

MTA Doktori Pályázat Bírálata

Dr. Szalai Csaba

Pamjav Horolma: Génekbe vésett vallomások: DNS-ujjlenyomat és őstörténet

Pamjav Horolma értekezése, igen érdekes, a hétköznapi embereket, bűnüldözőket és a genetikusokat egyaránt érdeklő témákkal foglalkozik. A dolgozatban két, egymással kapcsolódó témát mutat be, az igazságügyi genetikát és azt, hogy az Y kromoszóma, illetve a mitokondriális DNS vizsgálatából milyen következtetéseket lehet levonni az emberi történelemmel, népvándorlással, migrációval kapcsolatban.

Az értekezés rendkívül magas színvonalú, érdekfeszítő, szabatos, szép magyarsággal íródott, ami, tekintve, hogy a jelöltnek nem magyar az anyanyelve különösen dicséretes! Pamjav Horolma szakterületén nemzetközileg is elismert kutató, és a magas szintű szakmai hozzáértés az értekezés minden során érződik. Különösen kiemelkedőnek tartom, hogy az ilyen populációgenetikai vizsgálatokban mind genetikában, mind statisztikában és ezzel kapcsolatban valamilyen szinten matematikában is járatosnak kell lenni, és ez Pamjav Horolmánál magas színvonalon teljesül. Sőt, az utolsó fejezetben még az is kiderült, hogy a jelölt zenében is járatos, és a zenegenetika egyik úttörőjének számít!

Az értekezés **23 közleményen alapul**, amelyek közül Pamjav Horolma 5-ben első és 13-ban utolsó szerző. A közlemények összesített impakt faktora 87,194. A bíráló idejében Pamjav Horolma MTMT-ben található független idézettsége 427 volt. Ezek alapján megállapítom, hogy **a jelölt a tudományos publikációk és idézettség tekintetében teljesíti az MTA doktora cím elnyeréséhez szükséges feltételeket.**

Az értekezés elején található a gondosan összeállított rövidítések jegyzéke, melyet egy rövid előszó és köszönetnyilvánítás követ. Összesen 68 oldal az irodalmi háttérrel is tartalmazó bevezető rész. Az „Anyag és módszerek” rész 9 oldal, „Az eredmények és azok megvitatása” 93, melyet a 229 hivatkozást tartalmazó „Irodalomjegyzék” követ. Az értekezéshez egy melléklet is tartozik, amely tartalmazza a dolgozatban feldolgozott cikkek különlenyomatát. Ezek alapján megállapítom, hogy **az értekezés megfelel a formai követelményeknek.**

Az egész dolgozatra jellemző a világos fogalmazás, jól megszerkesztett, logikus felépítés, a kerek és értelmes mondatok. Helyesírási hibák, hibásan megfogalmazott mondatokat csak nagyon keveset találunk, ami nagyban hozzájárul a jó olvashatósághoz.

A **bevezetés** első felében egy rövid ismertetést olvashatunk az igazságügyi genetikáról, leginkább a nemi kromoszómák, illetve a mitokondriális DNS különböző genetikai markereinek használatáról. A bevezetés második felében az ember őstörténetéről és genetikai útjáról olvashatunk, illetve az Y kromoszóma és a mitokondriális DNS haplocsoportok elméletéről, nevezéktanáról, felhasználásáról. A leírás logikus felépítésű, tudományos szempontból kifogástalan és, ahogy az előbb is utaltam rá az egyes részek bonyolultsága ellenére is, jól érthető. A bonyolultságok főleg a haplocsoportok nevezéktanában van, és talán nem is a dolgozatban, ahol a dolgozat későbbi részének érthetőségéhez szükséges mélységben található a nevezéktan leírása, hanem inkább a szakirodalomban, ha az ember megpróbál, valaminek alaposabban utánanézni.

A dolgozatnak ebben a részében néhány kisebb, jelentéktelen pontatlanságot, ellentmondást találtam, aminek egy része az irodalomban is felfedezhető bizonytalanságra vezethető vissza. Ilyen például, hogy az értekezésben az olvasható, hogy a genomban található gének száma 25-30 ezer között van. A jelenlegi definíció alapján kétféle gént különböztetünk meg. Kódoló (ami átíródik fehérjére) és nem-kódoló (funkcionális RNS-t kódol, de nem íródik át fehérjére). A jelenlegi adatbázis alapján kódoló génből 20.418, nem-kódolóból 22.107 van a humán genomban.

Az értekezésben az olvasható, hogy két ember között a DNS állományukban átlagosan kevesebb, mint 0,3%, azaz <10 millió nukleotid különbség van. Ezzel szemben az általam ismert és az egyetemen is tanított adat, hogy két genom között 0,5% azaz kb. 15 millió nukleotid különbség van (Craig Venter két genomjának összehasonlításából), sőt ez az érték néhány egymástól régen elvált populáció között akár jelentősen magasabb is lehet.

