Genetic Evidence of African Slavery at the Beginning of the Trans-Atlantic Slave Trade

Rui Martiniano1, Catarina Coelho2, Maria Teresa Ferreira2,3,5, Maria João Neves2,4,5, Ron Pinhasi6,7 & Daniel G. Bradley1

1Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Ireland, 2Life Sciences Department, University of Coimbra, Coimbra, Portugal, 3Forensic Sciences Centre, Coimbra, Portugal, 4Dryas Arqueologia Lda., Coimbra, Portugal, 5Centro de Investigação em Antropologia e Saúde, Coimbra, Portugal, 6School of Archaeology, University College Dublin, Dublin, Ireland, 7UCD Conway Institute of Biomolecular and Biomedical Research, University College Dublin, Dublin, Ireland.

An archaeological excavation in Valle da Gafaria (Lagos, Portugal), revealed two contiguous burial places outside the medieval city walls, dating from the 15th–17th centuries AD: one was interpreted as a Leprosarium cemetery and the second as an urban discard deposit, where signs of violent, unceremonious burials suggested that these remains may belong to slaves captured in Africa by the Portuguese. We obtained random short autosomal sequence reads from seven individuals: two from the latter site and five from the Leprosarium and used these to call SNP identities and estimate ancestral affinities with modern reference data. The Leprosarium site samples were less preserved but gave some probability of both African and European ancestry. The two discard deposit burials each gave African affinity signals, which were further refined toward modern West African or Bantu genotyped samples. These data from distressed burials illustrate an African contribution to a low status stratum of Lagos society at a time when this port became a hub of the European trade in African slaves which formed a precursor to the transatlantic transfer of millions.
analysed here may be among the earliest victims of a tragic commerce that subsequently amplified to millions of forced transatlantic transfers.

Results

Sequencing results. We extracted DNA from nine bone samples from skeletons exhumed from Valle da Gafaria site in Lagos, Portugal. DNA extracts were then incorporated into NGS libraries, amplified with distinct indexes and pooled in equimolar content with 18 samples from other experiments. A partial MiSeq run yielded ~5.9 million reads containing indexes corresponding to the libraries prepared with the samples from the UDD and the Leprosarium cemetery. We trimmed adapter sequences and aligned the reads using BWA14 to the Human Reference Genome (hg19), filtering for base quality (q ≥ 15) and mapping quality (q ≥ 30), removing paralogs and selecting only uniquely aligned reads. We removed duplicate reads, which consisted of less than 0.25% across all samples. After read filtration we kept ~200 thousand reads mapping to the human reference genome in total.

The bone samples with highest endogenous human DNA content were Individuals 125 and 166 (7.6% and 9.99% respectively), both from the UDD site, and five other samples yielded human DNA at levels around 1% (0.57–1.88%) (Table 1). The lowest endogenous contents observed were for Individuals 25 and 65 (UDD), with 0.10% and 0.17%, and because of reduced numbers of reads after filtration (799 and 1657 respectively) these samples were removed from further analyses.

Cytosine deamination patterns. Cytosine deamination at the 5′-end of DNA fragments (leading to C to T changes) is a signal of postmortem chemical degradation suggesting authenticity in ancient DNA sequence6 and has been shown to increase with postmortem chemical degradation suggesting authenticity in of DNA fragments (leading to C to T changes) is a signal of.

| Burial Context | Sample | Bone sampled | Total Reads | Hits to hg19* | Fraction of Human Reads (%) | Genome Coverage [x] | Sex Determination** | Genotyped SNPs* |
|----------------|--------|--------------|-------------|--------------|----------------------------|---------------------|-----------------|-----------------|
| Leprosarium    | Ind. 5 | metatarsal    | 703096      | 10740        | 1.53                       | 0.00018             | XX              | 525             |
| Leprosarium    | Ind. 7 | metatarsal    | 866232      | 16300        | 1.88                       | 0.00027             | XX              | 762             |
| Leprosarium    | Ind. 34| Left Fibula   | 616433      | 3800         | 0.62                       | 0.00006             | XX              | 191             |
| Leprosarium    | Ind. 36| metatarsal    | 595411      | 9502         | 1.6                        | 0.00016             | XY              | 390             |
| Leprosarium    | Ind. 37| 4th lumbar vertebra | 1459685  | 8255         | 0.57                       | 0.00014             | XX              | 383             |
| UDD            | Ind. 25 | metatarsal    | 766371      | 799          | 0.1                        | 0.00001             | n/a             | 32              |
| UDD            | Ind. 65 | metatarsal    | 960013      | 1657         | 0.17                       | 0.00002             | n/a             | 60              |
| UDD            | Ind. 125 | metatarsal   | 597681      | 45414        | 7.6                        | 0.00077             | XX              | 2131            |
| UDD            | Ind. 166 | metatarsal   | 1076124     | 107507       | 9.99                       | 0.00182             | XX              | 4961            |

*Reads were filtered by base quality (q ≥ 15) and mapping quality (q ≥ 30). Duplicate reads were excluded and only unambiguously mapped reads were kept.

