

III. BÉLA KIRÁLY CSONTJAINAK TANÚSÁGA – AZ ÁRPÁD-HÁZ EREDETE

Bernert Zsolt¹, Fehér Tibor¹, Varga Dániel², Székely Gábor¹ és Németh Endre¹

¹Magyar Természettudományi Múzeum, Embertani Tár, Budapest, ²MTA Rényi Alfréd Matematikai Kutatóintézet, Budapest

Bernert Zs., Fehér T., Varga D., Székely G., Németh E.: Testimony from the Bones of Hungarian King Béla III – Origins of the Árpád Dynasty. *The role of applications based on machine learning is continuously growing in the industry, health sector/bioinformatics and scientific research. American researchers published a bit more than 10 years ago the first machine learning algorithms, which were able to safely predict Y-SNP based haplogroups from Y-STR data. The goal of the present study was to predict with machine learning algorithms the SNP-based subgroup of three ancient DNA samples (King Béla III and two Khazar samples) belonging to Y-DNA Haplogroup R1a, in order to predict their geographic origin and mutual genetic relatedness more accurately. This is the first study applying machine learning algorithms for researching Hungarian prehistory.*

Based on the Y-STR haplotype of King Béla III, we estimated with the machine learning algorithm in the first step that he belonged to the R1a-Z93 subgroup that is most common among Indo-Iranic and Turkic speaking peoples. The second step predicted that King Béla III belonged to the Z2123 subgroup of R1a-Z93. The Phylogenetic analysis showed King Béla III most likely belonged to the relatively rare YP451+ YP449- subgroup of Z2123, which practically only appears in the North Caucasus, especially among Karachays and Balkars.

Based on our results, we could hypothetically conclude that the Árpád Dynasty has common origin with one ethnic component of the Karachay people.

In our study we proved that it is possible to increase the accuracy of Y-DNA haplogroup prediction of historical aDNA samples with mathematical methods using contemporary Y-STR haplotypes. With the help of this method, larger historical aDNA studies could save a lot of research funds and DNA carrying out tailored deep SNP-testing of samples instead of using general SNaPshots.

Keywords: *Physical anthropology; Archaeogenetics; Machine learning; Y-SNP prediction; Y-STR; Gradient boosting; Árpád Dynasty; King Béla III; Hungarian ethnogenesis.*

Bevezetés

A magyar királyok, királynők, főemberek és családtagjaik legfontosabb temetkezési helye évszázadokig a székesfehérvári királyi bazilika volt, amelynek szervezett kirablása 1543-ban, a város török megszállásával kezdődött el. A mintegy másfél évszázados török hódoltság alatt az épület romba dőlt, kőanyagát fehérvári építkezéseken használták fel.

1848. december 5-én földmunkák során két, vörösmészke-lapokból összeállított sírkamrára bukkantak, amelyekben III. Béla király és első felesége, Anna királynő csontváza nyugodott. Ma a Budavári Nagyboldogasszony Templomban található maradványaik, amelyek az egyetlen hiteles és természettudományos módszerekkel vizsgált képviselői az egykori Árpád-háznak.

Álmos fejedelem és leszármazottai meghatározó szerepet töltek be a korai magyar történelemben. A honfoglalás sikere, a keresztény magyar állam megalapítása és első 300 éve elválaszthatatlanul összefonódik a dinasztiaival, amelyet – jóval később – Álmos fiáról Árpád-háznak neveztek el. Érthető, hogy az antropológusok és újabban a genetikával foglalkozó szakemberek egyaránt nagy figyelmet fordítottak – I. László koponyaereklyéje mellett – az egyetlen hiteles Árpád-házi király maradványainak vizsgálatára. A közelmúltban publikálta egy nemzetközi kutatócsoport III. Béla Y-STR haplotípusát és Y-SNP haplocsoportját (Olasz és mtsai 2018). A jelen tanulmány célja III. Béla alcsoportjának meghatározása a publikált haplotípus alapján a gépi tanulás és filogenetikai elemzések eszköztárával.

Kutatási előzmények

Az 1172 és 1196 között uralkodó III. Béla király csontmaradványait a megtalálásuk óta eltelt 170 évben számos alkalommal vizsgálták jeles orvosok, antropológusok és egyéb szakértők, köztük az első magyar antropológus professzor Török Aurél (Török 1894). A csontokról másolat, koponyájáról arcreekonstrukció is készült (Skultéty 2008). A maradványokat elemző kutatók több száz oldal terjedelmű írásai a következőképpen foglalhatók össze.

III. Béla király testmagassága meghaladta a 190 centimétert. Agykoponyája hosszú és széles, ez alacsony koponyaboltozattal és a szemöldökív mögött beszűkülő homlokcsonttal párosul, utóbbiak a mongoloid koponyák jellegzetességei. Az arckoponya magas és széles, az orr keskeny (1. ábra). A nagy, kerek szemüreg és a felső állcsont elődomborodó orrcsonti nyúlványa szintén a mongoloid nagyrassz örökségeként értékelhető. Számos ízületi kopás, a bal hüvelykujj törése és enyhe porckorongsérv árukodik a király fizikai aktivitásáról. A sérülések közül el kell különíteni a koponyatetőn megfigyelhető 8–9 milliméter átmérőjű jelképes trepanációt, amely szakrális célú szándékos beavatkozás volt (Éry és mtsai 2008). A jelképes koponyalékelés a felnőtt férfiak és nők között a honfoglaló magyarság körében mintegy 13% gyakorisággal fordult elő (Bernert és mtsai 2006). Az Árpád-korban a csontmaradványok tanúsága szerint a szokás lassan eltűnik.

III. Béla király archeogenetikai vizsgálatának eredményeit egy nemzetközi kutatócsoport publikálta (Olasz és mtsai 2018). A friss kutatási adatokból egyelőre annyit tudhatunk meg, hogy III. Béla király a Skandináviától Indiáig terjedő hatalmas földrajzi sávban kifejezetten gyakori R1a haplocsoportba tartozik (Underhill és mtsai 2015). Ez genetikai szempontból fontos és nem is meglepő eredmény. Az Árpádok apai vonalú eredetének – történelmileg is értelmezhető – feltérképezésében nem lehet megkerülni az R1a haplocsoport alcsoportokra finomítását. Az alcsoportok esetében általában már van elegendő adat ahhoz, hogy megállapítsuk azok mely kisebb régiókra, népcsoportokra jellemzőek (Underhill és mtsai 2015).

A jelen tanulmányban tehát egyetlen, bár a magyar őstörténet szempontjából kiemelten fontos család apai vonalának eredetével kapcsolatos kérdésekre keresünk válaszokat. A kérdéseket és a válaszokat egyaránt a DNS alapú genetika nyelvén fogalmazzuk meg. Témánk korlátainak megfelelően nem is törekszünk a magyar és más a magyarság kialakulásában szerepet játszó etnikai csoportok kialakulásával foglalkozó genetika tanulmányok teljes körű bemutatására. Mindenesetre felvillantunk néhány fontosabb populációgenetikai eredményt az elmúlt évtizedből. Kétségtelenül a

legizgalmasabb eredmények nagy, nemzetközi projektekből születtek. Régóta megoldatlan vagy nyitott társadalomtudományi kérdések megválaszolása érdekében fogták vállat az emberi génállományt ezekben a különösen izgalmas kutatásokban. Így a korai indoeurópaiak, az uráliak és török népek genetikai állományának feltérképezésére, és ezen keresztül az adott csoport legvalószínűbb őshazájának lokalizálására tettek kísérletet a kutatók.



1. ábra: III. Béla király (Török Aurél rajza)
Fig. 1: King Béla III (made by Aurél Török)

Az archeogenetikai kutatások alapján az indoeurópai népek őshazája valószínűleg a kelet-európai sztyeppéken (lényegében az Urál és a Dnyeszter folyók közti területen) lehetett, a késő rézkori, kora bronzkori Jamnaja-kultúrában, ahonnan egy intenzív népességmozgás mutatható ki Közép-Európába és Nyugat-Ázsiába (Allentoft és mtsai 2015, Haak és mtsai 2015).

A török nyelvek egyértelműen nyelvcserevel vagy domináns elitek által terjedtek el az Anatóliától Közép-Ázsián át Jakutföldig tartó hatalmas földrajzi térségben, mivel oly mértékben különbözött a mai török nyelvet beszélő csoportok génállománya (Yunusbayev és mtsai 2015). Autoszomális markerek segítségével mégis sikerült bebizonyítani, hogy népességmozgás is kísérte a török nyelvek terjedését, és a török nyelvű népek őshazája vélhetően a ma mongolok, burjátok, tuvaiak által lakott területen lehetett (Yunusbayev és mtsai 2015). Magyar szempontból is érdekes eredmény ez, mert két, a honfoglalókat vizsgáló tanulmány is rámutatott, hogy a honfoglalókat egy egyértelmű demográfiai impulzus érte pontosan erről a belső-ázsiai területről (Németh és mtsai 2017, Neparáczki és mtsai 2018). Mindez tökéletesen összhangban van a magyar nyelvben kimutatható erőteljes török hatással (Róna-Tas és Berta 2011). Új eredmény az is, hogy a legújabb eredmények megerősítik a korábbi, embertannal foglalkozó kutatók észrevételeit (Tóth 1965), miszerint a honfoglalók között valószínűleg a Kaukázus vidékéről érkezett csoportokat is találhatunk (Neparáczki és mtsai 2018).

