Detektion von langen identischen Haplotypen in alter DNA – Bestimmung von nahen und entfernten Verwandten

Harald Ringbauer

Summary

Detecting long indentical haplotypes in ancient DNA – Identifying close and distant relatives

Our new bioinformatic method called ancIBD makes it possible for the first time to find long identical DNA sequences between pairs of ancient DNA (aDNA). These so-called IBD segments (from »Identity by Descent«) must come from a common ancestor only a few generations away and are the direct genetic signals of biological relationship. Here we give an overview of the possible applications of ancIBD. For example, using IBD segments, it is possible to identify relatives up to the 6th degree – in contrast to the methods commonly used in aDNA, which only work up to the 3rd degree. Furthermore, the application of ancIBD makes it possible to measure genealogical connections between archaeological cultures within a few hundred years. We show the potential of our method by direct applications to published aDNA and find new biological connections and clusters in the West Eurasian Aeneolithic and the Bronze Age.

Zusammenfassung

Unsere neue bioinformatische Methode namens ancIBD ermöglicht es erstmals lange identische DNA-Sequenzen zwischen Paaren von alter DNA (aDNA) zu finden. Diese sogenannten IBD-Segmente (von »Identity by Descent«) müssen von gemeinsamen Vorfahren nur wenige Generationen entfernt stammen und sind die direkten genetischen Signale von biologischer Verwandtschaft. Wir geben hier eine Übersicht über die Anwendungsmöglichkeiten von ancIBD. So wird es etwa mittels IBD-Segmenten möglich, bis zum 6. Grad Verwandte nachzuweisen – im Gegensatz zu üblicherweise verwendeten Methoden in der aDNA, welche nur bis zum 3. Grad Verwandte identifizieren können. Darüber hinaus ermöglicht es die Anwendung von ancIBD, genealogische Zusammenhänge zwischen archäologischen Kulturen innerhalb weniger Jahrhunderte zu messen. Wir zeigen das Potential unserer Methode durch direkte Anwendungen auf publizierte aDNA und finden etwa neue Verbindungen und Cluster im westeurasischen Äneolithikum und in der Bronzezeit.

Dieser Artikel gibt einen Überblick über eine neue bioinformatische Methode zur Detektion langer identischer DNA-Segmente in der aDNA, also der DNA von Individuen, die vor hunderten oder tausenden Jahren gelebt haben. Die technischen Details und ersten Anwendungen beschreiben wir im Detail in einem Preprint (Ringbauer u. a. 2023). Der Fokus der vorliegenden Arbeit liegt auf einer Zusammenfassung der neuen Analyse-Möglichkeiten, die diese Methode eröffnet. Die Zielgruppe sind Leser, die an den direkten Anwendungen an aDNA interessiert sind.

Die Genome mancher Paare von Individuen teilen sich lange, beinahe identische DNA-Segmente. Diese werden oft »Identity-by-Descent« oder abgekürzt »IBD«, genannt, da sie von einem gemeinsamen Vorfahren vor nur wenigen Generationen ererbt worden sein müssen. Andernfalls wären die gemeinsamen Segmente schon zerstückelt worden, da bei jeder Generation die Genome zweier Eltern wie ein Mosaik vermischt werden. Nach nur wenigen Generationen bleiben keine langen IBD-Segmente mehr erhalten.

Daher sind lange IBD-Segmente *de facto* diagnostisch für biologische Verwandtschaften innerhalb weniger Generationen. So teilen sich nahe Verwandte eine große Anzahl von langen IBD-Segmenten (Abb. 1). Je weiter entfernt die Verwandtschaft, desto kürzer und weniger werden diese. Tatsächlich teilen sich Verwandte bis zum 6. Grad in den meisten Fällen mehrere lange IBD-Segmente, darüber hinaus werden nur einzelne oder meist gar keine IBD-Segmente geteilt – da so entfernte Verwandte sich in den meisten Fällen gar keine DNA mehr teilen. Hier meinen wir mit Grad, dass jede Verbindung zu einem Elternteil oder Kind je einen Grad addiert, während Geschwister ebenso einen Grad und Halbgeschwister zwei Grade addieren. Die genaue IBD-Segment-Verteilung ist stochastisch, da eine Rekombination zufällig entlang des Genoms passiert. Bis zum 3. Verwandtschaftsgrad treten aber klar definierte Cluster auf (vgl. Abb. 1b). Darüber hinaus kann man bis zum 6. Grad eindeutig Verwandtschaft anhand mehrerer langer IBD-Segmente nachweisen – die Cluster beginnen sich allerdings zu überlappen (vgl. Abb. 1c), oft ist für Verwandte 4–6. Grades daher keine eindeutige Bestimmung des exakten Grades möglich, und es verbleibt etwa ein Grad Unsicherheit. Das genetische Signal von IBD-Segmenten ist daher überaus spannend für Verwandtschaftsbestimmung und ebenso für demografische Analysen: Man erhält eine genetische Lupe für Verwandtschaften in der jüngeren Vergangenheit. Für moderne DNA-Daten existieren einige Methoden, diese IBD-Segmente zu bestimmen. Oft wurden diese Segmente benutzt, um mit diesem Signal wichtige demografische Parameter abzuschätzen. So wurde etwa dieses Signal verwendet, um genealogische Verbindungen aus ganz Europa über die letzten Jahrhunderte nachzuweisen und deren Intensität über diverse Zeitperioden zu ermitteln (Ralph/Coop 2013). Da geografisch weiter entfernte Gruppen immer weni-

