

# **A honfoglalók genetikai származásának és rokonsági viszonyainak vizsgálata archeogenetikai módszerekkel**

**Ph.D. értekezés**

**Szerző: Neparáczki Endre**

**Témavezető: Dr. Török Tibor**

**Társ-témavezető: Dr. Pálfi György**

**Biológia Doktori Iskola**

**Szegedi Tudományegyetem Természettudományi és Informatikai Kar**

**Genetikai Tanszék**

**Szeged**

**2017.**

# TARTALOMJEGYZÉK

<b>1. Bevezetés</b>	<b>6</b>
<hr/>	
1.1. <i>Az emberi genom genetikai változatai</i>	7
1.1.1. SNP	8
<hr/>	
1.1.2. Strukturális variánsok (SV)	9
<hr/>	
1.1.3. Mikroszatelliták (Short Tandem Repeat, STR)	9
1.2. <i>Genetikai tipizálás módszerei</i>	9
1.2.1. Restriktációs fragmenthossz polimorfizmus (RFLP)	9
<hr/>	
1.2.2. Short tandem repeats (STR)	10
<hr/>	
1.2.3. Single nukleotid polimorfizmus (SNP) - SNaPshot	10
<hr/>	
1.2.4. DNS szekvenálás	12
1.3. <i>A mitokondriális genom</i>	13
<hr/>	
1.4. <i>Az Y-kromoszóma</i>	14
<hr/>	
1.5. <i>Magyar populációk genetikai vizsgálatainak eddigi eredményei</i>	16
<hr/>	
1.6. <i>Archeogenetika</i>	17
<hr/>	
<b>2. Célkitűzések</b>	<b>22</b>
<hr/>	
<b>3. Anyagok és Módszerek</b>	<b>23</b>
<hr/>	
3.1. <i>Mintavétel</i>	23
<hr/>	
3.2. <i>A csontminták porítása</i>	24
3.2.1. Porítás Dremel kéziszerszámmal	24
<hr/>	
3.2.2. Porítás csontmalommal	25
3.3. <i>DNS kivonás</i>	25
3.3.1. DNS kivonás 1	26
<hr/>	
3.3.2. DNS kivonás 2	26

3.3.3.	DNS kivonás 3	26
3.3.4.	DNS kivonás 4	26
3.3.5.	DNS kivonás 5	27
3.4.	<i>PCR alapú haplotipizálás</i>	27
3.4.1.	mtDNS haplotipizálás	27
3.4.2.	Y kromoszóma haplocsoport meghatározás	27
3.4.3.	STR	28
3.4.4.	Az aDNS adatok hitelessége	29
3.5.	<i>NGS alapú haplotipizálás</i>	30
3.5.1.	Könyvtár készítés	30
3.5.2.	mtDNS dústítás és szekvenálás	31
3.5.3.	Genomi dústító kit	32
3.5.4.	NGS szekvencia kiértékelés	32
3.6.	<i>Adatelemzés adatbázisok felhasználásával</i>	33
3.6.1.	Populációgenetikai vizsgálat	33
3.6.2.	Filogenetikai vizsgálat	35
<b>4.</b>	<b><i>Eredmények</i></b>	<b>36</b>
4.1.	<i>PCR alapú eredmények</i>	36
4.1.1.	mtDNS haplotípusok	36
4.1.2.	A populációgenetikai analízis eredményei	37
4.1.3.	Filogenetikai eredmények	42
4.1.4.	Y-kromoszómális haplocsoportok	44
4.1.5.	Autoszómális STR eredmények	46
4.2.	<i>NGS alapú eredmények</i>	47
4.2.1.	mtDNS genom haplotípusok	47

4.2.2.	A PCR alapú eredmények felülvizsgálata az NGS eredmények alapján	50
4.2.3.	A teljes mtDNS genomok filogenetikai eredményei	55
<b>5.</b>	<b>Az eredmények értékelése</b>	<b>61</b>
5.1.	A PCR-alapú eredmények értékelése	61
5.2.	NGS eredmények összevetése PCR alapú eredményekkel	62
5.3.	Nagyfelbontású genetikai vizsgálatok	65
<b>6.</b>	<b>Köszönetnyilvánítás</b>	<b>70</b>
<b>7.</b>	<b>Irodalomjegyzék</b>	<b>72</b>
<b>8.</b>	<b>Saját közlemények listája</b>	<b>83</b>
<b>9.</b>	<b>Tartalmi összefoglaló</b>	<b>85</b>
<b>10.</b>	<b>Summary</b>	<b>89</b>
<b>11.</b>	<b>Rövidítések jegyzéke</b>	<b>93</b>
<b>12.</b>	<b>Függelék</b>	<b>94</b>

*„Így ér véget majd küldetésem, ha egyszer be tudom majd bizonyítani, hogy ellentétben a finn-magyar elmélet mellett kardoskodók megnyilatkozásaival, a magyar nép igenis Atilla népe.”*

- Kőrösi Csoma Sándor -

## 1. Bevezetés

A XXI. század elején a honfoglalók eredete és a magyarok származása még mindig vitatott kérdés. A honfoglalók 895/896-ban érkeztek Etelközből a Kárpát-medencébe (Róna-Tas, 1999), nagyrészt europid, kisebb részben europo-mongolid embertani jellegekkel (Éry, 1994; Fóthi, 1998), és régészeti hagyatékuk jól elkülöníthető a megelőző avar és későbbi Árpád-kor leleteitől (Révész, 2016). Mivel ebből az időből nem maradtak írásos emlékek nem tudhatjuk biztosan, hogy a honfoglalók milyen nyelven beszéltek, de az elfogadott elmélet a magyar nyelv Kárpát-medencébe érkezését egyértelműen hozzájuk köti.

Kijelenthető, hogy a korábban rendelkezésre álló források értelmezésében a történész, nyelvész és régész kutatók elérték a lehetséges határokat. A meglévő honfoglalás kori leletek és történeti adatok többsége feldolgozásra került, a lehetséges elméletek közül a történeti nyelvészet álláspontja vált elfogadottá, és ennek szellemében a magyar őstörténetet kérdése nyugvóponttra jutott.

A népcsoportok rokonsági és leszármazási viszonyainak kutatásában az utóbbi évtizedekben egyre nagyobb szerepet játszanak a genetikai vizsgálatok, melyek kezdetben ma élő emberekre korlátozódtak. A mai adatokból azonban csak közvetett módon következtethetünk az egykori őseinkre. Ezzel szemben a közelmúltban megjelent új tudományág, a régészeti genetika vagy archeogenetika közvetlen adatokkal szolgál a honfoglalók, vagy bármely ősi népesség származásáról, rokonsági viszonyairól. Tehát a régészeti genetika jelképesen „szóra tudja bírni a leleteket”, és ezzel megválaszolhat eddig nyitott kérdéseket is.

Doktori munkám célja a Szegeden korábban Raskó István által elkezdett archeogenetikai kutatások újraindítása, Török Tibor vezetésével. Témavezetőm és társ-témavezetőm külső támogatásával, első lépésként létre kellett hoznunk a későbbi kutatásoknak helyt adó SZTE Archeogenetikai Laboratóriumot az SZTE Genetikai és Embertani Tanszék helyiségeiben. Ez után elkezdhattük Karosi honfoglalás kori temetők archeogenetikai jellemzését. Munkám során olyan molekuláris biológiai módszereket alkalmaztam, amelyek az eddigieknél sokkal pontosabb anyai és apai ágú leszármazás vizsgálatot tesznek lehetővé.

## 1.1. Az emberi genom genetikai változatai

A genetikai rokonság megállapításához azokat a DNS szakaszokat vizsgálják, amelyek az egyének és populációk között eltérést mutatnak. Általánosságban elmondható, hogy a teljes genom DNS szekvenciája minden emberben 99,6 %-ban azonos (Levy és mtsai., 2007). Ennek ellenére nincs két genetikailag azonos ember, még az egypetűjű ikrek között is kimutathatóak apró genetikai különbségek, melyek az egyedfejlődés során létrejövő mutációk és kópia szám változatok (Bruder és mtsai., 2008).

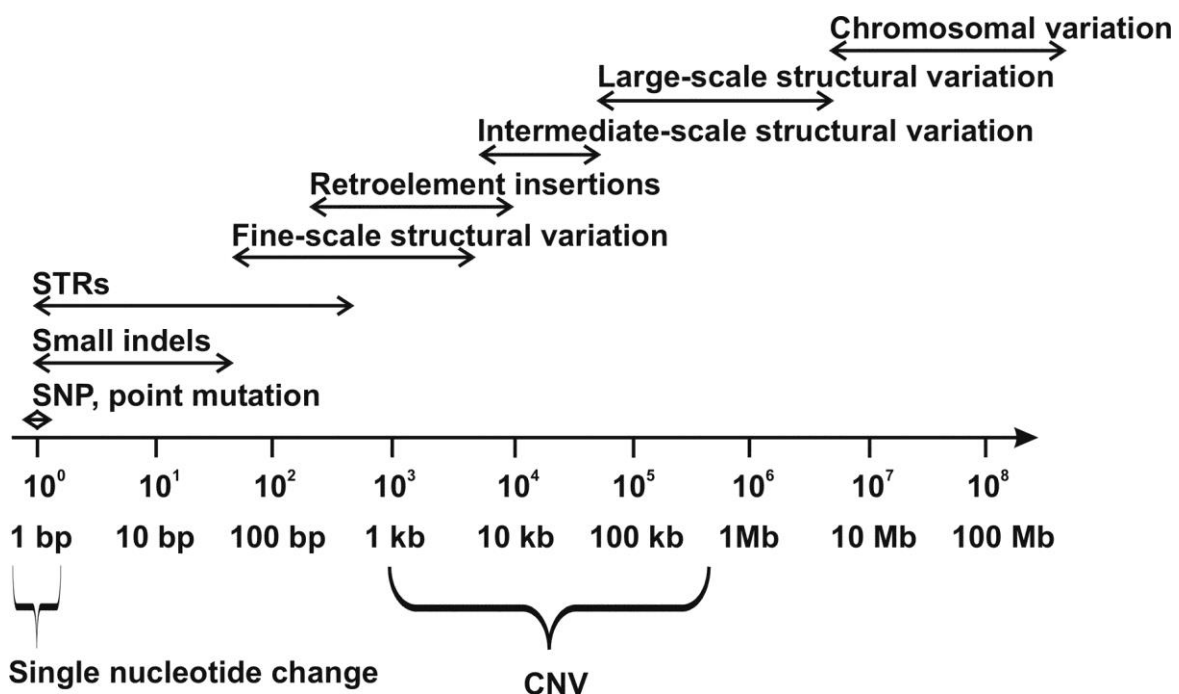
Azt, hogy két egyén genomja milyen mértékben különbözik egymástól, a nukleotid diverzitás érték adja meg. Ezt az értéket emberben 0,1%-tól (Jorde és Wooding, 2004) 0,4%-ig (Tishkoff és Kidd, 2004) becsülik. Mivel a humán genom több mint 3 milliárd ( $3 \times 10^9$ ) bázispárból áll, 0,3% nukleotid diverzitás értékkel számolva átlagosan 9 millió bázispár különbség lehet egyén és egyén között. Napjainkra 26 populációból több mint 1000 ember teljes genomját szekvenálták meg az 1000 genom projekt során (The 1000 Genomes Project Consortium, 2015), és azt találták, hogy egy-egy egyén genomja átlagosan 4,1-5 millió helyen különbözik a referencia genomtól. Ezen variánsok többsége SNP és rövid inszerció-deléció, de a nagyméretű szerkezeti átrendeződések sokkal nagyobb mértékű különbségeket okoznak az egyének között (1. ábra).

Az egyénre jellemző genetikai mintázatot genetikai ujjlenyomatnak hívjuk. Az egyénekre jellemző génváltozatok (allélek) az egyes populációkban eltérő, azokra jellemző gyakorisággal fordulnak elő. Ez lehetővé teszi, hogy a populációk rokonságára és származására is következtetést vonjunk le. Mivel az egyes génváltozatok gyakran jellegzetes földrajzi elterjedést mutatnak, nagyszámú élő és ásatag minta genetikai jellemzésével megismerhetővé válik az egykori populációk migrációja és meghatározható a különböző emberi csoportok biológiai értelemben vett rokonsági foka.

A sejtmagi DNS döntő hányada az átörökítés során erősen keveredik rekombinációval, ezért nem alkalmas leszármazási vonalak rekonstruálására. Ehhez olyan genetikai markerek vizsgálata alkalmas, melyeknél az anyai és az apai információ nem keveredik az egymást követő nemzedékekben. Ilyen nem rekombinálódó genetikai egység a mitokondriális DNS (mtDNS), amely a petesejtek mitokondriumai révén kizárólag anyai ágon öröklődik (Giles és mtsai., 1980), és melynek vizsgálatával a populációk anyai ági leszármazási vonalait követhetjük nyomon. A

másik ilyen egység, az apáról fiúra öröklődő Y-kromoszóma, melynek analizésével a populációk apai ági leszármazási viszonyait deríthetjük fel.

A genetikai változatok tanulmányozásának evolúciós és gyógyászati jelentősége is van. Mivel az egyes populációk különböznek a betegséget okozó, vagy betegségre hajlamosító allélokban, és az egyes gyógyszerek hatékonysága nagyban függ az egyén genetikai hátterétől, ezért a modern gyógyszer fejlesztés során egyre inkább figyelembe veszik a célpopuláció genetikai összetételét.



1. ábra Az egyes genetikai változatok méret eloszlás szerint csoportosítva.

### 1.1.1. SNP

Az SNP egyetlen nukleotid különbséget jelent két egyed között. A filogenetikai vizsgálatok során azokat az SNP-ket használják, amelynek változatossága a populációban 1%-nál gyakoribb. Az eddigi genom adatok szerint 10-30 millió SNP található a humán populációkban (The International HapMap Consortium, 2003), ez a leggyakoribb szekvencia variáns, amely az összes szekvencia különbségek közel 90%-át adja (Collins és mtsai., 1998). Az SNP-k többsége netruális vagyis nincs funkcionális hatása, csak 3-5% funkcionális, amelyek



genetikai markerként használhatók a genotípus-fenotípus összefüggések keresése során (Genome Wide Association Studies, GWAS), (Ke és mtsai., 2008).

### **1.1.2. Strukturális variánsok (SV)**

Ezek rövidebb-hosszabb szakaszon változtatják meg a nukleotid sorrendet vagy a kromoszómák szerkezetét. Ide tartoznak a kópia szám változatok (CNV), deléciók, inverziók, inszerciók és duplikációk. Ezek létrejöttében kiemelendő a transzpozonok és retrotranszpozonok szerepe, melyek nagy számban találhatóak a genomban, és mozgásuk során SV-kat hoznak létre. Egy tipikus emberi genom 2100-2500 SV-t tartalmaz, melynek átlagos megoszlása: 1000 nagy deléció, 160 CNV, 915 Alu inszerció, 128 L1 inszerció, 51 SVA inszerció, 4 NUMT és 10 inverzió (The 1000 Genomes Project Consortium, 2015).

### **1.1.3. Mikroszatelliták (Short Tandem Repeat, STR)**

Ezek nagyon rövid (2-5 bázispár) hosszúságú tandem ismétlődő DNS szakaszok, amelyek klaszterekben szétszóródva találhatóak az eukarióta genomban, és az ismétlődések száma nagy változatosságot mutat az egyének között. Eredetüket azzal magyarázzák, hogy az eredetileg spontán mutációval keletkező néhány nukleotid ismétlődés kialakulását követően a replikáció során a DNS szálak elcsúszhatnak, így a DNS polimeráz megtöbbszörözi ugyanazt a szekvencia darabot. Dinukleotid ismétlődéseknél 4-5, tri- ill. tetranukleotid ismétlődések esetén 2-3 meglévő ismétlődés elegendő az elcsúszáshoz. Az szál elcsúszással keletkező di-, tri- és tetranukleotid ismétlődés változatok a további replikációk során rögzülnek a genomban, és stabilan öröklődnek (Messier és mtsai., 1996).

## **1.2. Genetikai tipizálás módszerei**

### **1.2.1. Restriktív fragmenthossz polimorfizmus (RFLP)**

A módszer alapja, hogy a restriktív endonukleáz enzimek a DNS láncon meghatározott szekvencia sorrendeket ismernek fel, és azokat elhasítják. Ha mutáció miatt a vizsgált DNS szakasz szekvenciája úgy változik meg, hogy egy restriktív enzim felismerő helye eltűnik vagy új felismerő hely keletkezik, ez eltérő DNS fragment hosszaként észlelhető az emésztést követő elektroforézis mintázatban (Brown, 1980; Cann és mtsai., 1987). A módszer előnyei közé

tartozik, hogy technikailag viszonylag könnyen kivitelezhető, míg hátránya, hogy csak a vizsgált restrikciós enzimek felismerési helyeinek változását képes detektálni.

### **1.2.2. Short tandem repeats (STR)**

A mikroszatellitákat régóta használják genetikai markerként. Ezen vizsgálatok előnye, hogy nagyszámú STR-t párhuzamosan vizsgálva nagy valószínűséggel kimutatható egyedre jellemző mintázat. Az egyes mikroszatellita ismétlődések hosszát PCR reakciót követően gélelektroforézissel lehet mérni. Az STR mintázat genetikai ujjlenyomatként használható egyének genetikai azonosítására (Alberts és mtsai., 2014; Brinkmann és mtsai., 1998).

A módszer igazságügyi alkalmazása igen elterjedt (ezért végezzük ezen vizsgálatokat igazságügyi intézetekkel együttműködésben). Archaikus DNS vizsgálatra csak korlátozottan alkalmas módszer, mivel a degradált DNS fragment hossza gyakran nem elegendő a PCR reakcióhoz, és az STR kitekben rutinszerűen vizsgált nagyobb fragmentek (>200 bázispár) gyakran nem amplifikálhatók.

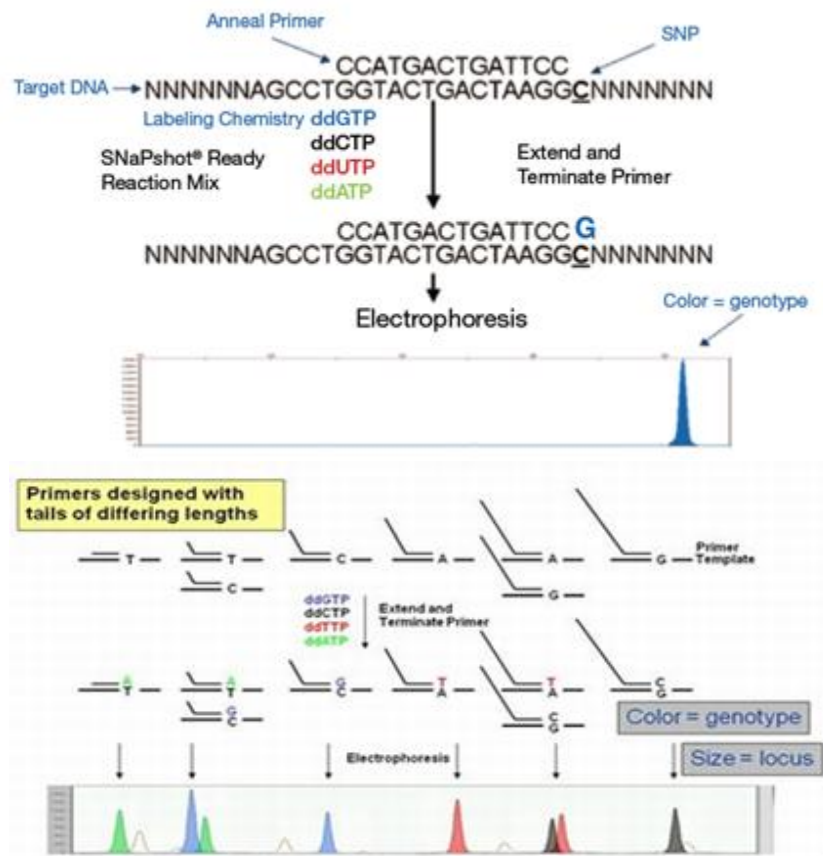
### **1.2.3. Single nukleotid polimorfizmus (SNP) - SNaPshot**

A teljes genom szekvenálás igen költséges módszer, ezért többnyire a csontmaradvány részleges genotípusát szokás meghatározni. Elsősorban azokra a DNS szakaszokra koncentrálunk, melyekben a populációk jellegzetes polimorfizmust mutatnak, melyek alapján az egyének eltérő úgynevezett haplotípusokba és haplocsoportokba sorolhatók. A haplotípus az egyén nem rekombinálódó DNS szakaszaira jellemző SNP mintázat, melynek kiemelkedő jelentősége van az mtDNS és Y-kromoszóma leszármazás vizsgálatában. A haplocsoport az azonos leszármazási vonalba tartozó haplotípusok összességét jelenti. A Humán Genom Projekt nagyszámú genom szekvencia összehasonlításával korábban meghatározta az mtDNS és az Y-kromoszóma vonalak leszármazási viszonyait, és kijelölte a haplocsoportba soroláshoz legalkalmasabb SNP pozíciókat.

A SNaPshot módszer arról kapta nevét, hogy képes egyetlen reakcióban nagyszámú SNP-t egyszerre meghatározni. A módszer multiplex PCR-t követő egy nukleotidos primer extenziós reakción alapszik. Az Euráziára jellemző leggyakoribb mtDNS haplocsoportokba történő besoroláshoz az mtDNS kódoló szakaszán 22 jól jellemzett polimorf pozíció szekvenciáját kell meghatározni (Haak és mtsai., 2010), és erre a GenoCore 22 nevű SNaPshot

módszer a legalkalmasabb. Az eurázsiai emberekben leggyakoribb 25 Y kromoszómás főbb haplocsoportokba való besorolásához szintén kidolgozták a 25 jellegzetes SNP pozíció meghatározásának SNaPshot módszereit (Haak és mtsai., 2010), melyet GenoY25 assay-nek neveztek el. Munkánk során mi is ezeket a módszereket használtuk.

A primer-extenzió (Single Base Extension = SBE) során egy primert hibridizálunk az amplifikált DNS-darabhoz, melynek 3' vége épp a kérdéses nukleotid előtt végződik. Ezután 4 fluoreszcensen jelölt dideoxy nukleotidot (ddNTP) és polimeráz enzimet adunk a reakcióhoz, így a kérdéses pozícióba egyetlen színnel jelölt nukleotid épül be (2. ábra). Mivel minden egyes SBE primer mérete különböző, így azok egy szekvenáló gélen szétválaszthatók, a fluoreszcens festék színe pedig elárulja, hogy az egyes SNP pozíciókba melyik bázis épült be (Der Sarkissian, 2011).



**2. ábra** A felső ábrán fluoreszcensen jelölt ddNTP-ék beépülését, majd annak elektroforézis utáni elektroferogram képét látjuk. Az alsó ábra azt szemlélteti, hogy hogyan lehet több reakció párhuzamosan végezni. Ebben a tervezett SBE primerek hossza különböző, amelyek PCR után a szekvenáló gélen elválaszthatók lesznek, így hossz alapján beazonosíthatóak az egyes vizsgált pozíciók, szín alapján pedig a SNP-t tudjuk meghatározni.

#### 1.2.4. DNS szekvenálás

A DNS-szekvenálás az a folyamat, melynek során meghatározzák a DNS molekula nukleotid sorrendjét. A DNS információtartalma a szekvenciájában található, ezért egy DNS szakasról annak teljes szekvenciája adja a legtöbb információt. Napjainkban a DNS-szekvenálás a molekuláris biológia egyik legmeghatározóbb eszközévé vált.

Az 1970-es évektől kezdve két módszert fejlesztettek ki, amelyek közül a Sanger-féle láncterminációs (enzimatis) módszer vált egyeduralmúvá. A fluoreszcens detektálás megjelenésével és automatizálásával könnyebbé és gyorsabbá vált a DNS-szekvenálás. Végül az ezredfordulón jelentek meg a teljesen más mechanizmuson alapuló, még nagyobb hatékonyságú és olcsóbb új-generációs szekvenáló módszerek (NGS, next-generation sequencing), amelyek az információszerezés mennyiségét tekintve újabb forradalmi változást indítottak el a molekuláris biológiában. Jelenleg ez utóbbi módszereket és az automata láncterminációs szekvenálást párhuzamosan használják a biológiában. Mivel a Sanger féle szekvenálás általánosan ismert, ennek részleteire nem térek ki, a továbbiakban csak az NGS módszereket tekintem át.

Az elmúlt évek során megjelent úgynevezett új-generációs szekvenálási technikák nagy előrelépést jelentettek a szekvenálás területén azáltal, hogy lehetővé teszik egyetlen kísérletben akár  $10^5$ - $10^6$  különböző DNS-molekula párhuzamos, gyors és automatizált leolvasását. Ezt a módszert nagy áteresztőképességű high-throughput sequencing (HTS) módszernek is nevezik. Ezen eljárások során nincs szükség a DNS-láncok időigényes méret szerinti elválasztására. Hátrányuk, hogy egy reakció során mindegyik molekulából csak viszonylag rövid, maximum néhány 100 nt. hosszúságú DNS-darab szekvenciája olvasható le.

A piroszekvenálást 1996-ban dolgozták ki, alapelve teljes mértékben eltér a Sanger-féle szekvenálástól. A piroszekvenálás a „szekvenálás szintézissel” elvén alapul, melynek során nagyszámú egyszálú DNS templátról enzimatisan komplementer szálakat szintetizálnak. A DNS populációt első lépésben egy mikrochipen rögzítik, és lokálisan amplifikálják. A nukleotid beépüléseket egy kemilumineszcens enzim segítségével detektálják. Amikor egy nukleotid beépül a DNS szálba, pirofoszfát (PPi) keletkezik, és ennek a mennyiségét mérik egy kapcsolt reakcióval, végső soron a luciferáz enzim helyi felvillanással jelzi ha egy pozícióban beépülés történt. Minden egyes lépésben csak egyféle nukleotidot adnak a rendszerhez, ezáltal csak az a molekula ad felvillanást ahol beépülés történt. Ha több azonos nukleotid épül be egyszerre, arra a fényintenzitás növekedéséből lehet következtetni. A felvillanás helyét a rögzített szekvenálandó

molekulák helyzetének megfelelően nagyfelbontású kamera érzékeli. Kemilumineszcencia csak komplementer nukleotid beépülése során következik be, a be nem épült nukleotidokat a következő ciklus előtt eltávolítják.

A második generációs szekvenátorok gyártása és fejlesztése különálló iparággá nőtte ki magát. Számos biotechnológiai cég kínál egymástól eltérő módszereken alapuló szekvenátorokat. Mi az SeQomics Kft. Illumina MiSeq platformját használtunk, együttműködés keretében. Ennek a módszerét a Solexa cég fejlesztette ki, melyet az Illumina cég később felvásárolt. Jellemzősége, hogy ez volt az első rövid leolvasásokat végző technika. Első lépésben a szekvenálni kívánt DNS-t (akár teljes genomot) apró, 100 bp-os darabokra fragmentálják. A dupla szálú DNS két végét kijavítják (ragados végek eltüntetése), és egy adeninnel toldják meg az 3' végeket. Ehhez egy timin túlnyúló véggel rendelkező ún. adapter DNS-t (meghatározott mesterséges szekvencia darab) ligálnak úgy, hogy a molekulák két végére két eltérő adapter kerüljön. Az adapter ligált DNS-darabokat NaOH-val denaturálják, majd a DNS populációt egy szekvenáló mikrochipre viszik fel, amely az adapterrel komplementer oligonukleotidokat (primerek) tartalmaz sűrűn kihorgonyozva. A DNS fragmentumok ezekhez hibridizálnak, és ezzel adott helyen rögzülnek. Ezután az ún. híd-amplifikációs (bridge amplification) módszerrel minden egyes molekulát helyben felsokszoroznak. Maga a szekvenálás itt is szintézissel történik. Az adapterhez egy primert hibridizálnak, melynek 3' vége a szekvenálandó molekulánál végződik. Itt a reakcióhoz egyszerre adják hozzá mind a négy eltérő fluoreszcens festékkel jelölt nukleotidot (reverzibilis terminátorok). A beépült nukleotid fluoreszcens jelét CCD kamerával detektálják. Minden egyes ciklus végén a nukleotid beépülése után kémiai úton levágják róla a fluoreszcens festéket és a 3' blokkolót, majd a be nem épült nukleotidokkal együtt lemossák. Ezt a folyamatot ismétlik ciklikusan 50-150 alkalommal, és a nukleotid sorrendeket a kamerához kapcsolt számítógép rakja össze lépésenként. Ezzel a módszerrel a humán genomot 2-3 nap alatt lehet megszekvenálni, ráadásul a költségek is alacsonyabbak mint a hagyományos módszernél.

### **1.3. A mitokondriális genom**

A sejtorganelumok közül a mitokondrium és a növényi kloroplasztisz rendelkezik önálló genommal. Az ember mitokondriális DNS-e (mtDNS) kettősszálú, cirkuláris, 16569 bázispárból áll, és 37 gént kódol. Az mtDNS nagyon kompakt, intronokat alig tartalmaz, a gének átfedhetnek

egymással. Nem kapcsolódnak hozzá hisztonok, melyek védelmet nyújtanának a mutagén hatások ellen. Nincsen excíziós repair rendszere, ami a pontmutációk eliminálásában venne részt, ezért viszonylag magas a mutációs rátája. A körülbelül 800 bázispár méretű hipervariábilis régió (HVR) nem kódol géneket csak szabályozó szerepe van, ezért az itt bekövetkező pontmutációkra kevésbé hat a szelekció, ennél fogva a HVR régió az mtDNS legpolimorfabb szakasza. A mtDNS mutációs rátája a nukleáris DNS-ének körülbelül tízszerese, a HVR szakaszé ennek is többszöröse. A haplotípusok elkülönítésére elsősorban a HVR-I szakaszt (nt 16024-16365) használják, de néhány haplocsoport (H és U) a HVR-II (nt 37-340) szekvenciákban is különbözik. Sejtípustól függően a sejtekben több száz vagy akár több ezer mitokondrium található, mivel ez a sejt energiatermeléséért felelős sejt szervecskéje. Egyetlen mitokondriumban 2-10 DNS molekula is lehet, így az mtDNS egy sejtben 1000-10000 kópiában van jelen (Fernández-Silva és mtsai., 2003). Magas kópiaszáma miatt az mtDNS megőrződése a régészeti maradványokban nagyságrendekkel jobb, mint a sejtenként csak egy kópiában jelen lévő genomi DNS-é. Archeogenetikai szempontból a magas kópiaszámon kívül a mitokondriális DNS másik jelentős tulajdonsága a maternális öröklődés, melynek köszönhetően információt szolgáltat az egyén anyai ágú leszármazási viszonyairól. Az mtDNS öröklődése során nem történik rekombináció, így a szekvenciában bekövetkező változások kizárólag a fokozatosan felhalmozódó mutációknak köszönhetőek.

#### **1.4. Az Y-kromoszóma**

Legfontosabb biológiai szerepe a nem meghatározása és a férfi fertilitás biztosítása. Az Y-kromoszóma haploid, csak a férfiakban van jelen, és apáról fiúra öröklődik. Az Y-kromoszóma 57227415 bázispár hosszú. A kromoszóma hosszának 95%-án nem játszódik le rekombináció a meiózis során az X- és az Y-kromoszóma között. Ezt a szakaszt az Y-kromoszóma nem rekombinálandó régiójának (Non-Recombining region of Y - NRY, Non-Recombining Portion Y - NRPY) hívják. Ezt a régiót mindkét oldalon a kromoszóma telomer szakaszain elhelyezkedő pszeudo-autoszómális régiók (kevesebb, mint 3 megabázisnyi szakasz a kromoszóma körülbelül 60 megabázisnyi hosszából) szegélyezik, melyek rekombinációja az X-kromoszómával a férfi meiózis gyakori és szabályszerű eseménye.

A rekombináció hiányának jelentősége, hogy a haplotípusok, vagyis az Y-kromoszómán található nukleotid változatok kombinációi általában érintetlenül adódnak tovább generációról

generációra. Más szavakkal az Y-kromoszóma nem rekombinálódó régiója egyetlen lókuszként öröklődik. Sajátos jellemzői következtében, melyek egyedivé teszik a kromoszómák között, az Y-kromoszóma hatékony eszköznek bizonyult a populációgenetikuskok számára a humán diverzitás tanulmányozásában és az apai ági leszármazási vonalak nyomon követésében. Mivel szekvenciájuk csak az idővel halmozódó mutációk révén változik, az autoszómális DNS-hez képest az Y-kromoszómák egy viszonylag egyszerűbb genetikai történetet őriznek, ugyanúgy használhatók az apai leszármazási vonalak rekonstruálására, mint az mtDNS az anyai ágon. Az Y-kromoszóma további sajátos jellemzője, hogy 1:1 nemi arány esetén a teljes populációban az Y- várható effektív populációmérete negyede bármelyik autoszómáénak, harmada az X-kromoszómáénak és egyenlő méretű a mitokondriális DNS-ével. Ennek megfelelően az Y-kromoszómális genetikai változatosságot (csakúgy, mint a mtDNS-ét) nagyobb mértékben érinti a genetikai sodródás hatása. A genetikai sodródás felgyorsítja a különböző populációkban az Y-kromoszómális leszármazási vonalak – és mtDNS vonalak - csoportjai közötti differenciációt, ezért a földrajzilag elkülönülő populációk között idővel jellegzetes eltérések jönnek létre (Jobling és Tyler-Smith, 2003). Ennek köszönhető, hogy az Y-kromoszóma – és az mtDNS – haplotípusok gyakorisága lényegesen nagyobb genetikai differenciákat mutat a populációk között, mint az autoszómális markereké. A populációk genetikai vizsgálata kimutatta, hogy az emberi faj genetikai változatosságának zömét populáción belül találjuk, és csak töredékét, 10-15%-át a populációk között, azonban ugyanez az arány az Y-kromoszóma elemzéseknél már 30-40%-nak adódott (Lewontin, 1972). A nagyobb genetikai eltérés nagyobb felbontást tesz lehetővé, ezért az a migrációk nyomon követésében az Y és mtDNS vizsgálatoknak kiemelkedő szerep jut. Az Y-kromoszóma populációk közti nagyfokú diverzitását a társadalmak patrilokalitása is magyarázza. Közismert, hogy a társadalmak körülbelül 70%-ánál amikor két ember egybekel, többnyire a nők váltanak lakóhelyet, ők költöznek a férjükhöz. Más szóval a férfiak genetikusan kevésbé mozgathatók. Ennek következményeként egy jóval homogénebb mitokondriális DNS elterjedési térkép alakult ki, míg az Y-kromoszómák egymástól függetlenül divergálódtak a különböző populációkban (Seielstad és mtsai., 1998). Ez az eredmény is mutatja, hogy az emberi kultúrának milyen hatalmas szerepe van/volt fajunk genetikai mintázatának kialakításában.

## 1.5. Magyar populációk genetikai vizsgálatának eddigi eredményei

A magyar, illetve magyar anyanyelvű népesség körében a populációgenetikai vizsgálatok kezdete elsősorban Czeizel Endre és Béres Judit nevéhez fűződik (Czeizel és mtsai., 1991; Guglielmino és mtsai., 2000; Guglielmino és Beres, 1996). A tanulmányok mind nukleáris genomi markerek, mind mtDNS és Y haplotípus RFLP vizsgálatának eredményeit foglalták össze. A kezdeti, csupán 24 genetikai marker alapján elvégzett kísérletsorozatban számos ellentmondásos, nehezen magyarázható eredményt kaptak. Így például az irániakhoz mind a székelyek, mind a csángók és a kiskunok nagyobb hasonlóságot mutattak, mint a jászok. Béres és mtsai. a magyarok finn-ugor genetikai rokonságának kérdését is vizsgálták Y kromoszómás és mtDNS markerek analízisével (Lahermo és mtsai., 1999), és eredményeik szerint a magyar és finnugor népesség teljesen elkülönül egymástól, amit későbbi hasonló vizsgálatok is megerősítettek (Guglielmino és mtsai., 1990; Semino és mtsai., 2000; Völgyi és mtsai., 2009).

A hazai aDNS kutatások 2000-ben kezdődtek a Szegedi Biológiai Központban, Raskó István kutatócsoportjában. Itthon nekik sikerült először ásatag DNS izolálniuk, továbbá mtDNS haplotipizálási módszert kidolgozniuk (Kalmár és mtsai., 2000). Vizsgálták a kunok eredetét (Bogácsi-Szabó és mtsai., 2005), a magyar nyelvű populációk és honfoglalók genetikai kapcsolatait (Tömöry és mtsai., 2007) és Y-kromoszómás vizsgálatot (Tat) is végeztek honfoglaló mintákon (Csányi és mtsai., 2008). 2003-tól az archeogenetikai vizsgálatokat áthelyezték Budapestre a Régészeti Intézetbe, ahol ezt követően a kutatási területen a napjainkig nem történt módszertani és tudományos előrelépés. A Budapesten végzett munkákból 2016 végén jelent meg az első publikáció (Csósz és mtsai., 2016), melyben a Tömöry és munkatársai által használt módszerekkel egy nagyobb mintaszámú honfoglalás kori és egy kis mintaszámú avar kori anyag mtDNS haplotípusát közölték.

Időközben egyre több recens magyar mintából készült publikáció látott napvilágot amelyekben mtDNS HVR régióját (Brandstätter és mtsai., 2007; Egyed és mtsai., 2007), Y kromoszómás mikroszatellita lókuszokat vizsgálták (Bíró és mtsai., 2009). Utóbbi cikk eredményeit a szerzők később újabb eredmények alapján felülbírálták (Bíró és mtsai., 2015). Ezen vizsgálatok mindegyike megerősítette, hogy a magyar népesség genetikailag a környező európai népességhez hasonlít, és nincs köze a finnugorokhoz. Emellett intenzíven kutatták a finnugor népekre jellemző Y-kromoszómás Tat-C csoportot alcsoportjainak meglétét a magyar és finnugor populációkban (Fehér és mtsai., 2014), és azt találták, hogy ennek egyik alcsoportja



(L1034) kb. 20%-os gyakoriságú a mansik között, és kb. 3%-os gyakoriságúnak mutatkozott a székeliek között, de ugyanez a marker megtalálható az üzbég, és baskír mintákban is, és teljesen hiányzik a többi magyar populációból.

Az idézett publikációk azt tanúsítják, hogy a magyarok genetikai rokonságának intenzíven kutatása ellenére a finnugor rokonság nem nyert megerősítést, a két népcsoportot legfeljebb egy nagyon vékony genetikai szál kapcsolja össze. Ezeket az eredményeket eddig azzal magyarázták, hogy a honfoglalást követően jelentősen lecserélődhetett a magyarság génkészlete. Az eddigi kutatások legnagyobb hiányossága, hogy a régészeti leletek genetikai vizsgálata abbamaradt, miközben a szakterületen forradalmi változások mentek végbe, melyek újgenerációs szekvenálással ma már lehetővé teszik a megbízható nagy felbontású genetikai vizsgálatokat is. Másrészt napjainkig teljesen hiányzik a magyarság népcsoportokra bontott részletes, nagy felbontású genom szintű vizsgálata is. Kutatásaink arra irányulnak, hogy ezeket a hiányokat pótoljuk, és ennek során elsőként a honfoglalók nagy felbontású mtDNS vizsgálatát végeztük el Szegeden, nagyszámú mintán.

## **1.6. Archeogenetika**

A régészeti leletekből kinyert DNS neve az ásatag DNS vagy archaikus DNS (ancient DNA, aDNA), melynek vizsgálata speciális módszereket igényel. A régészeti genetika születése a XX. század végére keltezhető, amikor Higuchi és mtsai., 1984-nak sikerült egy 150 éve kihalt állatból DNS-t izolálni és megszekvenálni. Pääbo, 1985-ben egy 2400 éves egyiptomi múmiából vont ki aDNS-t. 2013-ban egy német kutatócsoportnak sikerült egy több mint 300 ezer éves *Ursus deningeri* (medve) teljes mtDNS genomját rekonstruálni ezzel bizonyítván, hogy megfelelő körülmények között az aDNS sok százezer évig képes megőrződni (Dabney és mtsai., 2013). 2015 novemberében közölték a 110 ezer éves Denisova-i ember teljes genom szekvenciáját, melynek aDNS-ét egyetlen fogból vonták ki (Sawyer és mtsai., 2015).

Az aDNS szekvenciákból világossá vált a neandervölgyi és az anatómiailag modern ember közötti viszony (Juric és mtsai., 2016). Továbbá a neolitikus kori Európa génkészletének származási történetét is sikerült tisztázni (Brandt és mtsai., 2013; Haak és mtsai., 2010), majd az indoeurópai nyelvcsalád eredetét is sikerült genetikai adatokkal alátámasztani (Haak és mtsai., 2015).

Az élő gerincesek szervezetében a csontok 40%-a víz, míg a szárazanyag szerves-organikus és szervesetlen-anorganikus alkotórészekből áll. A csont anorganikus összetevői a szárazanyag tartalom 65%-át teszik ki ami nagyrészt kalcium-foszfátból, magnézium-foszfátból és kalcium-karbonátból áll. A csontszövet szerves állománya (a szárazanyag tartalom 35 %-a) osteocytákból, osteoblastokból, osteoclastokból, és 1-es típusú kollagén-fibrillumokból áll. A csontleletekből származó aDNS legnagyobb mennyiségben a kollagén rostok és kalcium só kristályok közé záródott osteocytákból származik, és megőrződése függ a csont eredeti szerkezetétől valamint a maradványt ért környezeti hatásoktól (hőmérséklet, pH, talajvíz, stb.; Campos és mtsai., 2012; Salamon és mtsai., 2005).

A fog kevésbé porózus mint a csont, ezért a modern DNS-el és bakteriális DNS-el való szennyeződésének esélye is alacsonyabb (Gilbert és mtsai., 2005). Adler és munkatársai kimutatták, hogy a gyökeret körbevevő vékony cement állomány (cementum) mtDNS tartalma ötször magasabb mint a dentiné (Adler és mtsai., 2011). A cementum tömött elmeszesedő szövet mely a foggyökér egészét beborítja. A dentin-cement határon lévő kis csatornában mineralizált cementoblast sejtek őrződnek meg, ezért a pulpa üreget (pulp cavity) körülvevő dentin az egyik legjobb aDNS forrás. Ezért vizsgálataink során lehetőleg a fog gyökeréből végeztük az aDNA extrakciót, és az elsősorban kalcium-foszfátot tartalmazó fogzománcot (enamel) eltávolítottuk.

A csontból és foggyökérből kinyert aDNS-nek csak kevesebb mint 1%-a endogén eredetű. Nemrég mutatták ki, hogy a legmagasabb endogén DNS tartalommal rendelkező csontszövet a halántékcsontról a koponyaalapján található sziklacsont (pars petrosa) (Pinhasi és mtsai., 2015), ezért az utóbbi években egyre inkább ezt használják aDNS forrásként.

A régészeti genetikában a legjelentősebb technikai nehézséget a minták csekély DNS tartalma és a DNS molekulák nagyfokú károsodottsága okozza. Ehhez kapcsolódó problémát jelent a minta modern DNS-el való beszennyezése, amely csak speciális laboratóriumi körülményekkel küszöbölhető ki.

Az aDNS munka speciális labor felszereltséget és technológiai fegyelmet követel. A labornak olyan épületben kell lennie, ahol nem folynak DNS vizsgálatok az élőlény ma élő változataival, hogy kizárható legyen a minták modern DNS-sel történő beszennyezése. Ez fokozottan érvényes a humán DNS-re, ahol a modern DNS szennyezés számtalan forrásból származhat, beleértve a régészt, antropológust, és a molekuláris biológust. Az aDNS laborban steril légfűtéssel enyhe túlnyomást kell létrehozni, hogy az ajtók nyitásakor megakadályozzuk

a kintről befelé irányuló légáramlást, amely szennyezések forrása lehet. A laborban minden munkafázis elkülönülő légtérben folyik, és az anyagok mozgata szigorúan egyirányú. Az anyagok és személyek először az aDNS laborba lépnek be, és csak ezt követően kerülhetnek át a külön épületben található molekuláris laborba. A molekuláris helyiségből az aDNS laborba nem kerülhet vissza anyag, mert azok már amplifikált humán DNS-sel szennyezettek lehetnek. Az aDNS laborba minden eszköz és anyag csak alapos DNS-mentesítés után kerülhet be, ami hipós törléssel és UV kezeléssel történik. A steril környezet fenntartása érdekében minden dolgozónak csuklyás overált, maszkot, cipővédőt és kesztyűt kell viselnie. Minden felületet le kell takarítani a munka elvégzése előtt és után. Elsőnek 3-10%-os hipóval mossuk a felületeket, majd 70%-os etanollal, hogy letisztítsuk a hipót és elkerüljük a korróziót, majd UV-sugárzás következik (Fulton, 2012). A felszínek, falak tisztántartása érdekében éjszakánként mindenütt UV lámpák működnek, és rendszeres hipós mosással kell tisztítani a helyiségek falait és padozatát. A DNS munkák és a PCR reakciók összemérése külön erre a célra szolgáló DNS-mentes PCR fülkék alatt történik. Cooper és Poinar, 2000-ben egy kilenc pontból álló listában foglalták össze a hiteles aDNS munka követelményeit, és az alább felsorolt kritériumok szigorú betartását javasolták, melyeket mi is fokozottan betartottunk:

- A munkafázisok térbeli elkülönítése: A pre-PCR labor kizárólag aDNS munkára használt steril labor lehet, a post-PCR labor lehetőleg külön épületben kapjon helyet.
- Negatív kontrollok minden lépéshez: Extrakciós és PCR kontrollok annak ellenőrzésére, hogy egyik munkafolyamat során sem kontaminálódott a minta, és a vegyszerek is DNS mentesek. Ha a negatív kontroll DNS-t tartalmaz, a kísérlet sikertelennek minősítendő. Pozitív kontroll használata nem ajánlott.
- Molekuláris adatok helyes értelmezése: A PCR sikere aDNS-ből elméletileg fordítottan arányos a fragment mérettel, 500-1000 bp fragmentek sikeres amplifikálásának valószínűsége elenyésző. Ha genomi DNS-t sikerül amplifikálni, elvárható, hogy a mtDNS is jól detektálható legyen. Az adatoknak filogenetikailag értelmezhetőnek kell lenniük.
- Reprodukálhatóság: Ugyanazon mintából független extrakciónak ugyanazon eredményt kell adnia más primerekkel megismételve is.
- A termék klónozása: A tipikus aDNS károsodások, tranzíciók kimutatására.