**Sex determination method11. Results shown in more detail in Supplementary Table S1. Individuals removed from posterior analysis because of insufficient endogenous DNA read number.

Sex estimation. Sex estimation can sometimes be crucial to understand certain archaeological contexts and it can be a very challenging task for anthropologists, especially when dealing with young individuals, bone degradation or absence of more sexually dimorphic bones such as the pelvis. A recently published method11, calibrated with modern and ancient sequence data, has shown that it is possible to confidently obtain this information by estimating the fraction of reads that map to the X- and Y-chromosomes. We applied this method to our data (Figure 2 and Supplementary Table S1) and determined that all samples were female, with the exception of Individual 36. Confidence intervals of sex determination for Sample 5 and 36 overlap the stringent boundaries of certainty (Female if Ry > 0.016, Male if Ry < 0.075, represented by gray shaded areas), but more sequence data would likely reveal that these assignments are correct.

Population genetics analysis. Despite the very low coverage (less than 0.002% of the genome, Table 1) across all samples, we still obtained sufficient overlap with known SNP positions genotyped in a range of worldwide population samples, allowing us to perform principal component analysis (PCA) to assess population diversity.
areas represent threshold for acceptance of assignment, calibrated with the number reads aligned to the Y- and X-chromosomes. Gray shaded regions represent threshold for acceptance of assignment, calibrated with the number reads aligned to the Y-chromosome divided by the sum of reads aligned to all autosomes. The principal component analysis defines loose population clusters that correspond to three African groups: North African Mozabites; Khoisan and Pygmy populations; and a grouping of Niger-Kordofanian-speaking populations such as those included Kenyan and South African Bantu populations plus West African Yoruba and Mandenka samples. When performing ADMIXTURE analysis, although a lower cross-validation error is obtained at \( k = 2 \) (Supplementary Figure S5b), this affinity becomes evident at \( k = 3 \) (Figure 3d). Separately displayed PCA (Supplementary Figure S3) and ADMIXTURE analyses (Supplementary Figure S4) for Leprosarium and UDD samples merged with the HGDP dataset are consistent.

**Investigation of modern human contamination.** We investigated the possibility of population affinity signals resulting from modern human contamination using PMDtools\(^6\) to compare analysis using only sequence reads with evidence of deamination (a modification prevalent in ancient DNA) to those of unfiltered data. In PCAs, as a result of the data reduction by filtration for deamination, complete separation of African and European reference clusters was not apparent, preventing confirmation of ancestral affinity for the less-preserved samples from the Leprosarium. Therefore, inference from these samples must carry a degree of caution. However, this analysis did have sufficient resolution to confirm African ancestry of samples 125 and 166 from the discard deposit (Supplementary Figure S6). Attribution of female sex to samples 7, 125 and 166 using only deaminated reads agreed with the previous analysis (Comparison of both analysis is shown in Supplementary Figure S7 and values obtained are shown in Supplementary Table S2). All other samples were left with under 1000 filtered reads which were insufficient to provide accurate estimates.

**mtDNA haplogroup assignment.** After aligning the filtered Next-Generation Sequencing Reads to the revised Cambridge Reference Sequence (rCRS)\(^17\), we obtained a low number of reads covering informative sites and we were not able to identify complete haplotypes. Nonetheless, we observe certain mutations in our data that point to the probability of African mtDNA lineages being present in the five samples (Supplementary Table S3) that have shown a greater affinity with African Populations in PCA and ADMIXTURE analysis. These were L3 sub lineages in samples 166 and 36; L3‘4’6 in 37; L2b1 in sample 125 and L1b1 in sample 7.

