

Szegedi Tudományegyetem

Természettudományi és Informatikai kar

Genetika Tanszék

37. OTDK Biológia Szekció helyi TDK forduló

**Dél-szibériai bronzkori kultúrák
archeogenetikai vizsgálata**

Csikós Gellért

Molekuláris Bionika Mérnök BSc

4. évfolyam

Témavezetők:

Dr. Török Tibor

Schütz Oszkár

egyetemi docens

PhD hallgató

Szeged

Tartalomjegyzék

Tartalmi összefoglaló	2
Irodalmi áttekintés	3
Az archaikus DNS (aDNS) vizsgálata	3
Labortervezés	3
Az Irmen és Karasuk kultúrák történeti háttere	6
Az Irmen kultúra	6
A Karasuk kultúra.....	7
Sargat-Gorokhovo horizont.....	8
Célkitűzés	9
Alkalmazott módszerek, felhasznált anyagok és eszközök.....	10
Felhasznált minták.....	10
aDNS extrakció	10
Könyvtárkészítés	11
Szekvenálás	14
PCA	15
Kromoszomális (genetikai) nem meghatározása.....	17
F-statisztika	18
Eredmények.....	20
Szekvenált adatok.....	20
PCA	22
F4-statisztika	25
Diskusszió.....	27
Köszönetnyilvánítás	28
Irodalomjegyzék.....	29

Tartalmi összefoglaló

Az uráli nyelvcsalád eredete továbbra is vitatott kérdés a nyelvészeti szakirodalomban. Egy nemrégiben megjelent publikációban (Grünthal és mtsai., 2022) azonban – a történettudomány, a nyelvészet és a régészet eszköztárát kombinálva – az uráli nyelvek elterjedését a korai bronzkorban az Altáj vidékéről nyugat felé vándorló Seima-Turbino kulturális horizonttal hozzák összefüggésbe.

A nyelvi adatok alapján az uráli ősnyelv a bronzkor végére gyorsan szétválhatott alágaira, amit valószínűleg az okozott, hogy az egyes csoportok a vándorlás során elkülönültek egymástól. Az Urál keleti előterét az ugor alágakhoz tartozó csoportok vehették birtokba, mivel ezek nagy része ma is ott él. A kora bronzkori Seima-Turbino vándorlást követően a késő bronzkorban ezen a területen az Irmen és Karasuk régészeti kultúrákat találjuk, majd a vaskorban a Sargat és Gorokhovo kultúrákat. Feltételezhetően ezek mindegyike már valamilyen ugor nyelven beszélhetett, és a honfoglaló magyarok őseit is ezek között kereshetjük. (Maróti és mtsai., 2022)

Az Irmen és Karasuk kultúrák embertani leletanyagának archeogenetikai vizsgálata lehetőséget nyújt ezek egymással, a vaskori Sargat-Gorokhovo kultúrákkal, valamint a kutatócsoportunkban korábban elemzett korai honfoglalókkal való összehasonlítására. Ezáltal pedig betekintést nyerhetünk az említett népeségek genetikai kapcsolataiba, valamint az uráli nyelvet beszélő csoportok közötti kapcsolati rendszerbe is.

Irodalmi áttekintés

Az archaikus DNS (aDNS) vizsgálata

Az archeogenetika alapkövének tekinthető Higuchi és munkatársai 1984-ben megjelent tanulmánya, melyben arról számoltak be, hogy egy múzeumban kiállított, a lovak családjába tartozó quagga (*Quagga quagga*) izomszövetéből sikeresen tudtak DNS-t kinyerni és szekvenálni (Higuchi és mtsai., 1984). Az újgenerációs szekvenálás felfedezése előtt nagyon hosszú és drága volt egy teljes genom szekvenálása, ezért akkoriban főként a rövidebb mitokondriális DNS-re (mtDNS), valamint a genomban található STR (short tandem repeat) régiókra összpontosítottak. A mtDNS több szempontból is megfelelő választásnak bizonyult, mert uniparentálisan, anyai úton öröklődik, valamint nagy kópiaszámban van jelen a sejtekben, így jobban megőrződik az archaikus mintákban. Az mtDNS adatokból is következtethetünk a vizsgált populációk földrajzi származására, genetikai keveredésére, valamint vándorlásaik útvonalaira. Mindazonáltal, az újgenerációs szekvenálás megjelenése az utóbbi évtizedekben a tudományág rohamos fejlődését tette lehetővé, mivel lehetőséget biztosít teljes archaikus genomok meghatározására, valamint az exogén eredetű szennyeződések kiszűrésére.

Labortervezés

Az aDNS-minták koruk és a földbe temetkezés gyakorlata miatt DNS-módosulásokat és exogén szennyeződések tartalmazzak, melyek hibás analízishez vezethetnek. Ennek elkerülésére a tudományos közösség szigorú szabályrendszert alakított ki a laborok működésére és a kezelt DNS-minták feldolgozására, tárolására vonatkozóan.

A legnagyobb veszélyt a modern DNS-szennyeződés jelenti. Az aDNS-mintát a régészeti feltárástól a DNS-kinyerésig tartó folyamatok során a feldolgozást végző személyek beszennyezhetik saját DNS-ükkel, a szennyező modern emberi DNS-szekvenciákat pedig nagyon nehéz elkülöníteni az aDNS-minta szekvenciáitól. A modern szennyezés steril laborkörnyezettel kerülhető el legjobban.

2000-ben Cooper és Poinar, majd 2013-ban Barnes és Shapiro jelentetett meg egy-egy értekezést az archaikus laborok optimális kialakításáról, hogy a munkafolyamat során minimalizálható legyen a modern DNS-szennyezés (Barnes és mtsai., 2013; Cooper & Poinar,

2000). Az SZTE Genetika Tanszék archeogenetikai kutatócsoportja a fenti cikkek felhasználásával dolgozta ki az alábbi protokollt:

Az aDNS kivonási és feldolgozási helyének elkülönítése: A kivonás során precíz, elővigyázatos munkával nagymértékben csökkenthetők a modern emberi DNS-szennyezések. Mindemellett az adatok analízisa során is van lehetőség elkülöníteni az aDNS-t a külső modern szennyező DNS-től megfelelő módszerek segítségével. Amennyiben több minta feldolgozása párhuzamosan fut, a keresztszennyezés problémája is felmerülhet. Ez mindenképpen kerülendő, ezért minden aDNS-minta feldolgozása után DNS-mentesíteni kell a munkaterületet, valamint a kivonáshoz használt eszközöket. Az aDNS-labort el kell különíteni az amplifikált aDNS-t feldolgozó molekuláris labortól, utóbbiból tilos közvetlenül visszamenni az előbbibe.

Negatív kontroll: Az aDNS-kivonás lépéseit követve mindig kell egy DNS-mentes negatív kontrollt is készíteni, amit az aDNS-t tartalmazó mintákkal együtt kell kezelni. A negatív kontroll segítségével megállapítható, hogy a folyamat mely részében kerülhetett szennyező modern DNS az aDNS-mintákba.

Megfelelő higiénia és óvintézkedések: Egy archaikus laboratórium létrehozása nagy precizitást igényel, jól meg kell tervezni, hogy a benne zajló munka során minimalizálni lehessen az exogén DNS-szennyezést és minél hatékonyabban lehessen haladni az aDNS-minták feldolgozásával. A labornak feltétlenül egy elkülönített helyen kell elhelyezkednie, az ott dolgozó személyeknek csak speciális steril ruházatban szabad belépniük. A laborban az aDNS-minták, fülkék és eszközök UV-kezelésen (260nm) esnek át, így a következő belépésnél sterilizált környezetben folytatódhat a feldolgozás. Emellett minden használat után az eszközöket hipóval és 3%-os hidrogén-peroxidos oldattal is meg kell tisztítani, ezzel biztosan megszabadulva a külső szennyeződésektől. Mivel a laboron kívülről beszálló levegő is hozhat magával szennyeződéseket, az aDNS-labornak steril légbefújással kell rendelkeznie, továbbá kis mértékű túlnyomást kell létrehozni benne, hogy elkerülhető legyen a levegő kintről beáramlása. Az aDNS-laborban a dupla kesztyű használata mellett a kesztyűk gyakori cseréje is ajánlott.

Reprodukálhatóság: Az eredményeknek laborfüggetlennek, azaz más helyszínen is reprodukálhatónak kell lenniük. A független egyezés igazolja az aDNS szekvenciájának szennyeződésmentes helyes sorrendjét. Ennek jelentőségét az újgenerációs szekvenálási

technológiák használata valamelyest csökkentette, mivel a vizsgált minták eredetisége szekvenciaanalízissel is igazolható.

aDNS minősége: Az aDNS-minták korukból adódóan rövidebb szekvenciákat tartalmaznak, fragmentáltságuk mértéke több környezeti jellemzőtől is függ: hőmérséklet, pH, páratartalom. Minél idősebb a minta, annál rosszabb a DNS megőrződése, ezért, ha az elemzett minta több helyen is jó minőségű, hosszabb szekvenciát tartalmaz, azok valószínűsíthetően külső szennyeződésből származnak.

Az Irmen és Karasuk kultúrák történeti háttere

Az Irmen és Karasuk kultúrák a késői bronzkor két jelentős régészeti kultúrája, amelyek Délnyugat-Szibéria területén alakultak ki (1. ábra). Mindkét népcsoport alapvető szerepet játszott a térség kulturális fejlődésének alakításában.



1. ábra: Az Irmen és Karasuk kultúrák földrajzi elhelyezkedése (Török, 2023)

Az Irmen kultúra

Az Irmen kultúra Nyugat-Szibéria erdős sztyeppe régiójában virágzott a késői bronzkor idején (i.e. 11–8. század). Központi területének az Ob-folyó felső szakasza tekinthető, ahol számos településüket és temetkezési helyüket tárták fel, ennek köszönhetően egyike azon nomád népeknek, melyek temetkezési szokásairól sokat tudunk ebből az időszakból.

