

VILNIAUS UNIVERSITETAS

Alina  
URNIKYTĖ

Lietuvos populiacijos genetinės  
struktūros ir evoliucinių veiksnių  
analizė, remiantis plataus masto  
genotipavimo duomenimis:  
praeitis ir dabartis

**DAKTARO DISERTACIJA**

Biomedicinos mokslai,  
medicina (06B)

---

VILNIUS 2018

Disertacija rengta 2014–2018 metais Vilniaus universiteto Medicinos fakulteto Biomedicinos mokslų instituto Žmogaus ir medicininės genetikos katedroje. Mokslinius tyrimus rėmė Lietuvos mokslo taryba, doktorantūra buvo finansuojama ES struktūrinių fondų lėšomis, 2018 m. buvo gauta stipendija už akademinis pasiekimus.

**Mokslinis vadovas**

**akad. prof. habil. dr. Vaidutis Kučinskas** (Vilniaus universitetas, biomedicinos mokslai, biologija – 01 B, medicina – 06 B)

VILNIUS UNIVERSITY

Alina  
URNIKYTĖ

Evaluation of Genetic Structure  
and Evolutionary Forces of the  
Lithuanian Population According  
to the High-Density Genotyping Data:  
Past and the Future

**DOCTORAL DISSERTATION**

Biomedical Sciences,  
Medicine (06 B)

---

VILNIUS 2018

This dissertation was written between 2014 and 2018 at the Department of Human and Medical Genetics, Biomedical Sciences Institute, Faculty of Medicine, Vilnius University. The research was supported Research Council of Lithuania, the doctoral studies were financed from the EU structural funds, and a scholarship was granted for academic accomplishments in 2018.

**Academic supervisor**

**Acad. Prof. Habil. Dr. Vaidutis Kučinskas** (Vilnius University, Biomedical Sciences, Biology 01 B, Medicine 06 B)

## TURINYS

SANTRUMPŲ SĄRAŠAS .....	7
ĮVADAS .....	9
1. LITERATŪROS APŽVALGA.....	15
1.1. Lietuva priešistoriniais laikais ir lietuvių kilmė.....	15
1.2. Populiacijos genetinė struktūra ir žmogaus genomo įvairovė.....	17
1.3. Gamtinė atranka ir žmogaus genetinė įvairovė.....	20
1.3.1. Genų dreifo teorija .....	20
1.3.2. Efektyvus populiacijos dydis .....	21
1.3.3. Teigiama gamtinė atranka.....	26
1.3.4. Gamtinės atrankos nustatymo metodai .....	30
1.3.4.1. Metodai, paremti haplotipų dažniu ir įvairove.....	30
1.3.4.2. Alelių dažnių spektru paremti metodai.....	30
1.3.4.3. Populiacijos diferenciacija paremti metodai.....	31
2. DARBO METODOLOGIJA.....	33
2.1. Tyrimo eiga.....	33
2.2. Tiriamieji .....	33
2.3. Genotipavimas .....	35
2.4. Duomenų paruošimas.....	36
2.4.1. Duomenų paruošimas populiacijos genetinės struktūros analizei .....	36
2.4.2. Duomenų paruošimas teigiamai gamtinei atrankai nustatyti.....	37
2.4.3. Duomenų paruošimas efektyviam populiacijos dydžiui nustatyti.....	38
2.5. Duomenų analizė .....	38
2.5.1. Populiacijos genetinės struktūros analizė .....	38
2.5.2. Kandidatinių teigiamos gamtinės atrankos veikiamų genomo sričių nustatymas.....	39
2.5.3. Ilgalaikio efektyvaus populiacijos dydžio ir divergencijos analizė.....	39
2.5.4. Trumpalaikio efektyvaus populiacijos dydžio analizė.....	41

3. TYRIMO REZULTATAI .....	42
3.1. Lietuvos populiacijos genetinės struktūros nustatymas ir įvertinimas kitų populiacijų atžvilgiu pagal plataus masto genotipavimo duomenis .....	42
3.2. Kandidatinių teigiamos gamtinės atrankos veikiamų genomo sričių nustatymas Lietuvos populiacijoje .....	49
3.3. Lietuvos ilgalaikio efektyvaus populiacijos dydžio, divergencijos laiko ir evoliucinių ryšių su kitomis pasaulio populiacijomis analizė .....	59
3.4. Praeities įvykių tarp Lietuvos populiacijos etnolingvistinių grupių rekonstravimas, remiantis efektyviu populiacijos dydžiu ir divergencijos analize.....	65
3.5. Lietuvos trumpalaikio efektyvaus populiacijos dydžio nustatymas bei $N_e$ ir $N$ santykio įvertinimas.....	69
3.6. Matematinų metodų lyginamoji analizė efektyviam populiacijos dydžiui nustatyti .....	73
4. REZULTATŲ APIBENDRINIMAS .....	76
IŠVADOS.....	82
LITERATŪROS SĄRAŠAS.....	83
PADĖKA.....	92
PRIEDAI .....	93
I PRIEDAS. Genominės DNR išskyrimo ir DNR koncentracijos bei švarumo nustatymo protokolai .....	93
II PRIEDAS. Genominės DNR genotipavimo <i>Illumina</i> VNP lustais protokolai .....	99
III PRIEDAS. Analizės metodai .....	120
IV PRIEDAS. Papildoma rezultatų medžiaga.....	124

## SANTRUMPŲ SĄRAŠAS

ASW – Afrikos kilmės pietvakarių JAV gyventojai (angl. *African Ancestry in Southwest US*)

AUS – australai iš Australijos

BEB – bengalai iš Bangladešo (angl. *Bengali from Bangladesh*)

CDX – dajai iš Sišuangbanos, Kinijos (angl. *Chinese Dai in Xishuangbanna, China*)

CEU – Šiaurės ar Vakarų Europos kilmės Jutos gyventojai (angl. *Utah Residents with Northern and Western European Ancestry*)

CHB – kinai iš Pekino miesto, Kinijos (angl. *Han Chinese in Beijing, China*)

CHD – kinai iš Denverio miesto, Kolorado (angl. *Chinese in Metropolitan Denver, Colorado*)

CHS – kinai iš Pietų Kinijos (angl. *Southern Han Chinese, China*)

CV – kryžminė duomenų kontrolė (angl. *Cross validation, CV*)

DEN – danai iš Danijos

DNR – deoksiribonukleorūgštis (angl. *DNA, Deoxyribonucleic Acid*)

ESN – esan gyventojai iš Nigerijos (angl. *Esan in Nigeria*)

FIN – suomiaai iš Suomijos (angl. *Finnish in Finland*)

$F_{ST}$  – F statistika

GBR – britai iš Anglijos ir Škotijos (angl. *British in England and Scotland*)

GIH – indų gudžaratų kilmės Hiustono gyventojai, Teksasas (angl. *Gujarati Indians in Houston, Texas*)

GWD – Vakarų Gambijos gyventojai (angl. *Gambian in Western Divisions in the Gambia*)

Ilgalaikis  $N_e$  – ilgalaikis efektyvus populiacijos dydis (angl. *Long-term effective population size*)

ITU – indų telugų kilmės Jungtinės Karalystės gyventojai (angl. *Indian Telugu from the UK*)

YRI – jorubai iš Ibadano, Nigerijos (angl. *Yoruba in Ibadan, Nigeria*)

JPT – japonai iš Tokijo, Japonijos (angl. *Japanese in Tokyo, Japan*)

K – protėvinių populiacijų skaičius (angl. *Number of ancestral populations*)

KHV – kinh tauta iš Hošimino miesto, Vietnamo (angl. *Kinh in Ho Chi Minh City, Vietnam*)

LD – nepusiausvyra sankiba (angl. *Linkage Disequilibrium*)

LWK – luchjai iš Vebujė, Kenijos (angl. *Luhya in Webuye, Kenya*)

MEX – meksikiečių kilmės Los Andželo gyventojai (angl. *Mexican Ancestry in Los Angeles, California*)

MKK – *Kinyawa* masajų populiacija iš Kenijos (angl. *Maasai in Kinyawa, Kenya*)

MSL – mendė iš Siera Leonės (angl. *Mende in Sierra Leone*)

mtDNR – mitochondrinė DNR

N – bendras individų skaičius populiacijoje (angl. *Census population size*)

*Ne* – efektyvus populiacijos dydis (angl. *Effective population size*)

NET – olandai iš Olandijos

PJL – pandžabiai iš Lahoro, Pakistano (angl. *Punjabi from Lahore, Pakistan*)

STU – Šri Lankos tamilų kilmės Jungtinės Karalystės gyventojai (angl. *Sri Lankan Tamil from the UK*)

PK – pagrindinė komponentė (angl. *Principal Component, PC*)

PKA – pagrindinių komponentių analizė (angl. *Principal Component Analysis, PCA*)

pr. m. e. – prieš mūsų erą

SWE – švedai iš Švedijos

Trumpalaikis *Ne* – trumpalaikis efektyvus populiacijos dydis (angl. *Recent effective population size*)

TSI – italai iš Toskanos regiono, Italijos (angl. *Toscani in Italia*)

UK – britai iš Jungtinės Karalystės

VNP – vieno nukleotido polimorfizmas (angl. *Single Nucleotide Polymorphism, SNP*)

VU MF BMI ŽMGK – Vilniaus universiteto Medicinos fakulteto Biomedicinos mokslų instituto Žmogaus ir medicininės genetikos katedra



## ĮVADAS

Dabartinė Lietuvos populiacija yra sudaryta iš ankstesnių baltų genčių atstovų. Tikėtina, kad dabartinės Lietuvos gyventojai išsaugojo senųjų protėvių genų fondą. Genomo įvairovės tyrimai palengvino žmogaus evoliucinės genetikos tyrimus. Kiekvienos populiacijos genetinė įvairovė – tai prisitaikymo prie aplinkos pokyčių garantas. Tiriant DNR sekos įvairovę, galima rekonstruoti žmogaus populiacijos evoliucinę istoriją, kilmę, struktūrą ar nustatyti evoliucijos veiksnius: gamtinę atranką, efektyvųjį populiacijos dydį ir divergencijos laiką tarp populiacijų. Demografiniai pokyčiai populiacijoje palieka pėdsaką, keisdami populiacijos genetinę įvairovę. Izoliuotos populiacijos išsaugo savo genetinį unikalumą, o migracija mažina genetinius populiacijų skirtumus. Taigi mūsų biologinė istorija yra užkoduota DNR sekoje.

Šiame darbe, taikant plataus masto vieno nukleotido polimorfizmą (VNP, angl. *SNP, Single Nucleotide Polymorphism*) genotipavimo duomenis, detaliai charakterizuota Lietuvos populiacijos vidinė genetinė struktūra ir nustatyti evoliucijos veiksniai: efektyvus populiacijos dydis, divergencijos laikas ir teigiama gamtinė atranka bei jos veikiamos specifinės genomo sritys Lietuvos populiacijai.

Darbas yra aktualus, kadangi tokie tyrimai ne tik papildo fundamentaliojo mokslo žinias, bet ir padeda geriau suprasti egzistuojančius ryšius tarp gamtinės atrankos ir ligų bei evoliucinius mechanizmus, dėl kurių polinkis sirgti tam tikromis ligomis yra skirtingas tiek individų, tiek populiacijos lygmeniu.

**Naujumas.** Šiame darbe, taikant plataus masto VNP genotipavimo duomenis ir ištyrus pakankamą skaičių asmenų, nustatyta detali Lietuvos populiacijos vidinė genetinė struktūra ir padėtis kitų Europos ir pasaulio populiacijų atžvilgiu. Darbo metu pirmą kartą nustatytos specifinės lietuvių populiacijai kandidatinių teigiamos gamtinės atrankos veikiamos genomo sritys, panaudojant plataus masto VNP genotipavimo duomenis, ir atlikta Lietuvos ilgalaikio ir trumpalaikio efektyvaus populiacijos dydžio, divergencijos laiko ir evoliucinių ryšių su kitomis pasaulio populiacijomis analizė. Remiantis nustatytu efektyviu populiacijos dydžiu ir divergencijos laiku, rekonstruoti praeities įvykiai tarp Lietuvos populiacijos etnolingvistinių grupių.

**Aktualumas.** Pirminiai Lietuvos populiacijos struktūros tyrimai buvo pagrįsti Y chromosomos, mikrosatelitų ir mitochondrinės DNR (mtDNR) duomenimis [1–3]. Šiuo metu metodų, skirtų populiacijos genetinių

duomenų gavybai, alternatyva yra naujos kartos sekoskaita, kuri išsprendžia daugybę ankstesnių metodikų apribojimų ir suteikia galimybę nustatyti naujus ir geriau suprasti jau žinomus mikroevoliucijos procesų mechanizmus, vykdant specifinių kohortų genomo variantų analizę. Darbo metu gauti rezultatai apie Lietuvos populiacijos genetinę struktūrą papildo jau turimas žinias ir suteikia galimybę atlikti palyginimo analizę bei daryti tvirtas išvadas. Šiame darbe, panaudojus Lietuvos genetinius duomenis kartu su 1 000 genomų projekto duomenimis, buvo nustatyta Lietuvos populiacijos genetinė struktūra kitų pasaulio populiacijų atžvilgiu, o tai leido įvertinti ne tik egzistuojančius skirtumus tarp populiacijų, bet ir mūsų populiacijos genetinį unikalumą [4]. Detali populiacijos genetinės struktūros analizė yra svarbi asociacijų tyrimuose, tiriant populiacijų praeitį, identifikuojant vietinės adaptacijos veikiamas genomo sritis bei priskiriant kilmę individo lygmeniu.

Gamtinė atranka yra pagrindinis evoliucijos veiksnys, galintis keisti genomo struktūrą. Mikroevoliuciniai ir aplinkos veiksniai lemia populiacijos gebėjimą prisitaikyti prie kintančių ir skirtingų aplinkos sąlygų. Teigiamos gamtinės atrankos nustatymas žmonių populiacijoje suteikia galimybę suprasti šiuolaikinių žmonių fenotipus ir gebėjimą prisitaikyti prie kintančių aplinkos sąlygų. Darbo metu nustatytos specifinės lietuvių populiacijai kandidatinės gamtinės atrankos veikiamos genomo sritys ne tik užpildo šioje srityje esančias žinių spragas, bet ir gali padėti nustatyti jų poveikį fenotipui bei identifikuoti adaptyvius alelius.

Nustatytas efektyvus populiacijos dydis (angl. *Effective Population Size, Ne*) gali padėti prognozuoti neutralios genetinės įvairovės praradimą ir pasiskirstymą, naudingų ir žalingų alelių fiksacijos tikimybę bei mažos populiacijos tinkamumą ir išgyvenamumą.

**Tęstinumas.** Didesnės skiriamosios gebos genetinės struktūros nustatymas, pavyzdžiui, *fineSTRUCTURE* algoritmu ir panaudojant viso genomo sekoskaitos duomenis, suteiktų galimybę išsamiau įvertinti Lietuvos populiacijos genetinę struktūrą tiek jos viduje, tiek kitų populiacijų atžvilgiu. Nustatytos specifinės lietuvių populiacijai kandidatinės teigiamos gamtinės atrankos veikiamos genomo sritys turėtų būti patvirtintos funkciniais metodais.

## Darbo tikslas

Lietuvos populiacijos genetinės struktūros ir evoliucinių veiksnių: efektyvaus populiacijos dydžio ir gamtinės atrankos, įvertinimas, taikant plataus masto vieno nukleotido polimorfizmo genotipavimo duomenis.

## Uždaviniai

1. Nustatyti ir įvertinti Lietuvos populiacijos genetinę struktūrą kitų populiacijų atžvilgiu, taikant plataus masto vieno nukleotido polimorfizmo genotipavimo duomenis.
2. Nustatyti specifines lietuvių populiacijai kandidatines teigiamos gamtinės atrankos veikiamas genomo sritis, panaudojant plataus masto vieno nukleotido polimorfizmo genotipavimo duomenis.
3. Atlikti Lietuvos ilgalaikio efektyvaus populiacijos dydžio, divergencijos laiko ir evoliucinių ryšių su kitomis pasaulio populiacijomis analizę, taikant plataus masto vieno nukleotido polimorfizmo genotipavimo duomenis.
4. Remiantis nustatytu ilgalaikiu efektyviu populiacijos dydžiu ir divergencijos laiku, rekonstruoti praeities įvykius tarp Lietuvos populiacijos etnolingvistinių grupių.
5. Nustatyti Lietuvos trumpalaikį efektyvų populiacijos dydį ir įvertinti efektyvaus populiacijos dydžio bei bendro individų skaičiaus populiacijoje santykį.

## Ginamieji teiginiai

1. Lietuvos populiacija yra homogeniška ir galimai pasižyminti specifinėmis genominėmis ypatybėmis kitų tirtų Europos populiacijų atžvilgiu.
2. Lietuvos populiacijai būdingos specifinės gamtinės atrankos veikiamos genomo sritys.
3. Lietuvos efektyvus populiacijos dydis ( $N_e$ ) yra mažesnis, palyginti su kitomis Europos populiacijomis.
4. Galimi aukštaičių ir žemaičių genetiniai skirtumai turi įtakos statistiškai reikšmingiems efektyvaus populiacijos dydžio ( $N_e$ ) skirtumams tarp grupių.
5. Lietuvos populiacija yra nedidelė, istorijoje patyrusi butelio kaklelio efekto padarinius ir ekspansiją, o tai galimai lemia mažą efektyvaus populiacijos dydžio ir bendro individų skaičiaus populiacijoje santykį ( $N_e/N$ ).

## Darbo aprobacija

Darbo rezultatai paskelbti Lietuvos ir užsienio leidiniuose, pristatyti mokslinėse konferencijose.

Disertacijos tema paskelbtų straipsnių sąrašas:

1. **Urnikytė A**, Molytė A, Kučinskas V. Recent effective population size estimated from segments of identity by descent in the Lithuanian population. *Anthropological Science* 2017; 125(2): 53–58. DOI: 10.1537/ase.170125.
2. Molyte A, **Urnikyte A**, Kučinskas V. A comparative analysis of mathematical methods for effective population size estimation. *Lietuvos matematikos rinkinys* 2016; Ser. A, 57: 53–58. ISSN 0132-2818.
3. **Urnikyte A**, Domarkiene I, Stoma S, Ambrozaityte L, Uktveryte I, Meskiene R, Kasiulevičius V, Burokiene N, Kučinskas V. CNV analysis in the Lithuanian population. *BMC Genetics* 2016, 4; 17(1): 64. DOI:10.1186/s12863-016-0373-6.

Stendiniai pranešimai:

1. **Urnikyte A**, Molyte A, Kučinskas V. Relationship between effective population size and inbreeding in the Lithuanian population. *The European Human Genetics Conference*, 2018 June 16–19, Milan, Italy. Abstract no. PS18.21, p. 332.
2. **Urnikytė A**, Molytė A, Kučinskas V, Kučinskienė ZA. Recent changes in contemporary effective population size from identical by descent segments. *American Society of Human Genetics*, 2017 October 17–21, Orlando, Florida. Abstract no. 2315T, p. 976.
3. **Urnikytė A**, Molytė A, Kučinskas V. Recent effective population size estimated from segments of identity by descent in the Lithuanian population. *The European Human Genetics Conference*, 2017 May 27–30, Copenhagen, Denmark. E-P18.14.
4. **Urnikytė A**, Molytė A, Pranckevičienė E, Kučinskas V. Inference of evolutionary relationships among human populations based on the estimates of effective population size. *American Society of Human Genetics*, 2016 October 18–22, Vancouver, Canada. Poster abstracts. Vancouver : ASHG, 2016, p. 491.
5. **Urnikytė A**, Molytė A, Kučinskas V. Recent Effective Population Size Estimated from Segments of Identity by Descent in the Lithuanian

Population. *The 3rd International Conference Evolutionary Medicine: Pre-Existing Mechanisms and Patterns of Current Health Issues*, 14th–19th of June, 2016, Vilnius University, Lithuania. P. 77, ISBN 978-609-459-720-6.

6. **Urnikytė A**, Molytė A, Pranckevičienė E, Kučinskas V. Demographic inference of the Lithuanian population. *The European Human Genetics Conference*, 2016 May 21–24, Barcelona, Spain. *European journal of human genetics*. London: Nature Publishing Group, 2016, 24(e-suppl. 1): 388. ISSN: 1018-4813.
7. **Urnikytė A**, Domarkienė I, Uktverytė I, Ambrozaitė L, Meskienė R, Kučinskas V. Genomic diversity and distribution of CNVs in Lithuanian population. *European Human Genetics Conference*, 2015 June 6–9, Glasgow, Scotland, United Kingdom. *European journal of human genetics*. London: Nature Publishing Group, 2015, 23(1): 332–333.

