

MIKROPRZESZŁOŚĆ

Badania specjalistyczne w archeologii



pod redakcją
Aldony Kurzawskiej i Iwony Sobkowiak-Tabaki



WYDZIAŁ
ARCHEOLOGII

MIKROPRZESZŁOŚĆ

Badania specjalistyczne w archeologii

pod redakcją

Aldony Kurzawskiej i Iwony Sobkowiak-Tabaki

Poznań 2021

Mikroprzeszłość
Badania specjalistyczne w archeologii

Recenzje:
dr hab. Maria Lityńska-Zajac, prof. IAE PAN
dr hab. Marek Nowak, prof. UJ

Redakcja:
Aldona Kurzawska
Iwona Sobkowiak-Tabaka

Opracowanie techniczne i skład komputerowy:
Bartłomiej Gruszka

Korekta językowa:
Agnieszka Gruszka

Projekt okładki i rycin poprzedzających rozdziały:
Przemysław Matejko

ISBN: 978-83-946591-8-9

<https://doi.org/10.14746/WA.2021.1.978-83-946591-8-9>

Monografia jest dostępna online w Repozytorium Uniwersytetu im A. Mickiewicza w Poznaniu
<https://repozytorium.amu.edu.pl/>

Wydział Archeologii
Uniwersytet im. Adama Mickiewicza w Poznaniu

Publikacja dofinansowana z Projektu Wydziału Archeologii nr DEC/19/WArch/2021

Copyright by Faculty of Archaeology Adam Mickiewicz University in Poznań and authors

Poznań 2021

Nakład:
200 egz.

SPIS TREŚCI

Przedmowa	5
Andrzej Michałowski	
Wprowadzenie	7
Aldona Kurzawska, Iwona Sobkowiak-Tabaka	
Palinologia	13
Piotr Kołaczek, Monika Karpińska-Kołaczek, Sambor Czerwiński, Katarzyna Marcisz, Mariusz Lamentowicz	
Archeobotanika	31
Magdalena Moskal-del Hoyo	
Dendroarcheologia	67
Henryk P. Dąbrowski	
Mikroskamieniałości okrzemkowe	89
Monika Rzodkiewicz	
Wioślarki	115
Izabela Zawiska	
Archeoentomologia	131
Marcin Kadej, Szymon Konwerski, Agata Hałuszko	
Archeomalakologia	155
Aldona Kurzawska	
Izotopy stabilne węgla ($\delta^{13}\text{C}$) i tlenu ($\delta^{18}\text{O}$) w archeomalakologii	181
Karina Apolinarska	
Archeozoologia	199
Jarosław Wilczyński	

Antropologia fizyczna	219
Dorota Lorkiewicz-Muszyńska, Julia Sobol, Wojciech Kociemba, Anna Hyrchała, Mariusz Glapiński	
Archeogenetyka	249
Maciej Chyleński	
Mikromorfologia	277
Karolina Leszczyńska, Michał Jankowiak	
Petroarcheologia	297
Piotr Gunia, Ewa Lisowska	
Surowce krzemionkowe – możliwości badań	315
Iwona Sobkowiak-Tabaka	
Traseologia	333
Katarzyna Pyżewicz	
Ceramika – badania petroarcheologiczne	353
Piotr Gunia, Marta Krueger, Ewa Lisowska	
Ceramika – badania osadów organicznych wnętrza naczyń	367
Marta Krueger	
Tekstylnia	387
Maria Cybulska, Anna Drązkowska	
Archeometalurgia	407
Marcin Biborski, Mateusz Biborski	
Mikroskopy stosowane w archeologii	431
Piotr Gunia, Ewa Lisowska, Aldona Kurzawska	
Ręczny spektrometr fluorescencji rentgenowskiej (XRF) w archeologii	443
Michał Krueger	
Wykaz autorów	451



Archeogenetyka

Maciej Chyleński

WPROWADZENIE

Z wszystkich dostępnych źródeł, z których uzyskujemy wiedzę na temat przeszłości człowieka, praca z kopalnym DNA jest jedną z najbardziej wymagających. Pomimo to kopalny DNA jest narzędziem, które może dostarczyć bezpośrednich danych na temat dawnych migracji, interakcji między populacjami pradziejowymi i struktury ich pokrewieństwa. Dlatego właśnie badania kopalnego DNA (określane mianem archeogenetyki, paleogenetyki lub archeologii biomolekularnej) są jedną z najszybciej rozwijających się gałęzi bioarcheologii. Dzięki szybkiemu rozwojowi technologii sekwencjonowania wysokoprzepustowego wzrasta ilość generowanych danych, które możemy wykorzystywać w badaniach DNA. W konsekwencji w ciągu ostatniej dekady przeszliśmy od badania fragmentów pojedynczych genomów mitochondrialnych do badania pełnych genomów jądrowych setek osobników. Stosowane metody, nadążając za tym rozwojem, szybko się zmieniają, co prowadzi do częstej dezaktualizacji informacji posiadanych przez część z potencjalnie zainteresowanych badaczy. Rozdział ten ma za zadanie przedstawić obecny stan wiedzy na temat badań kopalnego DNA oraz nakreślić potencjalne ścieżki rozwoju, choć niewykluczone, że sam też będzie wymagał regularnych aktualizacji.

DNA (kwas deoksyrybonukleinowy) występuje we wszystkich tkankach organizmów i w związku z tym jest wszechobecny tam, gdzie organizmy te występują (lub występowały). Jest obecne w samych organizmach, we wszystkich ich wydzielinach oraz – co dla nas istotne – w ich szczątkach. Jest przy tym cząsteczką unikatową dla każdego organizmu pozwalającą nam rozróżnić, w zależności od rozdzielczości danych, jakimi dysponujemy, poszczególne gatunki, populacje czy osobniki. Dane genomowe mogą pomóc w odpowiedzi na wiele pytań badawczych. Pytania te są bezpośrednio związane z właściwościami i charakterystyką kopalnego DNA, które z kolei wymuszają stosowanie specyficznej metodyki. Aby więc poznać zastosowania analiz kopalnego DNA, należy najpierw zrozumieć jego charakterystykę oraz metody badań, jakimi dysponujemy.

HISTORIA BADAŃ I CHARAKTERYSTYKA KOPALNEGO DNA

Historia badań kopalnego DNA liczy już sobie prawie 40 lat. Pierwsze próby izolacji DNA z eksponatów muzealnych podejmowane w 1984 r. (Higuchi i in. 1984). Ale aż do końca XX w. badania kopalnego DNA, stanowiły raczej ciekawostkę, jako że nie miały usystematyzowanej metodyki, zwłaszcza w kwestii potwierdzania autentyczności

otrzymywanych wyników. Skutkowało to niestety licznymi przypadkami, w których otrzymany wynik okazywał się skutkiem zanieczyszczenia badanego materiału współczesnym DNA. Dopiero w pierwszej dekadzie XXI w. wykrył się powoli zestaw zaleceń i reguł umożliwiających weryfikację autentyczności otrzymanego wyniku (Poinar i Cooper 2000). W tym samym czasie wzrastała też systematycznie ilość badanych materiałów, co pozwoliło z kolei wprowadzić metody statystyczne pomagające weryfikować hipotezy badawcze dotyczące przepływu genów czy migracji. Prawdziwy rozkwit w badaniach kopalnego DNA nastąpił jednak w drugiej dekadzie XXI w. wraz z rozwojem technologii sekwencjonowania wysokoprzepustowego mogących generować olbrzymie ilości danych genetycznych. Umożliwiło to analizowanie coraz większej ilości prób i stosowanie bardziej wyrafinowanych metod statystycznych, zarówno by weryfikować hipotezy badawcze, jak i autentyczność otrzymanywnych wyników.

WSPÓŁCZESNY DNA A KOPALNY DNA

Według niektórych definicji kopalny DNA (zwany też starożytnym DNA od *ancient DNA*) to DNA izolowany albo z materiałów zdeponowanych X lat temu. Przy czym X jest umowny i różny w zależności od autora definicji i może wynosić np. 75 lat (Graham 2007). Kopalny DNA może też być definiowany bardziej ogólnie, jako DNA izolowany ze szczątków niezdeponowanych z myślą o analizie genetycznej. W definicjach tych trudno jednak wskazać różnice między kopalnym DNA a DNA analizowanym w postępowaniach medycyny sądowej, tak jak czasami trudno w przypadku pochówków ludzkich wskazać granice między ekshumacją a wykopaliskami prowadzonymi przez archeologów.

W dużym uproszczeniu kopalny DNA to DNA silnie pofragmentowany i posiadający charakterystyczne zmiany w swojej strukturze, powstałe w wyniku procesów depozycyjnych i postdepozycyjnych zachodzących po śmierci (*post-mortem*) danego organizmu.

DNA jest polimerem, długim łańcuchem składającym się z 4 rodzajów podjednostek zwanych nukleotydami, różniących się zawartymi w nich zasadami azotowymi: adeniną, tyminą, guaniną i cytozyną, skrótowo określanymi odpowiednio jako A, T,

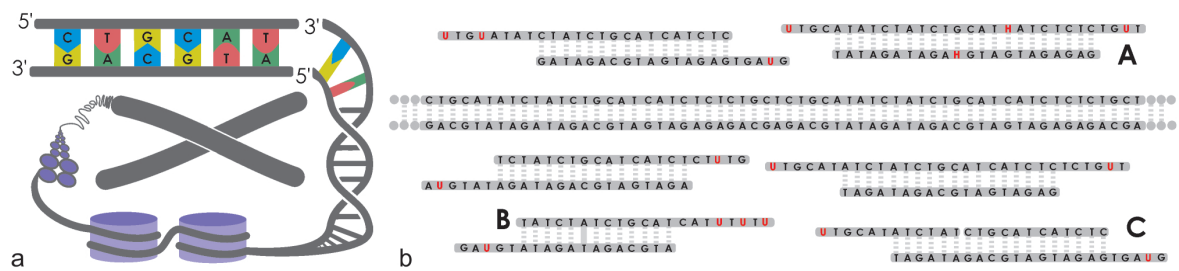
G i C. W organizmach DNA występuje najczęściej w postaci bardzo długich, liczących miliony nukleotydów cząsteczek. Poza tworzącymi łańcuch DNA silnymi wiązaniami kowalencyjnymi nukleotydy są w stanie też tworzyć słabe wiązania wodorowe między nukleotydami dwóch różnych łańcuchów, formując dwuniciowe cząsteczki zwinęte w ikoniczną podwójną helisę. Wiązania te występują jedynie w dwóch kombinacjach, adenina zawsze wiąże się z tyminą, a guanina z cytozyną. Z tego powodu dwuniciowe cząsteczki DNA składają się z łańcuchów komplementarnych względem siebie, w których A w jednym łańcuchu zawsze odpowiada T na drugim, a G występuje naprzeciw C (ryc. 1a).

Przeciętna komórka ludzka zawiera dwie pary 23 cząsteczek DNA, zwanych chromosomami, o łącznej długości około 3 miliardów pz (par zasad) każda. Dodatkowo każda komórka zawiera kilka do kilkudziesięciu tysięcy kopii cząsteczki DNA mitochondrialnego o długości 16 tysięcy pz. U wszystkich eukariontów (organizmów, których genom jest zawarty w jądrze komórkowym) te długie cząsteczki są w celach porządkowych, regulatorowych oraz (co istotne dla nas) ochronnych, skomasowane m.in. przez nawinięcie na kompleksy białkowe zwane histonami (ryc. 1a).

CHARAKTERYSTYKA FIZYKOCHEMICZNA CZĄSTECZEK KOPALNEGO DNA

Fragmentacja DNA zawartego w komórkach rozpoczyna się wraz z momentem śmierci organizmu. W pierwszej kolejności DNA ulega fragmentacji pod wpływem uwolnionych w wyniku śmierci komórkowej endogennych nukleaz, enzymów tnących kwasy nukleinowe (w tym DNA). W pierwszej kolejności cząsteczki DNA przecinane są w niechromonionych przez histony odcinkach i jako że na pojedynczy histon nawinięty jest fragment liczący 146 pz, na takiej długości fragment (oraz wielokrotność tej wartości) będzie wstępnie przecinany DNA.

Do degradacji DNA bardzo szybko mogą przystąpić mikroorganizmy (głównie bakterie), które rozkładają sam DNA, jak i chroniące go struktury (w tym histony). Procesy te dotyczą niemal wszystkich szczątków niezależnie od warunków środowiska i są odpowiedzialne za bardzo szybką utratę większości zawartego w tych szczątkach DNA (około 90% w ciągu pierwszego roku po depozycji



Ryc. 1. a – schematyczna struktura i sposób upakowania DNA w komórce eukariotycznej. Od prawej zgodnie z ruchem wskazówek zegara: podwójna nić DNA, DNA w postaci helisy, DNA nawinięte na histony, chromosom; b – schematyczny wygląd cząsteczek DNA w izolacji kopalnego DNA. Widoczne zmiany *post-mortem*: fragmentacja i deaminacja cytozyny w postaci czerwonych U oraz zanieczyszczenia współczesnym DNA w postaci długich i niemodyfikowanych cząsteczek. Dodatkowo pokazane są zmiany uniemożliwiające analizę DNA: zniszczenia spowodowane oksydacją (A), *cross-linking* pod wpływem światła UV (B) i przerywania jednej nici (C)



Ryc. 2. a – zęby o różnym (malejącym od lewej do prawej) stopniu wizualnego stanu zachowania. Jedynie ząb po prawej stronie rokuje dobrze dla izolacji kopalnego DNA; b – cała kość skroniowa z widoczną częścią skalistą; c – odseparowana część skalista kości skroniowej; d – przecięta część skalista z widocznymi strukturami ucha wewnętrznego

(Campos i in. 2012). Degradacja pozostałej ilości DNA postępuje pod wpływem czynników środowiskowych, zależnych od warunków depozycji danego organizmu. To właśnie te procesy, określane mianem rozkładu abiotycznego, odpowiedzialne są za przeważającą większość charakterystycznych zmian, jakim podlega DNA. Jest to przede wszystkim dalsza fragmentacja DNA. Także na tym etapie nie jest to proces przypadkowy, jako że odbywa się on najczęściej na drodze depurynacji hydrolytycznej, czyli przecinania łańcucha DNA w miejscu, w którym występują puryny (czyli G i A), prowadząc do powstania niesymetrycznych cząsteczek z tzw. lepkiymi końcami (ryc. 2). Co ciekawe, według

niektórych badań fragmentacja ta następuje stosunkowo szybko, tak że już po około 150 latach większość fragmentów ma średnią długość w granicach 40-80 pz (Sawyer i in. 2012). Tej długości fragmenty DNA mogą być chronione przed dalszą degradacją przez adsorpcję w macierzy albo tkanki kostnej albo sedymencie w miejscu depozycji (zwłaszcza zawartej w nim krzemionce).

Kolejnym istotnym abiotycznym procesem, jakiemu podlega DNA w wyniku depozycji, jest deaminacja nukleotydów (najczęściej cytozyny). Proces ten dotyczy najczęściej nukleotydów w jednokoniowych końcach 5' fragmentów DNA (powstałych w wyniku asymetrycznej fragmentacji) lub

nukleotydów z tymi końcami sąsiadujących. W wyniku tego procesu cytozyna zostaje przemieniona w uracyl-nukleotyd, który zastępuje tyminę w RNA (kwasie rybonukleinowym) i jako T odczytywany na drodze dalszych analiz. Gdy taki fragment DNA jest namnażany (krok niezbędny w celu dalszych jego analiz), w miejscu znajdującej się w oryginalnej sekwencji C występuje więc T, a na nici komplementarnej w miejscu G pojawia się A. Zmiany te następują stopniowo, skupiają się głównie na 10 ostatnich nukleotydach każdego fragmentu i w zależności od sprzyjających czynników i wpływającego czasu mogą dotyczyć nawet 40% znajdujących się tam cytozyn (Dabney i in. 2013).

