

This is a repository copy of *4000-year-old hair from the Middle Nile highlights unusual ancient DNA degradation pattern and a potential source of early eastern Africa pastoralists*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/194261/>

Version: Published Version

Article:

Wang, Ke, Bleasdale, Maddy, Le Moyne, Charles et al. (4 more authors) (2022) 4000-year-old hair from the Middle Nile highlights unusual ancient DNA degradation pattern and a potential source of early eastern Africa pastoralists. *Scientific Reports*. ISSN 2045-2322

<https://doi.org/10.1038/s41598-022-25384-y>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>



Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



OPEN

4000-year-old hair from the Middle Nile highlights unusual ancient DNA degradation pattern and a potential source of early eastern Africa pastoralists

Ke Wang^{1,2}, Madeleine Bleasdale^{3,4}, Charles Le Moyne^{4,5}, Cacia Freund¹, Johannes Krause¹, Nicole Boivin^{4,5} & Stephan Schiffels¹

Petrous bones and teeth are the skeletal elements most often targeted by researchers for ancient DNA (aDNA) extraction, and the sources of the majority of previously published ancient African genomes. However, the high temperature environments that characterise much of Africa often lead to poor preservation of skeletal remains. Here, we successfully reconstruct and analyse genome-wide data from the naturally mummified hair of a 4000-year-old individual from Sudan in northeastern Africa, after failed attempts at DNA extraction from teeth, petrous, and cranium of this and other individuals from the Kadruka cemeteries. We find that hair DNA extracted with an established single-stranded library protocol is unusually enriched in ultra-short DNA molecules and exhibits substantial interior molecular damage. The aDNA was nonetheless amenable to genetic analyses, which revealed that the genome is genetically indistinguishable from that of early Neolithic eastern African pastoralists located 2500 kms away. Our findings are consistent with established models for the southward dispersal of Middle Nile Valley pastoral populations to the Rift Valley of eastern Africa, and provide a possible genetic source population for this dispersal. Our study highlights the value of mummified hair as an alternate source of aDNA from regions with poor bone preservation.

The first complete ancient human mitochondrial genome, published in 2008, was generated from the frozen hair of an individual of the Saqqaq archaeological culture from Greenland, dated to 4750 years ago¹. Since then, the field of ancient DNA (aDNA) research has grown rapidly, with over 6000 ancient genomes published to date. This extraordinary growth has largely been based on analyses of ancient human bones and teeth, with their preservation widely regarded as superior to that of soft tissue and hair². Meanwhile, very few studies have attempted to extract aDNA from hair or soft tissue^{1,3}, especially from samples of significant age.

While only accounting for 1.7% of all published genomes, Africa is no exception to this trend. So far, aDNA studies focused on Africa have reported nearly 100 ancient whole genomes, reconstructed from aDNA preserved in petrous bones and teeth from sub-Saharan Africa. Over half of these studies focused on understanding the spread of pastoralism in ancient eastern Africa^{4,5}, in particular drawing on samples from present-day Kenya and Tanzania. aDNA studies have revealed multiple waves of human migrations associated with the spread of pastoralism, and complex subsequent admixture events with local hunter-gatherers in eastern Africa^{4,5}. Although there is broad consensus that early pastoralists in eastern Africa originated from population dispersals from the north, from regions such as Sudan and Ethiopia³⁻⁷, no aDNA data related to early pastoralists from these regions has yet been published.

Here, we extract DNA from five specimens (including petrous, teeth and hair) deriving from four individuals from middle Holocene cemeteries in the Dongola Reach of Upper Nubia, northern Sudan. Selected samples for this study represent three individuals from the early 7th millennium BP Neolithic site of Kadruka 21, and one Kerma period individual, directly dated to 4033 cal. BP, from Kadruka 1. Of all the tissues and samples tested,

¹Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. ²School of Life Sciences, Fudan University, Shanghai, China. ³Department of Archaeology, University of York, York, UK. ⁴Max Planck Institute for the Science of Human History, Jena, Germany. ⁵School of Social Science, The University of Queensland, Brisbane, Australia. ✉email: ke_wang@eva.mpg.de; stephan_schiffels@eva.mpg.de

only one sample of preserved hair, from the Kerma period individual, yielded a sufficient amount of authentic ancient human DNA to study. We argue that a critical prerequisite in this extraction was the first application of a state-of-the-art single-stranded library laboratory protocol on hair material, which is able to turn single-stranded DNA molecules into a sequencing library⁸. As we show in our analysis, sequence data generated from hair aDNA using this single-stranded library protocol features unique characteristics that pose challenges in downstream genetic analyses.

Our population genetic analysis of this hair sample from northern Sudan reveals a close genetic affinity of the Kerma period individual to early pastoralists from the Rift Valley in eastern Africa. Drawing on this evidence, as well as previously published proteomic evidence of dietary proteins extracted from the dental calculus of this individual, we conclude that our sample (“Sudan_Kadruka1_4000BP”) represents one of the earliest and most northerly ancient pastoralist populations to have been genetically characterised in Africa to date, as well as potentially a genetic source population for one of the earliest prehistoric pastoralist dispersals into eastern Africa.

Results

Archaeological context of Kadruka 1 SK68. Archaeological fieldwork in the Kadruka district of northern Sudan has revealed widespread past occupation of the alluvial plain and paleochannel systems to the east of the current Nile River channel⁹. In accordance with broader archaeological sequences for this region of the Middle Nile Valley^{10–12}, cultural deposits were primarily linked to the Neolithic (spanning the 7th millennium BP), as well as the more recent Kerma period (3450–4450 BP). Excavation programs in the Kadruka district focused on several cemeteries from these cultural periods, including Kadruka 1 and Kadruka 21. Reflecting local environmental fluctuations between dry and wet conditions associated with more intensive Nile floods during the Middle Holocene, Neolithic skeletal remains at these cemeteries are typically highly degraded^{13,14}. In contrast, progressive local aridification and floodplain contraction¹⁵, has facilitated enhanced preservation of organics, including hair and leather items, in the more recent Kerma period burials at Kadruka 1¹⁶.