#### Žodiniai pranešimai:

1. Molytė A (pranešėja), **Urnikytė A**, Kučinskas V. A comparative Analysis of Mathematical Methods for homogeneity estimation of the Lithuanian population. Pristatyti tyrimo rezultatai Lietuvos matematikų draugijos konferencijoje 2018 m. birželio 18–19 d. Kaune.
2. **Urnikyte A** (pranešėja), Mondal M, Bosch E, Molyte A, Kučinskas V. Detecting signatures of adaptive positive selection from high-density genotyping data in the Lithuanian population. Pristatyti tyrimo rezultatai „56th Polish and 14th International Conference Juvenes Pro Medicina“, 2018 m. gegužės 25–26 d. Lodzėje, Lenkijoje. Gautas III vietos apdovanojimas (PhD sekcija).
3. Molyte A (pranešėja), **Urnikyte A**, Kučinskas V. A Comparative Analysis Of Effective Population Size In Six Ethnolinguistic Groups Of The Lithuania Population. Pristatyti tyrimo rezultatai „XIV Baltic Congress of Laboratory Medicine“, 2018 m. gegužės 10–12 d. Vilniuje.
4. Molytė A, **Urnikytė A** (pranešėja), Kučinskas V. Efektyvaus populiacijos dydžio lyginamoji analizė tarp Lietuvos etnolingvistinių grupių. Pristatyti tyrimo rezultatai Lietuvos matematikos draugijos 58-ojoje konferencijoje, 2017 m. birželio 21–22 d. Vilniuje.
5. **Urnikytė A** (pranešėja). Lietuvos populiacijos evoliucinių ryšių analizė. Pristatyti tyrimo rezultatai Lietuvos mokslų akademijos Biologijos, medicinos ir geomokslų skyriaus jaunųjų mokslininkų konferencijoje „BIOATEITIS: gamtos ir gyvybės mokslų perspektyvos“, 2016 m.

- gruodžio 7 d. Gautas III vietos apdovanojimas (medicinos sekcija).
6. Molytė A, **Urnikytė A** (pranešėja), Kučinskas V. A comparative analysis of mathematical methods for effective population size estimation. Pristatyti tyrimo rezultatai Lietuvos matematikų draugijos LVII konferencijoje 2016 m. birželio 20–21 d. Vilniuje, Vilniaus Gedimino technikos universitete.
  7. **Urnikytė A** (pranešėja), Molytė A, Kučinskas V. Recent Effective Population Size Estimated from Segments of Identity by Descent in the Lithuanian Population. Pristatyti tyrimo rezultatai „The 3rd International Conference Evolutionary Medicine: Pre-Existing Mechanisms and Patterns of Current Health Issues“, 2016 m. birželio 14–19 d. Vilniuje, Vilniaus universitete. P. 77. ISBN 978-609-459-720-6.
  8. **Urnikytė A** (pranešėja). DNR kopijų skaičiaus pokyčių įvairovės ir pasiskirstymo analizė Lietuvos populiacijoje. Pristatyti tyrimo rezultatai konferencijoje „LITGEN: Lietuvos populiacijos genetinė įvairovė ir sandaros kitimai, susiję su evoliucija ir dažniausiai paplitusiomis ligomis“, 2015 m. kovo 6 d. Vilniuje.

Stažuotė:

*Pompeu Fabra* universiteto Eksperimentinių ir sveikatos mokslų departamento Evoliucinės populiacijų genetikos laboratorija (angl. *Evolutionary Population Genetics Lab of the Department of Experimental and Health Sciences at the Universitat Pompeu Fabra*), Barselona, Ispanija, nuo 2017 m. lapkričio 2 d. iki 2018 m. gegužės 11 d.

# 1. LITERATŪROS APŽVALGA

## 1.1. Lietuva priešistoriniais laikais ir lietuvių kilmė

Dabartinė Lietuvos populiacija yra sudaryta iš ankstesnių baltų genčių atstovų. Tikėtina, kad dabartinės Lietuvos gyventojai išsaugojo senųjų protėvių genų fondą. Lietuvą, kaip ir visą Pabaltijį, bent tris kartus buvo padengę ledynai. Paskutinis apledėjimas ėmė trauktis apie XIV tūkstantmetį pr. Kr., tačiau iš Lietuvos teritorijos jam išslinkti prirėikė apie 3 000 m. [5]. Pirmieji žmogaus pėdsakai pietų Pabaltijyje randami vėlyvojo paleolito periodu [5, 6]. Tai buvo medžiotojai, atėję iš Vidurio ir Vakarų Europos, kurie atsinešė dvi kultūras – Svidrų ir Pabaltijo Madeleino. Svidrų kultūros pėdsakų pietų Lietuvoje išliko iki neolito pradžios [5]. Pabaltijo Madeleino kultūra plito iš Vakarų Europos per Daniją ir Vokietiją. Šiaurės Europą pasiekė vėliausios jos atmainos. Daugiau duomenų yra apie žmogaus gyvenimą mezolito laikotarpiu: nuo 10 000 iki 4 000 metų pr. Kr., kai klimatas atšilo, gamta teikė pakankamai gėrybių, o ežerai – žuvų ir vandens paukščių [6]. Lemiamą reikšmę pietų ir rytų Pabaltijo mezolito kultūrai turėjo iš pietvakarių einanti Maglemosės tipo kultūra. Rytų Pabaltijyje ji sudarė sąlygas dviem vietinėms kultūroms susidaryti: mezolitinei Nemuno ir Kundos kultūrai, kurios peraugo į neolitinės. Atlikti Aukštumalos aukštapelkės archeologiniai tyrimai 2004 m. byloja apie pirmųjų gyventojų Lietuvos teritorijoje pasirodymą ir gyvenseną vėlyvojo ledynmečio ir ankstyvojo holoceno laikotarpiais [7]. Neolito laikotarpiu, kuris truko nuo 4 000 iki 2 000 m. pr. Kr., keičiasi žmonių gyvenimo būdas iš klajoklių į sėslų gyvenimą ir pradeda plėtotis primityvi žemdirbystė [6, 8]. Ankstyvojo neolito metu Lietuvoje paplito dvi kultūros: neolitinė Nemuno ir Narvos kultūra. Iš Narvos kultūros gyvenviečių ypač pažymėtina Šventoji prie Baltijos jūros, kuri laikoma svarbiausia šio laikotarpio Lietuvos gyvenvieta. Narvos kultūra savo šaknimis siekia Vakarų Europą, o Lietuvoje svarbiausios Nemuno kultūros gyvenvietės yra Varėnos rajone. Vidurinio neolito laikotarpiu atsiranda ryšių su dviem nevietinėmis neolitinėmis kultūromis – šukinės-duobelinės keramikos, ateiviai iš rytų (finougrų protėviai), ir rutulinių amforų kultūros, kuri Lietuvą užliejo tik iš dalies [5, 8]. Vėlyvojo neolito laikotarpiu prasideda virvelinės keramikos kultūra, kurią atnešė ateiviai iš pietvakarių, rytų Pabaltijyje pasirodę III tūkstantmečio pr. m. e. viduryje. Virvelinės keramikos kultūra yra viena pagrindinių indoeuropiečių bendrijai priklausiusių kultūrų, todėl su ja Pabaltijyje yra susietas ir baltų kilmės klausimas. R. Rimantienės teigimu, į Šiaurės Europą indoeuropiečiai atėję nuo Juodosios jūros su antrąja

pastoralistų (gyvulių augintojų) banga [9]. Lietuvoje pirmieji indoeuropiečiai pasirodė vakarų Lietuvoje su rutulinių amforų kultūra apie 3 000 m. pr. m. e. [9].

Taigi baltų tautų susiformavimą lėmė kelios skirtingos kultūrinės sistemos: priešindoeuropinė, arba senosios Europos, atkeliavusi iš Vakarų Europos, dabartinės Prancūzijos srities, ir indoeuropietiškoji, išsirutuliojusi iš Eurazijos stepių (kurganų) kultūros [6]. Indoeuropiečių kalbą ir visuomeninę patriarchalinę, patrilinarinę ir trijų klasių struktūrą į rytų Pabaltijį, Baltarusiją ir vidurio Rusiją apie 2 500 m. pr. m. e. atnešė virvelinės keramikos kultūros, suindoeuropietinti Vidurio Europos žmonės. Vietinės kultūros nebuvo išnaikintos ir paliko pėdsakų [9]. II tūkstantmetyje prieš mūsų erą šiaurės rytų ir rytų Lietuvoje vyravo brūkšniuotosios keramikos kultūra, kuri neabejotinai priklausė tiesioginiams lietuvių (aukštaičių) protėviams.

Baltai priklauso indoeuropiečių kalbų šeimai. Iš didelės šių kalbų šeimos – prūsų, lietuvių, kuršių, žiemgalių, sėlių ir latgalių (latvių) – iki šių dienų išliko tik lietuvių ir latvių kalbos. Kuršiai gyveno šiaurės vakarų Lietuvoje ir vakarų Latvijoje, žiemgaliai gyveno vidurio Lietuvos šiaurėje ir vidurio Latvijoje. Lietuviai, latviai ir sėliai, kurie gyveno dabartinėje Rytų Lietuvoje ir Latvijoje bei Baltarusijoje, rašytiniuose šaltiniuose minimi tik nuo XI amžiaus (žr. 1.1 pav.) [9]. Baltų, ypač lietuvių, kalbų konservatyvumas turėjo nemažą reikšmę indoeuropiečių prokalbei rekonstruoti.



1.1 pav. Baltų gentys XI–XII a. (<http://www.istorikas.lt/baltu-gentys-ix-xii-a/>)



## 1.2. Populiacijos genetinė struktūra ir žmogaus genomo įvairovė

Populiacija apibrėžiama kaip vienos rūšies individų grupė, gyvenanti tam tikroje teritorijoje ir galinti kryžmintis tarpusavyje. Populiacijų genetika – genetikos mokslo sritis, tirianti alelių dažnius bendrame genų fonde ir jų kitimą populiacijoje per kartas. Žmogaus genomas yra dinamiškas, jį keičia evoliucijos metu vykstantys DNR sekos pokyčiai, persitvarkymai, kurie perduodami kitoms kartoms ir gali įsitvirtinti žmonių grupėse ir populiacijose. Vykdyti didelės imties projektai: *Žmogaus genomo įvairovės projektas* (angl. *Human Genome Diversity Project, HGDP*), *Žmogaus genomo projektas*, tarptautinis *HapMap* projektas, *1 000 genomų projektas*, suteikė daug informacijos apie žmogaus genomo struktūrą, įvairovę ir funkcijas tiek tarp individų, tiek tarp populiacijų [4, 10–12].

Žmogaus genomo įvairovė pasireiškia nuo vieno nukleotido polimorfizmų iki didelių, mikroskopiškai matomų, chromosomos pokyčių (dydis >2–5 Mb). Bet kurių dviejų atsitiktinai pasirinktų asmenų DNR sekos identiškumas yra 99,9 proc., taigi skirtumai tarp individų vidutiniškai sudaro 0,1 proc. Daugelio tyrimų rezultatai parodė, kad apie 83–88 proc. genetinės įvairovės randama populiacijos viduje ir tik apie 9–13 proc. – tarp skirtingų populiacijų [13]. Kiekvienos populiacijos genetinė įvairovė – tai prisitaikymo prie aplinkos pokyčių garantas.

Tiriant DNR sekos įvairovę, galima rekonstruoti žmogaus populiacijos evoliucinę istoriją, kilmę, struktūrą ir išaiškinti individų ar populiacijų panašumus ir skirtumus. Demografiniai pokyčiai populiacijoje palieka pėdsaką, keisdami populiacijos genetinę įvairovę. Sumažėjęs populiacijos dydis mažina genetinę įvairovę, ir atvirkščiai. Izoliuotos populiacijos išsaugo savo genetinį unikalumą, o migracija mažina populiacijų genetinius skirtumus. Taigi galima teigti, kad mūsų istorija yra užkoduota DNR sekoje.

Tokiuose tyrimuose dažniausiai naudojamos mtDNR ir Y chromosomos sekos, VNP ar trumpos paeiliui pasikartojančios sekos, pasiskirsčiusios po visą genomą. Tiriant populiacijų divergencijos laiką, priemaišą (angl. *admixture*), migraciją, populiacijos dydžio pokyčius bei kitus demografinius procesus, pirmo pasirinkimo žymenys yra plataus masto VNP genotipavimo duomenys [14].

Populiacijos genetinės struktūros tyrimai naudojami siekiant išsiaiškinti panašumus ir skirtumus tarp vienos grupės individų ar tarp skirtingų individų grupių, ir veiksnius, kurie lemia tuos skirtumus. Tai vadinama biologine evoliucija, kuri gali vykti dėl gamtinės atrankos [15]. Žmogaus organizmą veikia nuolat kintantys aplinkos veiksniai, todėl

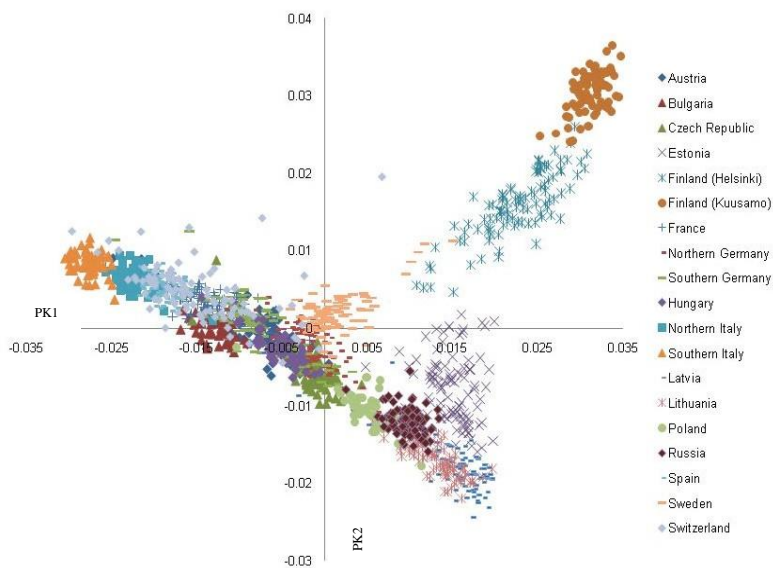
evoliucijos metu susiformavo daugybė prisitaikomųjų mechanizmų, palaikančių normalią organizmo veiklą. Nauji pokyčiai organizmui gali būti ir naudingi, ir žalingi. Pokyčiai, sudarantys sąlygas organizmui prisitaikyti prie aplinkos, perduodami iš kartos į kartą. Genų kitimas yra suderintas su viso genomo pokyčiais, kuriuos gamtinė atranka palaiko, jeigu jie suteikia privalumo [16].

Per pastarąjį dešimtmetį galimybės nustatyti detalesnę didesnės skiriamosios gebos populiacijų genetinę struktūrą stipriai pasikeitė, atsiradus didelio našumo genotipavimo technologijoms, kurios leidžia vienu metu atlikti šimtų tūkstančių ar milijono žymenų genotipavimą [17]. Detali populiacijos genetinės struktūros analizė yra svarbi asociacijų tyrimuose, tiriant populiacijų praeitį, identifikuojant vietinės adaptacijos veikiamas genomo sritis ir priskiriant kilmę individo lygmeniu [18–20].

Analizės būdas pasirenkamas priklausomai nuo turimų duomenų ir išsikeltų tikslų. Pastarąjį dešimtmetį populiacijos genetinei struktūrai nustatyti dažniausiai naudojamas tiesinės projekcijos metodas – pagrindinių komponenčių analizė (PKA) (angl. *Principal Component Analysis, PCA*) [20], taip pat modeliu paremti metodai, įdiegti programose STRUCTURE ir ADMIXTURE [21]. PKA metodas paremtas genetinių atstumų tarp porų individų skaičiavimu. Nustatomas duomenų genetinės įvairovės santykis. PKA pateikia globalius ryšius tarp populiacijų, remiantis lyginamų porų genetiniams atstumais. PKA taip pat naudojamas genetiniams atstumams tarp individų atvaizduoti. Modeliu paremti metodai, pvz., ADMIXTURE, tinkamesni rekonstruojant praeitį, individų kilmę, atsakant į klausimą, kuriai populiacijai individas priklauso. Skirtingi individai yra grupuojami į skirtingus klasterius, remiantis genetiniu panašumu ir pagal pasirinktą protėvinių populiacijų skaičių (K). Protėvinių populiacijų skaičius pasirenkamas pagal pradinę išsikeltą hipotezę, pvz., analizuojant dviejų populiacijų priemaišą, K bus 2. Neturint pradinės hipotezės, analizė atliekama su skirtingais K įverčiais, o tinkamiausias K pasirenkamas pagal mažiausią apskaičiuotą ADMIXTURE programos duomenų kontrolės (angl. *Cross validation, CV*) įvertį.

Vidurio ir Vakarų Europos kilmė ištirta gana išsamiai, didžiausia genetinė diferenciacija rasta tarp žemyno šiaurės ir pietų [22, 23]. J. Novembre su kolegomis ištyrė Europos populiacijų genetinę struktūrą, remdamasis 3 000 individų, genotipuotų pagal daugiau negu 500K žymenų, genetiniams duomenimis [23]. PKA rezultatai parodė, kad asmenys iš to paties geografinio regiono susigrupuoja ir taip atskiriamos skirtingos populiacijos. Taip pat nustatytas koreliacinis ryšys tarp tiriamų populiacijų genetinių ir geografinių atstumų. PK1 parodė šiaurės–šiaurės vakarų ar

pietų–pietryčių ašies genetinės variacijos gradientą. PK1 aprašė 0,30 proc., o PK2 – 0,15 proc. genetinės įvairovės [23]. Rytų Europos populiacijų genetinė struktūra išsamiai aprašyta M. Nelis ir kolegų darbe [24]. Tirti 3 112 individų iš 16 Europos populiacijų (95 individai iš Lietuvos populiacijos), genotipuotų pagal 270 000 VNP žymenis su *Illumina* 318K/370 CNV lustais, genetiniai duomenys. Atlikus Europos populiacijų PKA, EIGENSOFT programa nustatyta, kad PK1 aprašė 8,7 proc., PK2 – 4,9 proc. genetinės įvairovės, o tai parodo Europos populiacijų genetinį vientisumą. Įtraukus į analizę *HapMap* projekto asmenų iš CEU, YRI, CHB ir JPT populiacijų genetinius duomenis, nustatyta, kad PK1 aprašė 36,6 proc., o PK2 – 23,8 proc. genetinės įvairovės. Tarp tirtų Europos populiacijų išsiskyrė keturi regionai: 1) Suomija; 2) Baltijos šalys (estai, latviai ir lietuviai), Vakarų Rusija ir Lenkija; 3) Vidurio ir Vakarų Europa; 4) Italija. Pietų italai buvo labiausiai nutolę (žr. 1.2 pav.).



**1.2 pav.** Europos populiacijų genetinė struktūra, remiantis 273 464 VNP duomenimis [24]

Atlikus genetinių atstumų tarp porų populiacijų skaičiavimus  $F$  statistika ( $F_{ST}$ ), nustatyta nedidelė, tačiau statistiškai reikšminga koreliacija tarp genetinių ir geografinių atstumų ( $r^2 = 0,382$ ,  $p < 0,01$ ) [24].  $F_{ST}$  reikšmės tarp geografiškai artimų populiacijų buvo  $\leq 0,001$ , didžiausia  $F_{ST}$  reikšmė (0,023) nustatyta tarp pietryčių Italijos ir Suomijos. Tarp Vokietijos

ir Europos (CEU)  $F_{ST} = 0$ .  $F_{ST}$  reikšmės tarp Europos (CEU) ir Latvijos, Lietuvos, Estijos bei Vakarų Rusijos atitinkamai buvo 0,006, 0,005, 0,004 ir 0,004. Tarp lietuvių ir suomių iš Helsinkio  $F_{ST}$  reikšmė buvo 0,007, o tarp lietuvių ir suomių iš Kuusamo –  $F_{ST} = 0,013$ .  $F_{ST}$  tarp Lietuvos ir Šiaurės Italijos buvo 0,011, o tarp Lietuvos ir Pietų Italijos – 0,014.  $F_{ST}$  atstumai tarp Europos (CEU) ir Afrikos (YRI) buvo 0,153, tarp Europos (CEU) ir Japonijos (JPT) – 0,0111, tarp Europos (CEU) ir Kinijos (CHB) – 0,110. Remiantis geografiniais atstumais, nustatyti genetiniai atstumai tarp Suomijos ir Italijos, Rytų Baltijos regiono ir Lenkijos, Bulgarijos ir Vakarų Rusijos, Lenkijos, Lietuvos [24].

