Instytut Chemii Bioorganicznej
Polskiej Akademii Nauk
w Poznaniu

mgr Ireneusz Stolarek

Historia biologiczna populacji ludzkich zamieszkujących obszar współczesnej Polski w pierwszych wiekach naszej ery – interdyscyplinarne badania archeogenomiczne

Praca doktorska
wykonana pod kierunkiem
prof. dr hab. Marka Figlerowicza

Poznań, 2019

Niniejsza praca doktorska była finansowana z grantu
Narodowego Centrum Nauki SYMFONIA 2
“Dynastia i społeczeństwo państwa Piastów w świetle zintegrowanych badań
historycznych, antropologicznych i genomicznych”
nr 2014/12/W/NZ2/00466

Pragnę serdecznie podziękować mojemu Promotorowi,
Panu prof. dr hab. Markowi Figlerowiczowi
za opiekę, cenne wskazówki, liczne inspirujące dyskusje,
a także za wykazaną cierpliwość i wyrozumiałość

Dziękuję Pani dr. Luizie Handschuh
za zaangażowanie w realizowane wspólnie projekty
oraz bycie przykładem niezłomności i pracowitości

Dziękuję również wszystkim Pracownikom i Doktorantom
Europejskiego Centrum Bioinformatyki i Genomiki
za okazaną życzliwość i wszelką pomoc

Pracę dedykuję Mamie Annie

Niniejsza rozprawa doktorska składa się z następujących części:

STRESZCZENIE

W ostatnim czasie genetyka i genomika stały się nowymi siłami napędowymi badań nad przeszłością *Homo sapiens*. W konsekwencji, tworzona jest coraz bardziej szczegółowa mapa historii biologicznej człowieka. Pomimo wyraźnego postępu w rozumieniu procesów jakie zachodziły w Europie od początku procesu jej zaludniania, na mapie tej wciąż istnieją jednak liczne białe plamy. Od lat problemy związane z pochodzeniem i migracjami populacji człowieka we wczesnej epoce żelaza były i nadal pozostają przedmiotem wielu zaciekłych dyskusji. Niestety brak nowych niezależnych danych uniemożliwiał weryfikację istniejących hipotez.

Aby lepiej poznać historię biologiczną populacji człowieka żyjących w pierwszych wiekach n.e. na obszarze współczesnej Polski przeprowadziliśmy badania kopalnego DNA pobranego ze szczątków ponad 100 osób. Materiał do badań pochodził z cmentarzysk zlokalizowanych w zachodniej (populacja KOW-OVIA) oraz wschodniej (populacja MAS-VBIA) Polsce. Wbrew oczekiwaniom, obie badane grupy charakteryzowały się bardzo dużą zmiennością wewnątrzpopulacyjną, co świadczy, że nie były one małymi odizolowanymi społecznościami. Porównanie struktury genetycznej KOW-OVIA, MAS-VBIA oraz innych populacji historycznych Europy ujawniło, że obie badane grupy były blisko spokrewnione z populacją żyjącą w epoce żelaza na Półwyspie Jutlandzkim. Ponadto MAS-VBIA wykazywała bliskie związki z populacjami Stepu Pontyjskiego. Co ciekawe kobiety i mężczyźni z KOW-OVIA mieli odmienną historią genetyczną.

Dodatkowo wykazaliśmy, że analizowane próbki zawierają DNA archaicznych bakterii towarzyszących badanym osobom przed ich śmiercią. Zidentyfikowane wśród nich mikroorganizmy patogenne mogą być cennymi biomarkerami ubogacającym opis historii biologicznej populacji człowieka.

Podsumowując stwierdzić można, iż przeprowadzone badania są źródłem wielu nowych informacji o historii demograficznej populacji człowieka zamieszkujących obszar współczesnej Polski w pierwszych wiekach naszej ery. Zebrane dane dają podstawy zarówno do zweryfikowania szeregu istniejących jak i sformułowania nowych hipotez oraz narracji historycznych.

Wyniki uzyskane podczas realizacji mojej pracy doktorskiej opisane zostały w trzech pracach eksperymentalnych opublikowanych na łamach międzynarodowych czasopism znajdujących się na liście JCR.

ABSTRACT

Recently, genetics and genomics have become new drivers of research into the past of *Homo sapiens*. As a consequence, an increasingly detailed map of human biological history is being created. Despite clear progress in understanding the processes that have taken place in Europe since the beginning of its population process, there are still many white spots on this map. For years, problems related to the origin and migration of the human population in the early Iron Age have been and continue to be the subject of much fierce discussions. Unfortunately, the lack of new independent data prevented the verification of existing hypotheses.

To better understand the biological history of the human population living in the first centuries AD in the area of modern Poland, we have analyzed ancient DNA collected from the remains of over 100 people. The research material came from cemeteries located in western (population KOW-OVIA) and eastern (population MAS-VBIA) Poland. Contrary to expectations, both study groups were characterized by very high intra-population variability, which indicates that they were not small isolated communities. Comparison of the genetic structure of KOW-OVIA, MAS-VBIA and other historical populations of Europe revealed that both study groups were closely related to the population living in the Iron Age on the Jutland Peninsula. Moreover, MAS-VBIA showed close links with the Pontic Steppe populations. Interestingly, the KOW-OVIA women and men had a different genetic history.

In addition, we have shown that the analyzed samples contain DNA of archaic bacteria accompanying the subjects before their death. The pathogenic microorganisms identified among them can be valuable biomarkers enriching the description of the biological history of the human population.

To sum up, it can be stated that the conducted research is a source of much new information about the demographic history of the human population living in the area of modern Poland in the first centuries of our era. The collected data provide the basis for both verifying a number of existing and formulating new hypotheses and historical narratives.

The results obtained during the preparation of my dissertation were described in three experimental papers published in the international journals from the JCR list.

OPIS WYNIKÓW PRACY DOKTORSKIEJ

Wprowadzenie

Rozważania zaprezentowane w niniejszej rozprawie mieszczą się w głównym nurcie nowego, dynamicznie rozwijającego się, interdyscyplinarnego kierunku badawczego, wykorzystującego nowoczesne metody i techniki biologii molekularnej, jako narzędzia umożliwiające lepsze poznanie historii zarówno istniejących jak i wymarłych już gatunków roślin oraz zwierząt. Obiektem prowadzonych przeze mnie badań jest człowiek. Można więc powiedzieć, że praca ta poświęcona została historii biologicznej *Homo sapiens sapiens* (*H. sapiens s*). Cóż mieści w sobie tak zarysowany obszar badawczy? Myślę, że historię biologiczną *H. sapiens s* zdefiniować można jako dzieje gatunku od jego powstania po czasy obecne. Pierwszy człon - historia, odnosi się do dziejów, drugi - biologiczna, oznacza, że będą one rozpatrywane tylko w odniesieniu do biologii człowieka. W przypadku tym nie zdecydowałem się na użycie pokrewnego pojęcia, jakim jest historia ewolucyjna, ponieważ w pracy swojej koncentruję się na dziejach pojedynczego gatunku, nie zaś na procesach, które doprowadziły do jego powstania.

Każdy gatunek ma swoją historię biologiczną, jednak w przypadku człowieka jest ona szczególnie skomplikowana. Wynika to z faktu, że *H. sapiens s* jako jedyny gatunek zaczął opowiadać sam o sobie. Stąd też jego historia biologiczna może istotnie różnić się od klasycznie rozumianej historii. Ta ostatnia opisuje przede wszystkim przeszłość tworzonej przez człowieka cywilizacji. Z oczywistych względów dotyczy więc jedynie kilku ostatnich tysięcy lat i nie jest relacją zdarzeń, lecz subiektywnym przekazem twórcy na temat własnego dzieła.

Od kilkudziesięciu tysięcy lat człowiek jest jedynym żyjącym przedstawicielem rodzaju *Homo*. Co więcej, przez ostatnie tysiąclecia niezwykle mocno ukształtował się antropocentryczny obraz świata. Człowiek jawił się jako gatunek biologicznie osierocony, bez żadnych rodziców. Tymczasem jest to przekonanie całkowicie błędne. Wszystkie organizmy na Ziemi wywodzą się od jednego wspólnego przodka. Ewolucja stworzyła oraz ukształtowała całą obecną jak również wymarłą różnorodność życia. Podobnie jak w przypadku pojedynczego człowieka tak i w odniesieniu do gatunku, jesteśmy w stanie określić skąd pochodzi i kim są jego najbliżsi krewni.

Korzystając z zasad taksonomii biolodzy usystematyzowali różnorodność życia według hierarchicznej struktury, której podstawowymi elementami są gatunki. Przyjmuje się, że zwierzęta należą do tego samego gatunku jeśli jego przedstawiciele są w stanie wydać na świat płodne potomstwo. Gatunki, które mają wspólnego przodka są zgrupowane w jednostkę taksonomiczną zwaną rodzajem. Oba te fakty są uwzględnione w łacińskiej nazwie każdego organizmu, która w pierwszej części podaje jego rodzaj (*Homo* - człowiek), w drugiej gatunek (*sapiens* - rozumny). Rodzaje z kolei

są zebrane w rodziny. Wszyscy członkowie danej rodziny wywodzą swój rodowód od założycielskiego osobnika żeńskiego bądź męskiego. Ludzie są częścią rozbudowanej rodziny zwanej *Hominidae* (człowiekowate). Najbliżsi żyjący krewni człowieka to szympansy, goryle i orangutany. Spośród nich szympansy są najbliższe ewolucyjnie człowiekowi. Sześć milionów lat temu pojedyncza samica należąca do gatunku, z którego wyewoluował szympans i człowiek miała dwie córki, jedna została przodkiem szympansów, zaś druga jest prababką całej ludzkości. U podstaw tego podziału jednego gatunku na dwa potomne musiały leżeć zmiany w DNA.

Do niedawna jedynymi dostępnymi źródłami naszej wiedzy o zdarzeniach tworzących historię człowieka były analizy antropometryczne szczątków ludzkich oraz wszelkie pozostałe szeroko rozumiane znaleziska materialne stanowiące produkty kultur i technologii rozwijających się na przestrzeni wieków. Wadą uzyskiwanych w taki sposób informacji jest ich subiektywność, przede wszystkim, gdy nie dotyczą one bezpośrednio obiektu prowadzonych badań, a jedynie efektów jego działania. W konsekwencji powstała na ich podstawie wiedza jest spekulatywna, ma formę narracji i jest silnie uzależniona od erudycji oraz wyobraźni badacza.

Z upływem czasu kolejne dziedziny nauki zaczęły włączać się w badania przeszłości człowieka. Wprowadzając specyficzne dla siebie podejścia badawcze dostarczają nowych informacji oraz weryfikują istniejącą już wiedzę. Konsekwencją tych działań jest pełniejszy opis historii człowieka. W ostatnim czasie do grona dziedzin badających przeszłość dołączyły nauki korzystające z technik biologii molekularnej.

Archeogenomika

Pierwsze próby zastosowania technik biologii molekularnej do poznania historii biologicznej człowieka miały miejsce w latach 80' XX wieku, a więc ponad 40 lat temu. Autorem tej idei był włoski uczony Luigi Luca Cavalii-Sforza. Jego pionierskie prace skoncentrowane na analizie zmienności grup krwi w układzie AB0 zaowocowały obserwacjami dotyczącymi możliwości grupowania populacji człowieka w skali kontynentów[1]. Prawdziwa rewolucja w tego rodzaju badaniach nastąpił na przełomie wieków, wraz z rozwojem technik sekwencjonowania DNA. Pierwsze analizy genetyczne współczesnych populacji człowieka umożliwiały podzielenie *H. sapiens s* na 5 grup - Zachodni Euroazjaci, Wschodni Azjaci, rdzenni Amerykanie, Nowogwinejczycy oraz Afrykanie. Takie wykorzystanie zdobyczy biologii molekularnej do opisywania ludzkiej przeszłości określane jest terminem: archeogenetyka. Termin ten, funkcjonuje w literaturze także pod nazwą „paleogenetyka".

Bez wątpienia znajomość sekwencji genów oraz zasad dziedziczenia jest źródłem bezcennych danych, dotyczących naszych korzeni, związków między populacjami czy procesów ich migracji. Istnieje wiele spektakularnych przykładów, ukazujących, w jaki sposób archeogenetyka przyczyniła się do zweryfikowania hipotez

i poglądów odnoszących się do naszej przeszłości[2-4]. Obserwowany w ostatniej dekadzie dalszy rozwój technik izolacji, wzbogacania i sekwencjonowania DNA przyczynił się do znacznego zwiększenia skali analiz, zarówno pod względem ilości badanych osobników, jak i generowanych danych. W rezultacie archeogenetyka stopniowo zaczęła przekształcać się w archeogenomikę.

Cechą charakterystyczną badań łączących tradycyjnie rozumianą historią oraz historią biologiczną człowieka, jest wyjątkowa interdyscyplinarność. Efektywne funkcjonowanie w obrębie tak szeroko zarysowanego obszaru badawczego wymaga utworzenia dużych zespołów obejmujących swoją kompetencją takie dziedziny wiedzy, jak: historia, archeologia, antropologia, biologia molekularna, genomika i bioinformatyka. Ponadto poszczególni członkowie zespołu badawczego, oprócz tego, że są specjalistami w swojej dziedzinie, powinni posiadać minimum wiedzy niezbędnej do zrozumienia podstawowych zasad i reguł funkcjonowania pozostałych dziedzin. Jej brak prowadzić może do niewłaściwej interpretacji wyników uzyskanych w ramach innej dziedziny lub nadmiernych uproszczeń. Z drugiej strony pomijanie lub marginalizowanie ustaleń poczynionych w ramach, którejś z wymienionych wcześniej dziedzin prowadzi do niepełnego opisu przeszłości oraz powielania starych lub propagowania nowych błędnych opinii i hipotez. Zanim zatem przejdę do omówienia najnowszych osiągnięć archeogenomiki, chciałbym wprowadzić kilku podstawowych pojęć ułatwiających zrozumienie przedstawionych poniżej problemów.

Podstawowe pojęcia

Podstawowym obiektem zainteresowań archeogenomiki jest genom. W przypadku wszystkich organizmów eukariotycznych, w tym człowieka, termin genom odnosi się do całkowitego DNA zgromadzonego w jądrze komórkowym (tzw. genom jądrowy) oraz w mitochondriach (tzw. genom mitochondrialny, mtDNA). Genom jądrowy człowieka składa się z ok. 3,3 mld par nukleotydów zorganizowanych w struktury zwane chromosomami (22 diploidalne autosomy oraz 2 allosomy, czyli chromosomy płci X i Y). Występujący w każdej komórce, obok genomu jądrowego, zdecydowanie mniejszy genom mitochondrialny, składa się z ok. 16,5 tys. par nukleotydów[5].

Liczne zjawiska i procesy składające się na historię biologiczną gatunku nabierają nowego znaczenia i stają się bardziej zrozumiałe, gdy rozpatrywane są w kontekście mechanizmów ewolucji, przez pryzmat sił determinujących wewnątrz- i międzypopulacyjną różnorodność. Przyjmuje się, że głównymi czynnikami, które kształtują pulę genetyczną populacji są mutacje, dobór naturalny, dryf genetyczny oraz migracje[7].

Mutacje są wprowadzane do DNA w trakcie jego replikacji. Podczas podziału komórkowego następuje synteza dodatkowej kopii genomu, tak, by po podziale obie

komórki posiadały taką samą pulę DNA. Zazwyczaj jednak proces ten nie prowadzi do powstania dwóch identycznych zestawów chromosomów, lecz różniących się kilkoma nukleotydami. Dlatego też genomy poszczególnych komórek organizmu są od siebie nieco różne. Jeśli mutacje powstaną w genomie komórek rozrodczych (germinalnych), zostaną one przekazane kolejnym pokoleniom. Istnieje wiele form mutacji DNA. Najpowszechniejsze są substytucje pojedynczych nukleotydów. Jeśli w określonym miejscu genomu jakaś substytucja występuje w populacji z pewną arbitralnie wyznaczoną częstością (> 1%), mówimy wówczas o zjawisku zwanym polimorfizmem pojedynczego nukleotydu (SNP – ang. *single nucleotide polymorphism*). Średnie tempo wprowadzania mutacji do chromosomów autosomalnych (oraz dla chromosomu X), chromosomu Y i mtDNA jest różne i wynosi odpowiednio ok. 1-1,5 $\times 10^{-8}$, 3 $\times 10^{-8}$, 2,7-3 $\times 10^{-5}$ mutacji na jeden nukleotyd na jedno pokolenie[8-10]. Analizując proces pojawiania się nowych wariantów SNP, można zatem odtworzyć chronologię zdarzeń ewolucyjnych.

MtDNA występuje w każdej komórce w wielu kopiach. Dzięki temu, w odróżnieniu od DNA jądrowego, znacznie łatwiej pozyskać go z kopalnych szczątków. Co więcej mtDNA jest relatywnie krótki oraz nie podlega rekombinacji. W przypadku człowieka ważną cechą mtDNA jest specyficzny sposób jego dziedziczenia. Z pokolenia na pokolenie przekazywane są jedynie mitochondria żeńskich komórek płciowych. Oznacza to, że mtDNA zawsze pochodzi od matki. Z uwagi na powyższe możliwa jest analiza linii żeńskich składających się na daną populację. Podobnie genetycznie rodowód można badać w linii męskiej analizując mutacje występujące na nierekombinującym odcinku chromosomu Y.

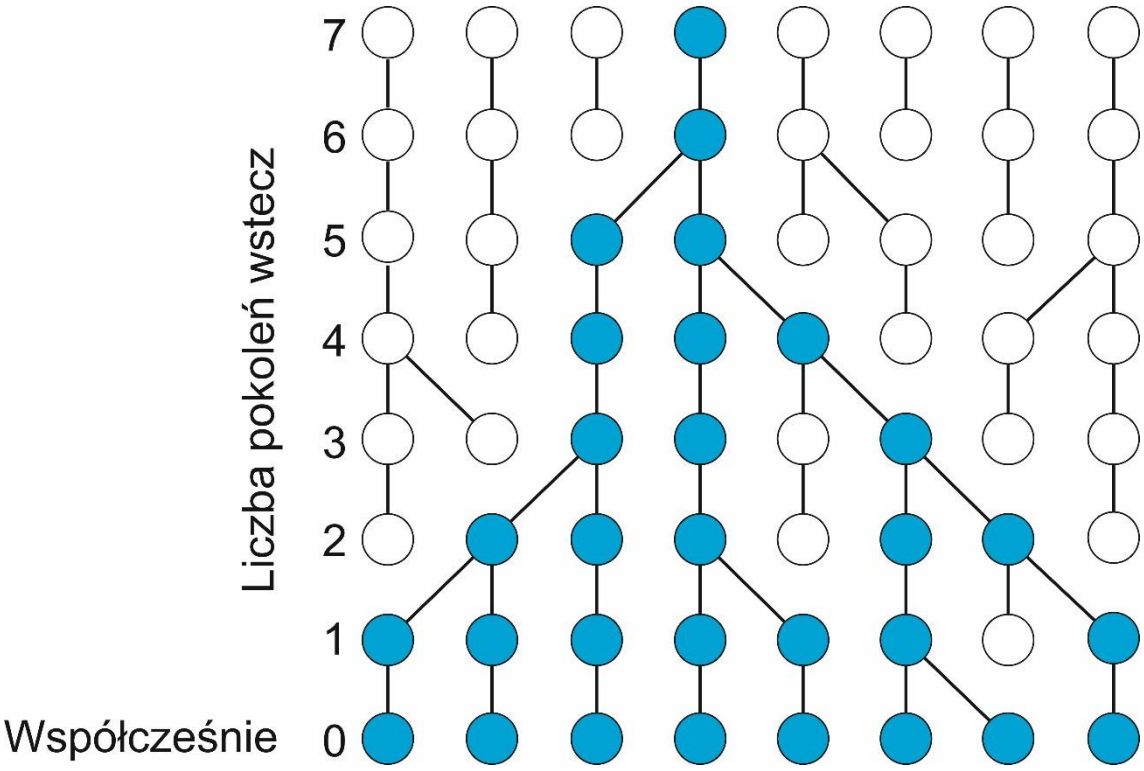
Oba elementy genomu nieulegające rekombinacji, czyli mtDNA oraz odcinek chromosomu Y są tak zwanymi haplotypami, czyli fragmentami genomu, dziedziczonymi w całości od jednego z rodziców. Występujące w obrębie tych sekwencji specyficzne wzory mutacji (SNP), są więc charakterystyczne dla poszczególnych linii żeńskich (mtDNA) oraz męskich (chromosom Y). Z tego względu zestaw SNP jest elementem jednoznacznie określającym dany haplotyp. Należy zaznaczyć, że oba te haplotypy znacznie różnią się od siebie długością. Haplotyp mtDNA ma długość ok, 16,5 tys. nukleotydów, przy czym najczęściej jest określany na podstawie odcinka 1,200 nukleotydów, zaś haplotyp chromosomu Y ma 30,000,000 nukleotydów.

Jedną z głównych koncepcji funkcjonujących w genetyce ewolucyjnej jest teoria koalescencji. Zakłada ona, że każdy allel (gen lub fragment DNA) wywodzi się od jednego wspólnego przodka (Rycina 1)[11]. Teoria koalescencji odnosi się do neutralnych lub prawie neutralnych alleli, występujących w populacji o stałej liczebności, w której kojarzenie par ma charakter losowy. Korzystając z zasad koalescencji możemy połączyć haplotypy, które mają wspólnego przodka w szersze zestawy zwane haplogrupami. Możliwym jest także wyznaczenie czasu powstania danej mutacji, a

więc i określenie wieku całej haplogrupy. Z uwagi na powyższe od wielu lat mutacje występujące w mtDNA i chromosomie Y stanowią główne markery wykorzystywane w genetyce populacyjnej i antropologii molekularnej.

Teoria koalescencji zakłada, że w przypadku każdego haplotypu w przeszłości musiał istnieć jeden wspólny przodek, który dał początek wszystkim występującym obecnie jego wariantom. Istniała zatem kiedyś na Ziemi mitochondrialna Ewa, która dała początek wszystkim współcześnie występującym mtDNA. Analogicznie, istniał kiedyś Y-chromosomowy Adam, który dał początek wszystkim współcześnie obserwowanym haplotypom chromosomu Y.

Kobiety współczesne mitochondrialnej Ewie, a także te żyjące przed nią, miały inny mtDNA. Z biegiem czasu te żeńskie rodowody wymarły i nie jest możliwe zaobserwowanie ich haplotypów we współczesnych populacjach. Niemniej jednak, nie możemy wykluczyć takiej możliwości, że obecnie żyje lub żył w przeszłości człowiek, który posiada/posiadał mtDNA jakiego jeszcze nigdy wcześniej nie zaobserwowaliśmy. Na podstawie zebranych dotychczas danych ustalono, że mitochondrialna Ewa żyła około 160 tysięcy lat temu (tya, ang. thousand years ago) w Afryce[12]. Y-chromosomowy Adam datowany jest na okres około 160-320 tya i żył prawdopodobnie w innym rejonie Afryki.



Rycina 1 Diagram obrazujący proces koalescencji alleli w populacji o stałej wielkości. k-alleli obserwowanych współcześnie (8) przechodzi serię zdarzeń koalescencyjnych. Teoretycznie proces obejmuje k-1 (7) populacji i prowadzi do wyłonienia tylko jednego allelu, który dał początek wszystkim obecnie istniejącym.

Często popełnianym błędem jest domniemanie, iż wspomniane osoby były pramatką i praojcem ludzkości. Tymczasem dały one początek jedynie rodowodom mtDNA oraz chromosomu Y, a zatem jedynie 2 haplotypom. W genomie jądrowym wyróżnić można tysiące analogicznych haplotypów, co oznacza, że zawiera on w sobie informacje o bardzo licznej grupie przodków. Typowy genom człowieka składa się z 47 łańcuchów dwuniciowego DNA odpowiadających 46 chromosomom oraz mtDNA (Rycina 2). Zakładając stałe tempo rekombinacji genomu jądrowego, (ok. 71 zdarzeń *crossing-over* podczas powstawania haploidalnych komórek rozrodczych) stwierdzić można, że w przypadku każdego osobnika 47 dwuniciowych DNA wyselekcjonowanych zostaje spośród 164 potencjalnych rodzicielskich wariantów: 46 niezrekombinowanych chromosomów od matki, 46 niezrekombinowanych chromosomów od ojca, 71 zrekombinowanych chromosomów pochodzących od matki lub ojca oraz mtDNA pochodzącego zawsze od matki. Analogicznie dwa pokolenia wstecz liczba potencjalnych wariantów dwuniciowego DNA, które mogą być odziedziczone przez daną osobę od 4 swoich dziadków wynosi już 327. Z oczywistych względów większość z potencjalnych wariantów jest zatem tracona z pokolenia na pokolenie. Z uwagi na powyższy fakt, materiał genetyczny niektórych przodków musi być w kolejnych pokoleniach całkowicie utracony. W przypadku 20 pokoleń ilość przodków jest już niemal tysiąc razy większa niż ilość odziedziczonych wariantów, jest zatem pewne, że materiał genetyczny dużej ich części nie występuje w ostatnim pokoleniu. Z tego powodu poszukiwanie wśród swoich antenatów dawno temu zmarłych historycznych postaci nie ma zwykle większego uzasadnienia, ponieważ prawdopodobieństwo, że ich DNA przetrwał przez dużą liczbę pokoleń jest astronomicznie małe.

Rycina 2. Diagram obrazujący dziedziczenie haplotypów. Cofając się w czasie liczba przodków, których posiada dana, osoba ulega podwojeniu z każdym pokoleniem. Jednak, liczba fragmentów DNA, które składają się na genom danej osoby przyrasta jedynie o 71 na pokolenie. Oznacza to, że cofając się o osiem lub więcej pokoleń, jest niemal pewnym, że dana osoba będzie mieć kilku przodków, od których nie odziedziczyła żadnego DNA.

Kopalny DNA

Dopiero w 2008 roku, Novembre i wsp. zademonstrowali, że badania DNA współczesnych ludzi nie pozwalają odtworzyć historii biologicznej całego gatunku.[13] Problem ten można jednak rozwiązać badając kopalny DNA (aDNA, ang. *ancient DNA*), czyli DNA pozyskany ze szczątków szkieletowych osób zmarłych. Pierwsze tego typu badania przeprowadzono w roku 1988. Analizie poddano DNA wyizolowany ze szczątków ludzkich datowanych na około 7 tya i pochodzących z Florydy[14]. Stwierdzono, że zidentyfikowany haplotyp mtDNA nie występuje współcześnie w puli genowej rdzennych Amerykanów, jest natomiast obecny u współczesnych Europejczyków, choć stosunkowo rzadko. Eksperyment ten dowiódł, że badania genetyczne kopalnych próbek stwarzają unikatową szansę na poznanie historii biologicznej populacji człowieka.

Badania aDNA nie są jednak łatwe, gdyż różni się on pod wieloma względami od DNA wyizolowanego z materiału pobranego od żywego osobnika. Po śmierci organizmu struktury komórkowe ulegają dezintegracji, a materiał genetyczny stopniowej dekompozycji[15]. Procesy degradacji DNA postępują najszybciej w

pierwszym roku po śmierci[16] głównie w wyniku dwóch procesów: fragmentacji łańcuchów DNA oraz modyfikacji tworzących je nukleotydów[17]. Dodatkowo szczątki są zasiedlane przez mikroby. Mogą one pochodzić zarówno z mikrobiomu badanej osoby jak i ze środowiska, w którym znalazły się szczątki. W konsekwencji izolowany aDNA jest pofragmentowany, zanieczyszczony obcym materiałem genetycznym oraz zawiera wiele zmodyfikowanych/uszkodzonych nukleotydów.

Fragmentacja łańcucha DNA spowodowana jest aktywnością nukleaz wyzwalanych w wyniku rozpadu struktur komórkowych lub wydzielanych przez zasiedlające szczątki mikroorganizmy. DNA jest także narażony na nieenzymatyczną hydrolizę, do której najczęściej dochodzi w miejscach depurynacji[15]. Fakt ten doskonale uwidaczniają odczyty uzyskane metodą NGS. Wynika z nich, że pierwszym nukleotydem, który w genomie referencyjnym poprzedza zmapowany koniec 5' odczytu jest zwykle puryna[18]. Konsekwencją fragmentacji jest powstanie odcinków DNA o długości od kilkudziesięciu do kilkuset nukleotydów.

Kolejną cechą aDNA jest specyficzna akumulacja modyfikacji, które mogą blokować jego amplifikację metodą PCR (ang. *blocking lesions*) lub prowadzić do błędnego wprowadzania nukleotydów podczas przygotowywania bibliotek sekwencyjnych (ang. *miscoding lesions*). Wykazano na przykład, że polimerazę blokują produkty oksydacji pirymidyn takie jak 5-hydroksy-5-metylohydantoiny oraz 5-hydroksyhydantoiny oraz wiązania krzyżowe (ang. *cross-links*) tworzące się pomiędzy nićmi DNA lub między DNA a innymi molekułami, np. białkami[17]. Błędne wprowadzenie nukleotydów nie wstrzymuje reakcji PCR, ale zmienia sekwencję nici syntetyzowanej *de novo* podczas namnażania matrycowego aDNA. Modyfikacją jakiej najczęściej ulegają nukleotydy w łańcuchu DNA jest deaminacja zmieniająca cytozynę na uracyl[15,19] (około 90% modyfikacji). Dochodzi do niej zwykle w obrębie końcowych jednoniciowych fragmentów aDNA (ang. *single-stranded overhangs*). Konsekwencją zmiany C>U, klasyfikowanej jako tranzycja typu 2 jest substytucja G>A w czasie pierwszej rundy amplifikacji matrycowego aDNA oraz substytucja C>T w drugiej rundzie amplifikacji aDNA metodą PCR. W porównaniu z modyfikacjami cytozyny, deaminacja puryn jest marginalna - konwersja adeniny do hipoksantyny, która paruje się z cytozyną zamiast tyminą, klasyfikowana jest jako tranzycja typu 1 (ok. 2-6% modyfikacji)[20,21]. Konwersja guaniny do ksantyny odbywa się na podobnym lub nawet niższym poziomie[15], natomiast tymina nie podlega deaminacji[22].

Archeogenomika w badaniach historii biologicznej człowieka

W ostatnich latach obserwujemy znaczący rozwój badań archeogenomicznych. Pierwszym czynnikiem, który miał wpływ na ów postęp były obserwacje, iż pewne fragmenty szczątków szkieletowych zawierają więcej endogennego aDNA niż inne. Początkowo DNA izolowano z kości długich lub z zębów. Obecnie, o ile to możliwe, DNA izoluje się z kości skalistej osłaniającej ucho wewnętrzne. Ilość DNA zachowana

w kości skalistej może być nawet 100 razy większa niż w zębach - drugim najczęściej wykorzystywanym w archeogenomice materiale kostnym[25].

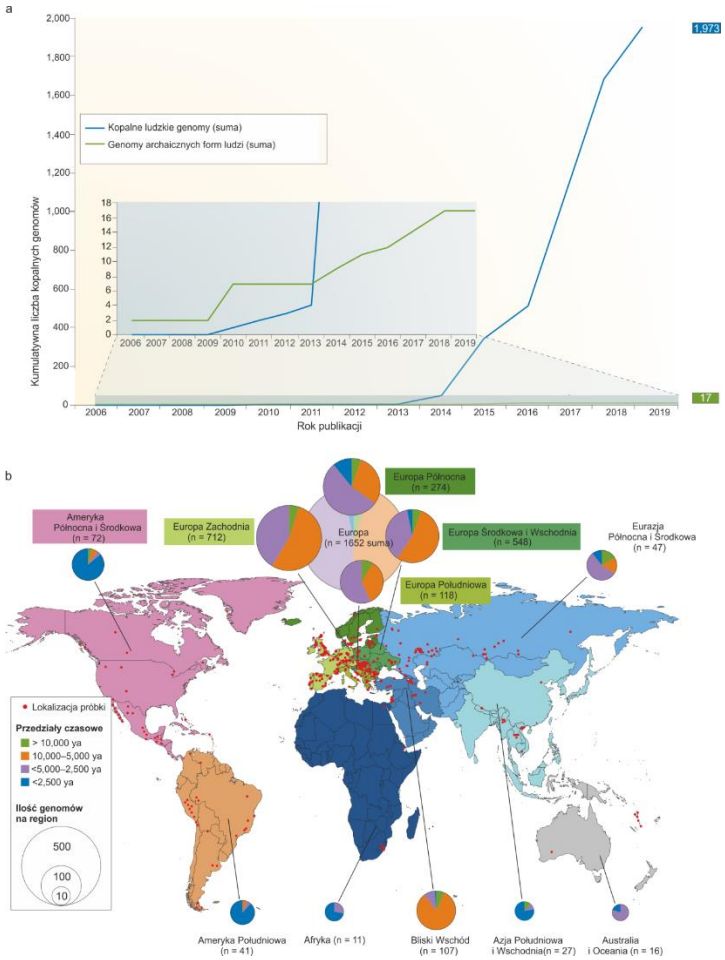
Kolejnym istotnym czynnikiem przyspieszającym rozwój archeogenomiki jest powstawanie nowych metod izolacji DNA. Szczególnie cenną jest wzbogacanie próbki w endogenny DNA (ang. *DNA capture*)[26]. Wykorzystuje się w niej biotynylowane sondy RNA, które są komplementarne do określonych fragmentów genomu. Sondy te hybrydyzują z wybranymi odcinkami matrycowego DNA, a następnie są odławiane za pomocą kulek magnetycznych pokrytych streptawidyną. Obecnie metodę tą najczęściej używa się do wzbogacania próbki we: (i) fragmenty pochodzące z genomu mitochondrialnego; (ii) fragmenty zawierające pozycje polimorficzne na chromosomie Y oraz (iii) wybrane fragmenty lub cały genom jądrowy.

Następnym szczególnie ważnym czynnikiem wpływającym na rozwój archeogenomiki było wprowadzenie NGS. Po raz pierwszy wykorzystano NGS w archeogenomice w roku 2006. Przeprowadzony wówczas eksperyment umożliwił poznanie fragmentu genomu Neandertalczyka o łącznej długości 1 mln nukleotydów[23]. Obecnie NGS oraz metody wzbogacania w endogenny aDNA są rutynowo wykorzystywane do ustalania pełnych sekwencji genomu jądrowego i mitochondrialnego wielu kopalnych osobników.

Od czasu pierwszego zastosowania NGS w archeogenomice wraz z rozwojem technologii i obniżeniem kosztów sekwencjonowania ilość danych genomicznych uzyskanych dla kopalnych szczątków szybko rośnie (Rycina 3). Obecnie znamy już sekwencje kopalnych genomów dla setek ludzi. Większość prac badawczych dotyczy szczątków pochodzących z regionu Eurazji[2,4,6,27], co podyktowane jest zarówno znacznie lepszym finansowaniem tamtejszych ośrodków naukowych jak również sprzyjającymi warunkami klimatycznymi umożliwiającymi zachowania aDNA. Znacząca ilość przebadanych szczątków szkieletowych datowana jest na czasy neolityczne (9 - 5 tya) oraz mezolityczne (14-10 tya). Coraz częściej badane szczątki pochodzą spoza Euroazji oraz z późniejszych okresów, takich jak epoki metali czy średniowiecze[28-33]. Co więcej w ramach badań historii biologicznej człowieka zaczęto wykorzystywać materiał genetyczny mikrobów towarzyszących kopalnym szczątkom ludzkim[34,35]. Rekonstrukcja sekwencji DNA wybranych patogennych bakterii umożliwia prześledzenie koewolucji człowieka i bakterii z nim związanych. Ponadto informacje dotyczące zmienności czynników wirulencji patogenów na przestrzeni wieków są cennym elementem badań epidemiologicznych. Ostatnie innowacje na polu archeogenomiki dotyczą badań kopalnych epigenomów oraz aDNA pochodzącego z osadów glebowych pobranych w miejscach, gdzie odkryto ślady bytowania przedstawicieli rodzaju *Homo*.

W ostatnim czasie zauważalny jest również znaczny postęp w zakresie podejść umożliwiających integrację danych pochodzących z badań historycznych, archeologicznych, antropologicznych i genetycznych. Stąd w procesie recenzowania manuskryptów dotyczących historii biologicznej coraz mocniej akcentowana jest

potrzeba całościowego opisu dyskutowanych problemów badawczych. W konsekwencji nierzadko obok wyznaczonych przez edytora recenzentów będących ekspertami z zakresu genetyki wybierane są także osoby, odpowiedzialne za zweryfikowanie poprawności opisu tła historycznego i archeologicznego. Dodatkowo w coraz liczniejszych publikacjach poszczególne stanowiska archeologiczne są charakteryzowane na podstawie dużej liczby przebadanych szczątków szkieletowych. Stan taki pozwala przejść z poziomu ogólnych rozważań o minionych przemianach demograficznych do badań skupionych wokół zagadnień związanych z czynnikami socjo-ekonomicznymi kształtującymi dawne społeczeństwa a nawet powiązaniami rodzinnymi pomiędzy badanymi osobami.



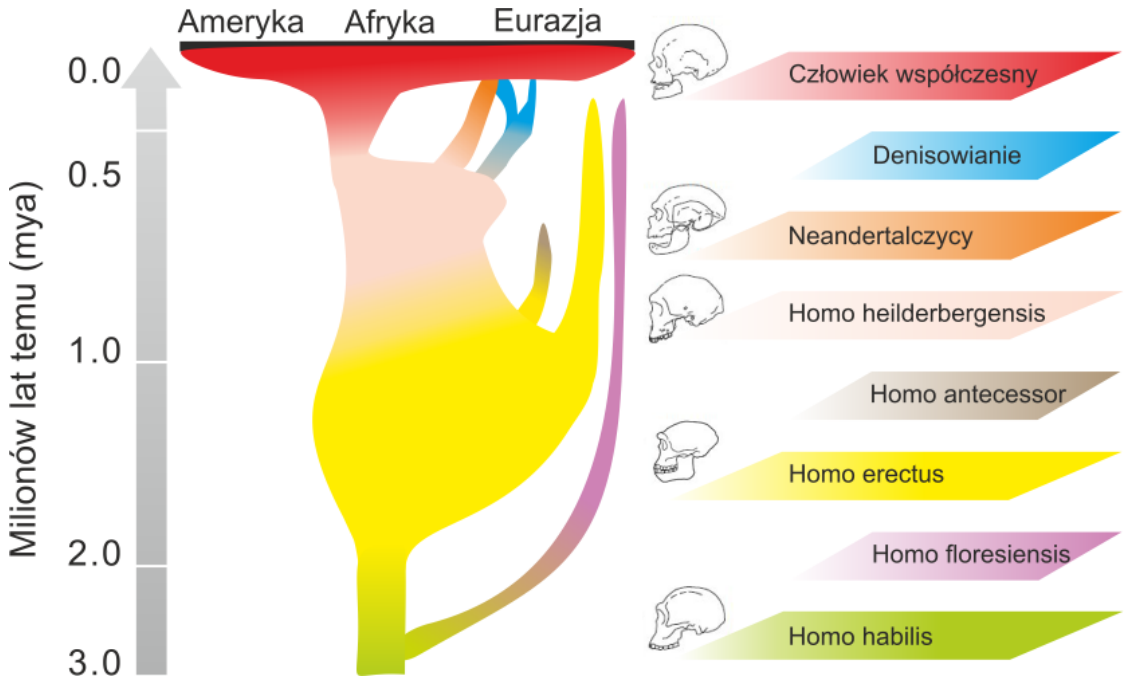
Rycina 2. Diagram obrazujący liczbę oraz rozmieszczenie w czasie i przestrzeni kopalnych szczątków ludzkich poddanych analizom genetycznym: a) Diagram obrazujący jak w ostatnich latach przyrastała liczba zbadanych pod względem genetycznym/genomicznym kopalnych szczątków współczesnych i archaicznych form człowieka; b) rozmieszczenie na mapie świata zbadanych szczątków anatomicznie współczesnych ludzi; dodatkowo kolorami oznaczono okres, z którego pochodziły badane szczątki.

Powstanie i pierwsze wędrówki *H. sapiens s*

Na ewolucję *H. sapiens s* składają się wszystkie procesy prowadzące do ukształtowania się anatomicznie współczesnego człowieka (AMH, ang. *Anatomically Modern Human*) jako jednego z przedstawicieli rzędu naczelnych. Do wydarzenia tego doszło najprawdopodobniej około 300 tya. Ważnym momentem na tej drodze było powstanie podrodziny *Homininae* (wszystkie człowiekowate za wyjątkiem orangutanów) liczącej sobie blisko 14 milionów lat. Nie ma wątpliwości, że kolebką wszystkich *Homininae* była Afryka, w której około 2-3 milionów lat temu pojawili się pierwsi przedstawiciele rodzaju *Homo*.

Rekonstrukcja relacji taksonomicznych pomiędzy przedstawicielami Homo oraz powiązanie poszczególnych etapów ich ewolucji z konkretnymi gatunkami jest niezwykle trudnym zadaniem. Ponieważ szczątki szkieletowe są często niekompletne, próbuje się tego dokonać na podstawie zaledwie kilku fragmentów kości. Bez znajomości sekwencji DNA wymarłych form *Homo* wszelkie rozważania paleoantropologiczne są zatem wielce spekulatywne, nawet jeśli dokonywane na dobrze zachowanych szkieletach. Dodatkowo, na podstawie zebranych dotąd informacji sądzić można, że powszechnym zjawiskiem było istnienie wielu współbieżnych sobie form *Homo*. Tym samym klasyczne pojmowanie procesu powstawania gatunku, jako odrębnej gałęzi drzewa ewolucyjnego wydaje się być błędne i powinno być zastąpione innym wyobrażeniem zakładającym, że relacje między gatunkami, zwłaszcza we wczesnych etapach ich specjacji, mają bardziej złożony charakter.

W skład podrodziny *Homininae* wchodzą wszystkie jej współczesne gatunki jak również wymarłe formy, z których najważniejszą dla ewolucji *H. sapiens s* był *Australopithecus* (Australopitek, w dosłownym tłumaczeniu małpa południowa). Jeden z jego gatunków blisko 3 miliony lat temu dał początek całemu rodzajowi *Homo* (Rycina 4). Obecnie najstarszym znanym przedstawicielem tego rodzaju jest *Homo habilis*, człowiek zręczny (2,100 - 1,500 tya). *Homo erectus* (1,900 - 140 tya), kolejny reprezentant rodzaju *Homo*, najprawdopodobniej jako pierwszy opanował sztukę posługiwania się ogniem oraz rozpoczął pierwsze wędrówki rodzaju *Homo* poza Afrykę. Na uwagę zasługuje fakt, że *Homo erectus* był najdłużej utrzymującą się formą *Homo* i przetrwał niemal 2 miliony lat. Następnym kamieniem milowym w ewolucji człowieka było pojawienie się w Afryce *Homo heidelbergensis* (600 - 300 tya), prawdopodobnego przodka *H. sapiens s* oraz anatomicznie różnych, archaicznych form *Homo sapiens* - neandertalczyków (430 - 40 tya) i denisowian (~40 tya).



Rycina 3. Drzewo genealogiczne przedstawiające ewolucję hominidów. Kolejne formy *Homo* są oznaczone kolorami. Diagram przedstawia ewolucję rodzaju *Homo* w czasie i przestrzeni (Ameryka, Afryka, Eurazja). Czas wyrażono w milionach lat temu (mya). Diagram ma charakter poglądowy i nie uwzględnia wszystkich, niekiedy kontrowersyjnych, poglądów i tez.

Do niedawna powszechnie przyjmowało się, że początków *H. sapiens s* należy doszukiwać się w Afryce Wschodniej. Tymczasem ostatnie znalezisko najstarszych szczątków szkieletowych przypisanych *H. sapiens s* pochodzi z Afryki Północnej, Jebel Irhoud, Maroko (~300 tya)[40]. Tym samym, pojedyncze odkrycie przesunęło szacowany czas powstania *H. sapiens s* o niemal 100 tysięcy lat, a także wymusiło zweryfikowanie hipotez dotyczących umiejscowienia kolebki ludzkości na mapie Afryki. Kolejne najstarsze znaleziska szkieletowe naszego gatunku pochodzą z Afryki Wschodniej i są młodsze o blisko 100 tysięcy lat: Omo Kibish (~195 tya) i Herto (~160 tya) w Etiopii oraz Ngaloba w Tanzanii (~120 tya). Rezultaty datowania szczątków wskazują, że w czasie, gdy w Afryce rozwijał się wczesny *Homo sapiens*, w jej południowej części żyła inna archaiczna forma *Homo naledi* (236 - 335 tya), zaś w Indonezji niezwykle długo przetrwał *Homo floresiensis* (190 - 50 tya). Z uwagi na fakt, że wiele z archaicznych form człowieka współistniało ze sobą wydaje się, że wszystkie przeszłe formy człowieka były częścią dużej, krzyżującej się ze sobą populacji. Zjawisko to w świetle najnowszych badań archeogenomicznych jawi się jako niepodważalny fakt[36-38].

