# Journal Pre-proof


Mitochondrial DNA variation in Sub-Saharan Africa: forensic data from a mixed West African sample, Côte d'Ivoire (Ivory Coast), and Rwanda

Tanja M.K. Göbel, Martin Bodner, Carlo Robino, Christa Augustin, Gabriela E. Huber, Michele Marra, Léon Mutesa, Serena Pasino, Alfredo Santovito, Bettina Zimmermann, Peter M. Schneider, Walther Parson

PII: S1872-4973(19)30249-2

DOI: https://doi.org/10.1016/j.fsigen.2019.102202

Reference: FSIGEN 102202

To appear in: *Forensic Science International: Genetics*

Received Date: 4 June 2019

Revised Date: 4 November 2019

Accepted Date: 6 November 2019

Please cite this article as: Göbel TMK, Bodner M, Robino C, Augustin C, Huber GE, Marra M, Mutesa L, Pasino S, Santovito A, Zimmermann B, Schneider PM, Parson W, Mitochondrial DNA variation in Sub-Saharan Africa: forensic data from a mixed West African sample, Côte d'Ivoire (Ivory Coast), and Rwanda, *Forensic Science International: Genetics* (2019), doi: https://doi.org/10.1016/j.fsigen.2019.102202

# Mitochondrial DNA variation in Sub-Saharan Africa: forensic data from a mixed West African sample, Côte d'Ivoire (Ivory Coast), and Rwanda

Tanja M.K. Göbel[a*], Martin Bodner[b*] martin.bodner@i-med.ac.at ✉, Carlo Robino[c], Christa Augustin[d], Gabriela E. Huber[b], Michele Marra[c], Léon Mutesa[e], Serena Pasino[c], Alfredo Santovito[f], Bettina Zimmermann[b], Peter M. Schneider[a], Walther Parson[b,g]

[a]Institute of Legal Medicine, Medical Faculty, University of Cologne, Cologne, Germany

[b]Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck, Austria

[c]Department of Public Health Sciences and Pediatrics, University of Turin, Turin, Italy

[d]Department of Legal Medicine, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[e]Center for Human Genetics, University of Rwanda, Kigali, Rwanda

[f]Department of Life Sciences and Systems Biology, University of Turin, Turin, Italy

[g]Forensic Science Program, The Pennsylvania State University, PA, USA

[*]equally contributing first authors

✉corresponding author at: Institute of Legal Medicine, Medical University of Innsbruck, Müllerstrasse 44/3, 6020 Innsbruck, Austria; Tel. +43 512 9003 70600, Fax +43 512 9003 73640,



## Highlights

- The first forensic mtDNA population sample for Côte d'Ivoire is presented

- The first forensic mtDNA population sample for Rwanda is presented
- A mixed forensic mtDNA population sample for West Africa is presented
- The sample may serve as database for several yet uncovered countries of West Africa
- A strategy for use of impaired but precious population samples is presented

**Abstract**

This study provides 398 novel complete mitochondrial control region sequences that augment the still underrepresented data from Africa by three datasets: a mixed West African sample set deriving from 12 countries (n=145) and datasets from Côte d'Ivoire (Ivory Coast) (n=100) as well as Rwanda (n=153). The analysis of mtDNA variation and genetic comparisons with published data revealed low random match probabilities in all three datasets and typical West African and East African diversity, respectively. Genetic parameters indicate that the presented mixed West African dataset may serve as first forensic mtDNA control region database for West Africa in general. In addition, a strategy for responsible forensic application of precious mtDNA population samples potentially containing close maternal relatives is outlined. The datasets will be uploaded to the forensic mtDNA database EMPOP (https://empop.online) upon publication.

**Keywords**

EMPOP, forensics, mtDNA database, West Africa, East Africa, quality control

## 1. Introduction

Mitochondrial DNA (mtDNA) is a powerful source of information in forensic applications, in particular when samples contain nuclear DNA of limited quality and quantity, or when even distant maternal kinship is investigated (e.g., [1-3]). However, its power is limited due to the lack of recombination and a strictly maternal inheritance. The significance of matching mtDNA sequences and the weight of evidence depends on the variation present in the relevant population(s) and is assessed by determining the frequency of the particular haplotype in an appropriate population database. To assess the rarity of profiles, high quality forensic population databases containing representative and quality-checked datasets have been established [3-5]. African mtDNA data, mirroring the genetic diversity in the cradle of humanity, are greatly underrepresented (not only) in forensic repositories. The forensic mtDNA database EMPOP [5], available at https://empop.online, in its current v.4/R12 holds 2,179 haplotypes from ten countries of Africa (5.1% of the total entries). In contrast, the databases for Asia, Europe, and the Americas are approximately four-, six- and ninefold larger than the African database, respectively, and 3,080 haplotypes are stored for US African Americans. The recommended minimum mtDNA range for population studies for forensic databasing (nps 16024-16569 1-576) [3] comprises the non-coding control region (CR) and is available for 1,980 haplotypes from Africa (90.9% of the African and 5.9% of all 33,447 CR haplotypes in the database, respectively). Although its Sub-Saharan part covers 74% of the surface area and houses 82% of the population of Africa (estimate for 2018) [6], only 52.5% of the African samples in EMPOP derive from this part of the continent. Thus, forensic application regarding Africa is clearly limited.