Modeliu paremta priemonės analizė, atlikta ADMIXTURE, STRUCTURE programomis pagal VNP genotipavimo duomenis, taip pat parodo, kad asmenys iš to paties geografinio regiono formuoja vieną klasterį, o populiacijos iš skirtingų regionų atsiskiria ir suformuoja atskirus klasterius [21, 25, 26].

I. Uktverytė 2014 m. savo disertacijoje analizavo lietuvių ir *HapMap3* projekto populiacijų priemonę [27]. Tirtos populiacijos geriausiai buvo aprašomos keturių grupių ( $K = 4$ ) modeliu. Pagrindinės grupės (Afrikos, Azijos, Pietų Azijos, Šiaurės Amerikos, Europos populiacijos) atsiskyrė. Didžiausia Azijos genetinės įvairovės dalis buvo stebima TSI populiacijoje, mažesnė – CEU, o mažiausia – tirtoje lietuvių populiacijoje.

Detalios Lietuvos populiacijos genetinės struktūros, modeliu paremtu metodu (ADMIXTURE), analizės aprašymo nėra.

Šiame darbe pirmą kartą analizuojama Lietuvos populiacijos genetinė struktūra ir padėtis 1 000 genomų projekto populiacijų atžvilgiu, naudojant plataus masto VNP genotipavimo duomenis ir PCA, ADMIXTURE analizės metodus.

### 1.3. Gamtinė atranka ir žmogaus genetinė įvairovė

#### 1.3.1. Genų dreifo teorija

Manoma, kad didžioji dalis genomo variantų yra neutralūs, nesuteikia biologinių privalumų nešiotojui. Tokiu atveju genų dreifas yra vienas svarbiausių veiksnių, galinčių nulemti tam tikro alelio išlikimą populiacijoje. Genų dreifas apibrėžiamas kaip atsitiktinis genų alelių dažnių svyravimas populiacijoje, tik atsitiktinumas nulemia, kurie aleliai bus perduodami kitai kartai [28]. Alelių dažnių svyravimai dažniausiai labiau būdingi mažoms populiacijoms. Veikiant genų dreifui, kai neveikia nei mutacijos, nei gamtinė

atranka, aleliai gali būti fiksuojami arba prarandami populiacijoje, taip sumažinant genetinę įvairovę. Tokiu atveju genų dreifas yra priklausomas tik nuo efektyvaus populiacijos dydžio.

### 1.3.2. Efektyvus populiacijos dydis

Alelių dažniai populiacijoje yra veikiami ne tik genų dreifo, gamtinės atrankos, bet ir demografinių procesų. Efektyvaus populiacijos dydžio ( $N_e$ ) sąvoką populiacijų genetikoje pirmą kartą pateikė Sewall'as Wright'as 1931 m., kaip būdą evoliuciniams pokyčiams, atsiradusiems dėl atsitiktinio alelių dažnio kitimo populiacijoje (genų dreifo), nustatyti [29].  $N_e$  gali būti apibrėžiamas kaip besidauginančių individų skaičius Wright'o-Fisher'io idealioje populiacijoje, kurioje genų dreifas bus tokio paties dydžio, kaip būtų vertinamas tikroje populiacijoje [30].

Idiali populiacija yra neriboto dydžio panmiksine populiacija (atsitiktinis populiacijos individų kryžminimasis), turinti pastovų dydį, vienodą, galinčių daugintis, vyrų ir moterų skaičių, nepersidengiančias kartas, atsitiktinį poros pasirinkimą, vienodas populiacijos individų reprodukcinės galimybes, joje neveikia gamtinė atranka ir mutacijos.

Pagrindinę  $N_e$  teoriją išplėtojo S. Wright'as, Moto Kimura ir James'as Crow, kurie pasiūlė skirtingus  $N_e$  apibrėžimus [29–31]. Dažniausiai naudojami apibrėžimai yra inbrydingo efektyvus dydis ir įvairovės efektyvus dydis, kurie matuoja genų dažnių pokyčių įvairovę iš kartos į kartą, tačiau praktikoje šie apibrėžimai dažniausiai nėra išskiriami ir vartojama bendra efektyvaus populiacijos dydžio sąvoka ( $N_e$ ).

Efektyvus populiacijos dydis yra svarbus molekulinės evoliucijos, augalų ir gyvūnų veisimo ir aplinkosaugos biologijos veiksnys [32].  $N_e$  sąveikaujant su kitais evoliucijos veiksniais, mutacijomis, gamtine atranka, migracija ir rekombinacija nulemia genetinės įvairovės pasiskirstymą ir kiekį populiacijoje.  $N_e$  gali padėti prognozuoti neutralios genetinės įvairovės praradimą ir pasiskirstymą, naudingų ir žalingų alelių fiksacijos tikimybę bei mažos populiacijos tinkamumą ir išgyvenamumą [33].  $N_e$  yra gerokai mažesnis už bendrą individų skaičių populiacijoje (angl. *Census population size*,  $N$ ) [34]. Rūšys, turinčios mažą  $N_e$ , pvz., žmonės, rodo sumažėjusią įvairovę ir sumažėjusį gamtinės atrankos efektyvumą, palyginti su kitomis rūšimis [35].  $N_e$  gali variuoti tarp skirtingų rūšies genomo sričių, pvz., autosomų, X ir Y chromosomose [36]. Vyrų didesnė reprodukcinė įvairovė, palyginti su moterų, mažina Y chromosomos  $N_e$ , palyginti su mtDNR, X chromosoma ir autosomomis. X chromosomos ir mtDNR  $N_e$  didesnis,

palyginti su autosomomis. Gamtinė atranka nulemia sumažėjusį  $N_e$  tose geno srityse, kuriose genetinė rekombinacija yra nedidelė [37].

$N_e$  gali būti nustatomas remiantis demografiniais ir genealoginiais duomenimis [37]. Tačiau surinkti visą šią informaciją populiacijos mastu yra sudėtingas uždavinys. Todėl  $N_e$  nustatymo pasirinkimo metodai remiasi genetiniais duomenimis [38]. Iki tol, kol atsirado plataus masto genotipavimo technologijos, daugelis tyrėjų  $N_e$  nustatyti rėmėsi individų, alelių dažnių pokyčiais, surinktais mažiausiai dviem skirtingais laiko momentais [38–40]. Plataus masto VNP genotipavimo duomenys paskatino naujų metodų, paremtų nepusiausvyra sankiba (angl. *Linkage Disequilibrium*, *LD*), kurie yra tikslesni ir informatyvesni, palyginti su pastaraisiais [41–43], kūrimą. Nepusiausvyra sankiba yra dviejų ar daugiau sričių alelių neatsitiktinė asociacija, kurios tam tikri haplotipai gali turėti didesnę dažnį, negu būtų tikimasi atsitiktiniu būdu [44]. Genetiniai metodai leidžia nustatyti praeities  $N_e$  nuo 200 iki 6 000 kartų, o demografiniai metodai suteikia informacijos apie  $N_e$  pokyčius ateityje [45, 46]. Taip pat reikia išskirti ilgalaikį  $N_e$  (angl. *Long-term effective population size*) ir trumpalaikį  $N_e$  (angl. *Recent effective population size*). Trumpalaikis  $N_e$  nurodo labai artimą, iki 200 kartų nuo dabarties, bendrą žmonių skaičiaus eksponentinį didėjimą populiacijoje, ilgalaikis  $N_e$  nurodo praeities, nuo 200 iki 6 000 kartų, efektyvų populiacijos dydį. Ilgalaikis  $N_e$  yra mažesnis už trumpalaikį  $N_e$ .

$N_e$  nustatymo metodas pasirenkamas priklausomai nuo išsikeltų tyrimo tikslo ir turimų duomenų.  $N_e$  kintančioje populiacijoje skaičiuojamas ne kaip aritmetinis, o kaip harmoninis vidurkis:  $N_e = \frac{t}{\sum(\frac{1}{N_i})}$ , čia  $N_i$  – besidauginančių individų skaičius  $i$  kartoje,  $t$  – kartų skaičius. Tai reiškia, kad mažas populiacijos dydis neproporcingai daro įtaką  $N_e$ . Artimoje praeityje išaugusių populiacijų ilgalaikis  $N_e$  išlieka nedidelis dėl praeityje sumažėjusio populiacijos dydžio, dėl butelio kaklelio ar įkūrėjo efekto.

Apskaičiuoti  $N_e$ , esant nelygiam lyčių santykiui, naudojama formulė:  $N_e = \frac{4N_m \times N_f}{(N_m + N_f)}$ , čia  $N_m$  – vyrų  $N_e$ ,  $N_f$  – moterų  $N_e$  [30].

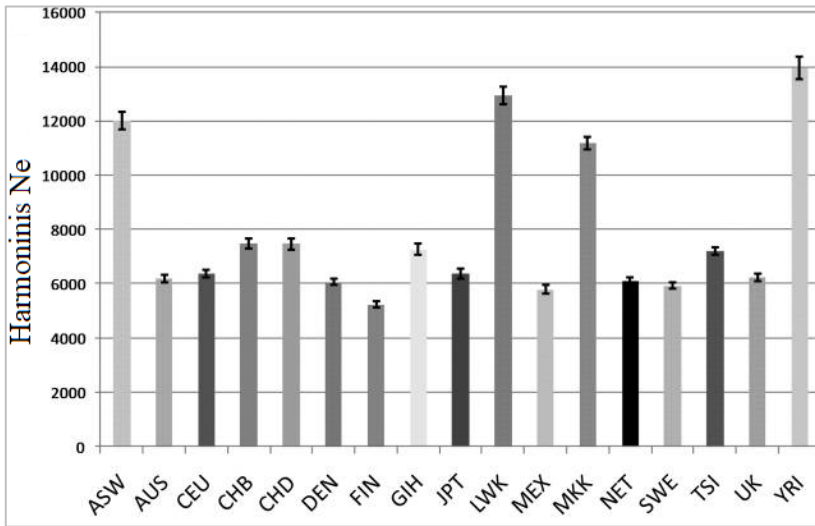
Kitas svarbus veiksnys, turintis įtakos  $N_e$ , yra vidutinis šeimos dydžio kintamumas. Šiuo atveju  $N_e$  apskaičiuojamas:  $N_e = \frac{4N-4}{(2+V)}$ , čia  $N$  – besidauginančių individų skaičius,  $V$  – palikuonių skaičiaus dispersija [47].

$N_e$  mažėjimui įtaką daro poravimasis tarp giminingų asmenų (inbrydingas), tai išreiškiama formule:  $\frac{1}{(1+F)}$ , čia  $F$  – inbrydingo koeficientas [48]. Inbrydingo laipsnis nustatomas pagal giminingumo koeficientą (angl.

*Coefficient of kinship*) – tikimybė, kad dviejų skirtingų individų abu aleliai yra identiški pagal kilmę. Pavyzdžiui, pirmos kartos pusbrolių santuokos  $F = 0,0625$ , antros kartos pusbrolių santuokos  $F = 0,0156$  [49]. *Ne* sumažėja, nes inbrydingas turi įtakos individo greitesnei motininių ir tėvinių alelių koalescencijai (angl. *coalescente*, lot. *coalesco* – susijungiu), palyginti su atsitiktiniu poravimusi [50].

Didelėse populacijose, kuriose neveikia gamtinė atranka, individai poruojasi atsitiktiniu būdu, aleliai yra nepriklausomi srityje ir tarp sričių, taip nulemiama Hardžio-Vainbergo pusiausvyra ir pusiausvyra sankiba. Tačiau riboto dydžio populacijose atsitiktinis genų dreifas nulemia asociaciją tarp vienos srities alelių ir skirtingų sričių alelių. Dėl to susidaro heterozigotų perteklius, o tai vėliau paskatina LD atsiradimą. Be genų dreifo, LD daro įtaką migracija ir gamtinė atranka [51]. LD tarp sričių nustatoma remiantis D statistika,  $D_{AB} = p_{AB} - p_A p_B$ , skirtumas tarp poros alelių – A ir B dažnių dviejose srityse ( $p_{AB}$ ) ir alelių dažnių ( $p_A$  ir  $p_B$ ) produktas. Koreliacijos koeficientas ( $r$ ) tarp  $p_A$  ir  $p_B$  yra apskaičiuojamas pagal formulę:  $r_{AB} = D_{AB} / (p_A(1 - p_A) p_B(1 - p_B))^{1/2}$  [52]. Jeigu dvi sritys yra neutralios ir populiacija izoliuota, o jos efektyvus dydis ilgą laiką pastovus, tada  $r_{AB}$  bus priklausoma nuo  $Ne$  ir rekombinacijos dažnio, kurį pažymime  $c$ , tarp sričių. Šiuo atveju susidaro pusiausvyra (angl. *Equilibrium*,  $E$ ), tada  $E(r_{AB}) = 0$  ir dispersija  $Vr_{AB} = E(r_{AB}^2) = (1 - c)^2 + \frac{c^2}{2Ne * c(2 - c) + 1/n}$ , čia  $c$  – rekombinacijos dažnis;  $r$  – alelių dažnių koreliacijos koeficientas tarp dviejų sričių, esančių LD;  $n$  – individų skaičius. Nepusiausvyra sankiba tarp dviejų genetinių sričių gali suteikti informacijos apie populiacijos dinamiką – efektyvaus populiacijos dydžio kitimą laike [53]. Genų dreifas didina nepusiausvyrą sankibą, o rekombinacija mažina. Taigi LD mažėja didėjant fiziniam atstumui tarp genetinių sričių. Žinant, kad LD priklauso tiek nuo  $Ne$ , tiek nuo rekombinacijos dažnio tarp sričių, daroma prielaida, kad LD esant dideliame atstumui tarp genetinių sričių atspindi artimos praeities  $Ne$ , o LD esant mažam genetiniame atstumui tarp genetinių sričių atspindi tolimes praeities  $Ne$  [53].

Remiantis DNR sekos įvairovės duomenimis, nustatytas ilgalaikis  $Ne$  žmonių populiacijoje varijuoja nuo ~10 000 ne Afrikos populacijose iki ~20 000 Afrikos populacijose [4, 53, 54].  $Ne$  gali varijuoti priklausomai nuo naudotų genetinių duomenų (VNP, viso genomo sekoskaitos duomenys) analizėje ir pasirinkto nustatymo metodo. B. P. McEvoy'aus ir kitų tyrėjų nustatytas  $Ne$  tarp skirtingų populiacijų varijavo nuo 13 900 (YRI) iki 5 200 (FIN) (žr. 1.3 pav.) [55].



**1.3 pav.** Tirtų populiacijų harmoninio  $N_e$  su 5 ir 95 procentilėmis pasiskirstymas [55]

$N_e$  įverčiai, nustatyti nepusiausvyra sankiba paremtu analizės metodu, panaudojant plataus masto VNP genotipavimo duomenis, yra mažesni, palyginti su kitais metodais [55]. To priežastis nėra išaiškinta, tačiau spėjama, kad dėl populiacijos butelio kaklelio efekto įtakos nepusiausvyrai sankibai. Viso genomo koalescencijos analizė naudoja vis sudėtingesnius genetinius modelius  $N_e$  nustatyti, vienas iš būdų yra paremtas Markovo koalescencijos analize (angl. *Pairwise Sequentially Markovian Coalescent Analysis, PSMC*), kuria galima nustatyti  $N_e$ , remiantis diploidinių genomų heterozigotiniais polimorfizmais [56]. Šiuo metodu nustatyta, kad tarp 100 tūkstančių ir 1 milijono metų  $N_e$  varijavo nuo 7 000 iki 15 000.

Žinant, kad  $N_e$  stipriai koreliuoja su  $N$  daugeliu atveju, naudinga ir įdomu nustatyti  $N_e$  ir  $N$  santykį populiacijoje [32, 57, 58].  $N_e$  ir  $N$  santykis gali suteikti informacijos apie ekologinius veiksnius, darančius įtaką  $N_e$  [57].  $N_e$  ir  $N$  santykio nustatymas nėra lengva užduotis, kadangi yra daug veiksnių, kurie turi įtakos šiam santykiui, pvz., persidengiančios kartos sumažina  $N_e$  25–75 proc.  $N$ . Stabiliose populiacijose  $N_e$  ir  $N$  santykis yra 0,5–1, o populiacijose, kuriose individų skaičius nepastovus, vidutinis  $N_e$  ir  $N$  santykis gali būti 0,10 [46, 59, 60].  $N_e$  ir  $N$  santykis taip pat varijuoja tarp rūšių, pvz., jūros rūšių  $N_e$  ir  $N$  santykis yra apie 0,0001, o sausumos rūšių – 0,14 [60]. Tačiau mažam  $N_e$  daugiausiai įtakos turi bendros populiacijos dydžio svyravimai [32].



Remiantis nustatytu  $N_e$  ir genetinių atstumų ( $F_{ST}$ ) dydžiais, galima nustatyti divergencijos laiką tarp tiriamų populiacijų. Pirmą didžiąją migracijos bangą iš Afrikos įvyko prieš ~40–60 tūkstančių metų. Šios migracijos palikuonys apgyvendino Europos, Azijos ir kitus pasaulio regionus [61]. I. Gronau su kitais mokslininkais nustatė vidutinį divergencijos laiką tarp Afrikos ir Eurazijos prieš ~50 000 (nuo 38 000 iki 64 000) metų, o Europos ir Rytų Azijos – prieš ~31–40 000 metų [60].

F. Tassi kartu su kitais mokslininkais nustatyto divergencijos laiko tarp Rytų Afrikos ir kitų pasaulio populiacijų, remiantis  $N_e$  ir  $F_{ST}$ , rezultatai pateikti 1.1 lentelėje [62].

**1.1 lentelė.** Nustatytas divergencijos laikas metais tarp Rytų Afrikos ir kitų pasaulio regionų

Populiacija	Vidurkis	95 proc. PI*
Europa	69 736	[62 664; 78 916]
Kaukazas	68 143	[60 080; 76 336]
Vakarų Azija	66 318	[60 187; 74 087]
Centrinė Azija	71 021	[65 785; 80 083]
Šiaurės Indija	70 23	[65 533; 79 438]
Pietų Indija	64 396	[59 931; 71 782]
Rytų Azija	87 432	[80 536; 97 802]
Pietų Azija	80 587	[73 452; 89 791]
Malaizija	71 622	[66 345; 81 344]

\*Pasikliautinis intervalas.

B. P. McEvoy’us kartu su kitais mokslininkais nustatė vidutinį divergencijos laiką tarp Afrikos ir Europos prieš ~48 000 metų, tarp Afrikos ir Rytų Azijos prieš ~66 000 metų, o tarp Rytų Azijos ir Europos prieš ~34 000 metų [55].

Divergencijos laikas tarp ne Afrikos populiacijų, nustatytas remiantis Y chromosoma ir mtDNR, yra atitinkamai prieš 40–50 tūkstančių metų ir prieš 52–60 tūkstančių metų [63, 64]. Remiantis fosilijomis ir archeologiniais tyrimų rezultatais, šiuolaikinis žmogus į Vakarų Europą ir Rytų Aziją atkeliavo prieš ~30–40 tūkstančių metų [65].

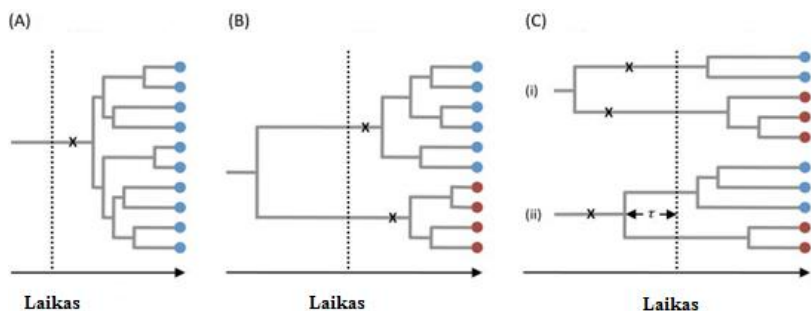
Šiame darbe pirmą kartą pateikta detali Lietuvos populiacijos ir šešių jos etnolingvistinių grupių  $N_e$  ir divergencijos analizė.

