



Mitokondriális DNS és mikroszatellita polimorfizmusok igazságügyi genetikai aspektusú vizsgálata a magyar népességben

DOKTORI ÉRTEKEZÉS

Készítette:

Egyed Balázs

ELTE TTK Biológia Doktori Iskola

Iskolavezető: dr. Erdei Anna

Klasszikus és molekuláris genetika Doktori Program

Programvezető: dr. Orosz László

Témavezető: dr. Lontainé dr. Santora Zsófia

igazságügyi szerológus, intézetigazgató
a biológia tudományok kandidátusa

Bűnügyi Szakértői és Kutatóintézet

2007

TARTALOMJEGYZÉK

1. BEVEZETÉS	4
1.1. AZ IGASZSÁGÜGYI GENETIKA ALAPJAI.....	6
1.1.1. Egyedi személyazonosítás – a genetikai profil	6
1.1.2. Az igazságügyi DNS laboratórium és az igazságügyi genetikus	6
1.1.3. A DNS mint törvényszéki bizonyíték	7
1.1.4. Mit jelent az egyezési valószínűség (matching probability)?	9
1.1.5. Az igazságügyi genetika rövid története	10
1.2. AZ IGASZSÁGÜGYBEN HASZNÁLTOS GENETIKAI MARKEREK.....	12
1.2.1. A mikroszatelliták általános jellemzése és nevezéktana	12
1.2.2. A mitokondriális DNS mint genetikai marker	15
1.2.3. DNS-profil és szekvencia adatbázisok.....	18
1.3. CÉLKITŰZÉSEK.....	20
1.3.1. A populációs minták összeállítása	21
1.3.2. Az analízisbe bevont STR markerek kiválasztása	22
1.3.3. A felmérések során detektált mikrovariáns STR allélek jellemzése	23
1.3.4. Az analízis STR markerek géndiverzitálásának és a polimorfizmus fokának megállapítása, allélgyakorisági táblázatok generálása	24
1.3.5. A vizsgált populációk Hardy-Weinberg egyensúlyi tesztelése az analízis STR markereken	24
1.3.6. A mtDNS teljes kontroll régió szekvencia-analízis laboratóriumi hitelesítése referencia adatbázisok generálása céljából.....	25
1.3.7. A mitokondriális DNS haplotípusok statisztikai analízise	26
1.3.8. A mtDNS szekvenciák haplocsoportosítása	26
1.3.9. A felmérésbe bevont magyar populációk genetikai strukturáltságának megállapítása.....	27
2. ANYAGOK ÉS MÓDSZEREK	28
2.1. A DNS KINYERÉSE A POPULÁCIÓS MINTÁKBÓL, KONCENTRÁCIÓ-MEGHATÁROZÁS.....	28
2.2. AZ STR LOKUSZOK PCR-AMPLIFIKÁCIÓJA ÉS ELLENŐRZÉSE GÉLELEKTROFORÉZISSSEL	29
2.3. STR GENOTÍPIZÁLÁS FRAGMENSANALÍZISSSEL	29
2.4. A MIKROVARIÁNS STR ALLÉLEK SZEKVENCIA-ANALÍZISE.....	30
2.5. A MTDNS KONTROLL RÉGIÓ PCR AMPLIFIKÁCIÓJA ÉS SZEKVENCIA-ANALÍZISE	30
2.6. A MTDNS HAPLOTÍPUSOK HAPLOCSOPORTOSÍTÁSA	32
2.7. POPULÁCIÓSTATISZTIKAI ÉS GENETIKAI ANALÍZISEK.....	33
2.7.1. Populáción belüli elemzések.....	33
2.7.1.1. STR markerekre kalkulált populációgenetikai alapértékek	33
2.7.1.2. A Hardy-Weinberg egyensúly (HWE) tesztelése, tesztsorozatok, Bonferroni-eljárás	35
2.7.1.3. A mtDNS haplotípusok populációgenetikai paraméterei	35
2.7.1.4. A mtDNS szekvenciapárok közötti eltérésének kalkulációja	36
2.7.2. A populációk összehasonlító statisztikai és genetikai elemzése	37
2.7.2.1. A populáció-párok homogenitás-vizsgálata G-statisztikával.....	37
2.7.2.2. Populációgenetika, F-statisztika, AMOVA	37
2.7.2.3. Hálózat analízis és migrációs ráta becslés	38
3. EREDMÉNYEK ÉS ÉRTÉKELES	39
3.1. AZ ANALÍZISBE BEVONT STR MARKEREK GENOTÍPIZÁLÁSA	39
3.2. MIKROVARIÁNS STR ALLÉLEK ELEMZÉSE	40
3.2.1. A D7S820 lókuszt 8.-1, 9.+1, 10 mikrovariáns alléljainak jellemzése.....	41
3.2.2. A D13S317 lókuszt egy ritka 7-es alléljának jellemzése	43
3.2.3. A D18S51 lókuszt 20.2 mikrovariáns alléljának jellemzése	44
3.2.4. Az FGA lókuszt 24.1 mikrovariáns alléljának jellemzése.....	45
3.3. A POPULÁCIÓS MINTÁKBAN GENERÁLT AUTOSZÓMÁS STR-PROFILOK STATISZTIKAI ÉS POPULÁCIÓGENETIKAI ELEMZÉSE	46
3.3.1. Allél- és genotípusgyakorisági értékek, lókusztok polimorfizmusának foka	46
3.3.2. A Hardy-Weinberg egyensúly tesztelése.....	51
3.3.3. A populációk genetikai struktúrája STR-profilok alapján	53
3.4. MITOKONDRIÁLIS DNS-PROFILOK GENERÁLÁSA A POPULÁCIÓS MINTÁKBAN	63
3.5. A POPULÁCIÓS MINTÁKBAN GENERÁLT MITOKONDRIÁLIS DNS-PROFILOK STATISZTIKAI ÉS POPULÁCIÓGENETIKAI ELEMZÉSE	64

3.5.1.	<i>Haplotípus-gyakoriságok és diverzitási paraméterek a populációkban</i>	64
3.5.2.	<i>Heteroplazmia a populációkban</i>	67
3.5.3.	<i>Az analizált populációk haplocsoport szerkezete</i>	69
3.5.4.	<i>Hálózat analízis és migrációs ráta az erdélyi populációkban</i>	71
3.5.5.	<i>Az analizált populációk genetikai struktúrája mtDNS profilok alapján</i>	73
3.6.	A POPULÁCIÓS ADATBÁZISOK IGAZSÁGÜGYI GENETIKAI CÉLÚ ALKALMAZHATÓSÁGA	78
4.	ÚJ MEGÁLLAPÍTÁSOK, KÖVETKEZTETÉSEK	81
5.	KÖSZÖNETNYILVÁNÍTÁS	89
6.	IRODALOMJEGYZÉK	90
7.	FÜGGELÉK	109
7.1.	ÁBRÁK	109
7.2.	TÁBLÁZATOK	120
7.3.	SAJÁT PUBLIKÁCIÓK	123
7.4.	RÖVIDÍTÉSEK JEGYZÉKE	131
8.	ÖSSZEFOGLALÁS	132
9.	SUMMARY	133

1. BEVEZETÉS

Az emberi populációk génállományát makroszkóposan vizsgálva az egyes közösségek, csoportok (szub-populációk) között csoportszintű jellegkülönbségek tárhatók fel, mely különbségek többek között a polimorf gének alléljeinek populációkon belüli gyakoriság eloszlásának eltéréseiben érhető tetten. A csoportszintű jellegek alapján tanulmányozható:

- a faj mint egy egész egység eredete és rokonsági viszonya a többi fajhoz.
- a populációk viszonya egymáshoz szigetek, szubrégiók, régiók és akár egész kontinensek szintjén (etnikumok, rasszok).

A makroszkópos szintet „mikroszkópra” cserélve és az egyén génállományára fókuszálva a következő kérdések merülnek fel.

- Hogyan használhatók fel legjobban a genetikai variációk egyedek azonosítására (az emberi faj esetében kb. 6,5 milliárd egyén esetében)?
- Mit tudhatunk meg az egyénről DNS állományának vizsgálata alapján? Fizikai jellegekről (biológiai nem, pigmentáció) és egyéb kategóriákról (pl. etnikai eredet).
- Hogyan dönthető el a genetikai variációk vizsgálata alapján, hogy kettő vagy több egyén egymással közeli rokonságban van-e (pl. szülő-gyermek viszony).
- A látókört kiszélesítve meghatározható-e, hogy egyes egyének, csoportok ugyanazt lezármazási vonalat képviselik, vagy ugyanazon genetikailag definiált csoportnak a tagjai-e („population history”)?

Az előbb feltett kérdések igen nagy társadalmi, szociális jelentőséggel is bírnak, mivel a kriminalisztikai célú DNS vizsgálatok és apasági tesztek ma már az igazságszolgáltatás és a szociális politika kulcsterületévé váltak. Az egyedek genetikájáról alkotott kérdések nem fogalmazhatóak meg elszigetelten önmagukban, mivel a kérdésekre adandó válaszok – pl. hogy két DNS minta ugyanazon személytől származik-e – abban a populációban megfigyelt genetikai variációk szerkezetétől és szintjétől függ, melybe a kérdéses személy is tartozik. Ezért az egyedi variációkat mindig populációs összefüggésbe kell helyezni, és megfelelő statisztikai módszert kell alkalmazni a különböző hipotézisek tesztelésére. Ez a statisztikai kiértékelés könnyen igen komplex is lehet, de egyúttal bűncselekmények vagy törvényszéki ügyek kapcsán laikus személyek számára is felfoghatónak és egyértelműnek kell hogy legyen. A statisztikai interpretáción túl további, gyakorlati szempontból is fontos kérdést jelent a DNS vizsgálatok során generált adatok minősége és hibagyakorisága. A keletkezett személyes genetikai adatok kezelése, védelme és törvényeknek megfelelő felhasználása pedig komoly etikai kérdéseket vet fel a DNS adatbázisok kapcsán.

A DNS alapú azonosítással és származástannal foglalkozó igazságügyi genetikát kihívásai, probléma felvetései, módszertani és technikai fejlesztései, eredményei nemcsak passzív felhasználóként, hanem aktív közreműködőként kapcsolják a genetikához mint alaptudományhoz és a jogalkalmazáshoz egyaránt, ugyanakkor nem tekinthető a kettő egyszerű öszszegződéseként. Az igazságügyi genetika önálló tudományterületként működik, és használja fel a genetika módszereit, tudományos eredményeit. Önálló voltát támasztja alá az is, hogy igazságügyi genetikai kutatások a genetika más területére vonatkozó információkat – pl. polimorfizmusok molekuláris struktúrája, DNS-szekvencia adatok, populációgenetikai adatok, stb. – is szolgáltatnak, melyek önmagukban ugyanakkor nem jelentik az igazságügyi genetika célját. Törvényszéki aspektusból nézve a genetika csak eszköz, melynek segítségével a jogi hipotézisek megerősíthetők vagy elvethetők. A tudományterület önállósodása – más, korábban határ- vagy multidiszciplináris tudományágak evolúciójával analóg módon – bizonyos tudományos koncepciók, besorolások és dogmák megváltozásával jár. Habár a változások érvényre jutása eltérő mértékű lehet, az igazságügyi genetika mint az alkalmazott tudományok autonóm tudományterülete napjainkra szinte minden országban az igazságszolgáltatás részévé vált (Brinkmann és Carracedo, 2003), és számos vonatkozásában kiszélesítette a bűncselekmények felderítésének és törvényszéki perek bizonyításának lehetőségét. Biológiai anyamradványok faji-, egyedi (személyi) eredetének meghatározása, bűnügyi helyszínek kapcsolatainak vizsgálata, ismeretlen személyazonosságú holttestek, maradványok azonosítása, leszármazási viszonyok elemzése mellett polgári jogi (pl. apasági teszt), történelmi, kegyeleti (pl. tömegsír) vizsgálatok céljára használható fel. További lehetőségek körvonalazódnak a technológia fejlődésével a kérdéses személyek megnyilvánuló tulajdonságainak (pl. nem, kor, szem- és bőrszín, testfelépítés, egészségi állapot stb.) hátrahagyott nyomokból – akár egyetlen sejtből (Oorschot és Jones, 1997) – történő kiderítésére (Pulker és mtsai, 2007), a megállapított adatok informatikai alapon történő rendszerezésére, adatbázissá szervezésére (bűnügyi DNS nyilvántartások), és nem emberi eredetű anyagok azonosítására (Budowle és mtsai, 2005a és 2005b). A lehetőségek és kockázatok objektív mérlegeléséhez tudományos és szakmai körű vitákra és konszenzusokra van szükség (National Commission of the Future of DNA Evidence, 2001), ami segíthet megelőzni és tompítani a potenciális társadalmi, politikai, vallási ellenérzéseket, jogi és etikai vitákat, melyeket napjainkban a genetikai módszerek más irányú felhasználása is – pl. klónozás – eredményez.

1.1. Az igazságügyi genetika alapjai

1.1.1. Egyedi személyazonosítás – a genetikai profil

Az elmúlt évek, évtizedek nemzetközi kutatásai (pl. Human Genom Project) felderítették a humán genom diverzitását és feltárták azokat a mutációs folyamatokat, melyek a szóban forgó diverzitás hátterében állnak. A kutatások eredményei szerint átlagosan 1250 bázispáronként várható egy szubsztitúciós eltérés (SNP – Single Nucleotide Polymorphism) két különböző eredetű genom kópia között (Jobling és mtsai, 2004), vagyis elméletileg néhány millió SNP típusú eltérés valószínűsíthető két diploid egyén genomjának összevetése során. A bázis szubsztitúciónál lényegesen kevesebb mikroszatellita és miniszatellita típusú variáció létezhet a humán genomban, és ez a több ezer már leírt szatellita DNS marker nagyrésze potenciális jelölt lehet független személyek közötti genetikai különbségek leírására a lényegesen nagyobb allélszámuknak és heterozigotizáció értékeküknek köszönhetően. Egy személy DNS állománya egy másiktól nagyon nagy számú eltérést is tartalmazhat. Ha ezek az eltérések, variációk viszonylag egyszerű módon meghatározhatók, kézenfekvő eljárást jelenthetnek személyek DNS mintából való azonosítására, és annak a konklúzív megállapítására, hogy két biológiai minta származhat-e egyazon személytől. Önmagában véve ugyanis egyetlen lokusz – DNS-szakasz – sem rendelkezik kizárólag egyetlen egyedre jellemző tulajdonsággal, azonban a megfelelő számú polimorf helyen megállapított allélok együttese – genetikai profil – „genetikai személyi számként” értelmezhető. A valós ügyek és a vizsgálatba bevont minták típusa, minősége adja meg azokat a praktikus okokat, melyek jelentősen befolyásolják a vizsgálatokhoz igénybe vett markerek számát és fajtáját.

1.1.2. Az igazságügyi DNS laboratórium és az igazságügyi genetikus

Az igazságügyi DNS laboratóriumban végzett munka során más szempontokat is figyelembe kell venni, mint egy kutató laboratóriumban:

- A személyazonosítás során alkalmazott DNS analitikai technikák és módszerek viszonylag lassan evolválódnak, mivel széleskörű elfogadottságuk a sikeres és megbízható törvényszéki alkalmazáson, a hatékony sztenderdizáltságon és a módszerek validálásán múlik. Nagyobb genetikai adatbázisok viszonylag nagy költséggel és hosszú időkifutással építhetők fel, ezért a már kész genetikai profilokat tartalmazó rendszerek megváltoztatása, újraépítése nem igazán kívánatos.

- A vizsgálatra kerülő DNS minták gyakran degradáltak vagy nyomnyi mennyiségűek, ezért az azonosításukhoz alkalmazott módszerek különösen érzékenyek, és emiatt a minták esetleges kontaminációjával is számolni kell.
- Genotipizálási hibák komoly következményekkel járhatnak, ezért az összes alkalmazott eljárásnak megbízhatónak, a vizsgálatoknak pontosan kivitelezettnek kell lenni.
- Mivel az összes laboratórium ugyanabból a markerkészletből merít, az alkalmazott metodikáknak széleskörű elfogadottsággal kell bírniuk. Újszerű, speciális és technikai-lag nehezen kivitelezhető eljárásokat nem könnyű elfogadtatni.
- Az alkalmazott metodikákat és a statisztikát érintő kérdések és viták gyakran a tárgyalóterem ellenséges és konzervatív légkörében kerülnek megvitatásra, nem pedig egy kevésbé terhelt tudományos ülésen.

Az igazságügyi genetikus szakértő és feladatának konkrét meghatározása a multidiszciplináris tudományterület kapcsán nehezen definiálható. Általánosságban elmondható, hogy legfőbb szerepét a biológiai minták, nyomok és anyagmaradványok vizsgálata jelentí természettudományos technikák objektív alkalmazásával, ezáltal igazolva egy jogi feltevést, egy adott bűncselekményt illetve az azzal fennálló kapcsolatot (Inman, 2001). A tudományos és a gyakorlati törvényszéki alkalmazás között azonban különbséget kell tenni. A kutatásban megfelelő és megbízhatónak tartott eljárásokat az óvatos, tartózkodó jogi szemléletmód nem minden esetben fogadja el törvényszéki szempontból is megbízhatóan alkalmazhatónak. A törvényszéki DNS vizsgálatok eredménye mindig vitatható lesz, és sokszor nem a technológia, sokkal inkább a gyakorló személy (szakértő?) azon rátermettsége, képessége miatt, hogy a vizsgálatokat körültekintően és helyesen vitelezze ki (Butler, 2005). A DNS-profil meghatározás ugyanis nem varázspálca, nem csallhatatlan törvényszéki eszköz, hanem egy összetett vizsgálati rendszer, mely ki van téve emberi hibáknak, és magában foglalja a mintakontamináció, mintacsere valamint a téves szakértői interpretáció lehetőségét is. A hiba-lehetőségek kiküszöbölésének céljából a bizonyítható – dokumentált – szakértés követelmény rendszerének kidolgozása a '90-es évek közepére megtörtént (TWGDAM, 1995). A minőség-irányítás igénye a DNS egyéb törvényszéki alkalmazásaira is kiterjed (Morling és mtsai, 2003, Jobling és Gill, 2004).

1.1.3. A DNS mint törvényszéki bizonyíték

Az egyedhez való kapcsolhatóság miatt a DNS-vizsgálatok elsősorban a személyi bizonyíték-bizonyítás eszközeként jelennek meg és tudatosultak a gyakorlatban. Habár a korábbi mód-

szerekhez képest a DNS alapú vizsgálatok számos előnnyel rendelkeznek, a genetikai tulajdonságok átörökítés esetén illetve az élőlényből való kikerülés vagy a halál beállta után megváltozhatnak (Parsons és Weeden, 1997). A DNS-molekula viszonylagos védettsége és stabilitása a technikai előnyökkel párosulva sok esetben a természetes bomlási folyamatok (Montagna, 1996) és a tévőlegesen megváltoztatott állapot ellenére (Sweet és Sweet, 1995) is megoldást jelenthet. A nemi kromoszómák markerei, így az Y-kromoszóma mint a humán genom kizárólag férfiakra jellemző szakasza (Corach és mtsai, 2001) az erőszakos bűncselekmények férfi elkövetőinek azonosítását tette még eredményesebbé. Ez leginkább azokban az eljárásokban érhető tetten, melyekben a kevert jellegű biológiai nyomok női eredetű sejtek túlsúlya miatt az autoszómás markerek analízise nem informatív. A mitokondriális DNS hipervariábilis szakaszainak átlapoló PCR alapú szekvencia-analízisével a DNS azonosítást a biológiai anyagmaradványok teljes körére kiterjesztették (Holland és Parsons, 1999). Az igazságügyi szempontból fontos megkülönböztetési – kizárási – esély részben a mitokondriális DNS maternális öröklődése miatti alacsonyabb voltát a sejtenkénti magasabb kópiaszám ellensúlyozza, ami szignifikánsan növeli a vizsgálatok sikerességi rátáját. Az extrémén kis mennyiségű DNS-t tartalmazó minták („low copy number DNA”) vizsgálhatósága a mikronyomok hagyományos értelmezését kimutatási érzékenység és a származás vonatkozásában is kiterjesztette (Gill és mtsai, 2000). A PCR technika sajátosságainak megfelelően rövid – kb. 100-150 bázispár – szakaszon belül is polimorf markerekkel (Balogh és mtsai, 2003) vagy az amplikonok méretének redukálásával (Butler és mtsai, 2003) növelhető a vizsgálati érzékenység.

A DNS vizsgálatok eredménye kétféle módon használható fel büntető- vagy törvényszéki eljárások során. Egyrészt a bírósági szakaszban a bíró az igazságügyi DNS szakértő véleményét mint bizonyítékot értékeli, a vád és a védelem pedig igyekszik az eredményeket minden aspektust figyelembe véve részletekbe menően megvizsgálni. Másrészt a DNS bizonyítékokat a büntetőeljárás során csak nyomozati információként felhasználva azok nem feltétlenül jutnak el a tárgyaló terembe (pl. bűnügyi DNS nyilvántartás adatai), mivel rendelkezésre állhatnak egyéb, nem DNS alapú közvetlen bizonyítékok is az ügyben. Mindazonáltal minden esetben figyelembe kell venni azt a tényt, hogy egy helyszíni DNS mintával tökéletesen megegyező gyanúsított DNS-profil nem feltétlenül jelenti a gyanúsított bűnösségét az adott ügyben, mivel az csak annyit demonstrál, hogy a kérdéses személy valamikor valamilyen kapcsolatban volt a helyszínnel.

1.1.4. Mit jelent az egyezési valószínűség (matching probability)?

Két DNS-profil azonosságát, vagyis a DNS bizonyíték súlyát, erejét, az úgynevezett egyezési valószínűséggel (idegen kifejezéssel match probability-vel) mérhetjük. Az egyezési valószínűség annak az esélye, hogy egy a populációból véletlenszerűen kiválasztott DNS-profil meg egyezik a kérdéses DNS-profillal (pl. helyszíni mintával). E definíció alapján felmerül néhány kérdés:

- Mekkora pontossággal kell megadni egy DNS-profil gyakoriságát? A konzervatív becslés szabálya szerint a védelem javára való tévedés elfogadható, de túlzott alá- vagy fölébecslés esetén a valóságtól elrugaszkodott egyezési valószínűségi értékek jöhetnek ki (Lander, 1989). A napjainkban alkalmazott mikroszatellita multiplexek tipikusan $1 : 10^{10}$ - 10^{18} egyezési valószínűségeket adnak.
- A kérdéses profilt milyen populációs adatokkal kell összevetni? Alapvető gyakorlat a „product rule” alkalmazása, vagyis az egyes lókuszokon kimutatott alléleknek a populációban megfigyelt gyakorisági értékeit összeszorozzuk. Amennyiben azonban számottevő linkage disequilibrium, vagy populáció struktúráldás figyelhető meg a populációban a számított profil gyakoriságok tévesek lehetnek.
- Lehetséges, hogy nem a gyanúsított, hanem egy közeli hozzátartozójának a DNS mintája került kimutatásra? Ez a kérdés azért merülhet fel, mert hozzátartozók nagyobb valószínűséggel rendelkeznek a gyanúsítottéval megegyező genotípussal, mint a random módon kiválasztott személyek. Emiatt ezzel a lehetőséggel is számolni kell az eredmények statisztikai interpretációja során.

Az egyezési valószínűség félreértésének tipikus esete a „vád tévedése” (prosecutor’s fallacy), vagyis amikor az egyezési valószínűséget összetévesztik annak az esélyével, hogy a szóban forgó DNS-profil a gyanúsítottól vagy a populáció egy másik független tagjától származik. A DNS-bizonyíték statisztikai kiértékelésének a gyakorlatban legáltalánosabban elfogadott módszere ugyanis a gyakoriságbecsléssel kombinált Bayes-elvű hipotézistesztesztelés (Evett és Weir, 1998, Garbolino és Taroni, 2002). A valószínűségszámítás Bayes-tételének esély formájú leírása szerint ugyanis az alapkérdést: pl. *Mekkora a valószínűsége annak, hogy a DNS-profil XY-től származik?* mindig legalább egy másik, ún. ellenhipotézis valószínűséghez kell viszonyítani: pl. *Mekkora a valószínűsége annak, hogy a DNS-profil az adott népességből véletlenszerűen kiválasztott másik személytől származik?* Az első kérdés általában a vád feltevésére, a második kérdés pedig általában a védelem hipotézisére vonatkozik. A hipotézis és ellenhipotézis teljesülésének feltételezésével számított valószínűségek hányadosát, a

valószínűségi hányadost (LR) adjuk meg az igazságügyi DNS-vizsgálatok statisztikai kiértékelése során. A vérrokonság (pl. apaság) genetikai tesztelése esetében a valószínűségi hányadost rokonsági (apasági) indexnek is hívják. Amennyiben a biológiai nyom csak egy személy genetikai anyagát tartalmazza, akkor a valószínűségi hányados – legegyszerűbb formájában – a DNS-profil egyezés valószínűségének reciprokával egyezik meg.

1.1.5. Az igazságügyi genetika rövid története

Egyedi személyazonosításra azok a biológiai tulajdonságok alkalmasak, amelyek az életkortól, a környezeti hatásoktól függetlenek, nem módosulnak és nem módosíthatók, ugyanakkor kellő mértékű változatosságot mutatnak ahhoz, hogy a vizsgálatok alapján az egyedek között különbséget lehessen tenni. A DNS egyetemes jellege lehetővé teszi, hogy akár testi-, akár szaporítósejteket tartalmazó biológiai minta vagy nyom azonos elvi és gyakorlati alapokon legyen vizsgálható. Habár a DNS molekula első leírása már 1868-ban megtörtént (F. Miescher), csak több évtizeddel ezután és apránként ismerték fel alapvető szerepét az öröklődésben. Körültekintő kutatások azóta igazolták szubsztanciális jellegét a genetikai információ generációról-generációra történő átadásában. Az 1980-as években a molekuláris genetikai technikák az első humán hipervariábilis DNS hosszpolimorfizmus (*DI4S1*) RFLP (*Restriction Fragment Length Polymorphism*) módszerrel való kimutatásával (De Martinville és mtsai, 1982, Balazs és mtsai, 1982) új távlatokat nyitottak a szakértői bizonyítás számára.

A polimeráz lánreakció (PCR) felfedezésének évében (Mullis és mtsai, 1986), 1985-ben került sor a DNS-vizsgálatok első törvényszéki alkalmazására egy nemi erőszaktevő és sorozatgyilkos azonosítására Angliában, bár még RFLP módszerrel (Jeffreys és mtsai, 1985a,b). A legnagyobbbrészt miniszatellita régiókban jelenlévő polimorf szakaszok – multilókuszos próbák – egyidejű vizsgálatával vonalkódszerű mintázat volt nyerhető, melyről a személyazonosítási lehetőség a DNS ujjlenyomat – „DNA fingerprinting” – elnevezést nyerte (Gill és mtsai, 1985). A gyakorlatban a multilókuszos próbákat gyorsan felváltották az egylókuszos próbák, melyek a szakirodalomban *VNTRs* – „Variable Number of Tandem Repeats” néven váltak ismertté, és az igazságügyi gyakorlatban a '80-as évek végére terjedtek el (Gill és Werrett, 1987). A PCR amplifikáció igazságügyi alkalmazásáról először Higuchi és munkatársai (1988) számoltak be, amikor egyetlen hajszálból kiindulva a *HLA-DQA1* lókusz szekvencia-polimorf régióját genotipizálták. A PCR termékek az RFLP-től eltérően diszkrét tulajdonságokként detektálhatók, egy PCR reakcióban egyidejűleg több lókusz is felsokszorozható, ami akár szekvencia-, akár hosszpolimorf lókuszok vizsgálatánál használható fel.

A mikroszatelliták, elterjedt néven *STR*-ek (*Short Tandem Repeat*) igazságügyben való alkalmazhatóságát a PCR-technika gyors elterjedése tette lehetővé (Polymeropoulos és mtsai, 1991). 1991-ben leírták az STR markerek fluoreszcens jelölését és detektálását (Edwards és mtsai, 1991), 1993-ban pedig kifejlesztették a genetikai nem meghatározását az Amelogenin lókuszt PCR-fragmensének fluoreszcens detektálásával (Sullivan és mtsai, 1993). A DNS polimeráz enzim csúszásából eredő, legfőképp a di- és trinukleotid ismétlődések tipizálásánál jelentkező bizonytalanságok miatt (Walsh és mtsai, 1996) az ezredfordulóra a 4-5 bázispáros repetíciókat tartalmazó lókuszok – 6-16 lókuszt együttes, multiplex analízisére alkalmas, az RFLP-ket is felülmúló megkülönböztetési valószínűséggel rendelkező – vizsgálati rendszerei váltak kereskedelmileg hozzáférhetővé. PCR termékek ciklus szekvenálásával és fluoreszcens jelölésével megnyílt az út a mitokondriális DNS (Anderson és mtsai, 1981) szekvencia polimorfizmusainak vizsgálata felé is (Sullivan és mtsai, 1992, Wilson és mtsai, 1995). A teljes mitokondriális genom polimorfizmusainak vizsgálatával (Parsons és Coble, 2001) jelentősége fokozódik, de a rendkívül érzékeny technika rendkívül nagy körütekintést is igényel (Herrnstadt és mtsai, 2003). A mitokondriális genom kódoló szakaszán lokalizált „*cytochrome b*” gén egy jól körülhatárolt szakasza genetikailag fixálódott fajspecifikus szekvenciát mutat, így az adott génszakasz szekvencia-elemzésével a biológiai minta faji eredete megállapítható lett (Parson és mtsai, 2000).

Hazánkban 1992-ben a Bűnügyi Szakértői és Kutatóintézetben elindultak az első igazságügyi célú DNS-vizsgálatok (Pádár és Woller, 1993), melynek eredményeként először egy emberölési ügyben született DNS vizsgálaton alapuló igazságügyi szakértői vélemény. 1995-1996 folyamán jelentek meg az első tudományos közlemények nemzetközi szaklapokban, melyek a magyarországi igazságügyi célú DNS vizsgálatok, populációs felmérések eredményeiről számoltak be (Woller és mtsai, 1995; Füredi és mtsai, 1995 és 1996; Csete és mtsai, 1996).

A nemzetközi szinten sztenderdizált és igazságügyi célra validált, populációs háttérvizsgálattal alátámasztott STR markerek magukba foglalják ma már az autoszómákat és az ivari kromoszómákat is, összességükben pedig alkalmassá váltak nemzeti- és nemzetközi adatbázisok kialakítására (Krenke és mtsai, 2002, Holt és mtsai, 2002). 1995-ben Nagy-Britanniában felállították a világ első bűnügyi DNS-profil nyilvántartását. Az Egyesült Államok Szövetségi Nyomozó Hivatala (FBI) által 1997-ben definiált 13 STR törzs-lókuszt nyilvántartó rendszerének (*CODIS*) elterjedése után megindult a nemzeti adatbázisok nemzetközi szerveződése (Interpol DNA Unit, 2003). Az évtized végére az STR vizsgálatokon alapuló igazságügyi genetikai szakértés Magyarországon is elterjedté vált (Kozma és mtsai, 1998; Lászik és mtsai,

2001), és 2000-ben hazánkban is hatályba lépett a bűnügyi DNS-profil nyilvántartást szabályozó törvény (1999. évi LXXXV. Törvény). A '90-es évek közepének nagy nemzetközi publicitással bíró bűnügyei – pl. O. J. Simpson-ügy – tovább fokozták a már DNS-vizsgálatokkal szemben támasztott elvárásokat (Council of Europe, 1993). Az ezredfordulóra felállt a nemzetközi online Y-STR haplotípus referencia adatbázis (Roewer és mtsai, 2001).

Az autoszómás és ivari kromoszómás hossz-polimorfizmusok, a mitokondriális DNS feltárt szekvencia-polimorfizmusai után a figyelem középpontjába a genomban előforduló több millió *SNP* analízise került (Phillips és mtsai, 2004). Habár limitált polimorfizmusuk miatt a megfelelő kizárási valószínűség eléréséhez az STR lókuszokhoz viszonyítva lényegesen több *SNP* egyidejű meghatározására van szükség (Chakraborty és mtsai, 1999), nagy számuknak és a rövid cél DNS-fragmensnek köszönhetően optimális megoldást jelenthetnek degradált anyagmaradványok vizsgálatához. További előnyükként említhető, hogy jól adaptálhatóak a legújabb automata vizsgálati platformokhoz (Sanchez és mtsai, 2006).

A mitokondriális DNS szekvencia adatbázisok minőségirányítási szempontú felülvizsgálata után (Bandelt és mtsai, 2001 és 2004; Yao és mtsai, 2004) 2006-ban elindult az EDNAP (European DNA Profiling Group) által életre hívott EMPOP adatbázis (Parson és Dür, 2007), melynek célja mitokondriális DNS kontroll régió referencia haplotípusok gyűjtése a világ minden részéről, az adatok minőség ellenőrzése és online hozzáférés biztosítása az alkalmazók számára. Az EMPOP adatbázisban 2004-től kezdődően mintegy 900 teljes mitokondriális DNS kontroll régió haplotípus került Magyarországról is feldolgozásra.

1.2. Az igazságügyben használatos genetikai markerek

1.2.1. A mikroszatelliták általános jellemzése és nevezéktana

A mikroszatellita lókuszok (hétköznapiabb nevükön STR-ek, Short Tandem Repeats) olyan nem átíródó DNS-régiók, ahol 2-6 bázispárból álló rövid szakaszok tandem módon ismétlődnek egymásután. A repeat egységek ismétlődéseinek száma az egyes egyének között nagy változatosságot mutat, ez a polimorf jelleg teszi őket alkalmassá genetikai személyazonosítási célokra (Butler, 2005). A mikroszatellita lókuszok eltérő alléljai egymástól hosszúságjukban különböznek, ezért ezeket a típusú polimorfizmusokat a hossz-polimorfizmusok közé sorolják. A hosszbeli változatokat kialakító tandem-repetíciók eredete (melyeket esetenként SSR-nek – „Simple Sequence Repeat”-nek is neveznek), és a genomban betöltött pontos funkciói

érdemben ma még nem teljesen tisztázódtak, de eltéréseik és öröklődő jellegük az egyedek megkülönböztetésére ad módot.

Az eukarióta genomban nagyon sok helyen találhatók ismétlődő szekvenciák, a mai napig polimorf mikroszatelliták ezreit azonosították a humán genomban is. A humán genom kb. 3%-át kitevő mikroszatellita lókuszok átlagosan tízezer nukleotidként figyelhetők meg (Edwards és mtsai, 1991). Kódoló szakaszokban intronok részeként és génen kívüli régiókban egyaránt megtalálhatók. Az ismétlődő egységek hosszának megfelelően di-, tri-, tetra-, penta-, hexa repeateket különböztetünk meg (Butler, 2005). Új STR markereket általában DNS-szekvencia adatbázisokban lefolytatott kereséssel (Weber és May, 1989) vagy molekuláris biológiai izolációs módszerek révén fedeznek fel (Edwards és mtsai, 1991). Az STR szekvenciák nem csak a repeatek számában, hanem a struktúrájukban is variabilitást mutatnak. A hossz-polimorfizmus esetenként szekvencia variabilitással párosul, mely alapján mikroszatellita DNS szekvenciák többféleképpen is csoportosíthatók (Weber, 1990, Urquhart és mtsai, 1994). Az azonos hosszúságú és szekvenciájú egységekből álló egyszerű repeatek mellett megkülönböztethetünk összetett, komplex, hipervariábilis szakaszokat. Az összetett repeatek két vagy több különböző, egymással szomszédos egyszerű repeatet tartalmaznak. A komplex repeatekben a különböző, változó hosszúságú repeat-egységeket változó méretű közbeiktatott szekvencia szakítja meg. Nem minden STR allél mérete feleltethető meg a repeat-ek egész számú többszöröseként. A repeat- vagy a repeateket határoló régióban egy vagy több megváltozott nukleotidot hordozó mikrovariáns allélok a konszenzus allél méretétől eltérhetnek. A komplex hipervariábilis repeatek szekvenciájukban és méretükben is eltérő, nem konszenzus allélek sokaságát tartalmazzák (Butler, 2005). A különböző allélváltozatok (Ellegren, 2004) kialakulása a molekuláris szinten egyetlen nukleotidot illetve az ismétlődések számát változtató mutációknak – DNS replikációs hibák (pl. a DNS polimeráz csúszása), a javító mechanizmus nem tökéletes működése – köszönhető (Nadir és mtsai, 1996).

Az STR markerek jelenleg kulcsszerepet töltenek be igazságügyi célú alkalmazások területén. Ezt nagy számuk és viszonylag magas fokú polimorfizmus fokuk mellett egyszerű PCR amplifikálhatóságuknak köszönhetik, melynek multiplex formában való alkalmazásával eddig használhatatlannak és ebből kifolyólag lényegtelennek tűnő biológiai anyagmaradványok is vizsgálhatóvá váltak az igazság kiderítése céljából. Az STR lókuszok multiplex PCR-alapú szimultán vizsgálata lehetőséget teremtett a genetikai személyazonosítás gyors (akár néhány órán belüli) kivitelezésére (Kimpton és mtsai, 1996) akár néhány pikogrammnyi, azaz egyetlen sejt genetikai állományát tartalmazó DNS mintából is (Findlay és mtsai, 1997), bár a minden tekintetből megbízható rutinszintű sorozatvizsgálatokhoz jelenleg ennél egy-két nagyság-

renddel több, kb. 100-1000 pikogramm örökítő anyag (kb. 20-200 sejt DNS-tartalma) szükséges. A viszonylag gyors és egyszerű kivitelezhetőség, a megbízható és szigorú minőségbiztosítási követelményeknek is eleget tevő vizsgálati stabilitás, az igazságügyi DNS laboratóriumok közötti kompatibilitás, a degradált, töredezett, kis mennyiségű mintáknál tapasztalt vizsgálati eredményességük és általában egyértelmű azonosíthatóságuk azok a jellemzők, melyek alapján az STR markerek a személyazonosítás fő genetikai markerévé váltak.

Napjainkig igazságügyi célokra mintegy száz humán polimorf mikroszatellitát jellemeztek és hitelesítettek (Ruitberg és mtsai, 2001; Butler, 2005), melyek esetében lókuszonként mintegy 6-30 allélváltozat figyelhető meg a humán népességben. Ezen markerek közül jelenleg 16 autoszómális marker képezi a különböző nemzeti bűnügyi DNS-profil nyilvántartások alapját, melyek között az Interpol, az ENFSI (European Network of Forensic Science Institutes) és FBI vizsgálati markerei is megtalálhatók. Az STR lókusztok nemzetközileg kompatibilis vizsgálata és hitelesítése Magyarországon is több éve zajlik (Füredi, 2003; Pádár, 2006).

Az STR allélek meghatározásához elengedhetetlen az alkalmazott kontrollok és a detektált tulajdonságok pontos definiálása, ami nemzetközileg egységesített nevezéktan segítségével történik. Az igazságügyi genetikai vizsgálatok során analizált STR lókusztok és alléljeik nevezéktanát a Nemzetközi Igazságügyi Genetikai Társaság (ISFG – International Society of Forensic Genetics) több ajánlásban is összegezte (Bär és mtsai, 1997; Olaisen és mtsai, 1998; Gill és mtsai, 1997, 2001; Gusmão és mtsai, 2006). A legfontosabb szabályok a következők:

- A szekvenciát mindig 5'→3' irányban kell olvasni, a tipizáláshoz használatos DNS szál pedig génben lokalizált marker esetében a kódoló szállal megegyező, nem kódoló szekvencia esetén az irodalomban eredetileg leírt, illetve a nyilvános adatbázisba először bejegyzett kell hogy legyen.
- A nevezéktan amennyire csak lehetséges kövesse a D#S# rendszer szabályait. Mivel számos nem standard elnevezés is használatos világszerte, ezek megváltoztatása nem célszerű, de az új lókusztok leírásánál már a D#S# rendszer használata tanácsolt annak érdekében, hogy elősegítsék az egységesítés folyamatát.

Az allélikus nevezéktanra vonatkozó ajánlások fő pontjai:

- Az elnevezés a teljes ismétlődési egységek (repeats) száma szerint történik.
- Amennyiben a szekvencia egy nem teljes repeatet is tartalmaz [mikrovariánsok, köztes (inter-) allélok], úgy a teljes egységek száma után egy ponttal elválasztva kell feltüntetni a hiányos repeat meglévő bázisainak számát.
- Megjelenhetnek köztes allélok az ún. flanking (határoló) régiók egyszerű inszerciós vagy deléciós mutációjának köszönhetően is.

1.2.2. A mitokondriális DNS mint genetikai marker

Az elmúlt években a mitokondriális DNS (mtDNS) analízis igazságügyi célú alkalmazása és jelentősége rendkívüli mértékben megnőtt, mivel olyan előnyös tulajdonságokkal is rendelkezik, melyekkel a nukleáris (sejtmagi) DNS állomány nem. A mtDNS átlagosan mintegy ezer példányban lelhető fel egy-egy szomatikus sejtben (Sato és Kuroiwa, 1991) a mitokondriumok kettős membránjával körülvéve, petesejtekben a példányszám akár a 100 000 kópiát is elérheti (Chen és mtsai, 1995) szemben a sejtmagi DNS két példányával (homológ kromoszómapárok). Ennek a tulajdonságának köszönhetően vizsgálati érzékenysége nagyságrendekkel nagyobb a nukleáris DNS-nél, és ennél fogva a nagymértékben degradálódott, bomlott állapotú biológiai minták (pl. csontmaradványok, szörképletek) azonosítására is alkalmas. Kizárólag maternálisan öröklődik (Giles és mtsai, 1980; Parsons és mtsai, 1997; Sutovsky és mtsai, 1999), nem rekombinálódik (Howell N, 1997; Kivisild és mtsai, 2000), mely tulajdonságok jelentősen megnövelik biológiai maradványok beazonosításához referencia mintaadóként számításba vehető személyek számát, hiszen az anyai ág összes élő személye ugyanazzal a mitokondriális DNS haplotípussal rendelkezik az esetleges mutációs eseményeket nem számítva. Ez az oka ugyanakkor annak is, hogy egy mitokondriális DNS haplotípus nem lehet egyedi, ami az egyezési valószínűségek szignifikanciáját csökkenti. A sejtmagi DNS állományhoz képest a mitokondrium kisebb mértékű DNS javító mechanizmussal rendelkezik, ami magasabb mutációs rátát eredményez. A mitokondriális DNS replikációját végző polimeráz enzim „proof-reading” hiányossága az osztódás során növeli a mutációk gyakoriságát, viszont a molekula cirkuláris jellege miatt kevésbé fogékony exonukleázok támadásának, ami segít megőrizni kompakt állapotát és ezzel együtt vizsgálhatóságát, azonosíthatóságát. A sejtmagi DNS-nél kb. tízszer nagyobb mutációs ráta eredményeképpen jöhetnek létre a különböző variációk az egyébként azonos anyai vonalakból. Az endoszimbionta elmélet (Gray, 1992) szerint evolúciósan kialakult mitokondriális genom a funkcionális protein kódoló génjeinek nagy részét elveszítette, illetve azok ma már a nukleáris genomban lokalizálódnak (Lang és mtsai, 1997).

A humán mitokondriális genom DNS szekvenciáját először 1981-ben írták le (Anderson és mtsai, 1981), melyet később újra analizáltak és pontosítottak (Andrews és mtsai, 1999), és az első leírója neve illetve a laboratórium helyszíne után szokás Anderson vagy Cambridge referencia szekvenciaként (CRS) nevezni. A mai humán mitokondriális genom egy ~16569 bp hosszú cirkuláris DNS molekula, ami egy pirimidin bázisokban gazdag „könnyű” és egy túl-

nyomórészt purin bázisokból álló „nehéz” szálból áll (CsCl gradiens centrifugálás által). A molekula 15447 bp hosszú kódoló szakaszán 37 gén lokalizálódik intronok nélkül összetömörítve, a kódoló régióból összesen 55 nukleotid nem íródik át. A 37 génből 22 db transzfer RNS-t kódol, 2 gén riboszómális RNS-t (12S és 16S rRNS), 13 gén pedig az oxidatív foszforilációban és ATP szintézisben közreműködő enzimeket. A mitokondriumok normális működéséhez szükséges összes többi gén a sejtmagi DNS-ben lokalizálódik. Számos kóros és kevésbé patogén pontmutációt, deléciót, duplikációt azonosítottak a mai napig a kódoló régióban, melyek jelentős hányada az öregedéssel is szoros összefüggést mutat (Arnheim és Cortopassi, 1992; Wallace, 1992). A molekulában természetes variációkként megfigyelhető egy-két bázist érintő inzerációs/deléciós mutációk miatt a molekula összes nukleotid száma egyénenként, sőt sejtenként, molekulánként is (hossz-heteroplazmia) változhat.

A mitokondriális DNS molekulát a kódoló régió kívül egy kb. 1122 bázispár hosszúságú nem kódoló szakasz az ún. kontroll régió alkotja, mely elnevezés az ebben a régióban lokalizálódó policisztronikus RNS transzkripció promotereire és a nehéz szál replikációs origójára utal. A kontroll régiót gyakran „D-loop” szakasznak is nevezik, utalva a DNS replikáció során elketronmikroszkóp alatt látható képére. Mivel a D-loop nem kódol génterméket kisebb rajta a szelekciós nyomás, ezért az egyedek között megfigyelt nukleotid variabilitás és polimorfizmus lényegesen nagyobb fokú mint a kódoló régióban. A standard referencia szekvencia sorozámozása önkényesen a kontroll régió közepén kezdődik kettéválasztva azt egy-egy 16024-16569 és 1-576 nukleotibázisok közötti szakaszra. Az igazságügyben általánosan vizsgált szekvenciák az ún. hipervariábilis régiók is ezekre a szakaszokra esnek: a hipervariábilis 1. régió (HV1) a 16024-16365 szakaszra, a hipervariábilis 2. régió (HV2) a 73-340 bázistartományra terjed ki, bár ezek a határok nem szigorúan meghatározottak és az alkalmazott vizsgálati módszertől és a vizsgáló laboratóriumtól függ.

