
Supplementary information

An ancient DNA perspective on the Russian conquest of Yakutia

In the format provided by the
authors and unedited

Section 1. Additional information about Yakut history, archaeological survey and ethics

<i>1.1 Background information on Yakutia, the Yakuts and the Russian conquest.....</i>	<i>2</i>
1.1.1 Global context	2
1.1.2 The Sakha Republic: a vast and diverse territory	2
1.1.3 On the use of the term “ <i>Yakut</i> ”	3
1.1.4 The Yakuts: a unique people	4
1.1.5 The origins of the Yakuts as viewed by themselves	4
1.1.6 The origins of the Yakuts as viewed by early explorers and travelers	5
1.1.7 The origin of the Yakuts as viewed by the Soviet school.....	5
1.1.8 The origin of the Yakuts and the archaeological record.....	6
1.1.9 The Russian conquest.....	8
1.1.10 Previous genetic and paleogenetic studies in Yakutia.....	9
1.1.11 Conclusions	10
<i>1.2 Archaeological surveys, stages and samples.....</i>	<i>10</i>
1.2.1 Rationale for the archaeological focus.....	10
1.2.2 Archaeological context: temporal stages and sampling	11
1.2.3 Stage 1: from the early 16 th century CE to the last decade of the 17 th century CE (~1500-1689 CE).....	13
1.2.4 Stage 2: the Yakut golden age, from the late 17 th to mid-18 th century CE (1689-1750).....	13
1.2.5 Stage 3: the second half of the 18 th century CE (1750-1800)	14
1.2.6 Stage 4: the 19 th century CE and the Christianization of Yakutia (1800-1922)	15
<i>1.3 Ethics and engagement with Indigenous Communities.....</i>	<i>15</i>
1.3.1 Global context	15
1.3.2 Legal and ethical aspects	15
1.3.3 Fieldwork, ethics committee and custom clearance.....	16
1.3.4 Student training	16
1.3.5 Public outreach and community engagement.....	17

1.1 Background information on Yakutia, the Yakuts and the Russian conquest

1.1.1 Global context

Siberia spreads over 13 million square kilometers and boasts an exceptional diversity of landscapes, climates, and populations. At the beginning of the 16th century CE (Common Era), the region played host to approximately 500 distinct tribal groups speaking over 120 languages¹. Among these groups, the Yakuts started to stand out as early as the 17th century CE and captured the attention of the first Western travelers (see Supplementary Information section 1.6). Their Turkic language, their economy, which was centered on cattle and horse breeding², and their communities structured into clans and tribes sharply contrasted with those from neighboring populations, consisting primarily of hunter-gatherers organized into small, nomadic groups, occasionally raising small reindeer herds, and speaking Siberian (Tungusic) languages.

The origins of the Yakuts have been a subject of enduring fascination and inquiry. During the Soviet era, new theories have emerged alongside the development of initial archaeological excavations. These studies gained significant momentum from 2004, with the establishment of collaborative partnership between the Yakut and French research teams involved in this study. This collaborative framework was strengthened through a diversity of initiatives, primarily the development of a joint International Research Laboratory from 2011, which was sponsored by CNRS (Centre National de la Recherche Scientifique, *i.e.*, National Centre for Scientific Research), Université de Toulouse (formerly, University Paul Sabatier, Toulouse, France), and the Northeastern Federal University in Yakutsk (NEFU) (see section 1.3.2 below).

1.1.2 The Sakha Republic: a vast and diverse territory

The Sakha Republic spans 3 million square kilometers, nearly the size of India. Its administrative boundaries have evolved over time, with the most recent changes introduced following the establishment of the Soviet Union in 1922, in an attempt to facilitate the ruling of a region located nearly 5,000 kilometers away from Moscow.

The Yakut people, also called the Yakuts, are found in the central region of the Sakha Republic, in the Vilyuy basin, and the basins of the Yana and Indigirka Rivers. These regions represented the focal point of the MAFSO archaeological campaigns between 2004 and 2019 (MAFSO stands for “Mission Archéologique Française en Sibérie Orientale”, *i.e.*, French Archaeological Mission in Eastern Siberia). The Yakut communities have remarkably adapted to extreme climatic conditions. In the coldest areas, such as the Yana and Indigirka basins, Yakut communities are centered around livestock breeding, primarily cattle and horses².

Other Indigenous groups inhabiting the region include the Evenks, Evens, Yukaghirs, Chukchi, and Dolgans. The region also plays host to villages populated by people of European descent, who, in some northern areas, are believed to have preceded the late arrival of the Yakuts. Some of them, are often referred to as the “*Old Believers*”, and harbor their own unique culture and traditions, adding to the diversity of the Sakha Republic. Other parts of the Republic are sparsely inhabited by groups that settled during the 19th century CE, the Soviet era, or more recently, drawn by exploiting abundant natural resources such as diamonds, gold, gas, oil, and timber. Altogether, the Yakuts represent the most common ethnic group in northeastern Siberia, totaling approximately 496,000 people^{3,4}.

1.1.3 On the use of the term “*Yakut*”

The concept of being a “*Yakut*”, along with other ethnic groups, did not exist amongst local groups at the time of first European contact. Terms such as “Dolgans,” “Yukaghirs,” or “Yakuts” were categories enforced by the Russian imperial power as tools of control. The earliest historical references to Yakutia as a territory with internal administrative subdivisions serve as a clear illustration of this process.

The Russian detachment led by Panteleimon Demidovich Pyanda, the first to reach the Lena River in 1620, vaguely referred to “*Indigenous populations*”, merely noticing the coexistence of nomadic and sedentary lifestyles⁵. In 1630, M. Vasiliev, in his report to the Tsar, elaborated a more complex picture, with Tunguses inhabiting the Chona River, Nanagir the Vilyuy River (and forming a nomadic population of approximately 700 individuals), and Dolgans and Yakuts settled along the Lena River up to the mouth of the Vilyuy. The Yakut leader was named Dygintcha, possibly a mispronunciation or an abbreviation for Tygyn. By 1633, A.F. Palizyn, the voivode (military leader) of Mangazeya (a merchant colony then recently founded in northwestern Siberia), sent a report to Moscow further describing the Lena territory, indicating that the inhabitants “*belonged to no one*”, and recommending the establishment of an administration for a “*new Mangazeya*” to control the area and explore its surroundings⁶. In 1634, the state requested additional information from the voivodes of Yenisei, Tobolsk, and Mangazeya, who confirmed that the Yakuts were under “*no jurisdiction*”⁷.

Between 1635 and 1639, the expedition of the “boyar’s son”, Voin Shakhov, collected yasak (tribute) at the mouth of the Vilyuy River from various groups. These included the Tagusy Yakuts, led by Chief Idykhtek, who had 50 men, as well as the pedestrian Yakuts of the Osugei, Kirik, Kokkui, and Kirikin clans, representing 125 Yakuts in total. They also took tribute from 12 members of the Tungus clans and 20 individuals from a Dolgan clan led by Chief Dygintcha⁷. This diversity of groups with distinct lifestyles challenges the idea of a unified nation in early 17th century CE Yakutia.

Moreover, it is highly likely that many fur trappers were attracted into the region during the fifteen years separating Pyanda’s (1620) and Shakhov’s (1635) expeditions. Local groups in this period started to appear as distinct yet interconnected entities, receiving a diversity of names, such as the “*Tagusy Yakuts*” and the “*pedestrian Yakuts*”. In fact, the term “*Yakut*” is thought to originate from the Tungus reindeer herders, who referred to them as “*eko*”, pronounced “*iako*”. The Cossacks transformed this to “*Yakut*” by altering the initial “*e*” sound, which is rarely found at the beginning of Russian words, to the more common “*Ya*”. The suffix “*out*” (“*ut*”) was subsequently added, as seen in other Siberian ethnonyms, such as the Teleouts and Lamouts.

Like for other populations, a single term—“*Yakut*”—cannot encapsulate the complex, heterogeneous, and dynamic populations inhabiting northeastern Siberia before and after the Russian conquest^{8,9}. It has, however, the practical advantage of improving readability and consistency throughout this study, as opposed to changing names for each period considered, as long as it is understood that the underlying group only started to become increasingly unified from 1689/1700 CE, and encompassed a diversity of cultural traits that changed through time. The individuals from the earliest phase, predating 1689/1700 CE, shared distinctive cultural characteristics, such as horsehide clothing (as opposed to local hunter-gatherers, who predominantly wore reindeer hides)¹⁰.

On April 27th, 1922, the Yakut Autonomous Soviet Socialist Republic (YASSR) was established. Following the collapse of the Soviet bloc, YASSR adopted the status of a Sovereign Republic in 1990 that the Yakuts decided to call the “Sakha Republic (Yakutia)”, to emphasize reference to their own language, sakha tȳla. By extension, the term “Sakha” also refers to the people speaking this language.

Today, the region is officially known as the Republic of Sakha (Yakutia), commonly abbreviated as RS(Y). While Yakuts represent the most abundant fraction of the population, Sakha citizens include a more diverse cultural and ethnic mosaic, encompassing Yakuts, Slavs, and members of other Indigenous groups.

1.1.4 The Yakuts: a unique people

The Yakuts inhabit one of the coldest regions on Earth, reaching below -60°C in winter and $+38^{\circ}\text{C}$ in summer¹¹. Despite sharing a latitude with hunter-gatherers who transitioned to reindeer herding between the 16th and 19th centuries CE, the Yakuts have developed a traditional lifestyle centered on cattle and horse breeding. Family potato fields provide an important supplementary food source, as do hunting and fishing in specific areas.

The Yakuts' connection to their horses is profound, extending well beyond their dietary, mobility, and economic roles to symbols of social status and spiritual significance. Horses remain outdoors, living freely in herds except for those used for riding or transport, which are kept near houses. This strongly contrasts with cattle, which are sheltered in stables during winter to survive the extremely cold environmental conditions. Horse herds typically consist of about ten animals, including a stallion, a dominant mare, other mares, and young horses². Horses occupy a central place in the Yakut culture: owning horses not only signified being a member of the Yakuts, but also being recognized as a prosperous member of the society. Rites of passage, religious practices, and many aspects of daily life were infused with this symbolism. The Yakuts thus belong to the broader “*horse civilization*”, a cultural and historical concept that refers to the societies in Central Asia and Mongolia that have historically relied on horses for transportation, warfare, and cultural practices. This civilization thrived from the Bronze Age and continues to endure today².

Before being recognized as a culture or a nation, the word “*Yakut*” refers, first and foremost, to a Turkic language, with rich historical significance, many regional variants¹², and approximately 40% of its vocabulary derived from Mongolic and Buryat languages¹³. There is evidence of intense linguistic contact with Mongolic-speaking groups dating from the time of the Great Mongol Empire¹⁴.

1.1.5 The origins of the Yakuts as viewed by themselves

The myth of “*Ellei Bootur*” is one of the many mythological narratives relating the origins of the Yakuts⁷. Ellei Bootur was a tall and robust man, sometimes described as a shaman, who is said to have migrated from the south, somewhere along the Lena River. He arrived in the land of “*Omogoi*”, a wealthy man with many children who lived off his herds along the riverbanks. Depending on the version of the myth, Ellei Bootur was born in Mongolia or Buryatia, and was the last survivor of his lineage. Other versions present him as Omogoi's younger brother, or a close friend, or even as a foreigner speaking a different language.

In central Yakutia, near the Vilyuy River, Ellei Bootur found a favorable location to settle. There, he met people, whose ancestors had also fled his native region centuries before him. Ellei Bootur was poor and started to work for Omogoi, herding livestock, fishing, hunting, and making hay. After several years of dedicated service, Omogoi offered Ellei Bootur the choice of one of his daughters as a wife. It is believed that Ellei Bootur chose the one who urinated so abundantly that she could emit steam, as a sign of great fertility, a choice that was questioned by Omogoi, who quarreled and offered a small dowry. Ellei Bootur formed a new household nearby, and over time, the couple prospered and had many children, while Omogoi suffered misfortunes. Some versions of the myth claim that Ellei Bootur used shaman's magic to weaken Omogoi, but others states that Omogoi finally accessed to heaven, or eventually reconciled with Ellei Bootur through his daughter. Ellei Bootur's descendants are thought to have formed the Yakut people, and are credited with the introduction of central elements of the Yakut culture. Some versions of the myth claim

that Ellei Bootur introduced new practices, such as preparing fermented mare's milk ("kymys"); organizing the grand festival of "Yhyakb" to honor the spirits; creating ritual songs, and; carving drinking vessels, all of which were previously unknown to Omogoi.

There is little doubt that the myth of Ellei Bootur reflects additions and reinterpretations from the 19th and 20th centuries CE, although the traveler I. Lindenau documented one of the earliest known versions around 1745¹⁵. It is, however, important to bear in mind that this myth is not exclusively Yakut in origin but shares similarities with the stories of reindeer herders from the Sayan Mountains, located on the southern border of Siberia and western Mongolia. Other accounts of the myth introduce "Tygyn", the leader of the Khangalassy clan—one of the earliest to establish contact with the Russians—as one of the many descendants of Ellei Bootur¹⁶. Tygyn is thought as the founding father of eight Yakut clans in central Yakutia, particularly in the Megino-Khangalassky region, all of which represent a significant part of the Yakut identity.

1.1.6 The origins of the Yakuts as viewed by early explorers and travelers

The origin of the Yakuts fascinated explorers and travelers alike as early as the first half of the 17th century CE. Unlike 19th-century ethnologists and 20th-century CE researchers, these early observers had access to first-hand Indigenous accounts. By the late 17th century CE, travelers such as the French F. Avril¹⁷, the Dutch E. Ides¹⁸ and N. Witsen¹⁹, and the German P.J. von Strahlenberg²⁰ hypothesized a southern origin for the Yakuts, based mostly on linguistic evidence and, at times, the physical appearance of the people in the region. Through a series of bold deductions²¹, they posited that the ancestral Yakut groups arrived in the territory of present-day Yakutia during the 13th or the 14th century CE²². E. Ides, who was part of the Russian embassy in China, was the first to suggest such an origin. His student, N. Witsen, had access, as early as 1664, to the first Russian and Cossack reports¹⁹, indicating a common origin for the Yakuts and the Buryats. These reports alleged that the Yakuts once lived in the Tobol region among the Kalmyk people, and stole livestock from the Buryats. The Buryats, who lived in the lake Baikal region, would have retaliated with an attack that forced the Yakuts to migrate further north to the Lena River region. E. Ides suggested that the arrival into the Lena River region occurred 200 years before the publication of his book in the late 17th century CE.

At the beginning of the 18th century CE, the Swedish researcher Y.I. Lindenau, who worked as a translator at the Academy of Sciences in Saint Petersburg, further developed this model using a combination of linguistic evidence and myths collected in the region. His work locates the homeland of the Yakut culture across the Transbaikal region and the upper Lena River, with the Yakuts arriving in their present-day territory under the leadership of "Badjer" and his sons, during the second half of the 16th century CE¹⁵.

Until the Soviet era, many ethnologists and linguists have revisited the question of the Yakuts' origins. Ethnologists, such as the Polish political exile V. Serochevskiy²³, expanded upon earlier hypotheses, incorporating broader, and sometimes more interpretive, versions of the myths²³. Linguists and members of the Yakut intelligentsia in the early 20th century CE demonstrated that the Yakut language was not merely a Turkic language with Mongolic borrowings, but suggested more complex origins²⁴. While the southern origin of the Yakuts remained undisputed, the identification of a single precise location proved challenging.

1.1.7 The origin of the Yakuts as viewed by the Soviet school

During the Soviet period, archaeological excavations flourished, and researchers compiled comprehensive ethnological records⁵. The discoveries of frozen tombs in the Altai in 1940-1950²⁵, which were readily

accessible and referenced, had a strong influence on Yakut archaeologists and ethnographers. These frozen tombs led to the adoption of a vocabulary directly referring to periods that were more than two millennia old and sometimes geographically distant by thousands of kilometers. Examples include “*Pazyryk*” types that referred to certain types of birch coverings on graves, or the “*Scytho-Siberian style*” describing many decorative artifacts²⁶.

In the 1930s, Ksenofontov contributed significantly to assimilating the Yakuts into the “*Sakha*” people, of Eastern Scythian origins, or even Xiongnu, for which a Russian archaeologist had excavated several sites a few years earlier^{27,28}. This trend continued throughout the Soviet era²⁹. A prestigious, direct continuity was implicitly assumed, and certain features closely resembling those found outside the Altai were overlooked, even though potentially meaningful. One such example is found in the double funerary chests typical of 18th-century CE graves, showing artifacts deposited between both coffins. The closest and most geographically relevant parallels are known in Xiongnu cemeteries from Mongolia, a tradition dating back to nearly two millennia, and possibly stemming from local Altaic or Chinese practice³⁰. Similarly, researchers regarded so-called “*chorons*”—wooden goblet vessels with ritual purposes (Fig. 1d)—as cultural elements unique to Yakuts. However, anthropomorphic Turkish steles from the migration period of Turkic-speaking ethnic groups (7th-10th centuries CE), found from Mongolia to Central Asia, frequently depict warriors holding goblets, reminiscent of *chorons*³¹.

Other distinguished Soviet scholars, such as A. P. Okladnikov (1908–1981), also delved into the history of the Yakuts. They naturally conducted their research within the framework of the Soviet school, which sought to understand the emergence of various peoples. Their ethnogenesis framework conceptualized fixed entities, such as the Yakuts, and defined the origins as successive proto-Yakut stages influenced by surrounding groups labeled according to contemporary populations: the “*Tungus*”, the “*Yukaghirs*”, and others. Adding to this complexity, the political and scientific environment of the time reshaped this approach, with the period between the 1930s to 1950s marked by significant censorship and alignment to the directives of the Communist Party. It was then unacceptable to suggest that Indigenous peoples might have foreign origins, *e.g.*, from Mongolia, and researchers were expected to emphasize local origins. As a consequence, S. Tokarev highlighted in 1945 the presence of Siberian elements to argue that the Yakuts descended from local populations that gradually integrated diverse Turkic elements arriving in Yakutia via various migratory routes³². This theory was rejected by A. P. Okladnikov in his major work⁵ from 1955 on the basis of novel archaeological, linguistic, and ethnological evidence. Okladnikov favored the Lake Baikal region as pivotal to the Yakuts’ formation, and recognized later contributions from Tungus, Chukchi, and Yukaghir components. Later on, Okladnikov’s intellectual trajectory appears to have been influenced by T. D. Lysenko (1898–1976), who dismissed genetics as a bourgeois science, and advocated the theory of the inheritance of acquired characteristics³³. Okladnikov revisited his theory on Yakuts’ origins, positing that a population originally breeding horses in more southern latitudes had undergone significant adaptation while settling north in Yakutia³⁴. The idea that the Yakuts could have transitioned into reindeer herding in more recent history gained prominence during this period.

1.1.8 The origin of the Yakuts and the archaeological record

Archaeological research on the origins of the Yakuts truly took off with the work of A. P. Okladnikov³⁵, and was followed by major contributions from A. Alexeev^{36,37} and A. Gogolev^{38–40}. Subsequently, and in collaboration with Russian colleagues, the nascent French-Yakut collaboration underlying this study led to the discovery of novel ancient tombs, the vast majority of which are presented here (see section 1.2.2), and advancements in dendrochronology as well as radiocarbon dating methods contributed to clarify chronological timelines⁴¹.

Current evidence suggests the presence of remnants of southern populations, between the 4th and 9th centuries CE, corresponding to very early exploration or settlement phases. It is noteworthy that A. N. Alexeev uncovered fragments of bows with arrowheads that closely resemble the weaponry of Central Asian nomads, including the military complex of the Türks in Mongolia²⁶, in the southwestern most range of present-day Yakutia. More specifically, the findings were distributed along the Tokko River, a tributary of the Olyokma River, which flows into the Lena River, and remains peripheral to the central Yakut homeland. Near the Lena River and approximately 200 kilometers south of central Yakutia, archaeologists have found the Shishkino petroglyphs⁴², which depict horsemen with banners potentially symbolizing clans and are believed to date from the same time period. Others petroglyphs are located further north, about 50 kilometers south of Yakutsk. They feature a large boat with numerous rowers. The petroglyphs, dating back to the Bronze Age through the 9th century CE, are associated with the northernmost known Turkic inscription^{43,44}. Determining the exact nature of these discoveries and petroglyphs, with their profound historical significance, is challenging.

The southern part of what is now the Sakha Republic, favorable for agriculture and somewhat comparable to the Lake Baikal region, may have been occupied early on by populations similar to those of northern Mongolia and the Baikal area. The Yakut homeland, beyond the so-called “*Lena Pillars*”, in more northern zones, was certainly explored early, including by horse riders like those associated with the Pokroskt site⁴⁵, which might have settled in favorable areas (the Pokroskt site is located in the largest alluvial plain of the Lena River). However, Yakut settlements appear to be more recent.

Two main cultural phases, or “*economic complexes*”, can be distinguished among archaeological settlements. These may be partially contemporaneous and may represent the direct ancestors of the Yakuts, or comprise some of the contributors of their modern gene pool.

The first was recognized as early as the 1940s by Okladnikov⁴⁶, and corresponds to the so-called “*small houses*” culture, with semi-subterranean dwellings, primarily located in the basins of the Yana and Indigirka rivers. While the dating remains uncertain and varies between the 13th and 16th centuries CE—with groups or cultural influences potentially extending until the 17th century CE—, archaeological contexts suggest a hunting and gathering economy, with no apparent connection to the later Yakut economy²⁶. According to A. P. Okladnikov, this phase corresponds at best to villages located “*precisely in the region where early 17th-century Russian written sources mention the Yakuts on foot*”⁴⁶, which describes overlapping geographic ranges but no common filiation.

The second main cultural phase is known as the Kulun-Atakh culture. It was defined by A. Gogolev^{38–40} to extend from central Yakutia to the border areas between central Yakutia and the Yana and Indigirka River basins. Two radiocarbon dates of items belonging to the earliest phases of this culture indicate a time period between the 10th and 13th centuries CE²⁶. Generally, Kulun-Atakh villages are found near so-called “*alas*”, which corresponds to vast meadows used for haymaking and grazing⁴⁷. Developing settlements at the proximity of an alas is a strategy that persisted in central Yakutia until the 20th century CE. The base of the economy seems to be mostly livestock oriented, including cattle and horses, although remains of wild mammals, birds, and fish have been discovered²⁶. The objects excavated testifying of people’s daily life show clear analogies with the cultural complexes found in the early Iron Age of present-day Yakutia, which may indicate a degree of cultural continuity²⁶.

The relationship between these archaeological cultures, and their respective relationships with those found further south, particularly the Kurumchi archaeological culture, remain unclear. Between the 5th and 10th centuries CE, the Kurumchi culture thrived in the north of present-day Mongolia, right in the areas located in the southern part of what is now the Sakha Republic (*i.e.*, downstream of the Selenga River, in the Tunka region, the Barguzin Valley around Lake Baikal, the Angara Valley, and the upper Lena River). The people of this culture, known as the Kurykans, practiced livestock farming, especially cattle, sheep, and

horses, and managed camels⁴⁸. According to analyses of petroglyphs, some members of this culture could have been familiar with writing⁴⁸. Decorative motifs characteristic of this culture are found on Yakut ceramic vessels as late as the 19th century CE, potentially indicating previous contact.

The earliest graves that have been excavated in present-day Yakutia date from the 15th to the 17th century CE. They predominantly involve male subjects buried on their backs, although some less common individuals were found buried on their sides, or with their lower limbs flexed⁴⁹. Some of those early graves, such as AtDaban0, are located on permafrost mounds (so-called “*pingos*”), which are reminiscent of the kurgans found in southern regions⁵⁰. Only three female graves from this period has been excavated so far. One of them belongs to the late 15th or early 16th century CE, and preserves Atlasovka, a crucial ancestral figure within the kinship networks reconstructed in this study (Fig. 4). This woman wore a headdress shaped like a hemispherical-topped hat. The headdress had a central opening with a rod to hold a plume and two decorated triangular plaques featuring complex raised patterns. Similar hemispherical-topped headdresses have been identified in the archaeological record between Lake Baikal and the mountains bordering Mongolia and China²⁶. Overall, the funerary practices of those subjects buried prior to the 17th century CE are characterized by a predominant cultural tradition, coexisting with seemingly diversified traditions^{7,49,51}.

1.1.9 The Russian conquest

Historical sources depict the Russian expansion into Yakutia as a process in which colonization through population settlement remained marginal^{52,53}. This is largely due to the extreme Yakutian climate, which hindered agriculture akin to that of the Russian plains or southern Siberia impossible. The alluvial plains of the Olyokminsk region, stretching over 50 kilometers along the left bank of the Lena River and covering approximately 500 square kilometers, is the only area remotely reminiscent of the peri-Baikal zone. It provided favorable conditions for agriculture, particularly wheat cultivation, to settlers of Russian origins, often called the “*Old Believers*”, who remained to this day isolated from Indigenous populations. The local sanitary conditions also often proved lethal for newcomers, as shown by the mortality rates of political exiles and prisoners sent to Yakutia in the early 20th century CE, estimated between 24% and 63% due to tuberculosis infections alone⁵⁴. Combined, the considerable limitations to agriculture and the sanitary conditions strongly limited the capacity of the Russian power to install demographically thriving colonies⁵⁵ (Fig. S1_1).

It is a well-accepted historical fact that the Russian imperial state initiated the conquest of Yakutia to exploit their abundant fur resources^{56,57}. However, control over these territories could not be achieved from Moscow and Saint Petersburg elites without maneuvering the complex local political landscape and establishing dynamic, strategic alliances with specific tribes⁷. While interactions were challenging and often violent when the first Cossacks arrived (1632–1689), alliances with Yakut leaders helped develop tax and administrative control of the local production. However, external control remained in practice limited, as Yakuts, often helped by local leaders, developed successful strategies to avoid taxes^{7,58,59}. This complex system of local alliances and tax avoidance strengthened the Yakut leadership, who established vast sedentary estates, maintaining direct ties with the tsar, but not necessarily with local administrators, and developed trade networks (circa 1689–1750), often to the detriment of the imperial authorities. This process has been coined “colonial neolithization”⁶⁰, and led to a golden age of relative prosperity and cultural development for the Yakuts at the expense of other hunter-gatherer neighbors, such as the Even.

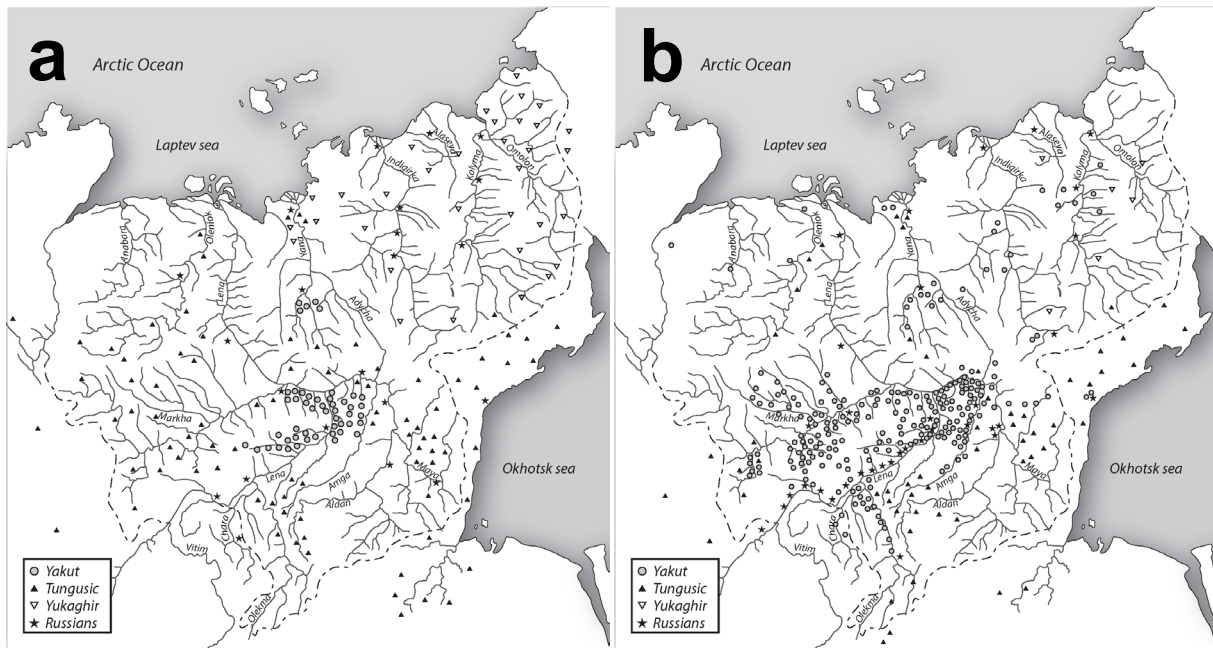


Figure S1.1. Maps showing the expansion of tribes and Russian settlement, according to the censuses of 1676 (a) and 1897 (b). We can observe that, except for a few forts during the period in question, Russian settlements always remained to the south of Yakut settlements. According to Jochelson, W. (1993) *The Yakuts*. The American Museum of Natural History. Volume XXXIII, 225p⁵⁵.

1.1.10 Previous genetic and paleogenetic studies in Yakutia

The first genetic studies focused on the origins of Yakut populations were based on uniparental loci (mitochondrial DNA, mtDNA) and the non-recombining portion of the Y chromosome (NRY), and emerged within the context of research on the peopling of the Americas and its connections with continental Asia^{61,62}. These analyses revealed a significant bottleneck in Yakut male lineages^{63,64}. Subsequent research, particularly by B. Pakendorf and her colleagues^{14,65}, has combined genetic and linguistic approaches to shed light on the southern Siberian origins of the Yakuts. It supported the hypothesis that the Yakuts represent a highly homogeneous population and revealed a pronounced genetic discontinuity with some of their neighboring groups, such as the Yukaghirs. While there was no conclusive evidence regarding admixture with Indigenous populations, the presence of a minor contribution could not be entirely ruled out based on mtDNA data alone. Y-chromosomal analyses further corroborated earlier findings of a significant bottleneck event in the Yakut population, the timing of which aligned well with the hypothesis that the Yakuts migrated northward under pressure from the Mongols⁵. Based on mtDNA analyses^{65,66}, a founder effect involving approximately 150 female individuals was proposed to have occurred a little over 1,000 years ago⁶⁷. This founder effect, involving both male and female individuals, could be the origin of genetic diseases that are typically rare but observed with greater frequency in Yakutia⁶⁸.

Further genetic studies based on autosomal SNP variation and an increasingly comprehensive data set of mitochondrial DNA and Y-chromosome sequences, established the genetic proximity between the Yakut populations and some of their other geographically and linguistically neighboring groups, particularly Tuvinians and Buryats and groups speaking Turkic South Siberian lineages⁶⁹. With increasing work, it became apparent that multiple admixture layers could be detected in most Siberian populations, indicative of significantly different admixture trajectories in each individual population⁷⁰. Paleogenomic researchers confirmed this by studying a few ancient individuals from Yakutia and west of Lake Baikal. These revealed

significant population transformations from the Late Upper Paleolithic to the Neolithic, and during the Bronze Age^{71,72}.

Going beyond the mere origin of individuals or the history of settlement, paleogenetic studies⁶⁰—fueled by the exceptional preservation of bodies in the permafrost^{45,73}—have delved into the establishment and distribution of the population from the 15th century CE to the 19th century CE⁷⁴, the enduring influence of paternal clans over four centuries of Yakut history⁷⁵, the identification of infectious diseases that plagued individuals^{76–78}, and the investigation of kinship relationships (while limited in resolution due to the challenges in amplifying autosomal STR markers^{79,80} designed for modern DNA).

1.1.11 Conclusions

Linguistic, archaeological, and paleogenetic data converge to place the Yakuts' origin in the peri-Baikal region. However, hypotheses regarding the timing of their foundation and arrival vary. Furthermore, given the cultural diversity observed during the period ~1500-1689 CE (Stage 1 as defined below in 1.2.2) and the cultural changes between the first and second stages, the question of an origin involving multiple genetic substrates and migrations arises. One of the major unresolved questions regards the level of admixture with an ancient local substrate. Beyond these elements related to settlement history, Yakutia offers an exciting opportunity to study the evolution of burial sites and the biological biographies of individuals buried at a crucial moment in human history.

1.2 Archaeological surveys, stages and samples

1.2.1 Rationale for the archaeological focus

No research project could endeavor to survey a territory of 3 million square kilometers, an area almost as large as India. To focus on relevant regions informing on the history of Yakuts, we first relied on maps published in 1933 by Jochelson⁵⁵, a Russian ethnographer, representing the area occupied by local populations and Russians between 1676 and 1897 (Fig. S1_1). Various ethnological texts, described in detail through the study from Crubézy and Nikolaeva (2017)⁷, also informed on those areas reflecting the Yakut history. Jochelson's maps largely relied on the tax collection records at the time. The 1676 map indicates the censuses of Yakuts in central Yakutia and the Vilyuy region only. The 1897 map reveals a Yakut expansion westward, northward, toward Verkhoyansk, and along the Indigirka River, movements also well-attested in ethnological texts. It is noteworthy that there were no formally individualized Russian settlements in 1676, as mercenaries moved between a few forts. By 1897, Russian villages, where the so-called “*Old Believers*” lived, were located south of the Lena River and were distinct from Yakut settlements.

Based on these foundations, we decided to conduct on-foot surveys of ~44,100 square kilometers (Fig. S1_2). Survey work took over 16 years and began in 2002. The survey areas included: Central Yakutia (the area with the densest Yakut settlement), the Vilyuy region, the Indigirka area, and Verkhoyansk, representing the four regions investigated in this study. The survey involved highly-trained teams of archaeologists, with high experience in fieldwork, as well as local populations.

Across the ~44,100 square kilometers surveyed, we excavated all graves that could be identified, whether isolated, or separated by only a few dozen meters. These graves encompassed all four archaeological

stages defined in 1.2.2, including Stage 4 (1800-1922) associated with increased Christianization. Overall, Christian cemeteries were scarce within the geographic areas surveyed. This contrasts with the region of Olokminsk (southern Yakutia), which is known for the absence of Yakut settlement in the past, and the high density of Christian cemeteries. These cemeteries were not excavated. The few Christian cemeteries identified in other areas, were generally limited to only a few graves, some of which were excavated.

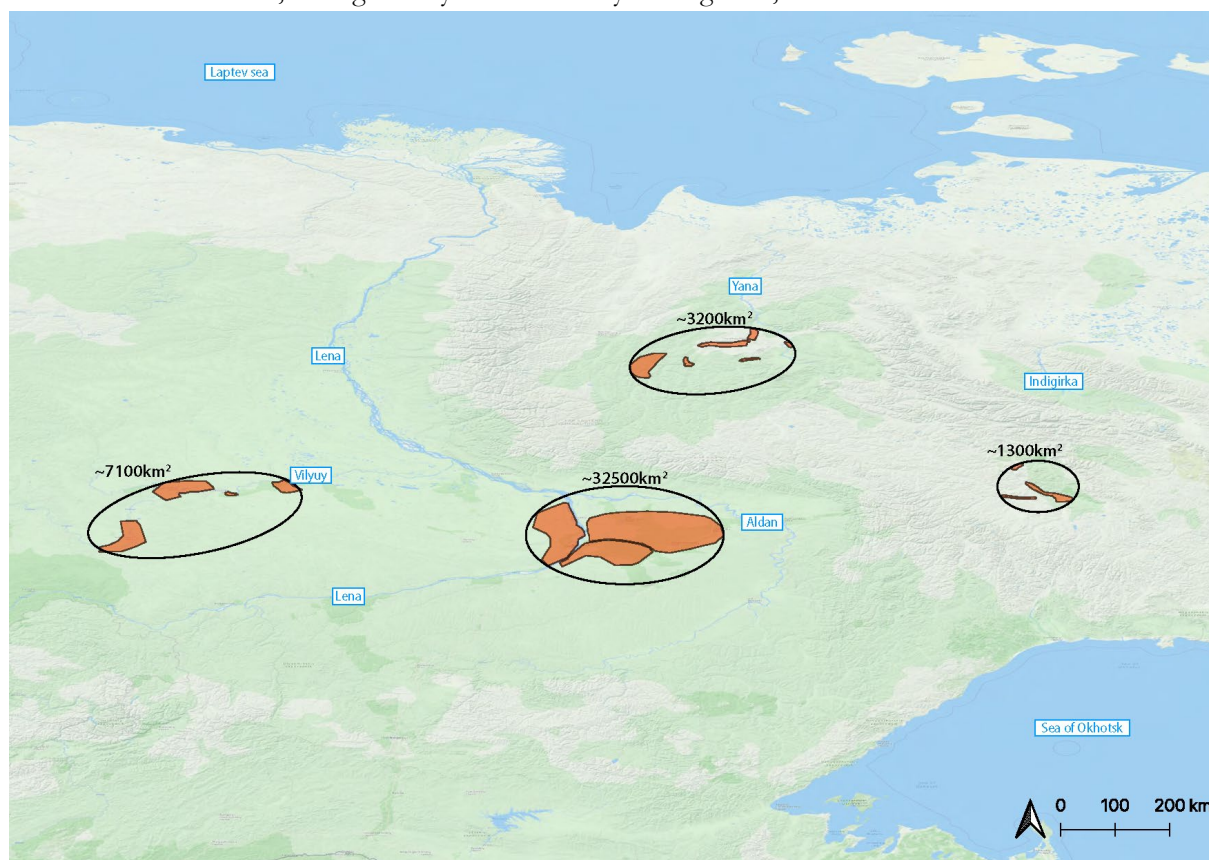


Figure S1_2. Map of surveyed areas in the Republic of Sakha.

