Hum Genet (2009) 126:395–410
DOI 10.1007/s00439-009-0683-0

ORIGINAL INVESTIGATION

# Ancient DNA provides new insights into the history of south Siberian Kurgan people

**Christine Keyser · Caroline Bouakaze · Eric Crubézy · Valery G. Nikolaev · Daniel Montagnon · Tatiana Reis · Bertrand Ludes**

Received: 6 February 2009 / Accepted: 6 May 2009 / Published online: 16 May 2009


**Abstract** To help unravel some of the early Eurasian steppe migration movements, we determined the Y-chromosomal and mitochondrial haplotypes and haplogroups of 26 ancient human specimens from the Krasnoyarsk area dated from between the middle of the second millennium BC. to the fourth century AD. In order to go further in the search of the geographic origin and physical traits of these south Siberian specimens, we also typed phenotype-informative single nucleotide polymorphisms. Our autosomal, Y-chromosomal and mitochondrial DNA analyses reveal that whereas few specimens seem to be related matrilineally or patrilineally, nearly all subjects belong to haplogroup R1a1-M17 which is thought to mark the eastward migration of the early Indo-Europeans. Our results also confirm that at the Bronze and Iron Ages, south Siberia was a region of overwhelmingly predominant European settlement, suggesting an eastward migration of Kurgan people across the Russo-Kazakh steppe. Finally, our data indicate that at the Bronze and Iron Age timeframe, south Siberians were blue (or green)-eyed, fair-skinned and light-haired people and that they might have played a role in the early development of the Tarim Basin civilization. To the best of our knowledge, no equivalent molecular analysis has been undertaken so far.

**Electronic supplementary material** The online version of this article (doi:10.1007/s00439-009-0683-0) contains supplementary material, which is available to authorized users.

C. Keyser (✉) · C. Bouakaze · D. Montagnon · B. Ludes
Laboratoire d'Anthropologie Moléculaire,
Institut de Médecine Légale, Université de Strasbourg,
11 rue Humann, 67085 Strasbourg Cedex, France
e-mail: Christine.Keyser@iml-ulp.u-strasbg.fr;
ckeyser@mageos.com

E. Crubézy
AMIS, CNRS, Université de Toulouse,
37 allées Jules Guesde, 31000 Toulouse, France

V. G. Nikolaev · T. Reis
State Medical University of Krasnoyarsk,
1 rue Partizana Zheleznyaka, 660022 Krasnoyarsk, Russia

## Introduction

Kurgans (Russian word for tumuli) are barrows characteristic of a culture arising on the steppes of southern Russia about 5000 BC and later spreading into eastern, central and northern Europe between 4400 and 2800 BC. The Kurgan culture is divided into different sub-cultures on the basis of the different kinds of graves under the barrows: pit-graves (Yamna), catacomb-graves (Katakomnaya) and timber-graves (Srubna). The westwards diffusion of this culture is sometimes equated with the appearance in eastern Europe of the Corded Ware culture and the introduction of Indo-European-speaking peoples (Gimbutas 1970).

In an attempt to reconstruct some of the population movements of ancient Kurgan people from the Eurasian steppes, the genetic background of 32 ancient human specimens from the Krasnoyarsk area in southern central Siberia (along the Yenisey River; Fig. 1) was characterized at the nuclear and mitochondrial DNA levels. Among these specimens, 10 were attributed to the Andronovo culture, 4 to the Karasuk culture, 12 to the Tagar culture and 6 to the Tachtyk one (Table 1). The Andronovo culture, related to the timber-grave group, appeared throughout the south Russian steppe, Kazakhstan and western central Asia during the second millenium BC. (Koryakova and Epimakhov 2007). The bearers of this Middle Bronze Age culture were strongly associated with the Indo-Iranians and credited with

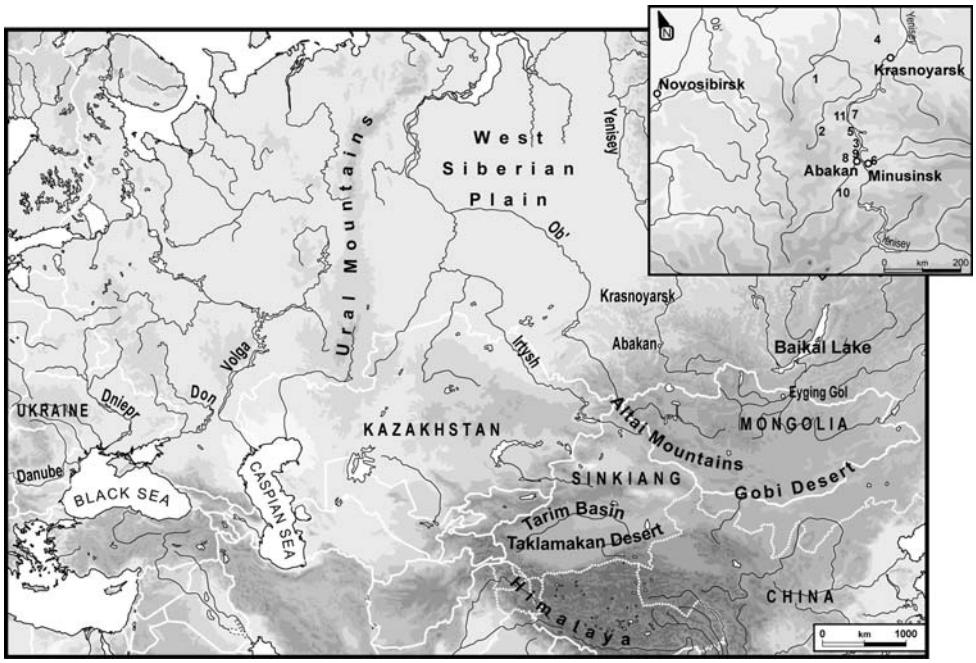


**Fig. 1** Map indicating the locations of the archaeological sites studied. *Numbers* refer to the burial sites noted in Table 1

the invention of the spoke-wheeled chariot (Lamberg-Karlovsky 2002). The Karasuk culture is a Late Bronze Age culture that succeeded the Andronovo culture in southern Siberia (late second millenium BC.). Karasuk people were farmers who practiced metallurgy on a large scale. They produced a realistic animal art, which probably contributed to the development of the later Scytho-Siberian animal art style. The Karasuk culture was replaced by the early Iron Age Tagar culture (first millenium BC.) which flourished in Khakassia (southern part of the Krasnoyarsk Krai) producing an art of animal motifs related to that of the Scythians of southern European Russia. On the Yenisey River, the Tagar culture was replaced by the Tashtyk culture, dating from the first to fourth century AD.

To investigate the history and origin of these ancient Krasnoyarsk specimens, two uniparentally inherited marker systems were analyzed. Indeed, apart from giving information about paternal and maternal lineages, both the non-recombining portion of the Y-chromosome (NRY) and the mitochondrial DNA (mtDNA) have proven to be good indicators of migration events in human population history (Underhill and Kivisild 2007). Autosomal short tandem repeats (STRs) were also typed to confirm conventional sexing and to assess possible parentage relationships and/or exogenous contamination. Finally, since the specimens under study are thought to have been "Caucasoid" (Kozintsev et al. 1999; Lebedynsky 2003; Moiseyev 2006), phenotype-informative single nucleotide polymorphisms (SNPs) were also tested.

To widen the geographic scale of our study, we determined the Y-chromosomal haplogroup of several Xiongnu specimens dated from the third century BC. to the second century AD. Xiongnu were nomadic tribes inhabiting the steppes north of China and controlling an empire stretching beyond the borders of modern-day Mongolia. We also performed Y-SNP typing of one Scytho-Siberian specimen from the Sebÿstei site in the Altaï Republic (Central Asia) dated from the middle of the fifth century BC. All these specimens were previously typed for autosomal and mtDNA polymorphisms (Keyser-Tracqui et al. 2003; Ricaut et al. 2004).