A mitokondriális DNS variánsok (haplotípusok) megállapítása és leírása a standard referencia szekvencia (Anderson vagy CRS) könnyű szálához viszonyítva történik, így pl. a 16126 nukleotid pozícióban észlelt citozin bázist 16126C jelzéssel írjuk le (vagy T16126C). Magától értetődve a referencia szekvenciától eltérést nem mutató nukleotid bázisokat nem kell jelölni. A mtDNS haplotípusok pontos meghatározása nemzetközileg egységesített nevezéktan segítségével történik. Az igazságügyi genetikai vizsgálatok során kapott DNS szekvenciák leírásának nevezéktanát, a vizsgálatok végzéséhez elengedhetetlen minőségbiztosítási követelményeket, a szekvencia eredmények interpretálásának szabályait az ISFG és az EDNAP (European DNA Profiling group) ajánlásokban, útmutatókban összegezte (Bär és

mtsai, 2000; Carracedo és mtsai, 2000; Tully és mtsai, 2001), melyek közül itt csak a legfontosabbakat emelem ki:

- Recens szennyeződések minimalizálása érdekében egy szigorú, több pontból álló előírást javasolnak betartani. Ezek közül a legfontosabbak:
 - Pre- és poszt PCR helységek térbeni elkülönítése.
 - A kérdéses és az összehasonlító mintákat elkülönítve kell vizsgálni.
 - A használati eszközök, vegyszerek, oldatok DNS mentesítése (pl. UV besugárzás).
 - Reagens és negatív kontrollok alkalmazása a teljes vizsgálati procedúrában.
 - Amennyiben lehetséges párhuzamos vizsgálatot kell végezni a kérdéses mintából.
 - A teljes laboratóriumi személyzet mitokondriális DNS szekvenciája rendelkezésre álljon a kapott eredmények ellenőrzése végett.
 - A laboratórium és teljes személyzete vegyen részt rendszeres jártassági és minőségellenőrzési teszteken.
- Az analizált minták haplotípusát mindig a referencia szekvenciához (CRS) viszonyítva kell megadni, a referenciától való eltéréseket lista formában felsorolva úgy, hogy megfeleljen az IUPAC (International Union of Pure and Applied Chemistry) kódjainak.
- Homopolimer szálak esetén (pl. C-stretchek) az ins/del mutációk pontos pozíciója nem ismert. Az általánosan elfogadott szabály szerint ilyenkor a homopolimer szál legvégén történt a variabilitást okozó mutáció (pl. a 311-315 C-stretch esetén 315.1C).
- Minden mintán mindkét (könnyű és nehéz) DNS szál szekvenálását el kell végezni a lehetséges szekvenálási hibák kiküszöbölése érdekében.
- Ha egy adott nukleotid pozícióban mindkét szálon a háttérjelektől tisztán elkülöníthetően kettő különböző bázis figyelhető meg azt heteroplazmiaként kell értelmezni és az IUPAC/IUB kódok szerint kell leírni.

Igazságügyi aspektusból a *heteroplazmia* jelensége az egyik leginkább problémás jellege a mitokondriális genomnak. Egyénen belül ugyanis a mitokondriális DNS állomány diverzifikálódhat, mivel az egyes DNS molekulák a sejtekben egymástól függetlenül replikálódnak (nincsenek kitéve a mitózisnak és a meiózisnak), egymással nem rekombinálódnak, és az osztódásuk során a mutációs ráta lényegesen nagyobb, mint a nukleáris genomé. Kettő (vagy több) mitokondriális DNS típus szubpopulációjának jelenlétét egy egyénen belül *heteroplazmiának* nevezzük. Ezek a szekvencia variánsok egymástól függetlenül replikálódnak és szegregálódnak. Bár a heteroplazmia kimutathatósága elsősorban az alkalmazott vizsgálati és detekciós metodikától (Hanekamp és mtsai, 1996; Steighner és mtsai,

1999; Brandstätter és Parson, 2003a) valamint a vizsgált szövettípustól függ (Bendall és mtsai, 1997; Hühne és mtsai, 1999), és a populációban megfigyelt relatív gyakorisága viszonylag alacsony: kb. 2-8 % (Comas és mtsai, 1995; Bendall és mtsai, 1996; Brandstätter és mtsai, 2004a), viszont mindenféleképpen számításba kell venni az igazságügyi azonosító vizsgálatok során, melyek eredményeinek interpretálását komplikálhatja. Bizonyos esetben egy-egy bizonyíték erősségét is növelheti, mint például az első igazságügyben - II. Miklós cár és családjának földi maradványainak azonosítása során - dokumentált pont mutációs heteroplazmia esetén (Gill és mtsai, 1994; Ivanov és mtsai, 1996). A szubsztitúciós heteroplazmiákon kívül gyakran megfigyelhető a policitozin szálak (mind a HV1 mind a HV2 régió) hossz-heteroplazmiája is (Bendall és Sykes, 1995, Marchington és mtsai, 1997). A heteroplazmiát illetően az a kérdés is felmerül, hogy egy adott személy milyen szinten heteroplazmiás. A többezer milliárd mtDNS molekula (amennyivel mindannyian rendelkezünk) nagy valószínűséggel különbözik egymástól valamilyen mértékben és arányban, viszont ezen eltérések kimutathatósága a különböző haplotípusok megoszlásától és a detektálás érzékenységétől függ.

Az igazságügyi genetikán kívül számos további tudományág tanulmányozza a mitokondriális DNS variációkat. Orvosi genetikával foglalkozó szakemberek számos betegség és kórkép okaként írták le a mtDNS valamilyen mutációját (Wallace és mtsai, 1999). Evolúcióbiológusok és filogenetikusok fajok közötti és belüli variációk vizsgálatával tárják fel a rokonsági viszonyokat. A széles nyilvánosság számára is izgalmas eredmények születtek a neandervölgyi ember maradványainak vizsgálata során (Krings és mtsai, 1997) vagy például a „mitokondriális Éva” meghatározásával (Cann és mtsai, 1987; Excoffier és Langaney, 1989; Watson és mtsai, 1997).

1.2.3. DNS-profil és szekvencia adatbázisok

A DNS vizsgálatok széleskörű kriminalisztikai felhasználásának általánosan kedvező tapasztalatai alapján egyes országok (pl. Nagy-Britannia, USA) már a 90-es évek elején megkezdték a bűnüldözést segítő helyi DNS adatbázisok létrehozását. Ezekkel párhuzamosan a molekuláris biológiai technikák fejlődése és a különböző biológiai kutatások eredményeképpen populációs DNS szekvencia adatbázisok jöttek létre, melyek célja elsősorban a további kutatások elősegítése volt. A lehetőségek és a szakmai elvárások nemzetközi szakmai szervezetekben történő egyeztetését követően az 1990-es évek közepétől kezdődően a világ számos országában kezdődött meg a nemzeti/nemzetközi bűnüldözést hatékonyan segítő különböző szintű

DNS adatbázisok felállítása. E folyamat intenzitását jól szemlélteti, hogy 2002-ig 41 országban indult el nemzeti bűnügyi DNS-adatbázis. Napjainkban a bennük tárolt információ döntő többségét az ötödik generációs – multiplex PCR és fluoreszcens detektáláson alapuló – technika, vagyis az STR alapú DNS-profil alkotás révén nyert adatok teszik ki. A DNS-adatbázis működése által elsősorban olyan rutin bűnügyek megoldásához nyújthat segítséget, amelyekben az elkövető ismeretlen. Ilyen esetekben gyakran marad hátra valamilyen DNS vizsgálatra alkalmas biológiai anyagmaradvány a bűncselekmény helyszínén.

A DNS-adatbázisok hatékony működésének elengedhetetlen feltétele a bennük tárolt információ megbízhatósága, bármilyen az adatbázisba bekerülő hiba súlyos következményekkel járhat. Az 1. típusba sorolt hibák esetén egy ártatlan személy tévesen keveredik gyanúba, a 2. típusú hiba pedig a „bizonyíték elvesztését” és így a valódi elkövető téves kizárását eredményezheti. Emiatt érthető meg, hogy a DNS-adatbázisok működtetése kapcsán rendkívül magas szintű minőségbiztosítási követelményrendszer betartására van szükség.

A hazai DNS-profil nyilvántartás jogszabályi feltételeit, illetve működési kereteit a bűnügyi nyilvántartásról és a hatósági erkölcsi bizonyítványról szóló 1999. évi LXXXV. törvény, valamint a törvény végrehajtásával kapcsolatban kiadott 7/2000. (II. 16.) és 8/2000 (II. 16.) BM-IM-PM rendelet határozza meg. A törvényi rendelkezések alapján meghatározott bűncselekmény kategóriákban a gyanúsított személy köteles DNS mintát szolgáltatni, és ezekből a személyi típusú mintákból nyert DNS-profilokat a bűncselekmények helyszínén fellelt mintákkal együtt egy közös bűnügyi célú DNS-adatbázisban kell elektronikus módon tárolni. A DNS-profilok kezelését, az automatikus kereséseket és lekérdezéseket egy erre a célra fejlesztett számítógépes program végzi, ami jelenleg Magyarországon az FBI által fejlesztett CODIS (Combined DNA Indexing System). A mind teljesebb szakmai felhasználhatóság érdekében egyre szorosabb együttműködésre lesz szükség a bűnügyi célú nyilvántartások és az eredmények szakmai interpretálásához nélkülözhetetlen web-alapú nemzetközi populációs DNS-adatbázisok között.

Populációs adatbázisok jelentős szerepet játszanak bűncselekmények vagy származástani ügyek kapcsán felmerülő genetikai profilok és haplotípusok előfordulási gyakoriságának becslésében. Ezekben az adatbázisokban is kiemelt fontosságú az adatok minősége annak érdekében, hogy a véletlen egyezésre egy megbízható becslést lehessen adni. 1994-ben indult el az európai populációk Y kromoszómás genetikai struktúráltságát felmérni hivatott nemzetközi program, amely mára kiszélesedett és világméretűvé vált, eredményeit az on-line hozzáférésű YHRD (Y Haplotypes Reference Database: <http://ystr.charite.de>, Willuweit és mtsai, 2007) populációs adatbázis tartalmazza. Az adatbázis jelenleg az Föld összes népességében

előforduló Y-STR haplotípusok legnagyobb gyűjteménye, a 2007. július 25-i állapot szerint 447 populációból mintegy 51.253 férfi 9-lókuszos, ún. minimál haplotípusát tartalmazza.

Az Y kromoszóma haplotípusokhoz hasonlóan, a mtDNS haplotípusok populáción belüli előfordulási gyakorisága sem állapítható meg az egyedi allélek (nukleotid variánsok) megfigyelt gyakoriságának multiplikációjával. A helyes módszer szerint a statisztikai interpretáció során azt kell megadni, hogy a kérdéses haplotípus hány példányban található meg az adatbázisban („*counting method*”; Parsons és Holland, 1999; Tully és mtsai, 2001). A 2006-ban indult EMPOP (EDNAP mtDNA Population Database) mitokondriális DNS adatbázis célja teljes kontroll régió referencia haplotípusok gyűjtése a világ minden populációjából, az adatok minőségellenőrzése és online hozzáférés biztosítása az alkalmazók számára (Parson és Dür, 2007). A mitokondriális DNS szekvencia adatok számos minőségellenőrzési lépésen keresztül, filogenetikai és matematikai vizsgálatok után kerülhetnek be az adatbázisba. Az EMPOP adatszisztemben (<http://www.empop.org>) a 2007. július 25-i állapot szerint 22 populációból 5173 személy teljes kontroll régió haplotípusa található meg, melyből 777 a magyar populációkból származó szekvencia.

1.3. Célkitűzések

Mikroszatellita markerek vizsgálatán alapuló igazságügyi genetikai szakértés egészen 1994-ig nem létezett Magyarországon, mint ahogy a magyar népesség genetikai összetételét sem vizsgálták igazságügyi genetikai aspektusból mindaddig STR polimorfizmusok bevonásával. A magyar népességre vonatkozóan csak a 90-es évek közepétől kezdődően kezdtek felállni az első STR markerekre épülő allél- és profilgyakorisági adatbázisok (Füredi és mtsai, 1995; Csete és mtsai, 1996; Kozma és mtsai, 1998), azonban akkor mégcsak néhány, korlátozott lókuszkészletre vonatkozóan. A mitokondriális DNS polimorfizmusain alapuló első igazságügyi genetikai vizsgálatok 2001-re datálhatók Magyarországon a Bűnügyi Szakértői és Kutatóintézetben, bár a módszer nemzetközi szakirodalomban való leírása és validálása a 90-es évek közepére megtörtént (Sullivan és mtsai, 1992; Piercy és mtsai, 1993; Wilson és mtsai, 1995ab). A 2000-es évek elején indultak el azok a nemzetközi kooperációban szervezett kutatások az AFDIL (Armed Forces DNA Identification Laboratory, Rockville, MD, USA) és az Innsbrucki Egyetem Igazságügyi Orvostani Intézetének (Gerichtsmedizin Innsbruck – GMI) bevonásával, melyek a magyar populáció mitokondriális DNS szintű genetikai heterogenitását célozta felmérni.

Munkám alapvető célja ezért humán multilokuszos STR-profilok és mitokondriális DNS teljes kontroll régió haplotípusok generálása és hitelesítése volt a magyar populációban. A populációs (szub-populációs) minták felmérésével kapott (STR és mitokondriális) DNS-profilok genetikai analízise által olyan magyar populációs referencia adatbázis(ok) létrehozása volt a célom, amelyek segítségével az STR markerek és a mtDNS szekvenciák analízisén alapuló DNS azonosító vizsgálatok bizonyító ereje populációstatisztikailag megbecsülhetővé válhat a magyarországi törvényszéki eljárások során.

1.3.1. A populációs minták összeállítása

Az igazságügyi genetikában alkalmazott DNS markerek hitelesítésének részét képezik a populációs felmérések. Allél- és profilgyakorisági adatbázisok nélkül az STR vizsgálatok, haplotípus adatbázisok nélkül a mitokondriális DNS analízis törvényszéki bizonyító erejének statisztikai interpretációja nem lehetséges. A nemzetközi ajánlások szerint globális vizsgálatok mellett szükség van regionális felmérésekre is. A felmérésekbe bevont populációs mintáknak lehetőség szerint minél jobban reprezentálnia kell a vizsgált népesség, szub-populáció genetikai összetételét, különös tekintettel azokra a populációkra, akik szeparáltabb közösségeket alkotnak (NRC, 1996; Tully és mtsai, 2001).

A célom mindenekelőtt egy olyan magyar populációs referencia adatbázis létrehozása volt, melynek segítségével az STR-profilalkotáson alapuló DNS azonosító vizsgálatok bizonyító ereje populációstatisztikailag megbecsülhető a magyarországi törvényszéki eljárások során. Ugyanez a referencia adatbázis képezheti az alapját egy hazai mitokondriális DNS haplotípus adatbázisnak is, mely további közösségekből származó, szub-populációs eredményekkel kiegészülve reprezentálhatja a magyar populáció genetikai heterogenitását. Mindezek nyomán a referencia populációs mintát a budapesti kevert (heterogén) populáció véletlenszerűen kiválasztott személyeiből kívántam összeállítani. Korábbi tanulmányok (Czeizel és mtsai, 1991; Guglielmino és Bérés, 1996) már felhívták a figyelmet arra, hogy a magyar népesség genetikailag erősen szeparált és egyáltalán nem homogén. Emiatt a tanulmány kapcsán az a kérdés vetődik fel, hogy a budapesti referencia populációs adatbázisból nyert adatok milyen feltételekkel és korlátokkal alkalmazhatók a magyar népességre. Ennek a tesztelésére további magyar, de nem feltétlenül csak magyarországi lokalizációjú populációs minta felmérését végeztük el az analízisbe bevont STR és mtDNS markerekkel, mindösszesen mintegy 1200 személyt bevonva a populációs felmérésbe.

A populációs minták gyűjtésében személyesen nem vettem részt, azt minden egyes mintacsoport esetében szakmailag kompetens (szakorvos) személyek végezték. A mintagyűjtés a mintaadó személyek beleegyezésével történt, laboratóriumunk a vér- és/vagy szájnyálkahártya-törlet mintákat név és egyéb személyazonosító adat nélkül, csak a nem feltüntetésével kapta meg a kollaboráló munkacsoportoktól. A populációs minták a következők voltak:

1. Budapesten gyűjtött, általános jellegűnek tekinthető populációs minta (223 személy, jelölése BuCa), mely referencia populációként szolgált a felmérésekben. A budapesti populációs minta kevert jellegénél fogva reprezentálhatja az átlag magyar népességet, mivel etnikai hovatartozás alapján nem szelektált.
2. Baranya megyében roma személyektől gyűjtött populációs minta (206 személy, jelölése BaRo).
3. A kelet-magyarországi (Debrecen) régióban roma személyektől gyűjtött populációs minta (110 személy, jelölése DeRo).
4. Budapesten élő askenázi zsidó személyektől gyűjtött populációs minta (178 személy, jelölése BuAs).
5. Csíkszeredában (Miercurea Ciuc, Románia) gyűjtött székely populációs minta (257 személy, jelölése Székely).
6. Gyimesfelsőlokon (Lunca de Sus, Románia) gyűjtött gyimesi csángó populációs minta (220 személy, jelölése Csángó)

Korábbi, a klasszikus vércsoportokra és fehérjemarkerekre irányuló vizsgálatok alapján a roma és az askenázi zsidó populáció mutatta a legnagyobb genetikai távolságot a többi magyar népcsoporttól (Guglielmino és Bérés, 1996; Czeizel, 2003). Munkám során ezt a kérdést DNS-markerek (STR és mtDNS) szintjén is elemezni szándékoztam. Másrészt, ha az egyes populációs adatbázisok között az STR-profil és mtDNS haplotípus gyakoriságok szintjén is szignifikáns különbségeket detektálunk, akkor ez a különbség lényeges hatással lehet az igazságügyi DNS-vizsgálatok statisztikai interpretációjára. A székely és csángó populációs minta felmérésünkbe való bevonása lehetőséget teremtett arra, hogy olyan magyar etnikumok is bekerüljenek a magyar népesség populációgenetikai felmérésébe, amelyek korábban nem vagy csak korlátozott mértékben kerültek feldolgozásra (Czeizel, 2003).

1.3.2. Az analízisbe bevont STR markerek kiválasztása

A mikroszatellita analízist olyan STR polimorfizmusok vizsgálatával szándékoztam lefolytatni, amelyek igazságügyi célú alkalmazhatóságát korábban letesztelték és igazságügyi geneti-

kai felhasználásra validálták (Hammond és mtsai, 1994; Lygo és mtsai, 1994; Micka és mtsai, 1996, 1999; Cotton és mtsai, 2000). Mindezek alapján olyan multiplex vizsgálati rendszer áll a populációgenetikai analízis rendelkezésére, melynek eredményei kompatibilisek és összevethetőek más, nemzetközi munkacsoportok eredményeivel, másrészt sikeresen applikálható a magyar igazságügyi genetikai szakértői gyakorlatba. Mindennek nemcsak tudományos, hanem szakmai-szakértői szempontból is rendkívül fontos vetülete van, mivel az igazságügyi DNS laboratóriumok közötti kompatibilitás a szakértői munka hitelességének egyik záloga.

A multiplex mikroszatellita analízis során 15 tetra-, és 2 pentanukleotid ismétlődést tartalmazó, autoszómás STR marker elemzését kívántam megvalósítani. A kiválasztás szempontjai között szerepelt: a marker nemzetközi hitelesítettsége, a lókuszt polimorfizmusának foka, a genetikai analízis kivitelezhetősége. A kiválasztás alapját képezte az is, hogy olyan lókusztokat alkalmazzunk, amelyek kromoszómás lokalizációjuk alapján gyakorlatilag nem kapcsolódtak. Az analízisbe bevont STR lókusztok a kromoszómák sorrendjében a következők voltak (zárójelben kromoszómás lokalizációjuk):

TPOX (2p25.3), *D2S1338* (2q35), *D3S1358* (3p21.31), *FGA* (4q28), *D5S818* (5q23.2), *CSFIPO* (5q33.1), *D7S820* (7q21.11), *D8S1179* (8q24.13), *TH01* (11p15.5), *VWA* (12p13.31), *D13S317* (13q31.1), *Penta E* (15q26.2), *D16S539* (16q24.1), *D18S51* (18q21.33), *D19S433* (19q12), *D21S11* (21q21.1), *Penta D* (21q22.3).

A Penta E és Penta D lókusztok pentamer mikroszatelliták, az összes többi tetramer. (Di- és trinukleotid repeatek analízisére ez a tanulmány nem vállalkozott, mivel az ilyen típusú STR markerek az analízis technológiai korlátai miatt igazságügyi genetikai célra nem validáltak, nem használatosak, ezért a jelen tanulmány számára nem relevánsak.) A vizsgálni kívánt markerek viszonylag nagy száma és a terjedelmi korlátok miatt az analizált lókusztok részletes jellemzésére itt nem kerül sor, azok részletes leírását az STRBase honlap (<http://www.cstl.nist.gov/biotech/strbase>) tartalmazza (Ruitberg és mtsai, 2001).

1.3.3. A felmérések során detektált mikrováriáns STR allélek jellemzése

Az igazságügyi genetikai vizsgálatok alapkövetelménye a genotípus pontos leírása (Bär és mtsai, 1997; NRC, 1997; Olaisen és mtsai, 1998). Az allélok pontos meghatározása az STR markerek esetében ún. referencia-allélok segítségével történik, vagyis az adott lókuszon a populációban tapasztalt tulajdonságok keverékét – alléllétra, allélkóktél – használjuk. Amennyiben egy ismeretlen allél nem feleltethető meg egyik referencia-allélnak sem, a tipizálandó allélt – lehetőleg szekvenálással – jellemezni kell (NRC, 1996; Bär és mtsai, 1997). A populá-

ciós felméréseink során alkalmazott nagy felbontású genotipizáló rendszernek köszönhetően több ritkán megfigyelhető, ún. mikrovariáns allélt figyeltünk meg, amelyek méretben nem egyeztek meg az alkalmazott referencia alléllétrák egyik fokával sem. Ezeket a ritkának számító és az észlelés időpontjában még le nem írt mikrovariáns allélok szekvencia-struktúráját DNS szekvenálással kívántam jellemezni.

1.3.4. Az analizált STR markerek géndiverzitásának és a polimorfizmus fokának megállapítása, allél-gyakorisági táblázatok generálása

Az STR profilegyezési valószínűség kalkulációjához számos paramétert és körülményt meg kell ismerni, amelyek közül a legfontosabbak: a vizsgált STR markerek polimorfizmusának foka, az allélok lókuszokon belüli és lókuszok közötti lehetséges asszociációja, ill. a teljes genetikai variancia megoszlása személyeken belül, a személyek között a populáción belül és a populációk között (Evett és Weir, 1998). A populációs minták analízise során nyert adatok ismeretében megállapítható, hogy a vizsgált markerek egyes alléljai milyen gyakorisággal fordulnak elő a populációkban, melyekből több, a polimorfizmus fokát jellemző paraméter – pl. heterozigotitás, megkülönböztető és kizáró erő – is meghatározható. A populációgenetikai elemzések első lépéseként az volt célszerű megállapítani, hogy a felmérésbe bevont STR markerek polimorfizmusának foka megfelelő mértékű-e és a kalkulált géndiverzitási értékek kellően magasak-e a magyarországi populációs minták esetében a hatékony személyazonosítás kivitelezése érdekében.

1.3.5. A vizsgált populációk Hardy-Weinberg egyensúlyi tesztelése az analizált STR markereken

A generált allél- és STR-profil gyakorisági adatbázisokat tesztelni kell arra nézve, hogy felhasználhatók-e a DNS-vizsgálat bizonyító erejének korrekt megbecsüléséhez. Az igazságügyi DNS-vizsgálatok statisztikai interpretációjának fő eleme a DNS-profil egyezési valószínűség meghatározása. Az egyezési valószínűség kalkulációjának alapmodelljét az a feltételezés szolgáltatja, hogy a DNS-profilokat alkotó allélok között nincs számottevő asszociáció (Evett és Weir, 1998). Hardy-Weinberg teszteléssel arra a kérdésre tudunk választ adni, hogy van-e jelentős mértékű allélikus asszociáció a vizsgált STR-ek alléljai között a lókuszokon belül az egyes populációkban. Az analízisbe bevont lókuszok közötti linkage egyensúlyi tesztelés hat-

ból négy populáció (BuCa, BaRo, DeRo, BuAs) vonatkozásában már megtörtént (Füredi, 2003), ezért a lókuszok közötti linkage equilibrium tesztelést nem terveztem elvégezni.

1.3.6. A mtDNS teljes kontroll régió szekvencia-analízis laboratóriumi hitelesítése referencia adatbázisok generálása céljából

Az igazságügyi célú mtDNS vizsgálat a célnak megfelelő populációs adatbázisokat igényel, amelyekkel megbecsülhető a kérdéses DNS-profilok (haplotípusok) gyakorisága az adott populációban. Ha egy eset kapcsán a bizonyíték és a gyanúsított mtDNS profilja megegyezik, a megfelelő referencia adatbázis alapján kalkulált haplotípus-gyakoriság egy konzervatív becslést adja meg annak a valószínűségnek, hogy a bizonyítékból kimutatott DNS-profil a gyanúsítottól, vagy a populáció egy véletlen személyétől származik. Az elmúlt évek populációgenetikai kutatásai jól meghatározott különbségeket tártak fel a különféle eredetű populációk mtDNS haplotípus megoszlásában (Stoneking és mtsai, 1991; Budowle és mtsai, 1999; Forster és mtsai, 2002), és ez az oka annak, hogy adott bizonyíték esetén a konkrét ügy egyéb körülményei is befolyásolják az alkalmazandó referencia adatbázist. Az eszményi referencia adatbázis leképezi a lehetséges mintaadók körének haplotípus megoszlását. Ez egy város vagy egy ország esetében a különféle populációcsoportok mitokondriális haplotípusait jelenti olyan arányban, mint amilyen az adott populációban található. Ennek az elvnek csak nagyon kevés adatbázis tud megfelelni és általában kivitelezhetetlen is a megvalósítás. Ezért inkább több, egymástól különböző releváns forrást reprezentáló adatbázisok alkalmazása az általános gyakorlat gyakorisági becslések kalkulálása során.

Munkám alapvető célja volt a vizsgálatba bevont magyar populációs mintákból a mitokondriális DNS teljes kontroll régióját felölelő haplotípusok generálása és hitelesítése, miáltal olyan komplex referencia adatbázis hozható létre, melynek segítségével lehetőség nyílik a mitokondriális DNS szekvenciák analízisén alapuló azonosítások egyezési valószínűségének populációstatisztikailag korrekt becslésére a magyarországi törvényszéki eljárások során. A célom az volt, hogy ezt oly módon hozzuk létre, hogy a létrejövő szekvencia adatbázisok mentesek legyenek a tipzálásból, adatbevitelből és transzportból, mintakeveredésből eredeztethető hibáktól, mely hamis és torz eredményekhez vezethet. Ennek érdekében a teljes kontroll régió PCR alapú szekvenálását végeztük el a populációs mintákon egy teljesen automatizált laboratóriumi eljárással (Brandstätter és mtsai, 2004b és 2007a), mely magába foglalja a robotizált PCR amplifikációt és szekvenálást, az elektronikus adattranszportot, valamint a szekvencia eredmények kettős kiértékelését ellenőrző supervisor személyt. A populációs min-

ták automata szekvencia-analízisét az amerikai légierő DNS azonosító laboratóriumában (AFDIL – Armed Forces DNA Identification Laboratory) és az Innsbrucki Egyetem Igazságügyi Orvostani Intézetének DNS laboratóriumában (GMI –Gerichtsmedizin Innsbruck) végeztük el. A populációs felméréssel párhuzamosan az automatizált szekvencia-analízis technológiája bevezetésre és hitelesítésre került a Bűnügyi Szakértői és Kutatóintézet DNS laboratóriumában.

1.3.7. A mitokondriális DNS haplotípusok statisztikai analízise

Annak érdekében, hogy a felmérésbe bevont populációk haplotípus adatai egymással összehasonlíthatók legyenek és azok populációkon belüli illetve populációk közötti eltérései analizálhatóvá váljanak statisztikai paraméterek kalkulálását kell elvégezni.

Célom volt mindenegybes populációra lebontva a véletlen egyezési valószínűségi (Random Match Probability – RMP) értékek és a szekvencia-párok közötti eltérések megoszlásának (mismatch distribution) megállapítása a vizsgált populációkon belül és a populációk között, egyéb európai országok (pl. Ausztria, Németország) adatait is bevonva az analízisbe. A budapesti referencia populáció adatait összevettem a SWGDAM (Scientific Working Group on DNA Analysis Methods) „kaukázusi” (Egyesült Államokbeli és Európai) személyek szekvenciáit tartalmazó adatbázisával (Allard és mtsai, 2002).

1.3.8. A mtDNS szekvenciák haplocsoportosítása

A mitokondriális DNS szekvenciák mintázatuk alapján haplocsoportba sorolhatók. A haplocsoportok azonos eredetű haplotípusok osztályait jelenti és a minták filogenetikai analíziséhez felhasználhatók (Wallace és mtsai, 1999; Finnilä és mtsai, 2001; Ruiz-Pesini és mtsai, 2004). Igazságügyi célú populációs adatbázisok haplocsoport szempontú analízise a szekvencia-adatok minőségellenőrzésének céljából is használatosak (Budowle és mtsai, 2003 és 2004).

Jelen tanulmányban a populációs mintákat meghatározott kontroll régió szekvenciájuk alapján kívántam haplocsoportosítani ún. haplocsoport-specifikus vagy haplocsoport-kapcsolt szekvencia polimorfizmusok alapján (Allard és mtsai, 2002).

1.3.9. A felmérésbe bevont magyar populációk genetikai strukturáltságának megállapítása

Az igazságügyi genetikai gyakorlatban is széleskörűen alkalmazott populációstatisztikai tesztek [G-statisztika, F-statisztika, molekuláris varianciaanalízis (AMOVA), hálózatanalízis] segítségével jellemezni lehet a felmért magyarországi populációk genetikai strukturáltságát. Tisztázásra várt, hogy a referenciaként deklarált budapesti populáció és a többi magyar szubpopuláció között mekkora a genetikai variancia az analizált STR lókuszokon és a mtDNS kontroll régió adatokban. Céлом volt annak a megállapítása is, hogy a magyar szubpopulációk és az analízisbe bevont más, európai populációk között mekkora a genetikai variabilitás.

Korábbi irodalmi adatok alapján (Czeizel és mtsai, 1991, Guglielmino és Béres, 1996) feltételezhető, hogy a magyar népesség szubstrukturáltsági foka igazságügyi genetikai aspektusból is relevanciával bír. Ezért figyelembe kell venni a vizsgált populációk autoszómás STR lókuszokon és mitokondriális DNS szekvenciákon mért genetikai varianciájának az igazságügyi statisztikai interpretációra gyakorolt hatását.

2. ANYAGOK ÉS MÓDSZEREK

Kutatómunkám során autoszómás mikroszatellita (STR) lókuszokon és a mitokondriális genom kontroll régióján végeztem felmérést hat magyar populációban, melyből négy populációs minta Magyarországon, kettő Romániában, Erdélyben került begyűjtésre. A felmérés lépései röviden a következők voltak:

- A DNS kinyerése a begyűjtött mintákból, a DNS koncentráció meghatározása
- Az STR lokuszok PCR-amplifikációja és ellenőrzése gélelektroforézissel
- STR genotipizálás DNS fragmens-analízissel
- A mikrovariáns STR allélek szekvencia-analízise
- A mtDNS kontroll régió PCR-amplifikációja és szekvencia-analízise
- A mtDNS haplotípusok haplocsoportosítása
- Populációstatisztikai és genetikai analízisek

2.1. A DNS kinyerése a populációs mintákból, koncentráció-meghatározás

A friss, nem hemolizált folyékony vérminták esetében a DNS feltárása a limfociták sejtmagjainak szukróz ozmózisos sokkal történő feldúsítását (Kanter és mtsai, 1986) követő proteolitikus sejtlízis (proteináz K), szerves extrakció (fenol-kloroform-izoamilalkohol) és ultraszűrés-koncentráálás (Microcon-100, Amicon, Millipore) módszerrel történt. A hemolizált folyékony vérminták illetve vérfoltok és szájnyalvákahártya-törlet minták esetében a DNS feltárását detergens tartalmú extrakciós pufferben történő proteolitikus sejtlízissel (proteináz K), szerves extrakcióval (fenol-kloroform-izoamilalkohol) és ultraszűrés-koncentráálással (Microcon-100, Amicon, Millipore) végeztük (Comey és mtsai, 1994).

A populációs felmérés során alkalmazott több lókuszos multiplex-PCR rendszerek meglehetősen érzékenyek a reakcióba bevitt templát DNS mennyiségére. Túl kevés bevitt DNS (<0,1 ng) allél- és/vagy lókuszkiesést eredményezhet, nagy mennyiségű templát aspecifikus és „off-scale” jelekkel zavarja a genotipizálást. Ezért a multiplex-STR PCR reakció körülményeit úgy optimalizálják, hogy viszonylag kis mennyiségű (0,5-2 ng) DNS-templát elegendő legyen a sikeres genotipizáláshoz. Az optimális DNS-mennyiség alkalmazása érdekében célszerű a DNS-extraktum koncentrációját meghatározni. A minták DNS-tartalmának szemikvantitatív meghatározására agaróz gélelektroforézist és etídium-bromid/SYBR Gold fluorofór festést alkalmaztunk standard protokoll szerint.

2.2. Az STR lokuszok PCR-amplifikációja és ellenőrzése gélelektroforézissel

A CODIS STR lokuszok (*CSF1PO*, *D3S1358*, *D5S818*, *D7S820*, *D8S1179*, *D13S317*, *D16S539*, *D18S51*, *D21S11*, *FGA*, *TH01*, *TPOX*, *VWA*) és az *Amelogenin* (genetikai nem meghatározását szolgáló) lokusz multiplex PCR-sokszorozását az AmpF/STR ProfilerPlus/Cofiler kitekkel végeztük el standard protokoll szerint (Applied Biosystems, 1997) GeneAmp PCR System 9600 (Applied Biosystems) készülékeken. Az Identifiler kit alkalmazásával (Applied Biosystems, 2001) a CODIS STR-eken kívül további kettő tetranukleotid mikroszatellita (*D2S1338* és *D19S433*), a PowerPlex 16[®] System felhasználásával pedig kettő pentanukleotid STR marker (*Penta E* és *Penta D*) került felsokszorozásra standard protokoll szerint (Promega Corporation, 2002). Az alkalmazott primerek szekvenciáit egyik marker esetében sem publikálták a gyártók. A PCR minták tipizálása előtt az amplifikálás hatékonyságának ellenőrzésére horizontális, natív (nem denaturáló), nagy felbontású, diszkontinuus poliakrilamid gélelektroforézist (nhd-PAGE) és ezüstfestéses vizualizációt alkalmaztunk (Budowle és mtsai, 1991).

2.3. STR genotipizálás fragmensanalízissel

Az STR lokuszok PCR termékeinek fragmens-analízise alapvetően három részből állt:

1. A PCR-rel felszaporított DNS-szakaszok méret szerinti elválasztása denaturáló kapilláris gélelektroforézissel ABIPrism[®] 310 és 3100 genetikai analizáló készülékeken (Applied Biosystems) az AmpF/STR ProfilerPlus/Cofiler/Identifiler (Applied Biosystems, 1997 és 2001) és PowerPlex 16[®] (Promega Corporation, 2002) kitek felhasználói kézikönyve szerint történt GeneScan[™]-500 ROX, GeneScan[™]-400 HD ROX, GeneScan[™]-500 LIZ (Applied Biosystems) és ILS-600 (Promega) belső méretstandardokat felhasználva.

2. Az elválasztott PCR-termékek detektálása lézer gerjesztett fluoreszcenciával fluoreszcensen jelölt primerek (nem publikált) által történt. Alkalmazott fluorofórok: 5-FAM, JOE, NED, ROX (ProfilerPlus/Cofiler), 6-FAM, VIC, NED, PET, LIZ (Identifiler), Fluorescein, TMR, JOE, CXR (PowerPlex[®] 16). A hossz-polimorf jellegű STR markerek jele a szeparálás során az allél típusára vonatkozó információnak felel meg.

3. A detektált adatok kiértékelése, azaz az egyes mintákhoz tartozó allélok típusának meghatározása (genotipizálás) félautomata módon GeneScan Analysis v3.1. és Genotyper v2.6.2 szoftverek (Applied Biosystems) alkalmazásával történt. A potenciális allélként

amplifikálódott fragmensek csoportosítására méretük alapján került sor. Az amplifikált DNS fragmensek méretének meghatározása („allele sizing”) GeneScan szoftverrel a mintákhoz kevert belső méretstandardok alapján történt. Ugyancsak a belső méretstandard szolgált az egyes injektálások között esetlegesen fellépő DNS-mozgékonyosság különbség („mobility shift”) kompenzálására is. A Genotyper szoftver a referencia allél-létra és a PCR-fragmens méretének összevetésével végzi a PCR-termékek automata genotipizálását („allele calling”). Az allélek meghatározásához referencia-allélokból összeállított „koktél”, ún. allél-létrát használtunk, melyek a multiplex-PCR rendszerekkel együtt megvásárolhatók.

2.4. A mikrovariáns STR allélek szekvencia-analízise

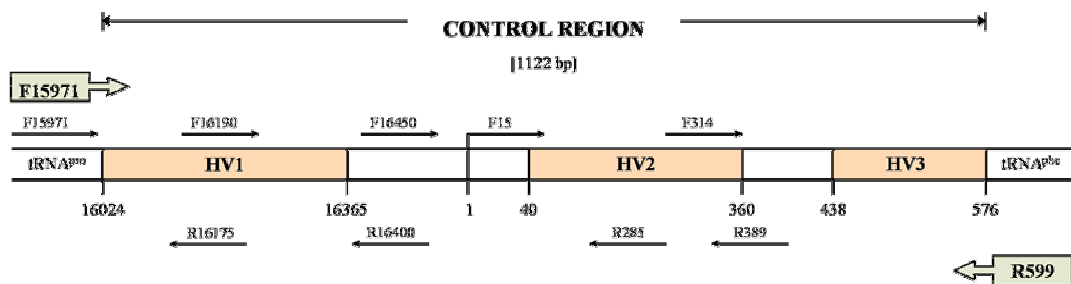
A populációs felmérés során megfigyelt több, ún. mikrovariáns allél (melyek nem vágtak egybe az alkalmazott referencia alléllétrák egyik fokával sem) pontos DNS szekvenciáját ciklus-szekvenálással határoztam meg. A mikrovariáns allélek PCR-amplifikációja publikált primerekkel történt (GenBank: <http://www.ncbi.nlm.nih.gov/Genbank/index.html>; STRBase: <http://www.cstl.nist.gov/biotech/strbase>), a sokszorozott fragmensek tisztítását Qiaquick PCR Purification kit (Qiagen, Hilden, Németország) vagy Centricon-100 oszlopok (Amicon, Millipore) alkalmazásával végeztem. A tisztított PCR termékek pGEM-T (Promega) vagy TOPO-TA (Invitrogen) vektorban kerültek klónozásra. A szelektált klónok szekvenálása PCR vagy M13 primerekkel a BigDye Terminator Cycle Sequencing Ready Reaction Kitek (Applied Biosystems, 2002 és 2003) felhasználásával standard protokoll szerint ABI Prism 310/3100 genetikai analizátor készüléken történt. A szekvenálást minden egyes allél esetében forward és reverse primerrel is elvégeztem. A kapott DNS szekvenciák kiértékelése, szerkesztése és illesztése a Sequencing Analysis Software 3.4.1 (Applied Biosystems) és Sequencher™ 4.1.2 (Gene Codes Corp.) komputerprogramok felhasználásával történt.

2.5. A mtDNS kontroll régió PCR amplifikációja és szekvencia-analízise

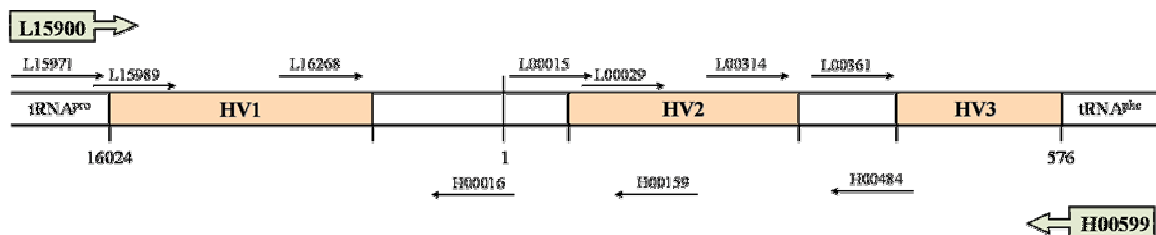
A populációs minták PCR alapú mtDNS teljes kontroll régió szekvenálását végeztük el egy teljesen automatizált, helyenként módosított publikált (Brandstätter és mtsai, 2004b és 2007a) laboratóriumi eljárással. A processzus magába foglalta a teljesen robotizált PCR amplifikációt és szekvenálást, valamint az elektronikus adattranszportot mindenféle manuális átírás nélkül. A szekvencia-adatok generálása során redundáns megközelítést alkalmaztunk annak érdekében, hogy minimalizáljuk a lehetséges hibákat a szekvencia kiértékelés során. A kontroll

régió szekvenálás során alkalmazott stratégia az alkalmazott PCR és szekvenáló primerekkel az 1.A-B. ábrán kerül bemutatásra.

1.A. ábra Az AFDIL laboratóriumban alkalmazott szekvenálási stratégia a BuCa és BaRo populációk felmérése során



1.B. ábra A GMI laboratóriumban alkalmazott szekvenálási stratégia a Székely és Csángó populációk felmérése során



Magyarázat: F/L = a könnyű szál primere (forward); R/H = a nehéz szál primere (reverse)
 PCR primerek dobozban, szekvenáló primerek nyíllal feltüntetve
 HV1, HV2, HV3: Hipervariábilis szegmens 1, 2, 3

A kontroll régió szekvenáláshoz felhasznált PCR és szekvenáló primerek, a rekációkhoz felhasznált vegyszerek és vizsgálati paraméterek publikálásra kerültek (Irwin és mtsai, 2006; Brandstätter és mtsai, 2004a és 2007a) ezért ezek leírásától itt eltekintek. A hatékony szekvenálás elősegítése érdekében a populációs minták feldolgozása 96 mintát tartalmazó tálcasorozatokkal történt, tálcánként négy negatív kontrollt alkalmazva mindkét, a felmérésben résztvevő laboratóriumban (AFDIL, GMI). A tálcasorozatok alkalmazása lehetővé tette multi-chanel pipetták és robotok alkalmazását a PCR és a szekvenálási reakciók előkészítésé-

ben és kivitelezésében (Corbett CAS-1200, Tecan Genesis 200, epMotion workstation) valamint a minták tisztítása során.

Az eredményként kapott nyers szekvencia adatok (raw data) kettős kiértékelése és illesztése a Sequencing Analysis Software v3.4.1 (Applied Biosystems), Sequencher™ 4.1.2 (GeneCodes Corp.) és SeqScape 2.0 komputerprogramok alkalmazásával történt két párhuzamosan dolgozó szakértő által. A konszenzus szekvenciák illesztése és összehasonlítása a javított Cambridge referencia szekvenciához (rCRS) történt (Anderson és mtsai, 1981; Andrews és mtsai, 1999) nemzetközi ajánlásoknak megfelelően (Carracedo és mtsai, 2000; Tully és mtsai, 2001). Az analízis során figyelembe vett szekvenciák a 16024-16569 és 1-576 nukleotid pozíciók közötti szakaszokat ölelték fel. Megfigyelt heteroplazmiás pozíciókat a megfelelő IUPAC/IUB kódoknak megfelelően jeleztük a hossz-heteroplazmiákat következetesen a pre-domináns nukleotiddal jelölve (Wilson és mtsai, 2002a és 2002b, Bandelt és Parson, 2007). A két szakértő általi független kiértékelés eredményét minden szekvencia esetében mutációs riportként exportáltuk egy LIMS-be (Laboratory Information Management System) integrált mtDNS adatok feldolgozására és statisztikai kiértékelésére szolgáló programba (Steinlechner és Parson, 2001), mely program az importált haplotípusok elektronikus összehasonlítását is elvégzi, melynek eredményét egy harmadik supervisor személy ellenőrzi. A kiértékelési sorozat végén a supervisor által jóváhagyott szekvenciák kerülnek át a mtDNS haplotípus adatbázisba, ahonnan a filogenetikai és populáció-statisztikai analízis céljára előkereshetők.

2.6. A mtDNS haplotípusok haplocsoportosítása

A populációs minták szekvencia-analízise során kapott mtDNS haplotípusok haplocsoportba való besorolása a kontroll régió szekvencia-motívumai alapján haplocsoport-specifikus vagy haplocsoporttal kapcsolt szekvencia polimorfizmusok szerint történt (Achilli és mtsai, 2004, 2005; Allard és mtsai, 2002; Derenko és mtsai, 2003; Finnilä és mtsai, 2001; Gresham és mtsai, 2001; Helgason és mtsai, 2000; Kalaydjieva és mtsai, 2001; Loogväli és mtsai, 2004; Malyarchuk és mtsai, 2002, 2003; Palanichamy és mtsai, 2004; Reidla és mtsai, 2003; Torroni és mtsai, 2006). Azon minták esetében, amelyeknél a kontroll régió szekvencia alapján nem volt egyértelmű a haplocsoportba való besorolás, az ún. „West-Eurasia-Plex” (Brandstätter és mtsai, 2003b) felhasználásával, vagy a kérdéses nukleotidpozíciók relatív evolúciós rátájának becslésével (Forster és mtsai, 2002) végeztük el a tipizálást.

2.7. Populációstatisztikai és genetikai analízisek

A populációkban megfigyelt polimorfizmusok populációgenetikai szempontból számos értékkel jellemezhetők, a jelen tanulmányban alkalmazott eljárások csak néhány, igazságügyi relevanciával is rendelkező érték vizsgálatára korlátozódtak. Az igazságügyi genetika a polimorfizmusok populációgenetikai jellemzéséhez számos egyszerű mérőszámot használ.

2.7.1. Populáción belüli elemzések

Annak jellemzésére, hogy a vizsgálatba bevont populációs minták milyen mértékben reprezentálják az adott populációt, valamint annak érdekében, hogy a törvényszéki gyakorlat igényelte válaszok a származás-leszármazás vonatkozásában hitelesen legyenek közölhetők, a kérdéses mintacsoportok tényleges szaporodásközösségként történő kezelése és elemzése szükséges. A nyert geno- és haplotípus adatokból a következő alapértékek kalkulálhatók.

2.7.1.1. STR markerekre kalkulált populációgenetikai alapértékek

A felmért populációkban megfigyelt allélgyakorisági adatok ismeretében a következő paramétereket számítottam ki az analizált STR lókuszok mindegyikén.

H(obs) – Observed Heterozigosity – a populációs mintában az adott lókuszon ténylegesen megfigyelt heterozigóták aránya.

H(exp) – Expected Heterozigosity – a populációs mintában a megfigyelt allélszámból kalkulált várt heterozigóta arány.

Autoszómás lókuszokon a *Nei-féle géndiverzitás* (h) tulajdonképp megfeleltethető a heterozigóták elfogulatlanul (unbiased) becsült arányának (H_{exp}):

$$h = n(1 - \sum x_i^2) / (n - 1) \quad \text{több lókuszra kombinálva} \quad h_{komb} = 1 - \prod (1 - h_i)$$

ahol n a kromoszómák száma, x_i az i -edik allél megfigyelt gyakorisága (Nei, 1973).

A Nei-féle géndiverzitás standard hibája $SE = \sqrt{2h(1-h)/n}$

PD – Power of Discrimination – az igazságügyi genetikában annak kifejezése, hogy két személyt a népességből véletlenszerűen kiválasztva átlagosan mekkora a valószínűsége annak, hogy a két személy a vizsgált lókuszon eltérő genotípust hordoz (*megkülönböztető erő*).