Beyond mere forensic significance, data on nature, spread and frequency of mtDNA variation are a prerequisite for many phylogenetic and phylogeographic insights. Exploring the extant and past matrilineal genetic structure is of particular interest on the continent of origin of

the anatomically modern human and starting point of all later migrations: Africa is the continent with the greatest expected mtDNA variation and most ancient lineages. Its mtDNA landscape has been shaped by both early and recent migrations since more than 150 thousand years (ky), affected by climatic, socio-economic, and political factors [7-10]. Despite all demographic shifts, some regional lineage accumulations can be noticed. The earliest split separated haplogroup L0, with sub-Saharan origin, mainly Southeast/East African spread and dated to 147.9 ky ago using complete mitogenomes, from the later evolving clades. These are L1 and L2 (mainly Southwest and Central/West African), L3 (that gave rise to all non-African mtDNA diversity and is broadly spread), as well as L4, L5 and L6 (mainly East African). L1-L6 are dated between 24.6 and 127.4 ky ago [10, 11]. To shed light on the mtDNA sequence variation in yet undescribed and underrepresented African populations, we present novel high-quality mtDNA CR data from Côte d'Ivoire (Ivory Coast), eleven countries of West Africa with small representations each, and Rwanda in East Africa. The datasets were compared to other African populations in order to identify phylogeographic patterns. The data will enrich the pool of publicly available African mtDNA sequence data of forensic quality and will be incorporated in EMPOP [5].

## 2. Materials and methods

### *2.1. Samples and DNA extraction*

This study combines 398 novel mtDNA haplotypes deriving from three independent sample sets:

***(i)*** Buccal swabs from 145 individuals with maternal ancestry from twelve West African countries were obtained after informed consent. This **West African** (WAF) dataset is maintained throughout this study due to the common acquisition history and the relatively

to very small number of individuals sampled per country, viz. Togo (n=34), Nigeria (n=28), Côte d'Ivoire (n=27), Ghana (n=22), Cameroon (n=11), Gambia (n=5), Guinea (n=5), Niger (n=4), Senegal (n=3), Benin (n=2), Burkina Faso (n=2) and Liberia (n=1). For one donor, only general West African maternal origin could be assured (Figure 1, Table S1). Ethnolinguistic information was not collected. Ethical approval was granted by the Ethics Commission of Cologne University's Faculty of Medicine (12-003/2012). Genomic DNA was extracted using the MagAttract DNA mini M48 kit on a BioRobot M48 workstation (Qiagen, Hilden, Germany) according to the manufacturer's recommendations;

***(ii)*** Further 153 buccal swabs from individuals living in **Rwanda** (RWA) (Figure 1) were obtained. No ethnolinguistic information was collected. The study was approved by the Rwanda National Ethics Committee (675/RNEC/2013). Genomic DNA was extracted using Chelex-100 [12];

***(iii)*** In addition, 100 whole blood samples were collected at the Ouangolodougou (Wangolodougou) hospital in Northern **Côte d'Ivoire** (Ivory Coast) (CIV) after informed consent. Their use for genetic studies was authorized by the hospital dean. The donors' self-reported ethnolinguistic affiliation was Dioula (n=49), Baoulé (n=16), Sénoufo (n=15), Bété (n=14), and Peulh (n=6), respectively (Figure 1, Table S2). Genomic DNA was extracted using the ChargeSwitch gDNA Normalized Buccal Cell kit (TFS) and a KingFisher mL magnetic particle processor (TFS).

### *2.2. MtDNA sequencing and haplogroup estimation*

For the West African and Rwandan samples, the entire CR was amplified, sequenced and interpreted following EMPOP recommendations for forensic mtDNA analysis [3, 13, 14] using BigDye Terminator v.1.1 and a 3100 ABI Prism Genetic Analyzer (Thermo Fisher Scientific

[TFS], Waltham, MA). Results were evaluated using Sequencher v4.8 (GeneCodes, Ann Arbor, MI). CR sequencing for the samples from Côte d'Ivoire was performed following earlier EMPOP recommendations [15] using BigDye Terminator v.3.1 and a 310 ABI Prism Genetic Analyzer (TFS). Resulting sequences were interpreted using SeqScape software (TFS). For reporting purposes, all 398 mtDNA haplotypes from the three subsets were aligned and compared to the revised Cambridge Reference Sequence (rCRS) [16]. SAM2 [17] was used for alignment and haplogroup estimation according to PhyloTree*mt*, build 17 [18], based on the variation and fluctuation observed in large numbers of verified haplotypes rather than strict decision trees. This approach is implemented in EMPOP [5]. The samples from Côte d'Ivoire were previously typed for X-STRs [19], and 230 mtDNA coding region SNPs of the West African sample set are included in a pan-African study [10].