Az Y haplocsoportokat ismertető fejezetben meg vannak adva a haplocsoportok magyarországi becsült gyakoriságai. Ha itt a becsült értékek alsó határát (azaz a kisebb számokat) összeadjuk 100%-nál magasabb értéket kapunk (112%). Hozzá kell tenni, hogy a saját eredményeknél részletesen, kördiagramban van a haplocsoportok százalékos eloszlása ábrázolva. A 21. ábrán sajnos a betűk és számok többsége olyan kicsi és elmosódott, hogy még nagyítóval is alig lehet őket elolvasni.

De ezek a kisebb pontatlanságok messze eltörpülnek, az egyébként precíz tudományos megfogalmazások mellett és a dolgozat értékét semmilyen mértékben nem csökkentik.

Az „**Anyag és módszer**” részben a „Minták és populációk” alfejezetben nem érthető a zárójelben található számok jelentése: 26395 magyar minta (4213, 254, 455, 21473). Egyébként a módszerek ismertetései a dolgozat eddigi magas színvonalának megfelelő minőségűek. Az utóbbi időben nagyot fejlődtek a DNS variációk kimutatásának technikái. A kezdeti Southern blottól, a multiplex PCR-n és fragmens analízisen, SNP mikroarray-n keresztül, lassan a legnagyobb felbontású DNS szekvenálás (NGS) is rutin technikává válik. Pamjav Horolma laboratóriuma az STR-eket multiplex PCR-rel és fragmens analízissel, az SNP-eket TaqMan próbával határozta meg. Bár az eredményeknél DNS szekvenálást is említ, az itt nem kerül leírásra. Részletes, bár a később használt módszereket nem teljesen lefedő leírás található az alkalmazott statisztikai módszerekről is.

Az eredmények részben először az igazságügyi genetikában végzett kutatások, vizsgálatok eredményeiről olvashatunk. Az itt végzett kutatások egyrészt alapkutatások, másrészt a manapság kedveltebb (jobban támogatott) alkalmazott kutatások közé tartoznak, hiszen az itt elért eredményeket közvetlenül lehet a bűnüldözésben és egyéb bírósági ügyekben alkalmazni. Egyik érdekes új eredmény hogy a D12S391 lokuszon 62 új allélt detektáltak, illetve egy, az addigiaknál jelentősen nagyobb magyar populáción, 26.395 főn mérték meg az STR lokuszok allélgyakoriságát, melyek közül 21.473 fogvatartott eredményei bekerültek a nemzeti bűnügyi adatbázisba. Megállapították, hogy a nagyobb mintaszám pontosabb valószínűségi hányados számítást tesz lehetővé.

A következő részben az X kromoszómán bekövetkező rekombinációkat vizsgálták 45 három generációs családon, 155 főn, 4 kapcsoltsági alcsoport között. Továbbá az X kromoszóma genotipizálásának alkalmazására mutat be példát apasági egy perben.

Ennek az alfejezetnek a további részeiben az Y kromoszóma és az mtDNS vizsgálatok igazságügyi alkalmazásáról van szó, ami bár érdekes igazából nem kutatás, ha jól láttam nem is született belőle tudományos cikk, így ezeket talán nem is lett volna fontos itt bemutatni.

A következő nagy és talán az értekezés legfontosabb fejezete az emberi őstörténet, migrációk kutatása az Y kromoszóma és az mtDNS variációk vizsgálata segítségével. Az első részben egy 2009-es publikációra utalva a magyarországi férfiak Y kromoszóma haplocsoportjainak eloszlását mutatja be. Itt egy kisebb hiányosság, hogy innen az irodalmi hivatkozás hiányzik, az csak később kerül említésre. Itt szintén előjön a nevezéktan problémája. Az itt ismertetett haplocsoportok és a bevezetőben bemutatottak

elnevezései között nem teljes az átfedés, emiatt sokszor nehéz a kettőt egymással megfeleltetni. Milyen mintán lettek meghatározva a bevezetőben leírt Y haplocsoportok magyarokra vonatkozó gyakoriságai? Ez ott nem derül ki, és van némi eltérés a két eredmény között. Pl., ha jól értem a Q haplocsoport a vizsgált magyar populációban hiányzik, az irodalomban leírtban 1-3% gyakoriság szerepel. Továbbá a 31. ábrán miért van jelölve a 0% gyakoriságú R2 haplocsoport, illetve mit jelentenek egyes haplocsoportoknál a *-k?

Számomra furcsa a 32. ábrán, a filogenetikai fán a „Yugoslavia” populáció. Ennyire homogénnek tekinthetők az egykori Jugoszlávia területén élő populációk, hogy populációgenetikai vizsgálatokban egyetlen populációként lehet rájuk utalni?