CZYNNIKI WPŁYWAJĄCE NA STAN ZACHOWANIA KOPALNEGO DNA

Badanie czynników wpływających na stan zachowania DNA jest trudne, jako że jest wiele potencjalnych zmiennych mogących występować w dowolnych kombinacjach i często połączonych ze sobą siecią wzajemnych zależności. Ponadto warunki depozycji mogą ulegać zmianie w wyniku działań antropogenicznych i środowiskowych i nie wszystkie z tych zmian jesteśmy w stanie prześledzić i wykryć. Część z tych czynników możemy badać lub szacować eksperymentalnie, jednak nie zawsze zaobserwowane w ten sposób zależności znajdują potwierdzenia w badanych materiałach. Mimo to dzięki posiadanej ilości danych jesteśmy w stanie wytypować te mające największe znaczenie.

Średnia temperatura

Temperatura wydaje się mieć największe znaczenie dla stanu zachowania DNA, zwłaszcza że pośrednio wpływa też na inne czynniki. W dużym uproszczeniu: im niższa temperatura, tym lepiej zachowuje się DNA, szacuje się, że wraz z spadającą temperaturą częstość depurynacji zmniejsza się logarytmicznie, o rząd wielkości co około 16°C (Lever i in. 2015). Negatywny wpływ na stan zachowania DNA mogą mieć nie tylko wysokie okresowe temperatury, ale też jej wahania prowadzące do cykli zamrażania i odmrażania.

pH gleby

Także wzrastające pH ma wpływ na zmniejszenie tempa depurynacji, dlatego gleby obojętne lub lekko zasadowe lepiej prognozują dla pozyskiwania wystarczającej ilości DNA, w przeciwieństwie do kwaśnych środowisk. Wyjaśnia to też, dlaczego tak trudno uzyskać DNA ze szczątków zdeponowanych w bardzo kwaśnych torfowiskach, mimo ich pozornie dobrego wizualnego stanu zachowania.

Obecność ciekłej wody

Woda jest niezbędna do zachodzenia reakcji związanych z rozkładem i degradacją DNA zarówno biotycznych, jak i abiotycznych. Sama woda nie jest czynnikiem rozkładającym, a jedynie środowiskiem, w którym te procesy się odbywają. Pozyskiwanie wystarczającej do analiz ilości DNA z materiałów pochodzących ze stanowisk podwodnych jest więc możliwe, choć trudne (np. Briggs 2020).

Obecność tlenu

Tlen, a zwłaszcza jego reaktywne formy, ma bardzo destrukcyjny wpływ na DNA, powodując modyfikacje nukleotydów, które uniemożliwiają namnażanie DNA, a w konsekwencji jego analizę. W teorii większość potencjalnych materiałów jest zdeponowana w środowisku beztlenowym, jednak zniszczenia dokonywane przez tlen (ale też światło UV) mogą być jedną z przyczyn, dla których ciężko otrzymać odpowiedniej jakości DNA ze szczątków zmumifikowanych lub eksponowanych przez jakiś czas przed depozycją.

Światło UV

Światło UV jest kolejnym czynnikiem o destrukcyjnym wpływie. Promienie UV powodują powstawanie trwałych wiązań między oboma łańcuchami DNA (tzw. *cross-linking*, ryc. 1b) uniemożliwiających ich rozplecenie, a w konsekwencji analizę. Wystawienie szczątków na bezpośrednie działanie światła słonecznego przez dłuższy czas (przed depozycją, w jej trakcie lub po niej) może uniemożliwić

uzyskanie odpowiednich ilości DNA do dalszych analiz.

wody opadowej (z drugiej strony może też ułatwić dostęp wody gruntowej).

Obecność soli

Pewne sole, jak np. NaCl, mogą mieć wpływ na zmniejszenie tempa depurynacji, ale też ułatwić adsorpcję DNA przez krzemionkę, a więc sprzyjać zachowaniu się DNA.

Obecność kwasów humusowych

Kwasy humusowe oraz fulwowe, taniny i polifenole związane są ze środowiskiem glebowym i powstają głównie na skutek rozkładu materiału roślinnego. Ich ilość zależy od rodzaju flory porastającej glebę, pH, temperatury, a także uwodnienia. Obecne w glebie nie tylko obniżają pH środowiska, ale są też inhibitorami enzymów, które wykorzystują się w analizach DNA.

Obróbka termiczna

Towarzyszące spalaniu, gotowaniu czy kremacji wysokie temperatury mają bardzo destrukcyjny wpływ na DNA. Jakkolwiek eksperymentalnie dowiedziono, że można DNA wyizolować ze szczątków kostnych poddanych temperaturze do 600°C (Harbeck i in. 2011). Jednak kombinacja spalania i czynników postdepozycyjnych sprawiają, że przypadki uzyskania nadających się do analizy ilości DNA z materiałów spalonych są nieliczne i co najwyżej dotyczą szczątków o słabym stopniu przepalenia.

Niektóre z powyższych czynników mogą być częściowo modyfikowane i bezpośrednio związane z zmiennymi takimi jak głębokość depozycji, obecność mikroorganizmów i wiek materiału.

Głębokość depozycji

Gleba przykrywająca szczątki może odgrywać rolę izolacyjną oraz ochronną przed czynnikami atmosferycznymi. Warstwa gleby nad szczątkami obniża wahania temperatury, chroniąc je przed ich ekstremalnymi wartościami. Ponadto, w zależności od głębokości depozycji, może ograniczyć dostęp dla

Obecność mikroorganizmów

Mikroorganizmy rozkładające materię organiczną (a więc są bezpośrednio odpowiedzialne za biotyczny rozkład DNA) są wszechobecne, ich ilość jest więc raczej pochodną warunków środowiskowych niż zmienną samą w sobie. Ponadto ich obecność prowadzi do uwalniania na przestrzeni wielu lat dużych ilości DNA pochodzenia głównie bakterieryjnego, który zanieczyszcza nam badane materiały.

Obecnie jedną z najważniejszych wartości używanych do określania jakości otrzymanego kopalnego DNA jest właśnie stosunek endogennego DNA badanych szczątków do zanieczyszczającego go DNA „środowiskowego”. Odróżnienie jednych fragmentów DNA od drugih nie przysparza trudności metodycznych (o ile przedmiotem naszych badań nie są mikroorganizmy blisko spokrewnionych z tymi występującymi w glebie), ale wymaga nakładu pracy i środków. Przyjmuje się, że do jakichkolwiek analiz nadają się materiały zawierające przynajmniej 1% endogennego DNA (oznacza to, że w przypadku takich prób 99% generowanych danych jest w zasadzie wyrzucanych).

Wiek materiału

Czas, jaki upłynął od momentu depozycji do pobrania próby do badań molekularnych, jest bez wątpienia czynnikiem limitującym. Opierając się na szacowanym teoretycznym tempie degradacji DNA, obliczono, że nie powinien on być w stanie przetrwać dłużej niż kilkaset tysięcy, ewentualnie, w stanie ciągłego zamrożenia, do dwóch milionów lat (Lindahl 1993). I rzeczywiście, najstarsze do tej pory poznane sekwencje DNA liczą sobie około 1 200 000 lat i pochodzą z wydobytych z wiecznej zmarzliny szczątków mamuta (van der Valk i in. 2021). Co ciekawe, nie stwierdzono znaczącej korelacji między wiekiem prób a długością sekwencji (jak już wspomniałem wcześniej, większość fragmentacji dokonuje się w pierwszym stuleciu). Wydaje się, że najmocniejsza zaobserwowana zależność (zakładając zbliżone warunki depozycji)

występuje między wiekiem próby a ilością zdeaminowanych cytozyn na końcach fragmentów DNA (Sawyer i in. 2012).

Co warto jednak zaznaczyć, często z różnych szczątków pochodzących z pozornie zbliżonego kontekstu można otrzymać zróżnicowane wyniki izolacji kopalnego DNA. Może to wynikać albo z tego, że część z badanych szczątków została poddana procesowi depozycyjnemu lub postdepozycyjnemu, który nam jako badaczom umyka (np. szczególne praktyki pogrzebowe). Innym potencjalnym wyjaśnieniem zróżnicowania stopnia zachowania DNA jest sugerowane przez niektórych badaczy znaczenie tzw. mikrośrodowiska panującego w bezpośrednim sąsiedztwie opróbkowanego materiału. Jak się okazuje, w przypadku pobierania wielu prób z tego samego materiału (w tym przypadku tej samej kości tego samego szkieletu), dla których wszystkie wymienione wcześniej zmienne środowiskowe powinny być bardzo zbliżone, można uzyskać bardzo zróżnicowane ilości nadającego się do analiz kopalnego DNA (Hagelberg i in. 1991).

Podsumowując więc, kopalny DNA (ryc. 1b) zawiera niewielkie ilości krótkich (40-80 bp) fragmentów najczęściej przeciętych asymetrycznie w miejscach, gdzie pierwotnie znajdowały się G i A ze zdeaminowanymi do uracylu cytozynami przy końcach fragmentów (wraz z innymi, rzadszymi modyfikacjami). Ponadto występuje w postaci mieszaniny fragmentów DNA pochodzących zarówno od badanego organizmu, jak i z występujących w środowisku mikroorganizmów. Cechy te dostarczają trudności w izolacji oraz analizach kopalnego DNA, ale też, co zobaczymy w następnych podrozdziałach, mogą służyć do weryfikacji autentyczności otrzymanych wyników.

METODYKA PRACY Z KOPALNYM DNA

Wszystkie kroki, jakie podejmujemy, pracując z kopalnym DNA, mają na celu uzyskanie izolatów DNA o najwyższej, z punktu widzenia badań kopalnego DNA, jakości. Oznacza to dążenie do uzyskania izolatów zawierających jak najmniejszy stosunek DNA środowiskowego do endogennego oraz ograniczanie ryzyka kontaminacji DNA wspólnym.

PREFEROWANE MATERIAŁY DO BADAŃ

DNA jest wszechobecne i znajduje się we wszystkich szczątkach organizmów. Jednak, jak można wywnioskować z poprzedniego podrozdziału, nie we wszystkich materiałach DNA zachowuje się równie dobrze. Najczęściej DNA izoluje się ze szczątków kostnych, preferowanie pochodzących z inhumacji lub na innej drodze przykrytych osadami (najlepiej obojętnymi lub zasadowymi, w klimacie umiarkowanym lub chłodniejszym), lub ze stanowisk jaskiniowych. Nie oznacza to, że nadającego się do analiz DNA nie można uzyskać ze szczątków zmumifikowanych, zdeponowanych w kwaśnej glebie lub torfie albo poddanych umiarkowanej obróbce termicznej (gotowanie, nadpalenie). W tych przypadkach należy się jednak liczyć ze znacznym zwiększeniem nakładów (zarówno pracy, jak i finansowych) oraz koniecznością pobrania większej ilości prób w celu zwiększenia prawdopodobieństwa uzyskania izolatów nadających się do dalszych analiz.

W przypadku materiałów kostnych preferowanymi elementami szkieletu są zęby (z zachowanym korzeniem) i części skaliste kości skroniowej, w przypadku braku tych elementów do badań nadają się też fragmenty trzonów kości długich zawierające kość zbitą. Celem jest wybranie fragmentów jak najmniej porowatych, których wnętrza są odcięte od środowiska zewnętrznego, niepokrytych żadnym nalotem i nieposiadających żadnych przebarwień. Początkowo za najlepsze źródło kopalnego DNA uważano zęby (a w zasadzie wnętrza korzeni zębów chronione przez nieprzepuszczalne, ale pod względem zawartości DNA jałowe, szkliwo). Najlepiej, by pobierane zęby były dobrze zachowane, w pełni wykształcone, nieporowate i nie popękane (ryc. 2a).

W ciągu ostatnich paru lat okazało się jednak, że jeszcze lepsze rezultaty można uzyskać, izolując DNA z elementów ucha wewnętrznego (Pinhasi i in. 2015). Elementy te są (przynajmniej u ssaków) tkanką kostną o największej gęstości, dodatkowo chronioną przez kość zbitą części skalistej kości skroniowej (ryc. 2). Zdaniem niektórych badaczy izolaty DNA uzyskane z tych elementów mogą zawierać kilkakrotnie do kilkasetkrotnie większe ilości endogennego DNA niż inne elementy szkieletu (Gamba i in. 2014; Pinhasi i in. 2015). Z drugiej strony wydaje się, że w przypadku względnie dobrze zachowanych szczątków, izolując DNA jedynie

z cementu zęba (warstwy zewnętrznej korzenia), z pominięciem zębiny (znajdującej się wewnątrz zęba, ale podobnie jak szkliwo dość jałowej), i miazgi (zawierającej dużo żywych komórek, ale otwartej i podatnej na działanie środowiska zewnętrznego), można osiągnąć podobne efekty jak z kości skalistej (Hansen i in. 2017). W przypadkach mniej optymalnych uważa się kość skalistą za bezpieczniejszy wybór.

Jednak, co istotne, nie tylko szczątki kostne mogą być źródłem DNA. Liczne badania wskazują, że zarówno włosy (Rasmussen i in. 2011; 2010), jak i koproliity (Gilbert i in. 2008) mogą dawać dobre wyniki w przypadku braku dostępu do materiałów szkieletowych. Są też badania pokazujące uzyskanie nadających się do analiz ilości DNA z fragmentów skorupki jaj ptasich (Oskam i in. 2010). Gdy interesuje nas flora bakteryjna (w tym obecność patogenów) i/lub dieta, można izolować DNA z kamienia nazębnego (Weyrich i in. 2015) lub z noszących ślady żucia fragmentów dziegiu (Kashuba i in. 2019). Z obu tych źródeł udaje się uzyskać również DNA ludzkie, a w przypadku dziegiu nawet w dużych ilościach i to przy całkowitym braku szczątków kostnych (Jensen i in. 2019).

Osobną poddziedziną są też badania DNA środowiskowego, w których to dokonujemy izolacji DNA z warstw (zarówno antropogenicznych, jak i naturalnych), by poznać skład gatunkowy obecnej w nich materii organicznej. DNA środowiskowe może pomóc wykryć obecność gatunków zwierząt i roślin trudno uchwytanych w mikro- i makroskopowych szczątkach roślin (np. roślin owadopylnych). Dane takie mogą nam dostarczyć informacji na temat paleośrodowiska, uzupełniając dane archeobotaniczne i archeozoologiczne (Willerslev i in. 2014) oraz gospodarki człowieka i jej wpływu na środowisko (Smith i in. 2015) (por. Kołaczek i in., w tym tomie).

Planując badania, trudno oszacować, jaki procent pobieranych prób będzie zawierał DNA w ilości i jakości nadającej się do dalszych analiz, w dużej mierze będzie to zależało od badanego stanowiska i może wahać się od kilku do około 80%. Szacunków nie ułatwia fakt, że badacze bardzo rzadko podają łączną ilość prób pobranych do badań, w publikacjach opisując tylko te, z których udało się pozyskać nadające się do dalszych analiz DNA. Ilość prób, które należy pobrać, zależy też oczywiście od

pytań badawczych. jakie chcemy zadać. i dostępnych danych porównawczych. Strategie próbkowania będą się różnić w zależności od tego, czy interesuje nas jedynie struktura krewniacza grupy pochowanej na jednym stanowisku (wtedy należy pobrać ich jak najwięcej z danego stanowiska), czy struktura i relacje danej grupy z innymi populacjami. W tym drugim przypadku należy zadać sobie dodatkowe pytania: czy nasza grupa jest jednorodna (czy przypadkiem nie podejrzewamy, że składają się na nią dwie oddzielne subpopulacje należące np. do różnych klas społecznych), czy istnieją już dane referencyjne dla populacji, które podejrzewamy o pokrewieństwo genetyczne z badaną grupą, jak dokładnej rozdzielczości strukturę chcemy zrekonstruować (czy interesują nas kontakty między potencjalnie odległymi od siebie grupami czy dużo bliższymi podgrupami reprezentującymi sąsiednie regiony osadnicze).