PM – Probability of Match – az 1-ből kivont PD értékkel egyezik meg (*egyezési valószínűség*).

$$PD = 1 - \sum x_i^2 \quad \text{több lókuszra kombinálva} \quad PD_{\text{komb}} = 1 - \prod (1 - PD_i)$$

$$PM = 1 - PD \quad \text{több lókuszra kombinálva} \quad pM_{\text{komb}} = \prod pM_i$$

ahol x_i az i -edik genotípus megfigyelt gyakorisága (Jones, 1972).

PE – Power of Exclusion – (*MEC – Mean Exclusion Chance*) az igazságügyi genetikában annak kifejezése, hogy a populációból egy egyedet véletlenszerűen kiválasztva mekkora az átlagos valószínűsége annak, hogy az egy szülő-utód relációból mint másik lehetséges szülő kizárható (*szülősségi kizáró erő*).

$$PE = 2\left(\sum p_i^2\right)^2 - \sum p_i^4 \quad \text{több lokuszra kombinálva} \quad PE_{\text{komb}} = 1 - \prod (1 - PE_i)$$

ahol p_i az i -edik allél megfigyelt gyakorisága (Ohno és mtsai, 1982).

PIC – polimorfizmus információ tartalom – az adott lókusz polimorfizmus fokának jelzőszáma (Botstein és mtsai, 1980):

$$PIC = 1 - \sum p_i^2 - \sum_{i=1}^n \sum_{j=i+1}^n 2p_i^2 p_j^2$$

A statisztikai számítások során a *minimális allélgyakorisági érték* ($p_{\min}$) használata lehetővé teszi, hogy populációs adatbázisokban a ritkán előforduló alléltípusok hiányos megfigyeléseit ellensúlyozhassuk (Budowle és mtsai, 1996). A minimális allélgyakoriság azt a küszöbértéket mutatja meg, amely fölött a megfigyelt allélok gyakorisága bizonyos (általában 95%-os) konfidenciával megbízhatónak tekinthető. A meg nem figyelt allélok gyakoriságát úgy lehet tekinteni, hogy az a $p_{\min}$ -nel egyenlő. Az STR lókuszokra ajánlott minimális allélgyakoriság statisztikai és populációgenetikai elveken alapul, a lókusz heterozigótaságának növekedésével értéke nő, a populációs minta méretének növekedésével pedig csökken (Chakraborty, 1992):

$$p_{\min} = 1 - \left(1 - CF^{\frac{1}{c}}\right)^{\frac{1}{2N}}$$

ahol CF a választott konfidencia-szint, c a gyakran előforduló alléltípusok száma, N pedig a populációs minta száma.

2.7.1.2. A Hardy-Weinberg egyensúly (HWE) tesztelése, tesztsorozatok, Bonferroni-eljárás

Az allélek lókuszon belüli függetlenségének vizsgálata, a Hardy-Weinberg egyensúly megléte a Fisher-féle exact tesztnek Guo és Thompson (1992) által továbbfejlesztett permutációs változatával történt, amelyet Schneider, Roessli és Excoffier tett hatékonyabbá és implementált *Arlequin* nevű populációgenetikai szoftverbe. A HWE tesztelése az *Arlequin* ver.2001 szoftver (Schneider és mtsai, 2000) használatával történt és az exact teszt során alkalmazott paraméterek a következők voltak: Markov-lánc lépés 1.000.000, dememorizáció lépés 1000.

Amennyiben a HWE tesztelés során az egyensúlyt a lókuszok összességén vizsgáljuk, akkor a lókuszok ugyanazon a hipotézisen elvégzett tesztsorozatban vesznek részt. Ebben az esetben a tesztsorozat szignifikancia-szintje (az ún. *kísérleti jellegű hibaráta*) annak a valószínűsége, hogy a tesztek közül legalább egy a nullhipotézis elvetését okozza, ha a nullhipotézis igaz. A kísérleti jellegű hibaráta (α') a következőképpen fejezhető ki:

$$\alpha' = 1 - (1 - \alpha)^L \approx L\alpha,$$

ahol α az egyedi teszt szignifikancia-szintje és L a tesztek száma.

A fenti procedúrát *Bonferroni eljárásnak* hívjuk (Weir, 1996). Annak érdekében, hogy elkerüljük a hipotézis hamis elvetését, minden egyedi tesztelést szorosabbá kell tenni a fenti képlet átalakított formája alapján:

$$\alpha = 1 - (1 - \alpha')^{\frac{1}{L}} \approx \alpha' / L$$

Pl. a 17 STR marker tesztsorozata esetén az $\alpha' = 0,05$ kísérleti hibarátához $\alpha = 0,0029$ egyedi szignifikancia-szintet kell alkalmazni.

2.7.1.3. A mtDNS haplotípusok populációgenetikai paraméterei

A megfigyelt mtDNS haplotípus adatok populáció-statisztikai kiértékelését a felmért populációk mindegyikére elvégeztem. A kalkulációk során a teljes kontroll régió (16024-16569, 1-576) szekvencia adatait vettem figyelembe, a policitozin szálak hossz-polimorfizmusait a 16193, 309 és 573 nukleotid pozíciókban kihagyva a számításokból. Az *Arlequin* ver.2001 szoftver (Schneider és mtsai, 2000) felhasználásával az összes populációs minta DNS szek-

venciájának beimportálása után a következő diverzitási paraméterek voltak kalkulálhatók az egyes populációkban:

Polimorf nukleotidok – a kontroll régióban a CRS referencia szekvenciához viszonyítva megfigyelt összes polimorf nukleotid pozíció.

Tranzíció/Transzverzió – a megfigyelt összes tranzíciós és transzverziós szubsztitúció.

Megfigyelt/Egyedi haplotípusok – a populációban megfigyelt összes különböző haplotípus, ebből mennyit lehetett kizárólag egyszer megfigyelni.

Random Match Probability (RMP) – annak kifejezése, hogy két személyt a népességből véletlenszerűen kiválasztva átlagosan mi a valószínűsége annak, hogy a két személy azonos haplotípussal rendelkezik (*véletlen egyezési valószínűség*). A véletlen egyezési valószínűség a haplotípus gyakoriságok négyzetösszegeként kalkulálható:

$$p = \sum x_i^2 \quad \text{ahol } x_i \text{ az } i\text{-edik haplotípus megfigyelt gyakorisága (Stoneking és mtsai, 1991).}$$

Genetikai diverzitás (hívják géndiverzitásnak is) – annak az átlagos esélye, hogy két haplotípust véletlenszerűen kiválasztva a mintából azok különbözőek. Kiszámítása ekvivalens a diploid adatok várt heterozigotizálásával, és annak haploid megfelelője:

$$H = n(1 - \sum x_i^2)/(n-1) \quad \text{ahol } x_i \text{ az } i\text{-edik haplotípus megfigyelt gyakorisága, } n \text{ a mintaszám.}$$

2.7.1.4. A mtDNS szekvenciapárok közötti eltérésének kalkulációja

A egyes haplotípuspárok között megfigyelt eltérések száma a populációkon belül és a populációk között is kalkulálható, melyhez a LISA (Future Technologies, Fairfax, VA, USA) és a PAUP* v4.0 (Swofford, 2002) számítógépes programokat vettük igénybe. A kapott adatokból a populáció diverzitására jellemző ún. „különbözőségi eloszlásgörbe” (*mismatch distribution*) generálható, melyben az i -edik oszlop megmutatja azon szekvencia-párok számát, amelyek i számú nukleotid pozícióban különböznek egymástól. Humán populációkban a recens demográfiai expanzió miatt általában unimodális eloszlásgörbe figyelhető meg, demográfiai egyensúlyban lévő populációkra általában a multimodális eloszlás jellemző (Rogers és Harpending, 1992). Ennek átlag értéke (*mean pairwise difference*) a populációra jellemző jelzőszám.

2.7.2. *A populációk összehasonlító statisztikai és genetikai elemzése*

Annak megállapítására, hogy a referencia adatbázisként rendelkezésre álló populációs minták mindegyike illetve valamelyike általános módon felhasználható-e származási valószínűség megállapításához a populációk összehasonlító elemzése szükséges.

2.7.2.1. *A populáció-párok homogenitás-vizsgálata G-statisztikával*

A populáció-párok homogenitásának statisztikai szignifikanciáját R×C szoftverrel (Carmody, Carleton University, Ottawa, Canada) teszteltük. Az R×C program két utas kontingencia táblázatok statisztikai szignifikanciáját teszteli. A szoftver a megfigyeléseket tartalmazó (eredeti) kontingencia táblázatból véletlenszerűen – az alapbeállítás szerint – 1000 új táblázatot replikál úgy, hogy az új táblázatok marginális összegei megegyezzenek az eredeti táblázat hasonló összegeivel. A program minden egyes táblázatra kiszámítja a G-statisztikát, majd végeredményként kiírja a random táblázatok azon hányadát (P), ahol a G-statisztika értéke meghaladta az eredeti táblázat hasonló értékét. Ha a P érték alacsony, akkor számolni kell azzal, hogy az eredeti táblázat sorai és oszlopai egymástól nem függetlenek; az adatok nem mutatnak elegendő homogenitást ahhoz, hogy összevonhassuk őket. A populációgenetika nyelvén ez azt jelenti, hogy adott szignifikancia-szint mellett (általában ez 5%) a populációk között szignifikáns különbséget találtunk. A szoftver egyik legfőbb erénye, hogy a kis számú megfigyeléseket tartalmazó vagy éppen üres cellákat (nincs megfigyelés) is képes tolerálni.

2.7.2.2. *Populációgenetika, F-statisztika, AMOVA*

A populációk genetikai strukturáltságának vizsgálatához, az STR és mtDNS genetikai profilok populációk közötti variabilitásának kalkulálásához hagyományos Wright-féle F-statisztika analízist (Wright, 1951; Weir és Cockerham, 1984) és molekuláris varianciaanalízist (AMOVA) használtunk (Excoffier és mtai, 1992; Michalakis és Excoffier, 1996). A populációs mintákra vonatkozó Wright-féle F-statisztika értékeket [F_{IS} , F_{ST} (θ)] a *Genetic Data Analysis (GDA v.1.1)* program segítségével (Lewis és Zaykin, 2001) az egyes lókuszokra nézve számítottuk ki. A populáció-párookra vonatkozó F_{ST} értékeket, az AMOVA analízist és azok statisztikai szignifikanciáját az *Arlequin ver.2001* (Schneider és mtai, 2000) szoftverrel határoztuk meg.

Az F -statisztikával lókuszonként és populációpáronként kapott F_{ST} értékek kifejezik, hogy a populációkban megfigyelt genetikai profilok eltérése alapján a totális genetikai variancia mekkora része tulajdonítható a populációk közötti varianciának. Az AMOVA elemzése során kapott Φ_{ST} (RST) érték, a Wright-féle F_{ST} érték molekuláris analógjaként az eltérő genotípusokat elválasztó molekuláris távolságok becslésével teszteli a populációkban megfigyelt genotípusok totális molekuláris varianciáját arra nézve, hogy annak hanyad része tulajdonítható a populációk közötti különbségnek. A Φ_{ST} értékekkel jellemzett populációk közötti variabilitás statisztikai szignifikanciáját szimulációval teszteltük. A szimulációt végző program az adott populáció-párban megfigyelt genotípusokat véletlenszerűen újra szétosztja a két populáció között. A szimulációval kapott P érték annak a gyakoriságát adja meg, hogy a replikációk hányad részében haladta meg a szimulációs Φ_{ST} érték az aktuális Φ_{ST} értéket. Ebben az esetben a $\{\Phi_{ST} = 0\}$ nullhipotézist teszteltük a $\{\Phi_{ST} > 0\}$ hipotézis ellenében. Ha $P < 0,05$, akkor $\alpha = 0,05$ szignifikancia-szint mellett a nullhipotézist elvethetjük, azaz molekuláris variancia-analízissel szignifikáns különbséget találtunk a két populáció között. Az előbbieket alapján minél nagyobb a Φ_{ST} érték, annál nagyobb a populációk közötti molekuláris távolság.

2.7.2.3. Hálózat analízis és migrációs ráta becslés

A median joining network analízist (Bandelt és mtsai, 1999) a Network 4.2 számítógépes program (Röhl A; © Fluxus Technology; Fluxus Engineering Web site: <http://www.fluxus-engineering.com>) felhasználásával végeztük el az összesen 360 erdélyi (Székely és Csángó) populációs mintán, az epsilon értéket 0-ra, a haplotípus gyakoriságot > 1 kritériumra beállítva. Az analízist megelőzően a szekvencia adatokból a policitozin szálak polimorfizmusait és a gyorsan mutálódó bázispozíciókat kiszűrtük.

A gyimesi csángó és a csíkszeredai székely populáció közötti migrációs rátára a MIGRATE (<http://popgen.csit.fsu.edu/Migrate-n.html>) program felhasználásával következtettünk a közös leszármazás teóriáját alkalmazva (Beerli és Felsenstein, 1999, 2001). A MIGRATE szoftver az effektív populáció méretet és az n számú populáció közötti múltbeli migrációs eseményeket becsüli meg, aszimmetrikus migrációs rátákkal és eltérő szubpopuláció méretekkel rendelkező migrációs mátrix modellt feltételezve. A programozás során a populációgenetikai paraméterek becslésére a maximum likelihood eljárást alkalmaztuk.

3. EREDMÉNYEK ÉS ÉRTÉKELÉS

A jelen tanulmányban tárgyalt magyar populációs minták mikroszatellita és mtDNS alapú genetikai feltérképezését több, egymásra épülő szakaszban hajtottuk végre (1. táblázat). Mindegyik vizsgálati szakasz egyúttal az adott markerek elemzésének a magyarországi laboratóriumi vizsgálatokba történő bevezetését, hitelesítését és esetleges továbbfejlesztését jelentette. Ezzel párhuzamosan a rendelkezésre álló populációs mintákon végrehajtottuk a statisztikai és genetikai analízist. Az eredményeink publikálására az egyes vizsgálati egységek lezárása után került sor, a tudományos közleményeket a Függelék 7.3. fejezete tartalmazza.

1. táblázat A populációs felmérés szakaszai

Időszak	ProfilerPlus	Identifiler	PowerPlex16	mtDNA	Referencia (Függelék)
1999-2000	BuCa, BaRo				7.3.1.; 7.3.2.
2001-2002			BuCa, BaRo, DeRo, BuAs	BuCa, BaRo	7.3.4; 7.3.6.
2003-2004		BuCa, Székely, Csángó	Székely, Csángó	Székely, Csángó	7.3.3.; 7.3.5.; 7.3.7; 7.3.8.
2005-2006				BuAs	Brandstätter et al., 2007c
ProfilerPlus:	D3S1358-D5S818-D7S820-D8S1179-D13S317-D18S51-D21S11-FGA-VWA-Amel multiplex				
Identifiler:	CSF1PO-D2S1338-D3S1358-D5S818-D7S820-D8S1179-D13S317-D16S539-D18S51-D19S433-D21S11-FGA-TH01-TPOX-VWA-Amel multiplex				
Powerplex 16:	CSF1PO-D3S1358-D5S818-D7S820-D8S1179-D13S317-D16S539-D18S51-D21S11-FGA-TH01-TPOX-VWA-PentaE-PentaD-Amel multiplex				
mtDNA:	mitokondriális DNS teljes kontroll régió (BuCa, BaRo, DeRo, BuAs, Székely, Csángó, populációs minták leírását lásd az 1.3.1. fejezetben)				

3.1. Az analízisbe bevont STR markerek genotipizálása

A genetikai azonosításokhoz hasonlóan az igazságügyi genetikai munka alapfeltétele a korrekt genotipizálás, ezért annak jelentőségét nem lehet eléggé hangsúlyozni. Az STR lókuszok igazságügyi célú genotipizálása során a tipizálandó minta alléljeit referencia allélekkel hasonlítjuk össze. A referencia alléleket allél-létrába gyűjtve a tipizálandó mintákkal szimultán alkalmazzuk, mint ahogy a Függelék 7.1.1. ábráján a felmérés során alkalmazott *ProfilerPlus* multiplex rendszer egy véletlen mintán való tipizálásával ezt bemutatom. Az ábrán az allélok

fragmensmérete balról jobbra növekszik, az alkalmazott különböző fluorofórokat eltérő szín jelöli. (Az *Identifiler* és *PowerPlex 16* rendszerek allél-létráinak bemutatásától helyhiány miatt itt eltekintek.)

Az alkalmazott automata genotipizálási rendszerek allélméret-meghatározásának pontossága (allele sizing accuracy, precision) az ABI Genetikai Analizátor készülékeken már máshol leírásra került (Füredi, 2003). Ezt röviden összefoglalva itt annyit említek, hogy az ABI Prism kapilláris elektroforetikus rendszerek allélméret-meghatározás pontossága mindegyik analizált multiplex STR rendszer esetében kielégítő eredményre vezetett. A kalkulált precizitási értékek gyakorlatilag megegyeztek a párhuzamosan publikált eredményekkel (Lazaruk és mtsai, 1998). A számított allél-méreték szórása 0,15 nukleotid méretnél kisebbnek adódott, ami alapján megállapítható, hogy a felmérés során alkalmazott fragmens-analizáló rendszer alkalmas hossz-polimorf allélek 1 nukleotid pontossággal történő genotipizálására.

A budapesti referencia populáció analízise (BuCa) a felmérésbe bevont összes multiplex STR rendszerrel (*ProfilerPlus*, *Identifiler*, *PowerPlex 16*) megtörtént (1. táblázat), a különböző rendszerekkel az átfedő markereken megfigyelt genotípusok nem mutattak eltérést az adott mintákon. A viszonylag nagy mintaszámnak köszönhetően (223 minta) ezt a felmérést egy kisebb léptékű konkordancia tesztelésnek is tekinthetjük, melynek eredménye egybevág a nagy mintaszámú adatbázisokon elvégzett nemzetközi STR primer konkordancia felmérésekkel (Budowle és Sprecher, 2001; Budowle és mtsai, 2001). A különböző multiplex STR rendszerekkel kapott egybevágó eredmények alapján megállapítható, hogy igazságügyi célú genetikai profil alkotásra a genotipizálás megbízhatósága szempontjából mindhárom alkalmazott rendszer egyaránt alkalmas.

3.2. Mikrovariáns STR allélek elemzése

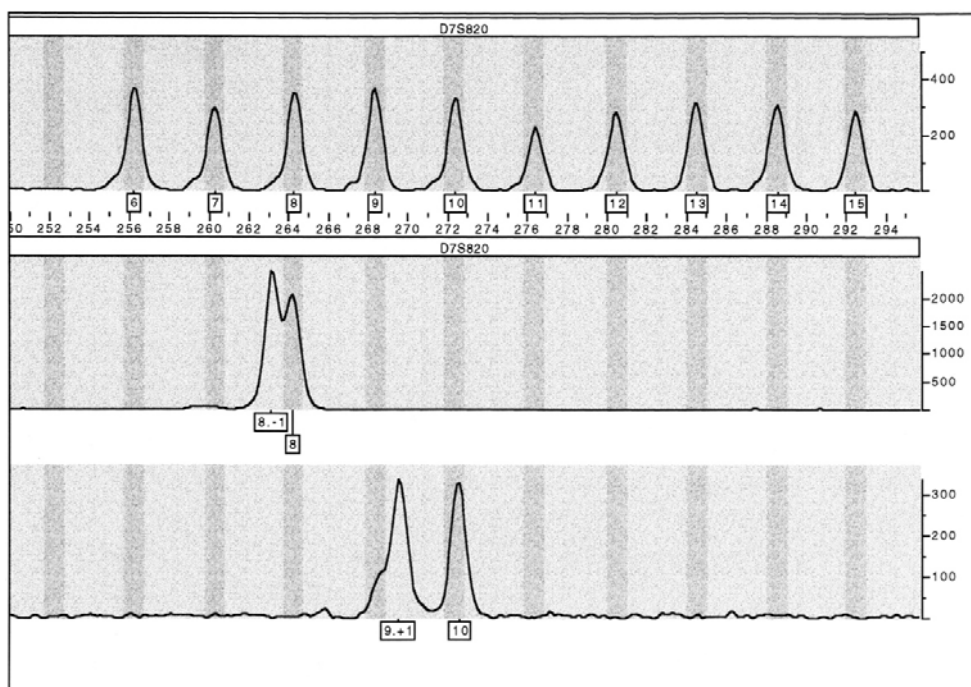
Az STR markerek populációs felmérése során az alkalmazott nagy felbontású genotipizáló rendszernek köszönhetően több, ritkán megfigyelhető, ún. mikrovariáns (nem konszenzus) allélt figyeltünk meg, amelyek méretben nem egyeztek meg az alkalmazott referencia alléllétrák egyik fokával sem, ezért ezeket intermedier alléleknek (vagy interalléleknek) is nevezik. Ezek a hazai és nemzetközi populációs felmérések eredményei alapján igen ritkának számító, az észlelés időpontjában még le nem írt allélek elnevezése elsősorban elektroforetikus eredetű és abból fakad, hogy elektroforetikus mobilitásuk általában eltér az allél-létrát alkotó és konszenzusnak tekintett allélek mobilitásától. Az alkalmazott genotipizálási rendszer nagy elválasztóképességének köszönhetően a mikrovariáns allélek

annak ellenére egyértelműen megkülönböztethetők voltak a konszenzus allélektől, hogy a köztük lévő méretbeli különbség többnyire csak 1 nukleotidnyi volt.

3.2.1. A D7S820 lókusz 8.-1, 9.+1, 10 mikrovariáns alléljainak jellemzése

Felméréseink során a D7S820 lókuszon két, különböző genotípussal rendelkező minta esetében figyeltünk meg a konszenzus allélmérettől eltérő interalléleket (2. ábra). A budapesti referencia populációs mintában (BuCa) egy férfi genotípusú mintából 9var/10 genotípust határoztunk meg (ahol a 9var elnevezés a mikrovariáns allélra utal). Érdekes módon egy kriminalisztikai ügy kapcsán (ami nem volt a populációs felmérés része) egy jugoszláv állampolgárságú, albán nemzetiségű férfi mintájában 8var/8 genotípust figyeltünk meg. A hasonló típusú mikrovariáns allélek szekvencia-analízisét végeztük el (Függelék 7.3.1. közlemény).

2. ábra A D7S820 lókuszon megfigyelt mikrovariáns allélok elektroferogramja



Mindkét heterozigóta genotípus PCR-termékét pGEM-T-ben (Promega) klónoztuk és az allélek legalább 2-2 klónját megszekvenáltuk. A mikrovariáns allélokat szekvencia struktúrájuk alapján 8.-1-es és 9.+1-es allélnak neveztük el (3. ábra).

3. ábra D7S820 mikrovariáns allélek fragmensmérete és szekvencia struktúrája (Referencia-szekvencia: GenBank ID G08616)

Allél	Fragmens méret	5' Flanking régió	Repeat régió	3' Flanking régió
8.-1	205 bp		- (GATA) ₈ -	 - (T) ₈ ATCT -
9.+1	211 bp		- (GATA) ₉ -	 - (T) ₁₀ ATCT -
10	214 bp		- (GATA) ₁₀ -	 - (T) ₈ AATCT -
12 (ref.)	222 bp		- (GATA) ₁₂ -	 - (T) ₉ ATCT -
		24 bp		13 bp 124 bp

A 9.+1 allélben 9 teljes repeat egység volt megfigyelhető a repeat régióban viszont egy T addícióval rendelkezett a 3' flanking régió (T)₉ blokkjában. A 8.-1 allél 8 teljes GATA repeattal bírt a repeat régióban, viszont a 3' flanking régióknak előbbiekben említett T homopolimer száljából egy T deletálódott. Érdekességképpen a 9.+1/10 genotípusú minta 10-es allélja is szekvencia-variációt mutatott, mivel egy T→A transzverzió volt detektálható a T-blokk végén ezzel egy (T)₈ szálát létrehozva (3. ábra). A (T)₉ homopolimer számban megfigyelt variáns szekvenciák azt sejtették, hogy ez a szakasz – magasabb mutációs rátája miatt – a D7S820 lókuszt polimorf része. Ezt a hipotézist a 3' flanking régió további, szülő-gyermek párokból származó mintáinak szekvencia-analízisével lehetne megerősíteni. Néhány Alu elemhez tartozó poli-A szakasról (a komplementer szálon poli-T) ugyanis köztudott a polimorfizmus (Economou és mtsai, 1990) és az is, hogy Alu szekvenciák gyakran összefüggésbe hozhatók rövid ismétlődő szakaszok (STR-ek) keletkezésével (Arcot és mtsai, 1995; Yandava és mtsai, 1997). Mindazonáltal a D7S820 lókuszt (T)₉ blokkjának 3' flanking régió szekvenciája nem mutatott homológiát Alu repetitív elemekkel. A BLAST programmal (Altschul és mtsai, 1997) ellenőriztük, hogy ez a szekvencia nem lehet-e a része egy retrotranszpozíció által átesett pszeudogénnek, de nem találtunk homológiát egyetlen egy akkor ismert génszekvenciával sem (keresést 1999.04.20-án elvégezve az NCBI <http://www.ncbi.nlm.nih.gov/BLAST> honlapján), viszont 100%-os homológiát kaptunk a humán DJ0649P17 PAC klónnal (GenBank AC004848). A teljes D7S820 lókuszt szekvenciája megtalálható volt a klónban, melyet a 7q11.23-q21 kromoszóma-szakaszra térképeztek, ezért lehetőség nyílt a D7S820 lókuszt kromoszómás helyének pontosabb meghatározására, amit korábban csak a 7. kromoszóma hosszabbik karjára (7q) lokalizáltak (Bouffard és mtsai, 1997).

3.2.2. A D13S317 lókuszt egy ritka 7-es alléljának jellemzése

Egy a populációs felmérés részét nem képező kriminalisztikai ügy kapcsán egy testvérpár mindkét tagjában D13S317 (7/12) genotípust mutattunk ki. Mivel a 7-es allélt addig nemzetközileg nem írták le és az allél-létrából is hiányzott, a ritka allél jellemzéséhez szekvenciaanalízist végeztünk el (Függelék 7.3.2. közlemény). Mivel a D13S317 lókuszt amplifikálásához használt *ProfilerPlus AmpFlSTR* multiplex-PCR rendszer primer szekvenciái nem nyilvánosak, publikált GenBank primer szekvenciával (GDB G09017) kíséreltük meg a ritka allél felszorozását. Meglepetésre a publikált primerek használatával 12/12 homozigóta genotípust kaptunk az előzőleg megfigyelt 7/12 helyett. A szekvencia-analízis során a 12-es allél szekvenciája a referencia-szekvenciával megegyezőnek bizonyult. Az amplifikálást egy másik gyártó általi multiplex-PCR rendszerrel megismételve (*Gamma STR Multiplex Kit*, Promega) ugyancsak 12/12 homozigóta genotípus adódott. Az amplifikált 12-es allélek fragmens-mérete a *ProfilerPlus* rendszer használata esetén 224 bp, a GDB primerek esetén 189 bp, míg a Gamma kit esetén 185 bp hosszúságú volt. Ez alapján azt lehetett sejteni, hogy a GDB primerek és a *Gamma* kitben alkalmazott primerek nagyban átfednek egymással, és a 7-es allélek kieséséért (drop-out) a primer szekvencia lehet a felelős. Ennek bizonyítására a forward primertől felfelé és a reverz primertől lefelé tervezett új primerekkel ismételtük meg a PCR-t, amelynek eredményeképpen a 7/12 genotípus újból megjelent. Mindkét allél több klónjának szekvenálása egy 4 nukleotidos (TGTC) szakasz deléciójára hívta fel a figyelmet a reverz primer kötődési helyén a 7-es allélben, egy 4 bp-os (ATCA) upstream inzercióval kombinálódva (4. ábra). A reverz primer-kötő szekvenciának e megfigyelt drasztikus módosulása valószínűsíthetőleg meggátolja a stabil templát/primer pár hibrid létrjöttét a PCR reakció során és ez a 7-es allél kieséséhez vezetett (Boutrand és mtsai, 2001).

4. ábra A D13S317 lókuszt vizsgált 7/12 alléljainak szekvencia struktúrája (aláhúzással jelölve a PCR-amplifikálásra használt GDB primerek; a klónozáshoz és szekvenáláshoz használt primerek kisbetűvel feltüntetve; a deletálódott és inzertálódott szekvencia-részletek vastag betűtípussal szedve)

D13S317 allel 12

5'-gggttgctggacatggtatcACAGAAGTCTGGGATGTGGA---N82---(ATCT)₁₂ATCAATC(ATCT)₃TTCTGTCTGTCTTTTTGGGC---N36---
gaccaacaattcaagctctc-3'

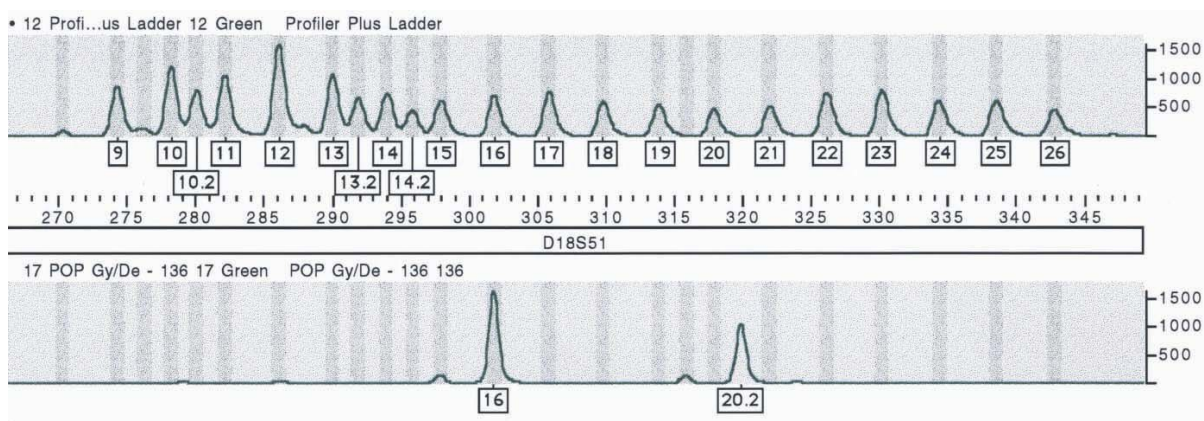
D13S317 allel 7 (variant)

5'-gggttgctggacatggtatcACAGAAGTCTGGGATGTGGA---N82---(ATCT)₇ **ATCAATCAATC**(ATCT)₃TTCTGTCTTTTTGGGC---N36---
gaccaacaattcaagctctc-3'

3.2.3. A D18S51 lókuszon 20.2 mikrovariáns alléljának jellemzése

A kelet-magyarországi Roma populációs minta (DeRo) multiplex STR felmérése során az egyik férfi genotípusú mintában a D18S51 lókuszon 16/20.2 genotípust detektáltunk (5. ábra).

5. ábra A D18S51 lókuszon megfigyelt mikrovariáns allélt tartalmazó genotípus



Az interallélt tartalmazó heterozigóta genotípusú mintát publikált primerekkel (STRBase, <http://www.cstl.nist.gov/biotech/strbase>) amplifikáltuk, a PCR-termékeket TOPO-TA vektorban (Invitrogen) klónoztuk és a szelektált klónokat megszekvenáltuk. Az intermediér méretű 20.2 allél forward és reverz szekvenálása egy (AA) dinukleotid deléciójára derített fényt a 21 teljes (AGAA) repeat-et követő 3' flanking régióban (6. ábra). A mikrovariáns allél DNS szekvenciáját a GenBank adatbázisba exportáltuk (Genbank ID: AY749635), és a nemzetközileg elfogadott nevezéktan szerint (Gill és mtsai, 1997) a helyes elnevezése 21.-2 allél.

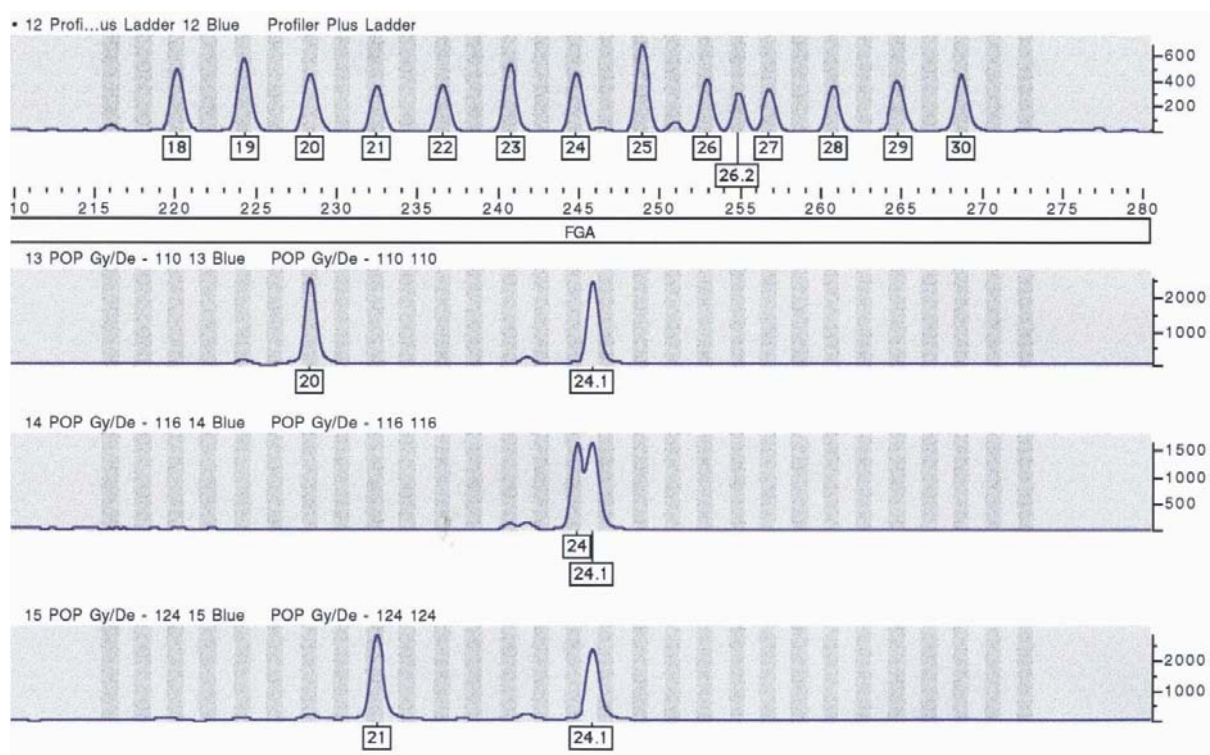
6. ábra A D18S51 allélek fragmensmérete és szekvencia struktúrája (Referencia-szekvencia: GenBank ID L18333; a deletált dinukleotid aláhúzással jelölve)

Allél	Fragmenthossz	5' flanking régió	Repeat régió	3' flanking régió
16	299 bp	██████████	- (AGAA)16 - <u>AAAG</u> AGAGAG -	██████████
20.2	317 bp	██████████	- (AGAA)21 - AG AGAGAG -	██████████
15*	295 bp	██████████	- (ATAG)15 - AAAG AGAGAG -	██████████

3.2.4. Az FGA lókuszt 24.1 mikrovariáns alléljának jellemzése

A kelet-magyarországi Roma populációs minta (DeRo) multiplex STR felmérése során négy mintában 24.1 típusú intermedier allél volt megfigyelhető az FGA lókuszon (7. ábra). A heterozigóta genotípusok 20/24.1, 24/24.1 és kettő mintában 21/24.1 voltak. A mikrovariáns allélt hordozó négy személy 15 STR lókuszos genetikai profiljának vizsgálata nem igazolta a mintaadó személyek közeli (szülő-gyermek) rokonságát.

7. ábra Az FGA lókuszon megfigyelt 24.1 mikrovariáns allélt tartalmazó genotípusok



Az interallélt tartalmazó heterozigóta genotípusú mintákat publikált primerekkel (Mills és mtsai, 1992) amplifikáltuk, a PCR-termékeket TOPO-TA vektorban (Invitrogen) klónoztuk és a szelektált klónokat megszekvenáltuk. Az intermedier méretű 24.1 allélek forward és reverz szekvenálása minden minta esetében egy T addíciójára derített fényt a komplex (CTTT) repeat struktúrát megelőző 5' flanking régióban. (A szekvencia struktúra ábrázolásától az egyszerű mutációs esemény miatt itt most eltekintek.) A szekvenálással jellemzett mikrovariáns allél DNS szekvenciáját a GenBank adatbázisba exportáltuk (Genbank ID: AY749636). Az eddigi megfigyelések alapján a 24.1 allél nagyon kis gyakorisággal fordul elő a populációs adatbázisokban. A 2005. június 10.-i lekérdezés időpontjában mindösszesen há-

rom példány megfigyelése és leírása történt a nemzetközi populációs adatbázisokban (STRBase, <http://www.cstl.nist.gov/biotech/strbase>). A 24.1 mikrováriáns allél jelenléte négy látszólag független mintában, valamint az allél eddig megfigyelt alacsony előfordulási gyakorisága a populációs adatbázisokban, lehetséges beltenyészetet vagy genetikai sodródást jelezhet a kelet-magyarországi roma szub-populációban.

3.3. A populációs mintákban generált autoszómás STR-profilok statisztikai és populációgenetikai elemzése

Az igazságügyi genetikai célú populációs felméréseink során hat populációs minta összesen 1194 személyének autoszóma STR-alapú genotipizálását végeztük el. A genetikai analízishez használt STR lókuszek körét úgy alakítottuk ki, hogy minden egyes minta legalább 13 lókuszon, az ún. CODIS törzs-markereken (CSF1PO-D3S1358-D5S818-D7S820-D8S1179-D13S317-D16S539-D18S51-D21S11-FGA-TH01-TPOX-VWA) meg legyen tipizálva. A budapesti referencia populáció (BuCa), valamint a székely és csángó populációs minták esetében ezenkívül további kettő tetra- (D2S1338, D19S433) és kettő pentanukleotid (Penta E, Penta D) repeat alapú mikroszatellita felmérését hajtottuk végre mindösszesen 700 személytől származó mintán. A két igazságügyben alkalmazott penta-STR marker (Penta E, Penta D) felmérését a baranya megyei (BaRo) és kelet-magyarországi (DeRo) roma populációs mintán, valamint a budapesti askenázi zsidó (BuAs) mintában is elvégeztük.

3.3.1. Allél- és genotípusgyakorisági értékek, lókuszek polimorfizmusának foka

Terjedelmi korlátok miatt a hat populációs minta összesen 1194 személyének – BuCa, Székely, Csángó populációk esetében 17 lókuszos, BaRo, DeRo, BuAs minták esetében 15 lókuszos – autoszómás STR profilját a doktori értekezés nyomtatott formájában nem közlöm. A DNS-profilokat tartalmazó hat táblázat a értekezéshez csatolt CD-rom Autoszómás STR profilok.xls nevű Excel fájlában található meg. A táblázatban minden minta esetében közlöm a vizsgált személy genetikai nemét, amely az Amelogenin lókuszt PCR-fragmens analízisének alapszik. A vizsgálat eredménye alapján az összes mintából 451 férfitől, 743 minta pedig nőtől származik. A genetikai szex és a megadott fenotípusos nem között nem tapasztaltunk eltérést.

A hat populációs minta felmérése során az analizált tizenhét mikroszatellita lókuszon a mikrovariánsokkal együtt összesen 202 alléltípust különböztettünk meg, melyből 26 bizonyult intermedier allélnak. A legtöbb alléltípust az FGA és Penta E lókuszokon detektáltuk, egyenként 20-20 db-ot, a legkevesebbet, hatot a TPOX markeren észleltünk. Az allélek repeat száma 5-35 között mozgott. [A Penta D marker 2.2-es alléljának szekvencia struktúrája 2007. augusztus 10.-i lekérdezés időpontjában még nem került publikálásra, ezért a pontos repeat számát nem ismerjük, annak ellenére, hogy a PowerPlex 16 rendszer alléllétrája tartalmazza ezt az allélt. A felmérésünk során a csángó mintában figyeltük meg egy személy esetében ezt a Penta D allélvariánst, melynek elektroforetikus mobilitása megegyezett az alléllétra 2.2 jelzésű alléljával. Nemzetközi populációs adatbázisok szerint a 2.2 allél előfordulási gyakorisága alacsony, általában $< 1\%$ (Steinlechner és mtsai, 2002; Paredes és mtsai, 2003)]. Az összes megfigyelt alléltípus mintegy 13%-a (26 db) bizonyult intermedier allélnak. Interallél alatt ebben az esetben azokat az allélvariánsokat értjük, amelyek elektroforetikus mobilitása különbözött az egész számú repeat-egységet tartalmazó allélokétól. Ebben az értelemben az interallél az alkalmazott multiplex rendszer alléllétrájának is része lehetett. A legtöbb interallélt (8 db) a D21S11 markeren figyeltük meg, de hasonlóan nagy számú (7 db) interallél volt megfigyelhető az FGA és D19S433 markereken. Egy-egy interallélt találtunk még a TH01, D18S51, D7S820, Penta D lókuszokon. Intermedier allélek jellemzően a komplex repeatet tartalmazó mikroszatellita lókuszokon alakulnak ki, ahol a különböző, változó hosszúságú repeat-egységeket változó méretű közbeiktatott szekvencia szakítja meg.

A hat populációs mintában az analízisbe bevont STR lókuszokon megfigyelt allélgyakorisági értékeket, és a markerek populációgenetikai alapértékeit a Függelék 7.3.1. és 7.3.3-5. publikációi tartalmazzák. A könnyebb vizuális összehasonlíthatóság érdekében az egyes lókuszokra vonatkozó allélgyakoriságokat diagramok formájában is feltüntettem (7.1.2-18. ábrák). Az ennek során alkalmazott görbített vonal-diagram természetesen nem jelenti azt, hogy a megfigyelt tulajdonságok (allél-gyakoriság) kontinuos jellegként értelmezhetők, viszont szemléletesen kiemeli az egyes populációk közötti különbségeket. Az allélfrekvencia-értékek és populációgenetikai alapadatok a mellékelt CD-rom Autószómás STR allélgyakoriság.xls Excel munkafüzetben is megtalálhatók.

Felmérésünk során a leggyakoribb allélnak a TPOX 8 bizonyult, amelynek frekvenciája a budapesti referencia mintában csaknem elérte a 60%-ot (7.1.13. ábra), de a többi populációs mintában is hasonlóan magas arányban volt reprezentálva ($>40\%$). E magas érték miatt a TPOX lókusz polimorfizmusának foka várhatóan a legkisebb a markerek között. Ezt jelzi a populációkban megfigyelt heterozigotizációs értékek összesítése is (Függelék 7.2.1. táblázat),

melyben a TPOX lókuszt rendelkezik abszolút értelemben a legkisebb heterozigotitással (BuCa: 60,5%) az analizált markerek és az összes populáció között is. A populációkon belül vizsgálva az egyes markerek heterozigotitási fokát hatból öt populációban a TPOX lókuszt mutatta a legalacsonyabb értéket; a BaRo mintában megfigyelt relatív magas arány (~71%) valószínűleg az allélgyakorisági diagramon is jól detektálható (7.1.13. ábra) eloszlásbeli különbségből fakad a 11-es allél javára a 8-as allél rovására. A BaRo mintához hasonló eltolódás figyelhető meg a csángó populációban is ezen a lókuszon, viszont érdekes módon a heterozigotitás emelkedése nélkül. A BaRo minta legalacsonyabb heterozigotitását a D7S820 lókuszon mutatta (62,1%), melynek allélgyakorisági diagramján (7.1.8. ábra) ugyancsak jelentős eloszlásbeli különbség volt detektálható a BaRo versus többi öt populációs minta viszonylatában. Mint ahogy a 3.3.2. fejezetben tárgyalásra kerül a BaRo populáció többek között e kettő markeren (TPOX, D7S820) eltérést mutat a HWE-től.

A populációs felmérésünk alapján az allélok a legkisebb maximális frekvencia-értékkel a BaRo személyek Penta E lókuszt oszlottak meg, ahol a két leggyakoribb allél (12/18) megfigyelt gyakorisága alig 14% volt (Függelék 7.1.15. ábra). A hat populációs mintát együtt vizsgálva a lókusztok közül ugyancsak a Penta E marker mutatta a legkisebb maximális allélgyakorisági értéket (19,1%). Ebből a szempontból a Penta E lókuszt az FGA mikroszatellita (24,8%) és a D18S51 marker (27,7%) követte. Ezek az eredmények azt sejtetik, hogy a magyar népességben a DNS-alapú személyazonosításhoz az analizált 17 STR marker közül a Penta E, FGA és D18S51 mikroszatelliták bírnak a legnagyobb igazságügyi megkülönböztető ill. individualizáló erővel.

A hat populációs minta allélgyakoriság-eloszlása szemmel láthatóan kisebb-nagyobb mértékben különbözött egymástól (Függelék 7.1.2-18. ábrák). Általánosságban elmondható, hogy a görbék lefutása a budapesti referencia (BuCa) és askenázi (BuAs) populáció, valamint a székelyek esetében egymásnak többé-kevésbé megfeleltethető, a roma populációk (BaRo, DeRo) és a csángó minta viszont ezektől több markeren is szignifikáns eltérést mutatnak. Jellemző módon a referencia populációban unimodális szerkezetű jelleggörbe bimodálissá, a bimodális trimodálissá változik a roma populációkban, mint ahogy az egyértelműen megfigyelhető a D3S1358 (7.1.2. ábra) és TH01 (7.1.3. ábra) markereken. Ezek a jellegbeli eltérések heterogén alapító hatásoknak és genetikai sodródási eseményeknek tudhatók be. Ezenkívül néhány allél gyakoriságában több markeren is (pl. D21S11, TH01) jelentős eltolódás figyelhető meg a roma populációkban a többi minta csoporthoz viszonyítva, ennek ellenére egyértelműen nem állapítható meg roma-specifikus eloszlási görbe vagy allél egyetlen egy analizált markeren sem.

A tanulmányba bevont 17 autoszómás marker minimális allélgyakoriságát és polimorfizmusának fokát kifejező jellemzők szélső értékeit a 2. táblázatban foglaltam össze. A szélsőértékek mellett a táblázat azt is bemutatja, hogy az adott legkisebbnek vagy legnagyobbnak bizonyult értéket mely populációs mintában és melyik markeren figyeltük meg.

2. táblázat A hat populációs mintán analizált 17 autoszómás STR lókuszt polimorfizmus fokának és a minimális allélgyakoriságok szélsőértékei

		Minimum	Maximum
PD	Érték	0,760	0,980
	Populáció	DeRo	BuAs
	Marker	TPOX	Penta E
PE	Érték	0,332	0,813
	Populáció	DeRo	BuAs
	Marker	TPOX	Penta E
PIC	Érték	0,528	0,900
	Populáció	DeRo	BuAs
	Marker	TPOX	Penta E
H_{obs}	Érték	0,605	0,904
	Populáció	BuCa	BuAs
	Marker	TPOX	Penta E
H_{exp}	Érték	0,586	0,910
	Populáció	BuCa	BuAs
	Marker	TPOX	Penta E
P_{min}	Érték	0,009	0,026
	Populáció	Szekely	DeRo
	Marker	TPOX	D18S51

A táblázatból egyértelműen kitűnik, hogy a legkisebb értéket minden esetben a TPOX, a legnagyobb értéket (a **P_{min}**-t leszámítva) pedig a Penta E lókuszon tapasztaltuk, ami megerősíti e két markernek az allélgyakorisági adatokból már sejthető individualizációs hatékonyságát (legkevésbé illetve leginkább hatékony markerek). Az igazságügyi személyazonosítási hatékonyságot (PD, PE, PIC) kifejező értékek esetében a minimumot mindig a kelet-magyarországi roma (DeRo) mintában, a maximum értéket pedig mindig a budapesti zsidó

(BuAs) populációban kaptuk. A minimális allélfrekvencia 0,9-2,6 % között mozgott. Mivel a kalkulált minimális allélgyakoriság a felmért populációs minta méretével fordítottan aránylik, a minta-pool mérete a DeRo populációban nagyobb, a székely populációban kisebb P_{min} értékhez vezetett. Az adatokból megállapítható (3. táblázat), hogy a populációs adatbázisunk allélgyakorisági értékei 1-2% fölött 95%-os konfidenciával megbízhatónak bizonyultak.