Szintén furcsa számomra, hogy az ábra alatti szövegben Spanyolország és Olaszország észak-európai országgént vannak említve.

Még egy megjegyzés, ami nem a jelölt munkáját érinti. Tudom, hogy a gyökér nélküli filogenetikai fákat így szokás ábrázolni, de talán informatívabb lenne, ha nemcsak a szemre bíznák az itt látható távolságok becslését, hanem a távolsághoz valamilyen egység is lenne rendelve, ahogy az később pl. az 50. ábrán is látható.

A dolgozat további része is magas színvonalú kutatásról, és alapos, tudományos gondolkodásról tanúskodik, számos érdekes és fontos eredménnyel. Én az értekezésből egy fejezetet hiányolok. Nincs egy összefoglaló rész, amiben a jelölt röviden, pontokba szedve összefoglalja a legfontosabb eredményeit.

Még egy, talán a mérnöki végzettségemre visszavezethető, számomra zavaró szokást jegyeznék meg az 53. ábrával kapcsolatban, ahol egy MDS térkép látható az Fst genetikai távolság alapján. Láttam, hogy a megjelent cikkben is így szerepel, és gondolom ez az ábrázolás általánosan elfogadott, de számomra innen nagyon hiányzik a tengelyek mellé írt definíció, azaz mit ábrázolunk minek a függvényében, ami az ábrát sokkal könnyebben értelmezhetővé tenné.

Nagyon érdekesnek találtam a zenegenetikáról szóló, utolsó fejezetet, amelyben a különböző népek népdalainak és genetikai hasonlóságainak összefüggéseit vizsgálták. A vizsgálat rendkívül kreatív és gyakorlatilag egy új tudományos diszciplína alapjait rakta le. A jelölt írta, hogy az interdiszciplinaritás miatt nehézségük volt a közlemény publikálásával. Én úgy gondolom, hogyha a cikket nem

Magyarországon, hanem mondjuk az USA-ban írták volna, nagy eséllyel a sokszor hasonló cikkeket publikáló Nature-ben jelenhetett volna meg.

Összefoglalva, a disszertációban leírtak és a mellékelt publikációk alapján, én **az alábbi eredményeket tartom a jelölt saját, a tudományt előremozdító, jelentős eredményeinek:**

1. Az igazságügyi genetikában a D12S391 lokuszon 62 új allélt detektáltak. Ennek jelentőségét mutatja, hogy 2017. januárjától ez a marker az FBI CODIS rendszerébe is bekerült.
2. Egy nagy magyar populáción meghatározták az igazságügyi személyazonosításnál használt STR lokuszok allélgyakoriságát, melyek közül 21.473 fogvatartott eredményei bekerültek a nemzeti bűnügyi adatbázisba, pontosítva ezzel személyazonosításnál a valószínűségi hányados számítását.
3. Több populációban meghatározták az Y kromoszóma haplocsoportjait és számos érdekes és új megállapítást tettek, pl.:
 - A kárpát-medencei magyar, székely és csángó populációk Y-kromoszómális haplocsoportjai eloszlásai közel azonosak egymással, valamint a közép- kelet-európai országok populációiével, de az elszigeteltebb székely és csángó populációk jobban megőrizték a dél-szibériai eredetű haplocsoportokat is.
 - Értékes megállapításokat tettek a magyar férfiak történelmi migrációjával kapcsolatban, pl. a magyarok 20%-a a neolitikumból, 8-10 ezer éve érkezett, 46% bronzkori bevándorló és csak 1% ugor.
 - A magyar férfiakat elhelyezték egy nemzetközi, 23 európai populációt tartalmazó filogenetikai fán. Megállapították például, hogy a magyar populáció a közép-európai klaszterbe tartozik, míg a finn populáció a magyaroktól nagyon távol helyezkedik el.
 - Megállapították, hogy az európai roma férfiak, bár megtalálható bennük a befogadó populációk genetikai nyomai is, még mindig hordozzák indiai őseikét is.
4. Nemzetközi szinten elsőként azonosították R1a haplocsoport két alcsoportját az R1a-Z280-at és az R1a-Z93-at. Ezek segítségével sikerült különbséget tenni az ázsiai és az európai R1a Y kromoszómák között.
5. Megállapították, hogy az Y-haplocsoportok alapján nagy a genetikai távolság a magyar és más finnugor népek között, viszont kapcsolatot találtak a legközelebbi nyelvi rokonaink a manysik és a magyarok között.

6. Megállapították, hogy a székelőknek a belső-ázsiai hunokhoz vagy az avarokhoz való speciális viszonyát a jelenlegi genetikai adatok nem támasztják alá.