Érdekes anomália jellemezte a recens uráli népek genetikai állományát. Míg az apai vonalakban az uráli népeket összekapcsolják az N haplocsoport különböző alcsoportjai, addig anyai vonalaikban az uráli népek távol állnak egymástól, és inkább a szomszédos

népekre hasonlítanak (Tambets és mtsai 2018). Ezért fontos eredmény, hogy az uráli népeket – a magyar kivételével – autoszomális markerek is összekapcsolják (Tambets és mtsai 2018). (A magyar minta egy 100-as budapesti minta volt.) Az azonosított autoszomális marker földrajzi előfordulása alapján az uráliak őshazája Szibériában lehetett (Tambets és mtsai 2018). Amennyiben figyelembe vesszük, hogy az uráliak migrációjával kapcsolatba hozott, apai vonalú N alcsoportok migrációja alapvetően kelet-nyugati irányban zajlott (Ilumae és mtsai 2016), és a vonatkozó autoszomális marker legkeletibb előfordulási csúcsa az Ob és Jenyiszej közötti területen található (Tambets és mtsai 2018), az Ob és Jenyiszej közötti területet sem lehet kizárni a potenciális uráli őshazák sorából. Említésre méltó párhuzam a genetikai és nyelvi alapon történő kutatások között, hogy Juha Janhunnen szintén az Ob és Jenyiszej közötti területet tartotta az uráli népek őshazájának nyelvészeti megfontolások alapján (Janhunnen 2014).

Az első, vélhetően finnugor népességeket vizsgáló archeogenetikai kutatás, a 3500 évvel ezelőtti Kola-félszigeten (Oroszország, Murmanszki terület) élt népesség (Bol'shoy Oleni Ostrov leletek) anyai vonalainak vizsgálata során, azt találta, hogy a népességben jelentős arányban található szibériai eredetű népesség. A Kola-félszigeti archív minták mtDNS haplotípus egyezés tekintetében a mai tuvaiakkal és azon belül a C haplocsoportban mutatják a legtöbb haplotípus egyezést (Sarkassian és mtsai 2013). Az újabb kutatások tovább pontosítják a képet a Kola-félszigeti archív mintákkal kapcsolatban (Lamnidis és mtsai 2018). Egyelőre 3 mintából sikerült elegendő DNS-t kinyerni az Y haplocsoportok meghatározásához, és háromból két minta tartozik az N-Tat haplocsoporthoz (Lamnidis és mtsai 2018). Az autoszomális vizsgálatok azt jelezték, hogy a 3500 évvel ezelőtt élt közösség tagjai átmenetet képeztek a mai nyugat-európai és kelet-ázsiai közösségek között, és legközelebb hozzájuk mai manysik, utána szölkupok és jukagírok egyes csoportjai álltak (Lamnidis és mtsai 2018). A tanulmány egyik fő állítása – egyelőre alacsony mintaszám mellett, hogy a 3500 évnél korábbi, baltikumi és fennoskandináv mintákban nem találtak szibériai komponenst és a későbbi mintákban pedig alacsonyabb volt az aránya. Úgy tűnik, hogy egy fontos, Szibériából érkező demográfiai impulzust sikerült azonosítani a kutatóknak, amely a mai számik (lappok) elődeihez köthető valószínűleg. Azt is valószínűsítik a szerzők, hogy a mai elterjedési területüknél jóval délebbre is éltek, és a délebben élő számi népesség egy balti-finn nyelvet beszélő közösséggel kölcsönhatásba lépve nyelvet váltott (Lamnidis és mtsai 2018). Mindez tökéletesen összhangban lenne a Baltikumra jellemző apai vonalú eredményekkel. Az észtek között az N haplocsoporton belül az N-VL29 alcsoport a domináns, a finnek és számik között azonban az N-Z1936 alcsoport és az N-VL29 alcsoport vegyesen fordul elő. Komoly következményekkel járó észrevétel még a kutatóktól, hogy a Kola-félszigeti archív mintákban szibériai és neolitikus európaiak genetikai öröksége is kimutatható, amelyek keveredését 3977(+/-77) évvel ezelőtre teszik a tanulmány szerzői (Lamnidis és mtsai 2018).

Térjünk most rá a kifejezetten magyar vonatkozású kutatásokra. Komoly előrelépés várható annak kapcsán, hogy elindultak az Urál vidéki „magyargyanús” temetők molekuláris biológiai vizsgálatai (Szeifert és mtsai 2018). A tanulmány egyik legizgalmasabb eredménye, hogy a tatárföldi Bolsije Tigani melletti „magyargyanús” temető (7–10. század) és a szintén tatárföldi, „magyargyanús” Csijalikszej-kultúrából (10–13. század) származó egyes minták között közeli rokoni szálak mutathatók ki (Szeifert és mtsai 2018). Ez az első populációgenetikai alapú tanulmány, amely közvetlen

bizonyítékokat szolgáltat az Urál vidéki „magyargyanús” temetők népességének tovább élésére a régióban.

Mindenesetre a honfoglalókat vizsgáló összes újabb, anyai vonalakat vizsgáló tanulmány egyöntetűen azt jelzi, hogy a honfoglalók legszorosabb recens kapcsolatai, Urál-vidéken élő népekhez vezetnek. Azon belül is két földrajzilag szomszédos és nyelvileg is rokon kpcsak népesség, a tatárok és baskírok között mutatható ki a legközvetlenebb kapcsolat a honfoglalókkal (Csósz és mtsai 2016, Németh és mtsai 2017, Neparáczi és mtsai 2018). Egy hajszálnyi különbség mégis van az eredmények között. Két tanulmány a tatárok és a honfoglalók (Csósz és mtsai 2016, Neparáczi és mtsai 2018), a harmadik a baskírok és honfoglalók közötti kapcsolatot látja szorosabbnak (Németh és mtsai 2017). Az eltérések mögött valószínűleg az eltérő matematikai módszerek, az eltérő baskír, tatár és honfoglaló adatbázisok állnak.

Új fejezetet nyithat a magyar őstörténeti kutatásokban, hogy egyes nemzetközi kutatásokban is megjelennek a magyar őstörténet számára fontos eredmények. Így például az, hogy elsősorban a baskírokban, másodsorban a tatárokban nem várt mértékben, recens finnugor népességekhez köthető komponens mutatható ki. A Bajkál-tó és a Balti-tenger közötti migrációs folyamatok autoszomális markerek alapján történő elemzésekor az derült ki, hogy a hantik és a baskírok között egy meglepően erős genetikai kapcsolat mutatható ki (Triska és mtsai 2017). A tanulmány szerzői – történeti okokból kiindulva – a baskírokba olvadt magyar népességgel magyarázzák a jelenséget (Triska és mtsai 2017). Az apai vonalú N haplocsoport alcsoportjainak Euráziában való elterjedését és gyakoriságát vizsgáló kutatások kiemelik, hogy a finneken, lappokban, északi oroszokban, obi-ugokban jellemző N-Z1936 alcsoportnak van egy váratlan gyakoriság csúcsa elsősorban a baskírok, másodsorban a tatárok között (Ilumae és mtsai 2016). Közép-Európában az N-Z1936 alcsoport előfordulása szórványos, de az a Kárpát-medence magyar etnikumú népességeire korlátozódik, és gyakorisági csúcsát a székelyek között éri el (Fehér és mtsai 2015). A székelyekben az N-Z1936 alcsoport egy alcsoportjának az N-L1034-nek az előfordulási aránya 4% (Fehér és mtsai 2015). Úgy tűnik, hogy a régészeti (Boldog és mtsai 2015) és történeti (Klima 2016) adatok mellett a populációgenetika is egyre több érvet szolgáltat amellett, hogy az Urál vidék egy fontos állomás lehetett a magyarság korai történetében.

Az Urál vidéki „magyargyanús” temetők kapcsán érdemes még megemlíteni, hogy bár az egyes temetők mtDNS haplocsoport-eloszlása viszonylag eltérő, az egyik legnagyobb Urálon túli „magyargyanús” temetőben, Uelgiben, 20 minta anyai vonalának feldolgozása után a C haplocsoport bizonyult a legnépesebb kelet-ázsiai eredetű haplocsoportnak (Szeifert és mtsai 2018). Ez azért is érdekes, mert a C haplocsoport a recens székely (Brandstätter és mtsai 2007), a magyar (Neparáczi és mtsai 2018), baskír (Bermisheva és mtsai 2002) és obi-ugor (Pimenoff és mtsai 2008) mintákban is a relatíve leggyakoribb kelet-ázsiai eredetű haplocsoport. Egyfajta anomália azonban az, hogy az egyik legújabb kutatás szerint egyes honfoglaló temetőkben mégis szinte marginális az előfordulása a C haplocsoportnak (Neparáczi és mtsai 2018).

Bár az apai vonalak legalább olyan informatívak biológiai (az Y-DNS több tízmillió nukleotidpárból, az mtDNS 16569 nukleotidpárból épül föl) és társadalomtudományi szempontból, az apai vonalak vizsgálata egyelőre háttérbe szorult a magyar őstörténeti kutatásokban. Így alapvetően a recens adatokból vagyunk kénytelen kiindulni, ha a magyar etnogenezist az apai vonalak mentén próbáljuk meg visszafejteni.

