

RESEARCH

Open Access



# Whole mitochondrial genome analysis of the Daur ethnic minority from Hulunbuir in the Inner Mongolia Autonomous Region of China

Chi-Zao Wang<sup>1,2,3</sup>, Xue-Er Yu<sup>4</sup>, Mei-Sen Shi<sup>5</sup>, Hui Li<sup>4,6\*</sup> and Shu-Hua Ma<sup>1,2,3\*</sup>

## Abstract

**Background:** Mitochondrial DNA (mtDNA) variations are often associated with bioenergetics, disease, and speciation and can be used to track the history of women. Although advances in massively parallel sequencing (MPS) technology have greatly promoted our understanding of the population's history (especially genome-wide data and whole Y chromosome sequencing), the whole mtDNA sequence of many important groups has not been fully studied. In this study, we employed whole mitogenomes of 209 healthy and unrelated individuals from the Daur group, a Mongolic-speaking representative population of the indigenous groups in the Heilongjiang River basin (also known as the Amur River basin).

**Results:** The dataset presented 127 distinct mtDNA haplotypes, resulting in a haplotype diversity of 0.9933. Most of haplotypes were assigned to eastern Eurasian-specific lineages, such as D4 (19.62%), B4 (9.09%), D5 (7.66%) and M7 (4.78%). Population comparisons showed that the Daurians do have certain connections with the ancient populations in the Heilongjiang River basin but the matrilineal genetic composition of the Daur group was also greatly influenced by other non-Mongolic groups from neighboring areas.

**Conclusions:** Collectively, the whole mtDNA data generated in the present study will augment the existing mtDNA database. Our study provides genetic links between the Daur population and the aborigine peoples from Siberia and the Amur-Ussuri Region. But on the whole, compared with other Mongolic-speaking groups, the modern Daur population is closer to the East Asian ancestry group.

**Keywords:** Daur, Mongolic-speaking groups, Whole mtDNA sequence, Heilongjiang River basin, Northern East Asia

## Background

The Daur minority is one of the important members of the Mongolic-speaking population. They originally lived on the north beach of the Heilongjiang River (Amur River) [1]. After the 17th century, they gradually moved to Hulunbuir, Qiqihar and other settlements on the south beach of the Heilongjiang River [2]. A small number of them even migrated to Tacheng Prefecture in Xinjiang Province with their families as government troops of the Qing Dynasty [3]. The Daur ethnic group has long intermingled with the Ewenk

\*Correspondence: shuhua6633@163.com ; LHCA@fudan.edu.cn

<sup>1</sup> Department of Radiology, The First Affiliated Hospital of Shantou University Medical College, Shantou 515041, China

<sup>4</sup> MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, 200438 Shanghai, China

Full list of author information is available at the end of the article



and Oroqen ethnic groups, two other officially recognized ethnic groups in China, who are members of the Tungusic-speaking ethnic populations. They were known as the Suolun in the Qing Dynasty and are now called the Three Minorities in Inner Mongolia [3].

The Daur ethnic group has appeared in many important genetic studies as one of the representatives of Mongolic-speaking populations and indigenous groups from the Heilongjiang River basin [4–6]. Our previous studies elaborated on the paternal phylogenetic relationship between the Daur group, other Mongolic-speaking populations and Tungusic-speaking populations (including the Aisin Gioro family) by analyzing their Y-STR genetic polymorphisms and whole Y-chromosome sequences [7–11]. Recent genome-wide studies have further revealed the high level of genetic continuity of indigenous populations from the Heilongjiang River basin (including the Daur ethnic group) over at least the last 14,000 years and their distinct phylogenetic position in the genetic structure of human populations in East Asia [6, 12, 13]. However, previous studies on the Daur group from the perspective of maternal inheritance were relatively limited in terms of sample size and merely based on partial sequence polymorphisms, such as hypervariable segments I and II (HVS-I and HVS-II, respectively) and the control region (CR) [14–16]. Two early genetic studies established a certain genetic relationship between the modern Daur group and the ancient Khitan, which is one of the most significant findings of ethnic studies in China [15, 16]. Therefore, expanding the sample size and introducing whole mitochondrial genome analyses will undoubtedly contribute to a more comprehensive understanding of the maternal genetic background of the Daurians.

In this study, the whole mitochondrial genomes of 209 healthy and unrelated Daur individuals from Northeast China were sequenced by massive parallel sequencing (MPS) on the HiSeq X Ten system (Illumina, San Diego, CA, USA). Based on the sequencing data, we analyzed the haplogroup distribution and genetic diversity of the maternal genetic structure of the Daur group. To shed more light on the genetic relationship of the Daur group with worldwide populations, especially other neighboring/linguistically close populations and some related ancient groups, we conducted comprehensive population genetic analyses via Principal Component Analysis (PCA) [17] and other methods.

## Methods

### Samples, DNA extraction and quantification

A cohort of 209 unrelated Daur individuals (84 females and 125 males) was collected after receiving informed

consent. The individuals were considered autochthonous if their ancestors had lived in Hulunbuir, Inner Mongolia Autonomous Region of China, for at least three generations. Written informed consent was obtained from all participants, and the ethics committee of School of Life Sciences, Fudan University, Shanghai, People's Republic of China approved this study.