### 1.3.3. Teigiama gamtinė atranka

Genų dreifas dažnai yra dominuojantis evoliucijos veiksnys, tačiau ne vienintelis, galintis keisti alelių dažnį populiacijoje. Nuo tada, kai Čarlzas Darvinas 1859 m. savo knygoje „Rūšių atsiradimas“ (*On the Origin of Species*) aprašė gamtinės atrankos teoriją, ji laikoma viena svarbiausių evoliucijos veiksnių. Remiantis Č. Darvino teorija, jeigu naujas paveldimas pokytis yra naudingas organizmui, jis bus palaikomas teigiamos gamtinės atrankos. Naudingas pokytis yra toks, kuris suteikia didesnes galimybes išgyventi ar daugintis ir yra perduodamas iš kartos į kartą. Alelių, lemiančių naudingą pokytį, dažnis sparčiai išauga populiacijoje, kol galiausiai tampa fiksuotas (dažnis 100 proc.). Gamtinės atrankos poveikis, keičiant naudingų alelių dažnį, priklauso nuo dviejų veiksnių: gamtinės atrankos stiprumo ir kartų skaičiaus, nuo kada gamtinė atranka prasidėjo. Gamtinė atranka veikia visą individo fenotipą, kurį lemia daugybė genetinių ir aplinkos veiksnių.

Neigiama gamtinė atranka eliminuoja naujus, žalingus variantus iš populiacijos, taigi, tikimasi, kad didžioji dalis žmogaus genomo yra veikiami šios atrankos tipo. Tai stebima egzozonuose, kuriuose genetinė įvairovė yra mažesnė, palyginti su intronine genomo dalimi. Taip pat nesinoniminiai VNP nustatomi rečiau negu sinoniminiai. Nesinoniminiai VNP, kurie paveikia baltymo funkciją, yra pašalinami neigiamos gamtinės atrankos dar prieš tokiems variantams pasiekiant didesnę alelio dažnį.

Geriausiai ištirtas gamtinės atrankos tipas yra žinomas kaip stipri atranka (angl. *Hard selective sweep*) – įvykusi nauja mutacija sukuria naudingą savybę viename individe, taigi gamtinė atranka veikia teigiamai tokį alelį, kurio dažnis didėja per kartas, kol pasiekia fiksaciją (žr. 1.4 pav., B) [66]. Kitas gamtinės atrankos tipas yra žinomas kaip silpna atranka (angl. *Soft selective sweep*), kai gamtinė atranka veikia vieno geno kelis alelius, kurie egzistuoja populiacijoje kurį laiką (angl. *Standing variation*) (žr. 1.4 pav., A) [66]. Dėl pasikeitusios aplinkos ar kitų veiksnių tie aleliai tampa naudingi ir jų dažnis pradeda didėti per kartas. Sudėtingesnis gamtinės atrankos tipas yra poligeninė adaptacija, kai gamtinė atranka veikia skirtingų genų alelius, kurie gali tapti naudingi pasikeitus aplinkai (žr. 1.4 pav., C). Lengviausiai nustatoma stipri atranka. Kiek sudėtingiau nustatyti silpnąją atranką, ypač tada, kai genetinis variantas turi didelį dažnį dar nepradėjus veikti gamtinei atrankai, o tai apsunkina tokio varianto atpažinimą ir interpretaciją. Poligeninės adaptacijos tipas yra nustatomas sunkiausiai [67].



**1.4 pav.** Stiprios (A) ir sunkios (B, C) atrankos tipai [66]. (A) Visi adaptyvūs aleliai (mėlynos spalvos apskritimai) susidarę dėl vienos naujos mutacijos (x). B – adaptyvūs aleliai susidarę dėl dviejų nepriklausomų naujų mutacijų (mėlynos ir raudonos spalvos apskritimai), C – adaptyvūs aleliai jau buvo susiformavę dėl įvykusios mutacijos prieš pradėdant veikti gamtinei atrankai

Yra nemažai publikuotų tyrimų apie gamtinės atrankos poveikį žmogaus genomui. Pirmieji tyrimai susiję su hemoglobino polimorfizmu. Heterozigotiniai pagal hemoglobiną (HbA/HbS) asmenys yra atsparesni maliarijai, palyginti su abiejų tipų homozigotiniais individais [68]. Panašūs rezultatai gauti analizuojant *G6PDG* geną individų, gyvenančių kraštuose, kur paplitusi maliarija [69]. Nustatyta, kad tam tikri genų variantai lemia individų atsparumą maliarijai. Pavyzdžiui, Duffy kraujo grupės genetinėje srityje, kuri turi tris alelius:  $FY^*A$ ,  $FY^*B$  ir  $FY^*O$ , nustatytas didelis  $FY^*O$  geno dažnis Užsachario Afrikoje.  $FY^*O$  alelio nešiotojai yra atsparūs maliarijos sukėlėjui *Plasmodium vivax* [70]. Tyrimo rezultatai parodė, kad  $FY^*O$  alelis yra palaikomas gamtinės atrankos Užsachario Afrikoje [70]. Nustatyta koreliacija tarp individo genotipo ir atsparumo tam tikroms infekcijoms [71, 72]. Kiti gamtinės atrankos veikiamų žmogaus geno sričių, patvirtintų funkciniais tyrimais, pavyzdžiai pateikiami 1.2 lentelėje.

**1.2 lentelė.** Gamtinės atrankos veikiami genai, patvirtinti funkciniais tyrimais

Genas	Sritis	Funkcija	Populiacija	Literatūros šaltinis
<i>FY</i>	1q21–q22	Atsparumas maliarijai	Afrika	Hamblin & Di Rienzo, 2000
<i>EDAR</i>	2q13	Plaukų, dantų, prakaito liaukų vystymasis	Azija	Sabeti ir kt., 2007
<i>LCT</i>	2q21	Laktazės persistavimas	Europa	Bersaglieri ir kt., 2004
<i>SLC45A2</i>	5p13.3	Odos pigmentacija	Europa	Sabeti ir kt., 2007
<i>SLC24A5</i>	15q21.1	Odos pigmentacija	Europa	Lamanson ir kt., 2005
<i>CYP3A5</i>	7q21.1	Jautrumas druskai	Europa, Azija	Thompson ir kt., 2004, 2006
<i>FOXP2</i>	7q31	Kalba	Pasaulis	Enard ir kt., 2002
<i>HBB</i>	11p15.5	Atsparumas maliarijai	Afrika	Ayodo ir kt., 2007

Genas	Sritis	Funkcija	Populiacija	Literatūros šaltinis
<i>CASP12</i>	11q22.3	Atsparumas sepsiui	Pasaulyje	Xue ir kt., 2006
<i>SLCP12</i>	15q21.1	Odos pigmentacija	Europa	Lamason ir kt., 2005
<i>ABCC11</i>	16q12.1	Ausų sieros sekrecija	Azija	Xue ir kt., 2009
<i>G6PD</i>	Xq28	Atsparumas maliarijai	Afrika	Tishkoff ir kt., 2001

Plataus masto VNP genotipavimo duomenys pastaruosius 17 metų buvo plačiai taikomi teigiamos gamtinės atrankos tyrimuose tiek Europoje, tiek visame pasaulyje (žr. 1.3 lentelę).

**1.3 lentelė.** Žmonių teigiamos gamtinės atrankos tyrimai Europoje ir visame pasaulyje

Šaltinis	*Metodas	**Žymenys
Huttley ir kt., 1999	Išplėstinis LD	STR
Akey ir kt., 2002	$F_{ST}$	SNP
Akey ir kt., 2002	$F_{ST}$	SNP
Payseur ir kt., 2002	SFS	STR
Kayser ir kt., 2003	$R_{ST}$	STR
Storz ir kt., 2004	$F_{ST}$ , SFS	STR
Shriver ir kt., 2004	$F_{ST}$	SNP
Carlson ir kt., 2005	SFS	SNP
Weir ir kt., 2005	$F_{ST}$	SNP
Benjamin 2006	iHS	SNP
Voight ir kt., 2006	iHS	SNP
Wang ir kt., 2006	LD	SNP
Kelley ir kt., 2006	SFS	SNP
Williamson ir kt., 2007	CLRT	SNP
<i>International HapMap Consortium, 2007</i>	iHS; EHH	SNP
Sabeti ir kt., 2007	iHS; XP-EHH	SNP
Tang ir kt., 2007	EHH; LRH	SNP
Kimura ir kt., 2007	Haplotipų homozigotiškumas	SNP
Olesyk ir kt., 2008	$F_{ST}$	SNP
Johansson & Gyllensten, 2008	$F_{ST}$	SNP
Kimura ir kt., 2008	LRH; SFS	SNP
Myles ir kt., 2008	$F_{ST}$	SNP
Amato ir kt., 2009	$F_{ST}$	SNP
Pickrell ir kt., 2009	iHS; XP-EHH; $F_{ST}$	SNP
Lopez Herraez ir kt., 2009	Rsb	SNP
Albrechtsen ir kt., 2010	Identiški pagal kilmę segmentai	SNP
Bigham ir kt., 2010	lnRH; WGRHLH; SFS	SNP; CNV
Simonson ir kt., 2010	iHS; XP-EHH	SNP
Beall ir kt., 2010	Alelių dažnių skirtumai	SNP
Lappalainen ir kt., 2010	iHS, LRH, EHH; $F_{ST}$	SNP

Šaltinis	*Metodas	**Žymenys
Chen ir kt., 2010	XP-CLR	SNP
Xu ir kt., 2010	iHS; XP-EHH; XP-CLR; $F_{ST}$	SNP
Metspalu ir kt., 2011	XP-EHH; iHS	SNP
Granka ir kt., 2012	iHS; XP-EHH	SNP
Piras ir kt., 2012	EHH; XP-EHH	SNP
Vernot ir kt., 2012	$F_{ST}$	SNP
Jarvis ir kt., 2012	$F_{ST}$ ; XP-EHH; iHS	SNP
Scheinfeldt ir kt., 2012	LSBL	SNP
Suo ir kt., 2007	iHS; XP-EHH	SNP
Migliano ir kt., 2013	iHS; XP-EHH	SNP
Raj ir kt., 2013	iHS; $F_{ST}$	SNP
Liu ir kt., 2014	iHS; XP-EHH	SNP
Bhatia ir kt., 2014	Alelių dažnių skirtumai	SNP
Colonna ir kt., 2014	iHS; XP-EHH; $F_{ST}$	SNP; indel
Sadler ir kt., 2015	SFS, iHS	SNP
Xuanyao Liu ir kt., 2015	XP-EHH	SNP
Vatsiou ir kt., 2016	iHS	SNP
Jonnalagadda ir kt., 2017	SFS, iHS; Rsb	SNP; LSBL
Liu ir kt., 2017	iHS; XP-EHH	SNP
Kelsey ir kt., 2018	iHS	SNP

\*iHS – integruota haplotipų statistika (angl. *Integrated Haplotype Statistics, iHS*); CLR – sudėtinės tikimybės santykis (angl. *Composite Likelihood Ratio, CLR*); LSBL – sričiai specifinis sekos ilgis (angl. *Locus-Specific Branch Length, LSBL*);  $\ln RH$  – logaritminis heterozigotų santykis (angl. *the log of the Ratio of Heterozygosities, lnRH*); WGLRH – viso genomo haplotipų testas (angl. *Whole Genome Long Range Haplotype tests, WGLRH*); EHH – išplėstinis haplotipų homozigotiškumas (angl. *Extended Haplotype Homozygosity, EHH*); XP-EHH – tarp populiacijų išplėstinis haplotipų homozigotiškumas (angl. *Cross Population Extended Haplotype Homozygosity, XP-EHH*); LRH – ilgų haplotipų testas (angl. *Long Range Haplotype test, LRH*); SFS – vietos dažnių spektro statistika (angl. *Site Frequency Spectrum statistic, SFS*);

\*\*CNV – kopijų skaičiaus pokytis (KSP, angl. *Copy Number Variation, CNV*); SNP – vieno nukleotido polimorfizmas (VNP, angl. *Single Nucleotide Polymorphism, SNP*); STR – trumpi tandeminiai pasikartojimai (TTP, angl. *short tandem repeats, STR*).

Šiame darbe pirmą kartą Lietuvoje analizuojama populiacijos teigiama gamtinė atranka, panaudojant plataus masto VNP genotipavimo duomenis ir tris statistinius metodus.

### 1.3.4. Gamtinės atrankos nustatymo metodai

#### 1.3.4.1. Metodai, paremti haplotipų dažniu ir įvairove

Haplotipas – vienos chromosomos alelių rinkinys, paveldimas kaip nedalomas vienetas. Tokie aleliai dažniausiai yra nepusiausvyroje sankiboje. Skirtumai tarp haplotipo dažnio ir įvairovės gali suteikti informacijos apie gamtinę atranką, kuri prasidėjo prieš <30 tūkstančių metų [73]. Didelio dažnio haplotipai, manoma, yra seni, nes reikia nemažai laiko, kad haplotipo dažnis padidėtų veikiant genų dreifui. Haplotipų, turinčių alelį, kuris buvo veikiamas netolimos gamtinės atrankos (angl. *Recent Natural Selection*), genetiniai atstumai yra daug didesni, palyginti su kitais tos srities panašaus dažnio haplotipais. Su laiku rekombinacija gali pertraukti haplotipą, taigi prarandama gamtinės atrankos genetinė žymė (angl. *Genetic signature of natural selection*).

Statistiniai metodai, galintys nustatyti tokio tipo gamtinę atranką, vadinami išplėstiniais haplotipų homozigotiškumo (angl. *Extended Haplotype Homozygosity*, EHH) testais, kuriais nustatomi ilgi, didelio dažnio haplotipai populiacijoje [71]. Kitas statistinis metodas, paremtas EHH, yra XP-EHH – tarp populiacijų išplėstinis haplotipų homozigotiškumas (angl. *Cross Population Extended Haplotype Homozygosity*, XP-EHH), kuriuo galime nustatyti populiacijai specifinius gamtinės atrankos veikiamus genetinius variantus. Šiai statistikai reikalinga referentinė populiacija [74].

Skirtingos tyrėjų grupės, analizuodamos gamtinę atranką, pritaikė EHH statistiką plataus masto VNP genotipavimo duomenims. Pavyzdžiui, P. C. Sabeti kartu su kitais tyrėjais, naudodamasis *HapMap2* projekto duomenimis, identifikavo apie 300 kandidatinių gamtinės atrankos veikiamų sričių, iš jų 22 yra stiprios kandidatinės sritys [74].

#### 1.3.4.2. Alelių dažnių spektru paremti metodai

Genetinių variantų, esančių skirtingų individų DNR sekose, dažniai skiriasi. Gamtinė atranka ir demografiniai procesai keičia genetinių variantų dažnio spektrą, tai vadinama alelių dažnių spektru (angl. *Allele Frequency Spectrum*, AFS). Pavyzdžiui, jeigu populiacijos dydis neseniai išaugo, daugelis naujai atsiradusių alelių yra neseni ir jų dažnis bus mažas. Tokioje populiacijoje bus retų alelių perteklius. Jeigu tam tikros srities haplotipas yra veikiamas gamtinės atrankos, tokio haplotipo dažnis bus didesnis, palyginti su kitais. Taigi haplotipo specifinis retų alelių perteklius gali būti stebimas tik toje srityje, bet ne kitur. Subpopuliacijos ir balansuojanti gamtinė atranka palaiko daugelį haplotipų ilgiau, negu būtų

tikimasi veikiant neutraliai evoliucijai. Taigi galiausiai susidaro vidutinio dažnio alelių perteklius.

Alelių dažnių spektru paremti statistiniai metodai remiasi prielaida, kad veikiant neutraliai evoliucijai skirtingi  $\theta$  (teorinis genetinės įvairovės nustatymo matas) įverčiai turi būti vienodi, čia  $\theta = 4Ne\mu$  ( $Ne$  – efektyvus populiacijos dydis;  $\mu$  – mutacijų dažnis kartoje). Tik kai kurie  $\theta$  įverčiai inkorporuoja informaciją apie alelių dažnius; segreguojančių sričių skaičius ( $S$ ) yra nepriklausomas nuo alelių dažnio, tačiau nukleotidų įvairovė ( $\pi$ ) yra priklausoma. Skirtumai tarp  $\theta$  įverčių, kurie inkorporuoja informaciją apie alelių dažnius, identifikuoja alelių dažnių spektro nukrypimus, veikiant neutraliai evoliucijai.

Vienas dažniausiai naudojamų statistinių metodų, paremtas alelių dažnių spektru, yra Tajima's D, kuris remiasi genetinės įvairovės sumažėjimu ir nustato senas gamtinės atrankos sritis, veikiamas apie 250 tūkstančių metų [75]. Tajima's D statistika lygina du įverčius: segreguojančių sričių skaičių ( $S$ ) ir vidutinį skirtumą tarp nukleotidų ( $\pi$ ). Esant nulinei hipotezei, tikimasi, kad  $E[\pi] = \theta$ , o  $E[S] = a_1\theta$ , čia  $a_1 = \sum_{i=1}^{n-1} 1/i$ , kur  $n$  – DNR sekų skaičius. Tajima's D apskaičiuojamas pagal formulę:  $D = (\pi - S/a_1)/\sqrt{V}$ , čia  $V = Var[\pi - S/a_1]$ . Vykstant neutraliai evoliucijai, aleliai yra pusiausvyroje ir Tajima's D lygi nuliui. Kitu atveju Tajima's D įverčių intervalas yra  $[-2; 2]$ . Reikšmingai teigiami įverčiai indikuoja populiacijos dydžio sumažėjimą ir (arba) balansuojančią atranką. Neigiami Tajima's D įverčiai parodo veikiančią teigiamą gamtinę atranką populiacijoje arba jos ekspansiją.

Alelių dažnių spektru paremtiems metodams tinkamiausi yra DNR sekoskaitos duomenys, tačiau plataus masto VNP genotipavimo duomenys taip pat tinkami analizuoti.

#### 1.3.4.3. Populiacijos diferenciacija paremti metodai

Gamtinė atranka gali veikti tam tikras genetines sritis vienoje populiacijoje, o kitoje ne, ypač tuomet, kai veikia skirtingi aplinkos veiksniai. Identifikuoti populiacijų skirtumai gali padėti nustatyti gamtinės atrankos veikiamas genomo sritis tam tikroje populiacijoje. Paprasčiausias būdas yra atlikti genomo sričių nukleotidų įvairovės palyginimo analizę tarp populiacijų. Jeigu VNP dažnis vienoje populiacijoje yra labai didelis, palyginti su kita populiacija, galima daryti prielaidą, kad to priežastis yra gamtinė atranka.

Atlikus 100 000 VNP alelių dažnio palyginimą tarp kinų (Han) ir tibetiečių, buvo nustatyti *EPAS1* geno du VNP su labai dideliu dažnių

skirtumu dėl gamtinės atrankos poveikio. Kiekvieno VNP alelių dažnio statistika gali būti užrašyta kaip  $|\Delta DAF|$ , alelio dažnio absoliutus skirtumas, lyginant dvi populiacijas. 1 000 genomų projekto metu nustatyti VNP su dideliu  $|\Delta DAF|$ , įskaitant VNP *EDAR* gene, taip pat *SLC24A5*, *DARC* ir *TLRI* genuose [13].

Dažniausiai naudojamas metodas diferenciacijai tarp populiacijų nustatyti yra  $F_{ST}$  statistika, kuri nustato gamtinės atrankos veikiamas sritis nuo <50 tūkstančių iki 75 tūkstančių metų [76, 77].

Šis metodas dažnai sulaukia kritikos, kadangi  $F_{ST}$  yra veikiamas populiacijos genetinės struktūros ir demografinių veiksnių. J. M. Akey'us kartu su kitais mokslininkais, naudodami  $F_{ST}$  metodą, nustatė daugiau kaip šimtą sričių, turinčių gamtinės atrankos žymes, ir reikšmingą lyginamų populiacijų diferenciaciją [78]. B. S. Weir'as kartu su kitais mokslininkais išanalizavo *Perlegan* (~1 mln. VNP), *HapMap* I projekto (~0,6 mln. VNP) duomenis ir nustatė, kad sričiai specifinių  $F_{ST}$  įverčių variabilumas yra per didelis. Nepaisant minėtų trūkumų, analizuojant daugybines nepriklausomas sritis, su tinkamais kriterijais identifikuojant išskirtis,  $F_{ST}$  gali būti puikus populiacijai specifinės gamtinės atrankos indikatorius [79].

Populiacijos diferenciacija paremti metodai dažniausiai yra naudojami kartu su metodais, paremtais haplotipų dažniu gamtinei atrankai nustatyti tam tikroje populiacijoje, palyginti su referentine populiacija. Šiame darbe buvo taikyta analogiška gamtinės atrankos analizės strategija.