Megállapítható, hogy a mai magyar népesség apai vonalainak összetétele nem különbözik lényegileg a szomszédos népektől (Völgyi és mtsai 2009), de néhány meglehetősen ritka és speciális komponens – elsősorban a székelyekben – mégis összeköti a magyarságot néhány távoli, szibériai és közép-ázsiai népcsoporttal (Bíró és mtsai 2015). Így az N-L1034 az obi-ugorokkal (Fehér és mtsai 2015), a R1a-Z93 a türkökkel vagy indo-irániakkal (Underhill és mtsai 2015) és a Q-L713 alcsoport pedig üzbégekkel és tien-shani hunokkal (Damgard és mtsai 2018, <https://www.yfull.com/tree/Q-L713/>).

Említésre méltó, hogy a Torgaj vidékén, alig 200 km-re az Urálon túli „magyargyanús” régészeti leletek helyszínétől él az argün közösség keretein belül egy magyar önelnevezéssel rendelkező etnikai csoport. Az apai vonalak vizsgálata azt jelezte, hogy a mai magyarokból és az Urálon túli magyar közösségből vett minták jelentősen eltérnek egymástól. A magyar minták 87%-át a magyarokból teljesen hiányzó, egész Euráziában meglehetősen ritka, a kazahokon belül argün migrációs markernek bizonyult G1 haplocsoport alkotta (Balanovsky és mtsai 2015). A magyar és a magyar csoport között egyetlen haplotípus egyezés volt, és ez a két minta egyaránt az N-Tat haplocsoportba tartozott (Bíró és mtsai 2009). Tehát amennyiben a magyar-magyar kapcsolatok valós történelmi alapon nyugszanak, akkor az valószínűsíthető, hogy a magyarok és magyarok őseinek elválásakor az uráli migrációs marker, az N-Tat haplocsoport előfordult, és viszonylag gyakori is lehetett mindkét népességben. Ez egyben azt is jelenti, hogy mindkét népesség jelentősen átalakult a feltételezett elválás óta demográfiai értelemben.

Mintavétel és módszertan

Olasz Juditék genetikai tanulmányukban nemcsak a haplocsoportot (Y-SNP), hanem III. Béla király Y-STR adatait is publikálták (Olasz és mtsai 2018). Mivel biológiai okokból összefüggés van az egy linkage-ben öröklődő alcsoport (Y-SNP) és az Y-STR adatok között, matematikai algoritmusok alkalmazásával – az Y-STR haplotípus alapján – megbecsülhető (Athey 2006), hogy III. Béla király melyik alcsoportba tartozik legnagyobb valószínűséggel. Ezzel az eljárással pontosabban lokalizálható az Árpád-ház apai ági eredete.

Annak eldöntése, hogy III. Béla király melyik R1a alcsoportba tartozhat a haplotípusa alapján, jellegzetes osztályozási feladat. A gyakorlott szakemberek az Y-STR haplotípus egy vagy több lókuszra jellemző STR értéke alapján gyakran igen jó eséllyel azonosítják vizsgált minta haplocsoportját vagy alcsoportját tapasztalataik alapján. III. Béla haplotípusa azonban nem ilyen. További mérések vagy matematikai és számítógépes eszköztár nélkül III. Béla esetében a tapasztalat sem segít jobban behatárolni, hogy az R1a haplocsoport mely alcsoportjába tartozik. Az ilyen „nehéz” osztályozási feladatok megoldására kézenfekvő a tanuló algoritmusok választása. Amerikai kutatók bizonyították be elsőként, hogy tanuló algoritmusokkal megbízhatóan jósolható meg haplocsoport Y-STR haplotípus alapján (Schlecht és mtsai 2008).

Tanuló algoritmusokat általában komplex adathalmazokon értelmezett objektumokon alkalmaznak a kutatók. Az ilyen komplex objektumok között azoknak az összetett összefüggéseknek a kiszámítása, amelyek összekapcsolják az egy osztályba tartozó objektumokat, vagyis a kapcsolatok azonosítása, másképpen nem megoldható. Maguk a tanuló algoritmusok a már előzetesen osztályokba sorolt objektum-halmazok elemzésével, azt a modellt keresik, amelyben leginkább illeszkednek az objektumok attribútumai az

objektumok címkéivel, azaz az objektumok osztályozásával (Tan és mtsai 2008). Ezt a lépést nevezzük tanulásnak. A már „betanított” algoritmus pedig egy nem osztályozott objektum hovatartozásáról megjósolja, hogy az objektum melyik osztályhoz tartozik. A „betanított” algoritmus megbízhatóságáról úgy kaphatunk képet, ha véletlenszerűen kiválasztunk néhány címkézett adatot, amelyeket kihagyunk a tanulási folyamatban, és a már „betanított” algoritmuson tesztelhetjük, hogy a becslt eredmény hány százalékban esik egybe a valóssal.

A tanuló algoritmusoknak számos fajtája van: döntési fán, neurális hálón, szupport vektorgépen, szabályokon stb. alapuló (Tan és mtsai 2008). Mi a rendelkezésre álló lehetőségek közül a döntési fákon alapuló gradient boosting technikát választottuk (Chen és Guestrin 2016). A haplotípusok alapján történő haplocsoport predikció matematikai értelemben azonos dimenziós, egészértékű vektorok osztályozását jelenti. Az alacsony dimenziós, belső struktúrával nem rendelkező vektorok felügyelt osztályozására a Gradient Boosted Trees eljárás a jelenleg legnépszerűbb módszer a gépi tanulás alkalmazóinak körében.

A Gradient Boosted Trees eljárás két fontos gépi tanulási gondolatra épül. Egyikük a döntési fa fogalma. Egy döntési fa csúcspontjai a levelek kivételével, bizonyos egyszerű, eldöntendő kérdéseket, míg a fa levelei már a címkéket vagy osztályokat reprezentálják. Egy adott objektum döntési fa által való elemzése során az objektum attribútumai alapján egymást követő kiértékelések sorozata után jutunk el a döntési fa egy adott leveléhez. Ez a levél azt jelzi, hogy a döntési sorozat alapján az adott objektumot melyik osztályhoz tartozónak becsülte meg az algoritmus. Esetünkben például egy döntési pont lehet: ha a DYS391 lokuszon az STR értéke 11-nél kisebb, akkor az adott döntési pontból az egyik következő döntési pontba, ha nem kisebb, akkor pedig egy másik, de szintén következő döntési pontba jutunk a feldolgozás során.

Az eljárás másik fontos összetevője a boosting fogalma, amely a gradient boosting algoritmus iteratív jellegére utal. Ennek lényege az leegyszerűsítve, hogy az algoritmus tanul a hibáiból az iteráció során, és egy adott határig folyamatosan növeli a becslési algoritmus pontosságát új döntési fákat generálva. Nevezetesen, ha az algoritmus az iteráció egy adott pontján épített egy döntési fát, akkor az kiértékeli a rendelkezésre álló már felcímkézett objektumokon, és a következő iterációs lépésben olyan döntési fát próbál építeni, amely a hibás osztályozásokat korrigálja. Ennek az iteratív korrekciós eljárásnak az eredménye egy döntési fákból épített erdő, amely a fák közti szavazással hozza meg döntését, és ezzel pontosabb előrejelzéseket tesz, mint amire az egyes fák külön-külön képesek lennének.

Az általunk használt gradient boosting eljárás fontos tulajdonsága, hogy véletlen elemeket is alkalmaz. Így a gyakorlatban nincs két azonos eredmény. Úgy kaphatunk képet a véletlen elemekből adódó bizonytalanságról, hogy sokszor lefuttatjuk a becslő algoritmusunkat. Egy adott futás esetében például következő volt az eredmény: Z2123 esélye 68,8%, a Z93* esélye 32,2%. Többféleképpen is kumulálhatjuk az egyes futások eredményeit. Mi a „többségi szavazás” megközelítést választottuk: 1000 futásból hány esetben szerepelt legnagyobb valószínűséggel az adott haplocsoport. Egy másik mérőszám a becslésünk megbízhatóságára, hogy a már osztályozott haplotípusok véletlenszerűen kiválasztott 10%-a nem vesz részt a tanulásban (held-out), és ezeken a haplotípusokon mérhető, mennyire megbízható az adott becslés az osztályozott adatokon. Hangsúlyozzuk ugyanakkor, hogy ez alapján formális statisztikai hipotézisvizsgálatot nem

végezhetünk az általunk vizsgált haplotípusról, mivel az nem ugyanabból a valószínűség eloszlásból vétett, mint a rendelkezésre álló osztályozott adat.

