Dynamic changes in genomic and social structures in third millennium BCE central Europe

Luka Papac 1*, Michal Ernée 2, Miroslav Dobeš 2, Michaela Langová 2, Adam B. Rohlach 1,3, Franziska Aron 1, Gunnar U. Neumann 1, Maria A. Spyrou 1, Nadin Rohland 6, Petr Velemínský 5, Martin Kuna 2, Hana Brzobohatá 2, Brendan Culeton 6, David Daněček 2,7, Alžběta Danielisová 3, Miluše Dobisíková 2, Josef Hložek 2,8, Douglas J. Kennett 9, Jana Klementová 7, Michal Kostka 10, Petr Krištof 8, Milan Kuchařík 11, Jana Kuljacevova Hlavová 12, Petr Limberský 7, Drahomíra Malýková 2, Lucia Mattiello 13, Monika Pecinovská 2, Katarina Petriščaková 10, Erika Průchová 14, Petra Stránska 15, Lubor Smejtek 13, Jaroslav Špaček 15, Radka Šumberová 2, Ondřej Švejcar 2, Martin Trefný 16, Miloš Vávra 13, Jan Kolář 17,18, Volker Heyd 19, Johannes Krause 1,20, Ron Pinhasi 21, David Reich 4,22,23,24, Stephan Schiffels 1, Wolfgang Haak 1,25*

Europe’s prehistory oversaw dynamic and complex interactions of diverse societies, hitherto unexplored at detailed regional scales. Studying 271 human genomes dated ~4900 to 1600 BCE from the European heartland, Bohemia, we reveal unprecedented genetic changes and social processes. Major migrations preceded the arrival of “steppe” ancestry, and at ~2800 BCE, three genetically and culturally differentiated groups coexisted. Corded Ware appeared by 2900 BCE, were initially genetically diverse, did not derive steppe ancestry from known Yamnaya, and assimilated females of diverse backgrounds. Both Corded Ware and Bell Beaker groups underwent dynamic changes, involving sharp reductions and complete replacements of Y-chromosomal diversity at ~2600 and ~2400 BCE, respectively, the latter accompanied by increased Neolithic-like ancestry. The Bronze Age saw new social organization emerge amid a ≥40% population turnover.

INTRODUCTION
Archaegenetics has revealed two major population turnovers in Europe within the past 10,000 years (1–5). The first, beginning in the seventh millennium BCE, was associated with expanding Neolithic farming communities from Anatolia (6, 7). European Early Neolithic farmers were initially genetically distinct from preceding hunter-gatherers (HG) and almost indistinguishable from Anatolian farmers (8–10), however incorporated HG ancestry into their gene pools over ensuing millennia (3, 11–13).

The second major turnover occurred in the early third millennium BCE with individuals of the Corded Ware (CW) culture (3, 4, 8). Of note, in what follows, we use the co-occurrence of human skeletal remains and markers of archaeological cultures (e.g., grave goods and body orientation) to denote an association between individuals and an archaeological culture (e.g., “CW individuals”), although this may not reflect a unified social entity. The CW represents a major cultural shift in central, northern, and northeastern Europe, bringing changes in economy, ideology, and mortuary practices (14–22). CW individuals were shown to be genetically distinct from culturally pre-CW people, having ~75% of their ancestry similar to Yamnaya individuals from the Pontic-Caspian steppe (3, 4, 23–27). This Yamnaya-like “steppe” ancestry then spread rapidly throughout Europe, reaching Britain, Ireland, the Iberian Peninsula, the Balearic Islands, Sardinia, and Sicily before the end of the third millennium BCE (5, 28–32).

Despite the importance of the third millennium BCE, our genetic understanding is mainly built upon studies with pan-European sampling strategies, with little emphasis on regional, high-resolution temporal transects (3–5, 8). Consequently, many temporal and geographic sampling gaps remain, resulting in limited knowledge about the processes at the level of the societies and communities and how cultural groups interacted, influenced, and gave rise to one another.

1Department of Archaegenetics, Max Planck Institute for the Science of Human History, 07745 Jena, Germany. 2Institute of Archaeology of the Czech Academy of Sciences, Prague, Letenská 4, 118 01, Czech Republic. 3ARC Centre of Excellence for Mathematical and Statistical Frontiers, School of Mathematical Sciences, The University of Adelaide, Adelaide, SA 5005, Australia. 4Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. 5Department of Anthropology, The National Museum, Prague, Cirkusová 1740, Prague 9, Horní Počernice, CZ 193 00, Czech Republic. 6Institutes of Energy and the Environment, Pennsylvania State University, University Park, PA 16802, USA. 7Central Bohemian Museum in Rostocky u Prahy, Zámek 1, Rostocky, CZ 252 63, Czech Republic. 8Department of Archaeology, Faculty of Philosophy and Arts, University of West Bohemia in Pilsen, Sedlackova 38, Pilsen, CZ 301 00, Czech Republic. 9Department of Anthropology, University of California, Santa Barbara, CA 93106, USA. 10Institute of Archaeology and Museology, Masaryk University, Arne Nováka 1, Brno 60200, Czech Republic. 11Institute of Preservation of Archaeological Heritage of Northwestern Bohemia, Jana Žižky 835, Most, CZ 434 01, Czech Republic. 12Central Bohemian Archaeological Heritage Institute, Nad Olšinyami 3/448, Prague 10, CZ 100 00, Czech Republic. 13Institute of Archaeology, Faculty of Arts, University of South Bohemia in České Budějovice, Braniborská 31A, CZ 370 05, České Budějovice, Czech Republic. 14The Municipal Museum in Čelákovice (formerly), Komenského 1646, Čelákovice, CZ 250 88, Czech Republic (private). 15Friedrich Alexander University Erlangen/Nürnberg, Kochstrasse 4/18, DE 91054 Erlangen, Germany. 16Department of Geography, Institute of Botany of the Academy of Sciences of the Czech Republic, Lidská 25/27, Brno 60200, Czech Republic. 17Institute of Archaeology and Museology, Masaryk University, Arne Nováka 1, Brno 60200, Czech Republic. 18Institute of Archeology, P.O. Box 59, Unioninkatu 38, 00014 University of Helsinki, Helsinki, Finland. 19Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany. 20Institute of Evolutionary Anthropology, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria. 21Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. 22Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA. 23Howard Hughes Medical Institute, Harvard Medical School, Boston, MA 02115, USA. 24School of Biological Sciences, The University of Adelaide, Adelaide, SA 5005, Australia.