Prospected zones are in orange, with cumulated area in ellipses. Historical texts indicated Yakut settlements in central Yakutia, particularly along the Lena River, Vilyuy, Indigirka, and Verkhoyansk, within the Yana River basin. The availability of old maps and historical documents enabled the identification of key early settlement areas. Experienced researchers have employed a systematic approach in exploring all these regions on foot to locate potential burial sites.

1.2.2 Archaeological context: temporal stages and sampling

Funerary practices in Yakutia evolved from the earliest burials of the transition between the 15th and 16th centuries CE, to 1922, the year marking the start of USSR, which radically transformed funerary practices⁷. A total of four main stages can be defined on the basis of the architecture of tombs and associated grave goods, correlated with radiocarbon dating and dendrochronological data⁴¹. These four stages are consistent across all excavated areas, with striking tomb similarities belonging to the same period, even when located 1,500+ kilometers apart. These changes reflect profound and far-reaching social, economic, and cultural transformations accompanying the time period investigated in this study (Extended Data Fig. 1).

No European influence is discernible in the tombs predating year 1689 (Stage 1), despite the presence of Europeans from 1632, which underscores local funerary practices devoid of exogenous influence, possibly reflecting the often-conflicting interactions between local and non-local populations⁸¹. In 1689, a Yakut clan formed an alliance with Cossack troops and triumphed at the Battle of Pokrovsk, overtaking all other Yakut clans and establishing dominance over a significant part of Yakutia⁷.

Concomitantly, a commercial post was established in Nerchinsk (Transbaikal region, southwestern Siberia), to foster trade and exchange between Russia and China⁴⁹. This commercial post would largely benefit Yakutia in the following decades⁷. During the following 60 years, Yakutia experienced a true golden age (1689-1750, Stage 2). Confident clan leaders monopolized the lucrative fur trade with China on the back of Russian authorities, which ushered a period of tremendous prosperity for the local elites^{7,59}. This period also marks the development of a new form of shamanism, also called familial⁸² or heterarchical⁸⁴, shamanism. This shamanistic practice is directly apparent in the archaeological record and restricted to female burials, although some tombs belonging to male chiefs can contain distinctive attributes reminiscent of shamanism (hence, their frequent interpretation as “*princes who were also shamans*”)⁷. Engravings found in the tombs indicate that this familial shamanism coexisted with another form of shamanism, known as ceremonial costume-based, or hierarchical, shamanism⁸⁴, and reflected the influence from various cultural traditions. The archaeological record also indicates that the ceremonial attire and shamanic accessories of hierarchical shamans became increasingly elaborate between the 18th and 19th centuries CE⁸². It is important to note that ethnographic accounts from the 19th century CE only partially align with the materiality of the archeological record regarding shamanism.

The period from 1750 to 1800 (Stage 3) corresponds to a period of economic reorganization under Russian state control⁵⁸. This was mainly the consequence of the relocation of the trade post between Russia and China, which now increased the distance separating Yakutia and the major commercial routes of the time⁴⁹. This period was, therefore, accompanied by a strong decline in regional prosperity⁷ but it also saw the first step towards Christianization, as apparent in the material culture deposited in a fraction of the burials. In the following period (Stage 4, 1800-1922), Yakutia fully integrated the Orthodox Russian world, and Christian symbols became omnipresent in burials; dietary habits and economic circuits were also reshaped during stage 4^{84,85}.

During the first two stages, the majority of the deceased were placed on open-air platforms, leaving no enduring trace in the archaeological record⁸². The material culture excavated with the buried minority (*e.g.*, clothes, associated grave goods) revealed, however, a complex social stratification, also apparent in the location of the burials in the landscape and the architecture of superstructures and coffins. Such funerary practices may reflect variation in familial preferences, although a significant proportion of those buried individuals correspond to persons whose return among the living was feared, such as respected leaders, shamans, suicides, or individuals who died a “bad death” during an epidemic or in tragic circumstances^{7,49}. Additionally, it cannot be ruled out that for some individuals, the practice of burying per se could represent the earliest perceptible signs of nascent Christianization, even in the absence of religious symbols⁸⁶.

The toponymy of old maps⁸⁷, and orally-transmitted traditions still known to local people, provide important insights into the social status of the deceased. For instance, local Yakuts specifically recall of certain tombs that belong to clan leaders who traveled to Moscow in the early 18th century CE, and were ennobled when meeting with the tsar. Such narratives were corroborated by the armorial engraving found on signet rings during excavations⁸⁸. Local accounts also formally designated the tombs of three female shamans, including two analyzed in this study (KyysOunouoga and Sordonokh). These tombs clearly stood out during excavation through their accumulation of symbolic elements, such as fingers sewn into sleeves, bindings cut, and pendant belts, otherwise known as “virgin belts”. These belts are characteristic of wedding⁸⁹ attire and assess the virginity of the bride, which could sometimes be confirmed during the autopsy of frozen bodies; they were also known to be worn by female shamans⁹⁰, particularly feared virgins, and their jingling was meant to repulse evil spirits^{7,49}.

The presence of similar belts helped identify other shamans than the three designated by local Yakuts, especially those located at the so-called “Arbre Chamanique” (AC) tomb (in English “the Shamanic Tree” tomb; archaeological Stage 2, central Yakutia). This site, where shamans are known to have gathered until recent times to perform animal sacrifices, features animal skins hanging from a tree. The AC tomb is

a multi-burial tomb that provides invaluable insights into the shamanic practices and beliefs during archaeological stage 2. It contained a “*tuktyé*”—i.e., a birch-bark box used to confine the spirits of those who had died unnatural deaths, or those who were feared for their potential malevolent come back (“*üeurs*”), or again the spirits of shamans who had died during rituals (“*kamlanie*”)⁸². A pendant belt was also found locked within a bag.

The adult woman buried there (AC1S3), for whom a relatively high inbreeding coefficient was estimated (Fig. 4), had her hair tied and covered with a fur cap. A fur was also shrouded on her face, which represents a common ritual practice at the time. However, the cap and shroud were overlaid with a so-called “mask-hat”⁹¹. This headdress featured two small horns at its apex, decorated with tiny blue, black, and white beads, and a leather strip with tiny perforations positioned over the eyes. Small blue beads adorned the edges of this strip. While the first discovery of this kind across Yakutia, a literature survey indicates that mask-hats were worn in at least two main circumstances: either by married women, since Yakut women were not allowed to show their faces to their fathers-in-law or older male relatives of their husband’s⁹², or by shamans on specific occasions^{91,93}. Interestingly, AC1S3 is the only adult woman excavated in Yakutia with such a mask, and the descriptions of these marital masks⁹² do not match the mask she wore. The position over a shroud indeed poorly aligns with an interpretation as a marital mask, which may better fit an interpretation as a shamanic mask, bearing in mind that female shamans played a central role in rituals such as “*Ysyakh*” (a celebration of nature’s renewal), and often wore elaborate costumes, decorated headdresses, and accessories symbolizing rebirth and transformation⁹³.

The case of the UsSergue1 burial (archaeological stage 3, central Yakutia near At Daban) is also worth mentioning in the context of shamanism. *Us Sergue* translates into “*the three posts for tying horses*”. Near the grave was a pit containing three horse sacrifices. While rare, this practice was also found in the Vilyuy region in the tomb of an adult female shaman wearing a virgin belt. Individual labelled UsSergue1, was identified as the great-granddaughter of AtDaban6 (Archaeological stage 1, central Yakutia). She was the second wife of *Mazary Bozhekkov*, one of the most prominent clan leaders who established political and economic relations with the Russians⁷⁵. Her tomb is one of the richest in Yakutia, comparable in both numbers and quality of items, to that of Eletchei1. Eletchei1 did not, however, wear a virgin belt, and unlike the other contemporaneous tombs, UsSergue1’s tomb retained all the attributes of an ancestral burial tradition. In the context of increasing Christianization at the time, this tomb illustrates the persistence of a familial shaman belonging to Yakutia’s most influential clan.

1.2.3 Stage 1: from the early 16th century CE to the last decade of the 17th century CE (~1500-1689 CE)

This stage includes essentially male burials, with the notable exception of Atlasovka (Fig. 1). Burial graves are often located on elevated terrain overlooking landscapes and lakes, although they are not consistently associated with extensive grasslands. These elevations, often formed by permafrost mounds, are sometimes so marked that an uninformed observer, such as the Yakuts in earlier times⁸⁷, might confuse them with kurgan burial mounds, which are typical of earlier South Siberian elites⁵⁰. Local 19th-century CE legends often describe these graves as those of hunter-warriors, family heads, or their children⁷.

1.2.4 Stage 2: the Yakut golden age, from the late 17th to mid-18th century CE (1689-1750)

Funerary practices reflect a golden age, with male and female burials typically associated with a whole array of social statuses and grave goods⁹⁴. Some burials contain indeed hundreds of thousands of beads, while others only include a few. Burials can be grouped into three main categories:

- Category 1: At Daban graves

Situated on a Quaternary terrace south of Yakutsk, overlooking the vast alluvial plain of the Lena River, this site represents a clan cemetery attributed to the Kangalaskii clan. The clan came to dominate part of Yakutia as it formed a strong alliance with the Russian power and separated male and female graves into distinct areas, especially at the location site called “At Daban”⁹⁵. In one prominent burial, a young girl wearing a virgin belt was found in a position overlooking the women’s section of the site⁹⁴. Several male subjects at this site exhibit familial ties at the first to third degree (*e.g.*, AtByran7, AtDaban4, AtDaban11, AtDaban0).

- Category 2: meander burials at the foot of At Daban

These graves reflect diverse cultural influences, and are found on terraces located above river meanders descending from Quaternary terraces toward the Lena River (*e.g.*, Oktiom and Oulakh). In the multiple burial of Oktiom, associated with a smallpox epidemic (presented in the study), the only ceramic discovered belongs to the so-called “small house” cultural tradition, an ancient feature²⁶ that may have persisted until the 17th century CE. In another meander, the grave Oulakh4 is strikingly similar to the “Tungus Buryat” burial found in the Barguzin Valley of Buryatia⁹⁴. Planks initially intended to line the interiors of *Ourassa* (summer dwellings resembling conical tents with birch bark exteriors and wooden interiors) covered the Oktiom and Oulakh5 burials⁹⁴. Some of these planks featured carvings of professional or hierarchical shamans, recognized by their fringed costumes⁸³. The associated archaeological evidence points to a complex society, blending local and external influences, where funerary rituals both express individual identities and collective affiliations. Of note, individuals buried in the same meander burial often share familial connections (*e.g.*, the multiple grave of Oktiom and Oulakh2, Oulakh4, and Oulakh5).

- Category 3: promontory graves

Promontory graves are scattered across the four regions studied, and mainly located on promontories, overlooking alas (*i.e.*, thermokarst lakes surrounded by pastures), or adjacent grazing plains. These locations reflect a time of social transformation in which group mobility decreased, and the concept of territorial ownership emerged⁷. The graves, overlooking family lands, symbolically affirmed territorial anchoring and authority. These sites have yielded burials of young girls or female shamans. These graves associated with an alas sometimes reveal kinship relationships around those of the same alas (Okhtoubout2 and AC1S2), but rarely beyond a few kilometers (Celysse and TungusKeul4). At Verkhoyansk, all buried individuals were related. Of note, a second-degree familial relationship was found between a heterarchical shaman (Sordonokh) and a hierarchical shaman (Kouranakh).

1.2.5 Stage 3: the second half of the 18th century CE (1750-1800)

This period corresponds to an era of economic decline in the region, following devastating smallpox and tuberculosis epidemics⁷⁸. This period, and the incapacity of shamans to protect against epidemics of a new kind, marked the decline of shamanic traditions, paving the way for Christianity⁷. Local narratives provide insights into the social structures of the time. For instance, the grave of Myran (Verkhoyansk), dating to the subsequent period, belonged to an elderly Christianized woman whose grave became the starting point of a Christian cemetery. This particular situation might be explained by the fact that she was the mother of a hierarchical shaman (Kouranakh) who had died as a young adult. His grave, featuring a solar disk on her chest and a headdress with a pendant covering his eyes, underscores a symbolic role⁹⁴. Missionaries often focused their efforts on christianizing the wives of elites, especially if their sons held shamanic roles. Interestingly, Kouranakh is a close relative of Sordonokh, a powerful heterarchical shamaness known from local accounts. Sordonokh wore a virgin belt and also possessed a solar mirror that the officiants had hidden under her clothing near her thigh⁹⁴.

1.2.6 Stage 4: the 19th century CE and the Christianization of Yakutia (1800-1922)

This period was associated to a shift in burial practices as most individuals were buried. The various funerary assemblages include: (1) cemeteries with graves, some of which feature funerary stelae⁹⁴ (*e.g.*, Indigirka: TouoraUrekh2 and TouoraUrekh 3); (2) clusters of two to three graves, which are unusual for earlier periods and can display distinctive architectural features on the surface, such as Christian crosses⁹⁴ (Indigirka: Omouk1 and Omouk3; Verkhoyansk: Byljazyk2), or not (Vilyuy: OulgoundaBerete2, OyogosseTumula2 and OyogosseTumula3), and; (3) isolated Christian graves, sometimes of newborns (Indigirka: Balagannakh2; Verkhoyansk: Khoumakhtaakh; Vilyuy: Tchenkeloeil), occasionally adults (Central Yakutia: Toutekh, Koudouk1, and Otchugoui).

1.3 Ethics and engagement with Indigenous Communities

1.3.1 Global context

While integrated into the Russian Federation, the autonomous Republic of Sakha stands out for its unique historical and cultural characteristics. From the early days of colonization, the Yakuts became subjects of the tsar on the same footing as Russian peasants but could assert their rights legally. This legal framework allowed them to preserve elements of their identity and adapt to shifting political landscapes. This resilience was especially evident during World War II, when the Republic of Sakha not only paid a heavy toll on the war effort, but also demonstrated its strategic importance by contributing generals to the Russian army. These particularities make comparisons with other conquest contexts around the world challenging, as the Russian conquest is without direct equivalent. Moreover, as USSR promoted Yakut students from working-class and rural backgrounds, the working class could access higher education. This is evident through our very collaboration with Yakut colleagues, who do not belong to the Yakut intelligentsia, nor to privileged families, but to what would be referred to as Indigenous communities in other contexts.

The research carried out in this study was developed as part of official bilateral collaborations between Russia and France, particularly between the Centre National de la Recherche Scientifique (CNRS, which stands for National Center for Scientific Research in English), and the universities of Toulouse (France) and Yakutsk (Republic of Sakha), and involved local partnership. The collaborative framework was supported by Yakut ethical agreements, with excavations led by Yakut specialists and authorized by the federal authorities of Russia. The following text describes the legal, ethical and collaborative framework underlying the present study.

1.3.2 Legal and ethical aspects

The CNRS (France) and the Université Paul Sabatier Toulouse III (now, Université de Toulouse, France) supported the development of the COSIE (COévolution homme/milieu en Sibérie Orientale, standing for Human–environment coevolution in Eastern Siberia) International Associated Laboratory from 2011 to 2019, and in 2020 of an International Research Laboratory (IRL). COSIE aimed at studying the coevolution between man and the environment in Eastern Siberia, mainly in Yakutia or Sakha Republic, during the last five centuries. It involved three French partners (CNRS, Université Paul Sabatier Toulouse III, as well as the Université de Strasbourg until 2019, and then Université de Paris from 2015), the Russian Foundation for Basic Research, the Northeastern Federal University in Yakutsk, and the Krasnoyarsk State Medical

University. The COSIE International Associated Laboratory was co-led by Prof. Crubézy (Université Paul Sabatier Toulouse III) and Prof. Anatoly Alexeev who served as the rector of the Northeastern Federal University (NEFU) in Yakutsk in 2011, and as the director of the interdisciplinary laboratory entitled “*Evolution of Nature and Humans in the North*”. The laboratory focused on the long-term interaction between humans and their environment, particularly over the past four centuries, when the arrival of Yakut and later European populations gave rise to epidemic diseases, some of which are re-emerging today. In addition to joint archaeological excavations, paleogenetics was proposed as a method to improve understanding of the history of human settlement, archaeological sites, and the evolution of infectious diseases. Knowledge transfer between French and Yakut partners was facilitated through workshops and the exchange of professors, researchers, and students. Official collaborative agreements between Université Paul Sabatier Toulouse III and NEFU were signed from 2011, and renewed twice for a five-year period, to facilitate exchange between both institutions.

In 2015, the collaborative framework was extended to laboratories specifically focused on Indigenous populations. This involved Dr. Anatoly Nikolaev, NEFU rector at the time, as well as Dr. Lilia Alekseeva (NEFU Museum of Archaeology and Ethnography), Dr. Sardana Fedorova (NEFU Molecular Biology Research Laboratory of the Institute of Natural Sciences), Prof. Olga Melnichuk (NEFU Institute of Modern Languages and Regional Studies), Dr. Maxim Cheprasov (NEFU Laboratory-Museum of the Mammoth at the Institute of Applied Ecology), and Dr. Natalya Popova (Institute for Humanities Research and Indigenous Studies of the North, under the Siberian Branch of the Russian Academy of Sciences). This collaboration later evolved to include the same partners in an international research project.

1.3.3 Fieldwork, ethics committee and custom clearance

Field archaeological excavations were led by Yakut colleagues both before (2004-2011), and after (2011-2019), the collaboration was established, in collaboration with the institutions mentioned above for the latter. In June 2012, ethical authorization for conducting ancient DNA analyses was granted by the Local Committee for Biomedical Ethics of the Federal State Budgetary Institution, which is referred to as the “Yakut Scientific Center for Comprehensive Medical Problems” of the Siberian Branch of the Russian Academy of Medical Sciences. The authorization aimed “*to collect information on the evolution of bacteria or viruses identified for the studied period, as well as to study the influence of socio-ecological factors on movement, adaptation, and dynamics of ancient populations*” (Excerpt from the Minutes, dated June 20, 2012, No. 31).

Subsequently, customs clearance certificates were issued by the rector of Yakutsk University to export scientific material, including archaeological remains. These materials were transported to France by the company Genomed, under customs resolution number RU/2019/05-41350/19.

1.3.4 Student training

Prof. Crubézy’s impact on the training of Yakut students was recognized in 2018 when he was appointed the same year as a foreign member of the Academy of Sciences of the Republic of Sakha and when NEFU awarded him a diploma in Yakutsk, for his 15-year commitment to research-based training. In addition to four to five students receiving training during each excavation campaign between 2011 and 2019, two Yakut students were granted scholarships from the French government to complete their doctoral theses, one of which was carried out under joint supervision between the Université Paul Sabatier Toulouse III (Prof. Crubézy) and NEFU (Prof. Nadezhda Pokatilova.). The first PhD thesis was defended on November 14th, 2016, by Dr. Dariya Nikolaeva and was entitled: “*Origine et évolution de la culture Sakha au sein de l’État russe (XVII^e siècle au XXI^e siècle)*” (“*Origin and Evolution of Sakha Culture within the Russian State (17th to 20th Century)*”).

The second PhD thesis was defended on March 23rd, 2021, by Dr Liubomira Romanova and was entitled: *“Évolution de l'alimentation et de l'économie chez les Yakoutes du XVII^e au début du XX^e siècle: Confrontations des données biologiques et culturelles”* (“*Evolution of Diet and Economy among the Yakuts from the 17th to the Early 20th Century: A Comparison of Biological and Cultural Data*”).

1.3.5 Public outreach and community engagement

Paramount to the COSIE programme was to informing the Yakut public about their research progress, including archaeological and genetic findings. Whether French researchers or Yakut researchers, all took part each year in radio broadcasts, or gave interviews in Yakut newspapers at the term of our joint excavations. Excavations provided multiple opportunities to engage with the inhabitants of the neighboring villages (*e.g.*, facilitating visits to excavation sites and addressing the questions raised by local residents), and local guides participated to archaeological prospects. Public engagement was carried out through a whole array of media, including:

(1) local TV interviews (*eg.* Dr Vincent Zvéni gorosky was interviewed on September 21st, 2016 in Bakhsy by “Tolon TV”; the programme was livestreamed, allowing for questions and answers from the audience, following translation by Georgii Romanov and Aisen Solovev; similar interview formats were followed on other channels, such as the local NVK Sakha channel on March 22nd, 2019 at the Yakut Science Center of Complex Medical Problems);

(2) WhatsApp campaigns (*e.g.*, in Khonuu, Momsky from October 31st to November 14th, 2016, or in Sasyr and the Magyr reindeer camp, Momsky, on March 2nd-19th 2019);

(3) lectures (*e.g.*, at the cultural center in Tolon on September 15th, 2017, when Dr. Vincent Zvéni gorosky and Georgii Romanov presented then ongoing research results);

(4) meetings with local associations (*e.g.*, on February 25th, 2019, with the president and members of the Public Organization Association of Indigenous small-numbered peoples of the North of the Sakha Republic (Yakutia), and the World Association of Reindeer Herders, with discussions involving Dr Mikhail Pogodaev, former president of the Even in Yakutia, Anastasia Shestakova, Dr Nicolas Bureau and Dr Vincent Zvéni gorosky);

(5) popular science books in French (*e.g.*, Crubézy, E., Nikolaeva, D. (2017). *Vainqueurs ou vaincus? L'énigme de la Iakoutie*, eds Odile Jacob [Defeaters or defeated? The Yakutia enigma]), Russian (Crubézy, E., Alexeev, A. *Mir Drevnikh Iakutov, Opyt Mezhdistsiplinarnykh Issledovaniï* [The World of the Ancient Yakuts: A Transdisciplinary Approach from the French/Yakut Expedition], University of the North East (Yakutsk, 2012)), and Yakut (*e.g.*, Popov V.V. (2018) *Sakha-Frantsuz arheologicheskaï non ètnograficheskaï èkspeditsiï iürdiönnërit, (2004-2017)* [Results of the Sakha-French Archaeological and Ethnographic Expedition (2004–2017), and; Popov V.V. (2023). *Süürbë èkspeditsiia (2004-2020)*. - [Twenty-Year Expedition (2004-2020)]);

(6) TV documentaries (*e.g.*, Jampolsky, M. *Les Mystères de Kyys La Chamane*. (Gédéon Programmes, France, 2007); this French TV documentary represents a long-standing testament to the unity and shared purpose of French and Yakut researchers, was broadcasted around the world, including on the Yakut national television, with Russian subtitles, on Christmas Eve, which represents some of the highest annual ratings);

(7) blogs (*e.g.*, on the history of Yakutia: <https://nikbara.ru/2021/08/23/o-proishozhdenii-i-passionarnosti-jakutskogo-naroda/>), websites (*e.g.*, interview of Prof. Crubézy on NEFU website: https://www.s-vfu.ru/news/detail.php?SECTION_ID=2268&ELEMENT_ID=78569), and YouTube

videos (e.g., in Yakut: <https://www.youtube.com/watch?v=vQ2AEUI7dzw>, and in Russian: www.youtube.com/watch?v=6cascHSrsVQ&t=223s);

(8) regular articles, reports and interviews by Vassily Popov, a researcher involved in field excavation, for mainstream newspapers. The latter were disseminated both in Yakut (e.g., through the *Kyyym* newspaper), and Russian (e.g., *Arheologicheskie raskopki v Verkhoyanskoy uluse (po materialam SFAĖĖ)* [Archaeological Excavations in the Verkhoyansk Ulus (Based on SFAĖĖ Materials)]; «Doch' piñanitsy-shamana ili arheologicheskii detektiv» (Arina Iākovleva) // gazeta «*Iakutsk vecherñiĭ*», 08.04.2016 g. / o pogrebenii Lepsei II, Verkhoyanskii raion [“The Daughter of a Drunken Shaman or an Archaeological Detective” (Arina Yakovleva) // Newspaper *Yakutsk Evening*, 08/04/2016 / About the Burial of Lepsey II, Verkhoyansk District; *iakutskom konskom sedle (po materialam SFAĖĖ)* [On the Yakut Horse Saddle (Based on SFAĖĖ Materials)]. 16 iūnia 2022 g.). The most recent of such articles was authored by Prof Rosalia Bravina (Institute for Humanities Research and Indigenous Studies of the North, Yakutsk, on May 31st, 2025, and was published in the leading local newspaper (“Sakha sire”). It can be accessed freely online at the following link: <https://sakha-sire.ru/arheologiya-uonna-genetika-noruottar-ikki-ardylarynaagy-byrajyakkka/>

(9) exhibitions. In an attempt to reach out to the broadest possible audience, photography exhibitions were organized in 2015 and 2020 at the Em. Yaroslavsky Museum in Yakutsk to document the work done during field excavations. These exhibitions toured the Churapchinsky and Khangalassky regions, following their first inauguration in the Yakutsk capital city. It is noteworthy that some of the archaeological material collected during field expeditions has now become part of the collections of the Archaeological Museum of Yakutsk University. Their exhibition from March 2023 featured funerary artifacts from several of the archaeological investigated in this study, including OyogosseTumula, Ieralakh, Lepsei2, Boulounniakh2, AC3, and Bougaryma. Finally, in 2019, the final year of the COSIE program, a significant event took place as part of the CNRS 80th anniversary. A major exhibition, entitled “*Toward the Origins of the Yakuts*” was held at the new museum in Yakutsk, *Historical Park “Russia – My History”* showcasing the findings gathered over 15 years of research in the region. Organized by L. Romanova, S. Duchesne, P. Gérard, and C. Petit, the exhibition showcased 100 exceptional artifacts uncovered over the past 17 years by Franco-Yakut missions and 30 informational panels featuring field photographs and reconstructions. The official inauguration took place on September 1st 2019, a gathering that transcended boundaries and brought together a diverse group of more than 250 people. The audience included officials from the university and the Academy of Sciences, as well as a representative of the President of the Republic of Yakutia and the former Minister of Culture. Presentations and lectures were delivered in the museum auditorium by Prof. E. Crubézy, Prof. A. Alexeev, and Dr L. Romanova. The event also included a display of reconstructed Yakut costumes, based on artifacts from excavations, presented on live mannequins accompanied by a group of traditional singers. It is estimated that the exhibition was seen by several thousand Yakuts.

References

1. Hartley, J. (2014). *Siberia: a history of the people* (Yale University Press).
2. Ferret, C. (2010). *Une civilisation du cheval. Les usages de l'équidé, de la steppe à la taïga* (Belin).
3. https://www.citypopulation.de/en/russia/admin/dalnevostochnyj_federaln/98__sacha/.
4. The Federal Service for State Statistics (Pocstat) <https://rosstat.gov.ru/>.
5. Okladnikov, A. (1955). *Yakutiya do prisoyedineniya k Russkomu gosudarstvu* [Yakutia before its incorporation into the Russian state]. In G. A. Vasilev (Ed.), *Istoriya Yakutskoy ASSR* [History of the Yakut ASSR] (Yakutsk: Yakutsk Book Publishing House).
6. Ivanov, V. (1974). *Istoricheskoe i etnograficheskoe issledovanie Iakutii XVII–XVIII vv.* [Historical and Ethnographic Study of Yakutia in the 17th–18th Centuries] (Nauka).
7. Crubézy, E., and Nikolaeva, D. (2017). *Vainqueurs ou vaincus? [Winners or losers?]* (Odile Jacob).
8. Anderson, D. (2000). *Identity and Ecology in Arctic Siberia: The Number One Reindeer Brigade* (Oxford University Press).
9. Anderson, D., Arzyutov, D., and Alymov, S. (2019). *Life Histories of Etnos Theory in Russia and Beyond* (Open Book Publishers).
10. Jochelson, W. (1926). *The Yukaghir and the Yukaghirized Tungus*. Preprint at E.J. Brill, G.E Stechert.
11. Kirillina, K., Tananaev, N., Savvinova, A., Lobanov, V., Fedorova, A., and Borisov, A. (2023). Climate change impacts the state of winter roads connecting indigenous communities: Case study of Sakha (Yakutia) Republic. *Clim Serv* 30, 100356. <https://doi.org/https://doi.org/10.1016/j.cliser.2023.100356>.
12. Balzer, M. (1994). Yakut. In *Encyclopedia of World Cultures*, Vol. VI, P. Friedrich and N. Diamond, eds. (GK Hall & Co), pp. 404–407.
13. Schöning, C. (1990). Classification problems of yakut. In *L'Asie centrale et ses voisins*, R. Dor, ed. (INALCO), pp. 91–102.
14. Pakendorf, B. (2007). Contact in the prehistory of the Sakha (Yakuts): Linguistic and genetic perspectives (Lot).
15. Lindenau, Â.I. (1983). *Opisanie narodov Sibiri -pervaya polovina 18 veka-* (Description of the peoples of Siberia ,The first half of the 18th century) - 1745. (Magadanskoe knižnoe izdatel'stvo).
16. Okladnikov, A. (1970). *Yakutia before it's incorporation into the russian state* (Mc Gill-Queen's University press.).
17. Avril, F. (1693). *Voyage en divers états d'Europe et d'Asie entrepris pour découvrir un nouveau chemin vers la Chine* [Journey through Various States of Europe and Asia Undertaken to Discover a New Route to China] (Boudot).
18. Ides, I., and Brand, A. (1967). *Zapiski o russkom posol'stve v Kitaj (1692-1695)* [Notes about Russian Embassy in China]. M. Glav. Red. Vost. Lit. 1967.
19. Witsen, N. (1692). *Noord en Oost Tartarye* [North and East Tartary] (Halma).

20. Strahlenberg, P. (1726). Vorbericht eines zum Druck verfertigten Werckes von der grossen Tatarey und dem Königreiche Siberien, mit einem Anhang von Grosz- Ruzsland. Worinnen von dem Autore die Einrichtung und vornehmsten Contenta desselben, vorgetragen werden [Preliminary Report] (Benjamin Gottlieb Schneider).
21. Nikolaeva, D. (2016). Origine et évolution de la culture Sakha au sein de l'Etat Russe -du XVIIème siècle-au XXème siècle- [Origin and Evolution of Sakha Culture within the Russian State - From the 17th to the 20th Century] (doctoral thesis, Université Paris-Saclay).
22. Antonov, E., Borisov, A., and Boyakova, S. (2003). Narod sakha ot veka k veku. Ocherki istorii [People of Sakha Through the Ages: Essays on History] (Nauka).
23. Serochevskiy, V.L. (1993). Âkuty. Opyt ètnografičeskogo issledovaniâ [The Yakuts: An Ethnographic Research Study], second edi (Assotsiatsiya Rossiyskaya Politicheskaya Entsiklopediya, Moscow).
24. Radlov, V. (1908). Die jakutsche Sprache in ihrem Verhältnisse zu den Türksprachen [The Yakut Language and Its Relation to the Turkic Languages] (Imperial Academy of Sciences of St. Petersburg).
25. Rudenko, S. (1953). Kul'tura naseleniia Gornogo Altaia v skifskoe vremia [The Culture of the Population of the Altai Mountains in the Scythian Era]. (Izdatel'stvo Akademii nauk SSSR).
26. Alexeev, A.N., Bravina, R.I., and Romanova, E.N. (2020). Istoriia Iakutii. Tom I [History of Yakutia, Vol.1] (Nauka).
27. Ksenofontov, G. (1992). Uraangkhai-sakhalar, Tom 2 [Uraanghai-Sakhalar, Volume 2] (National Publishing House of the Republic of Sakha (Yakutia)).
28. Ksenofontov, G. (1992). Uraangkhai-sakhalar, Tom 1 [Uraanghai-Sakhalar, Volume 1] (National Publishing House of the Republic of Sakha (Yakutia)).
29. Rudenko, S.I. (1962). Kul'tura hunnu Noin-Ulinskie kurgany [The tumuli of the Hunnu culture from Noin Ula] Izd-vo AN. (Publishers of the USSR Academy of Sciences).
30. Giscard, P.H., Tsagaan, T., and Crubézy, E. (2014). Le premier empire des steppes en Mongolie (Faton).
31. Khudiakov, I. (1987). Kochevniki Tsentral'noi Azii i ikh pogrebal'nye pamiatniki [Nomads of Central Asia and Their Funerary Monuments] (Nauka).
32. Tokarev, S.A. (1945). Obshchestvennyĭ stroĭ iakutov XVII-XVIII v. [The social structure of the Yakuts in the 17th to the 18th centuries] (Yakutskoe gosudarstvennoe izdanie), pp. 290–291.
33. Zhores, A. (1969). The Rise and Fall of T. D. Lysenko (Columbia University Press).
34. Okladnikov, A. (1970). Iakuty: Proiskhozhdenie naroda i ego kul'tury [The Yakuts: The Origin of the People and Their Culture] (Nauka).
35. Okladnikov, A. (1968). Istoriâ Sibiri s drevnejših vremen do naših dneĭ [History of Siberia from Ancient Times to the Present Day] (Nauka).
36. Alexeev, A. (1994). Proiskhozhdenie yakutskogo naroda (po novym arkheologicheskim materialam)[The Origin of the Yakut People (Based on New Archaeological Materials)] (Uchenye zapiski Yakutskogo universiteta, Yakutsk) pp. 131–154.

37. Alexeev, A. (1996). *Drevnyaya Yakutiya: Zheleznıy vek i epokha Srednevekov'ya* [Ancient Yakutia: From the Iron Age to the Middle Ages] (Izd. Akademii nauk Rossii, Sibirskoe otделение").
38. Gogolev, A. (1990). *Arkheologicheskie pamiatniki Yakutii pozdnego srednevekov'ya (XIV–XVIII vv.)*. [Archaeological Sites of Yakutia in the Late Middle Ages (14th–18th Centuries)] (Irkutsk University Press).
39. Gogolev, A. (1993). *Yakuty (problemy etnogeneza i formirovaniya kul'tury)* [Gogolev, A.I. The Yakuts (Problems of Ethnogenesis and the Formation of Culture).] (Yakutsk State University Press).
40. Gogolev, A. (2004). *Etnicheskaya istoriya narodov Yakutii (do nachala XX v.)* [The Ethnic History of the Peoples of Yakutia (Up to the Early 20th Century)] (Yakutsk State University Press).
41. Duchesne, S., Bravina, R., Popov, V., Kolodeznikov, S., Gérard, P., Myglan, V., Hochstrasser-Petit, Ch., Romanova, L., Petit, M., Kirianov, N., et al. (2020). Frozen graves of Yakutia, a chronological sequence. *Vestnik arkheologii, antropologii i etnografii* 4, 120–130. <https://doi.org/10.20874/2071-0437-2020-51-4-11>.
42. Okladnikov, A. (1959). *Shishkinskie pisanitsy – pamiatniki drevnei kul'tury Pribaikal'ia* [The Shishkino Petroglyphs – Monuments of Ancient Culture in the Baikal Region] (Irkutsk: Irkutsk book Publ).
43. Okladnikov, A., and Zaporozhskaia, V. (1972). *Petroglyfy srednei Leny*. [Petroglyphs of the Middle Lena] (Nauka).
44. Kotchmar, N. (1994). *Obraz shamana v naskal'nom iskusstve epokhi bronzy*. [The Image of the Shaman in Rock Art of the Bronze Age]. (Nauka).
45. Amory, S., Crubézy, E., Keyser, C., Alekseev, A.N., and Ludes, B. (2006). Early influence of the steppe tribes in the peopling of Siberia. *Hum Biol* 78, 531–549. <https://doi.org/10.1353/hub.2007.0001>.
46. Okladnikov, A. (1946). *Lenskie drevnosti. Otchet ob arkheologicheskikh issledovaniyakh na nizhnei Lene ot Zhiganska do Kumakh-Surta v 1942–1943* [Lena Antiquities. Report on archaeological investigations on the lower Lena from Zhigansk to Kumakh-Surta in 1942–1943].
47. Gadal, S., Zakharov, M., Kamičaitytė, J., and Danilov, Y. (2020). Alas Landscape Modeling by Remote Sensing Image Analysis and Geographic Ontology. Study case of Central Yakutia (Russia). In 6th International Conference on Geographic Information Systems Theory, Applications and Management, ATHENA Research & Innovation Information Technologies, ed. (Polytechnic Institute of Setúbal/IPS; Knowledge Systems Institute), pp. 112–118. <https://doi.org/10.5220/0009569101120118>.
48. Yildirim, E. (2020). Importance and Role of Kurykan People In Turkic History. *Journal of Turkology*, 287–302.
49. Crubézy, E., and Alexeev, A. (2007). *Chamane – Kyys, jeune fille des glaces* (Errance).
50. Gryaznov, M. (1969). *The ancient civilization of Southern Siberia* (Cowles Book Company).
51. Crubézy, E., and Alexeev, A. (2012). *Mir drevnikh iakutov: opyt mezhdistsiplinarnykh issledovaniy (po materialam sakha-frantsuzskoi arkheologicheskoi ekspeditsii)*. (University press Yakuts).