3. táblázat A populációkban megfigyelt minimális allélgyakoriság (P_{min}) szélsőértékek

	minimum	P_{min}	maximum
BuCa	0,0107		0,0126
BaRo	0,0116		0,0141
DeRo	0,0215		0,0260
BuAs	0,0138		0,0164
Szekely	0,0094		0,0113
Csango	0,0109		0,0129

Az STR markerek igazságügyi személyazonosítási hatékonyságát az egyes PCR-alapú multiplex genetikai analizáló rendszerekre összesítve is kifejeztük (Függelék 7.3.1-4. közlemények). Természetesen az individualizáció hatékonysága a szimultán vizsgált STR lókuszok számával arányosan növekedett. Az egyes populációkban a tizenhét (BaRo, DeRo, BuAs populációkban tizenöt) STR markerre kombinált pM (1-PD) és PE értékeket a Függelék 7.2.2-3. táblázatai tartalmazzák.

A rendelkezésünkre álló hat magyar populációs minta széleskörű igazságügyi genetikai STR analízise után megvizsgáltuk azt is, hogy van-e lényeges különbség az individualizáció várt hatékonyságában az egyes populációs minták között. Ennek érdekében a pM és PE értékeket minden populációs minta esetében a 13 CODIS törzs-lókuszra is kombináltuk. A kapott eredményeket a 4. táblázat mutatja be. A táblázatból leolvasható, hogy a 13 STR lókuszról álló CODIS-rendszerrel az igazságügyi személyazonosítás hatékonysága mindegyik populációban rendkívül magasnak és egymással közel azonos nagyságrendűnek bizonyult.

4. táblázat A hat populációs mintán analizált 13 CODIS STR törzs-lókuszra vonatkozó kombinált pM (1-PD) és PE értékek

	pM_{komb}	PE_{komb}
BuCa	2×10^{-15}	0,999994
BaRo	2×10^{-14}	0,999983
DeRo	9×10^{-15}	0,999991
BuAs	6×10^{-15}	0,999991
Szekely	4×10^{-15}	0,999993
Csango	8×10^{-15}	0,999989

3.3.2. A Hardy-Weinberg egyensúly tesztelése

Az allélek lókuszokon belüli esetlegesen meglévő asszociációjának mérésére mind a hat populációs mintában a teljes analizált lókusz-készleten (15 ill. 17 STR marker) az Exact-teszt alkalmazásával (Guo és Thomson, 1992) statisztikai szignifikancia tesztelést végeztünk. Az egyes lókuszokra vonatkozóan a tesztelésbe bevont allélek számát (n), és az Arlequin programmal (Schneider és mtsai, 2000) kapott eredményeket a 5. táblázatban összesítettem. A táblázat a mért valószínűségi értékeket (**P**) tartalmazza, az 5%-os szignifikancia-szintnél kisebb értékeket csillag (*) jellel jelöltem.

Mivel az analízisbe bevont lókuszok alléljai közötti linkage egyensúlyi tesztelés több populációs minta vonatkozásában már korábban megtörtént (Füredi, 2003), melynek során a kapcsoltsági egyensúly EM-alapú analízise az itt vizsgált teljes autoszómás lókuszkészletre is kiterjesztésre került, ezért a lókuszok közötti linkage equilibrium tesztelést nem ismételt meg. A megelőző felmérés alapján a BuCa, DeRo, BuAs populációs mintákban a valószínűségi értékek minden tesztben meghaladták az egyedi szignifikancia-szintet. A baranyai roma populáció esetében több lókuszpár is akadt, ahol az egyedi szignifikancia-szintnél kisebb valószínűségi érték volt mérhető, amit azonban az EM-alapú LD-analízis HWE érzékenysége is okozhatott (Füredi, 2003).

5. táblázat A Hardy-Weinberg egyensúly tesztelése Exact teszttel (*P*)

Marker	BuCa	BaRo	DeRo	BuAs	Székely	Csángó
D3S1358	0,850	0,033*	0,765	0,202	0,223	0,518
TH01	0,255	0,577	0,378	0,471	0,731	0,122
D21S11	0,694	0,337	0,112	0,726	0,566	0,031*
D18S51	1,000	0,296	0,144	0,108	0,208	0,163
PentaE	0,057	0,669	0,507	0,769	0,190	0,357
D5S818	0,643	0,736	0,998	0,415	0,074	0,033*
D13S317	0,874	0,235	0,016*	0,194	0,279	0,632
D7S820	0,341	0,011*	0,169	0,062	0,234	0,342
D16S539	0,222	0,242	0,550	0,437	0,349	0,944
CSF1PO	0,932	0,008*	1,000	0,597	0,491	0,323
PentaD	0,132	0,009*	0,550	0,625	0,481	0,610
VWA	0,098	0,512	0,264	0,496	0,484	0,435
D8S1179	0,364	0,004*	0,054	0,068	0,162	0,087
TPOX	0,899	0,045*	0,858	0,412	0,105	0,465
FGA	0,997	0,831	0,576	0,668	0,220	0,322
D2S1338	0,714	-	-	-	0,859	0,658
D19S433	0,328	-	-	-	0,393	0,092
n	446	412	220	356	514	440

* = Az Exact-teszt statisztikailag szignifikáns eltérést mutat a HWE-től a $p = 0,05$ szignifikancia szinten

Az Exact-teszt a HW egyensúlytól statisztikailag szignifikáns eltérést mutatott ki a roma populációkban és a csángó mintában, az érintett lokuszok a következők voltak:

- a BaRo minta esetében: D3S1358, D7S820, D8S1179, CSF1PO, Penta D, TPOX
- a DeRo minta esetében: D13S317
- a Csángó minta esetében: D5S818, D21S11

A kilenc lokusból hat esetében (D3S1358, D7S820, TPOX, D13S317, D5S818, D21S11) a HWE-től való eltérés nem volt jelentősen szignifikáns ($0,01 < P < 0,05$). Mindhárom populációs minta Exact-teszttel történő HWE elemzését a Bonferroni-eljárás (Weir, 1996) alapján $\alpha' = 0,05$ kísérleti hibarárával is elvégeztük. Ebben az esetben a BaRo és DeRo mintákra (15 teszt) a tesztsorozat egyedi szignifikancia szintje $\alpha = 0,0033$ volt, míg a csángó mintára (17 teszt) az $\alpha = 0,0029$ értéket adódott. A csángó populációban a 0,05 szignifikancia szinttől alacsonyabb értéket mutató mindkét lokusz (D5S818, D21S11) Exact-teszt analízise során $P > 0,01$ valószínűségi értéket kaptunk, ami meghaladta a tesztsorozatok egyedi szignifikancia-

szintjét ($\alpha = 0,0029$). A debreceni roma populációban (DeRo) a D13S317 lókuszt devianciáját ($P = 0,016$) a Bonferroni-eljárás nem erősítette meg ($\alpha = 0,0033$), hasonlóképpen a Baranya megyei roma (BaRo) minta esetében a D3S1358, D7S820, D8S1179, CSF1PO, Penta D, TPOX lókuszok Exact-teszt analízise során kapott valószínűségi értékek meghaladták a teszt-sorozatok egyedi szignifikancia szintjét. Mindezek alapján sem a roma mintákra, sem a csángó populációra nem vethetjük el a HWE meglétét, ha a azt a populációk teljes genomi szintjén vizsgáljuk.

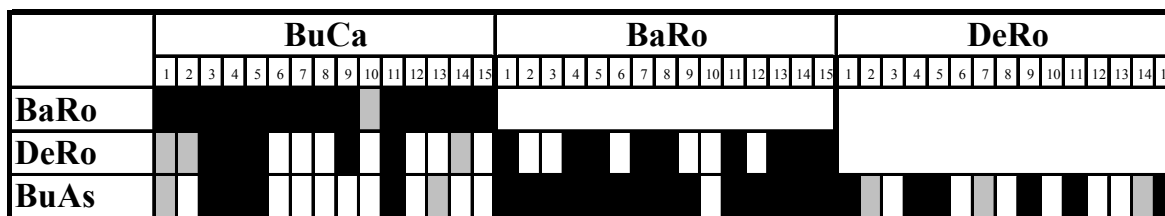
3.3.3. A populációk genetikai struktúrája STR-profilok alapján

Az értekezésben tárgyalt hat magyar populációs minta genetikai homogenitását ill. tagoltságát három különböző statisztikával igyekeztük meghatározni. A G-statisztika és a Wright-féle F-statisztika fenotípusos szinten, az allélgyakoriságok felhasználásával képez különböző értékeket és valószínűségeket. Az AMOVA (molekuláris varianciaanalízis) a Wright-féle F-statisztika molekuláris megfelelője (Excoffier és mtai, 1992) és mikroszatelliták esetén a stepwise mutációs modell (Ota és Kimura, 1973; Valdes és mtai, 1993) alkalmazásával az egyes allélok közötti távolságot molekuláris szinten, az STR lókuszt alkotó polimorf repeatek száma alapján határozza meg.

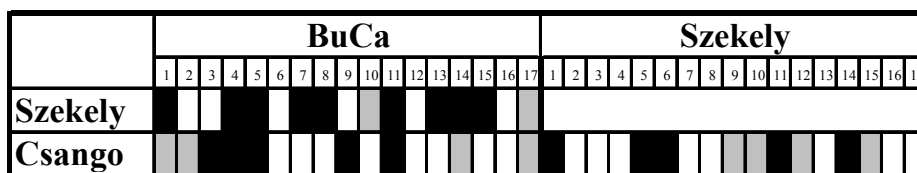
Korábbi publikációkban bemutatásra került a baranyai roma népesség allélgyakoriság adatainak a budapesti referencia populációval való összevetése (Függelék 7.3.1. közlemény), majd a kelet-magyarországi roma és a budapesti askenázi populáció bevonásával a négy magyarországi gyűjtésű populációs minta összehasonlító vizsgálatára került sor (Függelék 7.3.4. publikáció) az összes, itt tárgyalt autoszómás lókuszon, a D2S1338 és D19S433 markerek kivételével. Az egyes populációk között a lókuszpárokon belüli elemzések összesített eredményét a 8.a. ábra mutatja be. Az eredményeket az egyes lókuszokra vonatkozóan tüntettem fel, a G-statisztika esetében azt ábrázoltam, hogy a mért valószínűségi értékek kisebbek-e a felállított 5%-os ill. 1%-os szignifikancia-szintnél.

8. ábra G-statisztika analízis (*P*) a magyar populáció-párokban (A sorszámok a következő markereket jelölik: 1. D3S1358, 2. TH01, 3. D21S11, 4. D18S51, 5. Penta E, 6. D5S818, 7. D13S317, 8. D7S820, 9. D16S539, 10. CSF1PO, 11. Penta D, 12. VWA, 13. D8S1179, 14. TPOX, 15. FGA, 16. D2S1338, 17. D19S433)

a,



b,



Jelölések: ■ $P < 0,01$ ■ $0,01 \leq P \leq 0,05$ □ $P > 0,05$

A 8.a. ábra adataiból az első pillantásra kitűnik, hogy a négy magyarországi populációs minta egymáshoz viszonyított genetikai összetételében, a lókuszokon mért allélgyakoriságok eloszlása tekintetében szingifikáns különbségek vannak. A populáció-párok között mért allélgyakorisági értékek a legtöbb lókuszon és populáció-párban (61/90: ~ 2/3) szignifikáns eltérést mutatnak a G-statisztika elemzés alapján, a legnagyobb arányú eltérést a BuCa/BaRo populáció-pár (mind a 15 markeren) és a BaRo/BuAs populáció-páros (14 marker a 15-ből) adta. Figyelemre méltó volt ugyanakkor a roma populációk egymáshoz viszonyított nagyfokú heterogenitása (szignifikáns eltérés 9 markeren a 15-ből) a G-statisztika teszt alapján. A két budapesti (referencia és askenázi) populációs adatbázis között szintén deviációt mutattunk ki, de ez az eltérés hat lókuszra korlátozódott.

Az erdélyi populációk (székely és csángó) allélgyakoriság eloszlását mind a tizenhét vizsgálatba bevont autoszóma STR lókuszon összevetettük a budapesti referencia populáció adataival és egymással is (8.b. ábra) annak érdekében, hogy világos képet kapjunk arról milyen mértékben illeszkedik a magyarul beszélő, de nem magyarországi populációknak a genetikai állománya az átlag magyar népességhez (Függelék 7.3.5. közlemény). [A nagy földrajzi távolság (>500 km) és a történelmileg nem igazolható genetikai kapcsolat miatt a két roma és a

budapesti askenázi népe sséggel nem hasonlítottam össze az erdélyi populációk adatokat]. Érdekes módon a markerek többsége a BuCa/Székely és BuCa/Csángó populáció-párok viszonylatában is szignifikáns eltérést mutatott a G-statisztika teszt alapján (10 ill. 9 lókuszt), ami az erdélyi és a budapesti populációk részben eltérő genetikai szerkezetére utal. Az eltérésnek ez a mértéke a Székely/Csángó populáció-párban is megfigyelhető (9 lókuszt). Ez a diszkrepancia különösen szembe tünő a csángó populáció-párok esetében a TPOX lókuszon, mely markeren egy nyilvánvaló allélgyakoriságbeli eltolódás figyelhető meg a 11-es allél jávára a csángó mintacsoportban.

Az összesen kilenc elemzett populáció-párosban visszatérő módon a következő lókuszon tapasztaltunk legtöbbször szignifikáns eltérést: D3S1358, Penta E, Penta D (mind a kilenc populáció-párban); D18S51, TPOX (nyolc populáció-párban). Érdekes módon az intrapopulációs elemzés alapján legkevésbé informatívnak minősült TPOX marker csak a BuCa/BuAs populáció-párosban nem mutatott szignifikáns eltérést a G-statisztika teszt alapján. Ez a megfigyelés könnyen magyarázható a TPOX marker allélgyakoriság diagramjának (Függelék 7.1.13. ábra) szemrevételezésével: az eloszlásgörbe a BuCa/BuAs pár esetében fut együtt, az összes többi populációban jelentős (>20%) gyakoriságbeli eltolódás figyelhető meg.

A budapesti referencia populáció allélgyakorisági értékeit szomszédos országok népességeiben generált, publikált adatokkal is teszteltük. A vizsgálat a következő országok populációs adatait ölelte fel: Ausztria (Nussbaumer és mtsai, 2001), Románia (Barbarii és mtsai, 2004), Szlovénia (Drobnič és mtsai, 2001) és Szerbia (Keckarević és mtsai, 2005). Az összehasonlítás során kisebb ($0,05 > P > 0,001$) eltérés volt kimutatható néhány lókuszon (FGA, TPOX, D21S11, Penta E) a budapesti átlag népesség valamint az ausztriai és romániai populációk adatai között. Ennek oka valószínűleg a tanulmányba bevont osztrák és román népesség szub-struktúrálásából fakad, mivel mindkét külföldi populációs minta lokális jellegű szub-populációkban lett gyűjtve. Amikor ezekhez a szomszédos populációkhoz a magyar roma adatokat (Baranya megye és Kelet-Magyarország) valamint a budapesti askenázikat hasonlítottuk, az analízisbe bevont lókusztok többségén szignifikáns eltérés ($P < 10^{-3}$) volt kimutatható a G-statisztika teszt alapján (Függelék 7.3.4. publikáció).

A két erdélyi populáció allélgyakorisági értékeit publikált romániai (Anghel és mtsai, 2003; Barbarii és mtsai, 2004), szerb vajdasági (Veselinović és mtsai, 2004; Keckarević és mtsai, 2005) és egy török populáció (Akbasak és mtsai, 2001) adataival vetettük össze. A populáció-párok összehasonlítása során kisebb eltérést ($0,05 > P > 0,01$) találtunk két lókuszon (D7S820, D19S433) a székely-szerb populáció-párban, a csángó populáció viszont a vizsgált

lókuszok többségén szignifikáns különbséget mutatott mindhárom balkáni eredetű (román, szerb, török) populációs mintával, a legtöbb lókuszbeli diszkrepanciát (10 lókusz) a törökországi népességgel mutatva. A török populáció a székelyekkel szemben is több (6 lókusz), allélgyakorisági értékeit tekintve szignifikáns eltérést mutatott (Függelék 7.3.5. publikáció).

A G-statisztika teszt alapján az allélgyakoriságok szintjén a populációk között tapasztalt inhomogenitást F-statisztikával és AMOVA (Φ -statisztika) analízissel is teszteltük. A négy magyarországi populációs minta vonatkozásában (BuCa, BaRo, DeRo, BuAs) a populáció-párok között kalkulált Wright-féle egylókuszos F_{ST} indexek lényegesen magasabb értékűnek adódtak, mint az előzőleg európai populációkban megfigyeltek és publikáltak (Evett és mtsai, 1996; Budowle és Chakraborty, 2001b). Ezt a genetikai korrelációt a Φ -statisztika analízis (AMOVA) is megerősítette (9.b. ábra). A populáció-párok összehasonlítása során a legmagasabb F_{ST} értékeket a budapesti referencia és a baranya megyei roma populációk között figyeltük meg, a CSF1PO lókuszt kivéve minden egyes markeren szignifikáns eltérést detektálva a két populáció genotípus szerkezetében (9.a. ábra). A szignifikáns eltérést a Φ -statisztika analízis is megerősítette nyolc lókuszon. Hasonló mértékű variabilitás volt kimutatható ugyanezen roma és a budapesti askenázi populáció között is. Az F-statisztika analízis alapján figyelemre méltó inhomogenitás volt kimutatható a roma populáció-párokból (BaRo/DeRo, 9 lókuszon), ami összhangban van a G-statisztika teszt során a megfigyelt allélgyakoriságon alapuló eredménnyel. A roma populációkban megfigyelt inhomogenitást az allélok molekuláris szerkezetén alapuló AMOVA csak csekély mértékben támasztotta alá, mivel a két populációban megállapított Φ_{ST} mindössze négy lókuszon (D3S1358, D21S11, Penta E, TPOX) jelzett szignifikáns eltérést (9.b. ábra).

Mivel hasonlóan magas F_{ST} és Φ_{ST} értékeket ($> 0,01$) mindeddig csak eltérő történelmi-genetikai eredetű etnikai csoportok vagy geográfiaileg-genetikailag izolált populációk között mértek, a magyarországi populációban tapasztalt nagyfokú heterogenitás a vizsgált magyar szub-populációk diverz genetikai történelmével és/vagy izolációjukkal magyarázható.

10. ábra Hagyományos F-statisztika (F_{ST}) és AMOVA (Φ_{ST}) analízis eredménye a budapesti referencia valamint a csíkszeredai székely és a gyimesi csángó populációkban (A sorszámok a következő markereket jelölik: 1. D3S1358, 2. TH01, 3. D21S11, 4. D18S51, 5. Penta E, 6. D5S818, 7. D13S317, 8. D7S820, 9. D16S539, 10. CSF1PO, 11. Penta D, 12. VWA, 13. D8S1179, 14. TPOX, 15. FGA, 16. D2S1338, 17. D19S433)

a,

P	BuCa																	Szekely																	Csango																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
F_{ST}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																							
BuCa																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										</

b,

P	BuCa																	Szekely																	Csango																
Φ_{ST}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
BuCa																																																			
Szekely																																																			
Csango																																																			

Jelölések:  ($F_{ST}, \Phi_{ST} > 0,02; P < 0,1$)  $0,02 > (F_{ST}, \Phi_{ST}) > 0,01; 0,01 < P < 0,05$  ($F_{ST}, \Phi_{ST} < 0,01; P > 0,05$)

A populáció-párok férfi tagjainak Y-kromoszóma STR-profiljait F- és Φ -statisztika analízissel teszteltük (Függelék 7.3.4. publikáció). A haplotípus-párok között kalkulált F_{ST} értékek szignifikáns genetikai varianciát jeleztek a csángó populációs párokban (6. táblázat). Az AMOVA analízis molekuláris szinten is megerősítette ezt a genetikai diszkrepanciát a székely/csángó populáció-párban. A gyimesi csángó populációt érintő genetikai szubsztuktúráldás beltenyésztettségi és/vagy genetikai sodródás általi hatásokkal magyarázható egy ilyen zárt szubpopulációban.

6. táblázat Y-STR haplotípusokból kalkulált F- és Φ -statisztika értékek a budapesti referencia valamint a székely és csángó populációk között

	Székely/Csángó		Székely/BuCa		Csángó/BuCa		Székely/Csángó/BuCa	
	F_{ST}	Φ_{ST}	F_{ST}	Φ_{ST}	F_{ST}	Φ_{ST}	F_{ST}	Φ_{ST}
11 Y-STR	0,007*	0,040*	0,000	0,002	0,009*	0,005	0,010	0,014*

* = Az F- és Φ -statisztika értékek statisztikailag szignifikáns eltérést mutatnak a $P = 0,05$ szignifikancia szinten.

A populációk összehasonlítása során megvizsgáltuk azt is, hogy mely autoszóma STR lókuszok bírnak a legkevesebb és melyek a legtöbb varianciával a populáció-párok között. Mivel ebből a szempontból sok egymáshoz hasonló értéket kaptunk, ezért csak pár jellemző vonást ill. szélsőértéket emelek ki. (Az összehasonlításból a D2S1339 és D19S433 markereket kihagytam, mivel azok csak a BuCa/Székely/Csángó populációkban kerültek felmérésre). A vizsgált 15 autoszómas STR között nem volt egy olyan marker sem, amely legalább egy tesztben ne adott volna szignifikáns eltérést. A lókuszok közül a legkisebb varianciát az összesen lehetséges 3×9 db párosításból a D5S818, CSF1PO és TH01 lókuszokon tapasztaltuk 6, 7 ill. 8 alkalommal a lókuszok előbbi sorrendjében észleltünk szignifikáns eltérést. A legnagyobb varianciát a D3S1358 lókuszon detektáltuk 9, 8 és 4 db (összesen 21) tesztben mutatott inhomogenitással, melyet a D18S51 (8-8-4), Penta E (9-8-3) és TPOX (8-7-5) markerek követtek a 27 páros tesztből 20-ban mutatott diszkrepanciával. Hasonló mértékű populációk közötti variabilitást mutat a Penta D lókusz is (9-6-4) összesen 19 tesztben tapasztalt inhomogenitással. Szembetűnők voltak a lókuszok közötti ún. populáció „érzékeny” markerek: a TPOX lókusz a roma és csángó populációs párokban, az FGA lókusz az erdélyi populációkkal adott szinte minden teszten szignifikáns eltérést.

Annak érdekében, hogy a hat vizsgált populációs adatbázis allélgyakoriság adatait igazságügyi célra is felhasználhassuk, meghatároztuk a négy magyarországi és két erdélyi populációs adatbázis közös leszármazási (coancestry), populáción belüli beltenyésztési (inbreeding) és molekuláris variancia (AMOVA) együtthatóját [F_{ST} (θ), F_{IS} , Φ_{ST}] mindenegyres STR lókuszon. A megállapított statisztikai értékek a 7.a-b. táblázatban láthatóak.

7. táblázat

a, A négy magyarországi populációs minta 15 autoszóma STR lókuszra kalkulált F_{ST} (θ), F_{IS} és Φ_{ST} értékei

	D3S1358	VWA	FGA	D8S1179	D21S11	D18S51	D5S818	D13S317
F_{ST}	0,016	0,009	0,013	0,008	0,009	0,018	0,010	0,017
F_{IS}	0,007	≤ 0	≤ 0	0,028	0,007	0,018	≤ 0	0,040
Φ_{ST}	0,015	0,009	0,013	0,008	0,009	0,018	0,009	0,017

	D7S820	TH01	TPOX	CSF1PO	D16S539	PentaE	PentaD	15 lókusz
F_{ST}	0,033	0,006	0,020	0,003	0,011	0,009	0,027	0,014
F_{IS}	0,064	0,001	≤ 0	≤ 0	0,004	0,021	0,056	0,012
Φ_{ST}	0,033	0,006	0,020	0,003	0,011	0,009	0,028	0,012

b, A két erdélyi és a budapesti referencia (BuCa) populációs minta 17 autoszóma STR lókuszra kalkulált F_{ST} (θ), F_{IS} és Φ_{ST} értékei

	D3S1358	VWA	FGA	D8S1179	D21S11	D18S51	D5S818	D13S317	D7S820
F_{ST}	0,004	0,003	0,003	0,002	0,002	0,002	≤ 0	≤ 0	0,001
F_{IS}	0,008	0,008	0,004	≤ 0	0,013	≤ 0	0,013	0,024	≤ 0
Φ_{ST}	0,005	0,002	0,010	0,007	0,001	0,004	≤ 0	≤ 0	0,002

	TH01	TPOX	CSF1PO	D16S539	PentaE	PentaD	D2S1338	D19S433	17 lókusz
F_{ST}	≤ 0	0,013	0,002	0,002	0,004	0,007	≤ 0	0,001	0,003
F_{IS}	≤ 0	≤ 0	≤ 0	≤ 0	0,037	0,013	0,028	≤ 0	0,002
Φ_{ST}	≤ 0	0,011	0,002	≤ 0	≤ 0	0,002	≤ 0	≤ 0	0,002

A 7. táblázat F_{ST} -értékei alapján a hat vizsgált populációs mintában a 17 autoszóma STR marker közül a D7S820 lókusz mutatta a legnagyobb variabilitást a négy magyarországi (BuCa/BaRo/DeRo/BuAs) populáció között. Ugyancsak magas variációt kaptunk a TPOX és Penta D lókuszok esetében (7.a. táblázat). Az erdélyi populációkat is tartalmazó mintacsoportban (Székely/Csángó/BuCa) a TPOX lókusz mutat kifejezetten magas variabilitást (az összes variáció $> 1\%$ -a a populációk közötti variancia), ami megfelel a populáció-párokból tapasztalt eredményeknek (7.b. táblázat).

Az összes adatbázist és lókuszt együtt vizsgálva a CSF1PO mikroszatellita mutatta a legkisebb változékonyságot az adatbázisok között. Ez utóbbi megfigyelés összhangban van korábbi felmérések CSF1PO lókuszra vonatkozó eredményével (Deka és mtsai, 1995; Budowle és mtsai, 2001a; Budowle és Chakraborty, 2001). A CSF1PO mikroszatellita az 5. kromoszóma hosszú karján lévő növekedési faktor és növekedésifaktor-receptor génklaszter CSF-1R génjének 6. intronjában lokalizálódik. A CSF-1R gén transzmembrán glikoprotein terméke különböző típusú sejtek fontos tirozin-kináz enzime. Ez a génklaszter a humán populációkban szelekciós nyomás alatt állhat. A CSF1PO STR marker alacsony variabilitása annak a bizonyítéka lehet, hogy a CSF1PO lókuszon a természetes szelekció kiegyensúlyozó (stabilizáló) hatású (Deka és mtsai, 1995).

A 7. táblázat F_{IS} adatai arra engednek következtetni, hogy a populációkon belüli beltenyésztés foka nem teljesen elhanyagolható. A magasabb F_{IS} értéket elsősorban az egyéneknek a populáción belüli nem véletlenszerű párválasztása okozhatta (Evett és Weir, 1998).

A populációk páronkénti összehasonlítását a vizsgálatba bevont 15 ill. 17 STR lókusz együttes analízisével is elvégeztük. A 8. táblázat első két sora tartalmazza az összesen kilenc

analizált populáció-pár F_{ST} - és Φ_{ST} -alapú genetikai távolságmátrixát, míg a táblázat harmadik sora a Φ -statisztika szignifikanciáját jelöli. Az F-statisztika valószínűségi értékei (a populációs minták gametafázisának ismeretének hiányában) az *Arlequin* és *GDA* szoftverekkel ebben az esetben nem voltak meghatározhatók.

8. táblázat Az öt vizsgált magyarországi populációs minta páronkénti összehasonlítása a 15 ill. 17 STR lókuszból bevonásával

	BuCa	BuCa	BuCa	BaRo	BaRo	DeRo	BuCa	BuCa	Szekely
	BaRo	DeRo	BuAs	DeRo	BuAs	BuAs	Szekely	Csango	Csango
F_{st}	0,023	0,006	0,003	0,012	0,025	0,007	0,000	0,005	0,002
Φ_{st}	0,014	0,002	0,010	0,009	0,019	0,017	0,002	0,002	0,001
P_{AMOVA}	$< 10^{-3}$	0,100	$< 10^{-3}$	0,002	$< 10^{-3}$	$< 10^{-3}$	0,049	0,058	0,193

A populáció-párok között végzett AMOVA analízis nem tárt fel lényegi különbségeket a székely, csángó és a budapesti referencia populáció-párok között, ami megfelel a három populációs minta között az előbbieken elvégzett 17 lókuszos AMOVA eredményének (7.b. táblázat). Érdekes módon a kelet-magyarországi roma népesség (DeRo) és a budapesti referencia populáció (BuCa) tekintetében sem tárt fel szignifikáns eltérést az AMOVA, ugyanakkor a magyarországi roma populációk genotípus szerkezete között szignifikáns eltérés volt mérhető a 15 autoszóma STR összesített analízise során, ami megerősíti az F-statisztikával lókuszonként elvégzett tesztelés genetikai inhomogenitásra utaló eredményét (9.a. ábra). A többi populáció párosításban is szignifikáns különbséget jelzett a molekuláris varianciaanalízis. Kettő populáció-párban (BuCa/BaRo, BaRo/BuAs), melynek mindkét esetben egyik tagja a Baranya Roma népesség, a 15 lókuszból vonatkoztatott F-statisztika értékek meghaladták a 0,02 szintet.

A közelmúltban az USA ill. az európai népességben elvégzett igazságügyi populációs felmérések eredményei alapján a 0,02 szint meglehetősen magasnak számít (Budowle és mtsai, 2001a; Budowle és Chakraborty, 2001). Mivel a felmérések során nagyobb F_{ST} -értéket elsősorban izolált ill. eltérő genetikai eredetű népcsoportokban detektáltak, ezért megfigyeléseinket a budapesti referencia és askenázi, valamint a baranya megyei roma populáció eltérő genetikai múltja ill. izoláltsága magyarázhatja.

A tanulmányban tárgyalt populációk autoszóma STR allélszerkezetének összehasonlítására irányuló vizsgálataim során a budapesti referencia átlagnépesség (BuCa) allélgyakorisági értékeit további európai, kaukázusi rasszhoz tartozó referencia populációk allélfrekvencia

adataival vetettük össze (Függelék 7.3.1. és 7.3.3. közlemények). Ez az összevetés korántsem nevezhető teljesnek, mivel ezek az eredmények tizenöt lókuszra valamint hét külföldi populációs mintára vonatkoznak. Ugyanakkor úgy gondolom, hogy e vizsgálat konklúziója is segít alátámasztani a budapesti referencia adatbázisunk igazságügyi alkalmazásának hitelességét. A populációs összehasonlítások eredményét táblázatos formában közlöm (9. táblázat).

9. táblázat A budapesti referencia populációs adatbázis (BuCa) és hét európai kaukázusi populációs minta G-statisztikán alapuló összehasonlító elemzése (**P**). (A sor-számok a következő markereket jelölik: 1. D3S1358, 2. TH01, 3. D21S11, 4. D18S51, 5. Penta E, 6. D5S818, 7. D13S317, 8. D7S820, 9. D16S539, 10. CSF1PO, 11. Penta D, 12. VWA, 13. D8S1179, 14. TPOX, 15. FGA)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
osztrák ^a	NS	NS	NS	NS	0,02	NS	NS	NS	NS	NS	NS	NS	NS	<10 ⁻³	0,01
szlovén ^b	NS	NS	NS	NS	-	NS	NS	NS	-	-	-	NS	NS	-	NS
román ^c	NS	NS	10 ⁻³	NS	NS	NS	NS	NS	NS	NS	NS	NS	NS	10 ⁻³	0,03
szerb ^d	-	NS	-	-	-	NS	NS	NS	NS	NS	-	NS	NS	NS	-
olasz ^e	NS	NS	0,01	0,01	-	NS	NS	NS	-	-	-	NS	<10 ⁻³	-	0,03
spanyol 1 ^f	NS	NS	NS	NS	-	NS	NS	NS	-	-	-	NS	NS	-	NS
spanyol 2 ^g	-	NS	-	-	-	-	-	-	-	NS	-	NS	-	NS	-

NS = statisztikailag nem szignifikáns

- = nem teszteltük

^a Nussbaumer és mtsai, 2001

^b Zupanič és mtsai, 1998; Drobnič és mtsai, 2001

^c Anghel és mtsai, 2003; Barbarii és mtsai, 2004

^d Keckarević és mtsai, 2005

^e Garofano és mtsai, 1998

^f Entrala és mtsai, 1998

^g Martin és mtsai, 1995

Egy osztrák, egy szlovén, egy román, egy szerb, egy olasz és két spanyol populációs mintával való összehasonlításban a budapesti referencia populáció az osztrák, a román és az olasz mintától mutatott néhány markeren szignifikáns eltérést. Az osztrák és román populációk esetében kimutatott diszkrepancia a két külföldi mintapopuláció szub-struktúrálásából fakadhat, mivel mindkettő lokális jellegű szub-populációkban lett gyűjtve. Az olasz populációval szemben négy lókuszon megfigyelt deviáció közül csak a D8S1179 markeren detektáltunk statisztikailag nagyon jelentős eltérést. A budapesti referencia populációs mintánk allélgyakoriság-eloszlása tehát nagyfokú hasonlóságot mutatott más európai népcsoportokkal. Ezek az eredmények egybevágtak azzal a tudományos nézettel, amely szerint a kevert magyar

népesség az indoeurópai népekkel közeli genetikai rokonságban áll (Czeizel és mtsai, 1991; Guglielmino és Bérés, 1996; Guglielmino és mtsai, 2000).

3.4. Mitokondriális DNS-profilok generálása a populációs mintákban

A tanulmányom egyik alapvető célja volt mitokondriális DNS-profilok generálása és hitelesítése a vizsgálatba bevont magyar populációkban, ezáltal egy komplex haplotípus referencia adatbázis létrehozása, melynek segítségével lehetőség nyílik mitokondriális DNS szekvenciák analízisén alapuló azonosítások egyezési valószínűségének populációstatisztikailag korrekt becslésére a magyarországi törvényszéki eljárások során. Ennek érdekében hatból négy populációs mintán – a kelet-magyarországi roma (DeRo) populáció DNS mintái az analízis idejére elfogytak, a budapesti askenázi populáció (BuAs) analízise jelenleg is folyamatban van – a teljes kontroll régió szekvenciáját állapítottuk meg egy teljesen automatizált laboratóriumi eljárással (Brandstätter és mtsai, 2004b és 2007a) összesen 776 személy vonatkozásában. A szekvencia-analízis során teljesen robotizált PCR amplifikációt és szekvenálást alkalmaztunk, az adattranszport kizárólag elektronikus úton történt manuális átírás nélkül. A szekvencia-adatok generálása során redundáns megközelítést alkalmaztunk (vagyis több, egymással átfedő szekvenciák kontigjait képeztük) annak érdekében, hogy minimalizáljuk a lehetséges hibákat a szekvencia kiértékelés során. Az alkalmazott módszer a nullára redukálta az ún. „fantom rekombinációk” és hibás szekvenciák keletkezésének esélyét. Ezenkívül 776 magyar személy teljes mitokondriális DNS kontroll régió nukleotid szekvenciáját határoztuk meg; 211 személy az „átlagos” magyarnak tekintett budapesti referencia populációból származik, 205 személy a Baranya megyei roma populációt reprezentálja, 178 minta az erdélyi székely (Csíkszereda központtal), 182 pedig a gyimesi csángó magyar populációban lett gyűjtve. A kettős ellenőrzésen és filogenetikai analízisen átesett DNS szekvenciák haplotípusait táblázatos formában a tanulmány mellé csatolt CD-ROM mtDNS-profilok.doc dokumentuma tartalmazza és az EMPOP (<http://www.empop.org>) igazságügyi célú interneten elérhető mtDNS haplotípus adatbázisba is bekerültek. A DNS szekvenciák elektronikus formátumát a Genbank internetes adatbázisba (<http://www.ncbi.nlm.nih.gov/Genbank/index.html>) exportáltuk (azonosítási számok: DQ359478 – DQ359688, BuCa; DQ359273 – DQ359477, BaRo; EF185421 – EF185780, Székely-Csángó), ahol azok szabadon hozzáférhetőek. A budapesti askenázi populációban (BuAs) gyűjtött minták szekvencia-analízise megtörtént, jelenleg az adatok kiértékelése és publikálása van folyamatban.

3.5. A populációs mintákban generált mitokondriális DNS-profilok statisztikai és populációgenetikai elemzése

A populációs felmérés során megfigyelt mtDNS haplotípusok statisztikai kiértékelését és populációgenetikai elemzését mindenegybes szubpopulációban elvégeztük. A statisztikai felmérések során a teljes kontroll régió (16024-16569, 1-576) szekvencia adataival kalkuláltunk, a policitozin szála C inzercióit a 16193, 309 és 573 nukleotid pozíciókban kihagytuk a számításokból azok nem teljesen egyértelmű tipizálhatósága miatt.

3.5.1. Haplotípus-gyakoriságok és diverzitási paraméterek a populációkban

Az analízisbe bevont négy magyar populációs minta drámai módon különbözött egymástól mitokondriális DNS szinten. A populációkban megfigyelt diverzitási paramétereket a 10. táblázat tartalmazza. A 211 fős budapesti kaukázusi mintában 180 különböző haplotípus volt megkülönböztethető, melyből 167 bizonyult egyedinek. Mindezeket a haplotípusokat 183 variábilis nukleotid pozíció határozta meg. A haplotípus diverzitás kismértékben vagy jelentősen csökkent a többi populáció esetében. Míg a 178 székely személy mintájában 135 különböző (105 egyedi) haplotípust figyeltünk meg, melyet 176 polimorf bázis határozott meg, addig az azonos nagyságrendű, 182 személyből álló csángó populációban csak 84 különböző haplotípus volt kimutatható, melyből 59 bizonyult egyedinek, és ezeket a haplotípusokat 123 variábilis nukleotid bázis definiálta. Még ennél is homogénebbnek bizonyult a roma populáció, a 205 személyben megfigyelt 109 polimorf pont alapján 57 különböző haplotípus volt megkülönböztethető, melyből csak 33-at figyeltünk meg egyetlen alkalommal. A körülbelül azonos mintaméret ellenére a budapesti kevert populációs minta több mint háromszor annyi megfigyelt haplotípussal rendelkezik, mint a Baranya Roma populáció és ötször több egyedi haplotípussal, még akkor is, ha a HV1, HV2 és HV3 régió policitozin szálainak hossz-polimorfizmusait kihagyjuk a számításból. A gyimesi csángó és a baranya megyei roma populációkban a két másik populációhoz viszonyítva tapasztalt szignifikánsan kevesebb variábilis nukleotid pozíció és tranzíció genetikai izolációs hatásokra utalhat ezekben az etnikai csoportokban.

10. táblázat Mitokondriális DNS diverzitási mérőszámok az analízisbe bevont négy magyar populációs mintában [a teljes kontroll régió (16024-16569, 1-576) szekvencia adatai alapján, a C-szálok hossz-polimorfizmusait kihagyva a számításból]

Statisztikai paraméter	BuCa (n=211)	BaRo (n=205)	Szekely (n=178)	Csango (n=182)
Polimorf pontok	183	109	176	123
Tranzíciók	158	96	152	104
Transzverziók	12	3	11	7
Megfigyelt haplotípusok	180 (167 egyedi)	57 (33 egyedi)	135 (105 egyedi)	84 (59 egyedi)
Véletlen egyezési valószínűség (RMP)	1,00%	8,97%	1,04%	5,55%
Haplotípus-párok átlagos eltérése	9,5	9,8	10,8	10,7
Genetikai diverzitás	0,995	0,915	0,995	0,949

A székely populációban kalkulált genetikai diverzitás azonosnak adódott a budapesti referencia populációban számítottal ($H = 99,5\%$), ezzel szemben a csángó populációban már lényegesen alacsonyabb értéket kaptunk ($\sim 95\%$), míg a Baranya Roma populáció ennél is szignifikánsan kisebb diverzitási értékkel rendelkezik ($H = 91,5\%$). (A genetikai diverzitás ebben az esetben annak a valószínűsége, hogy a populációból véletlenszerűen kiválasztott két személy eltérő haplotípussal rendelkezik. Ekvivalens a diploid adatok számított heterizogizációjával.) Az intrapopulációs véletlen egyezési valószínűség (RMP) 1,00 %-os értéke a budapesti és a székely populációban megfelel az átlagos európai populációkban kalkulált adatoknak (Lutz és mtsai, 1998; Vanecek és mtsai, 2004; Brandstätter és mtsai, 2007a), a csángó populációban számított 5,55% és a roma mintákból kalkulált 8,97% viszont extrém magas értéknek számít, és mindeddig kuriózum a publikált populációs adatok körében.

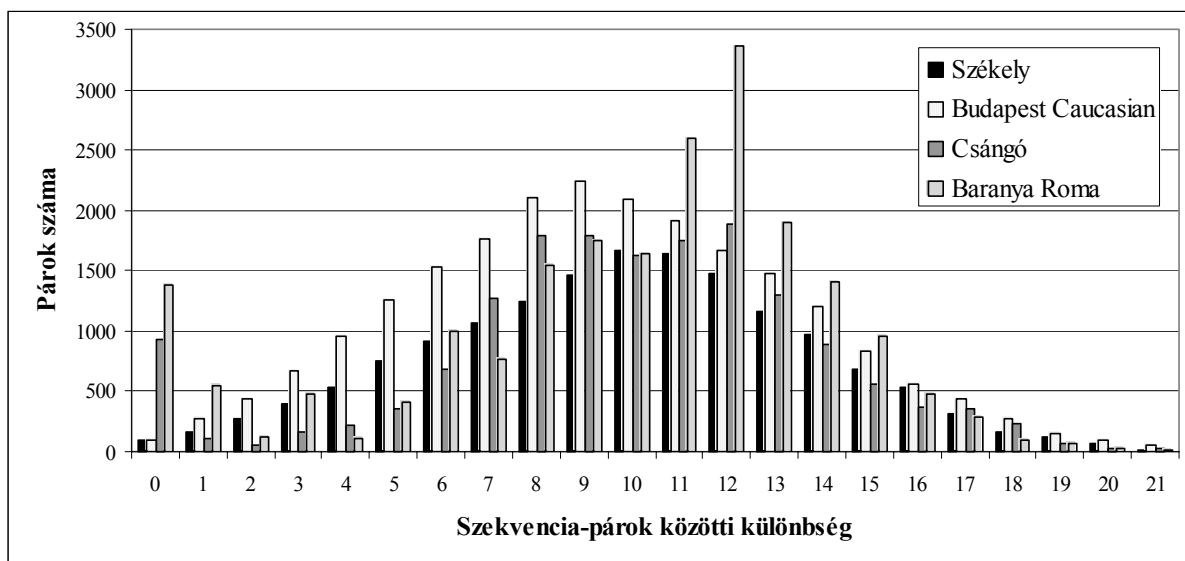
A leggyakrabban megfigyelt haplotípus mind a négy populációban más-más volt. (A populációkban megfigyelt mtDNS profilokat a dolgozat CD-mellékletének mtDNS-profilok.doc dokumentuma tartalmazza.) A budapesti referencia populációban megfigyelt leggyakoribb haplotípus a legelterjedtebb nyugat-eurázsiai haplotípussal (16519C-263G-315.1C) egyezik meg, és 14 BuCa személyben mutattuk ki (211 mintából 6,6%). Ez a haplotípus az európai eredetű H haplocsoportba (R0 főklaszter) sorolható. (A haplocsoport-osztályozást lásd a következő fejezetben.) A populációs felmérés során már megtipizált 776 személyből összesen 26 (3,3%) rendelkezett ezzel a haplotípussal, melyből kilencen a székely (5%), ketten a roma (0,9%) populációhoz tartoztak, egy személy (0,5%) pedig a csángók közül származott. Ezen-

kívül egyetlen egy haplotípus volt, mely mind a négy populációs csoportban előfordult, ez a 73G-263G-315.1C-16126C-16294T-16296T-16304C-16519C haplotípus (T2b haplocsoport), viszont lényegesen alacsonyabb gyakorisággal (1,4-1,9%).

A székelyek között (nem számítva a már előbb említett legelterjedtebb nyugat-eurázsiai típust) leggyakrabban a 16172C-16173T-16519C-263G-315.1C haplotípus fordult elő 3,9%-os aránnyal, mely ugyancsak az európai eredetű H haplocsoportba (R0 főklaszter) tartozik. A gyimesi csángó és a baranya megyei roma népesség haplotípus kompozíciója viszont feltűnő mértékben eltér egy átlagos nyugat-európai populáció képétől. A csángó személyek között megfigyelt leggyakoribb tulajdonság-együttes a 16224C-16311C-16519C-73G-146C-152C-263G-315.1C-498del haplotípus volt 19,2%-os prevalenciával (K1c haplocsoport), utána következett a 16069T-16126C-73G-185A-189G-263G-295T-315.1C-462T-489C haplotípus 8,2%-os gyakorisággal (J1 haplocsoport). A baranya megyei roma népességben a 73G-263G-315.1C-489C-524.1A-524.2C-16129A-16223T-16291T-16298C-16519C haplotípus (haplocsoport M5a) volt a leggyakoribb 28 személyben kimutatva (az analizált személyek 13,6%-a); ennek a szekvenciának egy kicsit módosult változata, amely csupán egy AC repeat inzercióban tér el az 524-es pozícióban az előző szekvenciától, mintegy 8%-ban volt megfigyelhető. A roma populáció mitokondriális DNS szintű homogenitására jellemző, hogy az öt leggyakrabban megfigyelt haplotípus ebben a népességben az összes profil mintegy 40%-át teszi ki. A roma és a csángó populációban megfigyelt jelenség – a leggyakoribb haplotípus a teljes populációs minta mintegy 1/5-ét teszi ki mindkét népességben – arra is utal, hogy ezekben a szub-populációkban a populációk közötti gén áramlás (gene flow) korlátozott, valószínűleg etnikai, nyelvi, kulturális, vallási különbségek miatt.

A DNS szekvenciák páronkénti összehasonlítását mindenegybes populációban elvégeztük. A szekvencia-párok között megfigyelt eltérések eloszlásgörbéje unimodális hisztogramnak adódott a budapesti referencia és a székely populációban, mely a szabályos normál eloszlást követte és 9-10 bázispár eltérésnél mutatta mindkét populációban a maximumot (8. ábra).