Genomic DNA was extracted from blood samples using a DP-318 Kit (Tiangen Biotechnology, Beijing, China) according to the manufacturer's protocol. The quantity of gDNA was measured with a NanoDrop ND-1000 (NanoDrop Technologies, Wilmington, DE, USA) according to the manufacturer's protocol. In consideration of the requirements of downstream processing, the gDNA was normalized to 0.1 ng/μL and stored at – 20 °C until amplification.

### Library construction and workflows for next-generation sequencing

DNA libraries were constructed using an MtDNA Library Preparation Kit 2.0 (Enlighten Biotech, Shanghai, China) and a WhoChrMT kit (Enlighten Biotech, Shanghai, China). PCR amplification was performed in a final volume of 30 μL containing 10 ng of template DNA, 5 μL RealCapChrMT Mix and 10 μL 3×EnzymeHF. Total reaction volumes were adjusted with nuclease-free water. The PCR was performed under the following conditions: enzyme activation for 3 min at 98 °C, 13 cycles of 20 s at 98 °C and 4 min at 58 °C, 7 cycles of 20 s at 98 °C and 1 min at 72 °C, 2 min at 72 °C followed by a 10 °C hold. The PCR products were purified with Agencourt AMPure XP beads (Beckman Coulter). Then, a second round of PCR amplification was carried out to introduce adapters and barcodes. The reaction volume (30 μL) was comprised of 10 μL 3×EnzymeHF, 18 μL nuclease-free water, 1 μL primer mix and 1 μL barcode mix. The PCR was performed under the following conditions: enzyme activation for 2 min at 98 °C, 7 cycles of 15 s at 98 °C, 15 s at 58 °C and 30 s min at 72 °C, extension for 2 min at 72 °C followed by a 10 °C hold. After purification, the libraries were pooled to a final concentration of 20 pM. Sequencing was performed on the Illumina HiSeq X Ten platform (Illumina, San Diego, CA, USA) with the corresponding Reagent Kit (PE150).

### Sequencing data analysis

The sequence data obtained from the Illumina HiSeq X Ten platform (Illumina, San Diego, CA, USA) were automatically analyzed by base recognition and converted into the original sequences in FASTQ format. First, redundant primers and indexes in the initial offline data were removed by cutadapt software [18]. Second,

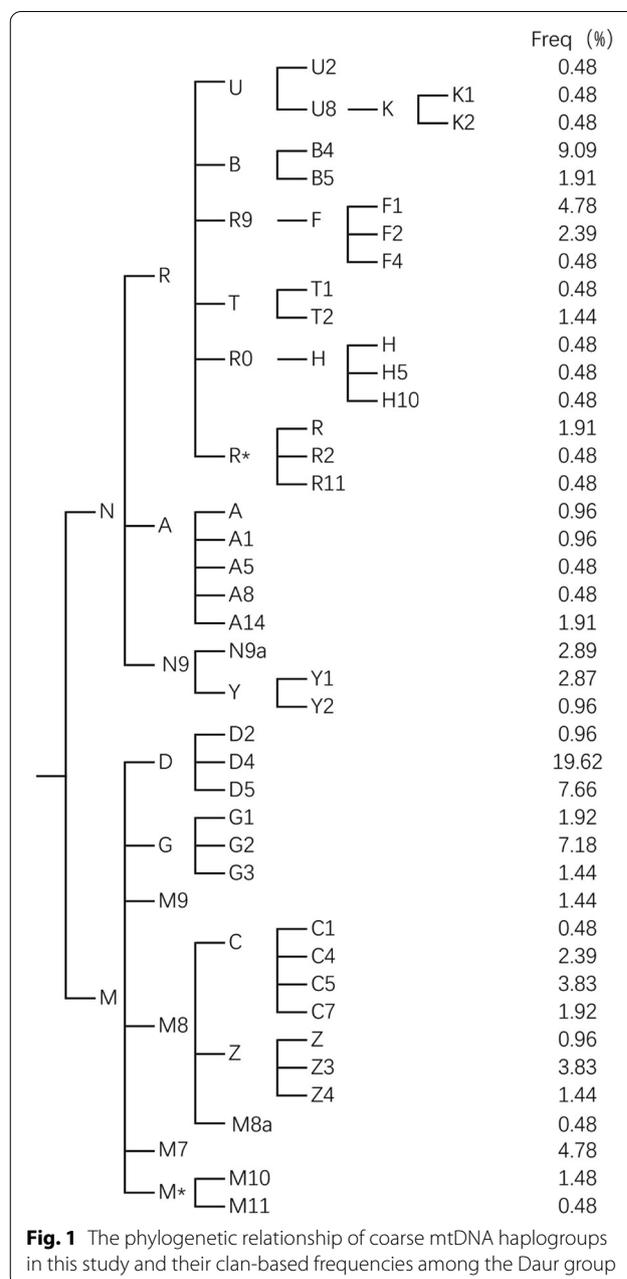
low-quality reads were filtered by Trimmomatic software [19]. To ensure the successful alignment of the loop amplification captured sequence, the final cleaned files were mapped to the revised Cambridge Reference Sequence [20] plus 64 bp (rCRS + 64 bp) using the Burrows-Wheeler Aligner [21] to generate the binary alignment/map (BAM) file. The sequences were also compared with the human reference genome hg19 to filter nuclear copies of mtDNA (NUMTs) [22]. We used Bedtools [23] to extract all reads that were successfully mapped to the HG19 reference genome from the BAM files in the previous step and then realigned them to rCRS + 64 bp to generate new BAM files using Bowtie2 software [24]. Then, SAMtools [25] and VarScan [26] were used to identify the mutation sites and output variants in VCF format files. Finally, BCFTools [25] was used to generate the consensus sequence (FASTA).