Jedynym sposobem ustaleniu relacji między poszczególnymi gatunkami i formami rodzaju *Homo* są badania aDNA. Bazują one na koncepcji zegara molekularnego oraz teorii koalescencji. Jak dowodzą najnowsze ustalenia ostatni wspólny przodek rodzajów *Homo* i *Pan* (szympans) datowany jest na ok. 7,000 - 5,000 tya. Na podstawie analizy genomu jądrowego współczesnych ludzi, przewiduje się, że ostatni wspólny przodek wszystkich *H. sapiens s* żył około 320 tya, podczas gdy

ostatni wspólny przodek człowieka współczesnego w linii matczynej około 160 tya (tzw. Mitochondrialna Ewa) a ojcowskiej 160 - 300 tya (tzw. Y-chromosomowy Adam). Zgodnie z oczekiwaniami wszystkie najstarsze haplogrupy markerów uniparentalnych (haplogrupa mtDNA L i haplogrupa Y A00) występowały u ludzi pochodzących z Afryki.

Powstanie i pierwszy okres istnienia AMH, opisać można za pomocą 4 możliwych scenariuszy. Pierwszym z nich jest „nowa hipoteza multiregionalna", według której kilka populacji AMH wyewoluowało niezależnie od siebie w różnych częściach Afryki. Drugi scenariusz zakłada jedno miejsce powstania AMH z późniejszą ekspansją tej populacji i ekstynkcją wszystkich pozostałych form nie-AMH. W trzecim modelu również proponuje się jedno miejsce powstania AMH, jednak z założeniem, że późniejszej ekspansji pierwotnej populacji towarzyszyło przetrwanie pewnych form nie-AMH. Ostatni zaproponowany scenariusz zakłada powstanie AMH w pojedynczej lokacji w Afryce Południowej a następnie jego migrację oraz krzyżowanie się z archaicznymi formami homininów w Afryce Północnej.

Również czas, szlaki oraz liczba wyjść AMH z Afryki są przedmiotem licznych dyskusji. Obecnie najszerzej wspierana jest hipoteza zakładająca dwie migracje AMH. Niedawne znalezisko szczątków AMH z Izraela (Misliya), datowanych na 200 tya wskazuje, że AMH opuścili Afrykę po raz pierwszy ponad 40 tysięcy lat wcześniej niż uprzednio przypuszczano na podstawie znalezisk ze Skhul oraz Quafzeh (czyli około 160 tya). Wydaje się, że pierwsza migracja AMH poza Afrykę wiodła przez Lewant. Lewant był też miejscem pierwszych kontaktów między AMH a neandertalczykami. Według obecnego stanu wiedzy neandertalczycy wyparli z regionu Lewantu pierwszych migrujących z Afryki AMH. Jednakże w kontekście odnalezienia w Chinach szczątków AMH, datowanych na 80 tya oraz w Arabii datowanych na 85 tya sądzić można, że określenie „wyjście z Afryki" niekoniecznie oznaczać musi pojedyncze wydarzanie. Był to raczej proces ciągły trwający dziesiątki tysięcy lat. Ponadto wspomniane powyżej znaleziska dowodzą, że początkowa ekspansja AMH nie ograniczała się do Bliskiego Wschodu.

Drugie wyjście AMH z Afryki jest zdarzeniem odrębnym od tych omówionych powyżej. Inaczej niż w przypadku pierwszego wyjścia z Afryki, drugie doprowadziło do ekstynkcji wszystkich wcześniejszych archaicznych form człowieka, tj. neandertalczyków, denisowian oraz *Homo floresiensis*. Na podstawie analiz czasu koalescencji wyznaczonych dla wybranych fragmentów genomów kopalnych osobników paleolitycznych łowców-zbieraczy ustalono, że drugie wyjście z Afryki miało miejsce około 70 tya.

Domieszki materiału genetycznego archaicznych form *Homo sapiens* w genomie AMH

Od niemal 40 tya człowiek współczesny jest jedynym żyjącym przedstawicielem rodzaju *Homo*. Jednak jeszcze 50 tya Euroazja była zamieszkana przez wiele grup archaicznych form człowieka. Najlepiej poznaną spośród nich są neandertalczycy. Populacja ta istniała między 430-39 tya i zajmowała obszar obejmujący Europę oraz część Azji (od Uzbekistan na południu po góry Ałtaj, na wschodzie). W rejonie gór Ałtaj przez tysiąclecia neandertalczycy współistnieli z nowo odkrytą formą archaicznych ludzi - denisowianami, którzy zamieszkiwali teren od gór Ałtaj aż po Oceanię. Badania archeogenomiczne dostarczyły w ostatnim czasie wielu bezcennych informacji na temat ewolucji i historii biologicznej AMH oraz obu wspomnianych archaicznych form człowieka.

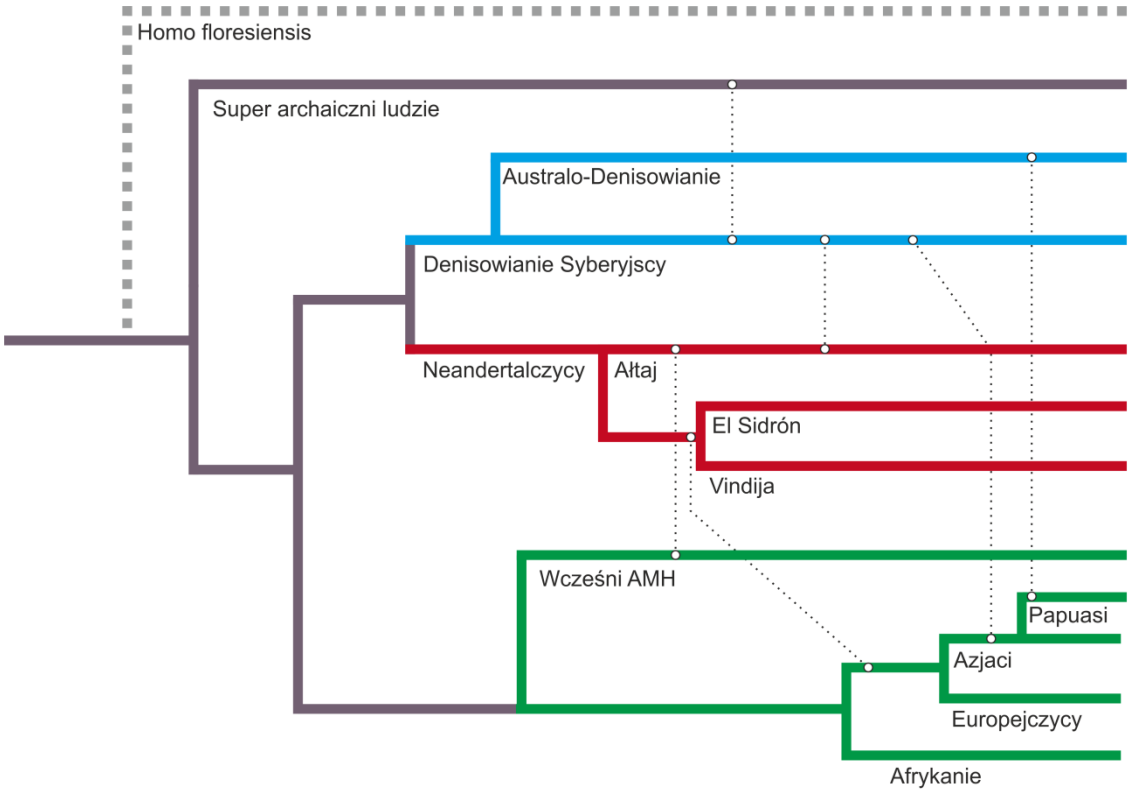
Analizując ilość mutacji w DNA jądrowym różniącą neandertalczyków, denisowian i współczesnych ludzi oraz wyznaczając tempo ich powstawania ustalono, że wspólny przodek AMH, neandertalczyków i denisowian żył ok. 770 - 550 tya. Zgoła odmiennych wyników dostarcza porównanie sekwencji mtDNA, gdzie linie matczyne współczesnych ludzi i neandertalczyków uległy separacji od siebie ok. 470 - 360 tya. Również relacje między neandertalczykami i denisowianami są odmienne w zależności od tego czy w analizach korzystamy z genomów jądrowych czy mtDNA. Na podstawie DNA jądrowego ustalono, że ostatni wspólny przodek obu archaicznych form człowieka żył ok. 470 - 380 tya. Tymczasem badania mtDNA sugerują, że linia matczyna denisowian oddzieliła się ok od tej prowadzącej do neandertalczyków i współcześnie żyjących ludzi około 1,000 - 800 tya. Należy jednak zaznaczyć, że obserwowane rozbieżności w opisie relacji ewolucyjnych pomiędzy trzema formami człowieka nie świadczą o wystąpieniu jakiejś istotnej wewnętrznej sprzeczności w proponowanym opisie dziejów archaicznych form człowieka. Wynikają one raczej z tego, że czasy koalescencji wyznaczone dla poszczególnych fragmentów tego samego genomu mogą być znacząco różne.

Stwierdzono ponadto, że mtDNA wczesnego neandertalczyka był bardziej podobny do mtDNA denisowian, niż do mtDNA późniejszych form neandertalczyka żyjących w Europie. Obserwacja ta pozwala przypuszczać, że ewolucja poszczególnych form człowieka nie przebiegała w sposób liniowy. Gdyby było inaczej bylibyśmy w stanie wskazać moment, w którym populacje uległy genetycznej separacji.

Mimo znalezienia szkieletów innych archaicznych form człowieka niż AMH, długo istniało przekonanie, iż nie dochodziło między nimi do krzyżowania a co za tym idzie do tzw. admiksji genowej. Tę ostatnią definiuje się jako przepływ genów między dwoma odrębnymi od siebie populacjami. Odcinki DNA odziedziczone w wyniku admiksji genowej są nazywane komponentami genetycznymi.

Aby określić ilość DNA pochodzącego od archaicznych form człowieka w genomach współczesnych ludzi bądź też stwierdzić zdarzenie admiksji genowej

między dwoma populacjami wykonuje się tzw. "test 4 populacji". Dla przykładu, aby ustalić czy doszło do admiksji genowej między przodkami współczesnej populacji człowieka a neandertalczykami należy dysponować sekwencjami genomowymi dwóch współczesnych populacji człowieka, neandertalczyka, oraz szympansa. Następnie ustala się, w których pozycjach obie sekwencje genomowe współczesnych populacji człowieka zawierają inny nukleotyd niż ten obecny w genomie szympansa. Następnie sprawdza się, czy w tych pozycjach któraś z badanych współczesnych populacji ma nukleotyd identyczny jak w sekwencji genomowej neandertalczyka. Jeśli tylko jedna ze współczesnych populacji człowieka ma w swojej sekwencji genomowej takie pozycje oznacza to, że doszło do admiksji genowej między tą populacją a neandertalczykami. Jeśli natomiast obie współczesne populacje człowieka miałyby nukleotyd zgodny z tym występującym w sekwencji genomowej neandertalczyka oznacza to, że obie pochodzą od jednej ancestralnej populacji, która oddzieliła się od przodków neandertalczyków wcześniej niż badane współczesne populacje oddzieliły się od siebie. Wówczas nie ma powodu, dla którego takie mutacje z większą częstością występowałyby w jednej z tych populacji. Jeżeli jednak neandertalczycy krzyżowali się z przodkami którejś ze współczesnych populacji, ich potomkowie powinni dzielić więcej mutacji z neandertalczykami niż pozostałe populacje człowieka. Wykorzystując powyższą metodę dostarczono dowodów, że AMH i archaiczne formy ludzi żyli obok siebie i krzyżowali przez tysiąclecia.



Rycina 4. Drzewo filogenetyczne archaicznych form Homo i anatomicznie współczesnych ludzi. Kolory odpowiadają poszczególnych formom Homo i Homo sapiens. Przerywanymi liniami oznaczono domieszki genowe między poszczególnymi populacjami.

Na podstawie przeprowadzonych analiz stwierdzono kilkukrotną admiksję genową między AMH i neandertalczykami. Do pierwszych krzyżowań doszło już ok. 100 tya. Dowodzi tego 2% domieszka genetyczna AMH w genomie neandertalczyka z regionu gór Ałtaj. Domieszki takowej nie stwierdzono u neandertalczyków zamieszkujących Europę. Zakłada się zatem, że w okresie 130 - 100 tya miała miejsce ekspansja neandertalczyków z Europy na wschód i krzyżowanie się z AMH, którzy prawdopodobnie należeli do grupy pierwszych wychodźców z Afryki. Populacja ta została nazwana jako „praEuroazjaci" (ang. *Basal Eurasians*). Jest to tzw. "populacja widmo" ponieważ nie dysponujemy żadnymi jej szczątkami szkieletowymi. Jej istnienia dowodzą jednak wyniki analiz kopalnych i współczesnych genomów.

Przepływ genów zachodził również od neandertalczyków do AMH. Domieszka genowa pochodząca od neandertalczyków stanowi około 0,3 - 2,6% genomu u wszystkich współczesnych nie-Afrykańczyków. Uważa się, że jest ona wynikiem drugiej ekspansji neandertalczyków na wschód oraz w rejon Lewantu ok. 70 tya. Jest to też prawdopodobne miejsce nabycia neandertalskiego DNA przez AMH, po jego drugim wyjściu z Afryki około 54 - 49 tya. Wyniki analiz wskazują, że ostatnia wymiana genów między AMH a neandertalczykami nastąpiła nie później niż 37 tya, co wiąże się z pierwszą obecnością AMH w Europie ~45 tya i ekstynkcją neandertalczyka ~39 tya. Świadczy o tym stwierdzona 6-9% domieszka genetyczna neandertalczyka w genomie Oasel, najstarszego europejskiego kopalnego genomu, który jest datowany na ok. 39 tya. Oznacza to, że neandertalczyk był jego przodkiem co najmniej 6 pokoleń wstecz.

Admiksja genowa miała także miejsce między AMH a denisowianami. Krzyżowanie się z tą archaiczną formą człowieka datowane jest na ok. 49 - 44 tya, a ilość DNA odziedziczonego po niej w genomach współczesnych ludzi wynosi od 0,1 - 6%, przy czym występuje ono tylko we współczesnych populacjach ludzi żyjących we Wschodniej i Południowej Azji, Oceanii oraz Australii. Najwyższy poziom tej domieszki wynosi 3 - 6% i jest obserwowany u rdzennych mieszkańców Nowej Gwinei. Analizując segmenty DNA pochodzące od denisowian w genomach współczesnych ludzi ustalono, że rozbieżność w ilości domieszki genetycznej pomiędzy Wschodnimi Azjatami a Nowo Gwinejczykami wynika z faktu, iż źródłami tych domieszek były dwie różne populacje denisowian. Grupa ta podzieliła się na dwie linie ewolucyjne około 400 - 280 tya. Pierwsza z nich żyła w obrębie gór Ałtaj (denisowianie syberyjscy), druga, która jest źródłem genetycznej domieszki u Nowo Gwinejczyków, żyła na obszarze Oceanii (australo-denisowianie). Australo-denisowianie również są tzw. "populacją widmo" - gdyż nie posiadamy żadnych ich szczątków szkieletowych, a ich istnienie jest postulowane na podstawie analiz archeogenomicznych. Czas rozdzielenia się obu linii ewolucyjnych denisowian (400 -280 tya) jest niewiele wcześniejszy niż czas, w którym żył ostatni wspólny przodek denisowian i neandertalczyków (470 - 380 tya). Ustalono, że pomiędzy tymi archaicznymi formami człowieka także doszło do admiksji genowej z około 1,8% przepływem genów od

wschodnich neandertalczyków do syberyjskich denisowian. Najnowszych dowodów wspierających wymianę genów między oboma formami człowieka dostarczyła poznana niedawno sekwencja genomowa osobnika, którego rodzicami byli neandertalczyk i denisowianin [41].

Biorąc pod uwagę 3-6% domieszkę denisowian oraz 2% domieszkę neandertalczyków, stwierdzić można, że 5 - 8% genomu populacji człowieka żyjącej na obszarze Melanezji pochodzi od archaicznych form człowieka.

Wyniki badań archeogenomicznych dostarczają również dowodów na istnienie innej populacji widmo, tzw. "superarchaicznych ludzi". Obecnie jest to najstarsza znana linia ewolucyjna człowieka, która uległa separacji od przodków człowieka współczesnego, neandertalczyków i denisowian ok. 1,400 - 900 tya, a ich 3 - 6 % wkład genetyczny został zidentyfikowany w genomach denisowian i datowany jest na ok. 65 tya.

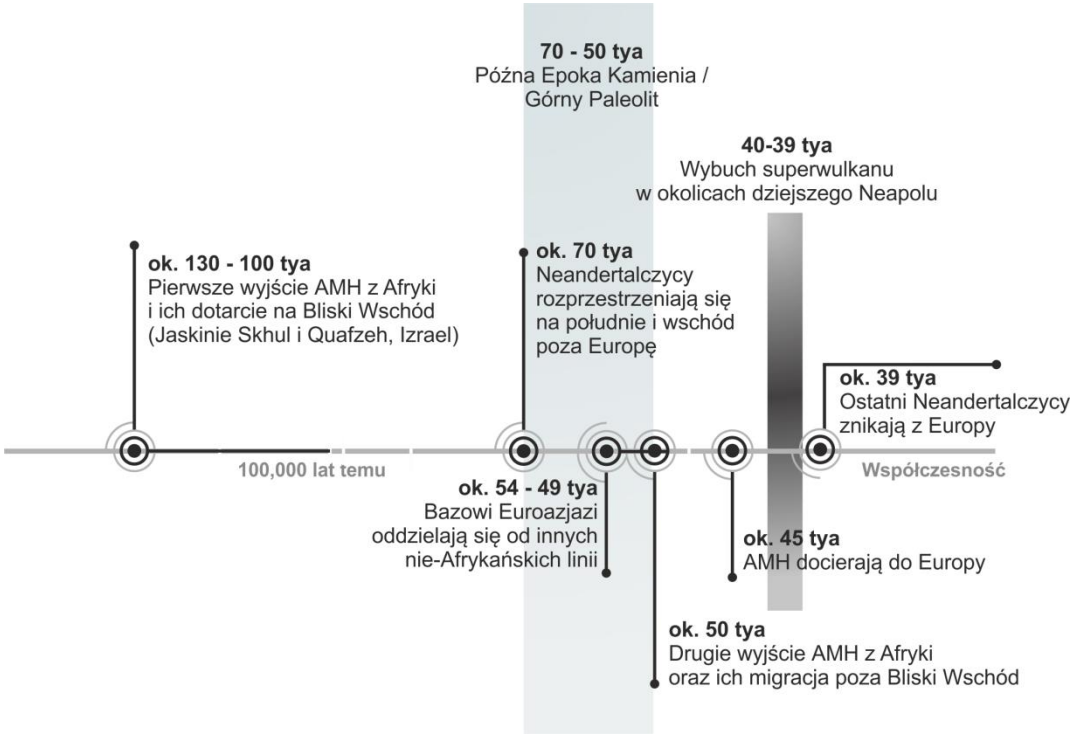
Przedstawione powyżej zdarzenia z historii biologicznej człowieka pokazują, że przepływ genów pomiędzy różnymi formami i grupami ludzi był powszechny. Co więcej przepływ genów pomiędzy tak różnymi anatomicznie i behawioralnie formami człowieka uwidacznia jakim problemem jest klasyfikacji jego archaicznych form i ich ewentualny podział na oddzielne gatunki.

AMH w Europie

Wiedzę o historii biologicznej pierwszych populacji AMH zamieszkujących obszar Euroazji nadal uznać można za szczątkową. Wynika to z faktu, iż badaniom archeogenomicznym jak dotąd poddane zostały jedynie pojedyncze osobniki, które obecnie traktowane są jako reprezentacyjne dla danej populacji. Niemniej jednak zastosowanie takiego podejścia doprowadziło do odkryć, których nie przewidywały żadne wcześniejsze narracje archeologiczne czy antropologiczne.

Od około 430 tya endogennymi mieszkańcami Europy byli neandertalczycy. Jak dowodzą wyniki badań archeologicznych i antropologicznych przybycie AMH poskutkowało radykalnymi przemianami demograficznymi. AMH przybili po raz pierwszy do Europy około 45 tya. Analizy próbek aDNA wyizolowanych od osobników żyjących na Syberii i w Rumunii przed 40-45 tya pokazują, że populacja ta (lub populacje te) określana jako paleolityczni łowcy-zbieracze nie ma znaczącego wkładu w strukturę genetyczną współczesnych Europejczyków. Około 39 tya doszło bowiem w Europie do wielkiej erupcji wulkanów, co doprowadziło do całkowitej ekstynkcji neandertalczyków i prawie całkowitej pierwszych europejskich łowców-zbieraczy. Po tym zdarzeniu, to jest około 37 tya, Europa została zaludniona ponownie najprawdopodobniej przez pojedynczą linię, której udało się przetrwać katastrofę ekologiczną. Dopiero osobniki reprezentujące tę populacje są bardziej spokrewnione ze współczesnymi Europejczykami niż Azjatami co świadczy, że ich materiał genetyczny przetrwał w Europie do obecnych czasów. Populacje łowców-zbieraczy

żyjące w Europie w okresie 37-33 tya związane były z kulturą oryniacką. Około 33 tya zostały one zastąpione przez genetycznie podobną populację związaną z kulturą grawecką. Tak utworzona struktura genetyczna trwała w Europie przez 11 tysięcy lat do nadejścia epoki lodowcowej to jest do około 22 tya. W tym samym czasie, tj. około 24-17 tya w rejonie jeziora Bajkał żyła odrębna populacja łowców-zbieraczy zwana ANE (ang. *Ancient Nort Eurasians*).

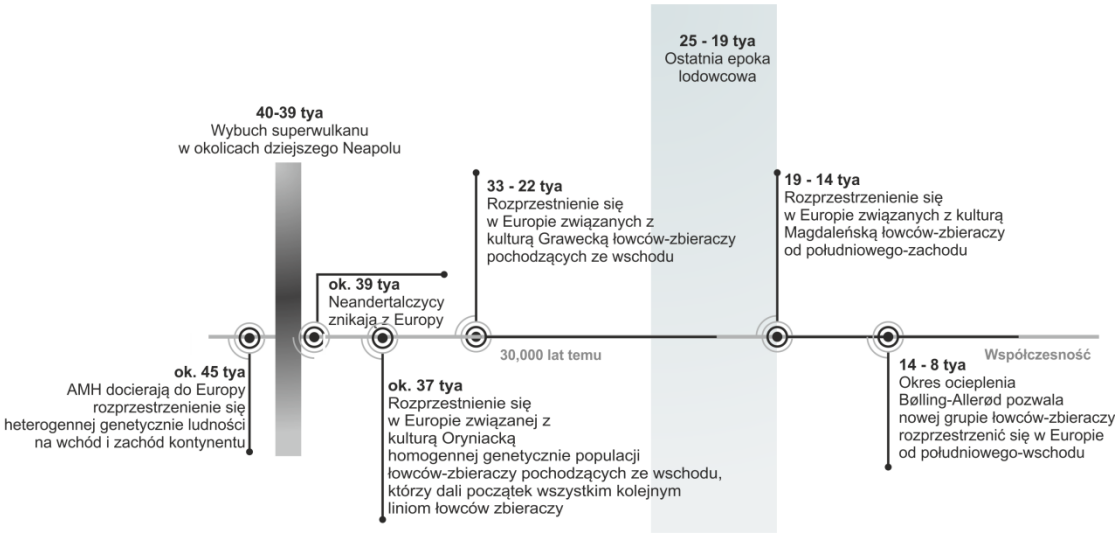


Rycina 6. Diagram obrazujący najważniejsze zdarzenia w historii biologicznej Homo sapiens s w okresie od pierwszego wyjścia AMH z Afryki do dotarcia do Europy. Czas wyrażono w tysiącach lat temu (tya).

Okres zlodowacenia przetrwały jedynie niewielkie grupy łowców-zbieraczy w rejonie dzisiejszej Hiszpanii. Kiedy około 19 tya lód zaczął się wycofywać doszło do rozprzestrzenienia się tej populacji w Europie Zachodniej, i powstania kultury magdaleńskiej. Co ciekawe tworząca ją populacja nie wywodziła się w prostej linii z grup związanych z kulturą grawecką, ale była blisko spokrewniona z wcześniejszą populacją związaną z kulturą oryniacką. Około 14 tya doszło do znaczącego ocieplenia klimatu. Okres ten zwany jest Bolling-Allerød. Z nim też związane jest kolejne istotne zdarzenie demograficzne w Europie, jakim była migracja łowców-zbieraczy z Bliskiego Wschodu.

W wyniku tych zdarzeń w Europie ukształtowały się cztery grupy łowców-zbieraczy: Zachodni Łowcy-Zbieracze (WHG, ang. *Western Hunter-Gatherers*), Kaukascy Łowcy-Zbieracze (CHG, ang. *Caucasus Hunter-Gatherers*), Wschodni Łowcy-Zbieracze (EHG, *Eastern Hunter-Gatherers*) oraz Skandynawscy Łowcy-Zbieracze (SHG, ang. *Scandinavian Hunter-Gatherers*). WHG dotarli do Europy wraz z

ekspansją w okresie Bolling-Allerød. Allele odziedziczone od tej grupy stanowią pierwszy główny komponent genetyczny obecny w puli genowej współczesnych Europejczyków. Pula genowa CHG powstała w wyniku admiksji między WHG oraz bazowymi Euroazjatami, zaś struktura genetyczna EHG została utworzona poprzez admiksję WHG oraz ANE. Ostatnia z grup łowców-zbieraczy, SHG powstała przez domieszkę genową WHG oraz EHG.



Rycina 7 Diagram obrazujący najważniejsze zdarzenia w historii biologicznej AMH po dotarciu do Europy w okresie od 45 tya do 8 tya. Czas wyrażono w tysiącach lat temu (tya).

Neolit – powstanie i rozwój rolnictwa

Rolnictwo zostało zapoczątkowane około 12-11 tya na terenie dzisiejszej Turcji i Syrii. Około 9 tya zaczęło się rozprzestrzeniać na zachód na obszar współczesnej Grecji oraz na wschód w rejon doliny Indus. W Europie proces wprowadzania rolnictwa był silnie związany z przemianami demograficznymi prowadzącymi do utworzenia nowej struktury genetycznej. Tak zwani wcześni rolnicy neolityczni (EEF, ang. *Early European Farmers* wywodzili się z dwóch różnych populacji. Grupa, która dotarła na Bałkany była spokrewniona z rolnikami z obszaru Anatolii, natomiast grupa osiadła w południowej Grecji była spokrewniona z rolnikami z Iranu.

Następnie rolnicy bałkańscy migrowali dalej dwoma ścieżkami: śródziemnomorską w kierunku Półwyspu Iberyjskiego oraz naddunajską do Europy Środkowej. W początkowych etapach zmiany demograficzne towarzyszące Neolityzacji przebiegały w odmienny sposób w różnych częściach Europy. W Europie Zachodniej i Środkowej dochodziło do postępującej admiksji genowej między rdzennymi WHG oraz nowo napływającymi EEF. Z kolei w Europie Wschodniej na

obszarach współczesnej Ukrainy i państw bałtyckich zachowana została ciągłość genetyczna. Powszechny w Europie Środkowej genetyczny komponent EEF jest nieobecny w próbkach pochodzących z Ukrainy i państw bałtyckich. Wyniki te dowodzą, że grupy łowiecko-zbierackie w północno-wschodniej Europie przejęły rolniczy tryb życia.

Do kolejnej istotnej zmiany w strukturze genetycznej Europy doszło około 6 tya w czasie Środkowego Neolitu. Zaobserwowano, ze populacje człowieka żyjące w tym okresie w Europie Środkowej charakteryzują się zwiększonym udziałem genetycznego komponentu WHG oraz typowych dla WHG haplogrup mtDNA. Zjawisko to jest wiązane z migracją ludności ze Skandynawii w obszar Europy Środkowej.

Późny Neolit i Epoka Brązu

Około 5 tya struktura genetyczna populacji żyjącej w Europie ponownie przeszła radykalną zmianę. Jej przyczyną był napływ ludności ze stepu indoeuropejskiego związanej z kulturą grobów jamowych (*Yamnaya*). Badania archeogenomiczne wskazują jednoznacznie, że *Yamnaya* stanowią źródło trzeciego głównego komponentu genetycznego obecnego w dzisiejszej puli genowej Europejczyków. Ludność *Yamnaya* była genetycznie różna od tej żyjącej wówczas w Europie. Ich przodkami były populacje zamieszkujące dzisiejsze tereny Armenii oraz Iranu. W konsekwencji *Yamnaya* byli niemal homogenną mieszanką tych dwóch grup, podczas gdy współcześni im mieszkańcy Europy byli wynikiem admiksji między EEF z Anatolii oraz WHG.

Bezpośrednich dowodów określających charakter zmian jakie zaszły w tym czasie w Europie dostarczyły analizy genomiczne ludzi związanych z kulturą ceramiki sznurowej (CWC, ang. *Corded Ware Culture*), która powstała około 4,5 tya. Struktura genetyczna populacji związanych z CWC wykazywała około 75% genetycznego komponentu *Yamnaya*, zaś resztę stanowiły obecne już w środkowym neolicie komponenty genetyczne związane z EEF i WHG. Ponadto analizy te pozwalają przypuszczać, że admiksja genetyczna między lokalnymi populacjami a *Yamnaya* była wynikiem jednej masywnej migracji a nie stałego napływu ludności ze wschodu.

Do ekspansji kultury CWC doszło w bardzo krótkim czasie na obszarze od Szwajcarii aż po europejską część Rosji. Badania archeogenomiczne pozwalają sformułować nową hipotezą wyjaśniającą ten fakt. U jej podstaw legły wyniki analiz aDNA mikrobów towarzyszących kopalnym szczątkom ludzkim pochodzącym z Europy oraz ze stepu. Zaobserwowano, ze znaczna część kopalnych osobników (7%) była zakażona chorobotwórczą bakterią *Yersinia pestis*. Blisko 700 lat temu drobnoustrój ten był odpowiedzialny za plagę czarnej śmierci - dżumy - która przyczyniła się do śmierci 1/3 populacji ówczesnej Europy, Chin i Indii. Chociaż genom *Y. pestis* datowany na okres około 5 tya nie zawiera kilku genów koniecznych do wystąpienia dżumy dymieniczej, to zawiera materiał genetyczny niezbędny do

wywołania dżumy płucnej. Możliwe jest zatem, że zajmujący się hodowlą bydła *Yamnaya* byli odporni na tę formę dżumy i zarazili chorobą nieodporne na nią populacje europejskie.

Dobrym przykładem ujawniającym ogromny potencjał archeogenomiki są badania dotyczące powstania i rozprzestrzenienia się w Europie kultury pucharów dzwonowatych (BBC, *ang. Bell Beaker Culture*). Według ustaleń archeologii BBC po raz pierwszy pojawiła się około 5 tya na Półwyspie Iberyjskim, skąd rozprzestrzeniła się do Europy Zachodniej, a następnie około 4,5 tya do Brytanii. Badania archeogenomiczne dowiodły, że kopalne osobniki związane z BBC pochodzące z Iberii nie różniły się genetycznie od poprzedzających je rolników neolitycznych. Obie populacje nie posiadały domieszki genetycznej *Yamnaya*. Kiedy jednak poddano analizie kopalne osobniki związane z BBC pochodzące z Francji, ich struktura genetyczna zawierała genetyczny komponent *Yamnaya*. Tym samym ustalono, że ekspansji BBC z Iberii w kierunku Europy Środkowej nie towarzyszyła migracja ludności. Analogiczne badania przeprowadzono na kopalnych osobnikach pochodzących z Brytanii. Genomy kopalnych rolników neolitycznych z Brytanii nie zawierały genetycznej domieszki *Yamnaya*, natomiast kopalne osobniki związane z BBC wywodziły znaczną część swojego genomu od *Yamnaya*. Dowodzi to, że w czasie ekspansji BBC do Brytanii, zjawisku temu towarzyszyła migracja ludności.

Archeogenomika w Polsce

Od pewnego już czasu do puli genowej europejskiej populacji człowieka przestały napływać nowe komponenty. Powstała struktura genetyczna ulegała jedynie zmianom ilościowym w wyniku lokalnych procesów demograficznych lub migracji o niewielkim zasięgu. Tym samym można powiedzieć, że od około 5 tysięcy lat struktura genetyczna europejskiej populacji człowieka składała się z tych samych elementów, zmieniają się tylko proporcje pomiędzy nimi. Z tego względu kolejnym istotnym wyzwaniem jest poznanie historii biologicznej populacji zamieszkujących poszczególne regiony Europy. Już pierwsze tego typu analizy wykazujące wyraźne różnice pomiędzy indywidualnymi populacjami dowiodły, że zbiorcza charakterystyka procesów demograficznych jest niewystarczająca dla opisania historii biologicznej danego regionu. Spostrzeżenie to dotyczy także populacji żyjących na obszarze współczesnej Polski.

Badania historii biologicznej ludzi zamieszkujących tereny dzisiejszej Polski były jak dotąd prowadzone w stosunkowo niewielkim zakresie. Pierwsze prace poruszające to zagadnienie powstały dopiero w ostatnich latach (Fernandes i wsp. 2018 oraz Tassi i wsp. 2017).

Najstarszy aDNA z obszaru dzisiejszej Polski znaleziono w północno-wschodniej części kraju (Dudka, Drestwo). Analizowany materiał pochodził od przedstawiciela populacji łączonej z kulturami subneolitycznymi (łowcy-zbieracze lecz produkujących ceramikę). W istocie osobnik ten posiadał typowy dla łowców-zbieraczy haplotyp mtDNA U5b1.

Niezwykle ciekawym obszarem z punktu widzenia badań procesów neolityzacji są Kujawy. Przebiegała tam bowiem granica oddzielająca południowe tereny, na których występowały rolnicze kultury neolityczne od północnych, gdzie występowały łowiecko-zbierackie kultury mezolityczne. Neolityczna społeczność tego regionu charakteryzowała się zwiększonym udział komponentu genetycznego WHG jeszcze przed okresem migracji do Europy Środkowej populacji ze Skandynawii. Co więcej obecność osób, bez genetycznego komponentu EEF świadczy o tym, że jeszcze w Środkowym Neolicie istniały enklawy autochtonicznych WHG.

W odróżnieniu od zachodniej części Europy Środkowej, na terytorium współczesnej Polski, 5 tya w strukturze genetycznej nie pojawił się komponent *Yamnaya*. Ówczesne populacje (bez genetycznego udziału *Yamnaya)* nie były związane z archeologiczną kulturą CWC, ale z kulturą amfor kulistych (GAC, *ang. Globular Amphora Culture*).

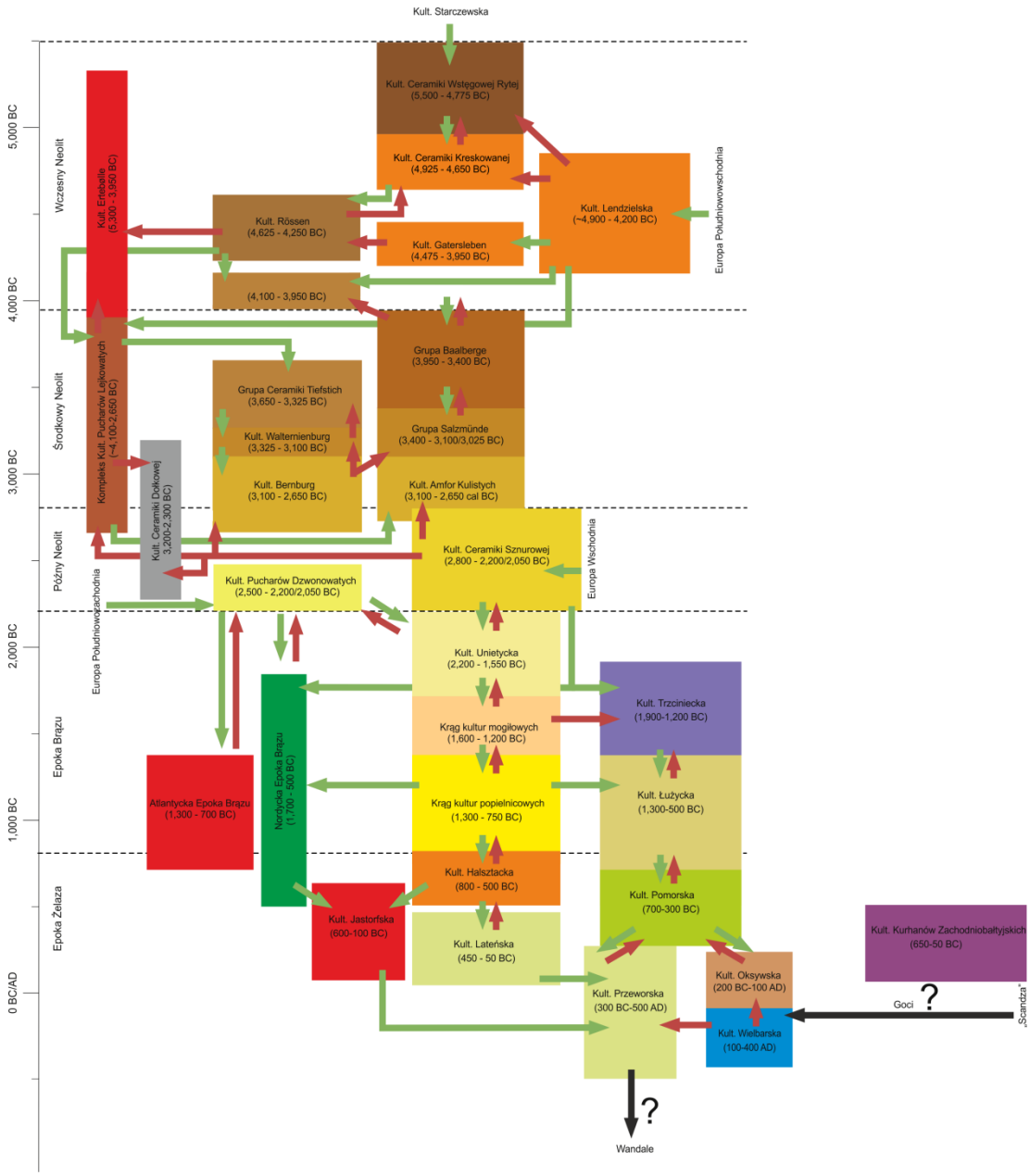
Począwszy od schyłku późnego neolitu stan naszej wiedzy o strukturze genetycznej populacji człowieka jest stosunkowo niewielki. Czas ten zgrubnie możemy podzielić na następujące części: epoka brązu, epoka żelaza, okres wędrówek ludów oraz średniowiecze.

Dotychczasowe badania archeologiczne sugerują, że w epoce brązu zarówno na obszarze współczesnej Polski jak również w innych częściach Europy istniało wielu kultur. Poznanie struktury genetycznej ówczesnych populacji człowieka jest jednak niezwykle trudne, ze względu na powszechnie panujący zwyczaj kremacji zwłok. Dopiero w epoce żelaza ponownie przywrócony został obrządek grzebania zmarłych. Jest to zatem pierwszy od późnego neolitu okres z którego stosunkowo łatwo dostępne są kopalne szczątki zawierające aDNA.

Po epoce żelaza nastał tzw. okres wędrówki ludów, nazywany przez antycznych Rzymian i Greków "inwazjami barbarzyńców". Przed rozwojem archeogenomiki dominowała teoria, iż były to masowe przemieszczenia ludności na obszarze Europy. W tej chwili powstaje wiele nowych hipotez poddających w wątpliwość wcześniejsze ustalenia. Badacze podejmują próby precyzyjnego określenia co tak naprawdę kryje się pod hasłem migracja. Czy określenie to faktycznie dotyczy masowego przemieszczania się ludności? A może dotyczy w wielu przypadkach jedynie przejęcia władzy przez nieliczne, ale dobrze zorganizowane grupy tworzące ówczesne elity społeczne. W tym kontekście niezwykle istotnym z punktu widzenia tworzenia się państwowości polskiej jest, trwający od dekad, spór dotyczący migracji Słowian.

Z tego też względu w mojej pracy doktorskiej zająłem się badaniami historii biologicznej populacji żyjących w pierwszych wiekach naszej ery na obszarze współczesnej Polski. Jak dotąd wiedza o żyjących wówczas ludziach pochodzi głównie z badań archeologicznych i antropologicznych. Nie dysponujemy żadnymi źródłami pisanymi dotyczącymi tego okresu, a badania archeogenomiczne są szczątkowe.

W pierwszych wiekach naszej ery na obszarze współczesnej Polski występowało szereg kultur archeologicznych. W okresie poprzedzającym w rejonie tym dominowała kultura Pomorska (około 2,7-2,4 tya). Wraz z jej stopniowym zanikiem zaczęła pojawiać się kultura Oksywska oraz Przeworska. Najstarsze stanowiska kultury Oksywskiej zidentyfikowano u ujścia Odry i Wisły. Tymczasem kultura przeworska objęła swym zasięgiem Śląsk, Wielkopolskę, Mazowsze, Podlasie i część Małopolski. Istnieje szereg hipotez łączących kulturę Przeworską z Wandalami. Struktura genetyczna ludności związanej z kulturą Przeworską nie została jak dotąd poznana. Kolejne przekształcenia w dorzeczach Odry i Wisły związane są z rozprzestrzenianiem się kultury wielbarskiej, która często kojarzona jest z wędrówkami Gotów.



Rycina 8. Diagram obrazujący występowanie i przemiany poszczególnych formacji, grup i kultur archeologicznych w Europie w okresie od wczesnego neolitu po epokę żelaza. Kolory odpowiadają poszczególnym kulturom archeologicznych. Strzałkami pokazano wzajemne relacje między poszczególnymi kulturami. Strzałki zielone oznaczają zapoczątkowanie danej kultury archeologicznej, strzałki czerwone oznaczają zastępowanie docelowej kultury archeologicznej. BC oznacza czas wyrażony w latach przed narodzeniem Chrystusa.

Cel pracy

Głównym celem mojej pracy doktorskiej było poznanie genomów mitochondrialnych przedstawicieli populacji człowieka zamieszkujących obszar współczesnej Polski w pierwszych wiekach naszej ery oraz wykorzystanie pozyskanych danych genomicznych do poszerzenie naszej wiedzy o historii biologicznej *H. sapiens s* w Europie Centralnej i Wschodniej.

Jego osiągniecie wymagało realizacji następujących zadań:

- identyfikacji cmentarzysk z epoki żelaza położonych na obszarach między Odrą a Wisłą oraz Wisłą a Bugiem – cmentarzyska te powinny być możliwie jak najlepiej scharakteryzowane pod względem archeologicznym i antropologicznym;

- datowania C14 wybranych szczątków ludzkich;

- izolacji DNA z kopalnych szczątków ludzkich datowanych na okres epoki żelaza

- sekwencjonowania DNA;

- przetworzenia surowych danych NGS;

- ustalenia autentyczności aDNA;

- wstępnej analizy przesiewowej badanych próbek w celu wybrania tych o największej zawartości endogennego DNA człowieka;

- oceny stopnia kontaminacji próbek w tym szczególnie współczesnym DNA człowieka;

- określenia genetycznej płci badanych osobników;

- wyznaczenia haplogrup i haplotypów mtDNA oraz złożenie pełnej sekwencji mtDNA;

- przeprowadzenia analiz populacyjnych:

    - wyznaczenia wewnątrzpopulacyjnej zmienności genetycznej badanych grup oraz jej porównanie z wewnątrzpopulacyjną zmiennością wyznaczoną dla współczesnych populacji człowieka izolowanych i nieizolowanych pod względem genetycznym,

- wyznaczenia związków genetycznych między badanymi populacjami a innymi populacjami zamieszkującymi Europę w okresie od jej pierwszego zasiedlenia po czasy współczesne,

- określenia genetycznego pochodzenia badanych populacji;

- przeprowadzenia analiz genomicznych mikroorganizmów towarzyszących kopalnym szczątkom ludzkim w kontekście zastosowania uzyskanych danych jako biomarkerów wspomagających badania procesów migracyjnych.