A recent study exploring the diet of individuals from Kadruka 21 and Kadruka 1 detected milk proteins attributable to domesticated cow (*Bovinae*) or sheep (*Ovis*) in the dental calculus of individual SK129 from Kadruka 21 and to goat (*Capra*) in the dental calculus of individual SK68 from Kadruka 1¹⁷. Furthermore, isotopic analyses of hair from individual SK68 (named “Sudan_Kadruka1_4000BP” in this study) broadly indicate a diet primarily composed of C₃-based resources (C₃ plants or animals consuming C₃ plants, $\delta^{13}\text{C}$ -17.0‰, $\delta^{15}\text{N}$ 12.0‰).

The early proliferation of herding economies in northeastern Africa, particularly apparent in the Kerma civilisation of Upper Nubia^{11,12}, has been proposed as a potential link to the dispersal of pastoral populations into eastern Africa^{5–7}, though there is as yet no published genetic evidence to support this migration model. Direct palaeoproteomic evidence for milk consumption¹⁷, together with the remains of cattle, sheep and goat found in grave assemblages¹⁸, identify individuals from Kadruka 1 and Kadruka 21 as belonging to early populations practicing pastoralism, making the sites ideal for archaeogenetic research seeking to examine early pastoralist dispersals in northeastern Africa.

Ancient DNA authentication and sequencing strategy. We screened five specimens deriving from four individuals from the Kadruka district of northern Sudan (Fig. 1a) for aDNA preservation, drawing on tooth (KDR001.A), hair (KDR001.B), petrous bone (KDR002, KDR004) and cranial (KDR003) samples. All specimens were excavated from archaeological contexts dating to the Neolithic and Kerma periods. The only sample that yielded detectable authentic aDNA was a lock of dark hair (127 mg) (Fig. 1b) from a Kerma period individual. We used a total of 27.5 cm of hair for aDNA extraction and 78.5 mg for radiocarbon dating. The hair sample from Kadruka 1 SK68 is directly dated to 3928–4139 calBP (Fig. 1c), contemporaneous to the Pastoral Neolithic period in eastern Africa.

To maximize the possibility of aDNA retrieval from shotgun sequencing, we applied double-stranded and single-stranded library protocols to the four extracts deriving from the Kadruka skeletal remains, but only applied the single-stranded library protocol to the hair sample, since this protocol typically results in higher yields than the double-stranded library approach for highly fragmented DNA. None of the aDNA libraries extracted from the skeletal elements showed authentic aDNA damage patterns, i.e., substitutions from cytosine to thymine (C->T) (Table S1). The hair sample (KDR001.B0101), however, provided an observed 17.5% C->T substitution rate at the first 5-prime base in sequencing reads (Table S1). Accordingly, we selected this library from the hair sample for deeper sequencing.

We found that the average read length was relatively short (33 base pair/bp, Table S1). This poses a challenge, since short DNA fragments can result in spurious alignments to the human reference genome, even if they do not originate from humans, but from, for example, microbes present in the burial environment. On the other hand, while long DNA fragments have high mapping certainty, they are more likely to originate from modern human DNA contamination (which typically consists of long DNA fragments). We therefore explored various read length cut-offs to yield as much authentic human aDNA as possible, while maintaining a low proportion of reads from presumed modern human contamination. For this purpose, we used two tools to assess both the rate of spurious alignments and the rate of modern human contaminants. First, we used SpAI²⁰ which uses simulations to estimate fractions of spurious and authentic alignments given certain read length cut-offs. For a cut-off length of 25 bp, SpAI estimated a spurious alignment fraction of 10% (Table S2). As the read length cut-off increases, the estimation for spurious alignments drops respectively (Table S2). Second, we used AuthenticCT²¹ to estimate the overall contamination level in the aligned fragments using base substitution patterns. We explored length cut-offs at 10, 25, 30 (custom setting) and 34 bp in the raw data processing steps, and summarised EAGER statistics and respective contamination estimates in Table S2.

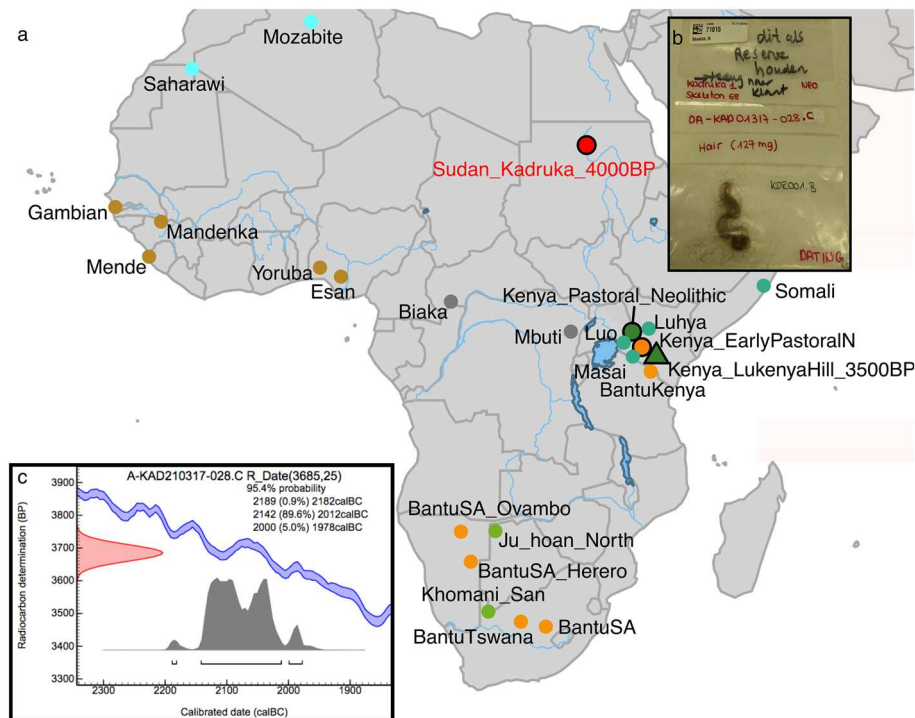


Figure 1. Sample details of the 4,000-year-old hair from Sudan. **(a)** Geographic location of Kadruka and modern African populations used for subsequent analyses. **(b)** Picture of the Kadruka hair sample. **(c)** C14 calibrated age (cal BC) of the Kadruka hair, plotted using IntCal13 calibration curve¹⁹.