11. ábra Szekvencia-párok között megfigyelt eltérések száma a populációkban



A szekvencia-párok átlagos eltérése 9,5 (BuCa) illetve 10,8 (Székely) volt (10. táblázat). Az eloszlásgörbe a csángó és a roma populációk esetében lényegesen eltért a normál eloszlástól, mivel mindkét populáció elemzése egyenetlen bimodális hisztogramot eredményezett egy-egy csúccsal a nulla és a 12 bázispár eltérésnél. A nulla különbségű, egymástól eltérést nem mutató szekvencia-párok nagy számára vezethető vissza a populációkban tapasztalt igen magas véletlen egyezési valószínűség (5,55 ill. ~9%). A 12 bázispárnál tapasztalt maximum érték valószínűleg a divergens mtDNS vonalak (pl. romáknál az M és H haplocsoportba tartozó minták) közötti többszörös páros összehasonlítás eredménye. Unimodális típusú eloszlásgörbék inkább nagyobb kiindulási pool-ból egyenletesen terjeszkedő populációkra jellemzők, míg az egyenetlen eloszlásra egy állandóbb populációmérettel rendelkező, és/vagy heterogén alapító hatásoknak kitett populációk a hajlamosak (Rogers és Harpending, 1992).

3.5.2. Heteroplazmia a populációkban

A szekvencia-analízis során igen alaposan megvizsgáltuk az összes minta szekvenciáját abból a szempontból is, hogy megfigyelhetők-e bennük pont mutációs, azaz heteroplazmia jelenségek. Mind a négy analizált populációban több példát is találtunk arra, hogy adott nukleotid pozícióban mindkét DNS-szál szekvenciája kevert bázisokat, valós heteroplazmiát tartalmaz. Szubsztitúciós heteroplazmia volt megfigyelhető 11 budapesti, 9 roma, 14 székely és 18 csángó személyben mindösszesen 24 különböző nukleotid pozícióban. Összesen 7 személyben figyeltünk meg heteroplazmiát a 16093 pozícióban (T16093Y), (mely pozíció egyébként is

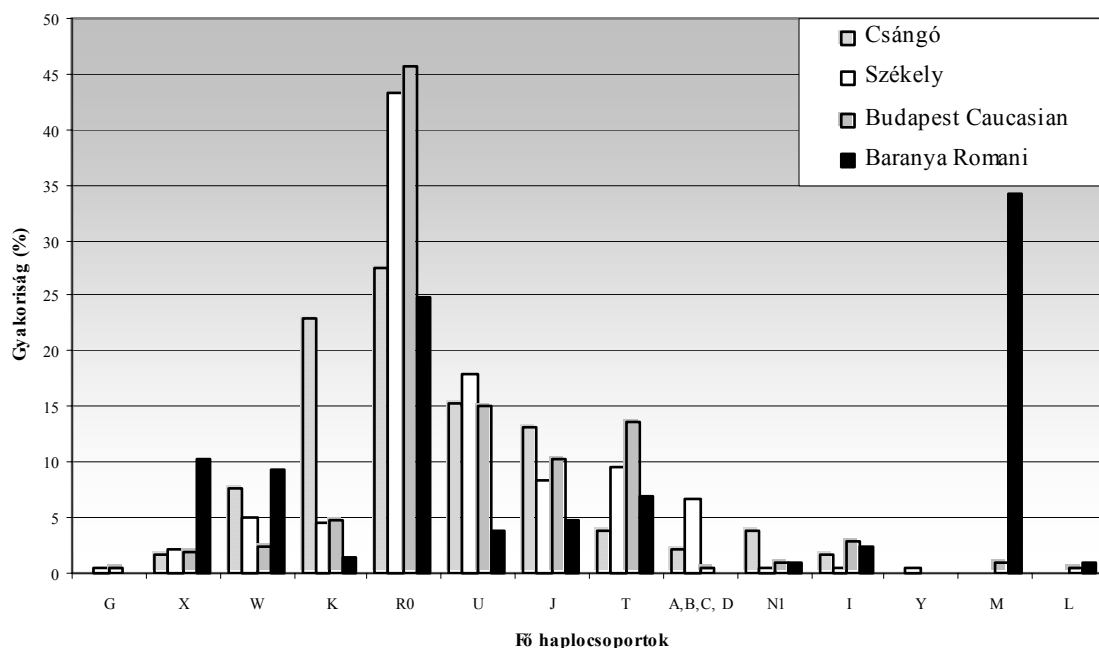
mutáció „hot-spot”-ként ismert a mitokondriális DNS-ben), 4 személyben a 16183 pozícióban (A16183M), két-két személyben a 16311 (T16311Y), a 152 (T152Y), a 183 (A183R) a 204 (T204Y) pozíciókban, a fennmaradó 25 személy pedig további 25 különböző nukleotid pozícióban volt heteroplazmiás. Az összes megszekvenált populációs minta (776) mintegy 6,7%-a (52 személy) mutatott kisebb-nagyobb mértékű heteroplazmiát. Ez az arány kissé magasabb annál, amit eddig vérmintákban tipikusan megfigyeltek (Melton, 2004), viszont konzisztens azokkal az eredményekkel, melyeket a legújabb adatbázis projektek kapcsán tapasztaltak (Brandstätter és mtsai, 2004a, 2004b, 2007a; Irwin-Ross és mtsai, 2006; Irwin és mtsai, 2007).

A pont-heteroplazmiák mellett természetesen hossz-heteroplazmiák is nagy gyakorisággal fordultak elő főleg a HV1 és HV2 régió policitozin szálaiban. A HV3 régió (AC)_n dinukleotid repeat szakasza az 523-as pozícióban több minta esetében is lehetséges hossz-heteroplazmiát mutatott. Más populációs adatbázisokban megfigyeltekhez hasonlóan (Bini és mtsai, 2003; Lutz és mtsai, 2000) populációs mintáink döntő többsége 5 db AC dinukleotid ismétlődést tartalmazott [(AC)₅], a fennmaradó mintákban pedig az (AC)_n repeat nagyfokú variabilitását tapasztaltuk. Az (AC)₅ ismétlődést tartalmazó minták esetében is többször megfigyelhető volt hossz-heteroplazmia (jóllehet extrém alacsony szinten), viszont az összes olyan mintában, mely 6 vagy ennél több AC ismétlődést tartalmazott, nyilvánvaló hossz-heteroplazmiát detektáltunk, melyek általában még kifejezettebbek voltak a 7 vagy ennél több predomináns AC ismétlődést tartalmazó minták esetében. A repeat szám növekedésével a minor komponensű molekulák aránya, és ezzel detektálhatóságuk is növekszik. Ez a megfigyelés összhangban áll az STR markerekről szerzett ismereteinkkel: a PCR amplifikáció során technológiai melléktermékként keletkező ún. „stutter” jelek aránya a repeat szám növelésével emelkedik (Butler, 2005). Az is elképzelhető azonban, hogy az AC ismétlődésekben valóban különböző méretű DNS-molekulákat, így valós hossz-heteroplazmiákat figyeltünk meg. Mindezidáig nagyon kevés populációs adat áll rendelkezésre az (AC)_n típusú heteroplazmiák pontos gyakoriságának megállapítására. Egy ázsiai populációban végzett tanulmány szerint fluoreszcensen jelölt PCR fragmensek analízise alapján 500 egyedből csak három személy bizonyult ténylegesen (AC)_n heteroplazmiásnak (Chung és mtsai, 2005). Ugyanebben a tanulmányban olvasható azt is, hogy hat vagy több AC ismétlődést tartalmazó minták szekvenciáinak klónozása és PCR amplifikációja igenis hajlamos „stutter” termékek létrehozására. Mindezek alapján úgy vélem, hogy egy erre a célra megfelelően kidolgozott klónozási/amplifikálási kísérlet tud konklúzív választ adni arra a kérdésre, hogy az (AC)_n szakaszon megfigyelt kevert DNS-molekulák valóban tényleges heteroplazmiának vagy PCR mellékterméknek minősíthetők.

3.5.3. Az analizált populációk haplocsoport szerkezete

Az azonos evolúciós eredetű (homológ) kontroll régió szekvenciák haplocsoportosítását haplocsoport-specifikus és haplocsoport-kapcsolt szekvencia motívumok alapján mind a négy populációs mintában elvégeztük. Azon minták esetében, melyekben a kontroll régió szekvencia maga nem volt elég informatív az egyértelmű haplocsoportosításhoz, a minták fő nyugat-eurázsiai haplocsoportokba való besorolása a kódoló régió multiplex SNP vizsgálatával történt (Brandstätter és mtsai, 2003b). A haplocsoportosítás összesített eredménye alapján a budapesti referencia, a baranya megyei roma, a csíkszeredai székely és a gyimesi csángó populációk haplocsoport szerkezete rendkívüli mértékben különbözött egymástól (12. ábra).

12. ábra A populációk mtDNS haplocsoport eloszlása



A budapesti referencia populáció haplocsoport eloszlása nagyon hasonló az átlagos nyugat-európai népességeket jellemző adatokra (Függelék 7.1.19. ábra). A BuCa populációs minta szekvenciáinak több mint 90%-a valamelyik nyugat-eurázsiai haplocsoportba sorolható be (Brandstätter és mtsai, 2003b), ezen belül az összes megfigyelt szekvencia mintegy 45%-a az R0 haplocsoport klasztert reprezentálja, mely magába foglalja az Európában leggyakoribb H haplocsoportot is, és a nyugat-európai eredetű népességek megkülönböztető jegye. Ennek a

populációnak a haplocsoport eloszlása nem mutat lényegi eltérést az FBI/SWGDAM kibővített (USA és Európai eredetű) kaukázusi mintákon alapuló adatbázisával szemben (Miller és Budowle, 2001; Monson és mtsai, 2002).

Az R0 klaszter túlsúlya figyelhető meg mindkét erdélyi populációs mintában is azzal a különbséggel, hogy míg a székely populációban ez a klaszter 43,3%-os gyakorisággal fordul elő (Függelék 7.1.21. ábra), addig a csángókban a szekvenciáknak csak 25,7%-a sorolható ide (Függelék 7.1.22. ábra), ami jócskán alacsonyabb az átlagos európai népcsoportokra jellemző 40-50%-os aránynál (Brandstätter és mtsai, 2003b, 2006). Az átlag nyugat-eurázsiai populációk második legnagyobb arányú haplocsoportja az U, mely normál esetben a minták mintegy 20%-át teszi ki. A székely népességben valóban az U haplocsoport volt a második leggyakoribb mintegy 18%-os prevalenciával, a gyimesi csángó populációban viszont az U haplocsoport 15,4%-os gyakoriságát jócskán felülmúlta a K haplocsoport megdöbbentően magas 23,1%-os aránya.

A teljes haplocsoport eloszlás a budapesti referencia és a székely populációk esetében általában jól követi az átlag európai sztenderdet, viszont számottevően eltér attól a gyimesi csángóknál. Ennek oka valószínűleg az itt élő emberek történelme, amely jelentős hatást gyakorolhatott genetikai állományukra is. A jelenleg itt élő emberek elődei több évszázaddal ezelőtt települtek le mai lakóhelyükre, és azóta is többé-kevésbé izoláltan éltek a környék magyar és román nyelvű népességétől, azokkal kevés genetikai kölcsönhatást kialakítva. A csángó populáció torzult haplocsoport eloszlása valószínűleg erre a szigetszerű genetikai izolációra utal, amit további nukleáris lókuszok vizsgálata is megerősít (Guglielmino és Beres, 1996).

Ázsiai népcsoportok inváziója és annak genetikai behatása világosan detektálható a székely népességben. Az ázsiai eredetű A, B, C, G és Y mitokondriális haplocsoportok relatíve nagy aránya (összesen 7,9%) ebben a népcsoportban a székely népséget a középkorban ért ázsiai eredetű nomád behatásokra utalhat.

A baranya megyei roma populáció haplocsoport eloszlása sokkal inkább az európai cigány népségekből megfigyeltre hasonlít (Gresham és mtsai, 2001) az M haplocsoport nagy arányú reprezentációjával, a maradék szekvencia pedig többnyire valamelyik nyugat-eurázsiai haplocsoportba sorolható be (Függelék 7.1.20. ábra). A Baranya Roma minták 35%-a az M haplocsoportba tartozik (magába foglalva a leggyakoribb M5a haplotípust is), viszont ezt a haplocsoportot csak hat különböző haplotípus reprezentálja ebben a populációban. Ezenkívül szinte az összes fennmaradó szekvencia besorolható valamelyik nyugat-eurázsiai (H, X, T, J, U, V, I, W, K) haplocsoportba, csupán három szekvencia (1,5%) tulajdonítható valamilyen egyéb ázsiai vagy afrikai jellegegyüttesnek, melyből kettő egyed az ázsiai N1b1

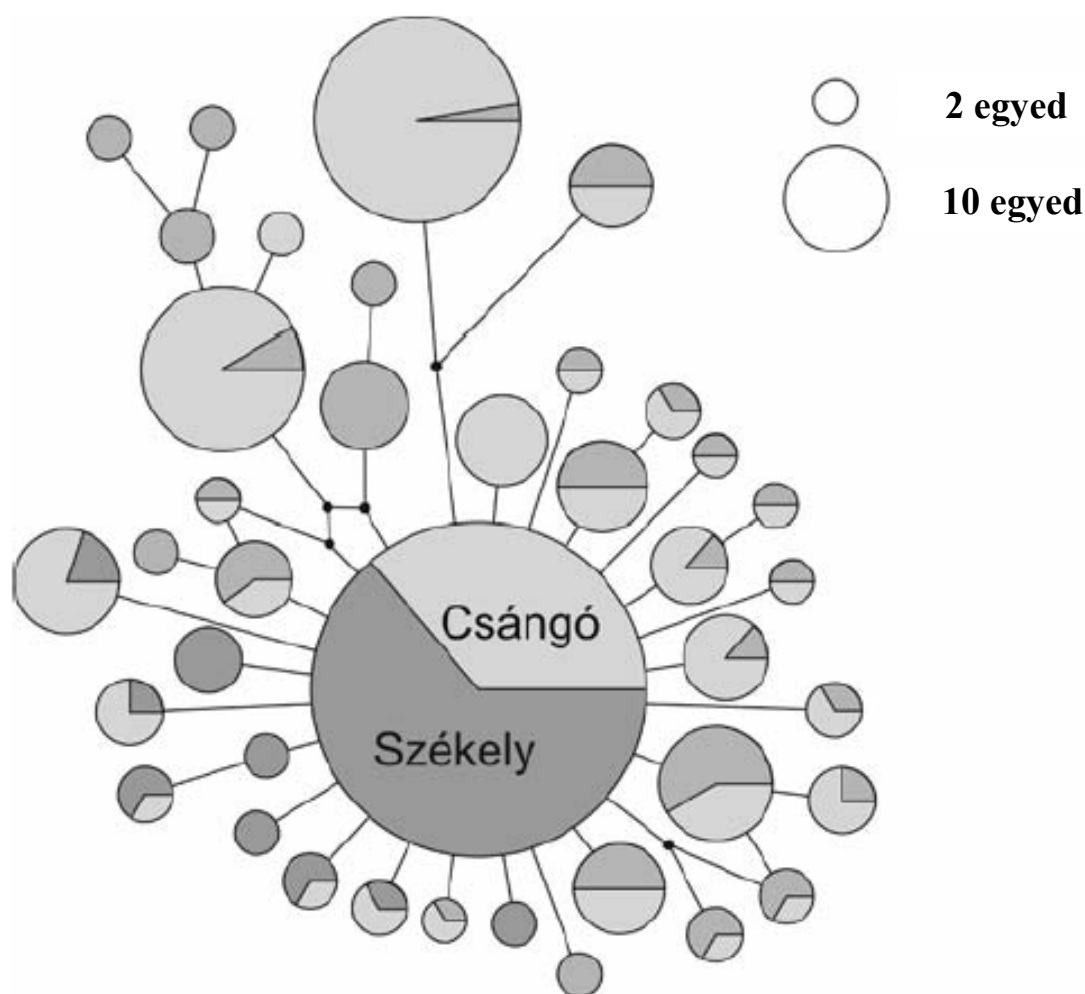
haplocsoportba, egy pedig az afrikai eredetű L3f haplocsoportba tartozik. A baranya megyei roma népesség mtDNS haplocsoport eloszlása nagymértékben hasonlít tizennégy, különböző eredetű európai roma populációban megfigyeltekre (Gresham és mtsai, 2001). Ennek a tizennégyes adatbázisnak az összesítése az M haplocsoport túlsúlyát (26,5%) és a maradék mtDNS vonalak nagyrészenek a nyugat-eurázsiai haplocsoportokba való besorolhatóságát mutatja az európai roma populációkban. A Baranya Roma populációban megfigyelt leggyakoribb haplotípus rendelkezik a legnagyobb prevalenciával a 14 európai roma populációban is, viszont lényegesen kisebb arányban mint a magyar populációs mintában (24,5% versus 10%). Ez a szekvencia a 73G-263G-315.1C-489C-524.1A-524.2C-16129A-16223T-16291T-16298C-16519C volt, amely az M5a szub-haplocsoportba tartozik és a Baranya Roma népesség M haplocsoportvonalainak 73%-át teszi ki. Ez a haplotípus az összesített európai roma populáció (Gresham és mtsai, 2001) M haplocsoportjába sorolt mintáinak csak 40%-át jelenti. Az európai összesített adatbázisban 8,4%-os gyakorisággal megfigyelt második leggyakoribb haplotípus (ugyancsak az M haplocsoport tagja) viszont egyszer sem volt megfigyelhető a baranya megyei roma populációs mintában.

3.5.4. Hálózat analízis és migrációs ráta az erdélyi populációkban

A median joining (MJ) hálózat-rekonstrukció során a számításba bevont haplotípusok gyakoriságát > 1 kritériumra beállítva, az eredetileg 87, egymástól különböző teljes kontroll régió haplotípust tartalmazó „pool” (melyekből a C-szál ins/del és az ún. „speedy” mutációkat előzőleg kiszűrtük) 39 aktív haplotípusra csökkent. (A MJ rekonstrukcióból ezzel kieső, csak egyetlen egy mintában megfigyelt haplotípusok mintegy 70%-a tartozott a székely, a maradék a csángó populációhoz).

A rekonstrukció során készült fa (13. ábra) a csángó populáció elkülönült szigetszerűségét szemléletesen jeleníti meg néhány, több megegyező haplotípust tartalmazó nagyobb csoporttal, melyek viszont lényegesen kisebb mértékben reprezentáltak a székely népességben. Ezzel szemben a különböző haplotípus-csoportok sokkal egyenletesebb méretbeli eloszlását mutatja a székely populáció, jelezve nagyobb genetikai diverzitását. A MJ rekonstrukció alapján a csángó eredetű haplotípusok nagyobb arányú beáramlása tételezhető fel a székely populáció irányába, mint az ellenkező irányba, ami felveti annak a lehetőségét, hogy a csángó és a székely népesség között aszimmetrikus a migráció.

13. ábra Median Joining Network analízis a székely és a csángó populációban. (A kördiagramok mérete a haplotípus gyakorisággal arányos. Sötétszürke: Székely rész; Világosszürke: Csángó rész.)



A MIGRATE program alkalmazásával maximum likelihood eljárással becsült populációgenetikai paraméter mindkét erdélyi populációs mintában alacsony értéket adott (16,234), ami a két populáció közötti génáramlás alacsony mértékét jelenti. A bevándorlási ráta (generációonkénti bevándorlók száma) a két populáció között aszimmetrikusnak nevezhető, mivel a székely populációból a csángók irányába történő bevándorlás 15,52 fő/generáció értékűnek bizonyult, míg a csángó → székely irányba ez a szám 23,25. Mindezek alapján úgy tűnik, hogy a bevándorlás a lényegesen kisebb méretű csángó populáció irányából a számszerűleg sokkal nagyobb székely populációba nagyobb arányú, mint fordított irányba. A migrációs rátában megfigyelt különbségek azzal magyarázhatók, hogy a gyimesi csángó emberek lényegesen elzárkózóbbak az idegen személyek/közösségek bevándorlásával szemben, mint a székelyek,

akik inkább hajlamosak a genetikai keveredésre. A székelyekben megfigyelt nagyobb genetikai diverzitás és az ázsiai eredetű mtDNS vonalak jelenléte is ezt a feltételezést erősíti meg.

Ezt a megfigyelést az előzőekben tárgyalt hálózat rekonstrukció is alátámasztja, mint ahogy a helyben lakó emberek elbeszélései is. A helyi közvélemény szerint ugyanis gazdasági-szociális okok miatt számos csángó ember vándorolt ki az elmúlt évszázadban Gyimesből Székelyföldre (sokan még tovább Magyarországra is), és ugyancsak sok gyimesi csángó lány házasodott Csíkszeredába, amivel tovább örökítették a mitokondriális DNS haplotípusukat a székely népességben felnövő leszármazottaiknak. A székelyek viszont csak kisebb számban, jórészt „fahérgallérosként” (tanárok, munkavezetők, igazgatók, stb.) telepedtek le Gyimesben. Ezenkívül a székelyeknek csak egy lassú, egyszerre mindig kismértékű, de folyamatos beáramlása igazolható Gyimesbe (legfőképp a mindenkori hatóságok előtt menekülve) az elmúlt egy-kétszáz évben.

3.5.5. Az analizált populációk genetikai struktúrája mtDNS profilok alapján

Az analízisbe bevont négy magyar populáció mitokondriális DNS szintű genetikai struktúrájának vizsgálatához, a mtDNS genetikai profilok populációk közötti variabilitásának kalkulálásához molekuláris varianciaanalízist (AMOVA) végeztünk el az *Arlequin ver.2000* (Schneider és mtsai, 2000) szoftver alkalmazásával. Az F-statisztika analízis során populációpáronként kapott F_{ST} értékek azt fejezik ki, hogy a populációk közötti és a populációkon belüli szekvenciapár eltérések alapján az összes genetikai variancia mekkora része tulajdonítható a populációk közötti varianciának. Minél nagyobb F_{ST} értéket kapunk, annál nagyobb a populációk közötti molekuláris távolság. Az F_{ST} értékkel jellemzett populációk közötti variabilitások statisztikai szignifikanciáját szimulációval ugyancsak teszteltük.

Az AMOVA analízis eredménye alapján a szekvencia-párok közötti összes variabilitás 94,8%-a köszönhető a populációkon belüli különbségeknek és a teljes variancia 5,2%-a vezethető vissza a populációk közötti eltérésekre (11.a. táblázat). Ez az eredmény rendkívüli módon eltér azoknak a hasonló kalkulációknak az adataitól, melyeket hat nyugat-európai populációban végeztek, és melynek során a számított variancia közel 100%-a volt az intrapopulációs különbségekre visszavezethető (Brandstätter és mtsai, 2007a). Az átlag nyugat-európai populációkban általában lényegesen kisebb az interpopulációs variancia aránya mint a jelen tanulmányban számított. A magyar szubpopulációk között megfigyelt nagyarányú variabilitás szignifikáns különbséget jelez a vizsgált szubpopulációk genetikai struktúrájában, melyet a szimulációs szignifikancia teszt $p < 0,001$ értéke is megerősít.

11. táblázat Az AMOVA eredménye a négy magyar populáció között

(a) Variancia eredmények (sz.f. = szabadsági fok)

Variancia forrása	sz.f.	Négyzetösszeg	Variancia elemek	Variancia százalék
Populációk között	3	169,318	0,26628 Va	5,24
Populációkon belül	773	3719,519	4,81180 Vb	94,76
Összesen	776	3888,837	5,07807	100,00

(b) Populáció-párok F_{ST} értékei (átló felett) és szignifikancia tesztelése (p , átló alatt)

	Gyimesi Csángó	Csíkszereda Székely	Budapesti Referencia	Baranya Roma
Gyimesi Csángó	-	0,024*	0,025*	0,101*
Csíkszereda Székely	0,000	-	0,001	0,070*
Budapest Referencia	0,000	0,240	-	0,081*
Baranya Roma	0,000	0,000	0,000	-

* = A populáció-párok között megfigyelt F_{ST} értékek szignifikáns eltérést mutatnak a $P = 0,05$ szignifikancia szinten.

A populáció-párok között kalkulált F_{ST} értékek alapján a baranya megyei roma populáció a legnagyobb genetikai távolságot mutatta az összes többi magyar szubpopuláció vonatkozásában (11.b. táblázat). A másik mintanépeség, melyik hasonló mértékű kifejezett különbséget mutat a többi populációhoz viszonyítva a gyimesi csángó népeség volt. A populáció-párok közti F_{ST} értékek szignifikancia becslése (P) során a következő null-hipotézist teszteltük: a populáció-párok nem különböznek egymástól genetikai struktúrájukban. Ez a null-hipotézis csak a Székely/BuCa populáció-pár vonatkozásában teljesült. A genetikai vizsgálatok alapján ugyanis a székely népeség mitokondriális DNS állományának genetikai struktúrája nagyon hasonlít (lényegi eltérést nem mutat) a Budapest városi kevert populációhoz, mely populáció viszont az átlagos nyugat-eurázsiai népeséggel mutat teljesen hasonló jellegeket (Függelék 7.3.6. közlemény). Az a megfigyelés, hogy a baranya megyei roma és a gyimesi csángó népeség genetikai kompozíciója feltűnő módon eltér a többi európai populációtól megerősítést nyert autoszómális és Y-kromoszóma specifikus STR markerek analízise által is (Függelék 7.3.4. és 7.3.5. közlemények).

Annak ellenére, hogy a magyar roma népesség mtDNS haplocsoport szerkezete lényegileg megegyezik a többi európai roma népességekben megfigyelttel (főleg ha őslakos európai populációkkal is összevetjük), alapvető haplotípus gyakoriságbeli különbségek mutathatók ki az összehasonlításba bevont roma populációk között (Gresham és mtsai, 2001). Megelőző molekuláris genetikai populációs tanulmányok (Kalaydjieva és mtsai, 2001) és a cigány népesség történelmének, szociális tagolódásának ismerete már előrejelezte a különböző eredetű, európai roma populációk genetikai szerkezetének jelentős, egymástól való eltérésének lehetőségét. Ezt a megfigyelést a populációs adatok potenciális igazságügyi alkalmazása miatt nem szabad figyelem nélkül hagyni. Mivel mind a mai napig nem áll rendelkezésre publikált, roma közösségekben készült teljes kontroll régió adatbázis, a különböző eredetű roma populációk közötti interpopulációs variabilitás felmérését a baranya megyei és az előzőekben már idézett európai roma populációk (Gresham és mtsai, 2001; Chaix és mtsai, 2004) HV1 szekvencia adataival végeztük el. Az összesen hat roma populációban kalkulált genetikai diverzitási paraméterek értékeit a 12. táblázatban foglaltam össze.

12. táblázat mtDNS diverzitási paraméterek különböző roma populációkban HV1 szekvencia adatok alapján (a HV1 régió ebben az esetben a 16024-16384 pozíciók közti szekvenciát jelenti)

	Lom (bolgár)	Rudari (bolgár)	Kalderash (bolgár)	Baranya (magyar)	Madrid (spanyol)	Turgovzi (bolgár)
Minta méret (N)	43	42	23	205	25	25
Genetikai diverzitás	0,948	0,913	0,949	0,914	0,973	0,977
Haplotípusok	18	15	15	43	11	19
Polimorf pontok	23	25	22	54	23	31
Haplotípus-párok átlagos eltérése	4,28	4,72	4,44	5,17	4,07	6,24

Az összehasonlításba bevont összes roma populáció lényegesen alacsonyabb genetikai diverzitási értékeket mutat, mint a budapesti referencia populáció ($H=0,995$). A magyar roma adatok mutatták az egyik legalacsonyabb genetikai diverzitást a többi roma populációs mintához viszonyítva is, mivel mindösszesen 43 különböző haplotípus volt kimutatható a 205 egyedi mintában (az összes minta $\sim 21\%$ -a; a többi roma populáció átlaga $\sim 50\%$). Ennek ellenére a szekvencia-párok egymástól való átlagos eltérése a magyar roma népességben kissé nagyobb volt, mint a többi roma populációs mintában, melyet kétszer több polimorf pozíció is határozott meg (bár ennek kétségkívül a lényegesen nagyobb mintaszám is az oka lehet).

13. táblázat F_{ST} értékek a baranya megyei roma és a többi európai roma populáció között HV1 szekvencia adatok alapján (16024-16384 nukleotid pozíciók között)

	Baranya Roma	Kalderash (bolgár)	Monteni (bolgár)	Lom (bolgár)	Turgovzi (bolgár)	Kalaidjii (bolgár)	Spanyol Roma
Baranya Roma	-	0,021*	0,088**	0,013*	0,035*	0,038*	0,100**
Kalderash		-	0,033*	0,004	0,001	0,034	0,085**
Monteni			-	0,058**	0,044**	0,095**	0,12**
Lom				-	0,011*	0,009	0,084**
Turgovzi					-	0,011*	0,068**
Kalaidjii						-	0,066*
Spanyol Roma							-

* Szignifikancia-teszt $p < 0,05$; ** Szignifikancia-teszt $p < 0,01$.

Hét európai roma populáció között elvégzett AMOVA analízis azt jelezte, hogy az összevont populációs mintában megfigyelt összes variancia 5,15%-a volt tulajdonítható a roma szubpopulációk közötti eltéréseknek, a maradék pedig az intrapopulációs variációt jelenti. A baranya megyei és a többi európai roma populáció között páronként kapott F_{ST} értékek a 0,014 és 0,100 közötti tartományba estek, és mindegyik populáció-pár F_{ST} értéke szignifikáns eltérést reprezentált a $p = 0,05$ szignifikancia szinten (13. táblázat). A nem magyar roma populációk közötti F_{ST} értékek a 0,001 és 0,095 közti tartományban szórtak (hasonlóan a magyar – nem-magyar roma párokhoz), és a roma populációk között elvégzett összesen 21 páros összehasonlítás során csupán négy populáció-pár között nem tapasztaltunk szignifikáns eltérést.

Annak érdekében, hogy a romák között megfigyelt genetikai szubstrukturáltság mértéke még szemléletesebb legyen, a budapesti referencia populáció szekvencia adatait AMOVA analízissel összevetettük az FBI/SWGDAM adatbázisban fellelhető négy európai eredetű (Ausztria, Franciaország, Görögország, USA “kaukázusi”) populáció adataival (Miller és Budowle, 2001; Monson és mtsai, 2002). Ezekben a populációkban az összes variancia csupán 0,02%-a volt tulajdonítható az interpopulációs különbségeknek (szemben a romák közötti 5,15%-kal). Az őslakos európai populációk között csak egyetlen populáció-pár F_{ST} értéke utalt szignifikáns eltérésre (magyar-francia: $F_{ST} = 0,004$, $p = 0,04$), mely eltérést viszont a Bonferroni-korrekciónak nem erősített meg.

A roma populációknak haplocsoport reprezentáció szinten megfigyelhető, azonos eredetre utaló hasonlósága ellenére megállapítható, hogy léteznek olyan faktorok is, amelyek korlátozták (vagy még ma is korlátozzák) az anyai génvonalak egyes roma közösségek közötti áramlását, még akkor is ha ezek a közösségek földrajzilag igen közel helyezkedtek/helyezkednek el egymáshoz (lásd még: Gresham és mtsai, 2001; Kalaydjieva és mtsai, 2001; Chaix és mtsai, 2004). A budapesti referencia populáció genetikai struktúráját tekintve viszont lényegében megegyezik a többi európai őslakos népességgel, aminek köszönhető az is, hogy a budapesti populációs mintacsoport szekvencia adatai beilleszthetők és együtt kezelhetők a kombinált európai mitokondriális DNS adatbázisokkal az igazságügyi célú összehasonlítások során.

Annak érdekében, hogy a két erdélyi populációs mintában megfigyelt mtDNS kontroll régió diverzitást egyéb szomszédos, dél-kelet európai populációkkal is összehasonlítsuk, a székely és csángó populáció szekvencia adatait egy-egy görög (Irwin-Ross és mtsai, 2006) és macedón (Zimmermann és mtsai, 2007) populációs adatbázissal is összevetettük. Érdekes módon a leggyakoribb csángó haplotípus (16224C-16311C-16519C-73G-146C-152C-263G-315.1C-498del) csak egyszer volt megfigyelhető a székely, a görög és a macedón mintákban, viszont egyszer sem volt jelen a budapesti referencia és a baranya megyei roma populációkban. Ez a megfigyelés is jó összhangban van az Y-kromoszóma STR haplotípusok populációk közötti eloszlásával, mivel a csángó és a székely populációban együttesen megfigyelt 134 különböző Y-STR haplotípusból csak 6 volt jelen mindkét populációs mintában (Függelék 7.3.5. publikáció). A H1 haplocsoportba besorolt leggyakoribb (5%) székely haplotípus viszont az összes többi populációban megtalálható (Buca: 6,1%, macedón: 3%, görög: 2,2%, BaRo: 1%) kisebb-nagyobb arányban. A populáció-párok között kalkulált standardizált géngyakoriságok variancia tartalmának (F_{ST}) összehasonlítása a csángó populációs minta mtDNS eloszlásának kifejezett különbségét mutatta az összes többi (székely, görög, macedón) populációs mintával szemben (14. táblázat).

14. táblázat Az erdélyi valamint a görög és macedón populáció-párok F_{ST} értékei (átló felett) és szignifikancia tesztelése (p , átló alatt)

	Csángó	Székely	Görög	Macedón
Csángó	-	0,026*	0,027*	0,027*
Székely	0,000	-	0,001	0,001
Görög	0,000	0,131	-	0,000
Macedón	0,000	0,178	0,373	-

* = A populáció-párok között számított F_{ST} értékek szignifikáns eltérést mutatnak a $P = 0,05$ szignifikancia szinten.

Ebben a speciális esetben inkább absztraktnak tűnő F_{ST} értékek jobban megérthetők és vizualizálhatók ha a haplocsoport eloszlás diagramokat szemrevételezzük: a gyimesi csángó populáció torzult haplocsoport eloszlása jelentősen eltér az átlagos nyugat-eurázsiai népességekben megfigyeltektől (9. ábra, Függelék 7.1.22. ábra). Az a megfigyelés, hogy a gyimesi csángó népcsoport genetikai struktúrája számottevő eltérést mutat az európai populációkkal szemben megerősítést nyert sejtmagi markerek főkomponens analízise (Guglielmino és Béres, 1996) és Y-kromoszóma specifikus STR markerek vizsgálata által is (Függelék 7.3.5. közlemény). Ezzel szemben a székely populáció mitokondriális DNS szintű genetikai struktúrája látszólag jobban illeszkedik a nyugat-eurázsiai változatokhoz.

A populáció-párok között kalkulált interpopulációs géngyakorisági variancia a 0,001 minimális (BuCa/Székely/Görög/Macedón populációkban) és a 0,101 maximális (BaRo/Csángó populációkban) értékek között változott. Az interpopulációs F_{ST} eredményeket összesítve az F_{ST} értékek megfeleltethetők azoknak az időszakoknak, melyek során az egyes populációk hipotetikus recens szétválása megtörtént, a legnagyobb értékkel ($>0,070$), ennek megfelelően a legkorábbi szeparációval a roma népesség esetében.

3.6. A populációs adatbázisok igazságügyi genetikai célú alkalmazhatósága

Kriminálisztikai és egyéb peres eljárások során az STR- és mitokondriális DNS-profilok populáción belüli relatív gyakoriságának mérésére populációs adatbázisok szolgálnak. Könnyű belátni, hogy a nem rekombinálódó haplotípus adatokat tartalmazó mtDNS adatbázisok méretének növelésével a mtDNS haplotípusok egyezési valószínűségének bizonyító ereje is nő. Ezért a cél minél nagyobb mtDNS adatbázisok generálása, amit genetikailag hasonló adatbá-

zisok összevonásával is elérhetünk. Ehhez viszont azt is garantálni kell, hogy a populációs adatbázisba kerülő populációk között nincs kimutatható jelentős genetikai differenciálódás.

Populációs felméréseink során megállapítottuk, hogy a budapesti referencia populációnk (BuCa) genetikai szerkezetét tekintve nagyon hasonló a többi nyugat-európai populációhoz, ezért nincsen szükség STR és mitokondriális DNS adatnak elkülönítésére az európai kombinált igazságügyi célú adatbázisokban. A székely populáció genetikai struktúrája – STR-profil és mtDNS haplotípus szinten – nem tér el jelentősen a budapesti kevert populációban megfigyelt adatoktól, így ugyancsak jól illeszkedik a nyugat-eurázsiai népességben megfigyelt jellegegyütteshez. Ezzel szemben a roma populációkban autoszóma STR és mitokondriális DNS szinten is, a gyimesi csángó populációban csak mtDNS szinten megfigyelt jelentős szubstrukturálódás miatt nem lenne helyénvaló ezeknek a populációknak a genetikai profil adatait más populációs adatokkal közös adatbázisba összevonni. Meg kell jegyeztem, hogy adatbázis szempontból a különböző eredetű roma populációk STR- és mtDNS-profil adatai sem vonhatók össze.

A különböző eredetű roma populációk mtDNS haplotípus adatait adatbázis szempontból tovább analizáltuk. Ennek során az egyes roma közösségekben megfigyelt leggyakoribb haplotípusok gyakoriságát megvizsgáltuk az összes többi roma szub-populációban és a szub-populációkból összevont ún. kombinált európai roma adatbázisban. Az esetek majdnem mindegyikében az egyik roma szub-populációban leggyakoribbként számontartott haplotípus gyakorisága lényegesen kevesebbnek adódott (sok esetben pedig nem is volt megfigyelhető), azaz alábecsültnék bizonyult a többi populációban. Ez alól csak a megegyező leggyakoribb haplotípussal rendelkező szub-populációk jelentettek kivételt. Megállapítható az is, hogy az általunk kombinált európai roma adatbázis alkalmazása során a szub-populációkban leggyakoribbként definiált haplotípusok gyakorisága jelentősen alulbecsültnék bizonyult. Mindezek alapján kijelenthetjük, hogy minden egyes szekvencia populáción belüli becsült gyakorisága akkor a legkonzervatívabb, ha azokat az egyes populációkban külön-külön kalkuláljuk. A roma népességben belül megfigyelt genetikai differenciálódás mértékéről további, nagyobb mintaszámú roma populációk felmérése átfogóbb képet adhat. Az ilyen típusú zárt közösségekből származó genetikai profil adatokat addig is külön adatbázisban kell tárolni, és extrém óvatossággal kell alkalmazni a genetikai szakvélemények statisztikai interpretálása során.

A magyar populációs adatbázisok összesített eredménye alapján további vizsgálatok igénye merülhet fel a különböző etnikai/történelmi eredetű magyar populációk (kisebbségek) átfogó genetikai felmérése céljából. Ennek során különösen megfontoltan kell megtervezni a populációs mintagyűjtést kisméretű, izolált közösségek esetén. Az értekezésben közölt, több

magyar populációból származó genetikai profilok ugyanakkor a már világszerte publikált populációs adatokkal összehasonlítva a humán migrációs hullámok mélységi elemzésére is alkalmasak lehetnek. A mitokondriális DNS szekvencia-adatok bekerültek az EMPOP on-line adatbázisba (www.empop.org), mely jelenleg a mtDNS teljes kontroll régió szekvenciák legnagyobb populációs adatbázisával rendelkezik. Az igazságügyi genetikai célú adatbázis főként nyugat-eurázsiai eredetű populációk szekvencia-adatait tartalmazza, de a közeljövőben a világ minden részéről származó, ellenőrzött szekvenciák befogadását tervezik.

4. ÚJ MEGÁLLAPÍTÁSOK, KÖVETKEZTETÉSEK

A doktori értekezésemben részletezett populációs felmérést azzal a célkitűzéssel indítottuk útjára, hogy polimorf autoszómális mikroszatellita lókuszok és a mitokondriális DNS kontroll régiójának tipizálása révén felállítsunk egy olyan populációs referencia adatbázist, ami megfelelőnek bizonyulhat a magyarországi igazságügyi genetikai szakértői vizsgálatok korrekt genetikai statisztikai interpretációjára. A populációs felmérés során a mitokondriális DNS polimorfizmusok automatizált genetikai analízisének bevezetésére és hitelesítésére is sor került a magyarországi igazságügyi genetikai gyakorlatba. Kereskedelmileg forgalmazott multiplex-PCR és fluoreszcens detektálási rendszerek lehetőségeit kiaknázva relatíve gyors, és nagyrészt automatizált módon generáltunk multilókuszos (15-17 STR markerre épülő) genetikai profilokat akár több száz vagy ezer populációs minta statisztikai elemzése céljából.

4.1. Az igazságügyi genetikai célú populációs felméréseink során hat magyar populációs minta összesen 1194 személyének autoszóma STR-alapú genotipizálását végeztük el. A genetikai analízishez használt STR lókuszok körét úgy alakítottuk ki, hogy minden egyes minta legalább 13 lókuszon, az ún. CODIS törzs-markereken (*CSF1PO*, *D3S1358*, *D5S818*, *D7S820*, *D8S1179*, *D13S317*, *D16S539*, *D18S51*, *D21S11*, *FGA*, *TH01*, *TPOX*, *VWA*) meg legyen tipizálva.

Referencia populációs mintánkat a budapesti kevert populációból véletlenszerűen kiválasztott 223 személy jelentette, akiket az összes kiválasztott marker (az előbbieken túl még a *D2S1338*, *D19S433*, *Penta E*, *Penta D* lókuszok) genotipizálásával jellemeztünk. A budapesti referencia populáció analízise a felmérésbe bevont és a laboratóriumunkba bevezetett, igazságügyi genetikai szempontból hitelesített összes multiplex STR rendszerrel (*ProfilerPlus*, *Identifiler*, *PowerPlex16*) megtörtént, a relatíve nagy mintaszámnak köszönhetően ezt a felmérést kisebb léptékű primer konkordancia tesztelésnek is tekinthetjük a kevert magyar populációban. A különböző multiplex rendszerekkel az összes mintán átfedő markereken kapott egyező genotípusok eredménye egybevág a nemzetközi adatbázisokon eddig elvégzett konkordancia felmérések eredményeivel.

Az igazságügyi aspektusú autoszóma STR analízisbe a világon elsőként vontunk be nem magyarországi, de magyar populációkat (székely és gyimesi csángó), melyeken az összes felmérésbe bevont mikroszatellita marker genotipizálását elvégeztük. Az erdélyi eredetű populációs minták mellett egy-egy dél-magyarországi (Baranya megye) és kelet-magyarországi

(Debrecen környéki) roma populáció, valamint egy budapesti askenázi zsidó populáció elemzésére is sor került 15 autoszómás STR markerre nézve.

4.2. Az igazságügyi genetikai vizsgálatok alapkövetelménye a genotípus pontos leírása. Az STR markerek esetében ez referencia-allélok segítségével történik. Ha az ismeretlen allél nem felel meg egyik referencia-allélnak sem (mikrovariáns), a tipizálandó allélt jellemezni kell.

Populációs felméréseink során négy mikroszatellita lókuszon (*D7S820*, *D13S317*, *D18S51*, *FGA*) összesen hatféle olyan allélvariánst detektáltunk, melyek egyszerű fragmensanalízissel nem voltak definiálhatók. A mikrovariáns STR alléleket PCR termékük klónozásával és szekvenálásával jellemeztük. A szekvencia-analízis során az STR lókusz polimorf ismétlődő szakaszát, valamint a repeatek határoló régióit érintő mutációkat is megfigyeltünk. A *D13S317* lókusz ritka 7-es alléljában egy 4 nukleotidos (TGTC) szakasz deléciója a reverz primer kötődési helyén az allél elvesztését (drop-out) okozta a módosított primerrel történő PCR-amplifikálás során. A *D7S820* lókuszon detektált mikrovariáns allélek pontos DNS szekvenciáinak leírásával lehetőség nyílt a *D7S820* lókusz kromoszóma lokalizációjának pontosabb meghatározására a 7q11.23-q21 kromoszóma-szakaszra. A kelet-magyarországi roma populáció több, egymástól látszólag független mintájában az *FGA* lókuszon leírt ugyanazon mikrovariáns allél jelenléte, az allél alacsony előfordulási gyakorisága beltenyészetet vagy genetikai sodródást jelezhet ebben a roma szub-populációban.

4.3. Az allél- és STR-profil gyakorisági adatbázisokat tesztelni kell arra nézve, hogy felhasználhatók-e a DNS-vizsgálat bizonyító erejének korrekt megbecsüléséhez. Az STR profil egyezési valószínűség kalkulációjához számos paramétert és körülményt meg kell ismerni, amelyek közül az egyik legfontosabb a vizsgált STR lókuszok polimorfizmusának foka.

Felmértük az analizált STR lókuszok polimorfizmusának fokát minden egyes populációs mintában. A 13 autoszómás (az ún. CODIS törzs-) lókuszra kombinált átlagos egyezési valószínűség (pM) $10^{-14} - 10^{-15}$ értékek között mozgott az egyes populációs mintákban. A kombinált átlagos szülősegi kizáró erő (PE) ugyanerre a lókusz-készletre vonatkozóan 0,999983-0,999994 értékek között variált. A 13 STR lókuszából álló CODIS-rendszerrel az igazságügyi személyazonosítás hatékonysága mindegyik populációban rendkívül magasnak és egymással közel azonos nagyságrendűnek bizonyult. Megállapítható, hogy a 13 autoszómás CODIS STR marker analízisével potenciálisan biztosítható a Magyarországon elkövetett bűncselekményekben érintett személyek individualizációja.

A markerek polimorfizmusának fokát jellemző paraméterek alapján a DNS-alapú személyazonosításhoz a magyar népességben a 17 STR marker közül a Penta E, FGA és D18S51 mikroszatelliták bírnak a legnagyobb intrapopulációs individualizációval, míg a TPOX lókuszt mutatta a legalacsonyabb értéket.

Az elemzett populációkat egymással összevetve, általában alacsonyabb diverzitási értékeket kaptunk a roma szub-populációkban és a csángóknál, míg a budapesti referencia és askenázi zsidó populációban valamint a székely népességben magasabb diverzitási értékek adódtak.

A hat populációs minta allélgyakoriság-eloszlása többé-kevésbé különbözött egymástól. Az allélgyakoriság-eloszlásokban megfigyelt jellegbeli eltérések heterogén alapító hatásokra és/vagy genetikai sodródási eseményekre utalhatnak a magyar szub-populációkban. Az allélgyakoriság diagramok alapján nem állapítható meg etnikum specifikus allél vagy eloszlásgörbe egyetlen egy markeren sem.

4.4. A lókuszon belüli potenciális allélikus asszociáció statisztikai tesztelése (Hardy-Weinberg equilibrium) eredményeként az egyensúlytól való szignifikáns különbséget kaptunk a baranyai roma populáció D3S1358, D7S820, D8S1179, CSF1PO, Penta D, TPOX lókuszeit, a kelet-magyarországi roma minta D13S317 lókusza, valamint a gyimesi csángó népesség D5S818, D21S11 lókuszeit esetében. Az összesen kilenc lókuszból hat esetben (D3S1358, D7S820, TPOX, D13S317, D5S818, D21S11) a HWE-től való eltérés nem volt jelentősen szignifikáns ($0,01 < P < 0,05$).