Vizsgálati eredmények

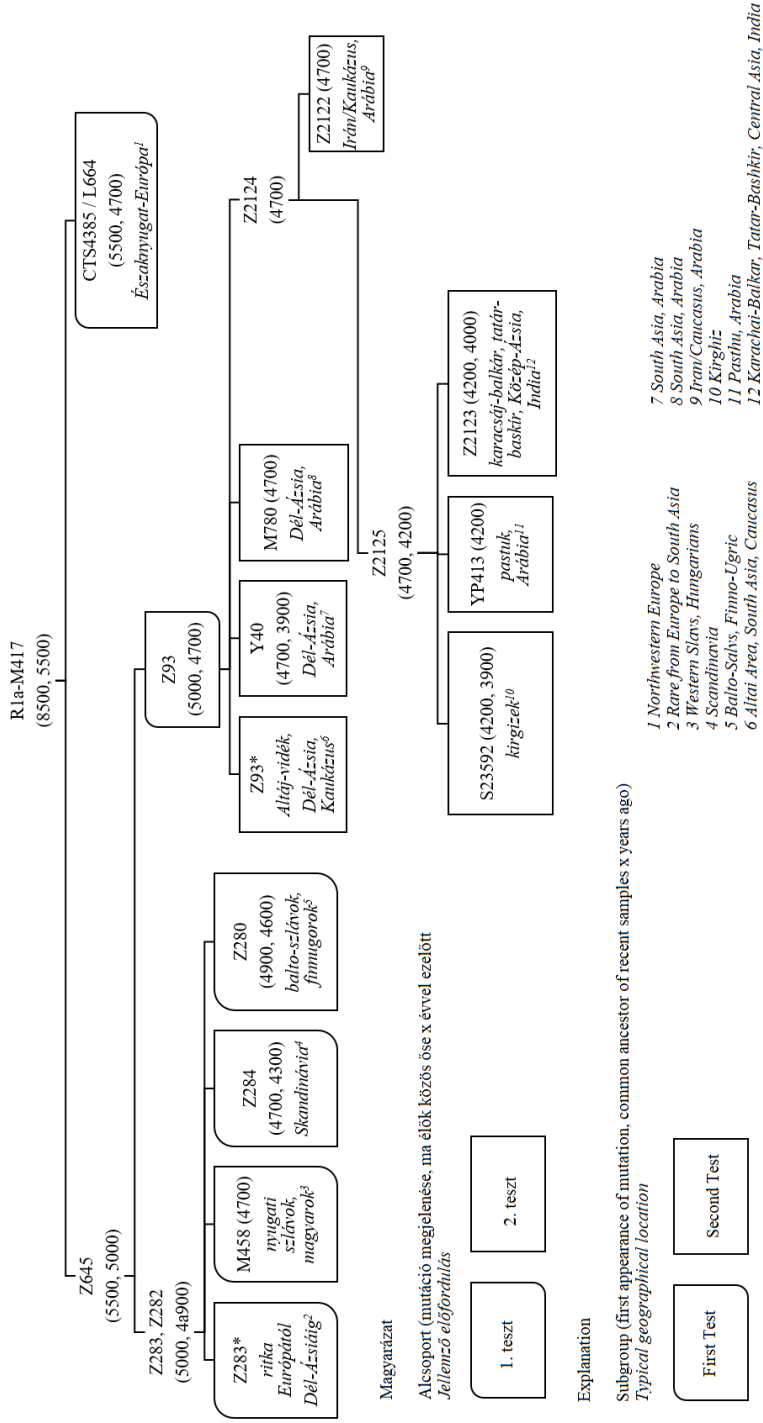
A munkacsapatunk által használt tanuló algoritmus betanításához a legnagyobb méretű és földrajzilag legszélesebb lefedésű publikus online R1a adatbázist használtuk föl, a Family Tree DNA adatbázisát (<https://www.familytreedna.com/groups/r-1a/about/results>). Első lépésben a teljes adatbázist öt nagyobb, földrajzilag is értelmezhető alcsoportra bontottuk: L664, M458, Z280, Z284, Z93. Ebben az első lépésben 17 STR-szakaszon 1610 emberrel lehetett összehasonlítani III. Béla STR adatait. A mérések (futások) 80,6%-ában a Z93, 13,9%-ában a Z280 és 5,5%-ban a Z284 volt a legesélyesebb alcsoport. Azaz a tanuló algoritmus III. Béla haplotípusát a legnagyobb eséllyel a Z93 haplocsoporthoz tartozónak ítélte meg. A becslés megbízhatósága a már osztályozott held-out adatokon 80%-os. A Z93 egy nagy alcsoport, jellemző haplocsoport a török és iráni nyelvet beszélők között, de előfordul indiaiak és arabok között is.

Második lépésben, az első körben legvalószínűbbnek mutatkozó Z93 csoportot bontottuk tovább. A Family Tree DNA adatbázisából 16 STR-szakaszon 434 ma élő, a Z93 haplocsoportba tartozó személlyel tudtuk összehasonlítani III. Béla Y-STR adatait. A Z93 csoporton belül hét alcsoportot különítettünk el, szem előtt tartva, hogy minden alcsoport az algoritmus értelmezéséhez szükséges elemszámmal rendelkezzen, valamint az eredmény földrajzi és etnikai szempontból is értelmezhető legyen. Kiegészítettük továbbá az adatbázisunkat két R1a-Z93 pozitívan tesztelt kazár archeogenetikai mintával (Klyosov és Faleeva 2017). Kíváncsiak voltunk, hogy becslő algoritmusunk milyen alcsoportot rendel majd ezekhez a haplotípusokhoz.

A hét alcsoport a Z93*, M780, Y40, Z2122, S23592, YP413, Z2123. A második lépésben III. Béla esetében a futások 80,5%-ában a Z2123, 10,8%-ában a Z93* és 6,2%-ában az S23592, 2,2%-ában pedig az M780 volt a legvalószínűbb alcsoport (2. ábra). Azaz az osztályozás második lépésében is volt a tanuló algoritmusunknak egy határozott választása, mégpedig a Z2123. Meglepetésünkre a két kazár a Z2122-es alcsoportba került. A becslés megbízhatósága a már osztályozott held-out adatokon 65%-os volt. Ez a megbízhatósági szint nem túl magas, de a későbbiekben ismertetett filogenetikai elemzések megerősítik a Z2123-as becslést.

Mai tudásunk alapján az ősapa, aki a Z2123 alcsoportot meghatározó mutációt hordozta, valamivel több, mint 4000 évvel ezelőtt élt, és megközelítőleg 4000 éve indult demográfiai expanzióinak (<https://www.yfull.com/tree/R-Z2123/>). A Z2123 alcsoport legkorábbi – archeogenetikai kutatások által megerősített – előfordulása a bronzkori Volga-vidékhez köthető, ahol egy, a gerendavázis régészeti kultúrához tartozó kurgánban eltemetett személyben azonosították (Mathieson és mtsai 2015). Tekintve, hogy csak pár száz év van a Z2123 demográfiai expanziójának kezdete, és az első azonosított Z2123 minta között, valószínű, hogy a Volga-vidék lehetett a Z2123 apai vonalak genetikai centruma is.

A Z2123 alcsoport mai előfordulása egyértelműen azt valószínűsíti, hogy az ebbe az alcsoportba tartozó férfiak részt vettek az indo-iráni népeknek a Volga és az Urál vidékről kelet felé Közép-Ázsia, Belső-Ázsia és India irányába tartó vándorlásában. Az 1. táblázatban láthatóak az R1a-Z93 csoporton belüli Z2123 alcsoport becslött gyakoriságának földrajzi eloszlása (Underhill és mtsai 2015) alapján.



2. ábra: Az R1a haplocsoport alcsoportjait meghatározó SNP-k filogenetikuss fája (Fehér Tibor munkája)
Fig. 2: The phylogenetic tree of SNPs that can be used for R1a haplogroup determination (made by Tibor Fehér)

1. táblázat. A Z2123 alcsoport gyakoriságának földrajzi (Underhill és mtsai 2015 alapján becsült) eloszlása
 Table 1. List of Z2123 frequency by population (estimation based on Underhill et al. 2015)

Populáció – Population	Nyelv – Language	Z2123 becsült arány – Z2123 estimated ratio (%)*	Régió – Region
Tatár (Tatársztán) – Tatarstan	Türk – Turkic	7,58	Volga-Urál – Volga-Ural
Kelet-Törökország – South Turkey	Türk – Turkic	4,33	Közel-Kelet – Near East
Egyesült Arab Emirátusok – United Arab Emirates	Arab	4,27	Közel-Kelet – Near East
Iráni (Kerman)	Indo-iráni – Indo-Iranian	8,57	Közel-Kelet – Near East
Dél-oszét – South Oset	Indo-iráni – Indo-Iranian	4,76	Kaukázus – Caucasus
Adige	Északnyugat-kaukázusi – NW Caucasian	7,50	Kaukázus – Caucasus
Karacsáj – Karachay	Türk – Turkic	18,84	Kaukázus – Caucasus
Abazin – Abasin	Északnyugat-kaukázusi – NW Caucasian	7,87	Kaukázus – Caucasus
Balkár – Balkar	Türk – Turkic	20,59	Kaukázus – Caucasus
Cserkesz – Cherkes	Északnyugat-kaukázusi – NW Caucasian	8,80	Kaukázus – Caucasus
Kabard	Északnyugat-kaukázusi – NW Caucasian	6,38	Kaukázus – Caucasus
Afganisztáni tádzsik – Afghanistan tajik	Indo-iráni – Indo-Iranian	10,71	Közép-Ázsia – Central-Asia
Afganisztáni türkmén – Afghanistan turkmen	Türk – Turkic	6,85	Közép-Ázsia – Central-Asia
Afganisztáni üzbég – Afghanistan usbeke	Türk – Turkic	12,80	Közép-Ázsia – Central-Asia
Kirgizisztán – Kirgyzstan	Türk – Turkic	6,15	Közép-Ázsia – Central-Asia
Közép-India – Middle India	Indo-iráni – Indo-Iranian	13,89	India – India

*: a forrásban közölt Z2125-ös adatokat vettük alapul, kivonva az összesből a kirgiz (S23592) és a pastu klasztert (YP413) – our estimation based on Z2125 subtracting subgroup S23592 and YP413 subgroups

Megállapítható, hogy a Z2123 alcsoport hatalmas területen fordul elő, mégis van négy egyértelmű régió, ahol előfordulási aránya kiugró: (1) a Volga-vidéken a tatároknál és baskiroknál, (2) a Kaukázus északi előterében a karacsájok és a balkárok között, (3) India központi területein és (4) Közép-Ázsiában egy Kelet-Anatóliától Kirgizisztánig tartó széles sávban. A négy statisztikai alapon azonosított terület közül három több szempontból is összeegyeztethető a honfoglaló magyarokra és az Árpádokra vonatkozó történelmi ismereteinkkel:

(1) A krónikákból Magna Hungariaként ismert terület, ahol Julianus barát a 13. században még magyarul beszélhetett a keleten maradt magyarokkal, a régészeti kutatások alapján részben átfed a mai Tatársztánnal (Klima 2016).