#### Haplogroup assignment and genetic diversity analysis

Sequencing performance was evaluated by read depth. The mtDNA haplogroups were determined using HaploGrep 2 [27] based on PhyloTree build 17 [28] and reconfirmed using the updated query engine (SAM2) built into EMPOP [29]. With reference to PhyloTree build 17, we constructed a simplified phylogenetic tree that showed the distribution of the coarse haplogroups. Haplotype diversities were calculated according to Nei's formula [30]. The discrimination capacity (DC) was also calculated as an important diversity parameter [31]. To show the differences in the genetic diversity of the different mitochondrial regions, haplotype-based analyses were repeated for the control region (CR, 16,024 to 576) and hypervariable segment I (HVS1, 16,024 to 16,488).

#### Population comparisons

To investigate the genetic relationship between the Daur group and other populations around the world, the whole mitochondrial genomes dataset was collected from 128 worldwide populations (Additional file 1: Table S1). In particular, we required the group size to be greater than 15 to avoid artificially low genetic diversity. Subsequently, the genetic background of the Daur group was analyzed by typical Principal Component Analysis (PCA) with the R statistical package (<https://www.r-project.org/>) based on haplogroup frequencies (Additional file 1: Table S2). *AncestryPainter* was used to illustrate the haplogroup sharing and ancestry composition of populations with a rounded and nice-looking graph [32]. For some mtDNA haplogroups of particular interest, the network analysis was constructed using the median-joining method in the Popart software [33, 34]. We also compared all Daur mitogenomes and published raw sequences (both ancient



**Fig. 1** The phylogenetic relationship of coarse mtDNA haplogroups in this study and their clan-based frequencies among the Daur group

and modern, Additional file 1: Tables S1, S3) in the hope of identifying some perfect matches.

## Results and discussion

### Sequence performance

The average mapped reads were 139,681 per sample, and the overall mean read depth was  $1260X \pm 422X$  (mean  $\pm$  SD) per individual. The variants recommended by EMPOP as well as the haplogroup information and the

mean sequencing depth of 209 Daur individuals are presented in Additional file 1: Table S4.

### Haplogroup distribution

Figure 1 presents a simplified phylogenetic tree that shows the distribution of the coarse haplogroups, and the detailed typing results are shown in Additional file 1: Table S4. In general, the matrilineal component of the Daur group was predominantly comprised of the eastern Eurasian-specific component (89.21%), represented by haplogroups D (28.24%), G (10.54%), B (10%), C (8.62%), R9 (7.65%), N9 (6.92%), Z (6.23%), A (4.79%), M7 (4.78%) and M9 (1.44%) [35, 36]. The remaining samples consisted of haplogroups U (1.44%), T (1.92%) and H (1.44%), which are generally confined to the European region [36, 37], and a few root types (R\* and M\*). Among these haplogroups, C and D have distinct Asian characteristics, and more than half of the northern Asian pool of human mtDNA is fragmented into their subclades [35, 38]. In the Daur population we studied, haplogroup C consisted of four sister subclades, C1 (0.48%), C4 (2.39%), C5 (3.83%) and C7 (1.92%), while haplogroup D consisted of three sister subclades, D2 (0.96%), D4 (19.62%) and D6 (7.66%). Notably, haplogroup D4 not only has a high frequency but also contains a total of 28 abundant downstream clades (Additional file 1: Table S4). Some subbranches of haplogroup D4 have very distinctive geographical distributions and are of great significance for the study of the demographic history of Asia [38, 39]. For example, haplogroup D4j (2.87% in this study) demonstrated a more southern geographic distribution, and haplogroup D4e4a (0.48% in this study) was mostly found in the Subarctic and Arctic regions [40]. According to previous studies, haplogroups B (10% in this study) and G (10.54% in this study) are also frequent in Mongolic-speaking groups [35, 41].

On the whole, the Daur population in this study embodies distinct regional and ethnic characteristics. Compared with earlier studies on Daur mtDNA [14–16], our research showed some changes in some haplogroup frequency distributions and detected some types that were not previously found in the Daur population (U, F, H, etc.), which could be attributed to the larger sample size and more advanced full mtDNA sequencing methods used in this study.