52. Crosby, A. (2009). *Ecological Imperialism: The Biological Expansion of Europe 900-1900* (second edition) (Cambridge University Press).
53. Acemoglu, D., Johnson, S., and Robinson, J.A. (2012). The Colonial Origins of Comparative Development: An Empirical Investigation: Reply. *American Economic Review* 102, 3077–3110. <https://doi.org/10.1257/aer.102.6.3077>.
54. Kabanovitch, R. (1952). Iz istorii bor'by s tuberkulezom [From the History of the Fight Against Tuberculosis] (Izdatel'stvo Akademii meditsinskikh nauk SSSR).
55. Jochelson, W. (1933). *The Yakut* (The American Museum of Natural History).
56. Fischer, R.H. (1943). *The russian fur trades, 1500-1700*. (University of California Press.).
57. Forsyth, J. (1992). *A history of the peoples of Siberia. Russia's north Asian colony 1581–1990*. (Cambridge University Press).
58. Platanova, N. (2007). Les commissions d'enquête, l'administration sibérienne et l'impôt sur les peuples autochtones en Russie au XVIIIe siècle [The commissions of inquiry, Siberian administration and tax on indigenous peoples in Russia in the 18th century]. *Histoire, économie & société*, 27–50 (Armand Colin).
59. Platanova, N. (2011). Le commerce des caravanes russes en Chine du XVIIe siècle à 1762 [The trade of Russian caravans in China from the 17th century to 1762]. *Histoire, économie & société*, 3–27 (Armand Colin).
60. Crubézy, E., Melnichuk, O., and Alexeev, A. (2020). Archaeology, Genetics And History 15 Years Of Research In Yakutia (2002 – 2017). *Vestnik arkheologii, antropologii i etnografii* 4, 110–119. <https://doi.org/10.1128/microbiolspec.PoH-0004-2014.Th>.
61. Torroni, A., Schurr, T.G., Cabell, M.F., Brown, M.D., Neel, J. V, Larsen, M., Smith, D.G., Vullo, C.M., and Wallace, D.C. (1993). Asian affinities and continental radiation of the four founding Native American mtDNAs. *Am J Hum Genet* 53, 563–590.
62. Karafet, T.M., Zegura, S.L., Posukh, O., Osipova, L., Bergen, A., Long, J., Goldman, D., Klitz, W., Harihara, S., de Knijff, P., et al. (1999). Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *Am J Hum Genet* 64, 817–831. <https://doi.org/10.1086/302282>.
63. Zerjal, T., Dashnyam, B., Pandya, A., Kayser, M., Roewer, L., Santos, F.R., Schiefenhövel, W., Fretwell, N., Jobling, M.A., Harihara, S., et al. (1997). Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am J Hum Genet* 60, 1174–1183.
64. Pakendorf, B., Morar, B., Tarskaia, L.A., Kayser, M., Soodyall, H., Rodewald, A., and Stoneking, M. (2002). Y-chromosomal evidence for a strong reduction in male population size of Yakuts. *Hum Genet* 110, 198–200. <https://doi.org/10.1007/s00439-001-0664-4>.
65. Pakendorf, B., Novgorodov, I.N., Osakovskij, V.L., Danilova, A.P., Protod'jakonov, A.P., and Stoneking, M. (2006). Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. *Hum Genet* 120, 334–353. <https://doi.org/10.1007/s00439-006-0213-2>.

66. Fedorova, S., Bermisheva, M., VILLEMS, R., Maksimova, N., and Khusnutdinova, E. (2003). Analysis of mitochondrial DNA lineages in Yakuts. *Mol Biol*, 544–553. <https://doi.org/10.1023/A:1025135326954>
67. Zlojutro, M., Tarskaia, L.A., Sorensen, M., Snodgrass, J.J., Leonard, W.R., and Crawford, M.H. (2009). Coalescent simulations of Yakut mtDNA variation suggest small founding population. *Am J Phys Anthropol* 139, 474–482. <https://doi.org/10.1002/ajpa.21003>.
68. Fedorova, S., and Barashkova, N. (2022). Issledovaniia populiatsionnoi genetiki naseleniia Respubliki Sakha (Iakutiia)[Population-Genetic Studies of the Population of the Sakha Republic (Yakutia)] (Nauka).
69. Fedorova, S.A., Reidla, M., Metspalu, E., Metspalu, M., Rootsi, S., Tambets, K., Trofimova, N., Zhadanov, S.I., Kashani, B.H., Olivieri, A., et al. (2013). Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): Implications for the peopling of Northeast Eurasia. *BMC Evol Biol* 13, 1. <https://doi.org/10.1186/1471-2148-13-127>.
70. Pugach, I., Matveev, R., Spitsyn, V., Makarov, S., Novgorodov, I., Osakovsky, V., Stoneking, M., and Pakendorf, B. (2016). The Complex Admixture History and Recent Southern Origins of Siberian Populations. *Mol Biol Evol* 33, 1777–1795. <https://doi.org/10.1093/molbev/msw055>.
71. Kılınç, G.M., Kashuba, N., Yaka, R., Sümer, A.P., Yüncü, E., Shergin, D., Ivanov, G.L., Kichigin, D., Pestereva, K., Volkov, D., et al. (2018). Investigating Holocene human population history in North Asia using ancient mitogenomes. *Sci Rep* 8, 8969. <https://doi.org/10.1038/s41598-018-27325-0>.
72. Kılınç, G.M., Kashuba, N., Koptekin, D., Bergfeldt, N., Dönertaş, H.M., Rodríguez-Varela, R., Shergin, D., Ivanov, G., Kichigin, D., Pestereva, K., et al. (2024). Human population dynamics and Yersinia pestis in ancient northeast Asia. *Sci Adv* 7, eabc4587. <https://doi.org/10.1126/sciadv.abc4587>.
73. Ricaut, F.X., Kolodetsnikov, S., Keyser-Tracqui, C., Alekseev, A.N., Crubézy, E., and Ludes, B. (2006). Molecular genetic analysis of 400-year-old human remains found in two Yakut burial sites. *Am J Phys Anthropol* 129, 55–63. <https://doi.org/10.1002/ajpa.20195>.
74. Crubézy, E., Amory, S., Keyser, C., Bouakaze, C., Bodner, M., Gibert, M., Röck, A., Parson, W., Alexeev, A., and Ludes, B. (2010). Human evolution in Siberia: From frozen bodies to ancient DNA. *BMC Evol Biol* 10, 25. <https://doi.org/10.1186/1471-2148-10-25>.
75. Zvéngorosky, V., Duchesne, S., Romanova, L., Gérard, P., Solovyev, A., Romanov, G., Barashkov, N., Fedorova, S., Ludes, B., Crubézy, E., et al. (2020). The genetic legacy of legendary and historical Siberian chieftains. *Commun Biol* 3, 1–9. <https://doi.org/10.1038/s42003-020-01307-3>.
76. Biagini, P., Thèves, C., Balaesque, P., Géraut, A., Cannet, C., Keyser, C., Nikolaeva, D., Gérard, P., Duchesne, S., Orlando, L., et al. (2012). Variola Virus in a 300-Year-Old Siberian Mummy. *New England Journal of Medicine* 376, 2057–2059. <https://doi.org/10.1056/NEJMc1212281>.
77. Thèves, C., Senescau, A., Vanin, S., Keyser, C., Ricaut, F.X., Alekseev, A.N., Dabernat, H., Ludes, B., Fabre, R., and Crubézy, E. (2011). Molecular identification of bacteria by total sequence screening: Determining the cause of death in ancient human subjects. *PLoS One* 6, e21733. <https://doi.org/10.1371/journal.pone.0021733>.

78. Dabernat, H., Thèves, C., Bouakaze, C., Nikolaeva, D., Keyser, C., Mokrousov, I., Géraut, A., Duchesne, S., Gérard, P., Alexeev, A.N., et al. (2014). Tuberculosis epidemiology and selection in an autochthonous Siberian population from the 16th-19th century. *PLoS One* 9. <https://doi.org/10.1371/journal.pone.0089877>.
79. Keyser, C., Hollard, C., Gonzalez, A., Fausser, J.L., Rivals, E., Alexeev, A.N., Ribéron, A., Crubézy, E., and Ludes, B. (2015). The ancient Yakuts: A population genetic enigma. *Philosophical Transactions of the Royal Society B: Biological Sciences* 370. <https://doi.org/10.1098/rstb.2013.0385>.
80. Zvéngorosky, V., Sabbagh, A., Gonzalez, A., Fausser, J.-L., Palstra, F., Romanov, G., Solovyev, A., Barashkov, N., Fedorova, S., Crubézy, É., et al. (2020). The limitations of kinship determinations using STR data in ill-defined populations. *Int J Legal Med* 134, 1981–1990. <https://doi.org/10.1007/s00414-020-02298-w>.
81. Gauthier, Y., and Garcia, A. (1996). L'exploration de la Sibérie [The exploration of Siberia]. (Actes Sud).
82. Crubézy, É., Gérard, P., Kirianov, N., Popov, V., Nikolaëva, D., Petit, C., Petit, M., Romanova, L., Keyser, Ch., Ludes, B., et al. (2017). The Relationship between Archaeology, Genetics, Ethnology and History. Reflections on a 2002-2015 Study in Yakutia (1632-1922). In *Clashes of Time. The Contemporary Past as a Challenge for Archaeology*, J.-M. Blaising, J. Driessen, J.-P. Legendre, and L. Olivier, eds. (Presses universitaires de Louvain), pp. 121–139.
83. Stepanoff, C. (2019). *Voyager dans l'invisible (La Découverte)*.
84. Macé, M., Richeval, C., Alcouffe, A., Romanova, L., Patrice, G., Duchesne, S., Cannet, C., Boyarskikh, I., Annie, G., Zvéngorosky, V., et al. (2021). At the Origins of Tobacco-Smoking and Tea Consumption in a Virgin Population (Yakutia, 1650-1900 A.D.): Comparison of Pharmacological, Histological, Economic and Cultural Data. *Biology (Basel)* 10, 1–26. <https://doi.org/10.3390/biology10121271>.
85. Crubézy, E., Duchesne, S., Razafindrazaka, H., Romanova, L., Gérard, P., Alcouffe, A., Esclassan, R., Melnichuk, O., Ushnitsky, I., Ludes, B., et al. (2021). Sucrose Is Not the Whole Story: Risk Factors and Oral Health at the Contact (Yakutia, Siberia-16th/19th). Preprint, <https://doi.org/10.3390/biology10100974> <https://doi.org/10.3390/biology10100974>.
86. Duchesne, S. (2020). *Pratiques funéraires, biologie humaine et diffusion culturelle en Iakoutie (16e-19e siècles)*. [Funerary practices, human biology and cultural diffusion in Yakutia (16th-19th centuries)] (Doctoral thesis, Université Paul Sabatier Toulouse III)
87. Romanova, L., and Crubézy, E. (2016). Sviaz' toponimiki i arkheologii: na primere iakutskikh toponimov [The Connection Between Toponymy and Archaeology: Using Yakut Toponyms as an Example]. *Filologicheskie nauki: voprosy teorii i praktiki*, Tambov 12, 146–149.
88. Crubézy, E., and Alexeev, A. (2012). *Mir drevnikh iakutov, opyt mezhdistsiplinarnykh issledovaniy*. [The World of Ancient Yakuts: An Interdisciplinary Research Approach (Based on Materials from the Sakha-French Archaeological Expedition)] (Yakutsk State University Press).
89. Petrova, S.I. (2010). Clothing in yakut traditional wedding rituals (19th–20th centuries). *Archaeology, Ethnology and Anthropology of Eurasia* 38, 106–110. <https://doi.org/https://doi.org/10.1016/j.aeae.2010.08.014>.

90. Petrova, S. (2002). Ritual'naia odezhda naroda Sakha [The Ritual Clothing of the Sakha People]. <https://doi.org/10.20874/2071-0437-2020-51-4-12>
91. Slepstov, E. (2008). Sviaz' atributiki shamanskogo oblachenii Yakutov s simbolikoi arkhainnykh obriadov. [The Connection Between the Attributes of Yakut Shamanic Attire and the Symbolism of Archaic Rituals.]. -Izvestiia Rossiiskogo gosudarstvennogo pedagogicheskogo universiteta im. A.I. Gertsena [Proceedings of the Russian State Pedagogical University Named After A.I. Herzen] pp. 94–98.
92. Petrova, S. (2014). Obychai izbeganiya u yakutov: traditsii i étnokul'turnye paralleli [The Custom of Avoidance Among the Yakuts: Traditions and Ethnocultural Parallels]. Vestnik Volgogradskogo gosudarstvennogo universiteta. [Bulletin of Volgograd State University] pp. 66–70.
93. Efimova, E. (2007). Evoliutsiia semantiki shamanskogo kostiuma Yakutov XVIII-XIX vekov [The Evolution of the Semantics of Yakut Shamanic Costume in the 18th–19th Centuries]. zvestiia Rossiiskogo gosudarstvennogo pedagogicheskogo universiteta im. A.I. Gertsena [Proceedings of the Russian State Pedagogical University Named After A.I. Herzen] pp. 69–74.
94. Duchesne, S., Romanova, L., Gérard, P., Nikolaeva, D., Alcouffe, A., Hochstrasser-Petit, C., Alekseeva, L., Petit, M., Dabernat, H., Géraut, A., et al. (2025). Beyond the Permafrost: Excavated Tombs of Eastern Siberia Archaeological data from the French-Sakha missions. Preprint at Zenodo, <https://doi.org/10.5281/zenodo.15017495>.
95. Romanova, L., Stépanoff, C., Telmon, N., and Crubézy, E. (2020). Health access inequities and magic medicine: the first ancient evidence? *The Lancet* 395, 1343–1344. [https://doi.org/10.1016/S0140-6736\(19\)31885-9](https://doi.org/10.1016/S0140-6736(19)31885-9).

Section 2. Supplementary methods, additional statistical analyses, including Figures S2_1 to S2_28

2.1.	<i>Dating, and archaeological sex estimations of the human remains</i>	27
2.2.	<i>DNA Extraction, DNA library building and indexing, sequencing</i>	27
2.3.	<i>Read-processing</i>	28
2.4.	<i>Post-mortem DNA damage</i>	29
2.5.	<i>Error rates, uniparental markers and contamination estimates</i>	32
2.6.	<i>Ploidy</i>	32
2.7.	<i>Imputation</i>	33
2.8.	<i>Kinship reconstruction</i>	34
2.9.	<i>Inbreeding and diversity estimates</i>	34
2.10.	<i>PCA</i>	35
2.11.	<i>ADMIXTURE</i>	38
2.12.	<i>Admixture modeling and dating</i>	40
2.13.	<i>ASCEND</i>	41
2.14.	<i>Oral Taxonomic Diversity</i>	44
2.15.	<i>Strain analyses</i>	49
2.16.	<i>Functional oral diversity</i>	54
2.17.	<i>Pathogen screening</i>	56
2.18.	<i>Variola strains phylogeny</i>	58
2.19.	<i>Ancient DNA methylation</i>	60
	Correlation with modern bone data	60
	Intragenomic quality test of DNA methylation	62
2.20.	<i>Statistical associations between cultural and non-cultural data</i>	66
	<i>References</i>	68

2.1. Dating, and archaeological sex estimations of the human remains

The French-Russian MAFSO excavations were carried out by a team of archaeologists, anthropobiologists, forensic scientists, and cultural artefact restorers. When frozen bodies were found, an autopsy was performed, as was the case for KyysOunuoga¹. When only skeletons were present, on-site anthropological analyses were performed, following dedicated protocols². The MAFSO corpus was also supplemented by the excavations led by a Russian team headed by Dr. Rosalia Bravina to increase temporal and geographic coverage. Chronological and cultural information derive from a variety of approaches, including dendrochronological dating applied to coffins and tomb superstructures, radiocarbon dating, and identification of imported Chinese and European goods, which became very frequent after 1689 CE³. A preliminary assessment of biological sex was carried out in the field based on morphological^{4,5} or morphometric markers (DSP)^{6,7}.

Archaeological remains were assigned to one of the four archaeological stages according to the methodology described by Duschene and colleagues³. The Iron Age Mokp individual was originally radiocarbon dated between 2,400 and 2,200 BP (see information about the “*Siberian grave of Pokrovsk*” in Amory et al.⁸). An independent radiocarbon date was obtained at the Keck AMS laboratory, University of California Irvine (USA), following the standard protocols of Librado and colleagues⁹, which confirmed the original report (UCIAMS-210906, 2025±20 uncal. BP, *i.e.*, 364-197 cal. BCE IntCal20¹⁰).

2.2. DNA extraction, DNA library building and indexing, sequencing

Samples were processed in the clean laboratory facilities at the Centre for Anthropobiology and Genomics of Toulouse (CAGT), University of Toulouse, France, or at the Centre of Excellence for GeoGenetics (CGG), University of Copenhagen, Denmark. Ancient DNA procedures aimed at limiting contamination by modern or amplified DNA were strictly followed. These included: separate buildings for the pre- and post-PCR amplification steps; disposable Personal Protective Equipment; and decontamination of all surfaces and tools using bleach or DNA AWAY between samples and between sessions. Each extraction, Uracil-Specific Excision Reagent (USER) treatment, library building and Polymerase Chain Reaction (PCR) amplification sessions included a blank reaction in which no sample (for the extraction step), or QIAGEN EB buffer (for the subsequent steps), was added to the reaction instead of a sample.

Bone and tooth samples: After gentle surface abrasion, a portion of the dense part of the bone samples was collected using a diamond wheel (Proxxon or Argofile instruments). For tooth samples, the cementum was isolated as recommended by Damgaard and colleagues¹¹. Whenever possible, samples were split in two halves, one half being stored to allow for further re-extraction. Samples were crushed into smaller fragments using a manual mortar or cutting pliers, or pulverized using a Retsch MM200 instrument, and placed in 5 mL Eppendorf LoBind tubes. DNA was extracted following a silica-column based method as described in Gamba et al.¹² (method Y), adapted from the original method of Yang et al.¹³. Samples were first pre-digested for 1h at 37°C in 4 mL extraction buffer (0.45M EDTA, 0.5% N-lauryl Sarcosyl, 0.25mg/mL proteinase K), with agitation. After centrifugation for 2 min at maximum speed, the supernatant was removed and the remaining undigested pellets were further incubated overnight at 42°C in 4 mL of fresh extraction buffer. In case when the samples were not fully dissolved, fresh proteinase K was added before incubating for an extra 24 h digestion. After centrifugation for 2 min at maximum speed, the supernatant was transferred on a 30 kDa Amicon Ultra-4 filter (Millipore), and centrifuged at 3,000 rpm until the volume

was reduced to 250 μ L. The supernatant was then purified on a MinElute column (QIAGEN) following manufacturer's instructions, and eluted in 60 μ L pre-heated elution buffer (QIAGEN EB, 0.05% Tween).

Calculus samples: Calculus samples were isolated as described in Sabin et al.¹⁴ and placed in 2 mL Eppendorf LoBind tubes. Samples labelled as "Name_C" in Extended Data Fig. 4A (e.g., E1etchei3_C_C_P4) were extracted for DNA following a protocol similar to the one used for bones and cementum, except that no 1h pre-digestion was performed and the digestion volume was limited to 1 mL. Samples labelled as "Name_CE" (e.g., E1etchei3_CE_C_P4) were subjected to an overnight digestion at 50°C in 555 μ L of a buffer consisting of 0.45 M EDTA, 1.8 mg/mL proteinase K and 9 mM DTT. The supernatant was further purified on a QIAGEN MinElute column and eluted in 40 μ L sterile water.

Soft tissue samples: Fragments of soft tissues (lung and muscle) were collected using sterile scalpels, cut into smaller fragments and placed in 2 mL Eppendorf LoBind tubes. Samples were digested in 1.11 mL of a buffer containing 0.45 M EDTA, 1.8 mg/mL proteinase K and 9 mM DTT, following an overnight incubation at 50°C with agitation. After 12 min centrifugation at 8,000 rpm, the supernatant was collected and purified on a silica column (MinElute, QIAGEN) following the manufacturer's instructions (40 μ L sterile water elution).

To remove Uracil residues and limit the impact of post-mortem damage on downstream analysis, the majority of DNA extracts were subjected to Uracil-Specific Excision Reagent (USER, New England Biolabs) treatment, consisting on an incubation of 22.8 μ L DNA extract with 7 μ L USER enzymatic mix, for 3h at 37°C.

Sequencing libraries were constructed from double-stranded DNA molecules, by ligation of universal (method S, as Single-index, similar to the method described in Gamba et al.¹⁵, adapted from Kircher and Meyer¹⁶) or indexed (method T, as Triple-index, as described in Fages et al.¹⁷) blunt-end adapters. For method T, a 7-nucleotide-long index (the sequence of which was obtained from Rohland et al.¹⁸) was present within both P5 and P7 adapters prior to ligation to DNA inserts. Such internal indices constitute the first seven bases of each Illumina Paired-End (PE) read.

DNA library pools were sequenced at CAGT on the Illumina MiniSeq instrument, at CGG on Illumina NextSeq, HiSeq2000, HiSeq2500 and HiSeq4000 instruments, or at CNRGH, Evry-Courcouronnes (France) on the Illumina HiSeqX instrument. The vast majority of the sequence data consisted of paired-end reads.

2.3. Read-processing

The demultiplexed fastq read pairs were processed using PALEOMIX¹⁹ bam_pipeline v1.2. Sequencing adaptors were trimmed (-mm 5) as well as poor-quality end, and paired-end reads were collapsed using AdapterRemoval²⁰ v2.3.1. All the resulting reads, and those remaining paired, were mapped against the hs37d5 reference genome, using Bowtie²¹, with local sensitive mapping parameters. The BAM alignment file was further filtered for alignment size and mapping quality superior or equal to 25 bp and 30 Phred-score, respectively. PCR duplicates were removed using Picard MarkDuplicates (<http://picard.sourceforge.net>), and realignment around indels was performed using GATK²². Sequencing statistics, as numbers of sequencing reads, endogenous DNA content and coverage are provided in Supplementary Tables 1a and 1b.

All the resulting alignments were merged into a single BAM alignment before being further pseudo-haploidized, random-sampling one read at positions characterized by one or more alignments. Pseudo-haploid genotypes were called using ANGSD²³ v.0.930 (htslib 1.9), skipping positions and/or reads showing

base and/or mapping Phred-quality scores strictly lower than 30 (`--doHaploCall 1 --doCounts 1 --minMapQ 30 --minQ 30 --remove_bads 1 --uniqueOnly 1`), and restricting calls for those 1,233,013 SNP positions forming the 1240k panel^{24,25}.

Soft tissues were analyzed for only one individual, AC2, and consisted of lung and muscle. They showed limited endogenous content, and provided extremely limited sequence data aligning against the human genome, resulting in 0.0000587-0.0003046-fold coverage, depending on the DNA library considered. The relatively limited DNA preservation observed in soft tissues match previous work on woolly mammoths²⁶. As a result, hard tissues, mostly petrosal bones and teeth, which are known to maximize DNA preservation^{11,12}, were preferentially analyzed.

2.4. Post-mortem DNA damage

The sequencing data underlying individual libraries constructed on DNA extracts from bones, teeth and soft tissues, were analyzed for the presence of post-mortem DNA damage (Supplementary Table 1b). Base compositional profiles within (1) the first and last ten nucleotide positions of read alignments against the human reference genome, and (2) the five reference positions preceding and following read alignments, were estimated using `mapDamage2 v2.2.1`²⁷, with the following parameters: `-l 50 -b 5 -a 5 -Q 20` (Figure S2_1). Considering the sequencing data generated from DNA extracts that were not treated by the USER enzymatic cocktail, the genomic position preceding (following) read alignments was generally enriched in purine (pyrimidine) residues. This aligns with depurination representing the major process driving DNA fragmentation post-mortem²⁸. In contrast, the sequencing data generated from DNA extracts that were treated by the USER enzymatic cocktail showed an enrichment of Cytosine (Guanine) residues at the genome reference position preceding (following) alignment starts. This is consistent with the cleavage of unmethylated Cytosine residues that were deaminated into Uracil residues post-mortem, which represents the specific target of the Uracil DNA Glycosylase, which is part of the USER enzymatic cocktail²⁹.

Nucleotide mis-incorporation profiles were estimated using `PMDtools v0.60`³⁰, with the following parameters (`-m 25 -q 20 -d --CpG --range 10 --platypus`). The sequencing data generated from DNA extracts that were not treated by the USER enzymatic cocktail showed an increase of C→T mis-incorporation rates towards alignment starts, mirrored by an increase of G→A mis-incorporation rates towards alignment ends. This aligns with Cytosine deamination representing one of the most important chemical drivers of post-mortem DNA decay²⁸. This pattern was considerably reduced when considering the sequencing data generated from DNA extracts that were treated by the USER enzymatic cocktail, in line with this treatment largely reducing the sequencing errors resulting from post-mortem DNA damage^{18,29} (Figure S2_2). Conditioning the analyses on CpG dinucleotide sites revealed the presence of an increase of C→T mis-incorporation rates towards alignment starts (and of G→A mis-incorporation rates towards alignment ends), including when considering the sequencing data generated following USER enzymatic treatment (Figure S2_2). This is in line with the deamination of methylated Cytosine residues, which are predominantly found in CpG contexts, forming Thymine residues (instead Uracil residues when affecting unmethylated Cytosine residues), which do not represent USER targets.

Combined, base compositional and nucleotide mis-incorporation profiles show the expected signatures of post-mortem DNA damage³¹, and, thus, indicate that the sequence data generated are authentic and ancient.

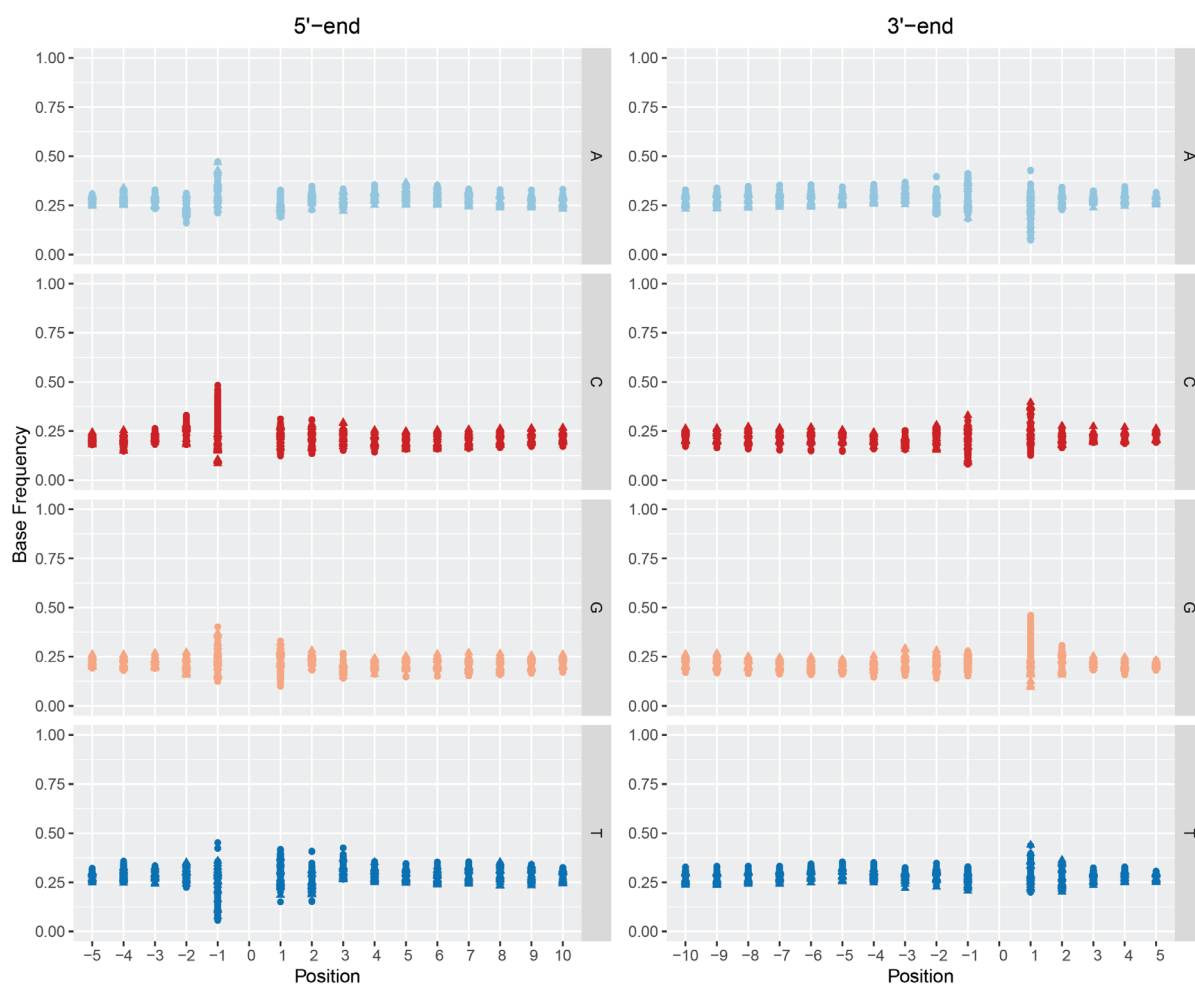
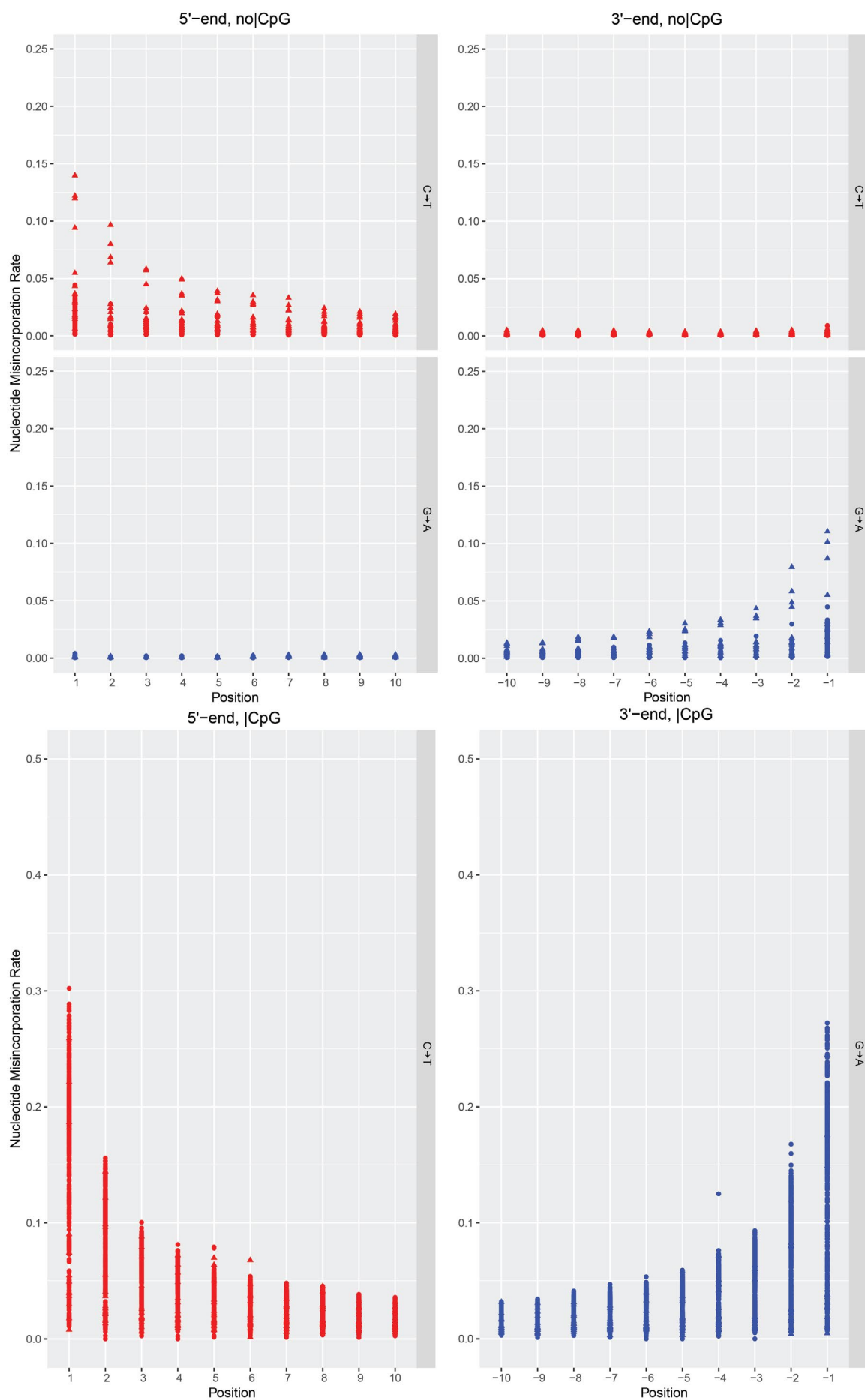


Figure S2_1. Base composition profiles.

Left: The frequency of each of the four nucleotides is calculated for the first ten (1 to 10) positions of read-to-reference alignments, as well as the five (-5 to -1) reference positions preceding alignment starts. *Right:* Same as Left, excepting that the last ten (-10 to -1) positions of read-to-reference alignments are shown, together with the five (1 to 5) reference positions following alignment starts. Circle: USER treatment. Triangle: no USER treatment.

Figure S2_2. (Next page) Nucleotide mis-incorporation profiles.

Left: C→T (top, red) and G→A (middle, blue) mis-incorporation rates within the first 10 nucleotide positions of read alignments. The C→T mis-incorporation rates shown at the bottom are estimated within CpG dinucleotide contexts. *Right:* C→T (top, red) and G→A (middle, blue) mis-incorporation rates within the last 10 positions of read alignments. The G→A mis-incorporation rates shown at the bottom are estimated within CpG dinucleotide contexts. Circle: USER treatment. Triangle: no USER treatment.



2.5. Error rates, uniparental markers and contamination estimates

Error rates were calculated using ANGSD²³ v.0.930 (htslib 1.9), and the methodology used in previous work³². Three-way alignments were used including an outgroup genome (the ancestral human genome), a high-quality genome (the genome from a modern Sardinian individual, sequenced to 39.4-fold average depth-of-coverage by Meyer et al.³³; Accession Nb. SS6004474), and each ancient genome individually. Relative error rates in the ancient genome, both globally and for each individual substitution class, were estimated by the relative difference between the branch length leading to the ancient genome and the high-quality modern genome.

A total of 46 women and 61 men were identified based on Ry ratio (Supplementary Table 1a). Mitochondrial haplotypes were called using haplogrep³⁴ (version 2.266), after aligning reads against the rCRS reference mitogenome (GenBank Accession Nb. NC_120920.1), and discarding those shorter than 25 bp. Read alignments were converted into a VCF file using bcftools³⁵ (version 1.8) mpileup, excluding reads with mapping and base qualities below 30, as well as the call (–ploidy 1) and filter modules, keeping sites covered by at least 5 independent reads. The resulting VCF file was then processed through haplogrep³⁴, calculating the best 100 hits. Contamination rates based on mitochondrial data were estimated using Schmutzi³⁶ and the same base quality threshold as above. Using heterozygosity measurements at polymorphic sites present on the X chromosome for males, contamination rates were also estimated for male samples following the methodology from Rasmussen et al.³⁷, implemented in ANGSD²³. Transition substitutions, and sites covered once or more than 200-times were discarded.

Mitochondrial contamination estimates were assessed within a 0%–5% confidence range for all but three individuals (Oulakh2, Koudouk3, and AtDaban4; Supplementary Table 1a). Oulakh2 had extremely limited amounts of data generated (mtDNA average depth-of-coverage = 1.37-fold). Koudouk3 had a ~7% mitochondrial contamination rate. For AtDaban4, the mitochondrial contamination estimates were within the range of 13% to 15%, while the X chromosome contamination remained very low (0.35%). The low coverage sample (Oulakh2) was disregarded in all other population genetic analyses. The other samples (Koudouk3 and AtDaban4) were included and projected together with the majority of the other samples in PCA analyses. These samples showed genetic profiles that were similar to those from the other individuals (*e.g.*, ADMIXTURE, and PCA), supporting minimal nuclear contamination. They were, however, excluded from those analyses conditioned on archaeological stages, as a conservative measure. Overall, nuclear contamination estimates were found to be limited (median=0.24%), and inferior to 0.4% (Supplementary Table 1a).

Y chromosome haplotypes were called using the Yleaf statistical package³⁸. Almost all the individuals from historical Yakutia analyzed belonged to haplogroup N1a1a1a1 (Supplementary Table 1a), except individuals Byljazyk3, OyogosseTumula1 and Khoumakhtaakh who carried haplogroups C2a1a1b1b, N1a2b1a, and R1b1a1b1, respectively. The Iron Age individual Mokp carried the N1a2b1 haplogroup and the Buryat individual carried the C2a1a2a haplogroup.