A régészek olyan irmeni kurgánokat, halomsírokat tártak fel, amelyek több sekély mélységű (20–25 cm mélységben elhelyezkedő) sírt tartalmaztak. Az elhunytakat jobb oldali, zsugorított testhelyzetben temették el úgy, hogy a fejüket déli irányban helyezték a sírba. Ez a temetkezési szokás a korábbi bronzkori hagyományokat tükrözi. Azonban a Tomsk-Ob régióban már a délnyugati tájolású temetkezés dominált, illetve hamvasztásra is volt példa, amely során az elhunytat részlegesen vagy teljesen elégették. A sírmellékletek között legtöbbször kerámiák, bronz ékszerek (fülbevalók, medálok), valamint esetenként kések és lándzsahegyek szerepeltek (N. L. Chlenova és mtsai., 1970).

Az Irmén kultúra része volt a Mezhevskaya-Irmén horizontnak, amely több, egymással összefüggő kultúrát foglalt magába a Nyugat-Uráltól az Altajig. Ezen a területen kulturális keveredés ment végbe az északkeletről érkező Karasuk kultúra hatásának köszönhetően.

A késői bronzkorban a gazdasági és technológiai fejlődésnek köszönhetően Nyugat-Szibériában dél és kelet felé népmozgás indult el, mely a környező kultúrák között feszültséget eredményezhetett. Emiatt jelenhettek meg az első árkokkal és bástyákkal megerősített települések (M.F. Kosarev és mtsai., 1966).

Az Irmén kultúrára jelentős hatást gyakorolt a Karasuk kultúra, hiszen az anyagi kultúrájában, különösen az ötvösművészeti tárgyak – például jellegzetes hajlított török és balták – stílusában megfigyelhetők a Karasuk jellemvonások. A bronzkor végével és a vaskor kezdetével az Irmén kultúra hozzájárult a később megjelenő Sargat kultúra kialakulásához, amely a vaskor elején Nyugat-Szibéria legjelentősebb kultúrájává vált. Ez az átmenet a regionális különbségek fokozódásával, valamint a meglévő és új hagyományok keveredésével járt. (Koryakova & Epimakhov, 2007)

A Karasuk kultúra

A Karasuk kultúra a késői bronzkorban (i.e. 1500–800) a dél-szibériai Minuszinszki-medencében alakult ki. Az Andronovo horizont egy késői leszármazottja, amely azonban jelentős külső hatások, különösen a Seima-Turbino horizont és a keleti – valószínűleg Kínához köthető – fazekas mesterségi tradíciók befolyása alatt fejlődött. Ezen befolyások is hozzájárultak a Karasuk kultúra sajátos fémművészeti stílusának kialakulásához, amely a különféle bronztárgyak – például hajlított kések, nyéllel rendelkező balták, kétélű török és dísz tárgyak – előállítására miatt vált híressé. (JETTMAR, 1970)

A Karasuk fémmegmunkálás jellegzetességei számos területen előkerültek, befolyásuk egészen az Urál-hegység nyugati részéig ért el. A Karasuk kultúra nyugati terjeszkedését jelzik az Urál déli részén felfedezett karakterisztikus hajlított bronztörök. (Chernykh, 2008) Az Irmén és Mezhevskaya kultúrákra gyakorolt hatása emellett a temetkezési tradíciókban, kerámiák és bronz tárgyak formájában mutatkozik meg, mely jellemzők az Ob-folyó felső szakaszán is megfigyelhetők, ahol a fentebb említett Irmén kultúra központja helyezkedett el.

A Karasuk temetkezés egyik érdekessége az állati maradványok jelenléte, melyek mennyisége összefüggésben állt az eltemetett egyén életkorával és nemével. A csecsemőt vagy

fiatalított tartalmazó sírokban egy, míg felnőttek esetében 2–3 állatáldozat volt megfigyelhető nemtől függően. Leggyakrabban bárány-, ritkábban szarvasmarha-, esetenként pedig lómaradványokat találtak a régészek. Az általában négyszögletes sírok az Irmen temetkezéshez hasonlóan a felszínhez közel helyezkedtek el. Az elhunytat a legtöbb esetben bal oldalra fektetve, a lábait kissé felhúzva helyezték a sírba, azonban a háton fekvő, nyújtott pozícióra is volt példa. Ez a temetkezési szokás a Karasuk kultúra után is folytatódott. Az elhunyt fejéhez közel egy kerámiaedényt helyeztek el. Emellett a női sírok általában kisebb ékszereket, hajdíszeket és késeket, míg a férfi sírok egyszerűbb késeket, fülbevalókat, ritkábban gyűrűket tartalmaztak (2. ábra). (Legrand, 2006)



2. ábra: Karasuk kések

Forrás: <https://www.shutterstock.com/hu/image-photo/bronze-age-knives-karasuksky-culture-south-1220681254>

Sargat-Gorokhovo horizont

A Sargat-Gorokhovo horizont a kora vaskorban (i.e. 5–3. század) volt meghatározó a Dél-Uráltól egészen az Irtis folyóig. Ezen két népcsoport biztosan keveredett egymással, a leletek intenzív kereskedelemről és vasmegmunkálásról tanúskodnak. Az ebből az időszakból feltárt településeket jellemző módon megerősítették, vélhetően a környező népekkel kialakult feszültségek miatt.

Célkitűzés

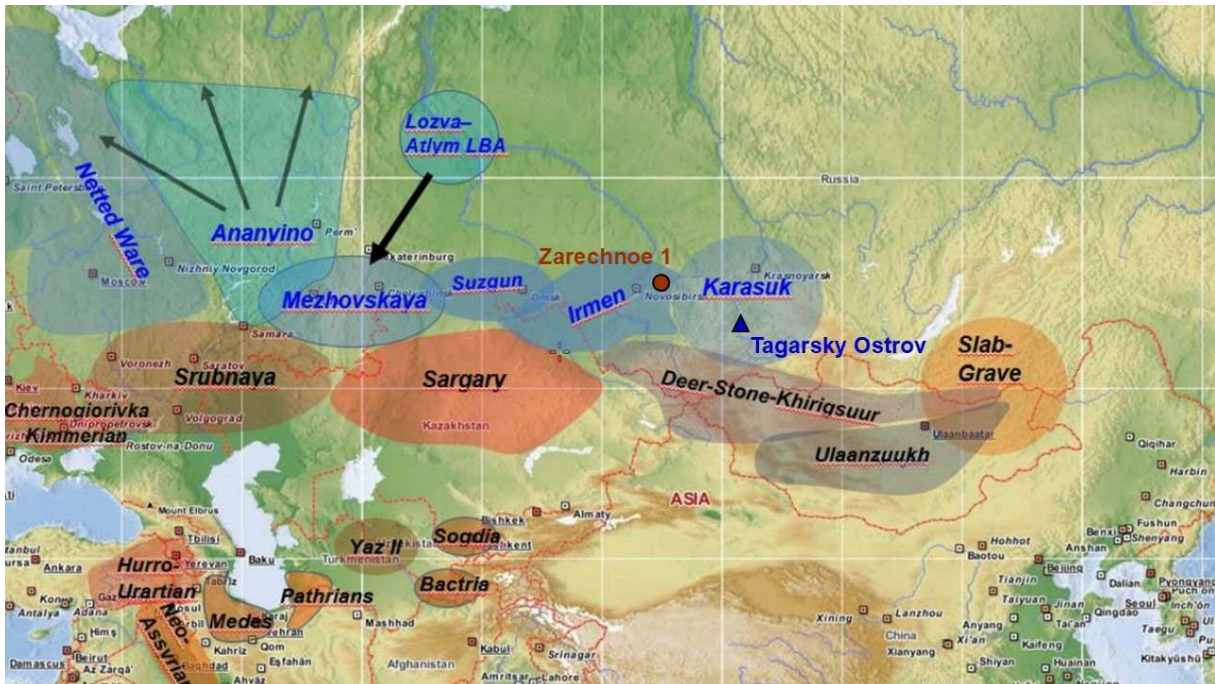
Jelen dolgozatban az Irmen és Karasuk kultúrkörbe tartozó temetkezések embertani leletanyagának archeogenetikai feldolgozását tűzzük ki célul. A teljes genomok kinyerését követően a szakirodalomban fellelhető legfrissebb genomelemző módszerek segítségével kívánjuk megvizsgálni a kultúrák embertani leletanyagának genetikai összetételét és a vaskori Sargat-Gorokhovo populációval mutatott esetleges kapcsolatait.

Az Irmen és Karasuk kultúrák egyaránt fontos szerepet játszottak a bronzkor és a vaskor közötti átmenet során Szibériában. A Karasuk kultúra a korszak technológiai fejlődésének kulcsfontosságú szereplője volt, míg az Irmen kultúra Nyugat-Szibéria kulturális fejlődését formálta, mely kihatással volt a későbbi Sargat kultúrára. Az említett három népcsoport között egyértelmű régészeti kapcsolat van, a megválaszolandó kérdés az, hogy volt-e genetikai kapcsolat is ezen kultúrák között és ha igen, akkor az milyen mértékű lehetett. A kutatásunk során ezen kérdésre kerestük a választ (Koryakova & Epimakhov, 2007).

Mint említettük, a Karasuk, Irmen és Sargat-Gorokhovo populációkat azért választottuk ki analízisünkhöz, mert a rendelkezésre álló genetikai és nyelvészeti adatok alapján jó okunk van azt feltételezni, hogy genetikailag hozzájárulhattak a honfoglaló magyarok kialakulásához, így ezen kultúrák genetikai vizsgálata hozzájárulhat a magyarok távoli őstörténetének tisztázásához.