*Corresponding author. Email: haak@shh.mpg.de (W.H.); papac@shh.mpg.de (L.P.)
In addition, the use of small sample sizes to represent supra-regional archaeological phenomena, as well as the resulting oversimplified culture-historical interpretations, has drawn criticisms from archaeologists (21, 33–40).

Unresolved questions concern the genetic and geographic origins of CW and Bell Beaker (BB) individuals, their relationship to one another and to Yamnaya individuals, as well as the origin of Early Bronze Age (EBA) Únětice individuals. Although it has been proposed that CW formed from a male-biased westward migration of genetically Yamnaya-like people (23, 41–44), no overlap in Y-chromosomal lineages (with the exception of a few nondiagnostic I2) has been found between the predominantly R1a-carrying CW and mainly R1b-Z2103-carrying Yamnaya males. Steppe ancestry is also present in BB individuals (5); however, they predominantly carry R1b-P312, a Y-lineage not yet found among CW or Yamnaya males. Therefore, despite their sharing of steppe ancestry (3, 4) and substantial chronological overlap (45), it is currently not possible to directly link Yamnaya, CW, and BB groups as paternal genealogical sources for one another, particularly noteworthy in light of steppe ancestry’s suggested male-driven spread (23, 41–43) and the proposed patriarchal/patrilocal social kinship systems of these three societies (46–48).

Crucial to understanding the cultural, social, and genetic transitions in third millennium BCE Europe are densely settled regions that attest to the (co)existence of societies attributed to pre-CW [Baden and Globular Amphora (GAC)], CW, BB, and EBA Únětice. Currently, no such region has been systematically studied from the archaeogenetic perspective. Situated in the heart of Europe and tightly nestled around the important Elbe river, the fertile lowlands of Bohemia, the western part of today’s Czech Republic, witnessed many major supra-regional archaeological phenomena (table S1, Fig. 1, fig. S1, and the Supplementary Materials). Dense agrarian settlement of Bohemia began after ~5400 BCE (49, 50) with the arrival of early Neolithic farmers (Linearbandkeramik-LBK, later Stichbandkeramik-STK, and Lengyel). They were succeeded by manifold societies of the Eneolithic (~4400 to 2200 BCE), associated with more than a dozen archaeological cultural groups including Jordanów, Michaelsberg, Funnelbeaker, Baden, Řivnáč, GAC, Early and Late CW, and BB (table S1) (50). The Eneolithic witnessed important innovations (metallurgy, the wheel, wagon and plough, fortified hillforts, and burial mounds) (51–53) and was succeeded by the globalized EBA Únětice culture, geographically centered around Bohemia.

In addition to material and technological developments, ideological changes, as manifested through mortuary behavior, are also evident (54). Although relatively common during the Funnelbeaker period (~3800 to 3400 BCE, n = ~100 known graves in Bohemia) (55), regular graves almost disappear from the succeeding Baden, Řivnáč, and GAC periods (Middle Eneolithic, ~3500 to 2800 BCE, n = ~20 in Bohemia) (56). Single graves, but now with strict gender differentiation in body position and grave goods, reappeared in abundance with CW from ~2900 BCE (n = ~1500 in Bohemia) (50, 57) and continued with BB (n = ~600 in Bohemia) from ~2500 BCE (58), who developed and maintained important differences from the preceding CW. The EBA Únětice culture (59, 60) continued with single graves (n = ~4000 to 5000 in Bohemia), but now again without gender differentiation in body position.

To better understand these transitions, we analyzed a high-resolution archaeogenetic time transect of 271 (206 newly reported and 65 previously published) individuals (Fig. 1, fig. S1, tables S2 to S4, and the Supplementary Materials) from the northern part of Bohemia. Through dense genetic sampling from geographically and temporally overlapping archaeological cultures, we aim to (i) address whether cultural changes in the Eneolithic and EBA of central Europe were driven by an influx of nonlocals, (ii) characterize the central European genetic diversity immediately prior the appearance of CW, (iii) date when individuals with Yamnaya-like steppe ancestry first appeared in central Europe and understand their genetic origin and social structure, (iv) characterize the nature and extent of biological exchange between the “locals” and “migrants” after the appearance of CW, and (v) identify social transformations linked to genetic and archaeological changes.

RESULTS

General sample overview

We screened 261 prehistoric individuals (table S3) from 37 sites (table S2) for ancient human DNA preservation, of which 219 individuals were enriched for 1,233,013 ancestry informative sites in the human genome (“1240k capture panel”) (8). After enrichment, individuals with fewer than 30,000 sites covered (on 1240k) or signs of contamination were removed (n = 13), resulting in a dataset of 206 newly reported individuals. We combined our dataset with 65 previously published individuals (5, 61, 62) from Bohemia (with >30,000 covered sites on 1240k; table S4) and wider (table S5), thereby extending the total number of published Bohemian Neolithic and pre-CW Eneolithic individuals from 7 to 58 (fig. S2), CW individuals from 7 to 54 (fig. S3), BB individuals from 40 to 64 (fig. S4), and EBA individuals from 11 to 95 (fig. S5). Crucially, we substantially expand the sample size of individuals around the time of CW formation (~3200 to 2600 BCE, from n = 1 to n = 50; Fig. 1B), i.e., the last pre-CW (Baden, Řivnáč, and GAC, from n = 0 to n = 18) and the early CW (from n = 1 to n = 32) individuals, allowing us to directly study the origin of CW in central Europe, the nature of their migration, and social interactions with coexisting pre-CW people. First-degree relatives were excluded from allele frequency–based analyses [f statistics, qpWave, qpAdm, Distribution of Ancestry Tracts of Evolutionary Signals (DATES), and Y chromosome analyses; table S4 and see Materials and Methods]. We also report 140 new radiocarbon dates to aid in finer temporal resolution, allowing us to study the genetic changes between early and late phases of important third millennium BCE cultural groups (e.g., CW, BB, and Únětice; tables S4 and S6).