### *2.3. Quality control*

All mtDNA haplotypes passed EMPOP quality control including plausibility, phylogenetic and quasi-median network inspection [5, 20]. Identity of samples and close maternal kinship of donors exhibiting identical haplotypes, excluding polycytosine stretch variation, were assessed and excluded from autosomal STR data for the West African and Côte d'Ivoire dataset as described in [21] using Familias [22]. X-STRs [19] were additionally considered for the latter dataset using Mendel [23]. For the donors from Rwanda, kinship could not be evaluated. Since datasets used for forensic mtDNA frequency estimates should generally not contain samples from close maternal relatives [21], a "full" Rwandan sample set including all samples and assuming absence of any close maternal kinship of donors ("**RWA_all**") is presented in this study in parallel, where appropriate, with a "minimal" set including only one representative for each unique CR haplotype ("**RWA_min**") to exclude potentially close

relatives. The forensically “ideal” representative mtDNA population sample can be assumed to lie in between the two extremes.

### *2.4. Forensic and population genetic calculations and comparisons*

The populations presented in this study were compared to all published African mtDNA datasets that had passed forensic quality control and were available in the complete CR range, excluding only European immigrant, tribal and West Eurasian outlier populations. Where appropriate, a dataset combining the WAF and CIV samples (“WAF+CIV”, n=245) was additionally considered to account for their common geographic origin. A total of 758 published haplotypes were used for comparisons along with the 398 novel haplotypes. Literature data included populations from Egypt (North Africa, n=277) [24], Ghana (West Africa, n=191) [25], Somalia (East Africa, n=190) [26] and Kenya (East Africa, n=100) [15]. For intra- and interpopulation comparisons, forensic and population genetic molecular diversity indices, as well as the analysis of molecular variance (AMOVA), were calculated using Arlequin v3.5.1.2 [27] from full CR data disregarding insertions due to length variation in polycytosine stretches, i.e. using nps 16024-16193 16194-309 310-573 574-576 as the range. Random match probability (RMP) and power of discrimination (haplotype diversity) were calculated as in [28]. To visualize the genetic relation of populations, a multi-dimensional scaling (MDS) plot (correspondence analysis) was created from pairwise $F_{ST}$ values using SPSS statistics v22 (IBM, Armonk, NY).

## 3. Results and discussion

### *3.1. Haplotype composition of the three novel datasets*

The **West African** sample (n=145) contained 132 different CR haplotypes (disregarding insertions at nps 16193, 309 and 573), 121 (91.7%) of which were unique. Hence, RMP was calculated as 0.79%, the power of discrimination as 99.9%, and the mean number of pairwise differences (MNPD) as 16.02 ± 7.18. Two similar haplotypes of haplogroup L1b were most frequent with three observations (2.1%) each: 73G 152C 182T 185T 195C 247A 263G 315.1C 357G 523DEL 524DEL 16126C 16187T 16189C 16223T 16264T 16270T 16278T 16311C 16519C relative to the rCRS [16], and this haplotype with the additional 16293G (Table 1, Table S1). In the **Côte d'Ivoire** sample (n=100), 91 different haplotypes were found, whereof 83 (91.2%) were unique. RMP was 1.20%, the power of discrimination reached 99.8%, and the MNPD was 14.09 ± 6.37. The most frequent haplotype with three observations (3.0%) was 73G 263G 315.1C 523DEL 524DEL 16124C 16223T 16278T 16362C 16527T falling into haplogroup L3b2 (Table 1, Table S2). In the dataset combining the WAF and CIV samples (n=245), 211 haplotypes were found, whereof 186 (88.2%) were unique. RMP was 0.57%, the power of discrimination was 99.8%, and the MNPD was 15.26 ± 6.84. Two haplotypes were found most frequent with four observations (1.6%) each: the latter of the L1b haplotypes found most frequent in the WAF sample (see above) and an L2c1 haplotype (73G 93G 146C 150T 152C 182T 195C 198T 263G 315.1C 325T 523DEL 524DEL 16223T 16278T 16318G 16390A).