Alkalmazott módszerek, felhasznált anyagok és eszközök

Felhasznált minták



3.ábra: Vizsgált minták, temetők elhelyezkedése

A minták a mai Oroszország területéről származnak. Az Irmen mintákat a Novoszibirszki területen a Zarechnoe 1 lelőhelyen tárták fel, míg a Karasuk maradványok a Minuszinszki medencében található Tagarsky Ostrov, lelőhelyről származnak.

Összesen 21 személy maradványait vizsgáltuk meg, 6 Irmen és 15 Karasuk kultúrába tartozó egyént. Az Irmen minták esetén 4, míg a Karasuk minták esetén 8 egyénből tudtunk sikeresen teljes genom szekvenciákat előállítani.

aDNS extrakció

Porítás: Az archeogenetikai vizsgálatok szempontjából az a szövet a legmegfelelőbb, amiben a legjobb minőségben és legnagyobb mennyiségben őrződik meg a DNS (Campos és mtsai., 2012). A kutatások kimutatták, hogy a rendkívül tömör, kemény csontszövetből lehet kinyerni a legmegfelelőbb aDNS-t, porítás útján. Mivel a halántékcsontról (*os temporale*) sziklacsonti része (*pars petrosa*) már újszülött korban eléri végső tömörségét, keménységét, így

ez a terület rendelkezik a legjobb DNS-megőrző képességgel. Ezen kívül elégséges minőségű és mennyiségű aDNS vonható ki a fogak gyökerét kívülről borító cementrétegből (*cementum*), valamint bizonyos esetekben a hosszú csöves csontok corticális állományából is.

Előemésztés: A csontminta porításával előállított csontporszemcsék felszíne még tartalmazhat szennyeződések, ami rövid EDTA és proteináz K előemésztéssel távolítható el.

Emésztés: Az előemésztett csontporszemcséket centrifugálással ülepitjük, majd a feloldott szennyeződések tartalmazó felülúszót eltávolítjuk. Ezt követően előre elkészített extrakciós pufferből 1,5ml-t mérünk a szemcsékre. A puffer megegyezik az előemésztő oldattal, de 1,35ml 0,5 M EDTA és 18,75μl 20mg/ml proteináz K emellett még 15μl 10%-os TritonX-100-at is tartalmaz, ami a csontporszemcsék eloszlatásában és szuszpenzióban tartásában játszik szerepet.

Szilika oszlophoz kötés: Az EDTA és Proteináz K teljesen feloldja a csontport, így az aDNS oldatba kerül. Az aDNS-oldathoz szilika szuszpenziót, valamint mintánként 3,58g tömény GuHCl-ot, 220μl 3M nátrium-acetátot, 3ml 100%-os izopropanolt és 3,75 μl Tween-20-at tartalmazó kötőpuffert adunk. Ebben a közegben az aDNS szelektíven kötődik a szilikához.

Mosás: A mosás DNS-mentes, 750μl 80%-os etil-alkohollal történik, melyet kétszer ismételünk, így eltávolítva a felesleges, ki nem kötődött szerves törmeléket.

DNS leoldás: A szilikára kikötött aDNS-t TE pufferrel (10 mM TRIS és 1mM EDTA elegye) oldjuk le, így egy tisztított aDNS-oldatot kapunk, ami szennyeződésektől mentes.

Koncentrációellenőrzés: Ehhez Qubit mérőeszközt alkalmazunk, amivel megállapítható, hogy van-e DNS a kivonatban. Az aDNS pontos mennyiségét Qubit Fluorometer segítségével határozzuk meg. A mérés fluorometrián alapul, mely során az oldaton keresztül átvilágított fény intenzitásából következtethetünk az oldatban található DNS koncentrációjára.

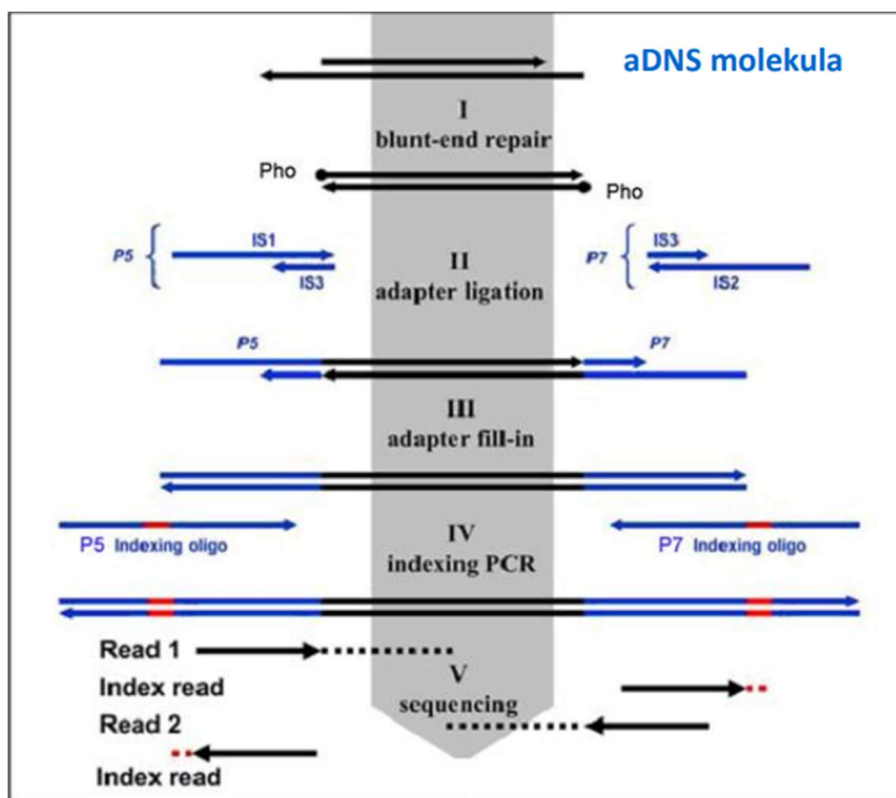
Könyvtárkészítés

Részleges UDG-kezelés: A régészeti embertani leletek DNS-szekvenciái az idő múlásával megváltoznak, mivel a talajban töltött hosszú idő folyamán a különböző fizikai, kémiai és biológiai hatások módosítják a szerkezetét. A legjelentősebb jól dokumentált változás

a citozin bázisok deaminációja uracillá. UDG enzim kezelésével a DNS egyik szálán az uracil pozíciók helyén abázikus helyek alakulnak ki, ezt a hiányt pedig a komplementer szálról javítjuk ki. Az UDG kezeléshez mintánként 6 μ l 10x-es Tango puffert, 0,24 μ l 25mM dNTP-t, 0,6 μ l 100mM ATP-t, 1,8 μ l 1 U/ μ l USER enzimet adtunk, majd vízzel 60 μ l-re egészítettük ki.

Blunt-end repair: A DNS szekvenciák túlnyúló 3' végeit mintánként 1,2 μ l 5U/ μ L T4 DNS polimerázzal leemészjtük ami egyben a túlnyúló 5' végeket is kiegészíti. Ezután a defoszforilálódott végeket 3 μ l 10 U/ μ L T4 polinukleotid kinázzal foszforiláljuk, így ligálható formára hozva a DNS-fragmentumokat.

Adapter ligálás: Az immáron foszforilált végű aDNS-re a 3. ábrán látható módon IS1-IS3 és IS2-IS3 univerzális adaptereket ligálunk. Mintánként 4 μ l 10x-es T4 DNS ligáz puffert, 4 μ l 50%-os PEG-4000-et, 1 μ l 50 μ M adapter mixet, 1 μ l 5U/ μ L T4 DNS ligázt alkalmaztunk.



4. ábra: Adapterligálás lépései
(Meyer & Kircher, 2010)

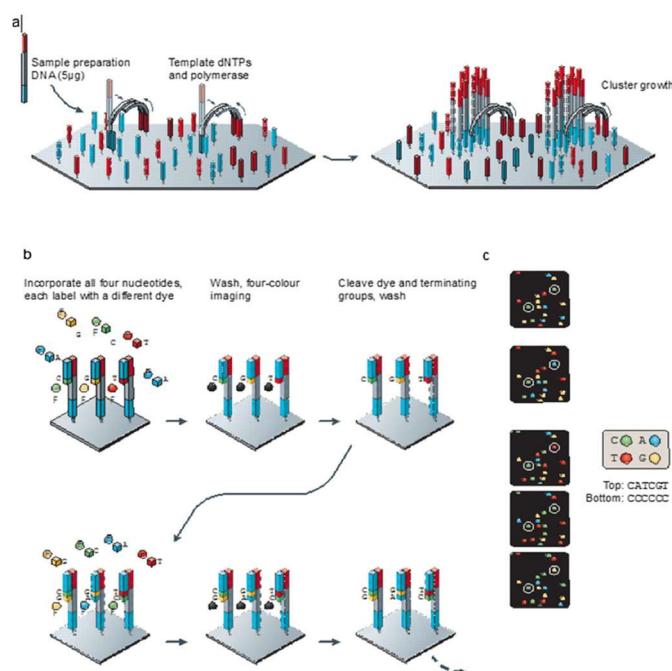
Adapter fill-in: A létrejött túlnyúló végeket 1,5 μ l 8U/ μ L Bst polimeráz enzim segítségével pótoljuk ki, ami az aDNS belső szakaszain található hibákat is kijavítja, így teljesen eltüntetve az egyesszálú töréseket és abázikus helyeket, emellett 4 μ l 10x-es ThermoPol reakció puffert és 0,4 μ l 25mM dNTP-t használtunk.

Előamplifikálás: A minták csekély DNS tartalma miatt az így előkészített aDNS mennyisége nem elegendő a szekvenáláshoz, ezért alacsony ciklusszámú PCR-reakcióval felamplifikáljuk.

