Doktori (Ph.D.) értekezés

A MIGRÁCIÓ HATÁSAI A ROMÁK GENETIKAI FELÉPÍTÉSÉRE

Bánfai Zsolt

A Doktori Iskola vezetője:

Dr. Gallyas Ferenc

egyetemi tanár

Témavezető:

Dr. Melegh Béla

egyetemi tanár

Pécs, 2021

Pécsi Tudományegyetem, Klinikai Központ, Orvosi Genetikai Intézet

Bevezetés

Teljesgenom-adatokon alapuló populációgenetika

Az információtechnológia fejlődése, úgy, mint a nagy teljesítményű számítástechnika, a digitális tárhelyek fejlesztése, az újgenerációs szekvenálás, a microarray-alapú genotipizálás lehetővé tette a kutatók számára az egyes populációkból származó szekvenciák és genotípus-adatok genomszintű összehasonlítását. Ez jelentős lépés volt a populációgenomika felé. A populációgenomika lehetővé teszi a populációgenetikai jelenségek nagy léptékű, sokkal összetettebb vizsgálatát, egyidejűleg több százezer lókusz vagy marker vizsgálatát. Lehetővé teszi a populációstruktúra, a populáció eredetének, és az olyan folyamatok vizsgálatát is, mint az egyes populációk keveredése és migrációja; ezáltal lehetővé téve a történelem eseményeinek feltárását, illetve jellemzését. A populációgenomikának klinikai felhasználásai is léteznek. Például lehetővé teszi annak vizsgálatát, hogy a genetikai variáció miként viszonyul a különböző betegségekhez és szindrómákhoz, és ez miként felelős az emberek egészségéért.

A szekvenálási és genotipizálási technológiák fejlődése felgyorsította az olyan kutatócsoportok megalakulását, amelyek célja az volt, hogy ezt a hatalmas mennyiségű kapott adatot képesek legyenek felhasználni populációgenomikai célokra, mind az emberi történelmet célzó, mind klinikai célú kutatásokhoz. Ezzel együtt e kutatócsoportok új matematikai és statisztikai módszerek kifejlesztését is célul tűzték ki, amelyek alkalmasak kezelni és felhasználni a nagymennyiségű genetikai adatban rejlő lehetőségeket, valamint leküzdeni a hatalmas adathalmazzal járó analitikai és számítási kihívásokat. Ezek a kutatócsoportok gyakran bárki számára elérhetővé teszik szoftvereiket kutatási célzattal. Ilyen kutatócsoportok többek között, a Stanford Egyetemen található Pritchard Lab, amelyet Jonathan Karl Pritchard vezet, a Tang Lab, amelyet szintén a Stanford Egyetemen dolgozó Hua Tang vezet, a Reich Lab, David Reich vezetésével amely a Harvard Medical School, a Broad Institute és a Massachusetts Institute of Technology együttműködésével jött létre, továbbá Daniel Lawson laboratóriuma a Bristoli Egyetemen, Brian Browning laboratóriuma a Washingtoni Egyetemen és Manfred Kayser kutatócsoportja a rotterdami Erasmus Egyetemen.

Az ezekhez hasonló kutatócsoportok folyamatos munkája lehetővé tette a teljes genom szekvenálásából származó adatok és a teljes genomra kiterjedő marker adatok kutatási célú felhasználását azáltal, hogy új populációstruktúra becslésre, leszármazás-vizsgálatára, haplotípusok azonosítására, állapot (identical-by-state, IBS) és leszármazás (identical-by-descent, IBD) alapján azonos DNS szegmensek detektálására, valamint genetikai keveredés elemzésére alkalmas számítógépes módszereket fejlesztettek ki. A populációtörténeti kutatások mellett a klinikai célú