The **Rwandan** dataset revealed 105 different haplotypes in the mtDNA CR range, 81 of which (77.1%) were unique also in the RWA_all sample (n=153). The most frequent haplotype was 73G 146C 153G 263G 315.1C 16223T 16311C 16354T 16399G falling into haplogroup L3h1a2a with nine occurrences (5.9%). Twenty-four haplotypes occurred more than once in the Rwandan dataset (range of occurrence: 2-7; mean: 3; median: 2). Thus, the RWA_min sample consisted of 105 haplotypes, excluding 48 potential close relatives. In the

RWA_all|RWA_min samples, RMP was 1.59%|0.95%, power of discrimination was 99.1%|100.0%, and the MNPD was calculated as 18.95 ± 8.44|20.02 ± 8.92 (Table 1, Table S3).

### 3.2. *Haplogroup composition of the three novel datasets*

The analysis of complete CR sequences revealed haplogroup spectra that were similar to published West and East African datasets, respectively (Table S4, Figure S1): The mixed **West African** sample (n=145) was composed of 37.9% L3 haplotypes (in descending frequency L3e, L3f, L3b, L3d, L3k), 37.2% L2 haplotypes (predominantly L2a; also L2b, L2c, L2d, L2e), 19.3% L1 haplotypes (L1b, L1c), 4.8% L0a1 haplotypes and a singleton (0.7%) L4b1a haplotype. The most common of the 61 CR haplogroups were L2a1 (14.5%) and L1b (5.5%) (Tables S1, Table S4). The **Côte d'Ivoire** sample (n=100) contained 45.0% L3 haplotypes (in descending frequency L3b, L3d, L3e, L3f, L3h, L3k), 38.0% L2 haplotypes, 15.0% L1 haplotypes, as well as singletons of haplogroup L0a1a and U6a5b (1.0% each). The first-subclade level patterns found within the L1 and L2 clades were identical to those reported above for the mixed West African sample. Of the 48 CR haplogroups found, the most frequent ones were L2a1 (12.0%) and L1b (7.0%) (Table S2, Table S4). Both the mixed West African and the Côte d'Ivoire sample sets thus showed a largely similar haplogroup composition as expected for West African populations, with differences only at the subclade level, reflecting their geographic proximity and partial overlap. The almost exclusive trisection into L3-, L2-, and L1-haplogroup clusters with predominance of L3 and L2 was likewise reported from other West African populations, where also observations of L0a (frequent in East Africa) and U6a haplotypes (discussed below) were made in similarly small proportions. Also, haplogroup L4 is more abundant in East Africa (see below) [25, 29] (Figure S1).

The **Rwandan** sample (listed here as RWA_all|RWA_min) revealed 35.9|29.5% L3 (comprising, in descending frequency, L3e, L3h, L3b, L3f, L3d, L3x, L3a), 23.7%|23.8% L0 (L0a, L0f), 15.0|18.1% L2 (L2a, L2b, L2d), 12.4|10.5% L4 (L4b, L4a), 5.9|7.6% L1 (L1c, L1b), and 2.6|3.8% L5 haplotypes (L5a, L5b), as well as haplotypes assigned to haplogroups M1a1 (2.6|3.8%), N1a1a (1.3|1.9%), and K1a (0.6|1.0%). Among the 55 found, the most frequent CR haplogroups were L4b2 (9.8|7.6%) and L3b (8.5%|5.7%) (Table S3, Table S4). This broader haplogroup pattern, mirrored also by higher MNPD (Table 1), is the expected East African [11, 25, 29] and highly similar to that reported from Kenya [15] (Figure S1). Clades not affected by identical haplotypes (i.e., absolute numbers of occurrences did not change between RWA_all and RWA_min) are underlined in Table S4.

Most of the non-L low-frequency lineages found in the novel datasets are highly informative about human history. Lineages M1(a) and U6(a) are explained as signals of ancient backflow into North Africa from the Mediterranean area in the Early Upper Paleolithic [30-33]. N1a1a is a low-frequency lineage with a relict distribution likely indicating a Pleistocene dispersal from Arabia [31, 32]. All are reported also in other East and North African populations [15, 26, 29, 34]. The single haplotype of K1a, a lineage found across West Eurasia according to EMPOP [5], might be attributable to more recent migration to Rwanda and is, intriguingly, shared with the dataset from Somalia [26]. Haplotypes assigned only to basal clades, most notably two L3* representatives in the West African dataset (Table S4), pinpoint the potential of additional sampling and coding region sequencing towards a more detailed haplogroup affiliation [35, 36].