**Discussion**

In this work we use low coverage next-generation sequencing data to address common issues encountered in archaeological contexts: sex estimation, sample preservation and population affinity. By pooling several samples in the same sequencing run and obtaining just a few tens of thousands sequencing reads per sample, we have estimated the sex of seven out of nine samples sequenced. Furthermore, we characterized cytosine deamination patterns within the data and conclude that they are broadly consistent with those expected from historical bone samples. The modest SNP data obtained for these samples was sufficient to perform PCA and ADMIXTURE analyses, tentatively suggesting a diversity of ancestral backgrounds in the Leprosarium cemetery, with two individuals presenting a greater similarity with modern European (5 and 34) and three with a degree of African ancestry (7, 36, 37). The latter seems less likely to have resulted from post-mortem contamination but the limited extent of these data were such that this may not formally be excluded. The superior preservation of the two African individuals (125 and 165) from the urban discard deposit burial site and resistance of their genetic signals to restriction of the data set to reads with showing deaminations encouraged investigation of their origins at a finer detail, showing affinity with Bantu-speaking groups and Western African Mandenka and Yoruba populations, as expected from historical records. A previous study\(^6\) has shown that African chromosomal segments in African-American individuals were most similar to Niger-Kordofanian-speaking populations such as those included here. Here, we did not achieve sufficient resolution in PCA and
Figure 3 | (A) Procrustes transformation of Principal Component Analysis combining UDD (lozenges in red) and Leprosarium (lozenges in yellow) samples merged with selected African, European and Asian populations from the 1000 Genomes Omni dataset. (B) ADMIXTURE Plot of the same data used for the PCA assuming k = 3. Population key: ACB, African Caribbeans in Barbados; ASW, Americans of African ancestry in SW USA; CEU, Utah residents (CEPH) with Northern and Western European ancestry; CHB, Han Chinese in Beijing, China; FIN, Finnish in Finland; GBR, British in England and Scotland; IBS, Iberian population in Spain; LWK, Luhya in Webuye, Kenya; TSI, Toscani in Italy; YRI, Yoruba in Ibadan, Nigeria. (C) Procrustes transformation of Principal Component Analysis combining both individuals from UDD (125 and 166, lozenges in red) merged with African samples from the HGDP dataset. (D) ADMIXTURE Plot of the same data used for the PCA assuming k = 3.
ADMIXTURE to distinguish between Bantu and non-Bantu Niger-Kordofanian populations. Mitochondrial DNA coverage was insufficient to ascertain complete haplotypes, but the base calls at polymorphisms that define sub lineages of haplogroups L1, 2 and 3 corroborate PCA and ADMIXTURE results and are compatible with the lineages identified in a previous study in an African slave cemetery in Brazil19. Future studies addressing the Atlantic Slave Trade would benefit immensely from publicly available dense genotypes or high coverage genomes from regions where the trade was most intense, such as Mauritania, Guinea, Senegal, Sierra Leone, Gambia, Angola and Mozambique. The evidence of African ancestry presented here, coupled with archaeology supporting these origins, violence and distressed burial suggests that the urban discard deposit human remains belonged to slaves brought to Maritime 15th/17th century Portugal; an early snapshot of the ignominious commerce that will become the Atlantic slave trade.

Methods

Sample preparation and DNA extraction. Sample cleaning, drilling, extraction and library preparation were performed in clean-room facilities at the Ancient DNA lab, Smurfit Institute, Trinity College Dublin (Ireland) which is exclusively dedicated to aDNA sample processing. Blank controls were incorporated in extractions, library preparation and PCR reactions to monitor the possibility of contamination. Bones were demineralized by UV light exposure and by removing its surface using a Dremel drill. The densest portion for the different bones was sawed off using a circular saw, and drilled until becoming powder. Extractions of 0.2 grams of bone preparation and PCR reactions to monitor the possibility of contamination. Bones library preparation were performed in clean-room facilities at the Ancient DNA lab, with AccuPrime Pfx Polymerase (Invitrogen) and a different indexing polymerase by incubating the libraries for 20 minutes at 80°C. Adaptors with the following changes: we used T4 DNA polymerase buffer (Thermo Scientific) instead of Tango buffer in the Blunt-End Repair step: replaced Solid Phase Reversible Immobilization (SPRI) purification with Minelute Purification, and instead of the final purification after Adapter Fill-in we heat inactivated Bst Polymerase by incubating the libraries for 20 minutes at 80°C.