Indexálás: A rövid univerzális adapterekkel rendelkező szekvenciákra még egy alacsony ciklusszámú PCR reakció segítségével egyedi P5-P7 „vonalkód” szekvenciákat ligálunk, melyek mintánként egyedi, 6–7 nukleotid hosszúságú indexáló oligót tartalmaznak. Ehhez pedig mintánként 19,5µl Accuprime Pfx Supermix-et, valamint 1µl 10mg/ml BSA-t használtunk.

Szekvenálás

Az aDNS-minták bázissorrendjét jellemző módon az Illumina platform alapú újgenerációs szekvenátorral határozzák meg. A folyamat alapját a szintézis közbeni szekvenálás adja, mely párhuzamosan történik a flow cellen (4. ábra). A flow cell az Illumina szekvenátor esetében egy kisméretű üveglap, melynek felületén a könyvtárkészítés lépésében használt adapterekkel komplementert képező oligók helyezkednek el nagy sűrűségben. A minta oldatot hígítást követően viszik fel a flow cellre, majd a kötődést követően lokálisan felamplifikálják, így klasztereket kialakítva a flow cellen. Minden klaszter egy meghatározott szekvenciájú DNS-darabot tartalmaz. A szekvenálás ciklusokban történik, melyek során a mintákhoz hozzáadott 4 különböző fluoreszcens pigmenttel jelölt nukleotid ciklusonként beépül a szekvenciába, ami egy adott hullámhosszon fényt bocsát ki. Ezeket a specifikus fényvillanásokat egy nagy felbontású CCD kamerával detektáljuk, így ciklusonként követni tudjuk a beépülő nukleotidokat, ezáltal pedig következtetni tudunk a DNS szekvenciájára.



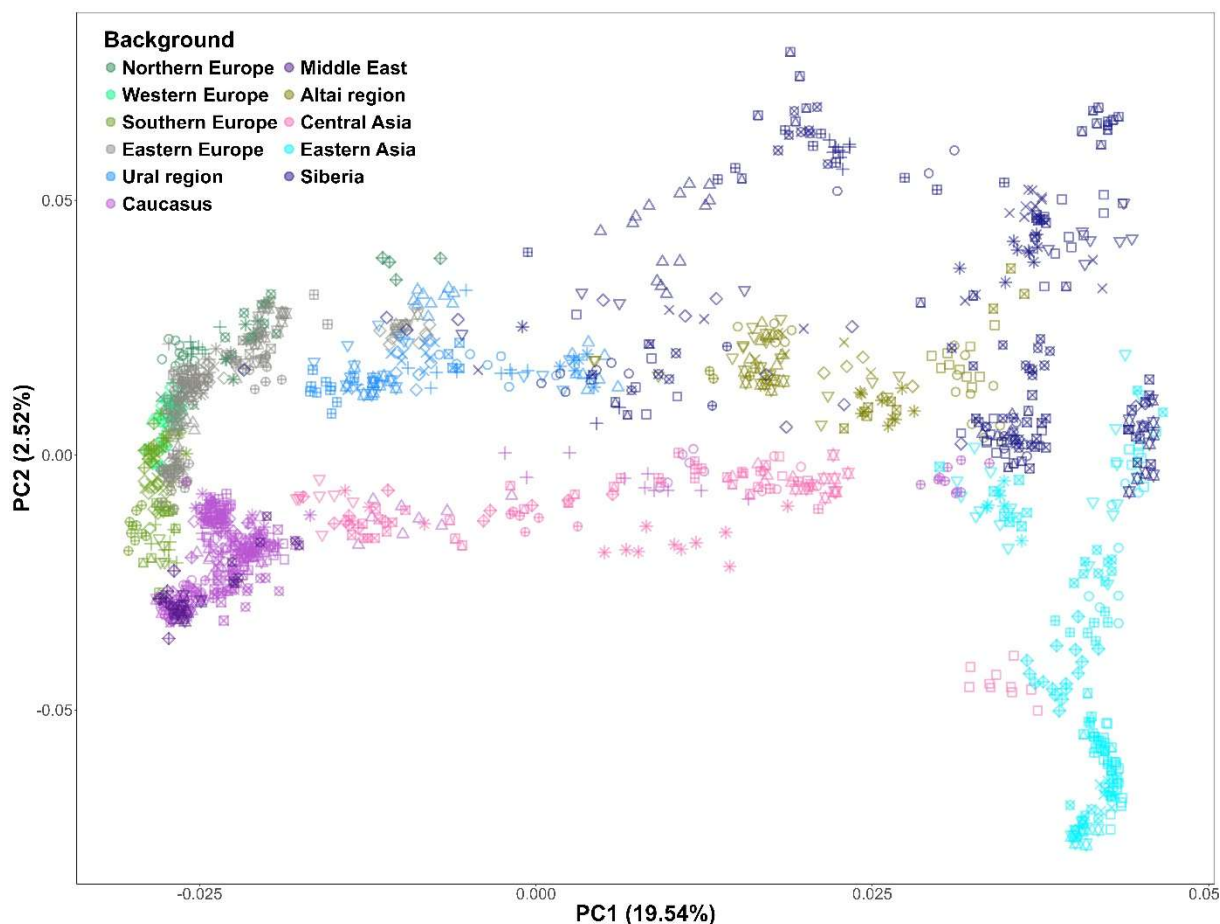
5. ábra: Illumina szekvenálás folyamata flow cellen

(Aydin, 2016)

PCA

PCA, másnéven főkomponens-analízis egy többváltozós adatredukciós statisztikai eljárás, mely egy sokváltozós mintakészlet dimenzióit csökkenti. Ehhez lineáris transzformációval olyan új szintetikus változókat (főkomponenseket) hoz létre, melyek az eredeti lókuszonkénti allélfrekvenciák különböző súlyú lineáris kombinációjából származnak. Matematikai szinten ez az adatok ortogonális lineáris transzformációját jelenti egy új koordináta-rendszerbe, melyben az adatpontokból nyerhető legnagyobb variancia az első tengelyen jelenik meg (első főkomponens, PC1), míg a második legnagyobb variancia a második tengelyen (PC2) jelenik meg és így tovább. A transzformáció célja kettős,. Egyrészt, az új szintetikus változók sokkal egyértelműbb adatot szolgáltatnak. A genomok közötti szisztematikus különbségek vizsgálata sokkal informatívabb, mint az eredeti allélfrekvencia-adatok vizsgálata. Másrészt, az adatpontok között magyarázott variancia mértékének csökkenő voltából fakadóan a sokdimenziós adathalmazt csak az első két vagy első három főkomponens mentén ábrázolva is vissza tudjuk adni az adathalmaz teljes varianciájának jelentős részét akár szabad szemmel is értelmezhető formában.

A genomi főkomponens-analízis egyik igen érdekes tulajdonsága, hogy a megjelenített genomok eloszlása nem véletlenszerű. Az első két főkomponens mentén ábrázolt szekvenciaadatok jellemzően reprodukálják a vizsgált egyének geolokációját (5. ábra). Ennek a jelenségnek egyszerűen az az oka, hogy az egyes emberek között mért genetikai távolság elsősorban a földrajzi távolságukkal arányos, így a genomok között mért legnagyobb variancia is ezt tükrözi.



6. ábra: Modern eurázsiai egyének főkomponens-analízisének eredménye (a genomok eloszlását az első és második főkomponens (PC1 és PC2) alapján ábrázoltuk).

Ha modern eurázsiai embereket ábrázolunk az első két főkomponensük mentén, akkor egy karakterisztikus eloszlást kapunk (6. ábra). Az egymáshoz hasonló genomok egymás közelébe kerülnek, és ezek többnyire közeli földrajzi lokációból származnak. Ezért az összes genom végső eloszlása szinte leképezi az eurázsiai kontinenst. Az első főkomponens (PC1) a kelet-nyugati földrajzi iránynak, míg a második főkomponens az észak-déli iránynak feleltethető meg. Ennek oka, hogy Eurázsia hosszabb, mint szélesebb, így a legnagyobb genetikai variancia a hossza mentén, a második legnagyobb pedig a szélessége mentén fog kialakulni. Az **6. ábrán** tehát az európai populációk a kép bal oldalán („nyugaton”), a kelet-ázsiai populációk pedig a kép jobb oldalán („keleten”) csoportosulnak. A két véglet között genetikai értelemben is átmenetet képező közép-ázsiai populációk vannak, melyek a szélső populációk stabil keverékét hordozó egyéneket tartalmaznak.

A főkomponens-analízis a gyakorlatban a következőképpen történik. Első lépésben egy $n \times m$ -es mátrixot hozunk létre, melyben n vizsgált ember és m marker található. Esetünkben

több százezer markert vizsgálunk, ezért $n \ll m$. Ebben a mátrixban legyen $C(i,j)$ egy érték, melynél i az ember, j az adott marker.

Kiszámoljuk, majd kivonjuk a $C(i,j)$ oszlopaiból az oszlopok átlagát ($\mu(j)$), illetve $p(j) = \frac{\mu(j)}{2}$, egy becslése a háttérben lévő allélfrekvenciának.

Majd egy következő lépéssel normalizáljuk az oszlopok varianciáját, mert egy SNP frekvenciaváltozása genetikai sodródás miatt generációnként $\sqrt{p(j)(1-p(j))}$ -vel történik. – így az adat oszlopai azonos varianciájúak lesznek (Patterson és mtsai., 2006). A következő mátrixot kapva:

$$M(i,j) = \frac{C(i,j) - \mu(j)}{\sqrt{p(j)(1-p(j))}}$$

Az $M(i,j)$ mátrixon szinguláris értékfelbontást alkalmazva kapunk egy olyan $n \times n$ -es kovariancia mátrixot ($K(i,j)$), mely megegyezik a vizsgált egyének számával.

