocha, J., & Fehn, A.-M. (2016). Genetics and demographic history of the Bantu. *eLS*, 1–9.

Russell, T., Silva, F., & Steele, J. (2014). Modelling the spread of farming in the bantu-speaking regions of africa: An archaeology-based phylogeography. *PLoS ONE*, *9*, e87854.

Sadr, K. (2015). Livestock first reached southern Africa in two separate events. *PLoS ONE*, *10*, e0134215.

Sanchez-Faddeev, H., Pijpe, J., van der Hulle, T., Meij, H. J., J van der Gaag, K., Slagboom, P. E., ... de Knijff, P. (2013). The influence of clan structure on the genetic variation in a single Ghanaian village. *European Journal of Human Genetics*, *21*, 1134–1139.

Schlebusch, C. M., Skoglund, P., Sjodin, P., Gattepaille, L. M., Hernandez, D., Jay, F., ... Jakobsson, M. (2012). Genomic variation in seven Khoe-San groups reveals adaptation and complex african history. *Science*, *338*, 374–379.

Snider, K., & Roberts, J. (2006). SIL comparative African wordlist. *Journal of West African Languages*, *31*, 73–122.

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., ... Richards, M. B. (2009). Correcting for purifying selection: An improved human mitochondrial molecular clock. *The American Journal of Human Genetics*, *84*, 740–759.

Soodyall, H., & Jenkins, T. (1993). Mitochondrial DNA polymorphisms in Negroid populations from Namibia: New light on the origins of the Dama, Herero and Ambo. *Annals of the Human Biology*, *20*, 477–485.

Swadesh, M. (1952). Lexicostatistic dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society*, *96*, 452–463.

Uren, C., Kim, M., Martin, A. R., Bobo, D., Gignoux, C. R., Helden, D. V., ... Henn, B. M. (2016). Fine-scale human population structure in southern Africa reflects ecological boundaries - Supplementary information. *Genetics*, *204*, 303–314.

Vansina, J. (1990). *Paths in the rainforests: Toward a history of political tradition in Equatorial Africa*. Madison: University of Wisconsin Press.

Vansina, J. (2004). *How societies are born: Governance in west Central africa before 1600*. Charlottesville: University of Virginia Press.

Vashro, L., & Cashdan, E. (2015). Spatial cognition, mobility, and reproductive success in northwestern Namibia. *Evolution and Human Behavior*, *36*, 123–129.

Vedder, H., & Inskeep, A. (2003). *Heinrich Vedder's the bergdama: An annotated translation of the German original with additional ethnographic material* (Vol. *2*). Köln: Rüdiger Köppe.

Veeramah, K. R., Wegmann, D., Woerner, A., Mendez, F. L., Watkins, J. C., Destro-Bisol, G., ... Hammer, M. F. (2012). An early divergence of KhoeSan ancestors from those of other modern humans is supported by an ABC-based analysis of autosomal resequencing data. *Molecular Biology and Evolution*, *29*, 617–630.

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York: Springer.

Vivelo, F. (1977). *The Herero of Western Botswana: Aspects of change in a group of Bantu-speaking cattle herders*. St. Paul: West Publishing.

Vossen, R. (1997). *Die Khoe Sprachen: Ein Beitrag zur Erforschung der Sprachgeschichte Afrikas*. Köln: Rüdiger Köppe.

Warmelo, N. V. (1951). *Notes on the Kaokoveld (South west Africa) and its people*. Pretoria: Government printer.

Wegmann, D., & Excoffier, L. (2010). Bayesian inference of the demographic history of chimpanzees. *Molecular Biology and Evolution*, *27*, 1425–1435.

Wegmann, D., Leuenberger, C., & Excoffier, L. (2009). Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood. *Genetics*, *182*, 1207–1218.

Wegmann, D., Leuenberger, C., Neuenschwander, S., & Excoffier, L. (2010). ABCtoolbox: A versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics*, *11*, 116.

Westphal, E. O. J. (1963). The linguistic prehistory of southern Africa: Bush, Kwadi, Hottentot, and Bantu linguistic relationships. *Africa*, *33*, 237–265.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Oliveira S, Fehn A-M, Aço T, et al. Matriclans shape populations: Insights from the Angolan Namib Desert into the maternal genetic history of southern Africa. *Am J Phys Anthropol*. 2017;00:1–18. https://doi.org/10.1002/ajpa.23378