## 2. DARBO METODOLOGIJA

### 2.1. Tyrimo eiga

Tyrimai, laikantis asmens duomenų apsaugos ir bioetikos reikalavimų, buvo vykdomi nuo 2014 m. spalio 1 d. iki 2017 m. spalio 31 d. Vilniaus universiteto Medicinos fakulteto Biomedicinos mokslų instituto Žmogaus ir medicininės genetikos katedroje (VU MF BMI ŽMGK), o nuo 2017 m. lapkričio 2 d. iki 2018 m. gegužės 11 d. – *Pompeu Fabra Universitat* Eksperimentinių ir sveikatos mokslų departamento Evoliucinės populiacijų genetikos laboratorijoje (angl. *Evolutionary Population Genetics Lab of the Department of Experimental and Health Sciences at the Universitat Pompeu Fabra*), Ispanijoje.

Disertacijos tyrimą sudaro dvi dalys: 1) Lietuvos populiacijos genetinės struktūros analizė ir 2) evoliucinių veiksnių: efektyvaus populiacijos dydžio ir gamtinės atrankos, analizė pagal plataus masto VNP genotipavimo duomenis. Tyrimo strategija ir darbo eiga pateikta 2.1 paveiksle.

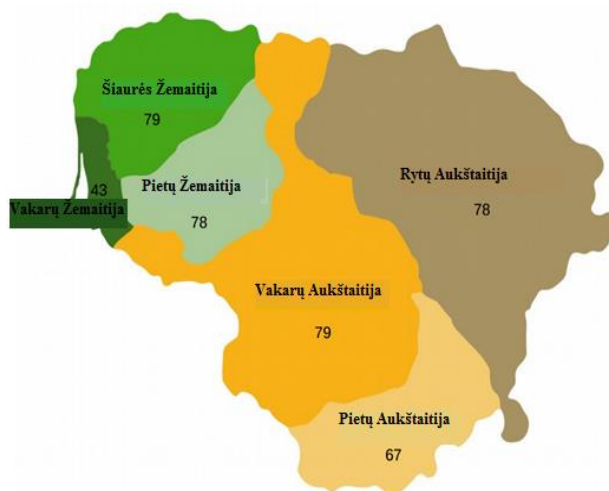
### 2.2. Tiriamieji

Pirmosios darbo dalies tiriamųjų imtį sudarė 294 asmenys iš bendros Lietuvos populiacijos, o antrosios darbo dalies imtį – 424 tiriamieji, papildžius pirmos darbo dalies tiriamųjų imtį 130 asmenų. Vidutinis tiriamųjų amžius – 53 metai. Tiriamųjų veninio kraujo ėminiai surinkti 2011–2013 m. LITGEN projekto (VP1-3.1-ŠMM-07-K-01-013) metu iš atsitiktinai pasirinktų asmenų, priklausančių šešioms Lietuvos etnolingvistinėms grupėms: trys aukštaičių grupės (Vakarų (n = 79), Pietų (n = 67), Rytų (n = 78)) ir trys žemaičių grupės (Šiaurės (n = 79), Pietų (n = 78) ir Vakarų (n = 43)), kurių tėvai ir seneliai gyveno toje pačioje etnolingvistinės grupės teritorijoje (žr. 2.2 pav.). Genominę DNR iš veninio kraujo ėminių išskyrė Vilniaus universiteto Medicinos fakulteto LITGEN projekto vykdytojai.

Disertacijos metu atliktas tiriamasis darbas yra 2011–2015 m. VU MF vykdyto projekto „Lietuvos populiacijos genetinė įvairovė ir sandaros kitimai, susiję su evoliucija ir dažniausiai paplitusiomis ligomis, akronimas LITGEN“ dalis.

Tiriamieji	LITGEN projekto (VP1-3.1-ŠMM-07-K-01-013) tiriamieji, veninio kraujo ėminiai surinkti 2011–2013 m.	
DNR	DNR išskyrimas iš kraujo leukocitų fenolio ir chloroformo ekstrakcijos metodu bei automatizuota DNR išskyrimo sistema <i>TECAN freedom EVO</i>	
Genotipavimas	<i>Illumina Infinium® HD</i> protokolas, 770K <i>HumanOmniExpress-12</i> v1.1 lustas, 719 666 VNP, 295 tiriamieji (atlikimo data – 2011–2014 m.)	<i>Illumina Infinium HTS</i> protokolas, <i>Infinium OmniExpress-24v1-2</i> lustas, 713 599 VNP, 144 tiriamasis (atlikimo data – 2017 m.)
Duomenų paruošimas	<i>GenomeStudio</i> v2011 <i>Plink</i> v1.07/ v1.9 <i>Perl</i> v5.22 <i>VCFtools</i>	
Duomenų analizė	Bioinformaciniai įrankiai: R v. 3.4.4 programinė įranga ( <i>NeON</i> paketas; <i>popGenome</i> , <i>Phangorn</i> biblioteka) 4P v. 1.0 IBDNe v. 04Sep15.e78 IBDseq v. r1206 EIGENSOFT v. 7.2.1 ADMIXTURE v. 1.3.0 KING v. 2.1 <i>Plink</i> v1.90 SHAPEIT2 <i>Selscan</i> v1.2.0a <i>VCFtools</i> v.0.1.13 ANNOVAR	

## 2.1 pav. Tyrimo strategija ir darbo eiga



**2.2 pav.** Lietuvos etnolingvistinių grupių žemėlapis ir tiriamųjų skaičiaus pasiskirstymas

Tyrimams, kurie buvo atliekami LITGEN projekto metu, 2011 m. gegužės 3 d. buvo gautas Vilniaus regioninio biomedicininų tyrimų etikos komiteto leidimas Nr. 158200-05-329-79. Taip pat buvo gauti visų tiriamųjų sutikimai dalyvauti LITGEN projekto tyrimuose.

### 2.3. Genotipavimas

Į tyrimą įtrauktų Lietuvos populiacijos asmenų genomine DNR buvo išskirta iš kraujo leukocitų fenolio ir chloroformo ekstrakcijos metodu ir automatizuota DNR išskyrimo sistema *TECAN freedom EVO*. Išskirtos DNR koncentracija ir švarumas nustatyti *NanoDropR ND-1000* spektrofotometru. Išsamūs abiem metodais atlikti DNR išskyrimo ir koncentracijos bei švarumo nustatymo protokolai pateikti I priede.

DNR mėginių plataus masto genotipavimas pagal vieno nukleotido polimorfizmo žymenis atliktas *Illumina Infinium<sup>®</sup> HD* ir *Illumina Infinium HTS* technologijos lustais. 295 asmenų DNR genotipuota *Illumina 770 K HumanOmniExpress-12 v1.1* (719 666 VNP) VNP lustais LITGEN projekto metu. 144 asmenų DNR genotipuota *Infinium OmniExpress-24v1-2* (713 599 VNP) lustais doktorantūros metu. Žymenų persidengimas tarp lustų – 707 138 VNP. *Illumina Infinium OmniExpress-24v1-2* lustai įsigyti iš doktorantūros, kuri buvo finansuojama pagal projektą „Doktorantūros studijų plėtra“ (Nr. 09.3.3-ESFA-V-711-01-0001), įgyvendinamą pagal 2014–2020 m. Europos Sąjungos fondų investicijų veiksmų programos 9 prioriteto „Visuomenės švietimas ir žmogiškųjų išteklių potencialo

didinimas“ priemonę 09.3.3-ESFA-V-711 „Mokslininkų ir kitų tyrėjų gebėjimų stiprinimas“, lėšų. Išsamūs plataus masto VNP genotipavimo protokolai pateikti II priede.

Genotipuotų duomenų pirminė analizė atlikta *GenomeStudio* v2011.1 programine įranga. Pirmiausia buvo sukurtas naujas analizės projektas *GenomeStudio* programoje, kurio metu buvo įkelti pirminiai genotipavimo duomenys (angl. *Raw data*) (žr. III priedo I dalį). Genotipuotų duomenų kokybės kontrolė pradedama klasterizuojant VNP. Toliau yra atmetami prastos kokybės mėginiai, kurie yra vertinami pagal genotipų kokybės santykį (angl. *Call Rate*). Jo įvertis turi būti >98 proc. 14 mėginių buvo pašalinti dėl prasto genotipų kokybės santykio. Naudojantis *GenomeStudio* programos papildomu įrankiu *PLINK Input Report Plug-in* v2.1.3, buvo sukurti įvesties failai (angl. *input*) .ped ir .map formatais, reikalingi tolesnei analizei (žr. III priedo II dalį).

## 2.4. Duomenų paruošimas

### 2.4.1. Duomenų paruošimas populiacijos genetinės struktūros analizei

Šioje tyrimų dalyje tiriamųjų imtis buvo padidinta 131 asmeniu. *GenomeStudio* programa buvo sukurti .map ir .ped formatų įvesties failai. Pašalintos 0, 23, 24 ir 25 chromosomos, naudojantis *plink* v1.90 programos komandine eilute *plink --file --chr 1-22 --recode --out* [80]. Toliau buvo atliktas 294 tiriamųjų genotipavimo duomenų sujungimas su 131 papildomai genotipuoto tiriamojo duomenimis, naudojantis *plink* v1.90 programos *--merge* komanda. Po sujungimo ir kokybės kontrolės *plink* v1.90 programa pagal 3.4.1 skyriuje aprašytus parametrus liko 424 individai iš 425 ir 532 836 VNP iš 589 752.

Atliktas VNP su nepusiausvyra sankiba ( $r^2 > 0,5$ ) pašalinimas (angl. *SNP pruning*) *plink* v1.07 programos komandine eilute *plink --file data --indep-pairwise 50 5 0,5*, čia lango dydis – 50 VNP, peršokimas – 5 VNP, koreliacijos koeficientas – 0,5. Po šios procedūros liko 232 752 VNP [80]. Šie duomenys buvo naudojami Lietuvos populiacijos struktūros analizėje EIGENSOFT v7.2 programos *SmartPCA* programiniu paketu išskirtims (angl. *Outliers*) nustatyti ir Lietuvos populiacijos kilmei įvertinti ADMIXTURE v1.3.0 programa [21, 81].

Lietuvos populiacijos duomenims su populiacijomis iš 1 000 genomų projekto sujungti pirmiausia iš .map formato buvo išfiltruotos failo eilutės *chr* ir *last position*, panaudojus pačios susikurtą *perl* scenarijų (angl. *script*).

Toliau buvo atliekamas Lietuvos populiacijos VNP duomenų pozicijų filtravimas iš 1 000 genomų projekto duomenų, panaudojant *vcftools* programos --*positions* komandą. Lietuvos populiacijos VNP duomenys sujungti su duomenimis iš 1 000 genomų projekto, t. y. su 20 populiacijų iš keturių pagrindinių geografinių regionų. Tai: Afrikos – jorubai iš Ibadano, Nigerija (angl. *Yoruba in Ibadan, Nigeria, YRI*), luhjai iš Webuye, Kenija (angl. *Luhya in Webuye, Kenya, LWK*), Gambijos populiacija iš Vakarų Gambijos (angl. *Gambian in Western Divisions in the Gambia, GWD*), mendės etninės grupės atstovai iš Siera Leonės (angl. *Mende in Sierra Leone, MSL*) ir esan gyventojai iš Nigerijos (angl. *Esan in Nigeria, ESN*); Europos – Šiaurės ar Vakarų Europos kilmės Jutos gyventojai (angl. *Utah Residents with Northern and Western European Ancestry, CEU*), italai iš Toskanos regiono, Italija (angl. *Tosceni in Italia, TSI*), suomiai iš Suomijos (angl. *Finnish in Finland, FIN*), britai iš Anglijos ir Škotijos (angl. *British in England and Scotland, GBR*) ir lietuviai iš Lietuvos (LT); Rytų Azija – kinai iš Pekino, Kinija (angl. *Han Chinese in Beijing, China, CHB*), japonai iš Tokijo, Japonija (angl. *Japanese in Tokyo, Japan, JPT*), kinai iš Pietų Kinijos (angl. *Southern Han Chinese, China, CHS*), dajai iš Sišuangbanos, Kinija (angl. *Chinese Dai in Xishuangbanna, China, CDX*) ir kinh tauta iš Hošimino miesto, Vietnamas (angl. *Kinh in Ho Chi Minh City, Vietnam, KHV*); Pietų Azija – indų gudžaratų kilmės Hiustono gyventojai, Teksasas (angl. *Gujarati Indians in Houston, Texas, GIH*), pandžabiai iš Lahoro, Pakistanas (angl. *Punjabi from Lahore, Pakistan, PJJ*), bengalai iš Bangladešo (angl. *Bengali from Bangladesh, BEB*), Šri Lankos tamilų kilmės gyventojai iš Jungtinės Karalystės (angl. *Sri Lankan Tamil from the UK, STU*) ir indų kilmės telugai, gyvenantys Jungtinėje Karalystėje (angl. *Indian Telugu from the UK, ITU*).

Sujungus duomenis, buvo atliktas VNP su nepusiausvyra sankiba ( $r^2 > 0,5$ ) pašalinimas *plink* v1.07 programa, po kurio liko 2 928 individai ir 264 946 VNP.

#### 2.4.2. Duomenų paruošimas teigiamai gamtinei atrankai nustatyti

Pirmiausia buvo atliktas Lietuvos populiacijos tiriamųjų giminingumo koeficiento (angl. *kinship*) ir inbrydingo koeficiento (F) nustatymas programomis *KING* v.2.1 ir *plink* v1.07 [80, 82]. Nustatytos neigiamos F reikšmės pakeistos į nulius, o tai galimai atspindi imties dydžio paklaidą (angl. *Sampling error*) [83]. Remiantis šios analizės ir PKA rezultatais, išskirtys ir (arba) giminingi asmenys iš tolesnės analizės buvo pašalinti.

Haplotipų formavimas atliktas *SHAPEIT2* programa [84]. Teigiamai gamtinei atrankai nustatyti naudoti Lietuvos populiacijos ir 1 000 genomų projekto suformuoti haplotipai pagal plataus masto VNP genotipavimo duomenis .vcf formatu.

### 2.4.3. Duomenų paruošimas efektyviam populiacijos dydžiui nustatyti

Pirmiausia 295 tiriamiesiems *GenomeStudio* programa buvo sukurti .map formato įvesties failai. Pašalintos 0, 23, 24 ir 25 chromosomos, panaudojant *plink* v1.90 programos komandinę eilutę *plink --file --chr 1-22 --recode --out*. Naujai sukurtoms .map ir .ped formato byloms atlikta kokybės kontrolė *plink* v1.07 programa pagal kokybės kontrolės įverčius: *--mind* 0,1 (individų, kuriems trūksta >10 proc. genotipo, pašalinimas); *--maf* 0,01 (VNP, kurių mažiausias alelio dažnis <0,01, pašalinimas); *--geno* 0,01 (VNP, kuriems trūksta >1 proc. genotipo, pašalinimas); *--hwe* 0,0001 (VNP, kurie neatitinka Hardžio-Vainbergo pusiausvyros dėsnio su p reikšme <10<sup>-4</sup>, pašalinimas). Pirmame analizės etape po kokybės kontrolės pašalintas vienas tiriamasis iš 295 ir 151 626 VNP iš 719 666. Galutiniai .ped ir .map formato failai konvertuoti *plink* v1.07 programa į dvejetainius failus su išvestimis .bed, .bim ir .fam, kurios reikalingos efektyviam populiacijos dydžiui nustatyti R programinės įrangos *NeON* paketu.

Lietuvos populiacijos divergencijos laikui nustatyti su kitomis pasaulio populiacijomis Lietuvos duomenys sujungti su HGDP-CEPH populiacijų duomenimis (žr. 4.1 lentelę) [10]. Duomenims sujungti naudotos *linux shell* komandos *cut* ir *join*, prieš tai konvertavus .ped ir .map formato failus į .vcf (angl. *Variant call format*) v4.2 formato failą.

## 2.5. Duomenų analizė

### 2.5.1. Populiacijos genetinės struktūros analizė

Lietuvos populiacijos struktūrai ir išskirtims nustatyti atlikta pagrindinių komponentų analizė (PKA) *EIGENSOFT* v7.2.1 programos *SmartPCA* programiniu paketu [81]. PKA taip pat atlikta sujungiant mūsų mėginius su referentiniais mėginiais iš 1 000 genomų projekto [4].

Kiekvieno individo genetinės kilmės įvertinimas atliktas *ADMIXTURE* v1.3.0 programa [21]. Analizuotas protėvinių populiacijų (K) skaičius nuo 2 iki 9 su 10 iteracijų. Geriausia K vertė nustatyta pagal kryžminės duomenų kontrolės (angl. *Cross validation*, CV) įvertį, apskaičiuotą *ADMIXTURE* v1.3.0 programa.

## 2.5.2. Kandidatinių teigiamos gamtinės atrankos veikiamų genomo sričių nustatymas

Genomo sritys, veikiamos netolimos gamtinės atrankos, nustatytos remiantis alelių fiksacijos indeksu ( $F_{ST}$ ) ir tarp populiacijų išplėstiniu haplotipų homozigotiškumo (XP-EHH) statistiniu metodu [74]. Abu metodai paremti diferenciacija tarp analizuojamos ir referentinės populiacijos. Analizė atlikta tarp porų populiacijų: LT ir CEU, LT ir FIN, LT ir YRI, CEU ir FIN, CEU ir YRI, FIN ir YRI.

Alelių  $F_{ST}$  įverčiai tarp analizuojamų populiacijų apskaičiuoti *VCFtools* v.0.1.13 programa [85]. Neigiami įverčiai pakeisti į nulius [83]. XP-EHH įvertis kiekvienam VNP tarp lyginamų populiacijų apskaičiuotas *Selscan* v1.2.0a programa [86]. Įverčiai  $>2$  (programos autorių rekomenduojama reikšmė) pasirinkti tolesnei analizei, kaip galimi kandidatai, identifikuojant gamtinės atrankos veikiamas genomo sritis. Diferencijuotoms genomo sritims identifikuoti pasirinktas išskirčių metodas pagal duomenų empirinį pasiskirstymą. Gamtinės atrankos veikiamas kandidatine genomo sritimi buvo laikoma ta, kuri buvo tarp 0,1 proc. išskirčių, srityje buvo  $\geq 2$  VNP ir bent vieno VNP  $F_{ST}$  p reikšmė  $<0,01$ .  $F_{ST}$  p reikšmės apskaičiuotos pagal kriterijų rangavimo metodą (angl. *Rank scores*) [87].

Senesni gamtinės atrankos signalai nustatyti Tajima's D neutralumo statistikos metodu. Tajima's D įverčiai apskaičiuoti R programinės įrangos *PopGenome* paketu, pasirenkant 100 kb slenkantį langą (angl. *Sliding window*) per visas autosomas su 10 kb peršokimo dydžiu [88]. Rezultatų išskirtys nustatytos pagal Tajima's D p reikšmių, apskaičiuotų pagal kriterijų rangavimo metodą, empirinį pasiskirstymą. Prieš tai slenkančių langų duomenys išrikiuoti didėjančia tvarka pagal Tajima's D įverčius. Rezultatai, kurių p reikšmės  $<0,01$ , įtraukti į tolesnę analizę.

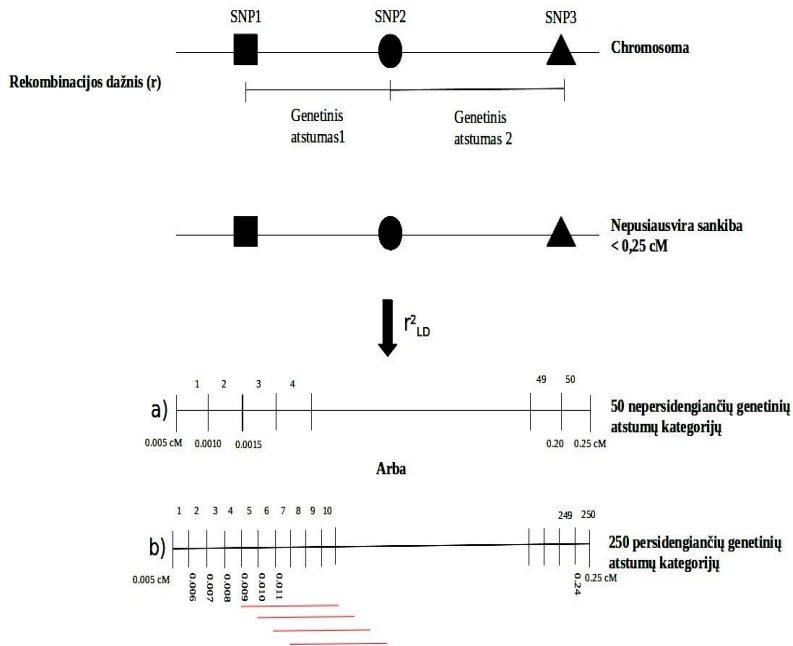
Anotacija atlikta ANNOVAR programa pagal *GRCh37 (hg19)*, *RefSeqGene*, *dbSNP147* ir CADD v1.3 [89, 90]. CADD yra įrankis, skirtas nustatyti žmogaus genomo vieno nukleotido variantų žalingumą.