Majd az eigenvektor felbontás segítségével a mátrixot sajátvektoraira (v) és sajátértékeire (λ) bontjuk fel.

$$K(i,j) \cdot v = \lambda \cdot v$$

ahol v iránya nem változik, λ viszont megadja, hogy a sajátvektor hossza mennyiben tér el, mikor a kovariancia mátrixsal szorozzuk. Minél nagyobb λ , annál távolabb helyezkednek el a főkomponensekre vetített pontjaink az origótól. Ennek segítségével megtalálhatjuk azt az egyenest az adatszettünkben, amire a pontjaink legjobban illeszkednek.

Kromoszomális (genetikai) nem meghatározása

Skoglund eljárását alkalmaztuk, mely az Y és X kromoszómához tartozó szekvenciák arányából következtet a genetikai nemre (Skoglund és mtsai., 2013). Ha ez az érték 0,77 feletti, akkor az alany férfi, ha 0,016 és 0,77 közötti, akkor nem meghatározható, ha pedig 0,016 alatti, akkor nő.

F-statisztika

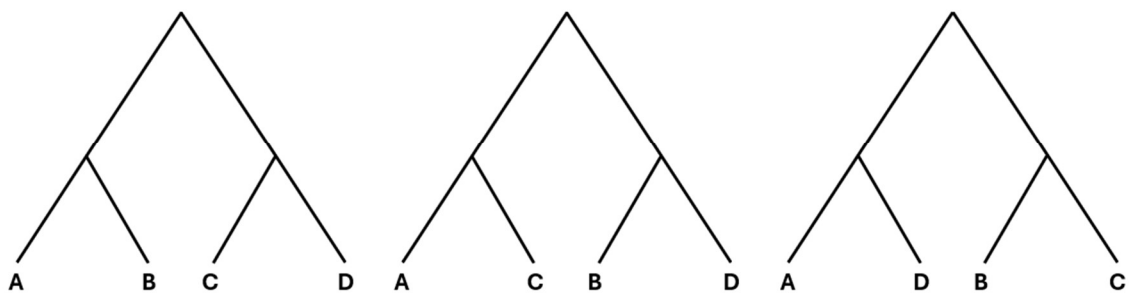
Az F-statisztikák genomok allélfrekvenciáinak statisztikai összehasonlításával következtetnek egyének vagy populációk leszármazási kapcsolatára, közös genomtörténetére. A módszer azon a feltevésen alapul, hogy a szaporodási közösségben álló egyének/populációk evolúciós múltjuk során – a közös genetikai sodródás révén („shared drift”) - azonos allélokat halmoznak fel. Az azonos allélok aránya statisztikailag szignifikáns módon meghatározható, és így következtethetünk a vizsgált alanyok genetikai távolságára.

Fontos megjegyezni azonban, hogy az emberi populációk leszármazási története igen bonyolult, vándorlásokkal és keveredésekkel tarkított. A populációk allélfrekvenciáinak viszonya sok mindentől függhet, így az F-statisztikák minden esetben hipotézis vizsgáló analízisek, melyek egy feltételezett leszármazási modell helyességét erősítik meg vagy vetik el. Az allélfrekvenciák korrelációját tudjuk vizsgálni 2, 3 vagy 4 populáció között, ebből eredően létezik F2, F3 és F4 statisztika. A vizsgált populációk számának függvényében, pedig megbizonyosodhatunk fa-szerű leszármazási viszonyok helyességéről („treeness test”), populációk evolúciós távolságáról vagy akár lehetséges keveredési események számáról, keveredési hozzájárulások arányáról is. (Lipson, 2020)

A dolgozatban 4 populáció közötti leszármazási kapcsolatokat vizsgáljuk, tehát az F4-statisztikát alkalmazzuk. Az F4-statisztika segítségével 4 populáció (A, B, C, D) allélfrekvenciáinak kovarianciáját vizsgáljuk, melyeket egy hipotetikus gyökértelen leszármazási fa segítségével képzelhetünk el (7. ábra). A statisztika $F4(A, B; C, D)$ értékét a következő módon számíthatjuk ki:

$$F4(A, B; C, D) = \frac{1}{J} \sum_{j=0}^J (a_j - b_j)(c_j - d_j)$$

, ahol a_j , b_j , c_j és d_j az A, B, C és D populációk j lokuszán mért allélfrekvenciái. Könnyen belátható, hogy a vizsgált négy populáció 3 féle módon viszonyulhat egymáshoz egy gyökértelen leszármazási fán (7. ábra).



7. ábra: Lehetséges filogenetikai fák nem keveredett populációk esetén

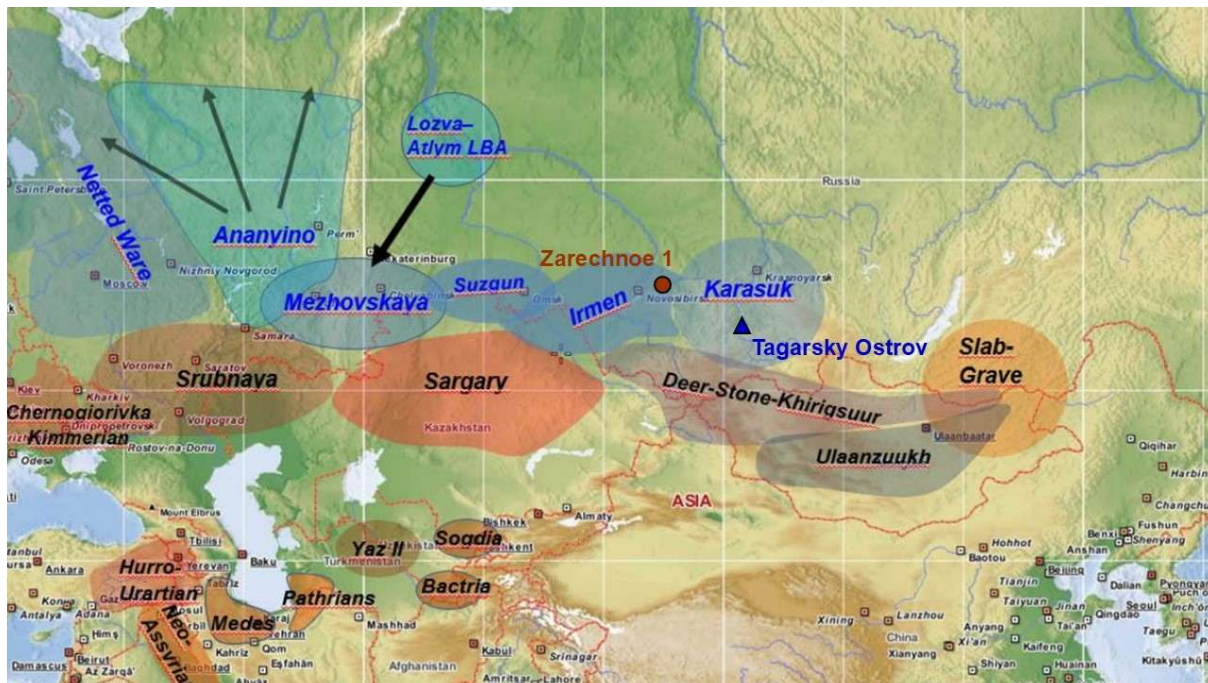
Az F4-statisztika lényege, hogy a számítás 0-t eredményez, ha a kiindulási hipotézisünk helyes és a műveleti sorrend megegyezik a vizsgált fa valós topológiájával. Tehát ha a vizsgált populációk leszármazása valóban A-B és C-D populációk találhatók egy ágon, úgy az $F4(A, B; C, D)$ műveleti sorrendű számítás szükségszerűen 0-t fog eredményezni (Reich és mtsai., 2009).

Az F4-statisztikának egy további fontos tulajdonsága, hogy 0-tól szignifikánsan különböző eredményt produkál, ha a vizsgált populációk valós leszármazási viszonya eltér a műveleti sorrendben felírt topológiától. Ez a tény felhasználható keveredési események azonosítására, sőt a keveredés irányának meghatározására is, ha olyan leszármazási topológiát vizsgálunk, mely részben ismert (pl. egy szimmetrikus külcsoport alkalmazásával). Ezt a megközelítést Greene és mtsai 2010 alkalmazta először, ahol eredetileg D-statisztikának nevezték el és egyének leszármazási viszonyainak vizsgálatára használták (Green és mtsai., 2010). Patterson és mtsai. 2012 azonban bebizonyították, hogy ez a megközelítés alapjaiban véve azonos az F4-statisztikával, így a gyakorlatban ma már nem tesznek különbséget a kettő között (Patterson és mtsai., 2012).

Az F4-statisztika eredményének statisztikai szignifikanciáját Z-érték számítással ellenőrizzük a genom vizsgáló analíziseknél rendszeresen használt „block-jackknife” módszer segítségével.