Read length filter	Nr. raw reads	Nr. reads prior mapping	Nr. Mapped reads (no qc)	Nr. Mapped reads (q30)	Nr. on-target reads (1240 k)	Nr. called 1240 K SNPs	Nr. MT reads	Mt/ NUC ratio	Nr. reads on X:Y (q30)	Damage 1st base 3' %	Damage 2nd base 3' %	Damage 1st base 5' %	Damage 2nd base 5' %	Median fragment length	AuthentiCT contam %
30 (default)	99,830,433	2,606,646	97,432	52,949	1113	952	172	329.78	1632:84	0.74	0.43	16.23	11.64	32	47.3 ± 2.4
25 (preferred)	99,830,433	4,680,356	576,384	231,040	5464	3336	656	224	6581:405	1.06	1.03	15.9	11.76	26	0.1 ± 0.3

Table 1. Eager statistics of deeper shotgun sequenced data from the hair sample KDR001.B0101.SG1.2 with different length filter cut-off. We performed contamination tests using AuthentiCT, which is designed for sequences from single-stranded libraries.

We find that at a length cut-off of 30 bp, 47.3 ± 2.4% of retrieved aDNA is likely of modern-human contaminant origin (Table 1). In comparison, a length cut-off of 25 bp yielded 4,680,356 mapped reads with 0.1 ± 0.3% contamination (Table 1). Together with the results from SpAL, we consider 25 bp a safe cut-off length for this library. Thus, we continued our downstream analyses with a 25 bp read length filter, ending up with 231,040 sequencing reads after mapping, from which we derived 3,336 pseudo-haploid allele calls on 1240 k SNP positions (Table 1).

Characteristics of the aDNA fragments from hair. Employing our final read length filter at 25 bp and additionally filtering for alignment mapping quality (Methods), we further explored alignment statistics. We find two unusual characteristics in the aDNA library generated from the Kadruka hair sample. The first is that the sample is enriched in unusually short DNA molecules, giving a median read length of 25 bp, compared to 44 bp for typical bone-derived shotgun aDNA in a previous African aDNA study using the same laboratory pipeline⁴. The second is that unusually high damage rates were observed in the interior of the DNA molecules from hair, while unusually low damage rates were seen in the exterior of the molecules. For instance, at the interior 10th bp from the 5' end, damage rates were 10% compared to 1% from typical bone-derived aDNA. While at the exterior, the 1st bp from the 5' end showed damage rates of 15%, compared to on average of 27%²² from typical bone-derived aDNA. These patterns are consistent with high degradation of DNA fragments in hair through intense sun exposure, already during the lifetime of the individual, which may result in hair containing largely

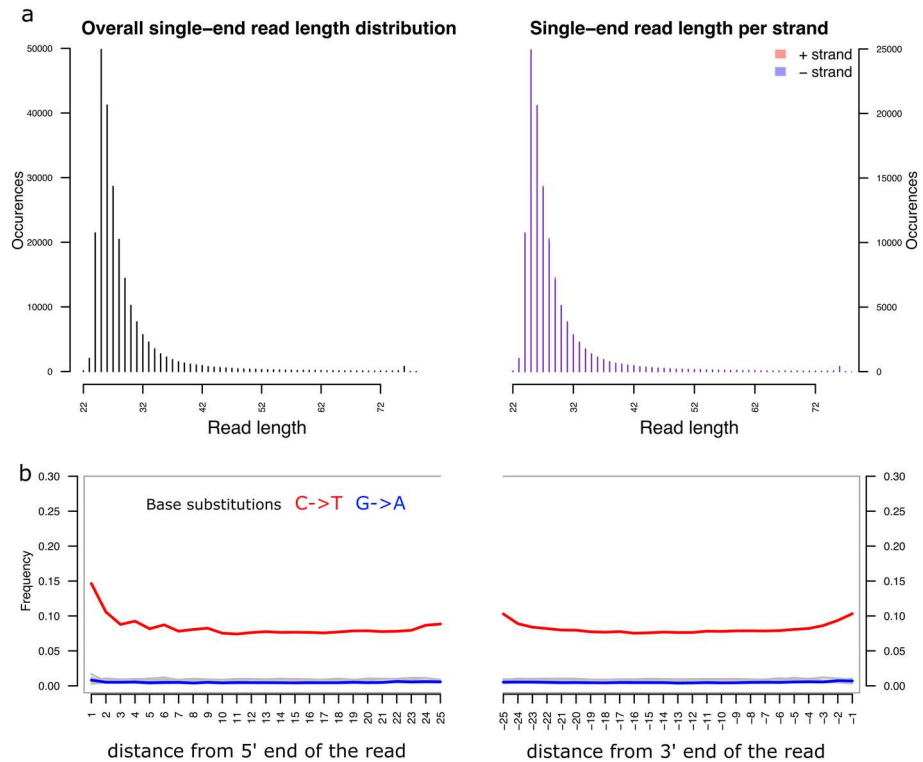


Figure 2. Characteristics of aDNA fragments from the hair sample. **(a)** Length distribution of shotgun sequencing reads mapped to the whole genome using read length filter cut-off at 25 bp in the step of adaptor removal. **(b)** High C-to-T substitution rates in the interior of aDNA fragments.

denatured single-stranded DNA fragments, as opposed to the more typically intact double-stranded fragments preserved in bone samples.