Libraries were amplified for 12 cycles in a separate room used only for PCR of ancient DNA, with AccuPrime Pfx Polymerase (Invitrogen) and a different indexing oligo for each sample so that multiple samples can be pooled and sequenced in the same sequencing run. With the purpose of screening for ancient DNA content, we pooled 27 samples of different origins and time periods, 9 of which belonged to the burial site in Lagos, Portugal. The resulting PCR product was purified with Minelute spin columns and eluted in 20 ul EB buffer. After DNA quantification using Quant-iT dsDNA HS Assay kit (Invitrogen, Oregon, USA) and with Agilent 2100 Bioanalyzer High Sensitivity DNA kit. Samples were pooled in equimolar concentrations by averaging the values obtained by both quantification methods and sequenced for 50 cycles, single-end reads mode, in a Illumina MiSeq instrument located in the Institute of Molecular Medicine (IMM), Trinity College Dublin, along with PhIX control at 1%.

Read alignment and filtering. Cutadapt23 was used to trim adaptor sequences present in the raw reads. These reads were then aligned to the human reference genome (UCSC hg19) using BWA5. The alignment parameters were set in a way that low quality bases ("q15") were trimmed from the 3’ end of the reads and seeding was disabled to improve accuracy ("-1 1024"). We then excluded reads with mapping quality inferior to 30 ("q 30") and removed duplicates using SAMtools22. We also selected reads containing the SAMtools flag X1 to filter for paralogs. The resulting bam file containing only confidently aligned reads was used to estimate the percentage of endogenous DNA present in the ancient samples by comparing the fraction of mapped reads with the total number of reads obtained with same barcode (Table 1).

Sample preservation and deamination patterns. To compare cytosine deamination patterns in the samples analysed in the present work, we downloaded sequence reads of a ~7,000-year-old Scandinavian Hunter-Gatherer20 and a 100-year-old Australian Aboriginal21 and aligned to the hg19 human reference genome with the same filtering criteria used for our data. After downsampling aligned reads to 100,000 (to match the number of aligned reads in Individual 166), we characterized cytosine deamination patterns using PMDTools and the results were plotted with a customized PMDTools R script.

Sex determination. A recently developed method for sex determination using NGS reads18 was employed in our samples using confidently aligned reads as a reference. Results are presented in Figure 2.

Principal component analysis. In order to compare our ancient samples to datasets of modern human populations, we identified bases in known SNP positions using Genome Gene Tool Kit (GATK) in Picard mode by providing an interval file ("-l stranded") for each modern human genotype dataset. Specifically, we used the 1000 Genomes dataset (ftp://1000genomes.ebi.ac.uk/vol1/ftp/technical/workings/20120131_omni_genotypes_and_intensities/) and genotypes from the Human Genome Diversity Project (HGDP; http://www.hgsnp.org/hgdp/) flipped to hg19 strand orientation. For Principal Component and admixture analysis we filtered our data in a similar way as described previously24. Briefly, we only included reads with bases in SNPs and low mapping quality of a base below 10 and 30, respectively. Potentially false mutations that may have been originated by cytosine deamination (C to T and G to A) were excluded from analysis and SNP data was converted to PLINK format files25. Because of the very low coverage of the data obtained, the vast majority of positions in the genome are covered by a single read only and, therefore, all genotypes were converted to binary. In the case of a chromosomal position being covered by more than one read only a nucleotide base was randomly chosen and included in the analysis. Likewise, and to avoid any bias in population variability, all heterozygous positions in human diversity datasets were converted to homoygous by randomly picking one of the alleles. PLINK was used to merge the ancient samples genotype data with Omni and HGDP datasets separately. Principal Component Analysis was done using SMARTPCA2, removing SNPs in LD with nearby SNPs (r squared > 0.2). We used the R package "vegan" (http://cran.r-project.org/web/packages/vegan/index.html) to do Procrustes transformation on the Principal Component coordinates. Because of the high-density genotyping in the Omni dataset, we were able to retrieve sufficient positions to perform Procrustes transformation on all ancient samples (Figure 3a).

Results are presented in Figure 2. For the HGDP dataset, the number of SNPs called is too low in the majority of samples to provide a clear clustering between populations. Therefore, for Procrustes transformation of the HGDP merged with historical DNA genotypes, we picked our two best samples in terms of endogenous DNA and number SNP positions identified (166 and 125) (Figure 3c). Principal Component Analysis was plotted with R version 2.14.3.

Estimation of modern human contamination. To access whether our data is being affected by modern human contamination, we used PMDTools to extract reads with evidence of deamination (PMD score 3) and compared Principal Component Analysis and sex determination using these reads with unfiltered data (PMD score -2).