- Független ismétlés: A minták egy részéből független laboratóriumban, más kutatók által kell tudni reprodukálni az eredményeket.
- Biokémiai megőrződés: Egyes biomolekulák megőrződése (például kollagén vagy aminosav racemizáció) összefüggést mutat a DNS megőrződésével, és valószínűsíti a DNS kihozatalát.
- Mennyiségi meghatározás: A kompetitív PCR és a valós-idejű (real time) PCR utal a reakcióban lévő kiinduló templát szálak mennyiségére. A szennyezés mentes aDNS-ben a nagyobb molekulák (150 bp) alacsony aránya jelentősen alacsonyabb a kisebb molekulákhoz (100 bp) képest.
- Társ-maradványok állapota: A mintával együtt eltemetett állati csontok vagy egyéb szerves maradványok hasonló minőségben kell, hogy megőrződjenek.

Ezeket a kritériumokat Llamas és mtsai., 2017-ben fogalmazták újra NGS vizsgálatok szempontjából:

- Negatív kontrollok használata: Extrakciós kontrollokat és templát nélküli PCR kontrollokat kell alkalmazni, hogy a labor esetleges DNS kontaminációját kimutathassuk (Green és mtsai., 2009; Kircher, 2012; Knapp és Hofreiter, 2010)
- Kísérletek ismétlése: Ugyanazon mintából több aDNS extraktum készítése, vagy azonos aDNS extraktumból több NGS könyvtár készítése (Meyer és mtsai., 2014; Orlando és mtsai., 2011). Ezen ismétlések elvégzése után az eredményeknek azonosnak kell lenniük.
- Fragment hossz: Az aDNS degradációja miatt minél rövidebbek a fragmentek annál nagyobb mennyiségben fordulnak elő, a fragmentek többsége 100 bp-nál rövidebb (Sawyer és mtsai., 2012).
- DNS megőrződés és a variánsok lefedettség mélysége: PCR termék klónozását kiváltja az NGS szekvenálás, de ettől még a kritériumoknak meg kell felelnie. Azaz meg kell határozni az endogén-exogén szekvenciák arányát, a DNS károsodásokat detektálni kell, mennyiségüket meghatározni, és a mintára jellemző variánsoknak többszörös lefedettségben kell megjelenniük a különböző egyedi szekvencia olvasatokban.

- Társ-maradványok: A vizsgált lelettel együtt feltárt nem humán minta kontrollként alkalmazásával a modern humán DNS kontamináció mértéke kimutatható (Rasmussen és mtsai., 2010; Skoglund és mtsai., 2012).
- Kontamináció becslés: Modern humán DNS kontamináció szintje megbecsülhető elhunyt humán NGS eredményekből azon szekvencia pozíciók alapján, amelyek jól jellemezték a recens humán populációkban. Ebben az esetben a haploid lókuszoknál (mtDNS, Y-kromoszóma), egy adott pontban egy allélt várunk. Egy másik allél jelenléte kontaminációra utal, és ebből ki lehet számítani a kontamináció mértékét (Fu és mtsai., 2013; Kousathanas és mtsai., 2016).
- Szekvencia variánsok validálása: A meghatározott haplotípusoknak illeszkedniük kell a filogenetikai fákba. A rosszul illeszkedő szekvenciák módszertani hibára, vagy eddig azonosítatlan filogenetikai vonalra utalnak, ezért ilyen esetekben a kísérleteket egy független laborban kell megismételni, különben előfordulhat, hogy hibás eredmények kerülnek bemutatásra. („Erratum for the Report “Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa” (previously titled “Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent”) by M. Gallego Llorente, E. R. Jon”, 2016; Llorente és mtsai., 2015).

## 2. Célkitűzések

Az elmúlt években a Szegedi Tudományegyetem Természettudományi és Informatikai karán, a Genetikai és az Embertani Tanszék együttműködésének keretében létrehoztunk SZTE Archeogenetikai Laboratóriumát, amely speciális archeogenetikai vizsgálatokra alkalmas. Első munkáim során célul tűztem ki az ásatag DNS izolálás módszertanának optimalizálását, és a legújabb haplotipizálási módszereket adaptálását. Ezt a célkitűzést néhány hónap alatt sikerült megvalósítanom. Rutinra tettünk szert az ásatag csontokból izolált mitokondriális DNS haplotípusának meghatározásában, melynek során a szokásos HVR szakaszok szekvenálását kiegészítettük a kódoló szakaszok SNP haplotípus meghatározásával a legújabb SNaPshot technika alkalmazásával. Ugyanezen módszerrel a jobb megtartású csontokból, meg tudtuk határozni az apai ág (Y-kromoszóma) haplocsoportját is.

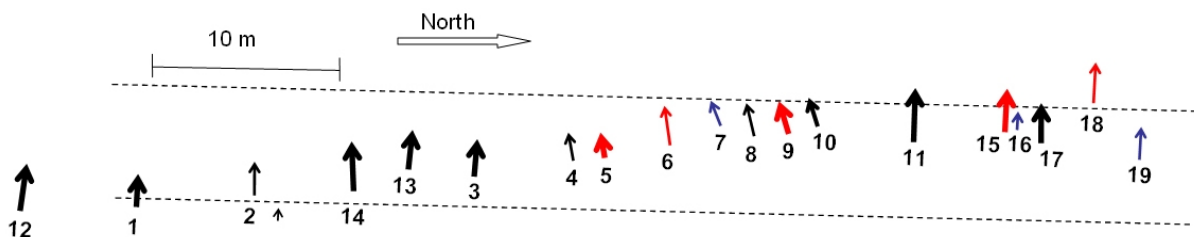
Az archeogenetikai kutatások területét az utóbbi néhány évben forradalmasította az újgenerációs szekvenálás (NGS) módszer alkalmazása. A laborban az elmúlt években elsajátítottuk az NGS könyvtár készítés módszerét is. Ezek a könyvtárak csak néhány % humán DNS-t tartalmaznak, ezért a szekvenálás előtt célszerű a humán DNS tartalmat emelni. A könyvtárakból sikerrel dúsítottuk a teljes mtDNS genomot, melyet ily módon elérhető áron tudunk megszekvenálni. Továbbá egy közel 300 SNP tartalmazó Y-kromoszómás dúsító kitet is összeállítottunk, melynek segítségével lehetővé vált az Y-kromoszóma dúsítása, és a férfi ágú haplocsoport nagy mélységű besorolása.

Ezt követően azt a célt tűztük ki, hogy a beállított módszerekkel egy kellően reprezentatív mintaszámon elvégezzük a honfoglalás kori minták nagyfelbontású genetikai vizsgálatát. Célunk között szerepelt néhány temető teljes anyagának vizsgálata, melyből a temetőn belüli és temetők közötti rokonsági viszonyok is felderíthetők. Az így kapott eredmények hozzájárulhatnak a honfoglalók származásának kiderítéséhez, valamint összehasonlítási alapul szolgálhatnak a további temetőkben és korszakokban később elvégzendő vizsgálatok számára.

### 3. Anyagok és Módszerek

#### 3.1. Mintavétel

Első kísérleteink során a Karos-III. számú temetőjének (3. ábra; Révész, 1996) 19 mintáját vizsgáltuk alacsony felbontású (HVR, SNaPshot) genetikai módszerekkel. Ennek során tömör csontokból, elsősorban combcsontból vettünk mintát. A későbbi NGS szekvenálásokhoz már foggyökereket is használtunk, ahol rendelkezésre állt.



**3. ábra** A Karos-III. temető sorba és csoportokba rendezett sírjai. A fekete nyilak a férfiakat, a pirosak a nőket, a kék pedig a gyerekeket jelölik. A vastag nyilak az idős korban elhunytakat jelölik.

A továbbiakban a Karos-I. és Karos-II. számú temetőre is kiterjesztettük a vizsgálatainkat, itt is foggyökerekből vettünk mintát, ahol ez rendelkezésre állt. Ebből a három temetőből az összes feltárt, fennmaradt leletet bevontuk a vizsgálatokba (Karos-I: 11 sír, Karos-II: 69 sír, Karos-III: 19 sír; Révész, 1996). Ezt követően néhány további honfoglalás kori temető anyagát - Kenéz-lő-Fazekaszug-II, Harta-Freifelt, Magyarhomoróg, Orosháza-Görbicstanya, Szabadkígyós-Pálliget és Sárrétudvari-Hízóföld (4. ábra) - is bevontuk a vizsgálatokba. A Kenéz-lő-Fazekaszug-II és Sárrétudvari-Hízóföld temetők esetében, ahol lehetett, már sziklacsontról készítettük az aDNS extrakciót.



**4. ábra** A vizsgált temetők elhelyezkedése.

## 3.2. A csontminták porítása

### 3.2.1. Porítás Dremel kéziszerszámmal

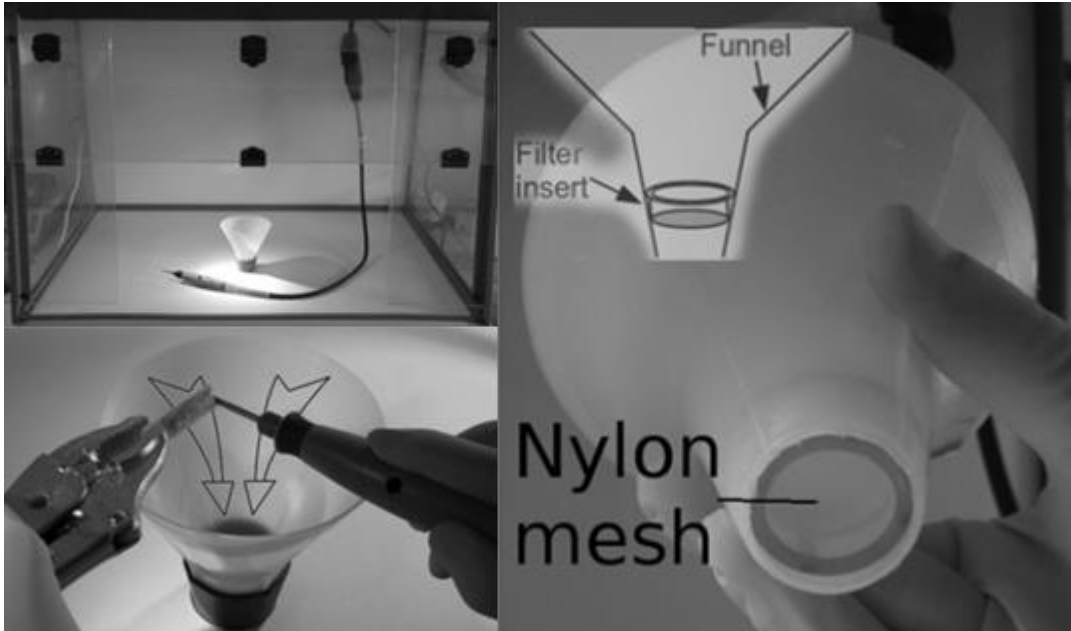
Benoit és munkatársai kidolgoztak egy olcsó, de hatékony csontporítási módszert, ami elszívó fülke alatt egy kisméretű fűrészárral történik (5. ábra; Benoit és mtsai., 2013). A csont és fogminták alufóliába csomagolva vagy légmentesen zárható tasakban kerülnek az aDNS laborba. A beléptető helyiségben a fóliát illetve tasakot hipóval áttöröljük, majd a csontokat, fogakat mindkét oldalról közelről UV kezeljük 20 percig. A kezelt csontok visszacsomagolva kerülnek az őrlőhelyiségbe.

A szennyezések elkerülésére a Dremel (DREMEL 3000) szerszám fűrészár mögötti részét gumikesztyűbe húzzuk. Bekapcsoljuk a légelszívót és ez alatt tisztítjuk meg a csont felszínét, körülbelül 1 mm vastagságban, 3\*2 cm-es felületen. A csontfelszín tisztítására Dremel ref. 9903 Wolfram-karbid vágófejet, a foggyökér finomabb tisztításához drótkéfés fejet használunk. A keletkező por a külső porszívó egységbe távozik.

A porítás előtt elszívó tölcserét cserélünk, és a tölcser alá beillesztünk egy mikroszűrőt (40 µm nylonháló). A Dremel-ben fejet cserélünk, és a ref. 570 számú karbid fugaeltávolítóval porítjuk a csontot. A tisztított felületen a fugaeltávolító fej oldalával, finom mozdulatokkal a tölcser fölött porítjuk a csontot vagy a foggyökereket, így a por a tölcserbe kerül és az átlag 100 µm-es méretű szemcsék a mikroszűrőben gyűlnek össze (5. ábra). Alacsony fordulatszám használata javasolt, mert magas hőmérséklet hatására károsodhat a DNS. Az aDNS kivonáshoz 200 mg-tól 1 g csontpor szükséges. Az elszívó kikapcsolása után csipesszel kivehető a mikroszűrő és a csontport egy tiszta feliratozott csőbe öntjük. Az extrakciós kontroll csövét a fülke alatt kinyitjuk, és becsukjuk. Ez a levegőben vagy a később használt vegyszerekben esetleg jelenlévő háttér DNS szennyezés kimutatására szolgál.

Ezt követően hipós vízzel, majd 3%-os hidrogén peroxiddal DNS mentesítjük, és UP vízzel öblítjük a munkafelületet, a tölcserét, a szűrőt, az elszívó fejet, a fűrészárakat, a csipeszt, és az elszívó csatlakozó csövét. A megtisztított szerszámokat 20 percig UV kezeljük, majd alufóliába csomagolva tároljuk. A fülkét szintén 20 percig UV kezeljük.





**5. ábra** Csontporítási módszer Dremel multifunkcionális szerszámmal (Benoit és mtsai., 2013)

### 3.2.2. Porítás csontmalommal

A megtisztított felületű csontból kisebb darabokat szelünk tiszta gyémánt koronggal, majd 2 x 20 percig UV kezeljük a felszíneket. A csontszeletet Dremel vágókoronggal aprítjuk a mozsárba. A malomba maximum 8 mm átmérőjű darabok tehetők, melyeket az őrlés során kb. 5 µm-es szemcseméretre porítunk. A csontporításra az irodalom szerint a rázó malmok a legmegfelelőbbek, mi ezek közül a Star Beater VWR golyósmalmot használjuk.

### 3.3. DNS kivonás

Miután a csontporítás megtörtént, minden mintából két vagy három DNS kivonatot készítettünk szilika szuszpenziós módszerrel (Rohland és Hofreiter, 2007). A csontporokra extrakciós puffert (EB) mértünk (0,45 M EDTA, 250 µg/ml Proteináz-K, 1% TritonX-100, 50 mM DTT). Az EB pufferben a csontpor feloldódik, 48 °C-on overnight forgatás során. Másnap a DNS oldatot centrifugáltuk, ami eltávolítja a fel nem oldódott frakciót. Ezután a DNS tartalmú felülúszóhoz kötő puffert (BB, melynek összetételén változtattuk az egyes kivonásoknál, amint azt a későbbiekben részletezem) és 150 µl szilika szuszpenziót adtunk. A kötő puffer pH-ját beállítottuk 4-6 közé, GuSCN BB esetén tömény sósavval, GuHCl BB esetén ecetsavval. A DNS kikötése a szilika szuszpenzióhoz 3 órán keresztül történt szobahőn, sötétben forgatva. Ezután a

szilika szuszpenziót a hozzá kötött DNS-sel centrifugáltuk, a pelletet kétszer mostuk 80% etanollal, végül a DNS-t 100 µl előmelegített TE-be eluáltuk. Az egyes kivonási módszerek eltéréseit alább részletezem:

### **3.3.1. DNS kivonás 1**

Ebben az esetben a DNS-t 200-400 mg femur vagy metatarsus csontporból vontuk ki, a porhoz 4 ml EB-t adtunk. A DNS kötés 16 ml BB (5 M GuSCN, 25 mM NaCl, 50 mM Tris) és szilika hozzáadásával történt.

### **3.3.2. DNS kivonás 2**

Itt a DNS kivonás 100-150 mg foggyökér csontporból készült, a porhoz 1 ml EB-t adtunk. A DNS kötés 5 ml BB (5 M GuSCN, 25 mM NaCl, 50 mM Tris) és szilika hozzáadásával történt.

### **3.3.3. DNS kivonás 3**

Ennek során a csontport előemésztettük, hogy eltávolítsuk a porszemcsék felszínén meglévő külső DNS szennyeződést, ami növeli az endogén DNS tartalom arányát (Damgaard és mtsai., 2015). A DNS kivonás 400-600 mg tömör csontból történt. A csontport először egy órán át mostuk 48 °C-on 8 ml 0,5 M EDTA-val, majd előemésztettük 20 percig 48 °C-on 4 ml 0,5 M EDTA és 100 µg/ml Proteináz K-val. Ezután a centrifugált üledékhez 3 ml EB-t adtunk, és 48 °C-on overnight forgattuk. Másnap centrifugáltuk, és a felülúszóból a DNS kötés 12 ml BB (5.2 M GuHCl, 100 mM NaOAc, 32% izopropanol) és szilika hozzáadásával történt.

### **3.3.4. DNS kivonás 4**

Ezt a módszert Bolzano-ban alkalmaztuk. Ebben az esetben a DNS kivonás 250 mg csontporból készült. A porhoz 1 ml EB-t adtunk (Triton-X és DTT nélkül), majd 48 °C-on overnight forgattuk, centrifugáltuk, és a felülúszóhoz 10 ml BB-t (4 M GuHCl, 32% izopropanol, 4% Tween-20, 100 mM NaOAc) és 100 µl szilika szuszpenziót adtunk. A mosó puffer összetétele: 125 mM NaCl, 10 mM Tris, 1 mM EDTA, és 50% etanol volt.

### **3.3.5. DNS kivonás 5**

Ezt használtuk minden NGS kísérlethez. A DNS kivonás 100 mg foggyökérből, sziklacsontból esetleg egyéb tömör csontból történt. A csontport minden esetben előemésztettük 1 ml 0,5 M EDTA és 100 µg/ml Proteináz K-val 30 percig 48°C-on. Ezután centrifugáltuk, és az előemésztett csontpor üledékhez 1 ml EB-t adtunk, majd overnight feltártuk. Másnap centrifugálást követően a felülúszóból a DNS-t 6 ml BB (5,83 M GuHCl, 105 mM NaOAc, 46,8% izopropanol, 0,06% Tween-20) és szilika hozzáadásával kötöttük ki.

## **3.4. PCR alapú haplotipizálás**

### **3.4.1. mtDNS haplotipizálás**

Az mtDNS HVR-I kontroll régiójának 413 bp hosszú szakaszát (np. 15,997–16,409) négy átfedő primer pár segítségével amplifikáltuk, továbbá a HVR-II régióból is amplifikáltunk egy 225 bp-os szakaszt (np. 172–327; Függelék 1. táblázat; Haak és mtsai., 2008). A PCR reakciót agaróz gélen ellenőriztük, és amennyiben megfelelő méretű amplifikátumot kaptunk - miközben a negatív kontroll nem adott jelet - a PCR termékeket ExoSAP (Exonukleáz-I és FastAP keveréke) reakcióval tisztítottuk a gyártó utasításai szerint (Thermo Scientific), majd elküldtük szekvenálásra. Minden mintából legalább két különböző DNS kivonatból ismételtük meg a szekvenálásokat.

A HVR szekvenáláson kívül mindegyik minta esetén legalább két DNS kivonatból meghatároztuk 22 kódoló régió SNP pozícióit is a GenoCoRe22 módszerrel a Haak és mtsai., 2010.-ben leírtaknak megfelelően (Függelék 2. és 3. táblázat). Minden PCR reakcióban GoTaq G2 Hot Start Polimerázt (Promega) és BSA-t (NEB) használtunk az AmpliTaq Gold (Applied Biosystems) és RSA helyett.

### **3.4.2. Y kromoszóma haplocsoport meghatározás**

Az Euráziában leggyakoribb Y haplocsoportokat a GenoY25 módszerrel azonosítottuk a Haak és munkatársai által leírtak szerint (Haak és mtsai., 2010). Az eredeti módszerhez képest csak a multiplex PCR reakció primer koncentrációit módosítottuk a következők szerint: Az M168, M2, M89, és M17 PCR primereket 0.03 µM-os; Az M304 és M9 primereket 0.04 µM-os, az M242, M45, M343, és M175 primereket 0.05 µM-os; és az M35 primert 0.1 µM-os

koncentrációban használtuk, mivel így a kapott PCR termékek koncentrációja kiegyenlítettebb lett (Függelék 4. táblázat). Itt is legalább két különböző DNS kivonatból végeztük el a kísérletet (Függelék 5. táblázat).

A multiplex PCR-ből gyakran hiányoztak egyes amplifikátumok, ezért a hiányzó vagy bizonytalan pozíciók meghatározására single-plex PCR reakciókat is összeállítottunk, melyekben minden esetben az adott primer párhoz optimalizáltuk a PCR reakció körülményeit. Minden single-plex reakciót 25 µl térfogatban végeztünk, ami 2 µl DNS extraktumot, 2.5 mM MgCl<sub>2</sub>-t, 6 µM primereket, 0.25 mM dNTP-ét, 1 mg/ml BSA-t, 1 U GoTaq G2 Hot Start Polimerázt és Flexi puffert tartalmazott. A primerek optimalizált annealing hőmérsékletei a következők voltak: az M170, M126, M304 primereké 62 °C, az S21 primeré 57 °C, és az M172 primeré 55 °C. A PCR program a következő lépésekből állt: 95 °C 2 perc denaturáció, majd 50 ciklus a következő lépésekkel: 95 °C 20 másodpercig, annealing 30 másodpercig, 72 °C 30 másodpercig, majd végső extenzió 72 °C 5 percig. Minden PCR reakció esetén templát nélküli negatív kontrollal zártuk ki, hogy a PCR termék nem külső DNS szennyeződésből származott. A PCR termékeket agaróz gélből tisztítottuk, majd az SNP pozíciót SBE reakcióval határoztuk meg, a GenoY25 módszerben leírtak szerint (Függelék 4. táblázat), de ez esetben reakciókhoz 1 mg/ml BSA-t is adtunk.

### **3.4.3. STR**

A temetőn belüli lehetséges rokonsági viszonyok megállapítására autoszómás STR, és Y-kromoszóma STR kísérleteket is végeztünk (Függelék 6. táblázat). Az Y-STR eredményekből az Y haplocsoportra is lehet következtetni, ami független módszerként szolgál a GenoY25 eredmények megerősítésére is. Az Y-kromoszóma amplifikálást a budapesti Bűnügyi Szakértői és Kutatóintézetben kivitelezték a PowerPlex Y23 kittel (Promega), ami 23 Y-STR lókuszt vizsgál. Az STR fragment méreteket és allélokat Genetic Analyserrel határozták meg (3130 és 3500 Life Technologies, Foster City, CA, USA), a GeneMapper ID-X v.1.4 szoftver segítségével. Az STR allélek alapján az Y-kromoszóma haplocsoport meghatározása az FTDNA Haplogroup Predictor Program-jával történt.

Az autoszómális STR analízist a Szegedi Tudományegyetem, Igazságügyi Orvostani Tanszékén a PowerPlex ESX 17 System kittel (Promega) végezték, a gyártó utasításait követve. Az STR fragmentek méretét ABI PRISM 310 Genetic Analyser (Life Technologies) kapilláris

elektroforézises berendezéssel határozták meg a GeneMapper ID v3.2 (Applied Biosystems) szoftver segítségével (Függelék 6. táblázat).

#### **3.4.4. Az aDNS adatok hitelessége**

Az ásatag DNS munkákat az erre célra kialakított steril légbefújással rendelkező, elkülönített helyiségekből álló aDNS laborban végeztük a Szegedi Tudományegyetem Genetikai Tanszékén, figyelembe véve a bevezetésben részletesen bemutatott követelményeket (Knapp és mtsai., 2012). A PCR reakció összeállítását követő „poszt-PCR” munkákat az épület távoli, elkülönített szárnyában, az SZTE Embertani Tanszék helyiségében kialakított molekuláris laborban végeztük.

A kísérletek során számos külső DNS szennyezést kiküszöbölő stratégiát alkalmaztunk (Champlot és mtsai., 2010): a vizet sugárkezeltek (5 kGy gamma sugárzással), az enzimeket, dNTP és primereket hőérzékeny hl-dsDNÁzzal kezeltük és a puffereket UV-val sugároztuk. Egy minta több DNS extraktumából is elvégeztük a haplotipizálást, valamint minden PCR reakció szennyezésmentességét templát nélküli PCR-rel ellenőriztük. Csak azokat az eredményeket fogadtuk el ahol a negatív kontroll nem adott PCR terméket.

Három minta esetében (15, 16, és 19) a DNS kivonást és a HVR haplotipizálást megismételtük az EURAC laborjában Bolzano-ban, ami a szegedivel azonos eredményt adott. A SNaPshottal meghatározott négy Y-kromoszómás haplocsoportból kettőt Y-STR analízissel is megerősítettünk.

Meghatároztuk a kísérletet végző kutatók haplotípusát, mint elsődleges lehetséges DNS szennyezés forrást, és azok egyik mintával sem mutattak egyezést. Nagyszámú, ma ázsiai elterjedtséget mutató mitokondriális haplotípust sikerült azonosítanunk, melyek nagyon ritkák a recens magyar populációkban, ezért nem származhattak szennyezésből. Mindezekből az következik, hogy az általunk meghatározott haplotípusok nem külső szennyeződésből származnak, hanem valóban a vizsgált mintákból kivont DNS-re jellemzők.

## 3.5. NGS alapú haplotipizálás

### 3.5.1. Könyvtár készítés

A könyvtár 50 µl DNS kivonatból indult ki, vagyis 50 mg fogporból kinyert DNS-ből. Ez átlag 0,1-0,8 ng/µl össz DNS-t tartalmazott, melynek csak 1-5%-a humán DNS, azaz egy szekvenáló könyvtár kb. 0,05–0,4 ng humán DNS-t, és 20-100-szor ennyi talajmikroba DNS-t tartalmaz. A DNS extraktumot először részleges uracil-DNS-glikoziláz (UDG) kezelésnek vetettünk alá, amely eltávolítja az aDNS-re jellemző nagymennyiségű deaminált citozin (=uracil) nagyrészt, melyek nélkül timinként jelentkeznének a végső szekvenciában. Ezután az aDNS-re jellemző sérült molekula végeket blunt-end repair reakcióval hoztuk ligálható állapotba (Rohland és mtsai., 2015). Ezután a javított DNS-t MinElute oszlopon (Qiagen) tisztítottuk, és adapter ligálással kettős szálú könyvtárat készítettünk azokból, a Meyer és Kircher, 2010-ben leírtaknak megfelelően. Ez első lépésben két rövidebb adapter ligálását jelenti, amit a megmaradó nickek és egyesszálú szakaszok „befoltozása” követ Bst polimerázzal az ún. adapter fill-in reakcióban. Ezután a könyvtárat előamplifikáltuk 2x50 µl PCR reakcióban, mely 800 nM-os IS7 és IS8 primereket, 200 µM dNTP mixet, 2 mM MgCl<sub>2</sub>, 0,02 U/µl GoTaq G2 Hot Start Polimerázt (Promega) és 1X GoTaq puffert tartalmazott. A PCR program a következő lépésekből állt: 96 °C 6 perc, 16 ciklus ezen lépésekből: 94 °C 30 másodperc, 58 °C 30 másodperc, 72 °C 30 másodperc, majd a végső extenzió 64 °C 10 perc. Ezt ismét MinElute tisztítás követi, melynek végén a könyvtárat 50 µl 55 °C-os EB pufferbe (Qiagen) eluáltuk, és a koncentrációjukat megmértük Qubit-al (Thermo Fisher Scientific). Amennyiben a könyvtárak koncentrációja 5 ng/µl alatt volt, újra amplifikáltuk azokat 5-12 ciklussal a koncentráció függvényében. Végül mindegyik preamplifikált könyvtár koncentrációja 10-50 ng/µl közé került, 50 µl térfogatban.

Ezt követően 50 ng preamplifikált könyvtárat kettős indexáltunk (Függelék 8. táblázat) Kircher és mtsai., 2012 szerint. Ez úgy zajlik, hogy egy újabb PCR reakcióban további adapter szekvenciákat illesztünk a molekulák végére, amely minden egyes mintánál egyedi szekvencia azonosítókat, „indexeket” is tartalmaz, ami a minták keverése és együttes amplifikálása, szekvenálása mellett is biztonsággal lehetővé teszi az egyes mintákhoz tartozó szekvenciák azonosítását. Az 50 µl-es indexáló PCR reakció 1000 nM P5 és P7 indexáló primereket és 1 x KAPA HiFi HotStart ReadyMix-et (Kapa Biosystems) tartalmazott. A PCR program a következő lépésekből állt: 98 °C 3 perc, 6 ciklus ezen lépésekből: 98 °C 20 másodperc, 66 °C 10 másodperc,

72 °C 15 másodperc, majd a végső extenzió 72 °C 30 másodperc. Az indexált könyvtárakat MinElute oszlopon tisztítottuk, majd a koncentrációjukat meghatároztuk Qubit-al, végül a fragment eloszlásukat Agilent 2200 TapeStation Genomic DNA ScreenTape-en ellenőriztük. Mivel a további lépések során a minták keverése jelentős költség és munka megtakarítást jelent, a kettős indexálás az eredmények megbízhatósága szempontjából nélkülözhetetlen.

### **3.5.2. mtDNS dúsítás és szekvenálás**

A hibridizációs dúsítás az endogén humán DNS arányának emelésére szolgál. Az mtDNS dúsításhoz biotinizált mtDNS csalikat készítettük 3 átfedő long-range PCR termékből (Maricic és mtsai., 2010) alapján, de a következő primer párokkal: L14759-H06378, L10870-H14799, L06363-H10888 (Haak és mtsai., 2010).

A dúsítást a Maricic és mtsai., 2010-ben írtaknak megfelelően végeztük, a következő módosításokkal: a mindegyik könyvtár molekula végén meglévő, egymással hibridizálni képes adapterek lefedésére az alább felsorolt 4 blokkoló oligót használtuk 3 µM-os koncentrációban:

BO1.P5.part1F: AATGATACGGCGACCACCGAGATCTACAC-Phosphate,

BO2.P5.part2F ACACTCTTTCCCTACACGACGCTCTTCCGATCT-Phosphate,

BO4.P7.part1 R GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-Phosphate és

BO6.P7.part2 R CAAGCAGAAGACGGCATACGAGAT-Phosphate.

Egy hibridizációs dúsítás során 300 ng biotinizált csalihoz 30 µl Dynabeads MyOne Streptavidin C1 mágneses gyöngyöt (Thermo Fisher Scientific) adtunk. Egy dúsításban 20 kettős indexált könyvtárat kevertünk össze (mindegyikből 300-300 ng-nyit), majd azokat MinElute oszlopon koncentráltuk. Ha kevesebb mintát dúsítottunk együtt, arányosan kevesebb csalit használtunk. A hibridizáció Agilent hibridizációs oldatban történt, 64 µl végtérfogatban, 2 napig 65 °C-on forgatva, melynek során a denaturált könyvtár molekulák adapter végeit a fenti blokkolók fedik, így az egyesszálú könyvtár molekulák homológ szekvenciái a streptavidin mágnesgyöngyökhöz kihorgonyozott biotinizált humán mtDNS csalihoz hibridizálnak. Mosás után a gyöngyökhöz kikötött dúsított könyvtárakat 20 µl vízben reszuszpendáltuk, majd a DNS leoldása a gyöngyökről 60 µl-es PCR reakcióban történt, mely 2000 nM IS5 és IS6 primereket és 1 x KAPA HiFi HotStart ReadyMix tartalmazott. A PCR program a következő lépésekből állt: 98 °C 1 perc, 10 ciklus a következők szerint: 98 °C 1 perc, 60 °C 30 másodperc, 72 °C 30 másodperc, majd a végső extenzió 72 °C 30 másodperc. Ez a PCR reakció kettős szálúvá írja az

egyes szálú dúsított molekulákat, és egyben olyan mennyiségűre amplifikálja azokat, hogy a MiSeq platformra felvihetők legyenek. A dúsított, felszorzott és összevont könyvtárakat MinElute oszlopon tisztítottuk és 15 µl EB pufferbe oldottuk le.

A könyvtárak koncentrációját Qubit-al határoztuk meg, a minőségüket Agilent 2200 TapeStation Genomic DNA ScreenTape-en ellenőriztük, és amennyiben a molekulák méret eloszlása és koncentrációja az elvárásoknak megfelelő volt ezután kerülhetett sor a szekvenálásra. Ezt a SeqOmics Kft. végezte, MiSeq szekvenáló platformon MiSeq Reagent Kit v3-el (Illumina, MS-102-3003), amely 2x150 bp paired-end szekvenciákat állít elő, vagyis a molekulákat mindkét végről megszekvenálja, és az adapterek leolvasása után a szekvenciákat egyedi mintánként csoportosítja.

### **3.5.3. Genomi dúsító kit**

Az Y-kromoszóma haplocsoportok meghatározása céljából az Y-kromoszóma informatív SNP-t tartalmazó szakaszait is dúsítottuk hibridizációval. A sokkal kevesebb kópiában jelenlévő genomi aDNS dúsítására RNS csalikot használtunk, mivel a DNS-RNS hibridizáció hatékonysága jobb. Összesen 191 db filogenetikailag informatív SNP választottunk ki, amely az összes főcsoport meghatározásához szükséges SNP-t tartalmazta, de az Eurázsiaiiban gyakori haplocsoportok esetén nagyszámú alcsoport SNP-jét is (Van Oven és mtsai., 2014; Függelék 9. táblázat). Az Y-kromoszómás SNP-kkel egyszerre néhány autoszómás SNP-t is dúsítottunk, ezek között megtalálható az egyének (és rokonok) azonosításához használható 52 SNPforID (Amigo és mtsai., 2008; Sanchez és mtsai., 2006), továbbá a laktóz intoleranciával kapcsolatba hozott SNP-ék (Enattah és mtsai., 2002), néhány haj, bőr, szemszín meghatározáshoz használatos marker (Edwards és mtsai., 2010; Valenzuela és mtsai., 2010) továbbá néhány rákkal kapcsolatba hozott genomi régiót (p53, K-ras, retinoblastoma stb.).

Az SNP-kre tervezett szintetikus csalik 120 nukleotid hosszú részt fednek le úgy, hogy ezt a szakaszt három átfedő 80 nukleotid hosszú részre vágjuk. A gyártó számítógéppel ellenőrizte a tervezett csalik minőségét és specifikusságát (MyBaits kit, MYcroarray, USA).

### **3.5.4. NGS szekvencia kiértékelés**

Először a szekvencia olvasatokról (read) a cutadapt szoftverrel (Martin, 2011) levágtuk az adaptereket. A read-ek minőségét FastQC programmal állapítottuk meg (Andrews, 2016). Az



adatkészletből eltávolítottuk a 25 nukleotidnál rövidebb read-eket, melyek kis méretük miatt rosszul illeszthetők. A vizsgálatra kész read-eket a humán GRCh37.75 mtDNS referencia genomra térképeztük a Burrows Wheeler Aligner (BWA) v0.7.9 szoftverrel (Li és Durbin, 2009). A GRCh37.75 human referencia genom az mtDNS revised Cambridge Reference Sequence-át (rCRS, NC\_012920.1) tartalmazza (Andrews és mtsai., 1999). Ezután eltávolítottuk a gyengébben illeszkedő read-eket. Megjegyzendő, hogy az esetleges sejtmagi eredetű rosszul illeszkedő mitokondium szekvenciák (NUMT) jelenléte az aDNS-ben elenyésző. Az illesztett szekvenciákból ún. BAM fájlok készültek, melyek szortírozása és jelölése a Samtools 1.1 (Li és mtsai., 2009) programmal készült. A szekvenciákban megjelenő PCR duplikátumokat a Picard Tools v 1.113 (Wysocker és mtsai., 2013) programmal távolítottuk el. Ezt követően megbecsültük az ásatag DNS-re jellemző DNS károsodások (elsősorban C>T, G>A tranzíciók) arányát a MapDamage 2.0 szoftverrel (Jónsson és mtsai., 2013). Ezek megléte elsősorban a molekula végeken bizonyítja a DNS archaikus mivoltát, és azért végeztünk csupán részleges UDG kezelést, hogy ezekből az aDNS-re jellemző „ujjlenyomatokból” maradjon néhány a molekulákon. Ezt követően leolvastuk a szekvenciát, melynek során a Freebayes (Garrison és Marth, 2012) program segítségével azonosítottuk a referencia genomtól eltérő SNP-eket, majd ezekből ún. variant call file-t (VCF) állítottunk elő, amely az összes eltérést listázza. Az így kigyűjtött variánsokat manuálisan is ellenőriztük az Integrative Genomics Viewer (IGV) program használatával (Robinson és mtsai., 2011; Thorvaldsdóttir és mtsai., 2013). Végül az ellenőrzésen átesett VCF fájlokból FASTA szekvencia fájlokat állítottunk elő a Genom Analysis Tool Kit (GATK v3.5) FastaAlternateReferenceMaker programmal (McKenna és mtsai., 2010).

### **3.6. Adatelemzés adatbázisok felhasználásával**

#### **3.6.1. Populációgenetikai vizsgálat**

Az NGS kísérleteket megelőző első adatsorunk egy alacsony felbontású haplotípus gyűjtemény volt, amely a Karosi III. számú temető populációjának HVR szekvenciáit és SNaPshot pozícióit tartalmazta. Az itt bemutatott kiértékelés erre az adatsorra vonatkozik. Az adatelemzés során azt feltételeztük, hogy a temető egyetlen populációt képvisel, ezért a karoshihoz genetikailag közel álló élő és kihalt populációk megkeresése céljából a temető haplocsoportjainak eloszlásához legjobban hasonlító populációkat kerestünk a fellelhető

adatbázisokból. Ehhez először összeállítottunk egy saját adatbázist, amely 35 Eurázsiaiából eddig azonosított archaikus (kihalt) populációt tartalmaz (mintaszám=1072). Ebben az adatbázisban a hasonló haplocsoport eloszlású rokon csoportokat összevontunk, például az európai kora és közép Neolit populációkat, valamint a késő Neolit és korai bronzkor populációit. A Kárpát-medencéből származó hasonló korú mintákat azonban nem vontuk össze. Az adatelemzésbe két, részben átfedő, korábban publikált honfoglaló populációt is belevettük (Bogácsi-Szabó és mtsai., 2008; Tömöry és mtsai., 2007). A ma élő népcsoportok 111 populációját tartalmazó adatbázisát a budapesti BSZKI munkatársai bocsátották rendelkezésünkre (mintaszám=20748; Függelék 7. táblázat). Ezt követően számítógéppel meghatároztuk az összes populáció haplocsoport eloszlását, melynek során 26 fő haplocsoportot vettünk figyelembe. A számítógépes algoritmusokat Juhász Zoltán fizikus, a KFKI munkatársa dolgozta ki, és az elemzést is ő végezte.

Az archaikus és recens populációk haplocsoport eloszlásának összehasonlítására a Self Organising Cloud (SOC) algoritmust alkalmaztuk. Ez egy olyan klasztereket (csoportokat) alkotó algoritmus, amely a 26 dimenziós (26 féle haplocsoportot tartalmazó) haplocsoport eloszlások hasonlósága alapján automatikusan csoportosítja a populációkat. Ezt az algoritmust korábban már sikerrel alkalmazták a genetika és a népzene tudomány területén (Juhász és mtsai., 2015; Pamjav és mtsai., 2013, 2012).

Emellett a rangsor analízis (iterative rank correlation = IRC) algoritmust is alkalmaztuk, amely a populációk egyes haplocsoportjainak gyakoriság értékei között keres összefüggést. Ennek az az értelme, hogy amennyiben több haplocsoport együttesen terjed egy közös földrajzi területről, akkor az várható, hogy a gyakoriság értékeik közötti összefüggés megmarad az utódpopulációkban is (Juhász és mtsai., 2016). Ezzel a módszerrel nem egyetlen haplocsoport eredetét és vándorlási mintázatát vizsgáljuk, hanem együttesen terjedő haplocsoport párokat, hármasokat, stb. azonosíthatunk. Az archaikus és recens populációk eloszlásait összevetve ezzel a módszerrel kideríthető, hogy egy adott haplocsoport eloszlás ősi, vagy közelmúltbeli keveredést tükröz-e, ugyanis amennyiben ugyanazt az eloszlást találjuk a mai populációban mint az archaikus mintában, akkor a keveredés valószínűleg az archaikus minta korát megelőzően történt.

### 3.6.2. Filogenetikai vizsgálat

Ezen vizsgálatok célja az, hogy a teljes adatbázisból az egyes karosi egyének haplotípusaihoz legközelebbi haplotípusokat összegyűjtve azok lehetséges leszármazási kapcsolatait ábrázolja egy törzsfán. Ehhez a Median Joining (M-J) Network algoritmust használtuk (Network 5.0.0.0 program, (Bandelt és mtsai., 1999), az elemzést a BSZKI munkatársa, Fehér Tibor végezte). A fá építése az egyes haplotípusok HVR-I régióiban található SNP variációk alapján történt.

Az NGS-sel meghatározott teljes mtDNS genomok filogenetikai vizsgálata a fentebb leírtól teljesen független adatelemzéssel történt. Ehhez egy újabb adatbázist kellett létrehoznunk, amely az NCBI adatbázisában fellelhető összes recens *Homo sapiens* mitokondrium genomot tartalmazta (mintaszám=32683), amit kiegészítettünk további publikálatlan magyar mintákkal (mintaszám=194) és az European Nucleotide Archives-ben valamint a szakirodalomban található összes ismert archaikus mtDNS genommal (mintaszám=458) is. Ez az adatbázis tehát az egész világról eddig ismert összes mtDNS genomot tartalmazza, és nagyon informatív, mivel a minták többségét úgy gyűjtötték, hogy abban az összes földrajzi régió és népcsoport képviselve legyen. Az egyes al-haplocsoportok jellegzetes földrajzi eloszlást mutatnak, és ezek alapján készült a híres „out of Africa” genetikai térkép is.

Az így összeállított adatbázisban szereplő összes szekvencia haplotípusát meghatároztuk a HaploFind programmal (Vianello és mtsai., 2013). Ezután ebből kiválasztottuk az összehasonlítható mintával megegyező és ahhoz közeli haplotípusú mintákat, majd az így kapott alcsoportokkal dolgoztunk tovább. Az egyes haplotípus csoportok szekvenciáit a MAFFT v 7 (Katoh és mtsai., 2002; Katoh és Standley, 2013) G-INS-1 progresszív módszerrel illesztettük. Ezen illesztéseket a MEGA v. 7.0.14 (Kumar és mtsai., 2016) programmal konvertáltuk át ún. NEXUS fájlformátumra, amely az illesztett szekvenciákból kigyűjti, és rendezi az eltéréseket. A NEXUS fájlformátumból a PopART programmal (Leigh és Bryant, 2015) készítettünk Median-Joining Network-öt (Bandelt és mtsai., 1999), amely nem más mint egy filogenetikai leszármazási fa, mely szemléletesen ábrázolja a hasonló szekvenciák egymástól való filogenetikai távolságát és legvalószínűbb leszármazási viszonyait. Végül az adatbázisból és a publikációkból egyenként visszakerestük a vizsgált mintánkkal megegyező, és ahhoz legközelebbi szekvenciák földrajzi származási helyét.

## 4. Eredmények

### 4.1. PCR alapú eredmények

#### 4.1.1. mtDNS haplotípusok

Első kísérleteinkben a Karos-III. temető 19 sírjának 17 leletéből tudtunk sikeresen mtDNS-t kivonni és meghatározni az egyének részleges anyai haplotípusát, a teljes HVR-I és a részleges HVR-II szakasz szekvenálásával, illetve a kódoló régió 22 SNP pontjának a meghatározásával. A 2. és 7. számú minták adatait kivettük az eredmények közül, mert a DNS rossz megőrződése miatt bizonytalan a besorolásuk. A 17 minta 7 fő haplocsoportba és 13 haplotípusba sorolható (1. táblázat).

**1. táblázat** A Karos-III temető 17 leletéből azonosított haplotípusok összefoglaló táblázata. Vizsgált szakaszok: HVR-I: nt. 16050-16400, HVR-II: nt. 190-309, és a kódoló régióban 22 SNP, a GenoCore22 szerint. A mutációk által definiált haplotípusokat a Haplogrep programmal határoztuk meg, amely a talált SNP-k alapján régiók alapján megad egy %-os értéket, ami azt jelzi, hogy a minta a meghatározott SNP-k alapján mekkora valószínűséggel tartozik az adott haplotípusba.

Minta (sírszám)	HVR-I mutációk (pozíció -16000)	HVR-II és kódoló régiós mutációk	Haplotípus	Haplogrep (%)
1.	183C 189C 217C	263G 7028T 9bp del 11719A 14766T	<b>B4</b>	100%
3.	362C	239C 263G	<b>H6</b>	100%
4.	069T 092C 126C 261T	228A 263G 295T 7028T 11719A 12612G 14766T	<b>J1c7</b>	100%
5.	183C 189C 217C	263G 7028T 9bp del 11719A 14766T	<b>B4</b>	100%
6.	189C	263G 7028T 9bp del 11719A 14766T	<b>B4'5</b>	100%
8.	051G 189C 362C	263G 7028T 11467G 11719A 14766T	<b>U2e</b>	95,89%
9.	051G	263G 7028T 11467G 11719A 14766T	<b>U2</b>	100%
10.	304C	263G	<b>H5</b>	100%
11.	189C 223T 278T	195C 257G 263G 6371T 7028T 11719A 12705T 14766T	<b>X2f</b>	100%
12.	183C 189C 223T 290T 319A	235G 263G 4248C 7028T 11719A 12705T 14766T	<b>A</b>	100%
13.	189C	263G 7028T 9bp del 11719A 14766T	<b>B4'5</b>	100%
14.	126C 163G 186T 189C 294T	195G 263G 7028T 11719A 13368A 14766T	<b>T1a</b>	100%
15.	069T 126C 362C	263G 295T 7028T 11719A 12612G 14766T	<b>J</b>	98,15%
16.	256T 270T	263G 7028T 11467G 11719A 14766T	<b>U5a</b>	100%
17.	362C	239C 263G	<b>H6</b>	100%
18.	126C 163G 186T 189C 294T	214G 263G 7028T 11719A 13368A 14766T	<b>T1a10a</b>	100%
19.	126C 163G 186T 189C 294T	214G 263G 7028T 11719A 13368A 14766T	<b>T1a10a</b>	100%
kutató	rCRS	n. d.	<b>H2a2a1</b>	100%

A temetőben a leggyakoribb haplocsoport a B mely az A-val együtt a karosi populáció 30%-át adja. Ezek a haplocsoportok Dél- és Kelet-Ázsiából terjedtek szét, és ma is ott a

leggyakoribbak. A haplocsoportok többsége (H, U, T, J, X) általános eurázsiai elterjedtségű, de két egyén a H6 alhaplocsoporthoz (3. és 17. minta) tartozik, ami szintén ázsiai kapcsolatot jelez. A H6 alhaplocsoportot a Közel-Keletről és a Kaukázus vidékéről származtatják, de ma Közép és Belső-Ázsiában a leggyakoribb (21%), különösképpen az Altaj vidékén (35%). A H6 Európában csak a bronzkor során jelent meg (Roostalu és mtsai. 2007), és csak kis gyakorisággal fordul elő.