### Genetic diversity analysis

Based on whole mtDNA sequence data, a total of 127 different haplotypes were identified from the 209 unrelated Daur samples, of which 81 (63.78%) were unique. Although close matrilineal relatives (first to three degrees) were excluded, 61.24% of the total samples still

shared haplotypes with others. It is worth noting that the haplotypes belonging to M7b1a1+(16,192), G2a1 and Z3d were shared by 6 individuals. Moreover, one haplogroup was shared between five individuals, seven were shared between four individuals, seven were shared between three individuals and 28 were shared between two individuals. The overall haplogroup diversity was calculated as 0.9933 with a discrimination capacity of 60.77%. Additional file 1: Table S5 summarizes the above results. Repeated analysis based on CR and HVS1 showed that whole mtDNA sequence data decreased the number of shared haplotypes and increased the number of unique haplotypes. This is reflected in the discriminatory capacity increasing from 53.11% with the HVS1 haplotypes and 54.55% with the CR haplotypes to 60.77% with the whole mtDNA sequence for the Daur samples (Additional file 1: Table S5). These results indicate that the whole mtDNA sequence data offer a high power of discrimination and can be useful for genetic investigation and maternal lineage research in the Daur minority.

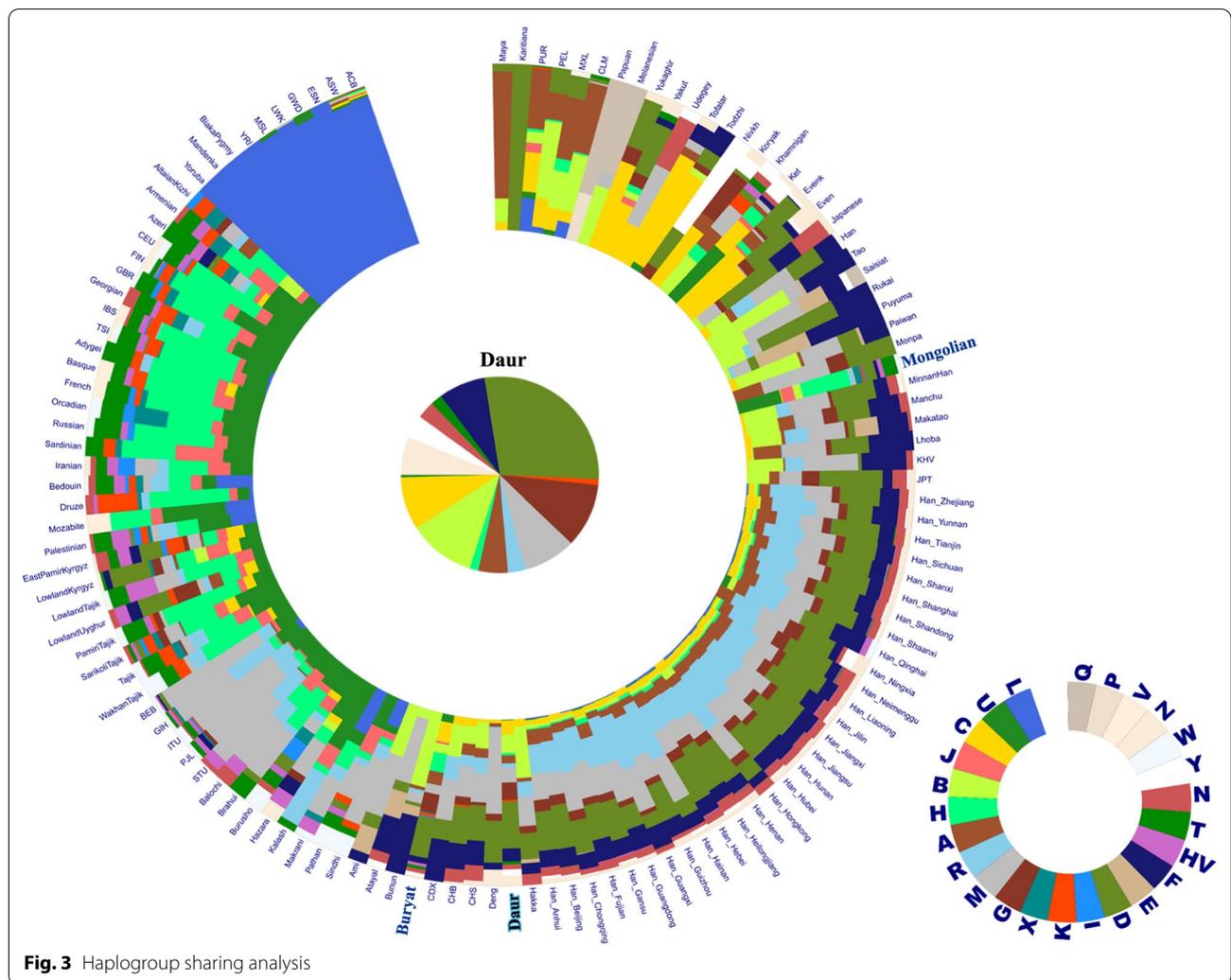
Of course, the genetic diversity of maternal genetic markers was slightly lower than that of paternal genetic markers, which is more due to the limitations of mitochondrial genetic markers themselves. In our previous study of genetic polymorphisms of 27 Yfiler® Plus loci in the Daur group, a total of 196 different haplotypes were observed in the sample of 203 Daur individuals, and the overall haplotype diversity was calculated as 0.9997 with a discrimination capacity of 0.9655 [7]. Our other two studies based on Y-STR/Y-SNP and Y-chromosome sequencing provided rich details on the paternal genetic diversity of the Daur group [8–10].