## Materials and methods

### Ancient human samples

DNA was extracted from 32 ancient human skeletons excavated from different kurgan sites of the Krasnoyarsk region in southern Central Siberia during the years 1964–2000. In this area, the average temperature is 20°C below zero in winter. Even if the graves under the kurgans were not frozen at the excavation, in summer the temperature at the graves level is never over a few degrees Celsius above or below zero. After being listed and arranged in cardboard boxes, skeletal remains were sent to the Krasnoyarsk State Medical Academy of Russia, at Krasnoyarsk University, where they have been stored in a dry and cold environment. In 2004, the cardboard boxes were opened for sampling by two members of our team and transferred to Strasbourg, France, under appropriate storage conditions. On arrival in the laboratory, samples were frozen until DNA extraction to ensure their good preservation. The location of the kurgans is indicated on Fig. 1; their associated culture, the time

**Table 1** Samples from the Krasnoyarsk region considered in the study

| Specimen | Code | Site/region/(map number) | Culture | Period | Dates | Sex |
|---|---|---|---|---|---|---|
| Bronze 1 | S07 | Tatarka cemetery, burial 64 Charypovsky region (1) | Andronovo | Middle Bronze Age | 1800–1400 BC | M |
| Bronze 2 | S08 | Tatarka cemetery, burial 55 Charypovsky region (1) | Andronovo | Middle Bronze Age | 1800–1400 BC | F |
| Bronze 3 | S09 | Solenoozernaïa IV, kourgane I, burial 3 Krasnoyarsk region (2) | Andronovo | Middle Bronze Age | 1800–1400 BC | M |
| Bronze 4 | S10 | Solenoozernaïa IV, kourgane I, burial 4 Krasnoyarsk region (2) | Andronovo | Middle Bronze Age | 1800–1400 BC | M |
| Bronze 5 | S11 | Solenoozernaïa I, burial 4 Krasnoyarsk region (2) | Andronovo | Middle Bronze Age | 1800–1400 BC | F |
| Bronze 6 | S12 | Solenoozernaïa I, burial 15 Krasnoyarsk region (2) | Andronovo | Middle Bronze Age | 1800–1400 BC | ? |
| Bronze 7 | S13 | Solenoozernaïa IV, kurgane I, burial 4 Krasnoyarsk region (2) | Andronovo | Middle Bronze Age | 1800–1400 BC | ? |
| Bronze 8 | S14 | Solenoozernaïa I, burial 4 Krasnoyarsk region (2) | Andronovo | Middle Bronze Age | 1800–1400 BC | F |
| Bronze 9 | S15 | Solenoozernaïa I, burial 29 Krasnoyarsk region (2) | Andronovo | Middle Bronze Age | 1800–1400 BC | ? |
| Bronze 10 | S16 | Oust-Abakansty, chief kurgan, Khakassia republic (3) | Andronovo | Middle Bronze Age | 1800–1400 BC | M |
| Karasuk 1 | S17 | Katcha, Drokino II, burial 1 Emelyanovsky region (4) | Karasuk | Late Bronze Age | 1400–800 BC | M |
| Karasuk 2 | S18 | Oust-Abakansty, kurgan IV, burial 1 Khakassia republic (3) | Karasuk | Late Bronze Age | 1400–800 BC | F |
| Karasuk 3 | S19 | Bogratsky, burial I Khakassia republic (5) | Karasuk | Late Bronze Age | 1400–800 BC | F |
| Karasuk 4 | S20 | Minoussinsk, Podgorny, burial 1 (6) | Karasuk | Late Bronze Age | 1400–800 BC | M |
| Tagar 1 | S21 | Novosselovsky region, Anach village, kurgan I, burial 3 (7) | Tagar | Iron Age | 800 BC–100 AD | ? |
| Tagar 2 | S22 | Novosselovsky region, Anach village, kurgan II, burial 4 (7) | Tagar | Iron Age | 800 BC–100 AD | F |
| Tagar 3 | S23 | Tchernogorsk, burial 1 Khakassia republic (8) | Tagar | Iron Age | 800 BC–100 AD | ? |
| Tagar 4 | S24 | Tchernogorsk, burial 6 Khakassia republic (7) | Tagar | Iron Age | 800 BC–100 AD | M |
| Tagar 5 | S25 | Oust-Abakansty, Khakassia republic (9) | Tagar | Iron Age | 800 BC–100 AD | M |
| Tagar 6 | S26 | Beysky region, burial 3 Khakassia republic (10) | Tagar | Iron Age | 800 BC–100 AD | M |
| Tagar 7 | S27 | Bogratsky region, kurgan 133, burial 3 Khakassia republic (11) | Tagar | Iron Age | 800 BC–100 AD | F |
| Tagar 8 | S28 | Bogratsky region, kurgan 133, burial 3 Khakassia republic (11) | Tagar | Iron Age | 800 BC–100 AD | M |
| Tagar 9 | S29 | Bogratsky region, kurgan 133, burial 3 Khakassia republic (11) | Tagar | Iron Age | 800 BC–100 AD | M |
| Tagar 10 | S30 | Bogratsky region, kurgan 133, burial 2 Khakassia republic (11) | Tagar | Iron Age | 800 BC–100 AD | ? |
| Tagar 11 | S31 | Bogratsky region, kurgan 133, burial 2 Khakassia republic (11) | Tagar | Iron Age | 800 BC–100 AD | ? |
| Tagar 12 | S32 | Bogratsky region, Abakano-Pérévoz II, burial 1, Khakassia republic (5) | Tagar | Iron Age | 800 BC–100 AD | M |
| Tachtyk 1 | S33 | Bogratsky region, Abakano-Pérévoz I, burial 5, Khakassia republic (5) | Tachtyk | Iron Age | 100–400 AD | F |

**Table 1** continued

| Specimen | Code | Site/region/(map number) | Culture | Period | Dates | Sex |
|---|---|---|---|---|---|---|
| Tachtyk 2 | S34 | Bogratsky region, Abakano-Pérévoz I, burial 4, Khakassia republic (5) | Tachtyk | Iron Age | 100–400 AD | F |
| Tachtyk 3 | S35 | Bogratsky region, Abakano-Pérévoz I, burial 2, Khakassia republic (5) | Tachtyk | Iron Age | 100–400 AD | F |
| Tachtyk 4 | S36 | Bogratsky region, Abakano-Pérévoz I, burial 4, Khakassia republic (5) | Tachtyk | Iron Age | 100–400 AD | F |
| Tachtyk 5 | S37 | Bogratsky region, Abakano-Pérévoz I, burial 4, Khakassia republic (5) | Tachtyk | Iron Age | 100–400 AD | F |
| Tachtyk 6 | S38 | Bogratsky region, Abakano-Pérévoz I, burial 4, Khakassia republic (5) | Tachtyk | Iron Age | 100–400 AD | F |

Question marks denote that the morphological sex determination was not possible or ambiguous. Map number, noted in bracket, indicates the location of the grave on Fig. 1

period, the estimated age and the morphological sex of the specimens studied are presented in Table 1. Some skeletons came from the same kurgan (e.g., S27–S31); one of them was excavated from a chief's kurgan (S16). Skeletal remains used for the DNA analyses were all long bone fragments. Moreover, the Y-SNPs typing was carried out on ten of the male Xiongnu specimens of the Egyin Gol valley necropolis (Keyser-Tracqui et al. 2003) as well as on a Scytho-Siberian skeleton from the Sebÿstei site (SEB 96K2) found in a kurgan of the Altaï Republic and dated from the middle of the fifth century BC. (Ricaut et al. 2004).

## DNA extraction

To eliminate surface contamination, the outer surface of the bones was removed to almost 2–3 mm of depth with a sanding machine (Dremel®, Breda, The Netherlands). Bone powder was generated using a column drill fitted with a surgical trepan. DNA was extracted from bone powders according to a previously published protocol (Keyser-Tracqui and Ludes 2005).

## Real time PCR quantification

Nuclear DNA quantitation was performed on an ABI Prism 7000 Sequence Detection System (Applied Biosystems, Foster City, CA, USA) using the Quantifiler® Human DNA Quantification Kit (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's protocol. In addition to the quantification of the nuclear DNA, the presence of PCR inhibitors was determined thanks to the co-amplification of an internal PCR control included in each reaction.

## Autosomal STR analysis

Autosomal STRs were amplified using the *AmpFl*STR® Profiler Plus™ Kit (Applied Biosystems, Foster City, CA, USA). Nine STRs (D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820) and the sex-determining marker amelogenin were simultaneously amplified. PCR reactions were performed according to the manufacturer's protocol, except for the 37 cycles used instead of the recommended 28, in a reaction volume of 10 µl thus reducing the volume of the DNA samples. Capillary electrophoresis was run on an ABI Prism 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) and data analysis was performed with the GeneMapper software (Applied Biosystems, Foster City, CA, USA). The parentage relationships between individuals were tested by pairwise comparison of the profiles.

## Y-chromosomal STR and SNP analysis

The DNAs of the male ancient specimens were analyzed at 17 Y-chromosomal STR loci [DYS19, DYS385a/b, DYS389I/II, DYS390, DYS391, DYS392, DYS393 (minimal haplotype), DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635 (Y GATA C4) and Y GATA H4] using the *AmpFl*STR® Y-filer™ PCR Amplification Kit (Applied Biosystems, Foster City, CA, USA). The experimental conditions were those recommended by the manufacturer except that 34 cycles were used instead of 30. STR products were analyzed on an ABI Prism 3100 Genetic Analyzer with GeneMapper software. The STR haplotypes obtained were individually compared to the Y chromosome Haplotype Reference Database (YHRD) (http://www.yhrd.org) (~57,000 9-loci haplotypes as of December 2008 database search) as well as to a private world Y-STR database maintained by us and containing data retrieved from the literature (~38,000 9-loci haplotypes). A two-step comparison procedure was applied: in a first step (since most of the data available for comparison do not include all the markers amplified in our study) comparison was made with the nine loci of the minimal haplotype; in a second step it was undertaken on the minimal haplotype plus other loci. No mismatch was allowed in the

comparative analysis: only exact matches were considered (even if many of the exact matches are likely "equal-by-state" and not "equal-by-descent").