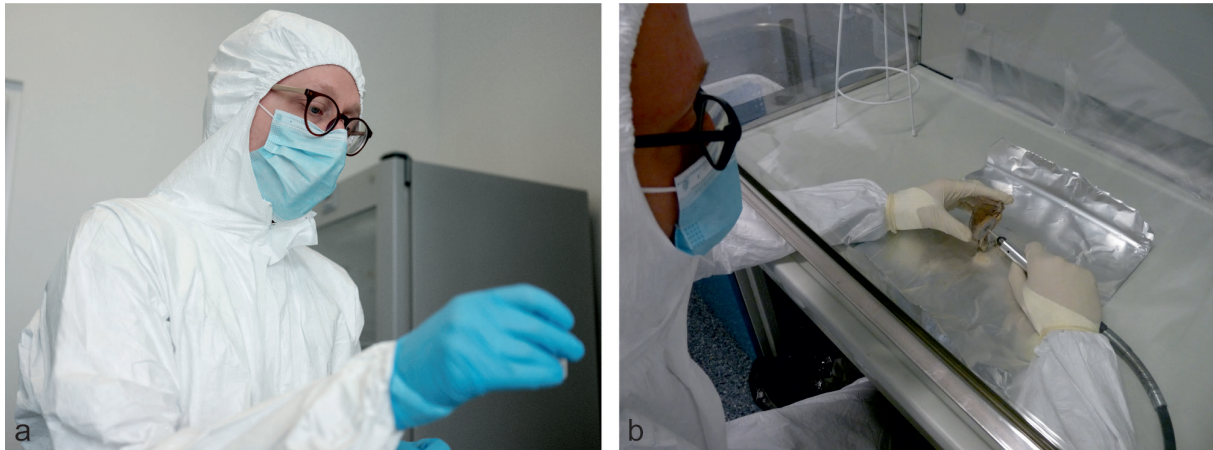
Planując takie badania „populacyjne” najlepiej dysponować wynikami wstępnymi, pozwalającymi oszacować procent osobników z dobrze zachowanym DNA. Tak by docelowo dążyć do minimum dziesięciu „dobrych” osobników (zwierających przynajmniej 5% DNA endogennego) dla każdej z badanych populacji (grup, podgrup i populacji referencyjnych).

POBRANIE PRÓB DO BADAŃ

Mając już wytypowane materiały, możemy przystąpić do ich pobrania oraz izolacji samego DNA. Od tego momentu do momentu namnożenia materiału genetycznego wszystkie procedury powinny być podporządkowane minimalizacji ryzyka kontaminacji współczesnym DNA.

KONTAMINACJA WSPÓŁCZESNYM DNA I SPOSOBY JEJ ZAPOBIEGANIA

Zanieczyszczenie współczesnym DNA było do niedawna, z uwagi na brak dobrych narzędzi do jego detekcji, przeszkodą w badaniach kopalnego DNA. Za najgroźniejsze uważa się zanieczyszczenie szczątków ludzkich przez DNA badaczy tych materiałów, począwszy od archeologów, przez antropologów fizycznych i muzealników po pracowników laboratoriów kopalnego DNA.



Ryc. 3. a – autor rozdziału w ubraniu ochronnym; b – proces przecinania części skalistej kości skroniowej. Fot. A. Wykrota (a), A. Juras (b)

Bezpośrednim źródłem zanieczyszczenia mogą być włosy, fragmenty naskórka, drobiny potu oraz oddech osób mających styczność z próbami. Współczesny DNA jest o tyle groźny, że metody amplifikacji materiału genetycznego mogą w preferowany sposób namnażać właśnie zanieczyszczenia. Największe zagrożenie wiąże się jednak z zanieczyszczeniami współczesnymi produktami namnażania DNA. Powstające w wyniku reakcji PCR (o której więcej poniżej) stężenia DNA są nieporównywalnie większe niż jego ilości w badanych materiałach i izolatach. Szacuję się, że jedna kropla aerozolu powstająca w wyniku zamknięcia próbówki zawierającej namnożony DNA może zawierać go więcej niż cały izolat. Dlatego też laboratoria kopalnego DNA, zwane potocznie czystymi laboratoriami (z ang. *clean lab*), muszą być fizycznie odseparowane od laboratoriów, w których pracuje się z namnożonym DNA, zwanych współczesnymi laboratoriami (z ang. *modern lab*). Separacja taka powinna polegać na:

- lokalizacji obu typów laboratoriów w innych budynkach (lub przynajmniej oddzielnych częściach budynku);
- instalacji osobnego, filtrowanego obiegu powietrza w czystym laboratorium;
- noszeniu ubrań ochronnych w laboratorium czystym: pełnych kombinezonów, maseczek, ochrony oczu, podwójnych rękawiczek, których spodnia warstwa nie powinna być zdejmowana w laboratorium, a wierzchnia zmieniana co próbę (ryc. 3a);
- utrzymywaniu laboratorium czystego w nadciśnieniu, z osobnym pomieszczeniem, służą służącą do przebrania się w strój ochronny;

- regularnej dekontaminacji wszystkich powierzchni, sprzętów i narzędzi;
- utrzymaniu ruchu jednokierunkowego między laboratoriami, możliwe jedynie przejścia od laboratorium czystego do współczesnego. Powrót do czystego laboratorium powinien być możliwy tylko następnego dnia, po uprzednim prysznicu i zmianie ubrań.

POBIERANIE PRÓB W TERENIE

Ekspozycje potencjalnych prób na kontakt z ludźmi należy ograniczyć jak najwcześniej i prowadzić wszystkie prace w maksymalnie sterylnych warunkach. Oczywiście bardzo rzadko możemy mieć kontrolę nad tymi czynnikami od momentu wykopalisk archeologicznych i większość badań dotyczy prób pobieranych z kolekcji muzealnych, które znalazły się w nich bez uwzględnienia możliwości badań kopalnego DNA. W przypadkach, gdy jednak mamy wpływ na proces pobierania prób jeszcze w terenie, należy przestrzegać (w ramach możliwości) paru wskazań:

- ograniczyć czas ekspozycji części szkieletu, z których planujemy pobrać próby do badań (najczęściej czaszki). Na czas ekspozycji reszty szkieletu do momentu jego dokumentacji i ekstrakcji można tą część przykryć materiałem lub folią;
- nigdy nie myć kości na mokro, jeśli zachodzi taka konieczność, można je oczyścić z osadu na sucho;
- jak najszybciej, po odpowiednim udokumentowaniu pobrać próby do badań, najlepiej jeszcze

w trakcie eksploracji obiektu, z którego pochodzą, w ostateczności po dokumentacji i opracowaniu materiałów przez specjalistę;

- zabezpieczyć wybrane materiały przed światłem słonecznym i wilgocią, np. zawinąć w czysty ręcznik papierowy do wysuszenia, a następnie w folię aluminiową i/lub woreczek strunowy;
- wszystkie powyższe czynności wykonywać przynajmniej w maseczce i rękawiczkach zmienianych regularnie (przynajmniej przy każdym przystąpieniu do danych materiałów);
- jeśli jednak próby do badań będziemy pobierać z kolekcji muzealnej, nie będziemy mieć wpływu (a często nawet wiedzy) na temat ilości osób oraz sposobu, w jaki się one obchodziły z danymi materiałami. Należy w takich przypadkach założyć, że najprawdopodobniej powierzchnia tych materiałów uległa kontaminacji. Nadal jednak, pobierając próby, powinniśmy zachowywać podobne środki ostrożności jak w przypadku robienia tego w trakcie wykopalisk.

WYBÓR WŁAŚCIWYCH FRAGMENTÓW MATERIAŁÓW SZKIELETOWYCH

W przypadku pobierania zębów, w pierwszej kolejności, kierujemy się ich wizualnym stanem zachowania (ryc. 2a). Pozostałymi kryteriami, które należy uwzględnić, mogą być:

- obecność antymerów (symetryczne odpowiedniki po drugiej stronie żuchwy lub szczęki), tak by jeden ząb z pary pozostawić do badań morfologicznych;
- łatwość ekstrakcji i jej destrukcyjny wpływ na przylegające struktury. Z zasady łatwiej pobiera się zęby jednokorzeniowe (siekacze, kły);
- koordynacja z innymi specjalistycznymi badaniami o charakterze destrukcyjnym. Teoretycznie analiz izotopowych szkliwa, izolacji DNA czy nawet datowania radiowęglowego można dokonać z tego samego zęba, jako że inne jego części mogą być wykorzystane do każdego z tych badań. W takich wypadkach, jeśli badania izotopowe są standaryzowane dla określonego zęba (np. pierwszego trzonowca), należy w celu ograniczenia destrukcyjności pobrać ten właśnie ząb (pod warunkiem, że jego stan zachowania jest dobry). Przy czym, z uwagi na ryzyko kontaminacji, należy w pierwszej kolejności przeprowadzić izolację DNA.

Pobieranie części skalistej kości skroniowej w zależności od stanu zachowania mózgowcowej może być bardziej złożone. Jeśli poszczególne kości nie są ze sobą połączone, wystarczy pobrać samą część skalistą, często zachowującą się jako osobny element, nawet przy bardzo słabym stanie zachowania czaszki (ryc. 2c) lub całą kość skroniową (ryc. 2b). Gdy jednak czaszka jest zachowana w całości, należy część skalistą wyciąć (np. przy użyciu wyposażonego w tarczki tnące elektronarzędzia).

Gdy natomiast zamierzamy izolować DNA z tkanki zbitych trzonów kości długich, należy najlepiej pobrać całe kości (mogą to być np. paliczki, kości śródstopia) albo, gdy uległy one fragmentacji w wyniku procesów postdepozycyjnych, ich najlepiej zachowane fragmenty.

Osobne protokoły stosuje się przy pobieraniu próbek z rdzeni glebowych, kamienia nazębnego czy warstw archeologicznych, należy jednak we wszystkich przypadkach zachować podobne względy ostrożności w celu minimalizacji ryzyka kontaminacji.

IZOLACJA KOPALNEGO DNA

Dekontaminacja pobranych prób

Po pobraniu próby należy ją jak najszybciej dostarczyć do czystego laboratorium, gdzie pobrane materiały należy umieścić w zamrażarce. Nie należy ich jednak zamrażać zaraz po pobraniu, jeśli nie jesteśmy w stanie zapewnić ich transportu w stanie zamrożonym (np. za pomocą przesyłki w suchym lodzie), jako że kolejne cykle zamrażania i rozmrażania mogą mieć destrukcyjny wpływ na DNA.

Niezależnie od sposobu pobrania prób w dalszej kolejności, już w czystych laboratoriach, stosuje się kroki mające za zadanie usunąć wszelkie zanieczyszczenia powierzchniowe. Można tego dokonać przy użyciu dowolnej kombinacji (lub najlepiej wszystkich) następujących metod:

- przemycie prób nisko stężonym (między 0,5% a 2%) roztworem podchlorynu sodu, w celu chemicznej degradacji znajdującego się na powierzchni próbek DNA. Gdy próby są niewielkich rozmiarów (np. zęby), można to zrobić przez zanurzenie w roztworze. Gdy są większe, można do przemycia użyć np. sterylnych jednorazowych wacików czy patyczków higienicznych;

- dekontaminację przez wystawienie na krótki czas (do 2 h) na silne promieniowanie UV (np. za pomocą specjalnych urządzeń zwanych *cross-linkerami*);
- usunięcie mechaniczne wierzchniej warstwy fragmentu, z którego bezpośrednio będziemy pobierali materiał do izolacji. Można tego dokonać np. jednorazowym skalpelem (przez skrobanie powierzchni), końcówką rozcierającą za pomocą elektronarzędzia (dekontaminując ją między użyciami) lub przez przecięcie fragmentu tarczką tnącą (znowu albo jednorazową, albo dekontaminowaną między użyciami) w celu odsłonięcia wnętrza fragmentu (ryc. 2d i 3b).

Rozdrobnienie materiału przed izolacją

Następnym krokiem jest albo rozdrobnienie części przeznaczanej do izolacji, albo jej odseparowanie od fragmentu kości w celu ekstrakcji DNA. Rozdrobnienia można dokonać bezpośrednio przy użyciu końcówek rozcierających i elektronarzędzia, celując w cement korzenia zęba lub widoczne po przecięciu kości skalistej struktury ucha wewnętrznego (ryc. 2d). Można też użyć mózdzierza i tłuczka po odseparowaniu od części niepożądanych (korony zęba, tkanki gąbczastej i zbitej wokół struktur ucha wewnętrznego). Oddzielenia od części niepożądanych można dokonać za pomocą tarczek tnących i elektronarzędzi lub pistoletu do piaskowania. Wszystkich te czynności należy wykonać, ograniczając nagrzewanie się materiału w wyniku tarcia, poprzez wykorzystywanie możliwie niskich obrotów elektronarzędzia i wiercenie interwałowe lub używając mózdzierza z opcją chłodzenia (tzw. *frezer-mill*).

Celem rozdrobnienia jest uzyskanie między 50 mg a 500 mg zhomogenizowanej do postaci drobnego proszku tkanki kostnej, z której następnie dokonamy ekstrakcji i izolacji DNA. Warto jednak nadmienić, że wyniki wstępne badań sugerują, że rozpuszczanie całych odciętych fragmentów (wymagające dłuższego czasu i większej objętości buforu ekstrakcyjnego) może dawać lepsze efekty niż rozpuszczanie proszku.

Ekstrakcja DNA

Proszek należy rozpuścić, doprowadzając do uwolnienia (ekstrakcji) DNA do roztworu. W tym celu

stosuje się około 1-2 ml buforu zawierającego składniki mające na celu rozpuścić zarówno frakcję mineralną (hydroksyapatyt), jak i organiczną (kolagen) kości. Reakcje prowadzi się w temperaturze 37-54°C do momentu rozpuszczenia proszku kostnego, zazwyczaj 24-48 godzin. Gdy jednak dokonujemy ekstrakcji z nierozdrobnionego fragmentu, ekstrakcja może potrwać do 5 dni, w czasie to których codziennie należy bufor ekstrakcyjny wymieniać i zbierać do późniejszej izolacji.

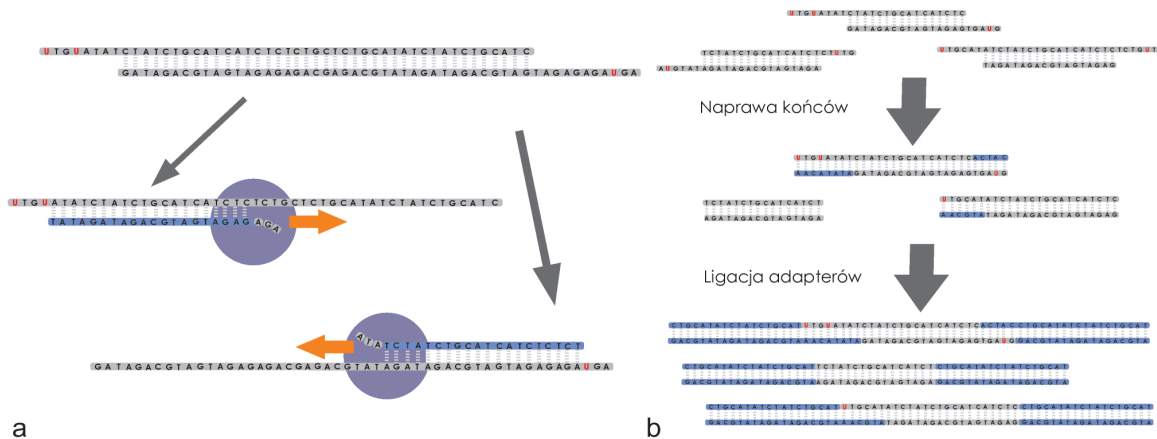
Część badaczy proponuje wykonywać krótką ekstrakcję wstępną, wychodząc z założenia, że endogenne DNA jest zamknięte i zaadsorbowane głęboko w macierzy tkanki kostnej, a zanieczyszczenia środowiskowe w większej ilości znajdują się w porach i przestrzeniach między tą macierzą (Damgaard i in. 2015). Według tej hipotezy pierwsza objętość ekstraktu, po 15-30 minutach inkubacji, zawiera przeważającą ilość DNA pochodzenia środowiskowego, można ją więc wyrzucić, wymieniając bufor, by zmaksymalizować ilość endogenego DNA w buforze docelowym.