analitikai módszerek (teljes genomra kiterjedő összehasonlító vizsgálatok – GWAS, GWA kutatás) is fejlődnek. Az egyik első átfogó szoftvercsomag a GWA kutatásokhoz a PLINK programcsomag volt. A PLINK tartalmazta az összes alapvető populációgenetikai statisztikai módszert, például a fixációs index (F_{st}) számítást, a Hardy-Weinberg egyensúly tesztelését, az endogámiai együtthatót, a Mendeli öröklődési hibák tesztjét és még sok más alapvető statisztikát, amelyeket a genotipizált biallélikus marker adatokon el lehet végezni. Ezenkívül képes kapcsoltság vizsgálatára, az IBD és az IBS DNS szegmensek azonosítására, episztatikus kölcsönhatások, populációstruktúra vizsgálatára, továbbá természetesen kvantitatív tulajdonságokra és eset-kontroll vizsgálatokra irányuló GWA tanulmányok elvégzésére is alkalmas. Természetesen sok fejlesztő klinikai és populációtörténeti célokat egyaránt figyelembe vesz, és mindkét célra fejleszt algoritmusokat. David Reich laboratóriumának munkatársai által kifejlesztett EIGENSOFT szoftvercsomag esetében két olyan populációstruktúrát vizsgáló módszer áll rendelkezésre, amelyek mindkét terület igényeinek megfelelnek.

A romák

A roma nép egy transznacionális, diaszpórikus etnikai csoport Dél-Ázsiából, amely számos társadalmilag és genetikailag eltérő alcsoportot foglal magába. A roma népesség méretét jelenleg 10-15 millióra becsülik, és többségük Európában, különösen Európa középső régiójában található. Jelentős számban találhatók meg továbbá az Ibériai-félszigeten is. Európán kívül a Kaukázus régiójában, a Közel-Keleten és Amerikában élnek. Genetikai összetételüket legalább egy alapító hatás, a genetikai sodródás és egyéb folyamatok alakították ki, ilyen például a differenciális keveredés elsősorban nyugat-eurázsiai népekkel. A roma alcsoportokat az indiaiakéhoz hasonló társadalmi szokások, egyfajta kasztrendszer okozta korlátozott génáramlás hozta létre, amely révén a klánnak is nevezett genetikailag zárt alcsoportok jöttek létre, ezáltal a roma lakosságban jellemző szubstruktúrát hozva létre. Az indiaiakhoz hasonlóan ezekben a roma alcsoportokban is magas a közeli rokonok házasságkötésének aránya. A romák genetikáját és genomját célzó humángenetikai kutatások számos mendeli öröklődést követő ritka betegséget okozó mutációt azonosítottak a roma népességben, amelyek semmilyen más etnikai csoportban nem fordulnak elő. Ennek oka lehet a már ismertetett speciális populációgenetikai események és jellemzők hatása, és amelyek révén a roma népesség genetikai és genomikai vizsgálata az orvosi genetika egyik fontos céljává vált. Korábban ezek a vizsgálatok bizonyos lókuszok és markerek összehasonlító statisztikai elemzésén alapultak, azonban a 2000-es években végbement technológiai fejlődés lehetővé tette a genom egészére kiterjedő orvosi genetikai eset-kontroll vizsgálatokat is.

Már a populációgenetikai vizsgálatok térhódítása előtt számos tanulmány készült származásukról, történelmi, nyelvi, antropológiai bizonyítékok alapján. Úgy gondolták, hogy a nyelvészet és az

antropológia hasonlóságokat fedez fel egyes indiai etnikai csoportok és a romák között. Az összehasonlító nyelvészet szerint az indiai nyelvek közül a hindi hasonlít leginkább a roma nyelvre. Az antropológiai kutatások mutattak rá először a romák indiaiakéhoz hasonló kasztrendszerére, és hogy több roma csoportban is megtalálható az endogámia.

Célkitűzések

- Indiától Európáig az egyes régiók populációinak roma leszármazáshoz való hozzájárulásának leírása
- A kaukázusi régió jelentőségének meghatározása a romák leszármazásában
- A proto-romák migrációja során bekövetkezett keveredési események jellemzése
- Kelet-Közép-Európa oszmán megszállásának a romákra és az Európa egyes populációira gyakorolt hatásainak vizsgálata

Anyagok és módszerek

Anyagok

Kutatásunk a résztvevők EDTA-val antikoagulált vérmintáiból izolált DNS-en alapul. A résztvevők Közép-Európából (Magyarországról vagy Magyarország szomszédságából) származó egyének voltak, akik legalább három generációra visszamenőleg tudatosan románnak vallották magukat. Minden résztvevő írásos beleegyezését adta a vizsgálatban való részvételhez. Mindannyian személyes szóbeli felvilágosításban részesültek a beleegyezési nyilatkozat aláírása előtt, amelyet a pécsi Regionális Kutatásetikai Bizottság hagyott jóvá ehhez a tanulmányhoz. A minták anonimizálásra kerültek. A kutatás a Helsinki Nyilatkozatban megfogalmazott elveket követi.