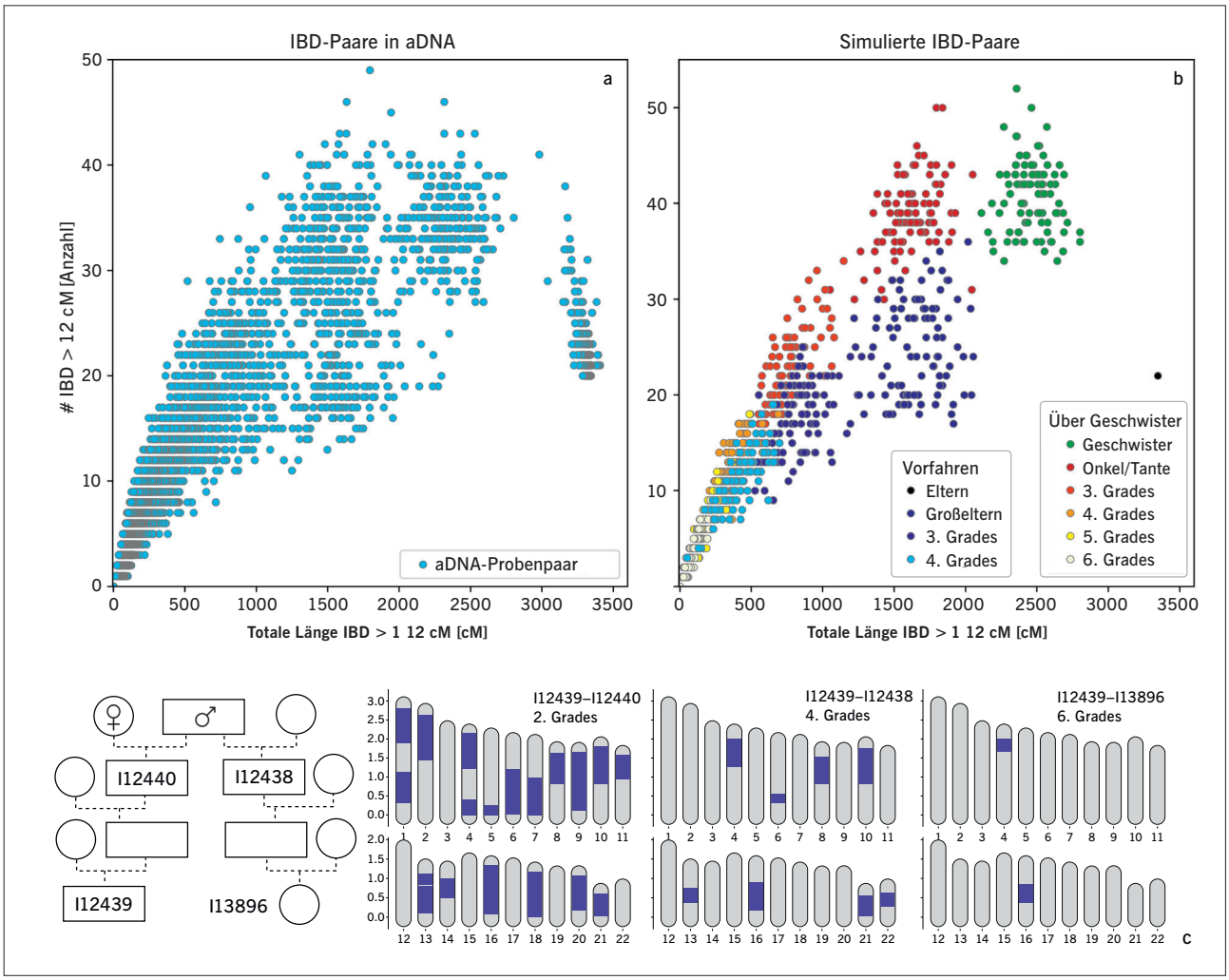


Abb. 1a–c Lange IBD-Segmente zwischen nahen Verwandten. a IBD-Segmente in einem großen aDNA-Datensatz analysiert mit ancIBD. Jeder Punkt stellt ein Paar von aDNA-Individuen dar. Es werden die Summe des Genoms und die Anzahl der langen IBD-Segmente (> 12 cM lang) dargestellt, je auf der X- und Y-Achse. Wir bemerken, dass viele Punkte 0 IBD-Segmente besitzen, aber hier nähere Verwandte der Fokus der Abbildung sind; b Simulierte IBD-Segmente zwischen verschiedenen Verwandten. Wie in a) aber wir haben die Verwandten simuliert (mit der Software PEDSIM, je 100 Replikate pro Verwandtschaftsgrad), und stellen die simulierten IBD-Segmente pro Paar dar; c Diese Abbildung stellt IBD-Segmente zwischen drei Paaren von nahen Verwandten dar. Wir zeigen Stellen im Genom (entlang der 22 Chromosome), an denen zwei Individuen IBD-Segmente besitzen (blaues Segment). Dieser Stammbaum ist Teil eines größeren rekonstruierten aDNA-Stammbaums aus Hazelton, Gloucestershire (England; Neolithikum) und wurde von C. Fowler u. a. (2022) publiziert.