Bohemia before Corded Ware (pre-CW, before ~2800 BCE)

We first assessed the genome-wide data by projecting the ancient individuals from Bohemia onto the first two axes of a principal components analysis (PCA) constructed from 1141 modern-day West Eurasian individuals (table S7). In the resulting PCA plot (Fig. 2A), all (n = 58) pre-CW individuals from Bohemia plot between Anatolia_Neolithic and Western HG (WHG), in close proximity to published culturally pre-CW individuals from central Europe (3, 8, 11, 13, 25). This suggests an absence of steppe ancestry, which we formally confirmed using qpAdm modeling (table S8), revealing that pre-CW individuals from Bohemia can be largely modeled as two-way mixtures of Anatolia_Neolithic and WHG (Fig. 3A, tables S8 and S9, and fig. S6). The percentage of HG ancestry is positively correlated with time (Spearman’s rank correlation r = 0.39, P < 0.004), showing that the previously reported trend of increasing HG ancestry during the Neolithic also took place in Bohemia (3, 11). We found this HG ancestry increase to be best modeled as a two-stage linear
process (Fig. 3A, table S8, and the Supplementary Materials), with an increase in HG ancestry during the fifth millennium BCE, followed by stasis (nonsignificant slope) thereafter (Fig. 3A and Supplementary Materials).

To gain insight into the process(es) by which HG ancestry was incorporated into the gene pool of pre-CW individuals from Bohemia, we used qpAdm to model each pre-CW cultural group as a three-way mixture of Anatolia_Neolithic, Loschbour, and Körös_HG as well as DATES to estimate the introgression date of incorporated HG ancestry (Fig. 3A and tables S10 and S11). Under a scenario of population continuity with sequential incorporation of HG ancestry, the mean date of introgression, as indicated by the gray intervals in
Fig. 3B (right), for succeeding cultures is expected to become more recent through time. Conversely, under population continuity without incorporation of further HG ancestry, the admixture date should be similar for successive cultural groups who have similar HG proportions.

Our results indicate two cultural transitions for which either of these expectations is not met. First, although Bohemian Jordanów and Funnelbeaker have similar amounts of WHG ancestry \( f^4(Mbuti.DG, WHG; Jordanów, Funnelbeaker) \approx 0 \); \( z \) score, 0.96], the estimated date of WHG introgression for Funnelbeaker is significantly earlier (5079 to 4748 BCE) than for Jordanów (4636 to 4310 BCE) (Fig. 3B and table S11), consistent with Bohemian Funnelbeaker individuals being derived from a different population (whose HG ancestry was incorporated further back in time), which superseded the Jordanów population in Bohemia. This transition between Jordanów and Funnelbeaker is corroborated by three additional observations. First, an \( f^4 \) statistic of the form \( f^4(Mbuti.DG, WHG; Jordanów, Funnelbeaker) \) is positive (\( z \) score, 3.14), revealing significantly greater genetic affinity of Bohemian Funnelbeaker to Funnelbeaker (Baalberge and Salzmünde) individuals from Saxony-Anhalt than to the preceding local Jordanów individuals. Conversely, \( f^4(Mbuti.DG, WHG; Bohemian-Funnelbeaker, Bohemia-Funnelbeaker) \) is consistent with \( 0 \) (\( z \) score, 1.03), suggesting phylogenetic cladality between Bohemian and German Funnelbeaker with respect to Bohemia-Jordanów individuals. Second, Bohemia-Jordanów individuals can be modeled as a two-way mixture of Anatolia Neolithic and Körös_HG but not Anatolia Neolithic and Loschbour, while the opposite is true for Bohemia-Funnelbeaker (table S12). This suggests different affinities, in addition to the different introgression dates, of the HG ancestries in Bohemian Jordanów and Funnelbeaker cultural groups. Third, \( q \)Wave does not support cladality between Bohemian Jordanów and Funnelbeaker (\( P = 0.00679 \)), while cladality between Bohemian and German Funnelbeaker cannot be rejected (\( P = 0.88 \); table S13). Together, these results indicate a largely (significantly more than 50%) nonlocal genetic origin of Bohemian Funnelbeaker individuals.

The second such case can be seen in the Řivnác to GAC cultural transition. GAC individuals carry the most HG ancestry among pre-CW cultural groups from Bohemia (25.7%, ±1.4), significantly more than Řivnác individuals [\( f^4(Mbuti.DG, WHG; Řivnác, GAC) \gg 0 \); \( z \) score, 4.46]. However, the estimated date of HG admixture in GAC is not later than in Řivnác individuals (Fig. 3B and table S11), suggesting that GAC individuals do not descend from a recent
mixture of Řivnáč and an HG source but instead constituted a recent, nonlocal incursion in Bohemia from a region that received more HG gene flow [e.g., Poland (13, 63)], in accordance with interpretations of archaeological evidence (56).

A distinct genetic origin for Řivnáč and GAC individuals is further supported by PCA and qpAdm modeling. From PCA, we find that with the exception of TUC003, Řivnáč and GAC individuals form distinct clouds (Fig. 3C). This is confirmed by qpAdm modeling where GAC individuals can be modeled as a mixture of Anatolia_Neolithic and Loschbour but not Anatolia_Neolithic and Körös_HG, while the opposite is true for Řivnáč individuals (table S14). Consequently, Řivnáč and GAC individuals are distinguishable based on the amount and source of HG ancestry, suggesting that Bohemia was inhabited by genetically differentiated groups of Řivnáč and GAC individuals at the time of CW appearance. The Řivnáč outlier (TUC003) also raises the interesting possibility of an individual born into a GAC but buried in a Řivnáč cultural context.

Among the 16 Řivnáč and GAC individuals who are contemporaneous with or postdate the appearance of CW in Bohemia (Fig. 1B), we find no detectable traces of steppe ancestry (Fig. 2A and table S8), suggesting that biological exchange from CW/Yamnaya into culturally pre-CW people (e.g., Řivnáč and GAC) was low, possibly non-existent. Steppe ancestry coappears with CW individuals in early third millennium BCE Bohemia.