Adatkészletek

A kutatásban szereplő közép-európai roma minták egyik része a Harvard Medical Schoollal folytatott nemzetközi együttműködésünkben keretében kerültek összegyűjtésre és genotipizálásra. A genotipizálást az Affymetrix Genome-Wide Human SNP 6.0 array chipen (n = 27; 726 016 SNP) végeztük. A mintakészlet másik része a Rotterdami Egyetem adatbázisából származnak (n = 152, 868 174 SNP), amelyet kérésre bocsátottak rendelkezésünkre. Ez az adatkészletet ugyanazon a microarray platformon került genotipizálásra, és összesen 179 mintát tartalmazott. A genotípushívást és a minőségellenőrzést az Affymetrix Power Tools parancssoros szoftvercsomagjában található Birdseedv2 algoritmus segítségével, az Affymetrix által ajánlott beállítások alkalmazásával végeztük el. A nyers genotípushívásokat egy házon belül készített szkript és a PLINK1.9 segítségével bináris PLINK formátumra konvertáltuk, és a genotípusadatok további szűrésre kerültek a genotípushívások

mintánkénti hiányosságai alapján. A hiányzó genotípusú SNP-ket eltávolítottuk az adatokból, a PLINK1.9 markerenkénti genotípushiányosságok azonosítására alkalmas algoritmusa segítségével. Ez azt jelenti, hogy minden olyan markert eltávolítottunk az adatkészletből, amely genotípusa nem található meg egységesen mind a 179 mintában.

A HapMap Phase 2 GRCh37 genetikai térkép felhasználásával a genetikai távolságot jellemző értékeket is hozzáadtuk a markeradatokhoz a PLINK segítségével, amely lehetővé tette a kapcsoltságon alapuló vizsgálatok elvégzését. A kapott adatkészlet 599 472 SNP-t tartalmazott a genotipizálást és az utólagos további minőségellenőrzést, illetve szűrést követően.

Az előzetes populációstruktúra vizsgálatok alapján a nem-roma európaiakkal erősen keveredett roma egyének az adatokból eltávolításra kerültek, amely 21 roma mintát érintett. Ez alapján 158 közép-európai roma mintát találtunk alkalmasnak további vizsgálati célokra.

Kutatásaink során több egyéb adatkészletet is alkalmaztunk, amelyek vagy saját adatkészletek, vagy szabadon hozzáférhető nyilvános online tárhelyekről származó, vagy kérésünkre rendelkezésünkre bocsátott adatok voltak. Továbbá olyan adatkészletet is alkalmaztunk, amelynek használata engedélyhez kötött.

Felhasználtunk mintákat a nyilvánosan és szabadon elérhető HGDP adatkészletből (n = 1044 57 populáció, 660 918 SNP Illumina 650 Y platformon genotipizálva). Az adatkészlet különböző populációk adatait tartalmazza a világ minden tájáról globálisan. Két további szabadon hozzáférhető adatkészletet is felhasználtunk az Estonian Biocentre adatállományaiból, amelyek két különböző tudományos publikációban kerültek leírásra, felhasználásra. Ezekre az adatkészletekre a továbbiakban a „kaukázusi adatok”-ként (n = 204, 13 populáció, 555 767 SNP) és a „zsidó adatok”-ként (n = 466, 39 populáció, 555 736 SNP) fogunk hivatkozni tükrözve ezzel az adott publikáció témáját, amelyben eredetileg leírásra kerültek. Egy indiai etnikai csoportokat tartalmazó, kérésünkre rendelkezésünkre bocsátott adatkészletét szintén alkalmaztunk (n = 121, 23 csoport, 524 053 SNP, Affymetrix 1 M és Illumina 650K platformokon genotipizálva), amelyet a Harvard Medical School hozott létre, és a indiai etnikai csoportok történetének felderítésére használtak fel. Az engedélyezett hozzáférést igénylő adatkészlet a POPRES adatkészlet volt az NCBI dbGaP adattárából (n = 4077, 57 populáció, 453 617 SNP Affymetrix 500 K platformon genotipizálva). Az adatkészlet a HGDP adatokhoz hasonlóan globálisan a világ minden tájáról származó populációk adatait tartalmazza, viszont vizsgálataink során csak európai és indiai mintákat használtunk fel belőle. Az előzetes szűrésen átesett, n = 238 egyénből és 898 723 SNP-ből álló magyar egyéneket tartalmazó adatokat szintén felhasználtuk bizonyos vizsgálatokban.