A set of 13 Y-chromosomal SNPs [M3, M9, M17, M45, M89, M173, M175, M216, M217, M242, 92R7, RPS4Y$_{711}$ (M130), and Tat (M46)] characterizing Asian and Amerindian populations were also tested (references for SNP selection are given in Bouakaze et al. 2007). These SNPs were amplified and analyzed according to a previously published SNaPshot® (Applied Biosystems, Foster City, CA, USA) minisequencing protocol (Bouakaze et al. 2007). The Y-SNP haplogroup nomenclature followed that of the Y Chromosome Consortium (Y Chromosome Consortium 2002, Jobling and Tyler-Smith 2003; Karafet et al. 2008).

### Mitochondrial DNA sequencing and SNP typing

A 381 bp sequence of the HVI region (positions 16009–16390 of the Cambridge Reference Sequence (CRS; Anderson et al. 1981, revised Andrews et al. 1999) was amplified in two overlapping fragments as previously described (Keyser-Tracqui et al. 2003). The cycle-sequencing reaction was performed with the BigDye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA). The products were detected on an ABI Prism 3100 automatic sequencer and analyzed with the Sequence Navigator Software package (Applied Biosystems, Foster City, CA, USA). Haplotypes were assigned to the different haplogroups using the "near matching" method (Yao et al. 2002).

To validate exact mtDNA haplogroup determination and allocate mtDNAs to particular haplogroups not clearly defined with the control region alone, we also typed haplogroup-tagging SNPs of the mitochondrial coding region (Fig. S1). Twenty-six SNPs were selected and combined in three multiplex reactions using SNaPshot® assays. Multiplex 1 and 2 included a selection of SNPs defining European and Asian haplogroups (UK, U, U2, U4, U5a1, HV, T, T1, N1a, N9a, A, F1, X, C, Z, D5, G2, H, H3), whereas multiplex 3 included nearly exclusively polymorphisms defining subclades inside haplogroup H (H1, H2, H2a, H5a, H6, H7, H8; Fig. S1).

The distribution of the different defined haplotypes among modern and ancient populations was investigated by exact sequence searches performed against ~40,400-mtDNA haplotypes collected from the literature and maintained in a personal database.

### Human pigmentation gene SNP analysis

Ten autosomal SNPs were selected because of either their association with normal human pigmentation variation, namely, eye, hair and skin color (rs12913832, rs1805007, rs1805008, rs7495174, rs6497268/rs4778241, rs11855019/rs4778138) or their previously reported allele frequency differences between populations of the world (rs1545397, rs16891982, rs2031526, rs1426654). These SNPs are located in six pigmentation candidate genes: HERC2, OCA2, MC1R, MATP/SLC24A2, DCT and Golden Gene/SLC24A5. The choice of these SNPs as well as detailed protocol for genotyping using the SNaPshot minisequencing methodology are described in Bouakaze et al. 2009.

### Measures against contaminations and validation of the data

Bearing in mind the critical issues of pre-laboratory contaminations encountered in most of ancient DNA (aDNA) studies (Sampietro et al. 2006), bone samples were collected with extensive precautions (e.g., they were handled with gloves by a reduced number of people). Moreover, to check for possible modern contamination, the DNA extracted from saliva samples of all people handling the material or working in the laboratory was genetically typed and then compared with the profiles or haplotypes of all ancient samples.

The precautions concerning the facilities, the laboratory ware and the reagents were thoroughly respected (laboratory dedicated to ancient DNA only, strict separation of pre- and post-PCR experimental areas, UV irradiation of the rooms and the laboratory ware between each experiment plus treatment of equipment and benches with DNA contamination-removal solution (DNA away), wearing appropriate protective clothing (lab coats, facemasks and double pairs of gloves), use of pipettes with aerosol resistant tips, systematic use of negative controls. Multiple independent DNA extractions and PCR amplifications were carried out for each sample. Moreover, each new DNA extraction was directly followed by an *AmpFl*STR® profiler Plus™ DNA amplification to ensure that the new extract was not contaminated and that the following amplifications (mtDNA, Y-chromosome, autosomes) will be performed on the same individual.

## Results

### Autosomal STR analysis

Of the 32 individual remains analyzed by multiplex amplification, 6 DNA samples [S12, S17, S20, S30, S31 and S38 (Table 1)] appeared severely degraded since no amplifiable product was obtained from at least three independent extracts. No inhibition was detected as indicated by the real time PCR quantification. The remaining extracted samples gave 26 more or less complete allelic profiles. Consensus data are reported in Table 2. The loci D13S317, D18S51,

**Table 2** Consensus allelic profiles of 26 of the specimens under study

| Burial | Amel. | D13S317 | D18S51 | D21S11 | D3S1358 | D5S818 | D7S820 | D8S1179 | FGA | VWA |
|---|---|---|---|---|---|---|---|---|---|---|
| S07 | XY | 8/14 | 16/18 | 30/32.2 | 15/16 | 7/11 | 10/12 | 13/15 | 21/21 | 15/16 |
| S08 | XX | 9/10 | 15/17 | 29/30 | 16/16 | 12/13 | 8/8 | 13/15 | 23/24 | 17/18 |
| S09 | XX | 11/11 | 17/17 | – | 14/15 | 11/11 | 11/11 | 11/14 | 22/22 | 16/16 |
| S10 | XY | 9?/12 | 13/15 | 30/31 | 17/17 | 10/11 | 11/11 | 13/13 | 21/23 | 17/17 |
| S11 | XX | 11/11 | 14/15 | 30/31 | 14/18 | 11/11 | 11/11 | 13/14 | 24/24 | 18/18 |
| S13 | XX | ? | ? | ? | 17/18 | 11/(12) | ? | 13/13 | ? | 16/18 |
| S14 | XX | ? | ? | ? | 15/16 | 11/12 | ? | 8/12 | 22/25 | 17/18 |
| S15 | XX | 11/11 | ? | 30/30 | 15/16 | 9/13 | ? | 13/14 | 23/24 | ? |
| S16 | XY | 9/12 | 12/16 | 32.2/32.2 | 16/17 | 11/12 | 10/12 | 10/10 | 21/23 | 16/20 |
| S18 | XX | 11/11 | 12/15 | 32.2/36 | 16/17 | 11/12 | ? | 13/13 | 22/22 | 18/18 |
| S19 | XX | 11/11 | 12/13 | 30/31 | 17/18 | 11/11 | 9/11 | 12/13 | 21/23 | 15/17 |
| S21 | XX | 9/12 | 17/17 | 28/31.2 | 15/19 | 11/12 | 9/10 | 12/12 | 23/24 | 17/18 |
| S22 | XX | 11/12 | 14/15 | 30/31 | 16/17 | 11/12 | 9/10 | 13/15 | 19/24 | 19/19 |
| S23 | XX | 8/11 | 16/16 | 30/30 | 16/18 | 12/13 | 10/10 | 14/16 | 23/24 | 17/18 |
| S24 | XY | 8/11 | 14/16 | 29/30 | 14/16 | 10/13 | 10/11 | 12/14 | 19/24 | 14/17 |
| S25 | XY | ? | ? | 29/30.2 | 15/17 | 12/12 | ? | 13/13 | 21/25 | 18/18 |
| S26 | XY | 11/12 | 12/15 | 30/32.2 | 14/16 | 12/12 | 10/12 | 10/13 | 23/24 | 18/19 |
| S27 | XX | 8/10 | 16/17 | 28/32.2 | 14/16 | 9/11 | 9/9 | 15/16 | 19/22 | 15/19 |
| S28 | XY | 8/11 | 17/18 | 31.2/32.2 | 16/16 | 9/14 | 11/11 | 13/14 | 24/25 | 18/18 |
| S29 | XY | 9/9 | 14/? | ? | 14/18 | 12/13 | 9/10 | 13/16 | ? | 14/19 |
| S32 | XY | 8/11 | ? | 29/32.2 | 15/18 | 9/11 | 11/11 | 14/15 | 21/22 | 16/17 |
| S33 | XX | 12/12 | 21/21 | 30/30.2 | 16/18 | 12/13 | 9/11 | 14/15 | 22/23 | 15/17 |
| S34 | XY | 10/12 | 12/16 | 30/32.2 | 15/17 | 7/12 | 9/11 | 10/13 | 21/24 | 17/19 |
| S35 | XX | 10/11 | 13/14 | 30/31.2 | 14/18 | 9/11 | 9/10 | 10/13 | 22/24 | 14/17 |
| S36 | XX | 11/12 | 14/15 | 28/32.2 | 14/15 | 11/11 | 8/11 | 14/16 | 25/25 | 15/19 |
| S37 | XX | 9/13 | 14/16 | 28/29 | 15/17 | 11/12 | 10/11 | 15/15 | 22/22 | 17/20 |

Question marks denote that alleles could not be clearly amplified for the locus in question. Consensus allelic profiles were built after obtention of a minimum of three DNA profiles for each specimen

D21S11 and D7S820 were often not amplified, probably because they are expressed in the higher molecular weight range. Such an inverse dependance of the amplification efficiency on the size of the segment to be amplified is typical of DNA retrieved from ancient remains and results from damage and degradation of the DNA (Smith et al. 2003).