We find the ratio of mitochondrial to nuclear DNA in the hair-derived DNA library to be relatively high (with a ratio at 224 at length cut-off 25 bp, Table S1), compared to typical rates with other tissues; for example the ratio is at 110 on average for petrous bones in a previous study using the same laboratory pipeline⁴. We investigated whether there is a notable difference in terms of aDNA preservation in nuclear DNA and mitochondria from the hair material. Specifically, we examined if the two idiosyncratic features of hair aDNA we describe above apply to both nuclear and mitochondrial DNA from hair. In Fig. 2 and Fig S2, we compared the read length distribution and average base substitution rates of reads mapped to the complete genome (i.e., nuclear and mitochondrial), the nuclear genome, and the mitochondrial genome. We find that both nuclear and mitochondrial DNA have high base substitution rates in the interior of sequence reads (Fig S2a, c), but reads mapped to mitochondria are relatively longer than reads mapped to autosomes (Fig S2b, d).

Given the tenfold reduction of mapped reads after applying a mapping quality filter (Table 1), we examined if the two features of hair aDNA observed here resulted from the mapping quality filter. We find that our mapping quality filter did not have a notable effect on the two features of ultra-short DNA fragment enrichment and high interior aDNA damage pattern (Fig S3).

Given the success of SNP capture techniques for poorly preserved human DNA²³, we also performed SNP capture for our hair-derived aDNA library. However, SNP capture did not provide an improvement over shotgun sequencing. Instead, we found the base substitution rates in the exterior and interior of reads are substantially lower in capture data, in comparison to the rate distribution in shotgun data (Fig. S1, Table S1), corroborating the fact of high contamination rate in the capture data ($42 \pm 3\%$ as estimated by *AuthentiCT*), likely due to capture preferentially targeting molecules without damage (due to more effective hybridization) than with damage. In addition, longer molecules are preferentially captured over short molecules.

Genetic affinity to early eastern African pastoralists. We performed Principal Components Analysis (PCA) and Outgroup- f_3 (Figs. 2, 3) to investigate the genetic ancestry of the individual (Sudan_Kadruka1_4000BP) from whom our hair sample derived, utilizing 3336 mapped reads overlapping with SNP positions from the Shotgun data, after read length filter at 25 bp of Sudan_Kadruka1_4000BP (Table S2). To maximise the resolution given the extremely low coverage and low number of called alleles, we used high-coverage modern African genomic data from the SGDP²⁴ and the HGDP²⁵, which includes all SNPs in 1240 k panel, instead of the commonly used Human Origin array data²⁶, for calculating Principal Components (PCs). We projected ancient Africans and ancient Near Easterners on the background of modern African groups²⁵. Although the number of available populations in SGDP and HGDP is limited, we observe clear separations of

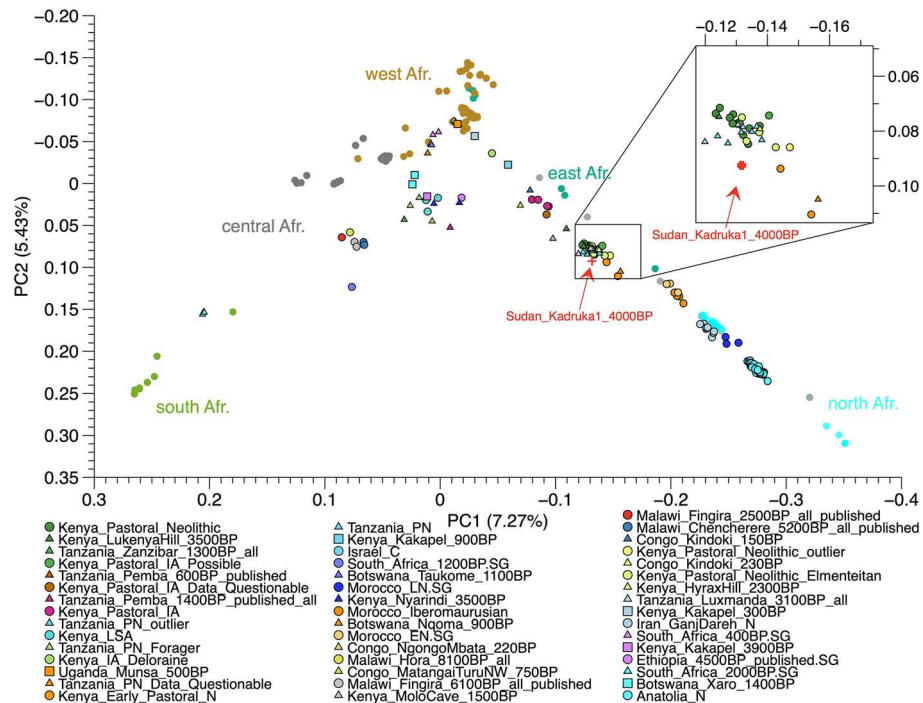


Figure 3. Genetic ancestry of Sudan_Kadruka1_4000BP. Principal Component Analysis (PCA) of African populations. We project Sudan_Kadruka1_4000BP into an African PCA (Table S3) with PCs calculated from modern Africans in SGDP²⁴. We use block jackknife strategy (taking-one-chromosome-out)²⁶ for error bar calculation of Sudan_Kadruka1_4000BP's location on the PCA.

African populations from different regions, with eastern/northern, southern and western African populations falling into the right, left and top corner of PC1/PC2 space, respectively.

The PCA shows Sudan_Kadruka1_4000BP located close to previously published early pastoralists in eastern Africa^{4,5}, such as Kenya_EarlyPastoral_N (3800–4000 calBP) and Kenya_Pastoral_Neolithic (1500–3000 calBP). Kenya_EarlyPastoral_N is a group of two pastoralist individuals dated to the early stage of the eastern African Pastoral Neolithic, both of whom are genetically derived from admixture between two early northeastern African-related ancestries from Sudan and Northern Africa/Levant⁵. To estimate the level of noise resulting from the sparsity of our SNP data, we computed a standard error for the projected PCs of Sudan_Kadruka1_4000BP using a block jackknife approach. Specifically, we computed pseudo-values by deleting each chromosome of the genotype data in turn and then used the resulting estimates from the remaining data as input for the weighted jackknife calculation²⁷. We find the standard errors of Sudan_Kadruka1_4000BP to be relatively small compared to overall genetic variation within Africa, which gives us confidence that the location calculated from the full data (Fig. 2a) is robust.