A romák keveredési folyamatait kétféle megközelítéssel vizsgáltuk, ezért két különböző fő adatkészletet hoztunk létre vizsgálatainkhoz. Az első megközelítés egy olyan adatkészlet létrehozása

volt, amely a roma nép migrációs útvonalára eső régiók populációit tartalmazta. Ezek a regionális csoportok az európaiak, a Kaukázus, a Közel-Kelet és Dél-Ázsia populációi voltak. Utóbbi csoportok pakisztáni és indiai populációkból állnak. Az adatkészlet létrehozásához a HGDP-adatokat, a Kaukázus és zsidó adatokat, illetve az indiai adatokat használtuk fel. A második megközelítésnél az Oszmán Birodalom hatását vizsgáltuk az oszmánok által megszállt országok népeire és a romákra nézve. Ezekben a vizsgálatokban európai, kaukázusi és közel-keleti mintákat vizsgáltunk, amelyekhez az előbb felsorolt adatkészleteken kívül a POPRES adatkészletet és a magyarokat tartalmazó adatkészletet is felhasználtuk.

Módszerek

A vizsgálatba bevont populációk szerkezetének és kapcsolatának tanulmányozásához az EIGENSOFT 6.01 szoftvercsomag SMARTPCA programját használtuk, amelyet még jelenleg is támogatnak, frissítenek, és amelyet a Harvard Medical School és a Broad Institute együttműködése keretében került kifejlesztésre. A legtöbb, ebben a tanulmányban alkalmazott, statisztikai módszereket alkalmazó szoftverrel ellentétben ez a szoftver matematikai, algoritmikus elven alapszik. A SMARTPCA főkomponens-analízist (PCA) végez egy allélfrekvencia mátrixon, a többdimenziós adatokat minimális információvesztéssel számszerűsíthető és értelmezhető, az adott kutatás szempontjából releváns dimenziókra szűkítve. A program a számított sajátértékek alapján szignifikancia-teszteket is végez. Segítségével a teljes genomra kiterjedő markeradatokban releváns mintázatokat tudunk megfigyelni, amelyek tükrözik a vizsgált etnikai csoportok kapcsolati viszonyait, és megadja a vizsgált populációk F_{st} mátrixát is. Ez az egyes etnikai csoportok páronként átlagos allél frekvencia-differenciációs mátrixa (fixációs index), és tükrözi a populációk egymáshoz való viszonyának mértékét.

A leszármazásbecslést és klaszteranalízist az ADMIXTURE segítségével végeztük, amely egy statisztikai módszeren alapuló klaszterező szoftver. Az ADMIXTURE szoftver egy ún. expectation maximization algoritmust használ a legnagyobb valószínűség becslés (maximum likelihood – ML - estimation) módszerének megvalósítására, amelynek célja a vizsgált populációk és adott hipotetikus ősök közötti kapcsolat mértékének meghatározása. A programba beépített keresztvalidációs módszerrel meghatározható a lehetséges közös ősök maximális száma (K értéke).

A TreeMix 1.13 algoritmus az ADMIXTURE-höz és számos más leszármazásbecslési algoritmushoz hasonlóan a STRUCTURE program működési elvén alapszik. A szoftver az autoszomális allélfrekvencia-adatok felhasználásával egy ML gráfot hoz létre. Segítségével a különböző etnikai csoportok kapcsolatára következtethetünk, mivel a program meg tudja becsülni az egyes vizsgált populációk esetleges szétválási és keveredési eseményeit, de alkalmazható a migrációs események azonosítására is.