2.6. Ploidy

To check that all of our individuals have a regular ploidy, we followed the methodology described by Sehnert et al.³⁹. We first obtained so-called “chromosome ratios” for chromosomes 21, 18, and 13. The chromosome ratios were calculated as the number of hits on the chromosome studied divided by the number of hits on a different chromosome or the sum of chromosomes³⁹. For chromosome 21, the chromosome ratio was calculated as $\text{chrom21}/\text{chrom9}$, vs $\text{chrom18}/\text{chrom8}$ for chromosome 18, or

chrom13/(chrom2+chrom3+chrom4+chrom5+chrom6) for chromosome 13. We then obtained a Normalized Chromosome Value (NCV) equal to the following value: $NCV = \frac{x-\mu}{\sigma}$, where x is the chromosome ratio for the individual and μ and σ are the mean and sd of the chromosome ratio for the entire dataset. No aneuploidy was detected (Figure S2_3), as all NCV values were below 4³⁹.

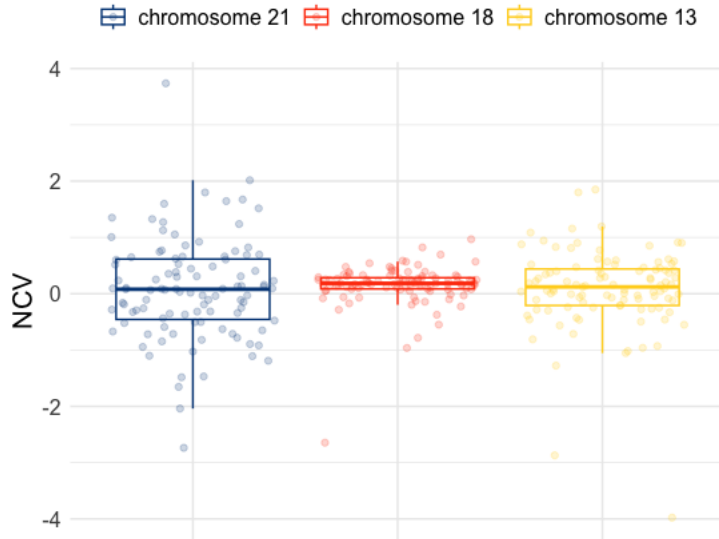


Figure S2_3. NCV values to test for aneuploidy for chromosomes 21, 18 and 13 for Yakuts (N=103).

Boxplots are defined as follows: minimum, maximum and center values represent the 25%, 75% and 50% quantiles, respectively; upper and lower whiskers are the maximum/minimum values of data within the 1.5 interquartile range over the 75th or under the 25th percentiles, respectively.

2.7. Imputation

We imputed a subset of genomes using GLIMPSE²⁴⁰ and the 1000 Genome Project⁴¹ panel as a reference dataset, following the instructions provided by the developers on the software website. To test for the minimal coverage needed to obtain accurate imputation, we selected four individuals, one from each of the four different archaeological stages, sequenced at high coverage (>15-fold): KenEbe8, AtDaban0, Kureleekh1 and TouoraUrek3. We down-sampled the data to 14 different average depth-of-coverage, ranging from 2-fold to 0.1-fold (*i.e.*, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.25, 1.5, 1.75, and 2-fold), and then imputed the resulting genotypes for the 1000 Genome Project panel⁴¹. Imputation accuracy was assessed by measuring the squared Pearson correlation between original and imputed genotypes, together with non-reference discordance rate, within minor allele frequency (MAF) bins and population-based variant groups. We found that a minimal coverage of 0.35-fold was necessary to keep the non-reference discordance rate below 6% and a squared Pearson correlation above 95% for MAF $\geq$ 5%. A total of 90 Yakut individuals (coverage>0.35-fold) were then imputed and filtered for MAF $\geq$ 5% and Genotype Probability $\geq$ 0.99 for all downstream analyses. The imputed individuals were combined with the phased 1000 Genome Project dataset⁴¹ for all downstream analyses, except those based on ChromoPainterv²⁴², and fineSTRUCTURE⁴², which required a liftover to the hg38 positions to include the matrix of phased genotypes released by Bergstöm et al.⁴³, as it comprised 20 modern Yakut individuals.

2.8. Kinship reconstruction

Relatedness between historical Yakuts (*i.e.*, all the individuals newly characterized in this study, except Mokp and the Buryat individual) was assessed on the non-imputed data using a combination of three complementary methodologies, READ2^{44,45}, lcMLkin⁴⁶ and TKGWV2⁴⁷ (Supplementary Tables 1r-t). For READ2^{44,45} and TKGWV2⁴⁷, the autosomal positions overlapping the 1240K dataset (see below) were used, restricting the former to $MAF \geq 1\%$. Prior to running lcMLkin⁴⁶, BAM files were transformed in VCF with the SNPbam2vcf.py script from VCFtools⁴⁸ (version 0.1.17), and filtered for sites located at last 100 kb apart and present at minimal 5% allelic frequencies (`-thin 100000 -maf 0.05` with plink v1.9⁴⁹). In these analyses, all the genomes, even the ones characterized at high-coverage were pseudo-diploidized, random-sampling one read per position using ANGSD (`-doHaploCall 1 -doCounts 1`). We carried out READ2⁴⁵ analyses on 105 individuals (excluding Lampa2 as showing minimal nuclear coverage, *i.e.*, 0.001X), and then filtered out the first-degree relationships estimated, if relying on less than 1,000 SNPs. Second-degree relationships supported by fewer than 2,000 SNPs were discarded, and the default filter of READ2⁴⁵ was used for third-degree relationships. READ2^{44,45} and TKGWV2⁴⁷ kinship analyses identified 14 first-degree relationships between pairs of individuals (Supplementary Tables 1rs). Three further pairs of first-degree male relatives (AtDaban4 *vs.* AtDaban11, Oktiom1Sujet1 *vs.* Oktiom1Sujet4, and; AtyyrMeite1 *vs.* Kureleekh1) also returned consistent results in all three analyses, corresponding to parent-offspring relationship according to READ2 and lcMLkin⁴⁶. Due to their different mitochondrial haplotypes, AtDaban11, Oktiom1Sujet1 and Kureleekh1 could be identified as the fathers of AtDaban4 (a young man, buried in Archaeological Stage 2, while AtDaban11 belonged to Archaeological Stage 1), Oktiom1Sujet4 (2-3 years old), and AtyyrMeite1, respectively. For the female Myran and the male Kouranakh1, READ2, and lcMLkin indicated a parent-offspring relationship and they shared the same mitochondrial haplogroup. Furthermore, all three methods identified 13 pairs of individuals related at the second degree, while READ2 and TKGWV2 agreed on six additional second-degree relationships, and TKGWV2 suggested five other possible pairs. Notably, two close families from two multiple graves (Oktiom and AC, where “AC” translates into “Shamanic Tree” in French), were reconstructed using uniparental markers, anthropological estimates of age-at-death, and estimated degrees of genetic relatedness (Extended Data Fig. 5a). The large cluster of related individuals in the Verkhoyansk region could be subdivided into two sub-clusters, connected by the single Tysarastaakh2b individual. None of the first- and second-degree relationships identified were between individuals from different regions.

Finally, Identify-By-Descent (IBD) content were calculated using ancIBD⁵⁰. As this software requires non-filtered VCF-formatted data, and performs an internal filtering procedure, we used the non-filtered VCF-formatted output of GLIMPSE2⁴⁰ as input for ancIBD. As recommended, the Yakut dataset was conditioned on the 1,240,000 SNPs from the 1240k dataset, for which ancIBD is optimized, and IBD sharing was screened for every pair with default settings⁵⁰. Analyses were run with default parameters, as pre-optimized for ancient DNA data. ancIBD only retains IBD block with a minimal SNP density of 220 SNPs per cM. We then categorized IBD blocks according to their size (*ie* ≥ 8 cM, ≥ 12 cM, ≥ 16 cM, and ≥ 20 cM) to calculate the total number of such blocks, and the genome fraction encompassing these categories (in Mb).

2.9. Inbreeding and diversity estimates

Runs-of-Homozygosity (ROH) were identified on the imputed dataset, using a set of 1000G biallelic transversions (TVs) with $MAF \geq 5\%$, removing any positions not fully covered (`--geno 0`). ROH were then calculated using Plink v1.9⁴⁹ with the following parameters: `--homozyg --homozyg-density 50 --homozyg-`

gap 100 --homozyg-kb 500 --homozyg-snp 50 --homozyg-window-het 1 --homozyg-window-snp 50 --homozyg-window-threshold 0.05. Short ROH segments (<1.6 Mb) inform on population size and ancient population history, while long ROH segments (≥ 1.6 Mb) reveal recent patterns of consanguinity. Genomic ROH fractions were calculated for two size bins: ≥ 1.6 Mb and <1.6 Mb. The genome fraction of short ROH was calculated according to the following formula: Total Length of ROH <1.6 Mb / (Total genome length considered - Total Length of ROH ≥ 1.6 Mb)⁵¹. Inbreeding scores were calculated with Plink v1.9⁴⁹ (-het), using transversions only and MAF $\geq 5\%$ (Supplementary Table 1a).

ROH segments for individuals with at least 400,000 SNPs on the pseudo-haploid 1240k panel, were assessed using HapROH⁵² with default parameters, using the 5,008 haplotypes from the 1000 Genome Project⁴¹ as a reference panel. ROH distribution was used as an indirect measure of genetic diversity and effective population size. This method detects ROH lengths larger than four centimorgans (cM), and up to 300 cM. The sum of the ROHs detected gives an overview of the parentage of each individual, as ROH length and parentage are positively correlated, and provides information on population size (*i.e.*, the greater the number of ROH segments, the smaller the population size). This analysis returned evidence of small amounts of inbreeding in almost all our individuals, with the exception of a single highly inbred individual (UsSerguei; Figure S2_4). It also returned estimates of the effective population size for each stage and region (Supplementary Table 1j).

Interestingly, HapROH differentiates really long ROH segments (20-300 cM), which helps discriminate the confounding effects of recent inbreeding versus small population sizes. This analysis showed that the high level of inbreeding detected in Mokp (Iron Age) is probably linked to a very small population size ($2N_e \sim 400$). The ancient individual from Buryatia (Tungus) showed no ROH segments, highlighting a different effective population size in Buryatia at the time than in 17th century CE Yakutia, despite similar material culture.

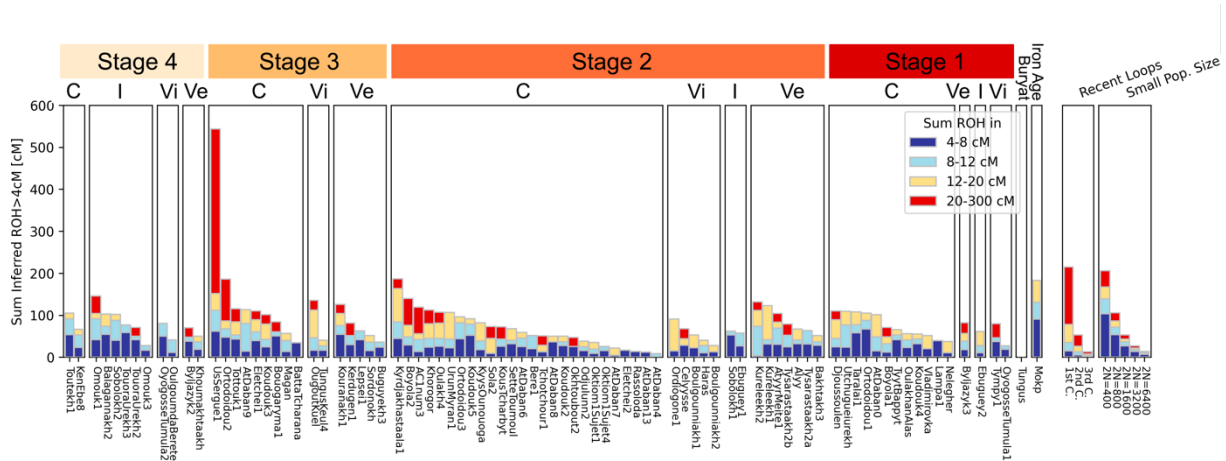


Figure S2_4. Visual inspection of runs of homozygosity (ROH) segments computed by hapROH for Yakuts individuals.

C= Central Yakutia, I=Indigirka, Vi=Vilyuy, Ve=Verkhoyansk. In the 'Recent Loops' box, 'C.' means cousins.

2.10. PCA

Principal Component Analysis (PCA) was carried out using the Human Origin reference panel for 597,573 autosomal genotypes. Genotypes were downloaded from the Allen Ancient DNA Resource v5 (AADR) website^{24,25}. We also included the genotypes of those Central Asian individuals with relevant genetic ancestry profiles reported by Zhang et al.⁵³. PCA was based on pseudo-haploid genotype calls for all the

individuals presented in this study and carried out using smartPCA from EIGENSOFT⁵⁴ version 7.2.170, projecting 913 ancient Eurasian and American individuals^{53,55–96} and 106 ancient Yakut individuals (coverage>0.02-fold) onto the principal components (PCs) obtained from in 2,761 Eurasian modern individuals^{33,41,57,60,73,75,77,80,87,91,92,97–104} (lsqproject: YES and shrinkmode: YES). Projections on the first two PCs are provided in Fig. 2a, while PC2 and PC3 are provided in Extended Data Fig. 3b.

To evaluate the impact of projection, in our fine-scale PCA comparison of modern and ancient Yakuts, we also performed a PCA using smartPCA from EIGENSOFT⁵⁴ version 7.2.170, but without any projection (Figure S2_5). We used the exact same set of individuals as those shown in Fig. 2a. No significant difference was observed between the two PCAs, with modern and historical Yakuts still plotting together without projection.

A third PCA was carried out to validate our imputation pipeline by confirming similar projections for imputed genotype data and pseudo-haploid data (Figure S2_6).

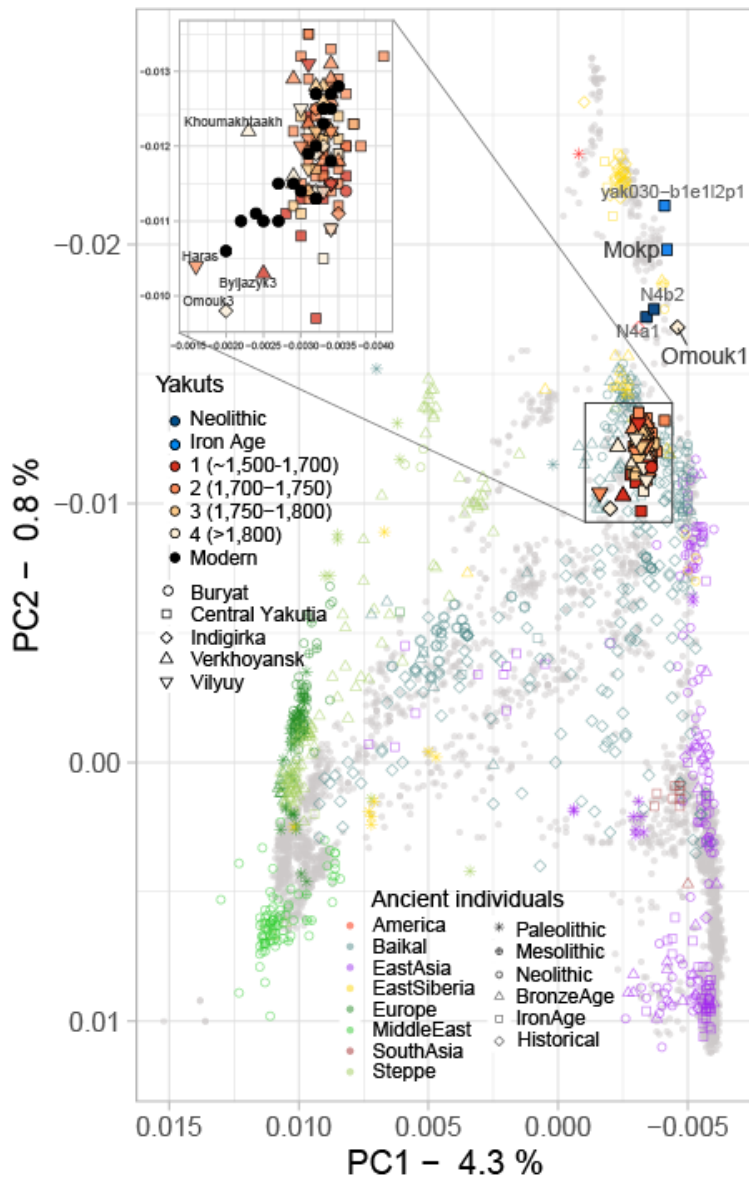


Figure S2_5. PCA performed on 597,573 SNPs for 2,761 modern Eurasian and American individuals, 892 ancient Eurasian and American individuals and 106 ancient Yakut individuals without projection.

All the modern individuals have been removed from the zoom panel, excepting the modern Yakuts (black dots). The percentage indicated on each PC reflects the explained fraction of genetic variation.

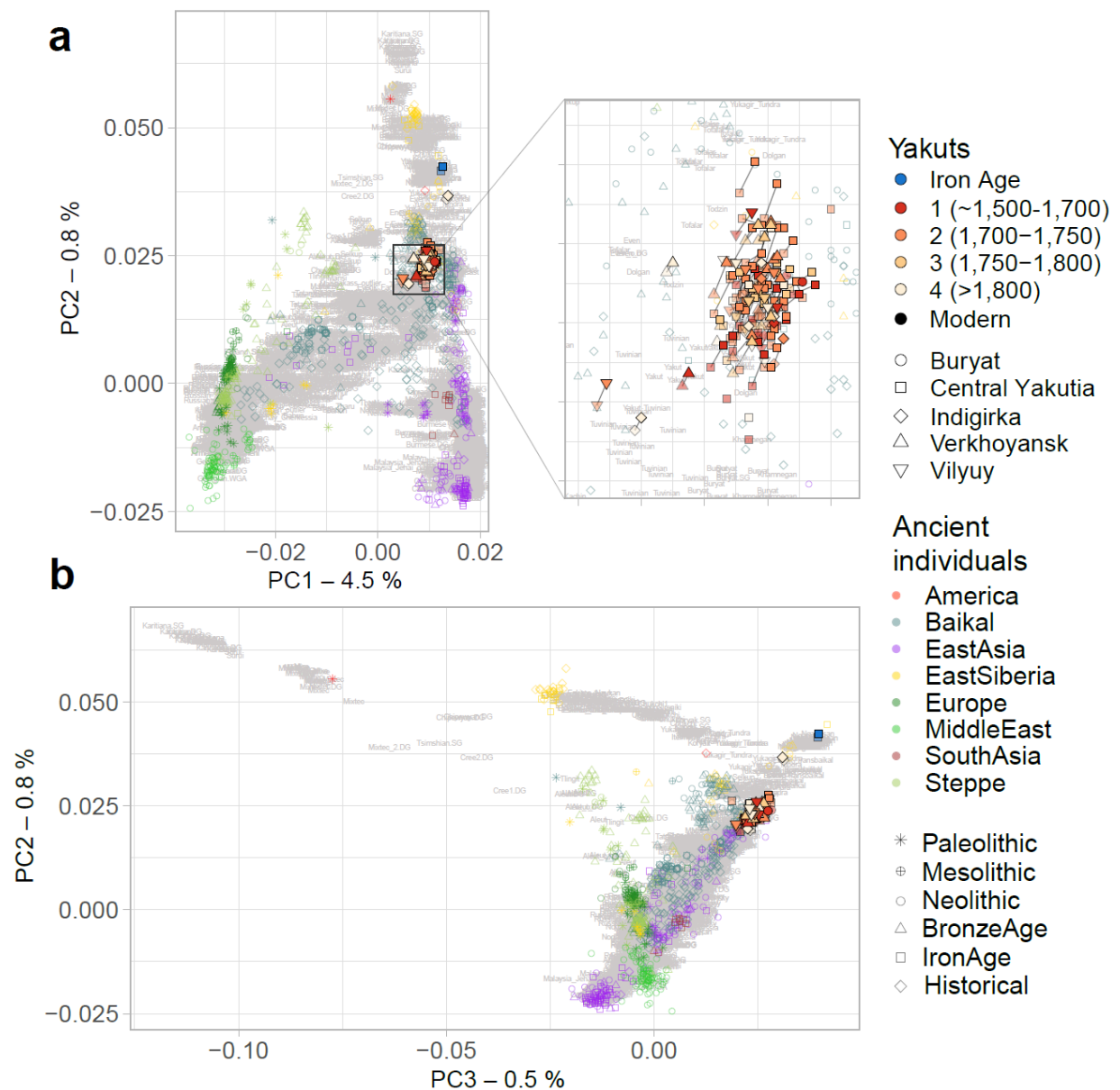


Figure S2_6. PCAs for the imputed dataset.

PCA calculated on 597,573 SNPs for 2,761 modern Eurasian and American individuals, with 892 ancient Eurasian and American individuals, 106 pseudo-haploid ancient Yakuts, and 90 imputed ancient Yakuts projected. Pseudo-haploid Yakut genomes are shown with transparency. The percentage indicated on each PC reflects the explained fraction of genetic variation. a) PC1 and PC2, with a zoom on the Yakut cluster. b) PC3 and PC2.

2.11. ADMIXTURE

Unsupervised ADMIXTURE¹⁰⁵ (version 1.3.0) analyses were carried out to estimate the proportions of genetic ancestries present in Yakuts (coverage $\geq$ 0.03-fold, pseudo-haploid), using autosomal positions as part of the 1240K HO panel, and a total of 3,639 Eurasian and American individuals. Sites were thinned for linkage disequilibrium with plink⁴⁹ (--indep-pairwise 200 25 0.4), resulting in a total of 327,582 SNPs. Confidence intervals were estimated from 100 bootstrap pseudo-replicates. Analyses were repeated 10 times using 10 random seeds in order to assess cross-validate convergence (Figure S2_7). Full ancestry profiles are provided in Figure S2_8 for the whole dataset.

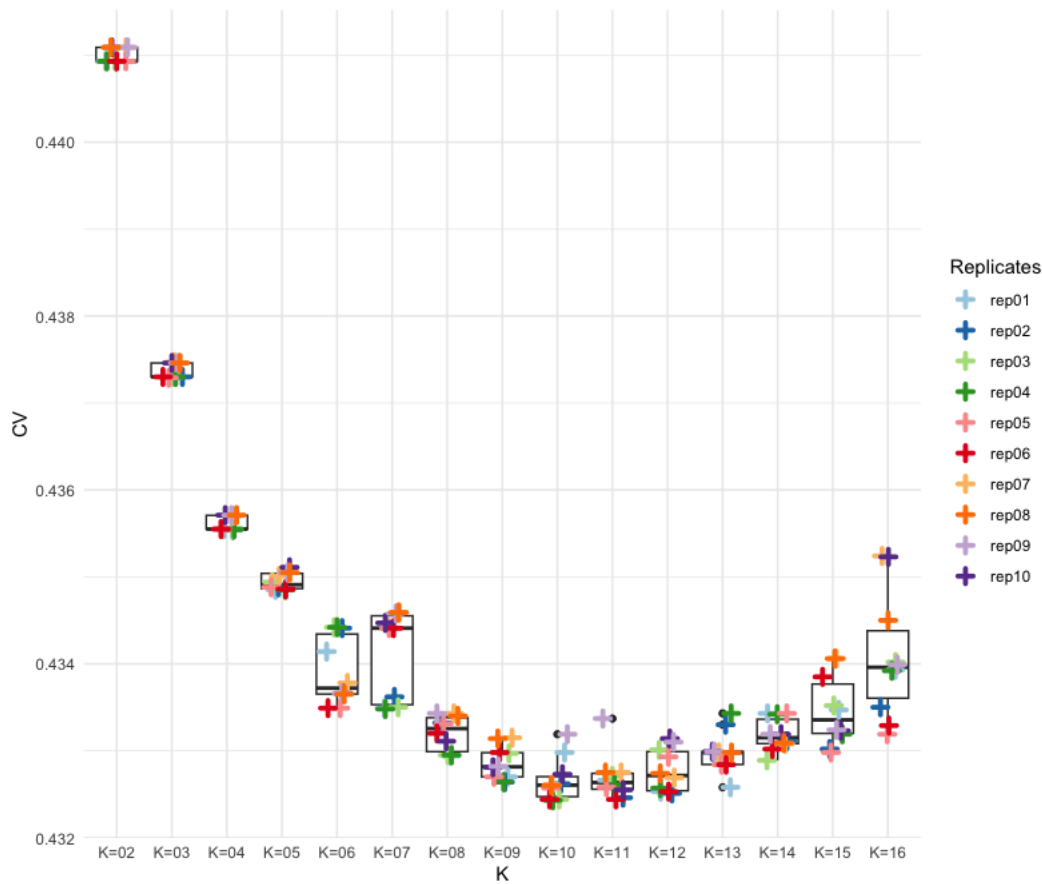
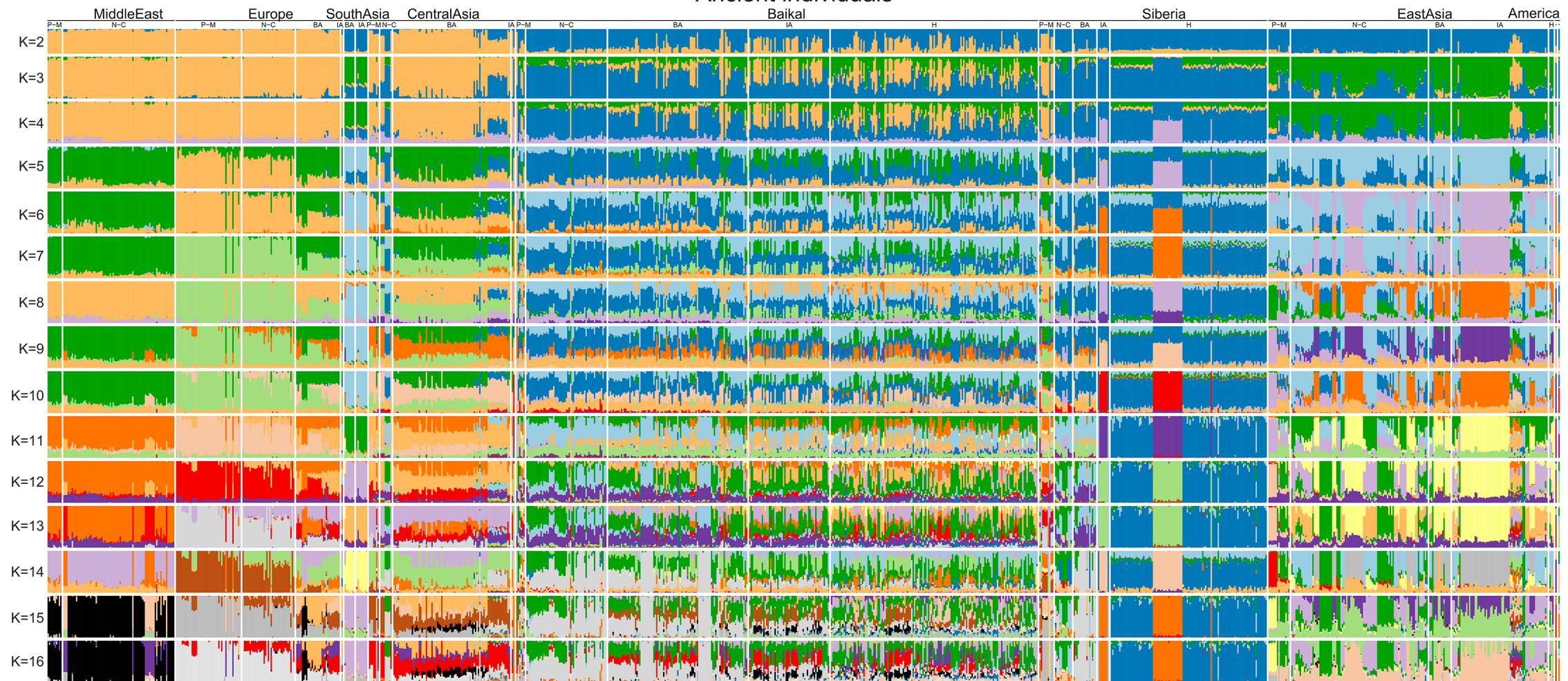


Figure S2_7. Cross-validation (CV) values for the ADMIXTURE¹⁰⁵ analyses.

K refers to the number of genetic ancestries considered in each analysis. Boxplots are defined as follows: minimum, maximum and center values represent the 25%, 75% and 50% quantiles, respectively; upper and lower whiskers are the maximum/minimum values of data within the 1.5 interquartile range over the 75th or under the 25th percentiles, respectively.

Figure S2_8. (Next page): Full ADMIXTURE¹⁰⁵ profile

Ancient individuals



Present-day individuals



2.12. Admixture modeling and dating

Admixture models for ancient Yakut individuals (coverage>0.1-fold) were assessed using the pseudo-haploid 1240K dataset and the HO dataset and qpAdm from Admixtools¹⁰⁰ (version 5.056). In all models tested, we first used Mbuti (N= 4), Ami (N=15), Mixe (N=6) modern individuals as outgroup (“right”) populations, together with Malta Ma-1 (N=1), Natufian (N=4), Turkey Neo (N=16), Villabruna (N=1) and Kostenki14 (N=1).

We initially sought to identify the most suitable source population to represent the Baikal_his source. To achieve this, we applied qpAdm on the 1240k dataset and the “rotating” method outlined by Harney et al. (2021)¹⁰⁶, testing various two-way admixture models between Yakutia_IA (N=2, comprising Mokp and yak03041), and one of the following populations: Russia_AngaraRiver_Medieval.SG⁷² (N=1), Mongolia_Sukhbaatar_Xiongnu (N=1) and Mongolia_Khuvsgul_MLBALateMedieval⁶⁷ individuals (N=2) and modern Buryat (Buryat.SG^{57,102}, N=4) (Supplementary Table 1g). Our results indicated that all models incorporating Yakutia_IA and another population from this set produced comparable outcomes. These models successfully fit the data and yielded plausible admixture proportions for most individuals, aside from a small subset of outliers. However, we observed a higher rate of infeasible or rejected models when using modern Buryat as a source (Supplementary Table 1g). To maximize statistical power, we then used all the individuals from these groups except the historical Buryat in the Baikal_his group for the final qpAdm models. Next, we investigated whether incorporating an additional western source would further improve model fit. We tested three-way admixture models, on the HO dataset (to include more modern populations), while continuing to rotate through the Baikal-related sources. The western sources considered included Russian, Polish, Bulgarian, and Czech (representing Slavic-speaking populations); Adygei, Abkhazian, Chechen, Lezgin, and North Ossetian (representing the North Caucasus); Mansi (as a representative of Uralic-speaking populations); and Altaian (as a proxy for Turko-Mongolic-speaking populations). The majority of these three-way models failed to fit the data (Supplementary Table 1g). In cases where they did, the two-way models without western sources already provided a strong fit, suggesting that western ancestry played only a minor role (Supplementary Table 1g). No other western sources outperformed the Russian group so we kept it as the western source for the final models.

The final qpAdm models were aimed at testing whether the Yakut genomic makeup was compatible with a two-way admixture from a local Siberian background (Yakutia_IA, N=2, comprising Mokp and yak03041), and another source, potentially representing two different ancestries from the Baikal region (Baikal_his (N=4)/Baikal_sib (N=11)), or Russia (Russian.SDG⁴³ N=24) (Supplementary Table 1f). Baikal sources were defined as Baikal_sib (N=11) and Baikal_his (N=4), with the former consisting of Mongolia_Khuvsgul_LateMedieval⁶⁷ (N=2), Mongolia_Dornod_LateMedieval⁶⁷ (N=7), and Mongolia_Khentii_LateMedieval⁶⁷ individuals (N=2), and the latter comprising Russia_AngaraRiver_Medieval.SG⁷² (N=1), Mongolia_Sukhbaatar_Xiongnu⁶⁷ (N=1) and Mongolia_Khuvsgul_MLBALateMedieval⁶⁷ individuals (N=2). The Baikal_sib populations were selected as they exhibited the closest ADMIXTURE ancestry profiles (Supplementary Information Section 2.11), and the Baikal_his were selected as explained above. Each ancient and modern individual from Yakutia was tested on the 1240k dataset for every combination of 2 or 3 populations putting the non-used population in the group of “right” populations¹⁰⁶ (Supplementary Table 1f).

Models were accepted if admixture proportion estimates ± 2 SE were within the [0, 1] interval, and when p -values ≥ 0.01 . Estimates outside the [0,1] interval are frequently expected under one-dimensional Isolation-By-Distance models, such as those presented by Harney et al.¹⁰⁶. However, recent work from Flegontova et al. (2025)¹⁰⁷ has revealed that estimates within the [0, 1] interval can be retrieved under more

realistic scenarios of Isolation-By-Distance involving two-dimensional landscapes, even for non-optimal sources. To reduce the risk of false positive model matches, we repeated qpAdm modeling using alternative sets of ‘right’ populations, each representing regionally and chronologically focused groups of interest. This approach was the only one practically feasible given that the sample size of the different sources could not be extended from data previously published, and that all detectable ancient graves within radii of just a few dozen kilometres have been systematically excavated by MAFSO in regions such as Verkhoyansk. The first set (‘set 1’) expanded our original reference set (‘set 0’) by including additional groups representing East Eurasian and West Siberian ancestries, including: WSHG (N=2), EEHG (N=5), Iran_GanjDareh_N (N=8), Russia_Ekven_OldBeringSea (N=16), Russia_BA_Okunevo.SG (N=14), and China_Upper_YR_LN (N=6). The second set (‘set 2’) was derived from ‘set 1’, but excluded Mbuti, Mixe, Villabruna, Ami, and Natufian, in order to reduce potential bias deriving from the presence of deep or divergent ancestries. Both sets yielded results consistent with those obtained using our initial set (‘set 0’). In particular, the genomic makeup of the majority of historical Yakut individuals could be successfully modelled as a two-way admixture between Yakutia_IA and Baikal_his sources, generally with greater admixture proportions from the local Indigenous population (Yakutia_IA).¹⁰⁶¹⁰⁷

We further applied DATES¹⁰⁸, using both the pseudo-haploid and imputed dataset, to two-way models to estimate the time of the admixture event between the local ancestry source (Yakutia_IA (N=2), extended with Nganasans (N=35), who show similar PCA placement and ADMIXTURE profiles, to obtain sufficient numbers of individuals), and Baikal populations (Supplementary Table 1h; Figure S2_9). As the confidence intervals using the Baikal_sib source were more restrained (Supplementary Table 1h), and Baikal_sib covered more individuals, the analyses including the Baikal_sib source were preferred (Fig. 3d). The time of admixture between a Russian source⁴³ and either an historical Yakut ancestry source (Yakut_his, comprising all newly sequenced Yakut individuals from the four stages, excluding related individuals and genetic outliers, N=92), or the local ancestry source (Nganasan/Yakut_IA, N=37), was also estimated for the imputed genomes of the PCA genetic outliers.

2.13. ASCEND

We used ASCEND¹⁰⁹ to assess the intensity and estimate the time for the bottleneck underlying the foundation of the Yakut gene pool. These analyses were repeated without specifying an outgroup, and choosing an outgroup (N=15) randomly from the populations present in our dataset. They were carried out for each archaeological stage or the whole group of ancient Yakuts, both for the pseudo-haploid and the imputed datasets, with the following parameters: binsize: 0.001, mindis: 0.001, maxdis: 0.3, maxpropsharingmissing: 1, minmaf: 0, usefft: YES and qbins: 100 (Supplementary Table 1i). The allele sharing correlation decay curve along with the fitted exponential model of our tests with outgroup are shown in Figure S2_10.