(2) A Z2123 alcsoport előfordulási gyakorisága csúcsát a Kaukázus északi vidékén éri el, a karacsájak és a balkárok között. Sajnos archeogenetikai kutatások hiányában nem dönthető el, hogy a Z2123-as alcsoport mikor és melyik népességmozgással került a Kaukázus északi előterébe. Nem tudjuk, hogy a török nyelvű karacsájak és balkárok előtt is jelen volt a Kaukázusban az alcsoport, vagy éppen ezekkel a török nyelvű népekkel érkezett oda. A kérdésnek kiemelt szerepe van vizsgálatunk szempontjából, mert a kipszak nyelvű karacsájak és balkárok csak a mongol korban, tehát a honfoglalás után érkeztek a térségbe. Történeti adatok, népzenei párhuzamok is rámutatnak a Kaukázus fontosságára a magyar őstörténet szempontjából (Sípos és Tavkul 2012). A genetikailag meghatározott biológiai jellegeket vizsgáló antropológiai kutatások már több évtizede jelezték, hogy a honfoglaló magyarságban nagy arányban találunk a Kaukázus északi előterére és annak tágabb környezetére, a vaskorig visszavezethető embertípusokat (Tóth 1965). A karosi vezérek anyai vonalai szintén a Kaukázus vidékéről származnak (Neparáczki 2018). Összességében, a korai magyarok és a Kaukázus térsége között különböző biológiai és kulturális szálak is kimutathatók, amelyek egymáshoz viszonyított időbelisége azonban még nem teljes mértékben tisztázott

(3) A földrajzilag legtávolabbi a közép-ázsiai terület. Az Emese álma-monda – amely az Árpádok eredetmondájának tekinthető – legközelebbi párhuzamai a perzsa Kúrosz és a szeldzsuk Oszmán csodás előjelekkel övezett születéséhez kapcsolódnak (B. Szabó és Sudár 2017). Azaz a genetikai kapcsolatokhoz hasonlóan az Árpádok eredetmondája is iráni és törökségi népességek irányába mutat. Különös formai párhuzam, hogy Oszmán és Ügyek (Álmos mondabeli édesapja) apósának a neve is feltűnően hasonlít egymáshoz: Edebali és Eunedbelia (Demény 1996). A hasonlóság, azért is meglepő, mert Oszmán lényegesen később élt, mint Ügyek. Említésre méltó, hogy bár turul szavunk török eredetű, a togrul (a turul török megfelelője) mint személynév és madárnév távolról sem volt általános a törökségi népek között sem térben, sem időben. Elsősorban a 9–14. század között, a különféle oguz csoportokra – például éppen a szeldzsukokra – volt jellemző a togrul személynévként és madárnévként való használata, amely Turkesztánból terjedt fokozatosan a meghódított nyugatabbi területeken (Sudár 2017).

Filogenetikai elemzésünk során három különböző eljárással készítettük el 115 az R1a-Z2123 alcsoportoz tartozó, a Family Tree DNA R1a projekt adatbázisában alcsoportokra tesztelt egyedi haplotípusok genetikai kapcsolatrendszerét. A genetikai hálózatot a Network 4.6.1.1 (Fluxus-Engineering) szoftver által implementált median-joining algoritmussal rajzoltuk meg. A median-joining algoritmus és Network 4.6.1.1 (Fluxus-Engineering) szoftverkészlet az egyik leggyakrabban használt vizualizációs eszköz apai és anyai vonalak kapcsolatrendszerének feltérképezésére.

Anyai ágak esetében csak SNP-eket, jellemzően egy adott szekvencia referenciától eltérő nukleotidokat, mutációkat táplálhatunk be a vizsgált minták esetében. Apai ágak esetében több lehetőség adódik, mert három különböző típusú input adatot is betáplálhatunk a Network rendszerének. Így Y-STR adatokat, Y-SNP adatokat, és Y-SNP és Y-STR adatokat együtt is betáplálhatunk. Értelemszerűen a valósághoz legközelebbi képet akkor kapjuk, amikor Y-SNP és Y-STR adatokat együtt vesszük figyelembe. Az Y-STR-ek önmagukban sajnos nem adnak biztos leszármazási magyarázatot, ugyanis visszafelé (back mutation) is változhatnak (ös: 13 → leszármazott: 12 → annak leszármazottja újra 13), így „téves találatok” is létrejöhetnek. Téves egyezés alatt azt értjük, hogy a két minta Y-STR adatai megegyeznek, de azok valójában csak távoli rokonai egymásnak, amelyet a minták SNP adataiban lévő eltérések bizonyítanak.

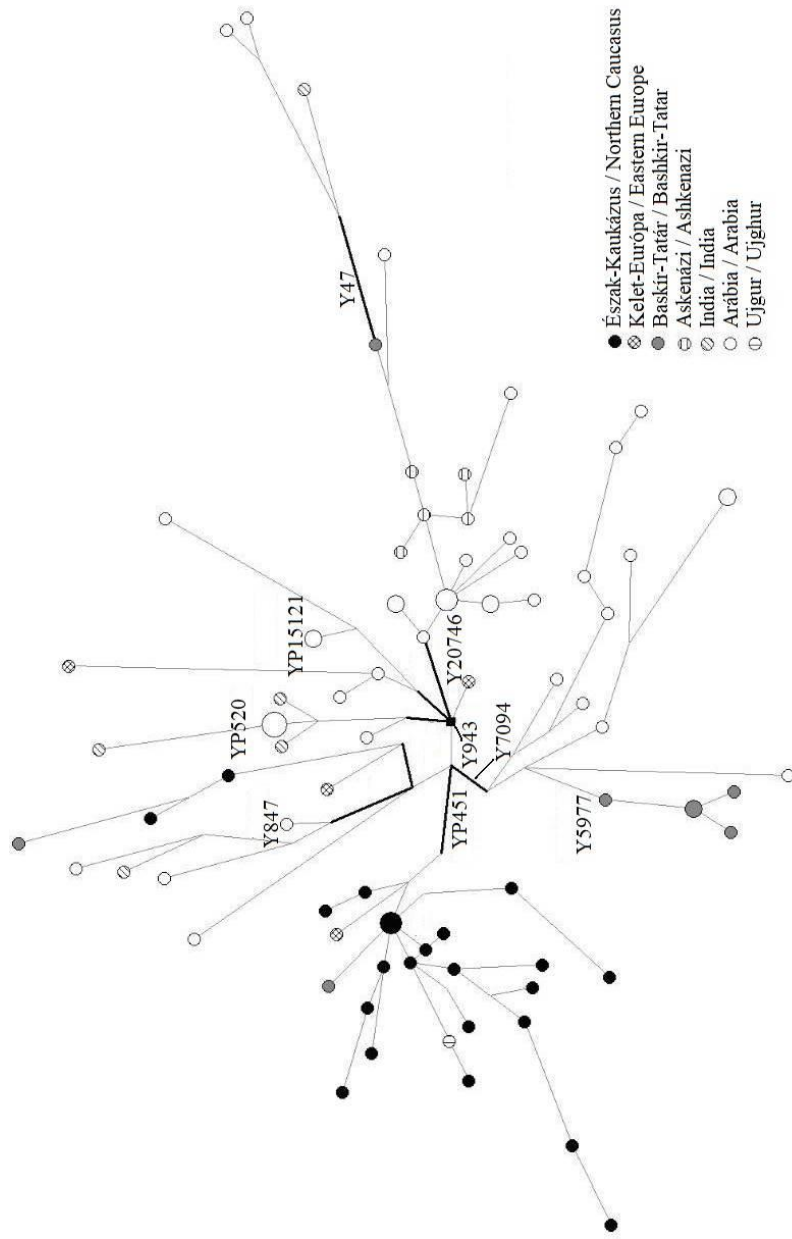
Különösen gyakran fordul elő ez olyan csoportok esetében, amikor egy szűk csoport leszármazottai intenzív demográfiai expanzióval terjedtek el hatalmas földrajzi területeken. Ilyenkor egymástól nagy földrajzi távolságban is azonos alapító Y-STR haplotípusok jelennek meg, és az egymástól térben és időben is távoli leszármazottak Y-STR adatai is lehetnek hasonlóak. Gyakori, hogy csak az Y-STR adatok alapján készült hálózatokban körök is megjelennek, amely szintén nem fordulhat elő a valóságban. Az ilyen téves találatok és hamis körök kiszűrése érdekében szükséges mutációk (SNP) adatok beiktatása is a networkbe, amely viszont csak két értéket vehet fel (0 vagy 1), attól függően, hogy az adott SNP-re pozitív vagy negatív az adott minta.

Az első esetben (3. ábra) az Y-STR adatok (14 lokuszon) mellett figyelembe vettük a Z2123 alcsoporton belüli SNP adatokat, alcsoportokat (Y2074, Y47, Y934, A1512, Y7094, Y874, YP451, YP520, BY307). Célunk az volt, hogy minél pontosabb képet kapjunk a Z2123 alcsoport belső struktúrájáról, az alcsoportok esetleges földrajzi összefüggésrendszeréről. Mivel a két magyar archeogenetikai minta alcsoportja ismeretlen volt a Z2123 alcsoporton belül, ezért értelemszerűen azok nem láthatóak a 3. ábrán. Vastag vonallal próbáltuk jelölni továbbá azokat a filogenetikai kapcsolatokat, ahol egyes fontosabb alcsoportot meghatározó mutációk megjelentek.