A keveredési események teszteléséhez az ADMIXTOOLS 4.1 szoftvercsomagban található programokat alkalmaztuk. Ezek a statisztikai algoritmusok a vizsgált populációk közötti allélfrekvencia-korrelációk mérésén alapulnak. Fejlesztői, a Broad Institute kutatói, f-statisztikaként hivatkoznak ezekre a módszerekre, azonban a 4 populációs tesztet a csomag D-statisztikaként valósítja meg.

Bizonyos tesztekben a Beagle 4.1 Refined IBD algoritmusát is alkalmaztuk. Az algoritmus IBD szegmenseket keres a romák és a vizsgált regionális csoportok populációi között. Az algoritmust főként a roma leszármazás összetevőinek becslésére használtuk. A mintapáronkénti átlagos IBD részesedést a romák (I populáció) és az egyes regionális populációk (J populáció) között a Beagle futás kimeneti adataiból a következő képlettel számoltuk ki:

$$Average\ pairwise\ IBD\ sharing = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} IBD_{ij}}{n \cdot m}$$

ahol az IBD_{ij} az i és j egyének között megosztott IBD szegmens hossza, n és m az I. és J populációban lévő egyének száma.

Mivel a hosszabb IBD szegmensek nagyobb száma két adott populáció közötti későbbi keveredésre utal, megvizsgáltuk, hogy van-e kimutatható különbség a romákban található IBD szegmensek hosszának megoszlásában a romák migrációs útvonalán található regionális populációkra nézve. Kiszámítottuk a különböző hosszúságú IBD szegmensek átlagos megoszlását a romák és a regionális populációkból származó egyénpárok között a Beagle kimeneti adatainak a segítségével. Az IBD szegmenseket hossz szerint osztályoztuk, a szegmensek számát az egyes hosszúsági osztályokban összesítettük, és elosztottuk az összes lehetséges egyénpár számával.

Ahogy az IBD hosszanalízissel, az ALDER 1.03 algoritmus segítségével is megpróbáltuk a romák és az egyes regionális populációk közötti génáramlás időrendjét meghatározni. Továbbá az ALDER algoritmus adott populációk keveredésének időpontját is képes megbecsülni. Az elődjéhez, a ROLLOFF-hoz hasonlóan az ALDER szintén a keveredés okozta kapcsoltság időbeni csökkenésének jelenségén alapul. Az algoritmus egy kevert célpopuláció SNP-i közötti korrelációs értékeket számítja ki a kiindulási populációk allélfrekvencia-különbségeivel súlyozva. A kiindulási populációk (pontosabban a kiindulási populációk helyettesítésére alkalmas recens populációk) az algoritmusnak referenciapopulációként szolgálnak. Az eredményeket a háttérkapcsoltság nagymértékben befolyásolja; ezért az algoritmus a referenciapopulációk allélfrekvenciáit használja fel a keveredési kapcsoltság jelének felerősítésére, ami elősegíti a háttérkapcsoltság kiszűrését. Az ALDER továbbfejlesztett algoritmusa kifinomult súlyozott kapcsoltsági statisztikákkal rendelkezik, és képes arra is, hogy teljes mértékben kiiktassa a háttérkapcsoltság által előidézett torzított becsléseket. További jelentős előny, hogy az algoritmus képes magát a célpopulációt referenciaként alkalmazni, ami gyakorlatilag torzításmentes statisztikákat eredményez.