A vezér (11. minta) X2f anyai haplotípusa külön említésre méltó, mivel ez a csoport valószínűleg kaukázusi eredetű, még ma is szinte csak itt fordul elő, ritka Kelet-Európában, Közép-Ázsiában és gyakorlatilag hiányzik a finnugor és a türk nyelvcsaládhoz tartozó Volga-Urál régió népeiből. A 10. minta a H5 haplocsoportba sorolható, amely hasonló földrajzi eloszlást mutat mint az X2f, leggyakoribb a Kaukázus régióban, és hiányzik a Volga-Uráli, finnugor és közép ázsiai populációkból (Roostalu és mtsai., 2007). Európában a szlovák és francia népek körében a leggyakoribb, de találtak ilyen mintát a neolitikus Kárpát-medencében, azaz nem szükségszerűen a honfoglalókkal érkezett.

Az U2 haplotípusnak egész Eurázsiaiban alacsony a gyakorisága (0,5-2%) (Maciamo, 2016), de a Karos III. számú temetőben két nem rokon személyt (8. és 9. minta) is találtuk, akik ebbe a csoportba tartoznak. Az U2 ma a legnagyobb gyakoriságot a Volga-Urál és az Észak-Kaukázus régiókban élő kis népcsoportok körében mutatja (Maciamo, 2016).

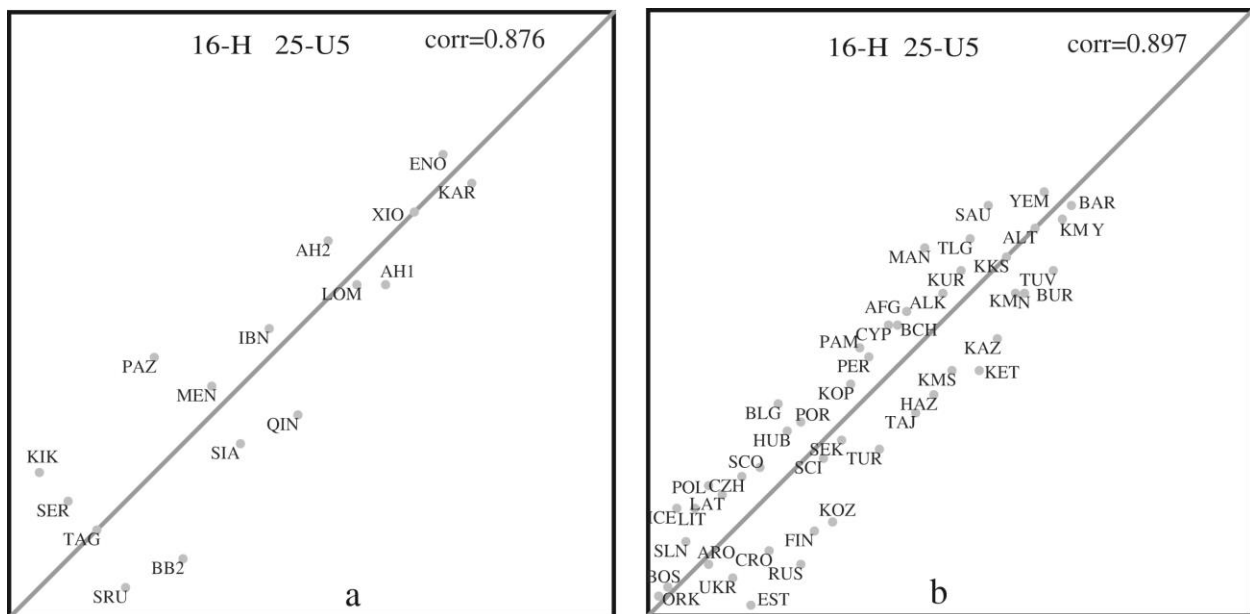
Az U2 és az U5 (16. minta) a legrégebbi európai haplotípusok közé tartozik, melyek a mezolitikus kor óta jelen vannak Eurázsiaiban, ezért alacsony a filogeográfiai információ tartalmuk. Ugyanez mondható el a J és T haplocsoportokról, melyek a korai Neolitikum óta jelen vannak Európában. A T1a haplocsoportba sorolható három személy (14., 18. és 19. minta) azt jelzi, hogy ez a haplocsoport gyakori lehetett a karosi magyarok között és a székelyek körében ma is nagyobb gyakoriságot mutat mint a körülöttük élő populációkban (2,25%, (Brandstätter és mtsai., 2007).

#### **4.1.2. A populációgenetikai analízis eredményei**

Ezen vizsgálat során azt feltételeztük, hogy a karosi népesség egyetlen populáció lehetett, és a hozzájuk leghasonlóbb összetételű népességeket kerestünk az ismert ősi és ma élő populációk között többféle számítógépes algoritmus alkalmazásával.

A 6. ábra az egyik használt algoritmus, az IRC (rangsor analízis) működését szemlélteti, amely haplocsoport gyakoriság összefüggéseket keres a populációk között. Az 6./a ábra 35 ősi

populációt tartalmazó archaikus adatbázisunkból 16-ot ábrázol, melyekben a H és U5 haplocsoport gyakorisága összefügg. Az x tengely mentén a H, míg az y tengely mentén az U5 haplocsoport gyakorisága csökken. A minták átló menti tömörülése jelzi, hogy a H és U5 gyakoriság értékek együtt változnak, azaz a populációkon belül együttesen csökken vagy nő az arányuk. A 6./b ábra azt mutatja, hogy a recens adatbázis 111 populációja közül 58-ban ugyanez a korreláció figyelhető meg. Mivel a H és U5 haplocsoportok gyakorisága mind az archaikus mind a recens populációkban korrelál ebből az következik, hogy valószínűleg már régóta együtt mozogtak, terjedtek, és a mai populációk jelentős része megőrizte az ősi populációkban létrejött haplocsoport eloszlási mintázatot. Az IRC analízissel az összes haplocsoport gyakorisági értékeit páronként összehasonlítva azt találtuk, hogy több haplocsoport között hasonló gyakorisági összefüggés mutatható ki, mely mintázat együttes terjedéssel már az ősi populációkban kialakult és máig megőrződött.

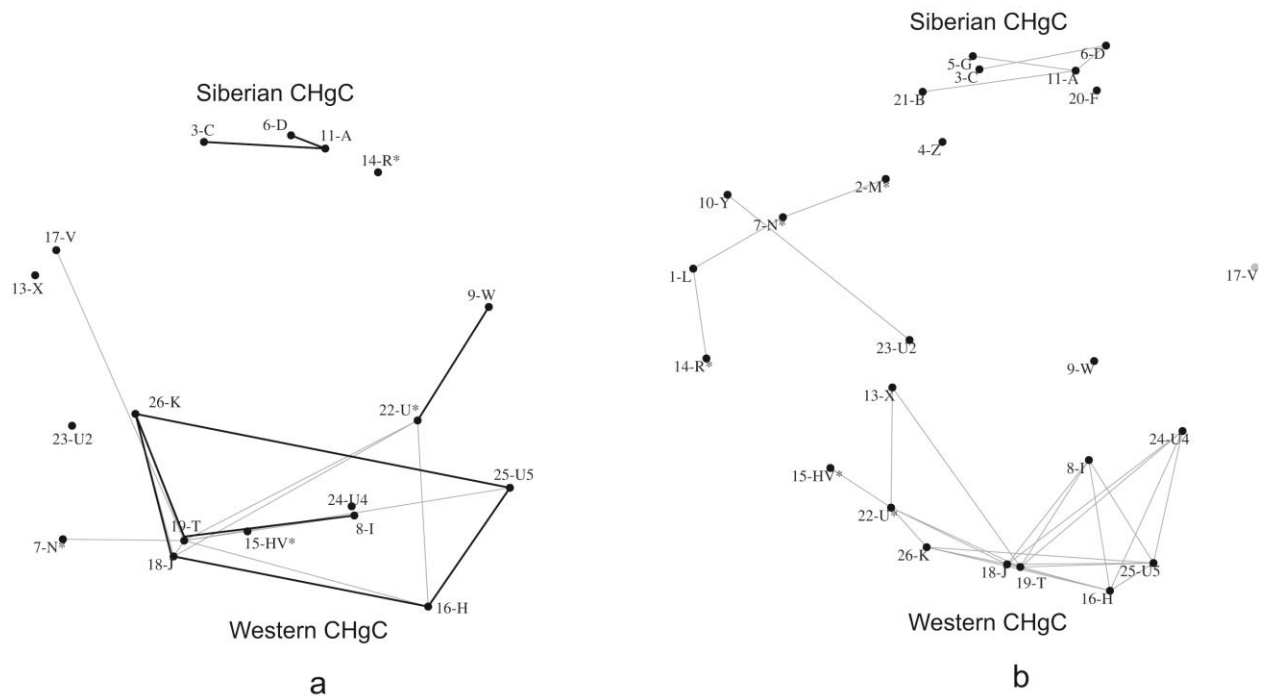


**6. ábra** A H és U5 haplocsoportok gyakorisági összefüggése archaikus (a) és recens (b) populációkban. Az x tengely mentén a H, míg az y tengely mentén az U5 haplocsoport gyakorisága csökken. Jól látható, hogy gyakorisági értékük együtt változik mind archaikus, mind recens populációk esetében. Az ábrán a pontok az egyes populációkat szimbolizálják, a 3 betűs rövidítések pedig a populációk nevének rövidítései (Függelék 7. táblázat).

A rangsor analízissel azonosított, egymással összefüggő gyakoriság értékeket mutató haplocsoportok kapcsolatát a 7. ábra összesíti. mutatja. Az egyes csoportok gyakoriság értékei közötti összefüggés mértékét MDS (Multiple Dimensional Scaling) algoritmussal ábrázoltuk, a legerősebben korreláló gyakoriságú haplocsoportokat vonalak kötik össze. A 7./a ábra az archaikus, a 7./b ábra pedig a recens populációkban talált haplocsoport gyakoriság

összefüggéseket mutatja. Megfigyelhető, hogy az archaikus mintáknál az U, H, HV, T, J és K haplocsoportok gyakoriság értéke szorosan összefügg, és ez az összefüggés a modern populációkban is megtalálható.

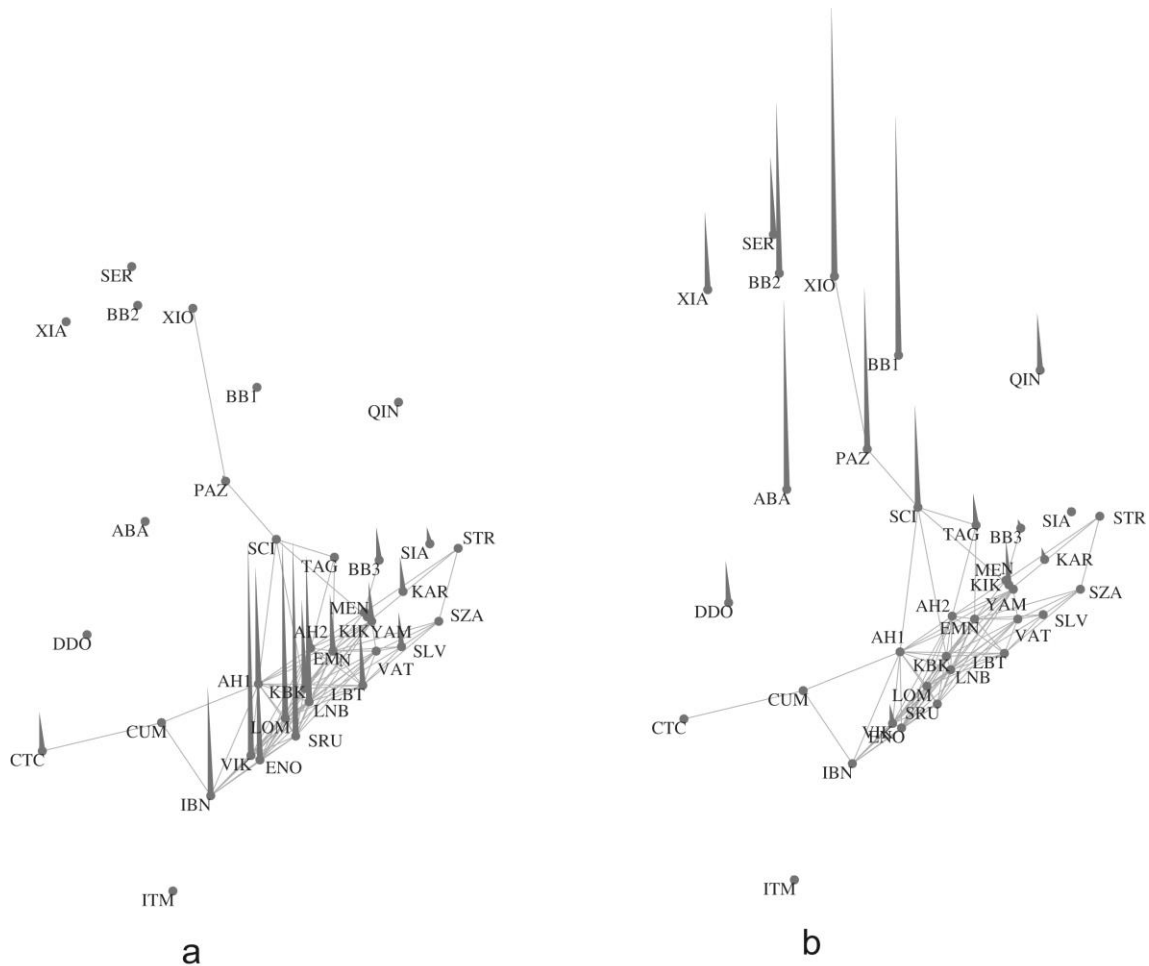
Ugyanígy összefüggést mutatnak egymással az ábrák felső részén egymással összekötött C, D és A haplocsoportok gyakorisági értékei, melyek ma tipikusan Ázsiára jellemzők. Az 7./a és 7./b ábra hasonlóságából az következik, hogy a mai populációk összetételét meghatározó vándorlási és keveredési folyamatok zömmel a vizsgált ősi populációk korát megelőzően, évezredekkel ezelőtt történhettek, amit a későbbi keveredések sem változtattak meg jelentősen. Mai elterjedtségük alapján a C, D, A, G, B és F haplocsoport klasztert szibériai klaszternek (Correlating Haplogroup Cluster= CHgC) neveztük el, míg az alsó U, H, HV, T, J és K együttest nyugati klaszternek (Függelék 1. ábra).



**7. ábra** A rangsor analízissel páronként mutatkozó haplocsoport gyakorisági összefüggések összesítő ábrája MDS algoritmussal ábrázolva. (a) archaikus és (b) recens populációkban mutatkozó gyakorisági összefüggések. A legmagasabb gyakoriság összefüggést mutató haplocsoportokat vonalak kötik össze. A vonallal összekötött haplocsoportok közös forrásból együtt terjedhettek a populációk között. A program két jól elkülönülő haplocsoport klasztert azonosított, amit szibériai és nyugati (európai) klaszternek neveztünk el.

Ezt követően az MDS algoritmussal kiszámítottuk és ábrázoltuk az ősi populációk haplocsoport eloszlása közötti távolságot (8. ábra). Minél hasonlóbb két populáció haplocsoport eloszlása annál közelebb található egymáshoz az ábrán, a legszorosabb kapcsolatokat összekötő vonalak jelzik. Az egyes populációkra rávetítettük a bennük fellelhető nyugati (8./a) és a

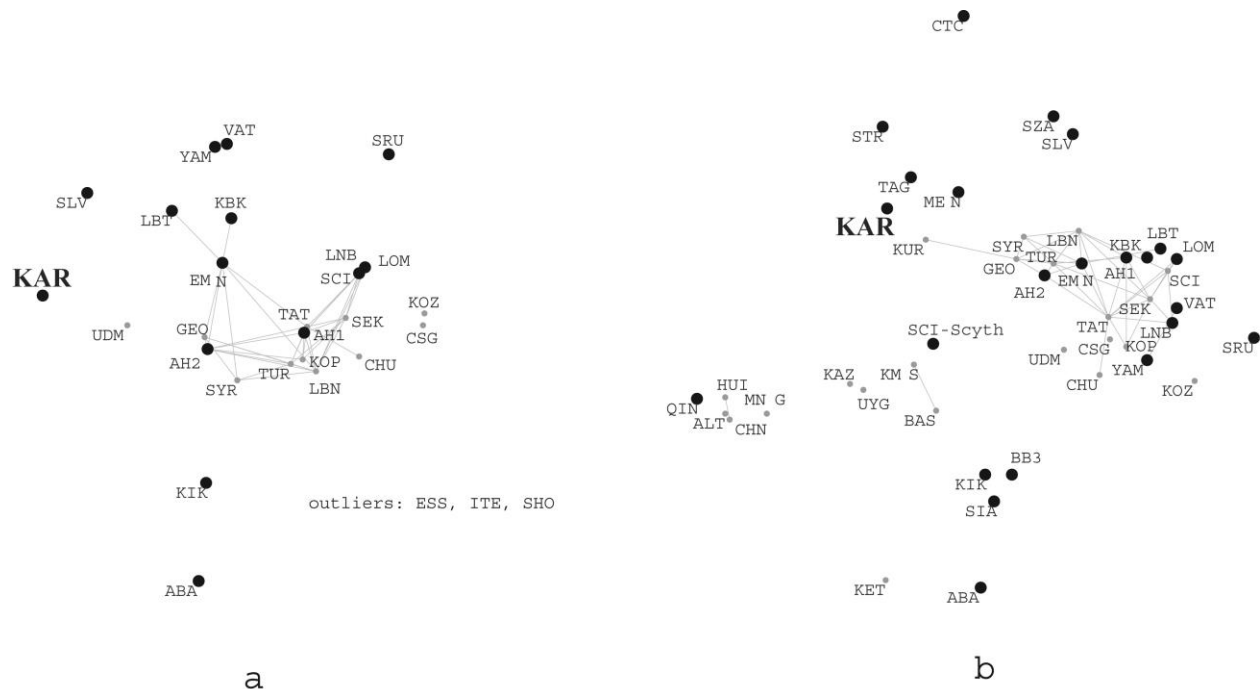
szibériai (8./b) klaszter gyakoriságokat is, melynek értéke a pontokból kinyúló csúcsokkal arányos. Az ábra alsó részén csoportosulnak az európai populációk, míg felső részén az ázsiai eredetűek. Látható, hogy a karosi populáció (KAR) a szibériai és európai eredetű génáramlás metszetén helyezkedik el, mivel kiegyenlítetten tartalmaz ázsiai és európai eredetű géneket. A karosi populációhoz legközelebb álló csoportok a 6000-3000 éve élt európai korai földművelők (STR=Starcevo), a Közel-keleti neolit (MEN), a Kárpát medencei Szakálhát kultúra (SZA), a Yamnyák (YAM, i. e. 3500-2300 Fekete tenger fölötti sztyepp), a Kazahsztán területén élt vaskori szkíta kurgánok (KIK, i.e. 800-600) népessége, és az ázsiai bronzkori szkíták a Tagar-Tachtyk kultúrából (TAG, i. e. 800 –i. u. 400). Szintén közelinek mutatkozik a közép-ázsiai bronzkori Sintastha kultúra (SIA, az első sztyeppeni nomádok) és a vaskori Baraba sztyeppé (BB3, dél Szibéria) népessége.



**8. ábra** Az ismert archaikus populációk haplocsoport eloszlási hasonlósága MDS algoritmussal ábrázolva. A hasonlóság arányos a pontok közötti távolsággal, a leghasonlóbb populációkat vonalak kötik össze. Emellett az (a) ábrán az egyes populációkra rajzolt oszlopok magassága jelzi a bennük található európai haplocsoportok gyakoriságát, míg a (b) ábrán az oszlopok ugyanígy az ázsiai haplocsoportok gyakoriságát jelzik ugyanazon a térképen.



A SOC algoritmus a haplocsoport eloszlások hasonlósága alapján 7 fő klaszterbe csoportosította az adatbázisunkban szereplő ősi és modern populációkat (Függelék 1. ábra). Ezek közül a karosi populáció a 2. klaszterbe tartozik amit a nyugati haplocsoportok dominanciája jellemez, de a szibériai csoportok jelenléte is számottevő. A 9./a ábrán a 2. klaszter MDS térképe látható, a karosiakhoz (KAR) legközelebbi populációkkal. A 9./a ábra annyiban különbözik a 8. ábrától, hogy csak a 2. klaszter populációit ábrázolja, és mind modern mind archaikus populációkat tartalmaz. Feltűnő, hogy a karosi populáció az aránylag magas szibériai komponensének köszönhetően kilóg a 2. klaszterből és 3. klaszterhez közelít. A 3. klaszterben a nyugati és szibériai haplocsoportok aránya kiegyenlítettebb. A karosi populáció kapcsolatainak pontos szemléltetésére a 2. és 3. klaszter populációit összevontan is ábráztuk (9./b ábra). Az MDS értékek alapján a karosi populáció legközelebbi rokonai a 2. klaszterből a grúz (GEO) és udmurt (UDM) populációk, míg a 3. klaszterből a kurd (KUR) populáció.



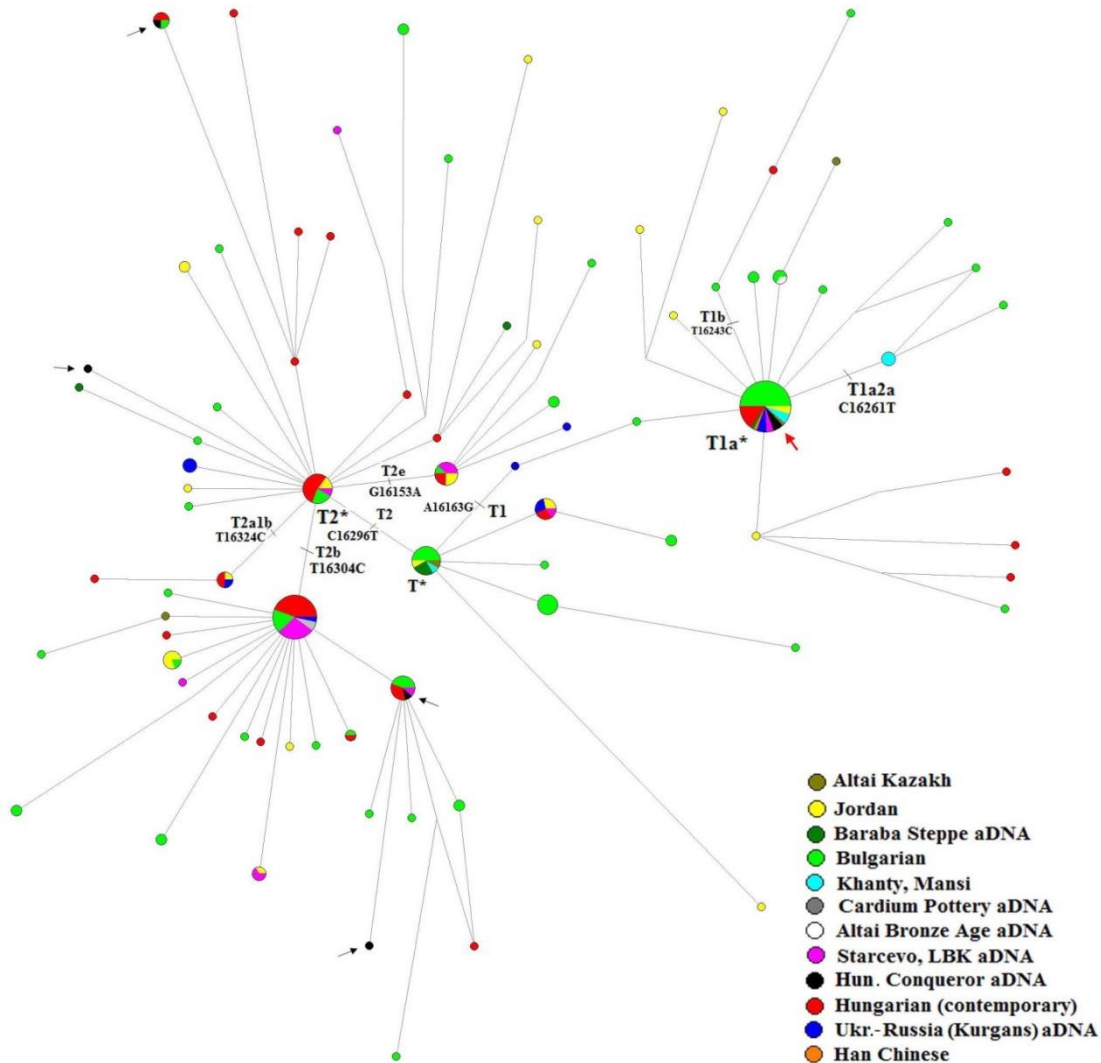
**9. ábra** A karosihoz leghasonlóbb haplocsoport összetételű populációk MDS térképe. A haplocsoport eloszlási hasonlóság arányos a pontok közötti távolsággal, a leghasonlóbb populációkat vonalak kötik össze. A nagy fekete pontok az archaikus populációkat, míg a kis szürkék a recenseket jelölik. 8./a: a 2-es klaszter populációinak MDS térképe, 8./b: a 2-es és 3-as klaszter együttes MDS térképe. Az egyes klasztereket a Függelék 3. ábra mutatja.

### 4.1.3. Filogenetikai eredmények

Alább csak a T és B haplocsoport filogenetikai hálózatát mutatom be, mivel csak ezek adtak a fentebb tárgyaltakhoz képest új információt a karosi egyének lehetséges genetikai kapcsolatairól.

A 10. ábra a T haplocsoport M-J hálózatát mutatja be, ami az adatbázisunkból kiválogatott T haplocsoportba sorolt minták HVR szekvencia sorrendje alapján készült. Az alapító T\* haplotípusba 12 minta esik. A T haplocsoportú karosi honfoglalók mind a T1a\* alapító haplotípusba tartoznak (piros nyíl), melybe további 43 archaikus és recens minta tartozik. (Archaikusok: 3 yamnaya kurgán, 1 Baraba sztyepp, 1 Starcevo neolit kori, 1 dunántúli vonaldíszes kerámia. Recens minták: 23 bolgár, 7 székely és csángó, 2 jordán, 2 mansi, 1 khanty, 1 altaji kazah, 1 han kínai) Ez az adatsor megerősíti a genetikai rokonságot a honfoglalók és a mai székelyek között, illetve a karosi populáció filogeográfiai kapcsolatát az ősi és mai dél-szibériai népekkel.

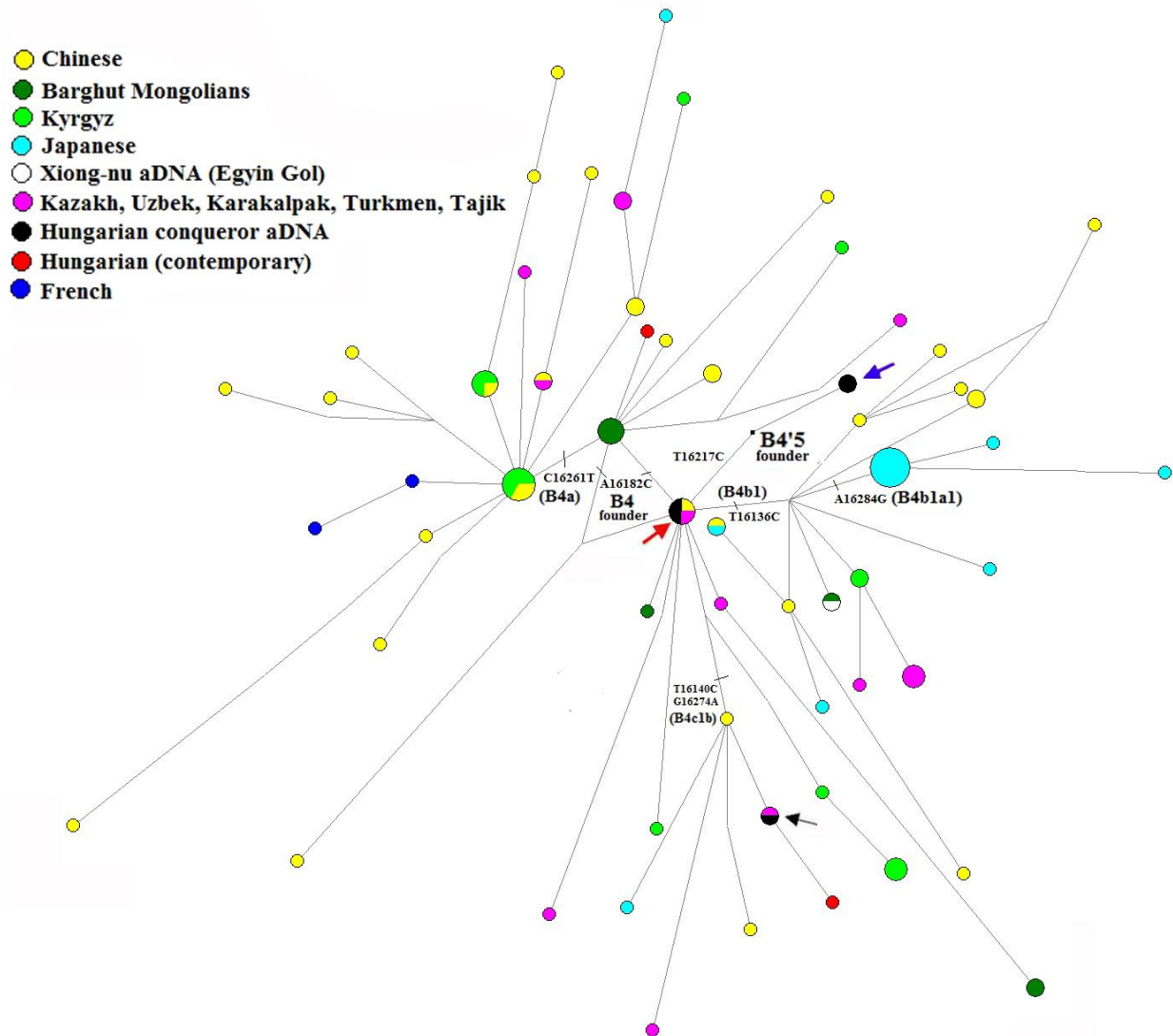
A T2b alapító haplotípusba 29 minta sorolható (13 recens magyar, 5 bolgár, 4 Neolit kori minta Starcevo-ból, 4 dunántúli vonaldíszes kerámia kultúrájú, 2 neolit kori kagylódíszes kultúrájú minta és 1 részkori kurgán minta). Tömöry (2007) négy T2b haplotípusba tartozó honfoglaló magyar mintát közölt (T2b, T2\*: T2b, T2e, T2a1b), ezeket fekete nyíl jelzi a 16. ábrán. A T2b csoport hiányzik Ázsiából, de megtalálható az ősi Duna menti és mediterrán neolit mintákban. Ez alapján a T1a haplotípusú karosi honfoglalók anyai ágon Dél-Szibériából származhatnak, míg a T2 csoportba sorolható minták inkább a neolit kori európai régióból.



**10. ábra** T haplocsoport Median Joining hálózata. A körök mérete arányos a mintaszámmal, a legkisebbek egy mintának felel meg. A különböző színek a listán felsorolt populációkból származó mintákat jelölik. A piros nyíl a karosi egyéneket mutatja, míg a fekete nyíl a Tömörö és munkatársa által közölt honfoglalókat jelzi. A karosi mintákkal azonos körbe eső, velük azonos HVR szekvenciájú egyének származási helyét a szövegben soroltuk fel.

A 11. ábra a B haplocsoport Median Joining hálózatát szemlélteti, ez adatbázisunkból kiválogatott B haplocsoportba sorolt minták HVR szekvencia sorrendje alapján készült. Az alapító B4 haplotípust 4 minta képviseli: 2 karosi honfoglaló (piros nyíl), 1 han kínai Liaoningból és 1 üzbég. A két karosi B4'5 egyén (kék nyíl) 2 mutáció távolságra található az alapító haplotípustól, ezekhez ma élő egyének nincsenek közel. Egy korábban publikált B haplotípusú honfoglaló (fekete nyíl) összeillik egy altaji kazah mintával, ami a B4c1b alhaplocsoportba tartozik. Ehhez a honfoglalóhoz 1 mutációnyi távolságra található egy recens magyar minta (piros körrel jelölve). Mivel Liaoning a tunguz nyelvű manchuk hazája, az üzbég

és kazah megfeleléseket is figyelembe véve az a legvalószínűbb, hogy a karosi B haplocsoport altáji, török eredetű.



11. ábra B haplocsoport Median Joining hálózata. A körök mérete arányos a haplotípus gyakoriságával, a legkisebb egy mintának felel meg. A különböző színek a listázott populációkat jelölik. A piros és kék nyilak mutatják a karosi populációt.

#### 4.1.4. Y-kromoszómális haplocsoportok

Az apai ágú leszármazás felderítése céljából végzett Y-kromoszómás SNP vizsgálataink (GenoY25) valamint Y-STR vizsgálataink néhány mintánál eredményre vezettek (Függelék 5. és 6. táblázat). Az Y kromoszóma darabok amplifikálása csak a legjobb megőrződésű aDNS mintákból lehetséges, mivel sejtmagi lókusztok alacsony kópiaszámban vannak jelen. Az anyagok és módszerekben leírtak szerint megpróbáltuk optimalizálni a reakciót, ennek ellenére

néhány multiplex primerpár alig működött (Függelék 5. táblázat); M17 terméke teljesen hiányzott minden mintából, az M35 terméke pedig nagyon gyenge jelet adott még recens mintából is. Az M172 termék esetében pedig erős háttér jelet kaptunk minden minta esetében, ami háttér szennyezésre utal, de ezt az M304 fő haplocsoport SNP allélja ezt minden esetben kizárja, ugyanis amennyiben az M172 alcsoport a referenciától eltérő SNP-t mutat, akkor értelemszerűen a főcsoport M304-nek is eltérést kellene mutatnia, de nem ez a helyzet. Ebből arra következtettünk, hogy az M172-es háttér jel valószínűleg nem humán eredetű háttér szennyeződésből ered, és valószínűleg ezzel magyarázható az alkalmanként megjelenő M216-os háttér is.

A hiányzó vagy bizonytalan SNP-eket néhány esetben single-plex PCR segítségével tudtuk pótolni, vagy megerősíteni (sárgával kiemelve a Függelék 5. táblázatban), amelyek alátámasztották a GenoY25 multiplex eredményeket. Végül a kiértékelés során az M172 pozíciót nem vettük figyelembe, így 4 mintának tudtuk meghatározni az Y haplocsoportját.

Az adatok alapján a Karos-III/1 és Karos-III/3 egyének az R1b1b1a haplocsoportba tartoznak, míg a Karos-III/12 és Karos-III/17 mintákat az I haplocsoportba tudtuk besorolni. A Karos-III/12 és Karos-III/17 minta esetében az I haplocsoportba tartozást a független Y-kromoszómális STR adatok is alátámasztották (Függelék 6. táblázat), amelyek még mélyebb besorolást is lehetővé tettek, az I2a alhaplocsoportba. Megjegyzendő, hogy az STR alapú Y haplocsoport besorolásnak van némi bizonytalansága, mivel az STR adatok és az SNP adatok között statisztikai alapon állították fel az összefüggést (Wang és mtsai., 2015).

Az STR vizsgálat PCR eredményei hiányosak, a nagyobb fragment méret tartományban nem adtak jelet (Függelék 6. táblázat), ezért csak részleges STR profilokat tudtuk készíteni, melyekből egyes lókuszek hiányoztak. Azonban aDNS esetében ez nem meglepő, mivel abban kevés nagyméretű fragment marad a degradáció miatt, így ez a hiány is megerősíti a DNS forrás szennyezés mentességét. Az STR vizsgálatokhoz a legjobb minőségű kivonatokat használtuk, abból is nagy mennyiségre volt szükség, így ezeket a kísérleteket sajnos nem tudtuk megismételni.

A Karos-III. számú temetőben nyugvó férfiak közel harmadának sikerült az apai ágú haplocsoportját meghatározni, és meglepő módon ezek mindegyike tipikusan európai elterjedettséget mutat (I2a és R1b). Az I-M170 mutáció által meghatározott haplocsoport egyike a legősibb Európai csoportoknak (Rootsi és mtsai., 2004), a Kárpát-medencében pedig már a

Neolitik korban kimutatták a jelenlétét (Gamba és mtsai., 2014). Ez a haplocsoport teljesen hiányzik Ázsiából. Az I2a csoportba tartozik a mai magyar férfiak 16,74%-a.

Az R1b-M269 mutáció által definiált haplocsoport ugyan eredetileg Ázsiából származik de ma az egyik leggyakoribb csoport Nyugat Európában, míg keletkezési helyén manapság csak nyomokban található (Myres és mtsai., 2011). R1b csoportba sorolható a mai magyar férfiak 18,1%-a. Mintáink ennek alcsoportjába, az R1b-S21 (U106) csoportba tartoznak, amely ma a német anyanyelvűekre jellemző alcsoport, és teljesen hiányzik Ázsiából. Ezen adatok alapján a Karos-III számú temető férfi tagjainak jelentős hányada apai ágon európai eredetű lehetett.

#### **4.1.5. Autoszómális STR eredmények**

A temetőn belüli lehetséges rokoni kapcsolatokat autoszómális STR vizsgálatokkal próbáltuk megfejteni, melyet az igazságügyi orvostanban alkalmaznak. Az mtDNS adataink alapján azonosított egyes vonalak között anyai ágú rokonság gyanítható, de ez nem elegendő bizonyíték. Így például azonos csoportba tartoznak a Karos-III/1 és Karos-III/5 (B4), a Karos-III/6 és Karos-III/13 (B4'5), a Karos-III/3 és Karos-III/17 (H6) és Karos-III/18 és Karos-III/19 (T1a10), akik rokonok lehetnek.

Ezen lehetőségek közül a hiányos STR eredmények (Függelék 6. táblázat) alapján is kizárható a rokonság a Karos-III/3 és Karos-III/17 között 3 vizsgált lókusznál (D21S11, D2S1338 és VWA), mivel ezekben a lókuszekben eltérő heterozigóta genotípust mutattak. Ezzel ellentétben a Karos-III/18 és Karos-III/19 mintánál meg tudtuk erősíteni a lehetséges rokoni kapcsolatot 5 STR lókusznál (D3S1358, D10S1248, D16S539, D2S441 és D12S391), melyeknél mindkét minta azonos allélokkal rendelkezett. A Karos-III/1 és Karos-III/5 valamint a Karos-III/6 és Karos-III/13 rokonságát a gyenge STR eredmények alapján ezzel a módszerrel nem tudtuk kizárni vagy megerősíteni.

Az azonos Y haplocsoportba tartozó egyének szintén rokonok lehetnek, de ez sem elegendő bizonyíték. A lehetséges férfiágú rokonságot sikerült kizárnunk két autoszómális (D8S1179 és D12S391) és egy Y kromoszómás (DYS390) STR adat alapján a Karos-III/12 és Karos-III/17 (I2a) egyének között. Ezzel szemben az STR adatok igazolták az apai ágú rokonságot a Karos-III/1 és Karos-III/3 egyének között, mégpedig 10 lókusznál (D3S1358, TH01, D21S11, D10S1248, D1S1656, D2S1338, D22S1045, VWA, D2S441 és D12S391). Ezt alátámasztják a GenoY25 eredmények is, mivel mindkét minta az R1b1b1a haplocsoportba

tartozik (Függelék 5. táblázat). Mivel mindkét egyén idős korban hunyt el (3. ábra), ők valószínűleg testvérek lehettek.

## 4.2. NGS alapú eredmények

### 4.2.1. mtDNS genom haplotípusok

A Karos-I temetőből 11, Karos-II temetőből 44, a Karos-III temetőből 18, a Sárrétudvari-Hízóföld temetőből 6, a Kenézlő-Fazekaszug-II temetőből 4, a Szegvár-Oromdülő temetőből 2 és a Magyarhomoróg, Orosháza-Görbicstanya, Szabadkígyós-Pálliget, Harta-Freifelt temetőkből 1-1 minta teljes mtDNS genom szekvenciáját határoztuk meg (2. táblázat). Ez a 89 minta tehát 3 temető csaknem teljes anyagát tartalmazza, ami nagyon reprezentatív mintaszámot jelent, a többi temetőből származó néhány minta eredményéből pedig arra tudunk következtetni, hogy azok mennyiben hasonlítanak, vagy térnek el a Karosiaktól. A nagy mintaszámú, nagy felbontású adatok alapján már sokkal nagyobb biztonsággal körvonalazható a honfoglalók vélhető származása.

Azonos mitokondrium genommal rendelkeznek, vagyis anyai ágú rokonság gyanítható a következő egyének között: Karos-I/1-Karos-I/2, a Karos-I/3-Karos-I/5, a Karos-I/10-Karos-I/1438, a Karos-II/9-Karos-II/60-Kenézlő-Fazekaszug-II/1027-Kenézlő-Fazekaszug-II/1045, a Karos-II/16-Karos-II/54, a Karos-II/21-Karos-II/22, a Karos-II/31-Szabadkígyós-Pálliget/7/anc4, a Karos-II/52-Karos-III/11 (vezérek), Karos-III/5-Karos-III/6, a Karos-III/7-Karos-III/8-Karos-III/9, a Karos-III/18-Karos-III/19. Meglepő módon a 3 szomszédos karosi temetőben a vezérek kivételével nem találtunk átfedő haplocsoportokat, vagyis a 3 temető népessége egymástól markánsan elkülönül. Az egyes karosi temetőkön belül vannak rokonok, de temetők között csak a vezérek lehettek testvérek (Karos-II/52-Karos-III/11). Talán ennél is meglepőbb, hogy a többi temetőkből származó kis mintaszám ellenére is több azonos mtDNS genommal rendelkező lehetséges rokont találtunk a Karosi minták és egyéb temetőkből származó leletek között. Ebből arra következtethetünk, hogy Karoson három hasonló összetételű (hun-germán lásd később), de eltérő azonosságtudatú törzs telepedett le, hisz egymással nem keveredtek. Ha ehhez hozzávesszük, hogy a távoli temetők között viszont közvetlen rokonokat találunk az a következtetés adódik, hogy a Kárpát-medence elfoglalását követően az egyes törzseket

megoszthatták, tagjaikat széttelepíthették, és az egyes területekre különböző törzsek töredékeit telepíthették egymás mellé.