Fig. 1a–c *Long IBD segments between close relatives. a IBD segments in a large aDNA dataset analyzed with ancIBD. Each point represents a pair of aDNA individuals. We plot the sum of the genome and the number of long IBD segments (> 12 cM long) on the x- and y-axes, respectively. We note that many points have 0 IBD segments, but here closer relatives are the focus of the figure; b Simulated IBD segments between different relatives. As in a) but we have simulated the distribution (using the software PEDSIM, 100 replicates per relatedness), and plot the true IBD segments; c This figure represents IBD segments between three pairs of close relatives. We show locations in the genome (along the 22 chromosomes) where two individuals have IBD segments (red segment). This family tree is part of a larger reconstructed aDNA family tree from Hazelton, Gloucestershire (England; Neolithic) and was compiled by C. Fowler et al. (2022) published.*

ger IBD-Segmente teilen, weil immer weniger gemeinsame Vorfahren vorhanden sind, sind IBD-Segmente auch ein ideales Signal um typische »Migrationsdistanzen« abzuschätzen (s. Ringbauer u. a. 2017). Da die »Hintergrundverwandtschaft« einer Population direkt mit der »effektiven« Populationsgröße zusammenhängt, wurden auch Analysemethoden entwickelt, die IBD-Segmente benutzen um Bevölkerungsgrößendynamiken über die letzten dutzenden Generationen abzuschätzen (vgl. Browning/Browning 2015).

Typische aDNA ist leider meist von weit geringerer Qualität und Sequenziertiefe als moderne DNA. Deswegen können moderne IBD-Detektionsmethoden nicht einfach direkt auf aDNA angewandt werden. So blieben bis jetzt typische aDNA-Verwandtschaftsbestimmungen nur bis zum maximal 3. Grad beschränkt, da nur durchschnittliche Statistiken über das ganze Genom gemittelt ausgenutzt werden können. Die Information von IBD-Segmenten, die oft nur einen kleinen Teil des Genoms ausmachen (vgl. Abb. 1c, zum 6. Grad Verwandte), geht darin unglücklicherweise komplett verloren.

Um das zu ändern und IBD-Segmente auch in alter DNA zu benutzen, haben wir eine neue bioinformatische Methode entwickelt, namens ancIBD, die im Detail in H. Ringbauer u. a. (2023) beschrieben ist und als Software-Paket veröffentlicht wurde. Kurz zusammengefasst gibt es zwei Schritte. Im ersten werden Lücken in der aDNA mithilfe moderner Genome gefüllt (sogenannte »Imputation«, wir benutzen die Software GLIMPSE; vgl. Rubinacci u. a. 2021). Im zweiten Schritt wenden wir dann ein von uns neu entwickeltes Hidden-Markov-

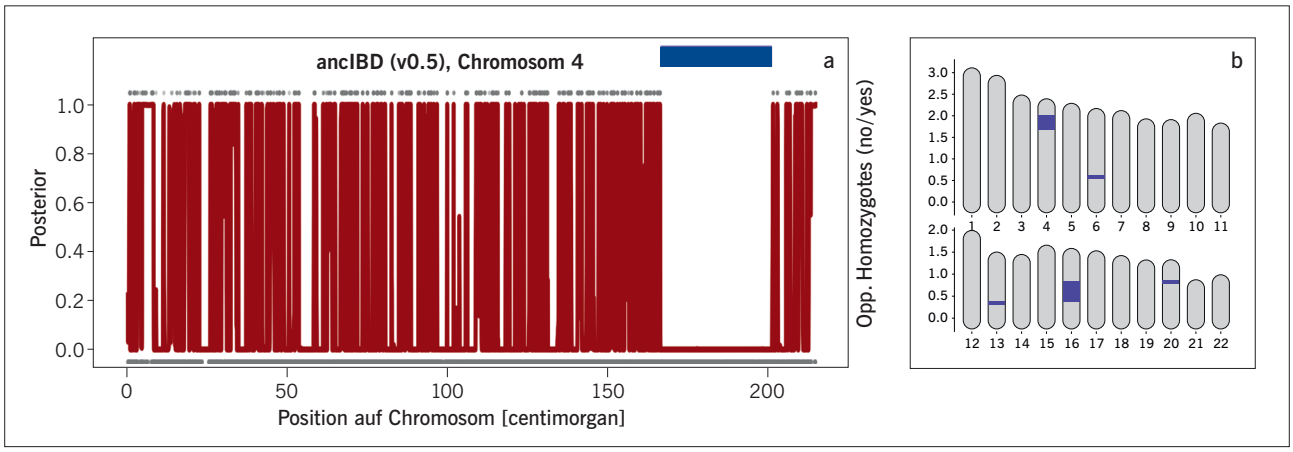


Abb. 2a–b Schematische Darstellung des Outputs von ancIBD. a Unser Algorithmus läuft entlang eines Chromosoms eines Paares von Individuen. Er findet lange genomische Sequenzen, wo beide Genome keine imputierten »Opposing Homozygotes« (graue Punkte oben) aufweisen. Basierend auf einem statistischen Posterior (rot) identifizieren wir diese Regionen als IBD-Segmente (blau); b Angewandt auf alle 22 Chromosome ergibt das alle IBD Segmente beider Individuen (blau).