A Bonferroni-eljárás alkalmazásával az Exact-teszt analízise során kapott valószínűségi értékek meghaladták a tesztsorozatok egyedi szignifikancia szintjét. Mindezek alapján sem a roma mintákra, sem a csángó populációra nem vethetjük el a HWE meglétét, ha a azt a populációk teljes genomi szintjén vizsgáljuk.

4.5. Magyarországon elsőként végeztük el a humán mitokondriális DNS teljes kontroll régió polimorfizmusainak populációs felmérését. Egy magyar mitokondriális DNS-profil referencia adatbázis létrehozása érdekében négy populációs mintán összesen 776 személy vonatkozásában – budapesti kevert, baranya megyei roma, csíkszeredai székely, gyimesi csángó – a mtDNS teljes kontroll régió szekvenciáját határoztuk meg egy teljesen automatizált laboratóriumi eljárással. A szekvencia-analízis során teljesen robotizált PCR amplifikációt és szekvenálást alkalmaztunk, az adattranszport kizárólag elektronikus úton történt. A szekvencia-adatok generálása során redundáns megközelítést alkalmaztunk annak érdekében, hogy

minimalizáljuk a lehetséges hibákat a szekvencia kiértékelés során. Az alkalmazott módszer nullára redukálta az ún. „fantom rekombinációk” és hibás szekvenciák adatbázisba való kerülésének esélyét.

Az igazságügyi aspektusú mitokondriális DNS analízisbe a világon elsőként vontuk be a roma populációt, valamint romániai területi elhelyezkedésű populációkat (székely és gyimesi csángó). A kettős ellenőrzésen és filogenetikai analízisen átesett DNS szekvenciák haplotípusai bekerültek az on-line hozzáférésű EMPOP igazságügyi célú mtDNS haplotípus adatbázisba (<http://www.empop.org>).

4.6. Az analízisbe bevont négy magyar populációs minta drámai módon különbözött egymástól mitokondriális DNS szinten. A budapesti kevert (referencia) népességben megfigyelt haplotípus diverzitás többé-kevésbé csökkent a többi populáció esetében. A körülbelül azonos mintaméret ellenére a budapesti kevert populációs minta több mint háromszor annyi különböző megfigyelt haplotípussal rendelkezik, mint a Baranya Roma populáció és ötször több egyedi haplotípussal, még akkor is, ha a policitozin szálak hossz-polimorfizmusait kihagytuk a számításból. Hasonlóképpen a csángó populációban szignifikánsan kevesebb különböző és egyedi haplotípus volt megfigyelhető a székely populációhoz képest, amellyel pedig szoros geográfiai szomszédságban és feltételezhetően genetikai kapcsolatban is áll. A gyimesi csángó és a baranya megyei roma populációkban a két másik populációhoz viszonyítottan tapasztalt szignifikánsan kevesebb megfigyelt haplotípus genetikai izolációs hatásokra utalhat ezekben az etnikai csoportokban.

A budapesti referencia és a székely populációban kalkulált genetikai diverzitási paraméterek megfelelnek az átlagos európai populációkban kalkulált adatoknak [géndiverzitás (H) $\sim 0,995$; intrapopulációs véletlen egyezési valószínűség (RMP) $\sim 1\%$], a csángó és a roma populációkban viszont a nemzetközi adatokhoz képest extrém alacsony diverzitási értékek adódtak [géndiverzitás (H) $< 0,95$; intrapopulációs véletlen egyezési valószínűség (RMP) $> 5\%$], ami mindeddig kuriózum a publikált nemzetközi populációs adatok körében.

A leggyakrabban megfigyelt haplotípus mind a négy populációban más és más volt. A budapesti referencia populációban megfigyelt leggyakoribb haplotípus a legelterjedtebb nyugat-eurázsiai haplotípussal egyezik meg, és kisebb-nagyobb gyakorisággal mind a négy populációban előfordult. Ezen kívül még egyetlen egy haplotípus fordult elő mind a négy populációs mintában, viszont lényegesen kisebb gyakorisággal. A gyimesi csángó és a baranya megyei roma népesség haplotípus kompozíciója feltűnő mértékben eltért az átlag nyugat-eurázsiai populáció képétől. A roma populáció mitokondriális DNS szintű intrapopulációs

homogenitására jellemző, hogy az öt leggyakrabban megfigyelt haplotípus ebben a népességben az összes profil mintegy 40%-át teszi ki. A roma és a csángó populációkban megfigyelt mitokondriális DNS szintű homogenitás arra is utalhat, hogy ezekben a populációkban az interpopulációs gén áramlás (gene flow) korlátozott.

4.7. A populációs minták mitokondriális DNS haplotípusainak leszármazási alapon történő csoportosítása alapján a budapesti referencia, a baranya megyei roma, a csíkszeredai székely és a gyimesi csángó populációk haplocsoport szerkezete rendkívüli mértékben különbözött egymástól. A budapesti referencia populáció haplocsoport eloszlása nagyon hasonló az átlagos nyugat-európai népeiségeket jellemzőre és ebben a populációban megfigyelt szekvenciák több mint 90%-a besorolható valamelyik nyugat-eurázsiai haplocsoportba.

A székely populáció haplocsoport eloszlása általában jól követi az átlag európai népeiségekre jellemző hisztogramokat, a gyimesi csángók viszont számottevően eltérnek attól. Ennek oka valószínűleg az itt élő emberek eredetében, történelmében keresendő, amely jelentős hatást gyakorolhatott genetikai állományukra is. A csángó populáció torzult haplocsoport eloszlása szigetszerű genetikai izolációra utal. A székely népeiséget a középkorban ért ázsiai eredetű nomád behatásra utalhat az ázsiai eredetű mitokondriális DNS haplocsoportok relatíve magas aránya (~ 8%) a székely populációban.

A baranya megyei roma populáció haplocsoport eloszlása sokkal inkább az európai cigány népeiségekben megfigyeltre hasonlít az M haplocsoport nagyon nagy arányú (>30%) reprezentációjával, a maradék szekvenciák többsége pedig általában besorolható valamelyik nyugat-eurázsiai haplocsoportba. A baranya megyei roma populációban megfigyelt leggyakrabban haplotípus bír a legnagyobb prevalenciával az összesített európai roma populációs mintában is.

4.8. A populációs minták autoszóma STR-profiljaiban és mtDNS haplotípus szerkezetében rejlő genetikai strukturáltság felmérése annak tisztázása érdekében történt, hogy populációs mintáink és a budapesti referencia-adatbázisunk allél- és haplotípus-gyakorisági értékei vajon milyen mértékben fogadhatók el hitelesnek a teljes magyar népeiségre nézve.

4.8.1. A hat populációs minta autoszóma STR szintű genetikai összetételében, a mikroszatellita lokuszokon mért allélgyakoriságok és allél-struktúrák eloszlása tekintetében szingifikáns különbségeket detektáltunk. Valamennyi statisztikai teszt a legtöbb szignifikáns eltérést a baranyai roma/budapesti referencia ill. a baranyai roma/budapesti askenázi populá-

ció-párosban mutatta. G- és F-statisztika alapján figyelemre méltó volt a roma populációk egymáshoz viszonyított nagyfokú heterogenitása is. A roma populációknak az allélgyakoriságok szintjén tapasztalt heterogenitását az allélok molekuláris szerkezetén alapuló AMOVA csak csekély mértékben támasztotta alá. A populáció-párok közül a legkisebb genetikai tagozódást autoszóma STR szinten a budapesti referencia/askenázi párosban kaptuk. A populáció-párok között kalkulált egylókuszos F_{ST} indexek lényegesen magasabb értékűnek adódtak a négy magyarországi populációs minta vonatkozásában, mint előzőleg európai populációkban megfigyeltek. Ezt a genetikai korrelációt a Φ -statisztika analízis (AMOVA) is megerősítette. Kettő populáció-párban (BuCa/BaRo, BaRo/BuAs) a 15 STR lókuszra vonatkoztatott F-statisztikai értékek meghaladták a 0,02 szintet. Mivel nemzetközi felmérések során hasonló mértékű F_{ST} értéket elsősorban izolált ill. eltérő genetikai eredetű népcsoportokban detektáltak, ezért eredményeink a magyarországi populációk eltérő genetikai múltjára ill. izoláltságára utalhatnak.

A markerek többsége a budapesti referencia populáció, a csíkszeredai székely és a gyimesi csángó populáció-párok viszonylatában is szignifikáns eltérést mutatott a G-statisztika teszt alapján ami az erdélyi és a budapesti kevert populáció részben eltérő genetikai szerkezetére utalhat. (Az eltérés a székely/csángó populáció-párban is megfigyelhető). A populáció-párok között kalkulált egylókuszos F_{ST} és Φ_{ST} (interpopulációs molekuláris variancia) értékek a lókuszok túlnyomó többségén nem erősítették meg az allélgyakoriság eloszlásban megfigyelt fenotípusos eltéréseket, ami alapján lényegesen kisebb mértékű szubsztuktúrálódás tételezhető fel a budapesti referencia, a csíkszeredai székely és a gyimesi csángó populációk között, mint a négy magyarországi populációban tapasztaltak.

A vizsgált 17 autoszómás STR között nem volt egyetlen egy marker sem, amely legalább egy tesztben ne adott volna szignifikáns eltérést. Azon 15 lókusz közül, amelyet mindegyik populációs mintán elemeztünk, a populáció-páronkénti összehasonlításban a legkisebb változatosságot a D5S818, CSF1PO és TH01 lókuszokon, a legnagyobb genetikai varianciát a D3S1358, D18S51, Penta E és TPOX lókuszokon detektáltuk. Ha a populációkat nem párokban, hanem együtt vizsgáltuk, a 17 STR marker közül a D7S820 lókusz mutatta a legnagyobb variabilitást a megfigyelt F_{ST} értékek alapján. Ugyancsak magas variációt kaptunk a TPOX és Penta D lókuszok esetében. A legkisebb változékonyságot az adatbázisok között a CSF1PO mikroszatellita mutatta. Az eredmények alapján az analizált 17 autoszómás STR marker közül a CSF1PO és TH01 lókusz bír a legkisebb, a D7S820 és TPOX lókusz rendelkezik a legnagyobb információ-tartalommal a magyar népességben a populációk elkülöníthetősége szempontjából. A hat populációban mért F_{IS} értékek nagyságából arra is lehet következtetni, hogy a

populációkon belüli beltenyésztés foka sem teljesen elhanyagolható, melyet elsősorban az egyéneknek a populáción belüli nem véletlenszerű párválasztása okozhatott.

Ha a populációk páronként történő AMOVA-elemzése során a 15 ill. 17 autoszómás lókuszból álló profilok molekuláris variancia tartalmát együtt vizsgáltuk, nem kaptunk jelentős különbséget a budapesti referencia/debreceni roma párban, valamint a budapesti referencia székely és csángó populációkkal alkotott populáció-párosai és a két erdélyi mintacsoport között. A többi párosításban azonban a molekuláris varianciaanalízis szignifikáns különbséget jelzett.

A budapesti referencia populáció egy osztrák, egy szlovén, egy román, egy szerb, egy olasz és két spanyol populációs mintával való, lókuszonkénti összehasonlításában a G-statisztika teszt alapján csak kevés eltérést tapasztaltunk. Budapesti referencia populációnk allélgyakoriság-eloszlása tehát nagyfokú hasonlóságot mutat más európai népcsoportokkal. Ezek az eredmények egybevágnak azzal a tudományos nézettel, amely szerint a kevert magyar népesség az indoeurópai népekkel közeli genetikai rokonságban áll.

4.8.2. Az eddig analizált négy magyar populáció (BuCa, BaRo, Székely, Csángó) mtDNS szekvencia-adatainak együttes AMOVA analízise alapján a szekvenciák közötti összes variabilitás 94,8%-a köszönhető a populációkon belüli különbözőségeknek és a teljes variancia 5,2%-a vezethető vissza az interpopulációs eltérésekre. Ez az eredmény rendkívüli módon eltér azoknak a hasonló kalkulációknak az adataitól, melyeket több nyugat-európai populációban végeztek. A magyar populációk között megfigyelt nagyarányú variancia szignifikáns különbséget jelez a vizsgált népessegek genetikai struktúrája között.

A populáció-párok között kalkulált F_{ST} értékek alapján a baranya megyei roma populáció mutatta a legnagyobb genetikai távolságot az összes többi magyar népcsoporttal szemben. A másik mintanépesség, mely hasonló mértékű szignifikáns különbséget mutatott a többi magyar populációhoz viszonyítva, a gyimesi csángó volt. A vizsgálatok alapján a székely népesség genetikai struktúrája nagyon hasonlít a budapesti referencia populációhoz, ami viszont az átlagos nyugat-eurázsiai népességgel mutat teljesen megegyező jelleget.

Annak ellenére, hogy a baranya megyei roma mtDNS adatok haplocsoport szerkezete lényegileg megegyezik az európai roma népességekben megfigyeltekkel, alapvető haplotípus gyakoriságbeli különbségek voltak kimutathatók a magyar és az európai roma populációk között. Ezt a genetikai különbséget megerősítette a baranya megyei és a többi európai roma populáció között páronként végzett F-statisztika rendkívül magas értékei. Az összevont roma populációs mintában megfigyelt összes variancia 5,15%-a volt tulajdonítható a roma

populációk közötti különbségeknek az AMOVA teszt alapján. A roma populációk haplocsoport reprezentáció szinten megfigyelhető hasonlósága ellenére megállapítható, hogy léteznek olyan faktorok, melyek korlátozzák az anyai génvonalak áramlását az egyes roma közösségek között.

A székely és csángó populáció mtDNS szekvencia adatait egy-egy görög és macedón populációs adatbázissal is összevetettük. A populáció-párok között kalkulált standardizált géngyakoriságok variancia tartalmának (F_{ST}) összehasonlítása a csángó populációs minta mtDNS struktúrájának kifejezett különbségét mutatta az összes többi populációval szemben. Ezzel szemben a székely populáció mitokondriális DNS szerkezete jól illeszkedik a nyugat-eurázsiai típusokba.

Hálózat-analízis alapján a gyimesi csángó eredetű haplotípusok nagyobb arányú beáramlására lehet következtetni a csíkszeredai székelyek irányába, mint ellenkezőleg, ami aszimmetrikus migrációs rátát tételez fel a két népesség között. Ezt a hipotézist megerősítette a migrációs ráta statisztikai becslése is, mely szerint a csángó bevándorlás a székely populációba nagyobb arányú, mint fordított irányba.

A budapesti referencia populáció mtDNS szintű genetikai struktúrája lényegében megegyezik a vegyes európai őslakos népesség szerkezetével, ami alapján kijelenthető, hogy a budapesti kevert populációs mintacsoport szekvencia adatai beilleszthetők az igazságügyi célú mitokondriális DNS adatbázisokba és együttesen alkalmazhatók az igazságügyi célú összehasonlítások során.

Populációs felmérésünk eredményeit összefoglalva megállapítható, hogy a magyar népesség genetikai szubstrukturáltsági foka lényeges hatással lehet a DNS-vizsgálat bizonyító erejének statisztikai kiértékelésére. Mivel hasonló adatokat mindeddig csak eltérő történelmi-genetikai eredetű etnikai csoportok vagy geográfiaileg-genetikailag izolált populációk között mértek, a magyar populációban tapasztalt nagyfokú heterogenitás a vizsgált magyar szubpopulációk diverz genetikai történelmével és/vagy izolációjukkal magyarázható. Eredményeink azt mutatják, hogy az alpopulációs teória alapján a magyarországi népcsoportokban élő személyek genetikai korrelációját figyelembe kell venni a DNS-profil egyezési valószínűség kalkulációja során annak érdekében, hogy a törvényszéki DNS-vizsgálatok bizonyító erejének interpretálása tévedéstől mentes legyen.

5. KÖSZÖNETNYILVÁNÍTÁS

Habár számomra e doktori munka elkészítése során a legnagyobb nehézséget annak végső fázisa, azaz a dolgozat elkészítése jelentette, nem szabad elfelejteni, hogy az értekezésemben bemutatott eredmények csapatmunkának köszönhetőek, és a dolgozat nem születhetett volna meg mások segítőkészsége, közreműködése nélkül. Mindezekért sokaknak tartozom köszönettel még akkor is, ha nevük említése figyelmetlenségem folytán elmarad.

Köszönöm témavezetőmnek, dr. Lontainé dr. Santora Zsófiának, a Bűnügyi Szakértői és Kutatóintézet igazgatójának, hogy minden kétségeim közepette támogatta azt az elképzelésem, hogy az intézet keretei között végzett kutatómunkából ez a dolgozat megszülethessen, és mindehhez biztosította a vizsgálatok feltételeit, felkarolta kezdeményezéseimet.

Köszönöm dr. Orosz Lászlónak, az ELTE TTK Genetika Tanszék vezetőjének, hogy mindvégig támogatta elképzeléseimet a kutatómunka során, és szakmai kapcsolatot jelentett az intézeti szakértői munka és az egyetem között.

Köszönöm barátaimnak és kollégáimnak, Pádár Zsoltnak, Füredi Sándornak, Woller Jánosnak és Horváth Zoltánnak az intézeti kutatómunkám során kapott rengeteg segítséget, és hogy idejüket nem kímélve mindig készséggel álltak rendelkezésemre.

Köszönettel tartozom a Bűnügyi Szakértői és Kutatóintézet Hemogenetikai Osztályának minden egykori és jelenlegi dolgozójának, akik a vizsgálatok kivitelezésében és az eredmények kiértékelésében jelentettek nélkülözhetetlen segítséget.

Köszönet illeti az Innsbrucki Egyetem Igazságügyi Orvostani Intézetét, névszerint Walther Parsont és Anita Brandstättert, valamint az Amerikai Légierő DNS Laboratóriumát (AFDIL), névszerint Jodi Irwint és Tom Parsonst, a mitokondriális DNS szekvenálások kapcsán nyújtott sok-sok szakmai segítségért és képzésért.

Köszönöm családomnak, hogy mindvégig kitartottak mellettem, türelmük és biztatásuk nélkül ez a dolgozat nem készült volna el.

Köszönet illeti mindazokat a személyeket és intézményeket, akik a vizsgálatok rendelkezésre bocsátották a populációs mintákat.

Végül szeretném megköszönni mindazon személyeknek az együttműködését, akik biológiai mintájukkal hozzájárultak az igazságügyi genetika magyarországi előrelépéséért. Nélkülük ezek az eredmények nem születhettek volna meg.

6. IRODALOMJEGYZÉK

- Achilli A**, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zeviani M, Briem E, Carelli V, Moral P, Dugoujon JM, Roostalu U, Loogväli EL, Kivisild T, Bandelt HJ, Richards M, Villems R, Santachiara-Benerecetti AS, Semino O, Torroni A **(2004)** The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. *Am J Hum Genet* 75:910–918
- Achilli A**, Rengo C, Battaglia V, Pala M, Olivieri A, Fornarino S, Magri C, Scozzari R, Babudri N, Santachiara-Benerecetti AS, Bandelt HJ, Semino O, Torroni A **(2005)** Saami and berbers – an unexpected mitochondrial DNA link. *Am J Hum Genet* 76:883–886
- Akbasak BS**, Budowle B, Reeder DJ, Redman J, Kline MC **(2001)** Turkish population data with the CODIS multiplex short tandem repeat loci. *Forensic Sci Int* 123:227-229
- Allard MW**, Miller K, Wilson M, Monson K, Budowle B **(2002)** Characterization of the Caucasian haplogroups present in the SWGDAM forensic mtDNA dataset for 1771 human control region sequences. *J Forensic Sci* 47:1215-1223
- Altschul SF**, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ **(1997)** Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Anderson S**, Bankier AT, Barrel BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG **(1981)** Sequence and organization of the human mitochondrial genome. *Nature* 290:457-65
- Andrews RM**, Kubacka I, Chinnery PE, Lightowlers RN, Turnbull DM, Howell N **(1999)** Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23:147
- Anghel A**, Marian C, Pitulescu M, Daba A, Sirbu IO, Rusu V, Budowle B **(2003)** Population genetic study of eight short tandem repeat loci CSF1PO, TPOX, TH01, F13A01, FESFPS, vWA, F13B and LPL in the Western Romanian population. *Forensic Sci Int* 131:218-219
- Applied Biosystems (1997)** AmpF/STR® Profiler™ PCR Amplification Kit User's Manual. Foster City, CA, USA
- Applied Biosystems (2001)** AmpF/STR® Identifiler™ PCR Amplification Kit User's Manual. Foster City, CA, USA

- Applied Biosystems (2002)** BigDye® Terminator v1.1 Cycle Sequencing Kit Protocol. Foster City, CA, USA
- Applied Biosystems (2003)** BigDye® Terminator v3.1 Cycle Sequencing Kit Protocol. Foster City, CA, USA
- Arcot SS, Wang Z, Weber JL, Deininger PL, Batzer MA (1995)** Alu repeats: a source for the genesis of primate microsatellites. *Genomics* 29:136-144
- Arnheim N, Cortopassi G (1992)** Deleterious mitochondrial DNA mutations accumulate in aging human tissues. *Mutat Res* 257:157
- Balazs I, Purrello M, Rubinstein P, Alhadeff B, Siniscalco M (1982)** Highly polymorphic DNA site *D14S1* maps to the region of Burkitt lymphoma translocation and is closely linked to the heavy chain gamma 1 immunoglobulin locus. *PNAS* 79:7395-7399
- Balogh MK, Burger J, Bender K, Schneider PM, Alt KW (2003)** STR genotyping and mtDNA sequencing of latent fingerprint on paper. *Forensic Sci Int* 137:188-195
- Bandelt HJ, Forster P, Röhl A (1999)** Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48.
- Bandelt HJ, Lahermo P, Richards M, Macaulay V (2001)** Detecting errors in mtDNA data by phylogenetic analysis. *Int J Legal Med* 115:64-69
- Bandelt HJ, Salas A, Lutz-Bonengel S (2004)** Artificial recombination in forensic mtDNA population databases. *Int J Legal Med* 118:267-73
- Bandelt HJ, Parson W (2007)** Consistent treatment of length variants in the human mtDNA control region: a reappraisal. *Int J Legal Med* DOI 10.1007/s00414-006-0151-5
- Barbarii LE, Rolf B, Constantinescu C, Hohoff C, Calistru P, Dermengiu D (2004)** Allele frequencies of 13 short tandem repeat (STR) loci in the Romanian population. *Forensic Sci Int* 141:171-174
- Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Lincoln P, Mayr WR, Olaisen B (1997)** DNA recommendations. Further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. International Society for Forensic Haemogenetics. *Int J Legal Med* 110:175-176
- Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Holland M, Lincoln P, Mayr W, Morling N, Olaisen B, Schneider PM, Tully G, Wilson M (2000)** DNA Commission of the International Society for Forensic Genetics: Guidelines for mitochondrial DNA typing. *Int J Legal Med* 113:193-196

- Beerli P, Felsenstein J (1999)** Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics* 152:763-773
- Beerli P, Felsenstein J (2001)** Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proc Natl Acad Sci USA* 98:4563-4568
- Bendall KE, Sykes BC (1995)** Length heteroplasmy in the first hypervariable segment of the human mitochondrial DNA control region. *Am J Hum Genet* 57:248
- Bendall KE, Macaulay VA, Baker JR, Sykes BC (1996)** Heteroplasmic point mutations in the human mtDNA control region. *Am J Hum Genet* 59:1276
- Bendall KE, Macaulay VA, Sykes BC (1997)** Variable levels of heteroplasmic point mutation in individual hair roots. *Am J Hum Genet* 61:1303
- Bini C, Ceccardi S, Luiselli D, Ferri G, Pelotti S, Colalongo C, Falconi M, Pappalardo G (2003)** Different informativeness of the three hypervariable mitochondrial DNA regions in the population of Bologna (Italy). *Forensic Sci Int* 135:48–52
- Botstein D, White RL, Skolnick M, Davis RW (1980)** Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet.* 32:314-331
- Bouffard GG, Iyver LM, Idol JR, Braden VV, Cunningham AF, Weintraub LA, Mohr-Tidwell RM, Peluso DC, Fulton RS, Leckie MP, Green ED (1997)** A collection of 1814 human chromosome 7 specific STSs. *Genome Res* 7: 59–64
- Boutrand L, Egyed B, Füredi S, Moomers N, Mertens G, Vandenberghe A (2001)** Variations in primer sequences are the origin of allele drop-out at loci D13S317 and CD4. *Int J Legal Med* 114:295-297
- Brandstätter A, Parson W (2003a)** Mitochondrial DNA heteroplasmy or artefacts – a matter of the amplification strategy? *Int J Legal Med* 117:180-184
- Brandstätter A, Parsons TJ, Parson W (2003b)** Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups. *Int J Legal Med* 117:291-298
- Brandstätter A, Niederstätter H, Parson W (2004a)** Monitoring the inheritance of heteroplasmy by computer-assisted detection of mixed basecalls in the entire human mitochondrial DNA control region. *Int J Legal Med* 118:47-54

- Brandstätter A**, Peterson CT, Irwin JA, Mpoke S, Koech DK, Parson W, Parsons TJ (2004b) Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database. *Int J Legal Med* 118:294-306
- Brandstätter A**, Salas A, Niederstätter H, Gassner C, Carracedo A, Parson W (2006) Dissection of mitochondrial superhaplogroup H using coding region SNPs. *Electrophoresis* 27:2541-2550
- Brandstätter A**, Niederstätter H, Pavlic M, Grubwieser P, Parson W (2007a) Generating population data for the EMPOP database - an overview of the mtDNA sequencing and data evaluation processes considering 273 Austrian control region sequences as example. *Forensic Sci Int* 166:164-75
- Brandstätter A**, Egyed B, Zimmermann B, Duftner N, Padar Z, Parson W (2007b) Migration rates and genetic structure of two Hungarian ethnic groups in Transylvania, Romania. *Ann Hum Genet* doi: 10.1111/j.1469-1809.2007.00371.x
- Brandstätter A**, Egyed B, Zimmermann B, Tordai A, Padar Z, Parson W (2007c) Mitochondrial DNA control region variation in Ashkenazi Jews from Hungary. *Forensic Sci Int Genetics* doi: 10.1016/j.fsigen.2007.07.006
- Brinkmann B, Carracedo A** (eds) (2003) *Progress in Forensic Genetics 9*. International Congress Series 1239. Elsevier, Amsterdam New York Tokyo
- Budowle B**, Chakraborty R, Giusti AM, Eisenberg AJ, Allen RC (1991) Analysis of the VNTR locus D1S80 by the PCR followed by high-resolution PAGE. *Am J Hum Genet*. 48:137-144.
- Budowle B**, Monson KL, Chakraborty R (1996) Estimating minimum allele frequencies for DNA profile frequency estimates for PCR-based loci. *Int J Legal Med* 108:173-176
- Budowle B**, Wilson MR, DiZinno JA, Stauffer C, Fasano MA, Holland MM, Monson KL (1999) Mitochondrial DNA regions HVI and HVII population data. *Forensic Sci Int* 103:23-35
- Budowle B, Chakraborty C** (2001) Population variation at the CODIS core short tandem repeat loci in Europeans. *Leg Med* 3:29-33
- Budowle B**, Shea B, Niezgoda S, Chakraborty C (2001a) CODIS STR loci data from 41 sample populations. *J Forensic Sci* 46:453-489

- Budowle B**, Masibay A, Anderson SJ, Barna C, Biega L, Brenneke S, Brown BL, Cramer J, DeGroot GA, Douglas D, Duceman B, Eastman A, Giles R, Hamill J, Haase DJ, Janssen DW, Kupferschmid TD, Lawton T, Lemire C, Llewellyn B, Moretti T, Neves J, Palaski C, Schueler S, Sgueglia J, Sprecher C, Tomsey C, Yet D **(2001b)** STR primer concordance study. *Forensic Sci Int* 124:47-54
- Budowle B, Sprecher CJ (2001)** Concordance study on population database samples using the PowerPlex 16 kit and AmpFISTR Profiler Plus kit and AmpFISTR COfiler kit. *J Forensic Sci* 46:637-41
- Budowle B**, Allard MW, Wilson MR, Chakraborty R **(2003)** Forensics and mitochondrial DNA: applications, debates, and foundations. *Annu Rev Genom Hum Genet* 4:119-41
- Budowle B**, Polanskey D, Allard MW, Chakraborty R **(2004)** Addressing the use of phylogenetics for identification of sequences in error in the SWGDAM mitochondrial DNA database. *J Forensic Sci* 49:1256-61
- Budowle B**, Garofano P, Hellman A, Ketchum M, Kanthaswamy S, Parson W, Haeringen W, Fain S, Broad T **(2005a)** Recommendations for animal DNA forensic and identity testing. *Int J Legal Med* 119:295-302
- Budowle B**, Murch R, Chakraborty R **(2005b)** Microbial forensics: the next forensic challenge. *Int J Legal Med* 119:317-330
- Butler JM**, Shen Y, McCord B **(2003)** The development of reduced size STR amplicons as tools for analysis of degraded DNA. *J Forensic Sci* 48:1054-64
- Butler JM (2005)** Forensic DNA typing. Biology technology, and genetics of STR markers. Second edition. Elsevier Academic Press, ISBN 0-12-147952-8
- Cann RL**, Stoneking M, Wilson AC **(1987)** Mitochondrial DNA and human evolution. *Nature* 325:31-36
- Carracedo A**, Bär W, Lincoln P, Mayr W, Morling N, Olaisen B, Schneider P, Budowle B, Brinkmann B, Gill P, Holland M, Tully G, Wilson M **(2000)** DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. *Forensic Sci Int* 110:79-85
- Chaix R**, Austerlitz F, Morar B, Kalaydjieva L, Heyer E **(2004)** Vlach Roma history: what do coalescent-based methods tell us? *Eur J Hum Genet* 12:285-292
- Chakraborty R (1992)** Sample size requirements for addressing the population genetic issues of forensic use of DNA typing. *Hum Biology* 64:141-159
- Chen X**, Prossner R, Simonetti S, Sadlock J, Jagiello G, Schon EA **(1995)** Rearranged mitochondrial genomes are present in human oocytes. *Am J Hum Genet* 57:239

- Chung U**, Lee HY, Yoo J, Park MJ, Shin K (2005) Mitochondrial DNA CA dinucleotide repeats in Koreans: the presence of length heteroplasmy. *Int J Legal Med* 119:50-53
- Comas D**, Paabo S, Bertranpetit J (1995) Heteroplasmy in the control region of human mitochondrial DNA. *Genome Res* 5:89
- Comey CT**, Koons BW, Presley KW, Smerick JB, Sobieralski CA, Stanley DM, Baechtel FS (1994) DNA extraction strategies for amplified fragment length polymorphism analysis. *J Forensic Sci* 39:1254-1269
- Cotton EA**, Allsop RF, Guest JL, Frazier RR, Koumi P, Callow IP, Seager A, Sparkes RL (2000) Validation of the AMPF/STR SGM plus system for use in forensic casework. *Forensic Sci Int* 112:151-61
- Council of Europe (1993)** The use of analysis of deoxyribonucleic acid (DNA) within the framework of the criminal justice system. Recommendation No. R (92) 1. Council of Europe Press
- Csete K**, Shurenkamp M, Varga T (1996) The STR system HumVWA and HumACTB2 in a Hungarian population. *Int J Legal Med* 108:316-317
- Czeizel A**, Benkmann H-G, Goedde HW (1991) Genetics of the Hungarian Population. Springer-Verlag, Berlin, Heidelberg, NewYork
- Czeizel E (2003)** A magyarság genetikája. Galenus Kiadó, Budapest
- Deka R**, Shriver MD, Yu LM, Ferrel RE, Chakraborty R (1995) Intra- and inter-population diversity at short tandem repeat loci in diverse populations of the world. *Electrophoresis* 16:1659-1664
- De Martinville B**, Wyman AR, White R, Francke U (1982) Assignment of the first Random Restriction Fragment Length Polymorphism (RFLP) Locus (*D14S1*) to a region of human chromosome 14. *Am J Hum Genet* 34:216-226
- Derenko MV**, Grzybowski T, Malyarchuk BA, Dambueva IK, Denisova GA, Czarny J, Dorzhu CM, Kakpakov VT, Miscicka-Sliwka D, Wozniak M, Zakharov IA (2003) Diversity of mitochondrial DNA lineages in south Siberia. *Ann Hum Genet* 67:391-411
- Drobnič K**, Regent A, Budowle B (2001) STR data for the AmpFISTR SGM plus from Slovenia. *Forensic Sci Int* 115:107-109
- Economou EP**, Bergen AW, Warren AC, Antonarakis SE (1990) The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome. *Proc Natl Acad Sci USA* 87:2951-2954

- Edwards A**, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am J Hum Genet* 49:746-756
- Egyed B**, Füredi S, Angyal M, Boutrand L, Vandenberghe A, Woller J, Pádár Z (2000) Analysis of eight STR loci in two Hungarian populations. *Int J Legal Med* 113:272-275
- Egyed B**, Csikai M, Füredi S, Pádár Z (2005) Population genetic data on the STR loci D2S1338, D19S433 and SE33 in Hungary. *J Forensic Sci* 50(3):720-721
- Egyed B**, Füredi S, Angyal M, Balogh I, Kalmar L, Pádár Z (2006a) Analysis of the population heterogeneity in Hungary using fifteen forensically informative STR markers. *Forensic Sci Int* 158:244-249
- Egyed B**, Füredi S, Pádár Z (2006b) Population genetic study in two Transylvanian populations using forensically informative autosomal and Y-chromosomal STR markers. *Forensic Sci Int* 158:244-249
- Egyed B**, Brandstätter A, Irwin JA, Padar Z, Parsons TJ, 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
- Ellegren H** (2004) Microsatellites: Simple sequences with complex evolution. *Nat Rev Genet* 5(6):435-45
- Entrala C**, Lorente M, Lorente JA, Carlos Alvarez J, Moretti T, Budowle B, Villanueva E (1998) Fluorescent multiplex analysis of nine STR loci: Spanish population data. *Forensic Sci Int* 98:179–183
- Evett IW**, Lambert JA, Buckleton JS, Weir BS (1996) Statistical analysis of a large file of STR profiles of British Caucasians to support forensic casework. *Int J Legal Med* 109:173-177
- Evett IW**, **Weir BS** (1998) Interpreting DNA evidence. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts
- Excoffier L**, **Langaney A** (1989) Origin and differentiation of human mitochondrial DNA. *Am J Hum Genet* 44:73-85
- Excoffier L**, Smouse P, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479-491
- Excoffier L**, **Slatkin M** (1998) Incorporating genotypes of relatives into a test of linkage disequilibrium. *Am J Hum Genet* 62:171-180

- Findlay I**, Taylor A, Quirke P, Fraizer R, Urquhart A (1997) DNA fingerprinting from single cells. *Nature* 389:555-556
- Finnilä S**, Lehtonen MS, Majamaa K (2001) Phylogenetic network for European mtDNA. *Am J Hum Genet* 68:1475-84
- Forster P**, Calí F, Röhl A, Metspalu E, D'Anna R, Mirisola M, De Leo G, Flugy A, Salerno A, Ayala G, Kouvatsi A, Villems R, Romano V (2002) Continental and subcontinental distributions of mtDNA control region types. *Int J Legal Med* 116:99-108
- Füredi S**, Woller J, Pádár Zs (1995) Hungarian population data for the STR systems TH01 and VWA. *Int J Legal Med* 108:48-49
- Füredi S**, Budowle B, Woller J, Pádár Zs (1996) Hungarian population data on six STR loci - HumVWFA31, HumTH01, HumCSF1PO, HumFES/FPS, HumTPOX, and HumHPRTB - derived using multiplex PCR amplification and manual typing. *Int J Legal Med* 109:100-101
- Füredi S**, Woller J, Pádár Zs (1997a) A population study of the STR loci HumLPL, HumF13B and HumF13A01 in Hungary. *Int J Legal Med* 110:107-108
- Füredi S**, Angyal M, Kozma Zs, Sétáló J, Woller J, Pádár Zs (1997b) Semi-automatic DNA profiling in a Hungarian Romany population using the STR loci HumVWFA31, HumTH01, HumTPOX and HumCSF1PO. *Int J Legal Med* 110:184-187
- Füredi S**, Kozma Zs, Woller J, Pádár Zs, Angyal M, Bajnóczky I, Nishi K (1998) Population genetic data on four STR loci in a Hungarian Romany population. *Int J Leg Med* 112:72-74
- Füredi S**, Woller J, Pádár Zs, Angyal M (1999) Y-STR haplotyping in two Hungarian populations. *Int J Legal Med* 113:38-42
- Füredi S** (2003) Humán polimorf mikroszatellita (short tandem repeat) lókuszok igazságügyi genetikai vizsgálata magyar populációkban. Doktori értekezés, ELTE TTK
- Garbolino P**, Taroni F (2002) Evaluation of scientific evidence using Bayesian networks. *Forensic Sci Int* 125:149-55
- Garofano L**, Pizzamiglio M, Vecchio C, Lago G, Floris T, D'Errico G, Brembilla G, Romano A, Budowle B (1998) Italian population data on thirteen short tandem repeat loci: HUMTH01, D21S11, D18S51, HUMVWA31, HUMFIBRA, D8S1179, HUMTPOX, HUMCSF1PO, D16S539, D7S820, D13S317, D5S818, D3S1358. *Forensic Sci Int* 97:53-60
- Giles RE**, Blanc H, Cann HM, Wallace DC (1980) Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci* 77:6715-6719

- Gill P, Jeffreys AJ, Werrett DJ (1985)** Forensic application of DNA “fingerprints”. *Nature* 318:577-79
- Gill P, Werrett DJ (1987)** Exclusion of a man charged with murder by DNA fingerprinting. *Forensic Sci Int* 35:145-48
- Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully G, Evett I, Hagelberg E, Sullivan K (1994)** Identification of the remains of the Romanov family by DNA analysis. *Nat Genet* 6:130
- Gill P, Brinkmann B, d’Aloja E, Andersen J, Bär W, Carracedo A, Dupuy B, Eriksen B, Jangblad M, Johnsson V, Kloosterman AD, Lincoln P, Morling N, Rand S, Sabatier M, Scheithauer R, Schneider P, Vide MC (1997)** Considerations from the European DNA profiling group (EDNAP) concerning STR nomenclature. *Forensic Sci Int* 87:185-192
- Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J (2000)** An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. *Forensic Sci Int* 112 :17-40
- Gill P, Brenner C, Brinkmann B, Budowle B, Carracedo A, Jobling MA, de Knijff P, Kayser M, Krawczak M, Mayr WR, Morling N, Olaisen B, Pascali V, Prinz M, Roewer L, Schneider PM, Sajantila A, Tyler-Smith C (2001)** DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs. *Int J Legal Med* 114:305-309
- Gray MW (1992)** The endosymbiont hypothesis revisited. *Int Rev Cytol* 141:233.
- Gresham D, Morar B, Underhill PA, Passarino G, Lin AA, Wise C, Angelicheva D, Calafell F, Oefner PJ, Shen P, Tournev I, de Pablo R, Kuinskaskas V, Perez-Lezaun A, Marushakova E, Popov V, Kalaydjieva L (2001)** Origins and Divergence of the Roma (Gypsies). *Am J Hum Genet* 69:1314–1331
- Guglielmino CR, Béres J (1996)** Genetic structure in relation to the history of Hungarian ethnic groups. *Human Biology* 68:335-355
- Guglielmino CR, Silvestri A de, Béres J (2000)** Probable ancestors of Hungarian ethnic groups: an admixture analysis. *Ann. Hum Genet.* 64:145-159
- Guo SW, Thompson EA (1992)** Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48:361-372
- Gusmão L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, Morling N, Prinz M, Roewer L, Tyler-Smith C, Schneider PM (2006)** DNA Commission of the

- International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. *Forensic Sci Int* 157:187-197
- Hanekamp JS**, Thilly WG, Chaudry MA (1996) Screening for human mitochondrial DNA polymorphisms with denaturing gradient gel electrophoresis. *Hum Genet* 98:243
- Hammond HA**, Jin L, Zhong Y, Caskey CT, Chakraborty R (1994) Evaluation of 13 short tandem repeat loci for use in personal identification applications. *Am J Hum Genet* 55:175-189
- Helgason A**, Sigureth AS, Gulcher JR, Ward R, Stefansson K (2000) mtDNA and the origin of the Icelanders: deciphering signals of recent population history. *Am J Hum Genet* 66:999-1016
- Herrnstadt C**, Preston G, Howell N (2003) Errors, phantoms and otherwise, in human mtDNA sequences. *Am J Hum Genet* 72:1585-86
- Higuchi R**, von Beroldingen CH, Sensabaugh GF, Erlich HA (1988) DNA typing from single hairs. *Nature* 322:543-546
- Holland MM**, **Parsons TJ** (1999) Mitochondrial DNA sequence analysis – Validation and use for forensic casework. *Forensic Sci Rev* 11:21-50
- Holt CL**, Buoncristiani M, Wallin JM, Nguyen T, Lazaruk KD, Walsh PS (2002) TWGDAM validation of AMPF/STR™ PCR amplification kits for forensic DNA casework. *J Forensic Sci* 47:66-96
- Howell N** (1997) mtDNA Recombination: What do in vitro data mean? *Am J Hum Genet* 61:18
- Hühne J**, Pfeiffer H, Waterkamp K, Brinkmann B (1999) Mitochondrial DNA in human hair shafts – existence of intra-individual differences? *Int J Legal Med* 112:172-175
- Inman K**, **Rudin N** (2001) Principles and practice of criminalistics. The profession of forensic science. CRC Press, ISBN 0-8493-8127-4
- Interpol DNA Unit** (2003) Global DNA Database Inquiry 2002: Results and Analysis. I.C.P.O. Interpol, General Secretariat
- Irwin J**, Egyed B, Saunier J, Szamosi G, O'callaghan J, Padar Z, Parsons TJ (2006) Hungarian mtDNA population databases from Budapest and the Baranya county Roma. *Int J Legal Med* DOI 10.1007/s00414-006-0128-4
- Irwin JA**, Saunier JL, Strouss KM, Sturk KA, Diegoli TM, Just RS, Coble MD, Parson W, Parsons TJ (2007) Development and expansion of high-quality control region databases to improve forensic mtDNA evidence interpretation. *Forensic Science International: Genetics* 1:154–157

- Irwin-Ross JA**, Saunier JL, Strouss K, Diegoli T, Sturk K, O'Callaghan JE, Paintner C, Williams HJ, Watson K, Just RS, Kovatsi L, Brandstätter A, Cariolou MA, Parson W, Parsons TJ (2006) Mitochondrial control region sequences from northern Greece and Greek Cypriots. *Int J Legal Med* DOI 10.1007/s00414-007-0173-7
- Ivanov PL**, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parsons TJ (1996) Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nat Genet* 12:417
- Jeffreys AJ**, Wilson V, Thein SL (1985a) Hypervariable minisatellite regions in human DNA. *Nature* 314:67-72
- Jeffreys AJ**, Wilson V, Thein SL (1985b) Individual-specific "fingerprints" of human DNA. *Nature* 316:76-79
- Jobling MA, Gill P** (2004) Encoded evidence: DNA in forensic analysis. *Nat Rev Genet* 5:739-51
- Jobling MA**, Hurles ME, Tyler-Smith C (2004) Human evolutionary genetics – Origins, peoples and disease. Garland Publishing, ISBN 08153 41857
- Jones DA** (1972) Blood samples: probability of discrimination. *J Forensic Sci Soc* 12:355-359
- Kalaydjieva L**, Calafell F, Jobling MA, Angelicheva D, de Knijff P, Rosser ZH, Hurles ME, Underhill PE, Tourné I, Marushiakova E, Vesselin P (2001) Patterns of inter- and intra-group genetic diversity in the Vlax Roma as revealed by Y chromosome and mitochondrial DNA lineages. *Eur J Hum Genet* 9:97-104
- Kanter E**, Baird M, Shaler R, Balazs I (1986) Analysis of restriction fragment length polymorphisms in deoxyribonucleic acid (DNA) recovered from dried bloodstains. *J Forensic Sci* 31:403-408
- Keckarević D**, Savić D, Keckarević M, Stevanović M, Tarasjev A, Čuljković B, Đarmati A, Vukosavić S, Romac S (2005) Population data on 14 STR loci from population of Serbia and Montenegro (new and renewed data). *Forensic Sci Int* 151:315-316
- Kimpton CP**, Oldroyd NJ, Watson SK és mtsai (1996) Validation of highly discriminating multiplex short tandem repeat amplification systems for individual identification. *Electrophoresis* 17:1283-1293
- Kivisild T**, Villems R, Jorde LB, Bamshad M, Kumar S, Hedrick P, Dowling T, Stoneking M, Parsons TJ, Irwin JA (2000) Questioning evidence for recombination in human mitochondrial DNA. *Science* 288:1931a

- Kozma Z**, Nagai A, Woller J, Füredi S, Sétáló J (1998) Fluoresce based co-amplification and automated detection of the STR loci HUMFIBRA and HUMD21S11 in a Hungarian Caucasian population sample. *Int J Legal Med* 111:103-104
- Krenke BE**, Tereba A, Anderson SJ, Buel E, Culhane S, Finis CJ, Tomsey CS, Zachetti JM, Sprecher CJ (2002) Validation of a 16-locus fluorescent multiplex system. *J Forensic Sci* 47:773-85
- Krings M**, Stone A, Schmitz RW, Krainitzki H, Stoneking M, Paabo S (1997) Neandertal DNA sequences and the origin of modern humans. *Cell* 90:19-30
- Lander ES** (1989) DNA fingerprinting on trial. *Nature* 339:501-505
- Lang BF**, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Gray MW (1997) An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* 387:493
- Lazaruk K**, Walsh PS, Oaks F, Gilbert D, Rosenblum BB, Menchen S, Sceibler D, Wentz HM, Holt C, Wallin J (1998) Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument. *Electrophoresis* 19:86-93
- Lászik A**, Sótonyi P, Rand S, Hohoff C (2001) Frequency data for the STR locus ACTBP2 (SE33) in eight populations. *Int J Legal Med* 115:94-96
- Lewis PO**, Zaykin D (2001) Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c). Free program distributed by the authors over the internet from <http://lewis.eeb.uconn.edu/lewishome/software.html>
- Loogväli EL**, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, Metspalu E, Tambets K, Reidla M, Tolk HV, Parik J, Pennarun E, Laos S, Lunkina A, Golubenko M, Barac L, Pericic M, Balanovsky OP, Gusar V, Khusnutdinova EK, Stepanov V, Puzyrev V, Rudan P, Balanovska EV, Grechanina E, Richard C, Moisan JP, Chaventre A, Anagnou NP, Pappa KI, Michalodimitrakakis EN, Claustres M, Golge M, Mikerezi I, Usanga E, Villems R (2004) Disuniting uniformity: a pied cladistic canvas of mtDNA haplogroup H in Eurasia. *Mol Biol Evol* 21:2012-2021
- Lutz S**, Weisser H-J, Heizmann J, Pollak S (1998) Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *Int J Legal Med* 111:67-77