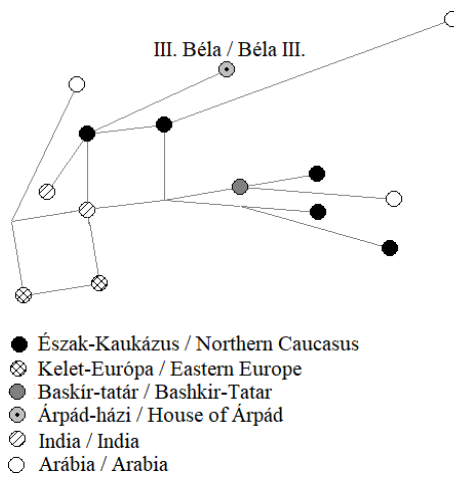
A második esetben (4. ábra) kizárólag az Y-STR haplotípusokat (14 lokuszon) használtuk föl a genetikai hálózat elkészítésére. Ez az ábrázolás a korábban elmondottak miatt kevésbé pontos, mint az előző, de bevonhattuk a vizsgálatba a magyar archeogenetikai mintát is. Ebben az esetben már csak arra voltunk kíváncsiak, hogy milyen haplotípusokkal szomszédos, milyen a magyar archeogenetikai minta mikrokörnyezete filogenetikai értelemben. Ennek megfelelően az eredmény csak a magyar archeogenetikai minta közvetlen környezetét mutatja meg, mert a teljes kapcsolatrendszer tekintetében a 3. ábra jóval pontosabb.

A harmadik esetben (5. ábra) összegeztük a két korábbi hálózat eredményeit, és ezek alapján azt tételeztük fel az előbbi két eset eredményei alapján, hogy a magyar minta az YP451-es alcsoportoz tartozik. És ezen észrevételek alapján rajzoltuk meg a kapcsolatrendszert az YP451 alcsoporton belül. Az YP451-es csoport meglehetősen korán, mintegy 4000 évvel ezelőtt vált el a többi Z2123 alcsoporttól, és nagyjából 2200 éve élt a mai YP451-ek alcsoportoz tartozó közös őse (<https://www.yfull.com/tree/R-YP451/>). Mai tudásunk alapján az YP451-es alcsoport szinte kizárólag a Kaukázus északi előterében, döntően a karacsájak és baskárok között fordul elő (<https://www.yfull.com/tree/R-YP451/>).

A 3. ábrán látható, hogy az SNP-k alapján az egyes alcsoportok földrajzi/etnikai értelemben is viszonylag jól elkülönülnek. A 3. ábra középpontjában az Y934 alapító haplotípus található (kis fekete négyzettel jelölve), tőle egy mutációra az egyetlen ismert oroszországi Y934* minta. Az ábra jobb oldalán helyezkednek el az Y934-es mutációra negatív Z2123 alcsoportok (Y47 és Y20746), míg a további irányban az Y934 alcsoportjai. A tatárok és baskírok között az Y5977 alcsoport a leggyakoribb. Szembeötlő a közép-ázsiai és a belső-ázsiai minták hiánya, amely talán annak köszönhető, hogy rendkívül kevés az egyszerre mély SNP felbontással és elegendő STR lokuszon vizsgált minta ezekből a régiókból. A rendelkezésre álló adatok alapján nem eldönthető, hogy az egyetlen dokumentált ujjur YP451+ (azon belül pedig YP450+) véletlen, vagy újabb YP451+ minták várhatóak az eddig kevésbé vizsgált belső- és közép-ázsiai régiókból.

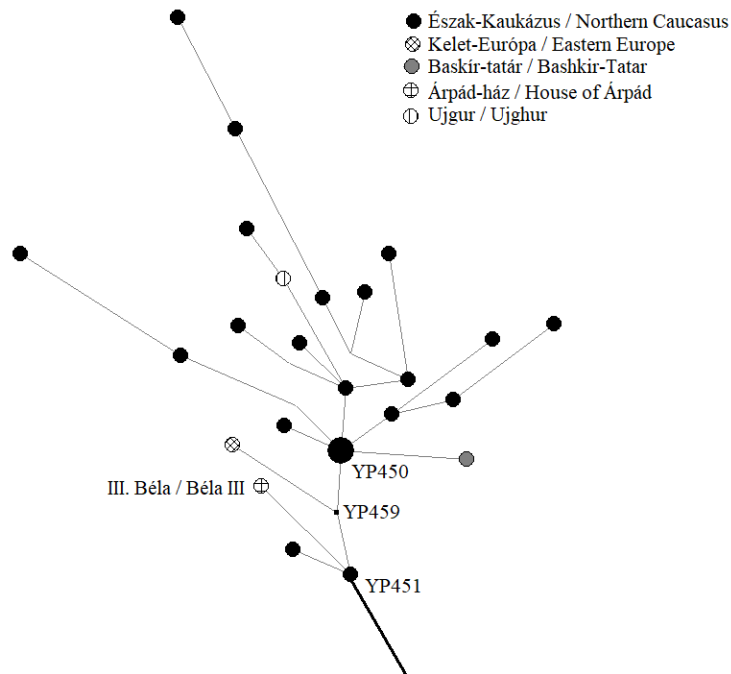


3. ábra: Az R1a-Z2123 alcsoportoz tartozó haplotípusok genetikai hálózata az Y-SNP-k és Y-STR-k alapján
 Fig. 3: Network of R1a-Z2123 haplotypes – taking into account Y-SNP subclades and Y-STR haplotypes



4. ábra: Az R1a-Z2123 alcsoportéhoz tartozó egyedi haplotípusok genetikai hálózata csak Y-STR adatok alapján, kiegészítve a magyar archeogenetikai mintával

Fig. 4: Network of R1a-YP451 haplotypes based on only Y-STR including the Hungarian archeogenetic sample



5. ábra: Az R1a-YP451 alcsoportéhoz tartozó egyedi haplotípusok genetikai hálózata Y-STR és Y-SNP adatok alapján, kiegészítve a magyar archeogenetikai mintával és a becsült Y-SNP-vel

Fig. 5: Network of R1a-YP451 haplotypes based on Y-STR és Y-SNP data, adding the Hungarian archeogenetic sample with its estimated Y-SNP

Meglepetésünkre a második esetben (4. ábra) III. Béla a karacsáj haplotípushoz, egy rendkívül ritka YP451+ és Y449- alcsoporthoz tartozó mintához kapcsolódott. Látható, hogy elvárásainknak megfelelően a tisztán Y-STR alapon megrajzolt genetikai hálózatnak már ezen a kis részleten is több pontatlanság van az előbbi ábrához képest. Két kör is megjelenik, amely a valóságban nem fordulhat elő. Illetve tisztán Y-STR adatok alapján olyan karacsáj és indiai minták kerülnek közel egymáshoz, amelyek valójában távol vannak egymástól az Y-SNP adatok alapján.

Az 5. ábrán az első két ábra (3. ábra és a 4. ábra) eredményei alapján próbáltuk megrajzolni az YP451 alcsoport szerkezetét kiegészítve a magyar archeogenetikai mintával. Látható, hogy az YP451 alcsoportot etnikai értelemben a karacsájok és balkárok, genetikai értelemben pedig az YP449-es alcsoport dominálja. Az Árpád-házi minta egy YP451+ és YP449- csoportba tartozó karacsáj mintához kapcsolódik közvetlenül. Ugyanebbe az YP451+ és YP449- alcsoportba még egy karacsáj minta tartozik.

Következtetések

Először is hangsúlyoznunk kell, hogy az Árpádok eredetére igazán megnyugtató válaszokat csak további mérések adhatnak. Reményeink szerint azonban becsléseinkkel támpontokat adhattunk a témával foglalkozó kutatóknak, hogy mely SNP-k vizsgálata lehet fontos az Árpádok eredetének vizsgálatakor.

Tanulmányunk gyakorlati jelentősége általánosságban is az, hogy az olcsóbb, kisebb felbontású archeogenetikai adatok matematikai módszerekkel tovább finomíthatók, mert a legvalószínűbb alcsoportokat meg lehet becsülni recens Y-STR haplotípusok alapján. Így nagyobb történeti szériák rokonsági kapcsolatainak elemzésekor átgondolt kutatási stratégia (mely mintákat érdemes mélyebben elemezni) mellett jelentős kutatási költség takarítható meg. Tanulmányunk egyben az első, amely a magyar őstörténeti kutatásokban tanuló algoritmusokat alkalmaz.

III. Béla Y-STR haplotípusa alapján tanuló algoritmus segítségével első lépésben azt becsültük, hogy III. Béla király a türk és indo-iráni nyelvet beszélők között gyakori R1a-Z93 alcsoporthoz tartozik. Tanuló algoritmusunk második lépésben a III. Bélát legnagyobb valószínűséggel egyaránt a Z2123 alcsoporthoz tartozónak becsülte. A szűkebb területen elterjedt Z2123 alcsoport elemzése lehetővé teszi a Z2123 alcsoporthoz tartozó minták és III. Béla haplotípusának filogenetikai elemzését.