**Corded Ware**

We report genomic data from the earliest CW individuals to date, including STD003 (northwestern Bohemia, 3010 to 2889 calibrated (cal) BCE), VL1076 (central Bohemia, 3018 to 2901 cal BCE), OBR003 (central Bohemia, 2911 to 2875 cal BCE), and PNL001 (eastern Bohemia, 2914 to 2879 cal BCE), showing that CW was widespread across Bohemia by 2900 BCE. The early radiocarbon dates are also supported by these individuals’ genetic profiles, who occupy the most extreme positions on PC2 (Fig. 2B), as expected under a scenario of the earliest CW being migrants from the east who mixed with locals, resulting in intermediate PC2 positions in later generations.

To explore the formation of the Bohemian CW gene pool, we grouped CW individuals with steppe ancestry and mean age > 2600 BCE (n = 27) into a Bohemia_CW_Early group and the rest (n = 21) into Bohemia_CW_Late (table S4). We found poor statistical support (P < 0.005) for modeling Bohemia_CW_Early as a two-way mixture of any known Yamnaya source and any local Bohemian or nonlocal pre-CW source from Poland, Ukraine, Hungary, or Germany (table S15). When using distal sources as proxies for the Neolithic ancestry (Anatolia_Neolithic and a range of HG sources), we found no strong support (P < 0.05) for all but one of the three-way distal models (table S16). However, this one statistically supported model results in a previously unobserved ratio of Neolithic ancestry in Europe (i.e., a Neolithic population of ~1:1 ratio of Anatolia_Neolithic:Sweden_Motala_HG). In addition, when modeling early CW individually as “standard” three-way mixtures of Anatolia_Neolithic, WHG, and Yamnaya_Samara (3), we find that in 37% (10 of 27) of cases, the model lacks strong support (P < 0.05) or infeasible; fig. S6 and table S9).

To explore why two-way proximal models between any Yamnaya and a European Neolithic source are insufficient in explaining Bohemia_CW_Early genetic diversity, we tried adding a third source to obtain better model fits. We find that when either one of Latvia_MN, Ukraine_Neolithic, or PittedWare is added as a source,
We observe a closer phylogenetic relationship between the Y chromosome lineages found in early CW and BB than in either late CW or Yamnaya and BB. R1b-L151 is the most common Y-lineage compared to autosomal genetic diversity, ruling out a population bottleneck as the likely cause. Our results suggest that the Y-lineage diversity in early CW males was supplanted by a nonrandom process [selection, social structure, or influx of nonlocal Y-haplogroups (64), possibly due to increased conflict between male-mediated patriline (65)]. We view that changes in social structure (e.g., an isolated mating network with strictly exclusive social norms) could be an alternative cause but would be difficult to distinguish in the underlying model parameters.

Bell Beaker

The earliest BB individuals occupy a similar position in PCA as CW individuals (Fig. 4B and fig. S7), suggesting a degree of genetic continuity. To explore the genetic origin of early BB individuals (Bohemia_BB_Early; mean date, 2400 BCE; n = 3), we modeled them as a two-way mixture between preceding and contemporaneous cultural groups. We found support for a local origin, although non-local alternatives cannot be ruled out (table S28). However, our Bohemia_BB_Early group consists of only three (female) individuals and is therefore likely limited in representativeness and resolution to discern source populations.

We find that late BB individuals (Bohemia_BB_Late; mean date, 2400 BCE; n = 56) carry significantly more Middle Eneolithic–like ancestry compared to Bohemia_BB_Early (table S29). To explore this genetic shift, we modeled the ancestry of Bohemia_BB_Late as a two-way mixture of Bohemia_BB_Early and local Middle Eneolithic sources (table S30), finding support for an additional ~20% local Middle Eneolithic–like ancestry in late compared to early BB.

We observe a closer phylogenetic relationship between the Y chromosome lineages found in early CW and BB than in either late CW or Yamnaya and BB. R1b-L151 is the most common Y-lineage among early CW males (6 of 11, 55%) and one branch ancestral to R1b-P312 (Fig. 4A), the dominant Y-lineage in BB (5). Although it is not possible to determine whether the P312 mutation(s) occurred in one of the early CW R1b-L151 males from Bohemia, we note that most Bohemian BB males are further derived at R1b-L2/S116 (R1b1a1a2b1), in contrast to BB males from England, several of
whom are derived at R1b-L21(R1b1a1a2c1), showing that English and Bohemian BB males cannot be descendants of one another, but rather diversified in parallel. A scenario of R1b-P312 originating somewhere between Bohemia and England, possibly in the vicinity of the Rhine (66, 67), followed by an expansion northwest and east is compatible with our current understanding of the phylogeography of ancient R1b-L151–derived lineages.

**EBA—Únětice culture**

The transition to the EBA in Bohemia is associated with a positive shift in the coordinates of PC2, relative to preceding late BBs (Fig. 4B, fig. S7, and table S31). Admixture $f_3$ statistics are most negative when EHG (Eastern HG) or WSHG (West Siberian HG) are used as a second source in addition to the geographically and temporally proximal Bohemia_BB_Late (table S32), suggesting a northeastern contribution to Bohemia_UNETICE_preClassical. To find a suitable proxy for a potential additional source population, we modeled Bohemia_UNETICE_preClassical as a two-way mixture of local Bohemia_BB_Late and various sources more positive on PC2 (table S33). We reject mixture models involving Bohemia_BB_Late and Yamnaya (Samara, $P = 5.3 \times 10^{-10}$; Kalmykia, $P = 5.8 \times 10^{-10}$; Ukraine, $P = 7.3 \times 10^{-12}$; and Caucasus, $P = 3.2 \times 10^{-15}$) or Bohemia_BB_Late.
and CW (early, \( P = 1.1 \times 10^{-4} \); late, \( P = 5.4 \times 10^{-6} \)). We fail to reject a two-way mixture model of 63.5% Bohemia_BB_Early and 36.5% Bohemia_BB_Late (\( P = 0.29 \)), suggesting a large (63.5%) contribution from an early BB lineage, which was largely unsampled during the late BB phase (2400 to 2200 BCE), but represents a potential new lineage at the dawn of the Bronze Age. The Y-chromosomal data suggest an even larger turnover. A decrease of Y-lineage R1b-P312 from 100% (in late BB) to 20% (in preclassical Únětice) implies a minimum 80% influx of new Y-lineages at the onset of the EBA.