### *3.3. Heteroplasmic positions*

Point heteroplasmies (all transitions) were observed at eleven different positions in ten samples of the West African dataset (204Y [twice], 16086Y, 16093Y [four occurrences], 16189Y, 16264Y, 16286Y, 16344Y, 16390R, 16400Y, 16526R, 16527Y) (Table S1) and eight positions in seven samples of the Rwanda (151Y, 152Y, 200R, 248R, 338Y, 16093Y, 16129R, 16172Y) (Table S3). All affected positions and the hotspots were in agreement with previous forensic reports [37] and EMPOP [5].

### *3.4. Genetic comparisons of populations*

In order to gain better insight into their genetic position, the three novel populations presented in this study (total n=398) were compared to published forensic African mtDNA datasets in the complete CR range excluding length variation in polycytosine stretches after nps 16193, 309 and 573. In total, 1,156 haplotypes were compared, including datasets from Egypt, Ghana, Somalia, and Kenya.

The **West African** sample shared altogether 26 (19.7%) of its haplotypes comprising 38 (26.2%) of the samples with the other datasets. Shared haplotypes were found in all populations, mainly in Ghana (15 shared haplotypes) and Côte d'Ivoire (this study; 12 shared haplotypes), four to five haplotypes in the other populations and only one in Somalia [26] (Table S5). Three and four of the haplotypes the mixed West African sample shared with the Ghana and the Côte d'Ivoire (this study) sample, respectively, derived from donors with maternal ancestry also in these countries, but only one haplotype exclusively (Table S1). The **Côte d'Ivoire** sample shared 16 (17.6%) of its haplotypes, comprising 19 samples (19.0%) with the other populations, mainly the West Africa sample (this study; 12 shared haplotypes) and Ghana (11 shared haplotypes), while the other populations shared up to three haplotypes, except Somalia that did not share any (Table S5). Consequently, the dataset

combining the samples from West Africa and Côte d'Ivoire shared 26 (12.3%) of its haplotypes, comprising 46 samples (18.8%), with the remaining datasets, mainly with Ghana (18 haplotypes). Five haplotypes were shared with the datasets from Rwanda (this study) and Kenya, four with Egypt and one with Somalia (Table S5). The **Rwandan** sample shared 15 (14.3%) of its haplotypes with other datasets. The shared proportion comprised 34 (22.2%)|15 (14.3%) samples when using the RWA_all|RWA_min dataset, respectively. All datasets shared up to four haplotypes with the Rwandan sample, except Kenya with nine shared haplotypes. An L3f1b1a haplotype was the only one found in each of the three novel datasets presented in this study. It was a singleton in all of them. When the WAF and CIV datasets were combined, five haplotypes were shared with the RWA datasets. No haplotype was shared between all populations included in the comparisons (Table S5).

AMOVA of the seven population samples (using the RWA_all sample for Rwanda) revealed that the vast majority of the observed variation in the mtDNA structure represented differences within populations (93.61%), while 6.39% were attributable to differences among them. When using the RWA_min sample instead, proportions changed to 93.65% and 6.35%, respectively. AMOVA results were statistically significant (Table S6). Pairwise $F_{ST}$ distances of the novel datasets were low in comparisons of the **West African** and **Côte d'Ivoire** samples with each other and Ghana, and intermediate in comparisons with Somalia, Kenya, and Rwanda. The same applied to the combined WAF+CIV dataset. For the **Rwandan** sample, $F_{ST}$ values were low with Kenya and intermediate with all other populations, except Egypt, that showed highest $F_{ST}$ values with all three novel datasets. The small genetic variance detected was significant for almost all comparisons. No significant difference in genetic structure was found between the West African and the Côte d'Ivoire as well as the Ghana samples, also when the first two were combined, and between the Rwanda and the

Kenya samples (discussed below) (Table S7). The visualization of $F_{ST}$ distances between populations in an MDS plot depicts the clusters of geographically close and genetically highly similar populations as indicated by the results from all genetic analyses outlined above. The novel populations from West Africa and Côte d'Ivoire (and their combination) clustered with the Ghana sample. Another cluster, yet distant from the former, was formed by both novel Rwandan and the Kenyan datasets; while the Somali sample was at intermediate distance from both clusters and the sample from Egypt appeared most distant from the two clusters while closer to the Somali sample (Figure S2). Thus, the correspondence analysis mirrors the geographic proximity of the populations. Along with the results described above, it likely reflects the common origin and high degree of recent or potentially ongoing migration and maternal gene flow among those geographically close populations. These factors led to a similar genetic composition, while the populations at greater geographic distance also revealed greater evolutionary distance and limited gene flow.