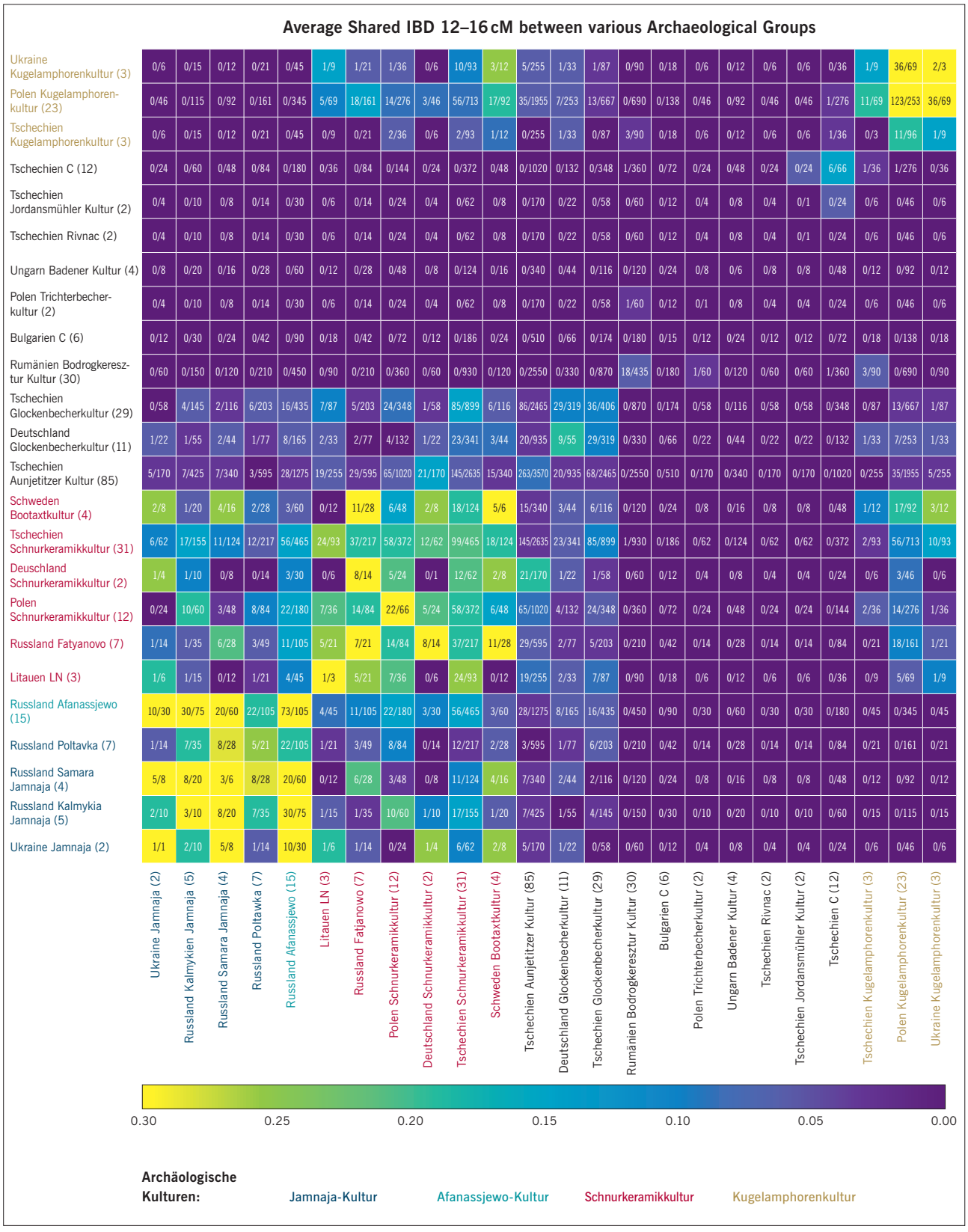
Fig. 2a–b *Schematic representation of the output of ancIBD. a Our algorithm runs along a chromosome of a pair of individuals. It finds long genomic sequences where both genomes have no imputed opposing homozygotes (grey dots above). Based on a statistical posterior (red), we identify these regions as IBD segments (blue); b Applied to all 22 chromosomes, this then results in all IBD segments of both individuals (blue).*

Model (HMM) an; dieses mathematische Model ist der Kern von ancIBD. Kurz zusammengefasst, suchen wir nach Stellen im Genom zweier Individuen, wo für lange Strecken keine sogenannten »Opposing Homozygotes« vorkommen, an denen die beiden Individuen je unterschiedliche Varianten homozygot (also doppelt) haben. Diese können in IBD-Segmenten nicht vorkommen, da immer zumindest ein Allel geteilt werden muss. Daher ist die Absenz von Opposing Homozygotes über lange genomische Strecken diagnostisch für IBD-Segmente (Abb. 2).