Despite the sparsity of the data, the PCA analyses conducted here clearly suggest a very close genetic relationship between Sudan_Kadruka1_4000BP and ancient eastern African pastoralist populations. To corroborate this finding, we also computed allele sharing rates with ancient populations from the Levant and Africa and present-day African populations at genomic sites where Sudan_Kadruka1_4000BP differs from the chimpanzee reference genome via outgroup- f_3 (Sudan_Kadruka1_4000BP, population X; Chimpanzee). Figure 4a shows that Sudan_Kadruka1_4000BP shares the highest genetic affinity with ancient Levantine groups, ancient northern and Eastern Africans and modern Africans from northern Sahara and the Horn of Africa. We computed pairwise comparisons employing f_4 (Sudan_Kadruka1_4000BP, Kenya_EarlyPastoral_N; population X, Chimp) to validate the close PC location between Sudan_Kadruka1_4000BP and Kenya_EarlyPastoral_N. Consistent with PCA location, f_4 -statistic result confirms the genetic cladality between Sudan_Kadruka1_4000BP and Kenya_EarlyPastoral_N given that none of the tested populations breaks cladality significantly, suggesting that these two individuals are indistinguishable in terms of allele frequencies (Fig. 4b). Additionally, we show that in f_4 (Kenya_EarlyPastoral_N, population X; Sudan_Kadruka1_4000BP, Chimp) results (Fig. 4c) that all tested ancient and modern African populations are either significantly positive (suggesting that they are less close to Kenya_EarlyPastoral_N compared to Sudan_Kadruka1_4000BP) or overlapping with zero, indicating equal genetic distance to both.

Discussion

In this study, we demonstrate the possibility of retrieving aDNA from ancient hair by reconstructing genome-wide data from a 4000-year-old hair sample in Sudan. We find that aDNA extracted from hair preserved in this extremely hot and arid environment features unusually short DNA fragments and unusually high interior aDNA damage patterns. We interpret such characteristics as a combined effect of the unique preservation environment

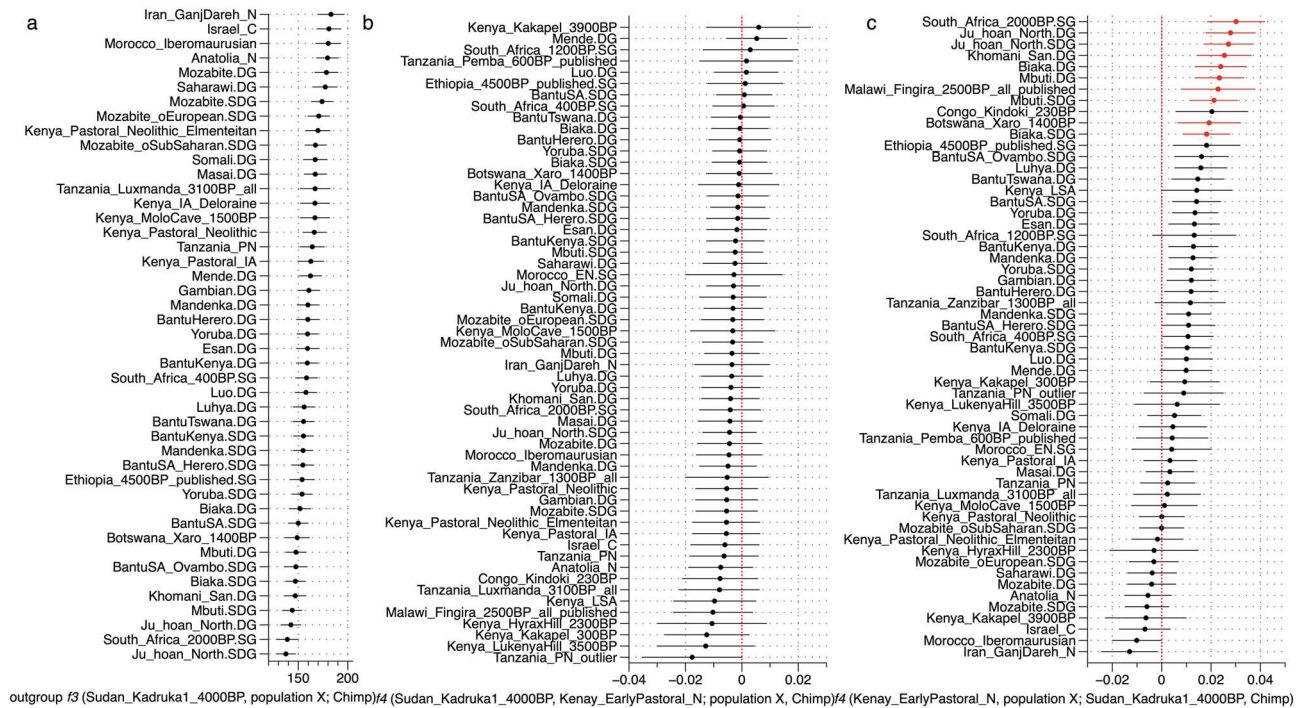


Figure 4. Genetic affinity with ancient African pastoralists. (a) f_3 (Sudan_Kadruka1_4000BP, population X; Chimp). (b) f_4 (Sudan_Kadruka1_4000BP, Kenya_EarlyPastoral_N; population X, Chimp). (c) f_4 (Kenya_EarlyPastoral_N, population X; Sudan_Kadruka1_4000BP, Chimp). Population X includes published ancient African and Near Eastern populations and modern African populations from SGDP and HGDP data sets (Table S3). We plot two standard error bars for f_3 and f_4 statistics shown here and highlight statistically significant tests (Z-score > 3) in red color.

in Sudan, of hair as a unique aDNA source, and of the single-stranded library protocol that we applied to generate our sequencing library. The high regional temperature likely leads to i) denaturation of double-stranded DNA molecules to single-stranded DNA, and ii) more frequent breakage of DNA molecules, resulting in shorter DNA fragments. Hence the predominance of short, single-stranded DNA in hair explains the excess C->T substitution rate in the interior of sequencing reads. Meanwhile, the arid preservation environment and the specific properties of hair potentially reduce enzymatic activity, resulting in reduced hydrolysis damage²⁸, preventing the commonly-observed high degradation at the 5' end of aDNA fragments and explaining the relatively low substitution rate at the 5' end of the ancient hair DNA molecules.