Eredmények

A populációstruktúra analízis és a leszármazásbecslési módszerek segítettek a romák eurázsiai kontextusba helyezésében, és megmutatták, hogy a vizsgált regionális populációk három nagy csoportot alkotnak különböző mértékben határozott klaszterekkel. Ezek a klaszterek az európaiak és a dél-ázsiaiak kissé laza csoportjai, valamint a közel-keleti és a kaukázusi szorosan csoportosuló népcsoportok voltak. A közép-ázsiaiakról és a dél-ázsiaiak néhány csoportjáról megállapítható, hogy viszonylag magas kelet-ázsiai leszármazásuk miatt külön csoportot alkotnak. Az eredmények tükrözték az egyes regionális csoportok tényleges földrajzi helyzetét. A roma mintákat e populációkon ábrázolva láthatjuk, hogy a romák erősen szétszóródnak Európa és Dél-Ázsia között, eurázsiai migrációjuk és alapvetően nomád jellegük miatt. A legtöbb roma egyén azonban Dél-Ázsia, a Kaukázus, a Közel-Kelet és Közép-Ázsia populációi között helyezkednek el egy szorosabb klasztert alkotva. A TreeMix megerősítette ezeket az eredményeket, a romákat a Közel-Kelet, a kaukázusi régió és Dél-Ázsia közé helyezte.

Az F_{st} számítások azt mutatták, hogy a romák legnagyobb mértékben a dél-ázsiai népességtől különböznek, és minél közelebb vannak a vizsgált populációk Európához, az F_{st} értéke annál inkább csökken. Ezek várható eredmények, és a kronológiai sorrendet tükrözik, amelyben a romák kapcsolatba léptek az egyes régiók lakosságával Északnyugat-Indiából Európa felé történő vándorlásuk során. A kaukázusi régiónál és Törökország populációinál azonban az F_{st} egy minimumát határoztuk meg, amely azt sugallja, hogy ez a régió fontos szerepet játszhat a romák leszármazásában. Tesztjeink továbbá azt is mutatják, hogy a romák még a szomszédos kaukázusi térséghez képest is figyelemre méltó török leszármazással rendelkezhetnek.

A populációstruktúra és a leszármazásbecslési vizsgálatok eredményei alapján feltételezett keverési eseményeket hivatalos keveredési tesztekkel is megvizsgáltuk, amelyek amellett, hogy a keveredések igazolásaként szolgálnak, meg képesek becsülni a keveredés arányát is. Erre a 4 populáció és az F_4-arány becslési teszteket alkalmaztuk. Ezek a tesztek megerősítették, hogy a Kaukázus, a Közel-Kelet és Közép-Ázsia népei valóban keveredtek a romák őseivel. Az F_4-arány becslés szerint továbbá az említett populációkból a romák leszármazási aránya jelentős. Az Admixture graph fitting eredményei szerint az általunk felállított roma leszármazási modell jól illeszkedik a rendelkezésre álló adatokra, ezzel megerősítve a romák és a kaukázusi régió populációinak feltételezett jelentős mértékű keveredését is. Az átlagos páronkénti IBD részesedés becslési eredmények azt mutatják, hogy a Kaukázus régió populációjának valamivel magasabb a romákkal közös IBD aránya, mint a közel-keleti és közép-ázsiai populációk esetén, és ez hasonló mértékű a dél-ázsiai populációk arányához. IBD hosszeloszlásra irányuló elemzéseink összhangban vannak az átlagos IBD részesedés adataival. Az eredmények azt mutatták, hogy a romáknak a Kaukázus régióval nagyobb a közös hosszú IBD szegmensek száma, mint

a közel-keleti és a közép-ázsiai népekkel. Ez arra is utalhat, hogy a Kaukázus populációinak keveredése a romákkal későbbi, és nagyobb mértékű.

Az oszmán hódítás kelet-közép-európai hatására vonatkozó vizsgálataink során igazolást nyert, hogy a romák és a törökök keveredtek egymással. Az oszmán hódoltság hatásának felmérése során a kelet-közép-európai populációk törökkel való keveredését is megvizsgáltuk a 4 populáció teszt alkalmazásával, amelynek segítségével közelebb kerülhetünk annak igazolásához, hogy Kelet-Közép-Európa rendelkezik az oszmán megszállás idejéből eredeztethető török leszármazással. Átlagos IBD részesedés becsléssel azt is megerősítettük, hogy ez a keveredés Kelet-Közép-Európa egykori oszmán megszállásából származhat. A romákban azonosított török leszármazás nagyobb, mint a törökökkel szomszédos területek populációitól eredeztethető leszármazása, illetve a kelet-közép-európaiak és a törökök átlagos közös IBD részesedésének becslése azt is megmutatta, hogy a kelet-közép-európai populációk átlagos közös IBD részesedése magasabb a törökökkel, mint más korábban oszmánok által megszállt régiók populációi esetén (Kaukázus, Közel-Kelet országai). A korábban oszmánok által megszállt területek populációi és a romák közötti átlagos IBD részesedési különbségek jelentőségének felmérése érdekében kiszámítottuk a törökök átlagos IBD részesedését a szardíniaiakkal is, amely csoport Európa többi részétől elkülönülve található Szardínia szigetén, ezért az oszmán hódoltság okozta demográfiai események kevésbé hatottak rájuk. A vizsgálat megerősítette, hogy Kelet-Közép-Európa oszmán megszállásának hatása kimutatható.