### Population comparisons

#### PCA

In our PCA results, 37.5% of the genetic variations were extracted by the first three components (Fig. 2). The African ancestry (AFR) populations can be separated clearly by PC1 and PC2, while the other large groups are closely related and even overlap. Our Daur population was clustered with other populations of the East Asian ancestry (EAS). As for the other two Mongolic-speaking populations, the Buryat population was located at the boundary between the groups of the East Asian ancestry (EAS) and the North Asia Ancestry (NAS), while the Mongolian population was closely related to groups of the West Asian Ancestry (WAS) and the Central Asian Ancestry (CAS). The distribution positions of the three Mongolic-speaking populations in the PCA map was roughly consistent with their geographic areas, which may reflect the maternal genetic contributions from



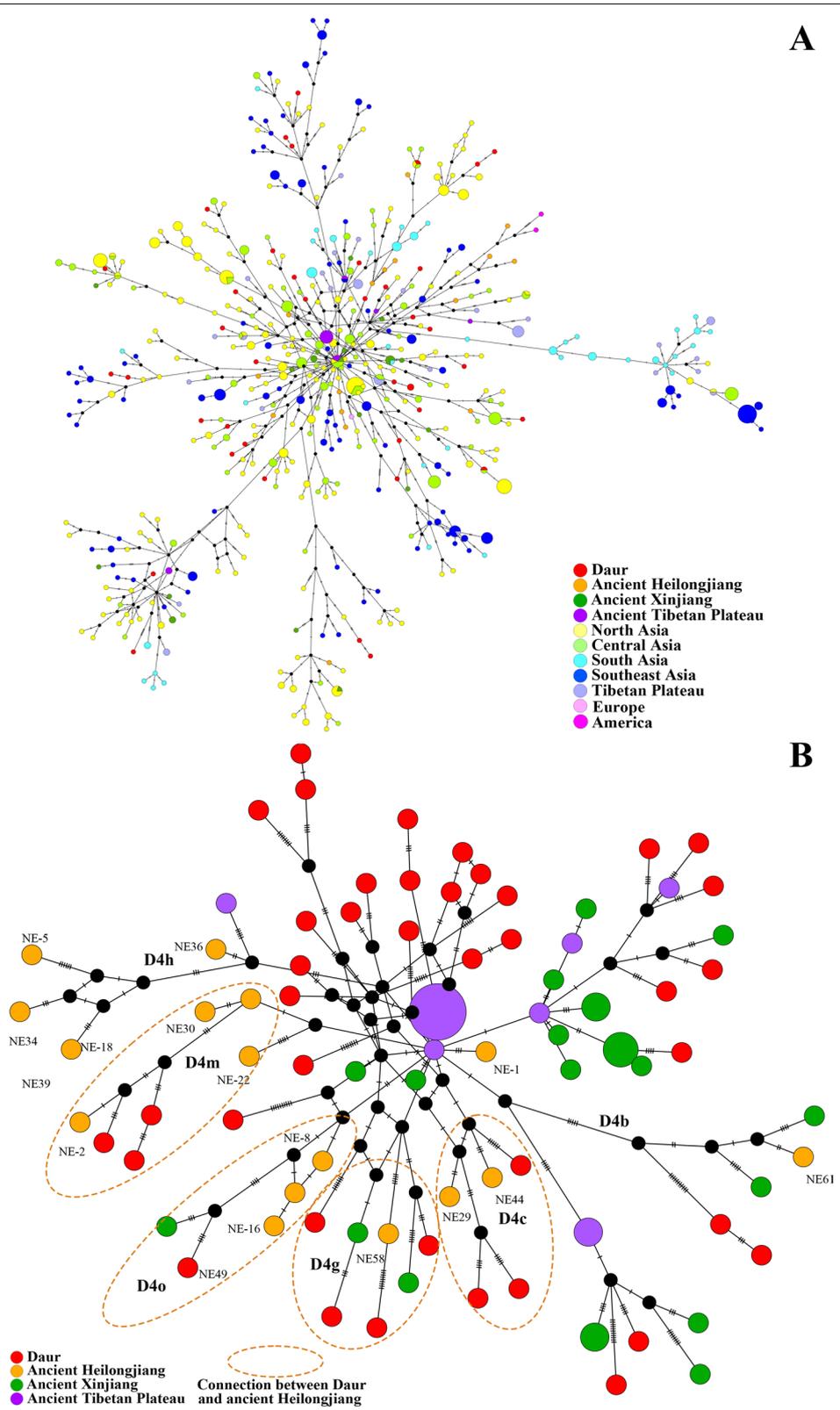


total haplogroups. Therefore, similar to the PCA results, the Daur population was closer to the eastern Eurasia groups, while the other two Mongolic-speaking populations were closer to the western Eurasia groups.