Eredmények

Szekvenált adatok



8. ábra: A vizsgált lelőhelyek földrajzi elhelyezkedése
(Török, 2023)

ID	Temető	Kultúra	Származási idő	Genetikai nem	Mitochondriális haplocsoport	Y haplocsoport	Átlagos lefedettség	Becsült mitokondriális szennyezettség (%)	Becsült X kromoszóma szennyezettség (%)
ZAR-3239	Zarechnoe 1, Novoszibirszki terület, Togucsin körzet	Irmen	I. e. 2. évezred	XY	A8a1	Q1a2a~	0,51	3	1,1
ZAR-3240	Zarechnoe 1, Novoszibirszki terület, Togucsin körzet	Irmen	I. e. 2. évezred	XY	U5a1l1	R1a1a1b2a2a3~	1,48	1	1,7
ZAR-3407	Zarechnoe 1, Novoszibirszki terület, Togucsin körzet	Irmen	I. e. 2. évezred	XX	I4b	-	0,93	1	-
ZAR-3412	Zarechnoe 1, Novoszibirszki terület, Togucsin körzet	Irmen	I. e. 2. évezred	XY	U4b1a1a	R1a1a1b2a~	1,93	1	1,1
TAG-1369	Tagarsky Ostrov, Krasnojarszki határterület, Minuszinszki-medence	Karasuk	I. e. 2. évezred vége - 1. évezred eleje	XX	U2e2	-	0,96	1	-
TAG-1370	Tagarsky Ostrov, Krasnojarszki határterület, Minuszinszki-medence	Karasuk	I. e. 2. évezred vége - 1. évezred eleje	XX	K1a+195	-	0,19	1	-
TAG-1371	Tagarsky Ostrov, Krasnojarszki határterület, Minuszinszki-medence	Karasuk	I. e. 2. évezred vége - 1. évezred eleje	XY	U2e1	R1a1a1b	0,52	2	1,4
TAG-1372	Tagarsky Ostrov, Krasnojarszki határterület, Minuszinszki-medence	Karasuk	I. e. 2. évezred vége - 1. évezred eleje	XX	U2e1h	-	0,77	4	-
TAG-1375	Tagarsky Ostrov, Krasnojarszki határterület, Minuszinszki-medence	Karasuk	I. e. 2. évezred vége - 1. évezred eleje	XY	R1b1	R1a1a1b2a~	0,68	1	1,4
TAG-1379	Tagarsky Ostrov, Krasnojarszki határterület, Minuszinszki-medence	Karasuk	I. e. 2. évezred vége - 1. évezred eleje	XX	C4a1a3	-	0,70	1	-
TAG-1382	Tagarsky Ostrov, Krasnojarszki határterület, Minuszinszki-medence	Karasuk	I. e. 2. évezred vége - 1. évezred eleje	XY	C5c	R1a1a1b2	1,04	1	0,7
TAG-1383	Tagarsky Ostrov, Krasnojarszki határterület, Minuszinszki-medence	Karasuk	I. e. 2. évezred vége - 1. évezred eleje	XY	U5a1a2a	R1a1a1b2a1~	1,16	1	0,9

1.táblázat: Megszekvenált Karasuk és Irmen minták adatai

A nemek meghatározásához Skoglund módszerét alkalmaztuk, mely alapján 5 nőt és 7 férfit tudtunk beazonosítani.

A vizsgált férfiak közül háromból 2 Irmen és mind a 4 Karasuk egyén ugyanazt a jellegzetes R1a1a1b2 Y haplocsoportot hordozta (1. táblázat, szürkével kiemelve).

Egy Irmen férfi tartalmazta a Q1a2a Y haplocsoportot, ami egy kelet-ázsiai elterjedt haplocsoport.

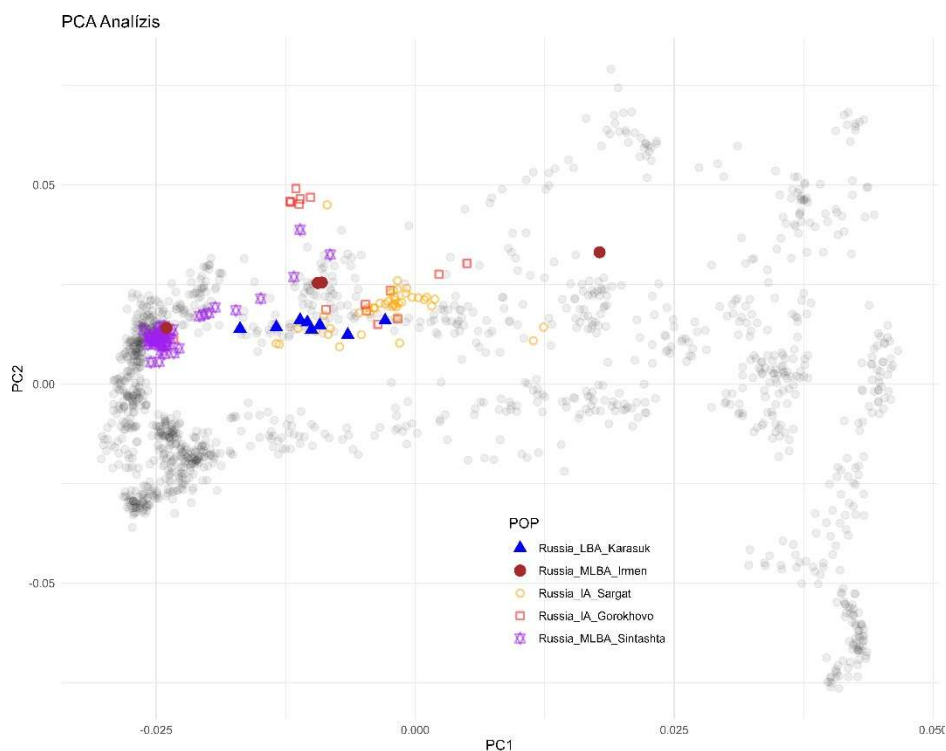
A kapott átlagos genomlefedettség értéke 0,91, majdnem eléri az 1x-es lefedettséget, ami azt jelenti, hogy átlagosan majdnem minden pozícióban egyszer volt egy olvasható nukleotid a szekvencián belül. Egyik mintának a lefedettségi értéke sem kevesebb, mint 0,5 – ezen küszöbérték felett alkalmasak a genomok statisztikai analízis elvégzésére.

PCA

Az analízist lefuttatva a megkapott PC1 és PC2 főkomponenseket (2. táblázat) R programozási nyelv segítségével ábrázoltam (7. ábra).

Master ID	POP (populáció)	Adatok eredete	PC1	PC2
TAG-1369	Russia_LBA_Karasuk	Sargat-Gorokhovo_project	-0,0092	0,0148
TAG-1370	Russia_LBA_Karasuk	Sargat-Gorokhovo_project	-0,0029	0,016
TAG-1371	Russia_LBA_Karasuk	Sargat-Gorokhovo_project	-0,0104	0,0155
TAG-1372	Russia_LBA_Karasuk	Sargat-Gorokhovo_project	-0,0134	0,0143
TAG-1375	Russia_LBA_Karasuk	Sargat-Gorokhovo_project	-0,0065	0,0124
TAG-1379	Russia_LBA_Karasuk	Sargat-Gorokhovo_project	-0,0111	0,0161
TAG-1382	Russia_LBA_Karasuk	Sargat-Gorokhovo_project	-0,01	0,0137
TAG-1383	Russia_LBA_Karasuk	Sargat-Gorokhovo_project	-0,0169	0,0139
ZAR-3239	Russia_MLBA_Irmen	Sargat-Gorokhovo_project	0,0178	0,0331
ZAR-3240	Russia_MLBA_Irmen	Sargat-Gorokhovo_project	-0,024	0,0141
ZAR-3407	Russia_MLBA_Irmen	Sargat-Gorokhovo_project	-0,009	0,0255
ZAR-3412	Russia_MLBA_Irmen	Sargat-Gorokhovo_project	-0,0094	0,0254

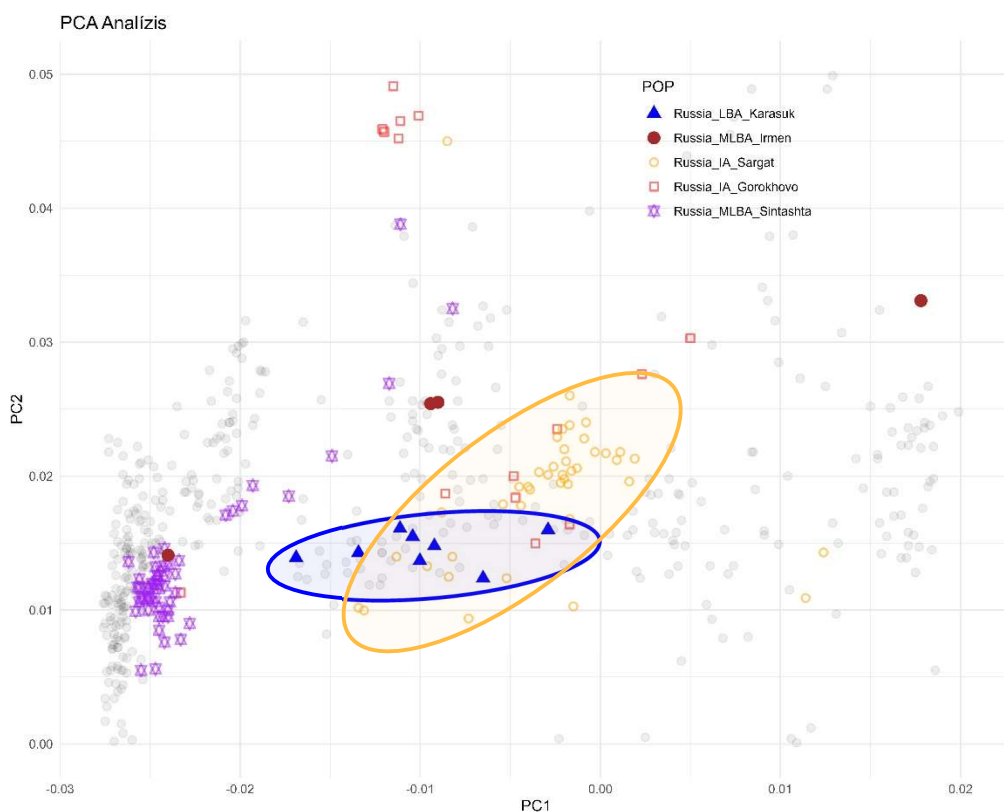
2. táblázat: Karasuk és Irmen kultúrák PCA-adatai