Krótkie omówienie wyników zaprezentowanych w publikacjach wchodzących w skład rozprawy doktorskiej

Zgodnie z przyjętymi założeniami wyselekcjonowano dwa dobrze scharakteryzowane pod względem archeologicznym i antropologicznym cmentarzyska zlokalizowane na obszarze między Odrą a Wisłą (Kowalewko, populacja KOW-OVIA) oraz Wisłą a Bugiem (Masłomęcz, populacja MAS-VBIA). Datowanie kopalnych szczątków ludzkich na podstawie danych archeologicznych oraz C14 potwierdziło, że badane osoby pochodziły z epoki żelaza. Osoby z populacja KOW-OVIA żyły w I-II wieku n.e., a z MAS-VBIA w II-IV wieku n.e.. DNA wyizolowany od 60 osób z grupy KOW-OVIA i 27 z grupy MAS-VBIA) sekwencjonowano metodą NGS. Uzyskane wyniki sekwencjonowania poddałem wielokierunkowym analizom bioinformatycznym. W rezultacie określiłem zawartość endogennego DNA człowieka w badanych próbkach, płeć badanych osób, poziom zanieczyszczenia innym ludzkim DNA, wyznaczyłem haplogrupy i haplotypy mtDNA oraz zrekonstruowałem pełnej długości genomy mtDNA. Następnie przeprowadziłem analizy z zakresu genetyki populacyjnej. Na ich podstawie wyznaczyłem wartości wewnątrzpopulacyjnej zmienności genetycznej oraz związki między badanymi populacjami i innymi zamieszkującymi Europę w okresie od jej pierwszego zasiedlenia przez AMH po czasy współczesne. Dodatkowo przeprowadziłem analizę mającą na celu zidentyfikowanie biomarkerów procesów migracyjnych w mikrobiomie towarzyszącym kopalnym szczątkom ludzkim.

W pierwszej z prac wchodzących w skład mojej rozprawy doktorskiej pt. „*A mosaic genetic structure of the human population living in the South Baltic region during the Iron Age*”, pokazałem, że KOW-OVIA nie jest izolowaną populacją i charakteryzuje się wewnątrzpopulacyjną różnorodnością genetyczną na poziomie współczesnych populacji europejskich. Struktura genetyczna KOW-OVIA jest najbardziej zbliżona do populacji żyjących w Europie Środkowej w czasie od co najmniej późnego neolitu. W szczególności bliskie związki genetyczne występowały między badaną populacją a populacją żyjąca w okresie epoki żelaza na Półwyspie Jutlandzkim (JIA). Co interesujące KOW-OVIA wykazywała bliskie związki genetyczne z kopalnymi populacjami zarówno z zachodniej jak i północno-środkowej Europy. Następnie zweryfikowałem hipotezę, czy obserwowane związki genetyczne KOW-OVIA z innymi kopalnymi populacjami mogą być wynikiem odmiennych historii genetycznych kobiet i mężczyzn. Ustaliłem, że kobiety i mężczyźni mieli odmienne historie genetyczne w linii matczynej. Kobiety były bliżej spokrewnione z populacjami rolników z wczesnego i środkowego neolitu, natomiast mężczyźni wykazywali bliskie związki genetyczne z populacjami późnego neolitu a w szczególności z populacją JIA.

W drugiej pracy pt. „*Goth migration induced changes in the matrilineal genetic structure of the central-east European population*” analizowaliśmy wpływ pierwszych migracji Gotów na strukturę genetyczną populacji żyjącej na obszarze współczesnej

Polski w czasie epoki żelaza, w tym szczególnie żyjącej między Wisłą a Bugiem MAS-VBIA. Ustaliliśmy, że populacja ta jest nieizolowaną grupą i odznacza się wysoką różnorodnością genetyczną na poziomie współczesnych populacji azjatyckich. W linii matczynej była najbardziej zbliżona do populacji wczesnego neolitu. Szczególnie bliskie związki genetyczne występowały pomiędzy MAS-VBIA oraz JIA i KOW-OVIA. Co bardzo interesujące w odniesieniu do populacji spoza Europy Środkowej, MAS-VBIA wykazywała bliskie związki genetyczne z *Yamnaya*.

W kolejnym etapie wyniki analiz genetycznych zostały skonfrontowane z istniejącymi narracjami historycznymi dotyczącymi powstania i pierwszych wędrówek Gotów oraz kultury Wielbarskiej. Wyniki przeprowadzonych przez mnie analiz wydają się być zgodne z hipotezą, w myśl której Masłomęcz był ważnym ośrodkiem władzy związanym z Gotami. Ponadto uprawdopodabniają Południową Skandynawię jako miejsce pochodzenia Gotów. Bliskie związki genetyczne między MAS-VBIA oraz współczesną jej KOW-OVIA wskazują na wspólną historię biologiczną obu grup, które posiadają również wspólny kontekst wielbarskiej materialnej kultury archeologicznej.

Na tej podstawie zaproponowaliśmy możliwy scenariusz pierwszych wędrówek Gotów przez tern współczesnej Polski. W pierwszym etapie Goci skolonizowali ujście Wisły w czasie ok. II-I wieku p.n. e.. Następnie przemieścili się na południe i weszli na tereny uprzednio zajmowane przez ludność związaną z kulturą przeworską (włączając Kowalewko, ok. I-II wieku n.e.). Charakter osiedleńczy Gotów w tym rejonie mógł być okresowy z uwagi na to, iż KOW-OVIA posiadała mozaikowatą strukturę genetyczną z powodu odmiennych historii genetycznych kobiet i mężczyzn. W kolejnym etapie Goci rozpoczęli wędrówkę wzdłuż Wisły a następnie rozprzestrzenili się na obszarze na wschód od Wisły do rzeki Buk. Z tego miejsca rozpoczęli wędrówkę w kierunku Morza Czarnego, gdzie doszło do admiksji genowej z populacjami stepu pontyjsko-kaspijskiego i powstania archeologicznej kultury Czerniachowskiej. Ostatecznie część ludności związanej z kulturą Czerniachowską przemieściła się z powrotem w rejon doliny hrubieszowskiej osiedlając się w pobliżu dzisiejszego Masłomęcza tworząc społeczność związaną z archeologiczną „grupą masłomęcką".

W ostatniej publikacji pt. „*Comprehensive analysis of microorganisms accompanying human archaeological remains*" dokonałem analizy mikrobiomu towarzyszącego kopalnym szczątkom ludzkim w celu zidentyfikowania potencjalnych biomarkerów migracji człowieka. Do analiz wykorzystałem DNA wyizolowane od 161 osób pochowanych na 7 cmentarzyskach datowanych na okres epoki żelaza i wczesnego średniowiecza.

W powyższej pracy wykazano, że mikrobiomy towarzyszące kopalnym szczątkom ludzkim zawierają mikroorganizmy związane z mikroflorą ludzi. Co istotne, bakterie te wykazują charakterystyczne wzory uszkodzeń typowe dla aDNA na poziomie zbliżonym do endogennego DNA ludzkiego pozyskanego z analizowanych szczątków. Jednocześnie poziom tych uszkodzeń jest wyższy niż dla bakterii klasyfikowanych jako środowiskowe. Z uwagi na powyższe mikroorganizmy

sklasyfikowane jako związane z człowiekiem zdają się reprezentować mikroflorę bakteryjną badanych osobników, która towarzyszyła im przed śmiercią (nie była skutkiem późniejszego zanieczyszczenia). Niektóre z bakterii związanych z człowiekiem to znane patogeny. Ich identyfikacja i badania ich ewolucji mogą być źródłem informacji ubogacającym opis historii biologicznej człowieka.

Podsumowanie

Biorąc pod uwagę przedstawione powyżej fakty stwierdzić można, iż przeprowadzone badania są źródłem wielu nowych informacji o historii demograficznej populacji człowieka zamieszkujących obszar współczesnej Polski w pierwszych wiekach naszej ery. Zebrane dane dają podstawy zarówno do zweryfikowania szeregu istniejących jak i sformułowania nowych hipotez oraz narracji historycznych.

Wyniki uzyskane podczas realizacji mojej pracy doktorskiej opisane zostały w trzech pracach eksperymentalnych opublikowanych na łamach międzynarodowych czasopism znajdujących się na liście JCR.

Bibliografia

Prace wchodzące w skład rozprawy doktorskiej zostały wyróżnione za pomocą podkreślenia.

1 Piazza, C.-S. P. M. A. *The History and Geography of Human Genes*. (Princeton University Press, 1994).

2 Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167-172, doi:10.1038/nature14507 (2015).

3 Brandt, G. *et al.* Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. *Science* **342**, 257-261, doi:10.1126/science.1241844 (2013).

4 Haak, W. *et al.* Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207-211, doi:10.1038/nature14317 (2015).

5 Brown, T. A. *Genomes*. (Bios Scientific Publishers : Wiley-Liss, 1999).

6 Mathieson, I. *et al.* Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499-503, doi:10.1038/nature16152 (2015).

7 Clark, H. (Sinauer Associates is an imprint of Oxford University Press, 2006).

8 Schneider, S. & Excoffier, L. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* **152**, 1079-1089 (1999).

9 Nachman, M. W. & Crowell, S. L. Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**, 297-304 (2000).

10 Xue, Y. *et al.* Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Current biology : CB* **19**, 1453-1457, doi:10.1016/j.cub.2009.07.032 (2009).

11 Rosenberg, N. A. & Nordborg, M. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature reviews. Genetics* **3**, 380-390, doi:10.1038/nrg795 (2002).

12 Poznik, G. D. *et al.* Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science* **341**, 562-565, doi:10.1126/science.1237619 (2013).

13 Novembre, J. *et al.* Genes mirror geography within Europe. *Nature* **456**, 98-101, doi:10.1038/nature07331 (2008).

14 Paabo, S., Gifford, J. A. & Wilson, A. C. Mitochondrial DNA sequences from a 7000-year old brain. *Nucleic acids research* **16**, 9775-9787, doi:10.1093/nar/16.20.9775 (1988).

15 Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715, doi:10.1038/362709a0 (1993).

16 Campos, P. F. *et al.* DNA in ancient bone - where is it located and how should we extract it? *Annals of anatomy = Anatomischer Anzeiger : official organ of the Anatomische Gesellschaft* **194**, 7-16, doi:10.1016/j.aanat.2011.07.003 (2012).

17 Dabney, J., Meyer, M. & Paabo, S. Ancient DNA damage. *Cold Spring Harbor perspectives in biology* **5**, doi:10.1101/cshperspect.a012567 (2013).

18 Briggs, A. W. *et al.* Patterns of damage in genomic DNA sequences from a Neandertal. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 14616-14621, doi:10.1073/pnas.0704665104 (2007).
19 Olivieri, C. *et al.* Characterization of nucleotide misincorporation patterns in the iceman's mitochondrial DNA. *PloS one* **5**, e8629, doi:10.1371/journal.pone.0008629 (2010).
20 Hansen, A., Willerslev, E., Wiuf, C., Mourier, T. & Arctander, P. Statistical evidence for miscoding lesions in ancient DNA templates. *Molecular biology and evolution* **18**, 262-265, doi:10.1093/oxfordjournals.molbev.a003800 (2001).
21 Gilbert, M. T. *et al.* Recharacterization of ancient DNA miscoding lesions: insights in the era of sequencing-by-synthesis. *Nucleic acids research* **35**, 1-10, doi:10.1093/nar/gkl483 (2007).
22 Lamers, R., Hayter, S. & Matheson, C. D. Postmortem miscoding lesions in sequence analysis of human ancient mitochondrial DNA. *Journal of molecular evolution* **68**, 40-55, doi:10.1007/s00239-008-9184-3 (2009).
23 Green, R. E. *et al.* Analysis of one million base pairs of Neanderthal DNA. *Nature* **444**, 330-336, doi:10.1038/nature05336 (2006).
24 Prufer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43-49, doi:10.1038/nature12886 (2014).
25 Pinhasi, R. *et al.* Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. *PloS one* **10**, e0129102, doi:10.1371/journal.pone.0129102 (2015).
26 Carpenter, M. L. *et al.* Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *American journal of human genetics* **93**, 852-864, doi:10.1016/j.ajhg.2013.10.002 (2013).
27 Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409-413, doi:10.1038/nature13673 (2014).
28 Amorim, C. E. G. *et al.* Understanding 6th-century barbarian social organization and migration through paleogenomics. *Nature communications* **9**, 3547, doi:10.1038/s41467-018-06024-4 (2018).
29 Stolarek, I. *et al.* Goth migration induced changes in the matrilineal genetic structure of the central-east European population. *Scientific reports* **9**, 6737, doi:10.1038/s41598-019-43183-w (2019).
30 Stolarek, I. *et al.* A mosaic genetic structure of the human population living in the South Baltic region during the Iron Age. *Scientific reports* **8**, 2455, doi:10.1038/s41598-018-20705-6 (2018).
31 Martiniano, R. *et al.* Genomic signals of migration and continuity in Britain before the Anglo-Saxons. *Nature communications* **7**, 10326, doi:10.1038/ncomms10326 (2016).
32 Schiffels, S. *et al.* Iron Age and Anglo-Saxon genomes from East England reveal British migration history. *Nature communications* **7**, 10408, doi:10.1038/ncomms10408 (2016).
33 Lipson, M. *et al.* Population Turnover in Remote Oceania Shortly after Initial Settlement. *Current biology : CB* **28**, 1157-1165 e1157, doi:10.1016/j.cub.2018.02.051 (2018).

34 Philips, A. *et al.* Comprehensive analysis of microorganisms accompanying human archaeological remains. *GigaScience* **6**, 1-13, doi:10.1093/gigascience/gix044 (2017).
35 Luhmann, N., Doerr, D. & Chauve, C. Comparative scaffolding and gap filling of ancient bacterial genomes applied to two ancient Yersinia pestis genomes. *Microbial genomics* **3**, e000123, doi:10.1099/mgen.0.000123 (2017).
36 Reich, D. *et al.* Denisova admixture and the first modern human dispersals into Southeast Asia and Oceania. *American journal of human genetics* **89**, 516-528, doi:10.1016/j.ajhg.2011.09.005 (2011).
37 Hajdinjak, M. *et al.* Reconstructing the genetic history of late Neanderthals. *Nature* **555**, 652-656, doi:10.1038/nature26151 (2018).
38 Reich, D. *et al.* Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053-1060, doi:10.1038/nature09710 (2010).
39 Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065-1093, doi:10.1534/genetics.112.145037 (2012).
40 Hublin, J. J. et al. New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. Nature 546, 289-292, doi:10.1038/nature22336 (2017).
41 Slon, V. *et al.* The genome of the offspring of a Neanderthal mother and a Denisovan father. *Nature* **561**, 113-116, doi:10.1038/s41586-018-0455-x (2018).

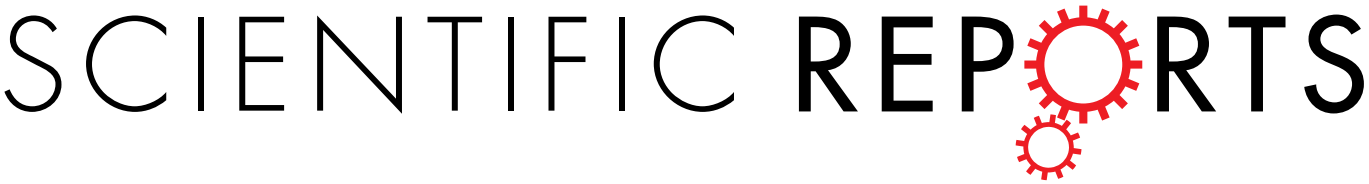
LISTA PUBLIKACJI WCHODZĄCYCH W SKŁAD ROZPRAWY DOKTORSKIEJ

1

Ireneusz Stolarek, Anna Juras, Luiza Handschuh, Malgorzata Marcinkowska-Swojak, Anna Philips, Michal Zenczak, Artur Dębski, Hanna Kóčka-Krenz, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

A mosaic genetic structure of the human population living in the South Baltic region during the Iron Age

Scientific Reports volume 8, Article number: 2455 (2018)
doi: 10.1038/s41598-018-20705-6



OPEN

A mosaic genetic structure of the human population living in the South Baltic region during the Iron Age

Received: 30 October 2017
Accepted: 23 January 2018
Published online: 06 February 2018

Ireneusz Stolarek[1], Anna Juras[2], Luiza Handschuh[1], Malgorzata Marcinkowska-Swojak[1], Anna Philips[1], Michal Zenczak[1], Artur Dębski[3], Hanna Kóčka-Krenz[3], Janusz Piontek[2], Piotr Kozlowski[1] & Marek Figlerowicz[1,4]

Despite the increase in our knowledge about the factors that shaped the genetic structure of the human population in Europe, the demographic processes that occurred during and after the Early Bronze Age (EBA) in Central-East Europe remain unclear. To fill the gap, we isolated and sequenced DNAs of 60 individuals from Kowalewko, a bi-ritual cemetery of the Iron Age (IA) Wielbark culture, located between the Oder and Vistula rivers (Kow-OVIA population). The collected data revealed high genetic diversity of Kow-OVIA, suggesting that it was not a small isolated population. Analyses of mtDNA haplogroup frequencies and genetic distances performed for Kow-OVIA and other ancient European populations showed that Kow-OVIA was most closely linked to the Jutland Iron Age (JIA) population. However, the relationship of both populations to the preceding Late Neolithic (LN) and EBA populations were different. We found that this phenomenon is most likely the consequence of the distinct genetic history observed for Kow-OVIA women and men. Females were related to the Early-Middle Neolithic farmers, whereas males were related to JIA and LN Bell Beakers. In general, our findings disclose the mechanisms that could underlie the formation of the local genetic substructures in the South Baltic region during the IA.

One of the consequences of the rapid development of DNA isolation, enrichment and sequencing technologies is currently observed: the substantial increase of our knowledge about the prehistory of anatomically modern humans (AMH). The first projects that focused on ancient DNA (aDNA) isolated from singular individuals gave a rough idea of when the first representatives of AMH appeared in Africa and how global human expansion occurred. Currently, the genetic studies on a population scale are a pressing need. Together with archaeological and anthropological research, genetic studies can provide new insight into the natural history of *Homo sapiens sapiens*. So far, such population scale studies have most frequently been concentrated on mitochondrial DNA (mtDNA) due to its high copy number, small size and lack of recombination. Consequently, the analyses of ancient mtDNA substantially contribute to unraveling the history of human settlement around the world.

The genetic structure of the European population has been shaped by a series of consecutive or partly concurrent processes that form the following chain of events: (i) Europe colonization by AMH during the Upper Paleolithic Period; (ii) the Late Glacial and Post-glacial recolonization; (iii) introduction of agriculture (Neolithization); (iv) genetic continuation through the Middle Neolithic (MN) and the progressive intermix of the Early European Farmers (EEF) with endogenous hunters-gatherers (HG); (v) the inflow of new genetic components from the southwest and southeast to Central Europe in the Late Neolithic (LN), and formation of separate cultures in the Early Bronze Age (EBA)[1–4].

The initial colonization of Europe by AMH began approximately forty-five thousand years ago (tya)[5–7]. This colonization was connected to the extinction of *Homo sapiens neanderthalensis* and a spread of AMH HG. One of the characteristics of the HG populations was the domination of mtDNA haplogroup U represented mainly by

[1]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland. [2]Institute of Anthropology, Faculty of Biology, Adam Mickiewicz University, Poznan, Poland. [3]Institute of Archaeology, Collegium Historicum, Adam Mickiewicz University, Poznan, Poland. [4]Institute of Computing Sciences, Poznan University of Technology, Poznan, Poland. Correspondence and requests for materials should be addressed to M.F. (email: marekf@ibch.poznan.pl)

U2, U4, U5a, U5b, and U8[1,8]. However, the Last Glacial Maximum (LGM), dated 27–16 tya, forced populations living in north Europe to leave this region. Approximately 19 tya, together with climate warming, a recolonization of Europe by indigenous HG was initiated[5,7]. They came from southwestern and eastern refugia where they survived the ice age. Interestingly, approximately 14.5 tya indigenous HG were replaced by the new incoming people, who diverged from the ancestral HG population prior to the beginning of the LGM (29 tya)[9]. Another major inflow of AMH to Europe occurred during the so-called Neolithic revolution, approximately 12 tya[1,4]. This period is inextricably linked with the introduction of agriculture (cultivation and domestication of animals) and a change from a nomadic to sedentary lifestyle. The Neolithic revolution, the so-called "Neolithization", most likely proceeded in two ways[10]. First, it resulted from cultural changes initiated in the Mediterranean region that spread towards the southwest part of the continent[3,11]. Second, it was connected with the inflow of the EEF from Central Anatolia along the Danube and towards Central Europe[6,12,13].

Neolithization was accompanied by the introduction of new mtDNA haplogroups (N1a, T2, K, J, HV, V, W, and X), currently called the "Neolithic package"[14]. Interestingly, during the MN, an increase in the occurrence of haplogroups typical for the HG was observed throughout Europe[2,15,16]. In the central part of the continent this phenomenon is believed to be connected with the expansion of the population associated with the Funnel Beaker Culture (TRB ger. Trichterrandbecherkultur)[1,17]. The TRB was mainly spread within the northern areas of Central Europe and Scandinavia.

Next, two events that significantly affected the mtDNA structure in Europe occurred in the LN. They both coexisted for approximately 300 years across Central Europe. The first one was a migration of the Yamnaya steppe herders from the Eurasian Steppe[15]. A consequence of the inflow was the formation and spread of the Corded Ware Culture (CWC)[15,18] and the occurrence of new mtDNA haplogroups (I, U2, T1, R) in Central and Eastern Europe[2]. The second event was the formation of the Bell Beaker Culture (BBC), most likely in Western Europe, and its propagation in Central Europe[2]. The most significant differences that have been identified so far between the BBC and CWC populations at the mtDNA level are in the lower frequency of I and U2 haplogroups and the significant prevalence of haplogroup H in BBC[2]. Other substantial changes in the mtDNA haplogroup frequency were found in the population associated with the Unetice Culture (UC) that appeared in Central Europe during the EBA; thus, right after the LN. The UC replaced the BBC and CWC; however, its population did not seem to be a direct genetic continuity of the population associated with the two former cultures. The analysis of mtDNA indicated closer similarity of the UC to CWC rather than to BBC[2].

Considering the above history, one can expect that the genetic profile of mtDNA of the present-day Central Europe populations should be especially highly influenced by the population related to the UC. However, this is not the case. The analysis of contemporary mtDNA showed that the population associated with the BBC had a more considerable impact on the genetic landscape of present-day Central Europe than the populations associated with the eastern CWC and UC[2]. This was despite the latter being chronologically closer to contemporary times than the BBC. These findings suggest that after the EBA, there were significant demographic changes in Central Europe and, consequently, changes in the genetic structure of the Central Europe population. To elucidate this problem, more detailed studies of the populations living in Central Europe between the EBA and present day are necessary. These studies require careful selection of samples with reference not only to time but also to geographical distribution. First, when the contemporary genetic structure of Central and Eastern Europe was formed will be identified. Presently, it is unknown whether this genetic structure was finally shaped by prehistorical or Early Medieval events. The second issue concerns the representativeness of the samples used to define the genetic structure of the Central European population in the EBA and after this period. Thus far, most of the ancient mtDNA analyses were based on samples from Mittelelbe-Saale in Saxony-Anhalt. Genetic studies of the prehistory of populations inhabiting the region of contemporary Poland are sparse. They are based on a very limited number of samples and focused usually on fragments of mtDNA[19–21].

To obtain a more thorough view of the demographic processes that affected the genetic profile of Central European populations after EBA, we analyzed mtDNA extracted from individuals living in the area between Oder and Vistula rivers (east of Saxony) at the time of the Iron Age (IA). From several burial grounds, we selected a particularly rich and well-characterized one, located in central Wielkopolska (also referred to as Greater Poland) in Kowalewko, near Poznan (52°35′24″N 16°46′44″E). According to the archaeological and anthropological studies, approximately 500 people who had lived on that area from 1 until 200 A.D. were buried there. Most of the burial sites have previously been thoroughly examined archaeologically and anthropologically[22]. Additionally, preliminary genetic[20] and metagenomic studies[23] of these remains have been performed.

Here, we present the first analysis of mtDNA obtained from a large group of individuals living between the Oder and Vistula rivers during the IA. We identified mtDNA haplogroups for 40 individuals, and sequenced a complete mtDNA genome for 33 individuals. The performed analyses revealed a high genetic diversity of the studied group. Interestingly, women and men from Kowalewko had a significantly different genetic history. The collected data shed new light on the processes that shaped the specific, local genetic substructure of the human population in the South Baltic region during the IA.

Results

aDNA extraction, sequencing and preliminary characteristics. The studied group included people whose remains were excavated from the IA Wielbark culture cemetery, located in Kowalewko (Supplementary Fig. S1). There were two major reasons for choosing this particular cemetery. First, it is located in the central part of the region that spreads between the Oder and Vistula rivers. The genetic structure of the population living in this region during the IA and before has been poorly characterized. Second, the cemetery was used by the local population for 200 years. Therefore, it may be assumed that the obtained genetic profile would characterize reasonably well the population that lived there for a considerable period of time. The study included 60 individuals, whose burial sites have been thoroughly characterized archaeologically and anthropologically (detailed

description in Supplementary Table S1). Because they represented the population living between the Oder and Vistula rivers during the IA, we called this group Kow-OVIA.

Radiocarbon dating (Supplementary Fig. S2 and Table S1) confirmed that the collected bone materials were from the first and second centuries (between AD 36 (+−30) and AD 181 (+−30)). aDNA was isolated from teeth, according to the procedure described by[24] and[25]. In addition, for three individuals, aDNA was isolated twice, each time from a different tooth (internal control). In total, 63 aDNA samples were obtained. The NGS libraries were successfully obtained for 60 samples (for 57 individuals, including two libraries for three individuals from whom two samples were extracted). To estimate the amount of human DNA, all of the obtained libraries were subjected to shallow sequencing (the average coverage of mtDNA genome was 3.62x and ranged between 1 and 11x). The average amount of human DNA in the examined samples was 12% and ranged from 0.1 to 91.9% (Supplementary Table S2). In all of the samples, human DNA showed a typical aDNA damage pattern (C > T and G > A alteration most frequently occurring at the 5′ and 3′ ends of the reads) (Supplementary Table S3). The data obtained from the shallow sequencing enabled mtDNA haplogroup definition for 23 individuals. Importantly, the haplogroups obtained for the samples isolated from the same individuals were identical. Furthermore, for 27 individuals, their genetic sex was identified. The obtained results are summarized in Supplementary Table S4.

The shallow sequencing allowed for determination of a full sequence of mtDNA genomes for 13 individuals. In addition, 23 samples (22 individuals) with high amounts of endogenous DNA (over 10%) were subjected to deep sequencing (the average coverage was 163x and ranged between 6x and 1092x). As a result, a full sequence of mtDNA genomes was determined for 22 additional individuals (Supplementary Table S4). In all cases, mtDNA haplogroups were also established. For 20 individuals, mtDNA haplogroups were determined twice, i.e., based on shallow and deep sequencing data. Importantly, in all cases, particular individuals were classified into the same haplogroups. Next we assessed the contamination levels in all samples. To this end, we used the software contamMix. Mean value for authenticity estimates was 96%. Two individuals (PCA0018 and PCA0063) had lower than average authenticity probability, therefore were excluded from subsequent analyzes (Supplementary Table 1). In total, the full sequence of mtDNA was determined for 33 individuals and the haplogroup was established for 40 individuals (Supplementary Table S4). All identified haplogroups had been found earlier in the populations living in Central Europe from the EBA to the IA and belonged to 9 haplogroups (H, N, I, J, K, T, U, W, and X) (Table 1).

Intrapopulation genetic diversity of Kow-OVIA. The intrapopulation diversity of Kow-OVIA was analyzed using the full length mtDNA sequences obtained for 33 individuals (Supplementary Table S4). In a first step, we used a Minimum Spanning Network (MSN) method[26] to identify and visualize differences between particular mitochondrial genomes (Fig. 1). The obtained graph was typical for the population with a high level of genetic diversity[27]. We identified 3 pairs of individuals with identical mtDNA sequences (PCA0059, PCA0031; PCA0049, PCA0027; PCA0060, PCA0036) and 4 pairs of individuals with only one different nucleotide in the mtDNA sequences (PCA0047, PCA0003; PCA0004, PCA0028; PCA0034, PCA0059; PCA0034, PCA0031). To exclude the possibility that the presence of two potentially maternally related individuals in relatively small dataset will affect the results of our consecutive analyses, we removed from the tested group three individuals - one from each pair with identical mtDNA sequences. For the comparison, the results obtained with the dataset containing potentially related individuals were shown in the Supplementary Figures S4–S6.

Next, we applied Arlequin ver. 3.5.1 to determine the levels of haplotype diversity (HD) based on two fragments of the HVS (hypervariable sequence) region (HVS-I between 16033 and 16365 np (nucleotide positions) and HVS-II between 73 and 340 np) (Supplementary Table S5), and nucleotide diversity (π) based on the fragment of the HVS-I region (between 16000 and 16410 np) (Supplementary Table S6). The HD level calculated for Kow-OVIA (1.0000 +/−0.0082) was similar to those currently observed for the contemporary, open European populations and higher (even when considering 95% confidence interval) than the HD obtained for the population considered isolated (Supplementary Table S5)[28]. The level of π determined for Kow-OVIA (0.007922 +/−0.004666) was only slightly lower than the average value obtained for the present-day European populations (0.009[29]) (Supplementary Table S6). Thus, the results of all three analyses, MSN, HD, and π, indicated that Kow-OVIA was not a genetically homogenous, isolated population.

Kow-OVIA relationships with other populations from European space-time. *mtDNA haplogroup frequency.* In the first step, we attempted to place Kow-OVIA in Central European space-time only. To this end, we compared mtDNA haplogroup frequencies in Kow-OVIA with other populations living in that region from the Mesolithic until the IA. The studied group, including Kow-OVIA and 13 other fossil populations (for details see Table 2 and Supplementary Table S7), was extended with the present-day Central Europe Metapopulation (CEM) (Supplementary Table S8). The obtained set of 15 populations called CEPT (Central European Population Transect) was subjected to unsupervised hierarchical clustering (Ward's method with Euclidean distance) (Fig. 2a). The obtained results generally complied with known chronology of the formation of the genetic structure of the Central European population. The first to be separated from other populations were the HG. Next, we observed a divergence of the Early Neolithic (EN) populations [the Starčevo Culture (STA), Linearbandkeramik in Transdanubia (LBKT), Linearbandkeramik population from Central Europe (LBK), and Schöningen Group (SCG)] and the Middle Neolithic (MN) populations [the Baalberge Culture (BAC)] from the LN/EBA populations [the Bernburg Culture (BEC), CWC, BBC, and UC]. Then, the dendrogram supported the division into LN/EBA populations together with Kow-OVIA and post-EBA populations [Jutland Iron Age (JIA), CEM]. We noticed that Rössen Culture (RSC) and Salzmünde Culture (SMC) did not group with other EN/MN populations, but occupied a basal position with respect to the LN/EBA clade.

To place the Kow-OVIA in a context broader than the Central European one, we removed CEM from the studied group and extended it with 16 populations that lived in other regions of Europe from the Mesolithic to the IA (Supplementary Table S7). The obtained set of 30 populations (European Population Transect, abbreviated

Sample	Haplogroup confidence	Haplogroup	Sex assignment**
Sample_PCA0001	0,8	W	(M)/—
Sample_PCA0002	1	H28a1	(M)/—
Sample_PCA0003	1	H5a1	(F)/—
Sample_PCA0004	1	U3a1a1	(F)/—
Sample_PCA0005	—	U5b*	(M)/—
Sample_PCA0006	1	U5a1d1	(F)/—
Sample_PCA0007	1	W1	(F)/—
Sample_PCA0013	1	J1c3	(—)/—
Sample_PCA0014	—	—	(F)/—
Sample_PCA0015	1	H1f1a	(M)/M
Sample_PCA0016	—	—	(—)/—
Sample_PCA0017	—	—	(M)/—
Sample_PCA0018	0,5	HV18	(M)/M
Sample_PCA0019	—	—	(F)/—
Sample_PCA0020	—	—	(—)/—
Sample_PCA0021	—	—	(—)/—
Sample_PCA0022	—	—	(F)/—
Sample_PCA0023	—	—	(—)/—
Sample_PCA0024	—	—	(F)/—
Sample_PCA0025	—	—	(F)/—
Sample_PCA0026	0,5	T2b16	(F)/F
Sample_PCA0027	1	H1a	(M)/M
Sample_PCA0028	1	U3a1a	(F)/F
Sample_PCA0029	1	X2c1	(—)/—
Sample_PCA0030	1	H2a2b	(F)/—
Sample_PCA0031	1	K2a	(F)/F
Sample_PCA0032	0,5	T2n	(F)/F
Sample_PCA0033	—	—	(F)/—
Sample_PCA0034	1	K2a	(F)/F
Sample_PCA0035/65	0,5	J2b1a5	(M)/M
Sample_PCA0036	1	U5b1d1	(M)/—
Sample_PCA0037	0,5	T2e	(M)/M
Sample_PCA0038	1	H1e1a	(—)/—
Sample_PCA0039	—	—	(—)/—
Sample_PCA0040	1	I4a	(M)/M
Sample_PCA0041	0,8	N	(M)/—
Sample_PCA0042	—	—	(—)/—
Sample_PCA0043	—	—	(—)/—
Sample_PCA0044	1	H5a1	(F)/F
Sample_PCA0045	1	U5b1d	(F)/F
Sample_PCA0046a_b	1	U8a1a1b	(M)/M
Sample_PCA0047	1	H5a1	(F)/F
Sample_PCA0048	—	—	(—)/—
Sample_PCA0049	1	H1a	(F)/F
Sample_PCA0050	1	H1e1a	(M)/M
Sample_PCA0051_a_b	0,4	K2a	(—)/F
Sample_PCA0052	1	U5a1a1	(F)/F
Sample_PCA0053	1	K1b2a	(F)/F
Sample_PCA0054	1	U3a1a	(—)/F
Sample_PCA0055	—	—	(—)/—
Sample_PCA0056	1	T2b6a	(F)/F
Sample_PCA0057	1	J1c7a	(F)/—
Sample_PCA0058	—	—	(—)/—
Sample_PCA0059	1	K2a	(F)/F
Sample_PCA0060	1	U5b1d1	(M)/M
Sample_PCA0061	1	H1ak	(—)/—

Continued

Sample	Haplogroup confidence	Haplogroup	Sex assignment**
Sample_PCA0062	1	U4a2	(M)/M
Sample_PCA0063	1	H2a5	(M)/M
Sample_PCA0064	—	—	(F)/–
Sample_PCA0066	1	H2a5	(M)/M

Table 1. Results of mtDNA haplogroup assignment. *Haplogroup assigned as in Juras, *et al.*[20]. **Anthropological assignment given in (), genetic sex assignment given after/.

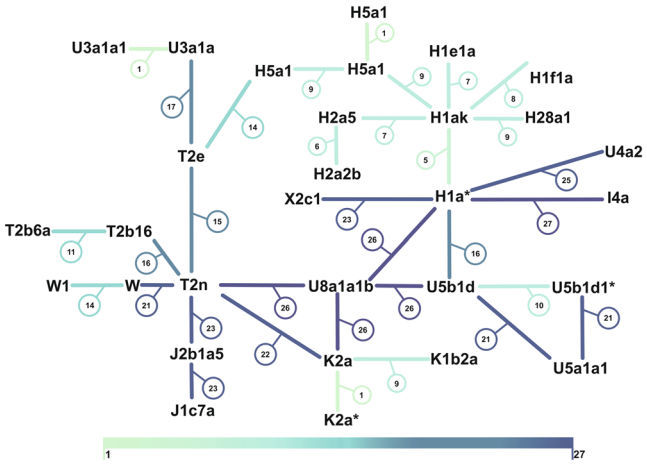


Figure 1. Minimal Spanning Network of 33 individuals from Kow-OVIA based on full mtDNA sequences. Each node corresponds to a haplotype determined for a unique mtDNA sequence. Numbers in circles show numbers of nucleotide differences between haplotypes. Edges are color-coded for the amount of nucleotide differences. The length of the edges is not informative. Stars mark haplotypes represented by two individuals with the identical mtDNA sequence.

EPT) as earlier was subjected to hierarchical clustering and additionally to PCA (Fig. 3a and b, Supplementary Table S9). Both analyses could distinguish one major group composed of EN/MN populations, however, excluding the Neolithic populations of the Iberian Peninsula (Neolithic Portugal (NPO) and Neolithic Basque Country (NBQ)). Within NPO and NBQ, the mtDNA haplogroup H was observed with a higher frequency than in the other Neolithic populations. LN populations were more scattered and their distribution seemed to be associated with the geographical location. For the populations inhabiting east peripheries of Europe [YAM, Bronze Age Kurgan samples from South Siberia (BAS), Iron Age Scythian (SCY) and Scytho-Siberian Pazyryk Culture (SSP)], the most characteristic were both the mtDNA haplogroups typical for Asia (mainly C, D) and a high prevalence of haplogroups U5a and a lack of U5b (except the population of Bronze Age Kazakhstan (BAK)) for almost the entire analyzed period of time. Interestingly, the closely located in time and space Kow-OVIA and JIA occupied the opposite side of the PCA diagram (Fig. 3b) than the preceding LN/EBA populations (the BEC, CWC, and UC). Such a placement resulted from the higher prevalence of the haplogroup H in Kow-OVIA and JIA compared to the BEC, CWC and UC. Alternatively, Kow-OVIA and JIA showed considerable differences in the prevalence of haplogroups typical for LN/EBA (I, R, T1, U2). They were less frequently observed in Kow-OVIA (2.7%) than in JIA (16.6%). As expected, the HG did not form a single group and differed from the other populations by a considerable portion of haplogroup U and a general lack of haplogroups associated with the "Neolithic package".

Our analyses also revealed very interesting and unreported earlier changes in the frequency of haplogroups U5a/U5b in the Central European populations (Supplementary Table S8). Initially, during the LN/EBA, the haplogroup U5a, characteristic for the populations from eastern glacial refugia, dominated. Then, during the IA, haplogroup U5b, typical for the populations from the western glacial refugia, was more frequent. At present, haplogroup U5a is again more often found in the CEM.

We also placed Kow-OVIA on a map showing the matrilineal genetic structure of the present-day human population. To this end, we performed a PCA of haplogroup frequencies of Kow-OVIA and 73 extant worldwide populations (Supplementary Fig. S3a and Table S10, Supplementary Material Text). Kow-OVIA was located within a range of European genetic diversity, in close proximity to the present Scandinavian populations (Norwegian and Swedish).

Genetic distances. The relationships between Kow-OVIA and the earlier and later European populations were also determined by measuring genetic distances among them. In the first step, we focused on populations from Central Europe only (from the CEPT group). Our analysis involved a fragment of mtDNA HVS-I sequence (the region between 16064 and 16400 np) (Fig. 2b and Supplementary Table S11). The established fixation indexes (Fst) showed that Kow-OVIA was most closely related to the BEC ($Fst = -0.01119$, $p = 0.74513$) and JIA

Abbreviation	Population	n
HGCN	Hunter Gatherers Central North Europe	23
HGSW	Hunter Gatherers South West Europe	13
HGE	Hunter Gatherers Eastern Europe	14
STA	Starčevo culture	44
LBKT	Linearbandkeramik culture in Transdanubia	39
LBK	Linearbandkeramik culture in Central Europe	102
RSC	Rössen culture	17
SCG	Schöningen group	33
BAC	Baalberge culture	19
SMC	Salzmünde culture	29
BEC	Bernburg culture	17
CWC	Corded Ware culture	44
BBC	Bell Beaker culture in Central Europe	35
UC	Unetice culture	94
PWC	Pitted Ware Culture	19
CAR	Cardial/Epicardial culture of the Iberian Penisula	18
NPO	Portuguese Neolithic population	17
NBQ	Neolithic population from Basque Country and Navarre	43
TRB	Funnel Beaker culture	10
TRE	Treilles culture	29
BAS	Bronze Age Siberia	11
BAK	Bronze Age Kazakhstan	8
RRBP	Gurgy 'Les Noisats' group	55
MIR	Iberian Chalcolithic El Mirador Burgos individuals	23
JIA	Jutland Iron Age	24
IIA	Iberian Iron Age population	27
SCY	Iron Age Scythian samples	16
SSP	Scytho-Siberian Pazyryk Culture	26
YAM	Yamnaya culture	41
Kow-OVIA	Oder Vistula Iron Age	40

Table 2. Published reference ancient mtDNA data and populations abbreviations. For list of references see Supplementary Table S7.

(Fst = −0.01054, p = 0.81012). CEM showed the lowest genetic distances to JIA (Fst = −0.00512, p = 0.75990), and Kow-OVIA (Fst = −0.00349, p = 0.74513).

In the second step, we determined the genetic distances between Kow-OVIA and populations living in the whole of Europe from the Mesolithic age to the IA. To this end, we selected 26 populations from the EPT group for which sequences of the mtDNA HVS-I region between 16064 and 16400 np (Supplementary Table S12) were known for a considerable number of individuals. The calculated genetic distances (Supplementary Table S12) again showed that Kow-OVIA was closest to the BEC (Fst = −0.01097, p = 0.72712) and then to the JIA (Fst = −0.01054, p = 0.80627). The results were visualized on a multidimensional scaling (MDS) plot (Fig. 4). As in the analysis of the haplogroup frequencies, we observed that although separated by a short distance, Kow-OVIA and JIA were both differently located regarding particular LN/EBA populations (the BBC, CWC, and UC). The distance between the Kow-OVIA and BBC (Fst = 0.00342; p = 0.36707) is five times smaller than for the JIA - BBC pair (Fst = 0.02426; p = 0.07196). Simultaneously, the relationship is opposite when one compares the Kow-OVIA and JIA with the CWC and UC (Fig. 4 and Supplementary Table S12). In addition, the performed analysis showed a previously unnoticed small genetic distance between the YAM and BBC (Fst = −0.00463, p = 0.58485), and the YAM and CWC (Fst = −0.00157, p = 0.50459).