**Network analysis**

As mentioned above, haplogroup D4 not only has a high frequency (19.62%) but also contains abundant downstream clades in the Daur samples. According to previous studies based on partial sequences, D4 is also the high-frequency type of several ancient ethnic groups in Northeast China [42–44]. In the latest genome-wide study of northern East Asia, D4 also accounted for the majority of the detected samples in the ancient Heilongjiang River basin (66.67%, 16/24) [13]. We collected relevant available full sequence data (Additional file 1: Table S3) and constructed networks (Fig. 4). In Fig. 4A, the Daur samples came from scattered sources, showing connections with multiple regions of Asia. When

we focused on the genetic connection between the Daur samples and ancient samples, we found that most samples from the ancient Heilongjiang River basin had close connections with samples of Daur (Fig. 4A, B), and concentrated in haplogroups D4m, D4o, D4g and D4c. Haplogroup D4h, another high-frequency type in ancient Heilongjiang River basin populations, has not been detected in the modern Daur group which also makes sense that D4h is a distinctive native American type that may not have been involved in the late demographic history of northern East Asia [45]. In other words, the network analysis shows that the Daurians do have certain connections with the ancient populations in the Heilong River basin, but in the development process of the Daurians, they also absorbed a large number of female population from other sources. As to whether the modern Daur group has the closest matrilineal genetic connection with the ancient Heilongjiang population, we will collect more complete mitochondrial



**Fig. 4** Median-joining haplogroup D4 networks. **A** Ancient and modern samples. **B** The modern Daur and ancient samples. Note 1: According to the naming method of the original author, samples from the ancient Heilongjiang River basin have been specially labeled [13]. The larger the number, the older the age. Note 2: Black internal nodes only represent data structure bifurcation points, not actual samples

**Table 1** Summary of perfected matched sequences

Haplogroup	A5c	A8a1
Perfect match	DMT040 and Buryat 643	DMT185 and Buryat 618 ald1 and sun30
The other raw sequences	Khamnigan 43	-

sequence data and carry out it in detail in follow-up studies.

### Perfectly matched sequences

After comparing all raw whole sequences (both ancient and modern), we found two perfect matches between the Daur mitogenomes and published Buryat sequences (Table 1), DMT040-Buryat 643 [46] in haplogroup A5c and DMT185-Buryat 618 [46] in haplogroup A8a1, respectively. Haplogroup A5c includes one raw sequence from the Khamnigan population [46], while haplogroup A8a1 includes another perfect match (ald1-sun30) formed in the Yakut population [47]. The discovery of these perfectly matched sequences reflects the genetic connection between the Daur population and other aborigine peoples from Siberia and the Amur-Ussuri Region. Although haplogroup A5c and A8a1 are not the dominant types in the present Daur population, they may be genetic traces from early ancestors and exist in low-frequency form. Of course, more perfectly matched sequences may be found in the future with the increasing abundance of whole mitochondrial genome data.

### Conclusions

The present study provided the first set of whole mitochondrial genome data of 209 Daur individuals residing in Northeast China. The investigation of the Daur maternal lineages revealed that the vast majority of haplogroups belong to the eastern Eurasian-specific component. Population analyses showed that the Daurians do have certain connections with the ancient populations in the Heilongjiang River basin but the matrilineal genetic composition of the Daur group was also greatly influenced by other non-Mongolic groups from neighboring areas. This study also shows that whole mitochondrial sequence data can improve the resolution and offers a high power of discrimination in maternal studies by comparison of whole and partial sequence data in genetic diversity and population comparative analyses. Overall, the mitogenomes generated in the present study will augment the existing Daur mtDNA database, which provides a deeper understanding of the genetic composition of the Daur group and could potentially be useful for

regional-specific and prerequisite references for forensic, genealogical, and evolutionary purposes.

### Abbreviations

MtDNA: Mitochondrial DNA; MPS: Massively parallel sequencing; HVS-I and HVS-II: Hypervariable segments I and II; CR: Control region; PCA: Principal component analysis; DC: Discrimination capacity; N-J: Neighbor-joining.

### Supplementary information

The online version contains supplementary material available at <https://doi.org/10.1186/s12862-022-02019-4>.

**Additional file 1: Table S1.** Reference groups used in PCA and haplogroup sharing analysis. **Table S2.** Coarse haplogroup frequencies. **Table S3.** References used in networks. **Table S4.** The detail information for the full mtDNA sequences observed in 209 Daur individuals. **Table S5.** Diversity indices for the Daur population obtained with different mtDNA regions.

### Acknowledgements

We thank all sample donors for their contributions to this work and all those who helped with sample collection. We are particularly grateful to Mr. A-Li Aola for his support of this study.

### Author contributions

CZW conceived the study and wrote the original draft. XEY performed calculations, interpreted results and contributed to writing the manuscript. CZW and XEY were major contributors in writing the manuscript. MSS and SHM helped in conceiving the study and contributed to the manuscript. MSS and HL provided the funding support. All authors read and approved the final manuscript.