Da unsere Methode zentral auf Imputation beruht, funktioniert sie für aDNA der letzten 45 000 Jahre, mit der Ausnahme von Subsahara-Afrika wo genetische Diversität besteht die mit unserem derzeitigen Referenzpanel nur mangelhaft imputiert werden kann. Allerdings macht Imputation die Methode auch relativ datenhungrig und funktioniert daher nicht für schlecht erhaltene aDNA. Unsere Experimente haben aber gezeigt, dass etwa die Hälfte der publizierten »genomweiten« aDNA-Daten erfolgreich für IBD gescreent werden kann. Um mittlere und längere IBD-

Abb. 3a–b a Beispiel zweier Verwandter in alter DNA, die ancIBD gefunden hat, ca. 3300 v. Chr. Es handelt es sich um zwei unpublizierte Individuen, eines aus dem Kontext der Usatovokultur (Publikation in Vorbereitung; nach persönlicher Kommunikation mit D. Reich) und eines aus einem Kontext der Maikop-Kultur (Wang u. a. 2019). Mehrere lange IBD-Segmente zeigen eindeutig, dass es sich um einen biologischen nahen Verwandten – wahrscheinlich 5. Grades (4.–6. Grad möglich) – handelt. Das würde etwa zwei Individuen entsprechen, die zweite Cousins sind, also zwei gemeinsame Urgroßeltern haben; b Geographische Koordinaten der beiden Verwandten. Ein Individuum wurde im heutigen südlichen Moldawien bestattet und kann mit der Usatovo Kultur zugeordnet werden (Publikation in Vorbereitung; nach persönlicher Kommunikation mit D. Reich), das andere wurde im Nordkaukasus bestattet und kann mit der Maikop-Kultur assoziiert werden (Wang u. a. 2019). Die geographische Distanz zwischen den beiden Orten beträgt 1150 km (Luftlinie).

Fig. 3a–b *a Example of two relatives in ancient DNA c. 3300 BC found by ancIBD. They are two unpublished individuals, one from a Usatove Culture context (publication in preparation; D. Reich, personal communication) and one from a Maykop Culture context (Wang et al. 2019). Several long IBD segments clearly show that there is a biological close relative – probably 5th degree (4th–6th degree possible) – involved. This would roughly correspond to two individuals who are second cousins, i.e. where two great-grandparents are in common; b Geographic coordinates of the two relatives. One individual was buried in what is now southern Moldova and can be associated with the Usatovo culture (publication in preparation; D. Reich, personal communication), the other was buried in the North Caucasus and can be associated with the Maikop culture (Wang et al. 2019). The geographical distance between the two sites is 1150 km (linear distance).*