Izolacja DNA

Z uzyskanego ekstraktu należy następnie wyizolować DNA, czyli oddzielić go od pozostałych składników ekstraktu. W biologii molekularnej dominuje kilka metod izolacji DNA, ale w badaniach kopalnego DNA wykorzystuje się w zasadzie różne warianty izolacji na podłożu krzemionkowym.

Metoda ta wykorzystuje właściwość DNA, którą już wspominaliśmy, czyli że w obecności niektórych soli absorbuje się on na powierzchni krystalicznej krzemionki. Krzemionka ta może w tej metodzie być unieruchomiona w postaci odpowiedniego filtra na kolumie, przez którą przepuszczamy poszczególne bufor w postaci drobnego proszku, który między poszczególnymi krokami zawirowujemy na dno próbki, w której prowadzimy izolację, lub opłaszczona na drobnych metalowych kulkach, które za pomocą magnesu unieruchamiamy na ściance próbki na czas wymiany buforów.

W wyniku izolacji otrzymujemy małą objętość (zazwyczaj między 40 µl a 100 µl) względnie czystego izolatu DNA. W zależności od protokołu wstępna objętość ekstraktu może być dużo większa (nawet do 5 ml), izolacja ma na celu również zagęszczenie roztworu, czasem w tym celu przed właściwą



Ryc. 4. a – pojedynczy cykl PCR. Aby polimeraza (zaznaczona na fioletowo) mogła działać, niezbędne jest, by sekwencje docelowe dla starterów (zaznaczonych na niebiesko) były dobrze zachowane; b – schemat tworzenia biblioteki genomowej

izolacją dokonuje się wstępnego przefiltrowania ekstraktu na filtrach celulozowych.

Namnożenie kopalnego DNA

Stężenia DNA w otrzymywanych izolatach są bardzo niskie, często poniżej poziomu detekcji metod wykorzystywanych standardowo w laboratoriach biologii molekularnej. Ponadto na stężenia te najczęściej składają się fragmenty DNA pochodzące od obecnych w środowisku mikroorganizmów. Niezależnie zresztą od pochodzenia izolatu, w celu analizy sekwencji DNA należy go najpierw namnożyć. Namnożenie może mieć na celu amplifikację tylko wybranych fragmentów genomu albo całego zawartego w izolacie DNA (tworzymy wtedy tzw. bibliotekę genomową). W obu przypadkach będziemy wykorzystywać cykliczną reakcję polimerazy (ang. *polymerase chain reaction*, PCR).

Reakcja PCR

Reakcja PCR jest podstawową metodą wykorzystywaną w genetyce i biologii molekularnej. Reakcja polega na cyklicznym zapętlaniu występujących w biologii komórki procesów w celu wykładniczego namnażania fragmentu DNA, tak by w krótkim czasie otrzymać miliony jego kopii. W dużym uproszczeniu metoda wykorzystuje enzym polimerazę, by po rozpleceniu dwuniciowej cząsteczki DNA (w wyniku działania wysokiej temperatury) stworzyć jej

dwie kopie, a następnie całość powtórzyć określoną ilość razy, za każdym razem otrzymując dwukrotną ilość fragmentów DNA. Do zajścia reakcji potrzebne jest kilka składników:

- polimeraza, a właściwie jej odmiana znaleziona w bakteriiach żyjących w gorących źródłach i w związku z tym działająca w wysokiej temperaturze potrzebnej do rozplecenia podwójnej nici DNA;
- deoksynukleotydy, czyli cegiełki, z których będziemy budować nowe kopie interesujących nas fragmentów;
- bufor tworzący optymalne warunki do przeprowadzenia reakcji;
- startery. Polimeraza wymaga dwuniciowego fragmentu, który rozpoznaje jako miejsce początku reakcji. W tym celu do reakcji dodajemy dwie krótkie (do około 25 pz) syntetyczne jednoniciowe cząsteczki DNA, zwane starterami, które flankują interesujący nas region. Gdy po wstępnym rozpleceniu nici obniżymy ponownie temperaturę, w pierwszej kolejności do obu pojedynczych nici przyłącza się krótsze fragmenty, tworząc dwie częściowo dwuniciowe cząsteczki które polimeraza następnie dokończy (ryc. 4).

Bezpośredni PCR wybranych fragmentów genomu zawierających interesujące nas informacje był dominującą metodą w pracy z kopalnym DNA do końca pierwszej dekady XXI w. Jednak wraz z rozwojem technik sekwencjonowania wysokopręstowego został wyparty przez metody oparte na bibliotekach genomowych.

Biblioteki genomowe

Biblioteka genomowa, zgodnie z definicją, to zbiór wszystkich fragmentów DNA w danej próbie (izolacie) umieszczonych na identycznym wektorze. W praktyce funkcję wektora pełnią najczęściej adaptery, czyli krótkie dwuniciowe fragmenty DNA przyklejone do końców wszystkich cząsteczek biblioteki. Sposobów przyklejenia (ligacji) tych fragmentów jest kilka, zarówno dostępnych w komercyjnych zestawach do tworzenia bibliotek (udostępnianych przez producentów sekwenatorów), jak i opracowanych przez samych badaczy. Większość tych protokołów składa się z kilku podobnych kroków, czasami wykonywanych jeden po drugim w następujących po sobie reakcjach, czasami zgrupowanych w jednej reakcji. Pierwszym krokiem jest „przytępienie” lepkich końców przez ich uzupełnienie lub przycięcie (najczęściej przycięcie końca 3' i uzupełnienie 5' za pomocą polimerazy T4), tak by wszystkie cząsteczki DNA były symetryczne i miały same tępe końce. Potem następuje ligacja samych adapterów, za pomocą enzymu ligazy. Na samym końcu bibliotekę amplifikujemy, wykorzystując sekwencje adapterów jako sekwencje docelowe dla starterów w PCR. Oprócz umożliwienia nam namnożenia całości zawartego w izolacie DNA adaptery pełnią inną ważną funkcję, pozwalając na przyłączenie się starterów w reakcji sekwencjonowania (działającej na podobnej zasadzie jak PCR).

PCR jest celowy, dąży do namnożenia z góry wybranego fragmentu. Może to być fragment zawierający polimorfizm odpowiedzialny za jakąś określoną cechę, np. zdolność trawienia laktozy (Mnich i in. 2018) albo fragment charakteryzujący się dużą zmiennością, a więc przydatny w badaniach populacyjnych, np. region hiperzmienny genomu mitochondrialnego (Juras i in. 2014). W wyniku sekwencjonowania wysokoprzepustowego w jednej reakcji dostajemy informację na temat przynajmniej dziesiątek tysięcy zmiennych. Z drugiej strony w metodach opartych na bibliotekach genomowych informacje, które otrzymujemy, są w dużej mierze losowe i im więcej obecnych w naszej próbie zanieczyszczeń pochodzenia środowiskowego, tym więcej otrzymywanych sekwencji jest nieprzydatnych (jeśli nie interesują nas obecne w środowisku mikroorganizmy). Jednak sama skala sekwencjonowania wysokoprzepustowego (sekwencjonowanie dziesiątek milionów fragmentów z każdego izolatu)

nawet w przypadku izolatów, które zawierają 95% fragmentów środowiskowych pozwala nam uzyskać tysiące informatywnych sekwencji.

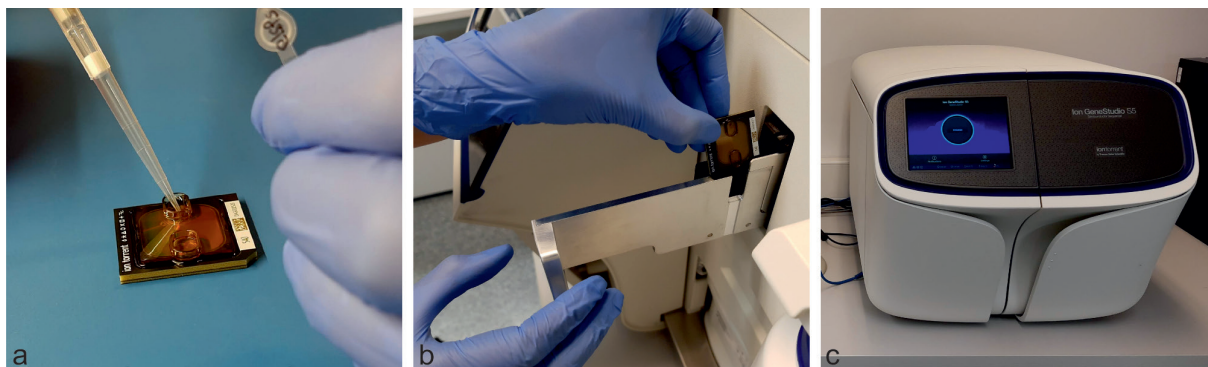
Ponadto bezpośredni PCR wychodzi z założenia, że docelowy fragment jest w całości zachowany, co w przypadku kopalnego DNA jest mało prawdopodobne. Szanse powodzenia możemy zwiększyć przez wybór bardzo krótkich sekwencji docelowych (pamiętając, że średnia długość fragmentów rzadko kiedy przekracza 80 pz), ale nadal musimy mieć nadzieję, że przynajmniej w którejś z cząsteczek interesujące nas fragmenty znajdują się w ich wnętrzu, a sekwencje docelowe dla starterów nie zawierają uniemożliwiających ich przyłączenie modyfikacji (co zważając, że te najczęściej będą się znajdować przy końcach cząsteczek DNA, jest mało prawdopodobne). W takim wypadku może dojść do sytuacji, w której to fragment pochodzący z zanieczyszczeń, a nie od badanych szczątków, będzie w preferowany sposób namnażany. Tymczasem metody oparte na bibliotekach genomowych prowadzą do namnożenia wszystkich cząsteczek (razem z częściowo lub całkowicie zachowanymi końcami), a proces produkcji bibliotek można zmodyfikować w ten sposób, by odsiewać długie cząsteczki pochodzące z dużą dozą prawdopodobieństwa z zanieczyszczeń.

Sekwencjonowanie DNA

Zakładając, że mamy już albo namnożone interesujące nas fragmenty, albo bibliotekę genomową, możemy się przenieść do „współczesnego laboratorium”, gdzie będziemy mogli poznać ich sekwencję. Zasadniczo możemy mówić o dwóch typach sekwencjonowania: Sangera (nazywanym tak od wynalazcy metody) i sekwencjonowni wysokoprzepustowym.

Sekwencjonowanie Sangera

Sekwencjonowanie Sangera polega na przeprowadzeniu specyficznej reakcji amplifikacji DNA, do której oprócz zwykłych deoksynukleotydów dodajemy też stopujące polimerazę i wybarwione barwnikami fluorescencyjnymi (każdy innym) dideoksynukleotydy. Jako że produktem wyjściowym reakcji PCR były miliony identycznych cząsteczek, statystycznie w wyniku reakcji sekwencjonowania



Ryc. 5. a – ładowanie za pomocą mikropipety biblioteki genomowej na chip do sekwencjonowania w technologii ION. b – instalacja chipa do sekwenatora. c – sekwenator ION Gene Studio S5

otrzymamy wiele wariantów produktu reakcji o różnej długości (a dokładniej tyle wariantów, jaką długość liczy nasz produkt PCR). Fragmenty te następnie należy rozdzielić pod względem długości, przy użyciu wąskich i długich kapilar wypełnionych polimerem stanowiącym przeszkodę dla DNA. Na końcu kapilary umieszczony jest detektor, który wykrywa, jakim barwnikiem wybarwione są kolejno przechodzące przez nią produkty reakcji sekwencjonowania. Z kolejności produktów, od najkrótszego (długości 1 pz) do najdłuższego (o długości równej wielkości analizowanego fragmentu) oraz obecnego w nich barwnika jesteśmy w stanie odtworzyć badaną sekwencję.

Sekwencjonowanie wysokoprzepustowe

Jest kilka technologii sekwencjonowania wysokoprzepustowego, ale w analizie sekwencji kopalnego DNA dominują głównie metody oparte na sekwencjonowaniu przez syntezę. W dużym skrócie metody te polegają na tym, by w jednym punkcie syntetyzować (znowu za pomocą polimerazy) wiele kopii pojedynczej cząsteczki DNA w taki sposób by przyłączenie każdego nukleotydu dawało wykrywalny sygnał. Rolę punktu może odgrywać dołek w mikrochipie (ryc. 5a) albo wydzielone pole na specjalnej płytce. Przyłączanie nukleotydu może być wykrywane albo przez zmianę pH, albo przez sygnał świetlny, gdy do nukleotydów przyłączone są fluorochromy. Cała idea sprowadza się do tego, by tych punktów (a więc zsekwencjonowanych w jednej reakcji fragmentów) było jak najwięcej, najlepiej setki milionów. Obecnie dominują dwie wiodące technologie: ION (od *Life Technologies*)

oraz metoda firmy *Illumina*. Niezależnie od producenta technologii sekwencjonowanie dotyczy względnie krótkich fragmentów (co w przypadku kopalnego DNA nie jest wadą a zaletą), do 200 pz, albo 400 pz w przypadku IONa i 75 pz, 100 pz albo 150 pz w przypadku *Illuminy*. Ilość sekwencji generowanych w jednej reakcji wynosi do 80 milionów w przypadku ION'a lub do 10 miliardów (sic!) w przypadku *Illuminy*.

CHARAKTERYSTYKA DANYCH POLIMORFIZMY DNA

By omówić specyfikę danych paleogenomicznych, warto przypomnieć główne rodzaje zmiennych (w genetyce nazywanych polimorfizmami) obecnych w DNA oraz to, w jaki sposób są one dziedziczone. To właśnie obecność (lub brak) oraz częstotści występowania wariantów (zwanych w genetyce allelami) różnicuje nam osobniki oraz populacje i jest celem sekwencjonowania.

Polimorfizmy pojedynczego nukleotydu

Najczęściej analizowanym przez badaczy kopalnego DNA rodzajem zmiennych są polimorfizmy pojedynczego nukleotydu (ang. *single nucleotide polymorphism*, SNP). Jak nazwa wskazuje, są to różnice w sekwencji, w których badane osoby, osobniki, gatunki, szczepy itp. różnią się pojedynczym nukleotydem. Może to oznaczać, że dany nukleotyd jest inny w zależności od allelu albo w niektórych wariantach danego nukleotydu nie ma. Najczęściej (ze względu

na częstość występowania, ale też łatwość badania) w danych paleogenomicznych mamy do czynienia z tym pierwszym przypadkiem. W organizmach diploidalnych (czyli posiadających dwa zestawy chromosomów) w genomie jądrowym każdy osobnik ma dwa warianty danego polimorfizmu, mogą być one albo takie same (mówimy wtedy, że dla danego polimorfizmu osobnik jest homozygotą), albo różne (wtedy jest heterozygotą). W rozmnażaniu płciowym potomstwo otrzymuje losowo po jednym wariantcie od każdego z rodziców. W perspektywie długiego czasu można założyć, że polimorfizmy dziedziczą się niezależnie od siebie. Dzieje się tak dzięki rekombinacji, procesowi, w którym w rozmnażaniu płciowym u każdego z rodziców dwa chromosomy z pary wymieniają się fragmentami. Im dalej od siebie na chromosomie znajdują się dwa polimorfizmy, tym bardziej prawdopodobne, że będą one dziedziczone niezależnie.