Turning points in the formation of the Central European genetic structure. To verify the existing hypothesis describing the formation of the genetic structure of the Central European population in the period between the EN and IA, we used the analysis of molecular variance (AMOVA). In the first step, we determined an optimal division of the populations from CEPT excluding the HG and CEM. The division was considered to be optimal if intragroup variability (Fsc) was minimal and intergroup variability (Fct) maximal. In our analysis, we took advantage of earlier observations made by[2] and[10] who demonstrated relatively high genetic homogeneity of the EN/MN population (STA, LBKT, LBK, RSC, SCG, BAC, SMC). Accordingly, we allocated all EN/MN populations to one group and analyzed 18 different combinations of population grouping (Supplementary Table S13). We found that Fsc and Fct were optimal if the studied populations were divided into three groups: EN/MN (the STA, LBKT, LBK, RSC, SCG, BAC, and SMC), LN/EBA/IA (the BEC, UC, and JIA), and LN/IA (the CWC, BBC, and Kow-OVIA) (Fsc 0.00274, p = 0.24891 + −0.00436; Fct 0.02553, p = 0.00000 + −0.00000). The obtained

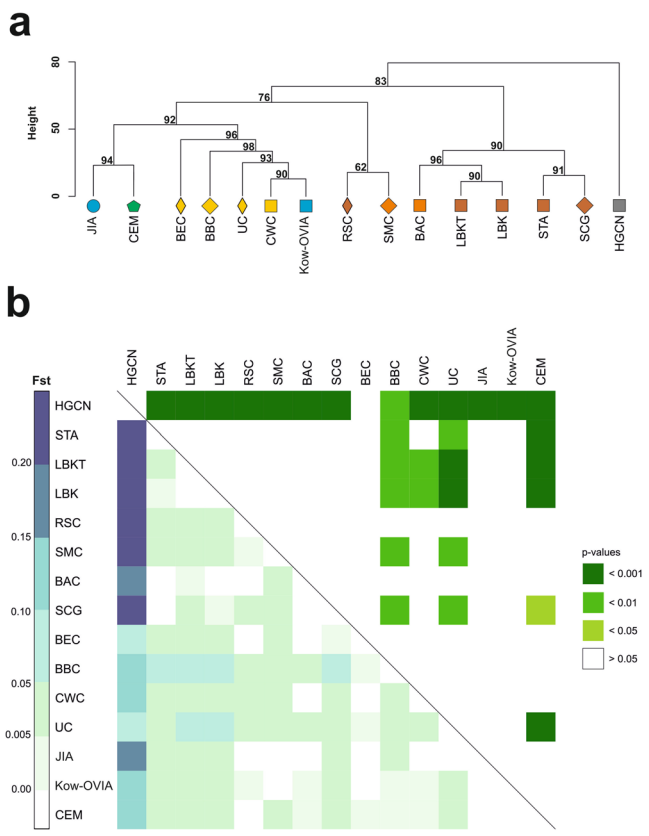


Figure 2. (**a**) Unsupervised hierarchical clustering with the Ward method and Euclidean distance on haplogroup frequencies for the CEPT populations. P-values of the clusters are given as the percent of reproduced clusters based on 10,000 bootstrap replicates. (**b**) Level plot of pairwise Fst values (below the diagonal) and corresponding p-values (above the diagonal) for the CEPT populations. P-values are based on 10,000 simulations and accounted for multiple comparisons with Benjamini-Hochberg corrections.

results suggest that the JIA was closely related to the North-Central Europe populations, whereas Kow-OVIA was related to the populations of both the Eastern and Western regions of Europe.

Next, we performed an analogous analysis but the studied group was extended by CEM (CEPT with excluded HG populations). Once again, we assumed that the populations of the EN/MN (STA, LBKT, LBK, RSC, SCG, BAC, and SMC) are a homogenous group, and we analyzed 33 combinations of the other fossil populations and CEM (Supplementary Table S14). The combination of 3 groups (the STA, LBKT, LBK, RSC, SCG, BAC, and SMC; the BEC, and UC); and the CWC, BBC, Kow-OVIA, JIA, and CEM) revealed better optimization of Fsc (0.00242, $p = 0.29426 + -0.00396$) and Fct (0.02166, $p = 0.00050 + -0.00022$) than for the two-group correlation (STA, LBKT, LBK, RSC, SCG, BAC, and SMC; and the BEC, CWC, BBC, UC, Kow-OVIA, JIA, and CEM) ($Fsc = 0.00593$, $p = 0.03545 + -0.00175$; $Fct = 0.02253$, $p = 0.00050 + -0.00022$). We also found that better optimization of the Fsc and Fct parameters was observed for the combinations where Kow-OVIA and JIA formed a group with the CEM, contrary to the combinations where the CEM was grouped with the LN/EBA populations (Supplementary Table S14). The results of the above analyses suggested that the contribution of the particular populations to the genetic structure of present-day Europe is not fully consistent with the current chronology of which populations inhabited Central Europe. The JIA, Kow-OVIA, BBC and CWC contributed more significantly to the genetic structure of present-day Europe than the BEC and UC.

Analysis of mtDNA lineages. To obtain more evidence to verify the above observations, we determined which populations gave rise to haplotypes found in Kow-OVIA (Supplementary Table S15). The same analysis of shared ancestral haplotypes[10] was performed for JIA (Supplementary Table S15). We observed a higher incidence of the HG mtDNA lineages in Kow-OVIA (9.68%) than in JIA (0%) and CEM (2%). Simultaneously, Kow-OVIA showed a lower incidence of the lineages originating from the LN/EBA (6.45%) compared to the JIA (29.16%) and CEM (13.62%).

Moreover, we performed a traditional analysis of shared haplotypes[30] (Supplementary Table S16), which demonstrated that approximately 29.4% of the haplotypes inherited by the Kow-OVIA were common for LN/EBA populations (BBC, CWC, and UC). These haplotypes came from haplogroups H and U5a. The haplotype that belonged to U5a, according to current knowledge, first occurred in the LN/EBA, whereas the H haplotypes showed continuation from the EEF (Starčevo). The JIA inherited only approximately 23% of the haplotypes shared by LN/EBA populations (BBC, CWC, UC). Approximately 41% of Kow-OVIA and 19% of JIA haplotypes were not found in any LN/EBA population although the latter preceded them.

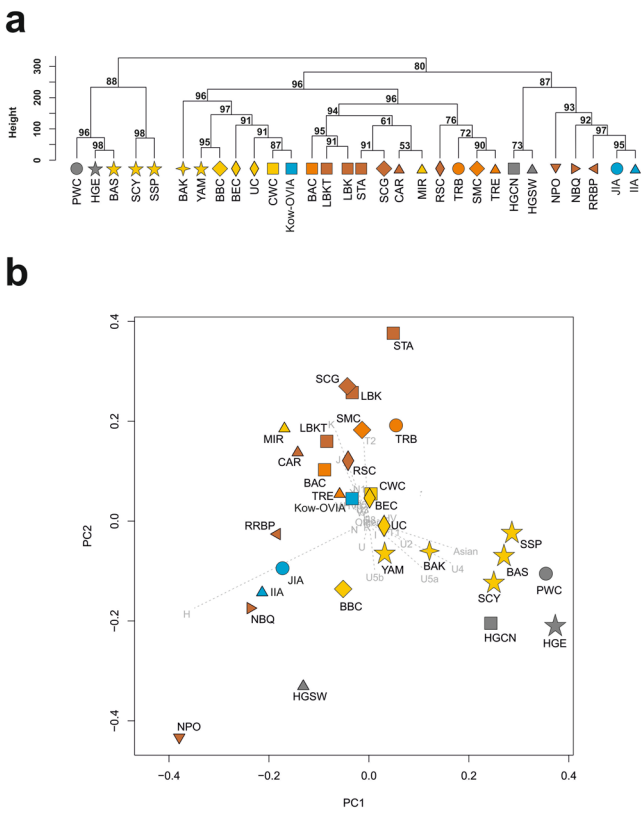


Figure 3. (**a**) Unsupervised hierarchical clustering with the Ward method and Euclidean distance on haplogroup frequencies for the EPT populations. P-values of the clusters are given as the percent of reproduced clusters based on 10,000 bootstrap replicates. (**b**) PCA on the haplogroup frequencies of EPT populations. Symbols indicate populations from Central Europe (squares and diamonds), Southern Scandinavia and Jutland Peninsula (circles), Iberian Peninsula (triangles), and East Europe/Asia (stars). Color shading of data points denotes Hunter-Gatherers (gray), Early Neolithic (brown), Middle Neolithic (orange), Late Neolithic/Early Bronze Age (yellow) and Iron Age (blue). The first two principal components of the PCA display 48.4% of the total genetic variation. Each haplogroup was superimposed as component loading vectors (gray dotted lines) proportionally to their contribution. Abbreviations: Central/North European Hunter-Gatherers (HGCN), Southwestern European Hunter-Gatherers (HGSW), East European Hunter-Gatherers (HGE), Starčevo Culture population (STA), Linearbandkeramik in Transdanubia (LBKT), Linearbandkeramik population from Central Europe (LBK), Rössen Culture (RSC), Schöningen Group (SCG), Baalberge Culture (BAC), Salzmünde Culture (SMC), Bernburg Culture (BEC), Corded Ware Culture (CWC), Bell Beaker Culture (BBC), Unetice Culture (UC), Pitted Ware culture (PWC), Funnel Beaker culture (TRB), Jutland Iron Age (JIA), Cardial/Epicardial culture of the Iberian Peninsula (CAR), Portuguese Neolithic population (NPO), Neolithic population from Basque Country and Navarre (NBQ), Iberian Chalcolithic El Mirador Cave individuals (MIR), individuals from Iberian Iron Age period (IIA), Treilles Culture (TRE), Gurgy 'Les Noisats' group (RRBP), Bronze Age Kurgan samples from South Siberia (BAS), Bronze Age Kazakhstan (BAK), Yamnaya (YAM), Iron Age Scythian (SCY), Scytho-Siberian Pazyryk Culture (SSP), Kowalewko Oder and Vistula Iron Age (Kow-OVIA).

Sex-linked factors affecting the genetic structure of Kow-OVIA. Finally, we attempted to determine whether the processes shaping the genetic structure of the Kow-OVIA population were biased by sex. To this end, based on anthropological and genetic sex assignment (Table 1), we divided the studied group into female and male subgroups – Kow-OVIA-F (17 individuals) and Kow-OVIA-M (14 individuals), respectively. Both subgroups were subjected to the analogical analyses as previously performed for the entire Kow-OVIA.

We found that the Kow-OVIA-F had a higher π value than the Kow-OVIA-M: 0.009185+/−0.005470 and 0.006987 0.004413, respectively (Supplementary Table S6). However, the difference was not statistically significant. Unsupervised hierarchical clustering involving CEPT showed that Kow-OVIA-M, similar to Kow-OVIA, grouped with the JIA and CEM whereas Kow-OVIA-F was placed in the diagram among the EN and MN populations (Fig. 5a and Supplementary Table S8).

PCA analysis of Kow-OVIA-F, Kow-OVIA-M and EPT populations (Fig. 5b and Supplementary Table S9) once again showed that Kow-OVIA-M was located similarly to Kow-OVIA, i.e., near the BBC and JIA due to considerable sharing of haplogroups H and U5b. However, the Kow-OVIA-F was grouped with the EN and MN populations, which was mainly a result of a high prevalence of the haplogroup K. The existence of genetic differences between females and males was also supported by the Fisher's test on the haplogroup frequencies ($p = 0.013$).

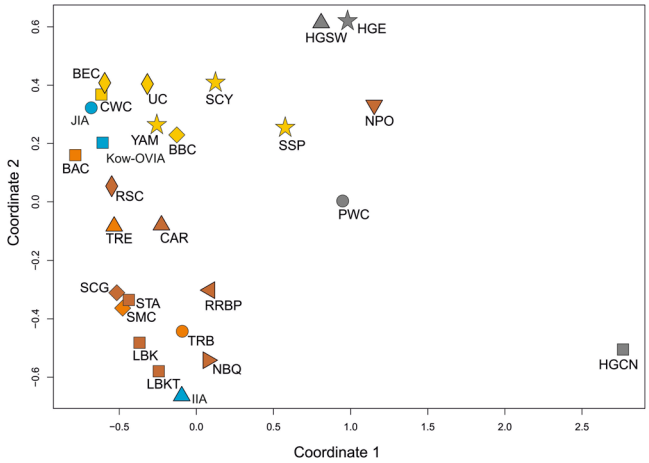


Figure 4. MDS plot of Slatkin's Fst values for EPT populations. Fst values were obtained for mtDNA HVS-I region (16064–16400 np). Symbols and color shading as in Fig. 2.

The PCA of haplogroup frequencies of Kow-OVIA-F/M and 73 extant worldwide populations again revealed high genetic differences between Kow-OVIA-F and Kow-OVIA-M (Supplementary Fig. S3b and Table S10). Although for both, the closest are contemporary European populations (PC1 and PC2), Kow-OVIA-F and Kow-OVIA-M occupy opposite sites on PC3 (in relation to the contemporary European populations).

Analysis of genetic distances between the Kow-OVIA-F/M and particular populations of the CEPT (Supplementary Table S17) also showed significant differences in the genetic history of the mtDNA lineages of male and female individuals. Kowalewko males showed the smallest genetic distance to the JIA (Fst = −0.01249, p = 0.74854), whereas females had smallest genetic distance to the BEC (Fst = −0.01598, p = 0.72236, respectively). The genetic distances between Kow-OVIA-F/M and BBC/CWC/UC/CEM were asymmetrical. The Kow-OVIA-M, as well as JIA, showed lower distances to CWC, UC and CEM than Kow-OVIA-F. Kow-OVIA-M was also located closer to BBC than Kow-OVIA-F (Fst = 0.00900, p = 0.41444, Fst = 0.01372, p = 0.39184, respectively). In this case, the distance between JIA and BBC (Fst = 0.02429, p = 0.13348) was twice as large as the distance between Kow-OVIA-F and BBC and much longer than between Kow-OVIA-M and BBC. The comparison of the Kow-OVIA-M and Kow-OVIA-F with older populations showed that: (i) the subgroup of women was genetically closer to the EN/MN populations (not including the BAC) than the male individuals; and (ii) the male subgroup was closer to the LN populations. We also verified the locations of Kow-OVIA-F and Kow-OVIA-M within the EPT set (Supplementary Table S18). The MDS analysis (Fig. 5c) demonstrated that the male individuals had smaller genetic distances to HG populations (HGCN, HGSW, HGE, and PWC) than did women and were closer to YAM than women (Fst = −0.00171, p = 0.48936; Fst = 0.01696, p = 0.21600, respectively). The analysis of shared haplotypes supported the described above tendencies - women from Kowalewko grouped with the EN and MN populations and were less connected with the BBC than men (Supplementary Table S19).

Discussion

Despite tremendous progress that has recently been made in archaeogenomic studies of European populations, our knowledge of the changes that occurred in the genetic structure of the peoples inhabiting Central Europe between the LN/EBA and Middle Ages is still very limited. One major problem is the scarcity of genetic data describing individuals who occupied the region east of the Oder river. To fill this gap, we performed an analysis of the mitochondrial genomes of a large group of people who lived in the area between the Oder and Vistula rivers (present-day western Poland) during the IA (the 1st and 2nd centuries AD). The studied group initially comprised 60 individuals; for 40 individuals, we assigned the mtDNA haplogroup, and for 33 of them, we determined the sequence of the entire mitochondrial genome. This dataset is particularly important to understand demographic processes that occurred in Central Europe during the IA and in the context of migrations between the 3rd and 6th centuries AD that are believed to have shaped the genetic landscape of contemporary Europe[31].

According to common knowledge, the region between the Oder and Vistula rivers 2 tya was densely wooded with single isolated human settlements scattered in the forest. It is thought that inhabitants of these settlements had very limited contact not only with the so-called "civilized world" but also with neighboring populations. Interestingly, the picture emerging from the latest archaeological studies of burials and artifacts seems to be different. Recently, it has been suggested that the settlement in Greater Poland was compact and regular, with the appearance and disappearance of the sites occupied by the arriving groups interacting with the surrounding populations[32] and references therein. Our results strongly support the hypothesis presented by[32]. The analyses of intrapopulation variability revealed that the studied group was not an isolated population. The calculated HD and π levels fell within the range of values characteristic for contemporary open European populations.

Based on our results of mtDNA haplogroup frequency analysis, we were able to update a phylogenetic tree describing the history of *H. sapiens* in Central Europe (see Fig. 2a). Compared with the previous phylogenetic tree[2], the new tree was expanded with the IA populations inhabiting the region between Oder and Vistula rivers

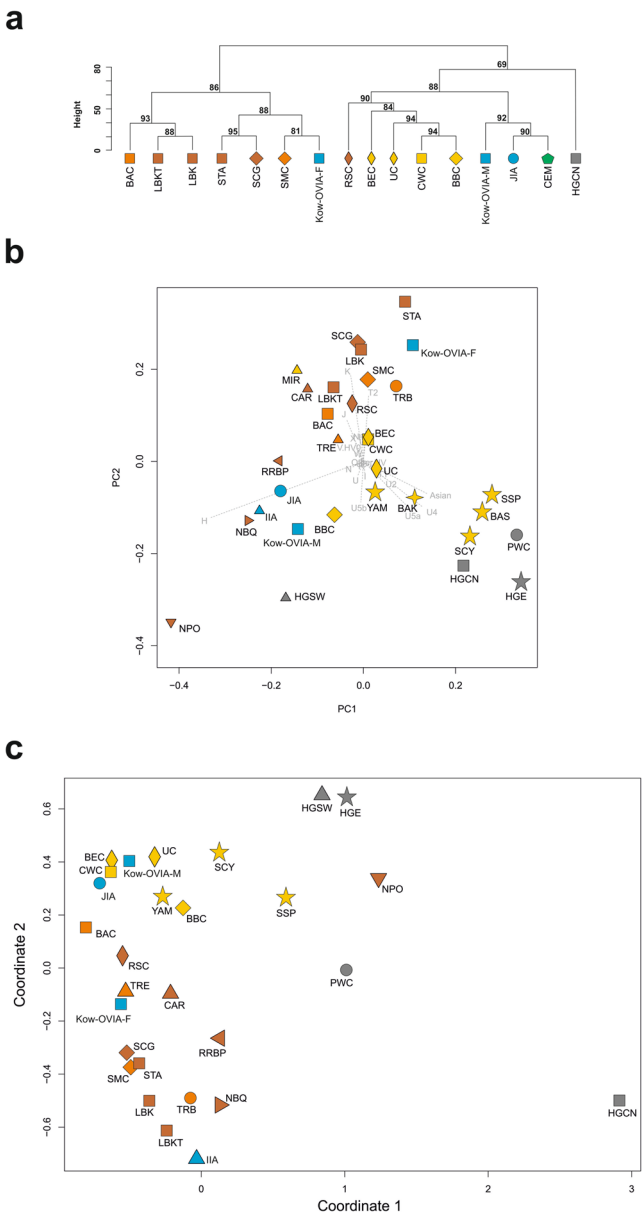


Figure 5. Analysis of haplogroup frequencies in the EPT populations, with Kow-OVIA female (Kow-OVIA-F) and male (Kow-OVIA-M) subgroups: (**a**) unsupervised hierarchical clustering with the Ward method and Euclidean distance, P-values of the clusters are given as the percent of reproduced clusters based on 10,000 bootstrap replicates; (**b**) PCA, each haplogroup was superimposed as component loading vectors (gray dotted lines) proportionally to their contribution; (**c**) MDS plot of Slatkin's Fst values, obtained for mtDNA HVS-I region (np 16064–16400). Symbols and color shading as in Fig. 2.

and the region of Jutland. In the earlier constructed tree, CEM was closest to BBC despite the latter coexisting in Central Europe with the CWC (LN) and followed by the UC (EBA). In the new tree, Kow-OVIA was clustered with the LN/EBA populations. The present-day European population (CEM) was clustered with the JIA that was contemporary to Kow-OVIA and inhabited Jutland.

Based on the results presented above one can assume that there are some discrepancies between the hierarchical clustering (Fig. 3a) and PCA (Fig. 3b) of haplogroup frequencies in the EPT (for example, the placement of the Kow-OVIA in relation to the Treilles Culture population (TRE)). These differences indicate that in contrast to dendrogram, the PCA plot (capturing only first two principal components) did not show the full spectrum of EPT variability. We explored this issue by plotting additional principal components (Supplementary Figure S7a and b). It demonstrated that their contribution to the explanation of the genetic variability of EPT populations was non-trivial.

Analysis of genetic distances (see Fig. 2b) showed that both JIA and Kow-OVIA, are the closest to the CEM. However, it should be mentioned that many of the resulting genetic proximities did not reach statistical significance at the alpha level 0.05 (mainly due to the multiple comparisons), thus they should be interpreted with

caution (Supplementary Table S11). Higher prevalence of the mtDNA haplogroup H in Kow-OVIA and JIA (its high level is also characteristic for the BBC) than in the preceding CWC and UC supports the hypothesis assuming significant demographic changes in Central Europe after the LN/EBA period[2]. This hypothesis is additionally strengthened by the results of AMOVA analysis indicating that there is some inconsistency between genetic distances and the chronology of the appearance of the studied populations in Central Europe, i.e., the older populations (BBC, CWC) contributed more to the genetic structure of CEM than the younger ones (UC).

Changes in the occurrence of mtDNA haplogroups U5a/U5b in Central Europe are also worth noting. At LN and EBA, the prevailing haplogroup was U5a for BBC/CWC/UC. Next, there was a dominance of U5b for the Kow-OVIA/JIA during IA and now U5a is again more popular (CEM). The first alteration in the U5a/U5b prevalence between the LN/EBA and the IA supports the hypothesis of demographic changes right after the LN, proposed by[2]. The second conversion indicated by our results suggests another crucial demographic event that should occur between the IA and present.

On the basis of the above observations, one may assume that in the IA, specific genetic substructures were formed in Central Europe. Because the demographic history of fossil populations often has a local character[33,34], it is worth considering the range of the observed changes. These considerations should also take into account the hypothesis on the migrations that most likely occurred between the 3rd and 6th century AD. In this context, it seems necessary to compare Kow-OVIA and JIA with other populations from the IA, in particular those located east of Vistula, and with the populations that inhabited this region during the Middle Ages.

The process of demographic change in Central and Northern Europe after the LN/EBA appears to have a complex nature. Genetic proximity of the Kow-OVIA and JIA is consistent with the Kow-OVIA's affiliation to the Wielbark archaeological culture, strictly connected with Baltic regions[32]. However, despite close genetic distance between Kow-OVIA and the JIA, PCA placed them asymmetrically in relation to other ancient populations (see Fig. 3b). According to MDS (Fig. 4), JIA was close to the North-East European populations (CWC, BEC, UC) whereas the relationships of Kow-OVIA with other populations (earlier and contemporary to Kow-OVIA) were not so obvious. Interestingly, a small genetic distance between the JIA and UC was correlated with a high prevalence of the mtDNA haplogroup I in both populations. This result is consistent with earlier hypotheses suggesting that the genetic structure of a contemporary Danish population was formed not later than in the IA[35]. Unfortunately, our knowledge of haplogroup I prevalence in the Nordic Bronze Age is still scarce because of the small number of analyzed individuals; thus, it cannot be unambiguously stated that the observed proximity of the JIA and UC was, indeed, a result of the demographic changes after the LN/EBA. However, the above conclusion postulating a close connection between JIA and UC is also supported by the result of shared haplotype analysis. We found that 25% of the ancestral haplotypes found in the JIA were first reported in the UC and were not common in any earlier population. To better understand this phenomenon, the analysis of Y chromosome haplogroups is required, as the Nordic Bronze Age is characterized by the occurrence of the Y-haplogroups I and I1[18], whereas the UC is mainly characterized by a higher prevalence of I2[13]. More detailed Y chromosome analyses involving a larger number of individuals would also shed more light on the process that resulted in a high prevalence of the mtDNA H haplogroup in the JIA, which is another signal of post-LN demographic changes[2]. In the case of Kow-OVIA, its genetic root in multiple European populations is evidenced by the fact that 55% of the ancestral haplotypes that were identified in this group were common to populations of the LN/EBA period (the CWC, BBC, and UC). For the JIA, the same origin is observed for only 19% of ancestral haplotypes.

Finally, we found that the genetic structures of female and male subpopulations of Kow-OVIA were significantly different. This fact cannot be explicitly determined based on the results of individual analyses; however, it is quite evident if one considers the whole set of data presented here including the Fisher test on haplogroup frequencies. The analyses of both mtDNA haplogroups and genetic distances indicated that women from Kowalewko were related closer to the EN/MN populations, and the men were closer to the CWC and UC. This observation may explain why the genetic relationships of Kow-OVIA with other ancient European populations were more complex and more difficult to define as it was in the case of JIA. In analyzing Kow-OVIA, we observed multiple overlapping effects of two subpopulations with different genetic affinities. One would speculate that the genetic profile of Kow-OVIA-F resulted from exogamy that was described for the CWC population[36]. This is, however, not the case. We found that the genetic differences between women and men were maintained for the entire observation period, i.e., for 200 years (approximately 8 generations). Such a composition of the genetic structure of Kow-OVIA could exist only if at least one subgroup (Kow-OVIA-F or -M) was periodically exchanged. It would further mean that Kowalewko played some specific roles in that region. According to the recent archaeological studies, the colonization pattern in IA Greater Poland could be linked with the existence of a centralized organization system[32]. Kowalewko could have been one of the important elements of this system. For example, it could have functioned as a garrison for the population closely associated with the JIA, such that warriors stayed in the garrison for only a few years and were then replaced by others. Other scenarios are also possible; however, verification of any hypothesis requires more detailed studies.

Materials and Methods

aDNA extraction and library preparation. For the purposes of this study, 63 samples (teeth) were obtained from 60 individuals from the Kowalewko archaeological site (Supplementary Table S1). After being transported to a clean aDNA lab, the teeth were cleaned with 5% NaOCl and rinsed with sterile water, followed by UV irradiation (254 nm) for 2 hours per site. The roots of the teeth were drilled using Dremel® drill bits. Bone powder (approximately 250 mg) was digested with proteinase K and DNA-containing extract was purified using a silica-based method following[24] and[37]. Genomic libraries were prepared following a protocol from[38] omitting the initial sonication step, due to the natural fragmentation of aDNA. Six separate PCR reactions were set up for each library. PCR amplifications were performed in 25 µl with 3 µl of the DNA library template, 12.5 µl of 1x AmpliTaq Gold® 360 Master Mix (Life Technologies, California), 0.5 µl of indexing primer (10 µM) and 0.5 µl of PCR primer

IS4 (10 µM) (Günther *et al*. 2015). The PCR profile was as follows: initial denaturation (94 °C, 12 min), 12–16 cycles of 94 °C (30 s), 60 °C (30 s), 72 °C (45 s) and final extension (72 °C, 10 min). PCR reactions for the same library were pooled and purified with AMPure® XP beads (Agencourt-Beckman Coulter) following[39]. Quality and size distribution of the libraries were verified with a High Sensitivity DNA kit and 2100 Bioanalyzer system (Agilent) while DNA concentration was determined with a Qubit fluorimeter and Qubit dsDNA HS Assay Kit (ThermoFisher Scientific), according to the manufacturers' protocols.

Next generation sequencing of aDNA libraries. For shallow sequencing, Genome Analyzer GAIIx (Illumina) and TruSeq SBS Kits v5-GA (Illumina) were applied. Seven to nine libraries were pooled per lane. On average, 4.75 mln 75 bp-long reads were collected per library. The libraries with the highest human DNA content were sent for deep sequencing to Macrogen Inc., Korea, with the use of HiSeq4000 (Illumina). On average, 87 mln 101 bp-long reads were collected per library on the HiSeq4000 (Illumina) (Supplementary Table S2).

Filtering, mapping and variant calling. Raw sequencing data (fastq files) were filtered with AdapterRemoval[40] by trimming missing nucleotides from both ends with the threshold of a minimum quality of 30 and a minimum length of 25 nucleotides. Filtered reads were aligned with BWA ver. 0.7.10[41] to the rCRS mitochondrial, and GRCH 37 reference genomes, with seed blocked for higher sensitivity and other parameters set as default, as suggested in[42]. Following the alignment, duplicate reads were removed with picard-tools ver. 1.117 MarkDuplicates. Read depth and coverage were assessed with samtools ver. 1.2[43] and bedtools ver. 1.2[44]. Consensus fasta sequences for haplogroup prediction and sequence analyses were generated with FreeBayes ver. 1.0.2–33-gdbb6160[45] calling the base supported by the 3/5 majority of reads (Supplementary Table S4).

Human DNA damage patterns. To examine data authenticity, we used mapDamage 2.0[46] to estimate human DNA damage parameters typical for aDNA for each sample: (i) λ, the fraction of nucleotides positioned in the single-stranded DNA overhang context; (ii) δs, C→T substitution rate in the single-stranded overhang context; and (iii) δd, C→T substitution rate in the double-stranded DNA context (Supplementary Table S3).

Contamination assessment. To estimate the level of contamination with contemporary human DNA or with other human aDNA we used the software contamMix to compare the alignment rates between estimated sample's consensus mtDNA sequence and 311 potential contaminant mtDNA sequences.

Analysis of intrapopulation genetic diversity. To assess intrapopulation diversity of Kow-OVIA, we used the MSN method on 33 full mtDNA sequences. HD was analyzed with Arlequin ver. 3.5.1[47] on two fragments of mtDNA HVS region (HVS-I, 16033–16365 np, and HVS-II 73–340 np) (Supplementary Table S5). π was calculated for the fragment of mtDNA HVS-I region (16000–16410 np) (Supplementary Table S6).

mtDNA haplogroup frequency analyses. Haplogroups were predicted with HaploFind[48] with respect to Phylotree build 17 (http://www.phylotree.org/)[49]. Only samples with a haplogroup score ≥ 0.8 and average coverage ≥ 3 were used in downstream analyses. To assess temporal mtDNA haplogroup frequency changes from Mesolithic to the present day, we performed Ward clustering with Euclidean distance on 23 haplogroups (H, H5, HV, HV0, V, I, J, K, N, N1a, R, T1, T2, U, U2, U3, U4, U5a, U5b, U8, W, X and others). We used a set of 40 samples from this study and 13 populations from Central/North Europe (Supplementary Table S7), and a generated Central European Metapopulation (CEM) composed of 500 random individuals sampled from the extant populations of Poland, Czech Republic, Germany and Austria, as in[2] (Supplementary Table S8).

To compare mtDNA variability of Kow-OVIA in a broader geographical context, we also applied unsupervised hierarchical clustering with the Ward method and Euclidean distance, and PCA on haplogroup frequencies of Kow-OVIA and published ancient mtDNA data from Mesolithic, Neolithic, Bronze Age and Iron Age populations from across Europe and West Asia (Supplementary Table S7). Haplogroups were divided into 25 groups present in ancient individuals (Asian [A, C, D, F, G, Z], N, N1a, I, J, W, X, R, HV, V/HV0, H, H5, T, T1, T2, J, U, U2, U3, U4, U5a, U5b, U8, K and others) (Supplementary Table S9).

To elucidate affinities of our samples in relation to present day populations, we performed a PCA analysis based on 23 haplogroup frequencies (Asia [A, C, D, F, G, Z], Africa [L], N1a, I, I1, W, X, HV, V/HV0, H, H5, T1, T2, J, U, U2, U3, U4, U5a, U5b, U8, K and others), with public data from 73 extant populations (Supplementary Table S10).

Cluster significance was tested by performing 10,000 permutations with the pvclust package in R ver. 3.3.0. PCA was conducted with prcomp of the vegan package in R ver. 3.3.0 (http://R-project.org).

Genetic distance analyses. For sequence based analyses, the longest mtDNA HVS-I fragment present in the biggest fraction of published samples was selected (16064–16400 np). Additionally, for newly reported samples from this study, no missing nucleotides were allowed in the selected HVS-I range, and at least 3x coverage was expected for 95% of the nucleotides. To examine genetic affinities on the sequence level, we calculated genetic distances (Fst)[50] between two sample sets: CEPT (Supplementary Table S11 and S17) and EPT (Supplementary Table S12 and S18), including only those individuals for which the 16064–16400 fragment of mtDNA HVS-I was present.

Pairwise and Slatkin's Fst[51] values were calculated in Arlequin ver. 3.5.1 for both datasets separately, with associated substitution model and gamma values selected with jModel test 0.1 AIC and BIC[52]. P values were calculated by performing 10,000 permutations. Genetic distances were visualized on an MDS plot with metaMDS function from the vegan package in R ver. 3.3.0.

Analysis of genetic structure. To examine if genetic affinities between particular populations from CEPT were a result of shared genetic structure, we conducted AMOVA analysis in 18 combinations of CEPT populations (excluding HG and CEM) (Supplementary Table S13). Next, we tested the genetic contributions of ancient populations to the extant mtDNA variability by analyzing AMOVA results from 33 combinations of CEPT populations (excluding HG) (Supplementary Table S14). Statistical significance was obtained by performing 10,000 permutation tests.

mtDNA lineage analyses. To further test genetic affinities and account for temporal succession of archaeological cultures in Central Europe, we conducted an analysis of shared ancestral haplotypes as described in[10] in CEPT (Supplementary Table S15) and classical shared haplotype analysis[30] in LN/EBA and IA populations (Supplementary Table S16 and S19).

For details, see the Supplementary Materials.

Availability of data and material. The data supporting the results of this article are available at SRA, BioProject PRJNA354503.

References

1. Bramanti, B. *et al*. Genetic discontinuity between local hunter-gatherers and central Europe's first farmers. *Science* **326**, 137–140, https://doi.org/10.1126/science.1176869 (2009).
2. Brandt, G. *et al*. Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. *Science* **342**, 257–261, https://doi.org/10.1126/science.1241844 (2013).
3. Soares, P. *et al*. The archaeogenetics of Europe. *Current biology: CB* **20**, R174–183, https://doi.org/10.1016/j.cub.2009.11.054 (2010).
4. Pinhasi, R., Thomas, M. G., Hofreiter, M., Currat, M. & Burger, J. The genetic history of Europeans. *Trends in genetics: TIG* **28**, 496–505, https://doi.org/10.1016/j.tig.2012.06.006 (2012).
5. Fu, Q. *et al*. The genetic history of Ice Age Europe. *Nature* **534**, 200–205, https://doi.org/10.1038/nature17993 (2016).
6. Lazaridis, I. *et al*. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413, https://doi.org/10.1038/nature13673 (2014).
7. Nielsen, R. *et al*. Tracing the peopling of the world through genomics. *Nature* **541**, 302–310, https://doi.org/10.1038/nature21347 (2017).
8. Fu, Q. *et al*. A revised timescale for human evolution based on ancient mitochondrial genomes. *Current biology: CB* **23**, 553–559, https://doi.org/10.1016/j.cub.2013.02.044 (2013).
9. Posth, C. *et al*. Pleistocene Mitochondrial Genomes Suggest a Single Major Dispersal of Non-Africans and a Late Glacial Population Turnover in Europe. *Current biology: CB* **26**, 827–833, https://doi.org/10.1016/j.cub.2016.01.037 (2016).
10. Szecsenyi-Nagy, A. *et al*. Tracing the genetic origin ofEurope's first farmers reveals insights into their social organization. P*roceedings. Biological sciences/The Royal Society***282**, https://doi.org/10.1098/rspb.2015.0339 (2015).
11. von Cramon-Taubadel, N. & Pinhasi, R. Craniometric data support a mosaic model of demic and cultural Neolithic diffusion to outlying regions ofEurope. *Proceedings. Biological sciences/The Royal Society* **278**, 2874–2880, https://doi.org/10.1098/rspb.2010.2678 (2011).
12. Hofmanova, Z. *et al*. Early farmers from across Europe directly descended from Neolithic Aegeans. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 6886–6891, https://doi.org/10.1073/pnas.1523951113 (2016).
13. Mathieson, I. *et al*. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503, https://doi.org/10.1038/nature16152 (2015).
14. Haak, W. *et al*. Ancient DNA from European early neolithic farmers reveals their near eastern affinities. *PLoS biology* **8**, e1000536, https://doi.org/10.1371/journal.pbio.1000536 (2010).
15. Haak, W. *et al*. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211, https://doi.org/10.1038/nature14317 (2015).
16. Lipson, M. *et al*. Parallel palaeogenomic transects reveal complex genetic history of early European farmers. *Nature* **551**, 368–372, https://doi.org/10.1038/nature24476 (2017).
17. Skoglund, P. *et al*. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* **336**, 466–469, https://doi.org/10.1126/science.1216304 (2012).
18. Allentoft, M. E. *et al*. Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172, https://doi.org/10.1038/nature14507 (2015).
19. Lorkiewicz, W. *et al*. Between the Baltic and Danubian Worlds: the genetic affinities of a Middle Neolithic population from central Poland. *PloS one* **10**, e0118316, https://doi.org/10.1371/journal.pone.0118316 (2015).
20. Juras, A. *et al*. Ancient DNA reveals matrilineal continuity in present-day Poland over the last two millennia. *PloS one* **9**, e110839, https://doi.org/10.1371/journal.pone.0110839 (2014).
21. Chylenski, M. *et al*. Late Danubian mitochondrial genomes shed light into the Neolithisation of Central Europe in the 5th millennium BC. *BMC evolutionary biology* **17**, 80, https://doi.org/10.1186/s12862-017-0924-0 (2017).
22. Skorupka, T. *Kowalewko 12. Cmentarzysko birytualne ludności kultury wielbarskiej (od połowy I w. n.e. do początku III w. n.e.)*. (Wydawnictwo Poznańskie, 2001).
23. Philips, A. *et al*. Comprehensive analysis of microorganisms accompanying human archaeological remains. *GigaScience*, https://doi.org/10.1093/gigascience/gix044 (2017).
24. Yang, D. Y., Eng, B., Waye, J. S., Dudar, J. C. & Saunders, S. R. Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *American journal of physical anthropology* **105**, 539–543, https://doi.org/10.1002/(SICI)1096-8644(199804)105:4 < 539::AID-AJPA10 > 3.0.CO;2-1 (1998).
25. Svensson, E. M. *et al*. Tracing genetic change over time using nuclear SNPs in ancient and modern cattle. *Animal genetics* **38**, 378–383, https://doi.org/10.1111/j.1365-2052.2007.01620.x (2007).
26. Excoffier, L., Smouse, P. E. & Quattro, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491 (1992).
27. Cassens, I., Mardulyn, P. & Milinkovitch, M. C. Evaluating intraspecific "network" construction methods using simulated sequence data: do existing algorithms outperform the global maximum parsimony approach? *Systematic biology* **54**, 363–372, https://doi.org/10.1080/10635150590945377 (2005).
28. Capocasa, M. *et al*. Detecting genetic isolation in human populations: a study of European language minorities. *PloS one* **8**, e56371, https://doi.org/10.1371/journal.pone.0056371 (2013).
29. Jorde, L. B. *et al*. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *American journal of human genetics* **66**, 979–988, https://doi.org/10.1086/302825 (2000).
30. Csakyova, V. *et al*. Maternal Genetic Composition of a Medieval Population from a Hungarian-Slavic Contact Zone in Central Europe. *PloS one* **11**, e0151206, https://doi.org/10.1371/journal.pone.0151206 (2016).
31. Vai, S. *et al*. Genealogical relationships between early medieval and modern inhabitants of Piedmont. *PloS one* **10**, e0116801, https://doi.org/10.1371/journal.pone.0116801 (2015).

32. Teska, M. & Michałowski, A. Connection between Wielkopolska and the Baltic Sea Region in the RomanIron Age. *Archaeologia Lituana* **14**, 63–77 (2013).
33. Martiniano, R. *et al*. The population genomics of archaeological transition in west Iberia: Investigation of ancient substructure using imputation and haplotype-based methods. *PLoS genetics* **13**, e1006852, https://doi.org/10.1371/journal.pgen.1006852 (2017).
34. Nunez, C. *et al*. Mitochondrial DNA Reveals the Trace of the Ancient Settlers of a Violently Devastated Late Bronze and Iron Ages Village. *PloS one* **11**, e0155342, https://doi.org/10.1371/journal.pone.0155342 (2016).
35. Melchior, L., Gilbert, M. T., Kivisild, T., Lynnerup, N. & Dissing, J. Rare mtDNA haplogroups and genetic differences in rich and poor Danish Iron-Age villages. *American journal of physical anthropology* **135**, 206–215, https://doi.org/10.1002/ajpa.20721 (2008).
36. Sjogren, K. G., Price, T. D. & Kristiansen, K. Diet and Mobility in the Corded Ware of Central Europe. *PloS one* **11**, e0155083, https://doi.org/10.1371/journal.pone.0155083 (2016).
37. Malmstrom, H. *et al*. More on contamination: the use of asymmetric molecular behavior to identify authentic ancient human DNA. *Molecular biology and evolution* **24**, 998–1004, https://doi.org/10.1093/molbev/msm015 (2007).
38. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor protocols* **2010**, pdbprot5448, https://doi.org/10.1101/pdb.prot5448 (2010).
39. Gunther, T. *et al*. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 11917–11922, https://doi.org/10.1073/pnas.1509851112 (2015).
40. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC research notes* **9**, 88, https://doi.org/10.1186/s13104-016-1900-2 (2016).
41. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595, https://doi.org/10.1093/bioinformatics/btp698 (2010).
42. Schubert, M. *et al*. Improving ancient DNA read mapping against modern reference genomes. *BMC genomics* **13**, 178, https://doi.org/10.1186/1471-2164-13-178 (2012).
43. Li, H. *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079, https://doi.org/10.1093/bioinformatics/btp352 (2009).
44. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842, https://doi.org/10.1093/bioinformatics/btq033 (2010).
45. Hwang, S., Kim, E., Lee, I. & Marcotte, E. M. Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Scientific reports* **5**, 17875, https://doi.org/10.1038/srep17875 (2015).
46. Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684, https://doi.org/10.1093/bioinformatics/btt193 (2013).
47. Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources* **10**, 564–567, https://doi.org/10.1111/j.1755-0998.2010.02847.x (2010).
48. Vianello, D. *et al*. HAPLOFIND: a new method for high-throughput mtDNA haplogroup assignment. *Human mutation* **34**, 1189–1194, https://doi.org/10.1002/humu.22356 (2013).
49. van Oven, M. Revision of the mtDNA tree and corresponding haplogroup nomenclature. *Proceedings of the National Academy of Sciences of the United States of America* **107**, E38-39; author reply e40-31, https://doi.org/10.1073/pnas.0915120107 (2010).
50. Wright, S. Genetical structure of populations. *Nature* **166**, 247–249 (1950).
51. Slatkin, M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457–462 (1995).
52. Posada, D. jModelTest: phylogenetic model averaging. *Molecular biology and evolution* **25**, 1253–1256, https://doi.org/10.1093/molbev/msn083 (2008).

Acknowledgements

We are grateful to Edvard Ehler from the Charles University in Prague, Prague, Czech Republic and the Institute of Anthropology, Adam Mickiewicz University, Poznan, Poland and Andrzej Michalowski from the Institute of Archaeology, Adam Mickiewicz University, Poznan, Poland, for stimulating discussion. We also thank to Anna Myszka and Dawid Trzciński from Institute of Anthropology, Adam Mickiewicz University, Poznań for providing samples. This work was supported by the Polish National Science Center [2014/12/W/NZ2/00466].