- Lutz S**, Wittig H, Weissner HJ, Heizmann J, Junge A, Dimo-Simonin N, Parson W, Edelmann J, Anslinger K, Jung S, Augustin C (2000) Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions? *Forensic Sci Int* 113:97-101
- Lygo JE**, Johnson PE, Holdaway DJ, Woodroffe S, Whittaker JP, Clayton TM, Kimpton CP, Gill P (1994) The validation of short tandem repeat loci for use in forensic casework. *Int J Legal Med* 107:77-89
- Malyarchuk BA**, Grzybowski T, Derenko MV, Czarny J, Wozniak M, Miscicka-Sliwka D (2002) Mitochondrial DNA variability in Poles and Russians. *Ann Hum Genet* 66:261-283
- Malyarchuk BA**, Grzybowski T, Derenko MV, Czarny J, Drobnic K, Miscicka-Sliwka D (2003) Mitochondrial DNA variability in Bosnians and Slovenians. *Ann Hum Genet* 67:412-425
- Marchington DR**, Hartshorne GM, Barlow D, Poulton J (1997) Homopolymeric tract heteroplasmy in mtDNA from tissues and single oocytes: support for a genetic bottleneck. *Am J Hum Genet* 60:408
- Martin P**, Alonso A, Budowle B, Albarrán C, García O, Sancho M (1995) Spanish population data on 7 tetrameric short tandem repeat loci. *Int J Legal Med* 108:145-149
- Melton T** (2004) Mitochondrial DNA heteroplasmy. *Forensic Sci Rev* 16:1-20
- Michalakis Y, Excoffier L** (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics* 142:1061-1064
- Micka KA**, Sprecher CJ, Lins AM, Theisen Comey C, Koons BW, Crouse C, Endean D, Pirelli K, Lee SB, Duda N, Ma M, Schumm JW (1996) Validation of multiplex polymorphic STR amplification sets developed for personal identification applications. *J Forensic Sci* 41:582-590
- Micka KA**, Amriott EA, Hockenberry TL, Sprecher CJ, Lins AM, Rabbach DR, Taylor JA, Bacher JW, Glidewell DE, Gibson SD, Crouse CA, Schumm JW (1999) TWGDAM validation of a nine-locus and a four-locus fluorescent STR multiplex system. *J Forensic Sci* 44:1243-57
- Miller KW, Budowle B** (2001) A compendium of human mitochondrial DNA control region: development of an international standard forensic database. *Croat Med J* 42:315-27
- Mills KA**, Even D, Murray JC (1992) Tetranucleotide repeat polymorphism at the human alpha fibrinogen locus (FGA). *Hum Mol Genet* 1:779

- Monson KL**, Miller KWP, Wilson MR, DiZinno JA, Budowle B (2002) The mtDNA population database: an integrated software and database resource for forensic comparison. *For Sci Comm* 4:2
Available at <http://www.fbi.gov/hq/lab/fsc/backissu/april2002/miller1.htm>
- Montagna CP**, (1996) The recovery of seminal components and DNA from the vagina of a homicide victim 34 days postmortem. *J Forensic Sci.* 41(4):700-02
- Morling N**, Allen R, Carracedo A, Geada H, Guidet F, Hallenberg C, Martin W, Mayr WR, Olaisen B, Pascali V, Schneider PM (2003) Recommendations on genetic investigations in paternity cases. *Int J Legal Med* 117:51-61
- Mullis K**, Faloona F, Scharf S, Saiki R, Horn G, Erlich H (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harb Symp Quant Biol* 1:263-73
- Nadir E**, Margalit H, Gallily T, Ben-Sasson SA (1996) Microsatellite spreading in the human genome: Evolutionary mechanisms and structural implications. *Proc Natl Acad Sci USA* 93(13):6470-75
- National Commission of the Future of DNA Evidence** (2001) The future of Forensic DNA testing : Prediction of the Research and Development Working Group. Washington
- Nei M** (1972) Genetic distance between populations. *Am Nat* 106:283-291
- Nei M** (1973) Analysis of gene diversity in subdivided populations. *Proc Natl Acad Sci USA* 70:3321-3323
- Nei M**, Roychoudhury AK (1974) Sampling variances of heterozygosity and genetic distance. *Genetics* 76:379-390
- NRC National Research Council** (1996) The evaluation of forensic DNA evidence. National Academy Press Washington, D.C.
- NRC National Research Council** (1997) The evaluation of forensic DNA evidence. *Proc Natl Acad Sci USA* 94:5498–5500
- Nussbaumer C**, Hanslik S, Fichtinger M, Bauer G (2001) STR data for the AmpFISTR SGM plus from a regional population of Austria. *Forensic Sci Int* 122:181-183
- Ohno Y**, Sebetan IM, Akaishi S (1982) A simple method for calculating the probability of excluding paternity with any number of codominant alleles. *Forensic Sci Int* 19:93-98
- Olaisen B**, Bar W, Brinkmann B, Budowle B, Carracedo A, Gill P, Lincoln P, Mayr WR, Rand S (1998) DNA recommendations 1997 of the International Society for Forensic Genetics. *Vox Sang* 74:61-63
- Oorschot RAH**, Jones MK (1997) DNA fingerprints from fingerprints. *Nature* 387:767

- Ota T, Kimura M (1973)** A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in finite population. *Genet Res* 22:201-204
- Pádár Z, Woller J (1993)** The first experiences of the application of DNA polymorphism in criminal cases in Hungary. *Proceedings of 1st International Symposium on Forensic Sciences, Liptovsky Jan (Slovak Republic)*
- Pádár Zs (2006)** Kutya eredetű anyagmaradványok igazságügyi genetikai vizsgálata. Doktori értekezés, SZIE ÁTK
- Palanichamy MG, Sun C, Agrawal S, Bandelt H-J, Kong Q-P, Khan F, Wang C-Y, Chaudhuri TK, Palla V, Zhang Y-P (2004)** Phylogeny of mitochondrial DNA macrohaplogroup N in India, based on complete sequencing: implications for the peopling of South Asia. *Am J Hum Genet* 75:966-978
- Paredes M, Crespillo M, Luque JA, Valverde JL (2003)** STR frequencies for the PowerPlex 16 System Kit in a population from Northeast Spain. *Forensic Sci Int* 135:75-8
- Parson W, Pegoraro K, Niederstatter H, Föger M, Steinlechner M (2000)** Species identification by means of the cytochrome b gene. *Int J Legal Med* 114:23-28
- Parson W, Dür A (2007)** EMPOP – A forensic mtDNA database. *FSI Genetics* 1:88-92
- Parsons TJ, Weedn VW (1997)** Preservation and Recovery of DNA in Postmortem Specimens and Trace Samples. In W.D. Haglund and M.H. Sorg (eds.): *Forensic Taphonomy. The Postmortem Fate of Human Remains*. CRC Press, New York London Tokyo
- Parsons TJ, Muniec DS, Sullivan K, Woodyatt N, Alliston-Greiner N, Wilson MR, Berry DL, Holland KA, Weedn VW, Gill P, Holland MM (1997)** A high observed substitution rate in the human mitochondrial DNA control region. *Nat Genet* 15:363
- Parsons TJ, Coble MD (2001)** Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial genome. *Croatian Med J* 42:304-09
- Phillips C, Lareu M, Sanchez J, Brion M, Sobrino B, Morling N, Schneider P, Syndercombe Court D, Carracedo A (2004)** Selecting single nucleotide polymorphisms for forensic application. *Prog in Forensic Genet* 10:18-20 International Congress Series 1261
- Piercy R, Sullivan KM, Benson N, Gill P (1993)** The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int J Legal Med* 106:85-90
- Polymeropoulos MH, Xiao H, Rath DS, Merrill CR (1991)** Tetranucleotide repeat polymorphisms at the human tyrosine hydroxylase gene. *Nucleic Acids Res* 19:3753

- Promega Corporation (2002)** PowerPlex® 16 System Technical Manual. Madison, WI, USA
- Pulker H, Lareau MV, Phillips C, Carracedo A (2007)** Finding genes that underlie physical traits of forensic interest using genetic tools. *FSI Genetics* 1: 100-104
- Reidla M, Kivisild T, Metspalu E, Kaldma K, Tambets K, Tolk HV, Parik J, Loogväli E, Derenko M, Malyarchuk B, Bermisheva M, Zhadanov S, Pennarun E, Gubina M, Golubenko M, Damba L, Fedorova S, Gusar V, Grechanina E, Mikerezi I, Moisan JP, Chaventré A, Khusnutdinova E, Osipova L, Stepanov V, Voevoda M, Achilli A, Rengo C, Rickards O, De Stefano GF, Papiha S, Beckman L, Janicijevic B, Rudan P, Anagnou N, Michalodimitrakis E, Koziel S, Usanga E, Geberhiwot T, Herrnstadt C, Howell N, Torroni A, Villems R (2003)** Origin and diffusion of mtDNA haplogroup X. *Am J Hum Genet* 73:1178-1190
- Roewer L, Krawczak M, Willuweit S, Nagy M, Alves C, Amorim A, Anslinger K, Augustin C, Betz A, Bosch E, Caglia A, Carracedo A, Corach D, Dekairelle A-F, Dobosz T, Dupuy BM, Füredi S, Gehrig C, Gusmao L, Henke J, Hidding M, Hohoff C, Hoste B, Jobling MA, Kargel HJ, de Knijff P, Lessig R, Liebeherr E, Lorente M, Martinez-Jarreta B, Nievas P, Nowak M, Parson W, Pascali VL, Penacino G, Ploski R, Rolf B, Sala A, Schmidt U, Schmitt C, Schneider PM, Szibor R, Teifel-Greding J, Kayser M (2001)** Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. *Forensic Sci Int* 118:106-113
- Rogers AR, Harpending H (1992)** Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol* 9:552-569
- Ruitberg CM, Reeder DJ, Butler JM (2001)** STRBase: a short tandem repeat DNA database for the human identity testing community. *Nucleic Acids Res* 29:320-322
- Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC (2004)** Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* 303:223-6
- Sanchez JJ, Phillips C, Børsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N (2006)** A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis* 27:1713-1724
- Satoh M, Kuroiwa T (1991)** Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. *Exp Cell Res* 196:137-140

- Schneider S**, Roesli D, Excoffier L (2000) Arlequin ver. 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland. URL: <http://lgb.unige.ch/arlequin>
- Steighner RJ**, Tully LA, Karjala JD, Coble MD, Holland MM (1999) Comparative identity and homogeneity testing of the mtDNA HV1 region using denaturing gradient gel electrophoresis. *J Forensic Sci* 44:1186-1198
- Steinlechner M, Parson W** (2001) Automation and high through-put for a DNA database laboratory: development of a laboratory information management system. *Croat Med J* 42:252-255
- Steinlechner M**, Grubwieser P, Scheithauer R, Parson W (2002) STR loci Penta D and Penta E: Austrian Caucasian population data. *Int J Legal Med* 116:174-5
- Stoneking M**, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA (1991) Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *Am J Hum Genet* 48:370-382
- Sullivan KM**, Hopgood R, Gill P (1992) Identification of human remains by amplification and automated sequencing of mitochondrial DNA. *Int J Legal Med* 105:83-86
- Sullivan KM**, Manucci A, Kimpton CP, Gill P (1993) A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. *Biotechniques* 15:636-8, 640-1
- Sutovsky P**, Moreno RD, Ramalho-Santos J, Dominko T, Simerly C, Schatten G (1999) Ubiquitin tag for sperm mitochondria. *Nature* 402:371-2
- Sweet DJ**, Sweet CH (1995) DNA analysis of dental pulp to link incinerated remains of homicide victim to crime scene. *J Forensic Sci.* 40:310-14
- Swofford** (2002) PAUP*: phylogenetic analysis using parsimony (* and other methods), 4.0.b10. Sinauer Associates, Sunderland, MA, USA
- Thomas MG**, Eeale ME, Jones AL, Richards M, Smith A, Redhead N, Torroni A, Scoyyari R, Gratrix F, Tarekegn A, Wilson JF, Capelli C, Bradman N, Goldstein DB (2002) Founding Mothers of Jewish Communities: Geographically Separated Jewish Groups Were Independently Founded by Very Few Female Ancestors. *Am J Hum Genet* 70:1411-1420
- Torroni A**, Achilli A, Macaulay V, Richards M, Bandelt HJ (2006) Harvesting the fruit of the human mtDNA tree. *Trends Genet* 22:339-345

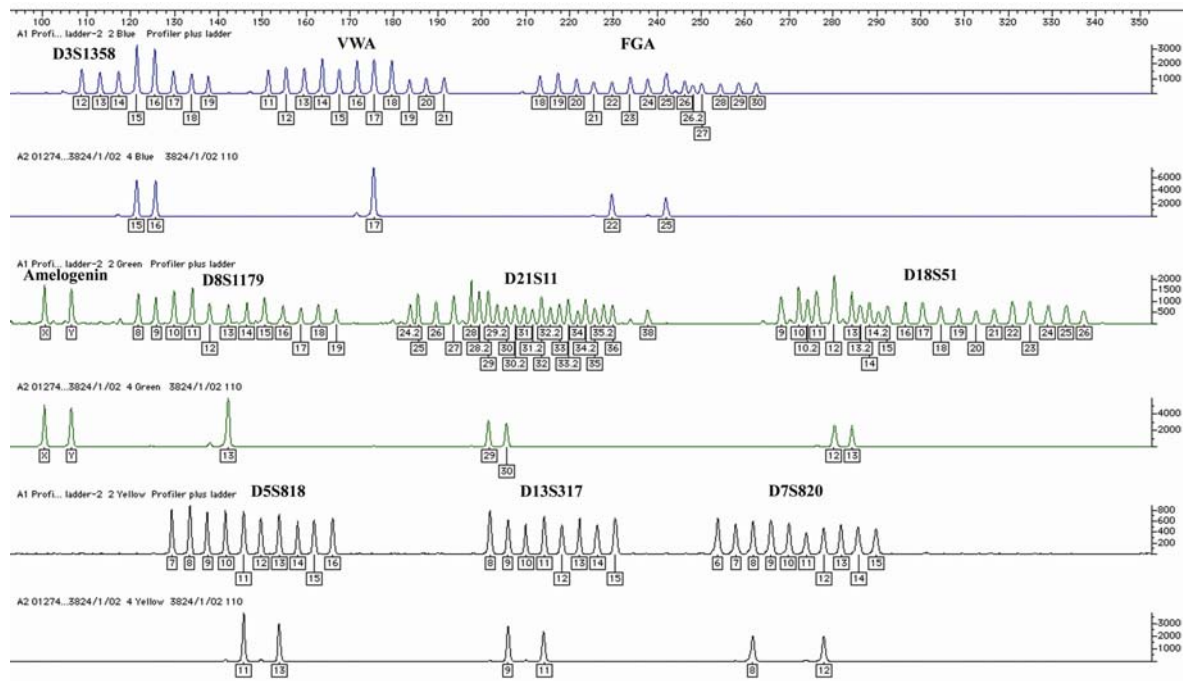
- Tully G**, Bär W, Brinkmann B, Carracedo A, Gill P, Morling N, Parson W, Schneider P (2001) Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. *Forensic Sci Int* 124:83-91
- TWGDAM (1995)** Established guidelines for a quality assurance program for DNA testing laboratories, including RFLP and PCR technologies. *Crime Lab Dig* 18:44-75
- Urquhart A**, Kimpton CP, Downes TJ, Gill P (1994) Variation in short tandem repeat sequences – a survey of twelve microsatellite loci for the use as forensic identification markers. *Int J Legal Med* 107:13-20
- Valdes AM**, Slatkin M, Freimer NB (1993) Allele frequencies at microsatellite loci: the stepwise mutational model revisited. *Genetics* 133:737-749
- Vanecek T**, Vorel F, Sip M (2004) Mitochondrial DNA D-loop hypervariable regions: Czech population data. *Int J Legal Med* 118:14-18
- Veselinović I**, Kubat M, Furač I, Škavić J, Martinović-Klarić I, Tasić M (2004) Allele frequencies of the 15 AmpFISTR Identifiler loci in the population of Vojvodina Province, Serbia and Montenegro. *Int J Legal Med* 118:184-186
- Wallace DC (1992)** Mitochondrial genetics: A paradigm for aging and degenerative diseases? *Science* 256:628
- Wallace DC**, Brown MD, Lott MT (1999) Mitochondrial DNA variation in human evolution and disease. *Gene* 238:211-30
- Walsh PS**, Fildes NJ, Reynolds R (1996) Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic Acids Res* 24:2807-12
- Watson E**, Forster P, Richards M, Bandelt HJ (1997) Mitochondrial footprints of human expansions in Africa. *Am J Hum Genet* 61:691-704
- Weber JL, May PE (1989)** Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Hum Genet* 44:388-396
- Weber JL (1990)** Human DNA polymorphisms based on length variation in simple sequence tandem repeats. *Genome analysis Vol.1*. Cold Spring Harbor Laboratory Press
- Weir BS, Cockerham CC (1984)** Estimating *F*-statistics for the analysis of population structure. *Evolution* 38:1358-1370
- Weir BS (1996)** *Genetic Data Analysis II*. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts

- Willuweit S**, Roewer L, on behalf of the International Forensic Y Chromosome User Group (2007) Y chromosome haplotype reference database (YHRD): Update. *Forensic Sci Int Genetics* 1:83-87
- Wilson MR**, Polanskey D, Butler J, DiZinno JA, Replogle J, Budowle B (1995a) Extraction, PCR amplification and sequencing of mitochondrial DNA from human hair shafts. *Biotechniques* 18:662-69
- Wilson MR**, DiZinno JA, Polanskey D, Replogle J, Budowle B (1995b) Validation of mitochondrial DNA sequencing for forensic casework analysis. *Int J Legal Med* 108:68-74
- Wilson MR**, Allard MW, Monson KL, Miller KWP, Budowle B (2002a) Recommendations for consistent treatment of length variants in the human mtDNA control region. *Forensic Sci Int* 129:35–42
- Wilson MR**, Allard MW, Monson KL, Miller KWP, Budowle B (2002b) Further discussion of the consistent treatment of length variants in the human mtDNA control region. *Forensic Sci Commun* 4:#4
- Woller J**, Furedi S, Padar Zs (1995) AMPFLP analysis of the VNTR loci D1S80 and ApoB in Hungary. *Int J Legal Med* 107:273-274
- Wright S** (1951) The genetical structure of populations. *Ann Eugen* 15:323-354
- Yandava CN**, Gastier JM, Pulido JC, Brody T, Sheffield V, Murray J, Buetow K, Duyck GM (1997) Characterization of Alu repeats that are associated with trinucleotide and tetranucleotide repeat microsatellites. *Genome Res* 7: 716-724
- Yao YG**, Bravi CM, Bandelt HJ (2004) A call for mtDNA data quality control in forensic science. *Forensic Sci Int* 141:1-6
- Zimmermann B**, Brandstätter A, Duftner N, Niederwieser D, Spiroski M, Arsov T, Parson W (2007) Mitochondrial DNA control region population data from Macedonia. *Forensic Science International: Genetics*, doi:10.1016/j.fsigen.2007.03.002
- Zupanic I**, Balazic J, Komel R (1998) Analysis of nine short tandem repeat (STR) loci in the Slovenian population. *Int J Legal Med* 111:248–250

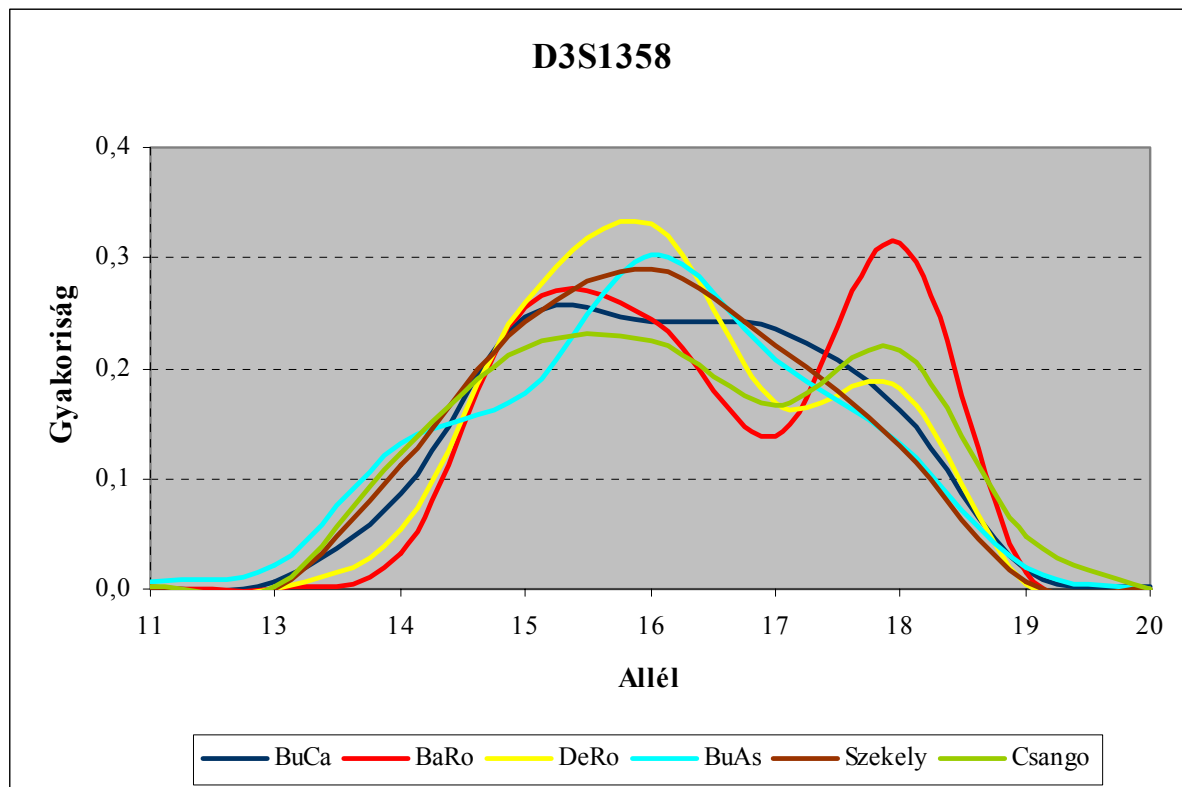
7. FÜGGELÉK

7.1. Ábrák

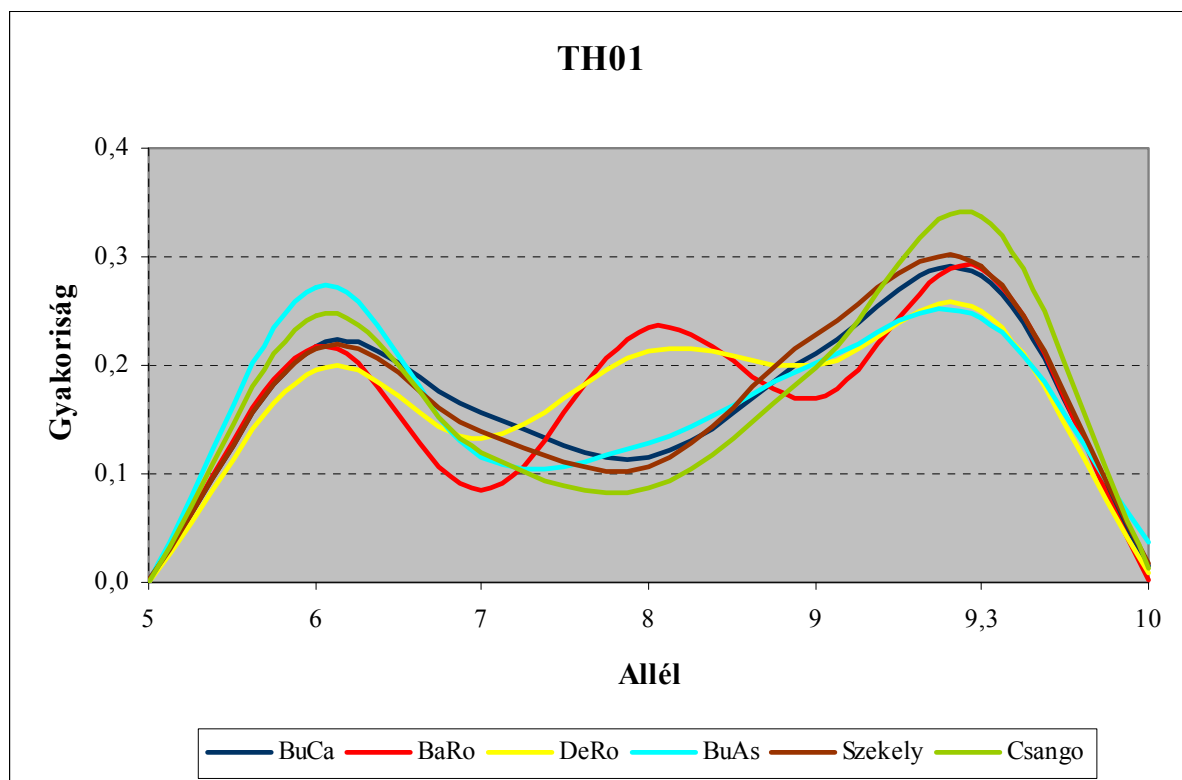
7.1.1. D3S1358-VWA-FGA-Amelogenin-D8S1179-D21S11-D18S51-D5S818-D13S317-D7S820 (Profiler Plus multiplex)



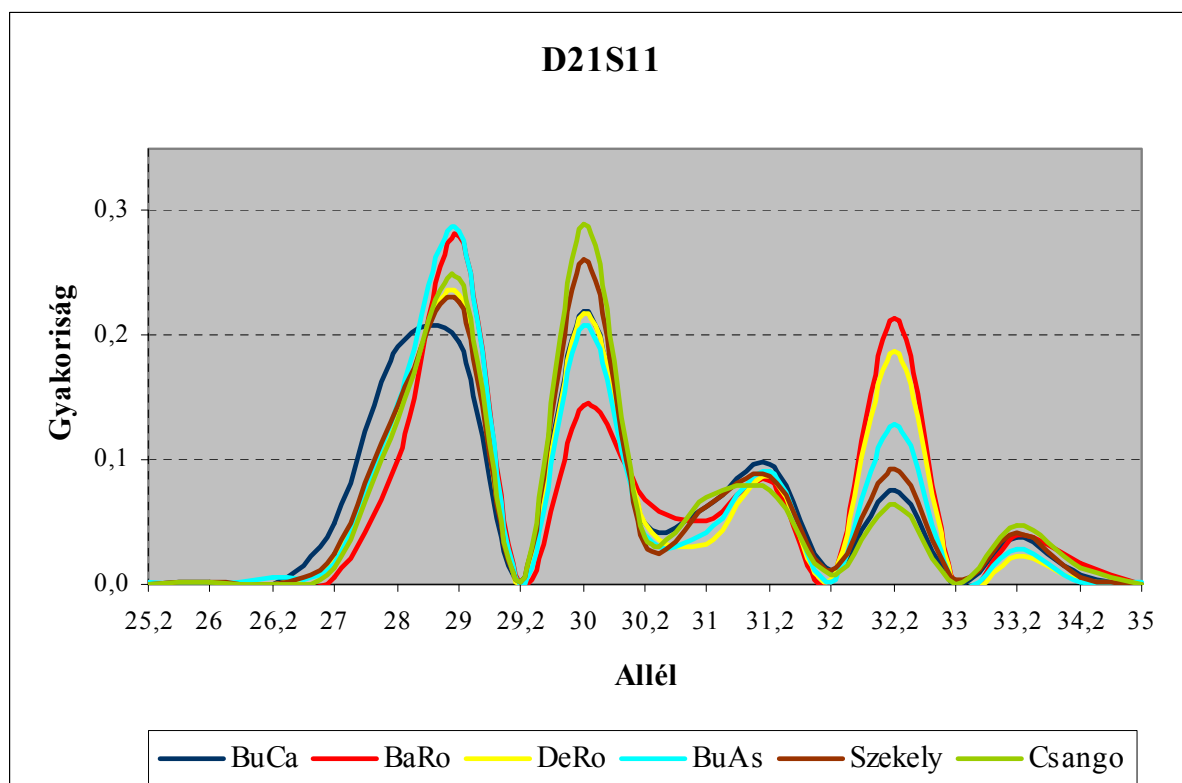
7.1.2. A D3S1358 lókuszt allélgyakoriság-eloszlása



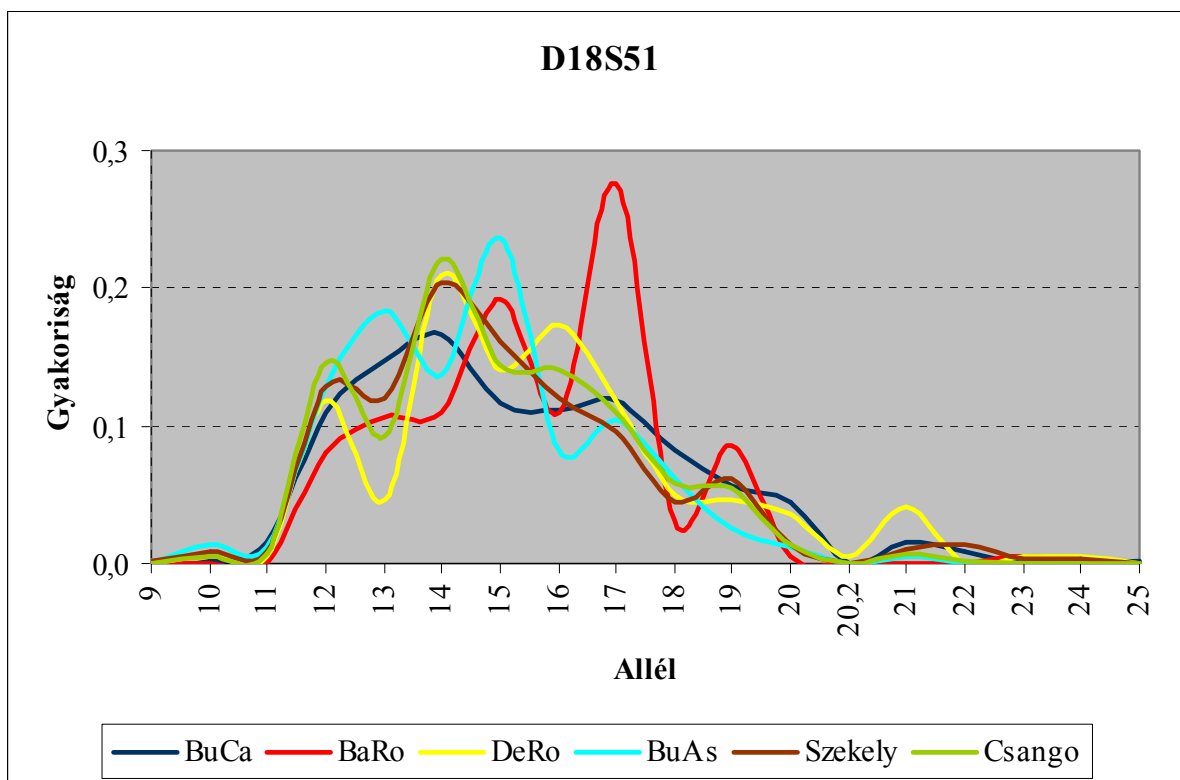
7.1.3. A TH01 lókusztól allélgyakoriság-eloszlása



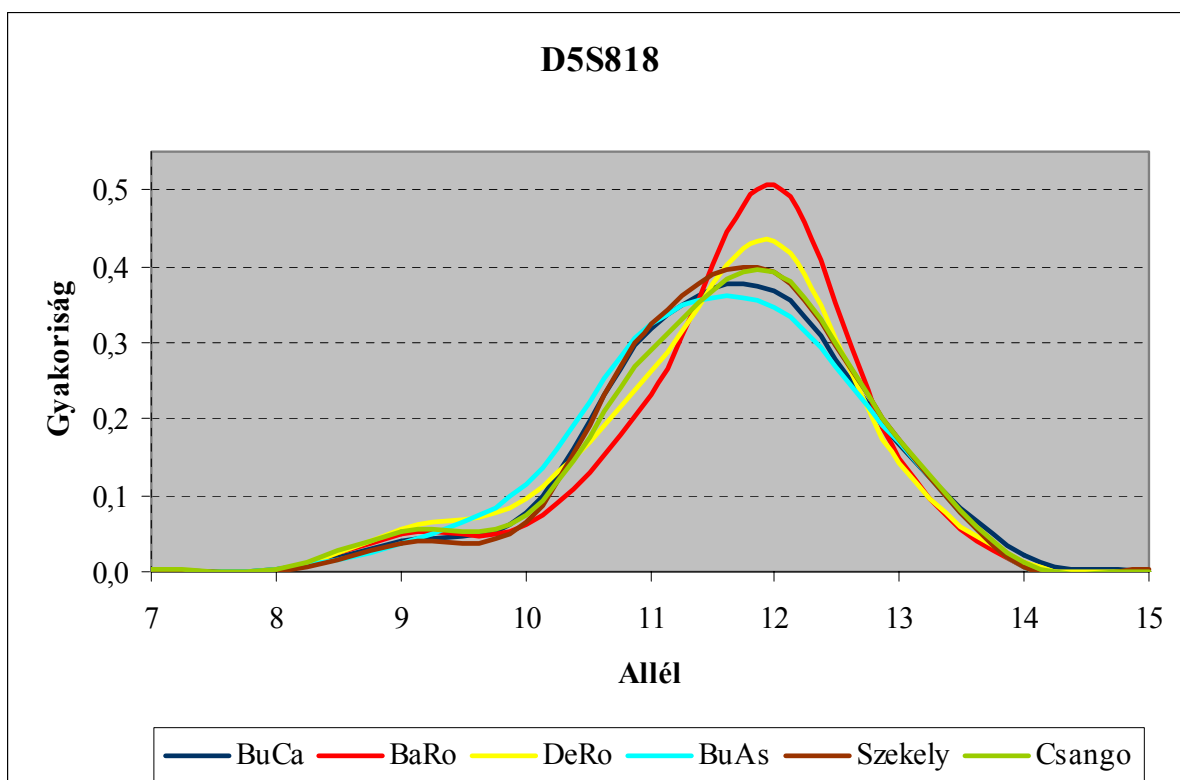
7.1.4. A D21S11 lókusztól allélgyakoriság-eloszlása



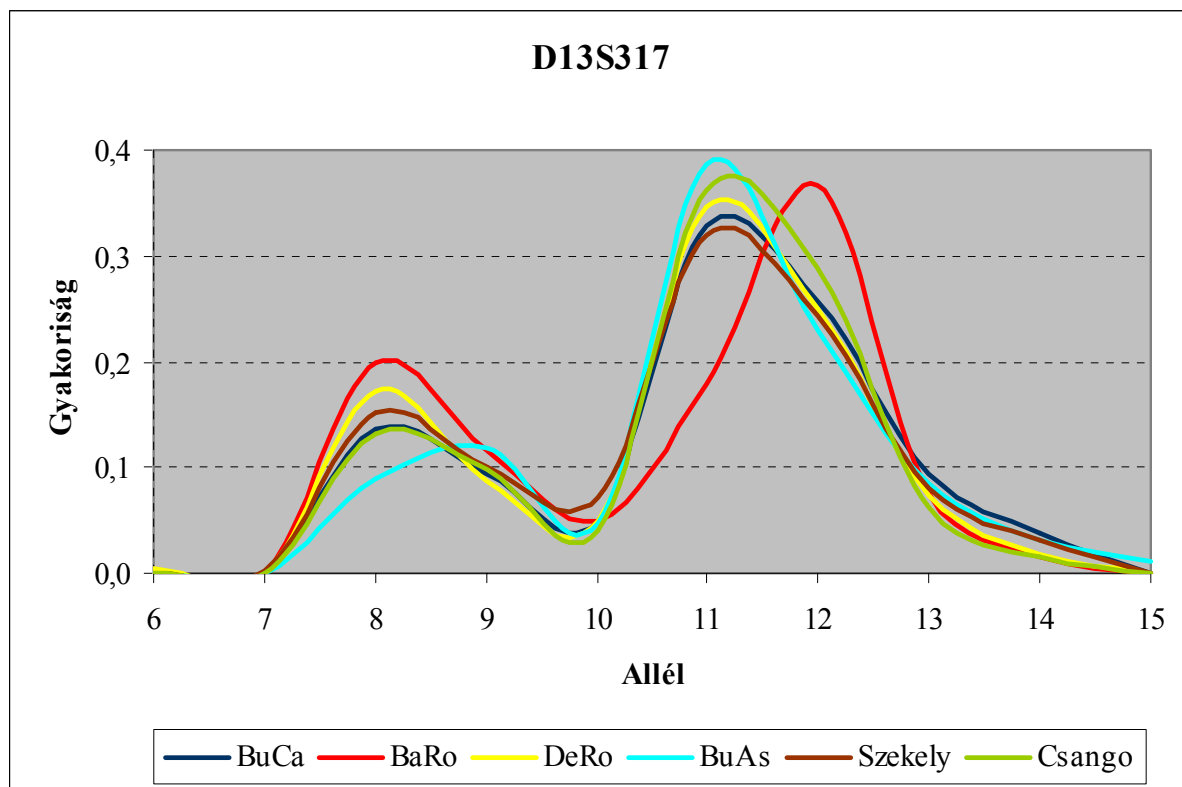
7.1.5. A D18S51 lókuszt allélgyakoriság-eloszlása



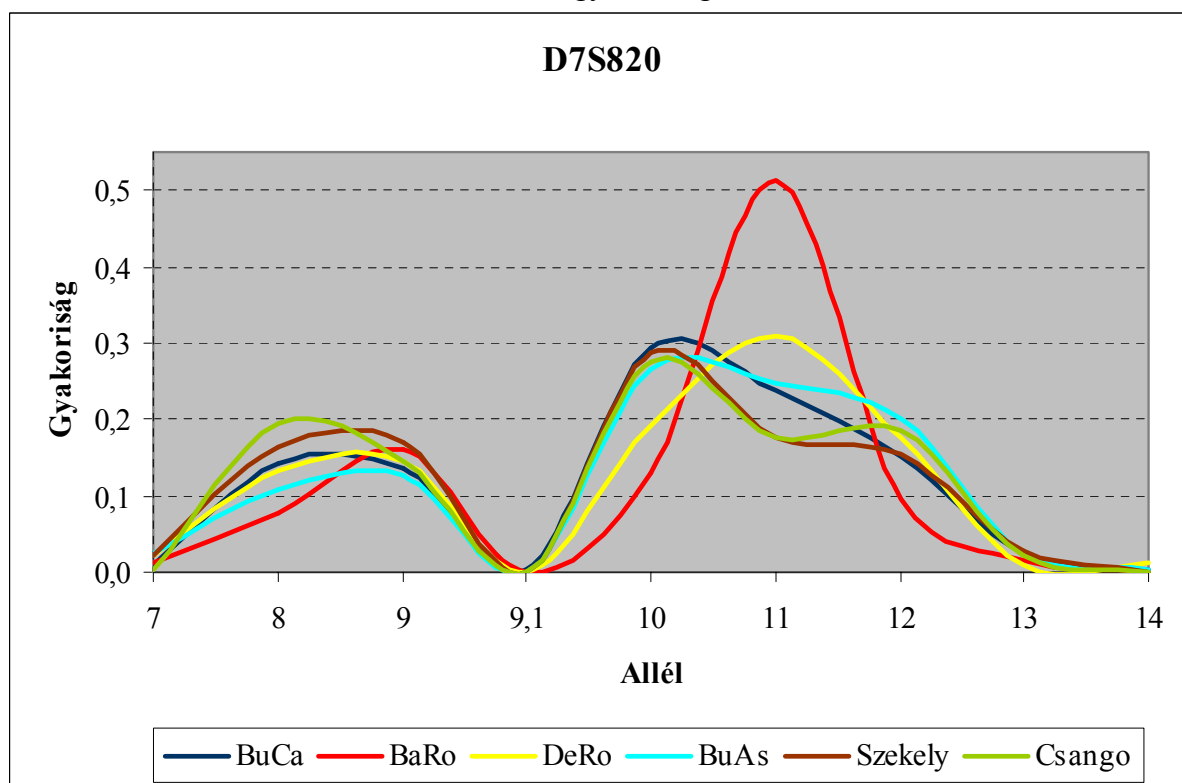
7.1.6. A D5S818 lókuszt allélgyakoriság-eloszlása



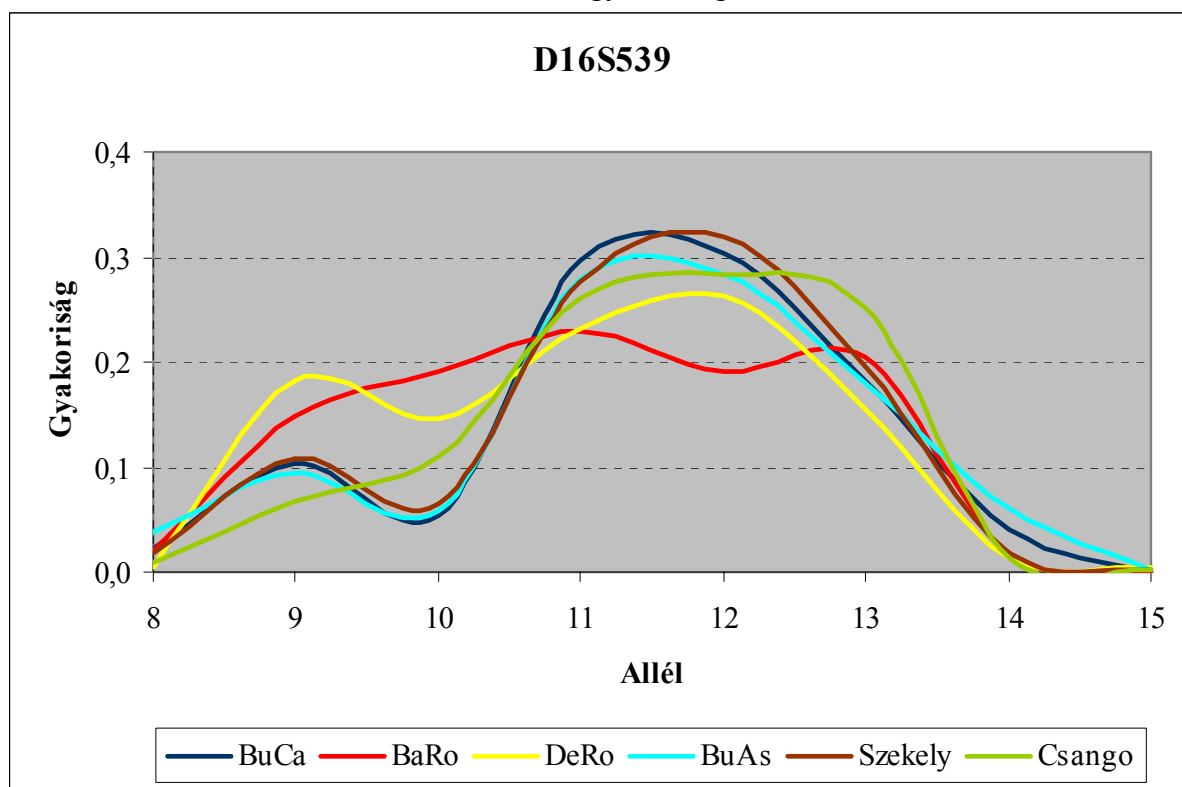
7.1.7. A D13S317 lókuszt allélgyakoriság-eloszlása



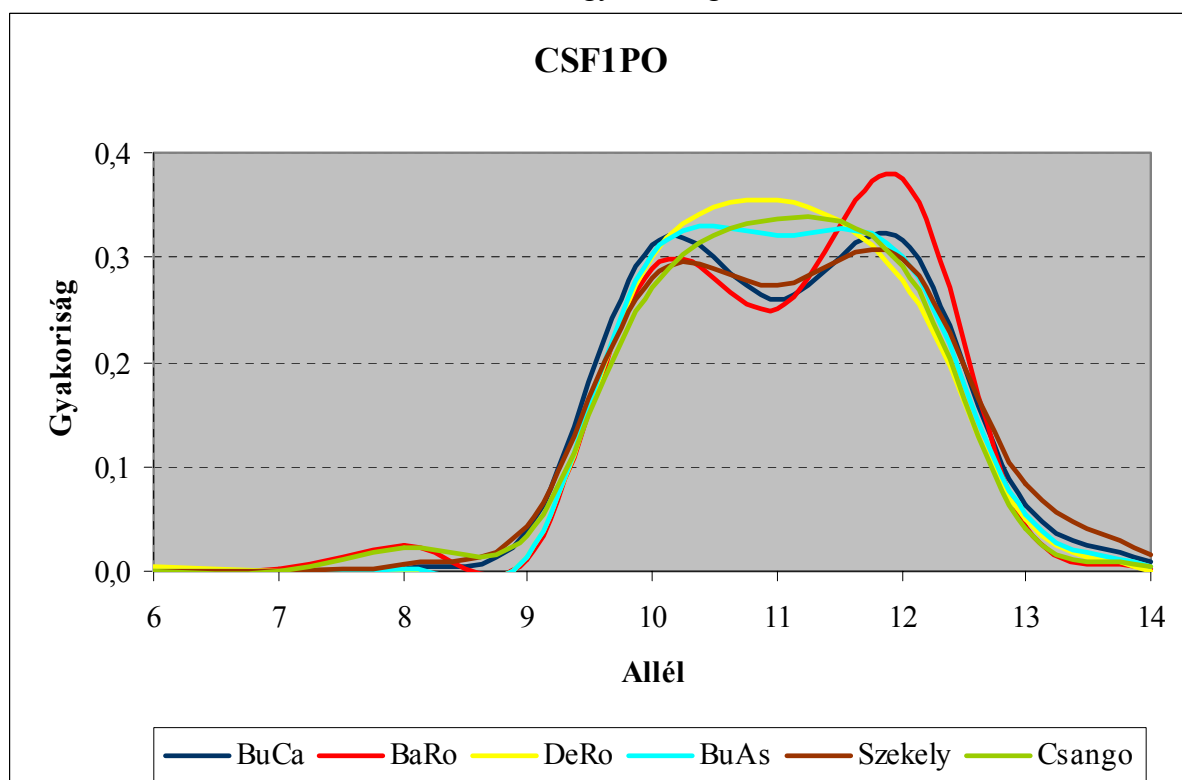
7.1.8. A D7S820 lókuszt allélgyakoriság-eloszlása