Morphological and molecular typing results for sex determination were in accordance with each other except for two specimens (S09 and S34); nevertheless, since the DNA profiles obtained for these two individuals were almost complete, it is highly probable that molecular results were the correct ones. Furthermore, the amelogenin locus allowed us to deduce the sex of four specimens for which morphological indicators of sex were absent (S13, S15, S21 and S23).

Comparison of the profiles in pairs revealed no first degree relatives, even for specimens unearthed from the same kurgan or even burial. It should be noted, nevertheless, that 35% of the DNA profiles were incomplete which might have hampered the detection of close familial kinship.

## Y-chromosomal STR and SNP analysis

To identify male lineages, an analysis of polymorphic markers located on the male-specific part of the Y-chromosome was performed. Seventeen Y-specific STRs were typed and used to construct haplotypes. All of the 10 male Siberian specimens were successfully typed at 14 loci (at least) and 6 different haplotypes were differentiated (Table 3). Three of them were shared between at least two specimens suggesting that some individuals could have belonged to the same paternal lineage (S10 and S16; S24, S34 and maybe S25; possibly S28 and S29) although buried in different kurgan and/or with no evidence of a close kinship link (cf., autosomal STRs). Pairwise comparisons of the haplotypes showed that except for S07, all the ancient male specimens bore closely related allelic profiles, differing at most at six loci (on the 17 tested) and always by one-step mutation only.

Haplogroup (hg) assignment, based on the Y-chromosomal SNP typing, revealed that except for S07, which was

**Table 3** Y-STR haplotypes and haplogroups determined for the ancient male specimens from the Krasnoyarsk region

| Specimen | DYS19 | DYS385 | DYS389I | DYS389II | DYS390 | DYS391 | DYS392 | DYS393 | DYS437 | DYS438 | DYS439 | DYS448 | DYS456 | DYS458 | DYS635 | YGATA | Haplogroup |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| S07 | 15 | 12/13 | 14 | 30 | 22 | 9 | 12 | 14 | 14 | 10 | 11 | 19 | 15 | 16 | 22 | 11 | C(×C3) |
| S10/S16 | 16 | **11/14** | 14 | 32 | 25 | **11** | **11** | **13** | **14** | **11** | **10** | **20** | **16** | **15** | **23** | 12 | R1a1 |
| S24/S34 | 17 | **11/14** | 13 | 31 | 24 | **11** | **11** | **13** | **14** | **11** | **10** | **20** | **16** | **15** | **23** | 13 | R1a1 |
| S25 | – | 11/14 | 13 | 31 | 24 | 11 | 11 | 13 | 14 | 11 | 10 | 20 | 16 | 15 | 23 | – | R1a1 |
| S26 | 16 | **11/14** | 13 | 31 | 24 | **11** | **11** | **13** | **14** | **11** | **10** | **20** | **16** | **15** | **23** | 13 | R1a1 |
| S28 | 16 | **11/14** | 14 | 31 | 25 | **11** | **11** | **13** | **14** | **11** | **10** | **20** | **16** | **15** | **23** | 12 | R1a1 |
| S29 | – | 11/14 | 14 | 31 | 25 | 11 | 11 | 13 | 14 | 11 | – | – | 16 | 15 | 23 | 12 | R1a1 |
| S32 | 17 | 11/14 | 13 | 31 | 24 | 11 | 12 | 13 | 14 | 11 | 10 | 20 | 16 | 15 | 23 | 13 | R1a1 |

– denote that alleles could not be amplified for the locus in question, alleles in bold define a motif shared by 4 of the 5 haplotypes belonging to R1a1 haplogroup

found to belong to hg C(×C3), all ancient specimens were affiliated to hg R1a1. This finding is in agreement with the relative similarity of the haplotypes mentioned above. While hg C has a distribution generally limited to populations of northern Eurasia, eastern Eurasia, Oceania, and the Americas, R1a1 is widely spread across Eurasia. It is found among western Eurasian, southern Asian, central Asian and Siberian populations. This haplogroup is thought to trace the migration patterns of the early Indo-Europeans, perhaps stemming from the Kurgan culture (Zerjal et al. 1999; Semino et al. 2000). The additional analysis performed on Xiongnu specimens revealed that whereas none of the specimens from the Egyin Gol valley bore this haplogroup, the Scytho-Siberian skeleton from the Sebÿstei site exhibited R1a1 haplogroup.

A search in the YHRD database as well as in our own databank revealed that none of the Y-STR haplotypes obtained from the south Siberian samples perfectly matched (at 17 loci) those included in the databases. Nevertheless, when not all loci were scored, matches were found for all samples except two (S07 and S32) for which even the search based on the 9-loci minimal haplotype was fruitless (Table 4). The S10/S16 haplotype matched the most frequent R1a1 haplotype (12 loci) seen in the south Siberian population of Derenko et al. (2006). This haplotype is notably found at high frequency in Altaians. It carries an allelic stucture 16-14-32-25-11-11-13 (DYS19-DYS389I-DYS389II-DYS390-DYS391-DYS392-DYS393) which is considered as a founder haplotype relative to southern Altaians (Kharkov et al. 2007). The S10/S16 haplotype is also found in eastern Europe (Hungary, Slovenia, Poland) as well as in Asia (Central Anatolia). The S24/S34 haplotype is mainly found in Poland and Germany. In Asia it is found in Anatolia, Armenia, Nepal and India. Haplotype of specimen S26 has a wide distribution since it appears in Europe as well as in western Asia, in Central Asia, in southern Asia and in southern Siberia. The allelic structure 16-24-11-11-13 (DYS19, DYS390, DYS391, DYS392, DYS393) found in this haplotype was described as the most frequent motif observed in a Ukrainian population by Kravchenko et al. (2002). According to these authors, this 5 Y-STR-loci haplotype might be an ancestral one. Haplotype S28 is the most frequently found in present-day populations. It is essentially carried by eastern and northern Europe individuals, as well as south Siberians. The S32 haplotype was not found in the databases even though it differs from the S24/S34 haplotype by only one-step mutation at locus DYS392. The S07 haplotype also did not appear in the YHRD database even when one mismatch was allowed in the minimal haplotype search. The current distribution pattern of all the Y-STR haplotypes found in our ancient sample is reported in the map on Fig. 2.

**Table 4** Results of the search against Y-STR databases

| Sample | Minimal haplotype | Minimal haplotype +2 to 7 additional loci |
|---|---|---|
| S07 | No match | No match |
| S10/S16 | 1 Hungarian[2]; 2 Slovenians, 2 Poles (YHRD) | 2 Tuvinians, 14 South Altaians[6]; 1 Turk[4]; 1 Hungarian[7]; 1 Pole[22]; 1 Cretan[16] |
| S24/S34 (S25) | 5 Poles[17,19]; 1 German[19]; 2 Poles, 1 Nepalese, 1 Turk, 1 German, 1 Armenian (YHRD) | 1 German[20]; 1 Pole[24]; 1 Indian[8] |
| S26 | 1 German[10]; 2 Poles[17,19]; 1 Serbian[23]; 1 Pole, 1 Kazakh, 1 Nepalese, 2 Germans, 1 Turk (YHRD) | 1 Turk[4]; 1 Chinese[15]; 1 Indo-Pakistani[1]; 8 South Siberians[6]; 1 Indian[5]; 5 Poles[24];1 German[20] |
| S28 (S29) | 2 Lithuanians[13,19]; 4 Latvians[13,19]; 6 Poles[17,19], 1 German[19]; 1 Byelorussian[18]; 1 Russian[14]; 2 Swedish[9,11]; 3 Serbians[23]; 1 Estonian[12], 1 Lithuanian[12]; 6 Slovenians, 13 Czechs, 4 Hungarians, 7 Germans, 3 Slovaks, 3 Ukrainians, 2 Croats, 1 Siberian Tuvan, 3 Poles, 3 Norwegian (YHRD) | 1 Austrian[2]; 1 Lithunian[18]; 1 Uigur[26]; 20 South Siberians[6]; 3 Poles[22]; 2 Malays[25]; 2 Indians[5]; 2 Russians[21] |
| S32 | No match | No match |

References are given in Table 1 in supplementary material


**Fig. 2** Current distribution pattern of the Y-STR haplotypes found in the ancient Siberians under study. Each *square* represents a present-day individual sharing the same Y-haplotype of an ancient specimen

Mitochondrial DNA analysis

The mtDNA haplotype of the 26 ancient individuals for whom genotypes were obtained, was determined by sequencing of the HVI region. To avoid ambiguous conclusions and to corroborate haplogroup assignment, all individuals were additionally typed for some specific mtDNA coding SNPs. Results are indicated in Table 5. A good agreement was found between coding and control region data except for three samples (S27, S29, S35) sharing a CRS HVI haplotype whose SNapShot assay coding SNPs allowed classification as hg U. Moreover, the SNapShot assay allowed us to obtain additional information regarding the phylogenetic refinement of two samples, S13 and S32, found to belong to subhgs H6 and H5a, respectively. Overall, 23 different haplotypes were distinguished and assigned to 16 different haplogroups. Twenty samples were found to belong to west Eurasian haplogroups (U2, U4, U5a1, T1, T3, T4, H5a, H6, HV, K, and I), whereas the 6 remaining samples were attributed to east Eurasian haplogroups (Z, G2a, C, F1b and N9a).