However, aware of the limited resolution of Bohemia_BB_Early (small sample size, low resolution, and large SEs), we explored alternative models for preclassical Únětice individuals. All model fits improve when Latvia_BA is included in the sources, resulting in two additional supported models (table S33). A three-way mixture of Bohemia_BB_Late, Bohemia_CW_Early, and Latvia_BA (\( P \) value of 0.086) not only supports a more conservative estimate of 47.7% population replacement but also accounts for the Y-chromosomal diversity found in preclassical Únětice, with R1b-P312 from Bohemia_BB_Late, R1b-U106 and I2 from Bohemia_CW_Early, and R1a-Z645 from Latvia_BA (Fig. 4A).

Although the geographic origin of this new ancestry cannot be precisely located, three observations offer clues. First, the Latvia_BA ancestry that improves all model fits (table S33) suggests an ultimate northeastern origin. Second, Y-haplogroup R1a-Z645 appears in Bohemia (and wider central Europe) for the first time at the beginning of the EBA, a lineage previously fixed in Baltic and common in Scandinavian CW males (23, 24), supporting a north/northeastern genetic contribution. Third, an Únětice genetic outlier (VLI051, male, Y-haplogroup R1a-Z645; table S34) resembles individuals from Bronze Age Latvia (Fig. 2D) (68), providing direct evidence for migrants from the northeast.

Fig. 5. Genetic distances of Early CW individuals from Vliněves. Pairwise \( F_{st} \) between the three highest (VLI076, VLI088, and VLI090; VLI_High) and three lowest (VLI008, VLI079, and VLI009; VLI_Low) early CW on PC2 from Vliněves in the context of modern European pairwise \( F_{st} \) (2 SE plotted; table S27). Inset: PCA of 1.141 modern West Eurasian individuals (gray points) on which early CW individuals from Vliněves (green symbols with black outline) were projected (see also Fig. 2B). The highly differentiated pairs of modern European populations labeled in the main figure are colored in the inset PCA.
We also detect a genetic shift in the transition from preclassical to classical Unětice, reflected in the decrease in PC2 coordinates for Unětice individuals dated after ~2000 BCE (Fig. 4B and fig. S7) and confirmed using qpWave (table S35) and f4 statistics (table S36). Bohemia_Unětice_Classical can be modeled as a mixture of Bohemia_Unětice_PreClassical and a local Eneolithic source (table S37). In contrast to the genetic shift between late BB and preclassical Unětice, the Y-lineage diversity remains similar throughout both Unětice phases, suggesting assimilation and subtler social changes.

**DISCUSSION**

The high-resolution genetic time transect in Bohemia, allowing early and late phases of cultural groups to be divided and studied separately (e.g., CW, BB, and Unětice), elucidates several major processes before and after the arrival of steppe ancestry (Fig. 6). Our dense sampling allows detection of novel, important, and perhaps “unexpected” changes within cultural groups (e.g., CW and BB), if they are seen through a strict cultural-historical lens. Previous studies have largely been interpreted as revealing major migrations at the beginning and end of the Neolithic (i.e., periods where the incoming groups were genetically very distinct); however, our results reveal additional large genetic turnovers. By sampling consecutive and partially contemporaneous cultural groups, we show that the spread of Funnelbeaker and GAC (69, 70), as well as the origin of Unětice, involved large genetic shifts over short time periods, likely explained by migrations.

We show that early CW were genetically exceptionally diverse, some resembling GAC and Yamnaya, with a few also falling outside of previously sampled central European Neolithic genetic diversity. Such a notably diverse signal is likely the result of the agglomeration of people from diverse cultural and linguistic backgrounds into an archaeologically similar but polyethnic or plural society. Important factors in ethnic identity include ancestry, history, ideology, and language (71, 72). The level of genetic differentiation (i.e., time since common ancestor) between early CW individuals with high and no steppe ancestry implies long biological isolation and hence different histories. The finding of GAC-like and Yamnaya-like genetic profiles in early CW suggests integration of people who came from ideologically diverse societies (i.e., neither GAC nor Yamnaya practiced strong gender differentiation in mortuary practices, unlike CW). It is likely that GAC and CW/Yamnaya individuals spoke different languages (3, 4, 43), meaning that early CW society in Bohemia encompassed people who had demonstrably different histories, likely originating from ideologically diverse cultures, who spoke different mother tongues.

The assimilation process of individuals without steppe ancestry into early CW society was female-biased (43). However, finding females also among individuals with the highest amounts of steppe ancestry (3 of 5; Fig. 2B) suggests that they were also well represented among migrating CW individuals [in contrast to (43)] or perhaps assimilated from nearby Yamnaya groups (e.g., Hungary). Finding individuals without steppe ancestry in early CW contexts (n = 4) is more common than individuals with steppe ancestry in pre-CW contexts (e.g., GAC, n = 0). This pattern of asymmetric gene flow between the contemporaneous GAC and CW may reflect newcomers (CW groups) having more benefit from incorporating people with important local knowledge (i.e., from pre-CW cultural contexts) into their communities. The archaeological record shows continuity of such knowledge (e.g., pottery production and lithic raw materials) in several regions (22, 67, 73).

FIG. 6. Schematic summary of the major processes that shaped the genetic and cultural diversity of Bohemia (red outline) over time. Arrows on maps indicate a general direction of influences rather than discrete routes of migration.
Vliněves is crucial for elucidating interactions between individuals with high and no steppe ancestry. This site yields the earliest dated CW (VLI076, 3018 to 2901 BCE) who is also genetically most differentiated from pre-CW individuals, while 20% (3 of 15) of the sampled early CW from Vliněves had no steppe ancestry. Intriguingly, we observe no archaeological differences between CW graves of individuals with and without steppe ancestry from two sites (Vliněves and Stadice; see the Supplementary Materials), suggesting full integration of genetically, and likely ethnically, diverse individuals within the same archaeological culture.