## 2.5.3. Ilgalaikio efektyvaus populiacijos dydžio ir divergencijos analizė

Lietuvos populiacijos ir jos šešių etnolingvistinių grupių efektyviam populiacijos dydžiui nustatyti naudotas R programinės įrangos *NeON* v1.0 paketas, kuris remiasi nepusiausvyros sankibos įvertinimu tarp autosomų VNP žymenų (žr. IV priedą, I) [91].

*NeON* paketui, kaip įvesties failams, reikalingi dvejetainiai (.bed, .bim, .fam) *plink* v1.07 programos failai. Trečiame .bim formato faile stulpelyje

pateikiama kiekvieno žymens pozicija genolapyje, kuri reikalinga nepusiausvyros sankibos koreliacijos koeficientui kvadratu ( $r_{LD}^2$ ) apskaičiuoti tarp žymenų. Efektyvus populiacijos dydis apskaičiuojamas:  $Ne \approx 1/(4c) * [(1/r_{LD}^2) - 2]$ , čia  $c$  – genetinis atstumas Morganais tarp žymenų. Programa, remiantis Mezzavilla'os-Ghirotto'o metodu, kuris buvo naudotas šiame darbe, sukuria 250 persidengiančių genetinių atstumų kategorijų kas 0,001 centimorganą (cM) nuo 0,005 iki 0,25 cM (žr. 2.3 pav., b).



**2.3 pav.** Skirtingi genetiniai metodai efektyviam populiacijos dydžiui nustatyti, remiantis nepusiausvyra sankiba: a – McEvoy'aus metodas, sukuriantis 50 nepersidengiančių genetinių atstumų kategorijų nuo 0,005 iki 0,25 cM; b – Mezzavilla'os-Ghirotto'o metodas

Kiekviena žymenų pora yra priskiriama vienai iš genetinių atstumų kategorijų. Efektyvus populiacijos dydis apskaičiuojamas kiekvienos chromosomos kiekvienai 250 kategorijų tam tikru laiko intervalu praeityje, pvz.,  $t = 1/(2c)$  kartai, čia  $c$  – vidutinis kiekvienos kategorijos genetinis atstumas [92]. Praeities  $Ne$  nustatomas nuo ~6 000 iki ~200 kartų, arba nuo 150 000 iki 5 000 m., karta – 25 metai, su pasikliautinaisiais intervalais [5 proc.; 95 proc.], kurių rezultatai vizualizuojami pritaikius  $NeON$  paketo funkciją  $Ne\_Plot$ . Ilgalaikis  $Ne$  apskaičiuojamas kaip harmoninis vidurkis su jo pasikliautiniu intervalu [29].



Divergencijos laikas ( $T$ ) tarp tirtų populiacijų apskaičiuotas pagal formulę  $T = \ln(1 - F_{ST}) / \ln(1 - 1/2 Ne)$ , *NeON* paketo *Tdverg* funkcija. Tam reikalinga genetinių atstumų ( $F_{ST}$ ) tarp analizuojamų populiacijų matrica (žr. IV priedo 1 lentelę) ir kiekvienos populiacijos ilgalaikio  $Ne$  sąrašo atskiras failas. Weir ir Cockheram genetiniai atstumai,  $F_{ST}$ , nustatyti 4P v1.0 programa [93].

HGDP-CEPH populiacijų  $Ne$  įverčius nustatė M. Mezzavilla ir S. Ghirrotto'as (*University of Trieste, Italy*) 2015 m. R programinės įrangos *NeON* paketu (žr. 4.1 lentelę) [91]. Duomenų vizualizavimas atliktas R programinės įrangos *Phangorn* biblioteka [94].

#### 2.5.4. Trumpalaikio efektyvaus populiacijos dydžio analizė

Trumpalaikis efektyvus populiacijos dydis nustatytas neparametriniu metodu, kuris naudoja identiškų pagal kilmę (angl. *Identical By Descent*, IBD) segmentus ir remiasi Wright'o-Fisher'io diskrečių kartų modeliu, įdiegtu į atviros prieigos programinį paketą *IBDNe* v.04Sep15.e78, publikuotą S. R. Browning'o ir B. L. Browning'o 2015 m. [29, 95]. Trumpalaikiam efektyviam populiacijos dydžiui nustatyti pasirinkta kita programa, kadangi *NeON* paketas nepritaikytas tokiems skaičiavimams. Šis metodas skaičiuoja tikėtiną laiko pasiskirstymą kartomis (karta – 25 metai), iki artimiausio bendro protėvio (angl. *Most Recent Common Ancestor*, TMRCA) pagal IBD segmentų ilgį. Pagal kiekvienam TMRCA priskirtų IBD segmentų kiekį yra apskaičiuojamas efektyvus populiacijos dydis konkrečiam TMRCA. Naudojant plataus masto VNP genotipavimo duomenis, trumpalaikis  $Ne$  apskaičiuojamas 50-iai kartų nuo nulinės kartos (dabarties).

IBD segmentai nustatyti *IBDseq* v.r1206 programiniu paketu, pasirinktas IBD segmentų ilgis – 7 cM [95]. Mažesni nei 7 cM IBD segmentai buvo pašalinti. Toliau nustatomi genetiniai atstumai panaudojant *HapMap* genolapį. Genominės sritys, kuriose nustatyta itin daug IBD segmentų, yra pašalinamos, dažniausiai tai yra centromerinės ir telomerinės sritys. Taip pat pašalinami giminingi asmenys.

*IBDNe* v.04Sep15.e78 programa trumpalaikiui  $Ne$  nustatyti nuskaito IBD segmentus, nustatytus *IBDseq* v.r1206 programa, atlieka IBD segmentų ir genominių sričių filtravimą ir apskaičiuoja trumpalaikį  $Ne$  su 95 proc. pasikliautinaisiais intervalais kiekvienai kartai (karta = 0, 1, 2, 3 ... 50). Detalus analizės protokolas pateiktas IV priedo II dalyje.

### 3. TYRIMO REZULTATAI

#### 3.1. Lietuvos populiacijos genetinės struktūros nustatymas ir įvertinimas kitų populiacijų atžvilgiu pagal plataus masto genotipavimo duomenis

Pirmiausia atlikta genetinių panašumų tarp Lietuvos populiacijos porų individų analizė, apskaičiuavus giminingumo ir inbrydingo (F) koeficientus (žr. 3.1 lentelę). Vidutinis giminingumo koeficientas tarp tirtų Lietuvos populiacijos asmenų nustatytas 0,00075, vidutinis F – 0,0022 [49]. Iš viso nustatyti 4 asmenys, kurių F buvo didesnis negu būdingas antros kartos pusbrolių santuokai (0,0156). Giminingi asmenys pašalinti iš tolesnės analizės.

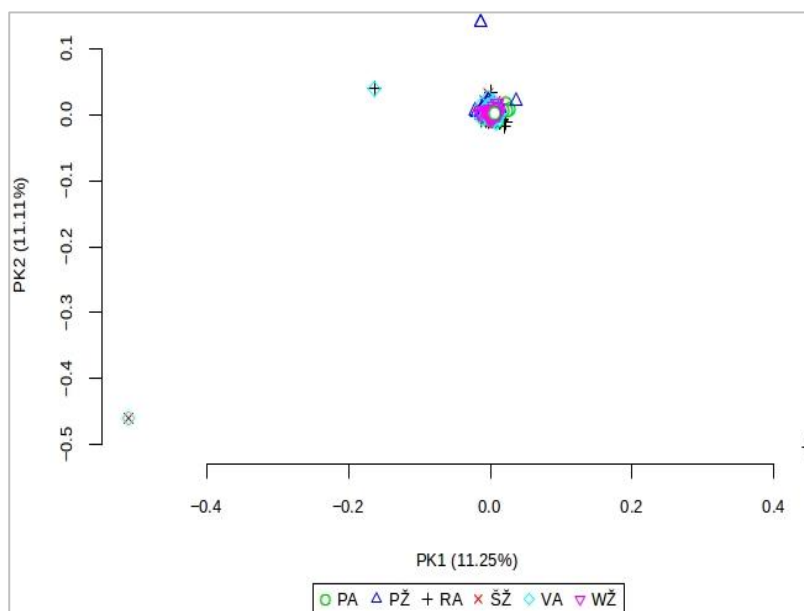
**3.1 lentelė.** Vidutiniai giminingumo ir inbrydingo koeficientai, nustatyti tarp Lietuvos populiacijos tirtų asmenų

Populiacija	Tiriamieji	Vidutinis giminingumo koeficientas	Vidutinis inbrydingo koeficientas (F)	Asmenys F > 0,0156
LT	424	0,00075	0,0022	4 (0,7 proc.)

Lietuvos populiacijos šešių etnolingvistinių grupių genetiniams ryšiams nustatyti atlikta PKA, panaudojus tolygiai po visą genomą pasiskirsčiusių 424 asmenų 232 752 VNP duomenis. Atlikus pirmų dviejų pagrindinių komponentų (PK) vizualizavimą, identifikuotos aštuonios išskirtys, kurios buvo pašalintos ir nevertintos tolesnėje analizėje. Pirmos dvi pagrindinės komponentės, kurios paaiškina 22,36 proc. genetinės įvairovės, parodė, kad asmenys iš šešių Lietuvos etnolingvistinių grupių suformuoja vieną bendrą klasterį. Gauti rezultatai parodo Lietuvos populiacijos homogeniškumą (žr. 3.1 pav.). Taip pat Lietuvos populiacijos etnolingvistinių grupių diferenciacija įvertinta apskaičiuavus  $F_{ST}$  atstumus (žr. 3.2 lentelę).

**3.2 lentelė.** Weir'o ir Cockerham'o  $F_{ST}$  reikšmės tarp tirtų Lietuvos populiacijos etnolingvistinių grupių. ŠŽ – Šiaurės žemaičiai, PŽ – Pietų žemaičiai, VŽ – Vakarų žemaičiai, VA – Vakarų aukštaičiai, PA – Pietų aukštaičiai, RA – Rytų aukštaičiai

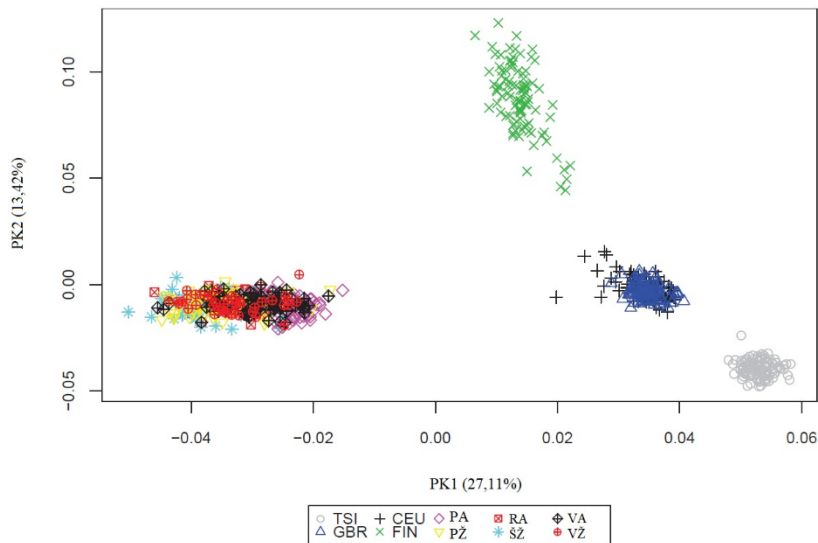
	PA	PŽ	RA	ŠŽ	VA	VŽ
PA	0	0,002586	0,002765	0,002738	0,002201	0,003884
PŽ	0,002586	0	0,002485	0,001886	0,00217	0,00333
RA	0,002765	0,002485	0	0,002236	0,001986	0,003764
ŠŽ	0,002738	0,001886	0,002236	0	0,002006	0,003293
VA	0,002201	0,00217	0,001986	0,002006	0	0,003641
VŽ	0,003884	0,00333	0,003764	0,003293	0,003641	0



**3.1 pav.** Pagrindinių komponentių analizė, panaudojus asmenų iš šešių Lietuvos populiacijos etnolingvistinių grupių genotipavimo duomenis. PA – Pietų aukštaičiai; PŽ – Pietų žemaičiai; RA – Rytų aukštaičiai; ŠŽ – Šiaurės žemaičiai; VA – Vakarų aukštaičiai; VŽ – Vakarų žemaičiai

Lietuvos populiacijos genetinei struktūrai nustatyti ir įvertinti Europos populiacijų atžvilgiu pagal plataus masto genotipavimo duomenis atlikta PKA, įtraukus 1 000 genomų projekto CEU (99 tiriamieji), TSI (109), FIN

(99) ir GBR (99) populiacijų duomenis (žr. 3.2 pav.) [4]. Analizėje naudoti 158 633 VNP, esantys pusiausvyroje sankiboje. Pirma PK, paaiškinanti 27,11 proc. genetinės įvairovės, atskiria lietuvius nuo kitų analizuotų Europos populiacijų, antra PK, paaiškinanti 13,42 proc. genetinės įvairovės, atskiria FIN populiaciją, kuri turi didžiausią duomenų sklaidą grafike, palyginti su kitomis populiacijomis. GBR su CEU formuoja vieną klasterį, TSI artimesnė CEU ir GBR, palyginti su LT ir FIN populiacijomis. Lietuvių populiacijos klasteris išsidėsto arčiau CEU ir GBR, palyginti su FIN ir TSI. Dėl skirtingo imties dydžio įtakos PK analizėje, ji buvo pakartota į analizę įtraukus 100 atsitiktiniu būdu atrinktų Lietuvos populiacijos individų (žr. IV priedo 1 pav.). Gauti rezultatai sutampa su pirminės PK analizės rezultatais. Skirtingas mėginių skaičius neturi įtakos gautiems rezultatams.



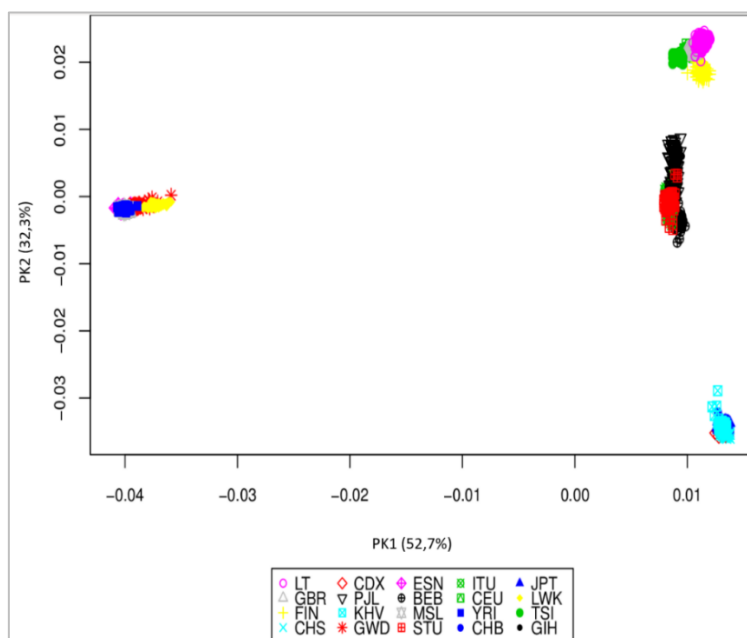
**3.2 pav.** Pagrindinių komponentių analizė, panaudojus LT ir keturių 1 000 genomų projekto populiacijų (GBR, FIN, TSI ir CEU) VNP duomenis

Analizuotų Europos populiacijų diferenciacija įvertinta apskaičiuotais  $F_{ST}$  atstumais (žr. 3.3 lentelę).  $F_{ST}$  reikšmės tarp LT ir CEU bei LT ir GBR buvo panašios – 0,006 ir 0,007.  $F_{ST}$  tarp LT ir FIN buvo 0,008, didžiausia diferenciacija nustatyta tarp LT ir TSI bei FIN ir TSI, kur  $F_{ST} = 0,011$ . Šie rezultatai koreliuoja su PKA rezultatais, kur Lietuvos populiacija artimesnė CEU ir GBR, didžiausi atstumai stebimi tarp LT ir TSI bei FIN ir TSI.

**3.3 lentelė.** Weir'o ir Cockerham'o  $F_{ST}$  reikšmės tarp tirtų Lietuvos ir 1 000 genomų projekto Europos populiacijų

	LT	CEU	GBR	FIN	TSI
LT	0				
CEU	0,006	0			
GBR	0,007	0,0003	0		
FIN	0,008	0,006	0,007	0	
TSI	0,011	0,003	0,004	0,011	0

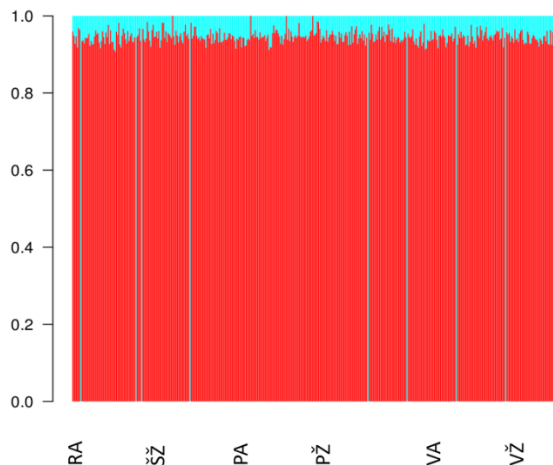
Lietuvos populiacijos genetinei struktūrai nustatyti ir įvertinti ne tik Europos, bet ir kitų pagrindinių geografinių regionų atžvilgiu, atlikta PKA, įtraukus 1 000 genomų projekto 19 populiacijų duomenis (žr. 3.3 pav.) [4]. Amerikos populiacijos neįtrauktos į analizę dėl sudėtinės genetinės kilmės. Analizėje naudoti 187 447 VNP, esantys pusiausvyroje sankiboje. PK1 paaikškino 52,71 proc., o PK2 – 32,3 proc. genetinės įvairovės ir parodė, kad visos analizuotos populiacijos klasterizuojasi atitinkamai pagal savo žemyną. Kaip ir buvo tikėtasi, Lietuvos populiacija formuoja klasterį su kitomis Europos populiacijomis (žr. 3.3 pav.).



**3.3 pav.** Pagrindinių komponentų analizė, panaudojus LT ir 1 000 genomų projekto 19 populiacijų VNP duomenis

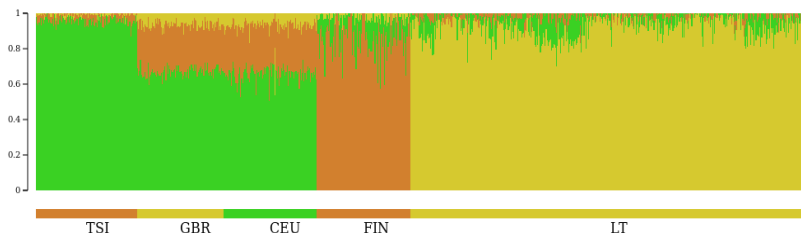
Lietuvos populiacijos struktūrai ir genetiniams komponentams nustatyti ir įvertinti atlikta priemaišos analizė ADMIXTURE programos klasterizavimo algoritmu, taikant didžiausio tikėtimumo metodą, analizuojant Lietuvos 424 individų iš šešių etnolingvistinių grupių 232 752 VNP duomenis (žr. 3.4 pav.). Tikslas – sugrupuoti skirtingus individus į K skirtingų klasterių, remiantis genetiniu panašumu. Skaičiavimai klasterizavimo algoritmu atliekami skirtingam klasterių skaičiui, priklausomai nuo analizuojamų populiacijų skaičiaus. Geriausias K pasirenkamas pagal mažiausią apskaičiuotą kryžminės duomenų kontrolės įvertį (CV) [21]. Kryžminis duomenų kontrolės įvertis apskaičiuojamas padalijant analizuojamus genotipus pagal nutylėjimą į  $v = 5$  lygias grupes. Atliekant vienos genotipų grupės skaičiavimus, kitos grupės genotipai laikinai yra užmaskuojami, o jų duomenys (G) naudojami apskaičiuoti įverčiams  $\theta = (Q, P)$ . Kiekvienas užmaskuotas genotipas  $g_{ij}$  apskaičiuojamas pagal formulę:  $\mu_{ij} = E[g_{ij}|Q, P] = 2 \sum_k q_{ik} P_{kj}$ , o prognozuojamo genotipo klaida apskaičiuojama pagal formulę:  $d(n_{ij}, \mu_{ij}) = n_{ij} \log(n_{ij}/\mu_{ij}) + (2 - n_{ij}) \log[(2 - n_{ij})/(2 - \mu_{ij})]$ .