To assess the present distribution of the mtDNA types found in the ancient sample, a search of their occurence among modern populations of Eurasia was carried out in

**Table 5** HVI haplotype and haplogroup attribution for each Krasnoyarsk specimen successfully analyzed and current distribution of the haplotypes

| Sample | HVI haplotype | MtHg | SNPHg | Occurrence of the haplotype in the world |
|---|---|---|---|---|
| S07/S14 | 356C | U4 | – | 2 Turks[7,38]; 6 Russians[1,19,33]; 1 Ukrainian[33]; 5 Mansi[10]; 18 Volga-Ural region individuals[2]; 2 Germans[40,48]; 17 Altai–Sayan region individuals[12,18,46]; 5 Bosnians[34]; 1 Slovenian[34]; 2 Uygur[54]; 1 Uzbek[54]; 3 Mongols[54,26,14]; 2 Chechens[42]; 15 Latvians[39,31]; 1 Albanian[3], 4 Macedonians[3,57], 3 Romanians[3], 4 Hungarians[4,22]; 1 Buryat[14]; 1 Finn[20]; 3 Poles[19]; 1 Lithuanian[31], 3 Seto[31], 3 Karelians[31], 3 Swedish[31]; 4 Slovaks[35]; 6 Greeks[23]; 3 Italians[49]; 1 ancient Hungarian[50] |
| S08 | 129A 185T 223T 224C 260T 298C | Z (Z1) | Z | 2 Mongols[28,29]; 1 Kazakh[8]; 8 Koryaks[45], 1 Itel'men[45]; 1 Russe[33]; 1 Ket[11], 2 Nganasans[11,52]; 3 Yukaghirs[52]; 11 Volga-Ural region individuals[2]; 1 Buryat[14], 1 Altaian[14], 2 Teleuts[14]; 1 Volot[19], 2 Karelians[31], 1 Swedish[31] |
| S09 | 126C 163G 186T 189C 294T | T1 | T1 | 2 Turks[8], 12 Italians[17,49]; 1 Chinese Han[55]; 1 Indian[0]; 1 Uzbek[54]; 5 Latvians[31,39]; 1 Mongol[26]; 8 Hungarians[4,22,50]; 3 Austrians[5]; 3 Altaians[14] 1 Chukchi[14]; 4 Finns[20]; 3 Germans[48]; 2 Greeks[28]; 3 Byelorussians[1];3 Estonians[31], 4 Lithuanians[31], 1 Seto[31], 5 Karelians[31], 8 Swedish[31]; 1 ancient Kazakhstan[30] and 1 ancient Xinjiang[17] specimens |
| S10 | 051G 092C 129C 183C 189C 362C | U2e | U2 | 1 Estonian[31]; 1 Uygur[54] |
| S11 | 126C 294T 324C | T4 | T | 2 Roma[24]; 1 Byelorussian[1],1 Hungarian[22]; 1 Greek[23] |
| S13 | 187T 362C | H | H6 | 1 Corsican[16] |
| S15 | 093C 224C 311C 319A | K2b | UK | 1 Hungarian[4]; 1 Austrian[5] |
| S16 | 192T 256T 270T | U5a1 | U5a1 | 1 Russian[1]; 2 Kets[11]; 1 Byryat[14]; 1 Khakassian[14]; 1 Mongol[26]; 1 Albanian[3]; 2 Macedonians[3]; 3 Romanians[3], 4 Greeks[23]; 3 Hungarians[4,22]; 2 Austrians[5], 2 Italians[49]; 1 Finn[20], 1 Lithuanian[31], 2 Swedish[31] |
| S18 | 114A 256T 270T 294T | U5a1 | U5a1 | 1 Northwestern European[44] |
| S19 | 356C 362C | U4* | U4 | 1 Indian[43]; 1 Hungarian[22]; 4 Volga-Ural region individuals[2]; 2 Altai–Sayan region individuals[13]; 1 Estonian[31], 1 Lithuanian[31]; 1 Bosnian[34], 2 Slovenians[34]; 1 Austrian[5]; 1 Greek[23]; 2 Italians[49] |
| S21/S22 | 126C 189C 292T 294T | T3 | T | 2 Sardinians[15] |
| S23 | 126C 189C 292T 294T 296T | T3 | T | Not found |
| S24 | 129A 223T 304C 391A | I (I4) | – | 1 Indian[27]; 2 Icelanders[21]; 1 Buryat[14]; 5 Swedish[31]; 5 Hungarians[22]; 1 ancient Scandinavian[36] |
| S25 | 093C 223T 227G 278T 362C | G2a | G2 | 1 Uygur[54]; 3 Koreans[14,29,41]; 2 Latvians[31,39] |
| S26 | 148T 223T 234T 288C 298C 327T | C | C | 1 Tuvinian[14] |
| S27 | CRS | H | U | Too frequent |
| S28 | 172C 179T 183C 189C 232A 249C 304C 311C | F1b | F1 | 2 Mongols[14,54] |
| S29 | CRS | H | U | Too frequent |
| S32 | 304T 319A | H (H5) | H5a | 1 Austrian[5] |
| S33 | 093C 129A 223T 298C 327T | C | C | 2 Kazakhs[8,54]; 6 Altaians[13]; 7 Buryats[13,46]; 15 Tuvinians[13,46];1 Uygur[54]; 2 Chineses[53,41]; 10 Evenks[14,46]; 1 Ulchi[46]; 6 Yakuts[47,14]; 2 Mongols[14,41]; 1 Kalmyk[14]; 2 Oroqen[41]; 1 Xiongnu[25] |
| S34 | 172C 311C | HV | HV | 1 Indian[43]; 1 Uzbek[54], 1 Mongol[54]; 1 Evenk[41]; 1 Kalmyk[14]; 1 Macedonian[57]; 2 Italians[49,51] |
| S35 | 093C 209C | H | U | 1 Estonian[31] |
| S36 | 223T 257A 261T | N9a | N9a | 5 Chinese[41,53,55]; 3 Koreans[32,41,56]; 2 Vietnameses[23] |
| S37 | 126C 163G 186T 189C 269G 294T 362C | T1 | T1 | Not found |

the literature and personally compiled database. Exact matches were observed for almost all sequences. Among those mostly prevalent was the hg U4 motif 356C (S07/S14) which was found in northern, eastern and southeastern European populations, as well as in Volga–Ural, Altai–Sayan and peri-Baïkal area populations. Note that this sequence occupies a central position on U4 phylogeny built by Malyarchuk (2004). It was also observed in an ancient Hungarian specimen from the tenth to eleventh century (Tömöry et al. 2007). The sub-hg U4 variant 356C–362C (S19) is present in northern, eastern and mediterranean Europe, in the Volga–Ural and Altai–Sayan regions as well as in southern India. Another sequence within hg U showing a wide geographic distribution is the S16-U5a1 haplotype which matches in southern Siberia, in Central Asia, as well as in northern, western, eastern and southeastern Europe. Conversely, the S18-U5a1 haplotype is most likely rare since found only once in a northwestern European, likewise the S10-U2 haplotype which has been described in one eastern European and one central Asian individual only. Specimens S27, S29 and S35 bear a CRS HVI sequence belonging to hg U. Such results have already been described in Russian and Byelorussian populations where hg U CRS sequences were found in a relatively high proportion (~35%; Belyaeva et al. 2003).

Within hg T, the most prevalent sequence type was that harbored by specimen S09. This sequence was described as the root sequence of hg T1 (Richards et al. 2000; Pike 2006). Its highest frequency is in western Eurasia (mainly the Baltic region) with occasional occurrences in eastern Eurasia. This founder haplotype was also observed in two ancient specimens, one from Kazakhstan (1400–1300 BC, Bronze Age; Lalueza-Fox et al. 2004), the other from a Xinjiang site in northwestern China (Gao et al. 2008). Surprisingly, the S37-haplotype, differing from the previous one by two additional mutations, was not observed in our database. The T3-haplotypes harbored by specimens S21, S22 and S23 are either rare or absent whereas the S11-T4 haplotype, although being described as a main founder cluster within haplogroup T (Richards et al. 2000), was not commonly found.

The H haplotypes observed in our ancient sample are uncommon in present-day populations since both S13-H6 and S32-H5a sequences were found in only one European individual. HV lineage are represented by one sequence (S34) found in India, in Central Asia and in southeastern Europe.

Specimen S24 was found to belong to haplogroup I, subclade I4. Exact matches to this I4 type were mainly found in northern and eastern Europe individuals. The fact that one ancient Scandinavian specimen (0–400 AD) bore this sequence gives direct evidence of the antique presence of such sequence in the north of Europe (Melchior et al. 2008). The K2 haplotype harboured by specimen S15 was observed only twice, in European samples (1 Hungarian and 1 Austrian).