Given these properties, the single-stranded library protocol we employed is uniquely well-suited as a sequencing tool in this case, maximising library complexity from heavily degraded and in large parts single-stranded source DNA fragments by building library molecules from both forward and backward strand DNA fragments regardless of the DNA strand orientation. In such cases, both short and long DNA fragments are sequenced effectively. Hence, our shotgun sequencing on the single-stranded library accumulates the short DNA fragments, as reflected in the read length distribution with a peak around short read lengths, and keeps the long fragments with interior damage, as reflected in the average high interior damage with C->T substitution rate over 10% throughout the sequence read (Fig. 2).

To our surprise, only the DNA from hair, rather than from the tooth of the same individual, showed signatures of authentic human aDNA preservation, as seen from the pattern of C->T substitutions, regardless of rather similar endogenous DNA content in tooth and hair (Table S1). From the hair sample, we obtained 231,040 mapped reads after applying mapping and base quality filters on nearly 100 million raw reads (Table 1), corresponding to 0.231% endogenous DNA from shotgun whole genome sequencing. One consequence of the relatively short DNA fragments in sequencing libraries built from hair is that the commonly used 1240 k SNP capture technique failed to enrich authentic human aDNA. Instead, in our case, 1240 k capture likely enriches for long non-deaminated sequences which are non-authentic aDNA (Fig S2). Consequently, we generated and analysed whole genome shotgun data throughout this study, which is less cost-effective than capture technology.

The very low coverage data resulting from this unique sample posed a challenge for downstream population genetic analyses. Nevertheless, we see from f_3 -statistics that this individual, from a rural agro-pastoral population linked with the Kerma culture of Upper Nubia¹⁶, shares close genetic affinity with Levantine groups. Moreover, we could show that this individual is genetically indistinguishable from early Pastoral Neolithic individuals dated to 4000BP living over 2500 km away in Kenya and Tanzania, even when correcting for relatively large standard errors in population genetic estimates due to the low coverage. This close relatedness to early pastoral populations in eastern Africa is consistent with archaeological evidence for the dispersal of herding populations southwards along the Nile River Valley following their establishment in the Kadruka region from the early 7th millennium BP^{29,30}, although we caution that inference from a single sample can at best be tentative.

The high affinity of the Kerma period individual from Kadruka 1 with Neolithic pastoral groups far to the south, and genetic indistinguishability of their sequenced DNA, would be consistent with this sample representing a possible genetic source population for the earliest eastern African pastoralists who settled in the Rift Valley. This in turn would point to a high degree of mobility of pastoralists between the Middle Nile Valley and present-day Kenya, potentially before or around the Pastoral Neolithic. It would imply that the southward dispersal of pastoralists from the Middle Nile Valley did not involve genetic exchange with pre-existing human groups along the migration route, particularly local foragers, and may therefore have been relatively rapid. Indeed, archaeological research suggests that dispersals were likely driven in part by increasing population pressure and regional aridification, with a marked period of aridity in Upper Nubia preceding the first appearance of pastoral groups in the Turkana Basin of eastern Africa at 5000 BP^{29–31}. These early eastern African pastoralists at sites like Lothagam in the Turkana Basin^{29,32}, while featuring regional forms of cultural expression, also broadly maintained traditions of monumental mortuary expression, as well as traditions of burial with elaborate grave goods for which the Neolithic and Kerma period cemeteries of the Middle Nile Valley, including Kadruka 1, are recognised⁹.

While much archaeogenetic research in Africa has focused on the Neolithic populations of eastern Africa, their source populations have, from a genetic standpoint, remained elusive. Highlighting the need for additional aDNA from ancient northeastern African populations, the aDNA recovered from ancient human hair in this study provides important new genetic data relevant to the detailed reconstruction of early pastoral dispersals from northeastern to eastern Africa. Our findings also point to the value of preserved ancient hair as a source of aDNA, particularly in regions where more typical aDNA archives are unavailable or poorly preserved, with exciting implications for future aDNA research.

Methods

aDNA extraction and C14 dating on the mummified hair. We sequenced five specimens (four individuals) in total from Kadruka 21 and Kadruka 1 for aDNA in dedicated clean rooms at the Max Planck Institute for the Science of Human History in Jena, Germany. We constructed both single-stranded and double-stranded DNA libraries for skeleton remains and single-stranded library for the hair sample following previously published protocols^{8,33}, and performed both whole genome shotgun and 1240 k-hybridization capture²³ to maximise the amount of aDNA.

We sampled from a mixed composition of hair for aDNA analyses and report the direct C14 date for the hair sample from separate additional sampling. A bulk hair sample was sampled for dating at the Centre for Isotope Research (CIO) Groningen in 2019. A sample of 78.5 mg of hair was chemically treated with 4% HCl, rinsed with decarbonized H₂O and dried before combustion. Combustion and graphitization of a subsample of the pre-treated hair (5.35 mg) was carried out according to the methods outlined in Dee et al. (2020)³⁴. The ¹⁴C measurements were carried out using an Ionplus AMS-MICADAS^{35,36}. C14 ages are calibrated to calendar years with OxCal version 4.3³⁷, using calibration curve IntCal13¹⁹.

Bioinformatic processing. We processed DNA sequences using the EAGER v1.92.56³⁸, with adaptors removed by AdapterRemoval v2³⁹ using flags *AdapterRemoval -threads 4 -trimms -trimqualities -preserve5p -minlength 25 -minquality 20 -minadapteroverlap 1*, reads mapped to hs37d5 by BWA v0.7.12⁴⁰ using flags *bwa aln -t 4 -n 0.01 -l 32*, and polymerase chain reaction duplicates removed by Dedup software v0.12.2³⁸. mapDamage v2.0.9⁴¹ was used to calculate substitution rate in read termini of sequenced DNA fragments.