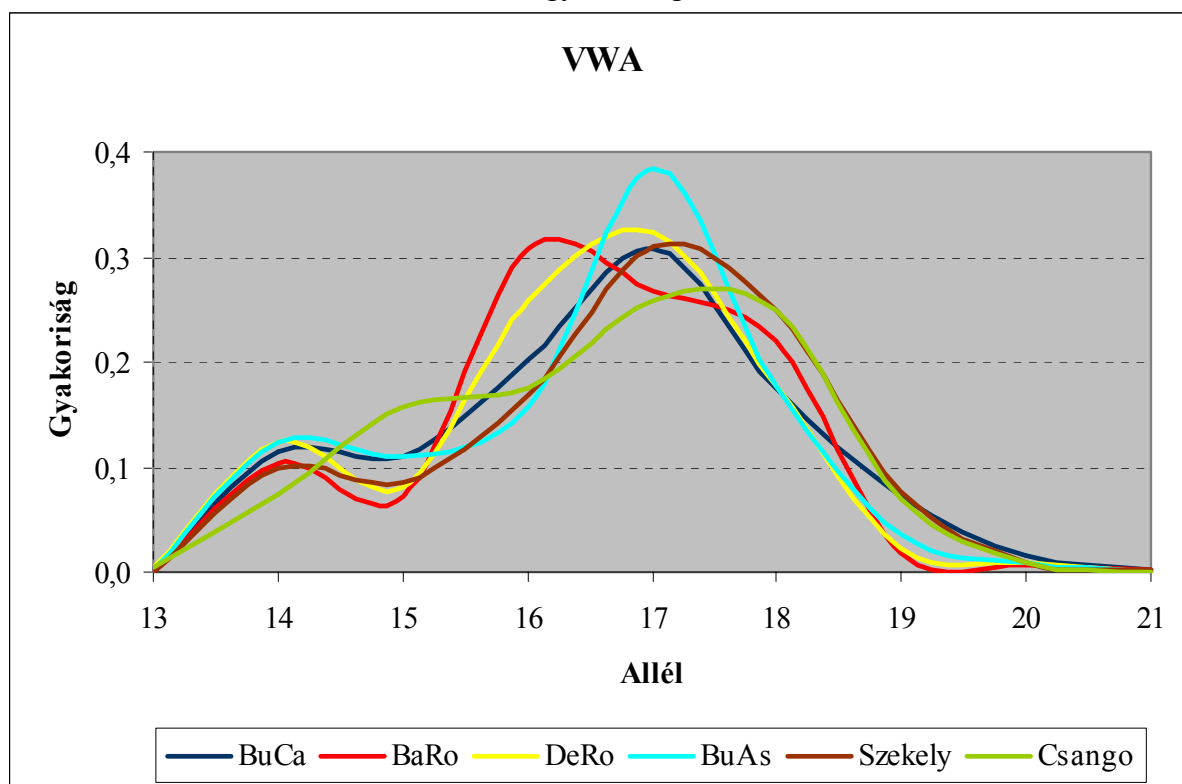
7.1.9. A D16S539 lókuszt allélgyakoriság-eloszlása



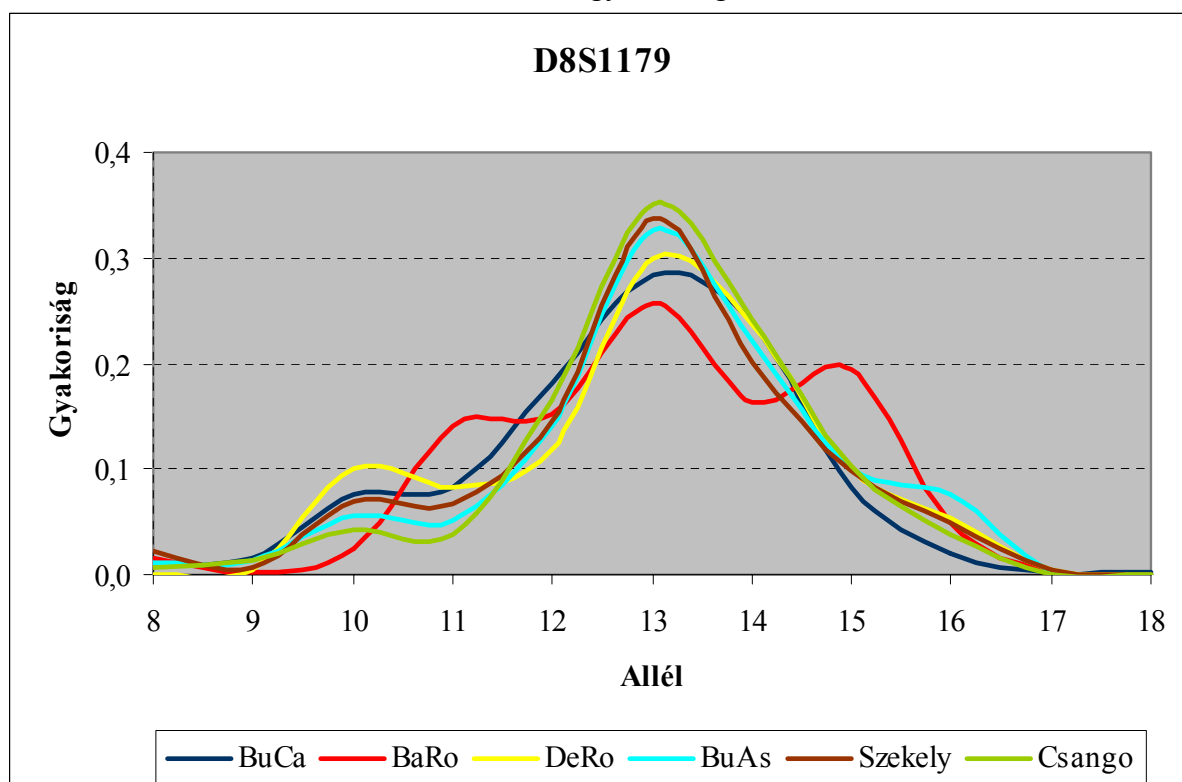
7.1.10. A CSF1PO lókuszt allélgyakoriság-eloszlása



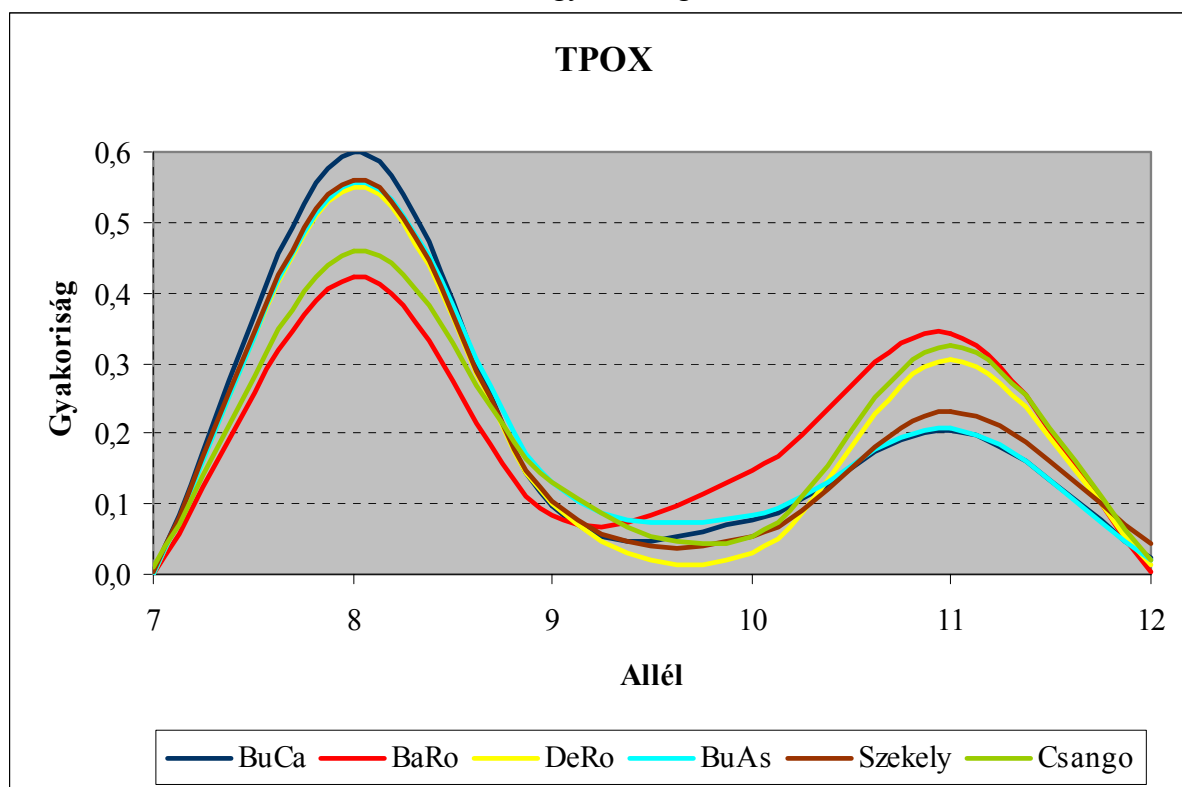
7.1.11. A VWA lókuszt allélgyakoriság-eloszlása



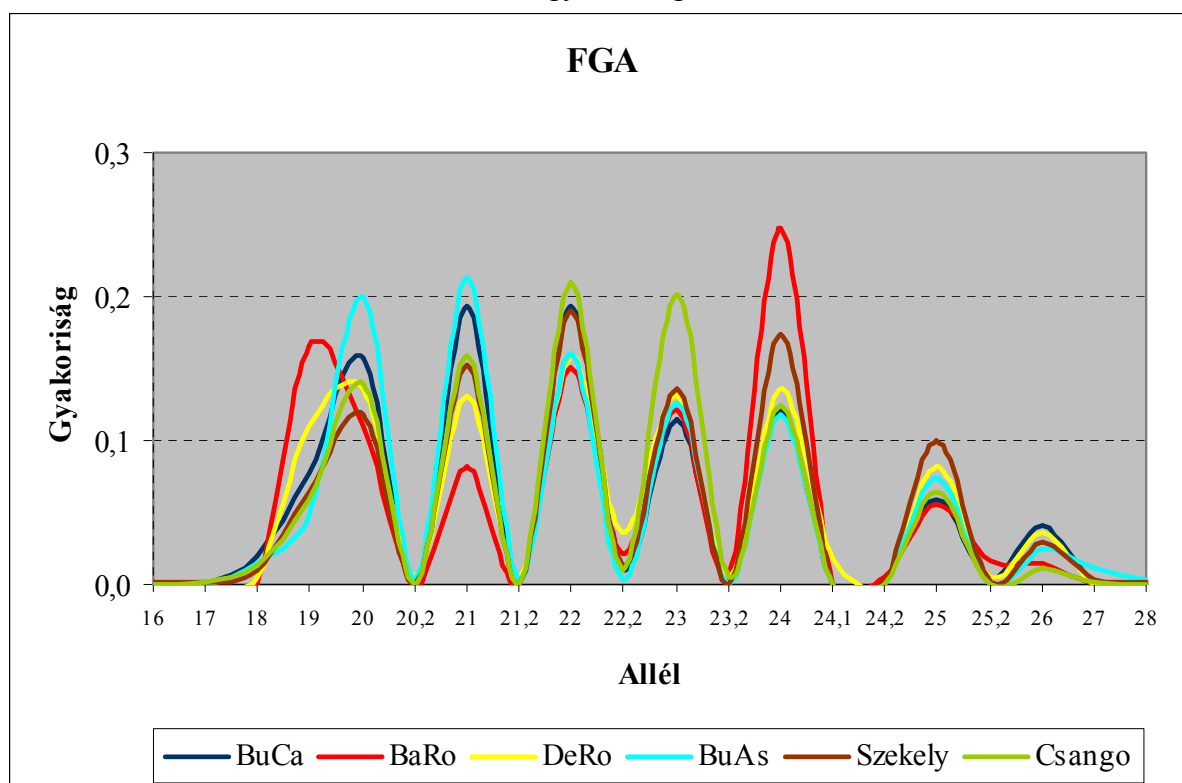
7.1.12. A D8S1179 lókuszt allélgyakoriság-eloszlása



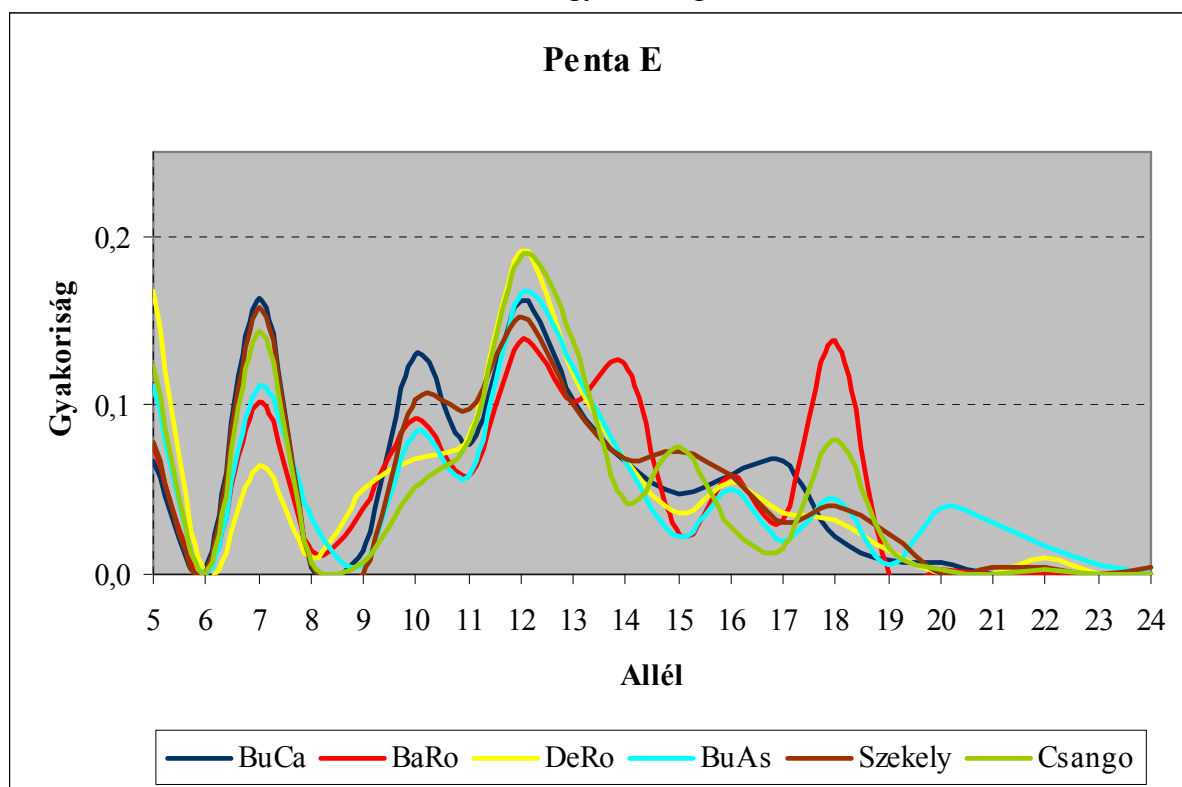
7.1.13. A TPOX lókuszt allélgyakoriság-eloszlása



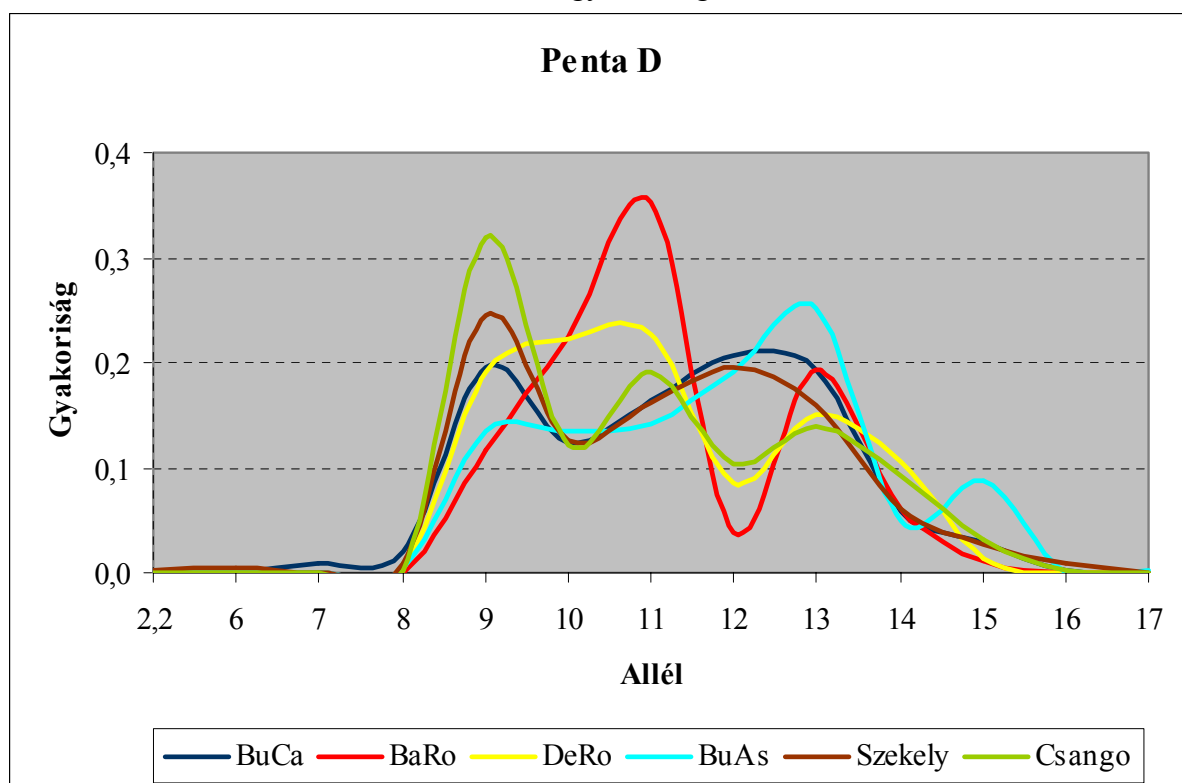
7.1.14. Az FGA lókuszt allélgyakoriság-eloszlása



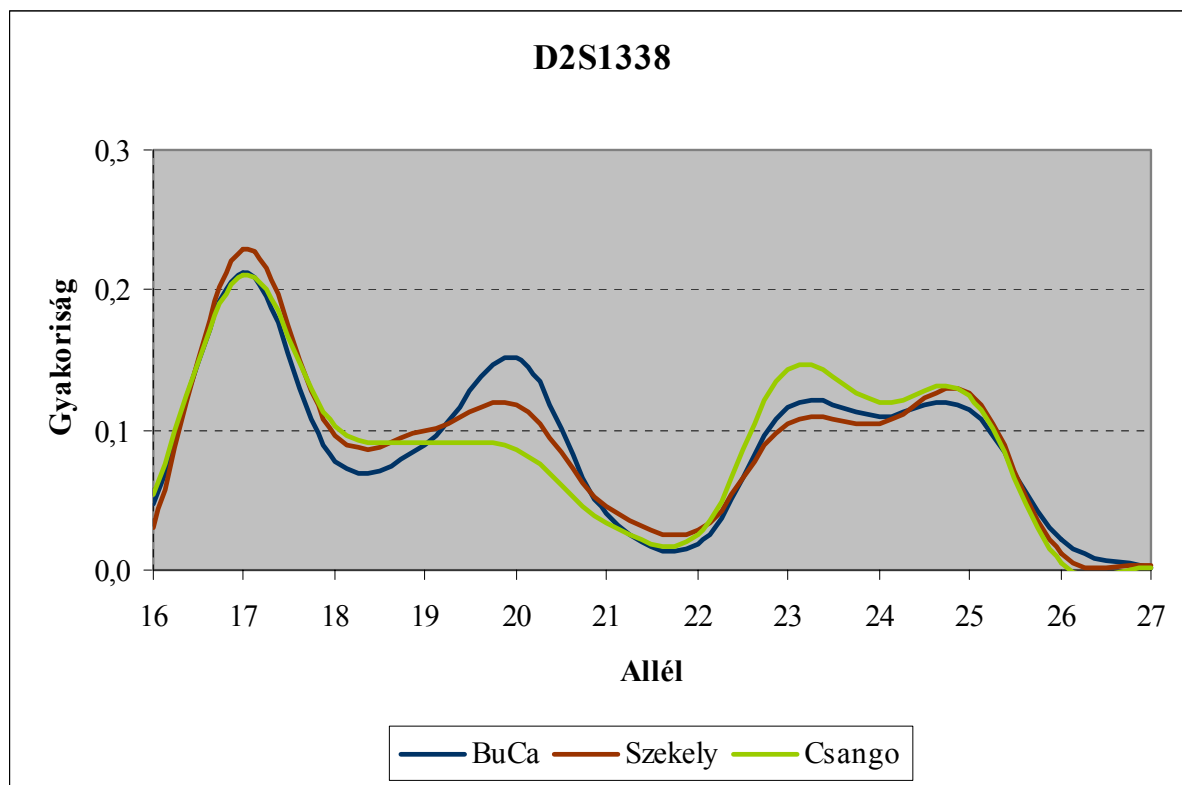
7.1.15. A Penta E lókusztól gyakoriság-eloszlása



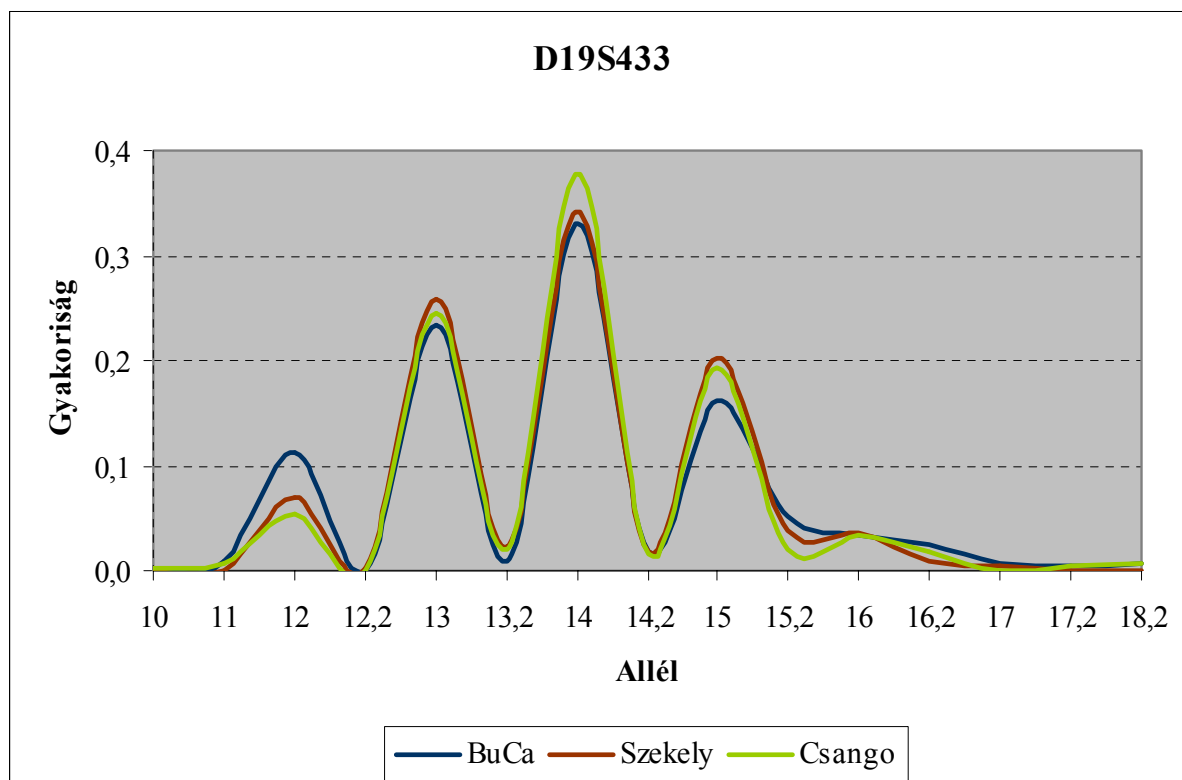
7.1.16. A Penta D lókusztól gyakoriság-eloszlása



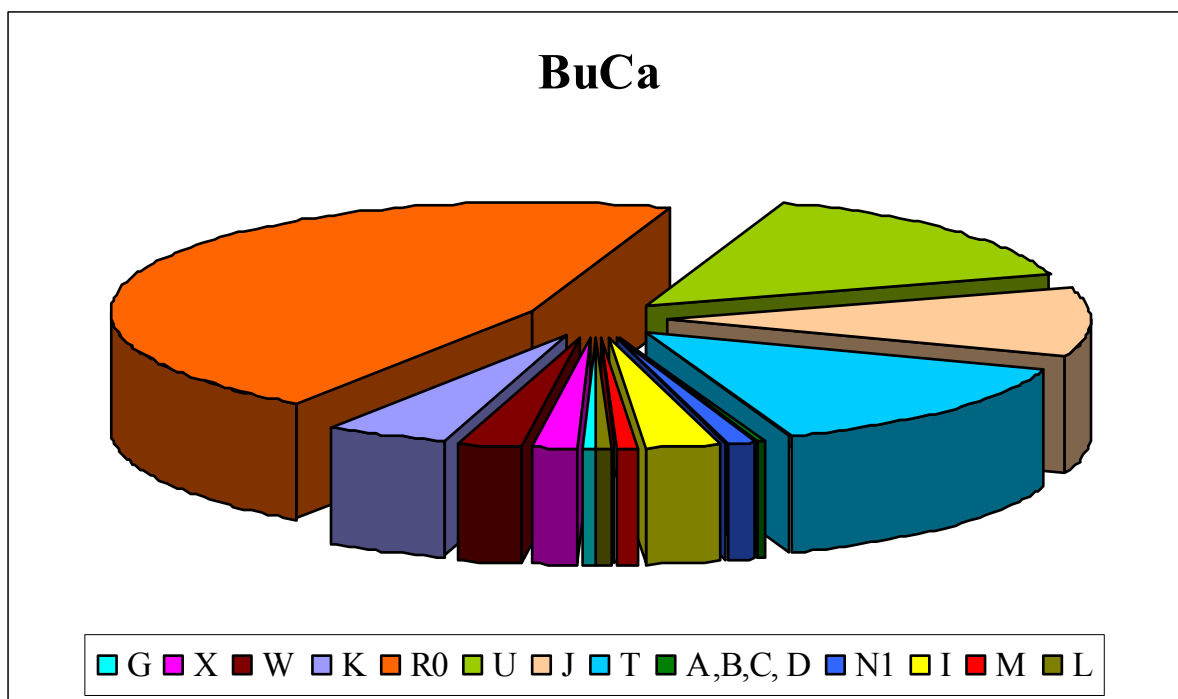
7.1.17. A D2S1338 lókusztól gyakoriság-eloszlása



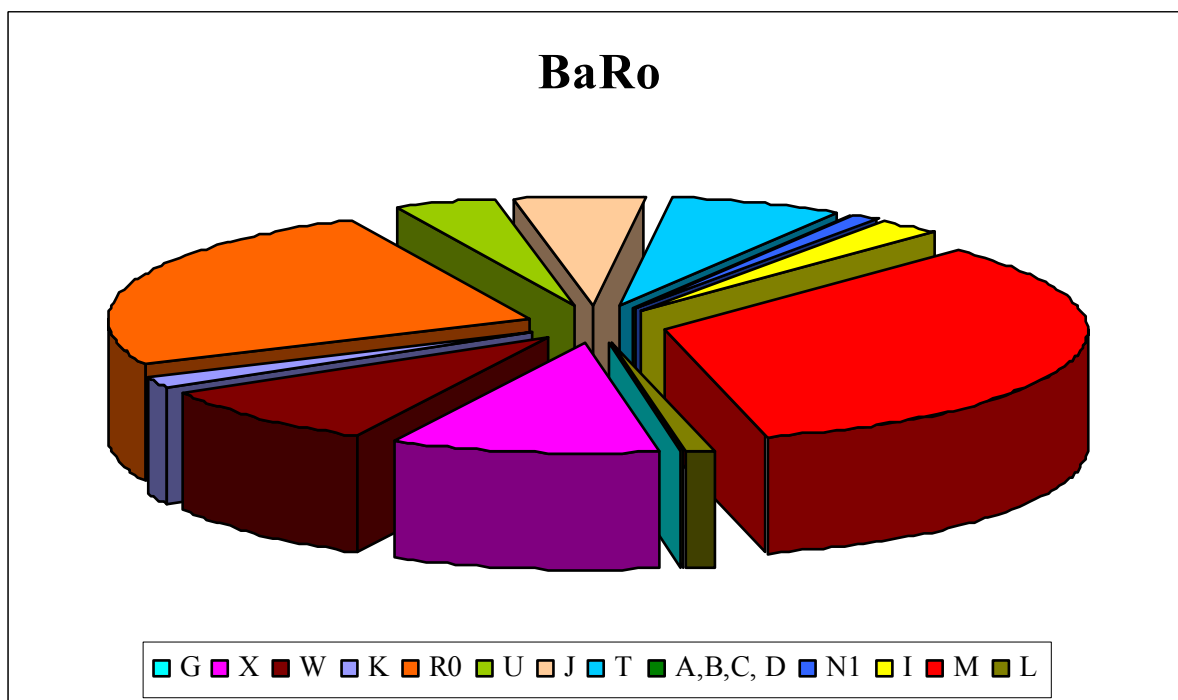
7.1.18. A D19S433 lókusztól gyakoriság-eloszlása



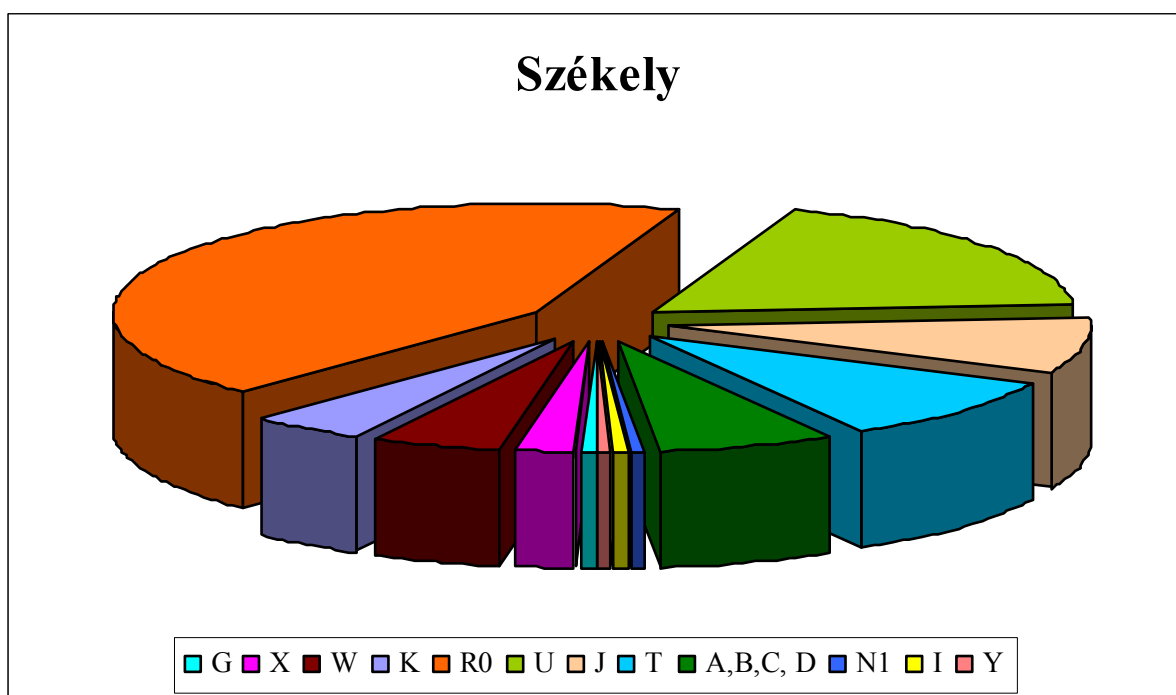
7.1.19. A BuCa populáció mtDNS haplocsoport-eloszlása



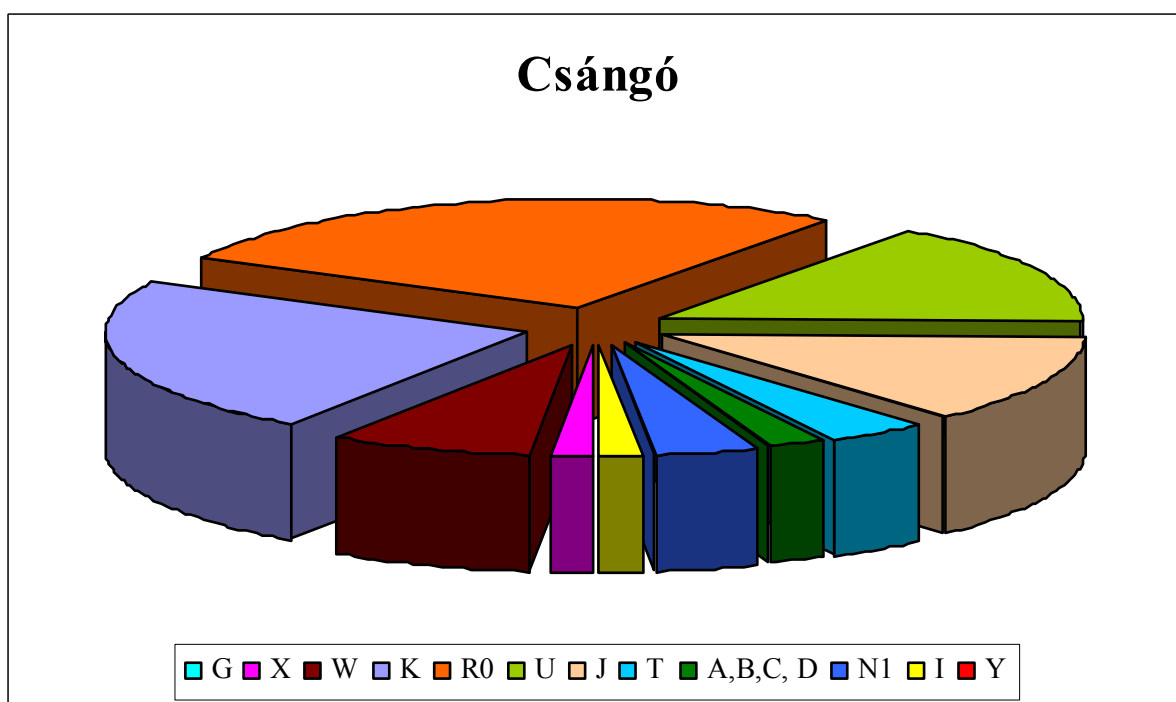
7.1.20. A BaRo populáció mtDNS haplocsoport-eloszlása



7.1.21. A Székely populáció mtDNS haplocsoport-eloszlása



7.1.22. A Csángó populáció mtDNS haplocsoport-eloszlása



7.2. Táblázatok

7.2.1. A populációkban megfigyelt heterozigotitás értékek

Marker	BuCa	BaRo	DeRo	BuAs	Szekely	Csango
CSF1PO	0,714	0,713	0,709	0,741	0,786	0,782
D3S1358	0,767	0,723	0,754	0,854	0,790	0,804
D5S818	0,731	0,655	0,736	0,747	0,685	0,723
D7S820	0,762	0,621	0,727	0,764	0,821	0,836
D8S1179	0,838	0,767	0,791	0,758	0,817	0,782
D13S317	0,785	0,714	0,672	0,775	0,762	0,741
D16S539	0,798	0,810	0,791	0,758	0,766	0,764
D18S51	0,847	0,830	0,900	0,831	0,903	0,877
D21S11	0,861	0,806	0,872	0,803	0,844	0,773
FGA	0,856	0,878	0,854	0,865	0,871	0,845
TH01	0,829	0,786	0,763	0,758	0,790	0,809
TPOX	0,605 ^b	0,709	0,618	0,657	0,634	0,618
VWA	0,834	0,728	0,854	0,803	0,794	0,759
Penta E	0,861	0,903	0,854	0,904 ^a	0,903	0,814
Penta D	0,830	0,694	0,745	0,798	0,840	0,773
D2S1338	0,847	-	-	-	0,844	0,863
D19S433	0,821	-	-	-	0,755	0,782

Magyarázat: ^a = megfigyelt legmagasabb heterozigotitás
^b = megfigyelt legalacsonyabb heterozigotitás
- = nem teszteltük

7.2.2. A populációkban megfigyelt PD (megkülönböztető erő) értékek

Marker	BuCa	BaRo	DeRo	BuAs	Szekely	Csango
CSF1PO	0,877	0,855	0,854	0,853	0,887	0,862
D3S1358	0,926	0,898	0,898	0,918	0,910	0,933
D5S818	0,879	0,841	0,877	0,881	0,859	0,867
D7S820	0,928	0,850	0,928	0,929	0,931	0,925
D8S1179	0,932	0,944	0,933	0,930	0,932	0,914
D13S317	0,924	0,917	0,916	0,914	0,927	0,905
D16S539	0,905	0,932	0,923	0,924	0,908	0,908
D18S51	0,975	0,952	0,959	0,957	0,964	0,958
D21S11	0,957	0,948	0,936	0,946	0,952	0,943
FGA	0,964	0,959	0,968	0,956	0,962	0,955
TH01	0,914	0,913	0,921	0,924	0,919	0,900
TPOX	0,785	0,821	0,760 ^b	0,810	0,800	0,837
VWA	0,926	0,909	0,902	0,911	0,924	0,935
Penta E	0,976	0,978	0,972	0,980 ^a	0,977	0,974
Penta D	0,945	0,909	0,943	0,950	0,946	0,938
D2S1338	0,970	-	-	-	0,970	0,968
D19S433	0,918	-	-	-	0,910	0,890
pM _{komb}	6,7×10 ⁻²¹	3,2×10 ⁻¹⁷	1,4×10 ⁻¹⁷	6,1×10 ⁻¹⁸	1,2×10 ⁻²⁰	4,5×10 ⁻²⁰

Magyarázat: ^a = megfigyelt legnagyobb PD érték
^b = megfigyelt legkisebb PD érték
- = nem teszteltük
pM_{komb} = kombinált egyezési valószínűség

7.2.3. A populációkban megfigyelt PE (apasági kizáró erő) értékek

Marker	BuCa	BaRo	DeRo	BuAs	Szekely	Csango
CSF1PO	0,484	0,455	0,441	0,451	0,517	0,478
D3S1358	0,585	0,527	0,532	0,603	0,566	0,619
D5S818	0,492	0,425	0,480	0,495	0,458	0,489
D7S820	0,598	0,467	0,603	0,604	0,616	0,601
D8S1179	0,626	0,647	0,632	0,623	0,625	0,577
D13S317	0,589	0,570	0,568	0,566	0,602	0,539
D16S539	0,562	0,618	0,598	0,597	0,553	0,553
D18S51	0,768	0,680	0,737	0,707	0,741	0,724
D21S11	0,706	0,668	0,666	0,660	0,679	0,654
FGA	0,722	0,704	0,758	0,700	0,723	0,693
TH01	0,585	0,566	0,591	0,592	0,578	0,547
TPOX	0,356	0,414	0,332 ^b	0,387	0,382	0,406
VWA	0,617	0,551	0,566	0,566	0,594	0,613
Penta E	0,786	0,800	0,784	0,813 ^a	0,796	0,769
Penta D	0,670	0,560	0,642	0,667	0,660	0,629
D2S1338	0,749	-	-	-	0,745	0,746
D19S433	0,605	-	-	-	0,559	0,544
PE _{komb}	0,99999996	0,99999855	0,99999933	0,99999942	0,99999994	0,99999989

Magyarázat: ^a = megfigyelt legnagyobb PE érték

^b = megfigyelt legkisebb PE érték

- = nem teszteltük

PE_{komb} = kombinált kizáró erő

7.3. Saját publikációk

- 7.3.1.** Egyed B, Füredi S, Angyal M, Boutrand L, Vandenberghe A, Woller J, Pádár Z (2000) Analysis of eight STR loci in two Hungarian populations. Int J Legal Med 113:272-275

- 7.3.2.** Boutrand L, Egyed B, Füredi S, Moomers N, Mertens G, Vandenberghe A (2001) Variations in primer sequences are the origin of allele drop-out at loci D13S317 and CD4. *Int J Legal Med* 114:295-297

- 7.3.3.** Egyed B, Csikai M, Füredi S, Pádár Z (2005) Population genetic data on the STR loci D2S1338, D19S433 and SE33 in Hungary. *J Forensic Sci* 50(3):720-721

- 7.3.4.** Egyed B, Füredi S, Angyal M, Balogh I, Kalmar L, Pádár Z (2006)
Analysis of the population heterogeneity in Hungary using fifteen
forensically informative STR markers. *Forensic Sci Int* 158:244-249

- 7.3.5.** Egyed B, Füredi S, Pádár Z (2006) Population genetic study in two Transylvanian populations using forensically informative autosomal and Y-chromosomal STR markers. *Forensic Sci Int* 164:257-265

- 7.3.6.** Irwin J, Egyed B, Saunier J, Szamosi G, O'callaghan J, Padar Z, Parsons TJ (2006) Hungarian mtDNA population databases from Budapest and the Baranya county Roma. *Int J Legal Med* DOI 10.1007/s00414-006-0128-4

- 7.3.7.** 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* doi: 10.1111/j.1469-1809.2007.00371.x

- 7.3.8.** Egyed B, Brandstätter A, Irwin JA, Padar Z, Parsons TJ, 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

7.4. Rövidítések jegyzéke

AMOVA	molekuláris varianciaanalízis (Analysis of Molecular Variance)
AFDIL	Amerikai Légerő DNS Azonosító Laboratóriuma (Armed Forces DNA Identification Laboratory)
BaRo	a Baranya megyei roma populációs mintánk
BuAs	a budapesti askenázi populációs mintánk
BuCa	a budapesti kevert jellegű populációs referencia mintánk
CODIS	Combined DNA Index System (az USA bűnügyi DNS-adatbázis rendszere)
CRS	Cambridge referencia szekvencia (elsőként publikált mtDNS szekvencia)
DeRo	a kelet-magyarországi roma populációs mintánk
EDNAP	igazságügyi genetikusok európai szervezete (European DNA Profiling Group)
EM	Expectation-Maximization algoritmus
EMPOP	európai mitokondriális DNS populációs adatbázis
GMI	Innsbrucki Egyetem Igazságügyi Orvostani Intézete (Gerichtsmedizin Innsbruck)
H	Nei-féle géndiverzitás
H _{obs}	megfigyelt heterozigótaság
HV1	mitokondriális DNS Hipervariábilis Régió 1
HV2	mitokondriális DNS Hipervariábilis Régió 2
HV3	mitokondriális DNS Hipervariábilis Régió 3
HWE	Hardy-Weinberg egyensúly
ISFG	Nemzetközi Igazságügyi Genetikai Társaság
LR	valószínűségi hányados
PD	igazságügyi megkülönböztető erő
PE	igazságügyi szülősi kizáró erő
PIC	polimorfizmus információ tartalom
pM	igazságügyi egyezési valószínűség
p _{min}	minimális allélgyakoriság
RFLP	restrikciós fragmens hossz-polimorfizmus
RMP	Véletlen egyezési valószínűség (Random Match Probability)
SNP	egynukleotid polimorfizmus (Single Nucleotide Polymorphism)
STR	Short Tandem Repeat
SWGAM	Scientific Working Group on DNA Analysis Methods

8. ÖSSZEFOGLALÁS

Allél- és profilgyakorisági adatbázisok nélkül az STR vizsgálatok, haplotípus adatbázisok nélkül a mitokondriális DNS analízis törvényszéki bizonyító erejének statisztikai interpretációja nem lehetséges. Doktori kutatómunkám célkitűzése ezért az volt, hogy autoszómás mikroszatellita lókuszok és a mitokondriális DNS kontroll régió polimorfizmusainak tipizálása révén felállítsunk egy olyan populációs referencia adatbázist, ami megfelelően bizonyulhat a magyarországi igazságügyi genetikai szakértői vizsgálatok korrekt statisztikai genetikai interpretációjára. Populációs felmérésünk hat magyar populációt, összesen 1194 populációs mintát ölelt fel. A populációs minták mindegyikének legalább 15 lókuszos STR-profilját állapítottuk meg, négy populáció 776 személyi mintájának pedig a mitokondriális DNS teljes kontroll régió profilját írtuk le.

A populációs felmérés során a mitokondriális DNS kontroll régió szekvencia-analízis automatizált bevezetésére és hitelesítésére került sor laboratóriumunkban magyarországi igazságügyi genetikai célokra. Kereskedelmi forgalomban elérhető multiplex-PCR és fluoreszcens detektálási rendszerek lehetőségeinek kiaknázásával relatíve gyors, és nagyrészt automatizált módon generáltunk multilókuszos (15-17 STR markerre épülő) genetikai profilokat több száz populációs minta statisztikai elemzése céljából.

A vizsgálatba bevont STR markerek polimorfizmusának populációs felmérésével megállapítottuk, hogy a 13 autoszómás CODIS STR marker analízisével potenciálisan biztosítható Magyarországon az igazságügyben igényelt szintű személyi individualizáció.

Megállapításra került, hogy a magyar népesség genetikai szubstrukturáltsági foka lényeges hatással lehet mind az STR, mind a mitokondriális DNS-vizsgálat bizonyító erejének statisztikai kiértékelésére. Mivel hasonló adatokat mindeddig csak eltérő történelmi-genetikai eredetű etnikai csoportok vagy geográfiailag-genetikailag izolált populációk között mértek, a magyar populációban tapasztalt nagyfokú heterogenitás a magyar népeiséget alkotó populációk diverz genetikai történelmével és/vagy izolációjával magyarázható.

A referencia populációként definiált budapesti kevert népeiségből származó mintacsoport genetikai struktúráját tekintve lényegében megegyezik az őshonos európai népeiséggel, mindazonáltal eredményeink azt mutatják, hogy az alpopulációs teória alapján a magyarországi népcsoportokban élő személyek genetikai korrelációját figyelembe kell venni a DNS-profil egyezési valószínűség kalkulációja során annak érdekében, hogy a törvényszéki DNS-vizsgálatok bizonyító erejének interpretálása tévedéstől mentes legyen.

9. SUMMARY

The strength of DNA evidence in legal procedure could only be interpreted with appropriate population genetic data of the investigated DNA polymorphisms. Therefore the main aim of this study was to introduce and validate genetic analysis of human microsatellite and mitochondrial DNA polymorphisms to the Hungarian forensic practice. In addition, our purpose was also to establish a Hungarian reference population database which could be used for interpreting the weight of DNA evidence presented at Hungarian court.

Six Hungarian populations with altogether 1194 individual samples have been involved into the population study. Each population sample was genotyped at least 15 autosomal STR loci generating multi-loci DNA-profiles and the sequence analysis of mitochondrial DNA entire control region of 776 individuals from four populations was performed.

In the course of the population survey a completely automated process of laboratory manipulations and data transfer was introduced and validated to sequence the entire control region of the mitochondrial DNA in the Hungarian Police DNA Laboratory. Using commercial available multiplex-PCR and fluorescence detection systems a relatively phast and automated generation of multi-loci DNA-profiles (15-17 STR markers) could be performed on hundreds of population samples.

Based on our findings during the survey of STR polymorphisms in the Hungarian populations, the 13 autosomal CODIS core loci involved in the study seemed to be suitable for forensic personal identification in Hungary.

Given the substructure level detected among the Hungarian populations can be a great importance from the point of view interpreting the strength of STR and mitochondrial DNA evidence at the Hungarian court. Since similar data have been observed until now only between ethnic groups with different historical/genetical origin or in geographically isolated populations, the great level of observed heterogeneity in the Hungarian population can be explained with the diverse genetic history and/or isolation of the Hungarian sub-populations.

The reference population sample collected from the mixed urban Budapest population has similar genetic character to the average autochthonous Europeans with normal genetic admixture and there appears to be no basis for treating the reference population separately from other European forensic databases. Based on the substructuring theory our findings suggest that the possibility of population differentiation in the Hungarian communities should be taken into account during the calculation of matching probabilities in the forensic genetic expert opinions to avoid statistical misinterpretation of DNA evidence.