The eastern Eurasian lineages are represented by sequences belonging to hgs N9a, Z, G2a, F1b and C. The Z haplotype observed in the S08 ancient specimen belong to subhg Z1. It is observed in northeastern Asians, in south Siberian populations as well as in Central Asia. It is also present among several populations of the Volga–Ural and Baltic Sea regions. The S25-G2a sequence has been observed in few but dispersed individuals (Koreans, Latvians and Uygur). The specimen S28 belongs to hg F1b with a motif reported in two Mongols only. Haplogroup C is represented by two sequences: one had a HVI motif observed in only one south Siberian individual (S26). The other had a HVI motif mainly found in Siberia and Central Asia (S33). Finally, the S36 specimen carries a N9a-haplotype identical to those described previously in East Asian individuals (Chinese, Koreans, Vietnamese). Figure 3 represents the current and past distribution of the overall mtDNA haplotypes found in our ancient Krasnoyarsk sample (except the CRS sequences). This distribution is similar to that depicted in Fig. 2 for the Y-chromosome, despite the sparser pattern of the mtDNA counterpart.

## Human pigmentation gene SNP analysis

In order to deepen the search of the geographic origin of the Siberian specimens under study, we typed SNPs located in human pigmentation genes. Ten SNP markers located in genes that have been described as accounting for variation in human hair, eye and skin color but also in ethnogeographic ancestry were thus selected and a minisequencing-based assay was developed on modern samples (Bouakaze et al. 2009). This assay was subsequently applied on the ancient Siberian samples so that complementary information provided by phenotype-associated SNPs could add to previous anthropological and genetic findings. The phenotype and ancestry of the ancient Siberian specimens under study are indicated in Table 6 (genotype details for each investigated marker is given in Bouakaze et al. 2009). Surprisingly, the typing of a SNP associated to eye color (rs12913832) shows that at least 60% (15/25) of the Siberian specimens had blue (or green) eyes (S27 cannot be tested because bone sample and DNA extract were used up). Moreover, the pigmentation SNP analysis showed that all except three specimens exhibited a European ancestry, even when they bore an Asian mtDNA haplotype as is the case for samples S25, S26, S28, S33 and S36, demonstrating the importance of studying both maternal and paternal lineages. These results also show that two individuals carrying the same mtDNA haplotype can be classified in opposite ethnogeographic groups as is the case for samples S07


**Fig. 3** Current distribution pattern of the mtDNA haplotypes found in the ancient Siberians under study (except the CRS sequence). Each *square* represents a present-day individual sharing the same mtDNA haplotype of an ancient specimen. *Cross* represents an ancient specimen different from those studied in this work

and S14: note that these two specimens belong to different paternal lineages since S07 is the single specimen carrying haplogroup C(×C3) but not haplogroup R1a1. Specimen S32 appears as having a mixed ancestry; curiously, this specimen exhibits Y-chromosome and mtDNA haplotypes virtually unknown in present-day populations. Most of the specimens seem to have been light-skinned people with blond or light brown hair.

## Discussion

With the present study, we aimed to unravel some of the early Eurasian steppe migration movements by analyzing paternal, maternal and autosomal genetic variation in Bronze and Iron Age anthropological remains recovered from the Krasnoyarsk area in southern Central Siberia. There is virtually no knowledge either about the origins and the history of these ancient south Siberian inhabitants or about the language(s) they may have spoken. Nevertheless, many scholars believe that these Kurgan people, and notably the bearers of the Andronovo culture, spoke a Proto-Indo-Iranian or a Proto-Iranian language (Lamberg-Karlovsky 2002). Moreover, the south Siberian tribes under study (Andronovo, Karasuk, Tagar) have been described as exhibiting pronounced Europoid features (Kozintsev et al. 1999; Lebedynsky 2003; Moiseyev 2006). These data raised questions as to where these people came from, which routes have been followed and to which extent they have contributed to the spread of the Indo-European language. In an attempt to answer these questions, we analyzed, in parallel, markers of distinct genetic systems.

The choice of autosomal STR markers as a first approach for analyzing the Siberian ancient remains was based not only on their high discriminatory power to investigate close familial relationship, but also on their ability to detect degraded and/or contaminated DNA. Twenty-six out of the 32 bone samples collected yielded amplifiable DNA and 65% of the genetic profiles were complete (Table 1). This high success rate suggested that the nucleic acids were well preserved and allowed us to envisage other single-copy nuclear genes analyses (i.e., pigmentation genes). No close kinship link was detected between the subjects under study even for those found in the same kurgan (S27–S31). Nevertheless, the Y-chromosomal analyses performed on the ten male specimens showed that S28 and S29 shared the same haplotype. These Y-chromosomal analyses, based on the combined use of STRs and SNPs, also revealed that, with the exception of one individual, all samples examined fall into hg R1a1.

Haplogroup R1a1 is defined by marker M173 plus M17 (Y Chromosome Consortium 2002; Jobling and Tyler-Smith 2003; Karafet et al. 2008) and has a widespread distribution area on the Eurasian continent. It is spread among western Eurasian (mostly eastern European and Volga–Ural populations), southern Asian (mainly India and Pakistan's populations), central Asian and Siberian populations (especially southern Siberians), whereas it is rather rare in

**Table 6** Most probable phenotype and biogeographical origin of each Krasnoyarsk specimen deduced from the 10 phenotype and ancestry autosomal markers SNPs analysis

| Sample | Eye color according to rs12913832 | Phenotype according to OCA2 diplotype | Ancestry according to rs1545397, rs16891982, rs2031526 |
|---|---|---|---|
| S07 | Brown | Blue or brown eye, dark brown hair, fair or medium skin | As |
| S08 | Brown | – | As |
| S09 | Blue | Blue or brown eye, blond or light brown hair, fair or medium skin | Eu |
| S10 | – | Blue or brown eye, brown hair | Eu |
| S11 | Blue | – | Eu |
| S13 | – | Blue or brown eye, blond or light brown hair, fair or medium skin | Eu |
| S14 | Brown | Blue or brown eye, brown hair, fair or medium skin | Eu |
| S15 | Blue | – | Eu |
| S16 | Blue | Blue or brown eye, blond or light brown hair, fair or medium skin | Eu |
| S18 | Blue | – | Eu |
| S19 | Blue | – | Eu |
| S21 | Blue | Blue or brown eye, brown hair, fair or medium skin | Eu |
| S22 | Brown | – | Eu |
| S23 | Blue | – | Eu |
| S24 | Blue | – | Eu |
| S25 | Blue | – | Eu |
| S26 | Brown | Blue or brown eye, blond or light brown hair, fair or medium skin | Eu |
| S28 | Brown | Blue or brown eye, blond or light brown hair, fair or medium skin | Eu |
| S29 | Blue | – | Eu |
| S32 | Brown | – | Mix |
| S33 | Blue | Blue or brown eye, blond or light brown hair, fair or medium skin | Eu |
| S34 | Blue | – | Eu |
| S35 | Brown | – | Eu |
| S36 | Blue | Blue or brown eye, blond or light brown hair, fair or medium skin | Eu |
| S37 | Blue | Blue or brown eye, blond or light brown hair, fair or medium skin | Eu |

Blue eye color phenotype is predictive of blue or green-eyed individuals (non brown eye color)

– indicates that the marker could not be amplified, *As* Asian, *Eu* European, *Mix* mix

East Asian populations. In western Eurasia, a clear north–east/south–west cline has been described (Rosser et al. 2000; Semino et al. 2000; Wells et al. 2001). Indeed, the R1a1 haplogroup frequency reaches a maximum in Poland, Hungary, and Ukraine and decreases in the direction of central and northern Europe. The same occurs in the southern direction, towards Anatolia and the Caucasus. These clinal frequency distributions have been associated with ancient population movements in Europe: according to Semino et al. (2000), the geographical distribution of the R1a1 haplogroup probably reflects the re-population of Europe after the last glacial maximum (~20–12 kya) from a refugium in eastern Europe, likely in Ukraine (Passarino et al. 2001). This postglacial spread might have been magnified by the movement of the Kurgan people from the north of the Caspian Sea in a much more recent timescale (Rosser et al. 2000; Semino et al. 2000). This "Kurgan people" expansion would have resulted in the spread of the Indo-European language as postulated by Gimbutas (1970). Thereby, R1a1 was viewed by some authors as the marker of the Indo-European contribution (Zerjal et al. 1999; Kharkov et al. 2004). According to Pericić et al. (2005), the present distribution pattern of the R1a1 haplogroup was probably also influenced by much later migratory events like massive Slavic migration from fifth century AD.

In our ancient sample, among the nine specimens carrying haplogroup R1a1, five different Y-chromosomal haplotypes were observed. Similarities were noted between these

haplotypes, particularly the motif 11/14-11-11-13-14-11-10-20-16-15-23 (in bold in Table 3) which is common to all of them except S32. This motif is typically an eastern European one since currently found in the Russian federation only (YHRD database). Matching haplotypes were found for all the R1a1-specimens except S32. Figure 2 shows that the current distribution pattern of the Y-STR haplotypes found in our ancient sample resembles that of R1a1. Indeed, they were observed at high frequencies in Slavic and Baltic populations (with peaks among Poland and Czech Republic) as well as in the indigenous populations of south Siberia. By contrast, they were only sporadically observed in central and east Asia and were absent in western Europe.