Polimorfizmy ilości powtórzeń

Zdecydowanie rzadziej wykorzystywanymi przez badaczy kopalnego DNA (mimo dominacji w biologii sądowej) są polimorfizmy ilości powtórzeń. Ponownie nazwa dobrze opisuje to złożone zjawisko, są to warianty charakteryzujące się zmienną ilością powtórzeń jakiegoś motywu w sekwencji DNA. Spośród kilku typów polimorfizmów ilości powtórzeń, najczęściej analizowanymi są krótkie powtórzenia tandemowe (ang. *short tandem repeats*, STR). Są to powtórzenia krótkich (2-4 bp) motywów następujących kilka (zazwyczaj od 5 do 20) razy. Oznacza to, że „siła rozdzielcza” STR jest zdecydowanie większa. Dużo niższe jest prawdopodobieństwo, że dwie osoby w sposób zupełnie przypadkowy dzielą ten sam z 15 możliwych niż z dwóch (jak w przypadku SNP) wariantów. Z tego powodu STR zdominowały badania z dziedziny medycyny sądowej. Jednakże ich użyteczność w badaniach kopalnego DNA jest ograniczona. STR są względnie długie, a ich badaniach wykorzystuje się metody oparte na bezpośrednim PCR. Pojedynczy STR ma większą moc statystyczną niż pojedynczy SNP, ale przy użyciu metod opartych na sekwencjonowaniu wysokoprzepustowym łatwiej jest uzyskać informacje na temat tysięcy SNP niż przy użyciu bezpośredniego PCR na temat kilku STR. Ponadto wykluczenie kontaminacji współczesnym DNA na

podstawie STR jest zadaniem zdecydowanie trudniejszym i mimo kilku zakończonych umiarkowanym sukcesem prób w poprzedniej dekadzie, obecnie raczej się od tej metody odchodzi.

HAPLOTYPY I HAPLOGRUPY

Pojęciem często używanym w genetyce populacyjnej, ale też w genealogii zarówno przez genetyków, jak i pasjonatów dziedziny, jest haplogrupa. By zrozumieć to pojęcie, musimy sobie przypomnieć, że człowiek (jak i wszystkie ssaki i większość kręgowców) ma genom diploidalny, ale jego fragmenty mogą być haploidalne (czyli posiadać tylko jedną kopię lub wariant). Dotyczy to przede wszystkim DNA mitochondrialnego i fragmentów chromosomu Y u osobników płci męskiej (u człowieka). Mitochondria jako centra energetyczne komórki, mają własny krótki genom (o długości 16569 pz). Mimo że w jednej komórce może być wiele mitochondriów (nawet tysiące), to zasadniczo wszystkie posiadają ten sam identyczny genom mitochondrialny (mtDNA). Mitochondria (a więc ich genomy) są u większości eukariotów dziedziczone w linii matczynej (z matki na potomstwo obu płci). Genom mitochondrialny w żaden sposób nie rekombinuje i za wyjątkiem losowo pojawiających się pod wpływem czynników środowiskowych, pojedynczych mutacji nie zmienia się z pokolenia na pokolenie. Genomy mitochondrialne różnią się od siebie polimorfizmami, przed wszystkim w postaci SNP, które nagromadziły się w poszczególnych liniach na drodze ewolucji. Taką linię ewolucyjną o określonym zestawie alleli w miejscach polimorficznych określamy mianem haplotypu. Grupę haplotypów o wspólnym pochodzeniu (znajdującą się na tej samej gałęzi drzewa filogenetycznego) nazywamy haplogrupą. Analogicznie, nie-rekombinujący region chromosomu Y (stanowiący 95% całego chromosomu) jest dziedziczony w niezmienionej postaci w linii męskiej (z ojca na potomstwo płci męskiej). Cały chromosom Y jest znacznie dłuższy niż genom mitochondrialny, liczy 58 milionów pz i zawiera w sobie znacznie większą ilość polimorfizmów różnych rodzajów.

Schemat dziedziczenia uniparentalnego (od jednego rodzica) oraz fakt dziedziczenia w niezmienionej postaci ma istotne implikacje dla badań genetycznych zarówno współczesnego, jak i kopalnego DNA:

- znając genotyp tylko części polimorfizmów, możemy przypisać danego osobnika do haplogrupy lub subhaplogrupy bez konieczności rekonstrukcji całego genomu. Początkowo badania kopalnego DNA polegały na sekwencjonowaniu tzw. regionu hiperzmiennego mtDNA, który to region, liczący sobie około 1000 pz, zawiera około 30% polimorfizmów obecnych w mtDNA. Poprzez sekwencjonowania kilku krótkich (poniżej 100 pz) produktów można uzyskać informacje z wystarczającej liczby polimorfizmów, by przypisać danej próbie haplogrupę. Jednak by poznać subhaplogrupę lub dokładny haplotyp (w celu np. wykluczenia pokrewieństwa w linii matczynej), niezbędne jest sekwencjonowanie całego genomu mitochondrialnego;
- możemy patrzeć na całe genomy haploidalne jako pojedyncze zmienne (a nie zbiory zmiennych). Przedmiotem analiz statystycznych są wtedy nie poszczególne polimorfizmy, a haplogrupy i haplotypy. Informatywne dla nas są wtedy nie tylko obecność (lub brak) poszczególnych haplogrup, ale częstości ich występowania w poszczególnych populacjach;
- nierekombinujące i dziedziczone uniparentalnie (w linii od jednego rodzica) markery niosą jednak z sobą ograniczoną ilość informacji. Rozpoznanie haplotypu mtDNA informuje nas o genetycznym wpływie ograniczonej ilości przodków, np. świadczy tylko o pojedynczym osobniku sprzed czterech pokoleń (praprababce od strony matki). Tymczasem w analogicznym przypadku dane uzyskane w wyniku analizy genomu jądrowego niosą ze sobą informacje o wszystkich 16 przodkach sprzed czterech pokoleń.

Ponadto genomy mitochondrialne, co ma niebagatelne znaczenie w badaniach kopalnego DNA, są obecne w komórkach w wielu (nawet tysiącach) kopiach. W związku z tym ilość fragmentów DNA pochodzących z genomu mitochondrialnego, mimo że genom ten jest stosunkowo niewielki, potrafi być bardzo duża. Skutkiem tego zrekonstruowanie haplogrupy lub haplotypu mitochondrialnego jest możliwe nawet w przypadku prób z bardzo słabym stopniem zachowania DNA, dla których uzyskanie, nadających się do analiz statystycznych ilości DNA jądrowego jest nieopłacalne.

Analiza danych

Analiza danych genetycznych stanowi osobną dziedzinę na styku biologii populacyjnej, statystyki i bioinformatyki i w przypadku badań kopalnego DNA składa się z kilku kroków: analiz sekwencji, mapowania i genotypowania.

Analizy sekwencji

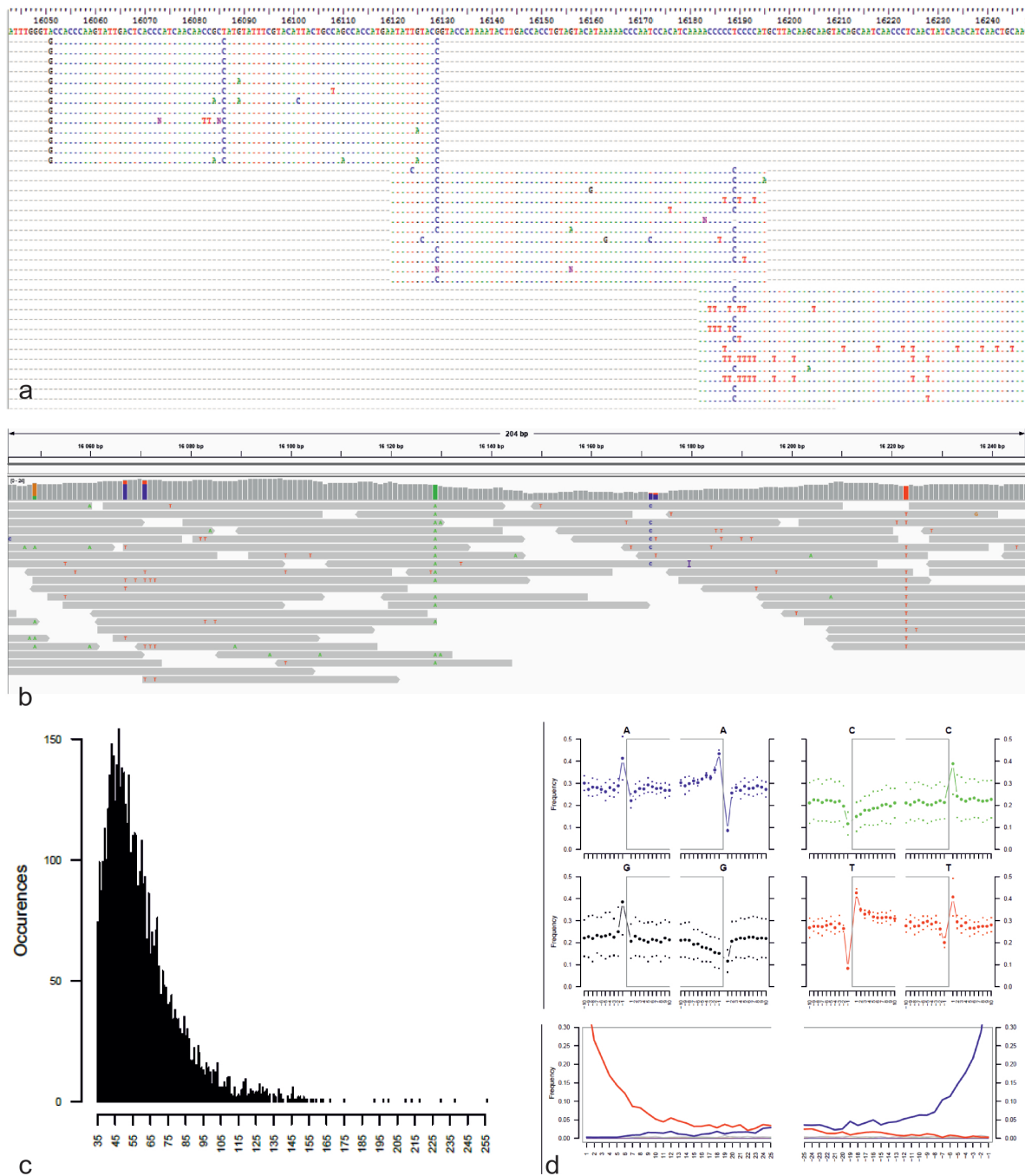
Bioinformatyczna obróbka danych sekwencjonowania istotna zwłaszcza w metodach wysokoprzepustowych, mająca za zadania odfiltrować sekwencje o niskiej jakości, artefakty, pozostałości sekwencji adapterów itp.

Mapowania do genomu(-ów) referencyjnych

W przypadku wyników uzyskanych na drodze sekwencjonowania produktów PCR uzyskane sekwencje można ręcznie porównać między sobą i sekwencją referencyjną nawet przy użyciu dowolnego edytora tekstowego, choć znacznie wygodniej wykorzystać do tego specjalne oprogramowanie. Wyniki sekwencjonowania wysokoprzepustowego wymagają specjalistycznego oprogramowania oraz dużej mocy obliczeniowej (najlepiej centrów superkomputerowych) w celu określenia, czy i w których miejscach miliony uzyskanych sekwencji pasują do sekwencji referencyjnej. W przypadku gdy interesuje nas jeden gatunek, którego genom znamy (jak wtedy, gdy badamy szczątki ludzkie), proces ten jest prostszy i szybszy niż w przypadku gdy badamy bioróżnorodność, np. fauny bakteryjnej kamienia nazębnego, lub w próbach środowiskowych. W takim przypadku musimy porównywać otrzymane sekwencje do całych baz danych (lub ich wycinków, jeśli interesuje nas tylko określona grupa organizmów).

Analizowania stopnia zanieczyszczenia prób

Istotny krok niezbędny do potwierdzenia autentyczności uzyskanych danych, wymagający jednak wcześniejszego zmapowania sekwencji do referencji, zostanie omówiony bardziej szczegółowo poniżej.



Ryc. 6. a – sekwencje regionu hiperzmiennego mtDNA uzyskane w wyniku sekwencjonowania metodą Sangera produktów wielu PCR (widoczne są tylko różnice względem sekwencji referencyjnej); b – ten sam fragment genomu mitochondrialnego, ale pokryty sekwencjami uzyskanymi w wyniku sekwencjonowania wysokoprzepustowego biblioteki genomowej (ponownie widoczne są tylko różnice względem referencji); c – wykres rozkładu długości fragmentów endogenego DNA obecnych w bibliotece genomowej; d – wzorce zmian *post-mortem* obliczone za pomocą programu mapDamage. U góry częstość nukleotydów w sąsiedztwie końców fragmentów DNA (częstsze A i G przy końcach 5' oraz C i T przy końcach 3'). Na dole częstość zmian C na T (na czerwono) i G na A (na niebiesko) względem ich pozycji fragmentach DNA (wzrastająca przy końcach tych fragmentów) (wyk. M. Chyleński)

Genotypowanie

Proces wyszukiwania polimorfizmów, które są obecne w naszych danych, oraz określenie alleli, które nasze badane próby posiadają w miejscach polimorficznych. Inaczej mówiąc, jest to wyszukiwanie zmiennych, których następnie użyjemy w analizach statystycznych i też wymaga osobnego podrozdziału w celu dokładniejszego omówienia.

Analizy statystyczne, populacyjne i inne

Zbiór metod i narzędzi, które mają za zadanie weryfikować nasze hipotezy badawcze, zostanie im poświęcony osobny podrozdział.

WYKLUCZANIE KONTAMINACJI WSPÓŁCZESNYM DNA

W metodach opartych na bezpośrednim PCR jedynym sposobem wykluczenia kontaminacji jest replikacja wyniku. Replikacja może polegać na kolejno:

- przeprowadzeniu kilku (min. 2) izolacji z pojedynczej próby;
- przeprowadzeniu kilku (min. 2) PCR z pojedynczego izolatu;
- sekwencjonowaniu kilku (min. 2) produktów z każdego PCR.

Jeśli we wszystkich przypadkach otrzymamy taki sam wynik i przy jego otrzymywaniu zastosowaliśmy wszystkie zalecenia wskazane do pracy z kopalnym DNA, to możemy uznać ostateczny wynik za autentyczny. Dodatkowym kryterium może być występowanie substytucji względem referencji C na T i G na A w niektórych sekwencjach, ich brak nie wyklucza jednak jednoznacznie, że sekwencje pochodzą z kopalnego DNA (ryc. 6a).

Sekwencjonowanie wysokoprzepustowe dostarcza nam sekwencji całych fragmentów DNA (ryc. 6B). Takie sekwencje zawierają w sobie dodatkowe, przydatne w wykrywaniu zanieczyszczeń informacje. Jako że danych jest duża ilość, można te informacje skwantyfikować i opracować statystycznie. Zmiany, o których tu mowa, to przede wszystkim:

- informacje o zmianach *post-mortem* względem sekwencji referencyjnej, przede wszystkim zmiany C na T i G na A blisko końców fragmentów DNA (ryc. 6d);

- rozkład długości fragmentów DNA pokazujący przewagę fragmentów krótkich i bardzo krótkich (ryc. 6c);
- wzorzec fragmentacji DNA pokazujący zwiększoną częstość puryn (A i G) w bezpośrednim sąsiedztwie końców 5' i odpowiadające im pirymidyny (C i T) przy końcach 3' (ryc. 6d).

Ponadto ilość sekwencji, jaką dysponujemy umożliwia oszacowanie stopnia kontaminacji poprzez przyjrzenie się sekwencjom zmapowanym do haploidalnych części badanych genomów, takich jak mtDNA i chromosom X (w przypadku osobników płci męskiej) w celu kwantyfikacji i analizy wszystkich niezgodności i niekonsekwencji w otrzymanych sekwencjach (Fu i in. 2013; Renaud i in. 2015).