**2. táblázat** A teljes mitokondrium genom alapján meghatározott haplotípusok összefoglaló táblázata. A temető melletti számok a sírszámot jelzik, ahol ez nem ismert ott a lelet leltári számát adtuk meg. Az leletek haplotípusainak vélhető származását a filogenetikai leszármazási fák alapján határoztuk meg. A Haplogrep program az SNP-k alapján meghatároz egy %-os értéket, ami azt jelenti, hogy mennyire biztos, hogy a minta a mutációs mintázata alapján az adott haplotípusba tartozik. Ahol ismert a lelet neve, ott feltüntettem (ffi=férfi, nő és gy=gyermek)

<b>Temető/sírszám vagy leltári szám/minta neve</b>	<b>Neme</b>	<b>Haplotípus</b>	<b>Haplogrep (%)</b>	<b>Haplotípus vélhető származása</b>
Karos-I/1	ffi	<i>H1b2</i>	100	skandináv-germán
Karos-I/2	ffi	<i>H1b2</i>	100	skandináv-germán
Karos-I/3	ffi	<i>D4j12</i>	95	ázsiai hun
Karos-I/5		<i>D4j12</i>	95	ázsiai hun
Karos-I/7	nő	<i>Y1a1</i>	98,06	ázsiai hun
Karos-I/10		<i>U3b1b</i>	80	szláv
Karos-I/12	nő	<i>U2e1b</i>	88,58	ázsiai hun
Karos-I/13	ffi	<i>U5b2b3</i>	93,9	egyéb európai
Karos-I/1438		<i>U3b1b</i>	96,08	szláv
Karos-I/3285		<i>K1c1</i>	95,06	skandináv-germán
Karos-I/3286		<i>T2b4h</i>	95,03	skandináv-germán
Karos-II/1	nő	<i>D4c2a</i>	89,24	ázsiai hun
Karos-II/6	ffi	<i>H5+709</i>	89,55	skandináv-germán
Karos-II/8	gy	<i>K1c1e</i>	98,34	ázsiai hun
Karos-II/9	ffi	<i>N1a1a1a1a</i>	91,94	ázsiai hun
Karos-II/10	ffi	<i>U5a1a1h</i>	100	skandináv-germán
Karos-II/11	ffi	<i>D2</i>	97,47	ázsiai hun
Karos-II/12	nő	<i>U3b</i>	94,72	Kaukázus (Közel-Kelet)
Karos-II/13	nő	<i>T1a1</i>	100	skandináv-germán
Karos-II/14	ffi	<i>F2a</i>	100	ázsiai hun
Karos-II/15	nő	<i>U4d2</i>	98,37	ázsiai hun
Karos-II/16	nő	<i>X2l</i>	97,58	Kaukázus (Közel-Kelet)
Karos-II/17	gy	<i>J1c2</i>	96,08	skandináv-germán
Karos-II/19	nő	<i>J1b1a1+146</i>	100	skandináv-germán
Karos-II/20	ffi	<i>C4</i>	100	ázsiai hun
Karos-II/21	gy	<i>H2a1+146</i>	100	skandináv-germán



Karos-II/22	gy	<b>H2a1+146</b>	98	skandináv-germán
Karos-II/23	nő	<b>T1a1</b>	95,2	skandináv-germán
Karos-II/25	nő	<b>F1a</b>	94,69	ázsiai hun
Karos-II/26	ffi	<b>T1a1</b>	100	skandináv-germán
Karos-II/27	nő	<b>H15b</b>	81,23	skandináv-germán
Karos-II/29	ffi	<b>J1b1a1+146</b>	96,28	skandináv-germán
Karos-II/30		<b>K1</b>	92,4	egyéb európai
Karos-II/31	gy	<b>N1a1a1a1</b>	92,89	ázsiai hun
Karos-II/32	nő	<b>T1a1</b>	96,3	skandináv-germán
Karos-II/36	ffi	<b>D4i</b>	97,27	ázsiai hun
Karos-II/37	nő	<b>G2a1d2</b>	95,69	ázsiai hun
Karos-II/41	ffi	<b>H35</b>	88,99	ázsiai hun
Karos-II/47	nő	<b>U4a1b1</b>	94,59	egyéb európai
Karos-II/48	gy	<b>K1c1</b>	92,52	skandináv-germán
Karos-II/50	ffi	<b>U5a2a2a</b>	97,59	skandináv-germán
Karos-II/51	ffi	<b>U4d2</b>	96,62	ázsiai hun
Karos-II/52	ffi	<b>X2f</b>	92,62	Kaukázus (Közel-Kelet)
Karos-II/53	nő	<b>H2a1c</b>	93,85	skandináv-germán
Karos-II/54	nő	<b>X2l</b>	97,58	Kaukázus (Közel-Kelet)
Karos-II/56	nő	<b>D4i2</b>	96,22	ázsiai hun
Karos-II/57	gy	<b>T2b</b>	95,99	skandináv-germán
Karos-II/58	nő	<b>U5a2b</b>	94,44	skandináv-germán
Karos-II/60	ffi	<b>N1a1a1a1a</b>	87,27	ázsiai hun
Karos-II/61	ffi	<b>U4d2</b>	96,62	ázsiai hun
Karos-II/68	nő	<b>U2e2a1a2</b>	95,33	skandináv-germán
Karos-II/70	nő	<b>X2c1</b>	99,19	skandináv-germán
Karos-II/71	nő	<b>H11a1</b>	93,31	ázsiai hun
Karos-II/72	nő	<b>H1b1</b>	87,64	skandináv-germán
Karos-II/73	gy	<b>D4b2a</b>	88,49	ázsiai hun
Karos-III/1	ffi	<b>B4d1</b>	87,09	ázsiai hun
Karos-III/3	ffi	<b>H6a1b</b>	92,02	skandináv-germán
Karos-III/4	ffi	<b>J1c7a</b>	95,95	skandináv-germán
Karos-III/5	nő	<b>B4d1</b>	91,07	ázsiai hun
Karos-III/6	nő	<b>B4d1</b>	91,07	ázsiai hun

Karos-III/7	gy	<b><i>U2e1b</i></b>	93,35	ázsiai hun
Karos-III/8	ffi	<b><i>U2e1b</i></b>	91,98	ázsiai hun
Karos-III/9	nő	<b><i>U2e1b</i></b>	92,72	ázsiai hun
Karos-III/10	ffi	<b><i>H5e1</i></b>	91,83	skandináv-germán
Karos-III/11	ffi	<b><i>X2f</i></b>	96,41	skandináv-germán
Karos-III/12	ffi	<b><i>A12</i></b>	93,67	ázsiai hun
Karos-III/13	ffi	<b><i>B4d1</i></b>	89,48	ázsiai hun
Karos-III/14	ffi	<b><i>T1a1b</i></b>	98,05	skandináv-germán
Karos-III/15	nő	<b><i>J2a1</i></b>	91,7	egyéb európai
Karos-III/16	gy	<b><i>U5a1a2a</i></b>	90,66	ázsiai hun
Karos-III/17	ffi	<b><i>H6a1a</i></b>	92,78	skandináv-germán
Karos-III/18	nő	<b><i>T1a10a</i></b>	94,78	skandináv-germán
Karos-III/19	gy	<b><i>T1a10a</i></b>	94,78	skandináv-germán
Magyarhomoróg/120/anc2		<b><i>H84</i></b>	85,43	Kaukázus (Közel-Kelet)
Orosháza-Görbiczstanya/2/anc3		<b><i>N1a1a1a1a</i></b>	89,89	ázsiai hun
Szabadvány-Pálliget/7/anc4		<b><i>N1a1a1a1</i></b>	91,1	ázsiai hun
Szegvár-Oromdűlő/412/anc5		<b><i>K1c1d</i></b>	97,18	skandináv-germán
Szegvár-Oromdűlő/593/anc6		<b><i>U5a2a1b</i></b>	97,78	egyéb európai
Sárrétudvari-Hízóföld/5/anc10	ffi	<b><i>I5a1a</i></b>	95,95	skandináv-germán
Sárrétudvari-Hízóföld/9/anc11	ffi	<b><i>T2b</i></b>	96,16	skandináv-germán
Sárrétudvari-Hízóföld/118/anc12	nő	<b><i>T2f1a1</i></b>	97,3	skandináv-germán
Sárrétudvari-Hízóföld/213/anc13	ffi	<b><i>J1c3g</i></b>	97,79	skandináv-germán
Harta-Freifelt/10/anc25		<b><i>H5e1a</i></b>	95,18	egyéb európai
Kenézlő-Fazekaszug-II/1027		<b><i>N1a1a1a1a</i></b>	86,53	ázsiai hun
Kenézlő-Fazekaszug-II/1044		<b><i>U5a1i</i></b>	90,4	skandináv-germán
Kenézlő-Fazekaszug-II/1045		<b><i>N1a1a1a1a</i></b>	87,09	ázsiai hun
Kenézlő-Fazekaszug-II/10936	ffi	<b><i>C4</i></b>	94,07	ázsiai hun
Sárrétudvari-Hízóföld/66	ffi	<b><i>H1c1</i></b>	83,3	skandináv-germán
Sárrétudvari-Hízóföld/103	nő	<b><i>C4a1</i></b>	91,39	ázsiai hun

#### 4.2.2. A PCR alapú eredmények felülvizsgálata az NGS eredmények alapján

A korábban általunk PCR alapú módszerrel jellemzett Karos-III. temetőből 15 mintát, valamint a Tömörly és mtsai., 2007-ben korábban közölt honfoglalókból 9 mintát, vagyis összesen 24 PCR alapú módszerrel haplotipizált mintát újra vizsgáltunk NGS szekvenálással (4.

táblázat). Ez alapján megítélhető a PCR alapú módszerek megbízhatósága, valamint jól összehasonlítható azon kétféle PCR alapú módszer, amit a korábbi aDNS vizsgálatok során alkalmaztak. A két módszer között az az elvi különbség, hogy a Csákyová és mtsai., 2016; Csősz és mtsai., 2016 ; Tömöry és mtsai., 2007-ekben először a HVR szekvenciákat olvasták le, majd az ebből kikövetkeztetett haplocsoportokhoz kerestek és határoztak meg néhány ezt megerősítő SNP-t a kódoló régióból. Ezzel szemben Neparáczki és mtsai., 2016-ban a HVR eredményektől függetlenül vizsgáltunk 22 kódoló régiós SNP-t, így a haplocsoport besorolás két független kísérlet egyeztetésével történt.

Az NGS eredmények hitelesítése céljából néhány minta esetén ugyanazon DNS kivonatból készítettünk részleges UDG kezelt és nem UDG kezelt könyvtárakat is. A kezeletlen könyvtár MapDamage profilja – melyből nem távolítottuk el az aDNS-re jellemző deaminált citozinokat - tipikus aDNS-re jellemző sérülés mintázatot mutatott (Függelék 2./a ábra; Rohland és mtsai., 2015), vagyis a szekvenciák többsége tele volt egyéni, csak az adott molekulánál megfigyelhető tranzíciókkal, miközben az átfedő olvasatokban ezek az eltérések kiegyenlítették egymást, és a végső olvasat ugyanazon szekvenciát adta mint a kezelt könyvtáré (Függelék 3. ábra). Ezzel szemben a részleges UDG kezelés a molekulák belső részéből eltávolította a tranzíciók nagyrészét, miközben a molekula végeken meghagyta az aDNS-re jellemző tranzíciós „ujjlenyomatot” (Függelék 2./b ábra). A könyvtárak átlagos fragment mérete 56 és 86 bp közé esett (3. táblázat), ami megfelel a degradált aDNS-re jellemző méret tartománynak (Sawyer és mtsai., 2012). Ezen az adatok alátámasztják, hogy a szekvencia olvasatok túlnyomó többsége endogén DNS molekulákból származik, de ezen adatokból megbecsülhető a külső kontamináció mértéke is oly módon, hogy a filogenetikai besorolás alapjául szolgáló SNP eltéréseknél meghatározzuk a konszenzustól eltérő SNP-t adó molekulák arányát (Fu és mtsai., 2013). A számítást elvégezve a kapott értékek mintáinkban nagyon alacsony kontaminációs szintet mutattak (3. táblázat). A konszenzus szekvenciák filogenetikai analízise (HaploGrep 2; Weissensteiner és mtsai., 2016) minden esetben ellentmondásmentes besorolást adott, ami a kontamináció mentesség további bizonyítéka. A végső konszenzus szekvenciákat feltöltöttük az NCBI GenBank adatbázisába, a következő elérhetőséggel: KY083702-KY083725.

**3. táblázat** Az NGS szekvenálási adatok részletei

temető/sírszám/minta név	mintavétel helye	össz readszám	rCRS-re térképezhető readok	egyedi readok száma	átlagos fragment hossz	a nukleotidok hány (%) van legalább 5x lefedve	becsült kontamináció (%)	(%) G-A tranzíciók a 3' végen (MapDamage)	(%) C-T tranzíciók az 5' végen (MapDamage)
Magyarhomoróg/120/anc2	fög	26152	12519	1656	55,97	89,62	0,00	6,11	7,81
Oroszláza-Görbics tanya/2/ anc3	combesont	85178	50555	3444	79,37	99,89	0,43	8,13	8,89
Szabadkigyós-Páljegy/7/anc4	fög	53176	27002	2854	63,85	97,46	0,45	8,19	9,22
Szegvár-Oromdűlő/412/anc5	fög	47760	23426	6150	60,07	99,95	0,75	5,49	6,40
Szegvár-Oromdűlő/593/anc6	fög	66798	35260	3841	58,09	98,64	2,08	6,47	8,70
Sárrétudvari-Hízófld/5/anc10	fög	17632	6664	2205	69,94	95,11	0,00	7,40	10,29
Sárrétudvari-Hízófld/118/anc12	fög	36214	14708	3394	63,96	98,81	0,36	8,75	11,53
Sárrétudvari-Hízófld/213/anc13	fög	42326	20383	4997	62,44	95,01	0,83	9,01	10,22
Harta-Freifeld/10/anc25	fög	135472	66169	4803	80,75	99,93	1,75	6,00	7,10
Karos-III/1	combesont	30830	5968	2414	81,44	98,07	0,00	13,09	11,48
Karos-III/3	combesont	58982	12858	3839	76,84	98,91	1,68	12,51	10,66
Karos-III/4	combesont	82194	22930	4862	68,71	99,98	0,33	12,93	11,98
Karos-III/5	metatarsus	394292	23740	10299	85,35	99,96	3,44	8,45	6,84
Karos-III/6	combesont	41054	3863	955	70,68	87,24	0,00	6,71	6,75
Karos-III/8	combesont	75724	26747	5938	64,26	99,67	1,79	11,37	11,08
Karos-III/10	combesont	67416	6371	1768	68,22	89,61	0,00	9,89	9,04
Karos-III/11	combesont	346124	4496	1569	79,62	93,91	1,32	12,78	11,42
Karos-III/12	combesont	52738	5843	2406	70,76	95,80	3,96	12,40	11,19
Karos-III/14	combesont	61346	16134	4532	69,29	99,69	1,82	14,53	14,13
Karos-III/15	combesont	422796	80431	13248	81,03	99,98	0,29	11,02	9,95
Karos-III/16	combesont	90950	23233	2851	75,65	99,14	1,34	11,34	11,12
Karos-III/17	combesont	43330	2626	1191	79,64	87,55	0,00	10,31	9,91
Karos-III/18	combesont	9184	3208	1596	68,48	90,49	0,00	15,42	12,11
Karos-III/19	fög	59102	30135	3244	69,07	98,78	0,00	6,39	6,55

Mivel az NGS szekvenciák olvasata közel 100%-osan megbízható, ezen szekvenciákból kigyűjtöttük azokat a HVR és kódoló régió SNP-eket, amelyeket a Neparáczi és mtsai., 2016; Tömör és mtsai., 2007-ekben PCR alapú módszerekkel korábban közöltek, és a 3. táblázatban összehasonlítottuk a kétféle módszerrel nyert eredményeket. A 3. táblázatból látható, hogy a Tömör és mtsai., 2007-ben közölt adatok közül az NGS szekvenálással újvizsgált 9 mintából 5 esetben a korábbi szekvenciákkal megegyező eredményt kaptunk, egy esetben a haplotipizálásba hiba csúszott, három esetben pedig teljesen más haplocsoportot határoztak meg. Az általunk korábban közölt szekvencia adatok közül a 15 újvizsgált szekvencia mindegyike ugyanazt a haplocsoportot adta, de ezek közül 8 esetben a haplotípus nem bizonyult helyesnek. Mindkét tanulmányban a hibák többségét a meglévő, de nem azonosított SNP-ék okozták, vagyis a PCR amplifikált DNS szekvencia olvasata nem volt megfelelő. Emellett a Tömör és mtsai., 2007-ben 3 olyan SNP-t is leolvastak, amely nem volt jelen a mintában (áthúzott nukleotid pozíciók a 4. táblázatban).

Elvileg az esetleges haplotípus eltérések a csontanyag tárolás-szállítás során történt összekeveréséből is származhatnak. Ezt a hibalehetőséget azonban ki tudtuk zárni azzal, hogy a két kísérletben eltérő eredményt mutató Szegvár-Oromdűlő/412/anc5 mintánál a Tömör és mtsai., 2007-ben korábban vizsgált combesontból (amelyen a mintavétel helye is látszott) és

ugyanazon minta foggyökeréből is készítettünk DNS kivonatot, melyek ugyanazt a szekvenciát adták.

Ezek az eredmények világosan jelzik, hogy a PCR alapú eredmények megbízhatósága korlátozott. Ennek az az oka, hogy az amplifikáció csupán néhány molekuláról indul, és az esetleges szennyező molekuláról induló amplifikátum könnyen többségbe kerülhet az endogén molekuláról induló társaival szemben. Az is kijelenthető, hogy a Nepraczkai és mtsai., 2016-ban használt módszer kisebb mértékű tévedéseket okozott (azaz a SNaPshot reakcióval kombinált HVR szekvenálás), ezért sokkal megbízhatóbb mint a Tömöry és mtsai., 2007-ben használt módszer (azaz HVR szekvenálás megerősítése kódoló régiós SNP-kkel).

4. táblázat Két eltérő PCR alapú módszer összehasonlítása NGS adatokkal. Sárgával a haplotipizálási hibákat, pirossal a téves haplocsoport besorolást emeltem ki.

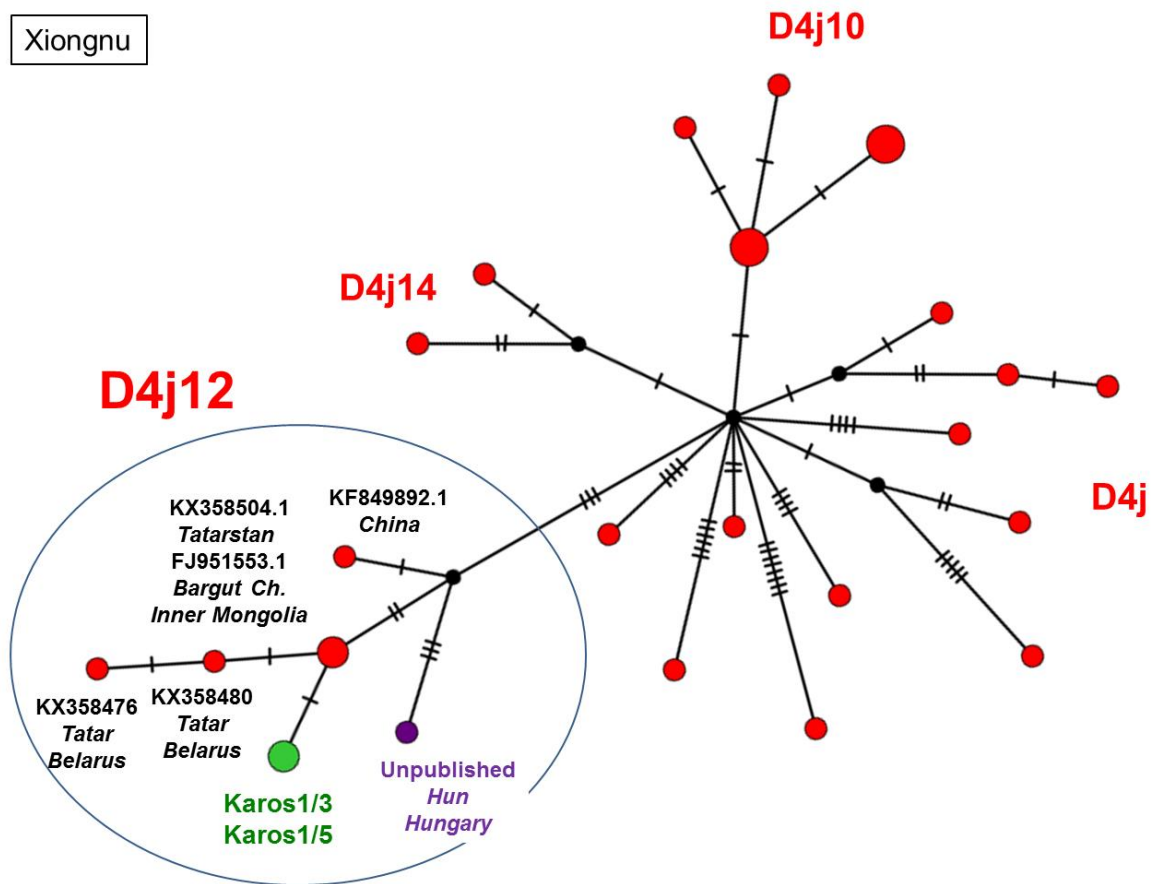
	cemetery/grave no. /sample name	HVR-I mutations found (position -16000)	HVR-II and coding region mutations studied / method	HVR-II and coding region mutations found	Hg described in the study (Hg with Haplogrep)	Haplotype identified by NGS in the present study	unnoticed SNP-s, or erroneously identified SNP-s in the region studied
Tömöry et al. 2007 mintái	Magyarhomoróg/120/anc2	CRS	73 7028 14766 /RFLP	-	H (H2a2a1)	<i>H84</i>	none
	Orosháza-Görbics tanya/2/anc3	147A 172C 183C 189C 223T 320T 355T	10238 / RFLP	10238C	N1a (L3e2b)	<i>N1a1a1a1a</i>	none
	Szabadkigyós-Pálliget/7/anc4	223T 356C	10400 12308 12705 / sequencing	12308G	U4 (U4a2b)	<i>N1a1a1a1</i>	16147A 16172C 16248T 16320T 16355T <del>16356C</del> 10398G <del>12308G</del> 12705T
	Szegvár-Oromdülő/412/anc5	CRS	73 7028 14766 /RFLP	14766T	H (R0)	<i>K1c1d</i>	16224C 16311C 73G 7028T
	Szegvár-Oromdülő/593/anc6	114A 192T 256T 270T 294T	12308 / sequencing	12308G	U5a1 (U5a2a)	<i>U5a2a1b</i>	none
	Sárrétudvari-Hízóföld/5/anc10	129A 148T 223T	10238 / RFLP 12705 / sequencing	10238C 12705T	I (N1)	<i>I5a1a</i>	16391A
	Sárrétudvari-Hízóföld/118/anc12	126C 182C 183C 189C 294T 296T 298C	9 bp del/ electrophoresis	9 bp del*	T (T2f1a)	<i>T2f1a1</i>	none
	Sárrétudvari-Hízóföld/213/anc13	311C	73 14766 / RFLP 11719 12308 12705 / sequencing	73G 11719A 14766T	R (R1)	<i>J1c3g</i>	16069T 16126C <del>16311C</del>
	Harta-Freifelt/10/anc25	294T 304C	73 7028 14766 /RFLP 10310 / sequencing	-	H (H5a4)	<i>H5e1a</i>	none
Neparáczi et al. 2016 mintái	Karos-III/1	183C 189C 217C	HVR-II: nt.190-309 sequenced and coding region 22 SNP-s of the GenoCoRe22 assay determined in all cases	263G 7028T 9bp del11719A 14766T	B4	<i>B4d1</i>	207A
	Karos-III/3	362C		239C 263G	H6	<i>H6a1b</i>	none
	Karos-III/4	069T 092C 126C 261T		228A 263G 295T 7028T 11719A 12612G 14766T	J1c7	<i>J1c7a</i>	none
	Karos-III/5	183C 189C 217C		263G 7028T 9bp del11719A 14766T	B4	<i>B4d1</i>	none
	Karos-III/6	189C		263G 7028T 9bp del11719A 14766T	B4'5	<i>B4d1</i>	16183C, 16217C
	Karos-III/8	051G 189C 362C		263G 7028T 11467G 11719A 14766T	U2e	<i>U2e1b</i>	217C, 16129C, 16256T,
	Karos-III/10	304C		263G	H5	<i>H5e1</i>	16189C 16294T
	Karos-III/11	189C 223T 278T		195C 257G 263G 6371T 7028T 11719A 12705T 14766T	X2f	<i>X2f</i>	16093C
	Karos-III/12	183C 189C 223T 290T 319A		235G 263G 4248C 7028T 11719A 12705T 14766T	A	<i>A12</i>	none
	Karos-III/14	126C 163G 186T 189C 294T		195G 263G 7028T 11719A 13368A 14766T	T1a	<i>T1a1b</i>	none
	Karos-III/15	069T 126C 362C		263G 295T 7028T 11719A 12612G 14766T	J	<i>J2a1</i>	16263 del
	Karos-III/16	256T 270T		263G 7028T 11467G 11719A 14766T	U5a	<i>U5a1a2a</i>	16399G
	Karos-III/17	362C		239C 263G	H6	<i>H6a1a</i>	none
	Karos-III/18	126C 163G 186T 189C 294T		214G 263G 7028T 11719A 13368A 14766T	T1a10a	<i>T1a10a</i>	none
Karos-III/19	126C 163G 186T 189C 294T	214G 263G 7028T 11719A 13368A 14766T	T1a10a	<i>T1a10a</i>	none		

### 4.2.3. A teljes mtDNS genomok filogenetikai eredményei

Ezen vizsgálat azt célozta, hogy az ismert adatbázisokból kikeressük az egyes honfoglalókhöz legközelebb eső szekvenciákat. Ismét hangsúlyozandó, hogy adatbázisunkban a világról eddig ismert összes modern és ősi mtDNS genom szekvencia megtalálható. Először meghatároztuk a mintánk haplotípusát (2. táblázat), majd adatbázisunkból kiválasztottuk az ezzel azonos és legközelebb álló haplotípusú mintákat (szekvenciákat). Az így összeállított csoportok mtDNS genom szekvenciáit egymáshoz illesztettük, majd ezek leszármazási viszonyait Median-Joining Network-el jelenítettük meg. Végül a filogenetikai törzsfákon a honfoglalókhoz közel eső szekvenciák földrajzi származását visszakerestük az adatbázisokból és az eredeti publikációkból (NCBI és mtDNA Community; 5. táblázat).

A vizsgált 89 minta 54 al-haplocsoportba volt sorolható, ezért 54 M-J Network leszármazási fát készítettünk. Ezek közül terjedelmi korlátok miatt csak négyet mutatok be. Az azonos szekvenciájú Karos-I/3 és Karos-I/5 minta a D4j12 haplotípusba tartozik (12. ábra), a hozzájuk legközelebb eső két, szintén azonos szekvenciájú recens minta (melyek ezért azonos körbe esnek) egyetlen nukleotid különbséget mutat, amit az összekötő vonalon lévő áthúzások száma jelez. Az egyik egy Belső Mongóliából származó bargut, a másik egy tatár egyéntől származik, és a a távolabbi találatok is ugyanazon népekből tatárokból és kínaiakból származnak. A ma Kelet-Európában élő tatárok Mongóliából származnak, és csak a tatárjárás idején kerültek mai helyükre. Fontos megemlíteni, hogy ugyanabba az alcsoportba tartozik egy ásatag hun minta Magyarországról (Vezér utcai lelet). Ezek alapján nagyon valószínű, hogy a Karos-I/3 és Karos-I/5 honfoglaló minták anyai vonalon Belső-Ázsiából származnak, és a hun találat azt is valószínűsíti, hogy ázsiai hun felmenővel (Xiongnu) rendelkeztek.

Xiongnu

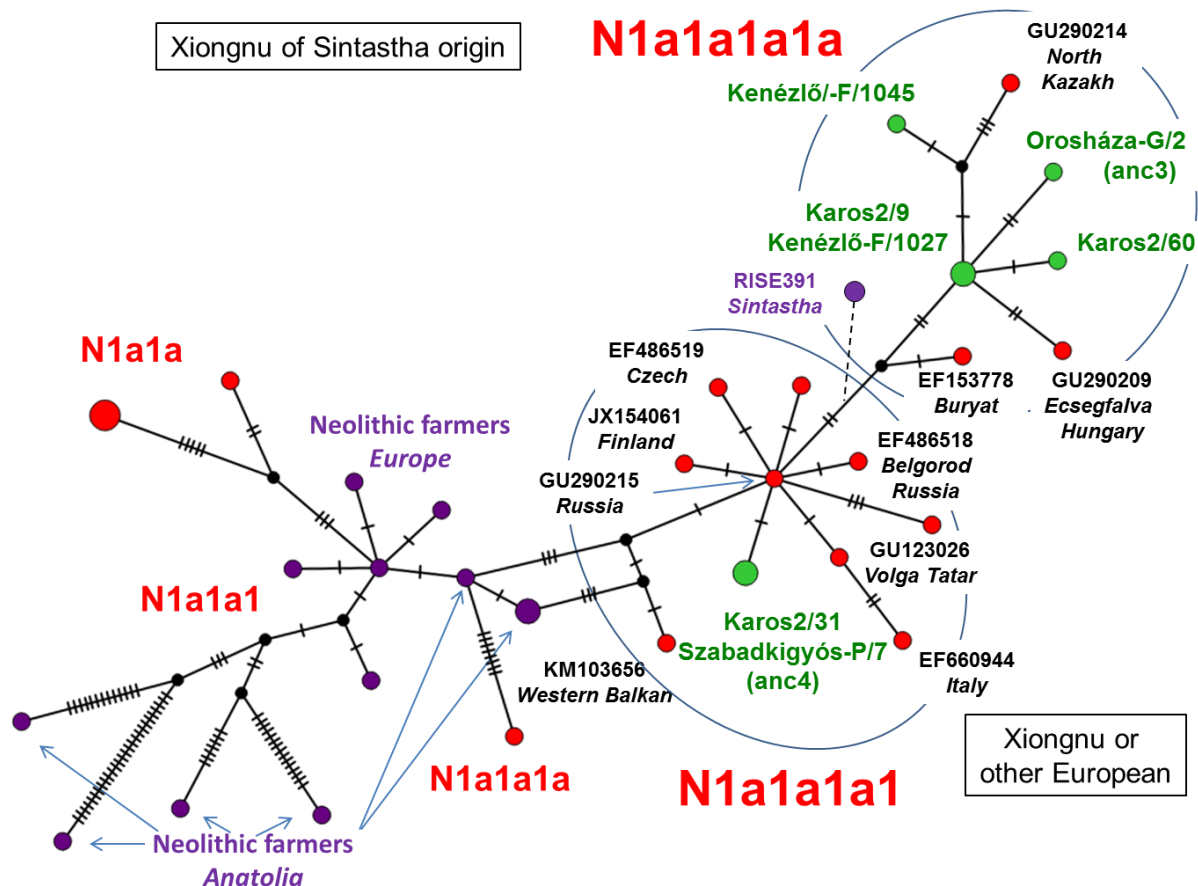


**12. ábra** A honfoglalókkal azonos D4j12 alcsoportba tartozó mintákat kék vonallal karikáztuk be. A piros körök recens mintákat, a lila körök ásatag mintákat, míg a zöld körök a vizsgált honfoglaló mintákat jelölik. A körök mérete arányos a bennük foglalt azonos szekvenciájú minták számával. Az összekötő vonalakon lévő áthúzások száma a két minta közötti nukleotid különbséget jelzi. A legközelebbi minták mellett feltüntettük a hozzájuk tartozó NCBI GenBank azonosító számokat és a származási helyet.

A 13. ábra az N1a1a1 al-haplocsoportba tartozó minták leszármazási viszonyait mutatja. Ez az ábra azt is érzékelteti, hogy az egyes al-haplocsoportok eltérő földrajzi régiókban találhatóak, vagyis a filogenetika finom felbontású filogeográfiai térképnek is megfelel, melyen az egyes alcsoportok elterjedési útvonalai is kirajzolódnak. Az N1a1a1 egy részletesen kutatott haplocsoport mivel igen ritka, és a neolitikus kor előtt teljesen hiányzott Európából. Recens mintákból már korábban ismerték ezt a haplocsoportot (Palanichamy és mtsai., 2010) mind Európából mind Ázsiából, de nem tudták eldönteni, hogy Ázsiába Európából, vagy a Közel-Keletről került át. Az ábráról leolvasható, hogy a neolitikus korban terjedt el a közel keletről a földművelés okozta népesség robbanással. Első európai megjelenését Neolitikus kori ásatag DNS mintákból mutatták ki (Haak és mtsai., 2015; Kilinc és mtsai., 2016). Eredményeink (13. ábra) azt mutatják, hogy az N1a1a1a1 típus a Kárpát-medencébe a honfoglalókkal érkezett. A 13. ábra

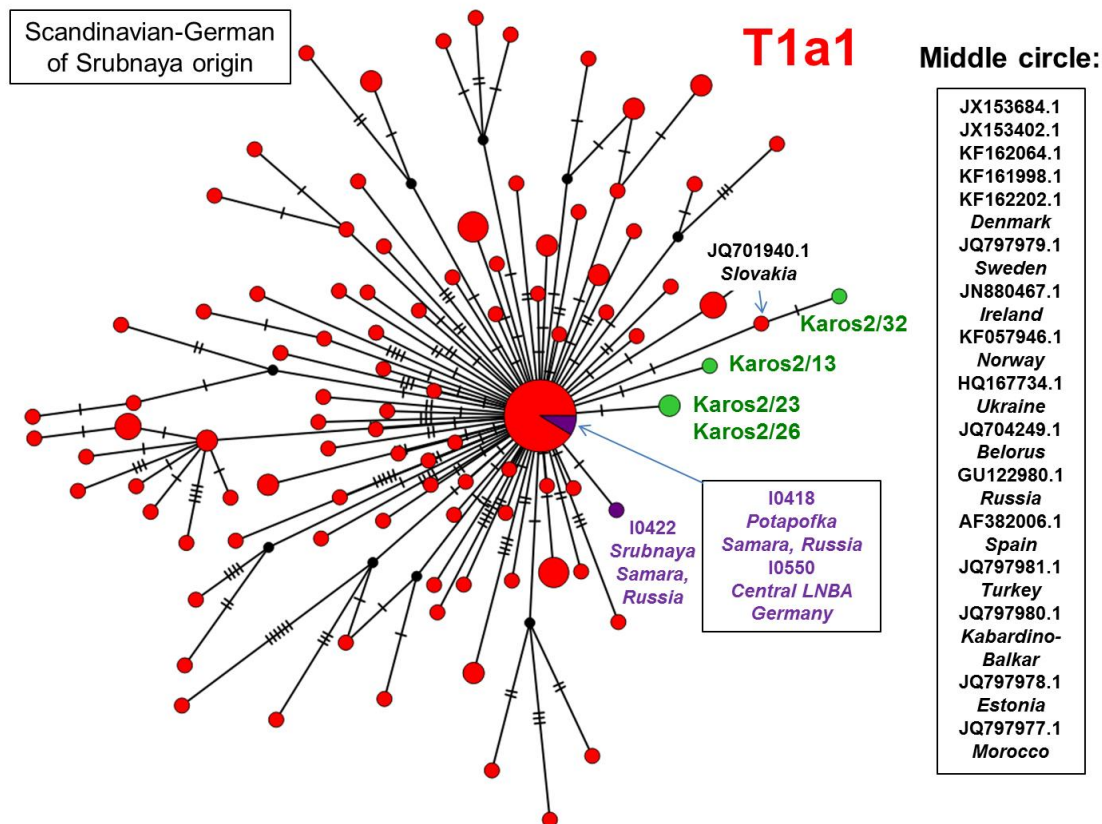


leszármazási viszonyai jól mutatják, hogy a honfoglalók ősei ezeket a géneket a Közel-Keletről kaphatták úgy, hogy a földművelés (és az ezzel járó génkészlet) egyszerre terjedt Európába, Kelet-Európába, majd onnan Ázsiába, és később az utóbbi ág Árpád népével visszatért a Kárpát medencébe. A Karos-II/31 és Szabadkigyós-Pálliget/7/anc4 honfoglalókhöz recens orosz, volgai tatár, cseh és finn minták mutattak közeli kapcsolatot, ami a kelet-európai ágak felel meg. A Karos-II/9, Karos-II/60, Kenézlő-Fazekaszug/1027, Kenézlő-Fazekaszug/1045 és Orosháza-Görbicstanya/2 honfoglaló mintákhoz pedig ma élő burját, kazah és egy magyar (Ecsegfalva) továbbá a Sintastha kultúrából származó ásatag lelet mutatott közeli kapcsolatot. Az N1a1a1a1 haplotípus a kelet-ázsiai ágat jelöli, melybe tartozó, i.e. 2000-re datálható korú Sintastha minta jelzi (Allentoft és mtsai., 2015), hogy az ázsiai ág már ekkor megjelent a közép-ázsiai lovas nomádokban.



**13. ábra** A honfoglalókkal azonos N1a1a1a1 és N1a1a1a1 alcsoportba tartozó mintákat kék vonallal karikáztuk be. A piros körök recens mintákat, a lila körök ásatag mintákat, míg a zöld körök a vizsgált honfoglaló mintákat jelölik. A körök mérete arányos a bennük foglalt azonos szekvenciájú minták számával. Az összekötő vonalakon lévő áthúzások száma a két minta közötti nukleotid különbséget jelzi. A legközelebbi minták mellett feltüntettük a hozzájuk tartozó NCBI GenBank azonosító számokat és a származási helyet.

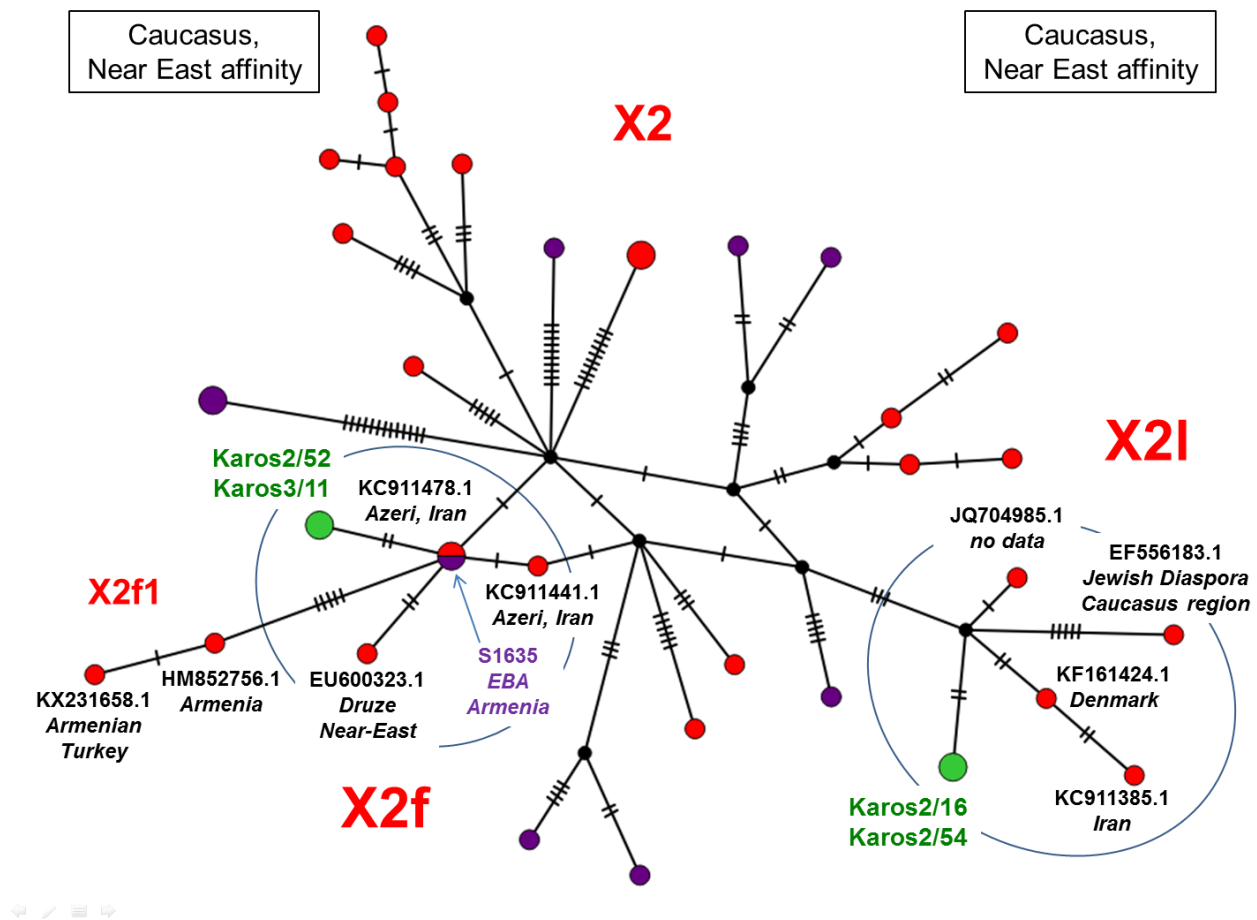
A honfoglaló minták szekvenciái közül 38-nak a legközelebbi szekvencia megfelelései elsősorban skandináv-germán népekből származó szekvenciák, erre mutatok be egy példát a 14. ábrán. Néhány haplotípus esetében viszonylag sok minta került egy alcsoportba, mivel nagyszámú ismert minta tartozik egyazon típusba, erre jó példa a T1a1 haplocsoport (14. ábra). Az ilyen népes alcsoportokban a leszármazási fán kirajzolódó központi kör nagyszámú azonos vagy csaknem azonos szekvenciájú mintát tartalmaz. A Karos-II/13, Karos-II/23, Karos-II/13 és a Karos-II/32 honfoglalókhöz legközelebbi recens szekvenciák dán, svéd, ír, norvég, ukrán, fehérorosz, orosz, spanyol, török, kabardino-balkar, litván és marokkói egyénekből származnak, továbbá egy ásatag lelet Oroszországból (Potapofka kultúra) és Németországból (késő neolitik/bronz kor). A Karos-II/32 honfoglalóhoz egy Szlovákia területéről származó recens minta áll legközelebb, de ebben az esetben gyanítható, hogy az egyén magyar felmenőkkel rendelkezett, hiszen Szlovákia közel ezer évig Magyarország egyik földrajzi régiója volt (Felvidék).



**14. ábra** A honfoglalókkal azonos T1a1 alcsoportba tartozó minták filogenetikai fája M-J Network algoritmussal kirajzolva. Itt az összes minta azonos alcsoportba tartozik, és az algoritmus láthatóan nehezen tudja strukturálni az igen hasonló szekvenciákat. A piros körök recens mintákat, a lila körök ásatag mintákat, míg a zöld körök a vizsgált honfoglaló mintákat jelölik. A körök mérete arányos a bennük foglalt azonos szekvenciájú minták számával. Az összekötő vonalakon lévő áthúzások száma a két minta közötti nukleotid különbséget jelzi. A legközelebbi minták mellett feltüntettük a hozzájuk tartozó NCBI GenBank azonosító számokat és a származási helyet.

A 15. ábrán egy olyan példát mutatok be, amely a honfoglalók egy részének kaukázusi, közel-keleti leszármazására utal. A Karos-II/52 és Karos-III/11 leletek az X2f haplotípusba tartoznak, amelyeket a mellékleteik alapján a régészek vezérnek határoztak meg. A hozzájuk legközelebb eső minta egy ásatag lelet Örményország területéről, amely korai Bronzkorból származik, de a legközelebbi recens minták is ugyanabból a földrajzi régióból származnak, iráni, drúz, örmény emberekből.

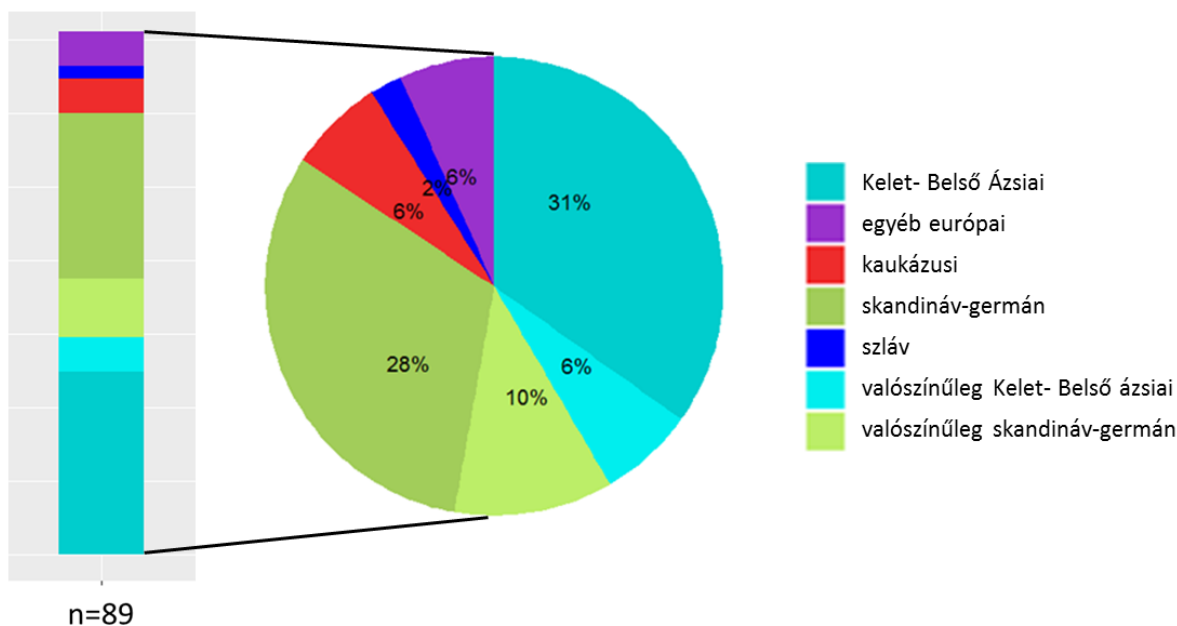
Az X2l haplotípusba tartozik a Karos-II/16 és Karos-II/54 honfoglaló lelet, melyek recens iráni, dán és kaukázusi zsidó diaszpórából származó mintákkal mutatnak közeli kapcsolatot. A dán találat ellenére ez a csoportot is inkább kaukázusi-közeli-keleti eredetűnek tűnik.



**15. ábra** A honfoglalókkal azonos X2f és X2l alcsoportba tartozó mintákat kék vonallal karikáztuk be. A piros körök recens mintákat, a lila körök ásatag mintákat, míg a zöld körök a vizsgált honfoglaló mintákat jelölik. A körök mérete arányos a bennük foglalt azonos szekvenciájú minták számával. Az összekötő vonalakon lévő áthúzások száma a két minta közötti nukleotid különbséget jelzi. A legközelebbi minták mellett feltüntettük a hozzájuk tartozó NCBI GenBank azonosító számokat és a származási helyét.

Az 54 filogenetikai fa alapján a fent bemutatott módon megállapítottuk az összes minta anyai vonalának valószínű származási helyét, amit a 16. ábrán összesítettem. A minták nagy

része egészen egyértelmű származást mutatott, 31%-uk Kelet-Belső Ázsiából, 28%-uk skandináv-germán népektől származó szekvenciákkal adott közeli megfelelést. A minták 6%-a mutatott nagy valószínűséggel kaukázusi-közel-keleti származást. Néhány esetben az adatok hiánya vagy ellentmondásos jellege miatt bizonytalanabb volt a földrajzi régió meghatározása, ilyenkor a legközelebbi szekvenciákat súlyozottan vettük figyelembe, különösen ha azokat néhány távolabbi szekvencia is megerősítette. A bizonytalanabb származású mintákat halvány színnel jelöltük az ábrán.



16. ábra Az általunk vizsgált 89 honfoglaló haplotípusainak valószínűsíthető eredete oszlop-, és kördiagramon

## 5. Az eredmények értékelése

Kísérleteinket a nemzetközi követelménynek megfelelő ásatag DNS laborban végeztük az archaikus DNS izolálás és szekvencia analízis korszerű módszereivel. Ez kezdetben az mtDNS HVR régió szekvenálásával, és 22 mtDNS kódoló régióban elhelyezkedő SNP vizsgálatával (GenoCoRe22) történt, amely munkánk kezdetekor a legjobb módszernek számított. A PCR alapú módszerrel az Y kromoszóma esetében kizárólag azt a 25 SNP-t vizsgáltunk (GenoY25), melyek a főcsoport besorolásához szükségesek. Ezt a metodikát Magyarországon először nekünk sikerült alkalmazni, ez a ún. klasszikus módszer legfejlettebb változata az archeogenetika területén.

A továbbiakban is igyekeztünk lépést tartani a szakterület élvonalával, ezért amint pénzügyi lehetőségünk nyílt rá áttértünk a dúsítással kombinált újgenerációs szekvenálásra. Első NGS munkáinkban újvizsgáltunk korábbi mintákat, és demonstráltuk a klasszikus és NGS módszer minősége közti különbséget. Kimutattuk, hogy az újgenerációs szekvenálással kapott eredmények tökéletesen megbízhatók, míg a PCR alapú haplotípus eredmények kevésbé biztosak.

### 5.1. A PCR-alapú eredmények értékelése

Első munkánk során a Karos-Eperjesszög III. számú kora honfoglalás kori temető 19 maradványát vizsgáltuk, melyből sikeresen meghatároztuk 17 lelet mitokondriális haplotípusát a HVR-I és HVR-II mtDNS szakaszok szekvenálásával, valamint a kódoló szakaszok 22 SNP pozícióját vizsgáló SNaPshot módszerrel. A szakmai elvárásoknak megfelelően a DNS kivonást és a haplotípus meghatározást többször, többféle módszerrel is megismételtük, és az egyes eredmények megerősítették egymást. Mind a HVR szekvenálás mind a 22 meghatározott kódoló SNP ugyanazt az eredményt adta, ezért a haplotípus és haplocsoport besorolásaink valós eredménynek tekinthetők. Munkamódszerünk kizárta a laboratóriumi humán DNS szennyezést, és a minták haplotípusa is különbözött a labor dolgozóinak haplotípusától.