We estimated the amounts of spurious alignments at various read length cut-off using SpAI (bioinf.eva.mpg.de/SpAI/). We tested three different read length filters (34, 25, 10 bp) to yield the maximum efficiency for endogenous aDNA while assuring contamination free. We used AuthenticCT²¹, which is designed for single-stranded library DNA sequences, to examine the contamination level. Finally, we applied a minimum read length filter of 25, and a minimum base quality of 30 and a minimum mapping quality of 30 for genotype calls. We call pseudo diploid genotypes by randomly sampling an allele for each site in the 1240 k panel using *Samtools v1.3*⁴² and *pileupCaller* (github.com/stschiff/sequenceTools).

Population genetics analyses. For downstream population genetic analyses, we combined our new data with previously published ancient African and Near Eastern genomes^{4,5,22,43–50}, together with 30 high-coverage modern African populations from the Simons Genome Diversity Project (SGDP)²⁴ and Human Genome Diversity Project (HGDP) datasets²⁵.

We performed PCA using *smartpca v16000*⁵¹ with “numthreads: 8” and “lsqproject: YES” option in the parameter file, and projected ancient individuals into the PC space calculated from modern African populations in SGDP (Table S3)²⁴. No shrinkmode option was applied, and the default outlier removal with up to 5 iterations was not altered. We applied a block jackknife approach for calculating standard errors (SE) of the PC location of Sudan_Kadruka1_4000BP⁵². We dropped one chromosome (i.e. block) each time and iterated for 22 chromosomes, to get an estimate of SE.

We calculated *f*₃ and *f*₄ statistics using *qp3pop v435* from the *AdmixTools v5.1* package²⁶ using options “*outgroupmode: YES*”, and *qpDstat v755* from the same package²⁶ using options “*f4mode: YES*” and “*prints: YES*”. We performed the Outgroup *f*₃ (Sudan_Kadruka1_4000BP, population X; Chimpanzee) test using Chimpanzee as the outgroup population. Population X represents published ancient African and Near Eastern populations and high-coverage modern populations from SGDP and HGDP datasets documented in Table S3. We report *f*₃ test results using more than 1000 overlapping SNPs. Given that Sudan_Kadruka1_4000BP shares high genetic affinity with early eastern African pastoral groups, we further tested the same list of population X using *f*₄ (Sudan_Kadruka1_4000BP, Kenya_EarlyPastoral_N; population X, Chimp) and *f*₄ (Kenya_EarlyPastoral_N,

population X; Sudan_Kadruka1_4000BP, Chimp) for the genetic cladility between Sudan_Kadruka1_4000BP and Kenya_EarlyPastoral_N.

Data availability

Genomic data newly generated in this study (BAM and original fastq file) are available through the European Nucleotide Archive (ENA), with accession number: PRJEB53198.