7. Az mtDNS haplocsoportok alapján megállapították, hogy a honfoglaló magyarok génállományának jelentős része közép-ázsiai és dél-szibériai génállományból származhatott, amely még mindig megtalálható a magyar népességben. A honfoglalók nyugatra vándorlásuk során különböző európai populációkkal is keveredtek. A honfoglaló magyarokhoz közel álló mai európai populációk indoeurópai nyelveket beszélnek.

8. Elindítottak egy új kutatási ágat, a zenegenetikát. Itt, 31 zenei kultúrát és a hozzájuk tartozó populációk genetikai hátterét vizsgálva megállapították, hogy a zenei hasonlóság és a genetikai hasonlóság között kapcsolat van, azaz a hasonló zenei kultúra a nagyon régi közös ősökre vezethető vissza. Viszont a szoros genetikai kapcsolat nem feltétlen indikátora a hasonló zenei kultúrának.

Az értekezéssel kapcsolatos kérdéseim:

1. Magyarországon milyen gyakran veszi igénybe a bíróság/rendőrség/ügyészség az igazságügyi genetikai technikákat? Milyen típusú ügyek a leggyakoribbak itt?

2. Jelenleg még mindig STR alapú a személyazonosítás, bár az SNP-alapú mérések szerintem egyszerűbbek, pontosabbak és nagyobb felbontásúak is. Mit gondol erről? Mi a tendencia, lesz-e valamikor áttérés másik technikára?

3. Az elmúlt években a technika fejlődésével és a személyes genomikai cégek elterjedésével több millió ember Y, és mtDNS haplocsoportjait állapították meg. Mennyire lehet ezeket beépíteni a korábbi, jóval kisebb számú minták értékeléseiből származó eredményekbe?

4. Hasonlóan, több millió személyből származó adat áll rendelkezésre nagy felbontású SNP mérésekből (pl. GWAS, személyes genomikai cégek). Ezek az eredmények mennyire módosítják, finomítják az Y kromoszómákból, mtDNS haplocsoportokból származó migrációs, rokonsági eredményeket?

5. Az emberiség manapság tapasztalható expanziója miatt évente >130 millió gyerek születik, azaz nagy eséllyel rengeteg új mutáció keletkezik pl. az Y kromoszómán is (a dolgozatban szerepel, hogy annak az esélye, hogy az Y kromoszómán már egy meglevő ősi szubsztitúció lokuszán, egy másik férfiban még egyszer mutáció történjen 1:57 millió, azaz gyakorlatilag esélytelen. Viszont ilyen sok gyerek születésével ez az esély már érzékelhető). Előfordult-e hogy valamelyik ősi haplocsoport

visszamutálódott? Leírnak-e új, haplocsoportokat (mint pl. E, vagy R), vagy mindig besorolják az újakat a régi csoportokba?

6. Az eredmények alapján a mai magyarok (illetve általában az európaiak is) túlnyomó többsége több 1000 éve itt élő emberek leszármazottai (pl. a magyarok 20%-a a neolitikumból, 8-10 ezer éve érkezett, 46% bronzkori bevándorló és csak 1% ugor). Ennyire kis hatással voltak a mai populációk összetételére az elmúlt kb. 2000 év, történelemből tanult, Ázsiából jövő népvándorlásai?

7. Végül egy, a dolgozattal összefüggő személyes kérdés, amely rámutat a nevezéktan bonyolultságára is. Engem egy genomikai cég az E-L29 (ami talán M84, vagy e1b1b1c1a?) Y kromoszómás haplocsoportba sorolt, ami a cég szerint az E-M123-ból leágazó E-M34-nek egy további leágazása, és állítólag elég gyakori, a vizsgált 5 millió emberből, minden 310. ebbe a haplocsoportba tartozik. Ennek ellenére gyakorlatilag semmi nem található róla az interneten/szakirodalomban. Hogyan lehet az ilyen haplocsoportokat beazonosítani?

Összességében megállapítom, hogy Pamjav Horolma szcientrometriai adatai megfelelőek, cikkei és a benyújtott értekezés magas színvonalúak. Az értekezéssel párhuzamosan benyújtott tézisfüzet jól foglalja össze az értekezés fő mondanivalóit. Pamjav Horolma kiemelkedő minőségű kutatásaival, fejlesztéseivel jelentős mértékben járult hozzá nemcsak a magyar, hanem a nemzetközi tudományos ismeretekhez jelentősen gyarapítva a tudomány szakot és hozzájárult a tudomány fejlődéséhez. **Ezek alapján a nyilvános vita kitézését és a mű elfogadását javaslom.**

2019. február 15.

Dr. Szalai Csaba MTA doktora

SE Genetikai Sejt és Immunbiológiai Intézet

Igazgató helyettes

Heim Pál Kórház