## 4. Conclusion

This study is a contribution in mitigating the underrepresentation of African data not only in forensic mitogenetics. The mixed **West African** sample has shown high genetic similarity to country-specific sets of the area (CIV in this study, [25]), confirming its mixed origin and supporting its representativeness. It thus may serve as general substitute forensic reference dataset for many West African countries where specific collections are not yet available. This study also presents the first forensic datasets depicting complete mtDNA CR sequences from **Côte d'Ivoire** and **Rwanda**. The applicability of the latter appears only slightly limited by potential maternal kinship, as indicated by statistics including both the full and the minimal versions of the dataset. This might serve as practical example of a responsible forensic

approach towards “convenience” datasets [21] to gain utmost scientifically valid output, namely when the population is not easily accessible for further sampling. In addition, these novel data contribute to an African etalon dataset that will assist future forensic quality control and the development of further continent-specific filters for quasi-median network analysis [20, 38]. The datasets will be uploaded to the EMPOP database (https://empop.online) [5] upon publication (EMP00793 [CIV]; EMP00042 [RWA]; EMP00082, EMP00299, EMP00300, EMP00794 [WAF]) and aid geography-, metapopulation-, as well as phylogeny-based queries. All sequences are also available from GenBank (https://www.ncbi.nlm.nih.gov/genbank) at accession numbers MK976923-MK977022 (CIV, Table S2), MK977023-MK977167 (WAF, Table S1), and MN018603-MN018755 (RWA, Table S3).

**Acknowledgements**

The authors are greatly indebted to all individuals that donated their DNA to research. The authors wish to thank current and former staff members at the Institute of Legal Medicine, Medical University of Innsbruck, for excellent technical assistance: Liane Fendt, Nicole Huber, Simone Nagl, Daniela Niederwieser, and Lisa Schnaller (in alphabetical order).

**Appendix A. Supplementary data**

Supplementary data associated with this article can be found, in the online version, at doi: …

**References**

[1] M.M. Holland, T.J. Parsons, Mitochondrial DNA Sequence Analysis - Validation and Use for Forensic Casework, Forensic Sci Rev, 11 (1999) 21-50.

[2] T.E. King, G.G. Fortes, P. Balaresque, M.G. Thomas, D. Balding, P. Maisano Delser, R. Neumann, W. Parson, M. Knapp, S. Walsh, L. Tonasso, J. Holt, M. Kayser, J. Appleby,

P. Forster, D. Ekserdjian, M. Hofreiter, K. Schurer, Identification of the remains of King Richard III, Nat Commun, 5 (2014) 5631.

[3] W. Parson, L. Gusmao, D.R. Hares, J.A. Irwin, W.R. Mayr, N. Morling, E. Pokorak, M. Prinz, A. Salas, P.M. Schneider, T.J. Parsons, DNA Commission of the ISFG: revised and extended guidelines for mitochondrial DNA typing, Forensic Sci Int Genet, 13 (2014) 134-142.

[4] J.A. Irwin, J.L. Saunier, K.M. Strouss, K.A. Sturk, T.M. Diegoli, R.S. Just, M.D. Coble, W. Parson, T.J. Parsons, Development and expansion of high-quality control region databases to improve forensic mtDNA evidence interpretation, Forensic Sci Int Genet, 1 (2007) 154-157.

[5] W. Parson, A. Dür, EMPOP-a forensic mtDNA database, Forensic Sci Int Genet, 1 (2007) 88-92.

[6] United Nations Department of Economic and Social Affairs, Statistics Division, Statistical Yearbook 2018, Sixty-first issue, United Nations, New York, 2018.

[7] A. Torroni, A. Achilli, V. Macaulay, M. Richards, H.J. Bandelt, Harvesting the fruit of the human mtDNA tree, Trends Genet, 22 (2006) 339-345.

[8] D.M. Behar, R. Villems, H. Soodyall, J. Blue-Smith, L. Pereira, E. Metspalu, R. Scozzari, H. Makkan, S. Tzur, D. Comas, J. Bertranpetit, L. Quintana-Murci, C. Tyler-Smith, R.S. Wells, S. Rosset, Genographic Consortium, The dawn of human matrilineal diversity, Am J Hum Genet, 82 (2008) 1130-1140.

[9] K. Shillington, History of Africa, 2nd edition, Palgrave Macmillan, New York, 2005.

[10] M. Cerezo, L. Gusmao, V. Cerny, N. Uddin, D. Syndercombe-Court, A. Gomez-Carballa, T. Göbel, P.M. Schneider, A. Salas, Comprehensive Analysis of Pan-African Mitochondrial DNA Variation Provides New Insights into Continental Variation and Demography, J Genet Genomics, 43 (2016) 133-143.

[11] T. Rito, D. Vieira, M. Silva, E. Conde-Sousa, L. Pereira, P. Mellars, M.B. Richards, P. Soares, A dispersal of Homo sapiens from southern to eastern Africa immediately preceded the out-of-Africa migration, Sci Rep, 9 (2019) 4728.