GENOTYPOWNIE

W przypadku bezpośredniego PCR najczęściej z góry znamy polimorfizmy, które analizujemy, mogą to być polimorfizmy odpowiadające za określoną cechę albo diagnostyczne dla określenia haplogrup (mtDNA i Y-DNA) lub przynależności gatunkowej. W przypadku danych pochodzących z sekwencjonowania wysokoprzepustowego do genotypowania należy użyć specjalistycznego oprogramowania wyszukującego polimorfizmy. Zakładając, że dysponujemy dużą ilością danych dla większej ilości prób reprezentujących określoną populację, możemy dokonywać tego tylko za pomocą wygenerowanych przez nas sekwencji. Można też (co stosuje się najczęściej w badaniach kopalnego DNA) wykorzystać listy polimorfizmów uzyskane dla referencyjnych populacji współczesnych. Genotypowanie danych kopalnych wiąże się jednak z trudnościami związanymi bezpośrednio z charakterystyką kopalnego DNA, a zwłaszcza z dwoma jej aspektami: fragmentarycznym charakterem danych oraz zmianami *post-mortem*.

FRAGMENTARYCZNY CHARAKTER DANYCH

Wbrew potocznie używanym i pojawiającym się w doniesieniach prasowych określeniom w badaniach kopalnego DNA bardzo rzadko uzyskujemy „całe genomy”. Genom jest pojęciem problematycznym, tak naprawdę do tej pory nie udało nam

się zrekonstruować w 100% całego dużego genomu eukariotycznego. Nawet najlepiej zbadany genom ludzki ma spore luki i mimo że minęło już ponad 15 lat od ogłoszenia wielkiego sukcesu, jakim było jego zsekwencjonowanie, dopiero w zeszłym roku poznaliśmy pierwszą kompletną sekwencję chromosomu (Miga i in. 2020). Sekwencjonowanie wysokoprzepustowe nadal jest kosztownym i czasochłonnym przedsięwzięciem, w związku z czym najczęściej ograniczamy je w momencie, gdy uzyskamy wystarczającą ilość danych, by wykorzystać je w planowanych analizach statystycznych.

Często więc sekwencje dla jednej próby, które nazywamy potocznie zrekonstruowanym genomem, będą pokrywały się z genomem referencyjnym w zakresie nie przekraczającym 10%. Pojęciem, którego najczęściej się używa do opisu ilości danych, jaką dysponujemy, jest „pokrycie genomu”. Pokrycie genomu równe 1x oznacza, że ilość nukleotydów, jaką zmapowaliśmy do genomu referencyjnego, jest równa wielkości tego genomu. Nie oznacza to, że zrekonstruowaliśmy cały genom tego osobnika, bo wiele miejsc w takim genomie będzie pokryte więcej niż jedną sekwencją i równocześnie o wielu jego częściach nie będziemy mieć żadnych informacji.

ZMIANY POST-MORTEM

Substytucje C na T i G na A będące skutkiem deaminacji cytozyny w pojedynczej sekwencji w miejscu polimorficznym są nie do odróżnienia od alleli występujących przyżyciowo. W celu rozwiązania tego problemu można zastosować kilka podejść (lub ich kombinacji):

- ignorować całkowicie polimorfizmy (zwane tranzycjami), w których allelami są C/T oraz A/G i używać do analiz tylko polimorfizmów (zwanych transwersjami), w których możliwymi allelami są G/T, G/C, A/T oraz A/C. Stosując to podejście, jako że tranzycje występują w genomie znacznie częściej niż transwersje, świadomie rezygnujemy z ponad 60% polimorfizmów, które moglibyśmy analizować;
- dążyć do wysokiego pokrycia genomu. Zmiany *post-mortem* występują losowo, jeśli dany polimorfizm będzie pokryty wieloma odczytami to, nawet jeśli w danym miejscu wystąpi tranzycja wywołana deaminacją cytozyny, statystyczna szansa na to, że pojawi się na więcej niż jednym odczycie,

jest bardzo niska. Pozostałe sekwencje zweryfikują rzeczywisty genotyp w danym miejscu (patrz też ryc. 6b). Podejście to wymaga jednak bardzo dobrego zachowania DNA albo dużych nakładów finansowych na sekwencjonowanie;

- usuwać zdeaminowane do uracylu cytozyny odpowiednimi enzymami w trakcie tworzenia biblioteki genomowej. Tak potraktowane biblioteki pozbawione są charakterystycznych dla kopalnego DNA zmian, co utrudnia potwierdzenie autentyczności otrzymanego wyniku. Dlatego w takich przypadkach najczęściej tworzy się przynajmniej dwie biblioteki, jedną niemodyfikowaną, którą wykorzystuje się do szacowania stopnia kontaminacji, i jedną modyfikowaną, z której można wykorzystać wszystkie polimorfizmy w trakcie analiz statystycznych;
- ucinając końce sekwencji przed genotypowaniem. Zmiany *post-mortem* znajdują się najczęściej przy i na końcach cząsteczek DNA, można więc, genotypując tranzycje, ignorować kilka (7 do 10) par zasad na końcach analizowanych sekwencji;
- uznawać i wykorzystywać tylko genotypy, które nie mogły powstać w wyniku zmian *post-mortem*. Czyli tylko C w polimorfizmie C/T i tylko G w polimorfizmie G/A.

ANALIZY STATYSTYCZNE

Zakładając, że mamy już nasze próby zgenotypowanie (względem wybranych danych referencyjnych), możemy przystąpić do analiz statystycznych służących do porównywania naszych prób między sobą oraz danymi referencyjnymi. Same analizy, które wybierzemy, będą zależały od pytań badawczych, na jakie chcemy odpowiedzieć. Na potrzeby tego rozdziału podzielimy je na: analizy pokrewieństwa, analizy populacyjne i analizy filogenetyczne.

ANALIZY POKREWIEŃSTWA

Analiz pokrewieństwa dokonuje się na danych pochodzących ze szczątków ludzkich w celu wyszukania pokrewieństwa genetycznego między porównywanymi próbami. Próby odpowiedzi na pytania dotyczące tego, czy i w jakim stopniu struktura przestrzenna pochówków odpowiada strukturze

pokrewieństwa pomiędzy osobami pochowanymi w danych obiektach, mają szerokie implikacje w rekonstrukcji struktury społecznej danej grupy lub populacji.

Zasady dziedziczenia alleli w rozmnażaniu płciowym (po jednym losowym allelu od każdego z rodziców) i wykorzystanie rachunku prawdopodobieństwa do określenia, czy dwie osoby mogą być z sobą spokrewnione, są metodą od dawna wykorzystywaną w badaniach genetycznych. Jednak w celu porównywania ze sobą danych pochodzących z sekwencjonowania wysokoprzepustowego kopalnego DNA, metody te należy odpowiednio zmodyfikować.

Czysto teoretycznie, w obecności całkowicie zrekonstruowanych genomów oraz odpowiednich danych referencyjnych, metody te są w stanie zrekonstruować pokrewieństwo nawet do 11 stopnia. Oczywiście w przypadku kopalnego DNA wynik ten jest nieosiągalny. Możliwe jest jednak z dużym prawdopodobieństwem wykrywanie pierwszego i drugiego stopnia nawet w przypadku słabego zachowania DNA w jednej z porównywanych prób przy paru zastrzeżeniach:

- częstość występowania poszczególnych alleli jest dla badanej populacji znana albo nie odbiega znacząco od częstości obecnych w populacjach współczesnych. Częstość alleli jest istotna, jako że dużo większą wagę ma fakt dzielenia tego samego allelu przez dwie osoby, jeśli w populacji występuje on rzadko);
- badana populacja charakteryzuje się niewielkim kojarzeniem krewniaczym lub jesteśmy w stanie obliczyć częstość tego zjawiska dla danej populacji.

Wymagane jest zmodyfikowanie metody, by uwzględniała niskie pokrycie i niekompletność danych. Stosowane w badaniach współczesnego DNA metody zakładają, że dysponujemy informacją o obu allelach, jakie posiadał każdy z porównywanych osobników w danym miejscu. Tymczasem w przypadku kopalnego DNA najczęściej dysponujemy tylko kilkoma odczytami dla każdego polimorfizmu, które rzadko kiedy reprezentują oba allele obecne w danym miejscu za życia. By wziąć na to poprawkę, można losowo zawsze wybierać tylko jeden allel obecny w danym miejscu (niezależnie, czy dysponujemy jednym czy dwoma) i rekompensować utratę informacji ilością analizowanych markerów.

Innym podejściem jest wykorzystanie haplotypów występujących lokalnie w genomie jądrowym,

zwanych czasami segmentami IBD (ang. *identical by descent*). Większość metod statystycznych zakłada, że polimorfizmy mogą być dziedziczone niezależnie, dzięki zjawisku rekombinacji. W rzeczywistości, jak pamiętamy, im bliżej siebie na genomie znajdują się dwa polimorfizmy, tym mniej jest to prawdopodobne. Sytuację tę, która jest problematyczna w standardowych metodach (i wymaga losowego wybierania jednego polimorfizmu z kilku sąsiadujących ze sobą), można wykorzystać zarówno w badaniach pokrewieństwa i populacyjnych szukając, kwantyfikując i mierząc segmenty IBD. Im bliżej ze sobą spokrewnione dwie osoby tym mniej będzie miejsc rekombinacji (których ilość kumuluje się wraz z następującymi pokoleniami) i tym dłuższe będą lokalne haplotypy. W celu szacowania pokrewieństwa nawet przy stosunkowo niewielkich ilościach danych można w ten sposób zweryfikować pokrewieństwo pierwszego i drugiego stopnia (Monroy Kuhn i in. 2018). Jak zwykle niekompletny charakter danych i zmiany *post-mortem* uniemożliwiają wykrycie dalszych stopni pokrewieństwa, choć stosowanie bardziej wyrafinowanych metod i genomów o wysokim pokryciu w niektórych przypadkach może pozwolić do wykrycia nawet piątego stopnia pokrewieństwa (Cassidy i in. 2020).

By jednak z dużym prawdopodobieństwem ocenić, czy dwie osoby były ze sobą spokrewnione w pierwszym lub drugim stopniu, szacuje się, że w zależności od metody wystarczy 5000-8000 nakładających się między porównywanymi próbami SNP, co jest wynikiem względnie łatwym do osiągnięcia. Nie oznacza to, że będziemy w stanie zweryfikować wszystkie pary osób w badanej grupie, np. nie będziemy w stanie porównać dwóch bardzo słabo zachowanych prób, ale możemy każdą z nich porównać z inną dobrze zachowaną, co także może dostarczyć nam informacji (lub zweryfikować niepewną informację) o ich wzajemnym pokrewieństwie.

By jednak dokładniej zweryfikować otrzymane wyniki i ustalić rodzaj pokrewieństwa danego stopnia, np. rozróżnić, czy pierwszy stopień oznacza rodzeństwo, czy rodzica i dziecko, niezbędne jest wykorzystanie danych dotyczących płci badanych osób, posiadanych przez nie haplogrup mtDNA i Y-DNA oraz ich wieku w chwili śmierci (ryc. 8b). Z uwagi zresztą na sposób dziedziczenia haplogrupy są w stanie całkowicie wykluczyć nam pokrewieństwo w linii matczynej.

ANALIZY POPULACYJNE

Analizy populacyjne mają odpowiadać na pytania, czy i w jakim stopniu badane populacje są ze sobą spokrewnione. Pomagają nam też weryfikować konkretne hipotezy oraz odpowiadać na pytania dotyczące charakteru przepływu genów między populacjami w przypadku jego wykrycia. Są to metody statystyczne, które służą do porównywania dużych ilości zmiennych, w celu np. grupowania i znalezienia alleli i polimorfizmów skorelowanych ze sobą. Przykładami takich analizy są analiza składowych głównych PCA (ang. *principal component analysis*) i skalowanie wielowymiarowe MDS (ang. *multidimensional scaling*) szeroko stosowane do analizowania wielu zmiennych, nie tylko w genetyce.

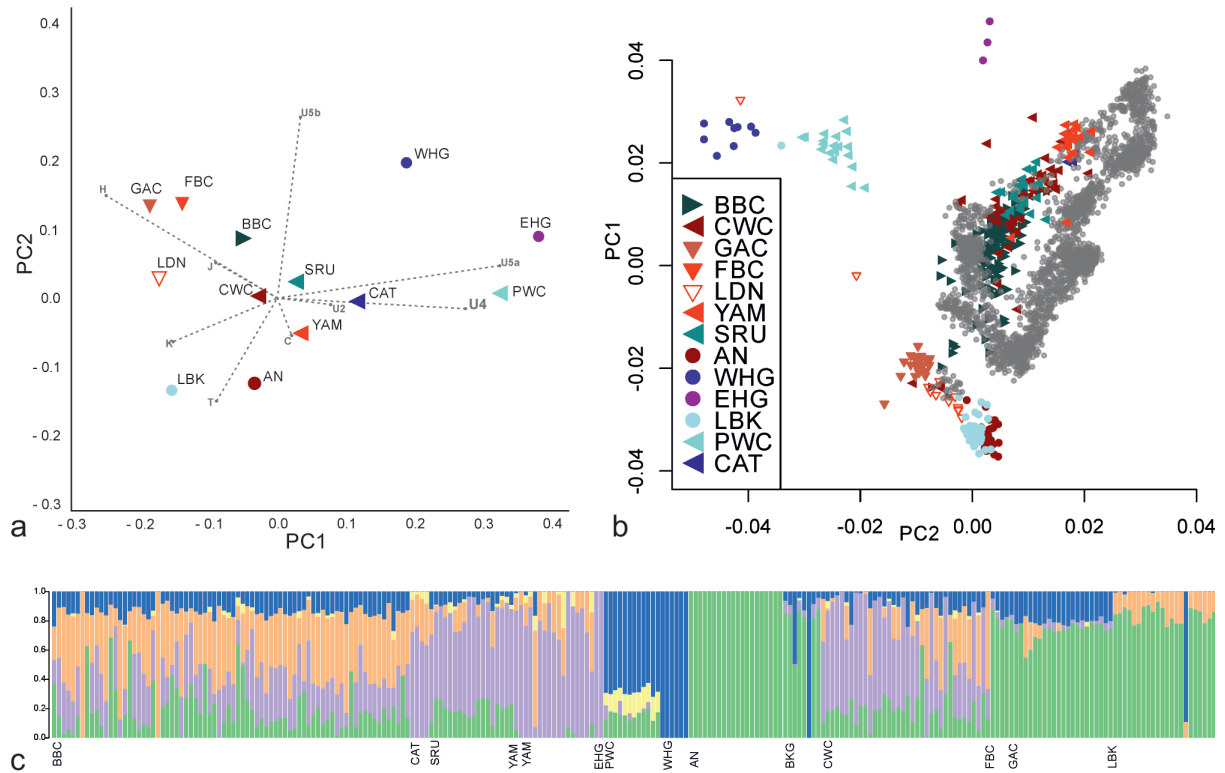
Najprostsze PCA można wykonać na danych zawierających informacje o częstości haplogrup. W takich wypadkach badanymi „obiektami” są populacje, a zmiennymi częstości występowania haplogrup w tych populacjach. Przy odpowiednim dobraniu prób relacje między populacjami powinny być zbliżone do tych obserwowanych w badaniach DNA jądrowego (patrz ryc. 7c i d). Jednak konieczność łączenia osobników w populację *a priori* sprawia, że analiza jest podatna na zaburzenia spowodowane małą liczbą osobników, obecnością osobników odstających (*outliers*) lub związane z dysproporcją płci w mieszających się populacjach (częstość haplogrup mtDNA się nie zmieni, jeśli elementem napływowym będą głównie mężczyźni).

Analizy genomów jądrowych, w których każdy osobnik jest badanym obiektem, a zmiennymi są polimorfizmy, ułatwia wykrycie takich błędów i prawidłowości. W rzeczywistości, z uwagi na niskie pokrycie genomu, rzadko kiedy jesteśmy w stanie porównywać kopalne próby bezpośrednio między sobą. Najczęściej wykorzystujemy w 100% zgenotypowane (dla danego zestawu SNP) próby współczesne, a następnie stosujemy zmodyfikowaną wersję oprogramowania do PCA, która porównuje nam osobno każdego osobnika badanego z danymi referencyjnymi, a następnie „uśrednia” uzyskane wartości, by odtworzyć relacje między badanymi próbami.