Author Contributions

I.S. conceived the study, participated in the study design, analyzed the data, created figures, discussed the results, and wrote the manuscript. A.J. extracted DNA, prepared NGS libraries and reviewed the manuscript. L.H. performed NGS sequencing, created figures, discussed the results and wrote the manuscript. M.M.-S., A.P. and M.Z. participated in discussions of the results and manuscript review. A.D. and H.K.-K. provided archaeological data and discussed the results. J.P. provided samples, discussed the results and reviewed the manuscript. P.K. participated in the study design, discussed the data and reviewed the manuscript. M.F. conceived the overall idea of the study, provided funding, participated in the study design, discussed the data and was responsible for the final version of the manuscript. All of the authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-20705-6.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

MATERIAŁY UZUPEŁNIAJĄCE DO PUBLIKACJI

Stolarek i wsp., Scientific Reports 2018

Spis tabel uzupełniających

Table S16. Shared Haplotypes Analysis of Central/North European LN/EBA and IA populations

Table S17. Fst and p-values for 14 ancient populations and CEM compared with Kow-OVIA-males, and Kow-OVIA-females. Presented are pair-wise and Slatkin's Fst values and pair-wise Fst p-values before and after post-hoc multiple comparisons adjustment by Benjamini-Hochberg correction

Table S18. Fst and p-values for 26 ancient populations compared with Kow-OVIA-males, and Kow-OVIA-females. Presented are pair-wise and Slatkin's Fst values and pair-wise Fst p-values before and after post-hoc multiple comparisons adjustment by Benjamini-Hochberg correction

Table S19. Shared Haplotypes Analysis (SHA) of OVIA and CEPT populations. Results shown as percent of shared HVS-I haplotypes between pairs of populations

Spis rysunków uzupełniających

Supplementary Figure S1. Location of Kowalewko and a scheme of the Kowalewko cemetery site 12, based on the Fig. 3 from the monograph by Tomasz Skorupka, Kowalewko 12. Biritual cemetery of a population of the Wielbark Culture (mid 1st to beginning of 3rd century AD), published in: Marek Chlodnicki [ed.], Archaeological rescue investigations along the gas transit pipeline, vol. II - Wielkopolska, part 3, Poznan 2001, generated using Corel Draw ver. 12.0, with the author permission. Sampled graves are marked with a red color. Europe and Poland maps were downloaded from Wikimedia Commons (https://commons.wikimedia.org), under the free license, and modified with Corel Draw ver. 12.0

Supplementary Figure S2. Results of radiocarbon (^{14}C) dating of Kow-OVIA individuals, supplied by the Poznan Radiocarbon Laboratory

Supplementary Figure S3. (a) 3D PCA on haplogroup frequencies of Kow-OVIA and 73 present day populations from Europe and the Near East. (b) 3D PCA on haplogroup frequencies of the same set of populations, with Kow-OVIA divided into female (Kow-OVIA-F) and male (Kow-OVIA-M) subgroups

Supplementary Figure S4. Unsupervised hierarchical clustering with the Ward method and Euclidean distance on haplogroup frequencies for the CEPT populations, where Kow-OVIA* contains potentially maternally related individuals. P-values of the clusters are given as the percent of reproduced clusters based on 10,000 bootstrap replicates

Supplementary Figure S5. (a) Unsupervised hierarchical clustering with the Ward method and Euclidean distance on haplogroup frequencies for the EPT populations. P-values of the clusters are given as the percent of reproduced clusters based on 10,000 bootstrap replicates. (b) PCA on the haplogroup frequencies of EPT populations. Kow-OVIA* contains potentially maternally related individuals

Supplementary Figure S6. MDS plot of Slatkin's Fst values for EPT populations. Fst values were obtained for mtDNA HVS-I region (16064-16400 np). Symbols and color shading as in Fig. 2

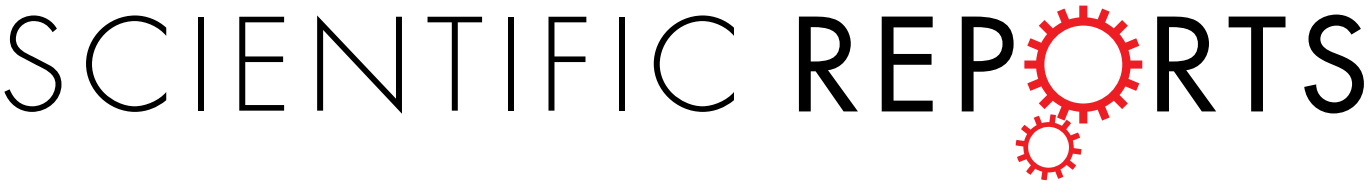
Supplementary Figure S7. PCAs on the haplogroup frequencies of EPT populations with plotted (a) PC2 and PC3, (b) PC3 and PC4

2

Ireneusz Stolarek, Luiza Handschuh, Anna Juras, Wioletta Nowaczewska, Hanna Kóčka-Krenz, Andrzej Michalowski, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Goth migration induced changes in the matrilineal genetic structure of the central-east European population

Scientific Reports volume 9, Article number: 6737 (2019)
doi: 10.1038/s41598-019-43183-w



OPEN

Goth migration induced changes in the matrilineal genetic structure of the central-east European population

Received: 24 October 2018
Accepted: 12 April 2019
Published online: 01 May 2019

I. Stolarek[1], L. Handschuh[1], A. Juras[2], W. Nowaczewska[3], H. Kóčka-Krenz[4], A. Michalowski[4], J. Piontek[2], P. Kozlowski[1] & M. Figlerowicz[1,5]

For years, the issues related to the origin of the Goths and their early migrations in the Iron Age have been a matter of hot debate among archaeologists. Unfortunately, the lack of new independent data has precluded the evaluation of the existing hypothesis. To overcome this problem, we initiated systematic studies of the populations inhabiting the contemporary territory of Poland during the Iron Age. Here, we present an analysis of mitochondrial DNA isolated from 27 individuals (collectively called the Mas-VBIA group) excavated from an Iron Age cemetery (dated to the 2nd-4th century A.D.) attributed to Goths and located near Masłomęcz, eastern Poland. We found that Mas-VBIA has similar genetic diversity to present-day Asian populations and higher diversity than that of contemporary Europeans. Our studies revealed close genetic links between the Mas-VBIA and two other Iron Age populations from the Jutland peninsula and from Kowalewko, located in western Poland. We disclosed the genetic connection between the Mas-VBIA and ancient Pontic-Caspian steppe groups. Similar connections were absent in the chronologically earlier Kowalewko and Jutland peninsula populations. The collected results seem to be consistent with the historical narrative that assumed that the Goths originated in southern Scandinavia; then, at least part of the Goth population moved south through the territory of contemporary Poland towards the Black Sea region, where they mixed with local populations and formed the Chernyakhov culture. Finally, a fraction of the Chernyakhov population returned to the southeast region of present-day Poland and established the archaeological formation called the "Masłomęcz group".

During the last decade, genetics and genomics have become new driving forces that have stimulated the rapid development of studies on the human past. Archaeogenomics has quickly evolved from analyses of single individuals to studies involving dozens of subjects[1–3]. As a result, an increasingly precise map of the genetic history of a human population is generated. Despite the progress in our understanding of the demographic processes that took place in Europe since its first peopling, the map still has a numerous blank spaces[4]. They are especially frequent in the case of the Early Bronze Age (EBA) and later periods, when more complex demographic and cultural events occurred[5–11]. Importantly, not all geographical regions have been sufficiently sampled[12]. The majority of data are confined to Central-West Europe, whereas Eastern Europe is scarcely represented[1].

Recently, the first efforts have been undertaken to determine the genetic roots of peoples living in Central-East and Eastern Europe[13]. The works of Chylenski *et al.* and Lorkiewicz *et al.*[14,15] suggest that the matrilineal genetic makeup of people living in the Vistula River Basin during the Early Neolithic (EN) and the Middle Neolithic (MN) better resembled the Funnel Beaker culture populations (TRB, ger. Trichter(-rand-)becherkultur), which were a mixture of Mesolithic Hunter Gatherers and Neolithic farmers, than the Linear Pottery culture-associated

[1]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland. [2]Department of Human Evolutionary Biology, Institute of Anthropology, Faculty of Biology, Adam Mickiewicz University, Umultowska 89, Poznan, 61-614, Poland. [3]Department of Human Biology, Faculty of Biological Sciences, Wroclaw University, Wroclaw, Poland. [4]Institute of Archaeology, Adam Mickiewicz University, Poznan, Poland. [5]Institute of Computing Sciences, Poznan University of Technology, Poznan, Poland. Correspondence and requests for materials should be addressed to M.F. (email: marekf@ibch.poznan.pl)

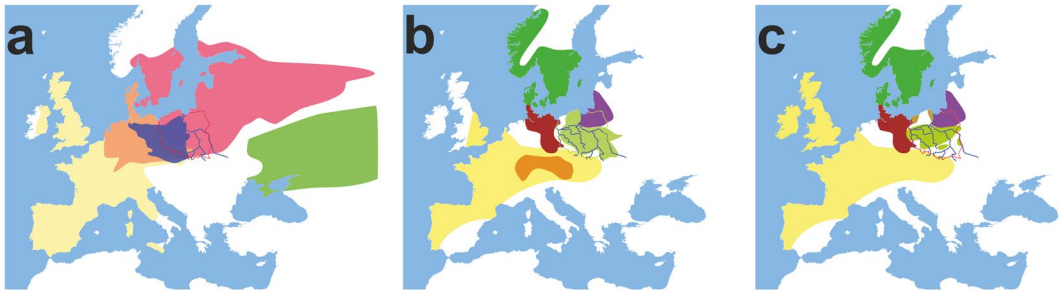


Figure 1. Geographical distribution of the archaeological cultures in Europe in the (**a**) LN/EBA: green, Yamnaya Culture 3300–2300 BC|pink, Corded Ware Culture 2900–2350 BC|yellow, Bell Beaker Culture 2900–1800 BC|orange, Corded Ware/Bell Beaker Culture|violet, Unetice Culture 2300–1600 BC; (**b**) BA/IA: green, Nordic Bronze Age 1700–500 BC|yellow, Hallstatt Culture 800–500 BC|orange, Hallstatt Culture Core|light green, Pomeranian Culture 700–300 BC|red, Jastorf Culture 600–100 BC|violet, Western Baltic Kurgans Culture 650–50BC; (**c**) IA prior to migrations period: green, Nordic Bronze Age 1700–500 BC|yellow, La Tene Culture 450–50 BC|light green, Przeworsk Culture 300 BC – 500 AD|red, Jastorf Culture 600–100 BC|violet, Western Baltic Kurgans Culture 650–50 BC|brown, Oksywie Culture 200 BC – 100 AD. Europe map by Roke, retrieved from https://commons.wikimedia.org/wiki/Category:Blank_maps_of_Europe#/media/File:BlankMap-Europe-v3.png, used under Creative Commons Attribution-ShareAlike 3.0 Unported (https://creativecommons.org/licenses/by-sa/3.0/deed.en), modified with Corel Draw ver. 12.0.

populations (LBK, ger. Linearbandkeramik), which were mostly composed of Neolithic farmers[13,16,17]. The studies of autosomal DNA of the Globular Amphora culture (GAC) population inhabiting the Vistula River Basin during the Late Neolithic (LN) showed no presence of Yamnaya steppe herder ancestry (YAM)[1,18]. At the same time, the Corded Ware culture (CWC) groups that later succeeded GAC have as much as 75% of their genetic makeup attributed to YAM (Fig. 1a)[7,17,19]. In particular, we know little about the genetic history of the populations that inhabited the territory of contemporary Poland in the Bronze Age (BA). During that time, cremation of the dead became a common custom; thus, our knowledge on these peoples is based almost exclusively on archaeological artifacts. They indicate that during the BA, several local cultures developed in this region[20]. All of them were, to a large extent, related to each other, and consequently, they are usually considered as one, called the Lusatian culture, existing from the EBA to the early Iron Age (IA). At the beginning of the IA (approximately 600 B.C.), an alternative to Lusatian culture, called the Pomeranian culture (PC), was developed. The PC dominated the region and spread between the Oder and Bug Rivers until the end of the 3rd century B.C. (Fig. 1b). During the period that immediately preceded the demographic events described in this article (the period between 1st and 2nd centuries B.C.), the lowland of contemporary central Poland was occupied by the Przeworsk culture (Fig. 1c), which replaced the earlier existing PC. The development of the Przeworsk Culture seems to be connected with a Vandal migration. In a similar period, in the regions of Middle Pomerania and Lower Powiśle (zone along the Baltic seashore), the Oksywie culture was established (see Fig. 1c).

The turn of the epochs (i.e., the end of 1st millennium B.C. and the beginning of the 1st millennium A.D.) was connected with further cultural and demographic transformations in the region that is present-day Poland[20]. A new grouping called the Wielbark culture was established, most likely under the influence of Goths and Gepids. To learn more about these processes, we have commenced systematic studies of the populations inhabiting the territory corresponding to present-day Poland in the first centuries AD. Our initial analyses were focused on a large group of individuals living in the region between Oder and Vistula Rivers in the IA (group called the Kow-OVIA)[8,21]. Biological material was collected from a biritual cemetery located in central Wielkopolska (also referred to as Greater Poland - contemporary western Poland) in Kowalewko. Former archeological and anthropological studies showed that approximately 500 people were buried there between 1 and 200 A.D. Our recently published report revealed a high genetic diversity of the Kow-OVIA[8]. Interestingly, women and men from Kowalewko had a significantly different genetic history. In general, the collected data showed how the interactions between newcomers (most likely Goths) and autochthonous communities derived from other cultural traditions could shape the new, local genetic substructure.

This paper is a continuation of our previous work and concentrates on the human population that lived between the Vistula and Bug Rivers (contemporary eastern Poland) in the IA. The studied group is younger than the Kow-OVIA (2nd-4th and 1st-2nd centuries A.D., respectively) and is believed to represent the next stage of the Goth migrations. The genetic makeup of individuals living at that time in this region has not yet been determined. The biological material was collected from one of the richest IA cemeteries located in east Poland near Masłomęcz village. The cemetery is attributed to the IA Wielbark culture, i.e., to the same culture as the recently characterized site at Kowalewko. This relation in terms of culture and time with geographical separation provided a unique opportunity to obtain the imprint of the demographic processes underlying the genetic variability of the population living during the IA in Central-East Europe. Our analysis involved 27 individuals. We assigned mitochondrial DNA (mtDNA) haplogroups and determined complete mtDNA genome sequences for all of them. We found a high intrapopulation genetic diversity of the studied group. It showed close matrilineal relationship with two other IA populations living, respectively, at the Jutland peninsula and between the Oder and Vistula Rivers. In contrast to our earlier findings concerning Kowalewko (contemporary Western Poland), we found no evidence

of distinct genetic history between females and males from Masłomęcz. This observation suggests that both populations were formed or functioned in different ways. Thus, the data presented here provide new insights into the processes that led to the formation of region-specific genetic substructures within the populations inhabiting the Vistula River Basin in the IA.

Results

DNA isolation, sequencing and preliminary characteristics of mtDNA. Our studies involved human skeletal remains excavated from the Wielbark culture cemetery of IA. The cemetery is located between the Vistula and Bug Rivers in southeast Poland, close to Masłomęcz village (50°72′39″N 23°89′27″E). The studied group, called the Mas-VBIA [abbreviation describing the exact location (Masłomęcz), geographical context (between Vistula and Bug rivers) and time (the IA)], included 27 individuals, whose burial sites previously had been thoroughly characterized archaeologically and anthropologically (Supplementary Table S1). Among others, it was earlier determined that human remains were from 2nd to 4thcenturies A.D., and the grave furnishings were characteristic of the Goths[22]. Radiocarbon dating of the selected samples (Supplementary Fig. S1 and Table S1) confirmed the archeological estimations.

DNA was isolated from teeth, in a dedicated aDNA laboratory, according to the procedure described by Yang and Svensson[23,24]. In total, 27 DNA samples were obtained. The NGS libraries were successfully generated for all samples. To estimate the amount of human DNA, all of the obtained libraries were subjected to shallow NGS sequencing. In order to determine the amount of endogenous human nuclear DNA and mtDNA content, fastq files were aligned to both GRCh37 and RSRS mtDNA reference sequences. The average coverage of mitochondrial genome was 4.7x and ranged between 0.3x and 22.1 × (Supplementary Table S2). In the case of all samples, human DNA showed typical aDNA damage patterns (C > T and G > A alteration most frequently occurring at the 5′ and 3′ ends of the reads, respectively) (Supplementary Table S3 and Supplementary Fig. 2).

Following the NGS screening, the DNA samples for which average coverage was below 5 for the 95% of the mtDNA genome were subjected to second round of sequencing. Additionally, the DNA samples with endogenous human DNA content <5% were subjected to mtDNA enrichment according to the procedure described by Carpenter[25] and then sequenced. As a result, we determined the mtDNA haplogroups and full sequences of the mitochondrial genomes for all 27 individuals (Supplementary Table S4). To test for possible human DNA contamination, we used contamMix[26] to estimate the rates of apparent heterozygosity in mtDNA. The average contamination was approximately 3% (Table 1). All identified haplogroups had been found previously in the populations living in Central Europe from the EBA to the IA, and belonged to 9 haplogroups: H, HV, N1, J, K, T, U, W, and X (Table 1). In addition, based on the anthropological analyses and the raw sequencing data, we determined the sex for 10 and 14 individuals, respectively. In 9 cases sex was assigned with both, anthropological and genetic methods that always gave concordant results (Tables 1 and S4).

Genetic homogeneity of the Mas-VBIA group. To assess whether the analyzed group of people formed a closed homogenous or opened diversified society, we determined the genetic diversity of the Mas-VBIA. To this end, we compared the full length mtDNA sequences obtained for all 27 individuals (Supplementary Table S4). We found 2 pairs of individuals who had identical mtDNA sequences (PCA0088, PCA0090; PCA0094, PCA0100). Interestingly, persons with the same mitochondrial genomes were buried far from each other (Supplementary Fig. 1). We visualized all mtDNA haplotypes from the MAS-VBIA population with the Median Joining Network (MJN) (Fig. 2).

Next, using Arlequin ver. 3.5.1 software[27], we determined the levels of haplotype diversity (HD) based on two fragments of the HVS (hypervariable sequence) region of mitochondrial genome (regions: HVS-I between nucleotide position (np) 16033 and 16365 and HVS-II between np 73 and 340) (Fig. 3a, Supplementary Table S5) and nucleotide diversity (π) based on the fragment of HVS-I region (between np 16000 and 16410) (Fig. 3b, Supplementary Table S6). The HD level calculated for the Mas-VBIA (1.0000 +/− 0.0113) was in the range of high values currently observed for the contemporary, non-isolated European populations. The level of π determined for the Mas-VBIA (0.016263 +/− 0.008846) was high and outside the range of values reported for the present-day European populations (0.009) and in the range of the values reported for contemporary Asian populations[28]. Thus, the results of all three analyses of MJN, HD and π showed that the Mas-VBIA was a genetically highly diverse population.

To exclude the possibility that the presence of two potentially maternally related individuals in the relatively small tested group will affect the results of our consecutive analyzes, we removed the two individuals from the tested group - one from each pair with identical mtDNA sequences.

Matrilineal ancestry of the Mas-VBIA and its contribution to the demographic history of Europe. *mtDNA haplogroup frequency*. mtDNA haplogroup H was the most common in the Mas-VBIA (32%); T2 and U5a were the second most common (16% each). To place the Mas-VBIA in the Central European space-time, we compared mtDNA haplogroup frequencies within the set of populations that lived in this region from the Mesolithic until the present (Central European Population Transect (CEPT), for details, see Materials and methods, Tables 2 and S7–S8). The results of unsupervised hierarchical clustering (Ward's method with Manhattan distance) of the CEPT (Fig. 4) were generally compatible with the known chronology of the formation of the genetic structure of the Central European population. The first to be separated from other populations were the Hunter Gatherers from Central and North Europe (HGCN). Next, we observed a divergence of the studied populations into 2 groups. The first one consisted of Early Neolithic (EN) populations [Starčevo culture (STA), Linearbandkeramik in Transdanubia (LBKT), Linearbandkeramik population from Central Europe (LBK), Rössen culture (RSC), Schöningen Group (SCG)] and MN populations [Baalberge culture (BAC), Salzmünde culture (SMC)]. The second group consisted of LN/EBA populations [BEC (Bernburg culture), CWC, Bell Beaker

Sample	Haplogroup	Anthropological sex	Genetic sex	mtDNA contamination
PCA0088	U3a1a	M	low chrX and chrY coverage	0.9898451
PCA0089	J1c3	F	F	0.9610449
PCA0090	U3a1a	—	F	0.9935336
PCA0091	U5a1b3	—	M	0.9964154
PCA0092	H16	F	F	0.9970401
PCA0093	T1a9	M	M	0.9048376
PCA0094	HV0f	F	F	0.9726827
PCA0095	H11a1	—	low chrX and chrY coverage	0.9812048
PCA0096	U4c1	—	low chrX and chrY coverage	0.9292417
PCA0097	T2a1a	—	low chrX and chrY coverage	0.988292
PCA0098	H1e2	—	low chrX and chrY coverage	0.9980056
PCA0099	H1cg	—	F	0.9373612
PCA0100	HV0f	M	M	0.9963028
PCA0101	U5a2b3	—	low chrX and chrY coverage	0.9964521
PCA0102	K1c1	M	M	0.9945874
PCA0103	H2a1a	F	F	0.9289728
PCA0104	H1a3	—	low chrX and chrY coverage	0.9814315
PCA0105	U5a2a1	—	F	0.9984908
PCA0106	T2b	—	low chrX and chrY coverage	0.9841855
PCA0107	U5a1b1e	—	low chrX and chrY coverage	0.9879122
PCA0108	T2b23	—	low chrX and chrY coverage	0.9255117
PCA0109	K1a27	—	F	0.9818328
PCA0110	H5e1b	M	M	0.9880837
PCA0111	N1a1a1a2	F	F	0.9538297
PCA0112	T2b2b	—	low chrX and chrY coverage	0.9910141
PCA0113	V	—	low chrX and chrY coverage	0.8358445
PCA0114	H7a1a	—	low chrX and chrY coverage	0.9844894

Table 1. Results of mtDNA haplogroup assignment.

culture (BBC), Unetice culture (UC)] and post-EBA populations [Jutland peninsula population from IA (JIA), Kow-OVIA, Mas-VBIA and Central European Metapopulation (CEM)]. Within this group, the JIA and CEM formed a separate sub-clade (alpha = 0.93). In addition, the remaining LN/EBA populations split into the two sub-clades. The first one was formed by the Mas-VBIA and BBC and the second by the BEC, UC, CWC and Kow-OVIA.

To look wider on the Mas-VBIA ancestry and its relationship with other contemporary populations, we assembled selected ancient populations in the European Population Transect dataset (EPT) (for details, see Materials and methods, Tables 2 and S7), which, as previously described, was subjected to hierarchical clustering and additionally to PCA of the mtDNA haplogroup frequencies (Fig. 5a–c, Supplementary Table S9). The results of EPT clustering were consistent with those obtained for the CEPT. The Mas-VBIA was placed within central European LN/EBA group and was related closely with the BBC and YAM. The Kow-OVIA was also placed within the central European LN/EBA group but formed a clade with the CWC, UC and BEC (Fig. 5a). First, the two principal components of the PCA plot (Fig. 5b) positioned the Mas-VBIA near the center of the graph, closest to the Kow-OVIA but also close to the CWC, BEC, UC and YAM. The 3rd and 4th principal components (Fig. 5c) placed the Mas-VBIA in close proximity to the YAM and BBC and revealed the Kow-OVIA's higher affinity to the BEC, CWC and UC, as supported by the hierarchical clustering.

To visualize the relationships of the Mas-VBIA group with contemporary humans, we placed it on a map showing the matrilineal genetic structure of present-day populations. The PCA of haplogroup frequencies observed in the Mas-VBIA and 73 extant worldwide populations (Supplementary Fig. S3, Supplementary Table S10) showed that the Mas-VBIA was located within a range of European genetic diversity.

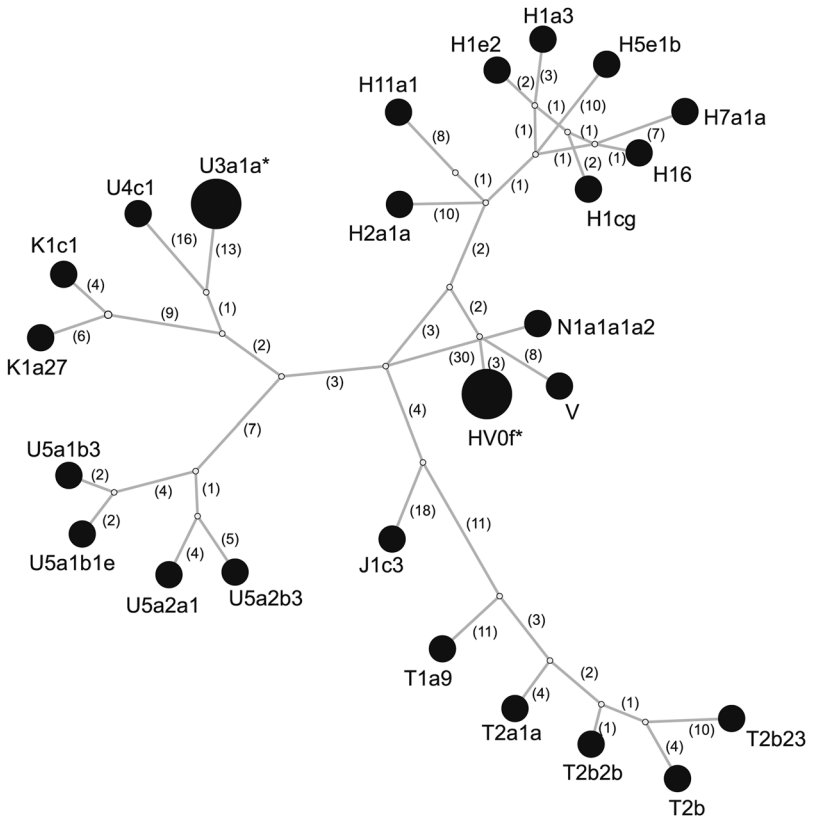


Figure 2. Median Joining Network of 27 individuals from Mas-VBIA based on full mtDNA sequences. Each node corresponds to a haplotype determined for a unique mtDNA sequence. Numbers in brackets show numbers of nucleotide differences between haplotypes. Stars mark haplotypes represented by two individuals with identical mtDNA sequence.

Genetic distances. The relationships between the Mas-VBIA and the earlier and later European populations were also determined by measuring the genetic distances among them. Genetic distances were determined based on a fragment of the mtDNA HVS-I sequence (the region between np 16064 and 16400) (Supplementary Table S11). In the first step, we concentrated on the CEPT (populations from Central Europe from Mesolithic to present-day). The established fixation indexes (Fst) showed that the Mas-VBIA was most closely related to the BEC (Fst = 0, p = 0.7608) and RSC (Fst = 0, p = 0.6529). Both the BEC and RSC share a common archaeological history connected with the spread of the TRB culture. Similarly to the Mas-VBIA, the Kow-OVIA had small genetic distances to the BEC (Fst = 0, p = 0.748) and RSC (Fst = 0.0002, p = 0.567). Unlike the Kow-OVIA, which was closely linked with the JIA, the Mas-VBIA had smaller genetic distances to the LN CWC and BBC populations (Fst = 0, p = 0.5872; Fst = 0, p = 0.67342, respectively) than to the JIA (Fst = 0.00915, p = 0.3768).

In the second step, we focused on the populations living in all of Europe from the Mesolithic to the IA. To this end, we selected 27 populations from the EPT, for which sequences of the mtDNA HVS-I region between np 16064 and 16400 (Table S12) were known for a considerable number of individuals. The calculated genetic distances (Supplementary Table S12) showed once again that the Mas-VBIA was the closest to the BEC (Fst = 0, p = 0.7418). Furthermore, the close connection to the Funnel Beaker cultural horizon was recapitulated with the low genetic distance between the Mas-VBIA and the TRB (Fst = 0, p = 0.6028). A small genetic distance was also found between the Mas-VBIA and the YAM (Fst = 0, p = 0.87892). At the same time, the genetic distance between the Mas-VBIA and the LBK was high (Fst = 0.02949, p = 0.0634). The results of the genetic distance analysis were visualized on the multidimensional scaling (MDS) plot (Fig. 6). In accordance with the established genetic distances, the MDS located the Mas-VBIA, CWC, BBC, BEC and YAM close together.

The Mas-VBIA in the context of the processes shaping the genetic structure of the Central European population. In the next stage of our studies, we attempted to determine how the Mas-VBIA fits into the existing scheme describing the process of the formation of the genetic structure of the Central European population. For this purpose, we performed an Analysis of Molecular Variance (AMOVA). In the first step, we determined an optimal division of the populations from CEPT, excluding the HGCN and CEM. The division was considered to be optimal if the intragroup variability (Fsc) was minimal and intergroup variability (Fct) was maximal. Considering earlier observations made by other authors[29,30], we allocated all EN/MN populations (STA, LBKT, LBK, RSC, SCG, BAC, SMC) to one group because of their high genetic homogeneity and analyzed 41 different combinations of populations' groupings (Supplementary Table S13). We found that Fsc and Fct were optimal if the studied populations were divided into three groups: (STA, LBKT, LBK, RSC, SCG, BAC,

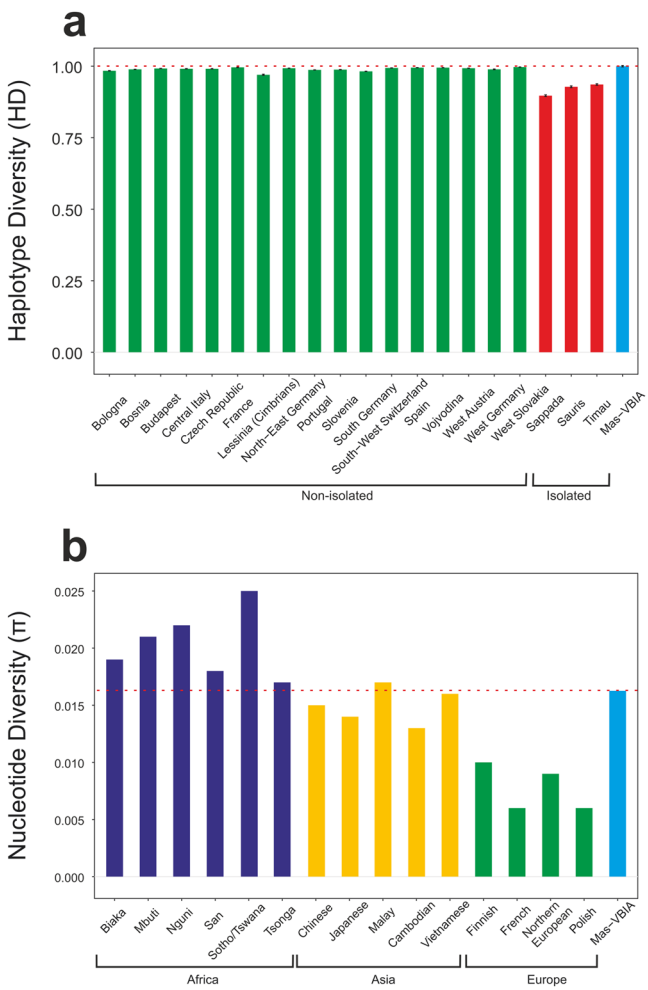


Figure 3. Intrapopulation genetic diversity estimates: (**a**) Haplotype Diversity; (**b**) Nucleotide Diversity. Red dotted line marks the estimate value of diversity of the Mas-VBIA.

SMC) + (CWC, UC, JIA) + (BEC, BBC, Mas-VBIA, Kow-OVIA). Importantly, we observed that the groupings were always better when the Mas-VBIA and Kow-OVIA were in the same group. This observation was not biased by the batch effect as samples from both cemeteries were sequenced separately. Moreover, particular samples were sequenced in multiple runs and in two independent laboratories. The obtained results suggest that the JIA had close relationships with the North-Central Europe populations (CWC and UC), whereas the Mas-VBIA was linked to the Funnel Beakers (e.g., BEC) and Bell Beakers.

Next, to determine to what extent particular populations contributed to the genetic structure of the contemporary Central European population, we performed an analogous analysis, but the studied group was extended by the CEM (CEPT with excluded HGCN). Once again, the populations of the EN/MN (STA, LBKT, LBK, RSC, SCG, BAC, SMC) were treated as one group, and we analyzed 91 combinations of the other fossil populations and the CEM (Supplementary Table S14). The combination of 3 groups (STA, LBKT, LBK, RSC, SCG, BAC, SMC) + (BEC, UC) + (CWC, BBC, Mas-VBIA, Kow-OVIA, JIA, CEM) showed the best optimization of Fsc and Fct values. Furthermore, clustering was always better if the CEM and IA populations did not form a separate group but were placed together with the CWC or BBC.

The results of the above analyses were consistent with our previous observations[8] suggesting that the contribution of the particular populations to the genetic structure of present-day Europe is not fully consistent with the chronology according to which these populations inhabited Central Europe. For example, the Mas-VBIA, Kow-OVIA, JIA, BBC and CWC contributed more significantly to the genetic structure of present-day Europe than the UC.

mtDNA lineages. To get better insight into the processes that shaped the genetic structure of the studied population, we determined which populations gave rise to the haplotypes found in the Mas-VBIA (Supplementary Table S15). Our analysis showed that 44% of the mtDNA haplotypes came from EN populations. In this group, mtDNA haplogroups H and T2 were most frequent. Haplotypes that appeared in the LN and EBA were not frequently detected in the Mas-VBIA (12% and 6%, respectively).

A consecutive analysis of shared haplotypes[31] (Supplementary Table S16) revealed that 78% of the haplotypes inherited by the Mas-VBIA from the LN/EBA or earlier populations were present in at least 1 LN/EBA population

Abbreviation	Population	Time period
HGCN	Hunter Gatherers Central North Europe	15400–4300 BP
HGSW	Hunter Gatherers South West Europe	12000–5000 BP
HGE	Hunter Gatherers Eastern Europe	15000–4500 BP
STA	Starčevo culture	8200–7450 BP
LBKT	Linearbandkeramik culture in Transdanubia	7600–6900 BP
LBK	Linearbandkeramik culture in Central Europe	7500–6900 BP
RSC	Rössen culture	6600–6200 BP
SCG	Schöningen group	6100–5900 BP
BAC	Baalberge culture	5900–5400 BP
SMC	Salzmünde culture	5400–5100 BP
BEC	Bernburg culture	5100–4600 BP
CWC	Corded Ware culture	4800–4200 BP
BBC	Bell Beaker culture in Central Europe	4500–4200 BP
UC	Unetice culture	4200–3150 BP
PWC	Pitted Ware Culture	5200–4300 BP
CAR	Cardial/Epicardial culture of the Iberian Penisula	7400–5700 BP
NPO	Portuguese Neolithic population	7200–5000 BP
NBQ	Neolithic population from Basque Country and Navarre	8100–7100 BP
TRB	Funnel Beaker culture	6300–4800 BP
TRE	Treilles culture	7000 BP
BAS	Bronze Age Siberia	4400–2800 BP
BAK	Bronze Age Kazakhstan	4000–2600 BP
RRBP	Gurgy 'Les Noisats' group	7000–6000 BP
MIR	Iberian Chalcolithic El Mirador Burgos individuals	4500–4050 BP
JIA	Jutland Iron Age	2500 BP - 400 AD
IIA	Iberian Iron Age population	2800 BP - 50 AD
SCY	Iron Age Scythian samples	2300–2600 BP
SSP	Scytho-Siberian Pazyryk Culture	2400–2300 BP
YAM	Yamnaya culture	5300–4600 BP
Kow-OVIA	Oder Vistula Iron Age	0–200 AD
Mas-VBIA	Vistula Bug Iron Age	100–300 AD

Table 2. Published reference ancient mtDNA data and populations abbreviations.

(BEC, CWC, BBC or UC), and 28% were present in all three CWC, BBC and UC populations. It was noteworthy that neither of the haplotypes inherited by the Mas-VBIA was uniquely shared only with the BEC. All haplotypes shared by the BEC and Mas-VBIA were also present in each of the remaining LN/EBA populations. Interestingly, though the Mas-VBIA showed a high genetic distance to the UC, it shared more haplotypes with the UC than with any other LN/EBA population. Additionally, the haplotypes shared uniquely with only one LN/EBA population were most frequently observed for the Mas-VBIA and UC. Here, however, it should be noticed that the UC is represented by a high number of individuals; thus, the chance of finding shared haplotypes was higher. Twenty-one percent of the haplotypes found in the Mas-VBIA were absent in any LN/EBA population.

Discussion

Despite intensive progress in genomic studies of ancient European populations, our knowledge on demographic processes that occurred in the Central-Eastern part of the continent after the Neolithic is still very limited. Until recently, there were no data describing the genetic makeup of people living in Central Europe east of the Oder River during the BA and IA. Archaeological and genomic findings suggested that on the turn of the LN and BA this region (contemporary Poland) was influenced by four major cultures: BBC, CWC, UC and YAM (Fig. 1a). However, due to the lack of reliable data, the actual range of these cultures, and consequently, the placement of the borders between them, as well as the later history of these populations, are difficult to determine.

To extend our knowledge on the issues described above, we initiated broad genetic studies of the populations inhabiting the territory of present-day Poland in the IA. In our earlier paper, we characterized the maternal genetic makeup of the individuals living in the west part of contemporary Poland[8]. In this report, we focused on the maternal genetic history of the Mas-VBIA group that lived in the eastern part of contemporary Poland (the region located between the Vistula and Bug Rivers) during the 2nd to 4th century A.D. We found that the intrapopulation diversity of the Mas-VBIA ($\pi = 0.016263$) was beyond expectations. It was as high, as is currently observed for Asian populations (π between 0.013 and 0.017), and exceeded the values typical for non-isolated European populations (π between 0.006 and 0.01), as well as the intrapopulation diversity previously determined for its western counterpart – the Kow-OVIA ($\pi = 0.0079$). However, it should be noted here that due to the post-mortem DNA damage, the diversity estimates could be inflated. Accordingly, one can hypothesize that people

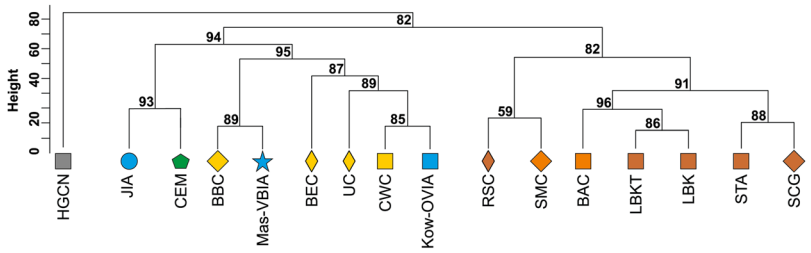


Figure 4. Unsupervised hierarchical clustering with Ward method and Manhattan distance of haplogroup frequencies for the CEPT populations. P-values of the clusters are given as the percent of reproduced clusters based on 10,000 bootstrap replicates. Symbols indicate populations from Central Europe (squares and diamonds), Southern Scandinavia and Jutland Peninsula (circles), and East Europe/Asia (stars). Color shading of data points denotes to Hunter-Gatherers (grey), Early Neolithic (brown), Middle Neolithic (orange), Late Neolithic/Early Bronze Age (yellow), Iron Age (blue), and the present-day Central Europe Metapopulation (CEM, green). Abbreviations: Central/North European Hunter-Gatherers (HGCN), Starčevo Culture population (STA), Linearbandkeramik in Transdanubia (LBKT), Linearbandkeramik population from Central Europe (LBK), Rössen Culture (RSC), Schöningen Group (SCG), Baalberge Culture (BAC), Salzmünde Culture (SMC), Bernburg Culture (BEC), Corded Ware Culture (CWC), Bell Beaker Culture (BBC), Unetice Culture (UC), Jutland Iron Age (JIA), Kowalewko Oder and Vistula Iron Age (Kow-OVIA), Masłomęcz Vistula and Bug Iron Age (Mas-VBIA).

living between the Vistula and Bug Rivers in the IA formed genetically diversified, non-isolated populations. The haplogroup frequencies observed for the Mas-VBIA were similar to those reported for the Kow-OVIA. Haplogroup H frequency was intermediate between CWC and BBC and lower than in CEM. Interestingly, the frequencies of the mtDNA haplogroups U5a (characteristic for Eastern Hunter Gatherer groups) and U5b (characteristic for Western Hunter Gatherer groups) were a significantly different in the Mas-VBIA and other IA populations studied so far (the JIA and Kow-OVIA). The Mas-VBIA was characterized by the absence of mtDNA haplogroup U5b and the high frequency of mtDNA haplogroup U5a (14.8%), whereas in the JIA and Kow-OVIA, mtDNA haplogroup U5b was more prevalent than U5a. As noticed previously[8], the domination of U5b over U5a in Central Europe was observed only during the IA. Therefore, either the process that led to the temporal increase of U5b mtDNA haplogroup frequency did not affect regions east of Vistula, or its effects were subsequently removed by another demographic event.

All high dimensional analyses of genetic diversity (PCA, hierarchical clustering, MDS) within the Mas-VBIA and other ancient and extant populations were consistent and showed that the Mas-VBIA was closely connected with the Kow-OVIA and had a similar genetic structure to that of Central European LN/EBA populations. Within the group of three IA Central European populations, the Mas-VBIA was the closest related with the Kow-OVIA, and the distance between the JIA and Mas-VBIA was shorter than that between the JIA and Kow-OVIA. Furthermore, the Mas-VBIA had close genetic connections with the YAM, located in the eastern parts of Europe. We did not observe these connections in either the Kow-OVIA or the JIA. Earlier, Haak *et al.*[32] showed that there are clear genetic links, even at the mtDNA level, between the central European CWC and BBC populations and the YAM. However, if close genetic connections between the Mas-VBIA and YAM were a result of the LN migratory events, the same should be observed in the remaining IA populations of the region. Therefore, we propose that the presence of the Pontic-Caspian steppe genetic component in the Mas-VBIA is a result of the migratory event or events that occurred in the IA period. Considering the high prevalence of the mtDNA haplogroup U5a in the Mas-VBIA, one can hypothesize that the direction of this movement was from east to west.

Another interesting observation was the genetic proximity between the BEC and virtually all subsequent populations (the CWC, BBC, UC, JIA, Kow-OVIA and Mas-VBIA). This proximity was most evident when shared mtDNA haplotypes were analyzed. We found that the haplotypes shared between all LN and IA populations were nearly always present in the BEC. The BEC is part of a TRB cultural horizon that spread from North to Central Europe during the MN in the form of the sequence of Danubian populations[14]. Thus, here we provide an additional line of evidence that Danubian Neolithic populations were a common genetic background of all Central-North and Central-East populations in Europe.

There are two major historical narratives describing the formation of the Wielbark culture. The first one links its development with Goth migrations[33], and the second, with the spread of the local Oksywie culture that appeared at the Baltic seashore at approximately the 2nd century B.C[20,34]. In addition, there are at least two hypotheses concerning the origin of Goths. One of them assumes that the homeland of the Goths was located in the southernmost part of the Germanic territories other than in Scandinavia[35]. Considering the results obtained for both the Kow-OVIA and Mas-VBIA, one can assume that they are, to a large extent, consistent with the postulated chronology of early migrations of Goths and their settlement in Central-East Europe. The high genetic diversity of the Mas-VBIA strongly corresponds with the suggested role of Masłomęcz at that time[22]. The archaeological findings indicate that in the 2nd and 3rd centuries A.D., it was one of the major cultural and political Goth centers. In addition, the genetic relationships reported here between the Mas-VBIA and both earlier characterized IA populations (the Kow-OVIA and JIA) support the opinion that southern Scandinavia was the homeland of the Goths. According to some authors, the process of their migration through the contemporary territory of Poland,

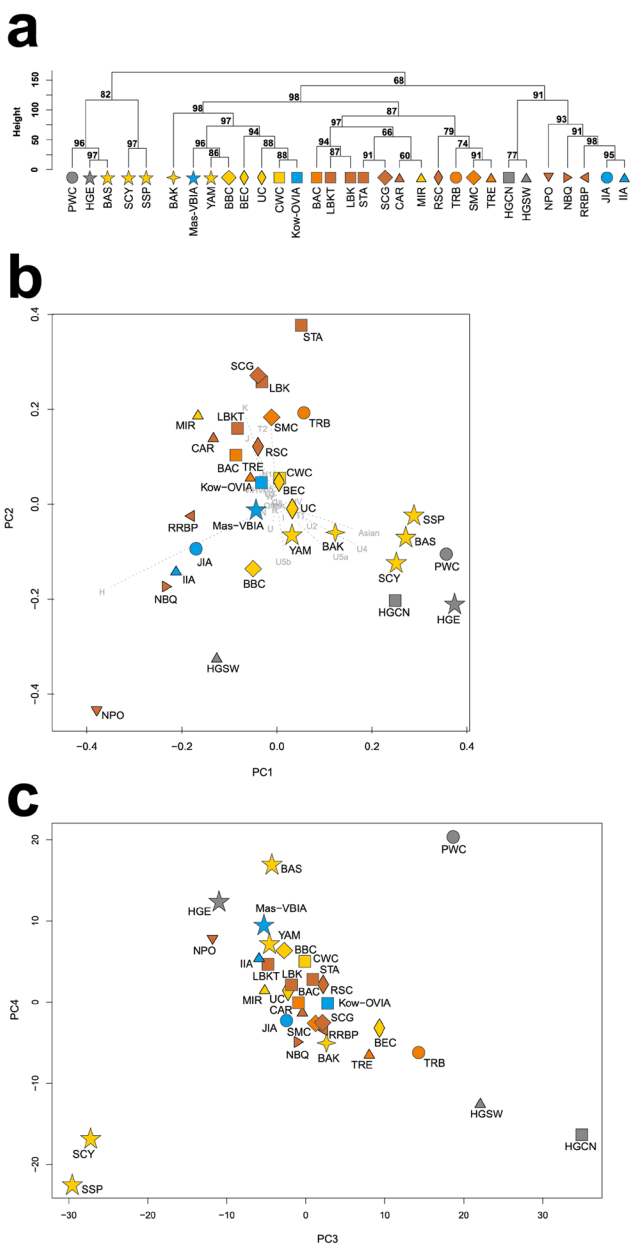


Figure 5. (**a**) Unsupervised hierarchical clustering with Ward method and Manhattan distance of haplogroup frequencies for the EPT populations. P-values of the clusters are given as the percent of reproduced clusters based on 10,000 bootstrap replicates; (**b**) Principal components 1 and 2 of the PCA on the haplogroup frequencies of EPT populations. Symbols indicate populations from Central Europe (squares and diamonds), Southern Scandinavia and Jutland Peninsula (circles), Iberian Peninsula (triangles), and East Europe/Asia (stars). Color shading of data points denotes to Hunter-Gatherers (grey), Early Neolithic (brown), Middle Neolithic (orange), Late Neolithic/Early Bronze Age (yellow) and Iron Age (blue). The first two principal components of the PCA display 48.4% of the total genetic variation. Each haplogroup was superimposed as component loading vectors (grey dotted lines) proportionally to their contribution. Abbreviations: Central/North European Hunter-Gatherers (HGCN), Southwestern European Hunter-Gatherers (HGSW), East European Hunter-Gatherers (HGE), Starčevo Culture population (STA), Linearbandkeramik in Transdanubia (LBKT), Linearbandkeramik population from Central Europe (LBK), Rössen Culture (RSC), Schöningen Group (SCG), Baalberge Culture (BAC), Salzmünde Culture (SMC), Bernburg Culture (BEC), Corded Ware Culture (CWC), Bell Beaker Culture (BBC), Unetice Culture (UC), Pitted Ware culture (PWC), Funnel Beaker culture (TRB), Jutland Iron Age (JIA), Cardial/Epicardial culture of the Iberian Peninsula (CAR), Portuguese Neolithic population (NPO), Neolithic population from Basque Country and Navarre (NBQ), Iberian Chalcolithic El Mirador Cave individuals (MIR), individuals from Iberian Iron Age period (IIA), Treilles Culture (TRE), Gurgy 'Les Noisats' group (RRBP), Bronze Age Kurgan samples from South Siberia (BAS), Bronze Age Kazakhstan (BAK), Yamnaya (YAM), Iron Age Scythian (SCY), Scytho-Siberian Pazyryk Culture (SSP), Kowalewko Oder and Vistula Iron Age (Kow-OVIA), Masłomęcz Vistula Bug Iron Age (Mas-VBIA); (**c**) Principal components 3 and 4 of the PCA on the haplogroup frequencies of EPT populations.