Az mtDNS szekvenciákat az adatbázisokban fellelhető hasonló szekvenciákhoz hasonlítva az anyai ágú felmenők leszármazási vonalaira, migrációjára lehet következtetni, mivel bizonyos haplocsoportok földrajzi helyhez és populációkhoz köthetők. Adatbázis elemzéseink alapján a karosi populáció ősi európai és ősi ázsiai eredetű populációk keveredéséből

származhatott. A populációgenetikai eredmények arra utaltak, hogy ez a keveredés valószínűleg i.e. 2000-700 táján, a bronzkor végén történhetett az Andronovo kultúra területén (mai Kazahsztán), mert a vizsgált populáció az innen származó ősi populációkkal és ma itt található modern populációkkal mutatta a leghasonlóbb haplocsoport megoszlást. A haplocsoportok egy része arra utalt, hogy a karosi honfoglalók ősei innen nyugat felé vándorolva további európai genetikai elemekkel keveredtek. Fontos hangsúlyozni, hogy a karosi populáció egy része biztosan nem származhatott Ázsiából, mivel a bennük kimutatott X2f és a H5 mitokondriális haplocsoportok valamint az I2a és R1b1b Y haplocsoportok szinte teljesen hiányoznak erről a területről. Populációgenetikai analízisünk eredménye alapján arra következtettünk, hogy a karosiak végső összetétele valahol a kelet-európai sztyeppén alakulhatott ki, és az X2f, H5 haplocsoportok elterjedtsége, valamint a T1a és J haplocsoportok gyakoriság adatai alapján arra következtettünk, hogy az ázsiai eredetű populációkkal keveredő népségek egy része a Kaukázus vidékéről származhatott.

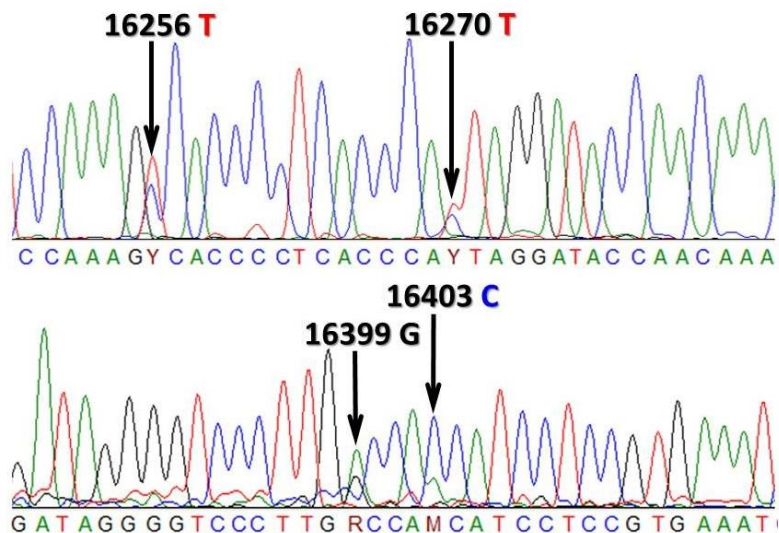
## **5.2. NGS eredmények összevetése PCR alapú eredményekkel**

A magas kópiaszámú mtDNS jobban megőrződik a régészeti maradványokban, mint az alacsony kópiaszámú magi DNS, ezért a legtöbb ásatag DNS szekvencia mitokondriumokból származik (Hofreiter és mtsai., 2001). A mitokondriumon belül a polimorf HVR kontroll régió egyedülálló mennyiségben tartalmaz filogenetikai információkat, ezért az mtDNS haplotipizálás során a HVR régió megszekvenálása vált az elsődleges célponttá. Azonban mivel a HVR polimorfizmusok nem elegendők a biztos haplocsoport besoroláshoz, ezért az egyértelmű besoroláshoz további kódoló régióban elhelyezkedő SNP-eket választottak ki (Behar és mtsai., 2007). Eleinte a kódoló SNP-eket egyesével határozták meg, RFLP-vel (Brown, 1980) vagy PCR klónok szekvenálásával, később egyre inkább teret nyert több kódoló SNP egyidejű vizsgálata multiplex PCR-rel kombinált SNaPshot módszerrel (Quintáns és mtsai., 2004; Salas és mtsai., 2005). Ezt a módszert fokozatosan adaptálták az archeogenetikai vizsgálatokban is (Bouakaze és mtsai., 2007; Haak és mtsai., 2010).

A kódoló SNP-eket egyesével meghatározni idő és költség igényes folyamat, ezért többnyire csak azokat vizsgálták, amelyek a HVR-I adatokkal meghatározott haplocsoport megerősítéséhez szükségesek. Ezzel a megközelítéssel az a fő probléma, hogy PCR termékekből készült aDNS szekvencia olvasatok egy része gyakran kontaminációból származik (Malmström

és mtsai., 2005). Amennyiben hibás a HVR szakasz olvasata, óhatatlanul rossz kódoló SNP-eket kezdenek el vizsgálni, ami végső soron rossz haplocsoport besoroláshoz vezethet. Ez lehet a magyarázata a Tömörý és mtsai., 2007-ben 3 hibás haplocsoport besorolásának is (4. táblázat). A GenoCoRe22 SNaPshot vizsgálat legnagyobb előnye, hogy az összes besoroláshoz fontos kódoló SNP-t vizsgálja függetlenül a HVR szekvenciától. A 22 kódoló SNP allél által meghatározott haplocsoportnak meg kell egyeznie a HVR szekvencia alapján meghatározott haplocsoporttal, így a haplocsoport besoroláshoz kétszeres kontrollunk van. Viszont ezzel a módszer sem mentesülünk a hibás haplotípus olvasatoktól. Ez a magyarázata a mi tanulmányunkban megjelent korrekt haplocsoport, de hibás haplotípus azonosításainak (4. táblázat; Neparáczi és mtsai., 2016).

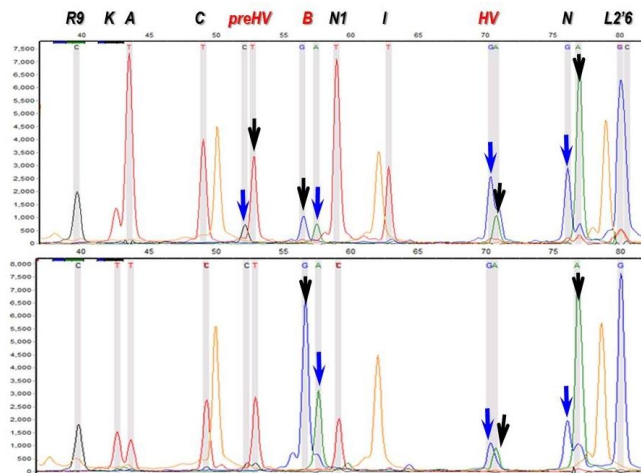
A PCR alapú aDNS szekvencia olvasatok problematikáját a 17. ábrán mutatom be. A PCR alapú szekvenciákban gyakoriak a kettős csúcsok, és ezek leolvasásakor a saját tanulmányunkban mindig a magasabb csúcsot vettük figyelembe a Neparáczi és mtsai., 2016-ban, amennyiben azok nem mondtak ellen a GenoCoRe22 adatoknak. A 17. ábrán bemutatott 16399 pozícióban a korrekt nukleotid is leolvasható lenne (A helyett G kellett volna olvasni, Táblázat 4.), de ezt nem vettük figyelembe, mert a G kisebb csúcsot adott, és ez a G nem volt szükséges SNP a haplocsoportba soroláshoz. Megjegyezendő, hogy a mellette elhelyezkedő dupla csúcs (16403 pozíció a 17. ábrán) esetében a jó olvasatot határoztuk meg mivel itt a magasabb csúcs volt a helyes nukleotid.



**17. ábra** A Karos-III/16 minta két HVR-I szekvencia kromatogramja a Neparáczi és mtsai., 2016-ból. A dupla csúcsokat nyilakkal jelöltük, az alsó szekvenciánál a jó olvasatok az alacsonyabb csúcsok lettek volna, amit az NGS eredmények alapján ismertünk meg.



A PCR alapú kódoló SNP meghatározás mindegyik módszerrel (RFLP, szekvenálás vagy SNaPshot) hasonló problémákkal szembesül, amit a 18. ábrán mutatok be. A multiplex PCR-rel amplifikált 22 mtDNS fragmentből két Single Base Extension (SBE) reakciót szokás összeállítani, ahol mindkét reakciók 11 haplocsoportot meghatározó SNP allél olvasatát adja. A 18. ábrán ugyanazon mintából készült két független SBE reakciót mutatok be, melyek számos kettős csúcsot tartalmaznak, melyek egyike bizonyosan kontaminációból származik. A SNaPshot reakciók ismétlésével néhány kettős olvasatot tisztázni lehet, így például az alsó elektroferogram alapján kizárható az ősi preHV allél, mivel ebben a pozícióban itt már csak egy csúcs (T) jelent meg. Ha ilyenfajta kizárás nem volt lehetséges, rutinszerűen a magasabb csúcsot választottuk, így a mint a B haplocsoportot definiáló kék csúcsot (G), vagy az N haplocsoportot definiáló zöld csúcsot (A) (18. ábra). Ezeket a döntéseket azonban mindig fenntartással kell kezelni, ezért például a B haplocsoportot definiáló 9 bp deléció megerősítettük single-plex PCR termék agaróz (metaphore) gélelektroforézisével is Neparáczki és mtsai., 2016-ban. Más esetekben a filogenetikai besorolás döntheti el, hogy melyik a valós csúcs (Cooper és Poinar, 2000), például a besorolás során a HV megelőzi a preHV-t, így amennyiben a preHV-t definiáló SNP egyértelműen eltér a referenciától, akkor értelemszerűen a HV-nek is el kell térnie. Ezért vettük figyelembe az alacsonyabb csúcsot (A) a HV esetében (18. ábra; Neparáczki és mtsai., 2016). A megbízható haplocsoport besoroláshoz SNaPshot reakciók ismétlése, azok összegzése továbbá a többször megszekvenált HVR olvasatok összevetése vezetett.



**18. ábra** A Karos-III/6 mintából származó két extraktum két SNaPshot SBE-II elektroferogram-ja Neparáczki és mtsai., 2016-ban. Az ábra tetején betűk jelentik a csúcsok alapján meghatározható haplocsoportot. A fekete betűkkel a színek által definiált ősi allélt, piros betűkkel a csúcsok alapján meghatározott származtatott allélt jelöltük. A nyilak a kettős csúcsokra mutatnak. Mivel minden festék eltérő mértékben módosítja a DNA futását így az azonos méretű de eltérő nukleotid beépülést tartalmazó csúcsok kissé eltérő pozícióba futnak. Fekete nyilak jelölik az általunk számításba vett csúcsokat, kékkel a mellőzött csúcsokat, melyek kontaminációból származhatnak. A narancssárga csúcsok a size standard-ek (GeneScan-120 LIZ, Applied Biosystems).



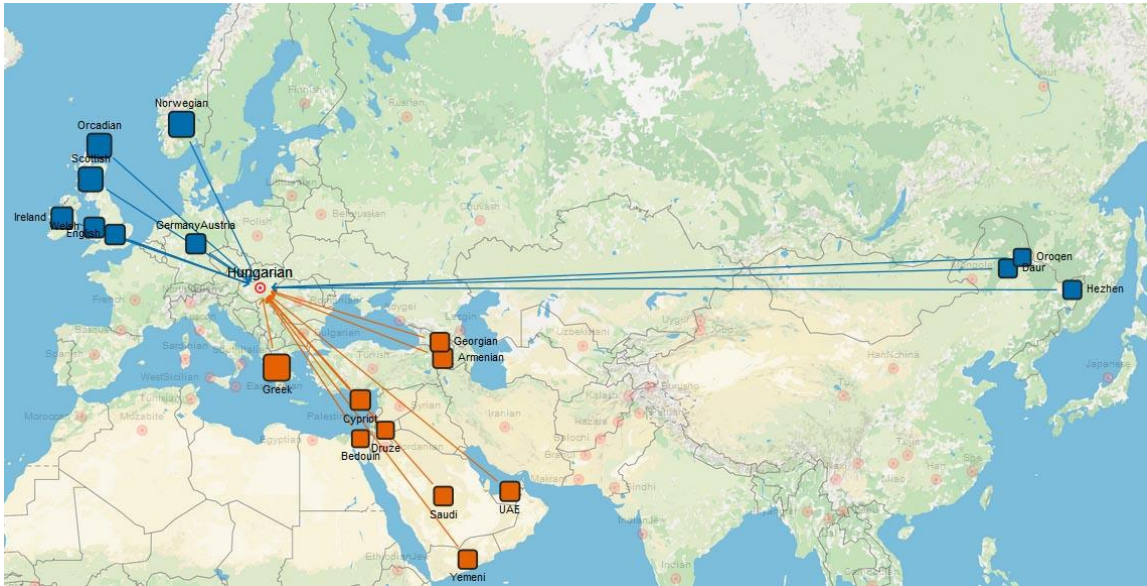
A vizsgálatokba bevont honfoglaló mintákat az 1930-90 évek között tárták fel, azóta sok kutató kezén megfordultak, kiknek többsége ma már lenyomozhatatlan. Nyilvánvaló, hogy a minták gyűjtése és tárolása során azok óhatatlanul kontaminálódtak. Tömöry és mtsai., 2007-ben a mintákat több temetőből válogatták össze és csak azokat publikálták, melyeknek legjobb volt a DNS megőrződése. Ennek ellenére módszerük hibákat eredményezett. A mi tanulmányunkban egy kis temető teljeskörű vizsgálata volt a cél (Neparáczi és mtsai., 2016), ezért nem volt lehetőség a minták között válogatni. Valószínűleg ennek is szerepe lehetett abban, hogy az alkalmazott megbízhatóbb módszer ellenére is hibásan olvastunk le néhány haplotípust. Fentiekből az következik, hogy a PCR alapú ásatag DNS haplotipizálás eredményeket nagyon óvatosan kell kezelni, pedig az aDNS tanulmányok többsége így készült (Gilbert és mtsai., 2005; Handt és mtsai., 1994; Richards és mtsai., 1995; Sampietro és mtsai., 2006). A hibás haplotípusok torzítják a statisztikai elemzéseket, így például az Fst számítást vagy a shared-haplotype analízist, melyeket a tanulmányok többségében rutinszerűen alkalmaztak (Csősz és mtsai., 2016; Tömöry és mtsai., 2007). A megbízhatóbb populációgenetikai vizsgálatokhoz hiteles, NGS-es szekvenáláson alapuló (vagy azzal megerősített) ásatag DNS adatokra van szükség.

### **5.3. Nagyfelbontású genetikai vizsgálatok**

A nagyfelbontású, teljes mtDNS genomokra alapozódó filogenetikai vizsgálatokkal lényegesen pontosabb képet tudunk alkotni a 89 honfoglaló anyai származásáról. A vizsgált honfoglalók eszerint anyai ágon nagyrészt két különböző, jól körvonalazható földrajzi területről származtak: Kelet-Belső Ázsiából, és Észak-Nyugat-Európából, továbbá egy kis részük a Kaukázus régióból, és a Közel-Keletről (16. ábra). Itt kell megemlítenünk, hogy az alacsony felbontású (HVR) vizsgálatok során tehát hibásan feltételeztük, hogy a karosi népesség egyetlen populáció lehet, mert a nagyfelbontású vizsgálatok szerint egyértelműen két eltérő eredetű populáció nem túl régi keveredéséből származott.

Következtetéseinket a ma élő népeken végzett genomi szekvenciákon alapuló úgynevezett Admixture vizsgálatok is megerősítik. Így például Hellenthal és mtsai., 2014-ben a ma élő magyarok genetikai összetételét vizsgálva megállapították, hogy a kelet-európai populációk közül a magyarok tartalmazzak legnagyobb méretű (4%) Ázsiából származó genom darabokat, melyek ma élő Oroqen, Mongol és Yakut genomokhoz állnak legközelebb. Az

Admixture vizsgálat a keveredés idejére is képes következtetni, és ezt a keveredést i.sz. 400-1000 közé datálták. Azt is kimutatták, hogy ezzel egyidőben jelentős skandináv génkészlet is bekerült a magyarokba, továbbá kaukázusi összetevőket is találtak. Ez teljes mértékben megfelel az általunk archaikus mintánkban talált összetevőknek (19. ábra).



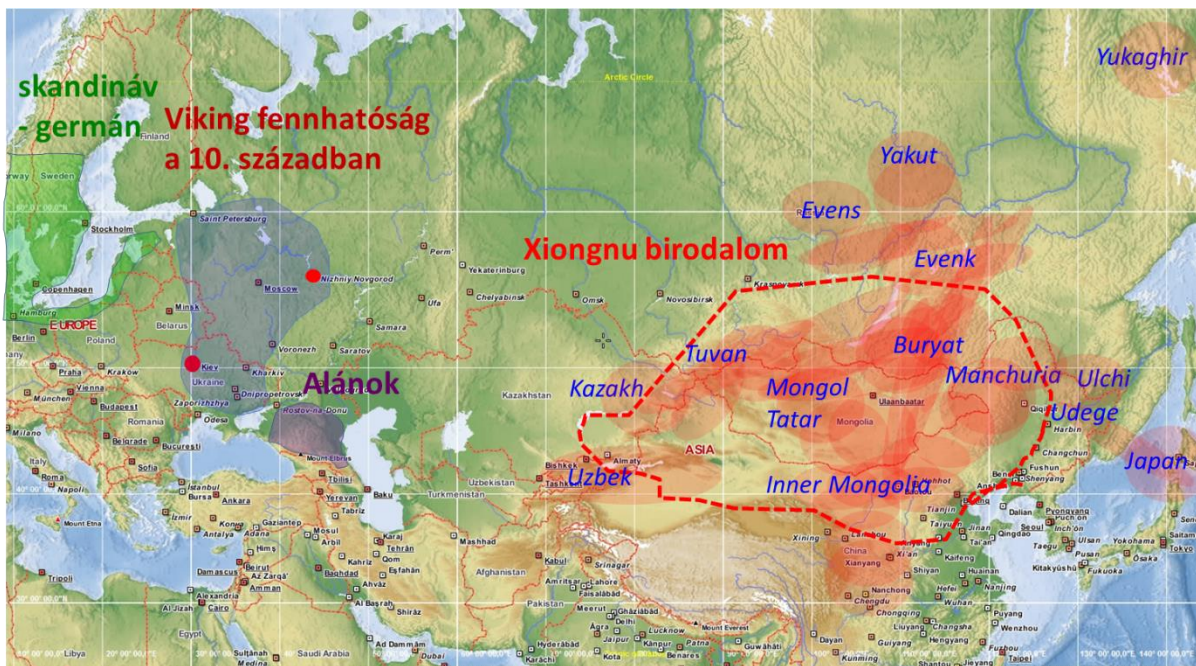
**19. ábra** A ma élő magyarok genetikai összetevői Admixtre analízis alapján (Hellenthal és mtsai., 2014). Az ázsiai és skandináv összetevők megjelenését egymással egyidőben, 400-1000 közé datálták. A mai magyarokban szintén kimutattak kaukázusi és Közél-keleti összetevőket is, de nem találtak finnugor kapcsolatot.

A hivatalosan elfogadott akadémiai álláspont szerint a magyarok finnugor eredetűek, azon belül is legközelebbi nyelvrokonaink a mansik és khantik, akikkel eddig genetikai leszármazási viszonyt is feltételeztek (Oktatási Hivatal, 2017). Épp ezért meglepő, hogy mintáinkban ezek a népcsoportok úgyszólván teljesen hiányoznak (a néhány finn megfelelés minden valószínűség szerint betelepült skandinávoktól származik), mansik legfeljebb elvétve a filogenetikailag távolabb eső oldalágakon jelennek meg, ami az mtDNS leszármazási fán több ezer éves távolságot jelent. Hellenthal és mtsai., 2014-ben sem találtak finnugor összetevőt a mai magyarokban.

Nemrég a legközelebbi nyelvrokonainknak számító khantik és mansik genetikai származását is vizsgálták genom szintű Admixture elemzéssel a Wong és mtsai., 2016-ban, akik kimutatták, hogy ezek a népek igen ősi európai génkészlettel rendelkeznek, melyhez 5-7 ezer évvel ezelőtt jelentős ázsiai génkészlet keveredett a mai evens és evenks populációk őseitől. Az ázsiai génkészletet egy viszonylag kisméretű populáció vihette be a finnugor népekbe, és az igen régi keveredésnek köszönhetően ezek a népek ma már többségükben eltérő alcsoportokba

tartoznak az azóta fellépő mutációk és genetikai sodródás következtében. Ugyanezzel magyarázható az eredetileg Kínából származó N-Tat Y kromoszóma haplocsoport jelenléte is a finnugor népekben (Rootsi és mtsai., 2007; Shi és mtsai., 2013).

Adataink összességében ellentmondanak a honfoglalók lehetséges finnugor származásának, és sokkal inkább összeegyeztethetők azok hun (Xiongnu=ázsiai hun) származásával. A Kelet-Ázsiai minták származási területe jól kirajzolja a Xiongnuk által egykor uralt területeket (20. ábra). Történeti források szerint ez a birodalom i.e. 50 körül szakadt ketté egy testvérviszály következtében, és a hunok egy része a mai Mongólia-Kína területén maradt (Csornai, 2007), másik részük nyugatra vándorolt. Ez jól egybevág az ismert archeogenetikai és antropológiai adatokkal, melyek szerint a mai Kazahsztán területén az ázsiai eredetű gének tömeges megjelenése pontosan ekkorra datálható (Ismagulov, 1982; Lalueza-Fox és mtsai., 2004). Mind a mongolok (Akim, 2016), mind egyes kazah törzsek a hunoktól származtatják magukat.



**20. ábra** A filogenetikai fák alapján az egyes honfoglalók kelet- belső-ázsiai találatainak földrajzi eloszlása látható halványpiros foltokkal. Ezek összessége jól kirajzolja az egykori ázsiai hun birodalmat (Xiongnu birodalom), amit piros szaggatott vonallal rajzoltam körbe.

Kínai írott források a Kr. u. 5. század közepéig írnak az ázsiai hunokról (Obrsuánszky, 2016), az európai hunok ekkor már eljutottak Európába és megtelepedtek a Kárpát-medencében. A kétféle hun népet a legtöbb kutató azonosítja (Bóna, 1993; Koch, 2007), de még mindig akadnak akik kételkednek az európai hunok xiongnu eredetében (Kelly, 2010). A mi adataink

közvetlen genetikai bizonyítéknak tekinthetők arra vonatkozóan, hogy az európai hunok Xiongnu eredetűek.

Adatainkat történeti források is támogatják, illetve adataink alapján számos eddig kétségbevonat hitelű történeti forrás hitelessége igazolható. Így például a középkori magyar krónikák szerint a “honfoglaló” magyarok őshazája Ázsia, akik testvéri-leszármazási viszonyban álltak a hunokkal. Krónikáink honfoglalás helyett a magyarok második bejöveteléről számolnak be (Bécsi Képes Krónika). IV. László udvarában Kézai Simon a megörökíti, hogy 1282-ben a magyar nemzetségfők magukat hun származásúnak tartják. A magyar néphagyomány (monda, népmese) is megőrizte a hun-magyar azonosságtudatot, emellett a mongol, kazah, török hagyomány is azt tartja, hogy a magyarok a hunok leszármazottai, és velük rokon nép, akik Belső-Ázsiából jöttek ki. Őseink Belső-Ázsiából a Kárpát-medencébe való jutása előtt mindig nagy kultúrnépek látókörében mozogtak, ennek köszönhetően jelentek meg nyomaik többek között kínai, tibeti, mongol, muszlim, görög írott forrásokban, melyek a hun és onogur (magyar) nép azonosságáról számolnak be (Thúry, 1897).

A genetikai eredmények alapján Árpád népe a kelet-európai sztyeppén maradt vagy oda visszavonuló hunok egyik ága lehetett. Ebből az következik, hogy a néphagyományok és a krónikák valós alapokon állnak, adataikat komolyan kell venni. Tehát nagy valószínűséggel a “második bejövetel” is valós alapokon áll, vagyis a magyar etnogenezis inentől fogva nem szűkíthető a “honfoglalókra”, hanem legalább az európai hun korig kiterjesztendő.

A honfoglalók genetikai összetétele arra utal, hogy a hun alaprég megőrzése mellett nagymértékben integrálták a Kelet-Európában talált germán elemeket, akiknek többsége adataink szerint skandináv eredetű. Ez alapján elsősorban a Ruszok (varégek, vikingek) jönnek számításba, akik a 9-10. században valóban a honfoglalók vonulási útvonalába eső területeket szálltak meg. A viking kapcsolatot támogató régészeti történeti adatok azonban csekélyek és ellentmondásosak. Kézai krónikája szerint például a Kijevnél legyőzött „oroszkok” közül sokan csatlakoztak a magyarokhoz. Egy generáció alatt azonban a vikingek nem válhattak lovas nomáddá, márpedig a vizsgált sírok tökéletes hun-germán genetikai-kulturális integrációt mutatnak, amely a germán elem sokkal régebbi integrációjára utal, továbbá a legyőzöttek nem csatlakozhattak ilyen magas arányban (40%).

A korabeli kelet európai sztyeppén skandináv-germán származású népek után kutatva a vikingeken kívül a gótokat találjuk, igaz pár 100 évvel korábban. Az európai hunok legyőzték a

gótokat valamint számos más germán törzset, és jelentős részüket integrálták birodalmukba. A gótok eredetileg is lovas nomád életmódot folytattak, így valószínűleg sokkal könnyebben keveredhettek a hunokkal mint a vikingek a honfoglalókkal. A Kárpát-medencéből a Fekete tenger vidékére visszaszorított hunok már egy erősen kevert népesség volt, és talán ezzel magyarázható a honfoglalók germán összetevője. A germán összetevő pontos eredetének tisztázása azonban további vizsgálatokat igényel.

A vezérek X2f haplocsoport kaukázusi eredetre utal, ez a terület pedig hosszú ideig az alánok fennhatósága alatt állt egészen addig, amíg a 370-es években a hunok hódoltatták őket. Az alánok egy része nyugatra menekült, más részük csatlakozott a hunokhoz. Valószínűleg ezt az integrációt örökölte meg a magyar mondai hagyomány, melyben Hunor és Magor elrabolják Dül alán király leányait. Ezek alapján úgy gondoljuk, hogy a kaukázusi haplocsoportok nagy valószínűséggel alán származásúak lehetnek.

Ugyan nem tartozik szorosan a témánkhoz, de szükségesnek tartjuk, hogy egy rövid kitekintést adjunk a magyar nyelv származására vonatkozó következtetésekről. Ezt azért tartjuk helyénvalónak, mivel a jelenlegi hivatalos álláspont a magyar nyelv Kárpát-medencébe érkezését egyértelműen a honfoglalókhoz köti. Eredményeink alapján ez a lehetőség csak akkor áll fenn, ha a magyar és hun nyelveket azonosnak tekintjük. A hunok nyelvét azonban a török nyelvek közé sorolják, és a nyelvészeti vizsgálatok igen komoly, ismeretlen néptől származó török eredetű szót azonosítottak nyelvünkben. Adataink alapján ez a réteg biztonsággal a hun nyelvből származtatható. A magyar nyelv finnugor rétegének származását pedig ezt követően csak a hunokat megelőző korok népeiben kereshetjük, akiknek kellően nagy populáció létszámmal kellett rendelkezniük ahhoz, hogy a germán, szláv, hun (török) népek tengerében is megőrizték nyelvüket. Szorosan ide kapcsolódik, hogy Wong és mtsai., 2016-ban kapcsolatot talált a mansi genomok és egy 6000 éves paleolit korú Kárpát-medencei genom között, ebben azt is leírták, hogy szerintük a nyelvi kapcsolat erre az időre nyúlhat vissza. Eredményeinknek ezért komoly kihatása van a történelem, régészet és nyelvészet jövőbeni kutatási irányaira.

## 6. Köszönetnyilvánítás

Köszönetet szeretnék mondani az opponenseimnek, és a bíráló bizottság tagjainak, a dolgozatomra fordított időért és hasznos észrevételeikért.

Köszönetemet szeretném kifejezni elsősorban témavezetőmnek, Dr. Török Tibornak, akihez kérdéseimmel mindig nyíltan fordulhattam, tanácsaival, tapasztalataival segített a munka során.

Szintén köszönettel tartozom Dr. Pálfi Györgynek, akinek ösztönző támogatása, baráti kapcsolatai nélkül nem jöhetett volna létre ez a dolgozat.

Hálás vagyok Genetikai Tanszék vezetőjének, Dr. Deák Péternek, továbbá a tanszékünk minden tagjának a támogatásért, hasznos tanácsokért.

Kiemelt köszönet illeti a szakdolgozóinkat, Kocsy Klaudiát, Tóth Gábor Endrét és Maár Kittit akik a kísérletek elvégzésében segítségemre voltak.

Szintén köszönetemet szeretném kifejezni Dr. Csányi Bernadettnek, aki megelőlegezett bizalommal segített a kutatásaink elvégzésében, és a rokonsági vizsgálatokat végezte.

Nagy köszönettel tartozunk az Dr. Albert Zink által vezetett labornak, főképp Dr. Frank Maixnernek, aki sok tanáccsal segítettek abban, hogy miként lehet létrehozni egy modern archeogenetikai labort továbbá segítettek az általuk használt technikák elsajátításában.

Dr. Raskó Istvánnak is köszönöm a támogatást, nélküle nem lett volna Szegeden archeogenetikai kutatás, és mind a mai napig szem előtt tartja ezen kutatások fontosságát.

Köszönöm Dr. Pamjav Horolmának, Dr. Juhász Zoltánnak és Dr. Fehér Tibornak, akik sokat segítettek nekünk az PCR alapú adatok elemzésékből.

Köszönöm Dr. Nagy Istvánnak, aki rugalmasan segített az új-generációs szekvenálásban, melyben nagy segítségünkre volt Dr. Bihari Péter, aki a MiSeq szekvenálás gyakorlati kivitelezését végezte.

Hatalmas köszönettel tartozom Dr. Maróti Zoltánnak, aki évekkal gyorsította meg az NGS szekvencia olvasatok adatelemzését. Továbbá Dr. Kalmár Tibornak is köszönöm az értékes tanácsait, észrevételeit.

Szeretném megköszönni Dr. Révész Lászlónak, akinek precíz régészeti feltárása lehetővé tette, hogy kutathassuk a honfoglalók eredetét. Itt szeretném megköszönni az összes régész és

antropológus munkáját, köztük Dr. Molnár Erikáét, Dr. Pap Ildikóét és Dr. Kustár Ágnesét, akik a mintákat adták, és akik nélkül nem jöhetett volna létre ez a munka.

Továbbá szeretném megköszönni az Szegedi Tudományegyetem Molekuláris Biológiai és Biokémia Tanszék valamint a Biotechnológiai Tanszék kollégáinak, hogy a kutatásaimhoz nélkülözhetetlen eszközeiket használhattam.



## 7. Irodalomjegyzék

- Adler, C.J., Haak, W., Donlon, D., Cooper, A., 2011. Survival and recovery of DNA from ancient teeth and bones. *J. Archaeol. Sci.* 38, 956–964. doi:10.1016/j.jas.2010.11.010
- Akim, H.G., 2016. Atila utódai még mindig imádkoznak az Arany Naphoz. T3 Kiadó, Sepsziszentgyörgy.
- Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., Walter, P., 2014. *Molecular Biology of the Cell* 6e, Garland Science. doi:10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C
- Allentoft, M.E., Sikora, M., Sjögren, K.-G., Rasmussen, S., Rasmussen, M., Stenderup, J., Damgaard, P.B., Schroeder, H., Ahlström, T., Vinner, L., Malaspinas, A.-S., Margaryan, A., Higham, T., Chivall, D., Lynnerup, N., Harvig, L., Baron, J., Casa, P. Della, Dąbrowski, P., Duffy, P.R., Ebel, A. V., Epimakhov, A., Frei, K., Furmanek, M., Gralak, T., Gromov, A., Gronkiewicz, S., Grupe, G., Hajdu, T., Jarysz, R., Khartanovich, V., Khokhlov, A., Kiss, V., Kolář, J., Kriiska, A., Lasak, I., Longhi, C., McGlynn, G., Merkevicius, A., Merkyte, I., Metspalu, M., Mkrtychyan, R., Moiseyev, V., Paja, L., Pálfi, G., Pokutta, D., Pospieszny, Ł., Price, T.D., Saag, L., Sablin, M., Shishlina, N., Smrčka, V., Soenov, V.I., Szeverényi, V., Tóth, G., Trifanova, S. V., Varul, L., Vicze, M., Yepiskoposyan, L., Zhitenev, V., Orlando, L., Sichevitz-Pontén, T., Brunak, S., Nielsen, R., Kristiansen, K., Willerslev, E., 2015. Population genomics of Bronze Age Eurasia. *Nature* 522, 167–172. doi:10.1038/nature14507
- Amigo, J., Phillips, C., Lareu, M., Carracedo, Á., 2008. The SNPforID browser: An online tool for query and display of frequency data from the SNPforID project. *Int. J. Legal Med.* 122, 435–440. doi:10.1007/s00414-008-0233-7
- Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., Howell, N., 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat. Genet.* 23, 147. doi:10.1038/13779
- Andrews, S., 2016. FastQC A Quality Control tool for High Throughput Sequence Data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Bandelt, H.-J., Forster, P., Rohl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37–48.
- Bécsi Képes Krónika, é. n.
- Behar, D.M., Rosset, S., Blue-Smith, J., Balanovsky, O., Tzur, S., Comas, D., Mitchell, R.J., Quintana-Murci, L., Tyler-Smith, C., Wells, R.S., 2007. The geographic project public participation mitochondrial DNA database. *PLoS Genet.* 3, 1083–1095. doi:10.1371/journal.pgen.0030104
- Benoit, J.-N., Quatrehomme, G., F Carle, G., Pognonec, P., 2013. An alternative procedure for extraction of DNA from ancient and weathered bone fragments. *Med. Sci. Law* 53, 100–6. doi:10.1258/msl.2012.012026
- Bíró, A., Fehér, T., Bárány, G., Pamjav, H., 2015. Testing Central and Inner Asian admixture among contemporary Hungarians. *Forensic Sci. Int. Genet.* 15, 121–126. doi:10.1016/j.fsigen.2014.11.007
- Bíró, A.Z., Zalán, A., Völgyi, A., Pamjav, H., 2009. A Y-chromosomal comparison of the Madjars (Kazakhstan) and the Magyars (Hungary). *Am. J. Phys. Anthropol.* 139, 305–310. doi:10.1002/ajpa.20984



- Bogácsi-Szabó, E., Csányi, B., Tömöry, G., Blazsó, P., Csósz, A., Kiss, D., Langó, P., Köhler, K., Raskó, I., 2008. Archeogenetikai vizsgálatok a Kárpát-medence X. századi népességén. *Magy. Tudomány* 10, 1204–1217.
- Bogácsi-Szabó, E., Kalmár, T., Csányi, B., Tömöry, G., Czibula, A., Priskin, K., Horváth, F., Downes, C.S., Raskó, I., 2005. Mitochondrial DNA of ancient Cumanians: culturally Asian steppe nomadic immigrants with substantially more western Eurasian mitochondrial DNA lineages. *Hum. Biol. an Int. Rec. Res.* 77, 639–662. doi:10.1353/hub.2006.0007
- Bóna, I., 1993. Hunok és nagykirályaik. Corvina, Budapest.
- Bouakaze, C., Keyser, C., Amory, S., Crubézy, E., Ludes, B., 2007. First successful assay of Y-SNP typing by SNaPshot minisequencing on ancient DNA. *Int. J. Legal Med.* 121, 493–499. doi:10.1007/s00414-007-0177-3
- Brandstätter, A., Egyed, B., Zimmermann, B., Duftner, N., Padar, Z., Parson, W., 2007. Migration rates and genetic structure of two Hungarian ethnic groups in Transylvania, Romania. *Ann. Hum. Genet.* 71, 791–803. doi:10.1111/j.1469-1809.2007.00371.x
- Brandt, G., Haak, W., Adler, C.J., Roth, C., Szécsényi-Nagy, A., Karimnia, S., Möller-Rieker, S., Meller, H., Ganslmeier, R., Friederich, S., Dresely, V., Nicklisch, N., Pickrell, J.K., Sirocko, F., Reich, D., Cooper, A., Alt, K.W., 2013. Ancient DNA reveals key stages in the formation of central European mitochondrial genetic diversity. *Science* (80-. ). 342, 257–261. doi:10.1126/science.1241844
- Brinkmann, B., Klintschar, M., Neuhuber, F., Hühne, J., Rolf, B., 1998. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am. J. Hum. Genet.* 62, 1408–1415. doi:10.1086/301869
- Brown, W.M., 1980. Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. *Proc. Natl. Acad. Sci. U. S. A.* 77, 3605–3609. doi:10.1073/pnas.77.6.3605
- Bruder, C.E.G., Piotrowski, A., Gijbers, A.A.C.J., Andersson, R., Erickson, S., Diaz de Ståhl, T., Menzel, U., Sandgren, J., von Tell, D., Poplawski, A., Crowley, M., Crasto, C., Partridge, E.C., Tiwari, H., Allison, D.B., Komorowski, J., van Ommen, G.J.B., Boomsma, D.I., Pedersen, N.L., den Dunnen, J.T., Wirdefeldt, K., Dumanski, J.P., 2008. Phenotypically Concordant and Discordant Monozygotic Twins Display Different DNA Copy-Number-Variation Profiles. *Am. J. Hum. Genet.* 82, 763–771. doi:10.1016/j.ajhg.2007.12.011
- 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? *Ann. Anat.* 194, 7–16. doi:10.1016/j.aanat.2011.07.003
- Cann, R.L., Stoneking, M., Wilson, A.C., 1987. Mitochondrial DNA and human evolution. *Nature.* doi:10.1038/325031a0
- Champlot, S., Berthelot, C., Pruvost, M., Andrew Bennett, E., Grange, T., Geigl, E.M., 2010. An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. *PLoS One* 5. doi:10.1371/journal.pone.0013042
- Collins, F.S., Brooks, L.D., Chakravarti, A., 1998. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* 8, 1229–1231. doi:10.1101/gr.8.12.1229
- Cooper, A., Poinar, H.N., 2000. Ancient DNA: do it right or not at all. *Science.* doi:10.1126/science.289.5482.1139b
- Czeizel, E., Benkman, H., Goedde, H.W., 1991. Genetics of the Hungarian population.
- Csákyová, V., Szécsényi-Nagy, A., Csosz, A., Nagy, M., Fusek, G., Langó, P., Bauer, M.,

- Mende, B.G., Makovický, P., Bauerová, M., 2016. Maternal genetic composition of a medieval population from a Hungarian-Slavic contact zone in central Europe. *PLoS One* 11. doi:10.1371/journal.pone.0151206
- Csányi, B., Bogácsi-Szabo, E., Tömöry, G., Czibula, Á., Priskin, K., Csősz, A., Mende, B., Langó, P., Csete, K., Zsolnai, A., Conant, E.K., Downes, C.S., Raskó, I., 2008. Y-chromosome analysis of ancient Hungarian and two modern Hungarian-speaking populations from the Carpathian Basin. *Ann. Hum. Genet.* 72, 519–534. doi:10.1111/j.1469-1809.2008.00440.x
- Csornai, K., 2007. Négy égtájon barbár csillag ragyog. László Gyula Történelmi és Kulturális Egyesület, Budapest.
- Csősz, A., Szécsényi-Nagy, A., Csakyova, V., Langó, P., Bódis, V., Köhler, K., Tömöry, G., Nagy, M., Mende, B.G., 2016. Maternal Genetic Ancestry and Legacy of 10th Century AD Hungarians. *Sci. Rep.* 6. doi:10.1038/srep33446
- Dabney, J., Knapp, M., Glocke, I., Gansauge, M.-T., Weihmann, A., Nickel, B., Valdiosera, C., García, N., Pääbo, S., Arsuaga, J.-L., Meyer, M., 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proc. Natl. Acad. Sci. U. S. A.* 110, 15758–63. doi:10.1073/pnas.1314445110
- 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. *Sci. Rep.* 5, 11184. doi:10.1038/srep11184
- Der Sarkissian, 2011. Mitochondrial DNA in ancient human populations of Europe. University of Adelaide.
- Edwards, M., Bigham, A., Tan, J., Li, S., Gozdzik, A., Ross, K., Jin, L., Parra, E.J., 2010. Association of the OCA2 polymorphism His615Arg with melanin content in East Asian populations: Further evidence of convergent evolution of skin pigmentation. *PLoS Genet.* 6. doi:10.1371/journal.pgen.1000867
- Egyed, B., Brandstatter, A., Irwin, J.A., Pádár, Z., Parsons, T.J., Parson, W., 2007. Mitochondrial control region sequence variations in the Hungarian population: Analysis of population samples from Hungary and from Transylvania (Romania). *Forensic Sci. Int. Genet.* 1, 158–162. doi:10.1016/j.fsigen.2007.03.001
- Enattah, N.S., Sahi, T., Savilahti, E., Terwilliger, J.D., Peltonen, L., Järvelä, I., 2002. Identification of a variant associated with adult-type hypolactasia. *Nat. Genet.* 30, 233–237. doi:10.1038/ng826
- Erratum for the Report “Ancient Ethiopian genome reveals extensive Eurasian admixture in Eastern Africa” (previously titled “Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent”) by M. Gallego Llorente, E. R. Jon, 2016. *Science* (80-. ). 351.
- Éry, K., 1994. A Kárpát-medence embertani képe a honfoglalás korában, in: *Honfoglalás és nyelvészet.* o. 217–224.
- Fehér, T., Németh, E., Vándor, A., Kornienko, I. V., Csáji, L.K., Pamjav, H., 2014. Y-SNP L1034: limited genetic link between Mansi and Hungarian-speaking populations. *Mol. Genet. Genomics* 290, 377–386. doi:10.1007/s00438-014-0925-2
- Fernández-Silva, P., Enriquez, J.A., Montoya, J., 2003. Replication and transcription of mammalian mitochondrial DNA. *Exp. Physiol.* 88, 41–56. doi:10.1113/eph8802514
- Fóthi, E., 1998. Összehasonlító embertani vizsgálat a Kárpát-medence népeinek ethnogeneziséhez. Embertani kapcsolatok a 6–8. században az eurázsiai steppe és a Kárpát-

- medence között. Móra Ferenc Múzeum Évkönyve Stud. Archaeol. 4, 497–521.
- Fu, Q., Mittnik, A., Johnson, P.L.F., Bos, K., Lari, M., Bollongino, R., Sun, C., Giemsch, L., Schmitz, R., Burger, J., Ronchitelli, A.M., Martini, F., Cremonesi, R.G., Svoboda, J., Bauer, P., Caramelli, D., Castellano, S., Reich, D., Pääbo, S., Krause, J., 2013. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* 23, 553–559. doi:10.1016/j.cub.2013.02.044
- Fulton, T.L., 2012. Setting Up an Ancient DNA Laboratory, in: Shapiro, B., Hofreiter, M. (Szerk.), *Ancient DNA: Methods and Protocols*. Humana Press, Totowa, NJ, o. 1–11. doi:10.1007/978-1-61779-516-9\_1
- Gamba, C., Jones, E.R., Teasdale, M.D., McLaughlin, R.L., Gonzalez-Fortes, G., Mattiangeli, V., Domboróczki, L., Kóvári, I., Pap, I., Anders, A., Whittle, A., Dani, J., Raczky, P., Higham, T.F.G., Hofreiter, M., Bradley, D.G., Pinhasi, R., 2014. Genome flux and stasis in a five millennium transect of European prehistory. *Nat. Commun.* 5, 5257. doi:10.1038/ncomms6257
- Garrison, E., Marth, G., 2012. Haplotype-based variant detection from short-read sequencing. arXiv Prepr. arXiv1207.3907 9. doi:arXiv:1207.3907 [q-bio.GN]
- Gilbert, M.T.P., Rudbeck, L., Willerslev, E., Hansen, A.J., Smith, C., Penkman, K.E.H., Prangenberg, K., Nielsen-Marsh, C.M., Jans, M.E., Arthur, P., Lynneruo, N., Turner-Walker, G., Biddle, M., Kjølbye, B., Collins, M.J., 2005. Biochemical and physical correlates of DNA contamination in archaeological human bones and teeth excavated at Matera, Italy. *J. Archaeol. Sci.* 32, 785–793. doi:10.1016/j.jas.2004.12.008
- Giles, R.E., Blanc, H., Cann, H.M., Wallace, D.C., 1980. Maternal inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci. U. S. A.* 77, 6715–9. doi:10.1073/pnas.77.11.6715
- Green, R.E., Briggs, A.W., Krause, J., Prüfer, K., Burbano, H. a, Siebauer, M., Lachmann, M., Pääbo, S., 2009. The Neandertal genome and ancient DNA authenticity. *EMBO J.* 28, 2494–2502. doi:10.1038/emboj.2009.222
- Guglielmino, C.R., Beres, J., 1996. Genetic structure in relation to the history of Hungarian ethnic groups. *Hum. Biol.* 68, 335–355.
- Guglielmino, C.R., De Silvestri, A., Beres, J., 2000. Probable ancestors of Hungarian ethnic groups: an admixture analysis. *Ann. Hum. Genet.* 64, 145–159. doi:doi:10.1017/S0003480000008010
- Guglielmino, C.R., Piazza, A., Menozzi, P., Cavalli??Sforza, L.L., 1990. Uralic genes in Europe. *Am. J. Phys. Anthropol.* 83, 57–68. doi:10.1002/ajpa.1330830107
- Haak, W., Balanovsky, O., Sanchez, J.J., Koshel, S., Zaporozhchenko, V., Adler, C.J., der Sarkissian, C.S.I., Brandt, G., Schwarz, C., Nicklisch, N., Dresely, V., Fritsch, B., Balanovska, E., VILLEMS, R., Meller, H., Alt, K.W., Cooper, A., 2010a. Ancient DNA from European early Neolithic farmers reveals their near eastern affinities. *PLoS Biol.* 8. doi:10.1371/journal.pbio.1000536
- Haak, W., Balanovsky, O., Sanchez, J.J., Koshel, S., Zaporozhchenko, V., Adler, C.J., der Sarkissian, C.S.I., Brandt, G., Schwarz, C., Nicklisch, N., Dresely, V., Fritsch, B., Balanovska, E., VILLEMS, R., Meller, H., Alt, K.W., Cooper, A., 2010b. Ancient DNA from European early Neolithic farmers reveals their near eastern affinities. *PLoS Biol.* 8. doi:10.1371/journal.pbio.1000536
- Haak, W., Brandt, G., Jong, H.N. d, Meyer, C., Ganslmeier, R., Heyd, V., Hawkesworth, C., Pike, A.W.G., Meller, H., Alt, K.W., 2008. Ancient DNA, Strontium isotopes, and