Received: 15 June 2022; Accepted: 29 November 2022

Published online: 03 December 2022

References

- Gilbert, M. T. P. *et al.* Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. *Science* **320**, 1787–1789 (2008).
- Pinhasi, R. *et al.* Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLoS ONE* **10**, e0129102 (2015).
- Zhang, M. *et al.* Comparative analysis of DNA extraction protocols for ancient soft tissue museum samples. *Zool Res* **42**, 280–286 (2021).
- Wang, K. *et al.* 2020 Ancient genomes reveal complex patterns of population movement, interaction, and replacement in sub-Saharan Africa. *Sci. Adv.* **6**, eaaz0183.
- Prendergast, M. E. *et al.* Ancient DNA reveals a multistep spread of the first herders into sub-Saharan Africa. *Science* <https://doi.org/10.1126/science.aaw6275> (2019).
- Ambrose S. East African Neolithic. In *Encyclopedia of Prehistory: Volume 1: Africa* (eds PN Peregrine, M Ember), pp. 97–109. Boston, MA: Springer US (2001).
- Lane, P. The 'moving frontier' and the transition to food production in Kenya. *Azania J. Br. Inst. East. Afr.* **39**, 243–264 (2004).
- Gansauge, M.-T. & Meyer, M. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nat. Protoc.* **8**, 737–748 (2013).
- Reinold, J. Kadruka and the Neolithic in the northern Dongola Reach. *Sudan & Nubia* **5**, 2–10 (2001).
- Bonnet, C. Excavations at the Nubian royal town of Kerma: 1975–91. *Antiquity* **66**, 611–625 (1992).
- Chaix, L. & Honegger, M. New data on animal exploitation from the Mesolithic to the Neolithic periods in northern Sudan. In *Climate Changes in Ancient Societies* (eds Kerner, S. *et al.*) 197–214 (Museum Tusulanum Press, 2015).
- Welsby, D. A. Life on the Desert Edge: 7000 Years of Settlement in the Northern Dongola Reach, Sudan. *Sudan Archaeological Research Society Publication 7. BAR International Series 980: Archaeopress* (2001).
- Maines, E. *Diversité biologique et archéologie de la mort: une approche populationnelle et culturelle du Néolithique soudanais (Haute-Nubie)* (Université Panthéon-Sorbonne, 2019).
- Maines, E. A tale of two taphonomies: assessing the contribution of taphonomic analysis to our understanding of Neolithic Sudanese burial sites using both the standard Anglophone and Francophone definitions. *Sour. Mater. Fieldwork Afr. Stud.* **3**, 29–67 (2021).
- Macklin, M. G. *et al.* A new model of river dynamics, hydroclimatic change and human settlement in the Nile Valley derived from meta-analysis of the Holocene fluvial archive. *Quat. Sci. Rev.* **130**, 109–123 (2015).
- Reinold, J. Les fouilles pré-et protohistoriques de la section française de la Direction des Antiquités du Soudan: les campagnes 1984–85 et 1985–86. *Archéologie du Nil moyen* **2**, 17–60 (1987).
- Bleasdale, M. *et al.* Ancient proteins provide evidence of dairy consumption in eastern Africa. *Nat. Commun.* **12**, 632 (2021).
- Chaix, L. Rapport préliminaire sur la faune du site de Kadruka 1. Soudan nord (Néolithique et Protohistorique). *Archéologie du Nil moyen* **2**, 61–62 (1987).
- Reimer, P. J. *et al.* IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. *Radiocarbon* **55**, 1869–1887 (2013).
- de Filippo, C., Meyer, M. & Prüfer, K. Quantifying and reducing spurious alignments for the analysis of ultra-short ancient DNA sequences. *BMC Biol.* **16**, 121 (2018).
- Peyrégne, S. & Peter, B. M. AuthenticIT: a model of ancient DNA damage to estimate the proportion of present-day DNA contamination. *Genome Biol.* **21**, 246 (2020).
- Skoglund, P. *et al.* Reconstructing prehistoric African population structure. *Cell* **171**, 59–71.e21 (2017).
- Fu, Q. *et al.* The genetic history of Ice Age Europe. *Nature* **534**, 200–205 (2016).
- Mallik, S. *et al.* The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature* **538**, 201–206 (2016).
- Bergström, A. *et al.* Insights into human genetic variation and population history from 929 diverse genomes. *Science* <https://doi.org/10.1126/science.aay5012> (2020).
- Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
- Busing, F. M. T. A., Meijer, E. & Van Der Leeden, R. Delete-m Jackknife for Unequal m. *Stat. Comput.* **9**, 3–8 (1999).
- Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).
- Sawchuk, E. A., Goldstein, S. T., Grillo, K. M. & Hildebrand, E. A. Cemeteries on a moving frontier: Mortuary practices and the spread of pastoralism from the Sahara into eastern Africa. *J. Anthropol. Archaeol.* **51**, 187–205 (2018).
- Honegger, M. & Williams, M. Human occupations and environmental changes in the Nile valley during the Holocene: The case of Kerma in Upper Nubia (northern Sudan). *Quatern. Sci. Rev.* **130**, 141–154 (2015).
- Williams, M. Holocene environments in Northeast Africa. In *The Oxford Handbook of Ancient Nubia* (eds G Emberling, BB Williams), New York: Oxford University Press (2021).
- Hildebrand, E. A. *et al.* A monumental cemetery built by eastern Africa's first herders near Lake Turkana, Kenya. *Proc. Natl. Acad. Sci. USA* **115**, 8942–8947 (2018).
- Dabney, J. *et al.* Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. USA* **110**, 15758–15763 (2013).
- Dee, M. W. *et al.* Radiocarbon dating at Groningen: New and Updated chemical pretreatment procedures. *Radiocarbon* **62**, 63–74 (2020).
- Synal, H.-A., Stocker, M. & Suter, M. MICADAS: A new compact radiocarbon AMS system. *Nucl. Instrum. Methods Phys. Res. B* **259**, 7–13 (2007).
- Salehpour, M., Håkansson, K., Possnert, G., Wacker, L. & Synal, H.-A. Performance report for the low energy compact radiocarbon accelerator mass spectrometer at Uppsala University. *Nucl. Instrum. Methods Phys. Res. B* **371**, 360–364 (2016).
- Ramsey, C. B. Methods for summarizing radiocarbon datasets. *Radiocarbon* **59**, 1809–1833 (2017).
- Peltzer, A. *et al.* EAGER: Efficient ancient genome reconstruction. *Genome Biol.* **17**, 60 (2016).
- Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).
- Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).

41. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F. & Orlando, L. mapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684 (2013).
42. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
43. Schlebusch, C. M. *et al.* Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. *Science* **358**, 652–655 (2017).
44. van de Loosdrecht, M. *et al.* Pleistocene North African genomes link Near Eastern and sub-Saharan African human populations. *Science* **360**, 548–552 (2018).
45. Fregel, R. *et al.* Ancient genomes from North Africa evidence prehistoric migrations to the Maghreb from both the Levant and Europe. *Proc. Natl. Acad. Sci. USA* **115**, 6774–6779 (2018).
46. Gallego Llorente, M. *et al.* Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa. *Science* **350**, 820–822 (2015).
47. Lazaridis, I. *et al.* Genomic insights into the origin of farming in the ancient Near East. *Nature* **536**, 419–424 (2016).
48. Harney, É. *et al.* Ancient DNA from Chalcolithic Israel reveals the role of population mixture in cultural transformation. *Nat. Commun.* **9**, 3336 (2018).
49. Narasimhan, V. M. *et al.* The formation of human populations in South and Central Asia. *Science* **365**, eaat7487 (2019).
50. Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503 (2015).
51. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
52. Meijer, E. & Van Der Leeden, R. Delete-m Jackknife for Unequal m. *Stat. Comput.* **9**, 3–8 (1999).

Acknowledgements

We thank Franziska Aron and Anja Wissgott for their help in the laboratory work. We thank Matthias Meyer, Kay Prüfer, Cesare de Filippo, and Stéphane Peyrégne for helpful discussions. We thank Jacques Reinold as director of the original fieldwork conducted at Kadruka 1 and Kadruka 21. We are also grateful to Jocelyne Desideri of the Anthropology Unit, University of Geneva for facilitating the sampling of the Kadruka material. Finally, we express our appreciation to *la Section française de la direction des antiquités du Soudan* and *la Commission des collections anthropologiques de la Faculté des sciences de l'Université de Genève* for supporting this research. Thanks also to the Centre for Isotope Research, University of Groningen for radiocarbon dating the sample.

Author contributions

S.S., N.B., J.K. and K.W. conceived the study. M.B. and C.L.M. collected and assembled skeletal material. C.L.M., M.B. and N.B. provided archaeological context. K.W., C.F., J.K. and S.S. produced data and performed genetic analyses. K.W., C.L.M., N.B. and S.S. wrote the main manuscript text with input from all co-authors. All authors reviewed the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-25384-y>.

Correspondence and requests for materials should be addressed to K.W. or S.S.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022