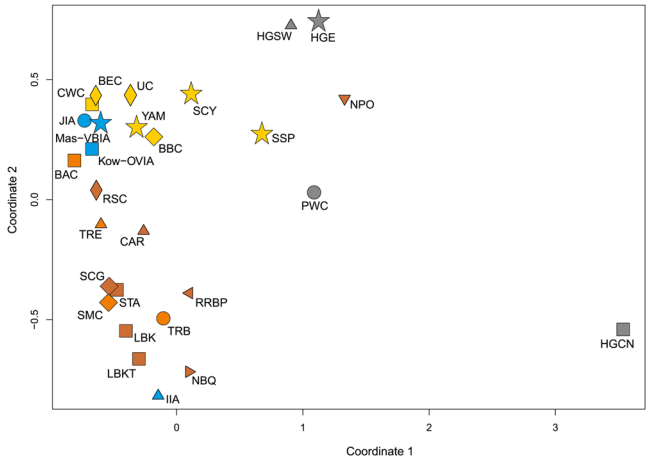


Figure 6. MDS plot of Slatkin's Fst values for EPT populations. Fst values were obtained for mtDNA HVS-I region (np 16064–16400).

which was connected with the spread of the Wielbark culture, can be divided into the following 6 stages, named with the letters A-F (for details, see Fig. 7a)[20]. In the first stage, the Goths colonized the mouth of the Vistula River during the 2nd and 1st centuries B.C. (stage A). Then, they spread west along the Baltic seashore (stage B). In stage C, the Goths moved south and ousted the Przeworsk culture from the northwest part of contemporary Poland, including Kowalewko (1st and 2nd centuries A.D.). The Goth's settlements established in this area could have a rather temporal character suggested by the fact that the genetic structure of the Kow-OVIA population was biased by sex. Males from the Kow-OVIA were closely related with the JIA and females with Neolithic farmers[8] (most likely they represented the local population related with the Przeworsk culture). During the next stage (D), the Goths spread east of the Vistula, and then they migrated along the Vistula and Bug Rivers towards the Black See and established the Chernyakhov culture by mixing with Pontic–Caspian steppe populations (stage E). Finally, a part of the Chernyakhov culture population moved back and established a large settlement near Masłomęcz called the "Masłomęcz group" (stage F, 2nd and 3rd centuries A.D.) (Fig. 7b). Such a scenario explains a large part of our findings well, including (i) the high prevalence of the mtDNA U5a haplogroup (characteristic for the eastern parts of Europe); (ii) the close genetic distance to the YAM observed for the Mas-VBIA and the lack of these characteristics in the cases of the Kow-OVIA and JIA; and (iii) the high genetic diversity of the Mas-VBIA. The latter indicates that the "Masłomęcz group", often considered the Goths, from a genetic standpoint, was a mixture of different populations. Unfortunately, the presented data cannot unequivocally verify both hypotheses on the Goth origin and their relationships with the Oksywie culture. Thus, we still do not know whether the Goths replaced the Oksywie culture or induced its formation. It is also possible that all the scenarios presented here contain some truth and describe different periods of the Goth population development and migrations. Thus, further studies are necessary to verify them.

Materials and Methods

aDNA extraction and library preparation. For the purposes of this study, 27 samples (teeth) were obtained from 27 individuals from the Masłomęcz archeological site (Supplementary Table S1). After being transported to a dedicated aDNA laboratory (at the Institute of Anthropology, Faculty of Biology, Adam Mickiewicz University, Poznan, Poland), the teeth were cleaned with 5% NaOCl and rinsed with sterile water, followed by UV irradiation (254 nm) for 2 hours for each site. The roots of the teeth were drilled using Dremel® drill bits. Bone powder (~250 mg) was digested with proteinase K, and DNA-containing extract was purified with the use of the silica-based method following Yang and Malmstrom[23,36]. Genomic libraries were prepared following the protocol of Meyer[37], with omission of the initial sonication step due to natural fragmentation of aDNA. In order to maximize the chance of amplification of each unique DNA template molecule, six separate PCR reactions were set up for each library. PCR amplifications were performed in 25 µl, with 3 µl of the DNA library template, 12.5 µl of 1x AmpliTaq Gold® 360 Master Mix (Life Technologies), 0.5 µl of indexing primer (10 µM) and 0.5 µl of PCR primer IS4 (10 µM)[38]. The PCR profile was as follows: initial denaturation (94 °C, 12 min), 12–16 cycles of 94 °C (30 s), 60 °C (30 s), 72 °C (45 s) and final extension (72 °C, 10 min). PCR reactions for the same library were pooled and purified with AMPure® XP beads (Agencourt-Beckman Coulter)[38]. The quality and size distribution of the libraries were verified with a High Sensitivity DNA kit and 2100 Bioanalyzer system (Agilent), while the DNA concentration was determined with a Qubit fluorimeter and Qubit dsDNA HS Assay Kit (ThermoFisher Scientific), according to the manufacturer's protocols.

Next generation sequencing (NGS) and enrichment of aDNA libraries. To screen the libraries, shallow NGS sequencing with the Genome Analyzer GAIIx (Illumina, California, USA) and TruSeq SBS Kits v5-GA (Illumina) was applied. Seven to eight libraries were pooled per lane. On average, 6.3 mln 75 bp-long reads were collected per library. Libraries with small human DNA content below 5% were enriched in mtDNA or whole

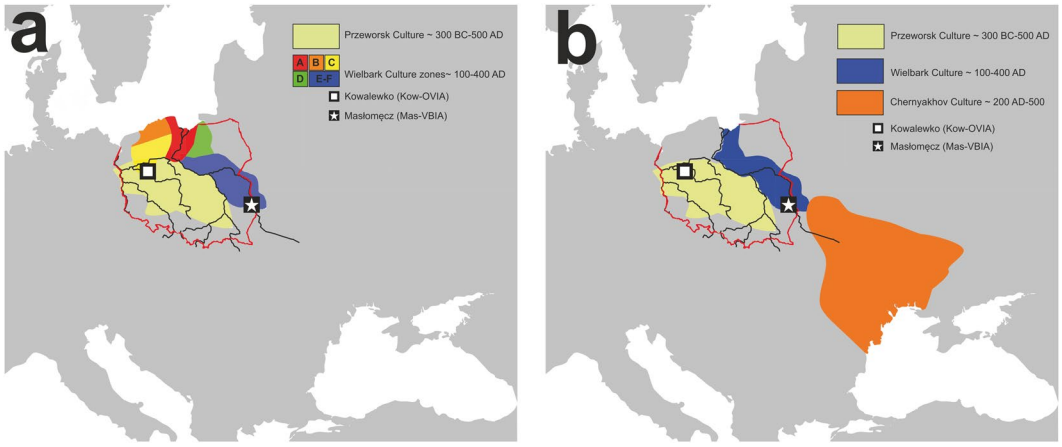


Figure 7. Geographical distribution of the archaeological cultures in contemporary Poland linked to the early migrations of the Goths: (**a**) formation stages of the Wielbark culture; (**b**) late stage of the Wielbark culture. Europe map by Roke, retrieved from https://commons.wikimedia.org/wiki/Category:Blank_maps_of_Europe#/media/File:BlankMap-Europe-v3.png, used under Creative Commons Attribution-ShareAlike 3.0 Unported (https://creativecommons.org/licenses/by-sa/3.0/deed.en), modified with Corel Draw ver. 12.0.

human genome sequences using MYbaits (Arbor Biosciences, Michigan, USA) and an in-solution hybridization capture procedure, according to the recommendations of the manufacturer. mtDNA enrichment was performed with MYTObaits (Arbor Biosciences), a set of 80-bp long tiling probes designed based on rCRS (Cambridge Reference Sequence). Two rounds of enrichment were applied, each followed by 14 amplification cycles. Enriched libraries were sequenced using GAIIx (Illumina) again. On average, 6 mln 75 bp-long reads were collected per mtDNA-enriched library and human genome content was in the range of 55–74% (Supplementary Table S2).

Filtering, mapping and variant calling. Raw sequencing data (fastq files) were filtered with AdapterRemoval[39] by trimming missing nucleotides from both ends with the threshold of minimum quality of 30 and minimum length of 25 nucleotides. Filtered reads were aligned with BWA ver. 0.7.10[40] to the rCRS mitochondrial and GRCH 37 reference genomes, with the seed blocked for higher sensitivity and other parameters set as default, as suggested by Schubert[41], with the command *bwa aln -l 1000 reference.fasta input.fastq*. To cope with the effects of DNA damage, we trimmed 3 bases from both read ends. Following the alignment, duplicate reads were removed with picard-tools ver. 1.117 MarkDuplicates. Read depth and coverage were assessed with samtools ver. 1.2[42] and bedtools ver. 1.2[43], respectively. Consensus fasta sequences for haplogroup prediction and sequence analyses were generated with FreeBayes ver. 1.0.2–33-gdbb6160. To assemble complete consensus mtDNA sequences we applied the following quality requirements: minimum coverage per base >=3, missing nucleotide count < 5%, no missing nucleotides in the HVS-I sequence, base call supported by the 3/5 majority of reads[44] (Supplementary Table S4).

Human DNA damage patterns. To examine data authenticity, we used mapDamage 2.0[45] and estimated whether the human DNA damage parameters for each sample were typical for aDNA: (i) λ, the fraction of nucleotides positioned in single-stranded DNA overhangs context, (ii) δs, C → T deamination probability in the single-stranded overhangs context, and (iii) δd, C → T deamination probability in the double-stranded DNA context (Supplementary Table S3).

Contamination assessment. To estimate the level of contamination with contemporary human DNA or with other human aDNA, we used the software contamMix[26]. According to its authors, the software uses 311 present-day human mtDNA sequences as potentially contaminating population and estimates the contamination by comparing the alignment rates between a particular sample's consensus mtDNA sequence and whole mtDNA sequences of 311 potential contaminants.

Genetic sex estimation. Genetic sex of each individual was estimated using the Ry method as described by[46]. The method is based on dividing the number of sequences mapped to the Y chromosome to the number of those mapped to X and Y chromosomes. Only sequences with mapping quality of minimum 30 were considered. Sex assignment was performed for samples with at least 3000 reads aligned to the sex chromosomes.

Analysis of intrapopulation genetic diversity. Haplotype diversity (HD) was analyzed with Arlequin ver. 3.5.1[27] on two fragments of mtDNA HVS (HVS-I, located between nucleotide positions (np) 16033–16365, and HVS-II np 73–340) (Supplementary Table S5). Nucleotide diversity (π) was calculated for the fragment of mtDNA HVS-I (np 16000–16410) (Supplementary Table S6). Furthermore, we visualized the intrapopulation diversity of the Mas-VBIA using the Median Joining Network (MJN) method on 27 full mtDNA sequences (Fig. 2).

mtDNA haplogroup frequency analyses. Haplogroups were assigned based on complete mtDNA sequences using HaploFind[47] with respect to Phylotree build 17 (http://www.phylotree.org/)[48]. Only samples with a haplogroup score ≥0.8 and average coverage ≥3 were used in downstream analyses. To assess changes in temporal mtDNA haplogroup frequency from the Mesolithic to the present day, we performed Ward clustering with Euclidean distance on 23 haplogroups (H, H5, HV, HV0, V, I, J, K, N, N1a, R, T1, T2, U, U2, U3, U4, U5a, U5b, U8, W, X and other). We used a set of 27 samples from this study and 14 ancient populations from Central/North Europe (Supplementary Table S7), and a generated Central European Metapopulation (CEM) composed of 500 random individuals sampled from extant populations of Poland, the Czech Republic, Germany and Austria, as in[30]. We called this group CEPT (Supplementary Table S8).

To compare the mtDNA variability of the Mas-VBIA in a broader geographical context, we also applied unsupervised hierarchical clustering with the Ward method and Euclidean distance and a Principal Components Analysis (PCA) on haplogroup frequencies of the Mas-VBIA and published ancient mtDNA data of populations from across Europe and West Asia. We called this group EPT (European Population Transect) (Supplementary Table S7). Haplogroups were divided into 24 groups present in ancient individuals (Asian [A, C, D, F, G, Z], N, N1a, I, J, W, X, R, HV, V/HV0, H, H5, T, T1, T2, J, U, U2, U3, U4, U5a, U5b, U8, K and other) (Supplementary Table S9).

To elucidate affinities of our samples in relation to present day populations, we performed a PCA analysis based on 23 haplogroup frequencies (Asia [A, C, D, F, G, Z], Africa [L], N1a, I, I1, W, X, HV, V/HV0, H, H5, T1, T2, J, U, U2, U3, U4, U5a, U5b, U8, K and other), with public data from 73 extant populations (Supplementary Table S10).

Cluster significance was tested by performing 10,000 permutations with the pvclust package in R ver. 3.3.0. A PCA was conducted with prcomp of the vegan package in R ver. 3.3.0 (http://R-project.org).

Genetic distance analyses. For sequence-based analyses, the longest mtDNA HVS-I fragment present in the biggest fraction of published samples was selected (np 16064–16400). Additionally, for newly reported samples from this study, no missing nucleotides were allowed in the selected HVS-I range, and at least 3x coverage was expected for 95% of the nucleotides. To examine genetic affinities on the sequence level, we calculated genetic distances (Fst)[49] between two sample sets: CEPT (Supplementary Tables S11 and S17) and EPT (Supplementary Tables S12 and S18), including only those individuals for which the 16064–16400 fragment of mtDNA HVS-I was present.

Pairwise and Slatkin's Fst[50] values were calculated in Arlequin ver. 3.5.1 for both datasets separately, with the associated substitution model and gamma values selected with jModel test 0.1 AIC and BIC[51]. P-values were calculated by performing 10,000 permutations. Genetic distances were visualized on an MDS plot with the metaMDS function from the vegan package in R ver. 3.3.0.

Analysis of genetic structure. To examine whether genetic affinities between particular populations from CEPT were a result of a shared genetic structure, we have conducted an Analysis of Molecular Variance (AMOVA) in 41 combinations of CEPT populations (excluding HGCN and CEM) (Supplementary Table S13). Next, we tested the genetic contributions of ancient populations to the extant mtDNA variability by analyzing AMOVA results from 91 combinations of CEPT populations (excluding HGCN) (Supplementary Table S14). Statistical significance was obtained by performing 10,000 permutation tests. An associated substitution model and gamma values were calculated separately for both scenarios with jModel test 0.1 AIC and BIC[51] (see Supplementary Information).

mtDNA lineage analyses. To further test genetic affinities and account for the temporal succession of archaeological cultures in Central Europe, we conducted an analysis of shared ancestral haplotypes as described in[29] in CEPT (Supplementary Table S15), and a classical shared haplotype analysis[31] in LN/EBA and IA populations (Supplementary Table S16).

Data Availability

The 27 novel complete mtDNA sequences supporting the results of this article are available in the National Center for Biotechnology Information (Genbank) under accession numbers MH492638-MH492664, http://www.ncbi.nlm.nih.gov/Genbank/.

References

1. Mathieson, I. *et al.* The genomic history of southeastern Europe. *Nature* **555**, 197–203, https://doi.org/10.1038/nature25778 (2018).
2. Mittnik, A. *et al.* The genetic prehistory of the Baltic Sea region. *Nature communications* **9**, 442, https://doi.org/10.1038/s41467-018-02825-9 (2018).
3. Olalde, I. *et al.* The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* **555**, 190–196, https://doi.org/10.1038/nature25738 (2018).
4. Reich, D. *Who We Are and How We Got Here.* (Oxford University Press, 2018).
5. Veeramah, K. R. *et al.* Population genomic analysis of elongated skulls reveals extensive female-biased immigration in Early Medieval Bavaria. *Proceedings of the National Academy of Sciences of the United States of America* **115**, 3494–3499, https://doi.org/10.1073/pnas.1719880115 (2018).
6. Juras, A. *et al.* Diverse origin of mitochondrial lineages in Iron Age Black Sea Scythians. *Scientific reports* **7**, 43950, https://doi.org/10.1038/srep43950 (2017).
7. Allentoft, M. E. *et al.* Population genomics of Bronze Age Eurasia. *Nature* **522**, 167–172, https://doi.org/10.1038/nature14507 (2015).
8. Stolarek, I. *et al.* A mosaic genetic structure of the human population living in the South Baltic region during the Iron Age. *Scientific reports* **8**, 2455, https://doi.org/10.1038/s41598-018-20705-6 (2018).
9. O'Sullivan, N. *et al.* Ancient genome-wide analyses infer kinship structure in an Early Medieval Alemannic graveyard. *Science advances* **4**, eaao1262, https://doi.org/10.1126/sciadv.aao1262 (2018).

10. Amorim, C. E. G. *et al*. Understanding 6th-century barbarian social organization and migration through paleogenomics. *Nature communications* **9**, 3547, https://doi.org/10.1038/s41467-018-06024-4 (2018).
11. Schiffels, S. *et al*. Iron Age and Anglo-Saxon genomes from East England reveal British migration history. *Nature communications* **7**, 10408, https://doi.org/10.1038/ncomms10408 (2016).
12. Hervella, M. *et al*. Ancient DNA from South-East Europe Reveals Different Events during Early and Middle Neolithic Influencing the European Genetic Heritage. *PloS one* **10**, e0128810, https://doi.org/10.1371/journal.pone.0128810 (2015).
13. Fernandes, D. M. *et al*. A genomic Neolithic time transect of hunter-farmer admixture in central Poland. *Scientific reports* **8**, 14879, https://doi.org/10.1038/s41598-018-33067-w (2018).
14. Chylenski, M. *et al*. Late Danubian mitochondrial genomes shed light into the Neolithisation of Central Europe in the 5th millennium BC. *BMC evolutionary biology* **17**, 80, https://doi.org/10.1186/s12862-017-0924-0 (2017).
15. Lorkiewicz, W. *et al*. Between the Baltic and Danubian Worlds: the genetic affinities of a Middle Neolithic population from central Poland. *PloS one* **10**, e0118316, https://doi.org/10.1371/journal.pone.0118316 (2015).
16. Mathieson, I. *et al*. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature* **528**, 499–503, https://doi.org/10.1038/nature16152 (2015).
17. Lazaridis, I. *et al*. Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413, https://doi.org/10.1038/nature13673 (2014).
18. Tassi, F. *et al*. Genome diversity in the Neolithic Globular Amphorae culture and the spread of Indo-European languages. *Proceedings. Biological sciences/The Royal Society* **284**, https://doi.org/10.1098/rspb.2017.1540 (2017).
19. Haak, W. *et al*. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* **522**, 207–211, https://doi.org/10.1038/nature14317 (2015).
20. Urbańczyk, P. *The Past Societies. Polish lands from the first evidence of human presence to the early middle ages*. Vol. 4 (Institute of Archaeology and Ethnology, Polish Academy of Sciences, 2016).
21. Philips, A. *et al*. Comprehensive analysis of microorganisms accompanying human archaeological remains. *GigaScience*, https://doi.org/10.1093/gigascience/gix044 (2017).
22. Kokowski, A. & M. O. *Grupa masłomęcka w okresie rzymskim: (III-IV wiek naszej ery)*. (Muzeum Okręgowe, 1987).
23. Yang, D. Y., Eng, B., Waye, J. S., Dudar, J. C. & Saunders, S. R. Technical note: improved DNA extraction from ancient bones using silica-based spin columns. *American journal of physical anthropology* **105**, 539–543, https://doi.org/10.1002/(SICI)1096-8644(199804)105:4<539::AID-AJPA10>3.0.CO;2-1 (1998).
24. Svensson, E. M. *et al*. Tracing genetic change over time using nuclear SNPs in ancient and modern cattle. *Animal genetics* **38**, 378–383, https://doi.org/10.1111/j.1365-2052.2007.01620.x (2007).
25. Carpenter, M. L. *et al*. Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. *American journal of human genetics* **93**, 852–864, https://doi.org/10.1016/j.ajhg.2013.10.002 (2013).
26. Fu, Q. *et al*. Genome sequence of a 45,000-year-old modern human from western Siberia. *Nature* **514**, 445–449, https://doi.org/10.1038/nature13810 (2014).
27. Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources* **10**, 564–567, https://doi.org/10.1111/j.1755-0998.2010.02847.x (2010).
28. Jorde, L. B. *et al*. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *American journal of human genetics* **66**, 979–988, https://doi.org/10.1086/302825 (2000).
29. Szecsenyi-Nagy, A. *et al*. Tracing the genetic origin of Europe's first farmers reveals insights into their social organization. *Proceedings. Biological sciences/The Royal Society* **282**, https://doi.org/10.1098/rspb.2015.0339 (2015).
30. Brandt, G. *et al*. Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. *Science* **342**, 257–261, https://doi.org/10.1126/science.1241844 (2013).
31. Csakyova, V. *et al*. Maternal Genetic Composition of a Medieval Population from a Hungarian-Slavic Contact Zone in Central Europe. *PloS one* **11**, e0151206, https://doi.org/10.1371/journal.pone.0151206 (2016).
32. Haak, W. *et al*. Ancient DNA from European early neolithic farmers reveals their near eastern affinities. *PLoS biology* **8**, e1000536, https://doi.org/10.1371/journal.pbio.1000536 (2010).
33. Gillett, A. Ethnogenesis: A Contested Model of Early Medieval Europe. *History Compass* **4**, 241–260 (2006).
34. Makiewicz, T. *The Archaeology of the Transit Gas Pipeline: The Goths in Greater Poland* (Poznan Archaeological Museum, 2012).
35. Kaliff; & Green. *Gothic Connections: Contacts Between Eastern Scandinavia and the Southern Baltic Coast 1000 Bc-500* (Uppsala: Uppsala university. Department of archaeology and ancient history, 2001).
36. Malmstrom, H. *et al*. More on contamination: the use of asymmetric molecular behavior to identify authentic ancient human DNA. *Molecular biology and evolution* **24**, 998–1004, https://doi.org/10.1093/molbev/msm015 (2007).
37. Meyer, M. & Kircher, M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor protocols* **2010**, pdb prot5448, https://doi.org/10.1101/pdb.prot5448 (2010).
38. Gunther, T. *et al*. Ancient genomes link early farmers from Atapuerca in Spain to modern-day Basques. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 11917–11922, https://doi.org/10.1073/pnas.1509851112 (2015).
39. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC research notes* **9**, 88, https://doi.org/10.1186/s13104-016-1900-2 (2016).
40. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595, https://doi.org/10.1093/bioinformatics/btp698 (2010).
41. Schubert, M. *et al*. Improving ancient DNA read mapping against modern reference genomes. *BMC genomics* **13**, 178, https://doi.org/10.1186/1471-2164-13-178 (2012).
42. Li, H. *et al*. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079, https://doi.org/10.1093/bioinformatics/btp352 (2009).
43. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842, https://doi.org/10.1093/bioinformatics/btq033 (2010).
44. Hwang, S., Kim, E., Lee, I. & Marcotte, E. M. Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Scientific reports* **5**, 17875, https://doi.org/10.1038/srep17875 (2015).
45. Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. & Orlando, L. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* **29**, 1682–1684, https://doi.org/10.1093/bioinformatics/btt193 (2013).
46. Skoglund, P. S. J. & Jakobsson, M. Accurate sex identification of ancient human remains using DNA shotgun sequencing. *Journal of Archaeological Science* **40**, https://doi.org/10.1016/j.jas.2013.07.004 (2013).
47. Vianello, D. *et al*. HAPLOFIND: a new method for high-throughput mtDNA haplogroup assignment. *Human mutation* **34**, 1189–1194, https://doi.org/10.1002/humu.22356 (2013).
48. van Oven, M. Revision of the mtDNA tree and corresponding haplogroup nomenclature. *Proceedings of the National Academy of Sciences of the United States of America* **107**, E38–39; author reply e40–31, https://doi.org/10.1073/pnas.0915120107 (2010).
49. Wright, S. Genetical structure of populations. *Nature* **166**, 247–249 (1950).
50. Slatkin, M. A measure of population subdivision based on microsatellite allele frequencies. *Genetics* **139**, 457–462 (1995).
51. Posada, D. jModelTest: phylogenetic model averaging. *Molecular biology and evolution* **25**, 1253–1256, https://doi.org/10.1093/molbev/msn083 (2008).

Acknowledgements
We are grateful to Anna Myszka and Dawid Trzciński from Institute of Anthropology, Adam Mickiewicz University, Poznań for providing samples. This work was supported by the Polish National Science Center [2014/12/W/NZ2/00466].

Author Contributions
I.S. participated in the study design, analyzed the data, created figures, discussed the results, and wrote the manuscript. L.H. prepared NGS libraries, performed NGS sequencing, discussed the results and wrote the manuscript. A.J. extracted DNA, prepared NGS libraries and reviewed the manuscript. W.N. provided samples and anthropological description of the skeletal remains. A.M. and H.K.-K. provided archaeological narratives and discussed the results. J.P. provided samples, discussed the results and reviewed the manuscript. P.K. participated in the study design, discussed the data and reviewed the manuscript. M.F. conceived the overall idea of the study, provided funding, participated in the study design, discussed the data and was responsible for the final version of the manuscript. All of the authors read and approved the final manuscript.

Additional Information
Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-43183-w.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

MATERIAŁY UZUPEŁNIAJĄCE DO PUBLIKACJI

Stolarek i wsp., Scientific Reports 2019

Spis tabel uzupełniających

Spis rysunków uzupełniających

Figure S1. Location of Masłomęcz and a scheme of the Masłomęcz cemetery site 15, based on the Fig. 36 from the monograph by Andrzej Kokowski, Grupa masłomęcka: z badań nad przemianami kultury Gotów w młodszym okresie rzymskim, published in: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej, Lublin 1995, and Fig. 1 from the monograph by Andrzej Kokowski, Cmentarzyska ludności grupy masłomęckiej, published in: Studia Antropologiczne V, Acta Universitatis Vratislaviensis No. 2050, Wrocław 1998, generated using Corel Draw ver. 12.0, with the author permission. Sampled graves are marked with a red color. Europe and Poland maps were downloaded from Wikimedia Commons (https://commons.wikimedia.org), under the free license, and modified with Corel Draw ver. 12.0

Figure S2. MapDamage 2.0 misincorporation plots showing damage patterns typical for aDNA

Fig. S3. a) 3D PCA of haplogroup frequencies of Mas-VBIA and 73 present day populations from Europe and Near East, b) PC1 and PC2 view, c) PC3 and PC4 view

3

Anna Philips, Ireneusz Stolarek, Bogna Kuczkowska, Anna Juras, Luiza Handschuh, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Comprehensive analysis of microorganisms accompanying human archaeological remains

Gigascience. 2017 Jul; 6(7): 1–13
doi: 10.1093/gigascience/gix044

GigaScience, 6, 2017, 1–13

doi: 10.1093/gigascience/gix044
Advance Access Publication Date: 13 June 2017
Research

RESEARCH

Comprehensive analysis of microorganisms accompanying human archaeological remains

Anna Philips[1], Ireneusz Stolarek[1], Bogna Kuczkowska[1], Anna Juras[2], Luiza Handschuh[1,3,4], Janusz Piontek[2], Piotr Kozlowski[1,4,*] and Marek Figlerowicz[1,5,*]

[1]European Center for Bioinformatics and Genomics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, 61-704, Poland, [2]Department of Human Evolutionary Biology, Institute of Anthropology, Faculty of Biology, Adam Mickiewicz University in Poznan, Poznan, 61-614, Poland, [3]Department of Hematology and Bone Marrow Transplantation, University of Medical Sciences, Poznan, 60-569, Poland, [4]Institute of Technology and Chemical Engineering, Poznan University of Technology, Poznan, 60-965, Poland and [5]Institute of Computing Science, Poznan University of Technology, Poznan, 60-965, Poland

*Correspondence address. Marek Figlerowicz, European Center for Bioinformatics and Genomics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, 61-704, Poland; Tel: +48-61-852-8919; E-mail: marekf@ibch.poznan.pl; Piotr Kozlowski, European Center for Bioinformatics and Genomics, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, 61-704, Poland; Tel: +48-61-852-8503; E-mail kozlowp@ibch.poznan.pl

Abstract

Metagenome analysis has become a common source of information about microbial communities that occupy a wide range of niches, including archaeological specimens. It has been shown that the vast majority of DNA extracted from ancient samples come from bacteria (presumably modern contaminants). However, characterization of microbial DNA accompanying human remains has never been done systematically for a wide range of different samples. We used metagenomic approaches to perform comparative analyses of microorganism communities present in 161 archaeological human remains. DNA samples were isolated from the teeth of human skeletons dated from 100 AD to 1200 AD. The skeletons were collected from 7 archaeological sites in Central Europe and stored under different conditions. The majority of identified microbes were ubiquitous environmental bacteria that most likely contaminated the host remains not long ago. We observed that the composition of microbial communities was sample-specific and not correlated with its temporal or geographical origin. Additionally, traces of bacteria and archaea typical for human oral/gut flora, as well as potential pathogens, were identified in two-thirds of the samples. The genetic material of human-related species, in contrast to the environmental species that accounted for the majority of identified bacteria, displayed DNA damage patterns comparable with endogenous human ancient DNA, which suggested that these microbes might have accompanied the individual before death. Our study showed that the microbiome observed in an individual sample is not reliant on the method or duration of sample storage. Moreover, shallow sequencing of DNA extracted from ancient specimens and subsequent bioinformatics analysis allowed both the identification of ancient microbial species, including potential pathogens, and their differentiation from contemporary species that colonized human remains more recently.

Keywords: microbiome; ancient DNA; NGS; metagenomics

Background

During the last 2 decades, a number of methods that permit isolation and sequencing of ancient DNA (aDNA) extracted from archaeological specimens have been elaborated. As a result, several complete genome sequences of long-dead organisms have been determined [1–5]. Typically, aDNA is sampled from teeth or bones as these are the densest tissues in vertebrates, which supports the preservation of aDNA in crystal aggregates [6, 7]. Ancient remains are usually deposited in soils for decades, so

Received: 14 March 2017; **Revised:** 9 May 2017; **Accepted:** 11 June 2017

DNA extracted is a mix of host DNA fragments and DNA from different organisms inhabiting the environment. To avoid the contamination that is usually present on bone/teeth surfaces (e.g., modern human, bacterial, fungal, or plant DNA), aDNA is sampled from interior parts, where the amount of aDNA is the highest. Despite applying rigorous DNA extraction protocols, the endogenous aDNA usually constitutes much less than 5% of the total extracted DNA, e.g., 1–5% for a Neanderthal [2] and 4% for a Mal'ta boy (24 000-year-old human) [8]. Of the remaining DNA, typically >95% is DNA of different microorganisms that have colonized the remains and have been acquired from the environment. When younger remains are considered (100–200 years old), the amount of endogenous aDNA is not much higher [9]; however, it is possible to obtain a sample containing even up to 70% of endogenous aDNA [4, 10]. This is because the preservation of DNA depends on many environmental factors [11, 12]. For example, cold temperatures [13, 14], microclimate of caves where remains have been buried [11], and swampy sediments [12] are known to enhance DNA stability. Moreover, it has been shown that the vast majority of DNA isolated from archaeological human remains belongs to bacteria that have colonized the remains [15, 16]. Bacteria amplify the porosity of bone and teeth [17, 18], making them more accessible to water, which may lead to so-called endogenous aDNA leaching [19] and replacement by exogenous DNA.

Some target enrichment procedures have been proposed to increase the amount of endogenous aDNA [20–24], and among them is the 2-step digestion method [14, 25, 26]. Interestingly, Orlando and colleagues showed that 2-step digestion does not influence the composition of bacterial communities (e.g., is the same in aDNA samples obtained after the first and second digestion runs) [9]. This observation suggests that niches exist deep within the bones and teeth. The environmental bacteria may reach these niches and preserve there.

Metagenome analysis has become a common source of information about microbial communities that occupy a wide range of ecosystems. Until today, environmental components [27] as well as flora of different human sites [28], e.g., oral [29, 30], skin [31], or intestinal [32–35], have been well characterized. In our study, we used this approach to analyze microorganisms that accompany archaeological human remains, which until now have not been exhaustively compared. Prior findings are limited to the rough identification of environmental bacteria [16] or concern a singular species, usually pathogenic. In the latter cases, the analyses were mostly undertaken after the identification of visible symptoms of past disease [36–38]. Efforts have also been undertaken to characterize human mummy intestinal [39] and colon [40] microbes, as well as the ancient oral microbiome [41–44]. They showed that aDNA of species that colonized the organism before death may be obtained. However, comprehensive characterization of microbial DNA accompanying human remains has never been done.

The current study was performed to characterize microorganisms associated with human archaeological remains. We used shotgun sequencing of DNA isolated from 161 human teeth collected from 7 archaeological sites dated from 100 AD to 1200 AD and stored under different conditions (e.g., museum or grave). For each individual sample, the microbiome was determined using Metagenomic Phylogenetic Analysis (MetaPhlAn2) based on multiple specific marker sequences derived from the genomes of microorganisms [45, 46]. Within this study, we focused on bacteria and archaea, which are known to constitute the majority of exogenous DNA in human archaeological remains [15, 16]. We checked whether microbial communities associated with specimens from different archaeological sites or of different ages were taxonomically and functionally distinct. We also attempted to identify microbes that may accompany the organism even before death and to distinguish bacteria/archaea that stem from postmortem contamination from those of original flora by studying their DNA damage patterns.

Data Description

We analyzed 161 human bone samples collected from 7 archaeological sites in Central Europe (Fig. 1A). As shown in Table 1, the samples differed by age (Roman Age group [KO and MZ] or Medieval group [GO, SI, NA, ME, and LO]) and by storage conditions (specimens that were in museum deposits for at least 20 years [long deposit: KO, MZ, SI, NA, and GO], relatively freshly discovered specimens [stored in museum deposit <5 years, short deposit: LO], or samples taken directly from an archaeological site [arch. site: ME]). Carbon isotope dating of the selected samples correlated well with dating based on archaeological analysis (see Supplementary Table S1).

Ancient DNA was always extracted from the roots of teeth. We drilled those parts of the roots that include both dentine and cementum. In all cases, enamel and cementum were preserved. Subsequently, all DNA samples were subjected to

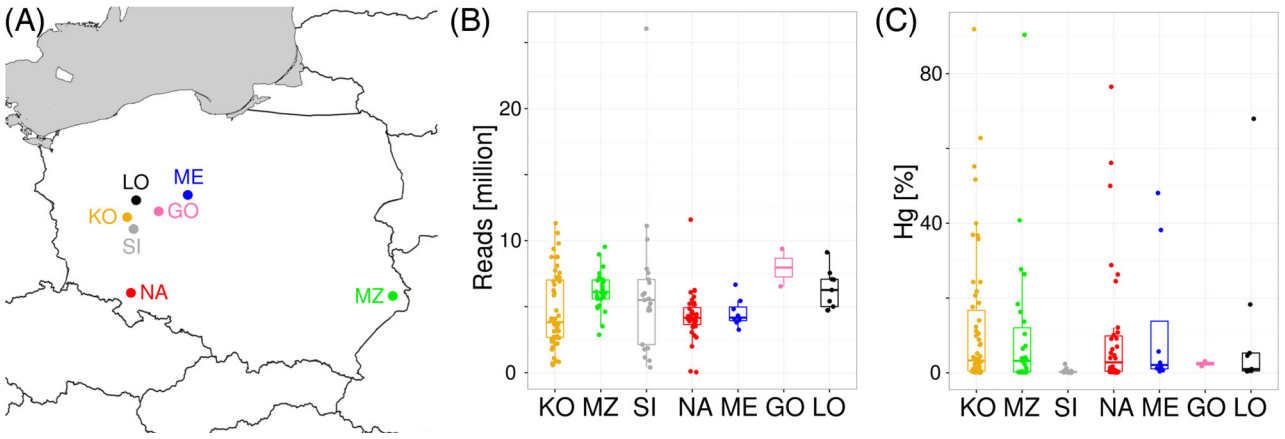


Figure 1: (**A**) The geographical positions of archaeological sites. KO and MZ are from the Roman Age group, and SI, NA, ME, GO, and LO are from the Medieval Group. Samples from ME were collected directly at the archaeological site. (**B**) Number of filtered reads (y-axis) per archaeological site (x-axis). (**C**) Percentage of reads mapped to the human genome (y-axis) per archaeological site (x-axis).

Table 1: Characteristics of the samples extracted from ancient human remains

Archaeological site	ID	Sample no.	Sample no. that passed selection	Dating	Date of excavation	Storage conditions	Sample type
Roman Age group							
Kowalewko	KO	58	48	100–300 AD	1990s	Long deposit	Tooth
Masłomęcz	MZ	27	24	200–400 AD	1970–1990	Long deposit	Tooth
Medieval group							
Sowinki	SI	21	19	1000–1100 AD	1980s	Long deposit	Tooth
Niemcza	NA	36	31	900–1000 AD	1960s	Long deposit	Tooth
Markowice	ME	8	8	1000–1200 AD	2014	Arch. site	Tooth
Gniezno	GO	2	2	1000–1200 AD	1980s	Long deposit	Tooth
Łęgowo	LO	9	8	1000–1200 AD	2013–2015	Short deposit	Tooth

shallow next-generation sequencing (NGS) with the usage of an Illumina single-end standard protocol (including blunt-end DNA repair) and 75 bp sequencing run. Altogether, 846.5 million reads were obtained. On average, 98.6% of reads passed trimming and quality filtration. After filtration, for 161 samples, the average number of reads per sample was 5 143 975 (median = 4 730 243; range = 34 857–26 055 295). In further analysis, we removed 8 samples that did not meet the arbitrary criterion of minimal raw reads number (<1 million). The average numbers of reads differ between archaeological sites (Kruskal-Wallis: $P = 0.0166$), but not between types of sample storage (Wilcoxon: $P = 0.2685$) or age (Wilcoxon: $P = 0.5607$) (Fig. 1B). Detailed information on each sample is summarized in Supplementary Table S1.

All reads were mapped to the reference human genome, and the percentage of human reads was determined for each sample. As shown in Fig. 1C, the fraction of human aDNA ranged from 0.01% to 91.9%; however, in most cases (100 samples), it was less than 5%. Nine samples had more than 50% human aDNA content. Differences in the amount of human aDNA content were observed for different archeological sites (Kruskal-Wallis: $P = 6.124e-05$), but not for freshly recovered and stored in museum samples (Wilcoxon: $P = 0.3160$). Marginal statistical significance was observed between older (KO, MZ) and younger (SI, NA, ME, GO, LO) samples (Wilcoxon: $P = 0.0467$), with a higher share of endogenous human DNA in older samples (average = 11.7% and 7.8%, median = 3.2% and 0.75%, for older and younger samples, respectively).

Analyses

Microbiomes of human archaeological remains

To characterize the microbiomes of analyzed archaeological samples, we used MetaPhlAn2. The program identifies bacteria/archaea, viruses/viroids, and unicellular eukaryotes using homology-based classification of NGS reads by alignment with predefined taxa-specific marker sequences [45]. The number of reads mapped to MetaPhlAn2 markers ranged from 708 (sample KO_014) to 95 950 (sample KO_006). Two samples with <1000 reads mapped to the marker sequences were removed from further analyses as the marker coverage is crucial for proper microorganism detection [46].

For the remaining 151 samples, our analyses (Fig. 2A) showed that the majority of reads mapped to bacterial or archaeal markers (76.4%) and 23.4% to virus/viroid markers. The remaining 0.2% constituted eukaryotes (present in 13 samples; 0.6–8.2%), which were subsequently identified as fungi, protists, or protozoa. The contributions of the particular types of microorganisms differed substantially between individual samples (in 12 samples, we found only bacteria; in sample KO_28, only viruses were identified) (Fig. 2C). However, these differences did not correlate with archaeological site (multivariate analysis of variance [MANOVA]: $P = 0.0532$) (Fig. 2B), sample age (MANOVA: $P = 0.2054$), or storage conditions (MANOVA: $P = 0.7672$).

The virus fraction varied from 0.1% to 99% between samples. Analysis of virus taxa showed that most of them were associated with plants; hence, we reasoned that they may have been acquired from the environment and were possibly indigenous flora. The most abundant viruses, *Dasheen mosaic virus* (58% of all identified viruses/viroids) and *Vicia cryptic virus* (26.7%), are both known to infect plants. Subsequently, 5 viruses and 1 viroid constituted less than 2.5% each of all identified viruses/viroids, and also all were found to be associated with plant genera (*Ageratum, Sauropus, Cichorium,* or *Malvastrum*). The remaining viruses were of low abundance (<1%) and were usually present in no more than a single sample. It is also noteworthy that we identified within our samples *Propionibacterium* phage—a double-stranded DNA (dsDNA) virus that is associated with oral microbiome [47, 48]. Detailed information on the microorganism composition in individual samples is available in Supplementary Table S2.

Characterization of bacteria and archaea in human archaeological remains

In the next step, we focused on the prokaryotic component of the analyzed microbiomes. We decided to exclude from this analysis samples with a very high fraction of viruses/viroids. As a result, 11 samples with fewer than 1000 reads mapping exclusively to bacterial/archaeal MetaPhlAn2 marker sequences were removed as they did not ensure a reliable microbiome profiling.