Šešių etnolingvistinių grupių K analizuotas nuo 2 iki 5. Pagal apskaičiuotus etnolingvistinių grupių CV ( $K = 2$ , CV = 0,52766;  $K = 3$ , CV = 0,53040;  $K = 4$ , CV = 0,53283;  $K = 5$ , CV = 0,53407) geriausiai aprašomas dviejų klasterių modelis. Priemaišos analizės rezultatai parodė Lietuvos šešių etnolingvistinių grupių vieną didžiausią bendrą genetinį komponentą (raudona spalva). Šiais rezultatais dar kartą yra patvirtinamas Lietuvos populiacijos homogeniškumas.



**3.4 pav.** Lietuvos šešių etnolingvistinių grupių priemaišos analizė, kai  $K = 2$ . Kiekviena vertikali linija reprezentuoja individo kilmės proporciją. Žalios vertikalios linijos yra išskirtys, kurios buvo nustatytos ir PKA metodu. RA – Rytų aukštaičiai; ŠŽ – Šiaurės žemaičiai; PA – Pietų aukštaičiai; PŽ – Pietų žemaičiai; VA – Vakarų aukštaičiai; VŽ – Vakarų žemaičiai

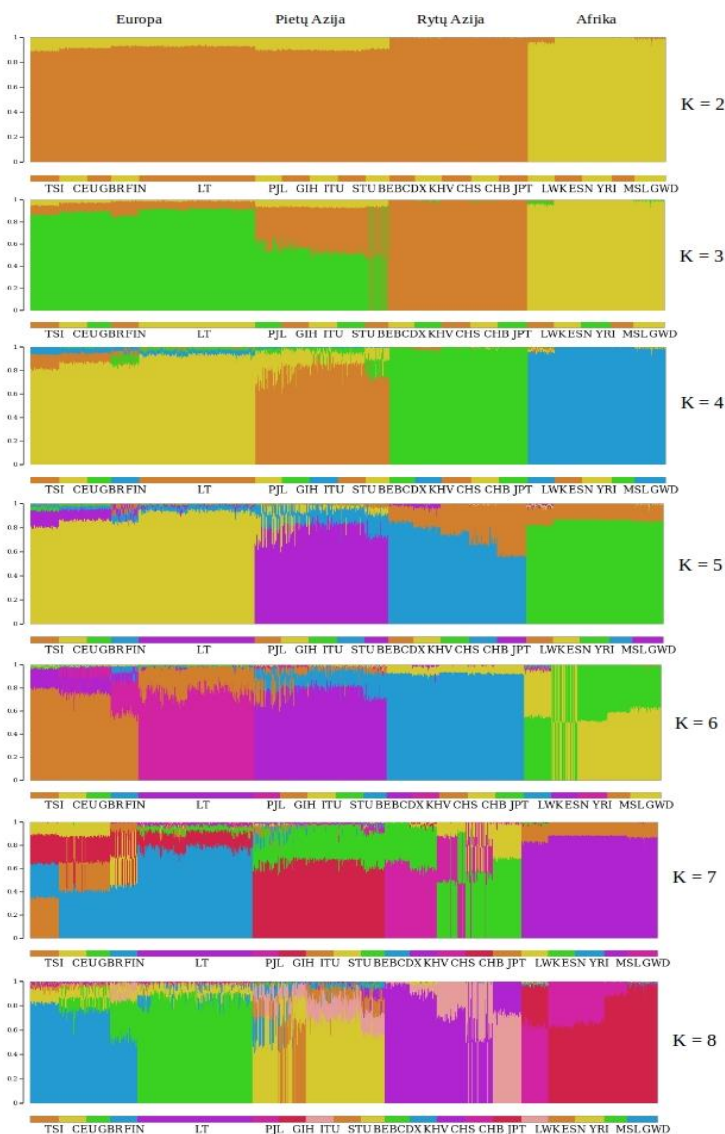
Lietuvos populiacijos priemaišos analizė atlikta Europos kontekste, įtraukus į analizę CEU, GBR, FIN ir TSI populiacijas iš 1 000 genomų projekto [4]. Pagal apskaičiuotus CV geriausiai aprašomas trijų klasterių modelis ( $CV = 0,54455$ ) (žr. 3.5 pav.). Kai  $K = 3$ , Lietuvos populiacijoje išsiskiria vienas pagrindinis genetinis komponentas (geltona spalva). CEU, GBR ir TSI populiacijoms tenka antras didžiausias genetinis komponentas (žalia spalva), kuris Lietuvos populiacijoje pasireiškia nedaug. FIN populiacijai tenka didžiausia dalis trečio genetinio komponento (ruda spalva), kuris stebimas ir kitose analizuotose Europos populiacijose (GBR, CEU), išskyrus Lietuvos populiaciją, kurioje šis komponentas pasireiškia nedideliu santykiu.



**3.5 pav.** Individų iš LT, CEU, GBR, FIN ir TSI populiacijos priemaišos analizė, kai  $K = 3$ . Kiekvienas individas reprezentuotas vertikalia spalvota linija, kurioje skirtingų spalvų segmentai atspindi individo kilmės proporciją, kilusią iš vienos  $K$  populiacijų

Globalus kilmės profilis nustatytas įtraukus į priemaišos analizę 1 000 geno projekto 19 populiacijų duomenis [4]. Analizuoti 2 317 asmenų iš 20 pasaulio populiacijų, įskaitant Lietuvos, 187 447 VNP genetiniai duomenys. Mažiausias apskaičiuotas CV parametro įvertis nustatytas aštuoniems kilmės komponentams ( $K = 2$ ,  $CV = 0,55796$ ;  $K = 3$ ,  $CV = 0,53075$ ;  $K = 4$ ,  $CV = 0,52453$ ;  $K = 5$ ,  $CV = 0,52410$ ;  $K = 6$ ,  $CV = 0,52339$ ;  $K = 7$ ,  $CV = 0,52309$ ;  $K = 8$ ,  $CV = 0,52275$ ;  $K = 9$ ,  $CV = 0,52316$ ). Kai  $K = 2$ , atsiskiria dvi populiacijų grupės: Afrika ir Rytų Azija (geltona spalva), Europa ir Pietų Azija (oranžinė spalva) (žr. 3.6 pav.). Kai  $K = 3$ , atsiranda naujas genetinis komponentas (žalia spalva), kuris išskiria Europą kartu su Pietų Azija nuo Rytų Azijos (oranžinė spalva) ir Afrikos (geltona spalva). Esant keturiems kilmės komponentams, naujas genetinis komponentas (geltona spalva) išskiria Europos populiacijas nuo Pietų Azijos populiacijų ir tik, esant  $K = 6$ , atsiranda naujas genetinis komponentas (fuksijos spalva), kurio didžiausias santykis išskirtas Lietuvos populiacijoje. Kai  $K = 8$  (mažiausias CV), Lietuvos populiacijoje atsiranda dominuojantis protėvinis genetinis komponentas (žalia spalva), kurio nedidelis santykis randamas ir kitose gretimose analizuotose Europos populiacijose (CEU, FIN, GBR). Taip pat Lietuvos populiacijoje stebimi Pietų Azijos ir Afrikos protėvių genetiniai komponentai. Dėl skirtingo

imties dydžio įtakos priemaišos analizėje ji buvo pakartota, į analizę įtraukus 120 atsitiktinių būdu atrinktų Lietuvos populiacijos individų (žr. IV priedo 2 pav.). Gauti rezultatai sutampa su pirminės priemaišos analizės rezultatais. Skirtingas mėginių skaičius neturi įtakos gautiems rezultatams.

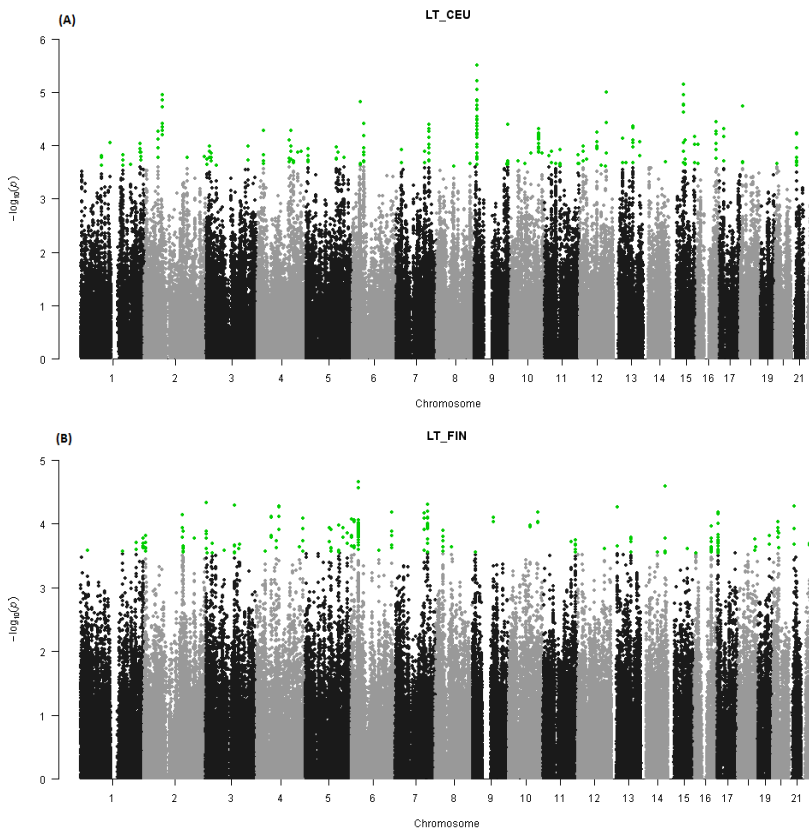


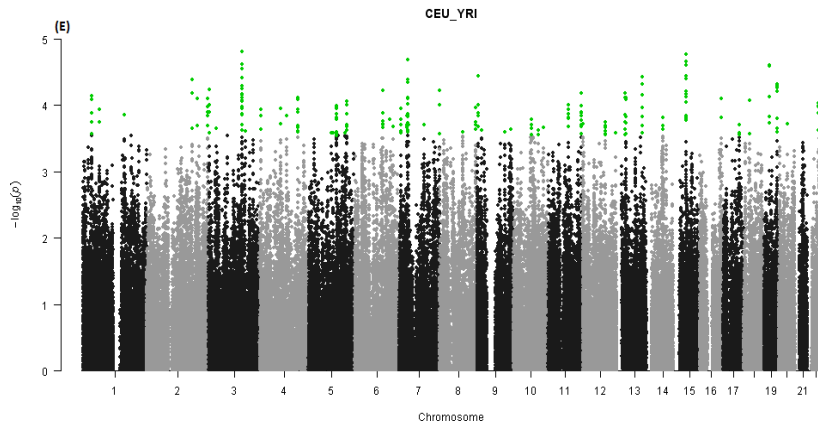
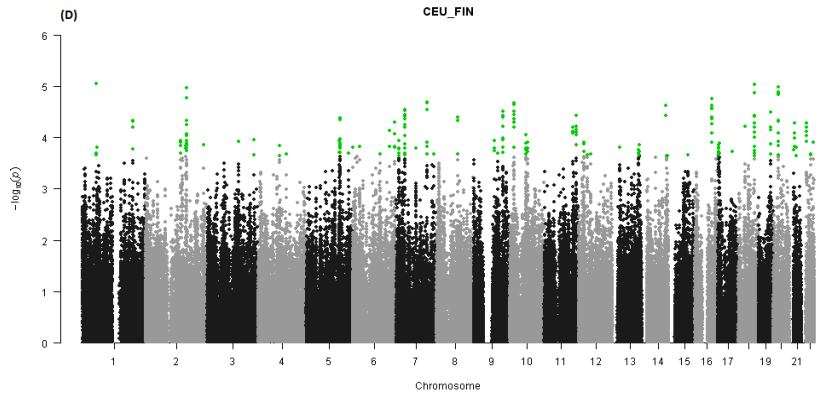
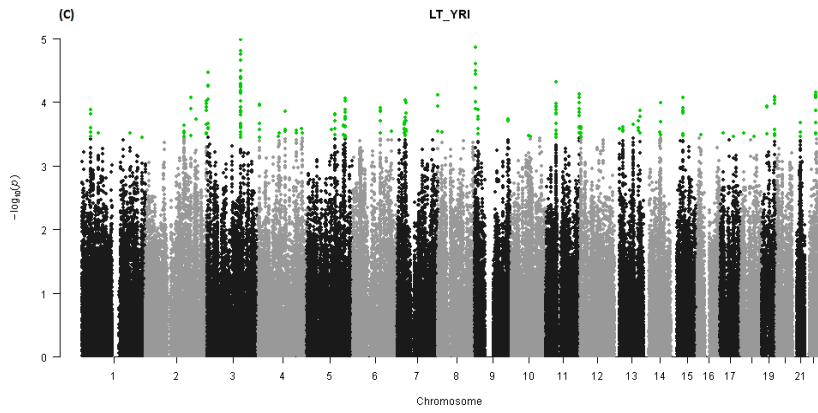
**3.6 pav.** Individų iš LT ir 1 000 genomų projekto 19 populiacijų priemaišos analizė, kai K yra nuo 2 iki 8. Kiekvienas individas reprezentuotas vertikalia spalvota linija, kurioje skirtingų spalvų segmentai atspindi individo kilmės proporciją, kilusią iš vienos K populiacijų

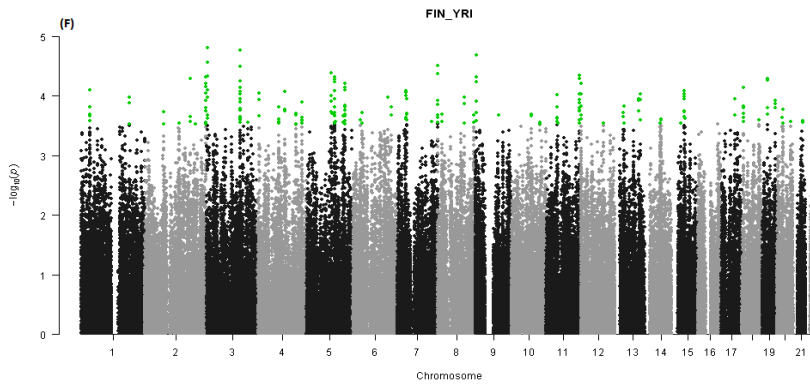


### 3.2. Kandidatinių teigiamos gamtinės atrankos veikiamų genomo sričių nustatymas Lietuvos populiacijoje

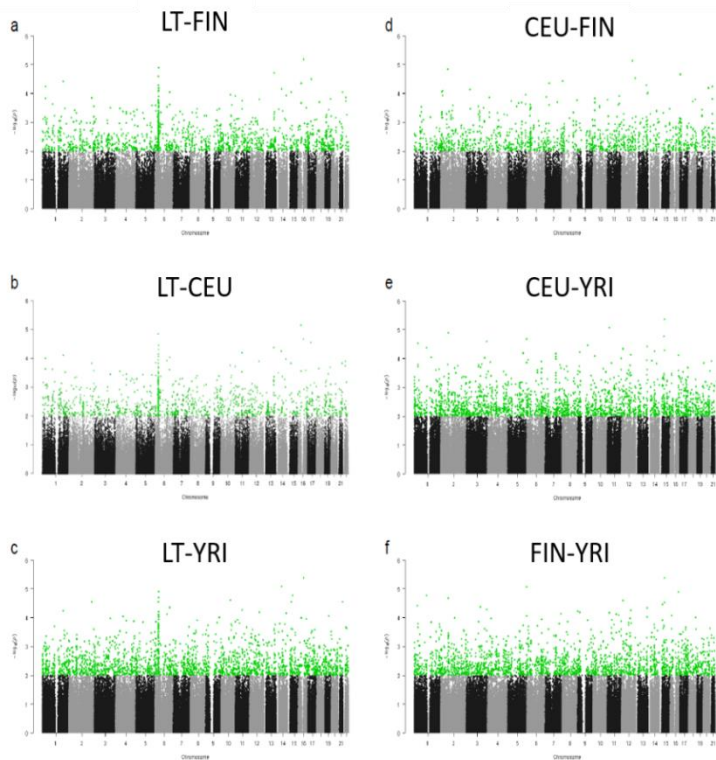
Tirtoje lietuvių populiacijoje nustatytos teigiamos gamtinės atrankos veikiamos genomo sritys, remiantis plataus masto VNP genotipavimo duomenimis. Specifinės Lietuvos populiacijai genomo sritys, veikiamos netolimos gamtinės atrankos, nustatytos  $F_{ST}$  ir XP-EHH statistiniais metodais tarp porų populiacijų: LT ir CEU, LT ir FIN, LT ir YRI, CEU ir FIN, CEU ir YRI, FIN ir YRI. Porų populiacijų nustatytų gamtinės atrankos signalų XP-EHH ir  $F_{ST}$  metodais pasiskirstymas pateiktas *Manhattan* grafike 3.7 ir 3.8 paveiksluose. Išsamūs rezultatai pateikti IV priedo 1 lentelėje. Gamtinės atrankos veikiama kandidatine genomo sritimi buvo laikoma ta, kuri buvo tarp 0,1 proc. XP-EHH įverčių išskirčių, srityje buvo  $\geq 2$  VNP ir bent vieno VNP  $F_{ST}$  p reikšmė buvo  $<0,01$  (žr. 3.4 lentelę). Analizuojant nustatytas gamtinės atrankos kandidatines sritis, daugiau dėmesio buvo skirta toms sritims, kurių genai susiję su žmonių adaptacija: atsparumu patogenams, dieta, klimatu ar pigmentacija [71].







**3.7 pav.** Teigiamos gamtinės atrankos nustatytų signalų XP-EHH metodu pasiskirstymas autosomose tarp porų populiacijų: LT ir CEU (A), LT ir FIN (B), LT ir YRI (C), CEU ir FIN (D), CEU ir YRI (E), FIN ir YRI (F). Žalios spalvos taškai nurodo VNP, kurie atsiranda 0,1 proc. nustatytų išskirčių



**3.8 pav.** Teigiamos gamtinės atrankos nustatytų signalų  $F_{ST}$  metodu pasiskirstymas autosomose tarp porų populiacijų: LT ir CEU (A), LT ir FIN (B), LT ir YRI (C), CEU ir FIN (D), CEU ir YRI (E), FIN ir YRI (F). Žali taškai nurodo VNP  $F_{ST}$ , kurių  $p < 0,01$

Iš viso Lietuvos populiacijoje nustatytos 32 kandidatinės genomo sritys, veikiamos artimos teigiamos gamtinės atrankos (žr. 3.4 lentelę). Daugelis genomo sričių buvo nustatytos lyginant Lietuvos populiaciją su CEU arba FIN populiacijomis. Keturios sritys nustatytos ir kitose analizuotose Europos populiacijose: trys jų bendros LT ir CEU, o viena – LT ir FIN (žr. 3.4 lentelę). Tarp bendrų signalų buvo *SLC24A5* genas (15 chromosoma), susijęs su šviesios odos pigmentacija, kuris yra taikinyš, veikiamas netolimos teigiamos gamtinės atrankos ne Afrikos populiacijose [94]. *SLC24A5* geno tarpgeninis variantas rs1834640, kuris atsakingas už odos pigmentaciją, buvo tarp šešių XP-EHH ir  $F_{ST}$  reikšmingų įverčių, nustatytų lyginant CEU ir YRI populiacijas.