[12] P.S. Walsh, D.A. Metzger, R. Higuchi, Chelex-100 as a Medium for Simple Extraction of DNA for PCR-Based Typing from Forensic Material, Biotechniques, 10 (1991) 506-513.

[13] W. Parson, H.J. Bandelt, Extended guidelines for mtDNA typing of population data in forensic science, Forensic Sci Int Genet, 1 (2007) 13-19.

[14] H.J. Bandelt, W. Parson, Consistent treatment of length variants in the human mtDNA control region: a reappraisal, Int J Legal Med, 122 (2008) 11-21.

[15] A. Brandstätter, C.T. Peterson, J.A. Irwin, S. Mpoke, D.K. Koech, W. Parson, T.J. Parsons, Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database, Int J Legal Med, 118 (2004) 294-306.

[16] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, Nat Genet, 23 (1999) 147-147.

[17] N. Huber, W. Parson, A. Dür, Next generation database search algorithm for forensic mitogenome analyses, Forensic Sci Int Genet, 37 (2018) 204-214.

[18] M. van Oven, M. Kayser, Updated Comprehensive Phylogenetic Tree of Global Human Mitochondrial DNA Variation, Hum Mutat, 30 (2009) E386-E394.

[19] S. Pasino, S. Caratti, M. Del Pero, A. Santovito, C. Torre, C. Robino, Allele and haplotype diversity of X-chromosomal STRs in Ivory Coast, International Journal of Legal Medicine, 125 (2011) 749-752.

[20] B. Zimmermann, A.W. Röck, A. Dür, W. Parson, Improved visibility of character conflicts in quasi-median networks with the EMPOP NETWORK software, Croat Med J, 55 (2014) 115-120.
[21] M. Bodner, J.A. Irwin, M.D. Coble, W. Parson, Inspecting close maternal relatedness: Towards better mtDNA population samples in forensic databases, Forensic Sci Int Genet, 5 (2011) 138-141.
[22] D. Kling, A.O. Tillmar, T. Egeland, Familias 3-Extensions and new functionality, Forensic Sci Int Genet, 13 (2014) 121-127.
[23] K. Lange, J.C. Papp, J.S. Sinsheimer, R. Sripracha, H. Zhou, E.M. Sobel, Mendel: the Swiss army knife of genetic analysis programs, Bioinformatics, 29 (2013) 1568-1570.
[24] J.L. Saunier, J.A. Irwin, K.M. Strouss, H. Ragab, K.A. Sturk, T.J. Parsons, Mitochondrial control region sequences from an Egyptian population sample, Forensic Sci Int Genet, 3 (2009) E97-E103.
[25] L. Fendt, A. Röck, B. Zimmermann, M. Bodner, T. Thye, F. Tschentscher, E. Owusu-Dabo, T.M. Göbel, P.M. Schneider, W. Parson, MtDNA diversity of Ghana: a forensic and phylogeographic view, Forensic Sci Int Genet, 6 (2012) 244-249.
[26] M. Mikkelsen, L. Fendt, A.W. Röck, B. Zimmermann, E. Rockenbauer, A.J. Hansen, W. Parson, N. Morling, Forensic and phylogeographic characterisation of mtDNA lineages from Somalia, Int J Legal Med, 126 (2012) 573-579.
[27] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Mol Ecol Resour, 10 (2010) 564-567.
[28] M. Stoneking, D. Hedgecock, R.G. Higuchi, L. Vigilant, H.A. Erlich, Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes, Am J Hum Genet, 48 (1991) 370-382.
[29] A. Salas, M. Richards, T. De la Fe, M.V. Lareu, B. Sobrino, P. Sanchez-Diz, V. Macaulay, A. Carracedo, The making of the African mtDNA landscape, Am J Hum Genet, 71 (2002) 1082-1111.
[30] A. Olivieri, A. Achilli, M. Pala, V. Battaglia, S. Fornarino, N. Al-Zahery, R. Scozzari, F. Cruciani, D.M. Behar, J.M. Dugoujon, C. Coudray, A.S. Santachiara-Benerecetti, O. Semino, H.J. Bandelt, A. Torroni, The mtDNA legacy of the Levantine early Upper Palaeolithic in Africa, Science, 314 (2006) 1767-1770.
[31] V. Fernandes, F. Alshamali, M. Alves, M.D. Costa, J.B. Pereira, N.M. Silva, L. Cherni, N. Harich, V. Cerny, P. Soares, M.B. Richards, L. Pereira, The Arabian cradle: mitochondrial relicts of the first steps along the southern route out of Africa, Am J Hum Genet, 90 (2012) 347-355.
[32] F. Gandini, A. Achilli, M. Pala, M. Bodner, S. Brandini, G. Huber, B. Egyed, L. Ferretti, A. Gomez-Carballa, A. Salas, R. Scozzari, F. Cruciani, A. Coppa, W. Parson, O. Semino, P. Soares, A. Torroni, M.B. Richards, A. Olivieri, Mapping human dispersals into the Horn of Africa from Arabian Ice Age refugia using mitogenomes, Sci Rep, 6 (2016) 25472.
[33] E. Pennarun, T. Kivisild, E. Metspalu, M. Metspalu, T. Reisberg, J.P. Moisan, D.M. Behar, S.C. Jones, R. Villems, Divorcing the Late Upper Palaeolithic demographic histories of mtDNA haplogroups M1 and U6 in Africa, BMC Evol Biol, 12 (2012) 234.
[34] A.M. Gonzalez, J.M. Larruga, K.K. Abu-Amero, Y. Shi, J. Pestano, V.M. Cabrera, Mitochondrial lineage M1 traces an early human backflow to Africa, BMC Genomics, 8 (2007) 223.