	Ukraine Kugelamphorenkultur (3)	Polen Kugelamphoren-kultur (23)	Tschechien Kugelamphorenkultur (3)	Tschechien C (12)	Tschechien Jordansmühler Kultur (2)	Tschechien Rivnac (2)	Ungarn Badener Kultur (4)	Polen Trichterbecher-kultur (2)	Bulgarien C (6)	Rumänien Bodrogkeresz-tur Kultur (30)	Tschechien Glockenbecherkultur (29)	Deutschland Glockenbecherkultur (11)	Tschechien Aunjetitzer Kultur (85)	Schweden Bootaxtkultur (4)	Tschechien Schnurkeramikkultur (31)	Deuschland Schnurkeramikkultur (2)	Polen Schnurkeramikkultur (12)	Russland Fatyanovo (7)	Litauen LN (3)	Russland Afanassjewo (15)	Russland Poltavka (7)	Russland Samara Jamnaja (4)	Russland Kalmykia Jamnaja (5)	Ukraine Jamnaja (2)
Ukraine Kugelamphorenkultur (3)	0/6	0/15	0/12	0/21	0/45	1/9	1/21	1/36	0/6	10/93	3/12	5/255	1/33	1/87	0/90	0/18	0/6	0/12	0/6	0/6	0/36	1/9	36/69	2/3
Polen Kugelamphoren-kultur (23)	0/46	0/115	0/92	0/161	0/345	5/69	18/161	14/276	3/46	56/713	17/92	35/1955	7/253	13/667	0/690	0/138	0/46	0/92	0/46	0/46	1/276	11/69	123/253	36/69
Tschechien Kugelamphorenkultur (3)	0/6	0/15	0/12	0/21	0/45	0/9	0/21	2/36	0/6	2/93	1/12	0/255	1/33	0/87	3/90	0/18	0/6	0/12	0/6	0/6	1/36	0/3	11/96	1/9
Tschechien C (12)	0/24	0/60	0/48	0/84	0/180	0/36	0/84	0/144	0/24	0/372	0/48	0/1020	0/132	0/348	1/360	0/72	0/24	0/48	0/24	0/24	6/66	1/36	1/276	0/36
Tschechien Jordansmühler Kultur (2)	0/4	0/10	0/8	0/14	0/30	0/6	0/14	0/24	0/4	0/62	0/8	0/170	0/22	0/58	0/60	0/12	0/4	0/8	0/4	0/1	0/24	0/6	0/46	0/6
Tschechien Rivnac (2)	0/4	0/10	0/8	0/14	0/30	0/6	0/14	0/24	0/4	0/62	0/8	0/170	0/22	0/58	0/60	0/12	0/4	0/8	0/4	0/1	0/24	0/6	0/46	0/6
Ungarn Badener Kultur (4)	0/8	0/20	0/16	0/28	0/60	0/12	0/28	0/48	0/8	0/124	0/16	0/340	0/44	0/116	0/120	0/24	0/8	0/6	0/8	0/8	0/48	0/12	0/92	0/12
Polen Trichterbecher-kultur (2)	0/4	0/10	0/8	0/14	0/30	0/6	0/14	0/24	0/4	0/62	0/8	0/170	0/22	0/58	1/60	0/12	0/1	0/8	0/4	0/4	0/24	0/6	0/46	0/6
Bulgarien C (6)	0/12	0/30	0/24	0/42	0/90	0/18	0/42	0/72	0/12	0/186	0/24	0/510	0/66	0/174	0/180	0/15	0/12	0/24	0/12	0/12	0/72	0/18	0/138	0/18
Rumänien Bodrogkeresz-tur Kultur (30)	0/60	0/150	0/120	0/210	0/450	0/90	0/210	0/360	0/60	0/930	0/120	0/2550	0/330	0/870	18/435	0/180	1/60	0/120	0/60	0/60	1/360	3/90	0/690	0/90
Tschechien Glockenbecherkultur (29)	0/58	4/145	2/116	6/203	16/435	7/87	5/203	24/348	1/58	85/899	6/116	86/2465	29/319	36/406	0/870	0/174	0/58	0/116	0/58	0/58	0/348	0/87	13/667	1/87
Deutschland Glockenbecherkultur (11)	1/22	1/55	2/44	1/77	8/165	2/33	2/77	4/132	1/22	23/341	3/44	20/935	9/55	29/319	0/330	0/66	0/22	0/44	0/22	0/22	0/132	1/33	7/253	1/33
Tschechien Aunjetitzer Kultur (85)	5/170	7/425	7/340	3/595	28/1275	19/255	29/595	65/1020	21/170	145/2635	15/340	263/3570	20/935	68/2465	0/2550	0/510	0/170	0/340	0/170	0/170	0/1020	0/255	35/1955	5/255
Schweden Bootaxtkultur (4)	2/8	1/20	4/16	2/28	3/60	0/12	11/28	6/48	2/8	18/124	5/6	15/340	3/44	6/116	0/120	0/24	0/8	0/16	0/8	0/8	0/48	1/12	17/92	3/12
Tschechien Schnurkeramikkultur (31)	6/62	17/155	11/124	12/217	56/465	24/93	37/217	58/372	12/62	99/465	18/124	145/2635	23/341	85/899	1/930	0/186	0/62	0/124	0/62	0/62	0/372	2/93	56/713	10/93
Deuschland Schnurkeramikkultur (2)	1/4	1/10	0/8	0/14	3/30	0/6	8/14	5/24	0/1	12/62	2/8	21/170	1/22	1/58	0/60	0/12	0/4	0/8	0/4	0/4	0/24	0/6	3/46	0/6
Polen Schnurkeramikkultur (12)	0/24	10/60	3/48	8/84	22/180	7/36	14/84	22/66	5/24	58/372	6/48	65/1020	4/132	24/348	0/360	0/72	0/24	0/48	0/24	0/24	0/144	2/36	14/276	1/36
Russland Fatyanovo (7)	1/14	1/35	6/28	3/49	11/105	5/21	7/21	14/84	8/14	37/217	11/28	29/595	2/77	5/203	0/210	0/42	0/14	0/28	0/14	0/14	0/84	0/21	18/161	1/21
Litauen LN (3)	1/6	1/15	0/12	1/21	4/45	1/3	5/21	7/36	0/6	24/93	0/12	19/255	2/33	7/87	0/90	0/18	0/6	0/12	0/6	0/6	0/36	0/9	5/69	1/9
Russland Afanassjewo (15)	10/30	30/75	20/60	22/105	73/105	4/45	11/105	22/180	3/30	56/465	3/60	28/1275	8/165	16/435	0/450	0/90	0/30	0/60	0/30	0/30	0/180	0/45	0/345	0/45
Russland Poltavka (7)	1/14	7/35	8/28	5/21	22/105	1/21	3/49	8/84	0/14	12/217	2/28	3/595	1/77	6/203	0/210	0/42	0/14	0/28	0/14	0/14	0/84	0/21	0/161	0/21
Russland Samara Jamnaja (4)	5/8	8/20	3/6	8/28	20/60	0/12	6/28	3/48	0/8	11/124	4/16	7/340	2/44	2/116	0/120	0/24	0/8	0/16	0/8	0/8	0/48	0/12	0/92	0/12
Russland Kalmykia Jamnaja (5)	2/10	3/10	8/20	7/35	30/75	1/15	1/35	10/60	1/10	17/155	1/20	7/425	1/55	4/145	0/150	0/30	0/10	0/20	0/10	0/10	0/60	0/15	0/115	0/15
Ukraine Jamnaja (2)	1/1	2/10	5/8	1/14	10/30	1/6	1/14	0/24	1/4	6/62	2/8	5/170	1/22	0/58	0/60	0/12	0/4	0/8	0/4	0/4	0/24	0/6	0/46	0/6