Finding Latvia MN-like ancestry in early CW, in conjunction with the absence of Y-chromosomal sharing between early CW and Yamnaya males, suggests a limited or indirect role of known Yamnaya in the origin and spread of CW to central Europe. Our results allude to either a northeast European Enolithic forest steppe contribution to early CW [a region consistent with some interpretations of the archaeological evidence (57)] or a hitherto unsampled steppe population who carried excess Latvia MN-like ancestry, a scenario that is less likely given the high degree of genetic homogeneity among 3000-BCE steppe groups [e.g., Yamnaya and Afanasievo separated by ~2500 km but genetically almost indistinguishable (4, 61)]. As much of 4000- to 2500-BCE (north)eastern Europe remains unsampled, inferring the precise geographic origin of early CW individuals remains elusive.

Since social kinship systems influence patterns of genetic diversity (13, 42, 48, 74), it is likely that several different kin systems existed in third millennium BCE central Europe. The highly diverse genetic profiles (both nuclear and Y-chromosomal) of early CW suggest a different social organization to late CW and BB, whose Y chromosome pattern is indicative of strict patrilineality. This suggests that different cultural groups, in addition to using various forms of material culture and mortuary practices, likely also conformed to different ideologies as expressed in their mating pattern and/or social organization. This is supported by the finding of completely non-overlapping Y chromosome variation between the partially contemporaneous late CW and BB, indicating a large degree of paternal mating isolation between these two groups, even when found at the same site (e.g., Vliněves).

The onset of the preclassical Únětice was accompanied by a ≥40% nuclear and ≥80% Y-chromosomal contribution ultimately originating from the northeast and breaking down the gender-differentiated mortuary practices and strict patrilineality of late CW and BB. This was neither evident in the burial customs nor in the material culture but could represent the underlying connection to the Baltics, the ultimate source of EBA amber in Bohemia associated with the later emerging Amber Road (75–77). Therefore, our results suggest two main periods (early CW and early Únětice) of genetic influence from the northeast, much of which remains unsampled in the European archaeogenetic record (e.g., Belarus).

Our results reveal a complex and highly dynamic history of Neolithic to EBA central Europe, during which migration and the movement of people facilitated abrupt genetic and social changes. Large-scale demic expansions occurred multiple times before and after the appearance of steppe ancestry in Europe. Early CW society was diverse and emerged amid a strong cultural and genetic transition, involving males and females of diverse origins and likely ethnicities. Genetic shifts occurred within CW, BB, and EBA societies despite continuity in material culture. Cultural affiliations played a major role in third millennium BCE social behaviors, which ultimately changed with the influx of new people over time. Although the impact of social processes is observable in patterns of genetic diversity, further interdisciplinary research is required to characterize the drivers of these changes, both at a micro- and macro-regional level.

**MATERIALS AND METHODS**

**Processing sites for the newly reported individuals**

Most (186 of 206, 90.2%) of the newly reported individuals were entirely processed at the Max Planck Institute for the Science of Human History in Jena, Germany, and the full details of sampling and ancient DNA wet laboratory work and bioinformatic processing are summarized in what follows. The individuals from the site Makotřasy were initially sampled and processed into powder at the University of Vienna, followed by subsequent laboratory work and bioinformatic and ancient DNA analysis at Harvard Medical School following previously described protocols (61).

**Sampling**

In total, 389 pars petrosa, teeth, and bones from 261 individuals were processed as part of this study. Upon introduction into the clean room facilities at the Max Planck Institute for the Science of Human History in Jena, Germany, all samples were wiped with 5% bleach and ultraviolet (UV) irradiated for 20 min on each side. Teeth were sampled by removing the crown followed by drilling into the pulp chamber to create bone powder. Pars petrosa were sampled by drilling into their dense region (78) to create bone powder. Between 50 and 100 mg of resulting bone powder from each sample were collected in different 2-ml Biopure tubes (one tube per sample) and used in subsequent DNA extraction.

**DNA extraction**

One milliliter of extraction buffer [containing 0.9 ml of 0.5 M EDTA, 0.025 ml of proteinase K (0.25 mg/ml), and 0.075 ml of UV high-performance liquid chromatography (HPLC) water] was added to Biopure tubes containing bone powder. Biopure tubes were then sealed with Parafilm and incubated overnight on a rotating wheel at 37°C. After incubation, Biopure tubes were spun for 2 min at 18,500 relative centrifugal force (rcf), separating the soluble from insoluble parts of the resulting solution. The soluble part was transferred to a 50-ml falcon tube containing 10 ml of binding buffer and 400 μl of sodium acetate (3 M, pH 5.2). Resulting mixture was transferred to high pure extender assembly (HPEA) falcon tube, which was centrifuged at 1500 rpm in a 50-ml Thermo Fisher Scientific TX-400 Swinging Bucket Rotor for 8 min. The column from each HPEA tube was removed and inserted into a fresh collection tube and centrifuged at 18,500 rcf for 2 min. A total of 450 μl of wash buffer from the high pure viral nucleic acid kit (HPVNAK) was added to each column, which was then centrifuged at 8000 rcf for 1 min. Columns were then removed and placed into new collection tubes. Another round of washing was performed whereby 450 μl of wash buffer from the HPVNAK was added to each column and centrifuged at 8000 rcf for 1 min. Columns containing washed DNA were then transferred to 1.5-ml siliconized tubes. Fifty microliters of TET (tris-EDTA + Tween 20) buffer were added to the center of columns, and the columns were then incubated at room temperature for 3 min and centrifuged at 18,500 rcf for 1 min. Another 50 μl of TET buffer was added to the center of columns, after which they were centrifuged once more at 18,500 rcf for 1 min. The resulting 100 μl of DNA extracts was stored at −20°C until further processing.
DNA libraries and in-solution capture