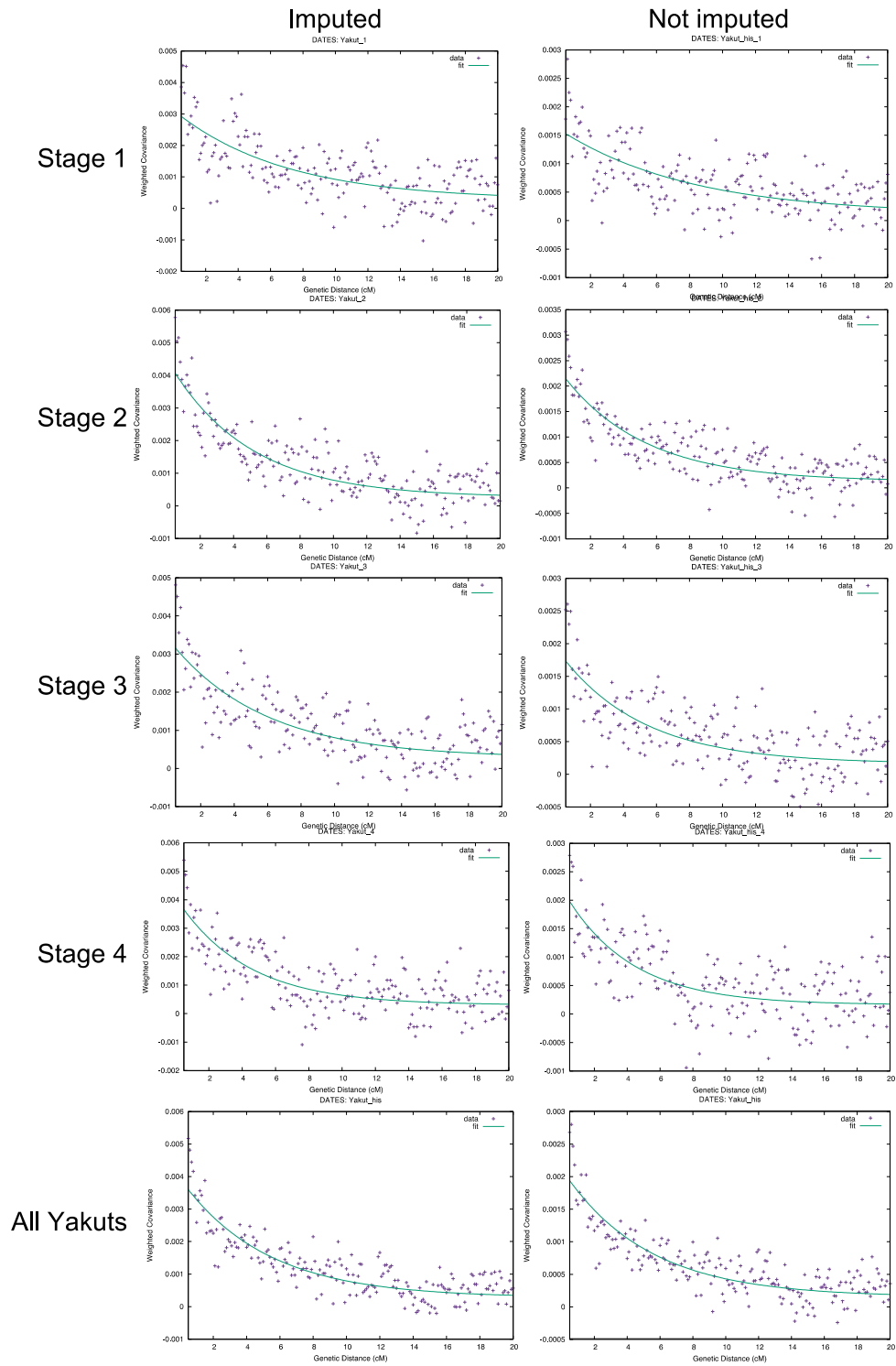


Figure S2_9. DATES¹⁰⁸ Weighted LD-decay for the imputed and pseudo-haploid dataset.

The admixture model is between Nganasan/Yakut IA and Baikul_sib to form Yakuts grouped by Stages, or considering all stages together.

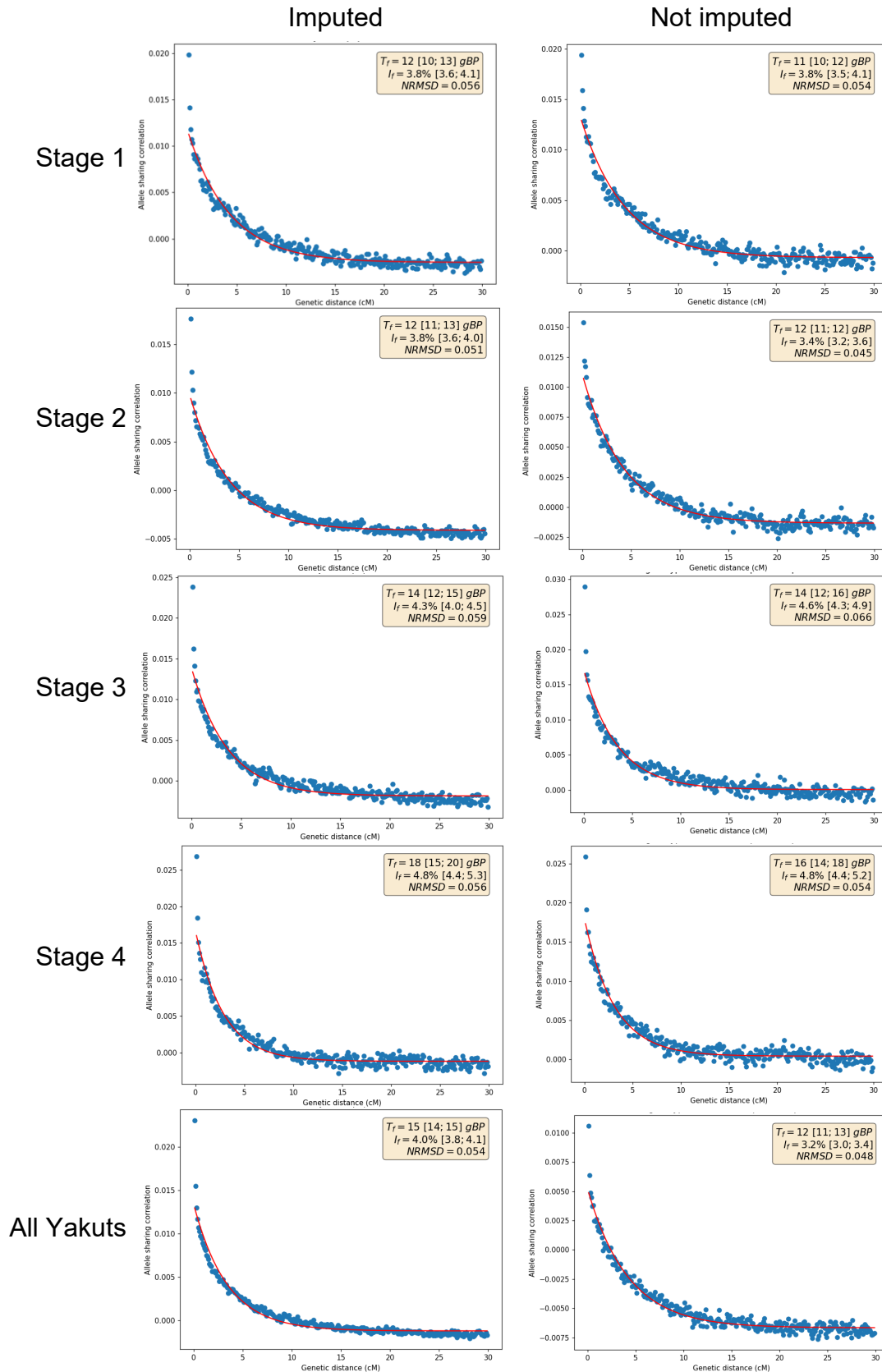


Figure S2_10. ASCEND¹⁰⁹ allele sharing correlation decay curve (blue points) along with the fitted exponential model.

In the top-right corner: the estimates of founder age (T_f) and intensity (I_f), with their associated 95% confidence intervals within brackets as well as the Normalized Root Mean Squared Deviation.

2.14. Oral Taxonomic Diversity

Microbial taxonomic profiles of 55 tooth extracts and 78 dental calculi were determined for each individual DNA library, restricting analyses to the fraction of collapsed reads (*i.e.*, “collapsed truncated” and “uncollapsed pairs” were disregarded). In order to avoid cross-mapping issues, we identified all the collapsed reads with no potential alignment against the human genome (hg37) and the human mitochondrial genome. To be conservative, we re-aligned all collapsed reads against those genomes using bwa¹¹⁰ (backtrack, version 0.6) with permissive parameters (no seed, -n: 0.01). Alignments with MQ>0 were considered to potentially derive from human genetic material and were, thus, disregarded in further steps. The resulting filtered bam files were reformatted as fastq files to be used for mapping.

Microbial read counts were obtained using MetaPhlAn4¹¹¹, discarding unclassified reads and reads too short. To ensure robust taxonomic profiling of microbial communities, we implemented a dynamic read length filtering strategy while running MetaPhlAn4. This approach accounted for variability in read length distributions across samples, enabling the selection of an appropriate minimum read length threshold for each sample, to avoid using low quality reads and, thus, false positives. For each sample, we first determined the most frequent read length (peak) and visually checked the distribution to account for anomalous read length distributions, where the primary peak did not correspond to the actual dominant read length. We then used as a minimal read length filter equal to the peak value minus ten, however, never including those reads shorter than 30 bp and longer than 70 bp. This resulted in the exclusion of two calculi and one tooth profiles for which no reads could be classified. This procedure was repeated on a panel of known sources that were used to assess the proportion of oral microbes contributing to each ancient DNA library. To achieve this, we ran Sourcetracker2¹¹², following data conversion into BIOM format, and conditioning analyses on species level (Extended Data Fig. 4a). Our comparative panel of known sources consisted of 10 profiles from soil samples, 10 profiles from non-oral bones serving as environmental contamination, 20 from dental plaque (10 for supplaque, and 10 for subplaque), 20 from guts (10 from rural, 10 from urban), 10 from skin, 18 modern calculus, 10 ancient calculi (see Supplementary Table 1k for details and references). Samples showing $\geq 25\%$ of oral sources were retained for further analyses, as such proportions were observed in oral samples previously analyzed and identified as authentic¹¹³. A total of 64 over 78 ancient dental calculi samples and 22 over 55 tooth samples passed this filter, while no DNA libraries constructed on lung and muscle extracts did, as expected. In cases where both a tooth and a calculus sample from the same individual passed filters, the profile maximizing oral microbial sources was kept, resulting in final dataset 74 individual oral microbiomes. The full MetaPhlAn4 profiles can be found in Supplementary Table 1l.

The most abundant bacterial taxa in the historical Yakut samples was *Actinomyces* dentalis. Other oral bacteria from the Actinobacteria or Proteobacteria phylum such as *Olsenella* sp oral taxon 80, *Desulfobulbus oralis* or *Lautropia mirabilis* were also commonly found (Figure S2_11).

Library-based read counts were then merged per individual, and bacterial taxa showing abundances lower than 1% were disregarded before carrying out Principal Coordinate Analysis (PCoA), based on Bray-Curtis distances (Fig. 3a). Species abundance for those microbes belonging to different bacterial complexes (red, orange, yellow, green and purple), as well as eight known oral pathogens, were measured and tested for potential shifts between archaeological stages (Kruskal-Wallis test; Fig. 3d and Figure S2_12). These analyses were repeated on a dataset restricted to calculus samples only. Here, we removed the individuals for whom only tooth samples were analyzed, and kept only the calculus data for those individuals associated with both tooth and calculus data. Alpha and beta diversity indexes, as well as abundance levels of bacterial complexes and pathogenic species showed the same temporal trends as analyses included both tissue types (Figure S2_13). Statistical assessment of differences in composition between region, period, sex and cultural

group also returned similar results, except for the significance between region, which was only marginally significant (PERMANOVA, p -value = 0.0532), probably due to reduced power (Supplementary Table 1m).

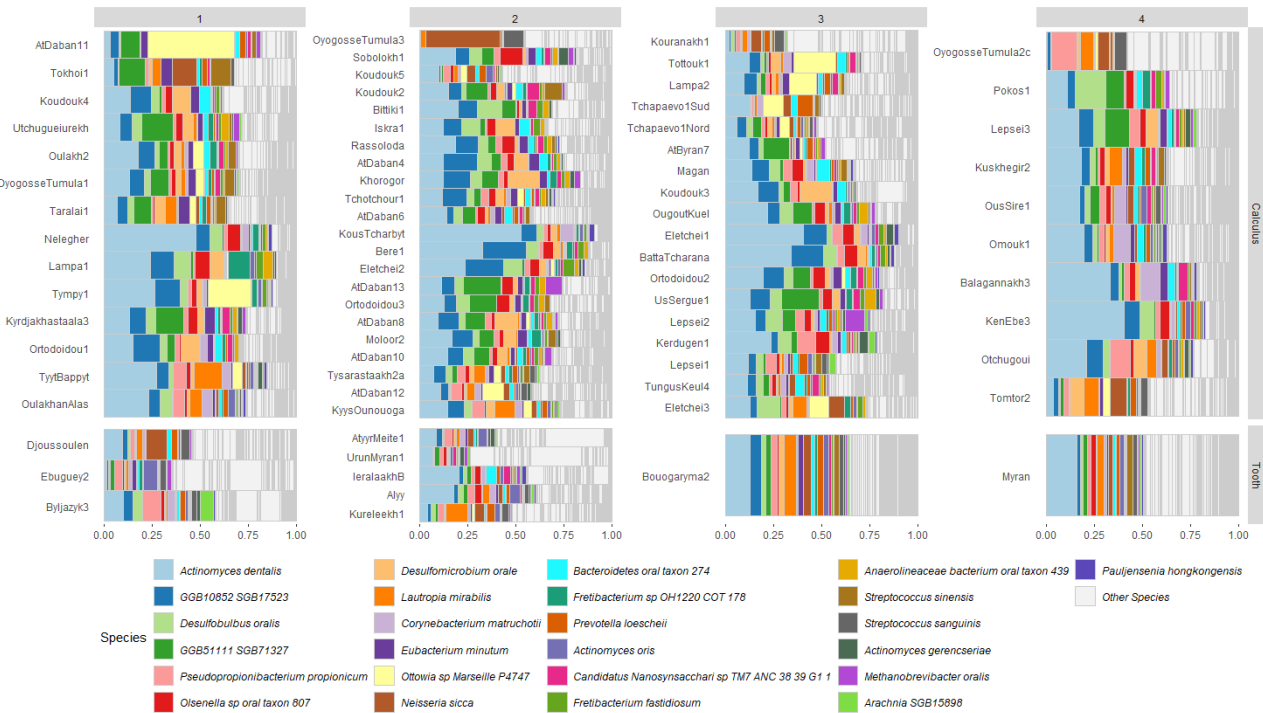


Figure S2_11. Microbial composition at the species level of the samples selected for analyses.

The 25 taxa that are most abundant across all samples are displayed in colors, while the other less abundant species identified are shown in light grey. Results are organized in four columns, representing each archaeological Stage (1 to 4), and separated by sample type (*i.e.*, dental calculus *vs.* tooth).

We performed a network analysis, following Quagliariello et al.¹¹⁴, to reveal subtle commonalities in the microbial compositions of the different samples that may have remained undetected in PCoA. We first assessed the presence of clusters within our dataset, reflecting genus-level patterns of similarities, on the basis of the goodness of clustering measured through 100 bootstraps. Here, the clusGap function from the cluster package (v.2.1.0) applied to a phyloseq (v.1.30.0)¹¹⁵ conversion of the matrix of read counts. This step revealed the optimal k number of clusters present in our dataset. We then used NetCoMi (v.1.0.2)¹¹⁶ to generate a network based on pairwise Aitchison distances calculated at the genus level, that we sparsified with the k -nearest neighbour (k -nn) algorithm. Finally, the sparsified network was subjected to hierarchical clustering, using the average linkage method and the k value obtained above (method = “average”; $k = 3$). A Pearson's Chi-squared test did not support any association in the distribution of individuals amongst clusters and archaeological stages (p -value=0.92, X-squared=1.9543, df=6; Figure S2_14).

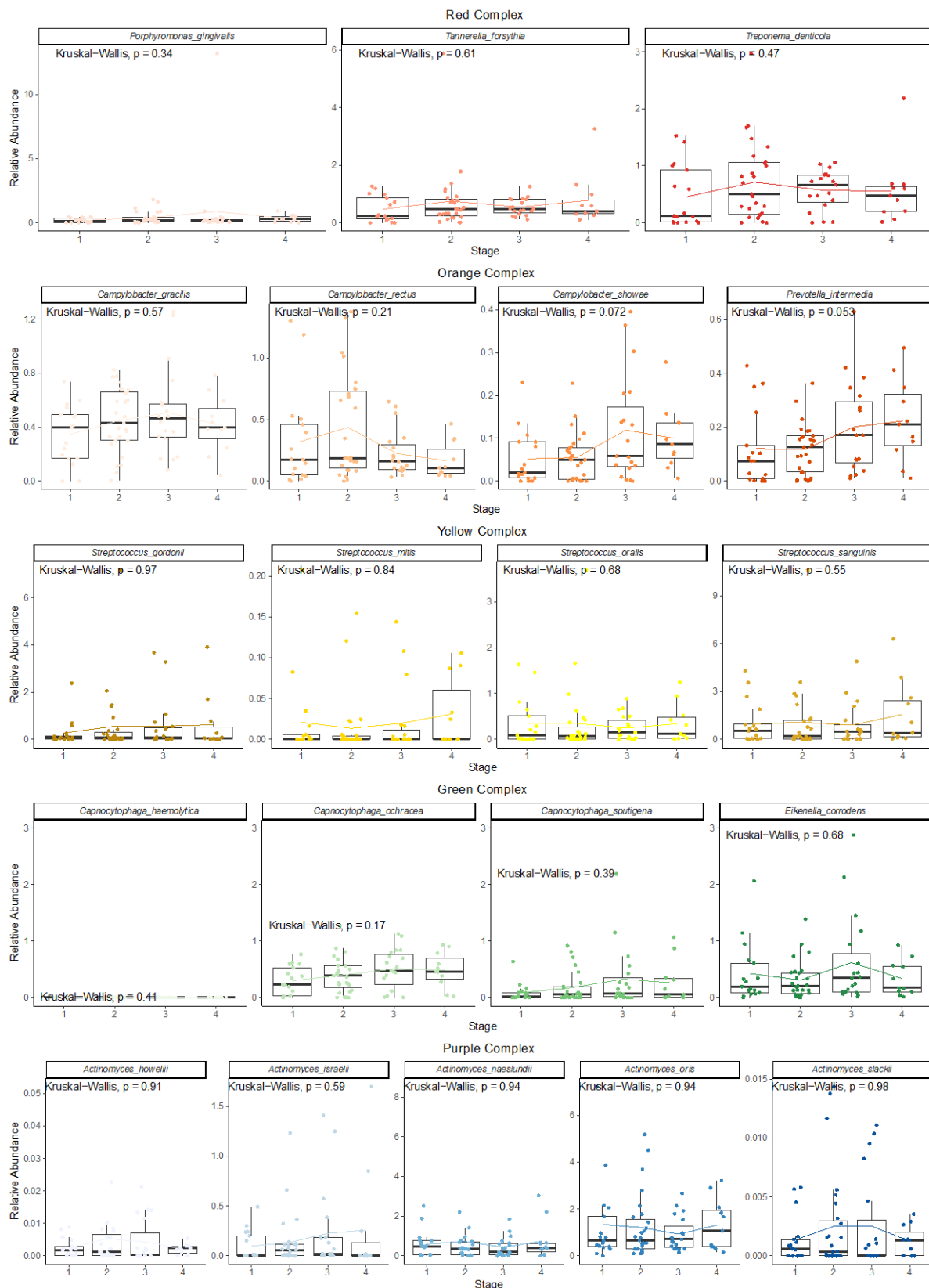


Figure S2.12. Relative species abundance for the microbes underlying five bacterial complexes.

Boxplots are defined as follows: minimum, maximum and center values represent the 25%, 75% and 50% quantiles, respectively; upper and lower whiskers are the maximum/minimum values of data within the 1.5 interquartile range over the 75th or under the 25th percentiles, respectively. Horizontal black segments show mean values for each species at each stage. For Stage 1, N= 17; Stage 2, N=27; Stage 3, N=19; and Stage 4, N=11.

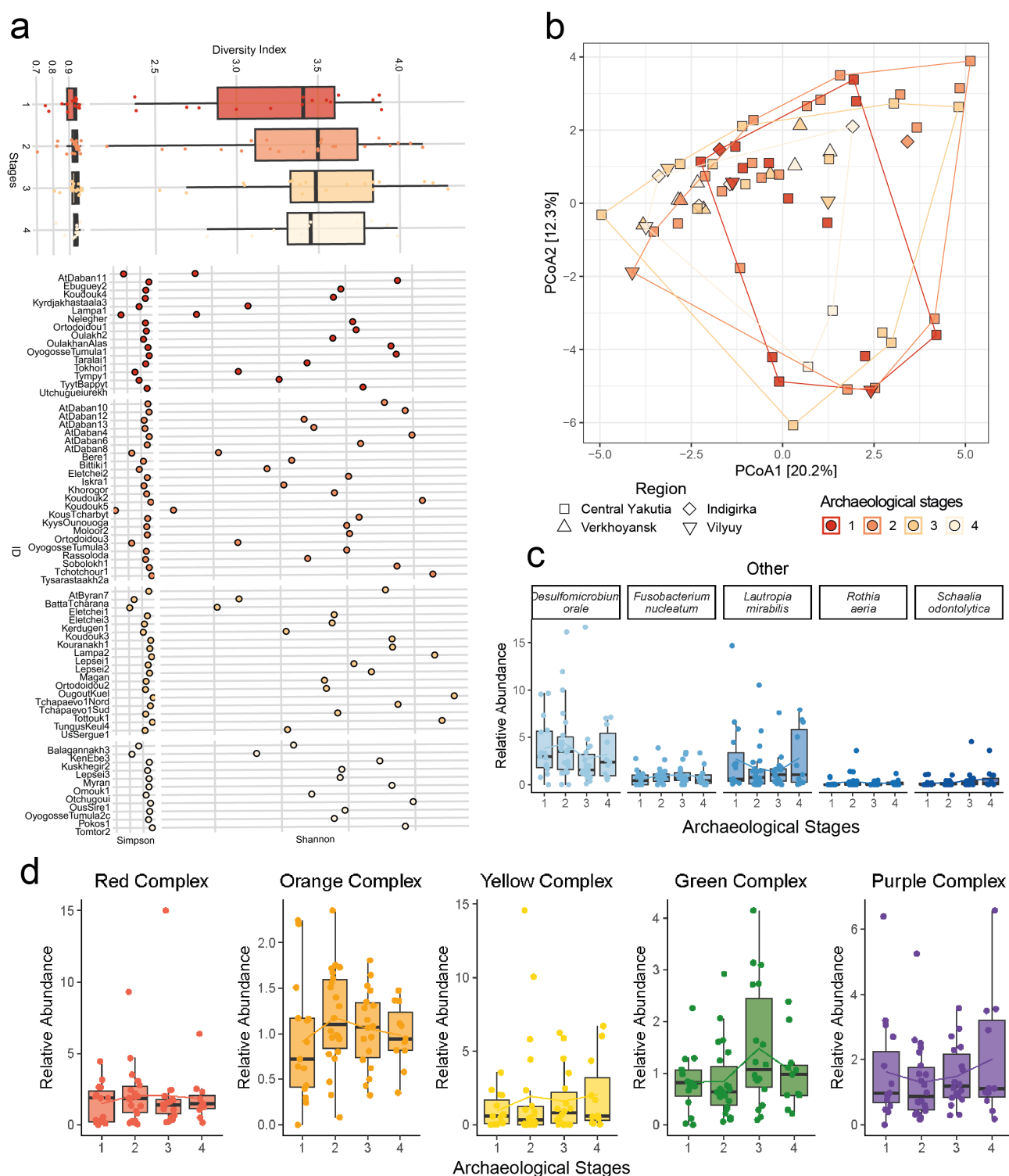
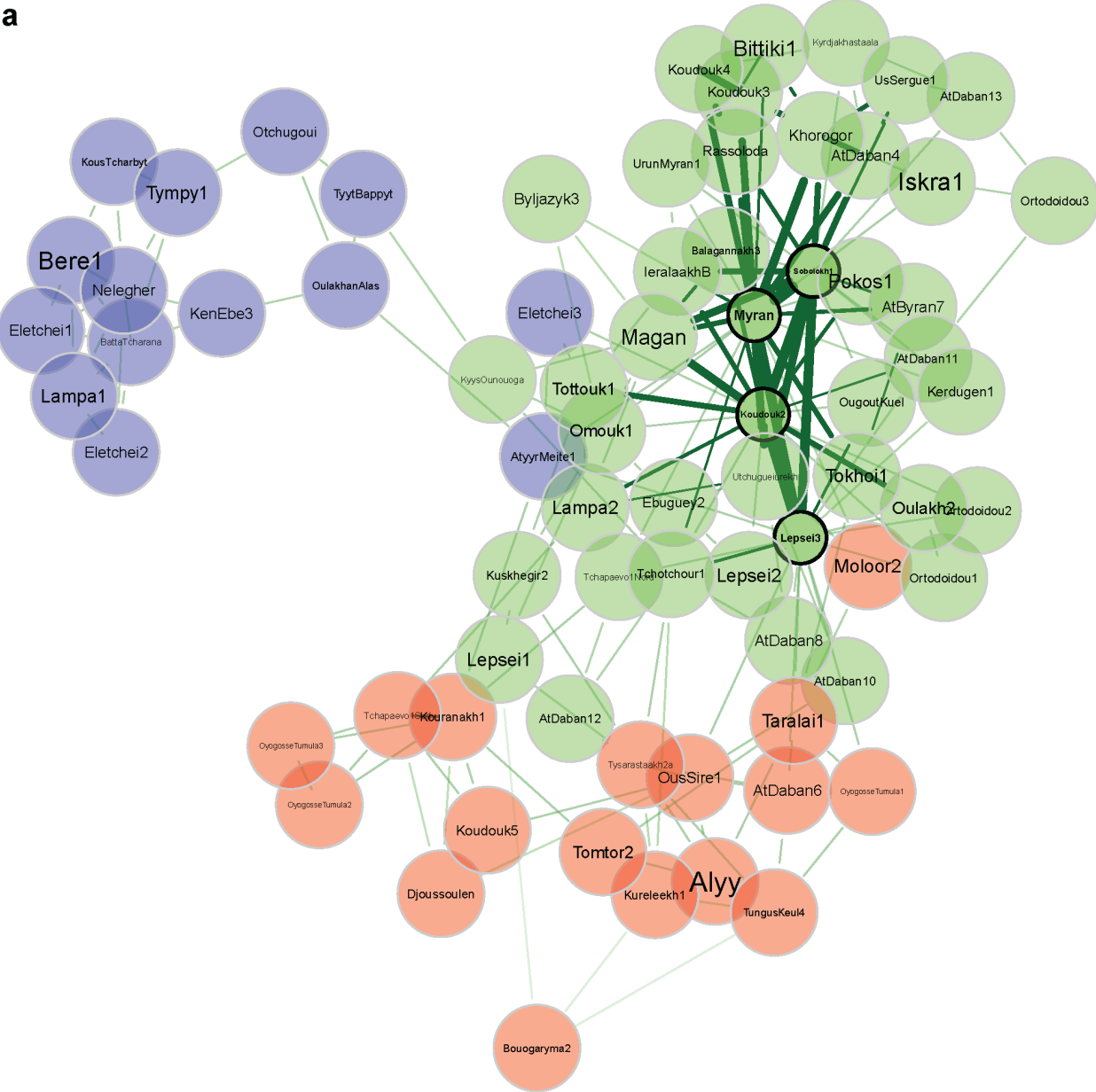


Figure S2_13. Taxonomical analyses for the dataset restricted to dental calculus only.

a) Diversity indexes for taxonomic diversity across the four archaeological stages. *Top*: Boxplot of both Shannon and Simpson diversity indexes for each Stage. *Bottom*: Per-individual diversity indexes. b) PCoA of 66 calculus samples carried out on Aitchison distances calculated on high-quality counts for 448 unique taxonomic species (abundance>1%) from the MetaPhlAn4 database. c) Relative species abundance of five oral pathogens. d) Relative species abundance of five bacterial complexes. Boxplots are defined as follows: minimum, maximum and center values represent the 25%, 75% and 50% quantiles, respectively; upper and lower whiskers are the maximum/minimum values of data within the 1.5 interquartile range over the 75th or under the 25th percentiles, respectively. Horizontal black segments show mean values for each species at each stage. For Stage 1, N= 15; Stage 2, N=22; Stage 3, N=18; Stage 4, N=9.

a



b Stages

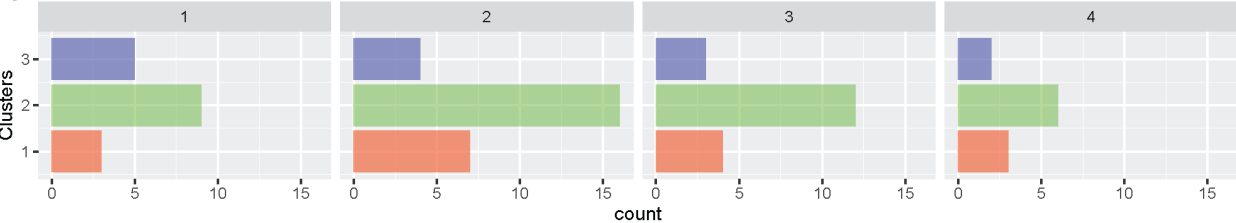


Figure S2_14. Network analysis of the Yakut oral microbiome taxonomic diversity.
a) Network and cluster. Edge thickness reflects their weight, which indicates non-negative similarities among samples. Edge length represents Aitchison distance between nodes. Colors represent the clusters identified by the hierarchical cluster analysis and black outlines indicate hubs. b) Distribution of the individuals from the four archaeological stages into the clusters identified by hierarchical cluster analysis.

2.15. Strain analyses

To investigate strain-level variation in the Yakut oral microbiome, we used StrainPhlAn^{411,117}, a population genomics tool undertaking metagenomic strain-level analysis. This analysis was carried out on the species belonging to the red complex and the eight most prevalent bacterial species in our dataset. In addition to the Yakut individuals presented in this study, we also included a set of reference samples from the AncientMetagenomeDir¹¹⁸ repository v24.09, to provide interpretative context. More specifically, we included Neanderthal oral metagenomic data as outgroup¹¹³, as well as oral metagenome data for 50 additional individuals dating back to the last 500 years at most, and originating from Eurasia and Africa^{113,119–123}. The description of the comparative dataset can be found in Supplementary Table 1n. The sequence data underlying our comparative panel were pre-processed following the same pipeline as that described above (Supplementary Information section 2.14) for each Yakut extract. StrainPhlAn4 was run with default parameters to extract species-specific MetaPhlAn4 markers per extract. We removed any sample with less than 80% of the markers covered, and any marker present in less than 80% of the samples. Those reads aligning against these markers were combined to form a multifasta alignment for phylogenetic reconstruction, adding the respective reference genome sequence. Phylogenetic reconstructions were carried out in the maximum-likelihood framework of in IQtree 1.6.12¹²⁴, with the best substitution model estimated using the Akaike Information Criterion (-m MFP), and node support assessed from 1,000 ultrafast bootstrap¹³⁴ pseudo-replicates (UFBoot) (each bootstrap tree optimized using a hill-climbing nearest neighbor interchange (NNI) search; -bb 1000 -bnni). Trees were then rooted using Neanderthal samples as outgroups when Neanderthal alignments could be characterized (*Actinomyces dentalis*, *Desulfobulbus oralis*, *Desulfomicrobium orale*, and *Lautropia mirabilis*), or using the tree midpoint otherwise (*Treponema denticola*, *Porphyromonas gingivalis*). In total, we successfully reconstructed phylogenies for six bacterial species providing strain-level information for at least one Yakut individual per archaeological stage, which helped test whether significant shifts in strains accompanied the Russian Conquest.

The individual strains identified amongst ancient Yakut individuals did not cluster by archaeological stages, nor formed monophyletic clades organized per individual region (Figs. S2_15-S2_17). For the four species in which sufficient sequence alignments were detected amongst our comparative non-Yakut historical human samples, these clustered together with strong node support. This indicates different strains circulating in those regions and historical Yakutia. The absence of stratification and significant clustering revealed no significant phylogenetic shifts during Russian Conquest, suggesting that, for the six species investigated, no new strain was introduced and ended up replacing those present locally.

Figure S2_15 (Next page). Strain-level analysis of species *Treponema denticola* and *Porphyromonas gingivalis*.

a) The Maximum Likelihood (ML) phylogenetic tree for *T. denticola* was reconstructed using IQtree 1.6.12¹²⁴, with the TVMe+ASC+R2 model identified as the best substitution model. Phylogenetic reconstruction was based on the positive hits identified in 24 Yakut individuals, plus the sequence of the reference genome. In total, the consensus sequence alignment included 126 StrainPhlAn biomarkers, spanning 1,209 nucleotide positions, with an average individual missingness of 20.68% (range=15.14-34.90%). **b)** The ML phylogenetic tree for *P. gingivalis* was reconstructed using IQtree 1.6.12¹²⁴, with the TVM+F+ASC+G4 model identified as the best substitution model. Phylogenetic reconstruction was based on the positive hits identified in 9 ancient Yakut individuals, supplemented with the data from one historical human, plus the sequence of the reference genome. In total, the consensus sequence alignment included 52 StrainPhlAn biomarkers, spanning 686 nucleotide positions, with an average individual missingness of 11.46% (range=1.02- 23.91%). Trees are rooted at mid-points. Nodes with ultrafast bootstrap support values of $\geq 95\%$ are marked in red, indicating highly-supported monophyletic clusters, although PhyloPhlAn strain analyses are commonly based on non-bootstrapped phylogenetic reconstructions¹²⁶.

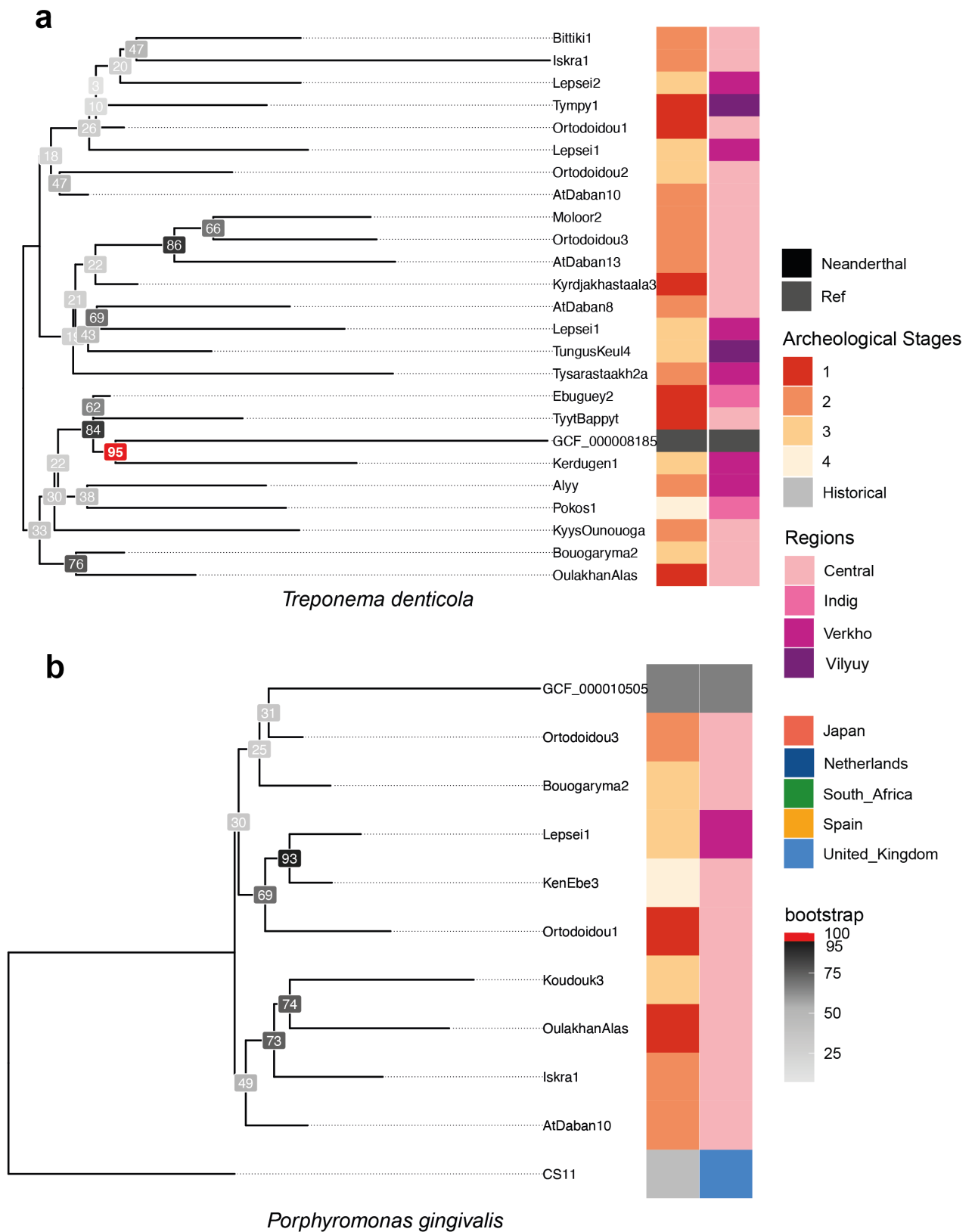


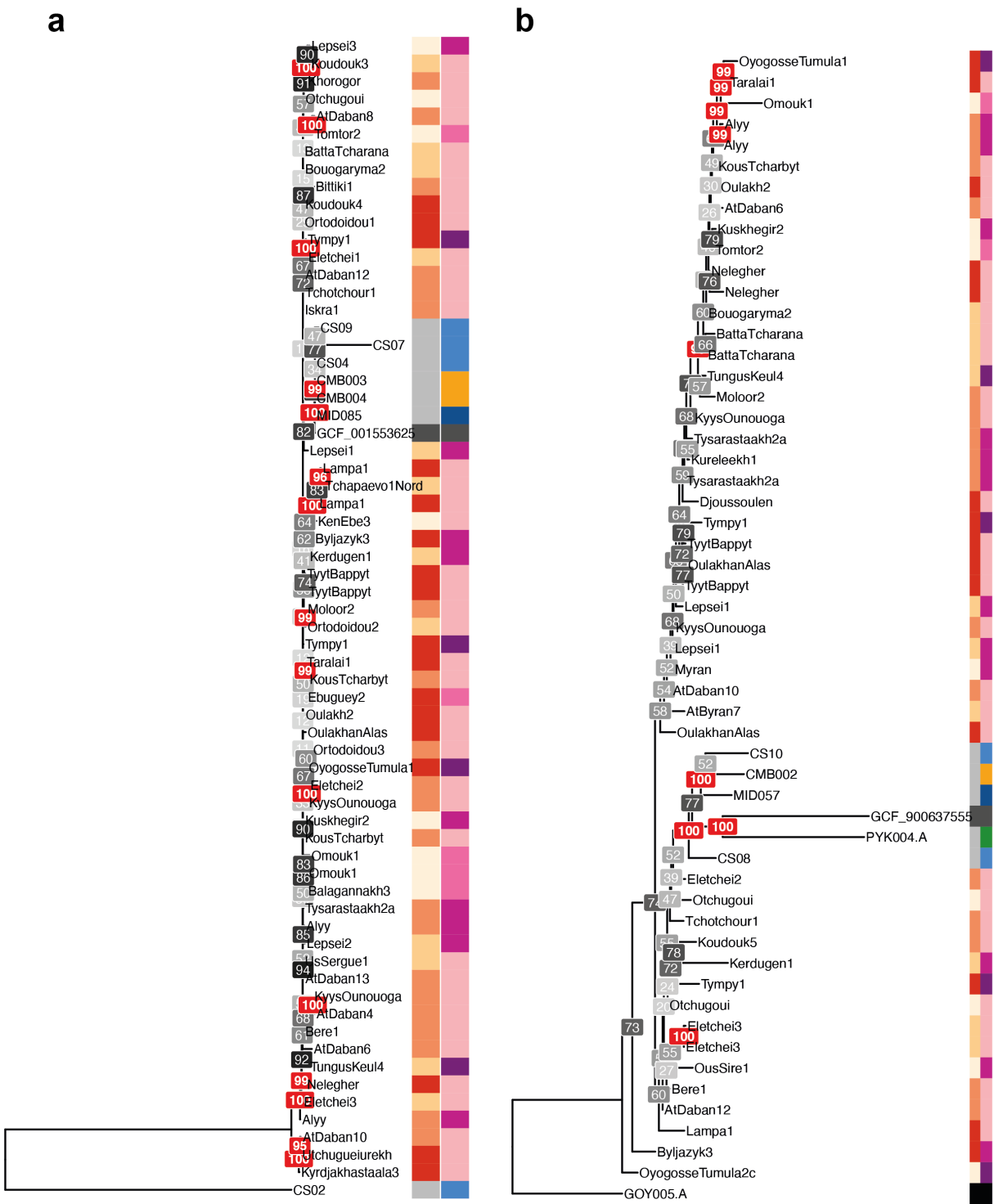


Figure S2_16 (Previous page). Strain-level analysis of *Actinomyces dentalis* and *Desulfobulbus oralis*.