[35] R.S. Just, M.K. Scheible, S.A. Fast, K. Sturk-Andreaggi, A.W. Röck, J.M. Bush, J.L. Higginbotham, M.A. Peck, J.D. Ring, G.E. Huber, C. Xavier, C. Strobl, E.A. Lyons, T.M. Diegoli, M. Bodner, L. Fendt, P. Kralj, S. Nagl, D. Niederwieser, B. Zimmermann, W. Parson, J.A. Irwin, Full mtGenome reference data: development and characterization of 588 forensic-quality haplotypes representing three U.S. populations, Forensic Sci Int Genet, 14 (2015) 141-155.

[36] M. Bodner, A. Iuvaro, C. Strobl, S. Nagl, G. Huber, S. Pelotti, D. Pettener, D. Luiselli, W. Parson, Helena, the hidden beauty: Resolving the most common West Eurasian mtDNA control region haplotype by massively parallel sequencing an Italian population sample, Forensic Sci Int Genet, 15 (2015) 21-26.

[37] J.A. Irwin, J.L. Saunier, H. Niederstätter, K.M. Strouss, K.A. Sturk, T.M. Diegoli, A. Brandstätter, W. Parson, T.J. Parsons, Investigation of heteroplasmy in the human mitochondrial DNA control region: a synthesis of observations from more than 5000 global population samples, J Mol Evol, 68 (2009) 516-527.

[38] B. Zimmermann, A. Röck, G. Huber, T. Krämer, P.M. Schneider, W. Parson, Application of a west Eurasian-specific filter for quasi-median network analysis: Sharpening the blade for mtDNA error detection, Forensic Sci Int Genet, 5 (2011) 133-137.

**Figure and table captions**

**Figure 1: Origin of sample sets used in this study.** The maps shows the continent of Africa (excluding islands) and its political borders. Samples from countries with assigned codes have been used. Novel datasets (bold codes) were included from Côte d'Ivoire (**CIV**), Rwanda (**RWA**) and twelve countries of West Africa (**mixed WAF**), viz. Benin (BEN), Burkina Faso (BFA), Cameroon (CMR), Côte d'Ivoire (CIV), Gambia (GMB), Ghana (GHA), Guinea (GIN), Liberia (LBR), Niger (NER), Nigeria (NGA), Senegal (SEN), and Togo (TGO). Published datasets used for comparisons (country codes underlined): Egypt (EGY) [24], Ghana (GHA) [25], Kenya (KEN) [15], and Somalia (SOM) [26].

**Table 1: Forensic parameters of three novel population datasets from West Africa, Côte d'Ivoire and Rwanda.**
*Footnote*: MtDNA range considered: nps 16024-16193 16194-309 310-573 574-576. The Rwanda dataset is listed as complete (RWA_all) and minimal (RWA_min) dataset.

**Forensic parameters of three novel population datasets. See text for details.**

| | West Africa | Côte d'Ivoire | West Africa + Côte d'Ivoire | Rwanda (full dataset) | Rwanda (minimal dataset) |
|---|---|---|---|---|---|
| **No. of samples** | 145 | 100 | 245 | 153 | 105 |
| **No. of haplotypes** | 132 | 91 | 211 | 105 | 105 |
| **No. of unique haplotypes** | 121 | 83 | 186 | 81 | 105 |
| **Power of discrimination [%]** | 99.9 | 99.8 | 99.8 | 99.1 | 100.0 |
| **Mean no. of pairwise differences** | 16.02 ± 7.18 | 14.09 ± 6.37 | 15.26 ± 6.84 | 18.95 ± 8.44 | 20.02 ± 8.92 |
| **RMP [%]** | 0.79 | 1.20 | 0.57 | 1.59 | 0.95 |