Twenty-five microliters of DNA extract was used for the construction of (in most cases) double-stranded uracil DNA glycosylase (UDG) half-treated DNA libraries. UDG repair was performed by adding DNA extract to a 25-μl mastermix containing 6 μl of 10× Buffer Tango, 6 μl of 10 mM adenosine 5′-triphosphate, 0.5 μl of bovine serum albumin (BSA; 20 mg/ml), 0.2 μl of 25 mM each deoxynucleoside triphosphate (dNTP), 3.6 μl of 1-U USER enzyme, and 8.7 μl of UV HPLC water. Resulting mixture was incubated at 37°C for 30 min followed by 12°C for 1 min. Inhibition of UDG treatment was achieved through the addition of 3.6 μl of 2-U uracil glycosylase inhibitor (UGI) to each tube followed by incubation at 37°C for 30 min and again at 12°C for 1 min. Blunt-end repair was performed by adding 3 μl of 10-U T4 polynucleotide kinase and 1.65 μl of 3-U T4 DNA polymerase and incubating the resulting mixture at 25°C for 20 min and then at 12°C for 10 min. Blunt-end repaired mixture was purified using a MinElute kit and eluted in 20 μl of elution buffer (EB) containing 0.05% Tween. Illumina adapters were ligated onto DNA molecules through the mixture of 18 μl of eluate from the previous step with 20 μl of 2× Quick Ligase Buffer, 1 μl of 10 μM Adapter Mix, and 1 μl of 5-U Quick Ligase. Resulting mixture was incubated at 22°C for 20 min followed by purification with a MinElute kit and elution in 22 μl of EB containing 0.05% Tween. Adapter fill in reaction was performed by adding 20 μl of eluate from the previous step to 4 μl of 10× isothermal buffer, 0.2 μl of 25 mM each dNTP, 2 μl of 8-U Bst 2.0 polymerase, and 13.8 μl of UV HPLC water, followed by incubation at 37°C for 30 min and 80°C for 10 min. Resulting libraries were stored at −20°C until further processing. Unique library-specific indexes were added to the 5′ and 3′ ends of molecules in each library through an indexing polymerase chain reaction (PCR). Each library was split into four separate indexing PCR reactions, which were carried out using 10 μl of 10× Pf Turbo buffer, 1.5 μl of BSA (20 mg/ml), 1 μl of 25 mM each dNTP, 1 μl of 2.5-U Pf Turbo polymerase, 73.5 μl of UV HPLC water, 2 μl of 10 μM P5 index, 2 μl of 10 μM P7 index, and 9 μl of DNA library. Amplification was achieved through an initial denaturation at 95°C for 2 min, followed by 10 cycles of 95°C for 30 s, 58°C for 30 s, and 72°C for 1 min, followed by 72°C for 10 min. Resulting indexed libraries of the same sample were pooled and purified using a MinElute purification kit. Purified libraries were quantified using quantitative PCR and amplified to 10^15 copies. Amplified libraries were shallow shotgun sequenced (~5 million reads) on an Illumina HiSeq or NextSeq platform to estimate the general human DNA content, presence of ancient DNA damage, and mitochondrial/nuclear coverage ratio. Libraries with >0.1% endogenous human DNA and >5% C-to-T misincorporations at the 5′ end were chosen for 1240k capture (8) and mitochondrial DNA (mtDNA) capture. In cases where more than one library from the same individual satisfied the criteria for capture, the better quality (higher endogenous DNA content) library was used for 1240k and mtDNA capture.

Sequencing

Postcapture libraries were single-end (75 cycles) or paired-end sequenced (2 × 50 or 2 × 75) on HiSeq or NextSeq Illumina platform to a depth of 20 to 50 million reads per library. Resulting sequence data were processed through EAGER (v1.92.38) (79). Illumina adapters were removed using AdapterRemoval (v2.2.0) (80), and in case of paired-end sequencing, corresponding reads from the same template molecule with a minimum of 11 base pairs (bp) of overlap were merged. Fastq files of merged and unmapped reads were concatenated, and reads shorter than 30 bp were discarded. Processed reads were mapped to the human reference genome (hg19) using BWA-aln and BWA-samse (v0.7.12) (81) applying maxdaff (-n) 0.01 and seeding turned off (-1 10000). Resulting bam files were sorted, and duplicate reads were removed using DeDup (v0.12.1) (https://github.com/apelzer/DeDup). Damageprofiler (v0.3.10) (https://github.com/Integrative-Transcriptomics/DamageProfiler) was used to calculate rates of misincorporation in read termini of DNA fragments in our captured libraries. BAM files had the last three bases from both 5′ and 3′ ends of reads and corresponding base quality scores masked for downstream analyses (82).

Sex determination and authentication

The genetic sex of each sample (bam file) was determined by calculating the normalized mean coverage on the X (mean X coverage/mean autosomal coverage) and Y (mean Y coverage/mean autosomal coverage) chromosomes (83). Samples with normalized mean Y coverage values greater than 0.2 were assigned male. Contamination was estimated in males by calculating the rate of heterozygosity on their X chromosome (84). In addition, we used Schmutzi to estimate the mitochondrial contamination in all libraries (85). Schmutzi was run on BAM files resulting from mapping 1240K capture sequencing data to the human mitochondrial reference genome. In cases where 1240k data were not enough to give an mtDNA contamination, we ran Schmutzi on the mtDNA capture data mapped to the human mitochondrial reference genome.

Genotyping

We used SAMTools (v1.3) (86) mpileup and pileupCaller from the sequenceTools (v1.4.0.2) package (https://github.com/stschiff/sequenceTools) to call pseudo diploid genotypes by sampling a random high-quality allele (base quality, ≥30; mapping quality, ≥30) from each of the 1240k sites (8). Newly generated genotype data for this study were then merged to a compiled dataset of previously published ancient and modern worldwide populations (https://reich.hms.harvard.edu/Allen-ancient-dna-resource-aadr-downloadable-genotypes-present-day-and-ancient-dna-data) (v42.2) using mergeit (v2450) from the EIGENSOFT package (https://github.com/DRreichLab/EIG).