Dobór referencji współczesnej ma tu duże znaczenie. Wykorzystanie populacji genetycznie bardzo odległej (np. afrykańskich w badaniach populacji europejskich) może spowodować, że analiza nie wykaże różnicowania pomiędzy badanymi grupami, jako że wszystkie będą jednakowo (bardzo) odległe od populacji referencyjnej. PCA jest dobrym narzędziem do wykrycia istotnych różnic

między badanymi populacjami i – jak zostało wykazane – wynik analizy (w postaci wykresu punktowego) jest stały nawet przy bardzo małej (około 10 tysięcy) ilości SNP (Allentoft i in. 2015). PCA pomaga odpowiedzieć na pytanie, czy badane populacje są do siebie podobne, czy różne pod względem genetycznym. Z wzajemnego położenia tych populacji względem siebie możemy wysnuwać wstępnie interpretacje, ale w celu dokładniejszej rekonstrukcji zdarzeń wymagane są bardziej dokładne analizy.

Procesy demograficzne są zazwyczaj złożone i rzadko polegają na pełnym zastąpieniu jednej populacji inną. Najczęściej, gdy dochodzi do zetknięcia się dwóch populacji, może dojść do ich mieszania się w różnych proporcjach. Do kwantyfikacji takich procesów służą narzędzia, takie jak ADMIXTURE, statystyki *f*, *qpAdm* i *qpGraph* (ryc. 8a). W dużym uproszczeniu metody te służą do określenia, jakie komponenty genetyczne i w jakich proporcjach są obecne w badanych populacjach. Czyli jakie inne populacje i w jakich proporcjach się połączyły, gdy powstała populacja, którą badamy. Niektóre z tych metod, np. *Admixture* mogą wymagać wykorzystania (tak jak w przypadku PCA) populacji współczesnych, inne, np. *qpAdm*, wręcz przeciwnie – wymagają dodania populacji genetycznie odległych jako grup obcych (*outgroups*), względem których nasze badane populacje będą porównywane. Czasami możemy narzucić potencjalne populacje źródłowe, a nawet uzyskać wyliczenie prawdopodobieństwa, z jakim proponowany przez nas model rzeczywiście opisuje obserwowane różnicowanie genetyczne. Przy użyciu takich metod udało się ustalić np. trzy główne populacje źródłowe odpowiedzialne za genetyczne różnicowanie populacji europejskich: mezolitycznych łowców zbieraczy, anatolijskich rolników i późnoneolityczny grupy ze Stepu Pontyjsko-Kaspijskiego utożsamiane z kulturą grobów jamowych (Allentoft i in. 2015; Haak i in. 2015). Obecnie tych właśnie populacji używa się jako równoznacznych z trzema głównymi komponentami genetycznymi wyszukiwanymi w analizach populacyjnych. Bardziej dokładne analizy i rosnąca liczba prób umożliwiły rekonstrukcje bardziej złożonej struktury komponentów genetycznych, np. rozróżnienia komponentów łowiecko-zbierackich wschodniego i zachodniego (Fu i in. 2016) czy rozbieżności komponentu stepowego na wschodni łowiecko-zbieracki, wczesnorolniczy irański i zachodniosyberyjski (Narasimhan i in. 2019). Metody te wymagają



Ryc. 7. Przykłady wyników analiz populacyjnych dla kilku populacji pradziejowych; a – wykres obrazujący wynik analizy PCA dokonanej na częstościach haplogrup mtDNA (dodatkowo na szaro zaznaczone są haplogrupy mające znaczenie dla obserwowanego wyniku); b – wynik PCA dla SNP genomów jądrowych osób reprezentujących te same populacje. Na szaro zaznaczone genomy współczesne, względem których była dokonywana analiza; c – wynik analizy Admixture dla tych samych osób ukazujący główne komponenty genetyczne obecne w analizowanych genomach

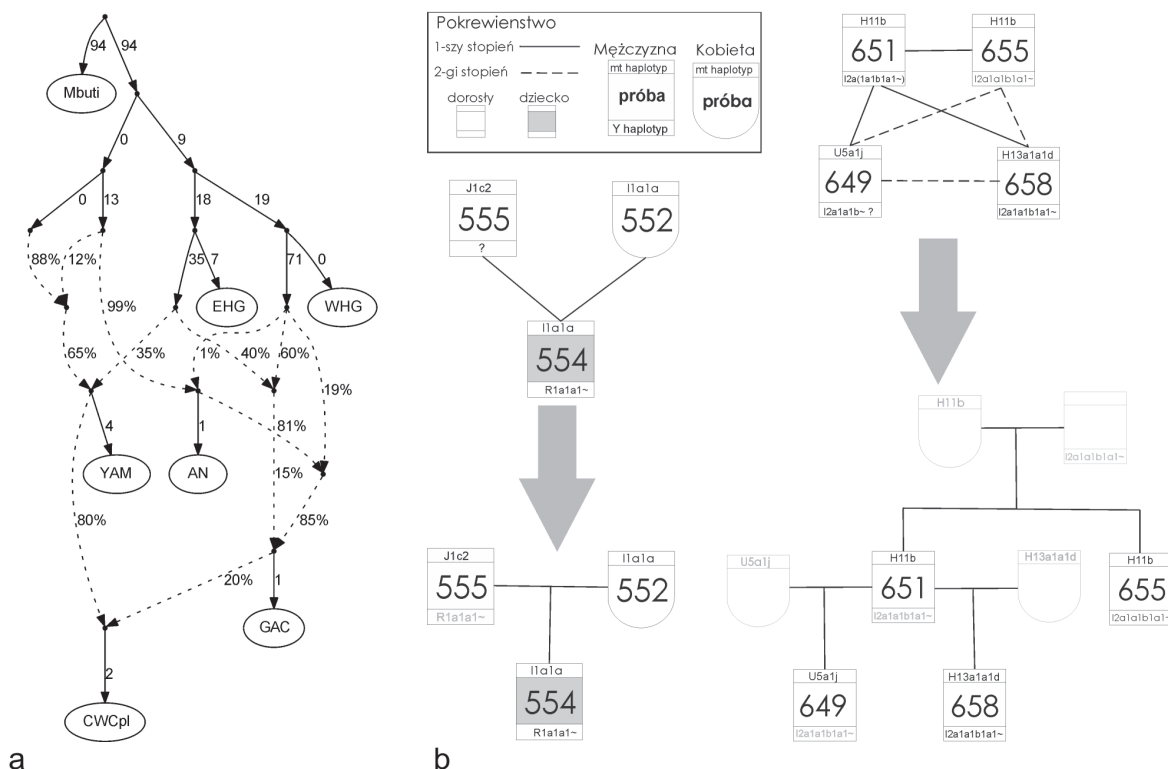
jednak większej ilości danych niż PCA, w postaci pokrycia genomu przynajmniej 0,1-0,2×.

Modelowania za pomocą qpAdm można użyć także do próby identyfikacji, które z potencjalnych populacji mogła być źródłem danego komponentu genetycznego. Na przykład w celu pokazania, że populacje stepowe związane z kulturą pucharów dzwonowatych mieszały się z lokalnymi społecznościami rolniczymi wraz z rozprzestrzenianiem się, a nie rozprzestrzeniały się już po zmieszaniu w miejscu swojej genezy (Olalde i in. 2018).

Sama identyfikacja komponentów genetycznych i ich źródła potrafi weryfikować wiele hipotez, zwłaszcza jeśli komponenty genetyczne są różne w mieszających się populacjach, ale interpretacja wyników może być problematyczna w sytuacji, gdy komponenty genetyczne w badanych populacjach są takie same. Z taką sytuacją mamy zawsze do czynienia w okresach pomiędzy wielkimi migracjami, na przykład w Europie od wczesnej epoki brązu, kiedy to na kontynent został wprowadzony ostatni znaczący

komponent genetyczny, mający swoje źródło na stepie. Zasadniczo od tego momentu wykrycie przepływu genów między „zhomogenizowanymi” populacjami europejskimi wymaga użycia dokładniejszych metod analizy. W takich sytuacjach na pomoc mogą przyjść, tak jak w przypadku badań pokrewieństwa, metody oparte na segmentach IBD, które z reguły powinny być dłuższe w populacjach ze sobą spokrewnionych (Cassidy i in. 2020; Margaryan i in. 2020). Takie podejścia wymagają niestety prób o wysokim pokryciu genomów, choć dla redukcji kosztów można do tych analiz wybrać tylko kilka reprezentatywnych dla populacji osobników z najlepiej zachowanym DNA. Problemатyczne jest również ilość dostępnych danych referencyjnych, jako że większość publikowanych genomów charakteryzuje się niskim pokryciem.

Przy użyciu analiz populacyjnych można dodatkowo spróbować odpowiedzieć na pytania związane z charakterem zachodzących procesów demograficznych.



Ryc. 8. a – przykład wyniku analizy qpGraph służącej do weryfikacji złożonych modeli demograficznych zawierających wiele następujących po sobie wydarzeń mieszania się populacji wraz z zrekonstruowanymi proporcjami, w jakich te populacje się łączyły; b – przykłady rekonstrukcji struktury pokrewieństwa osób pochowanych w pochówkach zbiorowych z epoki brązu. Warto zwrócić uwagę, jak uzupełnienie wyniku analizy stopnia pokrewieństwa o informacje o wieku płci i haplogrupach umożliwia rekonstrukcję drzew genealogicznych

Porównanie zmieniającej się zmienności na chromosomie X z tą obserwowaną na pozostałych chromosomach pozwoliła stwierdzić, że w migracji rolników z Bliskiego Wschodu do Europy na początku neolitu w równej mierze brały udział kobiety i mężczyźni, natomiast za napływ stepowego komponentu genetycznego na przełomie neolitu i epoki brązu byli odpowiedzialni głównie (w stosunku 7:1) mężczyźni (Goldberg i in. 2017). Ponadto narzędzia pozwalające oszacować, kiedy dany komponent genetyczny został wprowadzony do badanej populacji, wykazały, że migracja ze stepu miała charakter ciągłego procesu, a nie pojedynczego epizodu (Furtwängler i in. 2020).

ANALIZY FILOGENETYCZNE

Analizami filogenetycznymi określam wszystkie metody, które mają za zadanie odtworzenie tych relacji między badanymi próbkami. W rzeczywistości

część z opisanych w poprzednim podrozdziale analiz jest metodami filogenetycznymi (np. qpAdm), tutaj jednak skupię się na zastosowaniach związanych z odtwarzaniem drzew filogenetycznych w celu zbadania pochodzenia i relacji między gatunkami oraz odmianami i szczepami w obrębie gatunków. Metody te wykorzystują stopniowe gromadzenie się mutacji w badanych genomach wraz z upływem czasu. Obliczając różnice genetyczne pomiędzy badanymi próbkami w postaci pojawiających się nowych polimorfizmów oraz znając tempo zachodzących zmian (w oparciu o modele opracowane o osobniki o znanym datowaniu) i datowania poszczególnych osobników, jesteśmy w stanie zwiualizować relacje pomiędzy nimi w postaci drzewa filogenetycznego.

Drzewa takie tworzy się np. by poznać pochodzenie udomowionych gatunków zwierząt i roślin (porównując je do ich dzikich odmian o znanej dystrybucji geograficznej) lub szczepy patogenów odpowiedzialnych za konkretne epidemie. Czysto teoretycznie analizy te można dokonać (i często się

dokonywać) jedynie na współczesnych materiałach. Jednakże, jak się okazuje, dysponowanie danymi kopalnymi umożliwia datowanie obserwowanych procesów oraz wykrywanie wymarłych i zapomnianych odmian i szczepów.

W przypadku zwierząt najczęściej to mitochondrialny genom jest wykorzystywany do tworzenia drzew filogenetycznych z uwagi na sposób dziedziczenia i relatywnie szybkie i stałe tempo pojawiania się nowych mutacji. Analogiczną funkcję w przypadku roślin pełni DNA pochodzący z chloroplastów, organelli komórkowych odpowiedzialnych za fotosyntezę, posiadających swój genom w wielu kopiach w każdej komórce. Dzięki analizom kopalnego DNA w badaniach filogenetycznych poznaliśmy potencjalne pochodzenie i drogi udomowienia np. psów (Frantz i in. 2016; Thalmann i in. 2013), kotów (Ottoni i in. 2017), świni (Larson i in. 2007) czy pszenicy (Scott i in. 2019). Badania archeogenetyczne pozwoliły nam też zrozumieć, w jaki sposób palczki dzumy ewoluowały i wywoływały kolejne znane z historii pandemie (Rascovan i in. 2019; Rasmussen i in. 2015) oraz dostarczyły danych na temat pochodzenia i historii wywołującego kiłę krętka białego w Europie (Majander i in. 2020).

PODSUMOWANIE I PERSPEKTYWY

Jeśli warto zapamiętać jedną rzecz o badaniach kopalnego DNA to powinna to być ta, że dziedzina ta rozwija się niezwykle dynamicznie, ilość danych, które generujemy i prób, które badamy, przyrastają z roku na rok wykładniczo. Dokonywane na tych danych analizy pozwalają weryfikować rozmaite hipotezy badawcze dotyczące bioróżnorodności, pokrewieństwa i afiliacji genetycznych badanych populacji. Wymaga to jednak umiejętności dostosowania formułowanych pytań do dostępności materiału badawczego zarówno pod względem jego stanu zachowania (i warunków depozycji), jak i reprezentatywności dla potencjalnie różnych biologicznie populacji.

Przyrastająca ilość danych i spadające koszty sekwencjonowania pozwalają na analizy w coraz lepszej rozdzielczości, lepiej odpowiadające regionalnej

złożoności obrazu obserwowanego w kulturze materialnej. Badania o bardziej regionalnym charakterze, skupiające się na konkretnej grupie lub nawet stanowisku (Mittnik i in. 2019; Schroeder i in. 2019) są jedną z potencjalnych ścieżek rozwoju dyscypliny.

Kolejną obiecującą tematyką są badania kopalnego DNA obecnego w warstwach archeologicznych. Mogą one potencjalnie dostarczyć danych genetycznych przy braku makroszczątków i to nie tylko o środowisku i gospodarce, ale też samych twórcach danej warstwy (Slon i in. 2017).

Jednym z ciekawszych, dopiero rozwijanych zagadnień są badania metylacji DNA w populacjach pradziejowych. Metylacje są to modyfikacje DNA, które wpływają na ekspresję genów, dezaktywując lub aktywując określone fragmenty genomu w odpowiedzi na stres biologiczny lub zmiany środowiskowe. Jako że w wyniku zmian *post-mortem* cytozyny deaminują do uracylu, a metylowane cytozyny bezpośrednio do tyminy, porównując częstość zmian C na T między normalnymi bibliotekami a tymi, w których uracyl został usunięty, jesteśmy w stanie powiedzieć, które części genomu były metylowane. Potencjalnie może to pomóc odpowiedzieć na pytanie, w jaki sposób istotne zmiany (np. w stylu życia u zarania neolitu) wpłynęły na ekspresję genów (Hanghøj i Orlando 2019).

Nadal istnieją oczywiście spore luki w naszej wiedzy na temat genetycznego aspektu istotnych z punktu widzenia archeologii przemian i procesów. Niektóre okresy, jak przełom epoki brązu i żelaza, na terenach, na których dominowało ciepłopalenie, chwilowo pozostają poza zasięgiem badań archeogenetycznych (choć i tu może kiedyś uda się uzyskać się nadające do analiz ilości DNA, jeśli nie ze skremowanych szczątków, to może z warstw na osadach). Są też okresy, jak środek epoki brązu, dla których wydawało się, że dane genetyczne (jako że początkowo zdawały się charakteryzować niskim zróżnicowaniem) nie będą w stanie wniesić wiele nowych informacji. Jednak wraz z rosnącą rozdzielczością danych i metod jesteśmy w stanie otrzymać bardziej zniuansowany obraz zróżnicowania genetycznego dla tych populacji.