Segmente, die zumindest 12 cM lang sind, robust zu detektieren, benötigt es zumindest etwa 0,25 x durchschnittliches Coverage von Whole-Genome-Sequencing-Daten (WGS) und etwa 1 x Coverage auf 1240k oder Twist-Capture-Daten. Letztere Methode (Capture), wo gezielt aDNA von etwas mehr als einer Million Varianten angereichert wird, ist eine beliebte Technologie – der Großteil bisher produzierter alter Genome ist mittels Captures erstellt worden. Es ist eine besondere Stärke von ancIBD, auch diese Art von Daten gut verarbeiten zu können. Es sei auch gesagt, dass man ancIBD auch für schlechter erhaltene aDNA benutzen kann, der Anwender sollte sich allerdings bewusst sein, dass »False

Abb. 4 (linke Seite) Raten von langen IBD-Segmenten zwischen verschiedenen äneolithischen und bronzezeitlichen archäologischen Gruppen. Wir visualisieren Raten von IBD-Segmenten, die 12–16 cM lang sind, die wir in publizierter aDNA mittels ancIBD identifiziert haben. Jedes Quadrat stellt die Anzahl aller Paare dar, die solche IBD-Segmente aufweisen, dividiert durch alle möglichen Paare. Wenn überhaupt, hat nur ein relativ kleiner Anteil aller Paare zwischen zwei Gruppen ein langes IBD-Segment – aber diese durchschnittliche Rate, wie oft das passiert ist, ist höchst informativ. Sie ist diagnostisch für Verwandtschaften innerhalb weniger Jahrhunderte, IBD 12–16 cM müssen von gemeinsamen Vorfahren innerhalb etwa 20 Generationen stammen. Wir sehen, dass diverse Jamnajakulturen (und auch die zentralasiatische Afanassjewokultur – tausende Kilometer entfernt) einen klaren homogen Cluster bilden (beschriftet in Blau) und genauso verschiedene Schnurkeramische Kulturen, etwa die »Fatyanovo« aus Russland, »Schnurkeramik« aus verschiedenen zentralosteuropäischen Regionen und der Bootaxtkultur aus Schweden, einen weiteren Cluster bilden (beschriftet in Rot). Darüber hinaus sind zwischen den Jamnaja- und Schnurkeramik-Makro-Clustern etliche und konsistente Verbindungen ersichtlich. Interessanterweise hat die Schnurkeramikkultur zusätzlich generell starke Verbindungen zur Kugelamphorenkultur, einer Kultur aus dem Spätneolithikum, die noch keine substantielle Steppenabstammung aufweist. Da alle Schnurkeramikkulturen diese IBD-Verbindungen tragen, im klaren Gegensatz zu anderen spätneolithischen Kulturen noch ohne Steppenabstammung, impliziert dieses Signal, dass ein substantieller genetischer Beitrag der Kugelamphorenkultur »Mixture« sehr frühzeitig in der Entwicklung der Schnurkeramikkultur passiert sein muss und sich dann mit dieser Kultur ausgebreitet hat.

Fig. 4 (left page) *Rates of long IBD between different Eneolithic and Bronze Age archaeological groups. We visualize rates of IBD segments 12–16 cM long identified in published aDNA using ancIBD. Each square represents the number of all pairs that have such IBD segments; divided by all possible pairs. If anything, only a relatively small proportion of all pairs between two groups have a long IBD segment – but this average rate of how often this has happened is highly informative. It is diagnostic for kinship within a few hundred years, IBD 12–16 cM must come from a common ancestor within about 20 generations. We see that diverse Yamnaya Cultures (and also Central Asian Afanasievo Culture – thousands of kilometers away) form a clear homogeneous cluster (labeled in blue) – and also various Corded Ware cultures, e.g. »Fatyanovo« from Russia, »Corded Ware« from different Central-Eastern European regions, and »Battle Axe« from Sweden, form another cluster (labeled in red). In addition, several and consistent connections are evident between the Yamnaya and Corded Ware macro-clusters. Interestingly, Corded Ware generally has strong connections to Globular Amphora (Globular Amphora Culture), a Late Neolithic culture that still lacks substantial steppe ancestry. Since all Corded Ware cultures have these IBD connections, in clear contrast to other Late Neolithic cultures without steppe ancestry, this signal implies that substantial Globular Amphora »Mixture« must have happened very early in Corded Ware development, and then spread with that culture.*

Positives« (falsch positive IBD-Segmente) mehr werden, während die Power abfällt (manche Blöcke werden übersehen) – und zwar kontinuierlich je geringer das Coverage und je kürzer ein IBD-Segment ist. Längere Segmente sind generell einfacher zu finden als kürzere und funktionieren daher besser, auch bei weniger aDNA-Erhaltung, und je nach Fragestellungen sind auch höhere »False Positive Raten« vertretbar.