Mitochondrial and Y chromosome haplogroups

Mitochondrial haplogroups were called by mapping 1240k or mtDNA capture data to the human mitochondrial reference genome followed by creating pileups at each position (map quality and base quality filter of 30) and calling the most frequent base at each position. Resulting genotype information was converted to fasta files, and haplogroups were called using haplofind (87). Y chromosome haplogroups were called by mapping 1240k capture data to the whole human reference genome (hg19) followed by visual inspection of ancestral/derived alleles (after map quality and base quality filter of 30) at ISOGG (v15.58 April 2020) sites.

Principal components analysis

PCA was conducted using smartpca (v1600) from the Eigensoft package (https://github.com/DRreichLab/EIG). Principal components were calculated on the genotype data of modern West Eurasian individuals (table S7) (88–90). Ancient individuals were projected (lqproject: YES) onto the axes calculated from modern individuals.
“shrinkmode: YES” was used to account for artificial stretching of principal component axes between projected (ancient) individuals and modern individuals. FST values were also calculated in smartpca using “fst:precision: YES” and “inbreed: YES” parameters.

**Y haplogroup frequency simulations**

To investigate the process of Y-haplogroup inheritance in early and late CW groups, we simulated 100 realizations, assuming a generational time of 25 years, and analyzed the results using approximate Bayesian computation (ABC). For the ith realization, we assumed a constant population size of $N_t \sim U(10^2,10^4)$, with a starting a proportion of male offspring that R1a-M417(xZ645) individuals were having over this period. We then compared our observed number of per-generation R1a-M417(xZ645) to our simulated realizations using the rejection method and keeping the top closest 0.05% realizations (selected via cross-validation) to form samples from the joint posterior distributions for our simulation parameters. All ABC and cross-validation analyses were run in R using the abc package (94).

**SUPPLEMENTARY MATERIALS**

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/35/eabf6941/DC1

**REFERENCES AND NOTES**

44. I. Lazaridis, D. Reich, Failure to replicate a genetic signal for sex bias in the steppe.


72. G. Bailey, J. Peoples, M. Szmyt, Between the seas: Baltic-Pontic contact space in the 3rd millennium BC. Baltic-Pontic studies, Institute of Prehistory, 1999), vol. 8 of (Adam Mickiewicz University in Pozna


et al


M. Zápotocký, Čimburk und die Höhensiedlungen des Frühen und Älteren Äneolithikums in Böhmen (Prague, 2000).


120. M. Dobeš, J. Chochol, Kostrové pozůstatky z hrobu s kulovitými amforami v Blšanech u Mostu, 1999), pp. 7–43.

121. J. Chochol, Kostrové pozůstatky z hrobu s kulovitými amforami v Blšanech u Mostu. Architektonické nálezy ze severozápadních Čech, Most (1999), pp. 7–43.


123. J. Chochol, Kostrové pozůstatky z hrobu s kulovitými amforami v Blšanech u Mostu. Architektonické nálezy ze severozápadních Čech, Most (1999), pp. 7–43.


125. J. Chochol, Kostrové pozůstatky z hrobu s kulovitými amforami v Blšanech u Mostu. Architektonické nálezy ze severozápadních Čech, Most (1999), pp. 7–43.


128. J. Chochol, Kostrové pozůstatky z hrobu s kulovitými amforami v Blšanech u Mostu. Architektonické nálezy ze severozápadních Čech, Most (1999), pp. 7–43.


138. J. Chochol, Kostrové pozůstatky z hrobu s kulovitými amforami v Blšanech u Mostu. Architektonické nálezy ze severozápadních Čech, Most (1999), pp. 7–43.


140. J. Chochol, Kostrové pozůstatky z hrobu s kulovitými amforami v Blšanech u Mostu. Architektonické nálezy ze severozápadních Čech, Most (1999), pp. 7–43.
Acknowledgments: We thank S. Kaupova and E. Zavoninova for the help during sample collection; G. Brandt, A. Wissgott, E. Hoché, M. Himmel, and V. Wanka for the laboratory support; and S. Clayton and K. Prüfer for processing the raw sequencing data. We thank M. Furtuhl, H. Ringbauer, A. Childebayeva, and the population genetics group of the Department of Archaeoegenetics for suggestions and valuable feedback. We thank M. O'Reilly and H. Sell for graphics support. D.R. is an investigator of the Howard Hughes Medical Institute. Funding: This work was supported by the following: Max Planck Society; European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program under grant agreements no. 771234–PALEoRIDER (to W.H.) and no. 788616–YMPACT (to V.H.); John Templeton Foundation grant 61220 (to D.R.); Paul Allen Family Foundation (to D.R.); Czech Academy of Sciences award Premium Academiae (to M.E.); Institutional program RVO 67985912 of the Institute of Archaeology of the Czech Academy of Sciences; Czech Science Foundation grant 20–90709Y (to J.Ko.); and Czech Science Foundation grant 20–90709Y (to J.K.).


Downloaded from https://www.science.org on March 13, 2022
Dynamic changes in genomic and social structures in third millennium BCE central Europe

Luka Papac, Michal Ernée, Miroslav Dobeš, Michaela Langová, Adam B. Rohrlach, Franziska Aron, Gunnar U. Neumann, Maria A. Spyrou, Nadin Rohland, Petr Velemínský, Martin Kuna, Hana Brzobohatá, Brendan Culleton, David Dančík, Alžbeta Danielisová, Miluše Dobíšková, Josef Hloušek, Douglas J. Kennett, Jana Klementová, Michal Kostka, Petr Kříška, Milan Kuchař, Jaroslav Kuljavec, Páťo Limburský, Drahomíra Malyková, Lucia Mattiello, Monika Pecinovská, Katarina Petrišiková, Erika Prášková, Petra Stranská, Lubor Smejtek, Jaroslav Špaček, Šumberová, Ondřej Švejcar, Martin Trefný, Miloš Vávra, Jan Kolář, Volker Heyd, Johannes Krause, Ron Pinhasi, David Reich, Stephan Schiffels, Wolfgang Haak

Sci. Adv., 7 (35), eabi6941. • DOI: 10.1126/sciadv.abi6941