### Funding

This study was funded by the grants from the National Natural Science Foundation of China (Grants No. 81774395, 91731303); Natural Science Foundation of Guangdong Province (Grant No. 2019A1515011744); Grant for Key Disciplinary Project of Clinical Medicine under the Guangdong High-level University Development Program (Grant No. 002-18120302).

### Availability of data and materials

The 209 novel Daur complete mtDNA sequences have been uploaded to the Genome Sequence Archive (GSA) in the BIG Data Center (Members BIGDC 2017), Beijing Institute of Genomics (BIG), Chinese Academy of Sciences (<http://bigd.big.ac.cn/gsa-human>). The assigned accession of the submission is: HRA001624.

### Declarations

#### Ethics and consent to participate

Written informed consent was obtained from the all participants, and the ethics committee of School of Life Sciences, Fudan University, Shanghai, People's Republic of China approved this study. All methods were performed in accordance with the Declaration of Helsinki.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare no competing interests.

#### Author details

<sup>1</sup>Department of Radiology, The First Affiliated Hospital of Shantou University Medical College, Shantou 515041, China. <sup>2</sup>Shantou University Medical College, Shantou 515041, Guangdong, China. <sup>3</sup>Laboratory of Medical Molecular Imaging, The First Affiliated Hospital of Shantou University Medical College, No. 57 Changping Road, Shantou 515041, Guangdong, China. <sup>4</sup>MOE Key Laboratory

of Contemporary Anthropology, School of Life Sciences, Fudan University, 200438 Shanghai, China. <sup>5</sup>Criminal Justice College of China University of Political Science and Law, Beijing 100088, People's Republic of China. <sup>6</sup>Shanxi Academy of Advanced Research and Innovation, Fudan-Datong Institute of Chinese Origin, Datong 037006, China.