Die erste besonders spannende direkte Anwendung von ancIBD ist die Detektion von nahen Verwandten. Wir haben einen großen Datensatz von mehr als 10 000 publizierten und unpublizierten alten Genomen mit ancIBD durchleuchtet (vgl. Abb. 1a). Um nur ein aufschlussreiches Fallbeispiel herauszugreifen, haben wir nahe biologische Verwandte über mehr als 1200 km entfernt identifiziert (Abb. 3), beide stammen aus dem Äneolithikum, nördlich des Schwarzen Meeres, und sind etwas mehr als 5000 Jahre alt (basierend auf ^{14}C-Daten). Dabei handelt es sich höchstwahrscheinlich um Verwandte 5. Grades (auch 4.–6. Grad möglich), z. B. könnten es zweite Cousins (die sich zwei Urgroßeltern teilen), sein. Das zeigt eindeutig, dass zumindest ein Individuum innerhalb dieser Verwandtschaftskette mehrere hundert Kilometer zwischen Geburt und Bestattung zurückgelegt haben muss, was ein Fallbeispiel für hohe Mobilität im Äneolithikum ergibt.

Die zweite spannende Applikation sind sporadische einzelne IBD-Segmente. Ein einzelnes Paar mit einer solchen Verbindung hat noch wenig Aussagekraft, außer, dass eine biologische Verbindung in den letzten 5–50 Generationen gegeben sein muss (je länger das IBD-Segment, desto kürzer die obere Schranke). Wenn jedoch mehrere solcher Paare zwischen zwei Gruppen gegeben sind, kann man die generelle Häufigkeit solcher Verbindungen bestimmen – was, eingeteilt nach Längenkategorie, hochgradig aussagekräftig über die durchschnittliche biologische Verwandtschaft in jüngeren Zeiträumen ist. Um eine typische Anwendung zu skizzieren, haben wir wieder das westliche Eurasien im Äneolithikum und der Bronzezeit untersucht und Cluster sowie Verbindungen zwischen diversen archäologischen Kulturen identifiziert (Abb. 4; vgl. Legende).

Da das aDNA-Feld momentan rasant wächst und jedes Jahr beständig immer mehr alte Genome produziert, werden sogar noch quadratisch mehr Paare von Individuen analysierbar (in einem Datensatz der Größe n gibt es n*[n-1]/2 Paare). All diese Paare können nun mit unserer Methode ancIBD für lange IBD-Segmente gescreent werden. Damit werden rasant mehr Verwandte, auch über größere Distanzen, und Verbindungen zwischen vorgeschichtlichen Kulturen direkt messbar werden. Besonders spannend werden in naher Zukunft dann IBD-basierte demografische Analysen werden, die dann etwa Dynamiken von Populationsgrößen oder durchschnittliche individuelle Mobilität abschätzen können – auch in der Vorgeschichte. Wir hoffen daher, dass die hier vorgestellte Methode nur ein erster Schritt ist und gemeinsam mit neuen Proben und Verfahren das Fundament für eine neue Generation der aDNA-Erforschung eröffnet.

Literaturverzeichnis

Browning/Browning 2015
S. Browning/B. L. Browning, Accurate nonparametric estimation of recent effective population size from segments of Identity by Descent. Am. Journal Human Genetics 97,3, 2015, 404–418.

Fowler u. a. 2022
C. Fowler/I. Olalde/V. Cummings/I. Armit/L. Büster u. a., A high-resolution picture of kinship practices in an early Neolithic tomb. Nature 601,7894, 2022, 584–587.

Ralph/Coop 2013
P. Ralph/G. Coop, The Geography of Recent Genetic Ancestry across Europe. PLOS Biol. 11,5, 2013, e1001555, <https://doi.org/10.1371/journal.pbio.1001555> (10.07.2023).

Ringbauer u. a. 2017
H. Ringbauer/G. Coop/ N. H. Barton, Inferring Recent Demography from Isolation by Distance of Long Shared Sequence Blocks. Genetics 205,3, 2017, 1335–1351.

Ringbauer u. a. 2023
H. Ringbauer/Y. Huang/A. Akbari/S. Mallick/N. Patterson u. a., ancIBD-Screening for identity by descent segments in human ancient DNA. bioRxiv 2023, <https://doi.org/10.1101/ 2023.03.08.531671> (10.07.2023).

Rubinacci u. a. 2021
S. Rubinacci/D. M. Ribeiro/R. J. Hofmeister/O. Delaneau, Efficient phasing and imputation of low-coverage sequencing data using large reference panels. Nature Genetics 53,1, 2021, 120–126.

Wang u. a. 2019
Ch.-Ch. Wang/S. Reinhold/A. Kalmykov/A. Wissgott/G. Brandt u. a., Ancient human genome-wide data from a 3000-year interval in the Caucasus corresponds with eco-geographic regions. Nature Comm. 10, 2019, 590, <https://doi.org/10.1038/s41467-018-08220-8> (29.08.2023).

Abbildungsnachweis

1 nach Ringbauer u. a. 2023, 15 Fig. 3
2 Verf.
3 a Verf.
b A. Swieder, LDA; Kartengrundlage: GTOPO30/SRTM mit freundl. Genehmigung des U. S. Geological Survey (USGS), der National Aeronautics and Space Administration (NASA) und der National Geospatial-Intelligence Agency (NGA), public domain; Gewässer erstellt mit Natural Earth; freie Vektor- und Raster-Kartendaten @ naturalearthdata.com
4 Verf.

Anschrift

Dr. Harald Ringbauer
Max-Planck Institut für Evolutionäre Anthropologie
Deutscher Platz 6
04103 Leipzig
Deutschland
harald_ringbauer@eva.mpg.de
ORCID: https://orcid.org/0000-0002-4884-9682