A. The ML phylogenetic trees for *A. dentalis* was reconstructed using IQtree 1.6.12¹²⁴, with the TIM+F+ASC+R2 model identified as the best substitution model. Phylogenetic reconstruction was based on the positive hits identified in 84 ancient Yakut extracts, supplemented with the data from three Neanderthal outgroups and 19 historical humans, plus the sequence of the reference genome. In total, the consensus sequence alignment included 193 StrainPhlAn biomarkers, spanning 1,656 nucleotide positions, with an average individual missingness of 29.08% (range=7.61%- 36.59%). **B.** The Maximum Likelihood phylogenetic tree for *D. oralis* was reconstructed using IQtree 1.6.12¹²⁴, with the TVM+F+ASC+R4 model identified as the best substitution model. Phylogenetic reconstruction was based on the positive hits identified in 73 ancient Yakut extracts, supplemented with the data from six Neanderthal outgroups and 22 historical humans, plus the sequence of the reference genome. In total, the consensus sequence alignment included 196 StrainPhlAn biomarkers, spanning 4,340 nucleotide positions, with average individual missingness of 6.91% (range=1.47- 34.15%). Trees are rooted using the Neanderthal samples as outgroups. Nodes with ultrafast bootstrap support values of $\geq 95\%$ are marked in red, indicating highly-supported monophyletic clusters, although PhyloPhlAn strain analyses are commonly based on non-bootstrapped phylogenetic reconstructions¹²⁶. Individuals appearing more than once in the phylogeny correspond to individuals for which different archaeological materials were analysed, such as tooth roots and dental calculus.

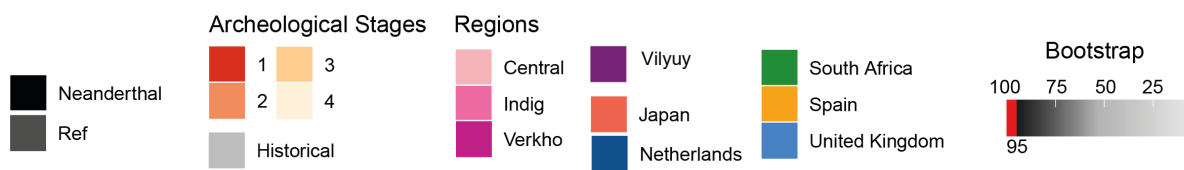
Figure S2_17 (Next page). Strain-level analysis of *Desulfomicrobium orale* and *Lautropia mirabilis*.

A. The Maximum Likelihood phylogenetic tree of *D. orale* was reconstructed using IQtree 1.6.12¹²⁴, with the TIM+F+ASC+R2 model identified as the best substitution model. Phylogenetic reconstruction was based on the positive hits identified in 58 ancient Yakut extracts, supplemented with the data from seven historical humans, plus the sequence of the reference genome. The tree is rooted at mid-point. In total, the consensus sequence alignment included 195 StrainPhlAn biomarkers, spanning 5,048 nucleotide positions, with an average individual missingness of 3.72% (range=0.00- 20.90 %). **B.** The Maximum Likelihood phylogenetic tree for *L. mirabilis* was reconstructed using IQtree 1.6.12¹²⁴, with the TVM+F+ASC+R4 model identified as the best substitution model. Phylogenetic reconstruction was based on the positive hits identified in 48 ancient Yakut extracts, supplemented with the data from one Neanderthal outgroup and six historical humans, plus the sequence of the reference genome. The tree is rooted using the Neanderthal sample as outgroup. In total, the consensus sequence alignment included 186 StrainPhlAn biomarkers, spanning 6,916 nucleotide positions, with an average individual missingness of 14.82% (range=0.10- 34.69%). Nodes with ultrafast bootstrap support values of $\geq 95\%$ are marked in red, indicating highly-supported monophyletic clusters, although PhyloPhlAn strain analyses are commonly based on non-bootstrapped phylogenetic reconstructions¹²⁶. Individuals appearing more than once in the phylogeny correspond to individuals for which different archaeological materials were analysed, such as tooth roots and dental calculus.



Desulfomicrobium orale

Lautropia mirabilis



269

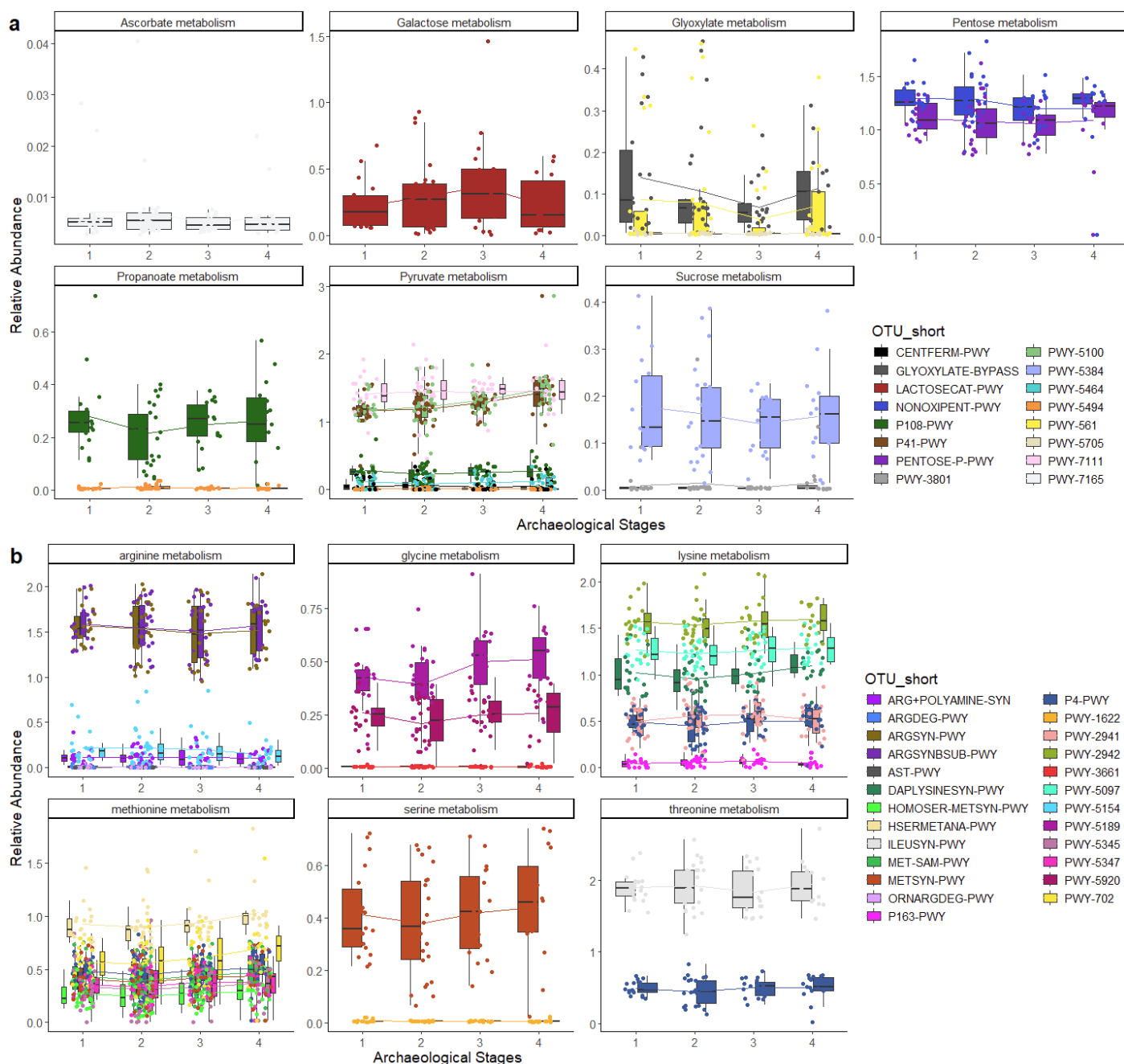
270

—

2.16. Functional oral diversity

The sequence data passing the Sourcetracker2 filters described above were also subjected to functional analyses, using the methodology implemented in HUMAnN 3.0¹²⁵, with default parameters. This step generated per-individuals functional profiles based on the UniRef90¹²⁷ and ChocoPhlAn (January 2023)^{127,128} databases, which were further joined by pathways, normalized by counts per millions, and centered log-ratio transformed as appropriate for compositional values, before conducting a PCA (Fig. 3b). Selected pathways associated with carbohydrate or amino acid metabolism were scrutinized for their relative abundances across individuals, and compared by archaeological stages, using a Kruskal-Wallis test (p -values ≥ 0.067 ; Fig. 3c and Figure S2_18).

These analyses were repeated on the dataset restricted to dental calculus samples only. Here, we removed the individuals with only tooth samples analyzed, and kept only the calculus data for those individuals associated with both tooth and calculus data. Beta-diversity and abundance levels of carbohydrate pathways showed the same temporal trends as analyses included both tissue types (Figure S2_19). Statistical assessment of differences in composition between region, period, sex and cultural group also returned similar results, as none were significant (Supplementary Table 1m).



288

289 **Figure S2_18. Relative abundance of the metabolic pathways across archaeological stages.**

290 a) Relative abundance of the detailed pathways underlying carbohydrate metabolism. b) Relative abundance of the

291 detailed pathways underlying amino-acid metabolism. Boxplots are defined as follows: minimum, maximum and center

292 values represent the 25%, 75% and 50% quantiles, respectively; upper and lower whiskers are the maximum/minimum

293 values of data within the 1.5 interquartile range over the 75th or under the 25th percentiles, respectively. Horizontal

294 black segments show mean values for each species at each stage. Horizontal black segments show mean values for

295 each species at each stage. For Stage 1, N= 14; Stage 2, N=24; Stage 3, N=19; and Stage 4, N=11.

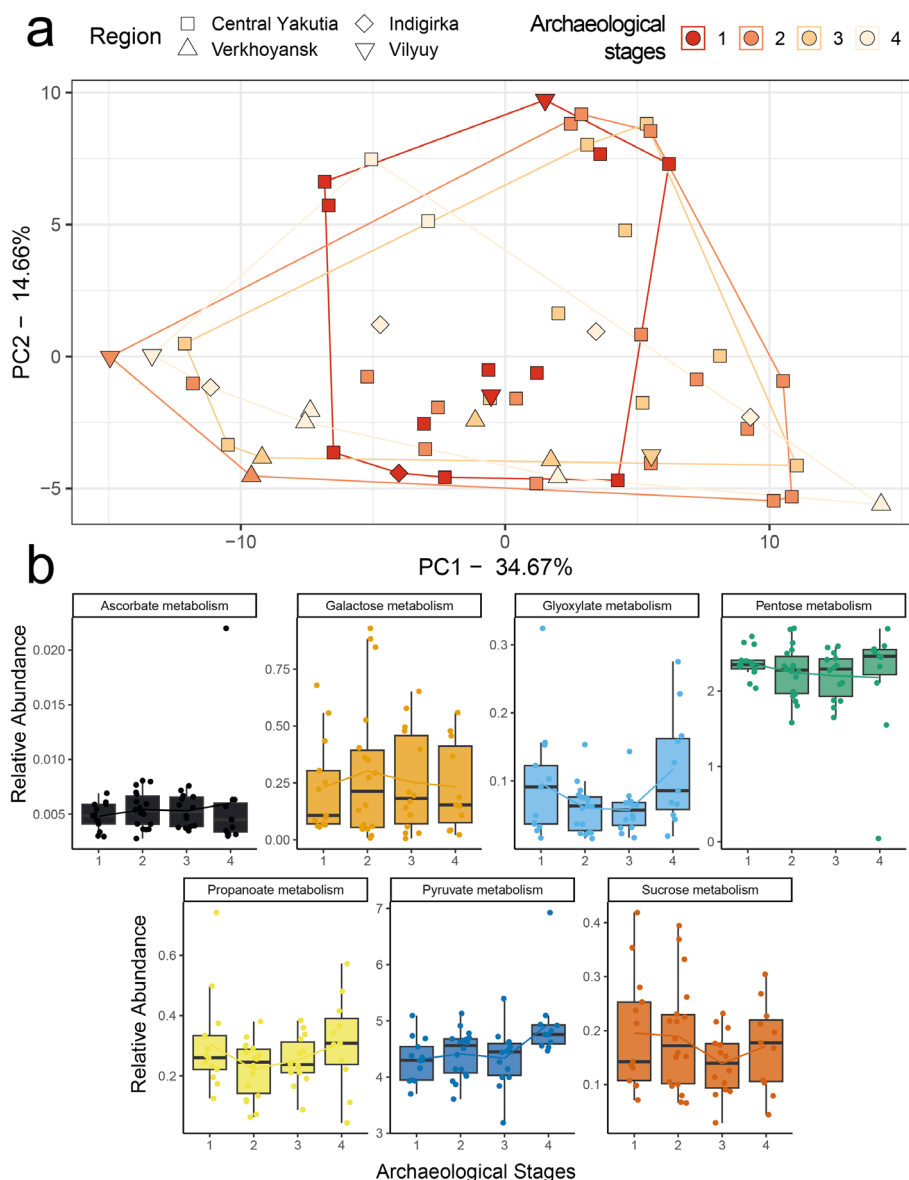


Figure S2_19. Functional analyses for the calculus only dataset.

a) PCA of 58 calculus samples carried out on Aitchison distances calculated on high-quality counts for 303 pathways, after a centered log-ratio transformation. b) Relative abundance of the pathways underlying carbohydrate metabolism. Boxplots are defined as follows: minimum, maximum and center values represent the 25%, 75% and 50% quantiles, respectively; upper and lower whiskers are the maximum/minimum values of data within the 1.5 interquartile range over the 75th or under the 25th percentiles, respectively. Horizontal black segments show mean values for each species at each stage. For Stage 1, N= 13; Stage 2, N=19; Stage 3, N=14; and Stage 4, N=11.

2.17. Pathogen screening

Reads aligned on the human genome (hg37) and on the human mitochondrial genome were filtered out. The resulting filtered fastq files were used for mapping against a selection of reference genomes from candidate pathogens (N=26; Figure S2_20a). This alignment step was carried out using PALEOMIX¹⁹ bam_pipeline v1.2. and bwa-0.6¹¹⁰ (backtrack, MinQuality: 30, no seed, -n: 0.1), which produced high quality

bam alignments that were removed for PCR duplicates. The number of aligned reads against each reference genome were counted per sample, and the average read-to-reference edit distance was also calculated. We considered a sample positive for the presence of any given pathogen as long as a minimal number of 100 high-quality alignments were identified, and the average edit distance was equal to or below 0.01. This conservative approach resulted in the identification of three individuals positive for *Variola major*, the etiologic agent of smallpox (AC1S2, AC1S3 and Rassoloda; Figure S2_20b).

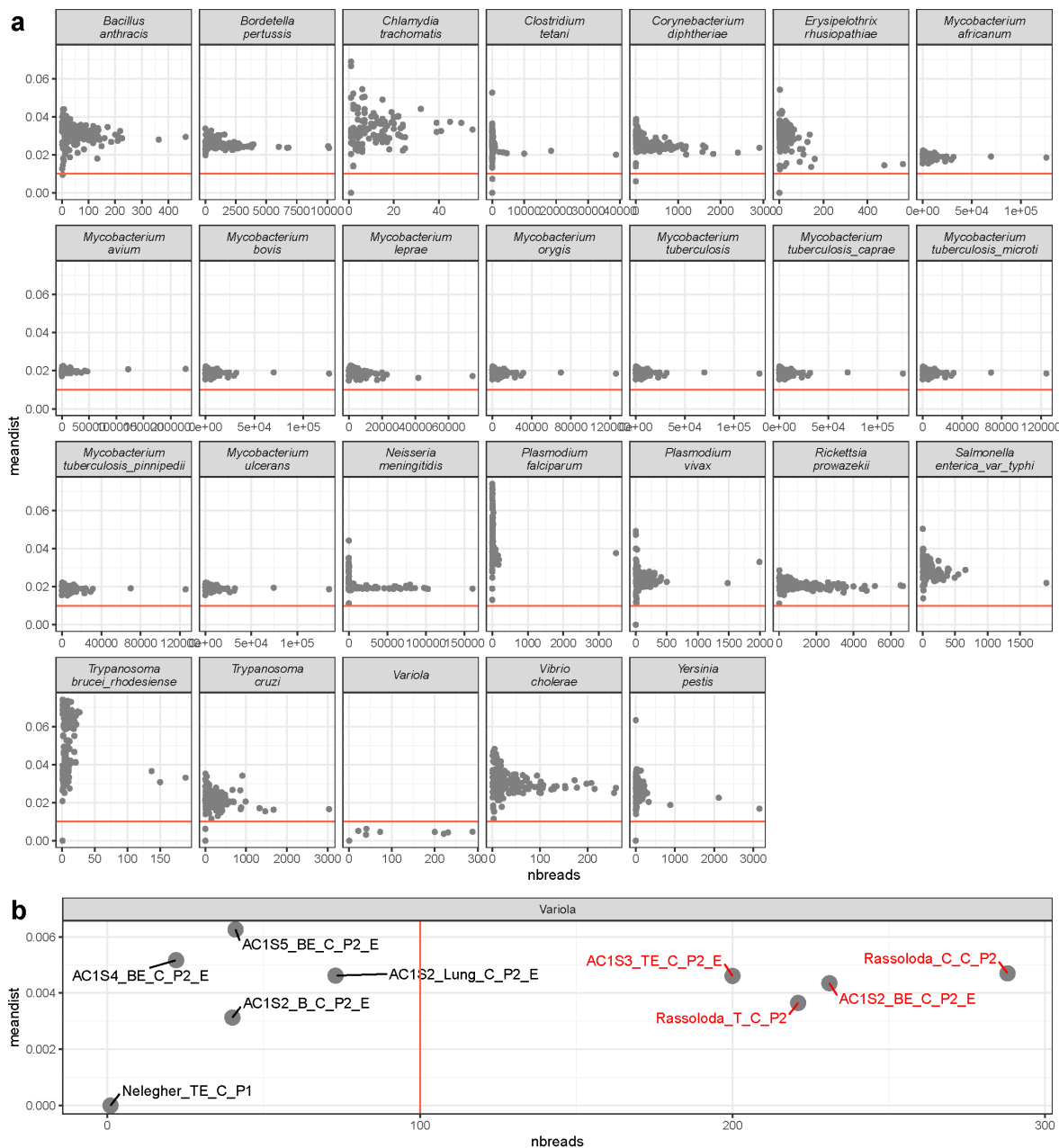


Figure S2_20. Results of read alignments against a selection of selected microbial pathogens. Meandist is the read-to-reference edit distance normalized by read length. Nbreads is the number of per-individual reads aligned against the pathogen genome considered. The red line is set for a mean distance of 0.01. **a)** All pathogens screened. **b)** Details of the individuals with reads aligned against the Variola virus. Samples with more than 100 reads and a mean distance below 0.01 were considered positive and shown in red.

2.18. Variola strains phylogeny

All the sequence data generated for the three individuals positive for smallpox were realigned against the VARV smallpox reference genome (Accession Nb. NC_001611.1), using PALEOMIX¹⁹ bam_pipeline v1.2. and bwa¹¹⁰ (backtrack, MinQuality: 30, no seed, -n: 0.1), which produced high quality bam alignments that were removed for PCR duplicates. While positive, AC1Num3 did not provide a sufficient number of reads (N=263) to proceed further with the rest of the analyses. Furthermore, genotypes were called using bcftools³⁵ mpileup and call modules, requiring a minimal base and mapping Phred quality-scores of 30, and considering the genome haploid. Low-quality genotypes (Phred quality-score<30), indels, and polymorphisms within 2 base-pairs from an indel were removed using bcftools³⁵ filter. The mapping statistics for the three positive individuals are given in Supplementary Table 1p.

To place the smallpox strains identified in the smallpox evolutionary tree, we applied the same procedure as above to the raw reads previously published for four ancient samples from the Northern Europe dating to the 7th to 10th century CE¹²⁹, and one sample from the 17th century CE in Lithuania¹³⁰. Additionally, the fasta sequence data corresponding to 45 smallpox genomes from the 20th century CE previously characterized were downloaded¹³¹ (Supplementary Table 1q), including the reference genome, one camelpox¹³² and one taterapox¹³¹ genome that were used as outgroups. The multifasta sequence data comprising a total of the 45 modern viral genomes were further aligned using MAFFT¹³³, with default parameters (--auto), visually checked, and manually corrected wherever appropriate. Gaps were added to the six ancient samples according to the gaps in the reference genome after the alignment procedure, and all fasta were merged to form a multifasta sequence of 52 viral genomes. Positions in which at least 50% of the sequences were covered were retained for Maximum Likelihood (ML) reconstruction in IQtree 1.6.12¹²⁴ (Fig. 3e). The best substitution model (-m MFP) was identified as TVM+F+ASC+G4, and node support was estimated from 1,000 ultrafast bootstrap¹³⁴ pseudo-replicates (UFBoot), with each bootstrap tree optimized using a hill-climbing nearest neighbor interchange (NNI) search based on bootstrap alignment (-bb 1000 -bnni). To ensure reproducibility, we also performed the exact same procedure ML tree reconstruction as above but based on the direct output of MAFFT¹³³ without visual checking and correcting of the modern viral genome alignment, with the ancient genomes added as described above. The resulting tree yielded the same topology for the well supported nodes (Figure S2_21).

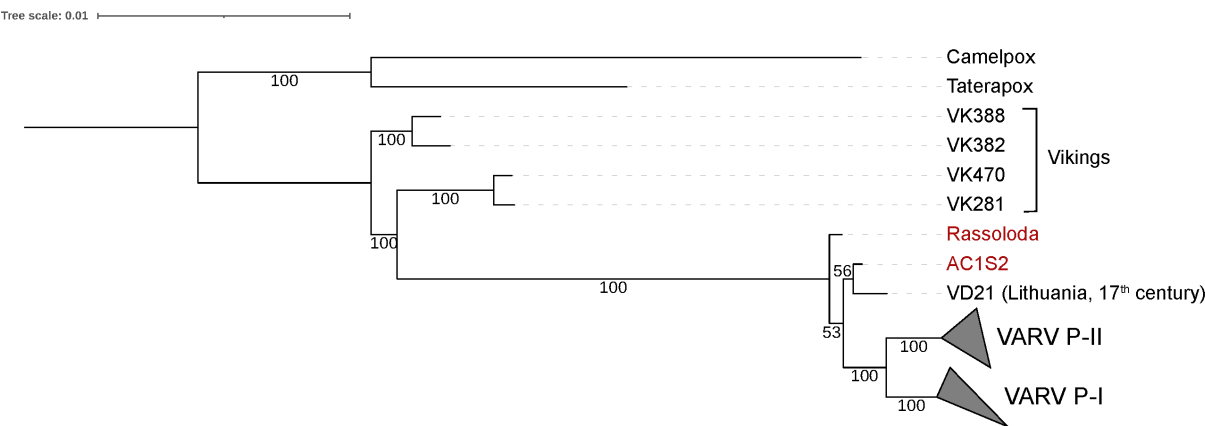


Figure S2_21. Maximum Likelihood phylogenetic tree of the two Yakut smallpox strains, based on raw MAFFT¹³³ alignment.

Node supports are based on 1,000 ultrafast bootstrap pseudo-replicates.

The positions of our two samples in the tree obtained was then tested against seven alternative tree conformations (Figure S2_22) by running an approximately unbiased (AU) topology test¹³⁵ with IQtree (-n 0 -zb 10000 -zw -au). The original topology was identified as the best topology amongst the alternatives tested, which assumed various affinities between the Yakut data and previously characterized ancient smallpox data (p -value=0.976 *vs.* ≤ 0.03).

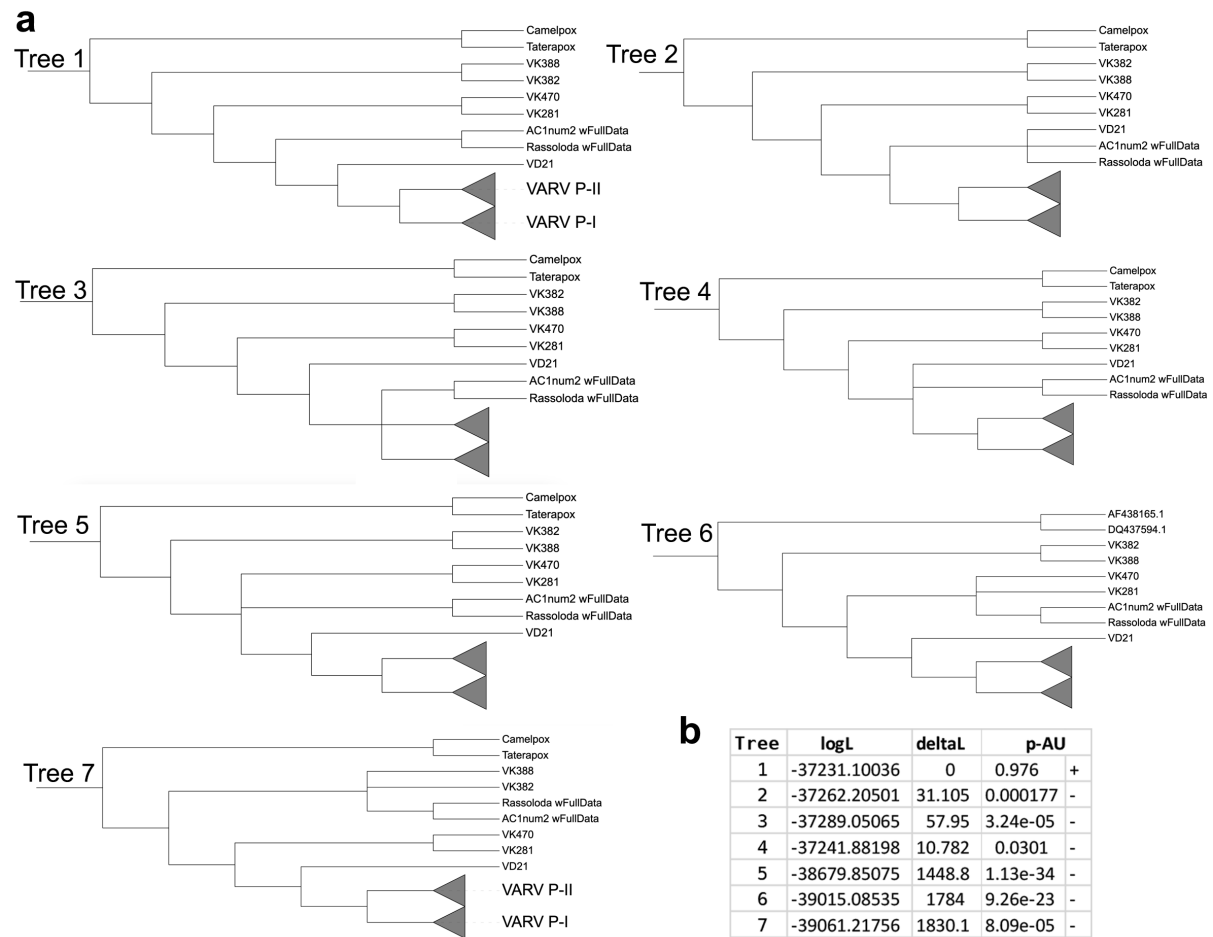


Figure S2_22. Details of the topological tests for the variola tree.

a) Seven alternative tree conformations. AC1num2 stand for AC1S2. VD21 is the Lithuanian strain from the 17th century CE previously reported by Duggan and colleagues¹³⁰. VK strains are the ones from the Viking age¹²⁹. b) Results of three different tests, providing maximal support for tree topology 1.

2.19. Ancient DNA methylation

Correlation with modern bone data

We used DamMet¹³⁶ to evaluate DNA methylation levels on the genomes from 21 individuals with coverage ≥ 9 -fold, as previous work established that relatively high coverage was needed to obtain reliable estimate. Overall, we followed the procedure previously described by Liu et al.¹³⁷ to identify the best combination of parameters for DamMet¹³⁶ DNA methylation inference. Briefly, three sizes of genomic windows were assessed (including 10, 25 and 50 CpGs, respectively), and the expected average cellular methylation levels (-M) was set between 70% and 95%, considering every value by increments of 5%. The minimal base and mapping qualities were 20 and 30 (Phred-scaled), respectively, and bases located further away than 30 bases from read termini were disregarded. The best combination of parameters was assessed as the one maximizing the R^2 coefficient of Spearman correlation to known DNA methylation values for bones at 450k CpGs, available from the literature for one female and three male individuals¹³⁸. Overlapping CpG windows were flagged and those presenting maximal CpG coverage were retained.

To further identify minimal depth requirements and assess the impact of the physical size of the genomic windows on the reliability of DNA methylation estimates, we measured Spearman correlation conditioning on windows of minimal depth 0-1000 reads (considering all possible value by increments of 100), and maximal sizes (500, 1,000, 5,000 and 1,0000 bp; Supplementary Table 1u). The average cellular methylation fraction (M) was found to have no impact on correlation levels (Figure S2_23), hence, a value of 75% was retained. Maximal correlation levels (0.38-0.8) were otherwise obtained for a maximum window size of 1,000 bp, windows of 25 CpGs, and a minimum depth of 400 reads per window. Four individuals (Odjullunn2, OugoutKuel, TouoraUrek2 and TouoraUrek3) presented low correlation scores (Spearman correlation $R^2 < 0.55$), in relation to a global underestimation of their DNA methylation levels, suggesting insufficient data. Those individuals were, thus, disregarded.

Despite encouraging correlation levels, two DNA methylation categories associated with scores of 0 and 1 were under-represented in the remaining samples (Figure S2_24a), in line with the work from Liu et al.¹³⁷. We therefore followed the mitigation procedure developed by those authors to improve ancient DNA methylation inference, using ~ 27.2 million CpGs in two modern bones published by Gokhman et al.¹³⁹ (Figure S2_24b).

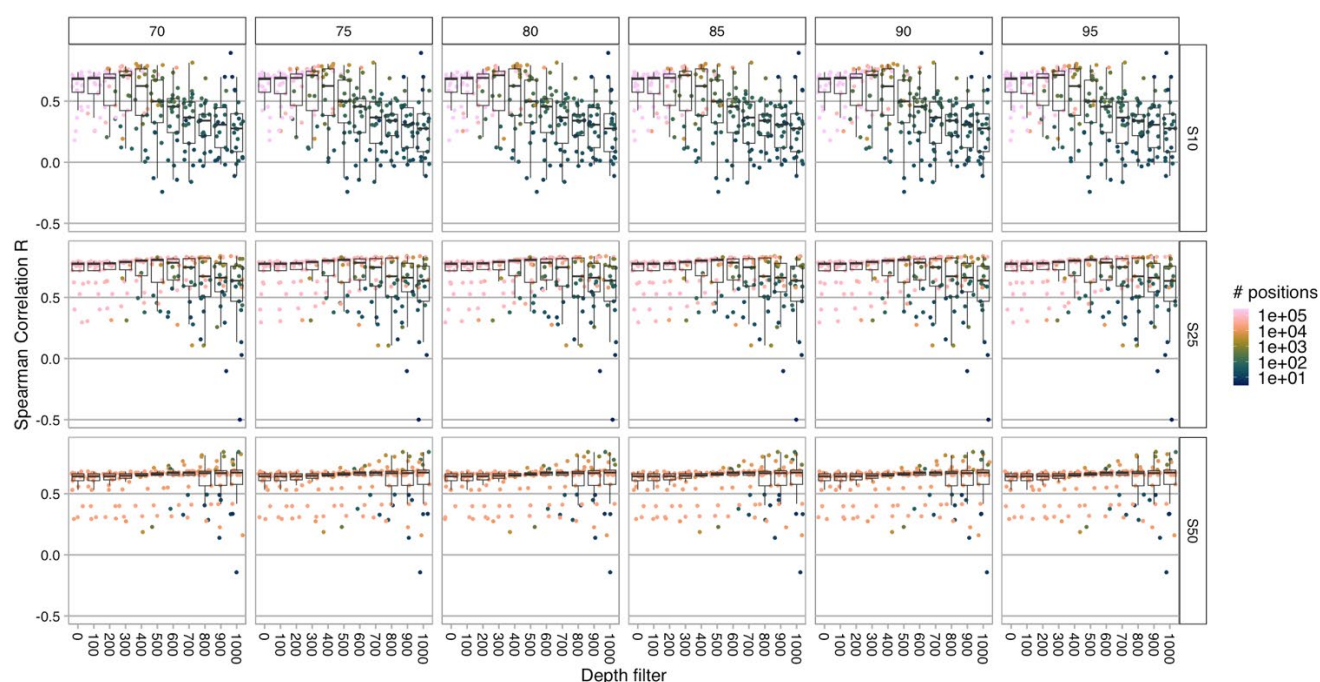


Figure S2_23. Impact of depth filter (x axis), number of CpGs per window (S10, S25, S50, for ten, 25 and 50 CpGs, respectively), and expected average cellular methylation levels (x strip = 70 to 95%) on the correlation between the DNA methylation values in the Yakuts individuals (N=21) and modern references.

The color gradient represents the number of CpGs positions used after removing windows larger than 1kb and considering ten minimal depth thresholds (x-axis). In this analysis, depth refers to the cumulated number of sequences covering all individual CpGs present in a window). The correlation measured between ancient and modern bone data provides an assessment of the quality of the inference, following Hanghøj and colleagues¹⁴⁰ and Liu and colleagues¹³⁷.