A filogenetikai elemzés alapján úgy tűnik, hogy III. Béla a YP451-es alcsoportba tartozik, amely túlnyomó többségében a karacsájok és balkárok között fordul elő. Az a recens karacsáj minta, amellyel III. Béla a legközelebbi kapcsolatot mutatja, a rendkívül ritka YP451+ és YP449- alcsoportba tartozik. Összesen két ilyen mintát ismerünk az Family Tree DNA százezres adatbázisából, és mind a kettő karacsáj etnikumhoz tartozik.

Mindez természetesen csak annyit jelent, hogy valószínűleg a karacsájok egyik etnikai komponenséhez köthető az Árpádok eredete. Nyitott kérdés azonban, hogy a 2200 éve demográfiai expanziónak induló YP451 alcsoport a karacsájok elődeivel érkezett a Kaukázus térségébe, vagy már a karacsájok érkezése előtt jelen volt egy olyan etnikumban, amely mára nyelvileg beolvadt a karacsájok közé. Ezért fontos megemlíteni, hogy a karacsájok etnogenezisében szkíták, hunok, alánok, bolgár-törökök és kipszakok is szerepet játszottak más népek mellett (Sípos és Tavkul 2012).

Irodalom

- 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., Aron, J., Della Casa, P., 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., Kolar, J., Kriiska, A., Lasak, I., Longhi, C., McGlynn, G., Merkevicius, A., Merkyte, I., Metspalu, M., Mkrtychyan, R., Moiseyev, V., Paja, L., Pálfi, Gy., Pokutta, D., Pospieszny, L., Price, T.D., Saag, L., Sablin, M., Shishlina, N., Smrcka, V., Soenov, V.I., Szeverényi, V., Tóth, G., Trifanova, S.V., Varul, L., Vicze, M., Yepiskoposyan, L., Zhitenev, V., Orlando, L., Sichert-Ponten, T., Brunak, S., Nielsen, R., Kristiansen, K., Willerslev, E. (2015): Population genomics of Bronze Age Eurasia. *Nature*, 522(7555): 167–172. DOI: [10.1038/nature14507](https://doi.org/10.1038/nature14507)
- Athey, T.W. (2006): Haplogroup Prediction from Y-STR Values using a Bayesian-allele-frequency approach. *Journal of Genetic Genealogy*, 2: 34–39.
- Balanovsky, O., Zhabagin, M., Agdzhoyan, A., Chukhryaeva, M., Zaporozhchenko, V., Utevska, O., Highnam, G., Sabitov, Z., Greenspan, E., Dibirova, K., Skhalyakho, R., Kuznetsova, M., Koshel, S., Balanovska, E. (2015): Deep phylogenetic analysis of haplogroup G1 provides estimates of SNP and STR mutation Rates on the human Y-chromosome and reveals migrations of Iranian speakers. *PLoS ONE*, 10(4): e0122968 DOI: 10.1371/journal.pone.0122968
- Bernert, Zs., Évinger, S., Fóthi, E. (2006): New symbolic trephination cases from Hungary. *Annales historico-naturales Musei nationalis hungarici*, 98: 177–183.
- Bermisheva, M.A., Tambets, K., Villems, R., Khusnutdinova, E.K. (2002): Diversity of mitochondrial DNA haplogroups in ethnic Populations of the Volga–Ural Region. *Molecular Biology*, 36(6): 802–812.
- Bíró, A.Z., Zalán, A., Völgyi, A., Pamjav, H. (2009): A Y-chromosomal comparison of the Madjars (Kazakhstan) and the Magyars (Hungary). *American Journal of Physical Anthropology*, 139(3): 305–310. DOI: [10.1002/ajpa.20984](https://doi.org/10.1002/ajpa.20984)
- Bíró, A., Fehér, T., Bárány, G., Pamjav, H. (2015): Testing Central and Inner Asian admixture among contemporary Hungarians. *Forensic Science International, Genetics*, 15: 121–126. DOI: [10.1016/j.fsigen.2014.11.007](https://doi.org/10.1016/j.fsigen.2014.11.007)
- Boldog, Z., Petkes, Zs., Sudár, B., Türk A. (2015): From the ancient homelands to the Carpathians – From the finds to the apparel. *Hungarian Archaeology*, 2015(3): 1–8.
- 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. *Annals of Human Genetics*, 71(6): 791–803. DOI: [10.1111/j.1469-1809.2007.00371.x](https://doi.org/10.1111/j.1469-1809.2007.00371.x)
- B. Szabó, J., Sudár, B. (2014): Vgec-ügyek – Egy elfeledett ősapá. In: Türk, A., Balogh, Cs., Major, B. (Szerk.) *Hadak Útján XXIV. A népvándorláskor fiatal kutatóinak XXIV. Konferenciája*, 2. Magyar Őstörténeti Témacsoport Kiadványok, Magyar Tudományos Akadémia Bölcsészettudományi Kutatóközpont, Budapest.
- Chen, T., Guestrin, C. (2016): XGBoost: A Scalable Tree Boosting System. In: Krishnapuram, B., Shah, M., Smola, A.J., Aggarwal, Ch.C., Shen, D., Rastogi, R. (2016): *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, San Francisco, USA. pp. 785–794.
- Csőszy, A., Szécsényi-Nagy, A., Csákyová, 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. *Scientific Reports*, 16(6): 33446. DOI: [10.1038/srep33446](https://doi.org/10.1038/srep33446)

- Damgaard, P., Marchi, N., Rasmussen, S., Peyrot, M., Renaud, G., Korneliussen, T., Moreno-Mayar, J.V., Pedersen, M.W., Goldberg, A., Usmanova, E., Baimukhanov, N., Loman, V., Hedeager, L., Pedersen, A.G., Nielsen, K., Afanasiev, G., Akmatov, K., Aldashev, A., Alpaslan, A., Baimbetov, G., Bazaliiskii, V.I., Beisenov, A., Boldbaatar, B., Boldgiv, B., Dorzhu, C., Ellingvag, S., Erdenebaatar, D., Dajani, R., Dmitriev, E., Evdokimov, V., Frei, K.M., Gromov, A., Goryachev, A., Hakonarson, H., Hegay, T., Khachatryan, Z., Khaskhanov, R., Kitov, E., Kolbina, A., Kubatbek, T., Lau, N., Margaryan, A., Merkyte, I., Mertz, I.V., Mertz, V.K., Mijiddorj, E., Moiyesev, V., Mukhtarova, G., Nurmukhanbetov, B., Orozbekova, Z., Panyushkina, I., Logvin, A., Sjögren, K.G., Štolcová, T., Taravella, A.M., Tashbaeva, K., Tkachev, A., Tulegenov, T., Voyakin, D., Yepiskoposyan, L., Undrakhbold, S., Varfolomeev, V., Weber, A., Sayres, M.A.W., Krادين, N., Allentoft, M.E., Orlando, L., Nielsen, R., Sikora, M., Heyer, E., Kristiansen, K., Willerslev, E. (2018): 137 ancient human genomes from across the Eurasian steppes. *Nature*, 557: 369–374. DOI: [10.1038/s41586-018-0094-2](https://doi.org/10.1038/s41586-018-0094-2)
- Demény, I.P. (1996): Emese álma. *Erdélyi Múzeum*, 58: 1–2.
- Éry, K., Marcsik, A., Nemeskéri, J., Szalai, Gy. (2008): Az épített sírok csontvázletelei, III. Béla király. In: Éry, K. (Szerk.) *A Székesfehérvári királyi Bazilika embertani leletei 1848–2002*. Balassi Kiadó, Budapest. pp. 37–67.
- Fehér, T., Németh, E., Vándor, A., Kornienko, I.V., Csáji, L.K., Pamjav, H. (2015): Y-SNP L1034: limited genetic link between Mansi and Hungarian-speaking populations. *Molecular Genetics and Genomics*, 290(1): 377–386. DOI: [10.1007/s00438-014-0925-2](https://doi.org/10.1007/s00438-014-0925-2)
- Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., Brandt, G., Nordenfelt, S., Harney, E., Stewardson, K., Fu, O., 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., 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](https://doi.org/10.1038/nature14317)
- Ilumae, A.M., Reidla, M., Chukhryaeva, M., Järve, M., Post, H., Karmin, M., Saag, L., Agdzhoyan, A., Kushniarevich, A., Litvinov, S., Ekomasova, N., Tambets, K., Metspalu, E., Khusainova, R., Yunusbayev, B., Khusnutdinova, E.K., Osipova, L.P., Fedorova, S., Utevska, O., Koshel, S., Balanovska, E., Behar, D.M., Balanovsky, O., Kivisild, T., Underhill, P.A., Villems, R., Rootsi, S. (2016): Human Y chromosome haplogroup N: A non-trivial time-resolved phylogeography that cuts across language families. *American Journal of Human Genetics*, 99(1): 163–173. DOI: [10.1016/j.ajhg.2016.05.025](https://doi.org/10.1016/j.ajhg.2016.05.025)
- Janhunen, J. (2014): A legkeletibb uráliak. In: Csúcs, S. (Szerk.) *Finnugor Világ*. Reguly Társaság, Solymár.
- Klima, L. (2016): *Jürkák, tormák, merják: Szemelvények a finnugor nyelvű népek történetének korai forrásaiból*. MTA Bölcsészettudományi Kutatóközpont, Budapest. p. 471.
- Klyosov, A.A., Faleeva, T. (2017): Excavated DNA from Two Khazar Burials. *Advances in Anthropology*, 7: 17–21. DOI: [10.4236/aa.2017.71002](https://doi.org/10.4236/aa.2017.71002)
- Lamnidis, T.C., Majander, K., Jeong, C., Salmela, E., Wessman, A., Moiseyev, V., Khartanovich, V., Balanovsky, O., Ongyerth, M., J., Pääbo, S., Onkamo, S., Haak, W., Krause, J., Schiffels, S. (2018): Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in Europe. *Nature Communications*, 9: 5018. DOI: [10.1038/s41467-018-07483-5](https://doi.org/10.1038/s41467-018-07483-5)
- Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Alpaslan Roodenberg, S., Harney, E., Stewardson, E., Fernandes, D., Novak, M., Sirak, K., Gamba, C., Jones, E.R., Llamas, B., Dryomov, S., Pickrell, J., Arsuaga, J.L., Bermudez de Castro, J.M., Carbonell, E., Gerritsen, F., Khokhlov, A., Kuznetsov, P., Lozano, M., Meller, H., Mochalov, O., Moiseyev, V., Guerra, M.A.R., Roodenberg, J., Verges, J.M., Krause, J., Cooper, A., Alt, K.W., Brown, D., Anthony, D., Lalueza-Fox, C., Haak, W., Pinhasi, R., Reich, D. (2015): Eight thousand years of natural selection in Europe. pp. Supplementary Information. Archaeological context for 83 newly reported ancient samples. 12–14. *bioRxiv* 016477. DOI: [10.1101/016477](https://doi.org/10.1101/016477)