- osteological analyses shed light on social and kinship organization of the Later Stone Age. *Pnas* 105, 18226–18231. doi:10.1073/pnas.0807592105
- Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., Fu, Q., Mittnik, A., Bánffy, E., Economou, C., Francken, M., Friederich, S., Pena, R.G., Hallgren, F., Khartanovich, V., Khokhlov, A., Kunst, M., Kuznetsov, P., Meller, H., Mochalov, O., Moiseyev, V., Nicklisch, N., Pichler, S.L., Risch, R., Rojo Guerra, M. a., Roth, C., Szécsényi-Nagy, A., Wahl, J., Meyer, M., Krause, J., Brown, D., Anthony, D., Cooper, A., Alt, K.W., Reich, D., 2015. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522, 207–211. doi:10.1038/nature14317
- Handt, O., Höss, M., Krings, M., Pääbo, S., 1994. Ancient DNA: Methodological challenges. *Experientia*. doi:10.1007/BF01921720
- Hellenthal, G., Busby, G.B.J., Band, G., Wilson, J.F., Capelli, C., Falush, D., Myers, S., 2014. A genetic atlas of human admixture history. *Science* 343, 747–51. doi:10.1126/science.1243518
- Higuchi, R., Bowman, B., Freiberger, M., Ryder, O. a, Wilson, a C., 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature* 312, 282–284. doi:10.1038/312282a0
- Hofreiter, M., Serre, D., Poinar, H.N., Kuch, M., Pääbo, S., 2001. Ancient DNA. *Nat. Rev. Genet.* 2, 353–9. doi:10.1038/35072071
- Ismagulov, O., 1982. Ethnic anthropology of Kazakhstan. Institute of History, Archeology and Ethnography of the Academy of Science of Kazakhstan.
- Jobling, M.A., Tyler-Smith, C., 2003. The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4, 598–612. doi:10.1038/nrg1124
- Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P.L.F., Orlando, L., 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29, 1682–4. doi:10.1093/bioinformatics/btt193
- Jorde, L.B., Wooding, S.P., 2004. Genetic variation, classification and ‘race’. *Nat. Genet.* 36, 528–533. doi:10.1038/ng1435
- Juhász, Z., Fehér, T., Bárány, G., Zalán, A., Németh, E., Pádár, Z., Pamjav, H., 2015. New clustering methods for population comparison on paternal lineages. *Mol. Genet. Genomics* 290, 767–784. doi:10.1007/s00438-014-0949-7
- Juhász, Z., Fehér, T., Németh, E., Pamjav, H., 2016. mtDNA analysis of 174 Eurasian populations using a new iterative rank correlation method. *Mol. Genet. Genomics* 291, 493–509. doi:10.1007/s00438-015-1084-9
- Juric, I., Aeschbacher, S., Coop, G., 2016. The Strength of Selection against Neanderthal Introgression. *PLoS Genet.* 12, e1006340. doi:10.1371/journal.pgen.1006340
- Kalmár, T., Bachrati, C.Z., Marcsik, A., Raskó, I., 2000. A simple and efficient method for PCR amplifiable DNA extraction from ancient bones. *Nucleic Acids Res.* 28, E67. doi:10.1093/nar/28.12.e67
- Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 30, 3059–3066. doi:10.1093/nar/gkf436
- Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi:10.1093/molbev/mst010

- Ke, X., Taylor, M.S., Cardon, L.R., 2008. Singleton SNPs in the human genome and implications for genome-wide association studies. *Eur. J. Hum. Genet.* 16, 506–515. doi:10.1038/sj.ejhg.5201987
- Kelly, C., 2010. *The End of Empire: Attila the Hun & the Fall of Rome*. W.W. Norton & Company Ltd.
- Kézai, S., é. n. *Gesta Hunnorum et Hungarorum*.
- Kilinc, G.M., Omrak, A., Ozer, F., Gunther, T., Buyukkarakaya, A.M., Bicakci, E., Baird, D., Donertas, H.M., Ghalichi, A., Yaka, R., Koptekin, D., Acan, S.C., Parvizi, P., Krzewinska, M., Daskalaki, E.A., Yuncu, E., Dagtas, N.D., Fairbairn, A., Pearson, J., Mustafaoglu, G., Erdal, Y.S., Cakan, Y.G., Togan, I., Somel, M., Stora, J., Jakobsson, M., Gotherstrom, A., 2016. The Demographic Development of the First Farmers in Anatolia. *Curr. Biol.* 26, 2659–2666. doi:10.1016/j.cub.2016.07.057
- Kircher, M., 2012. Analysis of high-throughput ancient DNA sequencing data. *Methods Mol. Biol.* 840, 197–228. doi:10.1007/978-1-61779-516-9\_23
- Kircher, M., Sawyer, S., Meyer, M., 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.* 40. doi:10.1093/nar/gkr771
- Knapp, M., Clarke, A.C., Horsburgh, K.A., Matisoo-Smith, E.A., 2012. Setting the stage - Building and working in an ancient DNA laboratory. *Ann. Anat.* 194, 3–6. doi:10.1016/j.aanat.2011.03.008
- Knapp, M., Hofreiter, M., 2010. Next generation sequencing of ancient DNA: Requirements, strategies and perspectives. *Genes (Basel)*. doi:10.3390/genes1020227
- Kniezsa, I., 1993. A megtelepedett magyarság népi alkata, in: Domanovszky, S. (Szerk.), *Magyar Művelődéstörténet*. Babits Kiadó.
- Koch, R., 2007. *Attila und die Hunnen*. Historischen Museum der Pfalz Speyer, Speyer.
- Kousathanas, A., Leuenberger, C., Link, V., Sell, C., Burger, J., Wegmann, D., 2016. Inferring heterozygosity from ancient and low coverage genomes, bioRxiv. doi:10.1101/046748
- Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* msw054. doi:10.1093/molbev/msw054
- Lahermo, P., Savontaus, M.L., Sistonen, P., Béres, J., de Knijff, P., Aula, P., Sajantila, A., 1999. Y chromosomal polymorphisms reveal founding lineages in the Finns and the Saami. *Eur. J. Hum. Genet.* 7, 447–458. doi:10.1038/sj.ejhg.5200316
- Lalueza-Fox, C., Sampietro, M.L., Gilbert, M.T.P., Castri, L., Facchini, F., Pettener, D., Bertranpetit, J., 2004. Unravelling migrations in the steppe: mitochondrial DNA sequences from ancient central Asians. *Proc. Biol. Sci.* 271, 941–7. doi:10.1098/rspb.2004.2698
- Leigh, J.W., Bryant, D., 2015. popart: full-feature software for haplotype network construction. *Methods Ecol Evol* n/a. doi:10.1111/2041-210x.12410
- Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., Denisov, G., Lin, Y., MacDonald, J.R., Pang, A.W.C., Shago, M., Stockwell, T.B., Tsiamouri, A., Bafna, V., Bansal, V., Kravitz, S.A., Busam, D.A., Beeson, K.Y., McIntosh, T.C., Remington, K.A., Abril, J.F., Gill, J., Borman, J., Rogers, Y.H., Frazier, M.E., Scherer, S.W., Strausberg, R.L., Venter, J.C., 2007. The diploid genome sequence of an individual human. *PLoS Biol.* 5, 2113–2144. doi:10.1371/journal.pbio.0050254
- Lewontin, R., 1972. The Apportionment of Human Diversity. *Evol. Biol.* 6, 381–398. doi:10.1007/978-1-4684-9063-3\_14
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler

- transform. *Bioinformatics* 25, 1754–1760. doi:10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi:10.1093/bioinformatics/btp352
- Llamas, B., Valverde, G., Fehren-Schmitz, L., Weyrich, L.S., Cooper, A., Haak, W., 2017. From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. *STAR Sci. Technol. Archaeol. Res.* 3, 1–14. doi:10.1080/20548923.2016.1258824
- Llorente, M.G., Jones, E.R., Eriksson, A., Siska, V., Arthur, K.W., Arthur, J.W., Curtis, M.C., Stock, J.T., Coltorti, M., Pieruccini, P., Stretton, S., Brock, F., Higham, T., Park, Y., Hofreiter, M., Bradley, D.G., Bhak, J., Pinhasi, R., Manica, A., 2015. Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. *Science* (80-. ). 350, 820–2. doi:10.1126/science.aad2879
- Maciamo, H., 2016. Haplogroup U2 (mtDNA) [WWW Document]. URL [http://www.eupedia.com/europe/Haplogroup\\_U2\\_mtDNA.shtml](http://www.eupedia.com/europe/Haplogroup_U2_mtDNA.shtml)
- Malmström, H., Storå, J., Dalén, L., Holmlund, G., Götherström, A., 2005. Extensive human DNA contamination in extracts from ancient dog bones and teeth. *Mol. Biol. Evol.* 22, 2040–2047. doi:10.1093/molbev/msi195
- Maricic, T., Whitten, M., Pääbo, S., 2010. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. *PLoS One* 5. doi:10.1371/journal.pone.0014004
- Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10–12. doi:10.14806/ej.17.1.200
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi:10.1101/gr.107524.110
- Messier, W., Li, S.H., Stewart, C.B., 1996. The birth of microsatellites. *Nature*. doi:10.1038/381483a0
- Meyer, M., Fu, Q., Aximu-Petri, A., Glocke, I., Nickel, B., Arsuaga, J.-L.L., Martínez, I., Gracia, A., de Castro, J.M.B., Carbonell, E., Paabo, S., Martínez, I., Gracia, A., de Castro, J.M.B., Carbonell, E., Pääbo, S., 2014. A mitochondrial genome sequence of a hominin from Sima de los Huesos. *Nature* 505, 403–6. doi:10.1038/nature12788
- Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb. Protoc.* 5. doi:10.1101/pdb.prot5448
- Myres, N.M., Rootsi, S., Lin, A. a, Järve, M., King, R.J., Kutuev, I., Cabrera, V.M., Khusnutdinova, E.K., Pshenichnov, A., Yunusbayev, B., Balanovsky, O., Balanovska, E., Rudan, P., Baldovic, M., Herrera, R.J., Chiaroni, J., Di Cristofaro, J., VILLEMS, R., Kivisild, T., Underhill, P. a, 2011. A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. *Eur. J. Hum. Genet.* 19, 95–101. doi:10.1038/ejhg.2010.146
- Neparáczki, E., Juhász, Z., Pamjav, H., Fehér, T., Csányi, B., Zink, A., Maixner, F., Pálfi, G., Molnár, E., Pap, I., Kustár, Á., Révész, L., Raskó, I., Török, T., 2016. Genetic structure of the early Hungarian conquerors inferred from mtDNA haplotypes and Y-chromosome haplogroups in a small cemetery. *Mol. Genet. Genomics*. doi:10.1007/s00438-016-1267-z
- Neparáczki, E., Kocsy, K., Toth, G.E., Maroti, Z., Kalmar, T., Bihari, P., Nagy, I., Palfi, G., Molnar, E., Rasko, I., Torok, T., 2016. Revising mtDNA haplotypes of the ancient

- Hungarian conquerors with next generation sequencing. *bioRxiv*. doi:10.1101/092239
- Obrsuánszky, B., 2016. Európa ura, Attila. Barót.
- Oktatási Hivatal, 2017. A magyar nép eredete, származása [WWW Document]. A középkor története. URL <http://tudasbazis.sulinet.hu/hu/tarsadalomtudomanyok/tortenelem/a-kozepkor-tortenete-476-1492/a-magyar-nep-tortenete-a-honfoglalasisig/a-magyar-nep-eredete-es-vandorlasi>
- Orlando, L., Ginolhac, A., Raghavan, M., Vilstrup, J., Rasmussen, M., Magnussen, K., Steinmann, K.E., Kapranov, P., Thompson, J.F., Zazula, G., Froese, D., Moltke, I., Shapiro, B., Hofreiter, M., Al-Rasheid, K.A.S., Gilbert, M.T.P., Willerslev, E., 2011. True single-molecule DNA sequencing of a pleistocene horse bone. *Genome Res.* 21, 1705–1719. doi:10.1101/gr.122747.111
- Pääbo, S., 1985. Molecular cloning of Ancient Egyptian mummy DNA. *Nature* 314, 644–645. doi:10.1038/314644a0
- Palanichamy, M.G., Zhang, C.-L., Mitra, B., Malyarchuk, B., Derenko, M., Chaudhuri, T.K., Zhang, Y.-P., 2010. Mitochondrial haplogroup N1a phylogeography, with implication to the origin of European farmers. *BMC Evol. Biol.* 10, 304. doi:10.1186/1471-2148-10-304
- Pamjav, H., Juhász, Z., Zalán, A., Németh, E., Damdin, B., 2012. A comparative phylogenetic study of genetics and folk music. *Mol. Genet. Genomics* 287, 337–349. doi:10.1007/s00438-012-0683-y
- Pamjav, H., Z.Juhász, Fehér, T., Bárány, G., Németh, E., 2013. Classification of the Y-haplogroup distributions of Western Eurasian populations using a self-learning algorithm. *Forensic Sci. Int. Genet. Suppl. Ser.* 4. doi:10.1016/j.fsigss.2013.10.101
- Pinhasi, R., Fernandes, D., Sirak, K., Novak, M., Connell, S., Alpaslan-Roodenberg, S., Gerritsen, F., Moiseyev, V., Gromov, A., Raczky, P., Anders, A., Pietruszewsky, M., Rollefson, G., Jovanovic, M., Trinhhoang, H., Bar-Oz, G., Oxenham, M., Matsumura, H., Hofreiter, M., 2015. Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLoS One* 10, 1–13. doi:10.1371/journal.pone.0129102
- Quintáns, B., Álvarez-Iglesias, V., Salas, A., Phillips, C., Lareu, M. V., Carracedo, A., 2004. Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. *Forensic Sci. Int.* 140, 251–257. doi:10.1016/j.forsciint.2003.12.005
- Rasmussen, M., Li, Y., Lindgreen, S., Skou Pedersen, J., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., Bertalan, M., Nielsen, K., Thomas, M., Gilbert, P., Wang, Y., Raghavan, M., Campos, P.F., Kamp, H.M., Wilson, A.S., Gledhill, A., Tridico, S., Bunce, M., Lorenzen, E.D., Binladen, J., Guo, X., Zhao, J., Zhang, X., Zhang, H., Li, Z., Chen, M., Orlando, L., Kristiansen, K., Bak, M., Tommerup, N., Bendixen, C., Osipova, L.P., Higham, T.F.G., Bronk, C., 10, R., Hansen, T. v. O., Nielsen, F.C., Crawford, M.H., Brunak, S., Sicheritz-Pontén, T., Villems, R., Nielsen, R., Krogh, A., Wang, J., Willerslev, E., Pedersen, J.S., Albrechtsen, A., Moltke, I., Metspalu, M., Metspalu, E., Kivisild, T., Gupta, R., Bertalan, M., Nielsen, K., Gilbert, M.T.P., Wang, Y., Raghavan, M., Campos, P.F., Kamp, H.M., Wilson, A.S., Gledhill, A., Tridico, S., Bunce, M., Lorenzen, E.D., Binladen, J., Guo, X., Zhao, J., Zhang, X., Zhang, H., Li, Z., Chen, M., Orlando, L., Kristiansen, K., Bak, M., Tommerup, N., Bendixen, C., Pierre, T.L., Grønnow, B., Meldgaard, M., Andreasen, C., Fedorova, S.A., Osipova, L.P., Higham, T.F.G., Ramsey, C.B., Hansen, T. v. O., Nielsen, F.C., Crawford, M.H., Brunak, S., Sicheritz-Pontén, T., Villems, R., Nielsen, R., Krogh, A., Wang, J., Willerslev, E., 2010. Ancient human genome

- sequence of an extinct Palaeo-Eskimo. *Nature* 463, 757–762. doi:10.1038/nature08835
- Révész, L., 2016. A honfoglalás és államalapítás kori temetők tanulságai. *Rubicon*.
- Révész, L., 1996. A karosi honfoglalás kori temetők.
- Richards, M.B., Sykes, B.C., Hedges, R.E.M., 1995. Authenticating DNA Extracted From Ancient Skeletal Remains. *J. Archaeol. Sci.* 22, 291–299. doi:10.1006/jasc.1995.0031
- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26. doi:10.1038/nbt.1754
- Rohland, N., Harney, E., Mallick, S., Nordenfelt, S., Reich, D., 2015. Partial uracil-DNA-glycosylase treatment for screening of ancient DNA. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 370, 20130624. doi:10.1098/rstb.2013.0624
- Rohland, N., Hofreiter, M., 2007. Ancient DNA extraction from bones and teeth. *Nat. Protoc.* 2, 1756–1762. doi:10.1038/nprot.2007.247
- Róna-Tas, A., 1999. Hungarians And Europe In The Early Middle Ages: An Introduction to Early Hungarian History. CEU Press.
- Roostalu, U., Kutuev, I., Loogvali, E.L., Metspalu, E., Tambets, K., Reidla, M., Khusnutdinova, E.K., Usanga, E., Kivisild, T., Villems, R., 2007. Origin and expansion of haplogroup H, the dominant human mitochondrial DNA lineage in west Eurasia: The Near Eastern and Caucasian perspective. *Mol. Biol. Evol.* 24, 436–448. doi:10.1093/molbev/msl173
- Rootsi, S., Magri, C., Kivisild, T., Benuzzi, G., Help, H., Bermisheva, M., Kutuev, I., Barac, L., Pericic, M., Balanovsky, O., Pshenichnov, A., Dion, D., Grobei, M., Zhivotovsky, L.A., Battaglia, V., Achilli, A., Al-Zahery, N., Parik, J., King, R., Cinnioğlu, C., Khusnutdinova, E., Rudan, P., Balanovska, E., Scheffrahn, W., Simonescu, M., Brehm, A., Goncalves, R., Rosa, A., Moisan, J.-P., Chaventre, A., Ferak, V., Füredi, S., Oefner, P.J., Shen, P., Beckman, L., Mikerezi, I., Terzić, R., Primorac, D., Cambon-Thomsen, A., Krumina, A., Torroni, A., Underhill, P.A., Santachiara-Benerecetti, A.S., Villems, R., Semino, O., 2004. Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am. J. Hum. Genet.* 75, 128–37. doi:10.1086/422196
- Rootsi, S., Zhivotovsky, L. a, Baldovic, M., Kayser, M., Kutuev, I. a, Khusainova, R., Bermisheva, M. a, Gubina, M., Fedorova, S. a, Ilumäe, A.-M., Khusnutdinova, E.K., Voevoda, M.I., Osipova, L.P., Stoneking, M., Lin, A. a, Ferak, V., Parik, J., Kivisild, T., Underhill, P. a, Villems, R., 2007. A counter-clockwise northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. *Eur. J. Hum. Genet.* 15, 204–11. doi:10.1038/sj.ejhg.5201748
- Salamon, M., Tuross, N., Arensburg, B., Weiner, S., 2005. Relatively well preserved DNA is present in the crystal aggregates of fossil bones. *Proc. Natl. Acad. Sci. U. S. A.* 102, 13783–13788. doi:10.1073/pnas.0503718102
- Salas, A., Quintáns, B., Álvarez-iglesias, V., 2005. SNaPshot Typing of Mitochondrial DNA Coding Region Variants, in: *Forensic DNA Typing Protocols*. o. 197–208.
- Sampietro, M.L., Gilbert, M.T.P., Lao, O., Caramelli, D., Lari, M., Bertranpetit, J., Lalueza-Fox, C., 2006. Tracking down human contamination in ancient human teeth. *Mol. Biol. Evol.* 23, 1801–1807. doi:10.1093/molbev/msl047
- Sanchez, J.J., Phillips, C., Børsting, C., Balogh, K., Bogus, M., Fondevila, M., Harrison, C.D., Musgrave-Brown, E., Salas, A., Syndercombe-Court, D., Schneider, P.M., Carracedo, A., Morling, N., 2006. A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis* 27, 1713–1724. doi:10.1002/elps.200500671



- 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. doi:10.1371/journal.pone.0034131
- Sawyer, S., Renaud, G., Viola, B., Hublin, J.-J.J., Gansauge, M.M.-T., Shunkov, M. V., Derevianko, A.P., Prüfer, K., Kelso, J., Pääbo, S., 2015. Nuclear and mitochondrial DNA sequences from two Denisovan individuals. *Proc. Natl. Acad. Sci.* 112, 2–6. doi:10.1073/pnas.1519905112
- Seielstad, M.T., Minch, E., Cavalli-Sforza, L.L., 1998. Genetic evidence for a higher female migration rate in humans. *Nat. Genet.* 20, 278–280. doi:10.1038/3088
- Semino, O., Passarino, G., Quintana-Murci, L., Liu, A., Béres, J., Czeizel, A., Santachiara-Benerecetti, A.S., 2000. MtDNA and Y chromosome polymorphisms in Hungary: inferences from the palaeolithic, neolithic and Uralic influences on the modern Hungarian gene pool. *Eur. J. Hum. Genet. EJHG* 8, 339–46. doi:10.1038/sj.ejhg.5200468
- Shi, H., Qi, X., Zhong, H., Peng, Y., Zhang, X., Ma, R.Z., Su, B., 2013. Genetic Evidence of an East Asian Origin and Paleolithic Northward Migration of Y-chromosome Haplogroup N. *PLoS One* 8. doi:10.1371/journal.pone.0066102
- Skoglund, P., Malmström, H., Raghavan, M., Storå, J., Hall, P., Willerslev, E., Gilbert, M.T.P., Götherström, A., Jakobsson, M., Menozzi, P., Piazza, A., Cavalli-Sforza, L., Novembre, J., Ramachandran, S., Stoneking, M., Krause, J., Bramanti, B., Malmström, H., Haak, W., Lacan, M., Rasmussen, M., Malmström, H., Malmström, H., Sjögren, K.-G., Price, T., Ahlström, T., Lunter, G., Goodson, M., Briggs, A.W., Patterson, N., Price, A.L., Reich, D., Li, J.Z., Surakka, I., Alshuler, D.M., Nelson, M.R., Durbin, R.M., Alexander, D.H., Novembre, J., Lange, K., Jakobsson, M., Rosenberg, N.A., Salmela, E., Krause, J., Orlando, L., Reich, D., Thangaraj, K., Patterson, N., Price, A.L., Singh, L., Skoglund, P., Götherström, A., Jakobsson, M., Yang, D.Y., Eng, B., Waye, J.S., Dudar, J.C., Saunders, S.R., Li, H., Green, R.E., Green, R.E., Zhang, Z., Schwartz, S., Wagner, L., Miller, W., Kampmann, M.L., Poinar, H.N., Kloss-Brandstätter, A., Oven, M. van, Kayser, M., Andrews, R.M., Scheet, P., Stephens, M., Pickrell, J.K., Novembre, J., Levy, S., Wang, J., Ahn, S.M., Pushkarev, D., Neff, N.F., Quake, S.R., Schuster, S.C., Wang, C., Engelhardt, B.E., Stephens, M., Li, H., Durbin, R., Rosenberg, N., García-Garcera, M., Fogg, M., Pearl, L., Connolly, B., Edgar, R.C., Keinan, A., Mullikin, J.C., Patterson, N., Reich, D., Reich, D., Creevey, C.J., McInerney, J.O., Jakobsson, M., Moorjani, P., 2012. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. *Science* 336, 466–9. doi:10.1126/science.1216304
- The 1000 Genomes Project Consortium, 2015. A global reference for human genetic variation. *Nature*. doi:10.1038/nature15393
- The International HapMap Consortium, X., 2003. The International HapMap Project. *Nature* 426, 789–796. doi:10.1038/nature02168
- Thorvaldsdóttir, H., Robinson, J.T., Mesirov, J.P., 2013. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* 14, 178–192. doi:10.1093/bib/bbs017
- Thúry, J., 1897. Krónikáink és a nemzeti hagyomány. *Irod. Közlemények* 275–296.
- Tishkoff, S. a, Kidd, K.K., 2004. Implications of biogeography of human populations for „race” and medicine. *Nat. Genet.* 36, S21–S27. doi:10.1038/ng1438
- Tömöry, G., Csányi, B., Bogácsi-Szabó, E., Kalmár, T., Czibula, Á., Csoosz, A., Priskin, K., Mende, B., Langó, P., Downes, C.S., Raskó, I., 2007. Comparison of maternal lineage and

- biogeographic analyses of ancient and modern Hungarian populations. *Am. J. Phys. Anthropol.* 134, 354–368. doi:10.1002/ajpa.20677
- Valenzuela, R.K., Henderson, M.S., Walsh, M.H., Garrison, N.A., Kelch, J.T., Cohen-Barak, O., Erickson, D.T., John Meaney, F., Bruce Walsh, J., Cheng, K.C., Ito, S., Wakamatsu, K., Frudakis, T., Thomas, M., Brilliant, M.H., 2010. Predicting phenotype from genotype: Normal pigmentation. *J. Forensic Sci.* 55, 315–322. doi:10.1111/j.1556-4029.2009.01317.x
- Van Oven, M., Van Geystelen, A., Kayser, M., Decorte, R., Larmuseau, M.H., 2014. Seeing the wood for the trees: A minimal reference phylogeny for the human Y chromosome. *Hum. Mutat.* 35, 187–191. doi:10.1002/humu.22468
- Vianello, D., Sevini, F., Castellani, G., Lomartire, L., Capri, M., Franceschi, C., 2013. HAPLOFIND: A new method for high-throughput mtDNA haplogroup assignment. *Hum. Mutat.* 34, 1189–1194. doi:10.1002/humu.22356
- Völgyi, A., Zalán, A., Szvetnik, E., Pamjav, H., 2009. Hungarian population data for 11 Y-STR and 49 Y-SNP markers. *Forensic Sci. Int. Genet.* 3. doi:10.1016/j.fsigen.2008.04.006
- Wang, C.-C., Wang, L.-X., Shrestha, R., Wen, S., Zhang, M., Tong, X., Jin, L., Li, H., 2015. Convergence of Y Chromosome STR Haplotypes from Different SNP Haplogroups Compromises Accuracy of Haplogroup Prediction. *J. Genet. Genomics.* doi:10.1016/j.jgg.2015.03.008
- Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A., Schönherr, S., 2016. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. *Nucleic Acids Res.* 44, W58-63. doi:10.1093/nar/gkw233
- Wong, E.H.M., Khrunin, A., Nichols, L., Pushkarev, D., Khokhrin, D., Verbenko, D., Evgrafov, O., Knowles, J., Novembre, J., Limborska, S., Valouev, A., 2016. Reconstructing genetic history of Siberian and Northeastern European populations. *Genome Res.* 27, 1–14. doi:10.1101/gr.202945.115
- Wysoker, A., Tibbetts, K., Fennell, T., 2013. Picard tools version 1.90. <http://picard.sourceforge.net>.

## 8. Saját közlemények listája

MTMT azonosító: 10030445

A dolgozat témájához kapcsolódó közlemények:

*Genetic data imply Xiongnu origin of the Hungarian Conquerors, who were considerably admixed with Germans of Scandinavian origin*

**Neparáczkai E**, Maróti Z, Kalmár T, Bihari P, Nagy I, Pálfi G, Fóthi E, Maráczi L, Raskó I, Török T

Manuscript in preparation

*Revising mtDNA haplotypes of the ancient Hungarian conquerors with next generation sequencing.*

**Neparáczkai E**, Kocsy K, Tóth GE, Maróti Z, Kalmár T, Bihari P, Nagy I, Pálfi G, Molnár E, Raskó I, Török T

PLoS One. 2017

PMID: Bírálattal alatt , **IF: 4.411**

*Genetic structure of the early Hungarian conquerors inferred from mtDNA haplotypes and Y-chromosome haplogroups in a small cemetery.*

**Neparáczkai E**, Juhász Z, Pamjav H, Fehér T, Csányi B, Zink A, Maixner F, Pálfi G, Molnár E, Pap I, Kustár Á, Révész L, Raskó I, Török T

Molecular Genetics and Genomics. 1: 14 p. Paper 10.1007/s00438-016-1267-z. 14 p. 2016

PMID: 27803981, **IF: 2.622**

Egyéb közlemények:

*Preliminary results from the paleomicrobiological studies of Mycobacterium tuberculosis infection in the Bácsalmás-Óalmás anthropological series.*

**Neparáczkai E**, Török T, Pósa A, Molnár E, Lovász G, Maixner F, Zink At, Dutour O, Pálfi G  
Acta Biologica Szegediensis 55:(1) pp. 41-45. 2011

*Skeletal manifestation of tuberculosis in a late medieval anthropological series from Serbia.*

Lovász G, Pálfi G, Marcsik A, Pósa A, **Neparáczky E**, Molnár E

Acta Biologica Szegediensis 54:(2) pp. 83-91. 2010

## 9. Tartalmi összefoglaló

Kutatásaink célja az volt, hogy folytassuk a Szegeden az MTA SZBK Genetikai Intézet korábbi igazgatója, Raskó István és munkacsoportja által elkezdett munkát, archeogenetikai eszközökkel fényt derítsünk a honfoglalók származására. Munkám kezdetén a Szegedi Tudományegyetem Genetikai Tanszékén a Genetikai és az Embertani Tanszék együttműködésének keretében létrehoztunk egy archeogenetikai vizsgálatokra alkalmas speciális laboratóriumot.

A csekély anyagi lehetőségek miatt sok eszközt és bűtort saját kezűleg készítettük el, és épp csak a legszükségesebb műszerekkel tudtuk felszerelni. Kezdetben a legolcsóbban megvalósítható csontporítási módszert állítottuk be, majd sok nehézség árán megoldottuk a szennyeződésmentes DNS kivonást. Az áttörést a gamma sugárkezelt desztillált víz használata hozta, mert mint kiderült a legnagyobb tisztaságúnak mondott víz is tartalmaz nyomokban emberi DNS-t. A gamma kezelt víz használatával a PCR reakciókban a negatív kontrollok még 52 ciklus után is sem mutattak szennyező DNS-re utaló jelet.

A kezdeti nehézségek után sikeresen optimalizáltuk az ásatag DNS izolálás módszertanát és adaptáltuk legújabb haplotipizálási módszereket. Rutinra tettünk szert az ásatag csontokból izolált mitokondriális DNS haplotípusának meghatározásában, melynek során a szokásos HVR szakaszok szekvenálását kiegészítettük a kódoló szakaszok SNP haplotípus meghatározásával a legújabb SNaPshot technika alkalmazásával (GenoCoRe22). Ugyanezen módszerrel a jobb megtartású csontokból, meg tudtuk határozni az apai ág (Y-kromoszóma) haplocsoportját is (GenoY25).

Mikor már sikeresen és biztosan tudtunk ásatag DNS-t izolálni, valamint a haplotípus azonosítására is volt működő módszerünk, a Karos-III számú temető mintáit kezdtük el jellemezni. Az egész szakirodalomban alig fordul elő egy teljes temető genetikai jellemzése, a honfoglaláskorból pedig ez volt az első ilyen munka. Mivel a Karos-III. temető kis mintaszámú (n=19) ez viszonylag könnyen kivitelezhető feladat volt. A temető 17 leletéből tudtunk elemezhető DNS-t kinyerni, és jellemezni. Eredményeink azt mutatták, hogy a népesség jelentős része közép-ázsiai eredetű, de viszonylag sok európai eredetű egyént is találtunk. Tehát a Karos-III számú temető kifelbontású genetikai vizsgálati eredményei gyakorlatilag nem különböztek

az eddig megjelent, nagyszámú temetőből válogatott mintákból nyert haplocsoport gyakoriság értékektől (Neparáczki és mtsai., 2016).

A populáció-genetikai vizsgálatok eredményei elsősorban közép-ázsiai népekkel mutattak hasonlóságot. Viszont a populáció-genetikai vizsgálatok jogos kritikája, hogy mit tekintünk egy populációnak. Ha Árpád népe 7 törzsből állt, akkor az a 7 törzs egy populáció vagy 7 különböző populáció esetleg 3 egymáshoz hasonló és 4 egymáshoz hasonló, végső soron két populáció? Ezt a kérdést csak úgy lehet megválaszolni, ha először az "esetleges" populációt alkotó összes minta leszármazását egyenként megvizsgáljuk, majd azokat összegezve megnézzük, hogy vannak-e azonos helyről származó egyének. Ezt az egyéni szekvenciák filogenetikai törzsfákba illesztésével lehet megoldani, de az alacsony felbontású (HVR) szekvencia adatoktól nem várhatunk pontos eredményt. Viszont a kezdeti aDNS filogenetikai vizsgálatok szinte kizárólag a HVR szakaszok alapján történtek. Ez a megközelítés még recens mintákból is fenntartással kezelendő, de ásatag mintánál méginkább, mivel a PCR alapú szekvencia olvasatok minden erőfeszítés ellenére sem pontosak.

Az archeogenetikai kutatásokat az utóbbi néhány évben forradalmasította az újgenerációs szekvenálás (NGS) módszere. Ezért a laborban bevezettük az NGS metodikát, melynek első lépéseként adaptáltuk az NGS könyvtár készítés módszerét aDNS molekulákból, részleges UDG kezeléssel kiegészítve. Ezek a könyvtárak csak néhány % humán DNS-t tartalmaznak, ezért a szekvenálás előtt célszerű a humán DNS tartalmat emelni. Ennek megvalósításához saját készítésű mtDNS dúsító csalit állítottunk össze, melynek segítségével könyvtárakból sikerrel dúsítottuk a teljes mtDNS genomot, amit immár elérhető áron tudunk megszekvenálni. Emellett összeállítottunk egy közel 300 SNP-t tartalmazó genomi dúsító kitet is, mellyel még csak előkísérleteket végeztünk.

Megvizsgáltuk, hogy az újgenerációs szekvenálással nyert adatokkal megegyeznek-e a klasszikus PCR alapú módszerekkel nyert szekvencia olvasatokkal. Eredményeink azt mutatták, hogy az HVR szekvenálás megerősítése néhány kódoló régiós SNP-vel még a haplocsoport besorolás terén sem zárja ki a tévedés lehetőségét, ezért a PCR módszerrel kapott adatokat fenntartással kell kezelni. (Neparáczki és mtsai., 2016)

Miután meggyőződünk az NGS módszer hatékonyságáról nagy mintaszámra terjesztettük ki vizsgálatainkat. Ennek eredményeképp 89 honfoglalás kori minta mtDNS genom szekvenciáját határoztuk meg a három Karos-Eperjesszögi temető majdnem összes leletéből,

továbbá 6 másik honfoglalás kori temető néhány leletéből. Azt találtuk, hogy az egyes karosi temetőkön belül lehetnek anyai ágú rokonok, viszont a 3 temető között a vezérek kivételével nincs átfedő haplocsoport. Ezzel szemben néhány további temetőben találtunk a karosi egyénekkal azonos szekvenciákat mutató mintákat. Ez arra utal, hogy a törzseket megoszthatták, és azok töredékeit telepíthették egymás mellé.

A jó minőségű mtDNS genom szekvenciákkal filogenetikai-filogeográfiai vizsgálatot végeztünk oly módon, hogy a szekvenciákat beillesztettük az ismert filogenetikai törzsfákba, és leolvastuk a fa azonos ágaira térképeződő minták földrajzi eredetét. Ezt a módszert a szakterületen rutinszerűen alkalmazzák a populációk származási és vándorlási útvonalainak felderítésére. A nagyfelbontású filogeográfiai vizsgálatok aránylag pontosan kirajzolták az egyes minták anyai vonalainak valószínű származási helyét. Végül 89 mtDNS genom elemzésével megállapítottuk, hogy a honfoglalók lehetséges őseinek 41,5 %-a xiongnu (ázsiai hun), 42,7 %-a skandináv-germán, 6,7 %-a kaukázusi (Közél-Keleti), 2,2 %-a szláv és 6,7 %-a egyéb európai származású lehetett (16. ábra). Adatainkat egyéb genetikai, antropológiai történeti adatok is alátámasztják.

Eredményeink támogatják a középkori krónikáink hitelét, melyek azt állítják, hogy a „honfoglalás” a magyarok második bejövetele, és a honfoglalók itt már magyar nyelvű népeiséget találtak. Adataink hihetővé teszik az elismerten eredetileg is magyar nyelvű székelyek hun hagyományát, és a honfoglalást megelőző jelenlétüket a Kárpát-medencében. Ezzel magyarázatot nyer a régészeti anyagból levont következtetés is, miszerint a honfoglalás kori népesség régészeti hagyatékából becsült csekély létszáma alapján nem adhatta alapját a 11. századi Árpád kori tömb magyarságnak (Kniezsa, 1993; Révész, 2016).

A finnugor-honfoglaló genetikai kapcsolat hiánya arra utal, hogy a magyar nyelv finnugor rétegeit nem lehet a honfoglalókhoz kötni, ellenben a magyar nyelvben meglévő masszív török nyelvi réteget nagy valószínűséggel hun eredetűnek tekinthetjük. Ez újabb kérdéseket és kutatási irányokat nyit a magyar nyelv lehetséges eredetéről, és rokonsági viszonyairól.

Eredményeink fontos eddig nyitott kérdéseket válaszolnak meg, és ezzel számos szakterületen újabb kutatási irányokat nyitnak. A magyarok eredetének kérdése továbbra sem tekinthető lezártnak, hiszen genetikai adataink és a régészeti adatok alapján valószínű, hogy a hun eredetű népesség már a honfoglalás korában is kisebbségben lehetett a Kárpát-medencében. A

magyarok eredetének tisztázása további genetikai, régészeti, történelmi, és nyelvészeti vizsgálatokat igényel. Szűkebb szakterületünket érintő további kérdések megválaszolásához el kell készíteni a kora Árpád kor, az avarok, később pedig a még korábbi korok nagy felbontású genetikai térképét. Továbbá szükséges lenne létrehozni a Kárpát medence recens lakosságát tájegységek szerint reprezentáló genomi adatbázist. Jelen munkánk csak egy hosszú kutatási program első lépésnek tekinthető.



## 10. Summary

Hungarians are the only non-Indo-European-speaking nation in Central Europe and according to current academic theory the Hungarian language arrived into the Carpathian Basin with the Hungarian Conquerors, although the genetic origin of the Hungarian conquerors has been obscure. The Conquest period is one of the most intensively studied area of Hungarian archeogenetic research, thus far all archeogenetic research has focused almost exclusively on this period. In Szeged several outstanding PhD. dissertations had been written on this topic. However, in recent years considerable methodological-technological development happened on the archeogenetic field, facilitating an increase both in sequence quantity and quality never seen before. The application of Next Generation Sequencing (NGS) resulted in very precise genetic classifications enabling deep haplotype analyses, whereby the genetic origin of the conquerors could be determined even more exactly.

The aim of this study was to continue the archeogenetics research in Szeged, which was initiated by István Raskó's research group in 2000. We have applied the latest methods in the aDNA field, optimized the ancient DNA extraction methods according to the latest developments, and used the most up to date haplotyping methods.

First we constructed a sterile aDNA laboratory in the Department of Genetics, University of Szeged, which satisfies the requirements established in the literature. Due the limited financial possibilities we had to prepare many of the tools and furniture with own hands, and could install just the most necessary instruments. Initially, the cheapest bone pulverization method could be adapted. At the beginning we had a lot of difficulties with the elimination of modern DNA contaminations. Finally this problem could be solved by using gamma-irradiated distilled water in the reagents, as it turned out, that water was the main major source of contamination, as even the highest purity commercial water contains traces of human DNA. With the use of gamma-treated water in the PCR reactions, negative controls showed no signs of contaminating DNA even after 52 cycles.

In our first experiments, we adapted the most up-to-date PCR-based haplotyping method, the so called Snapshot assay. With this assay 22 mtDNA coding region SNP-s (GenoCore22) or 25 Y chromosome SNP-s (GenoY25 in parallel) can be determined at the same time. We also

optimized the multiplex primer concentrations in the GenoY25 reactions, to obtain more uniform signal intensity for each primers.

Whithin the methods in hands, next we set out to genetically characterize the small Karos-III cemetery. Only few papers report the genetic profile of entire cemeteries in the aDNA literature have not been The whole literature does not occur in the genetic characterization of an entire cemetery, and as the Karos-III cemetery has small number of samples ( $n = 19$ ), it was relatively easy to carry out this goal. We have successfully amplified mtDNA from 17 samples, and determined their maternal haplotypes. Our results showed that a significant part of the Karos population had Central Asian origin, but approximately half of the samples were of European origin. In summary low-resolution genetic results of the Karos-III cemetery had slightly higher Asian component, but did not not differ significantly from previously published Hungarian conqueror haplogroup distributions.

Population genetic analysis the Karos-III data showed similarities with modern populations, which have a similar mixture of Asian and European haplogroups, mainly people in Central Asia. Population genetic analysis can be criticized on the basis of what can be regarded as a population. For example if the people of Árpád consisted of 7 tribes can they be regarded as one population, or 7 different populations? This question can only be answered after identifying the exact origin of each individuals by high resolution phylogenetic analysis, and consider one population only individuals derived from the same location . However haplogrouping based on (HVR) sequence data does not have enough resolution to identify the exact origin of the samples. Nevertheless all initial phylogenetic analysis of aDNA data were exclusively HVR based, and all population genetic analysis was carried out with this limited information. The results of such analysis are further doubted by the fact that PCR-based sequence reads of aDNA samples are quite inaccurate, as we have shown (Neparaczki et al., 2016).

In the second half of our research, we implemented next-generation sequencing, in order to obtain high quality reliable full length mtDNA sequences. By using the entire mtDNA genome in phylogenetic analysis, the resolution increases with an order of magnitude, (16 kb versus 200 bp), therefore these sequences are expected to give much more reliable results. We have sequenced 89 conqueror mtDNA genomes using NGS, and performed the phylogenetic analysis of each samples independently, by comparing their mtDNA genomes to all available genomes from databases. We found that some 40% of the conquerors had East Asian origin, where the

geographic origin of the best matching samples nicely corresponded to the region of the ancient Xiongnu empire. Other 40% of the samples had best matches with modern people from Scandinavia, Germany, or other regions in Western Europe. A smaller third group of the samples (6.7%) matched modern samples from the Caucasus region and the Near East. For a small proportion of samples, the exact origin could not be determined, but these samples were also of East or West European origin.

Our data implies that Hungarian conquerors assembled from three major sources before the conquest, corresponding to 3 major distinct populations. It follows that our initial population genetic analysis which considered the conquerors as a single population was not justified. Based on genetic and historical data half of the conqueror population had probably Xiongnu origin, corroborating the statement of medieval Hungarian chronicles, which all declare Hunnic origin of the Hungarians. The conquerors with Scandinavian-German genetic affinity had most probably Ostrogothic origin, as this group was reported to have been integrated into the European Hun Empire hundreds of years before the conquest. Interestingly this European component also support the Hun affinity of the Hungarian conquerors. Our data do not support the Finno-Ugric origin of the conquerors, therefore historical linguistic arguments will have to be reconsidered. The lack of Finno-Ugric genetic rather raises the possibility, that the language connection can also be indirect, which may have happened very long time (thousands of years) ago. Interestingly this conclusion is also supported by linguistic calculations, which estimate 2-5 thousand years between the divergence of Hungarian from its relative Finno-Ugric languages. In addition all linguistic analyses report a massive Turkic linguistic layer in Hungarian, which can now be directly linked to our Hun ancestors.

Our results restore the credibility of the medieval Hungarian chronicles, and necessitates reconsideration of much of current historical inferences. For example our results make credible the Hun tradition of the Hungarian speaking Seklers, and their presence in the pre-conquest Carpathian Basin. It follows that the Hungarian conquerors probably found Hungarian speaking population in the Carpathian Basin, as also reported in our chronicles. Finally, our results can also provide an explanation to some of the archaeological obscurities, for example archaeologist complain that the apparently low number of estimated conquest population could not provide the basis of the massive Hungarian speaking population of the Árpád-Age (11th century).

Finally our results do not answer the origin of modern Hungarians, and leave open quite a few questions. As the conquerors provided just one major component in the Hungarian ethnogenesis, the other components must be looked for in the numerous pre-conquest populations of the Carpathian basin. To further clarify Hungarian ethnogenesis high resolution genetic data are required from these populations, (Avars, Sarmatas, etc.) as well as from later Árpád-Age populations. It would also be necessary to create a representative recent Hungarian database possibly from full genome sequences, as a reference for comparison of archaic genomes. This requires tremendous further genetic research, and a full assembly of the archeogenetic map from different periods of the Carpathian basin. Our work can be considered as first step towards this goal.