Warto więc śledzić rozwój dziedziny, jaką są badania kopalnego DNA. Wraz z jej rozwojem wzrasta co prawda stopień skomplikowania metod, jakimi się posługuje, ale należy mieć nadzieję, że powstawanie zespołów i projektów interdyscyplinarnych oraz wzrastająca świadomość na temat realnych zastosowań badań kopalnego DNA ułatwią dialog i zrozumienie pomiędzy genetykami i archeologami.

BIBLIOGRAFIA

- Allentoft, M.E., Sikora, M., Sjogren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J. [...] Willerslev, E. 2015. Population genomics of Bronze Age Eurasia. *Nature* 522: 167–172.
- Briggs, L. 2020. Ancient DNA Research in Maritime and Underwater Archaeology: Pitfalls, Promise, and Future Directions. *Open Quaternary* 6(1): 3. <https://doi.org/10.5334/oq.71>
- Campos, P.F., Craig, O.E., Turner-Walker, G., Peacock, E., Willerslev, E., Gilbert, M.T.P. 2012. DNA in ancient bone – Where is it located and how should we extract it? *Annals of Anatomy - Anatomischer Anzeiger* 194(1): 7–16. <https://doi.org/10.1016/j.aanat.2011.07.003>
- Cassidy, L.M., Maoldúin, R.Ó., Kador, T., Lynch, A., Jones, C. [...] Bradley, D.G. 2020. A dynastic elite in monumental Neolithic society. *Nature* 582: 384–388. <https://doi.org/10.1038/s41586-020-2378-6>
- Chyleński, M., Ehler, E., Somel, M., Yaka, R., Krzewińska, M., [...] Juras, A., Marciniak, A. 2019. Ancient Mitochondrial Genomes Reveal the Absence of Maternal Kinship in the Burials of Çatalhöyük People and Their Genetic Affinities. *Genes* 10: 207. <https://doi.org/10.3390/genes10030207>
- Dabney, J., Meyer, M., Pääbo, S. 2013. Ancient DNA Damage. *Cold Spring Harbor Perspectives in Biology* 5: a012567. <https://doi.org/10.1101/cshperspect.a012567>
- Damgaard, P.B., Margaryan, A., Schroeder, H., Orlando, L., Willerslev, E., Allentoft, M.E. 2015. Improving access to endogenous DNA in ancient bones and teeth. *Scientific Reports* 5: 11184. <https://doi.org/10.1038/srep11184>
- Frantz, L.A.F., Mullin, V.E., Pionnier-Capitan, M., Lebrasseur, O., Ollivier, M., Perri, A., [...] Bradley, D.G., Larson, G. 2016. Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science* 352: 1228. <https://doi.org/10.1126/science.aaf3161>
- Fu, Q., Mittnik, A., Johnson, P.L.F., Bos, K., Lari, M., Bollongino, R., [...] Krause, J. 2013. A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes. *Current Biology* 23: 553–559. <https://doi.org/10.1016/j.cub.2013.02.044>
- Fu, Q., Posth, C., Hajdinjak, M., Petr, M., Mallick, S., Fernandes, D., [...] Krause, J., Pääbo, S., Reich, D. 2016. The genetic history of Ice Age Europe. *Nature* 534: 200.
- Furtwängler, A., Rohrlach, A.B., Lamnidis, T.C., Papac, L., Neumann, G.U., [...] Krause, J. 2020. Ancient genomes reveal social and genetic structure of Late Neolithic Switzerland. *Nature Communication* 11: 1915. <https://doi.org/10.1038/s41467-020-15560-x>
- Gamba, C., Jones, E.R., Teasdale, M.D., McLaughlin, R.L., Gonzalez-Fortes, G., [...] Bradley, D.G., Pinhasi, R. 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nature Communication* 5: 5257. <https://doi.org/10.1038/ncomms6257>
- Gilbert, M.T.P., Jenkins, D.L., Götherstrom, A., Naveran, N., Sanchez, J.J., Hofreiter, M., [...] Willerslev, E. 2008. DNA from Pre-Clovis Human Coprolites in Oregon, North America. *Science* 320: 786. <https://doi.org/10.1126/science.1154116>
- Goldberg, A., Günther, T., Rosenberg, N.A., Jakobsson, M. 2017. Ancient X chromosomes reveal contrasting sex bias in Neolithic and Bronze Age Eurasian migrations. *Proceedings of the National Academy of Sciences* 114(10): 2657–2662.
- Graham, E.A.M. 2007. DNA reviews: Ancient DNA. *Forensic Science, Medicine, and Pathology* 3: 221–225. <https://doi.org/10.1007/s12024-007-9009-5>
- Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., [...] Reich, D. 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522: 207–211.
- Hagelberg, E., Bell, L.S., Allen, T., Boyde, A., Jones, S.J., Clegg, J.B. 1991. Analysis of ancient bone DNA: techniques and applications. *Philosophical Transactions of the Royal Society of London B* 333: 399–407.
- Hanghøj, K., Orlando, L. 2019. Ancient Epigenomics, (w:) C. Lindqvist, O.P. Rajora (red.), *Paleogenomics: Genome-Scale Analysis of Ancient DNA*. Cham, 75–111.
- Hansen, H.B., Damgaard, P.B., Margaryan, A., Stenderup, J., Lynnerup, N., Willerslev, E., Allentoft, M.E. 2017. Comparing Ancient DNA Preservation in Petrous Bone and Tooth Cementum. *PLOS ONE* 12: e0170940. <https://doi.org/10.1371/journal.pone.0170940>
- Harbeck, M., Schleuder, R., Schneider, J., Wiechmann, I., Schmahl, W.W., Grupe, G. 2011. Research potential and limitations of trace analyses of cremated remains. *Forensic Science, Medicine, and Pathology* 204: 191–200. <https://doi.org/10.1016/j.forsciint.2010.06.004>
- Higuchi, R., Bowman, B., Freiburger, M., Ryder, O.A., Wilson, A.C. 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312: 282–284. <https://doi.org/10.1038/312282a0>
- Jensen, T.Z.T., Niemann, J., Iversen, K.H., Fotakis, A.K., Gopalakrishnan, S., Vågene, Å.J., Pedersen, M.W., [...] Schroeder, H. 2019. A 5700 year-old human genome and oral microbiome from chewed birch pitch. *Nature*

- Communication 10: 5520. <https://doi.org/10.1038/s41467-019-13549-9>
- Juras, A., Chyleński, M., Krenz-Niedbala, M., Malmström, H., Ehler, E., [...] Jakobsson, M., Dabert, M. 2017. Investigating kinship of Neolithic post-LBK human remains from Krusza Zamkowa, Poland using ancient DNA. *Forensic Science International Genetics* 26: 30–39. <https://doi.org/10.1016/j.fsigen.2016.10.008>
- Juras, A., Dabert, M., Kushniarevich, A., Malmström, H., Raghavan, M., Kosicki, J.Z., Metspalu, E., Willerslev, E., Piontek, J. 2014. Ancient DNA reveals matrilineal continuity in present-day Poland over the last two millennia. *Plos ONE* 9(10): e110839. <https://doi.org/10.1371/journal.pone.0110839>
- Kashuba, N., Kirdök, E., Damlien, H., Manninen, M.A., Nordqvist, B., Persson, P., Götherström, A. 2019. Ancient DNA from mastics solidifies connection between material culture and genetics of mesolithic hunter-gatherers in Scandinavia. *Communications Biology* 2: 185. <https://doi.org/10.1038/s42003-019-0399-1>
- Larson, G., Albarella, U., Dobney, K., Rowley-Conwy, P., Schibler, J., Tresset, A., Vigne, J.-D., Edwards, C.J., Schlumbaum, A., Dinu, A. 2007. Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Sciences of the United States of America* 104: 15276–15281.
- Lever, M.A., Rogers, K.L., Lloyd, K.G., Overmann, J., Schink, B., Thauer, R.K., Hoehler, T.M., Jørgensen, B.B. 2015. Life under extreme energy limitation: a synthesis of laboratory- and field-based investigations. *FEMS Microbiology Reviews* 39: 688–728. <https://doi.org/10.1093/femsre/fuv020>
- Lindahl, T. 1993. Instability and decay of the primary structure of DNA. *Nature* 362: 709–715. <https://doi.org/10.1038/362709a0>
- Majander, K., Pfrengle, S., Kocher, A., Neukamm, J., du Plessis, L., Pla-Díaz, M., [...] Krause, J., Schuenemann, V.J. 2020. Ancient Bacterial Genomes Reveal a High Diversity of *Treponema pallidum* Strains in Early Modern Europe. *Current Biology* 30: 3788–3803.e10. <https://doi.org/10.1016/j.cub.2020.07.058>
- Margaryan, A., Lawson, D.J., Sikora, M., Racimo, F., Rasmussen, S., Moltke, I., Cassidy, L.M., Jørsboe, E. [...] Nielsen, R., Werge, T., Willerslev, E. 2020. Population genomics of the Viking world. *Nature* 585: 390–396. <https://doi.org/10.1038/s41586-020-2688-8>
- Miga, K.H., Koren, S., Rhie, A., Vollger, M.R., Gershman, A., Bzikadze, A., [...] Phillippy, A.M. 2020. Telomere-to-telomere assembly of a complete human X chromosome. *Nature* 585: 79–84. <https://doi.org/10.1038/s41586-020-2547-7>
- Mittnik, A., Massy, K., Knipper, C., Wittenborn, F., Friedrich, R., [...] Stockhammer, P.W., Krause, J. 2019. Kinship-based social inequality in Bronze Age Europe. *Science* 366: 731. <https://doi.org/10.1126/science.aax6219>
- Mnich, B., Spinek, A.E., Chyleński, M., Sommerfeld, A., Dabert, M., Juras, A., Szostek, K. 2018. Analysis of LCT-13910 genotypes and bone mineral density in ancient skeletal materials. *PLOS ONE* 13: e0194966. <https://doi.org/10.1371/journal.pone.0194966>
- Monroy Kuhn, J.M., Jakobsson, M., Günther, T. 2018. Estimating genetic kin relationships in prehistoric populations. *PLOS ONE* 13: e0195491. <https://doi.org/10.1371/journal.pone.0195491>
- Narasimhan, V.M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., Mallick, S., Lazaridis, I. [...] Frachetti, M., Pinhasi, R., Reich, D. 2019. The formation of human populations in South and Central Asia. *Science* 365: eaat7487. <https://doi.org/10.1126/science.aat7487>
- Olalde, I., Brace, S., Allentoft, M.E., Armit, I., Kristiansen, K., [...] Reich, D. 2018. The Beaker phenomenon and the genomic transformation of northwest Europe. *Nature* 555: 190–196. <https://doi.org/10.1038/nature25738>
- Oskam, C.L., Haile, J., McLay, E., Rigby, P., Allentoft, M.E., [...] Bunce, M. 2010. Fossil avian eggshell preserves ancient DNA. *Proceedings of the Royal Society B Biological Sciences* 277: 1991–2000. <https://doi.org/10.1098/rspb.2009.2019>
- Ottoni, C., Van Neer, W., De Cupere, B., Daligault, J., Guimaraes, S. [...] Grange, T., Geigl, E.-M. 2017. The palaeogenetics of cat dispersal in the ancient world. *Nature Ecology & Evolution* 1: 0139. <https://doi.org/10.1038/s41559-017-0139>
- Pinhasi, R., Fernandes, D., Sirak, K., Novak, M., Connell, S. [...] H., Hofreiter, M. 2015. Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. *PLOS ONE* 10: e0129102. <https://doi.org/10.1371/journal.pone.0129102>
- Poinar, H.N., Cooper, A. 2000. Ancient DNA: do it right or not at all. *Science* 5482: 416.
- Rascovan, N., Sjögren, K.-G., Kristiansen, K., Nielsen, R., Willerslev, E., Desnues, C., Rasmussen, S. 2019. Emergence and Spread of Basal Lineages of *Yersinia pestis* during the Neolithic Decline. *Cell* 176: 295–305.e10. <https://doi.org/10.1016/j.cell.2018.11.005>
- Rasmussen, M., Guo, X., Wang, Y., Lohmueller, K.E., Rasmussen, S., [...] Willerslev, E. 2011. An Aboriginal Australian Genome Reveals Separate Human Dispersals into Asia. *Science* 334: 94–98. <https://doi.org/10.1126/science.1211177>
- Rasmussen, M., Li, Y., Lindgreen, S., Pedersen, J.S., Albrechtsen, A. [...] Willerslev, E. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463: 757–762. <https://doi.org/10.1038/nature08835>
- Rasmussen, S., Allentoft, M.E., Nielsen, K., Orlando, L., Sikora, M. [...] Willerslev, E. 2015. Early Divergent Strains of *Yersinia pestis* in Eurasia 5,000 Years Ago. *Cell* 163: 571–582. <https://doi.org/10.1016/j.cell.2015.10.009>
- Renaud, G., Slon, V., Duggan, A.T., Kelso, J. 2015. Schmutzi: estimation of contamination and endogenous

- mitochondrial consensus calling for ancient DNA. *Genome Biology* 16: 224. <https://doi.org/10.1186/s13059-015-0776-0>
- Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., Pääbo, S. 2012. Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. *PLOS ONE* 7: e34131. <https://doi.org/10.1371/journal.pone.0034131>
- Schroeder, H., Margaryan, A., Szmyt, M., Theulot, B., Włodarczak, P., [...] Allentoft, M.E. 2019. Unraveling ancestry, kinship, and violence in a Late Neolithic mass grave. *Proceedings of the National Academy of Sciences* 116: 10705. <https://doi.org/10.1073/pnas.1820210116>
- Scott, M.F., Botigué, L.R., Brace, S., Stevens, C.J., Mullin, V.E., Stevenson, A., Thomas, M.G., Fuller, D.Q., Mott, R. 2019. A 3,000-year-old Egyptian emmer wheat genome reveals dispersal and domestication history. *Nature Plants* 5: 1120–1128. <https://doi.org/10.1038/s41477-019-0534-5>
- Slon, V., Hopfe, C., Weiß, C.L., Mafessoni, F., de la Rasilla, M. [...] Meyer, M. 2017. Neandertal and Denisovan DNA from Pleistocene sediments. *Science* 356: 605. <https://doi.org/10.1126/science.aam9695>
- Smith, O., Momber, G., Bates, R., Garwood, P., Fitch, S., Palen, M., Gaffney, V., Allaby, R.G. 2015. Sedimentary DNA from a submerged site reveals wheat in the British Isles 8000 years ago. *Science* 347: 998. <https://doi.org/10.1126/science.1261278>
- Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V.J., Sawyer, S.K. [...] Wayne, R.K. 2013. Complete Mitochondrial Genomes of Ancient Canids Suggest a European Origin of Domestic Dogs. *Science* 342: 871. <https://doi.org/10.1126/science.1243650>
- van der Valk, T., Pečnerová, P., Díez-del-Molino, D., Bergström, A., Oppenheimer, J. [...] Dalén, L. 2021. Million-year-old DNA sheds light on the genomic history of mammoths. *Nature* 591: 265–269. <https://doi.org/10.1038/s41586-021-03224-9>
- Welker, F., Ramos-Madrigal, J., Gutenbrunner, P., Mackie, M., Tiwary, S. [...], C., Bermúdez de Castro, J.M., Willerslev, E., Cappellini, E. 2020. The dental proteome of Homo antecessor. *Nature* 580: 235–238. <https://doi.org/10.1038/s41586-020-2153-8>
- Weyrich, L.S., Dobney, K., Cooper, A. 2015. Ancient DNA analysis of dental calculus. Special Issue Ancient DNA. *Journal of Human Evolution* 79: 119–124. <https://doi.org/10.1016/j.jhevol.2014.06.018>
- Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M.E., Lorenzen, E.D., Vestergård, M., Gusarova, G., Haile, J., Craine, J., Gielly, L., Boessenkool, S., Epp, L.S. [...] Brochmann, C., Taberlet, P. 2014. Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature* 506: 47–51. <https://doi.org/10.1038/nature12921>