Hogy megbecsüljük a romák őseinek a kaukázusi régió, a Közel-Kelet, Közép- és Dél-Ázsia populációival való keveredésének közelítőleges időpontjait, az ALDER segítségével megvizsgáltuk az adatokban található keveredési kapcsoltságot. Az eredmények a vártnak megfelelően azt mutatják, hogy a romáknak a dél-ázsiaiakkal való keveredése a romák migrációjának legrégebbi keveredési eseménye. Az ALDER hasonló időpontra helyezte a romák őseinek a kaukázusi és közel-keleti népekkel való keveredését, de a kaukázusi csoportokkal való keveredés valamivel későbbi.

A romák és a kelet-közép európaiak törökökkel való keveredésének további megerősítése céljából szintén alkalmaztuk az ALDER-t. A romák esetében kapott keveredési dátum a közel-keleti populációkkal való feltételezett többszörös keveredés miatt kissé torzult, és a keveredést egy valamivel korábbi időpontra tette. A kelet-közép-európaiak esetén kapott keveredési időpont intervalluma megfelel annak az időintervallumnak, amikor az oszmánok Kelet-Közép-Európát megszállás alatt tartották. Az ALDER megerősítette a feltételezéseinket, miszerint a kelet-közép-európaiak török eredetű leszármazásának egy része a 150 éves oszmán jelenlétből származhat, és a romák török eredetű leszármazásának egy része szintén eredeztethető az oszmánok Európában való tartózkodásának idejéből.

Diszkusszió

A teljes genomra kiterjedő autoszomális marker adatok alkalmazásával meg tudtuk becsülni a Kaukázus hozzájárulását a romák leszármazásához, amely jelentősnek bizonyult leszármazásuk két fő forrásához, a dél-ázsiai és európai leszármazáshoz képest is. Vizsgálataink azt mutatják, hogy a kaukázusi régió lehet a roma származás harmadik legfontosabb forrása, figyelembe véve a migrációs útvonalat, amely magában foglalta a közép-ázsiai és a közel-keleti régiókat is. Eredményeink arra utalnak, hogy a Kaszpi-tenger és a Fekete-tenger területének jelentős a szerepe a romák genetikai örökségében, a régió fontos szerepet játszik mind migrációjukban, mind leszármazásukban.

Megerősítettük azt is, hogy az Oszmán Birodalom terjeszkedése Kelet-Közép-Európába rányomta bélyegét a helyi lakosságra, detektálható mértékben hozzájárulva azok török leszármazásához. A korábbi Y-kromoszóma markereken alapuló vizsgálatok szerint az R1b haplocsoport a nyugat-európaiak elsődleges jellemzője, és ennek a haplocsoportnak a jelentősége erősen csökken Kelet-Európa felé, ahol az R1a haplocsoport a domináns. A kelet-közép-európai etnikai csoportokban megközelítőleg azonos mértékben oszlanak meg az R1a és az R1b csoportok. Jelen kutatás teljes genomra kiterjedő autoszomális marker adatok vizsgálati eredményei ezekkel az eredményekkel egybehangoznak, kutatásunk az Y-haplocsoportokon alapuló vizsgálatok eredményeihez hasonló eredményekhez vezetett az európai populációk genetikai felépítésének vizsgálata során. A leszármazásbecslés, a keveredésvizsgálatok és az IBD-analízis elkülönítette az oszmán megszállás alatt állt közép-kelet-európai csoportokat Európa többi részétől, és megmutatta, hogy az autoszomális adatokban a közel-keleti térségből eredeztethető leszármazás is megfigyelhető. A szakirodalom szerint a kelet-közép-európaiak esetén jelentős az E3b és J haplocsoportok jelenléte, amely elsősorban a közel-keleti régióban elterjedt. Eredményeink megfelelnek ennek a megfigyelésnek, a közel-keleti leszármazás az oszmánok által megszállt európai területek populációjában mutatkozik a legnagyobb mértékben a teljes genomra kiterjedő autoszomális marker adatok vizsgálata esetén is. Fény derült továbbá arra is, hogy a kelet-közép-európai térségben élő romák nem csak az Európába irányuló vándorlásuk során, hanem az oszmánok európai jelenléte idején is szert tehettek török leszármazásra.