## 11. Rövidítések jegyzéke

aDNS	ásatag DNS
Alu inszerció	az ember leggyakoribb transzpozonja
BSA	Bovine Serum Albumin
CHgC	Correlating Haplogroup Cluster
CNV	copy-number variant
DNS	dezoxiribonukleinsav, az örökítő anyag
dNTP	Deoxynucleotide (dATP, dCTP, dGTP és dTTP)
DTT	Dithiothreitol
EDTA	Etilén-diamin-tetraecetsav
ExoSAP	Exonukláz-I és FastAP keveréké, PCR termék tisztítására alkalmas
GenoCoRe22	mitokondrium 22 kódoló régiós SNP-jére összeállított assay
GenoY25	Y-kromoszóma 25 SNP-jének vizsgálatára összeállított assay
GuHCl	Guanidine hydrochloride
GuSCN	Guanidine thiocyanate
GWAS	genome-wide asocieted studies
HL-dsDNase	Heat-Labile Double-Strand Specific DNase
HTS	High-Throughput Sequencing
HVR-I,-II	Hypervariable region-I,-II,
IRC	Iterative Rank Correlation
MDS	Multidimensional scaling
M-J Network	Median-Joining Network
mtDNS	mitokondrium DNS
NaOAc	Nátrium-acetát
NGS	Next-Generation Sequencing
np.	nucleotide position
NRPY	Non-Recombinig Portion Y
NRV	Non-Recombining region of Y
NUMT	nuclear mitochondrial DNA segment
rCRS	revised Cambridge Reference Sequence
RFLP	Restriktációs fragmenthossz polimorfizmus
RSA	Rabbit Serum Albumin
SBE	Single base extension
SNaPshot	SBE rekación alapulszik, ezzel több SNP-t lehet egy reakcióban vizsgálni
SNP	single nukleotid polymorphism
SOC	Self Organising Cloud
STR	Short Tandem Repeat
SV	strukturális variáns
SVA	SINE-VNTR-Alu elemek rövidítése
UDG	Uracil-DNA Glycosylase
VCF	variant call file

## 12. Függelék

lókuszt	név (Forward)	szekvencia (5'-3')	név (Reverse)	szekvencia (5'-3')
HVS-I	L15996	CTCCACCATTAGCACCCAAAGC	H16142	ATGTACTACAGGTGGTCAAG
HVS-I	L16117	TACATTACTGCCAGCCACCAT	H16233	GCTTTGGAGTTGCAGTTGATGTGT
HVS-I	L16209	CCCCATGCTTACAAGCAAGT	H16348	ATGGGGACGAGAAGGGATTTG
HVS-I	L16287	CACTAGGATACCAACAAACC	H16410	GCGGGATATTGATTTACCGG
HVS-II	L00172	ATCCTATTATTTATCGCACCTACG	H00327	TTGGCAGAGATGTGTTAAAGTGCT

**Függelék 1. táblázat** HVR szekvenálásokhoz felhasznált primerek

μM	Név	Szekvencia 5'-3'	Név	Szekvencia 5'-3'	Amplikon méret (bp)	Haplocop.	KR-SNP
0,022	L02727	AACACAGCAAGACGAGAAGACC	H02760	GGACCTGTGGGTTTGTAGGT	75	L2'3'4'6'7	2758
0,088	L03585	CCCCTCCCATAACCAAC	H03601	GGCTAGAATAAATAGGAGGCCTAGGTT	60	L	3594
0,051	L04237	TGATATGTCTCCATACCCATTACAA	H04253	CTTTATCAGACATATTTCTTAGGTTGAG	70	A	4248
0,051	L04578	TTACCTGAGTAGGCTAGAAATAAACA	H04618	GCAGCTTCTGTGGAACGAG	85	V,M3	4580
0,015	L05171	ACCCTACTACTATCTCGCACCTGA	H05204	CTAGGGAGAGGAGGTTGGAT	76	D	5178
0,011	L06363	ACCATCTTCTCCTACACCTAGCAG	H06378	GATGAAATTGATGGCCCCTAA	60	X	6371
0,051	L07003	GCAAACCTCATCACTAGACATCGTACT	H07029	CCTATTGATAGGCATAGTGGAAATG	77	H	7028
0,015	L08970	CATACTAGTATTATCGAAACCATCAGC	H09003	CTGCAGTAATGTTAGCGGTTAGG	83	W	8994
0,059	L10025	CTTTTAGTATAAATAGTACCGTTAACTTCAA	H10037	AAGTTATTACTCTTTTGAATGTTGCA	73	I	10034
0,059	L10228	TCCCTTCTCCATAAAAATCTTCTT	H10249	AGGAGGGCAATTTCTAGATCAAATA	70	N1	10238
0,037	L10382	AAGTCTGGCCTATGAGTGAACATA	H10421	TGAGTCGAAATCATTCTGTTTG	85	M	10400
0,029	L10548	GAATACTAGTATATCGCTCACACCTCA	H10558	GCGATAGTATTATCTTCTAGGCATAGTA	66	K	10550
0,051	L10870	CCACAGCCTAATTATTAGCATCATC	H10888	GCTAAATAGGTTGTTGATTTGG	67	N	10873
0,015	L11454	ATCGCTGGGTCATAGTACTTGC	H11479	TGAGTGTGAGGGTATTATACCATAG	73	U	11467
0,037	L11710	GGCGCAGTATTCTCATAATC	H11735	AGTTTGAGTTGCTAGCAGAAATAG	70	preHV	11719
0,037	L12611	CTACTTCTCCATAAATTCATCCCTGT	H12621	AATTCTATGATGGACCATGTAACG	60	J	12612
0,037	L12689	CAGACCCAACATTAATCAGTCTT	H12715	TGTTAGCGTAACTAAGATTAGTATGGT	78	R	12705
0,051	L13258	ATCGTAGCCTTCTCCACTTCAA	H13295	AGGAATGCTAGGTGTTGGT	80	C	13263
0,022	L13350	CACGCCTTCTCAAAGCCATA	H13372	GTTCATTGTTAAGGTTGTTGATGAT	67	T	13368
0,051	L14759	AGAACACCAATGACCCCAATAC	H14799	GGTGGGAGGTCGATGA	78	HV	14766
0,037	L13923	TTTCTCCAACATACTCGGATTCTAC	H13942	AGAAAGGCTAGATAGGGGATTGT	66	R9	13928
0,037	L08268	AATAGGGCCGTATTACCCATA	H08295	AGGTAATGCTAAGTTAGCTTTACAGTG	78	B	8280del



	mtDNA Haplosop.	Hg	13928_R9	3594_L3 <sup>4</sup>	10550_K	11467_U	4248_A	8994_W	13263_C	13368_T	11719_D <sup>2</sup> eIV_R0	8280a <sub>1</sub> _B	4580_Y	6371_X	10238_N <sub>1</sub>	7028_H	10034_L	5178_D	14766_H <sub>V</sub>	12705_R	10873_N	12612_J	2758_L2 <sup>6</sup>	10400_M			
Extract	rCRS (ős)	H	c	g	t	A	T	G	t	G	c	A	c	g	r	T	r	T	g	g	g	a	A	G	r		
	eltérés		g/t	a	e	G	C	A	e	A	t	G	t	a	C	a	C	t	a	a	a	g	G	A	a		
	1.		C	G	T	A	T	G	T	G	T	G	C	G	T	A	T	G	A	g/a	A	A	A	G	G		
2.	1.		C	G	T	A	T	G	T	G	T	G	C	G	T	A	T	G	A	g/a	A	A	A	G	G		
3.	1.		C	G	T	A	T	G	T	G	T	G	C	G	T	g/A	T	G	A	g/a	A	A	A	G	G		
	konszenzus 1.	B	C	G	T	A	T	G	T	G	T	G*	C	G	T	A	T	G	A	G	A	A	A	G	G		
1.	3.		C	G	T	A	T	G	T	G	T	G	C	G	T	g	T	G	G	G	A	A	A	G	G		
2.	3.		C	G	T	A	T	G	T	G	T	G	C	G	T	g	T	G	G	G	A	A	A	G	G		
3.	3.		C	G	T	A	T	G	T	G	T	G	C	G	T	g	T	G	G	G	A	A	A	G	G		
	konszenzus 3.	H	C	G	T	A	T	G	T	G	T	G	C	A*	C	G	T	G	G	G	A	A	A	G	G		
1.	4.		C	G	T	A	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G/a	G	G		
2.	4.		C	G	T	A	T	G	T	G	T	A	C	G	T	g/A	T	G	A	G	A	A	G	G	G		
	konszenzus 4.	J	C	G	T	A	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	G	G	G		
1.	5.		C	G	T	A	T	G	T	G	T	G	C	G	T	A	T	G	A	G	A	A	A	G	G		
2.	5.		C	G	T	A	T	G	T	G	T	G	A	C	T	A	T	G	G/a	G/a	A	A	g/a	G	G		
3.	5.		C	G	T	A	T	G	T	G	c/T	G/a	C	G	T	g/A	T	G	g/a	G	A	A	A	G	G		
	konszenzus 5.	B	C	G	T	A	T	G	T	G	T	G*	C	G	T	A	T	G	A	G	A	A	A	G	G		
1.	6.		C	G	T	A	T	G	T	G	T	G	C	G	T	A	T	G	A	G	A	A	A	G	G		
2.	6.		C	G	T	A	T	G	T	G	t/c	G/a	C	G	T	G	T	G	a/g	g/a	A	A	A	G	G		
3.	6.		C	G	T	A	T	G	T	G	T	G/a	C	G	T	g/A	T	G	g/A	G	A	A	A	G	G		
3.	6.		C	G	T	A	T	G	T	G	T	G	C	G	T	A	T	G	g/A	G	A	A	A	G	G		
	konszenzus 6.	B	C	G	T	A	T	G	T	G	T	G*	C	G	T	A	T	G	A	G	A	A	A	G	G		
1.	8.		C	G	T	A	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	A	G	G		
2.	8.		C	G	T	A	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	A	G	G		
3.	8.		C	G	T	A	T	G	T	G	T	A	C	G	T	G	T	G	a/g	G	A	A	A	G	G		
	konszenzus 8.	U	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	A	G	G		
1.	9.		C	G	T	A	T	G	T	G	T	A	C	G	T	A	T	g/T	A	G/a	G/a	A	A	G	G/a		
2.	9.		C	G	T	G/a	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	A	G	G		
3.	9.		C	G	T	G/a	T	G	T	G	T	g/a	C	G	T	A	T	G	A	G/a	g/A	A	A	G	G/a		
	konszenzus 9.	U	C	G	T	G	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	A	G	G		
1.	10.		C	G	T	A	T	G	T	G	C/t	A	C	G	T	G	T	G/a	G/a	G/a	A	A	A	G	G/a		
2.	10.		C	G	T	A	T	G	T	G	C/t	A	C	G	T	G	T	G/t	G/a	G	A	A	a/g	G	G/a		
3.	10.		C	G	T	A	T	G	T	G	c/t	A	C	G	T	g/a	T	G/t	g/a	G/a	A	A	A	G	G/a		
	konszenzus 10.	H	C	G	T	A	T	G	T	G	C	A	C	G	T	G	T	G	G	G	A	A	A	G	G		
1.	11.		C	G	T	A	T	G	T	G	T	g/a	C		T	A	T	G	A	a/g	A	A	g/a	G	G		
2.	11.		C	G	T	A	T	G	T	G	T	A	C	A	T	a/g	T	G	A	A	A	A	A	G	G		
3.	11.		C	G	T	g/a	T	G	T	G	T	A	C	g/a	T	a/g	T	G	a/g	A	a/g	g/A	A	A	G	G	
	konszenzus 11	X	C	G	T	A	T	G	T	G	T	A*	C	A	T	A	T	G	A	A	A	A	A	A	G	G	
1.	12.		C	G	T	A	C	G	T	G	T	A	C	G	T	a/g	T	G	a/g	g/a	A	A	g/A	A	G	G	
2.	12.		C	G	T	A	C	G	T	G	T	A	C	G	T	A	T	G	A	A	A	A	A	A	G	G	
	konszenzus 12	A	C	G	T	A	C	G	T	G	T	A	C	G	T	A	T	G	A	A	A	A	A	A	G	G	
1.	13.		C	G	T	A	T	G	T	T	g/a	C	G	T	A	T	G	A	G	A	A	A	A	A	G	G	
2.	13.		C	G	T	A	T	G	T	G	T	g/a	C	G	T	g/a	T	G	A	G	A	A	g/a	A	G	G	
	konszenzus 13	B	C	G	T	A	T	G	T	G	T	G*	C	G	T	A	T	G	A	G	A	A	A	A	G	G	
1.	14.		C	G	T	A	T	G	T		c/T	A	C	G	T	A	T	G	a/g	g/A	A	A	g/a	A	G	G	
2.	14.		C	G	T	A	T	G	T	A	T	A	C	G	T	A	T	G	A	G	A	A	A	A	G	G	
	konszenzus 14	T	C	G	T	A	T	G	T	A	T	A	C	G	T	A	T	G	A	G	A	A	A	A	G	G	
1.	15.		C	G	T	A	T	G	T	G	T	A	C	G	T	A				A	G	A	A	A	G	G	
2.	15.		C	G	T	A	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	A	A	G	G	
	consensus 15	J	C	G	T	A	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	A	A	G	G	
1.	16.		C	G	T	A	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	A	A	A	G	G
2.	16.		C	G	T	A	T	G	T	G	T	A	C	G	T	a/g	T	G	A	G	A	A	A	A	A	G	G
	konszenzus 16.	U	C	G	T	A	T	G	T	G	T	A	C	G	T	A	T	G	A	G	A	A	A	A	A	G	G
1.	17.		C	G	T	A	T	G	T	G	C	A	C	G	T	G	T	G	G	G	G	A	A	A	A	G	G
2.	17.		C	G	T	A	T	G	T	G	C	A	C	G	T	G	T	G	G	G	G	A	A	A	A	G	G
	konszenzus 17.	H	C	G	T	A	T	G	T	G	C	A	C	G	T	G	T	G	G	G	G	A	A	A	A	G	G
1.	18.		C	G	T	A	T	G	T	A	g/T	A	C	G	T	A	T	G	A	G	A	A	A	A	A	G	G
2.	18.		C	G	T	A	T	G	T	A	T	A	C	G	T	A	T	G	A	G	A	A	A	A	A	G	G
3.	18.		C	G	T	A	T	G	T	A	T	A	C	G	T	A	T	G	A	a/g	A	A	A	A	A	G	G
	konszenzus 18.	T	C	G	T	A	T	G	T	A	T	A	C	G	T	A	T	G	A	G	A	A	A	A	A	G	G
1.	19.		C	G	T	A	T	G	T	A	T	A	C	G	T	g/A	T	G	A	G	A	A	g/A	A	A	G	G
2.	19.		C	G	T	A	T	G	T	A	T	A	C	G	T	A	T	G	A	G	A	A	A	A	A	G	G
	konszenzus 19	T	C	G	T	A	T	G	T	A	T	A	C	G	T	A	T	G	A	G	A	A	A	A	A	G	G
	Saját haploportom	H	C	G	T	A	T	G	T	G	C	A	C	G	T	G	T	G	G	G	G	A	A	A	G	G	

Függelék 3. táblázat A GenoCore22 SNP tipizálás eredményei.







Population	n	Code	Region	Language/Period	L	M*	C	Z	G	D	N*	I	W	Y	A	S	X	R*	
Tuscany	322	TUS	Italy	IE, Romance	1,9%	0,6%	0,0%	0,0%	0,0%	0,3%	1,2%	0,9%	2,2%	0,0%	0,0%	0,0%	0,0%	2,5%	2,8%
Cyprus (Greeks)	79	CYP	Balkans	IE, Greek	3,8%	2,5%	0,0%	0,0%	0,0%	0,0%	0,0%	3,8%	3,8%	0,0%	0,0%	0,0%	3,8%	2,5%	
French	871	FRA	Western Europe	IE, Romance	0,9%	0,1%	0,0%	0,0%	0,0%	0,0%	1,3%	1,8%	1,1%	0,0%	0,0%	0,0%	1,1%	1,3%	
Greek	373	GRE	Balkans	IE, Greek	0,0%	0,8%	0,3%	0,0%	0,0%	0,0%	0,8%	2,4%	2,4%	0,0%	0,0%	0,0%	3,5%	2,4%	
Iraqi	51	IRQ	Middle East	AA, Semitic	0,0%	2,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	9,8%	
Jordan	290	JOR	Middle East	AA, Semitic	10,0%	1,4%	0,0%	0,0%	0,0%	0,3%	5,9%	1,4%	1,4%	0,0%	0,3%	0,0%	1,7%	3,8%	
Lebanese	980	LBN	Middle East	AA, Semitic	2,0%	1,4%	0,0%	0,0%	0,0%	0,2%	4,5%	2,1%	1,6%	0,0%	0,4%	0,0%	1,8%	6,3%	
Palestinian	120	PAL	Middle East	AA, Semitic	7,5%	3,3%	0,0%	0,0%	0,0%	0,0%	2,5%	0,8%	0,8%	0,0%	0,0%	0,0%	4,2%	5,8%	
Saudi Arabian	539	SAU	Middle East	AA, Semitic	10,0%	6,1%	0,0%	0,0%	0,2%	0,0%	7,4%	0,9%	1,1%	0,0%	0,0%	0,0%	2,8%	18,4%	
Syrian	234	SYR	Middle East	AA, Semitic	6,0%	0,9%	0,0%	0,0%	0,0%	0,0%	4,7%	0,9%	3,4%	0,0%	0,4%	0,0%	1,7%	5,1%	
Yemeni	300	YEM	Middle East	AA, Semitic	38,3%	5,3%	0,0%	0,0%	0,0%	0,0%	4,7%	0,0%	0,7%	0,0%	0,0%	0,0%	1,3%	12,0%	
India (Dravidian)	265	IND	South Asia	Dravidian	0,0%	64,5%	0,0%	0,0%	0,4%	0,0%	0,0%	0,0%	0,4%	0,0%	0,0%	0,0%	0,0%	19,8%	
India (Hindi)	400	INH	South Asia	IE, Indo-Iranian	0,0%	52,7%	0,7%	0,0%	0,0%	1,0%	0,0%	1,3%	3,7%	0,0%	0,3%	0,0%	0,7%	16,7%	
Bashkir	221	BAS	Volga-Ural	Altaic, Turkic	0,0%	1,4%	11,8%	0,9%	4,5%	9,0%	5,0%	1,4%	0,5%	0,5%	3,6%	0,0%	0,0%	0,0%	
Chuvash	55	CHU	Volga-Ural	Altaic, Turkic	0,0%	1,8%	1,8%	0,0%	0,0%	3,6%	1,8%	1,8%	0,0%	0,0%	1,8%	0,0%	0,0%	0,0%	
Komi Permyak	74	KOP	Volga-Ural	Finno-Permic	0,0%	0,0%	8,1%	0,0%	2,7%	5,4%	9,5%	2,7%	0,0%	0,0%	1,4%	0,0%	0,0%	2,7%	
Komi Zyryan	62	KOZ	Volga-Ural	Finno-Permic	0,0%	0,0%	0,0%	1,6%	1,6%	0,0%	0,0%	1,6%	0,0%	1,6%	0,0%	0,0%	0,0%	0,0%	
Mari	136	MRI	Volga-Ural	Finno-Permic	0,0%	0,7%	0,7%	2,9%	0,0%	1,5%	0,0%	0,7%	0,0%	0,0%	1,5%	0,0%	0,0%	0,0%	
Mordvin	102	MRD	Volga-Ural	Finno-Permic	0,0%	0,0%	2,0%	0,0%	0,0%	1,0%	0,0%	5,9%	0,0%	0,0%	0,0%	0,0%	0,0%	1,0%	
Tatar	228	TAT	Volga-Ural	Altaic, Turkic	0,0%	2,2%	1,8%	0,4%	1,8%	2,6%	3,1%	0,9%	1,8%	0,0%	3,1%	0,0%	0,0%	0,4%	
Udmurt	101	UDM	Volga-Ural	Finno-Permic	0,0%	0,0%	3,0%	5,0%	0,0%	11,9%	0,0%	0,0%	0,0%	0,0%	1,0%	0,0%	0,0%	6,9%	
Albanian	42	ALB	Balkans	IE, Albanian	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	9,5%	0,0%	2,4%	0,0%	0,0%	0,0%	0,0%	0,0%	
Aromuns	175	ARO	Balkans	IE, Romance	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,6%	1,1%	0,0%	0,0%	0,0%	0,0%	2,3%	1,1%	
Macedonian	37	MAC	Balkans	IE, Balto-Slavic	0,0%	0,0%	0,0%	0,0%	0,0%	2,7%	0,0%	2,7%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	
Romanians	105	ROM	Balkans	IE, Romance	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	1,9%	1,0%	1,0%	0,0%	0,0%	0,0%	4,8%	1,9%	
Konda Mansi	59	KMS	West Siberia	Ugric	0,0%	1,7%	15,2%	0,0%	6,8%	6,8%	0,0%	0,0%	0,0%	0,0%	5,1%	0,0%	0,0%	0,0%	
Altai-Kizhi	90	ALT	South Siberia	Altaic, Turkic	0,0%	7,7%	31,9%	2,2%	5,5%	8,9%	0,0%	1,1%	0,0%	0,0%	3,3%	0,0%	4,4%	1,1%	
Buryats	295	BUR	Transbaikal	Altaic, Mongolic	0,0%	5,0%	16,6%	1,4%	11,3%	34,8%	2,4%	0,3%	0,0%	1,4%	5,0%	0,0%	0,3%	0,3%	
East Evenks	45	EVE	Transbaikal	Altaic, Tungusic	0,0%	0,0%	62,0%	0,0%	2,2%	24,1%	0,0%	0,0%	0,0%	8,9%	2,2%	0,0%	0,0%	0,0%	
Kalmyks	110	KMY	North Caucasus	Altaic, Mongolic	0,0%	6,3%	10,9%	1,8%	8,2%	29,3%	4,5%	0,9%	0,0%	1,8%	3,6%	0,0%	0,0%	0,9%	
Khakass	57	KKS	South Siberia	Altaic, Turkic	0,0%	3,5%	19,3%	0,0%	1,8%	16,0%	0,0%	0,0%	0,0%	0,0%	3,5%	0,0%	0,0%	0,0%	
Khamnigans	99	KMN	Transbaikal	Altaic, Mongolic	0,0%	5,0%	16,2%	0,0%	10,1%	33,1%	1,0%	0,0%	0,0%	3,0%	5,0%	0,0%	0,0%	1,0%	
Koreans	103	KOR	East Asia	Korean	0,0%	16,5%	1,0%	0,0%	6,7%	39,8%	2,9%	0,0%	0,0%	1,0%	6,8%	0,0%	0,0%	0,0%	
Mongolians	47	MNG	Inner Asia	Altaic, Mongolic	0,0%	12,7%	17,0%	2,1%	10,6%	11,0%	4,2%	0,0%	0,0%	0,0%	13,0%	0,0%	0,0%	0,0%	
Shors	82	SHO	South Siberia	Altaic, Turkic	0,0%	3,7%	12,1%	1,2%	0,0%	12,2%	0,0%	2,4%	0,0%	0,0%	1,2%	0,0%	0,0%	0,0%	
Telengits	71	TLG	South Siberia	Altaic, Turkic	0,0%	4,2%	16,6%	0,0%	2,8%	21,0%	2,8%	1,4%	0,0%	0,0%	5,6%	0,0%	0,0%	4,2%	
Teleuts	53	TLT	South Siberia	Altaic, Turkic	0,0%	3,8%	28,2%	5,7%	0,0%	24,9%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	1,9%	0,0%	
Tuvans	105	TUV	South Siberia	Altaic, Turkic	0,0%	1,0%	50,0%	1,0%	6,7%	15,4%	1,9%	2,9%	0,0%	0,0%	1,0%	0,0%	0,0%	0,0%	
West Evenks	73	EVW	North Siberia	Altaic, Tungusic	0,0%	0,0%	48,4%	1,4%	4,1%	30,3%	0,0%	0,0%	0,0%	2,7%	4,1%	0,0%	0,0%	0,0%	
Yakuts	86	YAK	North Siberia	Altaic, Turkic	0,0%	0,0%	64,0%	0,0%	2,8%	22,4%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	
Persians	32	PER	Middle East	IE, Indo-Iranian	1,2%	9,7%	1,2%	0,0%	1,2%	2,4%	1,2%	3,7%	2,4%	0,0%	2,4%	0,0%	0,0%	3,6%	
Altai Kazakhs	98	ALK	South Siberia	Altaic, Turkic	0,0%	6,1%	8,2%	2,0%	2,0%	26,5%	2,0%	3,1%	0,0%	1,0%	3,1%	0,0%	0,0%	1,0%	
Barghuts	149	BAR	Inner Asia	Altaic, Mongolic	0,0%	4,0%	20,2%	2,7%	9,4%	35,5%	1,3%	0,0%	0,0%	0,7%	6,1%	0,0%	0,0%	0,7%	
Hazara	78	HAZ	Greater Iran	IE, Indo-Iranian	0,0%	20,5%	9,0%	3,8%	6,4%	5,2%	6,4%	0,0%	0,0%	0,0%	1,3%	0,0%	0,0%	0,0%	
Pashtos/Afghan	90	AFJ	Greater Iran	IE, Indo-Iranian	0,0%	13,3%	3,3%	3,3%	2,2%	1,1%	3,3%	0,0%	1,1%	0,0%	0,0%	0,0%	2,2%	11,1%	
Tajik	146	TAJ	Greater Iran	IE, Indo-Iranian	0,0%	13,0%	2,8%	0,0%	2,1%	4,1%	1,4%	1,4%	6,2%	1,4%	0,7%	0,0%	1,4%	1,4%	
Turkmen	75	TRM	Central Asia	Altaic, Turkic	0,0%	5,3%	5,4%	1,3%	5,3%	6,6%	5,3%	5,3%	2,7%	0,0%	0,0%	0,0%	0,0%	4,0%	
Uzbek	127	UZB	Central Asia	Altaic, Turkic	0,0%	17,2%	1,6%	1,6%	3,1%	4,8%	3,9%	0,8%	3,9%	0,0%	2,4%	0,0%	0,8%	3,1%	
Hungarian (Budapest)	210	HUB	Central Europe	Ugric	1,4%	1,0%	0,0%	0,0%	0,0%	1,0%	0,5%	2,9%	2,4%	0,0%	0,0%	0,0%	1,4%	0,5%	
English	262	ENG	British Isles	IE, Germanic	0,4%	0,4%	0,0%	0,0%	0,0%	0,0%	0,0%	3,8%	0,0%	0,0%	0,0%	0,0%	1,5%	0,0%	
Finnish	50	FIN	Scandinavia	Finno-Permic	0,0%	4,0%	0,0%	0,0%	0,0%	0,0%	0,0%	10,0%	10,0%	0,0%	0,0%	0,0%	0,0%	0,0%	
Galicia	135	GAL	Iberia	IE, Romance	2,9%	1,5%	0,0%	0,0%	0,0%	0,0%	0,0%	2,2%	0,0%	0,0%	0,0%	0,0%	0,7%	0,0%	
North German	140	GEN	Western Europe	IE, Germanic	0,7%	0,7%	0,0%	0,0%	0,0%	0,7%	0,7%	1,4%	0,0%	0,0%	0,0%	0,0%	0,7%	0,0%	
Scottish	874	SCO	British Isles	IE, Germanic	0,1%	0,8%	0,0%	0,0%	0,0%	0,0%	0,1%	4,2%	1,3%	0,0%	0,0%	0,0%	2,1%	0,0%	
South German	266	GES	Western Europe	IE, Germanic	0,0%	0,4%	0,0%	0,0%	0,0%	0,0%	0,8%	3,0%	1,1%	0,0%	0,0%	0,0%	1,1%	0,0%	
Romani (Gypsy)	275	GYP	Balkans	IE, Indo-Iranian	0,0%	26,5%	0,0%	0,0%	0,0%	0,0%	1,8%	1,8%	1,1%	0,0%	0,0%	0,0%	7,6%	0,0%	
Austria/Switzerland	187	ATS	Western Europe	IE, Germanic	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	2,1%	1,6%	0,0%	0,0%	0,0%	0,5%	0,0%	
Iceland	467	ICE	Scandinavia	IE, Germanic	0,0%	0,0%	0,4%	0,2%	0,0%	0,0%	0,0%	4,7%	0,2%	0,0%	0,0%	0,0%	1,5%	0,0%	
Ireland	128	IRE	British Isles	IE, Celtic	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	2,3%	2,3%	0,0%	0,0%	0,0%	0,0%	0,0%	
Orkney	152	ORK	British Isles	IE, Germanic	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	3,3%	2,0%	0,0%	0,0%	0,0%	7,2%	0,0%	
Scottish Isles	246	SCIL	British Isles	IE, Celtic	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	6,5%	0,4%	0,0%	0,4%	0,0%	2,0%	0,0%	
Saami	545	SAA	Scandinavia	Finno-Permic	0,0%	0,0%	0,0%	2,0%	0,0%	2,6%	0,0%	0,0%	1,5%	0,0%	0,0%	0,0%	0,0%	0,0%	
Bulgarians	855	BLG	Balkans	IE, Balto-Slavic	0,4%	0,2%	0,2%	0,1%	0,0%	0,4%	1,8%	1,2%	2,8%	0,0%	0,0%	0,0%	2,1%	0,9%	
Belarusian	267	BLR	Eastern Europe	IE, Balto-Slavic	0,0%	0,7%	1,1%	0,0%	0,4%	0,4%	3,4%	2,6%	3,4%	0,0%	0,0%	0,0%	1,1%	0,0%	
Ukrainians	18	UKR	Eastern Europe	IE, Balto-Slavic	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	5,6%	0,0%	
Bosnians	144	BOS	Balkans	IE, Balto-Slavic	0,7%	1,4%	0,0%	0,0%	0,0%	0,0%	0,7%	2,8%	1,4%	0,0%	0,0%	0,0%	1,4%	0,0%	
Poles	436	POL	Central Europe	IE, Balto-Slavic	0,2%	1,8%	0,0%	0,0%	0,0%	0,0%	0,5%	1,8%	3,7%	0,0%	0,0%	0,0%	1,8%	0,5%	
Slovenians	104	SLN	Central Europe	IE, Balto-Slavic	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	1,9%	4,8%	0,0%	0,0%	0,0%	0,0%	1,0%	0,0%	
Czechs	179	CZH	Central Europe	IE, Balto-Slavic	0,0%	1,1%	0,0%	0,0%	0,0%	0,6%	2,3%	2,8%	0,6%	0,0%	0,6%	0,0%	1,7%	0,0%	
Slovaks	207	SLK	Central Europe	IE, Balto-Slavic	1,0%	1,0%	0,0%	0,0%	0,0%	0,0%	0,5%	4,8%	2,9%	0,0%	0,0%	0,0%	1,4%	0,0%	
Japanese	169	JPN	East Asia	Japanese	0,0%	18,4%	1,1%	1,7%	11,2%	34,3%	4,7%	0,0%	0,6%	10,1%	0,0%	0,0%	0,0%	0,6%	
Russians	953	RUS	Eastern Europe	IE, Balto-Slavic	0,3%	0,2%	0,7%	0,1%	0,3%	0,5%	0,9%	2,6%	2,0%	0,0%	0,4%	0,0%	1,7%	0,4%	
Norwegian	74	NOR	Scandinavia	IE, Germanic	1,4%	0,0%	0,0%	0,0%	0,0%	0,0%	1,4%	4,1%	2,7%	0,0%	0,0%	0,0%	1,4%	0,0%	

HV*	H	V	J	T	F	B	U*	U2	U4	U5	K	SUM	Source
7,2%	39,1%	3,1%	5,6%	11,2%	0,0%	0,0%	5,9%	2,2%	1,9%	4,1%	6,8%	99,5%	Achili 2007
0,0%	22,8%	0,0%	6,3%	8,9%	0,0%	0,0%	12,7%	0,0%	2,5%	2,5%	22,8%	98,7%	Badro 2013
3,9%	45,4%	3,0%	7,9%	9,0%	0,1%	0,2%	2,8%	1,4%	2,5%	7,8%	8,4%	100,0%	Badro 2013
5,9%	42,1%	0,0%	10,2%	8,6%	0,0%	0,0%	6,7%	0,3%	3,8%	4,0%	5,6%	99,7%	Badro 2013
11,8%	15,7%	0,0%	21,6%	9,8%	0,0%	0,0%	15,7%	3,9%	2,0%	0,0%	7,8%	100,1%	Badro 2013
6,6%	26,2%	0,3%	8,6%	9,0%	0,0%	0,0%	17,6%	0,7%	0,3%	0,7%	3,8%	100,0%	Badro 2013
7,6%	29,9%	0,9%	9,4%	10,2%	0,0%	0,0%	11,9%	0,0%	0,0%	0,0%	9,2%	99,4%	Badro 2013
8,3%	25,0%	0,0%	15,0%	12,5%	0,0%	0,0%	10,0%	0,0%	0,0%	0,0%	4,2%	99,9%	Badro 2013
0,7%	8,7%	0,0%	21,5%	6,3%	0,0%	0,0%	8,3%	1,3%	0,6%	0,7%	4,1%	99,1%	Badro 2013
7,7%	24,4%	1,3%	11,5%	11,1%	0,0%	0,0%	10,3%	0,0%	0,0%	0,0%	10,3%	99,7%	Badro 2013
2,3%	4,7%	0,0%	14,0%	5,0%	0,0%	0,0%	3,7%	1,0%	0,7%	0,3%	6,0%	100,0%	Badro 2013
0,0%	1,2%	0,0%	0,4%	1,7%	0,0%	0,0%	0,0%	10,3%	0,0%	0,0%	0,4%	99,1%	Bamshad 2001
0,0%	2,3%	0,0%	0,7%	1,7%	2,7%	0,0%	0,0%	15,3%	0,0%	0,0%	0,0%	99,8%	Bamshad 2001
0,5%	12,2%	3,2%	3,2%	5,4%	6,3%	0,9%	0,5%	0,5%	12,7%	13,6%	1,4%	99,0%	Bermisheva 2002
0,0%	27,3%	7,3%	5,5%	3,6%	0,0%	0,0%	5,4%	0,0%	16,4%	14,5%	7,3%	99,9%	Bermisheva 2002
0,0%	32,4%	0,0%	4,1%	13,5%	0,0%	0,0%	1,4%	0,0%	9,5%	5,4%	1,4%	100,2%	Bermisheva 2002
0,0%	33,9%	0,0%	9,7%	12,9%	0,0%	0,0%	1,6%	0,0%	24,2%	9,7%	1,6%	100,0%	Bermisheva 2002
1,5%	40,4%	11,0%	7,4%	5,1%	0,0%	0,0%	0,0%	0,0%	10,3%	14,0%	2,2%	99,9%	Bermisheva 2002
1,0%	42,2%	4,9%	7,8%	7,9%	0,0%	0,0%	2,0%	6,9%	2,0%	15,7%	0,0%	100,3%	Bermisheva 2002
0,9%	30,7%	3,9%	7,5%	9,2%	0,0%	0,0%	5,3%	0,9%	7,0%	10,5%	5,7%	99,7%	Bermisheva 2002
0,0%	21,8%	0,0%	2,0%	23,8%	0,0%	0,0%	2,0%	9,9%	4,0%	8,9%	0,0%	100,2%	Bermisheva 2002
4,8%	50,0%	2,4%	4,8%	7,2%	0,0%	0,0%	2,4%	0,0%	4,8%	7,1%	4,8%	100,2%	Bosch 2006
1,7%	44,6%	3,4%	6,3%	16,6%	0,0%	0,0%	2,2%	0,0%	2,3%	10,9%	5,7%	98,7%	Bosch 2006
2,7%	45,9%	2,7%	5,4%	16,2%	0,0%	0,0%	8,1%	0,0%	5,4%	2,7%	5,4%	99,9%	Bosch 2006
5,7%	41,0%	1,9%	13,4%	8,6%	0,0%	0,0%	3,8%	1,9%	1,0%	3,8%	7,7%	99,1%	Bosch 2006
0,0%	11,9%	1,7%	10,2%	6,8%	1,7%	0,0%	5,1%	0,0%	13,6%	5,1%	5,1%	96,8%	Derbeneva 2002
0,0%	5,6%	0,0%	5,6%	0,0%	5,5%	4,4%	2,2%	0,0%	2,2%	1,1%	6,7%	99,4%	Derenko 2007
1,0%	6,8%	0,0%	0,7%	1,0%	3,1%	3,4%	1,0%	0,0%	1,0%	2,0%	1,4%	100,2%	Derenko 2007
0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	99,4%	Derenko 2007
5,5%	3,6%	0,0%	3,6%	1,8%	5,5%	3,6%	0,9%	1,8%	1,8%	0,9%	2,7%	99,9%	Derenko 2007
0,0%	7,0%	1,8%	5,3%	5,3%	24,3%	8,8%	0,0%	0,0%	1,8%	1,8%	0,0%	100,2%	Derenko 2007
2,0%	6,1%	0,0%	1,0%	1,0%	4,0%	7,1%	0,0%	0,0%	0,0%	2,0%	2,0%	99,6%	Derenko 2007
0,0%	0,0%	0,0%	0,0%	0,0%	4,9%	20,8%	0,0%	0,0%	0,0%	0,0%	0,0%	100,4%	Derenko 2007
0,0%	0,0%	0,0%	0,0%	0,0%	6,4%	15,3%	0,0%	0,0%	6,4%	0,0%	2,1%	100,8%	Derenko 2007
0,0%	11,0%	0,0%	6,1%	0,0%	41,2%	4,9%	0,0%	0,0%	2,4%	0,0%	1,2%	99,6%	Derenko 2007
1,4%	11,0%	0,0%	2,8%	5,6%	1,4%	11,0%	1,4%	2,8%	1,4%	1,4%	0,0%	98,8%	Derenko 2007
0,0%	7,5%	5,7%	0,0%	5,7%	5,7%	3,8%	0,0%	3,8%	1,9%	0,0%	1,9%	100,5%	Derenko 2007
1,0%	3,8%	0,0%	0,0%	0,0%	8,6%	1,9%	1,9%	0,0%	1,0%	1,9%	0,0%	100,0%	Derenko 2007
0,0%	1,4%	0,0%	0,0%	0,0%	1,4%	4,1%	0,0%	0,0%	0,0%	0,0%	2,7%	100,6%	Derenko 2007
0,0%	2,8%	0,0%	5,6%	0,0%	0,0%	2,8%	0,0%	0,0%	0,0%	0,0%	0,0%	100,4%	Derenko 2007
3,7%	26,0%	0,0%	4,9%	13,4%	0,0%	6,1%	6,1%	1,2%	0,0%	4,9%	3,7%	99,0%	Derenko 2007
3,1%	13,3%	0,0%	5,1%	4,0%	5,1%	4,1%	1,0%	1,0%	4,1%	2,0%	2,0%	99,9%	Derenko 2013
1,3%	2,0%	0,0%	0,0%	0,0%	3,4%	8,0%	2,0%	0,7%	0,0%	0,7%	1,3%	100,0%	Derenko 2013
6,4%	14,1%	0,0%	0,0%	0,0%	6,4%	1,3%	5,1%	3,9%	1,3%	6,4%	2,6%	100,1%	Di Cristofaro 2013
5,6%	20,0%	0,0%	3,3%	4,4%	1,1%	0,0%	6,7%	12,2%	1,1%	2,2%	2,2%	99,7%	Di Cristofaro 2013
9,6%	18,5%	0,0%	4,8%	4,4%	1,4%	1,4%	3,3%	8,9%	2,1%	6,9%	2,7%	99,9%	Di Cristofaro 2013
8,0%	22,7%	0,0%	2,7%	2,6%	2,6%	0,0%	1,3%	5,3%	5,4%	0,0%	5,3%	97,1%	Di Cristofaro 2013
11,0%	13,3%	0,0%	3,9%	3,9%	0,0%	1,6%	8,0%	3,9%	5,5%	0,8%	2,4%	97,5%	Di Cristofaro 2013
1,0%	40,5%	2,9%	10,0%	14,3%	0,0%	0,0%	3,4%	1,4%	2,9%	7,1%	4,8%	99,1%	Egyed 2007
0,0%	53,1%	3,1%	15,6%	6,5%	0,0%	0,0%	0,0%	0,0%	2,3%	6,2%	7,3%	100,2%	González 2003
0,0%	36,0%	12,0%	4,0%	8,0%	0,0%	0,0%	0,0%	2,0%	0,0%	10,0%	4,0%	100,0%	González 2003
0,0%	63,7%	5,2%	7,4%	3,6%	0,0%	0,0%	2,2%	1,5%	0,7%	3,0%	5,2%	99,8%	González 2003
0,0%	52,1%	5,7%	10,7%	8,5%	0,0%	0,0%	1,4%	0,7%	0,0%	7,1%	7,9%	99,7%	González 2003
0,0%	43,2%	3,5%	12,2%	10,2%	0,0%	0,0%	2,7%	0,8%	2,6%	8,3%	8,0%	100,1%	González 2003
0,0%	51,5%	4,5%	8,3%	9,5%	0,0%	0,0%	2,3%	0,4%	4,1%	7,6%	5,6%	100,2%	González 2003
0,0%	35,6%	0,0%	9,1%	2,2%	0,0%	0,0%	10,5%	0,0%	0,0%	2,2%	1,4%	99,8%	Gresham 2001
0,0%	54,0%	2,7%	10,2%	6,4%	0,0%	0,0%	1,1%	1,1%	3,7%	7,5%	9,1%	100,0%	Helgason 2001
0,0%	47,6%	1,7%	14,1%	10,1%	0,0%	0,0%	0,2%	0,0%	2,1%	9,4%	7,7%	99,9%	Helgason 2001
0,0%	47,7%	7,0%	14,1%	9,4%	0,0%	0,0%	0,0%	0,8%	2,3%	6,3%	7,8%	100,0%	Helgason 2001
0,0%	50,7%	1,3%	9,9%	5,9%	0,0%	0,0%	0,0%	0,0%	0,7%	11,8%	6,6%	99,4%	Helgason 2001
0,0%	34,6%	2,0%	14,6%	12,6%	0,0%	0,0%	4,5%	0,4%	0,4%	8,1%	13,4%	99,9%	Helgason 2001
0,0%	4,8%	41,3%	0,0%	0,6%	0,0%	0,0%	0,0%	0,0%	0,0%	46,1%	0,4%	99,3%	Ingman 2007
6,0%	41,9%	1,5%	7,9%	10,6%	0,0%	0,0%	4,2%	1,4%	3,9%	6,7%	5,9%	100,1%	Karachanak 2011
4,1%	37,1%	5,2%	9,0%	7,5%	0,0%	0,0%	3,0%	1,5%	3,8%	12,0%	3,7%	100,0%	Kushniarevich 2013
0,0%	44,4%	0,0%	5,6%	16,7%	0,0%	0,0%	0,0%	0,0%	11,1%	11,1%	5,6%	100,1%	Malyarchuk 2001
2,1%	47,9%	6,3%	6,9%	4,9%	0,0%	0,0%	2,1%	0,0%	5,6%	11,8%	4,2%	100,2%	Malyarchuk 2003
0,9%	45,2%	4,8%	7,8%	11,5%	0,0%	0,0%	1,4%	0,9%	5,1%	8,7%	3,4%	100,0%	Malyarchuk 2003
0,0%	47,1%	6,7%	9,6%	5,8%	0,0%	0,0%	1,9%	1,0%	5,8%	10,6%	3,9%	100,1%	Malyarchuk 2003
3,9%	44,2%	0,0%	11,7%	12,3%	0,0%	0,0%	2,8%	0,6%	2,8%	8,4%	3,9%	100,3%	Malyarchuk 2008
6,8%	45,2%	0,0%	9,2%	10,6%	0,0%	0,0%	1,0%	1,0%	5,3%	7,7%	1,4%	99,8%	Malyarchuk 2008
0,0%	0,0%	0,0%	0,0%	0,0%	5,3%	11,9%	0,0%	0,0%	0,0%	0,0%	0,0%	100,1%	Maruyama 2003
2,5%	40,5%	2,3%	9,7%	8,9%	0,1%	0,2%	4,1%	1,2%	4,5%	10,9%	4,4%	99,4%	Morozova 2011
0,0%	39,2%	5,4%	10,8%	12,2%	0,0%	0,0%	4,1%	0,0%	0,0%	13,5%	4,1%	100,3%	Passarino 2002