Model-based clustering. Using the aforementioned filtered datasets, we performed ADMIXTURE13 runs for values of k (ancestral populations) ranging from 2 to 10. Each run of ADMIXTURE for k = 2 to 10 has a coefficient variation (CV) value associated to it. CLUMPP24 version 1.1.2 was used to average clustering results between samples and obtain the best configuration of clusters across datasets with the fullSearch algorithm. Clustering results were visualised with distruct27 and are presented in Figure 3b and 3d.

Tentative mitochondrial DNA haplogroup identification. In order to identify mitochondrial DNA (mtDNA) haplogroups, we selected reads at least 25 bp long and used seqkit (https://github.com/lh3/seqkit) to trim the first and last 3 bp in sequencing reads and to deamination the effect of deaminations and select only those reads that were used to estimate the revised Cambridge Reference Sequence (cRS) and filtered as above. The final bam file was uploaded to MitoBamAnnotator28 which identifies mtDNA mutations which are then analysed with HaplOgrep29 identifying the most likely haplogroup to which a sample belongs.

1. Neves, M., Almeida, M. & Ferreira, M. Separados na vida e na morte: reato do tratamento mortuorio dado aos escravos africanos na cidade moderna de Lagos. XELB Rev. 10, 547–560 (2010).
2. Neves, M., Almeida, M. & Ferreira, M. História de um arrabalde durante os séculos XV e XVI: o "poço dos negros" em Lagos (Algarve, Portugal) e o seu contributo para o estudo dos escravos em Portugal, in A Herança do Infante 29–46 (2011).
3. Ferreira, M. T., Neves, M. J. & Wastlaperin, S. N. S. Lagos leprosarium (Portugal): evidences of disease. J. Archaeol. Sci. 40, 2298–2307 (2013).
4. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. (2010).
5. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–60 (2009).
6. Broherton, P. et al. Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem base conversions in ancient DNA. Proc. Natl. Acad. Sci. 131, 1318934111–(2014). doi:10.1073/pnas.1318934111.
7. Sawyer, S., Krause, J., Guschnoki, K., Savolainen, V. & Paabo, S. Temporal patterns of nucleotide miscorpirations and DNA fragmentation in ancient DNA. PLoS One 7, (2012).
8. Skoglund, P. et al. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proc. Natl. Acad. Sci. 131, 5717–28 (2007).
9. Rasmussen, M. et al. An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. Sci. 334, 94–98 (2011).
10. Skoglund, P. et al. Origins and Genetic Legacy of Neolithic Farmers and Hunter-Gatherers in Europe. Science (80-) 336, 466–469 (2012).
et al

26. Purcell, S. et al. The Sequence Alignment/Map (SAM) Format and SAMtools 1000 Genome Project Data Processing Subgroup. Bioinformatics 25, 2078–2079 (2009).

27. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* 2, 2074–2093 (2006).

28. R Development Core Team, R. R: A Language and Environment for Statistical Computing. *R Found. Stat. Comput.* 1, 409 (2011).

29. Zhidkov, I., Nagar, T., Mishmar, D. & Rubin, E. MitoBamAnnotator: A web-based tool for detecting and annotating heteroplasmy in human mitochondrial DNA sequences. *Mitochondrion* 11, 924–928 (2011).

30. Kloss-Brandstätter, A. et al. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum. Mutat.* 32, 25–32 (2011).

Acknowledgments

We thank Matthew Teasdale for comments on the manuscript, Cristina Gamba for discussion, Eppie Jones and Russell McLaughlin for the HGDP dataset preparation. Valeria Mattiageli helped generating the MiSeq data. This study has been financed by the BEAN project of the Marie Curie ITN (grant number 289966).

Author contributions

R.M., D.B. and R.P. conceived and designed the study. R.M. performed the experiments and data analysis. M.T.F. and M.J.N. excavated the skeletal human remains; M.T.F., C.C. and M.J.N. made the anthropological study of the skeletons. All authors contributed to the manuscript preparation.

Additional information

Accession codes: Sequence reads were uploaded to the Sequence Read Archive at the ENA (European Nucleotide Archive), study accession PRJEB6056. Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Martiniano, R. et al. Genetic Evidence of African Slavery at the Beginning of the Trans-Atlantic Slave Trade. *Sci. Rep.* 4, 5994; DOI:10.1038/srep05994 (2014).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/