- Németh, E., Csáky, V., Székely, G., Bernert, Zs., Fehér, T. (2017): Új filogenetikai mértékek és alkalmazásuk. *Anthropológiai Közlemények*, 58: 3–36. DOI: [10.20330/AnthropKozl.2017.58.3](https://doi.org/10.20330/AnthropKozl.2017.58.3)
- Neparáczki E., Maróti, Z., Kalmár, T., Kocsy, K., Maár, K., Bihari, P., Nagy, I., Fóthi, E., Pap, I., Kustár, Á., Pálfi, Gy., Raskó, I., Zink, A., Török, T. (2018): Mitogenomic data indicate admixture components of Central-Inner Asian and Srubnaya origin in the conquering Hungarians. *PLoS One*, 13(11): e0208295. DOI: [10.1371/journal.pone.0205920](https://doi.org/10.1371/journal.pone.0205920)
- Olasz, J., Seidenberg, V., Hummel, S., Szentirmay, Z., Szabados, Gy., Melegh, B., Kásler, M., (2018): DNA profiling of Hungarian King Béla III and other skeletal remains originating from the Royal Basilica of Székesfehérvár. *Archaeological and Anthropological Sciences*, 10: 1–13. DOI: [10.1007/s12520-018-0609-7](https://doi.org/10.1007/s12520-018-0609-7)
- Pimenoff, V.N., Comas, D., Palo, J.U., Vershubsky, G., Kozlov, A., Sajantila, A. (2008): Northwest Siberian Khanty and Mansi in the junction of West and East Eurasian gene pools as revealed by uniparental markers. *European Journal of Human Genetics*, 16(10): 1254–1264. DOI: [10.1038/ejhg.2008.101](https://doi.org/10.1038/ejhg.2008.101)
- Róna-Tas, A., Berta, Á (2011): *West Old Turkic. Turkic Loanwords in Hungarian*, 2. Harrassowitz, Wiesbaden.
- Sarkassian, C., Balanovsky, O., Brandt, G., Khartanovich, V., Buzhilova, A., Koshel, S., Zaporozhchenko, V., Gronenborn, D., Moiseyev, V., Kolpakov, E., Shumkin, V., Alt, K.W., Balanovska, E., Cooper, A., Haak, W. (2013): Ancient DNA reveals prehistoric gene-flow from Siberia in the complex human population history of North East Europe. *Plos Genetics*, 9(2): e1003296. DOI: [10.1371/journal.pgen.1003296](https://doi.org/10.1371/journal.pgen.1003296)
- Schlecht, J., Kaplan, M.E., Barnard, K., Karafet, T., Hammer, M.F., Merchant, N.C. (2008): Machine-Learning Approaches for Classifying Haplogroup from Y Chromosome STR Data. *PLoS Computational Biology*, 4(6): e1000093. DOI: [10.1371/journal.pcbi.1000093](https://doi.org/10.1371/journal.pcbi.1000093)
- Sípos, J., Tavkul, U. (2012): *A régi magyar népzene nyomában. A kaukázusi karacsajok népzeneje*. Zenetudományi Intézet – L'Harmattan, Budapest.
- Skultéty, Gy. (2008): Arcok a székesfehérvári királyi bazilikából. In: Éry, K. (Szerk.) *A Székesfehérvári királyi Bazilika embertani leletei 1848–2002*. Balassi Kiadó, Bp. pp. 177–184.
- Sudár, B. (2017): Turul szavunk művelődéstörténeti hátteréhez. *Történelmi Szemle*, 4: 605–619.
- Szeifert, B., Csákyová, V., Stégmár, B., Gerber, D., Egyed, B., Botalov, S.G., Goldina, R.D., Danich, A.V., Türk, A., Mende, B.G., Szécsényi-Nagy, A. (2018): Maternal genetic composition of early Medieval (6–10th centuries AD) populations lived in the Cis- and Trans-Ural and Volga-Kama region. *Archaeology of the Eurasian Steppes*, 6: in press.
- Tambets, K., Yunusbayev, B., Hudjashov, G., Ilumäe, A.M., Rootsi, S., Honkola, T., Vesakoski, O., Atkinson, Q., Skoglund, P., Kushniarevich, A., Litvinov, S., Reidla, M., Metspalu, E., Saag, L., Rantanen, T., Karmin, M., Parik, J., Zhadanov, S.I., Reisberg, T., Dibirova, K., Evseeva, I., Nelis, M., Klovins, J., Metspalu, A., Esko, T., Balanovsky, O., Balanovska, E., Khusnutdinova, E.K., Osipova, L.P., Voevoda, M., Villems, R., Kivisild, T., Metspalu, M. (2018): Genes reveal traces of common recent demographic history for most of the Uralic-speaking populations. *Genome Biology*, 19: 139. DOI: [10.1186/s13059-018-1522-1](https://doi.org/10.1186/s13059-018-1522-1)
- Tan, P.N., Steinbach, M., Kumar, V. (2008): *Introduction to Data Mining*. Pearson, UK, p. 792.
- Tóth, T. (1965): A honfoglaló magyarság ethnogenezisének problémája. *Anthropológiai Közlemények*, 9: 139–149.
- Török, A. (1894): Ásatok az Árpádok testereklýeinek embertani búvárlatához. *Értekezések a Természettudományok köréből*, 23(9): 565–630.
- Triska, P., Chekanov, N., Stepanov, V., Khusnutdinova, E.K., Kumar, G.P.A., Akhmetova, V., Babalyan, K., Boulygina, E., Kharkov, V., Gubina, M., Khidiyatova, I., Khitrinskaya, I., Khrameeva, E.E., Khusainova, R., Konovalova, N., Litvinov, S., Marusin, A., Mazur, A.M., Puzyrev, V., Ivanoshchuk, D., Spiridonova, M., Teslyuk, A., Tsygankova, S., Triska, M., Trofimova, N., Vajda, E., Balanovsky, O., Baranova, A., Skryabin, K., Tatarinova, T.V., Prokhortchouk, E. (2017): Between Lake Baikal and the Baltic Sea: genomic history of the gateway to Europe. *BMC Genetics*, 18(Suppl. 1): 110. DOI: [10.1186/s12863-017-0578-3](https://doi.org/10.1186/s12863-017-0578-3)

- Underhill, P.A., Poznik, G.D., Rootsi, S., Järve, M., Lin, A.A., Wang, J., Passarelli, B., Kanbar, J., Myres, N.M., King, R.J., Di Cristofaro, J., Sahakyan, H., Behar, D.M., Kushniarevich, A., Šarac, J., Šaric, T., Rudan, P., Pathak, A.K., Chaubey, G., Grugni, V., Semino, O., Yepiskoposyan, L., Bahmanimehr, A., Farjadian, S., Balanovsky, O., Khusnutdinova, E.K., Herrera, R.J., Chiaroni, J., Bustamante, C.D., Quake, S.R., Kivisild, T., Villems, R. (2015): The phylogenetic and geographic structure of Y-chromosome haplogroup R1a. *European Journal of Human Genetics*, 23: 124–131. DOI: [10.1038/ejhg.2014.50](https://doi.org/10.1038/ejhg.2014.50)
- Völgyi, A., Zalán, A., Szvetnik, E., Pamjav, H. (2009): Hungarian population data for 11 Y-STR and 49 Y-SNP markers. *Forensic Science International: Genetics*, 3(2): e27–28. DOI: [10.1016/j.fsigen.2008.04.006](https://doi.org/10.1016/j.fsigen.2008.04.006)
- Yunusbayev, B., Metspalu, M., Metspalu, E., Valeev, A., Litvinov, S., Valiev, R., Akhmetova, V., Balanovska, E., Balanovsky, O., Turdikulova, S. (2015): The Genetic legacy of the expansion of Turkic-speaking nomads across Eurasia. *PLoS Genetics*, 11(4): e1005068. DOI: [10.1371/journal.pgen.1005068](https://doi.org/10.1371/journal.pgen.1005068)

Levelezési cím: Bernert Zsolt
Mailing address: Embertani tár
Magyar Természettudományi Múzeum
Ludovika tér 2–6.
H-1082 Budapest
Hungary
bernert.zsolt@nhmus.hu