Portuguese	241	POR	Iberia	IE, Romance	7,0%	0,4%	0,0%	0,0%	0,0%	0,0%	0,0%	0,8%	1,2%	0,0%	0,0%	0,0%	2,1%	0,0%
Khanty	106	KHA	West Siberia	Ugric	0,0%	0,0%	10,4%	0,0%	1,9%	16,0%	0,9%	0,0%	0,0%	0,9%	0,0%	0,9%	0,0%	0,0%
Mansi	63	MAN	West Siberia	Ugric	0,0%	0,0%	20,6%	0,0%	1,6%	19,0%	3,2%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Estonians	409	EST	Eastern Europe	Finno-Permic	0,0%	0,0%	0,0%	0,0%	0,0%	0,2%	1,4%	1,0%	2,4%	0,0%	0,0%	0,0%	1,0%	0,0%
Latvians	299	LAT	Eastern Europe	IE, Balto-Slavic	0,0%	0,0%	0,0%	0,0%	0,3%	0,0%	0,0%	4,3%	4,0%	0,0%	0,0%	0,0%	0,3%	0,0%
Lithuanians	180	LIT	Eastern Europe	IE, Balto-Slavic	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	1,1%	3,9%	1,1%	0,0%	0,6%	0,0%	1,1%	0,0%
Azeri	40	AZR	Greater Iran	Altaic, Turkic	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	2,5%	5,0%	2,5%	0,0%	0,0%	0,0%	0,0%	0,0%
Balochi	39	BCH	Greater Iran	IE, Indo-Iranian	2,6%	33,3%	0,0%	0,0%	0,0%	0,0%	5,2%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	7,7%
Georgians	58	GEO	Caucasus	Kartvelian	0,0%	1,8%	0,0%	0,0%	0,0%	0,0%	6,9%	1,8%	5,2%	0,0%	0,0%	0,0%	8,6%	1,8%
Kalash	44	KAL	Greater Iran	IE, Indo-Iranian	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Kurdish	52	KUR	Greater Iran	IE, Indo-Iranian	0,0%	0,0%	5,8%	0,0%	0,0%	0,0%	1,9%	3,8%	3,8%	0,0%	1,9%	0,0%	3,9%	5,8%
Pakistani	100	PAK	South Asia	IE, Indo-Iranian	1,0%	47,0%	0,0%	1,0%	1,0%	0,0%	3,0%	0,0%	1,0%	0,0%	0,0%	1,0%	8,0%	0,0%
Shungan (Pamiri)	44	PAM	Greater Iran	IE, Indo-Iranian	0,0%	0,0%	18,2%	0,0%	0,0%	0,0%	0,0%	2,3%	4,5%	0,0%	2,3%	0,0%	2,3%	0,0%
Turkish	50	TUR	Middle East	Altaic, Turkic	0,0%	0,0%	0,0%	0,0%	2,0%	2,0%	2,0%	2,0%	0,0%	0,0%	4,0%	0,0%	6,0%	0,0%
Itelmen	47	ITE	North Siberia	Chukotko-K.	0,0%	0,0%	14,9%	6,4%	68,1%	0,0%	0,0%	0,0%	4,1%	6,4%	0,0%	0,0%	0,0%	0,0%
Ket	38	KET	North Siberia	Yeniseian	0,0%	0,0%	15,8%	2,6%	0,0%	2,6%	0,0%	0,0%	2,7%	0,0%	7,9%	0,0%	0,0%	0,0%
Koryak	155	KYK	North Siberia	Chukotko-K.	0,0%	0,0%	36,1%	5,8%	41,9%	1,3%	0,0%	0,0%	9,7%	5,2%	0,0%	0,0%	0,0%	0,0%
Negidal	33	NGD	East Siberia	Altaic, Tungusic	0,0%	0,0%	15,2%	0,0%	27,2%	24,3%	0,0%	0,0%	21,1%	0,0%	0,0%	0,0%	0,0%	0,0%
Nivkh	46	NIV	East Siberia	Nivkh	0,0%	0,0%	0,0%	0,0%	5,4%	28,6%	0,0%	0,0%	0,0%	66,1%	0,0%	0,0%	0,0%	0,0%
Udegey	56	UDG	East Siberia	Altaic, Tungusic	0,0%	28,3%	17,4%	0,0%	0,0%	0,0%	30,4%	0,0%	0,0%	8,7%	0,0%	15,2%	0,0%	0,0%
Ulchi	87	ULC	East Siberia	Altaic, Tungusic	0,0%	2,3%	13,7%	0,0%	11,4%	21,8%	6,9%	0,0%	0,0%	37,9%	0,0%	4,6%	0,0%	0,0%
Aleut Islanders	199	ALE	North America	Eskimo-Aleut	0,0%	0,0%	0,0%	0,0%	0,0%	71,9%	0,0%	0,0%	0,0%	28,2%	0,0%	0,0%	0,0%	0,0%
Canadian Eskimo	96	ESC	North America	Eskimo-Aleut	0,0%	0,0%	0,0%	0,0%	0,0%	12,5%	0,0%	0,0%	0,0%	87,5%	0,0%	0,0%	0,0%	0,0%
Chukchi	182	CHK	North Siberia	Chukotko-K.	0,0%	0,0%	18,8%	0,0%	26,7%	7,0%	0,0%	0,0%	0,0%	47,6%	0,0%	0,0%	0,0%	0,0%
Chuvantsi	32	CVN	North Siberia	Chukotko-K.	0,0%	0,0%	31,3%	6,3%	28,1%	9,4%	0,0%	0,0%	0,0%	25,1%	0,0%	0,0%	0,0%	0,0%
Greenland Eskimo	385	ESG	North America	Eskimo-Aleut	0,0%	0,0%	0,0%	0,0%	0,0%	4,0%	0,0%	0,0%	0,0%	96,1%	0,0%	0,0%	0,0%	0,0%
Siberian Eskimo	126	ESS	North Siberia	Eskimo-Aleut	0,0%	0,0%	0,0%	0,0%	0,0%	20,6%	0,0%	0,0%	0,0%	79,4%	0,0%	0,0%	0,0%	0,0%
Yukaghir	100	YUK	North Siberia	Uralic-Yukaghir	0,0%	0,0%	67,0%	3,0%	12,0%	18,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Ngasan	39	NGA	North Siberia	Samoyedic	0,0%	0,0%	51,3%	2,6%	0,0%	25,6%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Tibetan	155	TIB	Tibet-Yunnan	Tibeto-Burman	0,0%	42,6%	3,2%	4,5%	5,2%	18,7%	1,3%	0,0%	0,0%	0,0%	15,5%	0,0%	0,0%	0,0%
Han Chinese	251	CHN	East Asia	Sino-Tibetan	0,0%	20,7%	3,1%	2,0%	3,6%	21,5%	5,2%	0,0%	0,0%	1,2%	7,1%	0,0%	0,0%	2,4%
Hui Chinese	45	HUI	Inner Asia	Sino-Tibetan	0,0%	24,2%	2,2%	4,4%	8,8%	15,5%	0,0%	0,0%	0,0%	6,7%	0,0%	0,0%	2,2%	0,0%
Kazakh	93	KAZ	Central Asia	Altaic, Turkic	0,0%	9,5%	13,2%	11,3%	5,7%	13,2%	1,9%	1,9%	0,0%	0,0%	3,8%	0,0%	0,0%	0,0%
Kyrgyz	55	KYG	Central Asia	Altaic, Turkic	0,0%	7,4%	12,6%	1,1%	8,4%	20,0%	3,2%	0,0%	0,0%	4,2%	0,0%	0,0%	3,2%	0,0%
Uyghur	47	UYG	Inner Asia	Altaic, Turkic	0,0%	14,9%	6,4%	0,0%	12,7%	10,5%	0,0%	0,0%	6,4%	0,0%	4,3%	0,0%	0,0%	0,0%
Croatian	721	CRO	Balkans	IE, Balto-Slavic	0,0%	0,1%	0,0%	0,0%	0,0%	0,0%	0,1%	3,2%	3,3%	0,0%	0,0%	0,0%	1,1%	0,0%
Csángó	182	CSG	Central Europe	Ugric	0,0%	0,0%	0,5%	0,0%	0,0%	0,0%	3,8%	1,6%	7,7%	0,0%	1,6%	0,0%	1,6%	0,0%
Sekler	178	SEK	Central Europe	Ugric	0,0%	0,0%	5,1%	0,0%	0,6%	0,0%	0,6%	0,6%	5,1%	0,6%	1,1%	0,0%	2,2%	0,0%
<b>ancient populations</b>																		
Middle East Neolithic-BRA	28	MEN	Middle East	11840-1402	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	3,6%	3,6%	0,0%	0,0%	0,0%	0,0%	10,7%	10,7%
Iberian Neolithic	45	IBN	Iberia	10310-3160	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	2,2%	0,0%	0,0%	0,0%	0,0%	2,2%	0,0%
Cisbaikalian Neol.(Serovo)	15	SER	East Siberia	8000-4000	0,0%	0,0%	13,3%	13,3%	20,0%	46,7%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Early-Middle Neolithic	53	EMN	Europe	6000-3000	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	9,4%	0,0%	5,7%	0,0%	0,0%	0,0%	1,9%	0,0%
Starcevo	44	STR	Balkans	5700-5500	0,0%	0,0%	0,0%	0,0%	0,0%	6,8%	0,0%	4,6%	0,0%	0,0%	0,0%	6,8%	0,0%	0,0%
Dniepr-Donets Neolithic	17	DDO	Eastern Europe	5300-4700	0,0%	0,0%	35,3%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
LBK Transdanubia	39	LBT	Hungary	5200-4800	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	10,3%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Székálhát	33	SZA	Hungary	5200-4800	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	3,0%	0,0%	3,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Yamnaya, Afanasievo	49	YAM	Russia, Ukraine	5000-2700	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	2,0%	2,0%	6,1%	0,0%	0,0%	2,0%	0,0%	0,0%
Catacomb Kurgans	25	CTC	Ukraine	4700-2800	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	4,0%	0,0%	0,0%	0,0%	0,0%	0,0%	8,0%	0,0%
Enolithic kurgans	10	ENO	BG, MD, UA	4500-2000	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	10,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Baraba (UT-ODI-EK)	33	BB1	West Siberia	4000-1800	0,0%	0,0%	21,2%	18,2%	0,0%	24,2%	0,0%	0,0%	0,0%	9,1%	0,0%	0,0%	0,0%	0,0%
Late Neolithic-EBA Europe	56	LNB	Europe	3000-1600	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	3,6%	3,6%	0,0%	0,0%	0,0%	1,8%	0,0%	0,0%
Altai Bronze Age	12	ABA	South Siberia	2700-900	0,0%	0,0%	8,3%	0,0%	0,0%	33,3%	0,0%	0,0%	0,0%	8,3%	0,0%	0,0%	8,0%	0,0%
Tarim Basin Xiaohu	73	XIA	China	2515-1829	0,0%	4,1%	46,6%	0,0%	4,1%	19,2%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	5,5%	0,0%
Sintashta-Andronovo	41	SIA	Russia, Siberia	2300-1400	0,0%	0,0%	0,0%	5,0%	0,0%	0,0%	5,0%	0,0%	0,0%	5,0%	0,0%	0,0%	0,0%	0,0%
Bronze age (Vatya, Maros)	8	VAT	Hungary	2100-1600	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Baraba (LK-FYOD-LBB)	45	BB2	West Siberia	1800-1000	0,0%	0,0%	44,4%	0,0%	2,2%	0,0%	0,0%	0,0%	0,0%	15,6%	0,0%	0,0%	0,0%	0,0%
Srubnaya	14	SRU	Russia	1800-1200	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	7,1%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Bronze Age Kurgans	13	KBK	Kazakhstan	1400-1000	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	7,7%	7,7%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Baraba (Iron transition)	14	BB3	West Siberia	1000-800 BC	0,0%	0,0%	0,0%	0,0%	0,0%	7,1%	0,0%	0,0%	7,1%	0,0%	0,0%	0,0%	0,0%	0,0%
Iron Age Kurgans	13	KIK	Kazakhstan	800-600 BC	0,0%	15,4%	0,0%	0,0%	15,4%	0,0%	0,0%	0,0%	0,0%	23,1%	0,0%	0,0%	0,0%	0,0%
Tagar-Tachtyk	15	TAG	Russia	800 BC-400 AD	0,0%	0,0%	13,3%	0,0%	6,7%	0,0%	6,7%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Scythian Iron age	14	SCI	Russia	600-200 BC	0,0%	0,0%	7,1%	0,0%	0,0%	14,3%	0,0%	7,1%	0,0%	0,0%	7,1%	0,0%	0,0%	0,0%
Pazyryk Scytho-Siberian	25	PAZ	Mongolia, Russia	400-200 BC	0,0%	0,0%	12,0%	0,0%	4,0%	24,0%	4,0%	0,0%	0,0%	8,0%	0,0%	0,0%	0,0%	0,0%
Qin China aDNA	19	QIN	East Asia	221 BC-210 AD	0,0%	10,5%	5,3%	0,0%	5,3%	10,5%	21,1%	0,0%	0,0%	5,3%	0,0%	0,0%	0,0%	0,0%
Egyn Gol Xiongnu	46	XIO	Inner Asia	200 BC-200 AD	0,0%	4,3%	13,0%	0,0%	2,2%	41,3%	0,0%	0,0%	0,0%	17,4%	0,0%	0,0%	0,0%	0,0%
Lombard early medieval	40	LOM	Hungary, Italy	500-800 AD	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	5,0%	2,5%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Vikings	65	VIK	Norway	780-790 AD	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	4,6%	0,0%	7,7%	0,0%	0,0%	0,0%	0,0%
Ancient Hungarian Karos	17	KAR	Hungary	900-950 AD	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	5,9%	0,0%	5,9%	0,0%	0,0%
Hungarians 900 AD	27	AH2	Central Europe	900-1000 AD	0,0%	3,7%	0,0%	0,0%	0,0%	0,0%	7,4%	3,7%	0,0%	0,0%	0,0%	7,4%	7,4%	0,0%
Ancient Hungarian (10th cent)	67	AH1	Central Europe	900-1000 AD	0,0%	1,5%	1,5%	0,0%	0,0%	3,0%	6,0%	0,0%	0,0%	4,5%	0,0%	1,5%	3,0%	0,0%
Medieval Slavic	19	SLV	Slovakia	900-1200 AD	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	5,3%	0,0%
Italian medieval	27	ITM	Italy	900-1500	0,0%	3,7%	0,0%	0,0%	0,0%	3,7%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Cumanian	11	CUM	Hungary	1200-1300 AD	0,0%	0,0%	0,0%	0,0%	0,0%	9,1%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%

0,0%	40,7%	7,5%	6,6%	10,8%	0,0%	0,0%	7,8%	0,8%	1,7%	7,1%	5,4%	99,8%	Pereira 2000
0,0%	14,2%	0,0%	15,1%	4,7%	0,9%	0,0%	15,1%	0,9%	8,5%	9,4%	0,9%	99,9%	Pimenoff 2008
0,0%	14,3%	0,0%	14,3%	3,2%	0,0%	0,0%	3,2%	4,8%	14,3%	1,6%	0,0%	100,0%	Pimenoff 2008
1,5%	43,5%	3,9%	10,3%	7,8%	0,0%	0,0%	3,1%	1,5%	5,4%	14,2%	2,7%	99,9%	Pliss 2006
2,3%	44,5%	3,0%	6,4%	9,4%	0,0%	0,0%	1,7%	3,0%	9,4%	9,0%	2,3%	99,9%	Pliss 2006
1,7%	46,1%	5,6%	7,8%	10,0%	0,0%	0,0%	3,4%	0,0%	5,6%	9,4%	2,8%	100,2%	Pliss 2006
7,5%	30,0%	2,5%	10,0%	15,0%	0,0%	0,0%	20,0%	0,0%	0,0%	0,0%	5,0%	100,0%	Quintana-Murci 2004
10,3%	20,5%	0,0%	7,7%	0,0%	0,0%	0,0%	2,6%	2,6%	2,6%	2,6%	2,6%	100,3%	Quintana-Murci 2004
3,4%	22,4%	0,0%	8,6%	13,7%	0,0%	0,0%	0,0%	3,4%	3,4%	1,8%	8,6%	100,0%	Quintana-Murci 2004
27,2%	4,5%	0,0%	11,4%	4,5%	0,0%	0,0%	2,3%	15,9%	34,1%	0,0%	0,0%	99,9%	Quintana-Murci 2004
7,6%	11,5%	0,0%	7,7%	11,6%	0,0%	3,8%	13,5%	3,8%	0,0%	1,9%	11,5%	99,8%	Quintana-Murci 2004
6,0%	12,0%	0,0%	1,0%	1,0%	0,0%	0,0%	6,0%	11,0%	0,0%	0,0%	0,0%	100,0%	Quintana-Murci 2004
4,5%	29,5%	0,0%	4,5%	2,3%	0,0%	0,0%	4,6%	6,8%	4,5%	4,5%	9,1%	99,9%	Quintana-Murci 2004
10,0%	26,0%	0,0%	8,0%	8,0%	0,0%	0,0%	12,0%	2,0%	2,0%	8,0%	6,0%	100,0%	Quintana-Murci 2004
0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	99,9%	Starikovskaya 2005
0,0%	10,5%	0,0%	0,0%	0,0%	23,7%	0,0%	0,0%	0,0%	28,9%	5,3%	0,0%	100,0%	Starikovskaya 2005
0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%	Starikovskaya 2005
0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	12,1%	0,0%	0,0%	0,0%	0,0%	0,0%	99,9%	Starikovskaya 2005
0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,1%	Starikovskaya 2005
0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%	Starikovskaya 2005
0,0%	0,0%	0,0%	0,0%	0,0%	1,1%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	99,7%	Starikovskaya 2005
0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,1%	Volodko 2008
0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%	Volodko 2008
0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,1%	Volodko 2008
0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,2%	Volodko 2008
0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,1%	Volodko 2008
0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%	Volodko 2008
0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	3,9%	5,2%	0,0%	0,0%	0,0%	0,0%	100,1%	Wen 2004
0,4%	0,0%	0,0%	0,0%	0,4%	15,5%	16,8%	0,0%	0,0%	0,0%	0,0%	0,0%	99,9%	Yao 2002
0,0%	0,0%	0,0%	0,0%	4,4%	13,2%	17,7%	0,0%	0,0%	0,0%	0,0%	0,0%	99,3%	Yao 2004
5,7%	7,5%	1,9%	1,9%	7,6%	7,6%	3,8%	0,0%	0,0%	0,0%	3,8%	0,0%	100,3%	Yao 2004
6,3%	14,7%	0,0%	5,3%	1,1%	2,1%	6,3%	3,2%	0,0%	0,0%	1,1%	0,0%	100,1%	Yao 2004
4,2%	10,6%	0,0%	4,2%	2,1%	6,4%	2,1%	2,1%	6,4%	4,3%	0,0%	2,1%	99,7%	Yao 2004
3,7%	41,1%	5,2%	9,7%	7,4%	1,4%	0,0%	0,2%	1,7%	1,8%	10,3%	5,1%	95,4%	Pericic 2005
3,8%	23,6%	0,0%	13,2%	3,8%	0,0%	0,0%	3,2%	1,6%	1,6%	8,8%	23,1%	99,5%	Brandstatter 2007
9,6%	33,7%	0,0%	8,5%	9,5%	0,0%	0,6%	5,1%	1,7%	3,4%	7,9%	4,5%	100,4%	Brandstatter 2007
0,0%	17,9%	0,0%	3,6%	10,7%	0,0%	0,0%	14,3%	0,0%	3,6%	0,0%	21,4%	100,0%	Lazaridis 2016
2,2%	44,4%	0,0%	4,4%	2,2%	0,0%	0,0%	20,0%	0,0%	0,0%	13,3%	8,9%	99,8%	Hervella 2015
0,0%	0,0%	0,0%	0,0%	0,0%	6,7%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	100,0%	Trapezov 2012
3,8%	26,4%	3,8%	7,5%	20,8%	0,0%	0,0%	3,8%	0,0%	0,0%	7,5%	9,4%	100,0%	Haak 2015
2,3%	6,8%	6,8%	11,4%	22,7%	0,0%	0,0%	2,3%	0,0%	2,3%	0,0%	27,3%	100,1%	Szécsényi-Nagy 2014
0,0%	35,3%	0,0%	0,0%	17,6%	0,0%	0,0%	5,9%	0,0%	0,0%	5,9%	0,0%	100,0%	Newton 2011
2,6%	30,8%	2,6%	7,7%	28,2%	0,0%	0,0%	0,0%	2,6%	0,0%	2,6%	12,8%	100,2%	Szécsényi-Nagy 2014
3,0%	18,2%	6,1%	15,2%	33,4%	0,0%	0,0%	0,0%	0,0%	3,0%	0,0%	15,2%	100,1%	Szécsényi-Nagy 2014
0,0%	20,4%	0,0%	6,1%	20,4%	0,0%	0,0%	2,9%	2,0%	5,3%	26,5%	4,1%	100,0%	Haak 2015; Allentoft 20
0,0%	28,0%	0,0%	12,0%	0,0%	0,0%	0,0%	44,0%	0,0%	0,0%	4,0%	0,0%	100,0%	Wilde 2014
0,0%	40,0%	0,0%	0,0%	20,0%	0,0%	0,0%	10,0%	0,0%	0,0%	20,0%	0,0%	100,0%	Wilde 2014
0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	15,2%	3,0%	9,1%	0,0%	0,0%	100,0%	Molodin 2012
1,8%	33,9%	1,8%	5,4%	10,7%	0,0%	0,0%	0,0%	1,8%	5,4%	14,3%	16,1%	100,0%	Haak 2015; Allentoft 20
0,0%	25,0%	0,0%	0,0%	8,0%	0,0%	0,0%	0,0%	8,0%	0,0%	0,0%	0,0%	98,9%	Hollard 2014
0,0%	2,7%	0,0%	0,0%	1,4%	0,0%	1,4%	8,2%	1,4%	0,0%	2,7%	2,7%	100,0%	Li 2015
0,0%	10,0%	0,0%	10,0%	15,0%	0,0%	0,0%	0,0%	25,0%	15,0%	5,0%	5,0%	100,0%	Keyser 2009; Allentoft
0,0%	25,0%	0,0%	12,5%	25,0%	0,0%	0,0%	0,0%	0,0%	0,0%	25,0%	12,5%	100,0%	Allentoft 2015
0,0%	0,0%	0,0%	0,0%	15,6%	0,0%	0,0%	0,0%	0,0%	0,0%	22,2%	0,0%	100,0%	Molodin 2012
0,0%	35,7%	0,0%	7,1%	14,3%	0,0%	0,0%	0,0%	0,0%	0,0%	28,6%	7,1%	100,0%	Mathieson 2015
7,7%	30,8%	0,0%	0,0%	23,1%	0,0%	0,0%	7,7%	0,0%	0,0%	15,4%	0,0%	100,0%	Lalueza-Fox 2004
0,0%	14,3%	0,0%	7,1%	0,0%	0,0%	0,0%	14,3%	0,0%	21,4%	7,1%	21,4%	100,0%	Molodin 2012
7,7%	23,1%	0,0%	0,0%	7,7%	0,0%	0,0%	7,7%	0,0%	0,0%	0,0%	0,0%	100,0%	Lalueza-Fox 2004
6,7%	13,3%	0,0%	0,0%	26,7%	6,7%	0,0%	13,3%	0,0%	0,0%	0,0%	0,0%	100,0%	Keyser 2009
0,0%	14,3%	0,0%	0,0%	14,3%	7,1%	0,0%	7,1%	7,1%	0,0%	14,3%	0,0%	100,0%	Der Sarkissian 2011
8,0%	12,0%	0,0%	4,0%	0,0%	4,0%	0,0%	0,0%	0,0%	0,0%	20,0%	0,0%	100,0%	Ricaut 2004; Piliipenko
0,0%	0,0%	0,0%	0,0%	0,0%	26,3%	15,8%	0,0%	0,0%	0,0%	0,0%	0,0%	100,1%	Xu 2008
0,0%	0,0%	0,0%	4,3%	0,0%	8,7%	2,2%	0,0%	0,0%	0,0%	6,5%	0,0%	99,9%	Keyser 2003
2,5%	35,0%	0,0%	20,0%	12,5%	0,0%	0,0%	7,5%	2,5%	2,5%	7,5%	2,5%	100,0%	Alt 2014; Vai 2015
6,2%	40,0%	1,5%	9,2%	3,1%	0,0%	0,0%	4,6%	0,0%	3,1%	12,3%	7,7%	100,0%	Krzewińska 2015, Melc
0,0%	17,6%	0,0%	11,8%	17,6%	0,0%	23,5%	0,0%	11,8%	0,0%	5,9%	0,0%	100,0%	this study
3,7%	25,9%	3,7%	0,0%	14,8%	0,0%	3,7%	7,4%	0,0%	7,4%	3,7%	0,0%	99,9%	Tömöry 2007
1,5%	32,8%	3,0%	4,5%	11,9%	0,0%	1,5%	13,4%	0,0%	0,0%	10,4%	0,0%	100,0%	Bogácsi-Szabó 2008
0,0%	21,1%	0,0%	21,1%	31,6%	0,0%	0,0%	10,3%	0,0%	5,3%	0,0%	0,0%	99,8%	Csákyová 2016
0,0%	67,0%	0,0%	3,7%	15,0%	0,0%	0,0%	3,7%	0,0%	0,0%	0,0%	3,7%	100,5%	Guimaraes 2009
0,0%	36,4%	18,2%	9,1%	0,0%	0,0%	0,0%	27,3%	0,0%	0,0%	0,0%	0,0%	100,0%	Bogácsi-Szabó 2005

**Függelék 7. táblázat** Populációk haplocsoportot eloszlását mutató táblázat, melyben 111 recens (n= 20748) és 35 ásatag (n= 1072) Eurázsia populáció van, ezeken végeztük el a populációgenetikai analíziseket. A referenciák a jobb szélső oszlopban soroltak fel.

<b>P5 index</b>	<b>P5 adapter</b>	<b>P7 index</b>	<b>P7 adapter</b>
P5_iPCR-LP-1 TCGCAGG	cctgcga	Sol_iPCR-MPI-97 AATCTTC	gaagatt
P5_iPCR-LP-2 CTCTGCA	tgcaag	Sol_iPCR-MPI-98 ACCAACG	cggtggt
P5_iPCR-LP-3 CCTAGGT	acctagg	Sol_iPCR-MPI-99 AGATGGC	gccatct
P5_iPCR-LP-4 GGATCAA	ttgatcc	Sol_iPCR-MPI-100 CCAGGTT	aacctgg
P5_iPCR-LP-5 GCAAGAT	atcttgc	Sol_iPCR-MPI-101 CCGTTAG	ctaacgg
P5_iPCR-LP-6 ATGGAGA	tctccat	Sol_iPCR-MPI-102 CGCCTCT	agaggcg
P5_iPCR-LP-7 CTCGATG	catcgag	Sol_iPCR-MPI-103 CTTGCGG	ccgcaag
P5_iPCR-LP-8 GCTCGAA	ttcgagc	Sol_iPCR-MPI-104 GGCGGAG	ctccgcc
P5_iPCR-LP-9 ACCAACT	agttggt	Sol_iPCR-MPI-105 TGGACGT	acgtcca
P5_iPCR-LP-10 CCGGTAC	gtaccgg	Sol_iPCR-MPI-106 AACCATG	catggtt
P5_1-11 AACCAAG	cttggtt	P7_1-107 AACCAAG	cttggtt
P5_13-12 AAGATTC	gaatctt	P7_15-108 AAGCCTG	caggctt
P5_28-13 ACCAACT	agttggt	P7_26-109 ACCAACT	agttggt
P5_94-14 CATCGCG	cgcgatg	P7_77-110 CAACGAA	ttcgttg
P5_100-15 CCAGCGA	tcgctgg	P7_128-111 CGTTCCT	aggaacg
P5_181-16 GGTTGAC	gtcaacc	P7_153-112 GAGGTTG	caacctc
P5_185-17 GTAGAAT	attctac	P7_184-113 GGTTGAC	gtcaacc
P5_192-18 GTTGCTC	gagcaac	P7_212-114 TGAGAGA	tctctca
P5_215-19 TGCTCTA	tagagca	P7_214-115 TGCATGA	tcatgca
P5_229-20 TTGATGG	ccatcaa	P7_224-116 TTGGATC	gatccaa
P5_2-21 AACCGCA	tgcggtt	P7_2-117 AACCGCA	tgcggtt
P5_3-22 AACCTGC	gcaggtt	P7_3-118 AACCTGC	gcaggtt
P5_4-23 AACGACC	ggtcggtt	P7_6-119 AACGTTA	taacggtt
P5_5-24 AACGCAA	ttgcggtt	P7_10-120 AAGAACG	cggtctt
P5_6-25 AACGGTT	aaccggtt	P7_14-121 AAGCAGT	actgctt

P5_10-26 AAGAACG	cggtctt	P7_29-122 ACCGCTC	gagcgg
P5_32-27 ACCGTAT	atacgg	P7_44-123 AGAACCG	cggttct
P5_62-28 AGTAGCG	cgctact	P7_45-124 AGAAGAC	gtcttct
P5_64-29 AGTTCAA	ttgaact	P7_46-125 AGAATTA	taattct
P5_66-30 ATACGAA	ttcgtat	P7_52-126 AGATAGG	cctatct
P5_67-31 ATACTGG	ccagtat	P7_54-127 AGCAGGT	acctgct
P5_79-32 CAACCGG	ccggttg	P7_63-128 ATAACGT	acgttat
P5_80-33 CAACGAT	atcgttg	P7_84-129 CAGAGCA	tgctctg
P5_97-34 CCAAGCT	agcttg	P7_91-130 CATGCTC	gagcatg
P5_98-35 CCAATAA	ttattgg	P7_93-131 CCAAGTC	gacttgg
P5_101-36 CCATCAT	atgatgg	P7_94-132 CCAATAA	ttattgg
P5_106-37 CCGGCTC	gagccgg	P7_103-133 CCGTCCG	cggacgg
P5_107-38 CCGTTCT	agaacgg	P7_109-134 CCTTAAT	attaagg
P5_109-39 CCTACGG	ccgtagg	P7_113-135 CGACGGT	attaagg
P5_126-40 CGGTCGC	gcgaccg	P7_115-136 CGATTTCG	cgaatcg
P5_139-41 CTCTTGG	ccaagag	P7_126-137 CGGTCTC	gagaccg
P5_143-42 CTTAACG	cgttaag	P7_127-138 CGTATAT	atatacg
P5_145-43 GAACGGC	gccgttc	P7_140-139 CTGGCCT	aggccag
P5_151-44 GAGTCTT	aagactc	P7_143-140 CTTGGA	ttccaag
P5_164-45 GCGCTTA	taagcgc	P7_154-141 GAGTAAC	gttactc
P5_174-46 GGCCGAG	ctcggcc	P7_155-142 GATCGTC	gacgatc
P5_176-47 GGTAGTT	aactacc	P7_156-143 GATCTCG	cgagatc
P5_179-48 GGTTAGA	tctaacc	P7_163-144 GCATATT	aatatgc
P5_183-49 GTAATGA	tcattac	P7_164-145 GCATTGG	ccaatgc
P5_190-50 GTTATTG	caataac	P7_167-146 GCCTACG	cgtaggc

**Függelék 8. táblázat** Index kombinációk



OLIGO	CHROM	START	END
ASIP-PIGU	20	33218030	33218150
CTS11043	Y	22914919	22915039
CTS1211	Y	7271662	7271782
CTS4385	Y	15625939	15626059
DF27	Y	21380140	21380260
EXOC2	6	457688	457807
F1096	Y	8087513	8087633
F1206	Y	8440357	8440477
F1329	Y	8588971	8589091
F2930	Y	14108284	14108404
F3393	Y	23023914	23024034
F650	Y	23616946	23617066
F75	Y	7409949	7410069
HERC2	15	28365558	28365678
IRF4	6	396261	396381
K29	Y	6630393	6630513
KITLG	12	89328275	89328394
K-ras-probe-focus-G12	12	25398345	25398226
L1034	Y	21646338	21646458
L1259	Y	15615280	15615400
L131	Y	19372748	19372868
L1324	Y	8483601	8483721
L156	Y	17174681	17174801
L183	Y	8467076	8467196
L222	Y	16809418	16809538
L223	Y	6345088	6345208
L228	Y	7771298	7771418
L22	Y	8575949	8576069
L23	Y	6753451	6753571
L26	Y	22942837	22942957
L275	Y	19136828	19136948
L293	Y	9984962	9985082
L30	Y	15604839	15604959
L389	Y	28733041	28733161
L405	Y	28731857	28731977
L415	Y	2888603	2888723
L460	Y	7879355	7879475
L595	Y	21739591	21739711
L596	Y	14197571	14197691
L621	Y	18760021	18760141
L666	Y	7570756	7570876
L68	Y	18700090	18700210
L729	Y	19431548	19431668
L732	Y	18146925	18147045
L780	Y	21183583	21183703
L80	Y	14640655	14640775
L811	Y	21364648	21364768
Lactose-intolerance-1	2	136608586	136608705
Lactose-intolerance-2	2	136616694	136616813
LLY22g1_1	Y	23714589	23714709
M106	Y	21866364	21866484
M119	Y	21762625	21762745
M122	Y	21764614	21764734
M12	Y	7583420	7583540
M130	Y	2734794	2734914
M132	Y	21896201	21896321
M145	Y	21717148	21717268
M15	Y	21724639	21724759
M168	Y	14813931	14814051
M170	Y	14847732	14847852
M172	Y	14969574	14969694
M173	Y	15026364	15026484
M174	Y	14954220	14954340
M175	Y	15508646	15508766
M178	Y	21741695	21741815
M179	Y	14838640	14838760
M17	Y	21733108	21733228
M184	Y	14898103	14898223
M198	Y	15030692	15030812

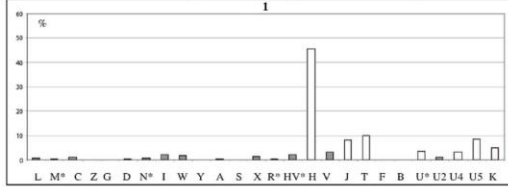
M201	Y	15027469	15027589
M203	Y	15591477	15591597
M207	Y	15581923	15582043
M20	Y	21733394	21733514
M214	Y	15471865	15471985
M215	Y	15467764	15467884
M216	Y	15437504	15437624
M217	Y	15437273	15437393
M2308	Y	7690122	7690242
M230	Y	15481312	15481432
M231	Y	15469664	15469784
M241	Y	15018399	15018519
M242	Y	15018522	15018642
M253	Y	15022647	15022767
M258	Y	15023304	15023424
M25	Y	21866604	21866724
M267	Y	22741758	22741878
M268	Y	22739241	22739361
M269	Y	22739307	22739427
M282	Y	21764371	21764491
M285	Y	22741680	22741800
M3035	Y	21940256	21940376
M304	Y	22749793	22749913
M306	Y	22750523	22750643
M317	Y	22754611	22754731
M335	Y	15026335	15026455
M33	Y	21740390	21740510
M342	Y	23243882	23244002
M343	Y	2887764	2887884
M346	Y	2887096	2887216
M347	Y	2877419	2877539
M353	Y	17309207	17309327
M356	Y	2888143	2888263
M3684	Y	14367121	14367241
M3727	Y	18776569	18776689
M377	Y	15027373	15027493
M38	Y	21742098	21742218
M40	Y	2663883	2664003
M410	Y	2751618	2751738
M412	Y	8502176	8502296
M415	Y	9170485	9170605
M417	Y	8533675	8533795
M420	Y	23473141	23473261
M423	Y	19096031	19096151
M427	Y	19092537	19092657
M429	Y	14031274	14031394
M438	Y	16638744	16638864
M448	Y	16520384	16520504
M458	Y	24366404	24366524
M45	Y	21867727	21867847
M46	Y	14922523	14922643
M479	Y	20834607	20834727
M497	Y	17419874	17419994
M4	Y	2744568	2744688
M522	Y	7173083	7173203
M526	Y	23550864	23550984
M529	Y	15654368	15654488
M530	Y	14286468	14286588
M55	Y	21872678	21872798
M578	Y	7202643	7202763
M69	Y	21893998	21894118
M73	Y	21888814	21888935
M74	Y	21889707	21889827
M75	Y	21890117	21890237
M89	Y	21917253	21917373
M96	Y	21778938	21779058
M9	Y	21730197	21730317
MCIR-hair-10	16	89985880	89985999
MCIR-hair-11	16	89986070	89986190
MCIR-hair-1	16	90052113	90052233

MC1R-hair-2	16	89986031	89986151
MC1R-hair-3	16	89986094	89986213
MC1R-hair-4	16	89986084	89986204
MC1R-hair-5	16	89985784	89985903
MC1R-hair-6	16	89985858	89985978
MC1R-hair-7	16	89986057	89986176
MC1R-hair-8	16	89986486	89986605
MC1R-hair-9	16	90052470	90052590
MEH2	Y	4925577	4925697
nme1-Nm23-H1-probe-focus	17	49239045	49239164
OCA2	15	28230258	28230378
P117	Y	14819633	14819753
P126	Y	21225710	21225830
P131	Y	15472803	15472923
P143	Y	14197807	14197927
P147	Y	21083360	21083480
P15	Y	23243966	23244086
P177	Y	14159786	14159906
P179	Y	14060248	14060368
P186	Y	7568508	7568628
P189	Y	14197917	14198037
P209	Y	19179275	19179395
P256	Y	8685170	8685290
P257	Y	14432868	14432988
P287	Y	22072037	22072157
P295	Y	7962971	7963091
P297	Y	18656448	18656568
P2	Y	21610771	21610891
P312	Y	22157251	22157371
P315	Y	6740005	6740125
P326	Y	8467230	8467350
P331	Y	16222501	16222621
P336	Y	9004861	9004981
P36	Y	14496381	14496501
P397	Y	28759585	28759705
P399	Y	13563023	13563143
P405	Y	8391048	8391168
P43	Y	21880913	21881033
P44	Y	14495191	14495311
p53-probe-1-focus-R282	17	7577154	7577035
p53-probe-2-focus-Y220	17	7578251	7578132
P60	Y	14484411	14484531
P79	Y	6740317	6740437
P91	Y	14850539	14850659
P99	Y	14886326	14886446
PF2824	Y	5253801	5253921
PF3146	Y	5688072	5688192
Retinoblastoma-exon-10	13	48941620	48941739
Retinoblastoma-exon-12	13	48947531	48947650
Retinoblastoma-exon-13	13	48951044	48951163
Retinoblastoma-exon-16	13	48954266	48954385
Retinoblastoma-exon-17	13	48955373	48955492
Retinoblastoma-exon-18	13	49027119	49027237
Retinoblastoma-exon-19	13	49030330	49030449
Retinoblastoma-exon-1	13	48877920	48878039
Retinoblastoma-exon-20	13	49033814	49033933
Retinoblastoma-exon-21	13	49037857	49037976
Retinoblastoma-exon-22	13	49039124	49039243
Retinoblastoma-exon-23	13	49039331	49039450
Retinoblastoma-exon-25	13	49050827	49050946
Retinoblastoma-exon-27	13	49054124	49054243
Retinoblastoma-exon-2	13	48881406	48881525
Retinoblastoma-exon-3	13	48916726	48916845
Retinoblastoma-exon-4	13	48919206	48919325
Retinoblastoma-exon-7	13	48934143	48934261
Retinoblastoma-exon-8	13	48936941	48937060
S17250	Y	15531294	15531414
SLC24A4	14	92773603	92773723
SLC24A4-hair	14	92801143	92801262
SLC24A5	15	48426424	48426544

SLC45A2-hair	5	33958899	33959018
SLC45A2-MATP	5	33951633	33951753
SNPforID-10	21	17710364	17710483
SNPforID-12	10	34755288	34755408
SNPforID-13	22	47836352	47836471
SNPforID-14	12	56603774	56603893
SNPforID-15	14	36170547	36170666
SNPforID-16	2	136616694	136616813
SNPforID-17	1	36768140	36768259
SNPforID-18	2	7149095	7149214
SNPforID-19	1	204790917	204791035
SNPforID-1	17	75551607	75551727
SNPforID-20	21	25672400	25672519
SNPforID-21	19	42410271	42410390
SNPforID-22	13	34864180	34864299
SNPforID-23	16	90105273	90105392
SNPforID-24	16	31079311	31079430
SNPforID-25	3	168644975	168645094
SNPforID-26	4	159181903	159182022
SNPforID-27	15	48426424	48426544
SNPforID-28	4	38803195	38803314
SNPforID-2	22	26350043	26350162
SNPforID-30	13	20901664	20901783
SNPforID-31	13	22374640	22374759
SNPforID-32	14	101142830	101142949
SNPforID-33	11	32424329	32424448
SNPforID-34	6	165045274	165045393
SNPforID-3	20	38849582	38849701
SNPforID-4	1	159174623	159174743
SNPforID-5	7	4456943	4457062
SNPforID-6	18	75432327	75432446
SNPforID-7	10	17064932	17065051
SNPforID-8	21	16685538	16685657
SNPforID-9	12	29369811	29369930
TYR	11	89010986	89011106
TYR-hair	11	88911636	88911755
TYRP1	9	12709245	12709364
U106	Y	8796018	8796138
U152	Y	15333089	15333209
U5	Y	23973534	23973654
U8	Y	14692167	14692287
V13	Y	6842203	6842323
V38	Y	6818231	6818351
V68	Y	17664711	17664831
V88	Y	4862801	4862921
VL29	Y	14570364	14570484
Z1936	Y	21463266	21463386
Z2124	Y	4521618	4521738
Z283	Y	21976243	21976363
Z284	Y	8717136	8717256
Z5857	Y	2759225	2759345
Z645	Y	8244985	8245105
Z647	Y	7682998	7683118
Z91	Y	16474733	16474853
Z93	Y	7552296	7552416

**Függelék 9. táblázat** Genomi RNS dúsitó kit SNP listája, kromoszómán lévő pozíciójával

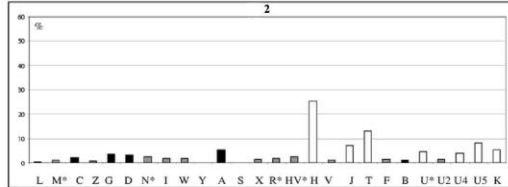
**Cluster 1:** Pure Western Hgs with a strong dominance of Hg H.



**Members:**

NOR, ICE, IRE, SCO, ORK, ENG, GEN, GES, FRA, AT'S, CZH, SLK, RUS, MRD, EST, LAT, LIT, UKR, HUB, TUS, GAL, CRO, POR, ROM, BLG, GRE, ARO, MAC, SLN, ALB, BOS, *ENO, IBN, ITM, VIK* and CUM as outlier.

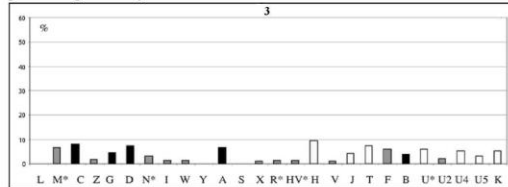
**Cluster 2:** Dominance of the Western CHgC with significantly less dominance of H, and appearance of some Siberian CHgC.



**Members:**

UDM, GEO, TAT, SEK, CSG, KOP, TUR, SYR, LBN, KOZ, *KAR, SLV, LBT, AH2, EMN, KBK, YAM, VAT, AH1, LNB, LOM, SCI, SRU, KIK, ABA.*

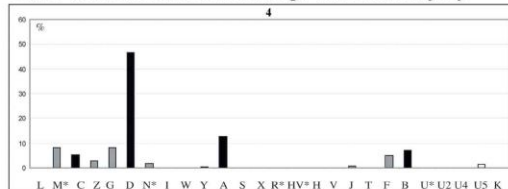
**Cluster 3:** Balanced composite of Western and Siberian CHgCs, dominance of H disappears. Hgs not belonging to either groups are also present significantly.



**Members:**

CHN, HUI, ALT, MNG, KAZ, UYG, BAS, KMS, KET, KUR, *QIN, BB3, MEN, TAG, SIA, STR, SZA,* and CTC as outlier.

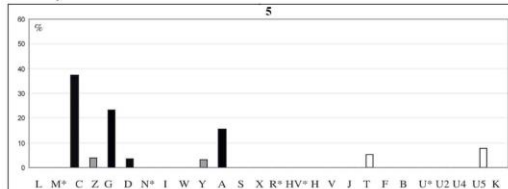
**Cluster 4:** Pure dominance of Siberian Hgs, with D and A majority.



**Members:**

JPN, KOR, ALE, *SER, XIO.*

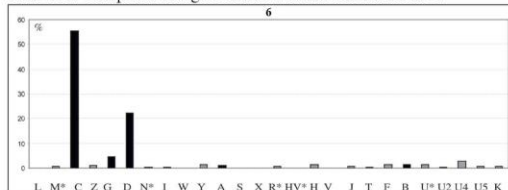
**Cluster 5:** Dominance of Siberian Hgs C, A and G (Latter is a definite member of the recent Siberian CHgC, but misses from the ancient version) with some Western T and U.



**Members:**

KYK, CVN, *BB2.*

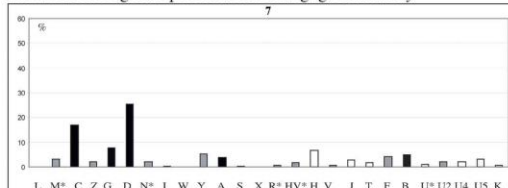
**Cluster 6:** Composite of Hgs C and D with the dominance of C.



**Members:**

YUK, YAK, EVE, TUV, EVW, NGA, *XIA.*

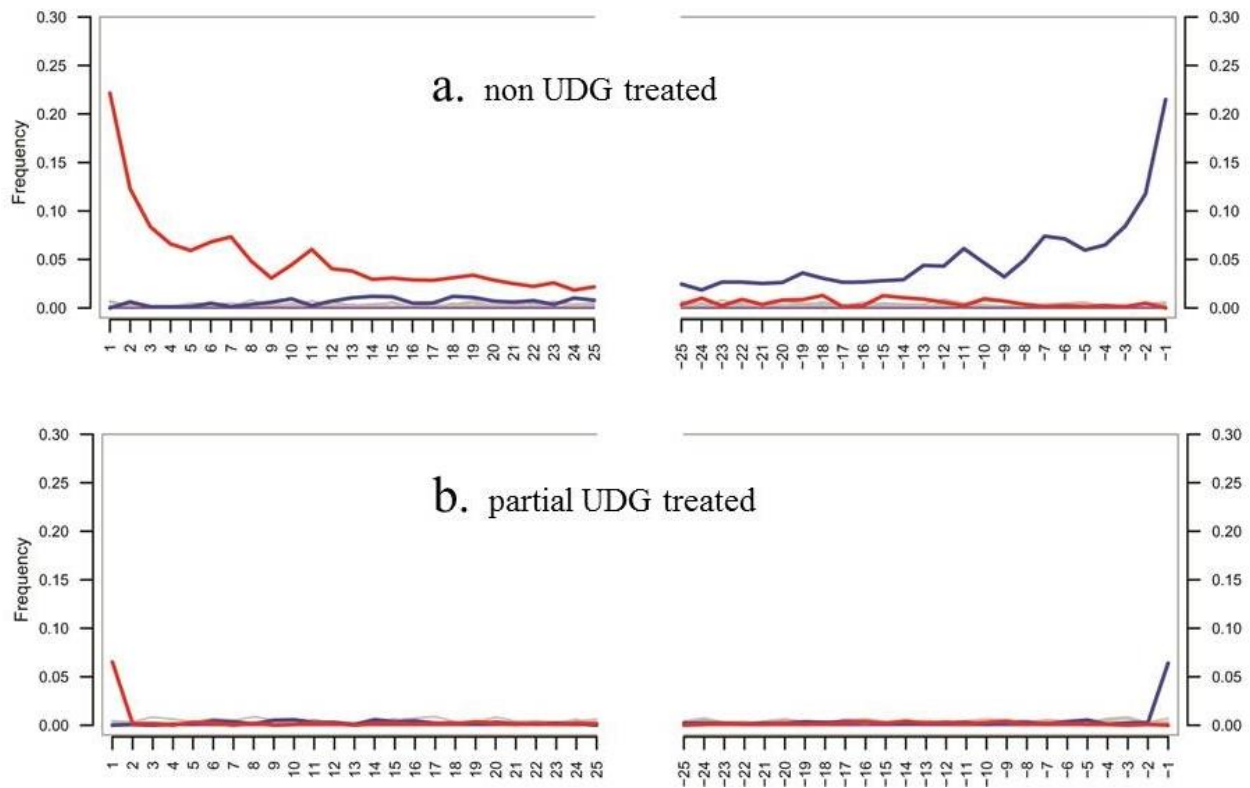
**Cluster 7:** Composite of Hgs C and D with the dominance of D, but A, B and Western CHgC are present as a non-negligible minority.



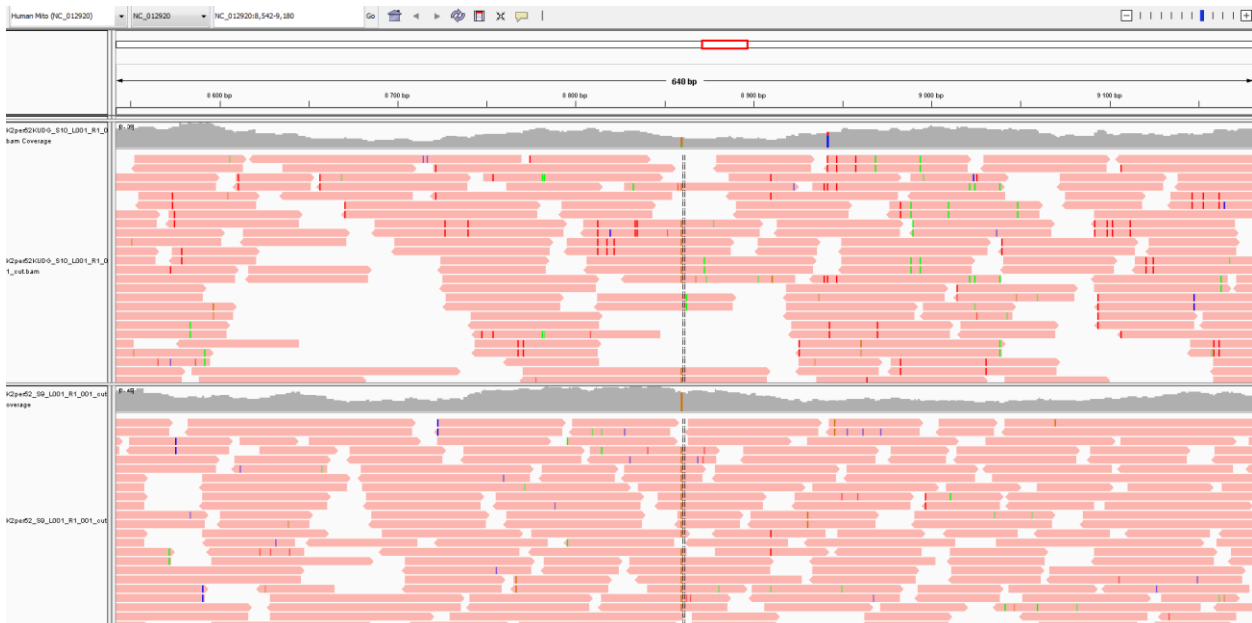
**Members:**

BAR, BUR, KMN, KMY, TLG, KYG, KKS, TLT, MAN, *PAZ, BBI.* Outliers are NGD and ULC.

**Függelék 1. ábra** SOC algoritmus által készített 7 kluszer, melyek a haplocsoport eloszlások alapján készültek.



**Függelék 2. ábra** Archaikus DNS jellemző sérülés mintázat a MapDamage 2.0 program segítségével készítve. a. nem UDG kezel könyvtár. b. Részleges UDG kezelt könyvtár ugyanazon kivonatból.



**Függelék 3. ábra** Az ábrán ugyanabban a kivonatból készült két könyvtár szekvencia olvasatainak a mitokondrium genomra illesztett egy részletét látjuk. A felső ábrán a könyvtár készítés első lépését kihagytuk, tehát nem végeztük el a részleges UDG kezelést, az alsó ábrán részleges UDG kezelést végeztünk. Jól látható a felső ábrán mennyivel több piros (Timin) és zöld (Adenin) vonal látható. Ezeket nagyrészt kijavítottuk az részleges UDG kezelés folyamán.