Altogether, 25 bacterial and 4 archaeal classes were identified in exogenous DNA of the analyzed samples, and among them, 6 bacterial classes accounted for >1% of identified bacteria/archaea. The most abundant classes were *Actinobacteria* (average = 57%; range = 0.18–98.9%), 3 classes of *Proteobacteria* (*Alphaproteobacteria* [average = 6%; range = 0–65.5%], *Betaproteobacteria* [average = 7%; range = 0–83.6%], *Gammaproteobacteria* [average = 12%; range = 0–95.4%]), *Acidobacteria* (average = 5%; range = 0–39.7%), and *Clostridia* (average = 4%; range =0–76.8%) (Fig. 3A). Although most of the bacteria belonging to the first 5 classes are typically found in the environment (wide range of soils, waters) [27, 49], some of their taxa were human flora components. For example, *Corynebacterium matruchotii* (*Actinobacteria*) [50] and *Lautropia mirabilis* (*Betaproteobacteria*) [51, 52] represented more than 5% of the DNA in 4 samples: KO_046b, NA_121, NA_123, LO_166 and KO_005, KO_006, KO_046b, LO_166, respectively (Supplementary Table S2). *Clostridia* and *Bacteroidetes* are

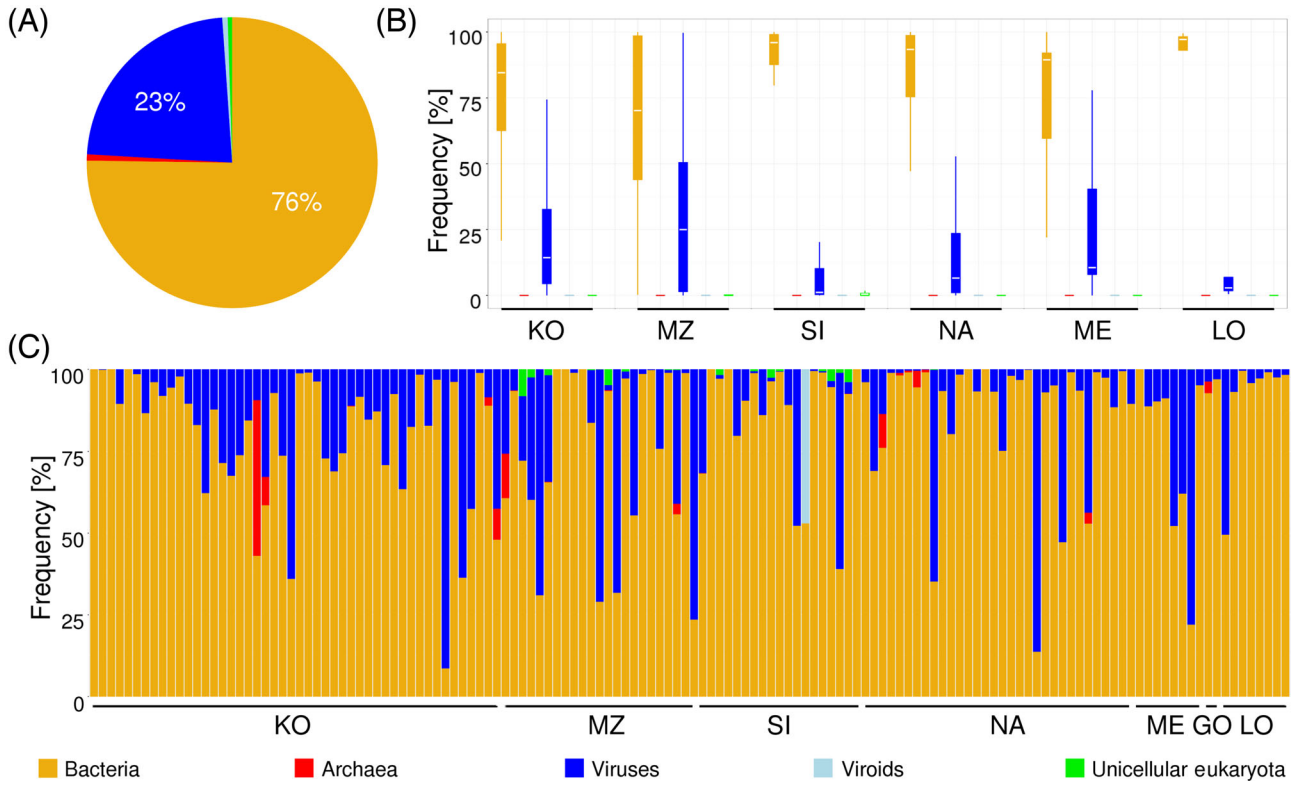


Figure 2: Microorganism kingdoms detected in analyzed archaeological samples. (**A**) Pie chart representing overall frequency of microorganism kingdoms in archaeological samples. (**B**) Box and whisker plot representing the distribution of frequencies of particular microorganism kingdoms in archaeological sites (GO not shown as it includes only 2 samples). (**C**) Stacked barplot indicating the frequency of microorganism kingdoms in a particular sample. Each bar represents an individual sample. Samples are ordered by the archeological sites. The color legend for all plots is shown at the bottom.

known to include many species inhabiting the human oral cavity or intestines [29]. Additionally, we found, in individual samples, markers characteristic for human pathogens, e.g., *Pseudoramibacter alactolyticus* (*Clostridia*) in sample MZ_88 [53] and *Bordetella parapertussis* (*Betaproteobacteria*) [54] in sample SI_084; *Clostridium sordellii* and *Clostridium tetani* (*Clostridia*) [55, 56] were found in 2 samples and 1 sample, respectively. Prokaryotic profiles differed substantially between individual samples (Fig. 3C) but did not differ between specific archaeological sites (MANOVA: $P = 0.3650$) (Fig. 3B), sample ages (MANOVA: $P = 0.3550$), or storage conditions (MANOVA: $P = 0.4729$). Similar high variation between individual samples and lack of specificity to archaeological sites was observed when prokaryotes were divided into groups based on gram +/- type (MANOVA: $P = 0.4364$) or oxygen requirements (aerobic, facultative aerobic, anaerobic, facultative anaerobic; MANOVA: $P = 0.5726$) (see Supplementary Figs S1 and S2).

The identification of singular prokaryotic taxa that are human- rather than environment-related motivated us to determine the fraction of microbes potentially associated with humans. All identified bacteria and archaea were divided on a genus level into 2 groups: environmental and human-related. The latter was further divided into 3 subgroups: oral, potential pathogens, and other (mostly gut). The genus characteristics were inferred based on the features of species identified by MetaPhlAn2. A genus was classified as human-related only if all species of this genus identified in our samples were human-related. The analysis showed that the majority (85.19%) of all bacteria/archaea were environmental (coming from soil and/or water); however, a substantial fraction of the investigated taxa (14.81%) were human-related, including 12.43% of microbes typical for human oral flora, 1.33% of potentially pathogenic bacteria, and 1.05% of other (see Fig. 4A and B). As shown in Fig. 4C, the fraction of human-related genera varied significantly among samples, and some of these genera constituted most of the exogenous DNA. Although the fraction of human-related genera did not differ significantly between archaeological sites (1-way ANOVA: $P = 0.7480$), it was noteworthy that this fraction was highest in NA, the archaeological site dated to the Middle Ages, from which the samples had been stored in a deposit for more than 20 years (see Fig. 4B). Interestingly, there was no relation between prevalence of human-related microbes and the levels of virus/viroid accumulation or the level of endogenous human aDNA (see Supplementary Table S1). The identification of human-related species in ancient remains raised the question of whether some of them accompanied the individual even before death.

Among all samples, the most frequent genera were the soil bacteria *Brevibacterium* (8.5% of all; present in 53 samples >1%; max. 71%) and *Kribbella* (8.4% of all; present in 60 samples >1%; max. 70%). The most abundant oral genera were *Bacteroidetes* (1.6% of all; present in 23 samples >1%; max. 28%), *Desulfobulbus* (1.4% of all; present in 25 samples >1%; max. 44%), and *Eubacterium* (1.4% of all; present in 20 samples >1%; max. 32%). *Methanobrevibacter* (0.8% of all; in 7 samples >1%; max. 34%), typically found in the human digestive system and in the oral cavity, was the most abundant taxon in the other human-related group as only *M. smithii* (human gut flora component) were identified in our samples (Supplementary Table S2). However, it must be pointed out that the genus *Methanobrevibacter* also contains

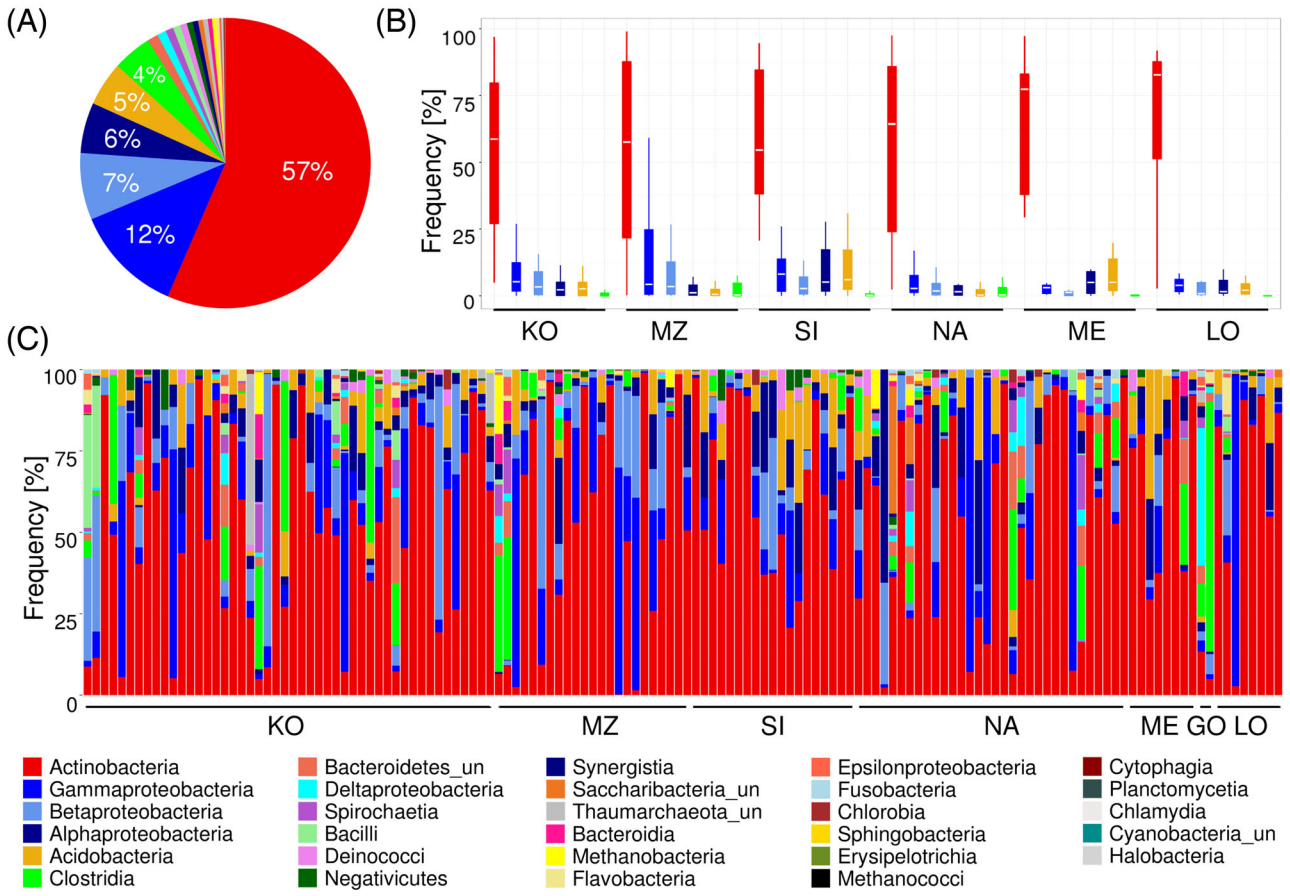


Figure 3: Bacterial and archaeal classes detected in analyzed archaeological samples. (**A**) Pie chart representing overall frequency of bacterial and archaeal classes in archaeological samples. (**B**) Box and whisker plot representing the distribution of frequencies of the 6 most abundant bacterial classes (present in at least 1%) of archaeological sites (GO not shown as it includes only 2 samples). (**C**) Stacked barplot indicating the frequency of bacterial and archaeal classes in a particular sample. Each stacked bar represents an individual sample. Samples are ordered by the archeological sites. The color legend for all plots is shown at the bottom.

species commonly found in the oral flora, e.g., *M. oralis*, which was not identified within analyzed samples; *Bordetella* was the most abundant taxon classified as a potential human pathogen (*B. pertussis* is known to cause pertussis; 1.2% of all; in 16 samples >1%; max. 60%).

In general, 89% of the analyzed prokaryotes were aerobic or facultative aerobic (Supplementary Fig. S1), and 63% were gram-positive (Supplementary Fig. S2). However, in the human-related group only (Table 2), the percentage of aerobic or facultative aerobic taxa was smaller (24%, 49%, and 54% for oral, pathogen, and other groups, respectively). Additionally, we found that the gram-negative prokaryotes dominated in the oral group (55%) and gram-positive prokaryotes in the other human-related group (70%). This slight dominance of gram-negative taxa in the oral group might be caused by lysozyme presence in an oral cavity that preferentially protects against gram-positive bacteria [57]. We additionally noticed that gram-negative species dominated (68%) in the potential pathogen group. These characteristics seem very useful for preliminary assessment of bacterial populations accompanying human remains.

To further investigate whether the prokaryotic profile permits classification of individual samples into specific groups (e.g., samples of similar age or storage conditions or samples from the same archaeological site), we performed Principal Coordinates Analysis (PCoA; Jaccard distance) on 4 taxonomic levels (class, family, genus, and species) (Fig. 5). Samples grouped into 1 big cluster in graphs created on all taxonomic levels. In the PCoA graphs generated on the family, genus, and species levels, there was 1 more significantly smaller cluster visible. Importantly, none of these clusters segregated samples according to the abovementioned features (age, storage, and site). Principal Component Analysis (PCA) (see Supplementary Fig. S3) and the Shannon diversity index (see Supplementary Table S1) again revealed high variation between individual samples at all analyzed taxonomic levels but did not show separation by sample source (species level, 1-way analysis of variance (ANOVA): $P = 0.5660$), sample age (species level, t-test: $P = 0.5535$), or storage type (species level, t-test: $P = 0.3516$). We also tested a hypothesis that the occurrence of some human-related or environmental bacteria might be associated with archeological sites. We performed PCA (Supplementary Fig. S4) and hierarchical clustering (Supplementary Fig. S5) on selected bacterial genera and found out that neither human-related nor environmental microbes segregated samples according to the archeological site, age, or storage type. Finally, we clustered samples based on 10-mer distances between exogenous reads (see the Methods section) and again observed no segregation according to the archeological site, age, or storage type (Supplementary Fig. S6).

In order to confirm that the major source of microbes observed in human archeological samples was the environment, we compared their microbiomes with the microbiomes of

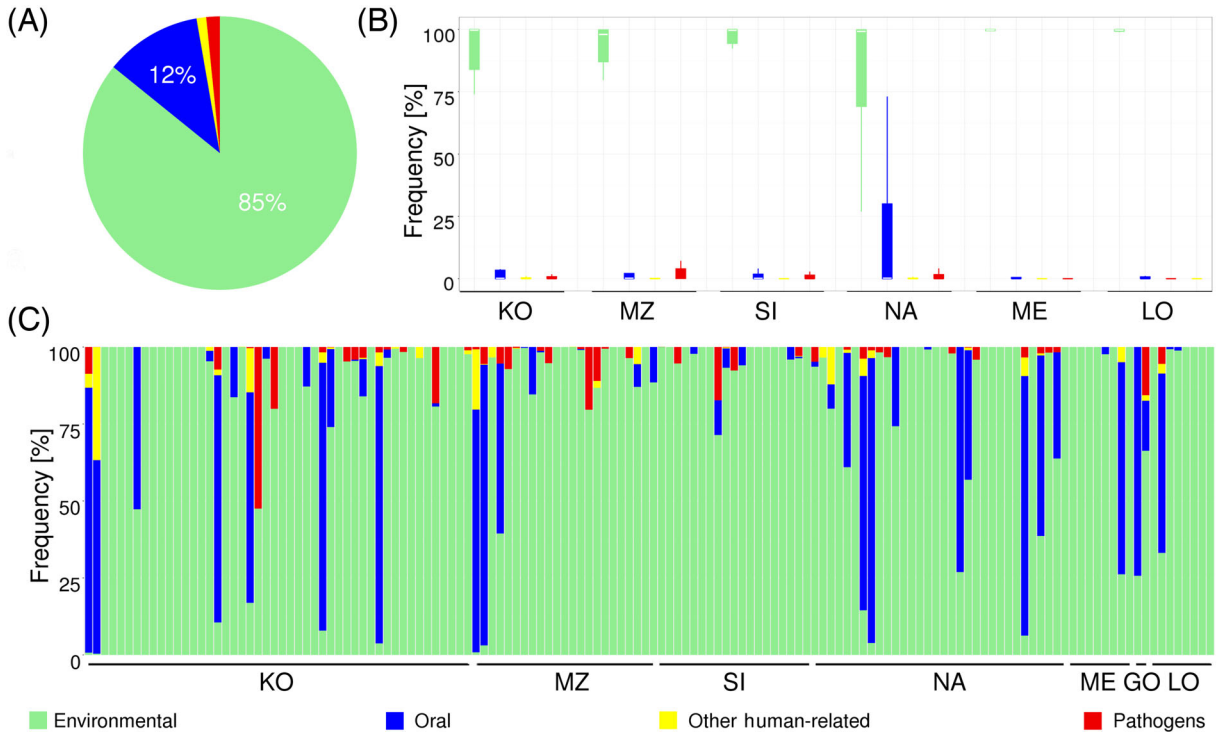


Figure 4: Bacterial and archaeal types (environmental [light green], oral [blue], other [yellow], and pathogenic [red]) detected in analyzed archaeological samples. (**A**) Pie chart representing overall frequency of bacterial and archaeal types in archaeological samples. (**B**) Box and whisker plot representing the distribution of frequencies of bacterial and archaeal types in archaeological sites (GO not shown as it includes only 2 samples). (**C**) Stacked barplot indicating the frequency of bacterial and archaeal types in a particular sample. Each stacked bar represents an individual sample. Samples are ordered by the archeological sites. The color legend for all plots is shown at the bottom.

Table 2: The percentage of bacteria/archaea of a given respiratory type (facultative [aerobic/anaerobic] and gram stain type [positive/negative] within environmental and human-related groups [oral, pathogenic, or other])

Group	(Facultative) anaerobic	(Facultative) aerobic	Gram-positive	Gram-negative
Environmental	4%	96%	66%	34%
Oral	76%	24%	45%	55%
Pathogenic	51%	49%	32%	68%
Other human-related	46%	54%	70%	30%

humans [58] and soils [27] by PCoA at the genus level (see Supplementary Fig. S7).

Validation of data obtained using shallow sequencing

All results presented above were obtained with the use of datasets generated by relatively shallow sequencing (on average ~5 million reads per sample). To check the reliability of our results, we determined for the selected samples to what extent the composition of microbiomes (on the class level) is affected by the depth of sequencing. For this analysis, we used 11 representative samples differentiated in terms of (i) filtered read numbers obtained in the shallow sequencing experiment (~2–8 million), (ii) number of reads mapped to the MetaPhlAn2 markers (~1000–60 000), and (iii) prokaryotic fraction (~10–90%). Eight samples were sequenced to the depth of ~50 million reads and 3 to the depth of ~100 million reads. Subsequently, we ran a MetaPhlAn2 profiling analysis on deep sequencing datasets. As expected, the total number of filtered reads as well as the number of reads mapping to the MetaPhlAn2 marker sequences increased significantly (about 9-fold); however, the Shannon diversity indexes and microbial compositions remained intact (correlation $R = 0.91–0.99$) (Fig. 6A; Supplementary Table S3). We obtained similar results when we analyzed 3 other taxonomic levels with somehow decreasing R with the depth of taxonomic level (average $R = 0.96, 0.90, 0.88, 0.78$ for class, family, genus, and species levels, respectively) (Fig. 6B; Supplementary Fig. S8). It is noteworthy that sample KO_030, second lowest in the number of raw reads, displayed very low correlation ($R = 0.35$) on a species level when results obtained based on shallow sequencing (~2.6 million reads) and deep sequencing (~47 million reads) were compared (Supplementary Fig. S8 and Supplementary Table S3). Overall, the correlation coefficient R and statistical significance values ($P < 0.0001$ in most cases) (see Supplementary Fig. S8) were still very high and confirmed that the microbial profiles obtained based on the shallow sequencing datasets are reliable and do not change significantly when datasets generated in much deeper sequencing are used to establish them.

Analysis of age-related aDNA damage patterns

Finally, to verify whether identified human-related prokaryotes are ancient species that colonized the human body before death or are modern contaminants, we analyzed the signatures of

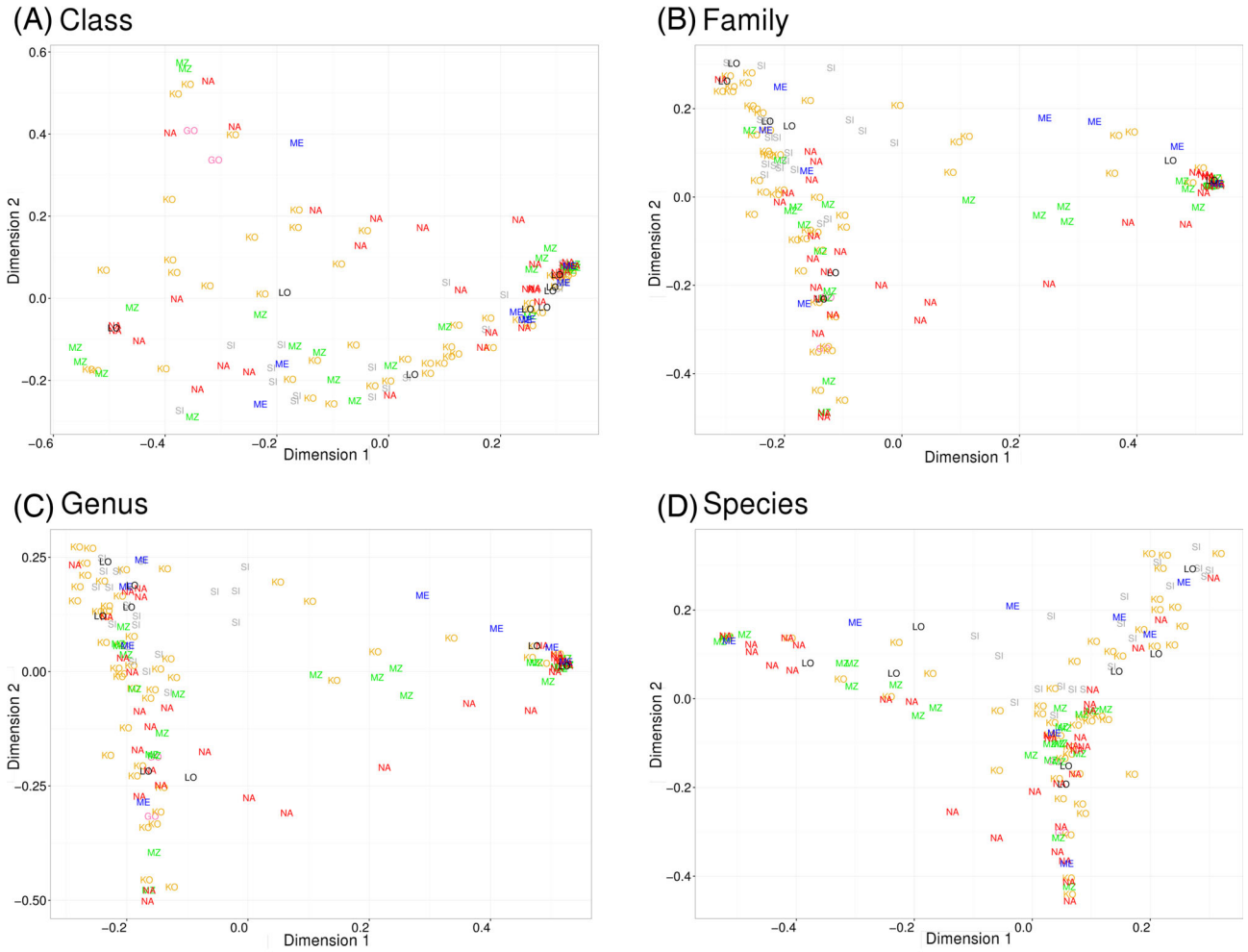


Figure 5: Principal coordinate analysis of microbial compositions at 4 taxonomic levels: (**A**) class, (**B**) family, (**C**) genus, and (**D**) species. Samples from certain archaeological sites are marked in different colors and labeled with an archaeological site ID.

age-related DNA damage. Age-related DNA damage was evaluated with the usage of mapDamage2.0 [59], which simulates the posterior distribution of (i) deamination in single-stranded DNA (δ_s), (ii) deamination in dsDNA (δ_d), and (iii) the level of DNA fragmentation (λ, represented as: 1/λ-1) [60, 61].

For this analysis, we used sequences of 77 complete genomes of the most representative prokaryotes of 313 identified in our samples (Supplementary Table S4). The 77 selected species constituted 93% of all identified bacteria/archaea, and each of the selected species accounted for at least 10% in at least 1 sample (Supplementary Table S4A). The remaining species represented only 7% of the total microbial DNA, and they typically accounted for less than 1% of an individual sample. Subsequently, for each sample, we mapped all reads against (i) all 77 selected genomes; (ii) a subset of 55 environmental bacteria genomes; (iii) a subset of 14 oral bacteria genomes; (iv) a subset of 3 gut bacteria and archaea genomes; (v) a subset of 5 potential pathogen genomes; and (vi) a subset consisting of all human-related bacterial genomes (22 genomes; oral, gut, and pathogens). Additionally, we mapped reads against a reference human genome to compare in each sample the level of DNA damage in human and microbial genomes. The comparison of DNA damage signatures in human and microbial DNA in individual samples is presented in Supplementary Fig. S9 and Supplementary Fig. S10.

As shown in Fig. 7, the average DNA damage determined for all 77 microbial genomes decreased with the increase in environmental bacteria fractions. Microbial DNA damage values differed significantly between samples with different fractions of environmental components (1-way ANOVA: (δ_s) $P = 0.0413$; (δ_d) $P = 0.0001$; (1/λ-1) $P < 0.0001$). The samples with the lowest (<25%) contribution of environmental bacteria displayed the highest level of microbial DNA damage (on average: $\delta_s = 0.2643$, $\delta_d = 0.0067$, 1/λ-1 = 2.7933), comparable with those observed for endogenous human aDNA (on average: $\delta_s = 0.3571$, $\delta_d = 0.0279$, 1/λ-1 = 1.6667). Noticeably, the damage of human aDNA did not depend on the amount of environmental bacteria in a sample (1-way ANOVA: (δ_s) $P = 0.8630$; (δ_d) $P = 0.3530$; (1/λ-1) $P = 0.4770$) (Fig. 7).

In the next step, for each sample, we calculated the DNA damage values separately for the following groups of bacterial species: (i) environmental; (ii) all human-related; (iii) oral; (iv) gut; and (v) potential pathogens. We compared these values with corresponding values determined for the endogenous human aDNA in the same sample. As is shown in Fig. 8, the highest differences between the levels of human and microbial DNA damage were observed for environmental bacteria that showed very little DNA damage (on average: $\Delta\delta_s = 0.1767$; $\Delta\delta_d = 0.0264$; Δ(1/λ-1) = 1.3089). It is also shown in Fig. 8 that the DNA damage of

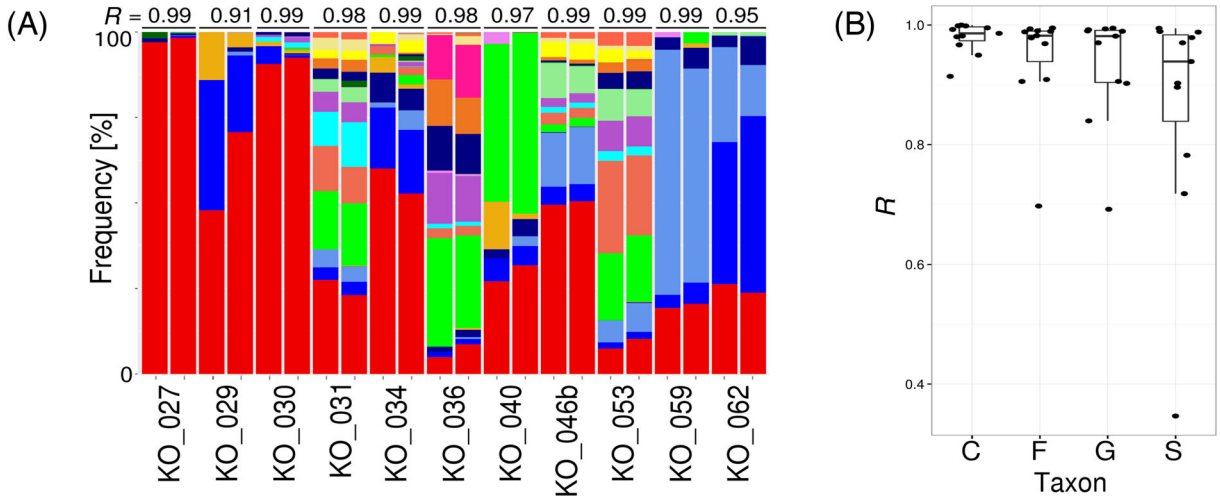


Figure 6: (**A**) Comparison of bacterial and archaeal profiles (stacked barplot) on the class level based on shallow and deep sequencing of the selected 11 samples (sample ID is indicated on the x-axis; the first bar in a pair is shallow, and the second is deep sequencing). The correlation coefficient *R* is placed above each shallow/deep stacked bar pair. The color legend is the same as in Fig. 3. (**B**) Correlation *R* values (y-axis) for shallow and deep sequencing pairs on different taxonomic levels (C: class; F: family; G: genus; S: species).

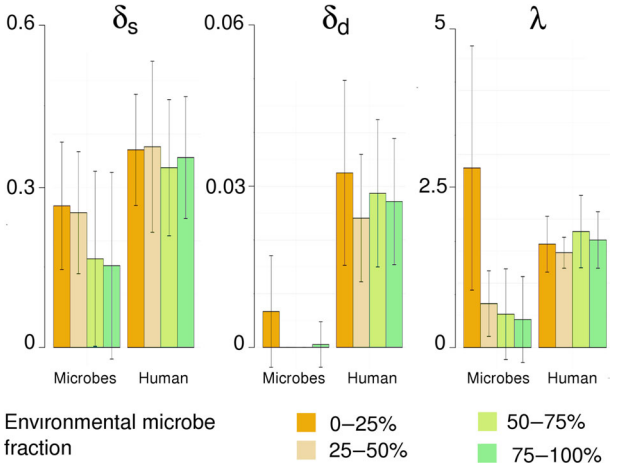


Figure 7: The DNA damage in samples with different fractions of environmental bacteria/archaea. Barplots indicating deamination rate in single-stranded DNA overhangs (δ_s) and double-stranded DNA fragments (δ_d) in microbial (left-hand site) and human DNA (right-hand site), grouped based on the fraction of environmental bacteria/archaea in the sample and the length of single-stranded DNA overhangs (λ, expressed as: $1/\lambda$-1) calculated for 77 representative bacteria/archaea and endogenous human aDNA. Samples were grouped based on the fraction of environmental bacteria/archaea in a sample (0–25%, 25–50%, 50–75%, and 75–100%).

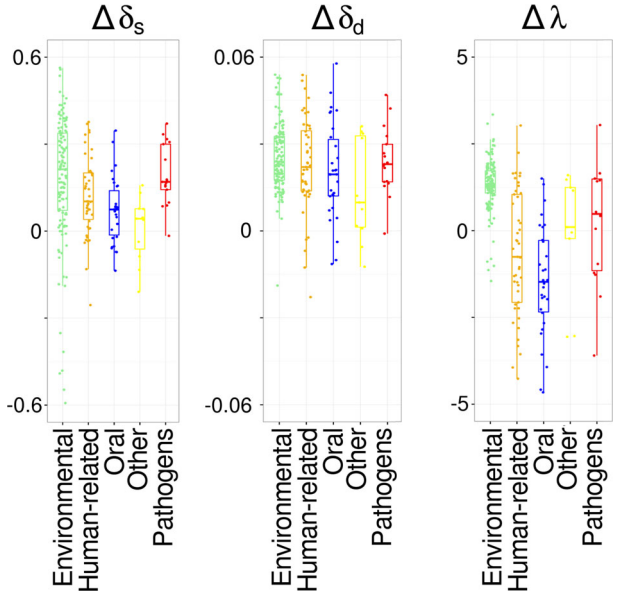


Figure 8: The differences of DNA damage levels ($\Delta\delta_s$, $\Delta\delta_d$, $\Delta\lambda$, expressed as: $\Delta(1/\lambda\text{-}1)$) of bacteria/archaea species belonging to the 5 groups (environmental, all human-related, oral, gut, and pathogen) in comparison to damage levels in human aDNA. Boxes, whiskers, and dots represent the distribution of differences in DNA damage levels of particular bacterial/archaeal groups. Each dot represents the difference in an individual sample. The color legend is the same as in Fig. 4 (all human-related species are in orange).

human-related species is similar to that observed for human aDNA (average: $\Delta\delta_s = 0.1224$; $\Delta\delta_d = 0.0278$; $\Delta(1/\lambda\text{-}1) = -0.8805$). The variations in the obtained values may result from different rates of microbial DNA decay as well as from misclassification of some microbial species.

Actinobacteria, as well as all classes known to be non-spore-forming, are more durable than other bacteria [62]. Thus, we used *Actinobacteria* (the most abundant class in our study) to analyze whether the differences between environmental and human-related species in DNA damage levels were influenced by different rates of damage in various microbe types. Within the human-related group (oral), we identified 3 species belonging to *Actinobacteria*, present in 12 samples in >5%. Within the environmental group, we identified 12 species, present in 104 samples as >5%. The DNA damage pattern comparison again showed a higher damage rate in human-related rather than environmental *Actinobacteria* (t-test: (δ_s) $P = 0.0091$; (δ_d) $P = 0.0299$; ($1/\lambda$-1) $P = 0.0004$) (Fig. 9). This finding confirmed that the larger accumulation of DNA damage observed for human-related species was not microbe type–specific. Therefore, different DNA damage levels in environmental and human-related bacteria did not result from differences in the stability of bacterial genomes but from their age.

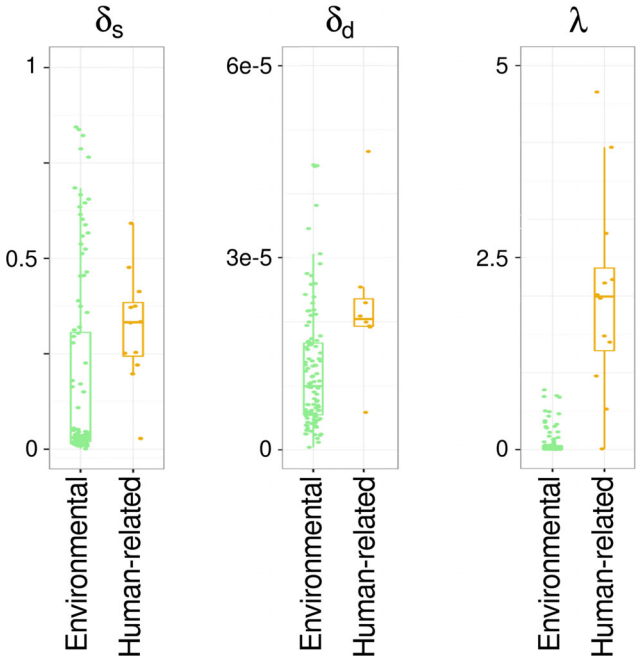


Figure 9: DNA damage level (δ_s, δ_d, $1/\lambda$-1) in environmental and all human-related *Actinobacteria* species. Boxes, whiskers, and dots represent the distribution of DNA damage levels in particular samples. The color legend is the same as in Fig. 4 (all human-related species are in orange).

Discussion

This study represents one of the most comprehensive analyses of the microbiomes that accompany ancient human skeletal remains. Accordingly, the analyzed DNA could come from (i) microorganisms that formed the human microbiome and existed in the human organism before death or (ii) environmental species that contaminated human remains or participated in the body's decomposition process.

In this study, we analyzed 161 datasets (total sequencing > 63 Bbp) collected from 7 different archaeological sites. We employed a novel approach based on a clade-specific genes analysis (MetaPhlAn2) [63]. This method relies on the database of marker sequences derived from whole genomes that unequivocally allows for the identification of microbial taxa down to the species level. Moreover, this method works not only for prokaryotes but also for all unicellular organisms and viruses. In contrast, a traditional approach based on the analysis of a singular 16S rRNA marker gene [64] is limited to the identification of bacteria/archaea at the genus level at most, thus being less accurate [65]. The applied methodology allowed us to determine the amount and type of viruses and fungi as well as bacteria and archaea in the analyzed samples. Notably, we showed that shallow sequencing (the average number of reads per analyzed sample was ~5 million) permitted retrieval of reliable microorganism profiles. The result, validated using deeper sequencing (to 50–100 million reads), confirmed that our findings from shallow sequencing were trustworthy, although it has to be noted that the accuracy slightly decreased with the taxonomic levels (Fig. 6).

The thorough analyses of all microorganisms as well as only prokaryotes revealed that there are substantial differences between individual samples, but the differences were not characteristic for particular sample types. We showed that there was no correlation between the composition of microbial population and geographical place, sample age, or storage history. It has to be noted, however, that our results do not exclude completely the effect of storage on microbial composition. Such an effect may exist, but it is too low to be detected due to very high variation in microbial composition between individual samples. On the other hand, the high variance may suggest that pores in the teeth constitute independent variable micro-environments, some easily accessible to an exogenous DNA, while others not (or temporarily not), which promotes the stochastic and unique microbial composition. Moreover, the comparison between museum specimens (more than 20 years from excavation) and relatively freshly sampled materials suggested that the treatment applied before storage (e.g., washing) and storage itself do not influence the microorganism composition in tooth niches. Most likely, the migration of bacteria or a diffusion of microbial DNA and other microorganisms must be most intense when the remains are in direct contact with soil or water and negligible when placed in a relatively sterile environment, such as a museum deposit. These findings are of a certain importance as they indicate that studying ancient microbiome museum specimens may be as good as studying freshly discovered specimens.

Overall, we identified 25 microbial classes; the genetic material of 6 of them comprised more than 1% of all bacterial and archaeal DNA (Fig. 3A). Most of identified genera were ubiquitous bacteria belonging to the *Actinobacteria* class, such as *Brevibacterium, Kribbella, Actinoplanes,* and *Streptosporangium,* which are typically found in a wide range of soils and waters (Fig. 4). The obtained results are in line with previous findings [15, 16, 49], as well as with the common notion that DNA contamination of fossil remains comes from the soil and water. In addition, in some samples, we identified a substantial portion of microbes associated with the human body, mainly with the oral cavity, belonging predominantly to the *Clostridia* (*Eubacterium*, *Pseudoramibacter*), *Actinobacteria* (*Propionibacterium, Corynebacterium, Actinomyces*), and *Bacteroidia* (*Tannerella*) classes. Moreover, we identified 2 bacterial and 1 archaeal genera typical of the human digestive system: *Neisseria*, *Escherichia* (*Proteobacteria* class), and *Methanobrevibacter* (*Methanobacteria* class), as well as 4 potential human pathogens: *Bordetella*, *Stenotrophomonas*, *Bartonella* (*Proteobacteria* class), and *Clostridium* (*Clostridia* class).

The analyses of viruses present in the aDNA samples revealed that the substantial fraction of them accounted for 2 plant RNA viruses, whose genomes are composed of ssRNA: *Dasheen mosaic virus* (58% of all identified viruses/viroids) and *Vicia cryptic virus* (26.7%). As our NGS library preparation protocol was not designed for RNA sequencing (lack of the reverse transcription step), this result is rather unexpected and has to be interpreted with caution. Identification of RNA viruses may be potentially explained by (i) unintended reverse transcription of viral RNA either by some environmental reverse transcriptase or by DNA polymerase used for the NGS library preparation (DNA polymerase can display residual activity on RNA template, especially if the latter is in a relatively high concentration); or (ii) missmapping of some reads to markers of RNA viruses and consequently microbial misclassification. The second possibility may be enhanced by the very high genetic variability of RNA viruses. Thus, further studies are required to solve this problem.

DNA damage pattern analysis of the identified environmental and human-related microbes showed that the DNA of human-related species had significantly higher numbers of C → T and G → A substitutions, which are typical of aDNA. Moreover, their damage levels were comparable with those observed for endogenous human aDNA in the corresponding samples (see Supplementary Figs S9 and S10). According to the

assumption that environmental microbes colonized archaeological bones relatively recently, DNA of environmental microbes displays a minimal amount of aDNA characteristic signatures. There is a possible bias caused by different dynamics of post-mortem DNA modifications in various bacteria types [66]. It has been shown that non-spore-forming *Actinobacteria* are more durable than endospore formers such as *Bacillaceae* and *Clostridiaceae* [62]. The DNA damage analysis within the *Actinobacteria* class only revealed that human-related *Actinobacteria* species manifested aDNA damage patterns and that the environmental species showed the opposite pattern. This additional analysis supported our results and showed that the different levels of aDNA damage in environmental and human-related groups were not caused by the differences in bacterial genome stability. This also suggested that the identified human-related species may truly accompany the individual even before death. For environmental components, it seems that their DNA is relatively young and must have been acquired recently. One possible explanation is that some niches in the teeth are open and DNA exchange occurs continuously with the environment, whereas other niches are hardly accessible, so only endogenous species may reach and be preserved in these niches.

Many human pathogens belong to the same genera as environmental species [67]. For example, *Bordetella bronchiseptica* can survive in the environment and is present in a wide range of animals [68, 69]. The genus *Bordetella* also contains species that are commonly found in the environment, such as *B. petrii*. *Clostridium tetani* is known to be the causative agent of tetanus, but it is often found in soils and participates in the body's decomposition process. Hence, the identification of potential pathogens in body remains may not certainly mean that the individuals were infected with the bacterium before death. In fact, our analyses revealed that the DNA of some of the identified potential pathogens showed the DNA damage degree closer to the damage of environmental microbes than to the damage of human-associated ones.

We showed that identification of candidate bacteria/archaea species accompanying the organism before death is possible using standard aDNA extraction protocols and shallow shotgun sequencing. The use of microbial markers derived from whole genomes is crucial as aDNA typically lacks huge blocks of information and using only the 16S rRNA gene as a marker may be not sufficient.

Our results indicated that not only fresh samples but also museum specimens seem to be good sources of ancient microbial DNA. Moreover, this methodology may be employed for screening remains without visible signs of disease, which provides the huge possibility of finding ancient pathogens for further analysis. In particular, this may provide additional knowledge to the fields of epidemiology and bacterial population genomics, allowing for the investigation of the rate of bacterial evolution, and may even bring forth some information on the ancient human diet.

Potential Implications

Here, we showed that the composition of the microbiome of archeological remains is highly variable but does not show any evident correlation with the method or duration of sample storage. That opens up a possibility to study on a wide range the microbiomes present in human and also non-human remains. We also demonstrated that it is possible to obtain reliable profiles of microbiomes from single-end shallow next-generation sequencing that allow the cutting of time and costs for any microbiome study. The presented procedures might be used as a first step of ancient pathogen identification, especially when a large set of samples with no apparent infection symptoms is considered. Finally, our studies revealed that by analyzing the DNA damage pattern, one can identify the putative ancient microorganisms present in the microbiome of archeological remains.

Methods

Experimental procedures

DNA extraction from teeth was performed in the ancient DNA laboratory at the Faculty of Biology, Adam Mickiewicz University, Poznan, Poland. To avoid contamination that might be introduced through laboratory manipulations, all reagents used for DNA purification (buffers, water) and small plastic materials were UV irradiated (254 nm) for 1 hour. The surface of the teeth was cleaned with 0.5–5% NaOCl, rinsed with sterile and UV-irradiated water, and exposed to UV (254 nm) for 2 hours per each site. Following UV irradiation, the roots of the teeth were drilled using Dremel®, and bone powder was collected to sterile tubes (2 ml) and digested for 48 hours at 56°C in a buffer containing EDTA, UREA, and proteinase K, as described in Juras et al. [70]. After digestion, DNA was purified using the MinElute kit (QIAGEN, RRID:SCR_008539) according to Yang et al. [71] and Malmstrom et al. [72]. Genomic libraries preparation was performed as described in Meyer and Kircher [73]. The protocol comprised a blunt-end repair step. A single-stranded DNA overhanging 5′- and 3′-ends was filled in or removed by T4 DNA polymerase. Typical T4 DNA polymerase removes 3′-overhangs and fills in 5′-overhangs. Shallow sequencing was conducted following the Illumina single-end standard protocol on GAIIx using a 75-bp sequencing run. Deep sequencing was conducted following the Illumina pair-end standard protocol on GAIIx using a 100-bp sequencing run.

Contamination control

DNA contamination from the laboratory environment and reagents was controlled through setting up negative controls during DNA extraction, genomic libraries preparation, and amplification in parallel with the samples at all experimental steps. DNA concentrations in negative controls were undetectable with Qubit dsDNA HS Assay (Thermo Fisher Scientific) and Bioanalyzer 2100 HS DNA Assay (Agilent), implying concentrations below 0.01 ng/uL. Concentrations of the libraries built from ancient human teeth were between 1.1 and 125.5 ng/uL (on average, 18.76 ng/uL). The amount of DNA in negative controls was at least 100-fold lower than for ancient samples and was not subjected to the sequencing.