Received: 13 August 2021 Accepted: 27 April 2022

Published online: 18 May 2022

## References

- Narangoa L, Cribb R. Historical Atlas of Northeast Asia, 1590–2010: Korea, Manchuria, Mongolia, Eastern Siberia. New York: Columbia University Press; 2014.
- Dmytryshyn B, Crownhart-Vaughan EAP, Vaughan T. Russia's conquest of Siberia, 1558–1700: a documentary record, vol. 1. Portland: The Press of the Oregon Historical Society; 1985.
- Aola A-L. The western expedition and defense of Solon. China: Minzu University of China Press; 2017. In Chinese.
- Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, et al. The genetic legacy of the Mongols. *Am J Hum Genet*. 2003;72(3):717–21.
- Xue Y, Zerjal T, Bao W, Zhu S, Shu Q, Xu J, et al. Male demography in East Asia: a north-south contrast in human population expansion times. *Genetics*. 2006;172(4):2431–9.
- Wang CC, Yeh HY, Popov AN, Zhang HQ, Matsumura H, Sirak K, et al. Genomic insights into the formation of human populations in East Asia. *Nature*. 2021;591(7850):413–9.
- Wang CZ, Su MJ, Li Y, Chen L, Jin X, Wen SQ, et al. Genetic polymorphisms of 27 Yfiler® Plus loci in the Daur and Mongolian ethnic minorities from Hulunbuir of Inner Mongolia Autonomous Region, China. *Forensic Sci Int Genet*. 2019;40:e252–5.
- Wei LH, Yan S, Yu G, Huang YZ, Yao DL, Li SL, et al. Genetic trail for the early migrations of Aisin Gioro, the imperial house of the Qing dynasty. *J Hum Genet*. 2017;62(3):407–11.
- Wang CZ, Wei LH, Wang LX, Wen SQ, Yu XE, Shi MS, et al. Relating Clans Ao and Aisin Gioro from northeast China by whole Y-chromosome sequencing. *J Hum Genet*. 2019;64(8):775–80.
- Liu BL, Ma PC, Wang CZ, Yan S, Yao HB, Li YL, et al. Paternal origin of Tungusic-speaking populations: insights from the updated phylogenetic tree of Y-chromosome haplogroup C2a-M86. *Am J Hum Biol*. 2021;33(2):e23462.
- Wang CZ, Shi MS, Li H. The origin of Daur from the perspective of molecular anthropology. *J North Minzu Univ*. 2018;5:110–7 (In Chinese).
- Siska V, Jones ER, Jeon S, Bhak Y, Kim HM, Cho YS, et al. Genome-wide data from two early Neolithic East Asian individuals dating to 7700 years ago. *Sci Adv*. 2017;3(2):e1601877.
- Mao X, Zhang H, Qiao S, Liu Y, Chang F, Xie P, et al. The deep population history of northern East Asia from the Late Pleistocene to the Holocene. *Cell*. 2021;184(12):3256–66.
- Kong QP, Yao YG, Liu M, Shen SP, Chen C, Zhu CL, et al. Mitochondrial DNA sequence polymorphisms of five ethnic populations from northern China. *Hum Genet*. 2003;113(5):391–405 (In Chinese).
- Wu DY, Ma SS, Liu CY, Yang HM, Liu FZ, Chen ZC, et al. Study on the molecular archaeology of Khitan ancient cadavers. *J Yunnan Univ (Nat Sci Edn)*. 1999;53:300 In Chinese.
- Xu Y, Zhang XL, Zhang QC, Cui YQ, Zhou H, Zhu H. Genetic relationship between Ancient Khitan and Modern Daur. *J Jilin Univ Sci Edn*. 2006;06:997–1000 (In Chinese).
- Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet*. 2006;2(12):e190.
- Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet J*. 2011;17(1).
- Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
- Andrews RM, Kubacka I, Chinnery PF, Lightowler RN, Turnbull DM, Howell N. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet*. 1999;23(2):147.
- Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589–95.
- Just RS, Irwin JA, Parson W. Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing. *Forensic Sci Int Genet*. 2015;18:131–9.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2.
- Langmead B, Salzberg SL, Langmead B. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357–9.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
- Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, et al. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics*. 2009;25(17):2283–5.
- Weissensteiner H, Pacher D, Kloss-Brandstatter A, Forer L, Specht G, Bandelt HJ, et al. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res*. 2016;44(W1):W58–63.
- van Oven M, Kayser M. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat*. 2009;30(2):E386–94.
- Huber N, Parson W, Dur A. Next generation database search algorithm for forensic mitogenome analyses. *Forensic Sci Int Genet*. 2018;37:204–14.
- Clegg MT. Molecular evolution: molecular evolutionary genetics. *Science*. 1987;235(4788):599.
- Ip SCY, Lin SW, Lam TT. Haplotype data of 27 Y-STR loci in Hong Kong Chinese. *Forensic Sci Int Genet*. 2019;38:e14–5.
- Feng Q, Lu D, Xu S. AncestryPainter: A graphic program for displaying ancestry composition of populations and individuals. *Genom Proteom Bioinf*. 2018;16(5):382–5.
- Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*. 1999;16(1):37–48.
- Leigh JW, Bryant D. PopART: full-feature software for haplotype network construction. *Methods Ecol Evol*. 2015;6(9).
- Derenko M, Malyarchuk B, Denisova G, Perkova M, Rogalla U, Grzybowski T, et al. Complete mitochondrial DNA analysis of eastern Eurasian haplogroups rarely found in populations of northern Asia and eastern Europe. *PLoS One*. 2012;7(2):e32179.
- Palanichamy MG, Mitra B, Zhang CL, Debnath M, Li GM, Wang HW, et al. West Eurasian mtDNA lineages in India: an insight into the spread of the Dravidian language and the origins of the caste system. *Hum Genet*. 2015;134(6):637–47.
- Derenko M, Malyarchuk B, Denisova G, Perkova M, Litvinov A, Grzybowski T, et al. Western Eurasian ancestry in modern Siberians based on mitochondrial data. *BMC Evol Biol*. 2014;14:217.
- Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Rogalla U, Perkova M, et al. Origin and post-glacial dispersal of mitochondrial DNA haplogroups C and D in northern Asia. *PLoS One*. 2010;5(12):e15214.
- Li YC, Ye WJ, Jiang CG, Zeng Z, Tian JY, Yang LQ, et al. River valleys shaped the maternal genetic landscape of Han Chinese. *Mol Biol Evol*. 2019;36(8):1643–52.
- Ko MS, Chen CY, Fu Q, Delfin F, Ko YC. Early austronesians: into and out of Taiwan. *Am J Hum Genet*. 2014;94(3):426–36.
- Lan Q, Xie T, Jin X, Fang Y, Mei S, Yang G, et al. MtDNA polymorphism analyses in the Chinese Mongolian group: efficiency evaluation and further matrilineal genetic structure exploration. *Mol Genet Genomic Med*. 2019;7(10):e00934.
- Wang H, Liu W, Yuqin FU, Zhang X, Zhou H, Zhu H. Molecular biological analysis of remains from Jiangjougou Cemetery in Inner Mongolia. *Prog Nat Sci*. 2006;16(7):727–31.
- Molecular genetic analysis of remains from Lamadong cemetery, Liaoning, China. Wiley Subscription Services, Inc, A Wiley Company. 2007;134(3):404–11.
- Yu C, Xie L, Zhang X, Hui Z, Hong Z. Genetic analysis on Tuoba Xianbei remains excavated from Qilang Mountain Cemetery in Qahar Right Wing Middle Banner of Inner Mongolia. *FEBS Lett*. 2006;580(26):6242–6.
- Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, Olivieri A, et al. Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Curr Biol*. 2009;19(1):1–8.
- Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Dambueva I, Perkova M, et al. Phylogeographic analysis of mitochondrial DNA in northern Asian populations. *Am J Hum Genet*. 2007;81(5):1025–41.

47. Duggan AT, Whitten M, Wiebe V, Crawford M, Butthof A, Spitsyn V, et al. Investigating the prehistory of Tungusic peoples of Siberia and the Amur-Ussuri region with complete mtDNA genome sequences and Y-chromosomal markers. *PLoS One*. 2013;8(12):e83570.

### **Publisher's note**

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

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