Összegzés

- IBD analízis alkalmazásával sikeresen megbecsültük az egyes régiók, Dél-Ázsia, a Közel-Kelet (és Közép-Ázsia), a Kaukázus és Európa populációinak hozzájárulásának arányát a romák leszármazásához, tehát sikeresen megvizsgáltuk e régiók roma migrációban betöltött szerepének jelentőségét.

- Az f-statisztikák segítségével megerősítettük, hogy a romák kaukázusi leszármazása kimutatható, jelentős, továbbá különböző statisztikai megközelítésekkel jellemzésre kerültek.

- Az f-statisztikák és a keveredési kapcsoltságon alapuló módszerek segítségével további megerősítésre és jellemzésre kerültek a romák vándorlása során az egyes regionális populációkkal történt keveredési események, ezáltal populációgenetikai megközelítéssel is megerősítettük a romák a fő migrációs útjának koncepcióját.

- Populációstruktúra vizsgálatokat, leszármazásbecslést, keveredési vizsgálatokat, IBD és kapcsoltságon alapuló elemzéseket alkalmazva megállapítottuk, hogy az oszmán megszállás a kelet-közép-európai populációkra és romákra gyakorolt genetikai hatása kimutatható, ami egyben azt is jelenti, hogy a Közel-Kelet a roma migrációban még kisebb szerepet játszik, mint arra korábbi becsléseink utaltak, kiemelve a szomszédos Kaukázus régió jelentőségét a roma leszármazásban és a roma migrációban.

A tézis alapjául szolgáló tudományos publikációk:

1. **Revealing the impact of the Caucasus region on the genetic legacy of Romani people from genome-wide data**
   Z Bánfai, V Ádám, E Pöstyéni, G Büki, M Czakó, A Miseta, B Melegh.
   PLoS One. 2018 Sep 10;13(9):e0202890. doi: 10.1371/journal.pone.0202890

   Impakt faktor: 2.776

2. **Revealing the genetic impact of the Ottoman occupation on ethnic groups of East-Central Europe and on the Roma population of the area**
   Z Bánfai, BI Melegh, K Sumegi, K Hadzsiev, A Miseta, M Kásler, B Melegh
   Front Genet. 2019 Jun 13;10:558. doi: 10.3389/fgene.2019.00558

   Impakt faktor: 3.517

Egyéb publikációk (10)

Impakt faktor: ∑ 55.57

Összesített impakt faktor: 61.863

Köszönetnyilvánítás

Hálával tartozom témavezetőmnek, Melegh Béla Professzor Úrnak, aki felkeltette érdeklődésemet a humángenetika és a humán populációgenetika iránt, és lehetővé tette számomra, hogy az Interdiszciplináris Orvostudományok Doktori Iskolához Ph.D. hallgatóként csatlakozzak, illetve szakmai tevékenységemet végigkísérte, kutatói munkámat irányította, és segítette.

Külön köszönet az Orvosi Genetikai Intézet minden munkatársának és Ph.D. hallgatójának, különösen Sümegi Katalinnak, Szabó Andrásnak, Büki Gergelynek, Maász Anitának.

Hálás vagyok az Orvosi Genetikai Intézet minden asszisztensének, akik hozzáértő, lelkiismeretes munkájukkal és szakmai tapasztalatukkal segítették munkámat.

Végül hálával tartozom a családomnak, akik végtelen és megértő türelmükkel és támogatásukkal lehetővé tették ezt a munkát.