Regarding the mtDNA analyses, our findings indicate that the ancient Krasnoyarsk mtDNA pool harbored both western and eastern Eurasian lineages. Nevertheless, most of the retrieved sequences ($n = 20$, 77%) belong to western Eurasian mtDNA haplogroups (HV, H, T, I, U and K). The eastern Eurasian lineages (23% of the sequences) were represented by haplogroups or subhapologroups C, Z, G2a, F1b and N9a. The western Eurasian contribution to the ancient mtDNA pool reached 90% for the Bronze Age and decreased to 67% for the Iron Age. Thus, despite a small sample size, our data suggests a temporal pattern which is in agreement with the view that west Eurasian populations predominated in the Krasnoyarsk region during the Bronze Age, whereas Asian component began to increase from the Iron Age on. This result is similar to that obtained in the ancient DNA study of Lalueza-Fox et al. (2004) who showed that all Kazakh sample specimens from before thirteenth to seventh centuries BC belonged to European lineages. After that time, there was an influx of East Asian sequences which are thought to have coexisted with the prior west Eurasian genetic substratum.

As shown in Table 5, and particularly in Fig. 3, the current distribution of the ancient mtDNA haplotypes can be broadly divided into three different geographic poles. The first is represented roughly by eastern and northern Europe, the second by the Volga–Ural region and the third by southern Siberia. It is interesting to note that the distribution of the paternal and maternal lineages is close. Indeed, except for the Volga–Ural region, both maps overlap. This would mean that the story of women matches well that of men. In other words, the migrations in which south Siberian specimens were involved seemed to be "whole-population movements" rather than "war-like movements" involving the men only. The fact that East Asian mtDNA sequences appeared at the Iron Age could signify that once settled, migrants of supposed European ancestry began to establish relationships with groups coming from the east and to take Asian women as wives. Moreover, the relative high diversity of the mtDNA gene pool observed in the ancient specimens indicates that numerous populations carrying different mtDNA variants were involved in the formation of southern Siberian populations, even reflecting long-distant movements. It would not have presented any major difficulty for Bronze Age and Early Iron Age peoples to range from one end of Eurasia to the other within some centuries. Historical records and archaeology attest that nomadic groups moved across Eurasia from North of the Black sea, through Central and Inner Asia, to northeast Asia in a matter of centuries (Mair 2005). Some of them are described in Chinese historiography as horse-riding, Caucasian-looking, Indo-European-speaking people and are sometimes referred as the "Kurgan Culture" (Zerjal et al. 2002). Paleogeographic studies provide material which suggests that climate change, particularly in the eastern regions of the steppes, was among the causes of these population movements (Van Geel et al. 2004).

If we consider that there is a correspondence between the overall distribution of haplotypes and haplogroups and past human movements, it seems that the European or Caucasoid component observed in the ancient Siberian sample may originate from East European populations. Moreover, it is likely that some mtDNA lineages were carried to southern Siberia from the Volga–Ural region. Incidentally, in the fifth century BC, Herodotus mentioned transit trade occurring in Central Asia along a route that stretched from the Urals in the west to the Altai and the Minusinsk Basin in the east (Hemphill and Mallory 2004). In Altai, the presence of the R1a1 haplogroup in the middle of the fifth century BC is confirmed by the sample SEB 96K2 of Ricaut et al. (2004) which was found to belong to this Y-haplogroup. The boundary of the eastern European influence seems to be fixed at the peri-Baikal area since no R1a1 haplogroup was found in the Xiongnu specimens of the Northern border of Mongolia.

According to the "Kurgan hypothesis" of Marija Gimbutas, nomadic peoples of the Volga steppe region, assumed to speak a Proto-Indo-European language, infiltrated Europe in three waves between 4400 and 2800 BC. Around 4400 BC, Kurgan people from the lower Dnieper and lower Volga regions began moving along the Black Sea littoral into the Danube Basin. They migrated in the Central Balkans and further into Central Europe. During the middle of the fourth millennium BC, the Kurgan culture in the North Pontic Region continued to develop. People travelled across western Ukraine north of the Carpathian Mountains to Poland and Central Germany. They also moved southwest into eastern Romania. Shortly after 3000 BC, the third Kurgan wave (Yamna people), originating once more from the Volga steppe, spread from Central Europe to Northwest Germany, the east Baltic area, southern Scandinavia, the upper Dnieper basin and Central Russia. These three waves of migrations might explain the distribution of mtDNA and

Y-chromosome lineages observed in the present work (Figs. 2, 3).

The Andronovo culture was preceded by the Afanasievo one, which is held to share the closest similarities with the Yamna culture found in the Pontic-Caspian region (Hemphill and Mallory 2004). An eastward migration of the Yamna-derived Afanasievo populations in the Eastern steppe thus provides a possible explanation for the appearance of a European component in the gene pool of ancient south Siberians.

Whereas archaeological records are inconclusive about the anthropological traits characteristic of ancient Siberians, our data deduced from the analysis of human pigmentation gene SNPs seems consistent with the fact that most of them had blue (green) eyes. Indeed, among the SNPs tested was rs12913832, a single DNA variation within a regulatory element of HERC2 gene which is associated to blue eye color in humans. This polymorphism, together with the diplotypes obtained from variations of the OCA2 locus (major contributor to the human eye color variation) showed that at least 60% of the ancient Siberian specimens under study had blue (or green) eyes. Such color phenotype is, according to Eiberg et al. (2008), caused by a founder mutation which most likely originated 6-10 kya from a region around the Black sea, near modern-day Ukraine or Turkey and then diffused into Northern Europe. Our data also suggest that south Siberian specimens might have had blond or light brown hair and fair skin and that they were of European ancestry, a result which appears as evident as those of uniparental markers.

Interestingly, the haplotype of specimen S09 matches that of an ancient specimen from the Yuansha site (Taklamakan desert, Xinjiang Province, northwestern China) and dated back to 2,135 ± 50 years (Gao et al. 2008), suggesting genetic relationships between Andronovo populations and those of the Xinjiang. The Bronze Age inhabitants of the Xinjiang were intrigued at their "Caucasoid" physical appearance and putative "European" origins (Mallory and Mair 2000). Two hypotheses have been offered by archaeologists to account for the origins of these Bronze Age people believed to have spoken an Indo-European language called Tocharian and depicted as possessing red or blonde hair, long noses and blue or green eyes: the "steppe hypothesis" and the "Bactrian oasis hypothesis". Proponents of the latter assert that settlement of the Xinjiang came from sedentary based population of the Oxus civilisation found in Uzbekistan, Afghanistan and Turkmenistan, whereas proponents of the "steppe hypothesis" maintain that the Tarim region experienced a colonization attributed to Afanasievo and Andronovo populations who migrated to Xinjiang from the Altai–Minusinsk regions north of the Tarim Basin (Hemphill and Mallory 2004). Our results corroborate the "steppe hypothesis".

An essential aspect of the present work is the confidence in the validity of our data. Indeed, the field of ancient DNA studies is fraught with technical pitfalls and needs stringent criteria to ensure the reliability of results, particularly when human remains are studied (Hofreiter et al., 2001). In this study, extensive precautions (described in the "Materials and methods") were taken to avoid the amplification of contaminating contemporary DNA molecules. Despite the fact that not all reported criteria of authenticity could be met (Cooper and Poinar 2000), the possibility that our data arose from contaminating DNA was considered highly unlikely. Of course, the reasons as to why some criteria were not adopted have to be explained (Gilbert et al. 2005; Bandelt 2005). Amplified products were not subjected to cloning and amino-acid racemization for the following reasons: (1) an indirect procedure (DNA profiling) was used to assess contamination and biochemical preservation of the DNA samples. Indeed, the (partial) allelic profiles obtained in this study were not mixtures of different individuals' DNA and none of them matched any of those involved in the handling of the bones or DNA samples. Moreover, these DNA profiles were different between each other and testified to the good quality of the extracted DNA (preservation of the nuclear DNA) while showing evidence of its antiquity (allelic drop out, inverse correlation between amplification efficiency and length of the amplification product). Thus, the cold climatic conditions encountered in Siberia had undoubtedly protected the recovered specimens against DNA degradation; (2) the mtDNA sequences were determined based on at least two (often three) independent DNA extracts and PCRs performed on both strands of the DNA; this strategy is costly but efficient in terms of reliability; (3) the quality of the sequences was, in general, comparable to those produced from modern DNA (sharp peaks and little to no background); (4) the mtDNA sequences of poor quality have not been taken into account in this work (samples S12, S17, S30 and S31). Additional criteria of authenticity were considered in this study: results of both sex typing methods (morphologic and genetic) were in accordance with each other (except twice); there was concordance between mtDNA HVI haplotype and haplogroup-defining SNPs along the coding region and also between Y-haplotype and haplogroups. The fact that the mtDNA analysis in our ancient sample revealed the presence of founder mitochondrial lineages as well as of sequences found in other ancient specimens might also be an indication of their phylogenetic antiquity.