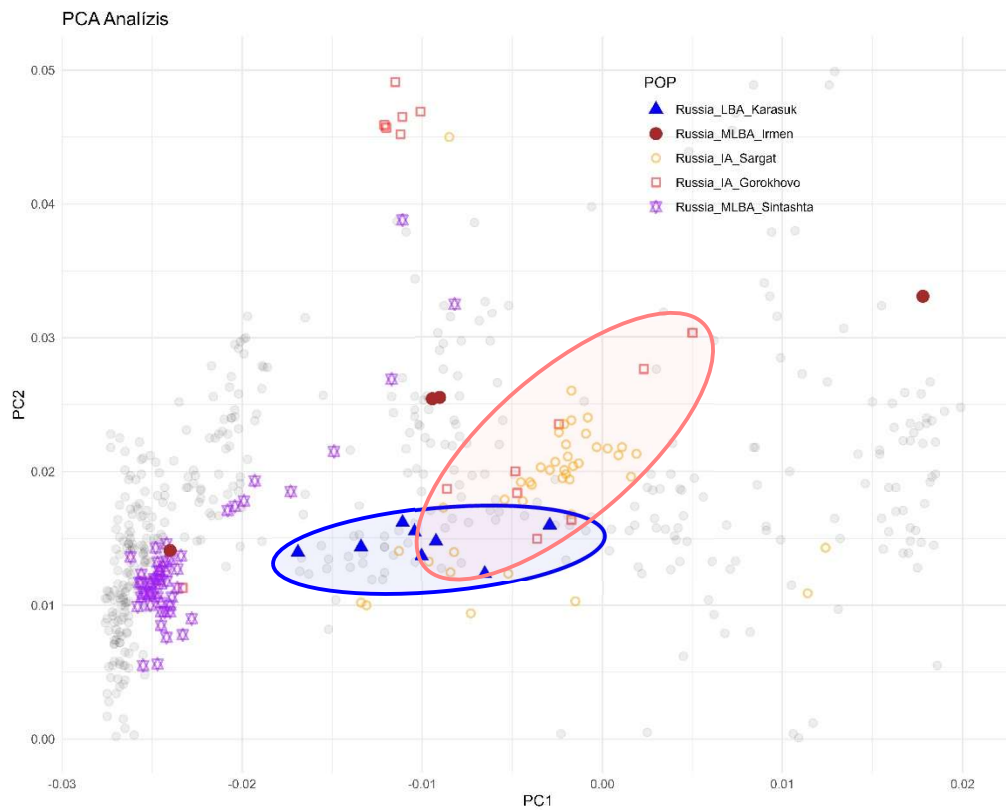
9. ábra: A PCA-analízis eredménye

A Karasuk és Irmen kultúrák genetikai viszonyait Allentoft és munkatársai 2015-ben a Nature folyóiratban (Allentoft és mtsai., 2015) és Narasimhan és munkatársai 2019-ben a Science folyóiratban (Narasimhan és mtsai., 2019) leközölt Sargat-Gorokhovo, valamint Sintashta adataival hasonlítottam össze, a háttérrel képező modern genomok listáját pedig függelékként a dolgozatomhoz csatoltam (1.függelék). A Sintashta populációt azért vettem be az elemzésekbe, mert Narasimhan és munkatársai 2019-ben publikált cikke (Narasimhan és mtsai., 2019) alapján ez a kultúra is nagyban hozzájárult a bronzkori népességek genetikai és kulturális alakulásához.

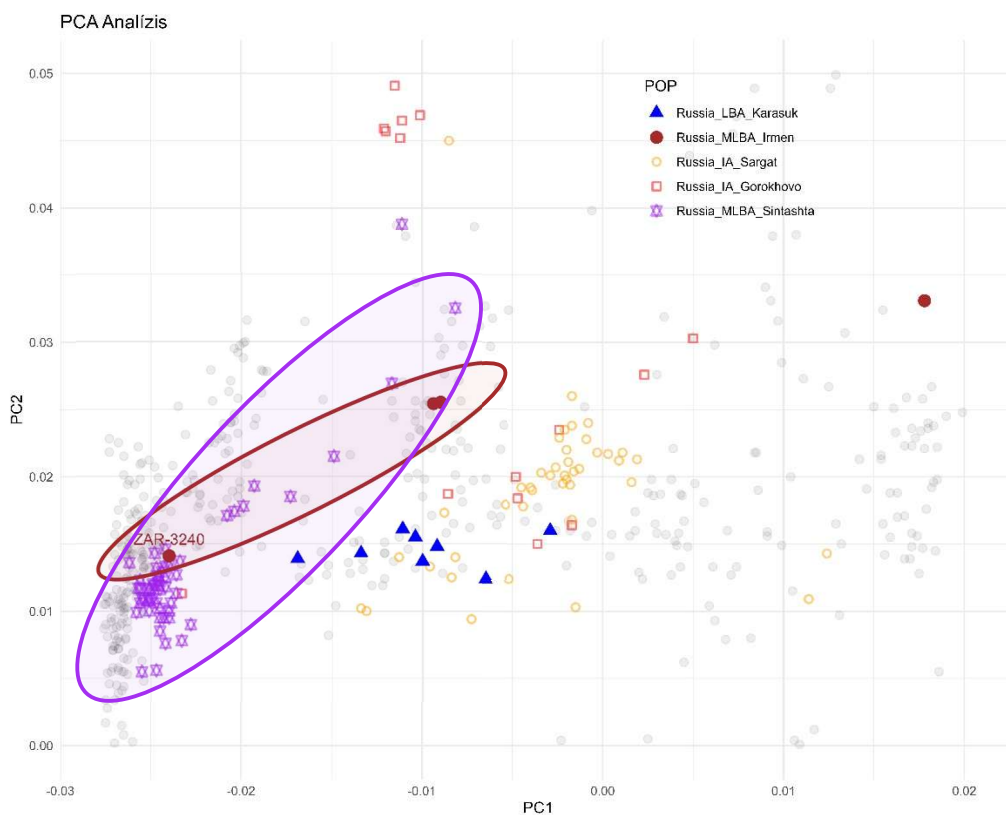
A PCA ábrán (7. és 10. ábra) az Irmen minták egymástól távol vannak, vagyis genetikailag nagyon heterogének, nem mutatnak szignifikáns hasonlóságot más ismert populációkkal. Habár a Karasuk minták is heterogének, egy jobban körülírható, homogénebb csoportba térképeződnek, és átfedésben vannak a Sargat és Gorokhovo mintákkal. Erősebb nagyításban (8. és 9. ábra) jobban látható a minták eloszlása, melyek kisebb-nagyobb genetikai klaszterekbe csoportosulnak.



10. ábra: Karasuk és Sargat kalszterek összehasonlítása



11.ábra: Karasuk és Gorokhovo klaszterek összehasonlítása

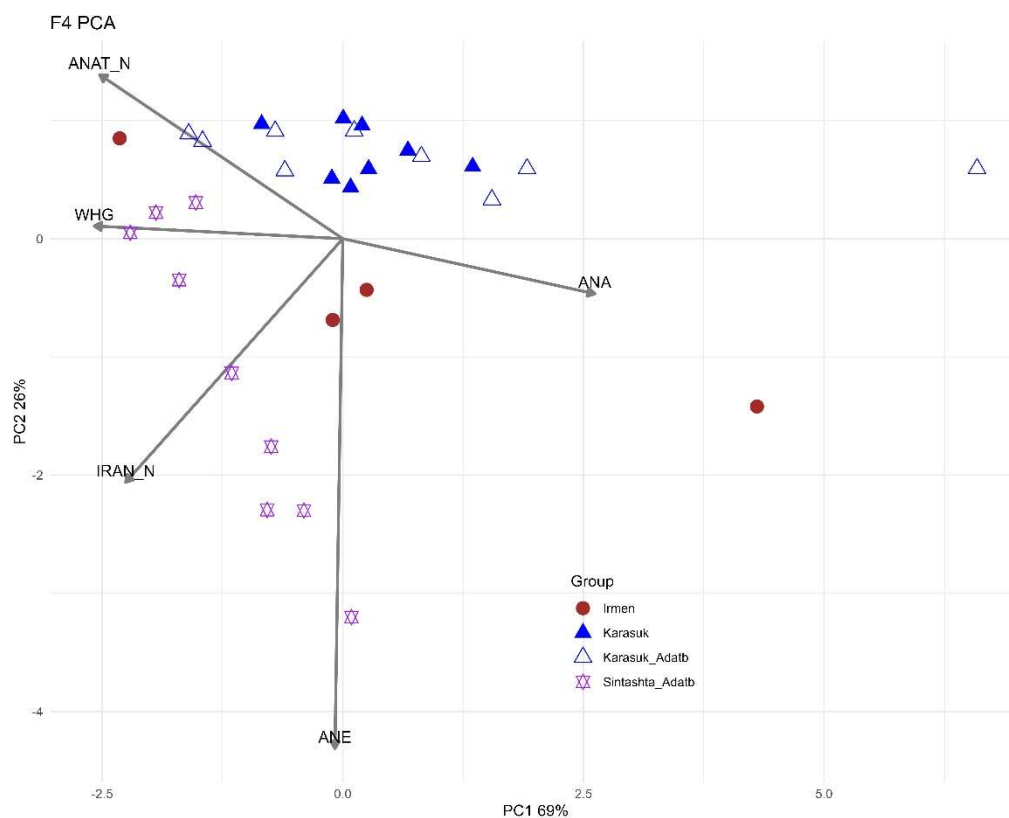


12.ábra: PCA az Irmen minták feltüntetésével

F4-statisztika

Öt független F4-statisztikát futtattunk az Irmen és Karasuk egyéneken, valamint azon Sintashta egyéneken, akik a PCA-n az Irmenhez hasonló pozíciót foglaltak el. A vizsgálat a következő F4 összeállításon alapult: F4(Ethiopia_4500BP, Referencia; Russia_MLBA_Sintashta, Test). Az Ethiopia_4500BP egy szimmetrikus külcsoport, mely minden Afrikán kívüli populációhoz azonos módon viszonyult. Erre a populációra azért volt szükség, mivel egyszerűsíti a leszármazási fánk topológiáját és lehetővé teszi, hogy a statisztika eredményéből (pozitív vagy negatív) következtessünk a genetikai affinitás irányára. A Referencia helyén különböző őspopulációkat cserélgettünk. Ezek egymástól független ősi genetikai komponenseket reprezentálnak, melyekről ismert, hogy nagyban hozzájárultak az eurázsiai népek kialakulásához és jellegzetes földrajzi elterjedéssel rendelkeznek. Ezek a következők voltak: **WHG** – Western Hunter Gatherer, ősi európai halász-vadászok; **ANA** – Ancient North Asian, kelet-ázsiai halász-vadászok; **ANE** – Ancient North Eurasian, Észak és Közép-Euráziában elterjed halász-vadászok; **IRAN_N** – Iran Neolithic, Dél-Euráziában elterjed ősi földművesek; **ANAT_N** – Európában elterjed ősi földművesek. A Russia_MLBA_Sintashta a régió alaplakosságát reprezentálja, melyből a vizsgált egyének génkészletük jelentős részét meríthették, míg a Test helyén a vizsgált egyének szerepeltek.