Bioinformatics procedures

All reads were trimmed, and adapters were removed using the AdapterRemoval tool (AdapterRemoval, RRID:SCR_011834) [74]. The minimal length of reads was set to 25, and the minimal base quality was set to 30.

To investigate the composition of microbial communities in each sample, we used the MetaPhlAn2 program with default settings (MetaPhlAn, RRID:SCR_004915) [46]. To avoid bias in the assessment of microorganism abundance, we mapped (using Bowtie2 [Bowtie2, RRID:SCR_005476] [75] and the recommended sensitive global alignment strategy) all reads against

the MetaPhlAn2 markers database and removed PCR duplicates with Picard MarkDuplicates tool 1.82 (Picard, RRID:SCR_006525). Next, we ran MetaPhlAn2 with the option "-a" to determine all taxonomic levels.

To assess the amount of endogenous DNA, reads were mapped against human nuclear (hg19) [76] and complete mitochondrial genomes (GenBank Accession no. NC 012920.1) [77].

To investigate aDNA damage patterns, we employed mapDamage2.0 with the default settings (mapDamage, RRID:SCR_001240) [59]. All plots were generated using R 3.3.2 ggplot2 package (ggplot2, RRID:SCR_014601).

Statistical analysis

Shannon diversity, PCA, and PCoA on 4 taxonomic levels (class, genus, family, species) were run in R (functions: diversity(), prcomp(), and pcoa(), respectively) for all identified microorganisms and for bacteria/archaea only. PCoA was run on the Jaccard, and Bray-Curtis distance tables were calculated from the taxon abundance. To determine whether low-abundance taxa (<1%) may have influenced the analysis, we also ran PCoA without them (data not shown). To determine if *k*-mers of exogenous reads might segregate samples according to their age, storage, or archeological site, we followed the approach described in Dubinkina et al. [78].

To test if certain groups displayed statistically significant differences, we applied a 1-way ANOVA, followed by a Tukey HSD and a t-test (R functions: aov(), TukeyHSD(), t.test()), as well as the following non-parametric tests: Kruskal-Wallis and Wilcoxon (R functions: kruskal.test(), wilcox.test()).

Correlation *R* was calculated as a Pearson correlation coefficient.

Additional files

Supplementary Table S1. Summarized information on NGS datasets used within this study. The first column from the left lists sample IDs. Column 2 comprises information on C14 dating of selected samples. Column 3 and column 4 describe the depth of sequencing (number of raw and filtered reads). Columns 5 and 6 describe the reads that map to the human genome (number, percentage). Columns 7 and 8 describe reads mapping to the Metaphlan2 markers DB (number, percentage). Column 9 describes the number of reads that mapped to the prokaryotic markers only. Columns 10–16 describe the percentage (within a sample) of viruses/viroids, eukaryote, all prokaryote, environmental prokaryote, oral prokaryote, other human-related prokaryote, and potential pathogens, respectively. Columns 17–22 describe the number of identified bacterial/archaeal taxa and Shannon index on class, family, and species level, respectively.

Supplementary Table S2. Summarized information on bacterial/archaeal taxa (column 1) identified within samples (columns 5–165). Column 2–4 describe taxon gram stain type, respiratory type, and its typical habitat, respectively.

Supplementary Table S3. The information on 11 samples used for the validation of results obtained in a shallow sequencing experiment. The first column from the left lists sample IDs. Column 2 describes the total number of filtered reads. Columns 3 and 4 describe the reads that mapped to the Metaphlan2 markers DB (number, percentage). Column 5 describes the percentage of prokaryote identified in a sample. Columns 6 and 7 describe the number of bacterial/archaeal classes and the Shannon index.

Supplementary Table S4. Summarized information on 77 bacterial and archaeal species (column 1) selected for aDNA damage analysis. Columns 2–4 describe species gram stain type, respiratory type, and its habitat, respectively. Columns 5–12 describe the number of samples in which the species were present in more % than the threshold (80%, 70%, 60%, 50%, 40%, 30%, 20%, 10%, 1%, respectively). Column 13 describes the maximal percentage of a species observed. Column 14 describes the overall percentage of a species in all samples. Columns 15–154 describe the species percentage in an individual sample. A) Table summarizes the number of samples with species present in more than the threshold and their percentage with respect to the all the identified species.

aDNA_microorganisms_Figlerowicz_Supplementary_Figures.pdf

aDNA_microorganisms_Figlerowicz_Supplementary_Tables.xlsx

aDNA_microorganisms_Figlerowicz_Supplementary_Tables_leg.docx

Abbreviations

aDNA ancient DNA
dsDNA double-stranded DNA
NGS next-generation sequencing.

Acknowledgment

We thank Wioletta Nowaczewska for providing samples from Masłomęcz.

Funding

This work was supported by polish National Science Center (2014/12/W/NZ2/00466).

Availability of supporting data and materials

Other data further supporting this work can be found in the *GigaScience* repository, GigaDB [79]. The datasets supporting the conclusions of this article are available in the National Center for Biotechnology Information Sequence Read Archive repository, SRP093814 [80].

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

A.P. conceived the study, participated in the study design, analyzed the data, discussed the results, and wrote the manuscript; I.S. participated in the statistical analysis and figure preparation and submitted the datasets to the Sequence Read Archive; B.K. ran preliminary Metaphlan2 analysis; A.J. extracted DNA and participated in NGS library preparation; L.H. prepared NGS libraries and ran NGS; J.P. participated in results and discussion; P.K. participated in the study design, analyzed and discussed the data, and participated in drafting the manuscript; M.F. conceived the overall idea of the study, participated in the study design, analyzed and discussed the data, coordinated studies, and was responsible for the final version of the manuscript; all authors read and approved the final manuscript.

References

1. Fu Q, Posth C, Hajdinjak M et al. The genetic history of Ice Age Europe. Nature 2016;**534**(7606):200–5.
2. Green RE, Krause J, Briggs AW et al. A draft sequence of the Neandertal genome. Science 2010;**328**(5979):710–22.
3. Rasmussen M, Li Y, Lindgreen S et al. Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature 2010;**463**(7282):757–62.
4. Meyer M, Kircher M, Gansauge MT et al. A high-coverage genome sequence from an archaic Denisovan individual. Science 2012;**338**(6104):222–6.
5. Librado P, Fages A, Gaunitz C et al. The Evolutionary Origin and Genetic Makeup of Domestic Horses. Genetics 2016;**204**(2):423–34.
6. Malmstrom H, Stora J, Dalen L et al. Extensive human DNA contamination in extracts from ancient dog bones and teeth. Mol Biol Evol 2005;**22**(10):2040–7.
7. Salamon M, Tuross N, Arensburg B et al. Relatively well preserved DNA is present in the crystal aggregates of fossil bones. Proc Natl Acad Sci U S A 2005;**102**(39):13783–8.
8. Raghavan M, Skoglund P, Graf KE et al. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature 2014;**505**(7481):87–91.
9. Der Sarkissian C, Ermini L, Jonsson H et al. Shotgun microbial profiling of fossil remains. Mol Ecol 2014;**23**(7): 1780–98.
10. Reich D, Green RE, Kircher M et al. Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 2010;**468**(7327):1053–60.
11. Ovchinnikov IV, Gotherstrom A, Romanova GP et al. Molecular analysis of Neanderthal DNA from the northern Caucasus. Nature 2000;**404**(6777):490–3.
12. Lawlor DA, Dickel CD, Hauswirth WW et al. Ancient HLA genes from 7,500-year-old archaeological remains. Nature 1991;**349**(6312):785–8.
13. Smith CI, Chamberlain AT, Riley MS et al. Neanderthal DNA. Not just old but old and cold? Nature 2001;**410**(6830):771–2.
14. Schwarz C, Debruyne R, Kuch M et al. New insights from old bones: DNA preservation and degradation in permafrost preserved mammoth remains. Nucleic Acids Res 2009;**37**(10):3215–29.
15. Poinar HN, Schwarz C, Qi J et al. Metagenomics to paleogenomics: large-scale sequencing of mammoth DNA. Science 2006;**311**(5759):392–4.
16. Noonan JP, Hofreiter M, Smith D et al. Genomic sequencing of Pleistocene cave bears. Science 2005;**309**(5734):597–9.
17. Sampietro ML, Gilbert MT, Lao O et al. Tracking down human contamination in ancient human teeth. Mol Biol Evol 2006;**23**(9):1801–7.
18. Jans MME, Nielsen-Marsh CM, Smith CI et al. Characterisation of microbial attack on archaeological bone. J Archaeol Sci 2004;**31**(1):87–95.
19. Haile J, Holdaway R, Oliver K et al. Ancient DNA chronology within sediment deposits: are paleobiological reconstructions possible and is DNA leaching a factor? Mol Biol Evol 2007;**24**(4):982–9.
20. Carpenter ML, Buenrostro JD, Valdiosera C et al. Pulling out the 1%: whole-genome capture for the targeted enrichment of ancient DNA sequencing libraries. Am J Hum Genet 2013;**93**(5):852–64.
21. Schuenemann VJ, Singh P, Mendum TA et al. Genome-wide comparison of medieval and modern *Mycobacterium leprae*. Science 2013;**341**(6142):179–83.
22. Gansauge MT, Meyer M. Selective enrichment of damaged DNA molecules for ancient genome sequencing. Genome Res 2014;**24**(9):1543–9.
23. Avila-Arcos MC, Cappellini E, Romero-Navarro JA et al. Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. Sci Rep 2011;**1**:74.
24. Cruz-Davalos DI, Llamas B, Gaunitz C et al. Experimental conditions improving in-solution target enrichment for ancient DNA. Mol Ecol Resour 2017;**17**(3):508–22.
25. Orlando L, Ginolhac A, Raghavan M et al. True single-molecule DNA sequencing of a pleistocene horse bone. Genome Res 2011;**21**(10):1705–19.
26. Ginolhac A, Vilstrup J, Stenderup J et al. Improving the performance of true single molecule sequencing for ancient DNA. BMC Genomics 2012;**13**:177.
27. Fierer N, Leff JW, Adams BJ et al. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proc Natl Acad Sci U S A 2012;**109**(52):21390–5.
28. Ding T, Schloss PD. Dynamics and associations of microbial community types across the human body. Nature 2014;**509**(7500):357–60.
29. Wade WG. The oral microbiome in health and disease. Pharmacol Res 2013;**69**(1):137–43.
30. Xu X, He J, Xue J et al. Oral cavity contains distinct niches with dynamic microbial communities. Environ Microbiol 2015;**17**(3):699–710.
31. Ferretti P, Farina S, Cristofolini M et al. Experimental metagenomics and ribosomal profiling of the human skin microbiome. Exp Dermatol 2017;**26**(3):211–19.
32. O'Toole PW, Jeffery IB. Gut microbiota and aging. Science 2015;**350**(6265):1214–5.
33. Zhernakova A, Kurilshikov A, Bonder MJ et al. Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. Science 2016;**352**(6285):565–9.
34. Falony G, Joossens M, Vieira-Silva S et al. Population-level analysis of gut microbiome variation. Science 2016;**352**(6285):560–4.
35. Donaldson GP, Lee SM, Mazmanian SK. Gut biogeography of the bacterial microbiota. Nat Rev Microbiol 2016;**14**(1):20–32.
36. Rasmussen S, Allentoft ME, Nielsen K et al. Early divergent strains of *Yersinia pestis* in Eurasia 5,000 years ago. Cell 2015;**163**(3):571–82.
37. Maixner F, Krause-Kyora B, Turaev D et al. The 5300-year-old *Helicobacter pylori* genome of the Iceman. Science 2016;**351**(6269):162–5.
38. Seifert L, Wiechmann I, Harbeck M et al. Genotyping *Yersinia pestis* in historical plague: evidence for long-term persistence of *Y. pestis* in Europe from the 14th to the 17th century. PLoS One 2016;**11**(1):e0145194.
39. Rollo F, Ermini L, Luciani S et al. Studies on the preservation of the intestinal microbiota's DNA in human mummies from cold environments. Med Secoli 2006;**18**(3):725–40.
40. Ubaldi M, Luciani S, Marota I et al. Sequence analysis of bacterial DNA in the colon of an Andean mummy. Am J Phys Anthropol 1998;**107**(3):285–95.
41. Weyrich LS, Dobney K, Cooper A. Ancient DNA analysis of dental calculus. J Hum Evol 2015;**79**:119–24.
42. Warinner C, Speller C, Collins MJ et al. Ancient human microbiomes. J Hum Evol 2015;**79**:125–36.
43. Warinner C, Rodrigues JF, Vyas R et al. Pathogens and host immunity in the ancient human oral cavity. Nat Genet 2014;**46**(4):336–44.

44. Adler CJ, Dobney K, Weyrich LS et al. Sequencing ancient calcified dental plaque shows changes in oral microbiota with dietary shifts of the Neolithic and Industrial revolutions. Nat Genet 2013;**45**(4):450–5.
45. Segata N, Izard J, Waldron L et al. Metagenomic biomarker discovery and explanation. Genome Biol 2011;**12**(6):R60.
46. Truong DT, Franzosa EA, Tickle TL et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods 2015;**12**(10):902–3.
47. Pride DT, Salzman J, Haynes M et al. Evidence of a robust resident bacteriophage population revealed through analysis of the human salivary virome. ISME J 2012;**6**(5):915–26.
48. Willner D, Furlan M, Schmieder R et al. Metagenomic detection of phage-encoded platelet-binding factors in the human oral cavity. Proc Natl Acad Sci U S A 2011;**108(suppl 1)**:4547–53.
49. Metcalf JL, Xu ZZ, Weiss S et al. Microbial community assembly and metabolic function during mammalian corpse decomposition. Science 2016;**351**(6269):158–62.
50. Tsuzukibashi O, Uchibori S, Shinozaki-Kuwahara N et al. A selective medium for the isolation of *Corynebacterium* species in oral cavities. J Microbiol Methods 2014;**104**:67–71.
51. Colombo AP, Boches SK, Cotton SL et al. Comparisons of subgingival microbial profiles of refractory periodontitis, severe periodontitis, and periodontal health using the human oral microbe identification microarray. J Periodontol 2009;**80**(9):1421–32.
52. Shaddox LM, Huang H, Lin T et al. Microbiological characterization in children with aggressive periodontitis. J Dent Res 2012;**91**(10):927–33.
53. Antunes HS, Rocas IN, Alves FR et al. Total and specific bacterial levels in the apical root canal system of teeth with post-treatment apical periodontitis. J Endod 2015;**41**(7):1037–42.
54. Javed S, Said F, Eqani SA et al. *Bordetella parapertussis* outbreak in Bisham, Pakistan in 2009–2010: fallout of the 9/11 syndrome. Epidemiol Infect 2015;**143**(12):2619–23.
55. Vidor C, Awad M, Lyras D. Antibiotic resistance, virulence factors and genetics of *Clostridium sordellii*. Res Microbiol 2015;**166**(4):368–74.
56. Hanif H, Anjum A, Ali N et al. Isolation and antibiogram of *Clostridium tetani* from clinically diagnosed tetanus patients. Am J Trop Med Hyg 2015;**93**(4):752–6.
57. Endersen L, Coffey A, Ross RP et al. Characterisation of the antibacterial properties of a bacterial derived peptidoglycan hydrolase (LysCs4), active against *C. sakazakii* and other Gram-negative food-related pathogens. Int J Food Microbiol 2015;**215**:79–85.
58. Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature 2012;**486**(7402):207–14.
59. Jonsson H, Ginolhac A, Schubert M et al. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 2013;**29**(13):1682–4.
60. Briggs AW, Stenzel U, Johnson PL et al. Patterns of damage in genomic DNA sequences from a Neandertal. Proc Natl Acad Sci U S A 2007;**104**(37):14616–21.
61. Sawyer S, Krause J, Guschanski K et al. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PLoS One 2012;**7**(3):e34131.
62. Willerslev E, Hansen AJ, Ronn R et al. Long-term persistence of bacterial DNA. Curr Biol 2004;**14**(1):R9–10.
63. Schubert M, Ermini L, Der Sarkissian C et al. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. Nat Protoc 2014;**9**(5):1056–82.
64. DeSantis TZ, Hugenholtz P, Larsen N et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 2006;**72**(7):5069–72.
65. Wang Q, Garrity GM, Tiedje JM et al. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 2007;**73**(16):5261–7.
66. Setlow P. Mechanisms for the prevention of damage to DNA in spores of *Bacillus* species. Annu Rev Microbiol 1995;**49**:29–54.
67. Bouwman AS, Kennedy SL, Muller R et al. Genotype of a historic strain of *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A 2012;**109**(45):18511–6.
68. Weyrich LS, Rolin OY, Muse SJ et al. A type VI secretion system encoding locus is required for *Bordetella bronchiseptica* immunomodulation and persistence in vivo. PLoS One 2012;**7**(10):e45892.
69. Bendor L, Weyrich LS, Linz B et al. Type six secretion system of *Bordetella bronchiseptica* and adaptive immune components limit intracellular survival during infection. PLoS One 2015;**10**(10):e0140743.
70. Juras A, Chylenski M, Krenz-Niedbala M et al. Investigating kinship of Neolithic post-LBK human remains from Krusza Zamkowa, Poland using ancient DNA. Forensic Sci Int Genet 2017;**26**:30–39.
71. Yang DY, Eng B, Waye JS et al. Technical note: improved DNA extraction from ancient bones using silica-based spin columns. Am J Phys Anthropol 1998;**105**(4):539–43.
72. Malmstrom H, Svensson EM, Gilbert MT et al. More on contamination: the use of asymmetric molecular behavior to identify authentic ancient human DNA. Mol Biol Evol 2007;**24**(4):998–1004.
73. Meyer M, Kircher M. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor Protoc 2010;**2010**(6): pdb prot5448.
74. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes 2016;**9**:88.
75. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;**9**(4):357–9.
76. Meyer LR, Zweig AS, Hinrichs AS et al. The UCSC Genome Browser database: extensions and updates 2013. Nucleic Acids Res 2013;**41**(database issue):D64–69.
77. Andrews RM, Kubacka I, Chinnery PF et al. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat Genet 1999;**23**(2):147.
78. Dubinkina VB, Ischenko DS, Ulyantsev VI et al. Assessment of k-mer spectrum applicability for metagenomic dissimilarity analysis. BMC Bioinformatics 2016;**17**:38.
79. Philips A, Stolarek I, Kuczkowska B, *et al.* Supporting data for "Comprehensive analysis of microorganisms accompanying human archaeological remains." Gigascience Database 2017. http://dx.doi.org/10.5524/100310.
80. Sequence Read Archive. https://www.ncbi.nlm.nih.gov/sra/?term=SRP093814. Accessed 24 Nov 2016.

MATERIAŁY UZUPEŁNIAJĄCE DO PUBLIKACJI

Philips i wsp., GigaScience 2017

Spis tabel uzupełniających

Supplementary Table 1. Summarized information on NGS datasets used within this study. The first column form the left lists samples IDs. Column 2 comprises information on C14 dating of selected samples. Column 3 and column 4 describe the depth of sequencing (number of raw and filtered reads). Columns 5 and 6 describe the reads that map to the human genome (number, percentage). Columns 7 and 8 describe reads mapping to the Metaphlan2 markers DB (number, percentage). Column 9 describes the number of reads that mapped to the prokaryotic markers only. Columns 10-16 describe the percentage (within a sample) of viruses/viroids, eukaryote, all prokaryote, environmental prokaryote, oral prokaryote, other human-related prokaryote and potential pathogens, respectively. Columns 17-22 describe the number of identified bacterial/archaeal taxa and Shannon index on class, family and species level respectively

Supplementary Table 2. Summarized information on bacterial/archaeal taxa (Column 1) identified within samples (Columns 5-165). Column 2-4 describe taxon gram stain type, respiratory type and its typical habitat, respectively

Supplementary Table 3. The information on 11 samples used for the validation of results obtained in shallow sequencing experiment. The first column form the left lists samples IDs. Column 2 describes the total number of filtered reads. Column 3 and 4 describe the reads that mapped to the Metaphlan2 markers DB (number, percentage). Column 5 describes the percentage of prokaryote identified in a sample. Column 6 and 7 describe the number of bacterial/archaeal classes and Shannon index

Supplementary Table 4. Summarized information on 77 bacterial and archaeal species (Column 1) selected for aDNA damage analysis. Columns 2-4 describe species gram stain type, respiratory type and its habitat, respectively. Columns 5-12 describe the number of samples in which the species were present in more % than the threshold (80%, 70%, 60%, 50%, 40%, 30%, 20%, 10%, 1% respectively). Column 13 describes the maximal percentage of a species observed. Column 14 describes the overall

percentage of a species in all samples. Columns 15-154 describe the species percentage in an individual samples. A) Table summarizes the number of samples with species present in more than the threshold and their percentage in respect to the all of identified species

Spis rysunków uzupełniających

Supplementary Figure 1. Bacterial/archaeal respiratory types detected in analyzed archaeological samples. A) Pieplot representing overall frequency bacterial/archaeal respiratory types in archaeological samples; B) box and whiskers plot representing the distribution of frequencies of particular bacterial/archaeal respiratory type in archaeological sites (GO not shown as includes only 2 samples); and C) stacked barplot indicating the frequency of bacterial/archaeal respiratory types in a particular sample. Each bar represents an individual sample. Samples are ordered by the archeological sites. The color legend for all plots is shown at the bottom

Supplementary Figure 2. Bacterial/archaeal gram stain types detected in analyzed archaeological samples. A) Pieplot representing overall frequency bacterial/archaeal gram stain types in archaeological samples; B) box and whiskers plot representing the distribution of frequencies of particular bacterial/archaeal gram stain type in archaeological sites (GO not shown as includes only 2 samples); and C) stacked barplot indicating the frequency of bacterial/archaeal gram stain types in a particular sample. Each bar represents an individual sample. Samples are ordered by the archeological sites. The color legend for all plots is shown at the bottom

Supplementary Figure 3. Principal component analysis (PCA) of microbial compositions at genus taxonomic level. Left-hand side: Principal components 1 and 2. Right-hand side: Principal components 1 and 3. Samples from certain archaeological sites are marked in different colors and labelled with archaeological site ID

Supplementary Figure 4. Principal component analysis (PCA) of microbial compositions at genus taxonomic level: A) Human-related genera B) Environmental genera. Left-hand side: Principal components 1 and 2. Right-hand side: Principal components 1 and 3. Samples from certain archaeological sites are marked in different colors and labelled with archaeological site ID

Supplementary Figure 5. Hierarchical clustering of samples (Manhattan distances, cluster method: average) at the genus taxonomic level A) Human-related genera B) Environmental genera. Samples from certain archaeological sites are marked in different colors and labeled with archaeological site ID

Supplementary Figure 6. Principal component analysis (PCA) of exogenous 10-mer reads. Left-hand side: Principal components 1 and 2. Right-hand side: Principal components 1 and 3

Supplementary Figure 7. Principal coordinate analysis (PCoA) of microbial compositions at genus taxonomic level in various human associated microbiomes (mouth – brown squares, nose – red squares, skin – green squares, vagina – purple squares, stool – blue squares), soils (light green circles) and ancient human remains (black labels). Left-hand side: Principal components 1 and 2. Right-hand side: Principal components 1 and 3

Supplementary Figure 8. Comparison of bacterial and archaeal profiles (stacked barplot) on A) class B) family C) genus D) species levels based on shallow and deep sequencing of the selected 11 samples (Sample ID is indicated on the x axis). The correlation coefficient R is placed above each shallow/deep stacked bar pair. The bar colors represent different classes, families, genera or species, respectively. * $0.01>p>0.001$; ** $p=0.09360$; nothing: $p<0.0001$

Supplementary Figure 9. The differences of DNA damage levels (δs) of human and bacterial/archaeal DNA in individual samples. Human DNA is in black, All bacterial/archaeal DNA is in grey and bacterial/archaeal DNA belonging to the environmental group is in green, all human-related in orange, oral in blue, other in yellow and pathogen in red

Supplementary Figure 10. The differences of DNA damage levels (λ , expressed as: $1/\lambda$-1) of human and bacterial/archaeal DNA in individual samples. Human DNA is in black, All bacterial/archaeal DNA is in grey and bacterial/archaeal DNA belonging to

the environmental group is in green, all human-related in orange, oral in blue, other in yellow and pathogen in red

OŚWIADCZENIA WSPÓŁAUTORÓW

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327 NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 26.05.2019

dr Luiza Handschuh

Pracownia Genomiki

Instytut Chemii Bioorganicznej PAN

Ul. Z. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

A mosaic genetic structure of the human population living in the South Baltic region during the Iron Age

Ireneusz Stolarek, Anna Juras, Luiza Handschuh, Malgorzata Marcinkowska-Swojak, Anna Philips, Michal Zenczak, Artur Dębski, Hanna Kóčka-Krenz, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Scientific Reports volume 8, Article number: 2455 (2018)

doi: 10.1038/s41598-018-20705-6

Oświadczam, iż w trakcie powstawania powyższej publikacji moją rolą było wykonanie sekwencjonowania NGS, praca nad rysunkami, dyskusja uzyskanych wyników oraz pisanie manuskryptu.

dr Luiza Handschuh

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327 NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 26.05.2019

dr Luiza Handschuh

Pracownia Genomiki

Instytut Chemii Bioorganicznej PAN

Ul. Z. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Goth migration induced changes in the matrilineal genetic structure of the central-east European population

Ireneusz Stolarek, Luiza Handschuh, Anna Juras, Wioletta Nowaczewska, Hanna Kóčka-Krenz, Andrzej Michalowski, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Scientific Reports volume 9, Article number: 6737 (2019)

doi: 10.1038/s41598-019-43183-w

Oświadczam, iż w trakcie powstawania powyższej publikacji moją rolą było przygotowanie bibliotek NGS, sekwencjonowanie NGS, dyskusja nad wynikami oraz pisanie manuskryptu.

dr Luiza Handschuh

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327 NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 26.05.2019

dr Luiza Handschuh

Pracownia Genomiki

Instytut Chemii Bioorganicznej PAN

Ul. Z. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Comprehensive analysis of microorganisms accompanying human archaeological remains

Anna Philips, Ireneusz Stolarek, Bogna Kuczkowska, Anna Juras, Luiza Handschuh, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Gigascience. 2017 Jul; 6(7): 1–13

doi: 10.1093/gigascience/gix044

Oświadczam, iż w trakcie powstawania powyższej publikacji moją rolą było przygotowanie bibliotek NGS oraz sekwencjonowanie NGS.

dr Luiza Handschuh

Poznań, 26.05.2019

Dr inż. Anna Juras
Zakład Biologii Ewolucyjnej Człowieka
Instytut Antropologii
Wydział Biologii
Uniwersytet im. A. Mickiewicza w Poznaniu

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

A mosaic genetic structure of the human population living in the South Baltic region during the Iron Age

Ireneusz Stolarek, Anna Juras, Luiza Handschuh, Malgorzata Marcinkowska-Swojak, Anna Philips, Michal Zenczak, Artur Dębski, Hanna Kóčka-Krenz, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Scientific Reports volume 8, Article number: 2455 (2018)

doi: 10.1038/s41598-018-20705-6

Oświadczam iż w trakcie powstawania powyższej publikacji moją rolą była ekstrakcja DNA, przygotowanie bibliotek NGS oraz krytyczne czytanie manuskryptu.

Anna Juras

Dr inż. Anna Juras

Poznań, 26.05.2019

Dr inż. Anna Juras
Zakład Biologii Ewolucyjnej Człowieka
Instytut Antropologii
Wydział Biologii
Uniwersytet im. A. Mickiewicza w Poznaniu

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Goth migration induced changes in the matrilineal genetic structure of the central-east European population

Ireneusz Stolarek, Luiza Handschuh, Anna Juras, Wioletta Nowaczewska, Hanna Kóčka-Krenz, Andrzej Michalowski, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Scientific Reports volume 9, Article number: 6737 (2019)

doi: 10.1038/s41598-019-43183-w

Oświadczam iż w trakcie powstawania powyższej publikacji moją rolą była ekstrakcja DNA, przygotowanie bibliotek NGS oraz krytyczne czytanie manuskryptu.

Dr inż. Anna Juras

Poznań, 26.05.2019

Dr inż. Anna Juras
Zakład Biologii Ewolucyjnej Człowieka
Instytut Antropologii
Wydział Biologii
Uniwersytet im. A. Mickiewicza w Poznaniu

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Comprehensive analysis of microorganisms accompanying human archaeological remains

Anna Philips, Ireneusz Stolarek, Bogna Kuczkowska, Anna Juras, Luiza Handschuh, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Gigascience. 2017 Jul; 6(7): 1–13

doi: 10.1093/gigascience/gix044

Oświadczam iż w trakcie powstawania powyższej publikacji moją rolą była ekstrakcja DNA oraz uczestniczenie w przygotowaniu bibliotek NGS.

Dr inż. Anna Juras

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327, NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 26.05.2019

dr Anna Philips

Zakład Biologii Molekularnej i Systemowej

Instytut Chemii Bioorganicznej PAN

Ul. Z. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Comprehensive analysis of microorganisms accompanying human archaeological remains

Anna Philips, Ireneusz Stolarek, Bogna Kuczkowska, Anna Juras, Luiza Handschuh, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Gigascience. 2017 Jul; 6(7): 1–13

doi: 10.1093/gigascience/gix044

Oświadczam iż w trakcie powstawania powyższej publikacji moją rolą było stworzenie ogólnej idei pracy, uczestnictwo w planowaniu prac badawczych, analiza danych, dyskusja nad wynikami oraz pisanie manuskryptu.

A. Philips

dr Anna Philips

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327, NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 26.05.2019

dr Anna Philips

Zakład Biologii Molekularnej i Systemowej

Instytut Chemii Bioorganicznej PAN

Ul. Z. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

A mosaic genetic structure of the human population living in the South Baltic region during the Iron Age

Ireneusz Stolarek, Anna Juras, Luiza Handschuh, Malgorzata Marcinkowska-Swojak, Anna Philips, Michal Zenczak, Artur Dębski, Hanna Kóčka-Krenz, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Scientific Reports volume 8, Article number: 2455 (2018)

doi: 10.1038/s41598-018-20705-6

Oświadczam iż w trakcie powstawania powyższej publikacji moją rolą była dyskusja nad wynikami oraz krytyczne czytanie manuskryptu.

A. Philips

dr Anna Philips

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327 NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 26.05.2019

dr Małgorzata Marcinkowska-Swojak

Zakład Biologii Molekularnej i Systemowej

Instytut Chemii Bioorganicznej PAN

Ul. Z. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

A mosaic genetic structure of the human population living in the South Baltic region during the Iron Age

Ireneusz Stolarek, Anna Juras, Luiza Handschuh, Malgorzata Marcinkowska-Swojak, Anna Philips, Michal Zenczak, Artur Dębski, Hanna Kóčka-Krenz, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Scientific Reports volume 8, Article number: 2455 (2018)

doi: 10.1038/s41598-018-20705-6

Oświadczam, iż w trakcie powstawania powyższej publikacji moją rolą była dyskusja nad wynikami oraz krytyczne czytanie manuskryptu.

dr Małgorzata Marcinkowska-Swojak

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327 NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 26.05.2019

mgr inż. Michał Zeńczak

Zakład Biologii Molekularnej i Systemowej

Instytut Chemii Bioorganicznej PAN

Ul. Z. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

A mosaic genetic structure of the human population living in the South Baltic region during the Iron Age

Ireneusz Stolarek, Anna Juras, Luiza Handschuh, Malgorzata Marcinkowska-Swojak, Anna Philips, Michal Zenczak, Artur Dębski, Hanna Kóčka-Krenz, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Scientific Reports volume 8, Article number: 2455 (2018)

doi: 10.1038/s41598-018-20705-6

Oświadczam iż w trakcie powstawania powyższej publikacji moją rolą była dyskusja nad wynikami oraz krytyczne czytanie manuskryptu.

mgr inż. Michał Zeńczak

Michał Zeńczak

Poznań, 26.05.2019

mgr Bogna Kuczkowska

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Comprehensive analysis of microorganisms accompanying human archaeological remains

Anna Philips, Ireneusz Stolarek, Bogna Kuczkowska, Anna Juras, Luiza Handschuh, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Gigascience. 2017 Jul; 6(7): 1–13

doi: 10.1093/gigascience/gix044

Oświadczam iż w trakcie powstawania powyższej publikacji moją rolą było przeprowadzenie wstępnej analizy za pomocą narzędzia Metaphlan2.

Bogna Kuczkowska

mgr Bogna Kuczkowska

Poznań, 26.05.2019

Artur Dębski

Instytut Archeologii

Uniwersytetu im. Adama Mickiewicza

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

A mosaic genetic structure of the human population living in the South Baltic region during the Iron Age

Ireneusz Stolarek, Anna Juras, Luiza Handschuh, Malgorzata Marcinkowska-Swojak, Anna Philips, Michal Zenczak, Artur Dębski, Hanna Kóčka-Krenz, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Scientific Reports volume 8, Article number: 2455 (2018)

doi: 10.1038/s41598-018-20705-6

Oświadczam iż w trakcie powstawania powyższej publikacji moją rolą było dostarczenie danych archeologicznych oraz dyskusja nad wynikami.

Artur Dębski

A. Dębski

Poznań, 26.05.2019

prof. dr hab. Hanna Kóčka-Krenz
Instytut Archeologii
Uniwersytetu im. Adama Mickiewicza

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

A mosaic genetic structure of the human population living in the South Baltic region during the Iron Age

Ireneusz Stolarek, Anna Juras, Luiza Handschuh, Malgorzata Marcinkowska-Swojak, Anna Philips, Michal Zenczak, Artur Dębski, Hanna Kóčka-Krenz, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Scientific Reports volume 8, Article number: 2455 (2018)

doi: 10.1038/s41598-018-20705-6

Oświadczam, iż w trakcie powstawania powyższej publikacji moją rolą było dostarczenie danych archeologicznych oraz dyskusja nad wynikami.

prof. dr hab. Hanna Kóčka-Krenz

Poznań, 26.05.2019

prof. dr hab. Hanna Kóčka-Krenz

Instytut Archeologii
Uniwersytetu im. Adama Mickiewicza

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Goth migration induced changes in the matrilineal genetic structure of the central-east European population

Ireneusz Stolarek, Luiza Handschuh, Anna Juras, Wioletta Nowaczewska, Hanna Kóčka-Krenz, Andrzej Michalowski, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Scientific Reports volume 9, Article number: 6737 (2019)

doi: 10.1038/s41598-019-43183-w

Oświadczam iż w trakcie powstawania powyższej publikacji moją rolą było dostarczenie narracji archeologicznej oraz dyskusja nad wynikami.

prof. dr hab. Hanna Kóčka-Krenz

Poznań, 26.05.2019

dr hab. Andrzej Michałowski, prof. UAM

Instytut Archeologii UAM

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Goth migration induced changes in the matrilineal genetic structure of the central-east European population

Ireneusz Stolarek, Luiza Handschuh, Anna Juras, Wioletta Nowaczewska, Hanna Kóčka-Krenz, Andrzej Michalowski, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Scientific Reports volume 9, Article number: 6737 (2019)

doi: 10.1038/s41598-019-43183-w

Oświadczam iż w trakcie powstawania powyższej publikacji moją rolą było dostarczenie narracji archeologicznych oraz dyskusja nad uzyskanymi wynikami.

/dr hab. Andrzej Michałowski, prof. UAM/

Poznań, 26.05.2019

dr Wioletta Nowaczewska

Katedra Biologii Człowieka UWr

Ul. Kuźnicza 35, 50-138

Wrocław

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Goth migration induced changes in the matrilineal genetic structure of the central-east European population

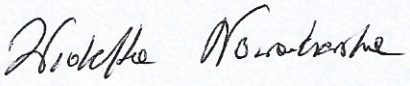
Ireneusz Stolarek, Luiza Handschuh, Anna Juras, Wioletta Nowaczewska, Hanna Kóčka-Krenz, Andrzej Michalowski, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Scientific Reports volume 9, Article number: 6737 (2019)

doi: 10.1038/s41598-019-43183-w

Oświadczam, iż w trakcie powstawania powyższej publikacji moją rolą było dostarczenie próbek oraz przygotowanie opracowania antropologicznego szczątków kostnych.

dr Wioletta Nowaczewska



INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327 NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 26.05.2019

prof. dr hab. Piotr Kozłowski

Zakład Genetyki Molekularnej

Instytut Chemii Bioorganicznej PAN

Ul. Z. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

A mosaic genetic structure of the human population living in the South Baltic region during the Iron Age

Ireneusz Stolarek, Anna Juras, Luiza Handschuh, Malgorzata Marcinkowska-Swojak, Anna Philips, Michal Zenczak, Artur Dębski, Hanna Kóčka-Krenz, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Scientific Reports volume 8, Article number: 2455 (2018)

doi: 10.1038/s41598-018-20705-6

Oświadczam iż w trakcie powstawania powyższej publikacji moją rolą było uczestnictwo w planowaniu prac badawczych, dyskusja nad wynikami oraz krytyczne czytanie manuskryptu.

prof. dr hab. Piotr Kozłowski

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327 NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 26.05.2019

prof. dr hab. Piotr Kozłowski

Zakład Genetyki Molekularnej

Instytut Chemii Bioorganicznej PAN

Ul. Z. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Goth migration induced changes in the matrilineal genetic structure of the central-east European population

Ireneusz Stolarek, Luiza Handschuh, Anna Juras, Wioletta Nowaczewska, Hanna Kóčka-Krenz, Andrzej Michalowski, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Scientific Reports volume 9, Article number: 6737 (2019)

doi: 10.1038/s41598-019-43183-w

Oświadczam iż w trakcie powstawania powyższej publikacji moją rolą było uczestnictwo w planowaniu prac badawczych, dyskusja nad wynikami oraz krytyczne czytanie manuskryptu.

prof. dr hab. Piotr Kozłowski

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327 NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 26.05.2019

prof. dr hab. Piotr Kozłowski

Zakład Genetyki Molekularnej

Instytut Chemii Bioorganicznej PAN

Ul. Z. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Comprehensive analysis of microorganisms accompanying human archaeological remains

Anna Philips, Ireneusz Stolarek, Bogna Kuczkowska, Anna Juras, Luiza Handschuh, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Gigascience. 2017 Jul; 6(7): 1–13

doi: 10.1093/gigascience/gix044

Oświadczam iż w trakcie powstawania powyższej publikacji moją rolą było uczestnictwo w planowaniu prac badawczych, analiza danych, dyskusja nad wynikami oraz pisanie manuskryptu.

prof. dr hab. Piotr Kozłowski

Poznań, 26.05.2019

prof. dr hab. Janusz Piontek

Zakład Biologii Ewolucyjnej Człowieka
Instytut Antropologii
Wydział Biologii
Uniwersytet im. A. Mickiewicza w Poznaniu

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Goth migration induced changes in the matrilineal genetic structure of the central-east European population

Ireneusz Stolarek, Luiza Handschuh, Anna Juras, Wioletta Nowaczewska, Hanna Kóčka-Krenz, Andrzej Michalowski, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Scientific Reports volume 9, Article number: 6737 (2019)

doi: 10.1038/s41598-019-43183-w

Oświadczam iż w trakcie powstawania powyższej publikacji moją rolą było dostarczenie próbek, dyskusja nad wynikami oraz krytyczne czytanie manuskryptu.

prof. dr hab. Janusz Piontek

Poznań, 26.05.2019

prof. dr hab. Janusz Piontek

Zakład Biologii Ewolucyjnej Człowieka
Instytut Antropologii
Wydział Biologii
Uniwersytet im. A. Mickiewicza w Poznaniu

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

A mosaic genetic structure of the human population living in the South Baltic region during the Iron Age

Ireneusz Stolarek, Anna Juras, Luiza Handschuh, Malgorzata Marcinkowska-Swojak, Anna Philips, Michal Zenczak, Artur Dębski, Hanna Kóčka-Krenz, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Scientific Reports volume 8, Article number: 2455 (2018)

doi: 10.1038/s41598-018-20705-6

Oświadczam iż w trakcie powstawania powyższej publikacji moją rolą było dostarczenie próbek, dyskusja nad wynikami oraz krytyczne czytanie manuskryptu.

prof. dr hab. Janusz Piontek

Poznań, 26.05.2019

prof. dr hab. Janusz Piontek

Zakład Biologii Ewolucyjnej Człowieka
Instytut Antropologii
Wydział Biologii
Uniwersytet im. A. Mickiewicza w Poznaniu

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Comprehensive analysis of microorganisms accompanying human archaeological remains

Anna Philips, Ireneusz Stolarek, Bogna Kuczkowska, Anna Juras, Luiza Handschuh, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Gigascience. 2017 Jul; 6(7): 1–13

doi: 10.1093/gigascience/gix044

Oświadczam iż w trakcie powstawania powyższej publikacji moją rolą było uczestnictwo w uzyskaniu wyników oraz dyskusja nad nimi.

prof. dr hab. Janusz Piontek

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327 NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 26.05.2019

prof. dr hab. Marek Figlerowicz

Zakład Biologii Molekularnej i Systemowej

Instytut Chemii Bioorganicznej PAN

Ul. Z. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Comprehensive analysis of microorganisms accompanying human archaeological remains

Anna Philips, Ireneusz Stolarek, Bogna Kuczkowska, Anna Juras, Luiza Handschuh, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Gigascience. 2017 Jul; 6(7): 1–13

doi: 10.1093/gigascience/gix044

Oświadczam iż w trakcie powstawania powyższej publikacji moją rolą było stworzenie ogólnej idei pracy, uczestnictwo w planowaniu prac badawczych, analiza danych i dyskusja nad wynikami, koordynowanie prac oraz odpowiedzialność za ostateczną wersję manuskryptu.

prof. dr hab. Marek Figlerowicz

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327 NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 26.05.2019

prof. dr hab. Marek Figlerowicz

Zakład Biologii Molekularnej i Systemowej

Instytut Chemii Bioorganicznej PAN

Ul. Z. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

Goth migration induced changes in the matrilineal genetic structure of the central-east European population

Ireneusz Stolarek, Luiza Handschuh, Anna Juras, Wioletta Nowaczewska, Hanna Kóčka-Krenz, Andrzej Michalowski, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Scientific Reports volume 9, Article number: 6737 (2019)

doi: 10.1038/s41598-019-43183-w

Oświadczam iż w trakcie powstawania powyższej publikacji moją rolą było stworzenie ogólnej idei pracy, zdobycie finansowania, uczestnictwo w planowaniu prac badawczych, dyskusja nad wynikami oraz odpowiedzialność za ostateczną wersję manuskryptu.

prof. dr hab. Marek Figlerowicz

INSTYTUT CHEMII BIOORGANICZNEJ
Polskiej Akademii Nauk

ul. Noskowskiego 12/14, 61-704 Poznań
tel.: centrala 61 852 85 03, sekretariat 61 852 89 19
faks: 61 852 05 32, e-mail: ibch@ibch.poznan.pl
REGON 000849327 NIP 777-00-02-062
http://www.ibch.poznan.pl

Poznań, 26.05.2019

prof. dr hab. Marek Figlerowicz

Zakład Biologii Molekularnej i Systemowej

Instytut Chemii Bioorganicznej PAN

Ul. Z. Noskowskiego 12/14

61-704 Poznań

OŚWIADCZENIE O WSPÓŁAUTORSTWIE

A mosaic genetic structure of the human population living in the South Baltic region during the Iron Age

Ireneusz Stolarek, Anna Juras, Luiza Handschuh, Malgorzata Marcinkowska-Swojak, Anna Philips, Michal Zenczak, Artur Dębski, Hanna Kóčka-Krenz, Janusz Piontek, Piotr Kozlowski, Marek Figlerowicz

Scientific Reports volume 8, Article number: 2455 (2018)

doi: 10.1038/s41598-018-20705-6

Oświadczam iż w trakcie powstawania powyższej publikacji moją rolą było stworzenie ogólnej idei pracy, zdobycie finansowania, uczestnictwo w planowaniu prac badawczych, dyskusja nad wynikami oraz odpowiedzialność za ostateczną wersję manuskryptu.

prof. dr hab. Marek Figlerowicz