To conclude, in this work we demonstrated that some carriers of the Kurgan culture, believed to be Indo-European speakers, were also carriers of the R1a1 haplogroup. These data lend further support to the idea that R1a1 might be a marker to the migration patterns of the early Indo-Europeans, an idea also supported by the recent article of

Haak et al. (2008) in which individuals of the Corded Ware Culture, a culture commonly associated with Indo-European, might bore R1a1 Y-chromosome (as we deduced from their Y-STR typing results). The modern distribution of lineages is the outcome of many millennia of population movements and therefore the assumption of a Proto-Indo-European speaker's homeland in Kurgan region should be taken with great caution. Nevertheless, our study opens possibilities for new debates. We also showed for the first time that Bronze and Iron Ages south Siberian populations displayed "European" physical appearance, thus corroborating physical anthropological records. Another conclusion that can tentatively be inferred from the data presented here is that the Andronovo culture might be the eastern spread of the Kurgan culture and might be related to Tocharian speakers in the Tarim Basin.

**Acknowledgments** The authors are grateful to Prisca Blandin-Frappin, Aurélie Marchet and Sarah Romac for their skilled technical assistance. They also thank Marie Lacan for pertinent comments during the redaction of the paper, Sophie Lienart for her work and the reviewers for constructive suggestions that help improve the manuscript.

## References

Anderson S, Bankier AT, Barrell BG et al (1981) Sequence and organisation of the human mitochondrial genome. Nature 290:457–465

Andrews RM, Kubacka I, Chinnery PF et al (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 23:147

Bandelt HJ (2005) Mosaics of ancient mitochondrial DNA: positive indicators of nonauthenticity. Eur J Hum Genet 13:1106–1112

Belyaeva O, Bermisheva M, Khrunin A et al (2003) Mitochondrial DNA variations in Russian and Belorussian populations. Hum Biol 75:647–660

Bouakaze C, Keyser C, Amory S et al (2007) First successful assay of Y-SNP typing by SnaPshot minisequencing on ancient DNA. Int J Legal Med 121:493–499

Bouakaze C, Keyser C, Crubézy E et al (2009) Pigment phenotype and biogeographical ancestry from ancient skeletal remains: inferences from multiplexed autosomal SNP analysis. Int J Legal Med (in press)

Cooper A, Poinar HN (2000) Ancient DNA: do it right or not at all. Science 289:1139

Derenko M, Malyarchuk B, Denisova GA et al (2006) Contrasting patterns of Y-chromosome variation in South Siberian populations from Baikal and Altai–Sayan regions. Hum Genet 118:591–604

Eiberg H, Troelsen J, Nielsen M et al (2008) Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. Hum Genet 123:177–187

Gao S, Cui Y, Yang Y et al (2008) Mitochondrial DNA analysis of human remains from the Yuansha site in Xinjiang. China Sci China Ser C Life Sci 51:205–213

Gilbert TP, Bandelt HJ, Hofreiter M et al (2005) Assessing ancient DNA studies. Trends Ecol Evol 20:541–544

Gimbutas M (1970) Proto-Indo-European culture: the Kurgan culture during the fifth, fourth and third millennia BC. In: Cardona G, Hoenigswald HM, Seen AM (eds) Indo-European and Indo-Europeans. University of Pennsylvania Press, Philadelphia, pp 155–195

Haak W, Brandt G, de Jong HN et al (2008) Ancient DNA, strontium isotopes, and osteological analyses shed light on social and kinship organization of the Later Stone Age. Proc Natl Acad Sci USA 47:18226–18231

Hemphill BE, Mallory JP (2004) Horse-mounted invaders from the Russo-Kazakh steppe or agricultural colonists from western Central Asia? A craniometric investigation of the Bronze Age settlement of Xinjiang. Am J Phys Anthrop 124:199–222

Hofreiter M, Serre D, Poinar HN et al (2001) Ancient DNA. Nat Rev Genet 2:353–359

Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4:598–612

Karafet TM, Mendez FL, Meilerman MB et al (2008) New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Res 18:830–838

Keyser-Tracqui C, Ludes B (2005) Methods for the study of ancient DNA. Methods Mol Biol 297:253–264

Keyser-Tracqui C, Crubézy E, Ludes B (2003) Nuclear and mitochondrial DNA analysis of a 2,000-year-old necropolis in the Egyin Gol valley of Mongolia. Am J Hum Genet 73:247–260

Kharkov VN, Stepanov VA, Borinskaya SA et al (2004) Gene pool structure of eastern Ukrainians as inferred from the Y-chromosome haplogroups. Russ J Genet 40:415–421

Kharkov VN, Stepanov VA, Medvedeva OF et al (2007) Gene pool differences between northern and southern Altaians inferred from the data on Y-chromosomal haplogroups. Russ J Genet 43:551–562

Koryakova LN, Epimakhov AV (2007) The Urals and Western Siberia in the Bronze and Iron Ages. Cambridge University Press

Kozintsev AG, Gromov AV, Moiseyev VG (1999) Collateral relatives of American Indians among the Bronze Age populations of Siberia? Am J Phys Anthropol 108:193–204

Kravchenko SA, Slominsky PA, Bets LA et al (2002) Polymorphism of STR loci of the Y chromosome in three populations of eastern slavs from Belarus, Russia, and Ukraine. Russ J Genet 38:80–86

Lalueza-Fox C, Sampietro ML, Gilbert MTP et al (2004) Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient Central Asians. Proc R Soc Lond B 271:941–947

Lamberg-Karlovsky CC (2002) Archaeology and language. The Indo-Iranians. Curr Anthropol 43:63–75

Lebedynsky I (2003) Les nomades. Errance, Paris

Mair VH (2005) Gene, geography and glottochronology: The Tarim Basin during late prehistory and history. Journal of Indo-European Monograph Series, 50. Institute for the study of man, Washington, DC

Mallory JP, Mair VH (2000) The Tarim mummies. Thames and Hudson, London

Malyarchuk BA (2004) Differentiation of the mitochondrial subhaplogroup U4 in the populations of Eastern Europe, Ural, and Western Siberia: implication to the genetic history of the uralic populations. Russ J Genet 40:1281–1287

Melchior L, Gilbert MT, Kivisild T et al (2008) Rare mtDNA haplogroups and genetic differences in rich and poor Danish Iron-Age villages. Am J Phys Anthropol 135:206–215

Moiseyev VG (2006) Nonmetric traits in early iron age cranial series from western and southern Siberia. Archaeol Ethnol Anthropol Eurasia 25:145–152

Passarino G, Semino O, Magri C et al (2001) The 49a, f haplotype 11 is a new marker of the EU19 lineage that traces migrations from northern regions of the Black Sea. Hum Immunol 62:922–932

Perićić M, Lauc LB, Klarić IM et al (2005) High-resolution phylogenetic analysis of southeastern Europe traces major episodes of paternal gene flow among Slavic populations. Mol Biol Evol 22:1964–1975

Pike DA (2006) Phylogenetic networks for the human mtDNA haplogroup T. J Genet Geneal 2:1–11

Ricaut FX, Keyser-Tracqui C, Cammaert L et al (2004) Genetic analysis and ethnic affinities from two Scytho-Siberian skeletons. Am J Phys Anthropol 123:351–360

Richards M, Macaulay V, Hickey E et al (2000) Tracing European founder lineages in the Near Eastern mtDNA pool. Am J Hum Genet 67:1251–1276

Rosser ZH, Zerjal T, Hurles ME et al (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography rather than by language. Am J Hum Genet 67:1526–1543

Sampietro ML, Gilbert MTP, Lao O et al (2006) Tracking down human contamination in ancient human teeth. Mol Biol Evol 23:1801–1807

Semino O, Passarino G, Oefner PJ et al (2000) The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. Science 290:1155–1159

Smith CI, Chamberlain AT, Riley MS et al (2003) The thermal history of human fossils and the likelihood of successful DNA amplification. J Hum Evol 45:203–217

Tömöry G, Csányi B, Bogácsi-Szabó E et al (2007) Comparison of maternal lineage and biogeographic analyses of ancient and modern Hungarian populations. Am J Phys Anthropol 134:354–368

Underhill PA, Kivisild T (2007) Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. Annu Rev Genet 41:539–564

Van Geel B, Bokovenko NA, Burova ND et al (2004) Climate change and the expansion of the Scythian culture after 850 BC: a hypothesis. J Archaeol Sci 31:1735–1742

Wells RS, Yuldasheva N, Ruzibakiev R et al (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. Proc Natl Acad Sci USA 98:10244–10249

Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genom Res 12:339–348

Yao YG, Kong QP, Bandelt HJ et al (2002) Phylogeographic differentiation of mitochondrial DNA in Han Chinese. Am J Hum Genet 70:635–651

Zerjal T, Pandya A, Santos FR (1999) The use of Y-chromosomal DNA variation to investigate population history: recent male spread in Asia and Europe. In: Papiha SS, Deka R, Chakraborty R et al (eds) Genomic diversity: applications in human population genetics. Plenum, New York, pp 91–101

Zerjal T, Wells RS, Yuldasheva N et al (2002) A genetic landscape reshaped by recent events: Y-chromosomal insights into Central Asia. Am J Hum Genet 71:466–482