A kapott eredményeket egy mátrixba összevontuk és Főkomponens Analízist futtattunk rajtuk, majd ezt ábráztuk (13.ábra), így egyszerre vizsgálni lehet az Irmen és Karasuk populációk viszonyait. A populációkhoz tartozó vektorok a PCA során számolt sajátvektorok.



13.ábra: PCA ábra az f4 statisztika eredményeiből

Az ábrázolásból látszik, hogy a korábbi PCA analízishez képest az Irmen egyének jobban elkülönülnek a kilógó Sintashta mintáktól, hasonló mértékű átfedés már nem látható a klaszterek között. A Karasuk minták megmaradtak az előző analízishez hasonlóan egy homogén klasztert alkotva.

Diszkusszió

A szekvenálási adatok között az Y kromoszómához tartozó R1a1a1b2 haplocsoport az Irmen és Karasuk mintáknál egyaránt a leggyakoribbnak mutatkozott. Ezen haplocsoport a Sargat és Gorokhovo populációkban is gyakori volt (A.S. Pilipenko és mtsai, 2017), mely arra enged következtetni, hogy az Irmen és Karasuk kultúrák hozzájárultak a Sargat populáció genetikai összetételéhez. Ez egy jellegzetes ázsiai Y haplocsoport, amely a kora bronzkori sztyeppén terjedt el, és a vaskori ázsiai szkíta csoportokra is jellemző volt.

Egy Irmen férfi (ZAR-3240) tartalmazta a Q1a2a Y haplocsoportot, mely egy kelet-ázsiai haplocsoport, ezért a PCA-n ez az egyén a jobb oldalra térképeződött.

A PCA ábrákon azt látjuk, hogy átfedés van a Karasuk és Sargat között, valamint kisebb átfedés van a Karasuk és a Gorokhovo között is, tehát a Karasuk minták genetikailag hasonlóak a Sargathoz és a Gorokhovoval is vannak közös motívumaik. Ez alapján genetikai kapcsolat feltételezhető a két csoport között, de ennek bizonyítása még további elemzéseket igényel.

Az Irmen genomok sokkal heterogénebbek, kevésbé megfigyelhető olyan homogén klaszter, mint a Karasuknál, azonban egy minta (ZAR-3240) közel helyezkedik el a Sintashta genomokhoz, és másik 2 minta (minták azonosítója) is a Sintashta halmazból kinyúló, úgynevezett „outlierek” közelébe esik. Ez alapján az Irmen népesség Sintashta származása tűnik legvalószínűbbnek, de ez is további vizsgálatokat igényel. Eddigi adataink arra utalnak, hogy az Irmen népesség jelentősen különbözött a Karasuktól.

Az F4 statisztikák összevonásával futtatott PCA alapján a korábban valószínűnek tartott Irmen – Sintashta származás kevésbé látszik elfogadhatónak, mert a vizsgált klaszterek iránya kevésbé fedett egymással. Feltehetőleg egy eddig nem vizsgált populációhoz köthető az Irmenek származása.

Köszönetnyilvánítás

Köszönöm témavezetőimnek, Dr. Török Tibor egyetemi docensnek és Schütz Oszkár PhD hallgatónak, akik türelmükkel, szakmai tudásukkal, valamint motiválásukkal hozzásegítettek dolgozatom megírásához.

Köszönöm Dr. Maróti Zoltán tudományos főmunkatársnak, aki segített a PCA analízis lefuttatásában.

Köszönöm Varga Gergely István tudományos munkatársnak, aki készséggel segített a labormunka elsajátításában.

Irodalomjegyzék

- Allentoft, M. E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P. B., Schroeder, H., Ahlström, T., & Vinner, L. (2015). Population genomics of bronze age Eurasia. *Nature*, 522(7555), 167–172.
- Aydin, S. (2016). Microbial sequencing methods for monitoring of anaerobic treatment of antibiotics to optimize performance and prevent system failure. *Applied Microbiology and Biotechnology*, 100. <https://doi.org/10.1007/s00253-016-7533-5>
- Barnes, I., Barnett, R., & Shapiro, B. (2013). VERTEBRATE STUDIES | Ancient DNA. In S. A. Elias & C. J. Mock (Szerk.), *Encyclopedia of Quaternary Science (Second Edition)* (Second Edition, o. 719–722). Elsevier. <https://doi.org/10.1016/B978-0-444-53643-3.00255-7>
- 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>
- Chernykh, E. (2008). The “Steppe Belt” of stockbreeding cultures in Eurasia during the Early Metal Age. *Trabajos de Prehistoria*, 65. <https://doi.org/10.3989/tp.2008.08004>
- Cooper, A., & Poinar, H. N. (2000). Ancient DNA: do it right or not at all. *Science (New York, N.Y.)*, 289(5482), 1139. <https://doi.org/10.1126/science.289.5482.1139b>
- Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H.-Y., Hansen, N. F., Durand, E. Y., Malaspinas, A.-S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., ... Pääbo, S. (2010). A Draft Sequence of the Neandertal Genome. *Science*, 328(5979), 710–722. <https://doi.org/10.1126/science.1188021>

- 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(5991), 282–284. <https://doi.org/10.1038/312282a0>
- JETTMAR, K. (1970). CROSS-DATING IN CENTRAL ASIA: The Chronology of the Karasuk and the Scythian Periods. *Central Asiatic Journal*, 14(4), 253–276.
- Koryakova, L., & Epimakhov, A. V. (2007). *The Urals and Western Siberia in the Bronze and Iron Ages*. Cambridge University Press.
- Legrand, S. (2006). The emergence of the Scythians: Bronze Age to Iron Age in South Siberia. *Antiquity*, 80(310), 843–859. <https://doi.org/10.1017/S0003598X00094461>
- Lipson, M. (2020). Applying f4-statistics and admixture graphs: Theory and examples. *Molecular Ecology Resources*, 20(6), 1658–1667. <https://doi.org/10.1111/1755-0998.13230>
- Maróti, Z., Neparáczi, E., Schütz, O., Maár, K., Varga, G. I. B., Kovács, B., Kalmár, T., Nyerki, E., Nagy, I., Latinovics, D., Tihanyi, B., Marcsik, A., Pálfi, G., Bernert, Z., Gallina, Z., Horváth, C., Varga, S., Költő, L., Raskó, I., ... Török, T. (2022). The genetic origin of Huns, Avars, and conquering Hungarians. *Current Biology*, 32(13), 2858-2870.e7. <https://doi.org/10.1016/j.cub.2022.04.093>
- Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, 2010(6), pdb.prot5448. <https://doi.org/10.1101/pdb.prot5448>
- Narasimhan, V. M., Patterson, N., Moorjani, P., Rohland, N., Bernardos, R., Mallick, S., Lazaridis, I., Nakatsuka, N., Olalde, I., Lipson, M., Kim, A. M., Olivieri, L. M., Coppa, A., Vidale, M., Mallory, J., Moiseyev, V., Kitov, E., Monge, J., Adamski, N., ... Reich, D. (2019). The formation of human populations in South and Central Asia. *Science*, 365(6457), eaat7487. <https://doi.org/10.1126/science.aat7487>

- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient Admixture in Human History. *Genetics*, 192(3), 1065–1093. <https://doi.org/10.1534/genetics.112.145037>
- Patterson, N., Price, A. L., & Reich, D. (2006). Population Structure and Eigenanalysis. *PLOS Genetics*, 2(12), 1–20. <https://doi.org/10.1371/journal.pgen.0020190>
- Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, 461(7263), 489–494. <https://doi.org/10.1038/nature08365>
- Skoglund, P., Storå, J., Götherström, A., & Jakobsson, M. (2013). Accurate sex identification of ancient human remains using DNA shotgun sequencing. *Journal of Archaeological Science*, 40(12), 4477–4482. <https://doi.org/10.1016/j.jas.2013.07.004>
- Török, T. (2023). Integrating Linguistic, Archaeological and Genetic Perspectives Unfold the Origin of Ugrians. *Genes*, 14(7). <https://doi.org/10.3390/genes14071345>
- Pilipenko A.S., Cherdantsev S.V., Trapezov R.O., Molodin V.I., Kobeleva L.S., Pozdnyakov D.V., Polosmak N.V. (2017) Kinship Analysis of Human Remains from the Sargat Mounds, Baraba Forest-Steppe, Western Siberia. *Archaeology, Ethnology & Anthropology of Eurasia*. 2017;45(4):132-142.
- Kosarev M. F. Bronze Age of Western Siberia. M., 1981. 278 c. 290. Kosarev M. F. On the origin of the Irmen culture// *Monuments of the Stone and Bronze Ages of Eurasia*. M., 1966. C. 169-175
- Chlenova N. L. Dating of the Irmen culture // *Problems of chronology and cultural affiliation of archaeological monuments of Western Siberia*. Tomsk, 1970. C. 133-149.