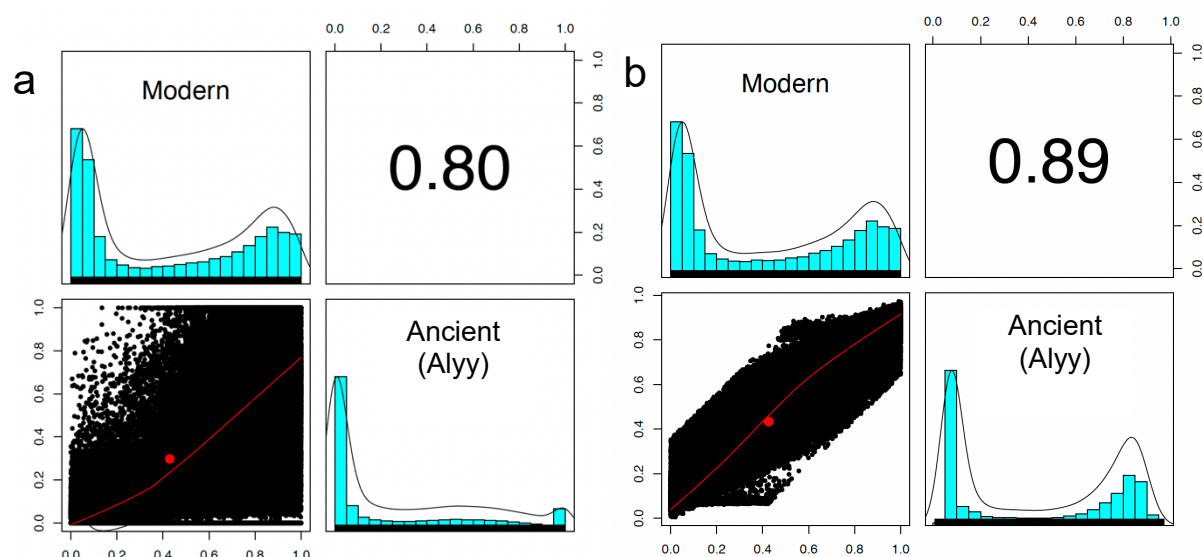


Figure S2_24. Correlation of the DNA methylation values (F) for Alyy before and after numerical transformation.

Distribution of DNA methylation values averaged across modern individuals and raw F values predicted by DamMet¹³⁶ for sample Alyy, and their Spearman correlation coefficients R2, before (a) and after (b) the two-step mathematical transformation of ancient DNA methylation values described in the STAR Methods of the Liu and colleagues¹³⁷ publication.

Intragenomic quality test of DNA methylation

The validity of DNA methylation inference was also assessed by checking for the presence of well-established patterns along the genome, following method by Hanghøj and colleagues¹⁴⁰. For example, we looked at CpG Islands and their flanking shores (0–2 kb) and shelves (2–4 kb) (shelves are associated with higher DNA methylation levels than shores, and shores are associated with higher DNA methylation levels than CpG Islands¹⁴¹), as well as exon and intron regions (the former show lower DNA methylation levels than the latter¹⁴²) (Figure S2_25 and Figure S2_26). We also explored DNA methylation patterns within the 2-kb regions flanking CTCF binding sites, which display periodic variation of DNA methylation levels when bound to CTCF^{143,144}. For the CTCF binding sites, the transformed DNA methylation scores were normalized over the background using LOESS function in R¹⁴⁵ (span=0.13, representing the value minimizing residual standard errors in a range of possible values between 0.01 and 0.5), for signal denoising according to Hanghøj and colleagues¹⁴⁰. The transformed DNA methylation scores monotonically increased moving away from the occupied CTCF binding sites to the most external boundaries of the region. This trend was removed by obtaining position-specific average of transformed DNA methylation scores using LOESS function in R (span =0.3) that we subtracted to the denoised transformed DNA methylation scores values (Figure S2_27ac). Periodicity was assessed using spectral density analyses, using the spectrum function from the stats package in R¹⁴⁵ (Figure S2_27bd). The DNA methylation profile observed for the Otchugoui individual at CTCF sites, as well as at the CpGs islands and intron/exons was not on par with expectation and were, thus, disregarded (Figures S2_25-S2_27).

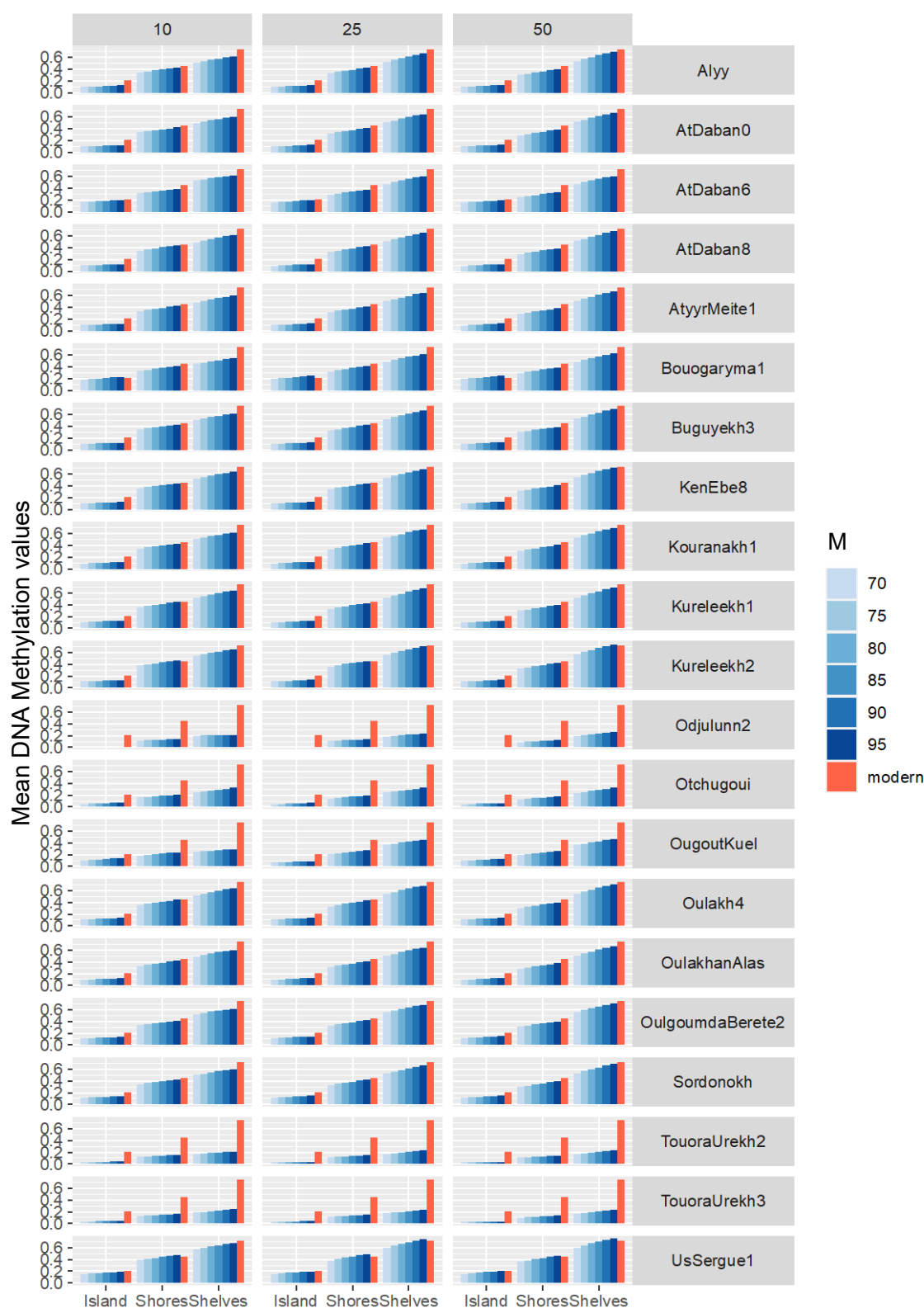


Figure S2_25. Genomic regions with known DNA methylation levels including CpG islands, their shores and shelves for all Yakuts with DNA methylation data.

Modern (red) is provided as a comparative baseline (the average across 3 males is given for male Yakut and the modern female value is given for female Yakut). DNA methylation values (F) are calculated on windows of a maximum size of 1kb and with a minimum depth of 400 reads, but with different numbers of CpGs per window (S10, S25, and S50, for ten, 25 and 50 CpGs), and different expected average cellular DNA methylation levels (M; blue shades; 70 to 95%).

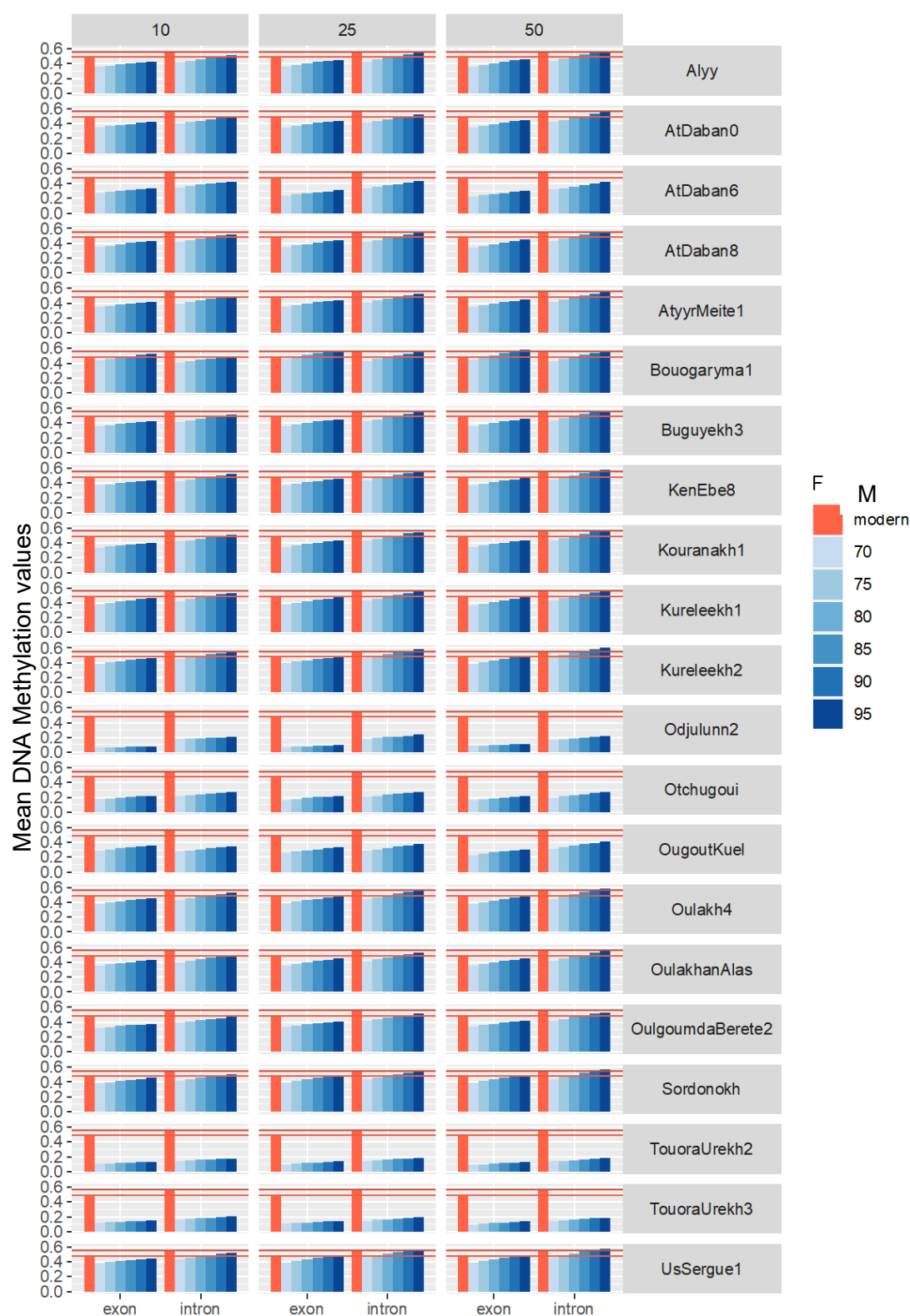


Figure S2_26. Exonic and intronic regions with known differentiated DNA methylation levels for all Yakuts with DNA methylation data.

Modern (red bar and orange lines) is provided as a comparative baseline (the average across 3 males is given for male Yakut and the modern female value is given for female Yakut). DNA methylation values (F) are calculated on windows of a maximum size of 1kb and with a minimum depth of 400 reads, but with different numbers of CpGs per window (S10, S25, and S50, for ten, 25 and 50 CpGs), and different expected average cellular DNA methylation levels (M; blue shades; 70 to 95%).

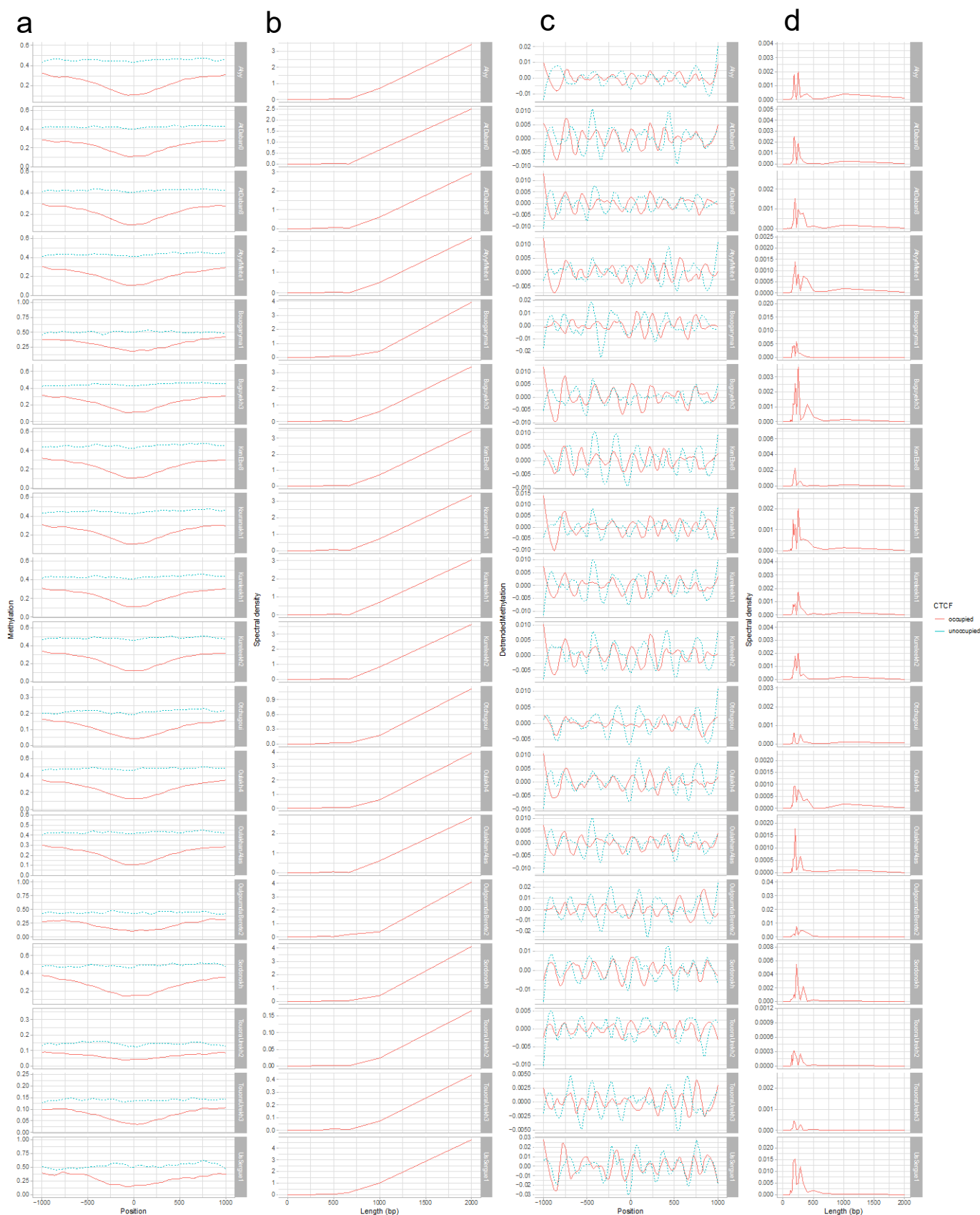


Figure S2_27. DNA methylation profiles around CTCF binding sites.

The profiles recovered at occupied and unoccupied CTCF-binding sites¹⁴⁴ are reported in red and blue, respectively.

a) Raw DNA methylation values at CTCF regions (transparent) and normalized over the background values (solid). b)

Spectral density plot for F values. c) & d) Same as a) and b), after detrending.

2.20. Statistical associations between cultural and non-cultural data

MAFSO excavations revealed an extensive array of cultural items associated with every single burial (Supplementary Table 1c). We aimed at testing whether any such item showed non-random associations with DNA-based estimates of kinship and sex, as well as with other variables, such as archaeological stages, regions and age-at-death, all provided by the archaeological contexts (for a total of five categories). To achieve this, we generated a presence-absence matrix summarizing the characteristics of each burial (Supplementary Table 1c), and calculated pairwise Bray-Curtis between individuals. The Ward agglomeration algorithm was applied to squared root of the distance matrix to identify clusters of individuals possibly matching the five categories described above (Figure S2_28). To test whether the distribution of these distances calculated between pairs of individuals in a category (sex, region, archaeological stages) was significantly different to random permutations of individuals across categories, we used ANALyses Of SIMilarities analysis (ANOSIM, *anosim* from the *vegan* package¹⁴⁶ in R¹⁴⁵) and a permutational multivariate analysis of variance (PERMANOVA; *adonis2* from the *vegan* package¹⁴⁶ in R¹⁴⁵) (Fig. 4b, Supplementary Table 1v).

Moreover, we binned individuals into four additional categories defining wealth, leadership, christianity and shamanism, on the basis of the collection of cultural goods find in their burials. More precisely, individuals fell into the “wealth” category if they possessed at least one of the following items: a Russian ceramic teapot (one specimen), a knife with an enameled handle (a scarce Russian specimen), a knife with bronze decorations handle, a pearl-embellished knife sheath, an axe, a bronze cauldron, a lead object (tripod), a bronze bell, an horse comb, a personal belongings box, a silver comb, a fly-whisk, a headband, a belt decorated with pearls or metal ornaments, pearl-embellished boots, a decorated coat, some red wool (a very rare red wool garment), silk (used on the garments-inserts/dress/shirt), a torque, and any number of bracelets and rings³. “Leadership” characterized individuals associated with at least one of the following items: a choron, horse offerings or horse-related objects (stirrup, saddle, or bit)¹⁴⁷. The “Christianity” category discriminated those individuals according to the presence of a cross or a casket of anthropomorphic shape^{3,148}. Finally, “shamanism” was defined following previous archaeological work in Yakutia¹⁴⁹ and restricted to: (i) the adults whose arms or legs were restrained; (ii) those who displayed a virgin belt; iii) those showing one of these items in the following list (command stick, planted birch, hidden sun disk, small bells, clothing in reverse order, shroud on face, sable band, long-hair for male), or; (iv) for the Sobolokh1 female, who was dressed in a manner forbidden to the women of her time. Children presenting such criteria, as well as adults associated with only power command sticks or buried in coffins with shaman engravings, were considered uncertain. Other individuals are described as uncertain, based on the general appearance of the tomb and the evaluations from anthropologist experts, either because the clues present are not sufficiently robust, or because the absence of clues does not definitely rule out the shamanic character. To test whether the similarity of the oral microbiome distances (both taxonomic and functional), the genetic distances and DNA methylation distances (Bray-Curtis) between groups in these four categories, as well as for sex, region and archaeological stages, was lower than to the similarity within each group, we used ANOSIM and PERMANOVA (Fig. 4b, Supplementary Table 1v).

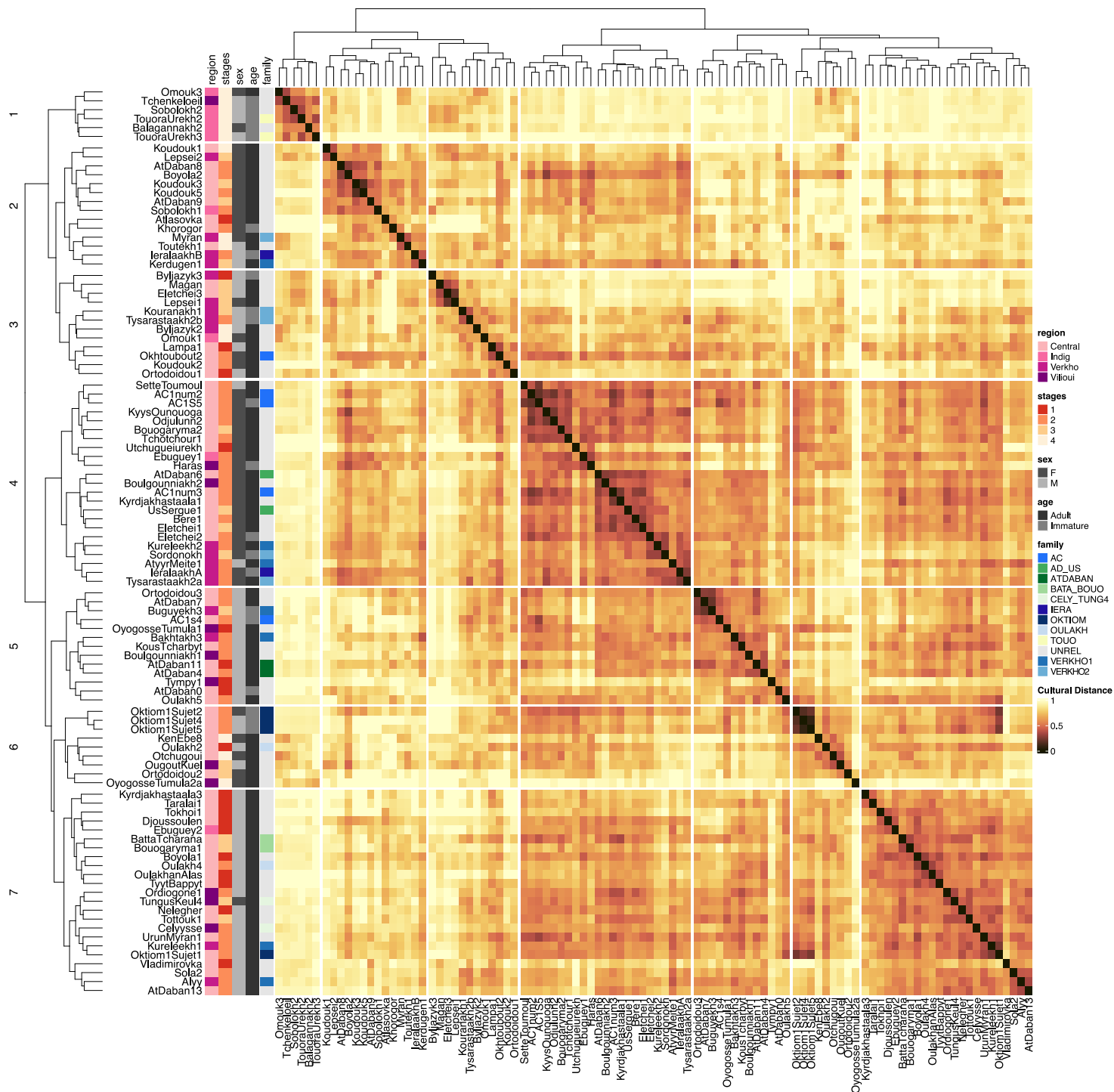


Figure S2_28. Heatmap of the Bray-Curtis cultural distances between pairs of Yakuts from ~1500 to 1900 CE, with other parameters highlighted.

The tree is the result of a Ward agglomeration method on the square root of the Bray-Curtis distances.

References

1. Crubézy, E., and Alexeev, A. (2007). Chamane, Kyys jeune fille des glaces [Shaman, Kyys young girl of the ice] (Errance).
2. Duday, H., Courtaud, P., Crubezy, É., Sellier, P., and Tillier, A.-M. (1990). L'Anthropologie « de terrain » : reconnaissance et interprétation des gestes funéraires. *Bull Mem Soc Anthropol Paris* 2, 29–49. <https://doi.org/10.3406/bmsap.1990.1740>.
3. Duchesne, S., Bravina, R., Popov, V., Kolodeznikov, S., Gérard, P., Myglan, V., Hochstrasser-Petit, Ch., Romanova, L., Petit, M., Kirianov, N., et al. (2020). Frozen graves of Yakutia, a chronological sequence. *Vestnik arkheologii, antropologii i etnografii* 4, 120–130. <https://doi.org/10.20874/2071-0437-2020-51-4-11>.
4. Brůžek, J. (1991). Fiabilité des procédés de détermination du sexe à partir de l'os coxal : implications à l'étude du dimorphisme sexuel de l'homme fossile. *Bull Mem Soc Anthropol Paris* 4, 67-104.
5. Brůžek, J. (2002). A method for visual determination of sex, using the human hip bone. *Am J Phys Anthropol* 117, 157–168. <https://doi.org/10.1002/ajpa.10012>.
6. Brůžek, J., Santos, F., Dutailly, B., Murail, P., and Cunha, E. (2017). Validation and reliability of the sex estimation of the human os coxae using freely available DSP2 software for bioarchaeology and forensic anthropology. *Am J Phys Anthropol* 164, 440–449. <https://doi.org/10.1002/ajpa.23282>.
7. Murail, P., Brůžek, J., Houët, F., and Cunha, E. (2005). DSP: A tool for probabilistic sex diagnosis using worldwide variability in hip-bone measurements. *Bull Mem Soc Anthropol Paris* 17, 167–176. <https://doi.org/10.4000/bmsap.1157>
8. Amory, S., Crubézy, E., Keyser, C., Alekseev, A.N., and Ludes, B. (2006). Early Influence of the Steppe Tribes in the Peopling of Siberia. *Hum Biol* 78, 531–549. <https://doi.org/10.1353/HUB.2007.0001>.
9. Librado, P., Tressières, G., Chauvey, L., Fages, A., Khan, N., Schiavinato, S., Calvière-Tonasso, L., Kusliy, M.A., Gaunitz, C., Liu, X., et al. (2024). Widespread horse-based mobility arose around 2200 bce in Eurasia. *Nature* 2024 631:8022 631, 819–825. <https://doi.org/10.1038/s41586-024-07597-5>.
10. Reimer, P.J., Austin, W.E.N., Bard, E., Bayliss, A., Blackwell, P.G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R.L., Friedrich, M., et al. (2020). The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP). *Radiocarbon* 62, 725–757. <https://doi.org/10.1017/RDC.2020.41>.
11. Damgaard, P.B., Margaryan, A., Schroeder, H., Orlando, L., Willerslev, E., and Allentoft, M.E. (2015). Improving access to endogenous DNA in ancient bones and teeth. *Sci Rep* 5, 11184. <https://doi.org/10.1038/srep11184>.
12. Gamba, C., Jones, E.R., Teasdale, M.D., McLaughlin, R.L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kovári, I., Pap, I., Anders, A., et al. (2014). Genome flux and stasis in a five millennium transect of European prehistory. *Nature Communications* 2014 5:1 5, 1–9. <https://doi.org/10.1038/ncomms6257>.

- 546 13. Yang, D.Y., Eng, B., Wayne, J.S., Duda, J.C., and Saunders, S.R. (1998). Improved DNA extraction
547 from ancient bones using silica-based spin columns. *Am J Phys Anthropol* 105, 539–543.
548 [https://doi.org/10.1002/\(SICI\)1096-8644\(199804\)105:4<539::AID-](https://doi.org/10.1002/(SICI)1096-8644(199804)105:4<539::AID-AJPA10>3.0.CO;2-1)
549 [AJPA10>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1096-8644(199804)105:4<539::AID-AJPA10>3.0.CO;2-1).
- 550 14. Sabin, S., and Yates, J.A.F. (2020). Dental Calculus Field-Sampling Protocol (Sabin version) v2.
551 [protocols.io. https://doi.org/10.17504/PROTOCOLS.IO.BQECMTAW](https://doi.org/10.17504/PROTOCOLS.IO.BQECMTAW).
- 552 15. Gamba, C., Hanghoj, K., Gaunitz, C., Alfarhan, A.H., Alquraishi, S.A., Al-Rasheid, K.A.S.,
553 Bradley, D.G., and Orlando, L. (2016). Comparing the performance of three ancient DNA
554 extraction methods for high-throughput sequencing. *Mol Ecol Resour* 16, 459–469.
555 <https://doi.org/10.1111/1755-0998.12470>.
- 556 16. Meyer, M., and Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed
557 target capture and sequencing. *Cold Spring Harb Protoc* 5.
558 <https://doi.org/10.1101/PDB.PROT5448>.
- 559 17. Fages, A., Hanghoj, K., Khan, N., Gaunitz, C., Seguin-Orlando, A., Leonardi, M., McCrory
560 Constantz, C., Gamba, C., Al-Rasheid, K.A.S., Albizuri, S., et al. (2019). Tracking Five Millennia of
561 Horse Management with Extensive Ancient Genome Time Series. *Cell* 177, 1419–1435.e31.
562 <https://doi.org/10.1016/j.cell.2019.03.049>.
- 563 18. Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., and Reich, D. (2015). Partial uracil–DNA–
564 glycosylase treatment for screening of ancient DNA. *Philosophical Transactions of the Royal*
565 *Society B: Biological Sciences* 370, 20130624. <https://doi.org/10.1098/rstb.2013.0624>.
- 566 19. Schubert, M., Ermini, L., Sarkissian, C. Der, Jónsson, H., Ginolhac, A., Schaefer, R., Martin, M.D.,
567 Fernández, R., Kircher, M., McCue, M., et al. (2014). Characterization of ancient and modern
568 genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX.
569 *Nature Protocols* 2014 9:5 9, 1056–1082. <https://doi.org/10.1038/nprot.2014.063>.
- 570 20. Schubert, M., Lindgreen, S., and Orlando, L. (2016). AdapterRemoval v2: rapid adapter trimming,
571 identification, and read merging. *BMC Res Notes* 9, 88. [https://doi.org/10.1186/s13104-016-](https://doi.org/10.1186/s13104-016-1900-2)
572 [1900-2](https://doi.org/10.1186/s13104-016-1900-2).
- 573 21. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature*
574 *Methods* 2012 9:4 9, 357–359. <https://doi.org/10.1038/nmeth.1923>.
- 575 22. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K.,
576 Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: A MapReduce
577 framework for analyzing next-generation DNA sequencing data. *Genome Res* 20, 1297–1303.
578 <https://doi.org/10.1101/gr.107524.110>.
- 579 23. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation
580 Sequencing Data. *BMC Bioinformatics* 15, 356. <https://doi.org/10.1186/s12859-014-0356-4>.
- 581 24. Mallick, S., and Reich, D. (2023). The Allen Ancient DNA Resource (AADR): A curated
582 compendium of ancient human genomes, V5. Preprint at Harvard Dataverse,
583 <https://doi.org/doi/10.7910/DVN/FFIDCW> <https://doi.org/doi/10.7910/DVN/FFIDCW>.
- 584 25. Mallick, S., Micco, A., Mah, M., Ringbauer, H., Lazaridis, I., Olalde, I., Patterson, N., and Reich, D.
585 (2024). The Allen Ancient DNA Resource (AADR) a curated compendium of ancient human
586 genomes. *Scientific Data* 2024 11:1 11, 1–10. <https://doi.org/10.1038/s41597-024-03031-7>.

26. Seguin-Orlando, A., Gamba, C., Sarkissian, C. Der, Ermini, L., Louvel, G., Boulygina, E., Sokolov, A., Nedoluzhko, A., Lorenzen, E.D., Lopez, P., et al. (2015). Pros and cons of methylation-based enrichment methods for ancient DNA. *Sci Rep* 5. <https://doi.org/10.1038/SREP11826>.
27. Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., and Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29, 1682–1684. <https://doi.org/10.1093/bioinformatics/btt193>.
28. Briggs, A.W., Stenzel, U., Johnson, P.L.F., Green, R.E., Kelso, J., Prufer, K., Meyer, M., Krause, J., Ronan, M.T., Lachmann, M., et al. (2007). Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences* 104, 14616–14621. <https://doi.org/10.1073/pnas.0704665104>.
29. Briggs, A.W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., and Pääbo, S. (2009). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Res* 38, e87. <https://doi.org/10.1093/NAR/GKP1163>.
30. Skoglund, P., Northoff, B.H., Shunkov, M. V., Derevianko, A.P., Pääbo, S., Krause, J., and Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proceedings of the National Academy of Sciences* 111. <https://doi.org/10.1073/pnas.1318934111>.
31. Orlando, L., Allaby, R., Skoglund, P., Der Sarkissian, C., Stockhammer, P.W., Ávila-Arcos, M.C., Fu, Q., Krause, J., Willerslev, E., Stone, A.C., et al. (2021). Ancient DNA analysis. *Nature Reviews Methods Primers* 2021 1:1 1, 1–26. <https://doi.org/10.1038/s43586-020-00011-0>.
32. Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., Cappellini, E., Petersen, B., Moltke, I., et al. (2013). Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499, 74–78. <https://doi.org/10.1038/nature12323>.
33. Meyer, M., Kircher, M., Gansauge, M.-T., Li, H., Racimo, F., Mallick, S., Schraiber, J.G., Jay, F., Prufer, K., de Filippo, C., et al. (2012). A High-Coverage Genome Sequence from an Archaic Denisovan Individual. *Science* (1979) 338, 222–226. <https://doi.org/10.1126/science.1224344>.
34. Kloss-Brandstätter, A., Pacher, D., Schönherr, S., Weissensteiner, H., Binna, R., Specht, G., and Kronenberg, F. (2011). HaploGrep: A fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. *Hum Mutat* 32, 25–32. <https://doi.org/10.1002/humu.21382>.
35. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., and Davies, R.M. (2021). Twelve years of SAMtools and BCFtools. *Gigascience* 10, 1–4. <https://doi.org/10.1093/GIGASCIENCE/GIAB008>.
36. Renaud, G., Slon, V., Duggan, A.T., and Kelso, J. (2015). Schmutzi: Estimation of contamination and endogenous mitochondrial consensus calling for ancient DNA. *Genome Biol* 16, 1–18. <https://doi.org/10.1186/S13059-015-0776-0/TABLES/5>.
37. Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K.E., Rasmussen, S., Albrechtsen, A., Skotte, L., Lindgreen, S., Metspalu, M., Jombart, T., et al. (2011). An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science* (1979) 334, 94–98. <https://doi.org/10.1126/science.1211177>.

38. Ralf, A., Montiel González, D., Zhong, K., and Kayser, M. (2018). Yleaf: Software for Human Y-Chromosomal Haplogroup Inference from Next-Generation Sequencing Data. *Mol Biol Evol* 35, 1291–1294. <https://doi.org/10.1093/molbev/msy032>.
39. Sehnert, A.J., Rhees, B., Comstock, D., De Feo, E., Heilek, G., Burke, J., and Rava, R.P. (2011). Optimal Detection of Fetal Chromosomal Abnormalities by Massively Parallel DNA Sequencing of Cell-Free Fetal DNA from Maternal Blood. *Clin Chem* 57, 1042–1049. <https://doi.org/10.1373/CLINCHEM.2011.165910>.
40. Rubinacci, S., Ribeiro, D.M., Hofmeister, R.J., and Delaneau, O. (2021). Efficient phasing and imputation of low-coverage sequencing data using large reference panels. *Nat Genet* 53, 120–126. <https://doi.org/10.1038/s41588-020-00756-0>.
41. The 1000 Genomes Project Consortium, Auton, A., Abecasis, G.R., Altshuler, D.M., Durbin, R.M., Bentley, D.R., Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. <https://doi.org/10.1038/nature15393>.
42. Lawson, D.J., Hellenthal, G., Myers, S., and Falush, D. (2012). Inference of Population Structure using Dense Haplotype Data. *PLoS Genet* 8, e1002453. <https://doi.org/10.1371/journal.pgen.1002453>.
43. Bergström, A., McCarthy, S.A., Hui, R., Almarri, M.A., Ayub, Q., Danecek, P., Chen, Y., Felkel, S., Hallast, P., Kamm, J., et al. (2020). Insights into human genetic variation and population history from 929 diverse genomes. *Science* (1979) 367. <https://doi.org/10.1126/science.aay5012>.
44. Kuhn, J.M.M., Jakobsson, M., and Günther, T. (2018). Estimating genetic kin relationships in prehistoric populations. *PLoS One* 13. <https://doi.org/10.1371/journal.pone.0195491>.
45. Alaçamlı, E., Naidoo, T., Aktürk, Ş., Güler, M.N., Mapelli, I., Vural, K.B., Somel, M., Malmström, H., and Günther, T. (2024). READv2: Advanced and user-friendly detection of biological relatedness in archaeogenomics. *bioRxiv*, 2024.01.23.576660. <https://doi.org/10.1101/2024.01.23.576660>.
46. Lipatov, M., Sanjeev, K., Patro, R., and Veeramah, K.R. (2015). Maximum Likelihood Estimation of Biological Relatedness from Low Coverage Sequencing Data. *bioRxiv*, 023374. <https://doi.org/10.1101/023374>.
47. Fernandes, D.M., Cheronet, O., Gelabert, P., and Pinhasi, R. (2021). TKGWV2: an ancient DNA relatedness pipeline for ultra-low coverage whole genome shotgun data. *Sci Rep* 11. <https://doi.org/10.1038/s41598-021-00581-3>.
48. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. <https://doi.org/10.1093/BIOINFORMATICS/BTR330>.
49. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* 81, 559–575. <https://doi.org/10.1086/519795>.
50. Ringbauer, H., Huang, Y., Akbari, A., Mallick, S., Olalde, I., Patterson, N., and Reich, D. (2023). Accurate detection of identity-by-descent segments in human ancient DNA. *Nature Genetics* 2023 56:1 56, 143–151. <https://doi.org/10.1038/s41588-023-01582-w>.

51. Cassidy, L.M., Maoldúin, R., Kador, T., Lynch, A., Jones, C., Woodman, P.C., Murphy, E., Ramsey, G., Dowd, M., Noonan, A., et al. (2020). A dynastic elite in monumental Neolithic society. *Nature* 582, 384–388. <https://doi.org/10.1038/s41586-020-2378-6>.
52. Ringbauer, H., Novembre, J., and Steinrücken, M. (2021). Parental relatedness through time revealed by runs of homozygosity in ancient DNA. *Nat Commun* 12. <https://doi.org/10.1038/s41467-021-25289-w>.
53. Zhang, F., Ning, C., Scott, A., Fu, Q., Bjørn, R., Li, W., Wei, D., Wang, W., Fan, L., Abuduresule, I., et al. (2021). The genomic origins of the Bronze Age Tarim Basin mummies. *Nature*, 0–3. <https://doi.org/10.1038/s41586-021-04052-7>.
54. Patterson, N., Price, A.L., and Reich, D. (2006). Population Structure and Eigenanalysis. *PLoS Genet* 2, e190. <https://doi.org/10.1371/journal.pgen.0020190>.
55. Allentoft, M.E., Sikora, M., Sjögren, K.G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., et al. (2015). Population genomics of Bronze Age Eurasia. *Nature* 522, 167–172. <https://doi.org/10.1038/nature14507>.
56. Broushaki, F., Thomas, M.G., Link, V., López, S., van Dorp, L., Kirsanow, K., Hofmanová, Z., Diekmann, Y., Cassidy, L.M., Díez-del-Molino, D., et al. (2016). Early Neolithic genomes from the eastern Fertile Crescent. *Science* (1979) 353, 499–503. <https://doi.org/10.1126/science.aaf7943>.
57. de Barros Damgaard, P., Martiniano, R., Kamm, J., Moreno-Mayar, J.V., Kroonen, G., Peyrot, M., Barjamovic, G., Rasmussen, S., Zacho, C., Baimukhanov, N., et al. (2018). The first horse herders and the impact of early Bronze Age steppe expansions into Asia. *Science* (1979) 360, eaar7711. <https://doi.org/10.1126/science.aar7711>.
58. Damgaard, P. de B., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliussen, T., Moreno-Mayar, J.V., Pedersen, M.W., Goldberg, A., Usmanova, E., et al. (2018). 137 ancient human genomes from across the Eurasian steppes. *Nature* 557, 369–374. <https://doi.org/10.1038/s41586-018-0094-2>.
59. Feldman, M., Fernández-Domínguez, E., Reynolds, L., Baird, D., Pearson, J., HersHKovitz, I., May, H., Goring-Morris, N., Benz, M., Gresky, J., et al. (2019). Late Pleistocene human genome suggests a local origin for the first farmers of central Anatolia. *Nat Commun* 10, 1218. <https://doi.org/10.1038/s41467-019-09209-7>.
60. Flegontov, P., Altınışık, N.E., Changmai, P., Rohland, N., Mallick, S., Adamski, N., Bolnick, D.A., Broomandkhoshbacht, N., Candilio, F., Culleton, B.J., et al. (2019). Palaeo-Eskimo genetic ancestry and the peopling of Chukotka and North America. *Nature* 570, 236–240. <https://doi.org/10.1038/s41586-019-1251-y>.
61. Fu, Q., Hajdinjak, M., Moldovan, O.T., Constantin, S., Mallick, S., Skoglund, P., Patterson, N., Rohland, N., Lazaridis, I., Nickel, B., et al. (2015). An early modern human from Romania with a recent Neanderthal ancestor. *Nature* 524, 216–219. <https://doi.org/10.1038/NATURE14558>.
62. Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., Furtwängler, A., Haak, W., Meyer, M., Mittnik, A., et al. (2016). The genetic history of Ice Age Europe. *Nature* 534, 200–205. <https://doi.org/10.1038/nature17993>.
63. Hofmanová, Z., Kreutzer, S., Hellenthal, G., Sell, C., Diekmann, Y., Díez-Del-Molino, D., Van Dorp, L., López, S., Kousathanas, A., Link, V., et al. (2016). Early farmers from across Europe

- directly descended from Neolithic Aegeans. *Proc Natl Acad Sci U S A* *113*, 6886–6891.
https://doi.org/10.1073/PNAS.1523951113/SUPPL_FILE/PNAS.1523951113.SD03.XLSX.
64. Günther, T., Malmström, H., Svensson, E.M., Omrak, A., Sánchez-Quinto, F., Kılınç, G.M., Krzewińska, M., Eriksson, G., Fraser, M., Edlund, H., et al. (2018). Population genomics of Mesolithic Scandinavia: Investigating early postglacial migration routes and high-latitude adaptation. *PLoS Biol* *16*, e2003703. <https://doi.org/10.1371/JOURNAL.PBIO.2003703>.
 65. Järve, M., Saag, L., Scheib, C.L., Pathak, A.K., Montinaro, F., Pagani, L., Flores, R., Guellil, M., Saag, L., Tambets, K., et al. (2019). Shifts in the Genetic Landscape of the Western Eurasian Steppe Associated with the Beginning and End of the Scythian Dominance. *Current Biology* *29*, 2430–2441. <https://doi.org/10.1016/j.cub.2019.06.019>.
 66. Jeong, C., Ozga, A.T., Witonsky, D.B., Malmström, H., Edlund, H., Hofman, C.A., Hagan, R.W., Jakobsson, M., Lewis, C.M., Aldenderfer, M.S., et al. (2016). Long-term genetic stability and a high-altitude East Asian origin for the peoples of the high valleys of the Himalayan arc. *Proceedings of the National Academy of Sciences* *113*, 7485–7490.
<https://doi.org/10.1073/pnas.1520844113>.
 67. Jeong, C., Wang, K., Wilkin, S., Taylor, W.T.T., Miller, B.K., Bemmman, J.H., Stahl, R., Chiovelli, C., Knolle, F., Ulzibayar, S., et al. (2020). A Dynamic 6,000-Year Genetic History of Eurasia's Eastern Steppe. *Cell* *183*, 890–904.e29. <https://doi.org/10.1016/j.cell.2020.10.015>.
 68. Jones, E.R., Gonzalez-Fortes, G., Connell, S., Siska, V., Eriksson, A., Martiniano, R., McLaughlin, R.L., Gallego Llorente, M., Cassidy, L.M., Gamba, C., et al. (2015). Upper Palaeolithic genomes reveal deep roots of modern Eurasians. *Nat Commun* *6*, 1–8.
<https://doi.org/10.1038/ncomms9912>.
 69. Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., et al. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* *528*, 499–503. <https://doi.org/10.1038/nature16152>.
 70. Pinhasi, R., Zarina, G., Moiseyev, V., Lightfoot, E., Nigst, P.R., Manica, A., Pinhasi, R., and Bradley, D.G. (2017). The Neolithic Transition in the Baltic Was Not Driven by Admixture with Early European Farmers. *Current Biology* *27*, 576–582.
<https://doi.org/10.1016/j.cub.2016.12.060>.
 71. Kılınç, G.M., Omrak, A., Özer, F., Günther, T., Büyükkarakaya, A.M., Bıçakçı, E., Baird, D., Dönertaş, H.M., Ghalichi, A., Yaka, R., et al. (2016). The Demographic Development of the First Farmers in Anatolia. *Current Biology* *26*, 2659–2666. <https://doi.org/10.1016/j.cub.2016.07.057>.
 72. Kılınç, G.M., Kashuba, N., Koptekin, D., Bergfeldt, N., Dönertaş, H.M., Rodríguez-Varela, R., Shergin, D., Ivanov, G., Kichigin, D., Pestereva, K., et al. (2021). Human population dynamics and *Yersinia pestis* in ancient northeast Asia. *Sci Adv* *7*, 1–14. <https://doi.org/10.1126/sciadv.abc4587>.
 73. Lazaridis, I.P., Nadel, D., Rollefson, G., Merrett, D.C., Rohland, N., Mallick, S., Fernandes, D., Novak, M., Gamarra, B., Sirak, K., et al. (2016). Genomic insights into the origin of farming in the ancient Near East. *Nature* *536*, 419–424. <https://doi.org/10.1038/nature19310>.
 74. Olalde, I., Brace, S., Allentoft, M.E., Armit, I., Kristiansen, K., Booth, T., Rohland, N., Mallick, S., Szécsényi-Nagy, A., Mittnik, A., et al. (2018). The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* *555*, 190–196. <https://doi.org/10.1038/nature25738>.

- 750 75. Lazaridis, I., Patterson, N., Mittnik, A., Renaud, G., Mallick, S., Kirsanow, K., Sudmant, P.H.,
751 Schraiber, J.G., Castellano, S., Lipson, M., et al. (2013). Ancient human genomes suggest three
752 ancestral populations for present-day Europeans. *Nature* 513, 409–413.
753 <https://doi.org/10.1038/nature13673>.
- 754 76. Lipson, M., Szécsényi-Nagy, A., Mallick, S., Pósa, A., Stégmár, B., Keerl, V., Rohland, N.,
755 Stewardson, K., Ferry, M., Michel, M., et al. (2017). Parallel palaeogenomic transects reveal
756 complex genetic history of early European farmers. *Nature* 551, 368–372.
757 <https://doi.org/10.1038/nature24476>.
- 758 77. Mao, X., Zhang, H., Qiao, S., Liu, Y., Chang, F., Xie, P., Zhang, M., Wang, T., Li, M., Cao, P., et
759 al. (2021). The deep population history of northern East Asia from the Late Pleistocene to the
760 Holocene. *Cell* 184, 3256–3266.e13. <https://doi.org/10.1016/j.cell.2021.04.040>.
- 761 78. Wang, C.-C., Reinhold, S., Kalmykov, A., Wissgott, A., Brandt, G., Jeong, C., Cheronet, O., Ferry,
762 M., Harney, E., Keating, D., et al. (2019). Ancient human genome-wide data from a 3000-year
763 interval in the Caucasus corresponds with eco-geographic regions. *Nat Commun* 10, 590.
764 <https://doi.org/10.1038/s41467-018-08220-8>.
- 765 79. Mathieson, I., Alpaslan-Roodenberg, S., Posth, C., Szécsényi-Nagy, A., Rohland, N., Mallick, S.,
766 Olalde, I., Broomandkhoshbacht, N., Candilio, F., Cheronet, O., et al. (2018). The genomic history
767 of southeastern Europe. *Nature* 555, 197–203. <https://doi.org/10.1038/nature25778>.
- 768 80. McColl, H., Racimo, F., Vinner, L., Demeter, F., Gakuhari, T., Moreno-Mayar, J.V., van Driem,
769 G., Gram Wilken, U., Seguin-Orlando, A., de la Fuente Castro, C., et al. (2018). The prehistoric
770 peopling of Southeast Asia. *Science* (1979) 361, 88–92. <https://doi.org/10.1126/science.aat3628>.
- 771 81. Mittnik, A., Wang, C.-C., Pfrengle, S., Daubaras, M., Zariņa, G., Hallgren, F., Allmäe, R.,
772 Khartanovich, V., Moiseyev, V., Törv, M., et al. (2018). The genetic prehistory of the Baltic Sea
773 region. *Nat Commun* 9, 442. <https://doi.org/10.1038/s41467-018-02825-9>.
- 774 82. Narasimhan, V.M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., Mallick, S., Lazaridis,
775 I., Nakatsuka, N., Olalde, I., Lipson, M., et al. (2019). The formation of human populations in
776 South and Central Asia. *Science* (1979) 365, eaat7487. <https://doi.org/10.1126/science.aat7487>.
- 777 83. Ning, C., Li, T., Wang, K., Zhang, F., Li, T., Wu, X., Gao, S., Zhang, Q., Zhang, H., Hudson, M.J.,
778 et al. (2020). Ancient genomes from northern China suggest links between subsistence changes
779 and human migration. *Nature Communications* 2020 11:1 11, 1–9.
780 <https://doi.org/10.1038/s41467-020-16557-2>.
- 781 84. Ning, C., Wang, C.C., Gao, S., Yang, Y., Zhang, X., Wu, X., Zhang, F., Nie, Z., Tang, Y.,
782 Robbeets, M., et al. (2019). Ancient Genomes Reveal Yamnaya-Related Ancestry and a Potential
783 Source of Indo-European Speakers in Iron Age Tianshan. *Current Biology* 29, 2526–2532.e4.
784 [https://doi.org/10.1016/J.CUB.2019.06.044/ATTACHMENT/6613EC96-18A0-42FF-81A6-](https://doi.org/10.1016/J.CUB.2019.06.044/ATTACHMENT/6613EC96-18A0-42FF-81A6-A70A7D54DDAF/MMC4.PDF)
785 [A70A7D54DDAF/MMC4.PDF](https://doi.org/10.1016/J.CUB.2019.06.044/ATTACHMENT/6613EC96-18A0-42FF-81A6-A70A7D54DDAF/MMC4.PDF).
- 786 85. Omrak, A., Günther, T., Valdiosera, C., Svensson, E.M., Malmström, H., Kiesewetter, H.,
787 Aylward, W., Storå, J., Jakobsson, M., and Götherström, A. (2016). Genomic Evidence Establishes
788 Anatolia as the Source of the European Neolithic Gene Pool. *Current Biology* 26, 270–275.
789 <https://doi.org/10.1016/j.cub.2015.12.019>.
- 790 86. Skoglund, P., Malmstrom, H., Omrak, A., Raghavan, M., Valdiosera, C., Gunther, T., Hall, P.,
791 Tambets, K., Parik, J., Sjogren, K.-G., et al. (2014). Genomic Diversity and Admixture Differs for

- 792 Stone-Age Scandinavian Foragers and Farmers. *Science* (1979) *344*, 747–750.
793 <https://doi.org/10.1126/science.1253448>.
- 794 87. Raghavan, M., Skoglund, P., Graf, K.E., Metspalu, M., Albrechtsen, A., Moltke, I., Rasmussen, S.,
795 Stafford Jr, T.W., Orlando, L., Metspalu, E., et al. (2014). Upper Palaeolithic Siberian genome
796 reveals dual ancestry of Native Americans. *Nature* *505*, 87–91.
797 <https://doi.org/10.1038/nature12736>.
- 798 88. Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J.S., Albrechtsen, A., Moltke, I., Metspalu, M.,
799 Metspalu, E., Kivisild, T., Gupta, R., et al. (2010). Ancient human genome sequence of an extinct
800 Palaeo-Eskimo. *Nature* *463*, 757–762. <https://doi.org/10.1038/nature08835>.
- 801 89. Sikora, M., Pitulko, V. V., Sousa, V.C., Allentoft, M.E., Vinner, L., Rasmussen, S., Margaryan, A.,
802 de Barros Damgaard, P., de la Fuente, C., Renaud, G., et al. (2019). The population history of
803 northeastern Siberia since the Pleistocene. *Nature* *570*, 182–188. [https://doi.org/10.1038/s41586-](https://doi.org/10.1038/s41586-019-1279-z)
804 [019-1279-z](https://doi.org/10.1038/s41586-019-1279-z).
- 805 90. Sikora, M., Seguin-Orlando, A., Sousa, V.C., Albrechtsen, A., Korneliussen, T., Ko, A.,
806 Rasmussen, S., Dupanloup, I., Nigst, P.R., Bosch, M.D., et al. (2017). Ancient genomes show
807 social and reproductive behavior of early Upper Paleolithic foragers. *Science* (1979) *358*, 659–662.
808 <https://doi.org/10.1126/science.aao1807>.
- 809 91. Rasmussen, M., Anzick, S.L., Waters, M.R., Skoglund, P., Degiorgio, M., Stafford, T.W.,
810 Rasmussen, S., Moltke, I., Albrechtsen, A., Doyle, S.M., et al. (2014). The genome of a Late
811 Pleistocene human from a Clovis burial site in western Montana. *Nature* *2014* *506*:7487 *506*, 225–
812 229. <https://doi.org/10.1038/nature13025>.
- 813 92. Wang, C.-C., Yeh, H.-Y., Popov, A.N., Zhang, H.-Q., Matsumura, H., Sirak, K., Cheronet, O.,
814 Kovalev, A., Rohland, N., Kim, A.M., et al. (2021). Genomic insights into the formation of human
815 populations in East Asia. *Nature* *591*, 413–419. <https://doi.org/10.1038/s41586-021-03336-2>.
- 816 93. Yaka, R., Mapelli, I., Kaptan, D., Doğu, A., Chyleński, M., Erdal, Ö.D., Koptekin, D., Vural, K.B.,
817 Bayliss, A., Mazzucato, C., et al. (2021). Variable kinship patterns in Neolithic Anatolia revealed by
818 ancient genomes. *Current Biology* *31*, 1–14. <https://doi.org/10.1016/j.cub.2021.03.050>.
- 819 94. Yu, H., Spyrou, M.A., Karapetian, M., Shnaider, S., Radzevičiūtė, R., Nägele, K., Neumann, G.U.,
820 Penske, S., Zech, J., Lucas, M., et al. (2020). Paleolithic to Bronze Age Siberians Reveal
821 Connections with First Americans and across Eurasia. *Cell* *181*, 1232–1245.e20.
822 <https://doi.org/10.1016/j.cell.2020.04.037>.
- 823 95. Yang, M.A., Gao, X., Theunert, C., Tong, H., Aximu-Petri, A., Nickel, B., Slatkin, M., Meyer, M.,
824 Pääbo, S., Kelso, J., et al. (2017). 40,000-Year-Old Individual from Asia Provides Insight into Early
825 Population Structure in Eurasia. *Current Biology* *27*, 3202–3208.e9.
826 <https://doi.org/10.1016/j.cub.2017.09.030>.
- 827 96. Yang, M.A., Fan, X., Sun, B., Chen, C., Lang, J., Ko, Y.C., Tsang, C.H., Chiu, H., Wang, T., Bao,
828 Q., et al. (2020). Ancient DNA indicates human population shifts and admixture in northern and
829 southern China. *Science* (1979) *369*, 282–288.
830 https://doi.org/10.1126/SCIENCE.ABA0909/SUPPL_FILE/PAPV2.PDF.
- 831 97. Jeong, C., Balanovsky, O., Lukianova, E., Kahbatkyzy, N., Flegontov, P., Zaporozhchenko, V.,
832 Immel, A., Wang, C.-C., Ixan, O., Khussainova, E., et al. (2019). The genetic history of admixture
833 across inner Eurasia. *Nat Ecol Evol* *3*, 966–976. <https://doi.org/10.1038/s41559-019-0878-2>.

- 834 98. Lamnidis, T.C., Majander, K., Jeong, C., Salmela, E., Wessman, A., Moiseyev, V., Khartanovich,
835 V., Balanovsky, O., Ongyerth, M., Weihmann, A., et al. (2018). Ancient Fennoscandian genomes
836 reveal origin and spread of Siberian ancestry in Europe. *Nat Commun* 9, 5018.
837 <https://doi.org/10.1038/s41467-018-07483-5>.
- 838 99. Mallick, S., Li, H., Lipson, M., Mathieson, I., Gymrek, M., Racimo, F., Zhao, M., Chennagiri, N.,
839 Nordenfelt, S., Tandon, A., et al. (2016). The Simons Genome Diversity Project: 300 genomes
840 from 142 diverse populations. *Nature* 538, 201–206. <https://doi.org/10.1038/nature18964>.
- 841 100. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster,
842 T., and Reich, D. (2012). Ancient Admixture in Human History. *Genetics* 192, 1065–1093.
843 <https://doi.org/10.1534/genetics.112.145037>.
- 844 101. Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G.,
845 Sudmant, P.H., de Filippo, C., et al. (2014). The complete genome sequence of a Neanderthal from
846 the Altai Mountains. *Nature* 505, 43–49. <https://doi.org/10.1038/nature12886>.
- 847 102. Raghavan, M., Steinrücken, M., Harris, K., Schiffels, S., Rasmussen, S., DeGiorgio, M.,
848 Albrechtsen, A., Valdiosera, C., Ávila-Arcos, M.C., Malaspinas, A.S., et al. (2015). Genomic
849 evidence for the Pleistocene and recent population history of Native Americans. *Science* (1979)
850 349. https://doi.org/10.1126/SCIENCE.AAB3884/SUPPL_FILE/RAGHAVAN-SM.PDF.
- 851 103. Skoglund, P., Posth, C., Sirak, K., Spriggs, M., Valentin, F., Bedford, S., Clark, G.R., Reepmeyer,
852 C., Petchey, F., Fernandes, D., et al. (2016). Genomic insights into the peopling of the Southwest
853 Pacific. *Nature* 2016 538:7626 538, 510–513. <https://doi.org/10.1038/nature19844>.
- 854 104. Skoglund, P., Mallick, S., Bortolini, M.C., Chennagiri, N., Hünemeier, T., Petzl-Erler, M.L.,
855 Salzano, F.M., Patterson, N., and Reich, D. (2015). Genetic evidence for two founding populations
856 of the Americas. *Nature* 2015 525:7567 525, 104–108. <https://doi.org/10.1038/nature14895>.
- 857 105. Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in
858 unrelated individuals. *Genome Res* 19, 1655–1664. <https://doi.org/10.1101/gr.094052.109>.
- 859 106. Harney, É., Patterson, N., Reich, D., and Wakeley, J. (2021). Assessing the performance of
860 qpAdm: a statistical tool for studying population admixture. *Genetics* 217.
861 <https://doi.org/10.1093/genetics/iyaa045>.
- 862 107. Flegontova, O., Işıldak, U., Yüncü, E., Williams, M.P., Huber, C.D., Kočí, J., Vyazov, L.A.,
863 Changmai, P., and Flegontov, P. (2025). Performance of qpAdm-based screens for genetic
864 admixture on graph-shaped histories and stepping stone landscapes. *Genetics* 230, 47.
865 <https://doi.org/10.1093/GENETICS/IYAF047>.
- 866 108. Chintalapati, M., Patterson, N., and Moorjani, P. (2022). The spatiotemporal patterns of major
867 human admixture events during the European Holocene. *Elife* 11.
868 <https://doi.org/10.7554/ELIFE.77625>.
- 869 109. Tournebize, R., Chu, G., and Moorjani, P. (2022). Reconstructing the history of founder events
870 using genome-wide patterns of allele sharing across individuals. *PLoS Genet* 18, e1010243.
871 <https://doi.org/10.1371/JOURNAL.PGEN.1010243>.
- 872 110. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler
873 transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.

- 874 111. Blanco-Míguez, A., Beghini, F., Cumbo, F., McIver, L.J., Thompson, K.N., Zolfo, M., Manghi, P.,
875 Dubois, L., Huang, K.D., Thomas, A.M., et al. (2023). Extending and improving metagenomic
876 taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nature Biotechnology* 2023
877 41:11 41, 1633–1644. <https://doi.org/10.1038/s41587-023-01688-w>.
- 878 112. Knights, D., Kuczynski, J., Charlson, E.S., Zaneveld, J., Mozer, M.C., Collman, R.G., Bushman,
879 F.D., Knight, R., and Kelley, S.T. (2011). Bayesian community-wide culture-independent microbial
880 source tracking. *Nature Methods* 2011 8:9 8, 761–763. <https://doi.org/10.1038/nmeth.1650>.
- 881 113. Yates, J.A.F., Velsko, I.M., Aron, F., Posth, C., and Hofman, C.A. (2021). The evolution and
882 changing ecology of the African hominid oral microbiome. *PNAS* 118, 1–11.
883 <https://doi.org/10.1073/pnas.2021655118>.
- 884 114. Quagliariello, A., Modi, A., Innocenti, G., Zaro, V., Conati Barbaro, C., Ronchitelli, A., Boschin,
885 F., Cavazzuti, C., Dellù, E., Radina, F., et al. (2022). Ancient oral microbiomes support gradual
886 Neolithic dietary shifts towards agriculture. *Nature Communications* 2022 13:1 13, 1–14.
887 <https://doi.org/10.1038/s41467-022-34416-0>.
- 888 115. McMurdie, P.J., and Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive
889 Analysis and Graphics of Microbiome Census Data. *PLoS One* 8, e61217.
890 <https://doi.org/10.1371/JOURNAL.PONE.0061217>.
- 891 116. Peschel, S., Müller, C.L., Von Mutius, E., Boulesteix, A.L., and Depner, M. (2021). NetCoMi:
892 network construction and comparison for microbiome data in R. *Brief Bioinform* 22, 1–18.
893 <https://doi.org/10.1093/BIB/BBAA290>.
- 894 117. Truong, D.T., Tett, A., Pasolli, E., Huttenhower, C., and Segata, N. (2017). Microbial strain-level
895 population structure and genetic diversity from metagenomes. *Genome Res* 27, 626–638.
896 <https://doi.org/10.1101/GR.216242.116>.
- 897 118. Yates, J.A.F., Andrades Valtueña, A., Vågene, Å.J., Cribdon, B., Velsko, I.M., Borry, M., Bravo-
898 Lopez, M.J., Fernandez-Guerra, A., Green, E.J., Ramachandran, S.L., et al. (2021). Community-
899 curated and standardised metadata of published ancient metagenomic samples with
900 AncientMetagenomeDir. *Sci Data* 8, 1–8. <https://doi.org/10.1038/s41597-021-00816-y>.
- 901 119. Weyrich, L.S., Duchene, S., Soubrier, J., Arriola, L., Llamas, B., Breen, J., Morris, A.G., Alt, K.W.,
902 Caramelli, D., Dresely, V., et al. (2017). Neanderthal behaviour, diet, and disease inferred from
903 ancient DNA in dental calculus. *Nature* 2017 544:7650 544, 357–361.
904 <https://doi.org/10.1038/nature21674>.
- 905 120. Eisenhofer, R., Kanzawa-Kiriyama, H., Shinoda, K.I., and Weyrich, L.S. (2020). Investigating the
906 demographic history of Japan using ancient oral microbiota. *Philosophical Transactions of the*
907 *Royal Society B* 375. <https://doi.org/10.1098/RSTB.2019.0578>.
- 908 121. Kazarina, A., Petersone-gordina, E., Kimsis, J., Kuzmicka, J., Zayakin, P., Griškjans, Ž., Gerhards,
909 G., and Ranka, R. (2021). The Postmedieval Latvian Oral Microbiome in the Context of Modern
910 Dental Calculus and Modern Dental Plaque Microbial Profiles. *Genes* 2021, Vol. 12, Page 309 12,
911 309. <https://doi.org/10.3390/GENES12020309>.
- 912 122. Velsko, I.M., Semerau, L., Inskip, S.A., García-Collado, M.I., Ziesemer, K., Ruber, M.S., Benítez
913 De Lugo Enrich, L., Molero García, J.M., Valle, D.G., Peña Ruiz, A.C., et al. (2022). Ancient
914 dental calculus preserves signatures of biofilm succession and interindividual variation independent
915 of dental pathology. *PNAS Nexus* 1, 1–14. <https://doi.org/10.1093/PNASNEXUS/PGAC148>.

916 123. Velsko, I.M., Fellows Yates, J.A., Aron, F., Hagan, R.W., Frantz, L.A.F., Loe, L., Martinez, J.B.R.,
917 Chaves, E., Gosden, C., Larson, G., et al. (2019). Microbial differences between dental plaque and
918 historic dental calculus are related to oral biofilm maturation stage. *Microbiome* 7, 1–20.
919 <https://doi.org/10.1186/S40168-019-0717-3/FIGURES/9>.

920 124. Nguyen, L.T., Schmidt, H.A., Von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A Fast and
921 Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol*
922 32, 268–274. <https://doi.org/10.1093/MOLBEV/MSU300>.

923 125. Beghini, F., McIver, L.J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., Mailyan, A.,
924 Manghi, P., Scholz, M., Thomas, A.M., et al. (2021). Integrating taxonomic, functional, and strain-
925 level profiling of diverse microbial communities with biobakery 3. *Elife* 10.
926 <https://doi.org/10.7554/ELIFE.65088>.

927 126. Asnicar, F., Thomas, A.M., Beghini, F., Mengoni, C., Manara, S., Manghi, P., Zhu, Q., Bolzan, M.,
928 Cumbo, F., May, U., et al. (2020). Precise phylogenetic analysis of microbial isolates and genomes
929 from metagenomes using PhyloPhlAn 3.0. *Nature Communications* 2020 11:1 11, 1–10.
930 <https://doi.org/10.1038/s41467-020-16366-7>.

931 127. Suzek, B.E., Wang, Y., Huang, H., McGarvey, P.B., and Wu, C.H. (2015). UniRef clusters: a
932 comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics*
933 31, 926–932. <https://doi.org/10.1093/BIOINFORMATICS/BTU739>.

934 128. Franzosa, E.A., McIver, L.J., Rahnavard, G., Thompson, L.R., Schirmer, M., Weingart, G., Lipson,
935 K.S., Knight, R., Caporaso, J.G., Segata, N., et al. (2018). Species-level functional profiling of
936 metagenomes and metatranscriptomes. *Nature Methods* 2018 15:11 15, 962–968.
937 <https://doi.org/10.1038/s41592-018-0176-y>.

938 129. Mühlemann, B., Vinner, L., Margaryan, A., Wilhelmson, H., Castro, C.D.L.F., Allentoft, M.E.,
939 Damgaard, P.D.B., Hansen, A.J., Nielsen, S.H., Strand, L.M., et al. (2020). Diverse variola virus
940 (smallpox) strains were widespread in northern Europe in the Viking Age. *Science* (1979) 369.
941 https://doi.org/10.1126/SCIENCE.AAW8977/SUPPL_FILE/AAW8977_TABLES8.XLSX.

942 130. Duggan, A.T., Perdomo, M.F., Piombino-Mascali, D., Marciniak, S., Poinar, D., Emery, M. V.,
943 Buchmann, J.P., Duchêne, S., Jankauskas, R., Humphreys, M., et al. (2016). 17th Century Variola
944 Virus Reveals the Recent History of Smallpox. *Current Biology* 26, 3407–3412.
945 <https://doi.org/10.1016/j.cub.2016.10.061>.

946 131. Esposito, J.J., Sammons, S.A., Frace, A.M., Osborne, J.D., Glsen-Rasmussen, M., Zhang, M.,
947 Govil, D., Damon, I.K., Kline, R., Laker, M., et al. (2006). Genome sequence diversity and clues to
948 the evolution of variola (smallpox) virus. *Science* (1979) 313, 807–812.
949 https://doi.org/10.1126/SCIENCE.1125134/SUPPL_FILE/ESPOSITO_SOM_REV_TITLE.PDF.

951 132. Afonso, C.L., Tulman, E.R., Lu, Z., Zsak, L., Sandybaev, N.T., Kerembekova, U.Z., Zaitsev, V.L.,
952 Kutish, G.F., and Rock, D.L. (2002). The Genome of Camelpox Virus. *Virology* 295, 1–9.
953 <https://doi.org/10.1006/VIRO.2001.1343>.

954 133. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7:
955 Improvements in performance and usability. *Mol Biol Evol* 30, 772–780.
956 <https://doi.org/10.1093/MOLBEV/MST010>.

- 957 134. Hoang, D.T., Chernomor, O., Von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2:
958 Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* 35, 518–522.
959 <https://doi.org/10.1093/MOLBEV/MSX281>.
- 960 135. Shimodaira, H. (2002). An Approximately Unbiased Test of Phylogenetic Tree Selection. *Syst Biol*
961 51, 492–508. <https://doi.org/10.1080/10635150290069913>.
- 962 136. Hanghøj, K., Renaud, G., Albrechtsen, A., and Orlando, L. (2019). DamMet: Ancient methylome
963 mapping accounting for errors, true variants, and post-mortem DNA damage. *Gigascience* 8, 1–6.
964 <https://doi.org/10.1093/gigascience/giz025>.
- 965 137. Liu, X., Seguin-Orlando, A., Chauvey, L., Tressières, G., Schiavinato, S., Tonasso-Calvière, L.,
966 Aury, J.M., Perdereau, A., Wagner, S., Clavel, P., et al. (2023). DNA methylation-based profiling of
967 horse archaeological remains for age-at-death and castration. *iScience* 26, 106144.
968 <https://doi.org/10.1016/J.ISCI.2023.106144>.
- 969 138. Lokk, K., Modhukur, V., Rajashekar, B., Märtens, K., Mägi, R., Kolde, R., Koltšina, M., Nilsson,
970 T.K., Vilo, J., Salumets, A., et al. (2014). DNA methylome profiling of human tissues identifies
971 global and tissue-specific methylation patterns. *Genome Biol* 15, 1–14.
972 <https://doi.org/10.1186/GB-2014-15-4-R54/TABLES/4>.
- 973 139. Gokhman, D., Lavi, E., Prüfer, K., Fraga, M.F., Riancho, J.A., Kelso, J., Pääbo, S., Meshorer, E.,
974 and Carmel, L. (2014). Reconstructing the DNA methylation maps of the neandertal and the
975 Denisovan. *Science* (1979) 344, 523–527.
976 https://doi.org/10.1126/SCIENCE.1250368/SUPPL_FILE/GOKHMAN.DATAS2.ZIP.
- 977 140. Hanghøj, K., Seguin-Orlando, A., Schubert, M., Madsen, T., Pedersen, J.S., Willerslev, E., and
978 Orlando, L. (2016). Fast, Accurate and Automatic Ancient Nucleosome and Methylation Maps
979 with epiPALEOMIX. *Mol Biol Evol* 33, 3284–3298.
980 <https://doi.org/10.1093/MOLBEV/MSW184>.
- 981 141. Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev* 25,
982 1010–1022. <https://doi.org/10.1101/GAD.2037511>.
- 983 142. Laurent, L., Wong, E., Li, G., Huynh, T., Tsirigos, A., Ong, C.T., Low, H.M., Sung, K.W.K.,
984 Rigoutsos, I., Loring, J., et al. (2010). Dynamic changes in the human methylome during
985 differentiation. *Genome Res* 20, 320–331. <https://doi.org/10.1101/GR.101907.109>.
- 986 143. Kelly, T.K., Liu, Y., Lay, F.D., Liang, G., Berman, B.P., and Jones, P.A. (2012). Genome-wide
987 mapping of nucleosome positioning and DNA methylation within individual DNA molecules.
988 *Genome Res* 22, 2497–2506. <https://doi.org/10.1101/GR.143008.112>.
- 989 144. Fu, Y., Sinha, M., Peterson, C.L., and Weng, Z. (2008). The insulator binding protein CTCF
990 positions 20 nucleosomes around its binding sites across the human genome. *PLoS Genet* 4.
991 <https://doi.org/10.1371/JOURNAL.PGEN.1000138>.
- 992 145. R Core Team (2018). R: A Language and Environment for Statistical Computing. Preprint.
- 993 146. Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B.,
994 Solymos, P., Stevens, M.H.H., Szoecs, E., et al. (2022). vegan: Community Ecology Package.
995 Preprint.
- 996 147. Crubézy, E., and Nikolaeva, D. (2017). Vainqueurs ou vaincus? [Winners or losers?] (Odile Jacob).

- 997 148. Duchesne, S. (2020). Pratiques funéraires, biologie humaine et diffusion culturelle en Iakoutie (16e
998 -19e siècles) [Funerary practices, human biology and cultural diffusion in Yakutia (16th-19th
999 centuries)]. (Doctoral thesis, Université Paul Sabatier Toulouse III)
- 1000 149. Crubézy, E., Gérard, P., Kirianov, N., Popov, V., Nikolaïeva, N., Petit, C., Petit, M., Romanova,
1001 L., Keyser, C., Ludes, B., et al. (2017). The Relationship Between Archaeology, Genetics,
1002 Ethnology and History. In *Clashes of Time*, and L. O. Edited by J.-M. Blaising, J. Driessen, J.-P.
1003 Legendre, ed. (Presses universitaires de Louvain), pp. 121–138.
- 1004