**3.4 lentelė.** Kandidatinės gamtinės atrankos veikiamos genomo sritys, nustatytos  $F_{ST}$  ir XP-EHH metodais Lietuvos populiacijoje

Genomo koordinatės	VNP <sup>1</sup>	Genai	Populiacija
chr1:35484861–35635453	3	<i>ZMYM6;ZMYM1; SFPQ;</i>	LT ir YRI
chr1:36549157–36562551	3	<i>TEKT2;COL8A2</i>	LT ir YRI
chr1:231866461–231908347	3	<i>TSNAX-DISCI</i>	LT ir CEU
chr1:245363423–247401645	4	<i>KIF26B</i>	LT ir FIN
chr2:56070352–56161538	4	<i>PNPT1;EFEMP1</i>	LT ir CEU
chr3:2347409–2355167	5	<i>CNTN4</i>	LT ir FIN
chr3:112511598–112816192	5	<i>LOC101929694; CD200RIL</i>	LT ir FIN
chr3:130243249–131445951	2	<i>COL6A5;COL6A6; CPNE4</i>	LT ir FIN
chr4:106552414–106594625	3	<i>ARHGEF38;</i>	LT ir YRI arba FIN ir YRI
chr4:89629000–89669852	4	<i>HERC3;FAM13A</i>	LT ir FIN
chr4:181637915–181691738	4	<i>NONE;LINC00290</i>	LT ir FIN
chr5:10073956–10109791	4	<i>LOC285692; FAM173B</i>	LT ir CEU
chr5:133177588–133638034	4	<i>FSTL4;WSPAR; CDKL3</i>	LT ir FIN
chr6:11710405–11717934	4	<i>TMEM170B;ADTRP</i>	LT ir FIN
chr6:27811815–28096280	42	<i>HIST1H2BN; HIST1H2AL; HIST1H1B;HIST1H4L;HIST1H3J; HIST1H2BO;OR2B2;OR2B2;OR2 B6; ZNF165;</i>	LT ir FIN

Genomo koordinatės	VNP <sup>1</sup>	Genai	Populiacija
		<i>ZSCAN12P1;</i> <i>ZSCAN16-AS1;GPX5</i>	
chr6:46596286–46834685	8	<i>CYP39A1;PLA2G7;</i> <i>MEP1A;ADGRF5</i>	LT ir CEU
chr7:30276161–30415102	5	<i>MTURN;ZNRF2;</i> <i>DKFZP586I1420;</i> <i>LINC01176</i>	LT ir YRI
chr7:112994204–113020044	4	<i>LINC00998;</i> <i>PPP1R3A</i>	LT ir FIN
chr7:129187430–129350170	11	<i>SMKRI1;NRF1</i>	LT ir CEU
chr9:12352971–12537279	29	<i>PTPRD-AS2;TYRP1</i>	LT ir CEU
chr10:116362835–116377962	4	<i>ABLIM1</i>	LT ir FIN
chr11:129832270–130645506	11	<i>PRDM10;LINC00167;APLP2;C11</i> <i>orf44;</i> <i>LOC100507431</i>	LT ir YRI arba CEU ir YRI
chr11:41897826–42307166	10	<i>LINC01499;LOC100507205;</i> <i>;HNRNPKP3</i>	LT ir YRI
chr12:68640583–68654780	6	<i>IL26;IL22</i>	LT ir CEU
chr12:103326468–103394540	5	<i>PAH;ASCL1</i>	LT ir CEU
chr15:48389924–48686175	8	<i>LINC01491;SLC24A5;SLC12A1;D</i> <i>UT;</i> <i>FBN1</i>	LT ir YRI arba CEU ir YRI
chr15:51639821–51673630	9	<i>GLDN</i>	LT ir CEU
chr15:58579956–58588116	4	<i>AQP9;LIPC</i>	LT ir CEU
chr16:66962411–67028008	7	<i>RRAD;FAM96B;</i> <i>CES2;CES3;CES4A</i>	LT ir FIN
chr17:4722606–4885982	15	<i>PLD2;MINK1;CHRNE;GPI1BA;SL</i> <i>C25A1;RNF167;ENO3;</i> <i>SPAG7;CAMTA2</i>	LT ir FIN
chr20:19702049–19706617	6	<i>SLC24A3;RIN2</i>	LT ir FIN arba CEU ir FIN
chr21:21389233–21568418	8	<i>MIR548XHG;</i> <i>LINC00320</i>	LT ir CEU

<sup>1</sup>Reikšmingi VNP, esantys genomo srityje, kuri buvo nustatyta XP-EHH metodu, ir yra tarp 0,1 proc. išskirčių bei mažiausiai vieno VNP.  $F_{ST}$  p reikšmė < 0,01.

Tarp nustatytų specifinių lietuvių populiacijai kandidatinių teigiamos gamtinės atrankos veikiamų genomo sričių stipriausias signalas, 284 kb dydžio, identifiukuotas 6 chromosomos 27811815–28096280 srityje, analizuojant LT ir FIN populiacijas. Genai šioje srityje priklauso histonų ir

uoslės receptorių genų šeimai. Atlikus variantų anotaciją, nustatytos dvi nesinoniminės pakaitos *OR2B6* gene: rs7767176 (exon1:c.G349A) ir rs9380030 (exon1:c.A809G) su XP-EHH įverčiais 3,8. rs7767176  $F_{ST}$  lygi nuliui, o rs9380030  $F_{ST}$  lygi 0,006, rs7767176 CADD nustatytas nereikšmingas, rs9380030 CADD įvertis lygus 27.

Kitas stiprus signalas, ~225 kb dydžio, nustatytas 3 chromosomoje, lyginant LT ir YRI populiacijas. Srityje yra *COL6A5* ir *COL6A6* genai, koduojantys VI tipo klageno alfa 5 ir alfa 6 grandines. Nustatyta *COL6A5* geno nesinoniminė pakaita (rs12488457), turinti CADD įvertį 23,2, kuri buvo tarp reikšmingų XP-EHH ir  $F_{ST}$  įverčių. Nors šis signalas kaip reikšmingas buvo nustatytas lyginant LT ir YRI, Europos populiacijose šio alelio dažnis yra aukštas (>0,748), tačiau YRI populiacijoje žemas (~0,013). Dar vienas signalas, lyginant LT ir YRI populiacijas, rastas 1 chromosomoje su dviem variantais (rs274750 ir rs274752), esančiais *COL8A2* geno 3' UTR srityje, kurie XP-EHH ir  $F_{ST}$  metodais buvo nustatyti kaip reikšmingos išskirtys. *COL8A2* genas koduoja VIII tipo klageno alfa 2 grandinę. Signalas buvo reikšmingas tik lyginant LT ir YRI populiacijas, tačiau pakitę (angl. *derived*) rs274750 ir rs274752 variantų C ir A aleliai yra fiksuoti (~1) tirtose trijose Europos populiacijose ir randami vidutiniu dažniu (0,5) YRI populiacijoje. *COL6A5* geno polimorfizmai asocijuoti su kūno masės indeksu ir odos fenotipu, pvz., egzema, o mutacijos šiame gene asocijuotos su ragenos distrofijomis [96, 97].

Nesinoniminė pakaita nustatyta *PLA2G7* gene (rs1805018, exon7:c.T593C), kuri yra 6 chromosomos 46596286–46834685 srityje. Šis genas lemia polinkį sirgti astma – viena dažniausių lėtinių kvėpavimo takų ligų, pasireiškiančių vaikams ir jauniems asmenims.

Kitas stiprus signalas nustatytas 9 chromosomos tarpgeninėje srityje (12352971–12537279), kurioje iš vienos pusės yra *PTPRD-AS2* genas, kitoje pusėje – *TYRPI* genas. *TYRPI* genas koduoja melanosomų fermentą, kuris dalyvauja melanino biosintezėje. Šis genas atsakingas už šviesios odos pigmentaciją ir yra genas kandidatas, veikiamas adaptyvios gamtinės atrankos tik Europos populiacijose [98]. Tačiau tarp nustatytų reikšmingų VNP išskirčių srityje nebuvo susijusių su tam tikra funkcija (nesinoniminiai variantai).

Nesinoniminius VNP nustatytas *MINK1* gene (rs2302319), serino / treonino kinazė, kuri neigiamai reguliuoja Ras ir Rap2 signalo transdukciją, kontroliuojant neuronų struktūrą ir AMPA receptoriaus veiklą. Taip pat nesinoniminius VNP nustatytas *ENO3* gene (rs238238), kurio funkcija – skersaruožių raumenų vystymasis ir regeneracija, taip pat gali būti atsakingas už Glikogeno kaupimosi ligą (GSD13), metabolinį sutrikimą, būdingas mialgijas, bendrą raumenų silpnumą ir nuovargį [99].

Taip pat buvo nustatyti du signalai, susiję su imunitetu: vienas 6 chromosomoje su *IL26* ir *IL22* interleukino genais, o kitas – 12 chromosomoje su *BRD2* ir *HLA* genais.

Atsižvelgiant į empirinį p reikšmių, mažesnių negu 0,01, pasiskirstymą, identifikuotos 36 genomo sritys, turinčios ekstremaliai neigiamus Tajima's D įverčius, kaip galimos Lietuvos populiacijos kandidatinės gamtinės atrankos sritys (žr. 4.5 lentelę). Taip pat Tajima's D analizė atlikta CEU ir FIN populiacijoms, kad būtų galima palyginti ir nustatyti Lietuvos populiacijos specifines sritis (žr. 3.9 pav.). Iš viso identifikuota 15 specifinių Lietuvos populiacijos sričių.

**3.5 lentelė.** Kandidatinės gamtinės atrankos sritys, nustatytos Tajima's D metodu

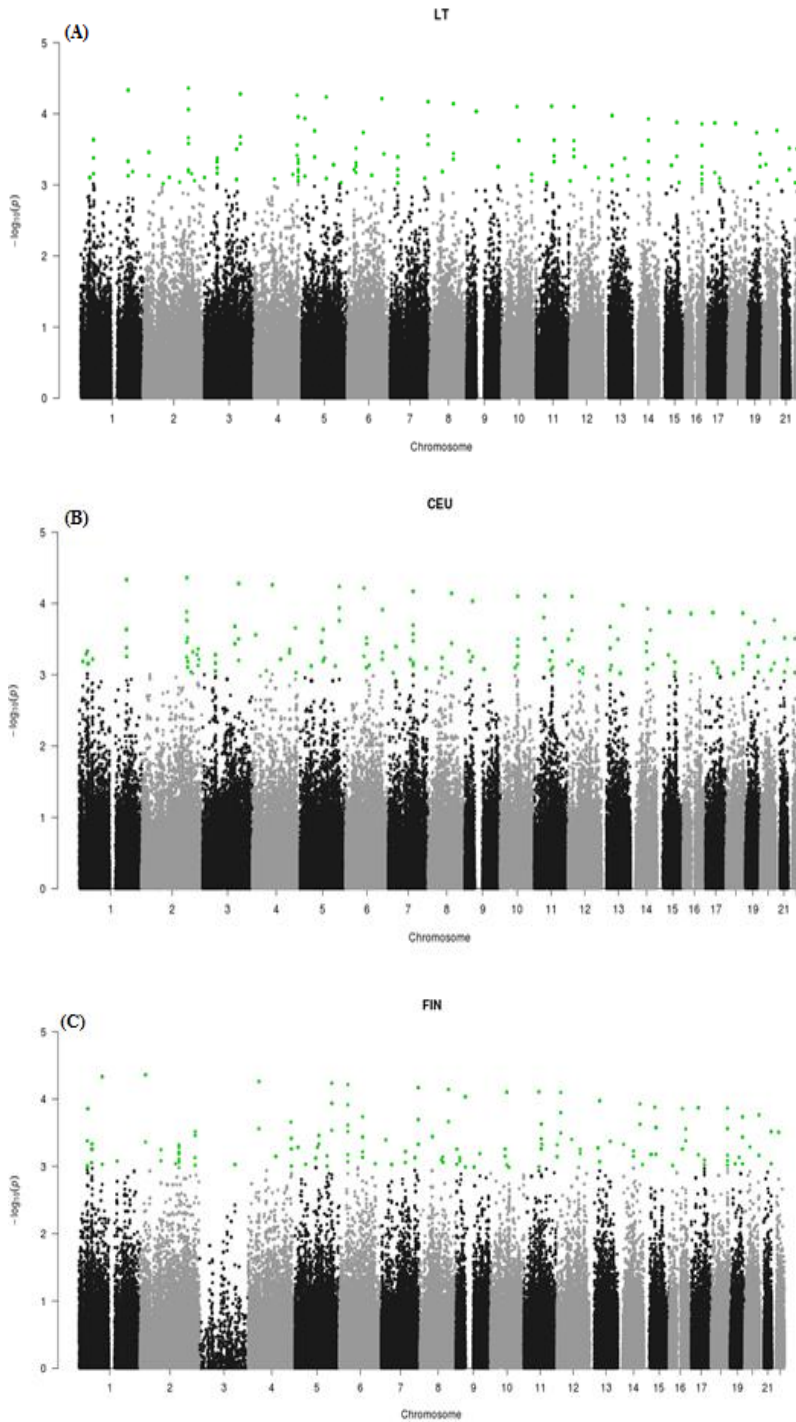
Genomo koordinatės	Langas*	P reikšmė	Sritis	Genai	Populiacija
chr1:35818960–35948959	4 (22)	0,0008	egzoninė	<i>KIAA0319L</i> ; <i>ZMYM4</i>	LT
chr1:49988960–50728959	6 (10)	0,0003	egzoninė	<i>AGBL4</i> ; <i>ELAVL4</i>	LT
chr1:188788960–188968959	9 (17)	0,0003	intergeninė	<i>LINC01037</i> ; <i>BRINP3</i>	LT; CEU
chr2:21728675–21888674	7 (7)	0,0004	intergeninė	<i>TDRD15</i> ; <i>LINC01822</i>	LT; FIN
chr2:179468675–179648674	9 (10)	0,0003	egzoninė	<i>TTN</i>	LT; CEU
chr3:50343412–51893411	4 (17)	0,0006	egzoninė	<i>C3orf18</i> ; <i>CACNA2D2</i> ; <i>CISH</i> ; <i>CYB561D2</i> ; <i>DCAF1</i> ; <i>DOCK3</i> ; <i>GRM2</i> ; <i>HEMK1</i> ; <i>HYAL2</i> ; <i>IQCF3</i> ; <i>IQCF6</i> ; <i>MANF</i> ; <i>MAPKAPK3</i> ; <i>NPRL2</i> ; <i>RAD54L2</i> ; <i>RASSF1</i> ; <i>RBM15B</i> ; <i>TEX264</i> ; <i>TMEM115</i> ; <i>TUSC2</i> ; <i>ZMYND10</i>	LT; CEU

Genomo koordinatės	Langas*	P reikšmė	Sritis	Genai	Populiacija
chr3:128763412–128903411	5 (13)	0,0006	egzoninė	<i>CNBP;GP9;ISY1;ISY1-RAB43;RAB43</i>	LT;CEU;FIN
chr3:143543412–143683411	5 (8)	0,0001	egzoninė	<i>SLC9A9</i>	LT;CEU;FIN
chr4:171930684–172360683	4 (9)	0,0002	ncRNA_egzoninė	<i>LINC02431;MIR6082</i>	LT
chr4:176190684–176400683	11 (15)	0,0005	intergeninė	<i>ADAM29;GPM6A</i>	LT; FIN
chr5:50531164–50691163	6 (7)	0,0003	egzoninė	<i>ISL1</i>	LT
chr5:126311164–126441163	4 (7)	0,0005	egzoninė	<i>C5orf63</i>	LT
chr6:35265879–35395878	4 (7)	0,0005	egzoninė	<i>DEF6;PPARD</i>	LT; FIN
chr6:97855879–98005878	6 (6)	0,0007	ncRNA_egzoninė	<i>MIR548H3</i>	LT
chr7:30280729–30470728	8 (16)	0,0006	egzoninė	<i>NOD1;ZNR2</i>	LT
chr7:151730729–151870728	5 (7)	0,0002	egzoninė	<i>GALNT11;KMT2C</i>	LT; CEU
chr8:48621077–48801076	9 (20)	0,0007	egzoninė	<i>CEBPD;PRKDC;SPIDR</i>	LT
chr8:93731077–93901076	8 (9)	0,0002	egzoninė	<i>TRIQQ</i>	LT; CEU
chr9:38474202–38614201	5 (6)	0,0001	egzoninė	<i>ANKRD18A</i>	LT; FIN
chr9:125434202–125574201	5 (8)	0,0006	egzoninė	<i>OR1K1;ORIL3;ORIL4;ORIL6;OR5C1</i>	LT
chr10:66065709–66215708	6 (7)	0,0002	intergeninė	<i>REEP3;ANXA2P3</i>	LT; FIN
chr10:118195709–118325708	4 (11)	0,0008	egzoninė	<i>PNLIP;PNLIPRP3</i>	LT
chr11:71554229–71734228	9 (11)	0,0004	egzoninė	<i>DEFB131B;IL18BP;NUMA1;RNF121</i>	LT



<b>Genomo koordinatės</b>	<b>Langas*</b>	<b>P reikšmė</b>	<b>Sritis</b>	<b>Genai</b>	<b>Populiacija</b>
chr12:1296235–1426234	4 (6)	0,0009	egzoninė	<i>ERC1</i>	LT; CEU; FIN
chr12:15736235–15886234	6 (11)	0,0003	egzoninė	<i>EPS8; PTPRO</i>	LT; CEU; FIN
chr13:34108565–34278564	8 (11)	0,0005	UTR5	<i>STARD13</i>	LT
chr14:64046743–64236742	10 (13)	0,0005	egzoninė	<i>SGPP1; WDR89</i>	LT
chr15:48357093–48477092	3 (4)	0,0005	egzoninė	<i>MYEF2; SLC24A5</i>	LT; CEU; FIN
chr15:69617093–69737092	3 (6)	0,0002	egzoninė	<i>KIF23; PAQR5</i>	LT; CEU
chr16:67321264–67501263	7 (20)	0,0005	egzoninė	<i>ATP6V0D1; HSD11B2; KCTD19; LRRC36; PLEKHG4; TPPP3; ZDHHC1</i>	LT; CEU
chr17:29252345–29392344	5 (5)	0,0002	egzoninė	<i>ADAP2; LOC107984974; RNF135</i>	LT; CEU; FIN
chr18:30389383–30559382	8 (11)	0,0001	egzoninė	<i>CCDC178</i>	LT; CEU
chr19:50580913–50700912	3 (4)	0,0005	egzoninė	<i>IZUMO2</i>	LT
chr20:58399095–58539094	3 (8)	0,0006	egzoninė	<i>CDH26; FAM217B; PHACTR3; PPP1R3D; SYCP2</i>	LT; FIN
chr21:44859932–44989931	3 (8)	0,0005	egzoninė	<i>HSF2BP</i>	LT

\* Langų skaičius, esantis tarp 0,1 proc. Tajima's D p reikšmių išskirčių. Skliausteliuose nurodytas bendras langų skaičius toje genetinėje srityje.



**3.9 pav.** Teigiamos gamtinės atrankos nustatytų signalų Tajima's D neutralumo statistika pasiskirstymas autosomose LT (A), CEU (B) ir FIN (C) populiacijose. Žali taškai nurodo reikšmingus Tajima's D įverčius, kurių  $p < 0,01$

Atlikus anotaciją, nustatytas daugiau negu vienas genas daugelyje genetinių sričių. Vidutinis genetinių sričių dydis – ~217 kb, viena jų buvo 1,5 Mb (chr3:50343412–51893411), taip pat pirmos chromosomos 49988960–50728959 sritis buvo 739 kb dydžio, o 4 chromosomoje – ~429 kb dydžio sritis.

Lietuvos populiacijos specifiniai stipriausi signalai nustatyti pirmoje chromosomoje (35818960–5948959), kurioje buvo 22 langai srityje, o p reikšmė – 0,0008. Srityje randami *KIAA0319L* ir *ZMYM4* genai. Kitas stiprus signalas, turintis 20 langų, nustatytas 8 chromosomos 48621077–8801076 srityje su *CEBPD*, *PRKDC*, *SPIDR* genais. Taip pat stiprus signalas (16 langų) identifikuotas ir 7 chromosomoje (30280729–30470728) su *NOD1* ir *ZNRF2* genais.

### 3.3. Lietuvos ilgalaikio efektyvaus populiacijos dydžio, divergencijos laiko ir evoliucinių ryšių su kitomis pasaulio populiacijomis analizė

Nustatyti evoliuciniai ryšiai tarp Lietuvos ir 22-ųjų referentinių HGDP ir CEPH populiacijų (žr. 3.6 lentelę), rekonstravus du vienus svarbiausių žmogaus evoliucijoje parametrus – efektyvų populiacijos dydį ( $N_e$ ) ir divergencijos laiką [10]. Analizė atlikta pagal nepusiausvyros sankibos struktūrą, remiantis autosomų VNP duomenimis, panaudojus R programinės įrangos *NeON* paketą [91].

Analizuotą lietuvių populiacijos grupę sudarė 295 nesusiję asmenys iš bendros Lietuvos populiacijos, kuriems buvo atliktas plataus masto VNP genotipavimas pagal 719 666 VNP žymenis. Po kokybės kontrolės analizei naudoti 294 tiriamųjų 568 040 autosomų VNP duomenys.

Palyginimui naudotas HGDP ir CEPH populiacijų  $N_e$  reikšmės 2015 m. apskaičiavo *NeON* paketo autoriai M. Mezzavilla ir S. Ghirrotto'as (*University of Trieste, Italy*) (žr. 3.6 lentelę).

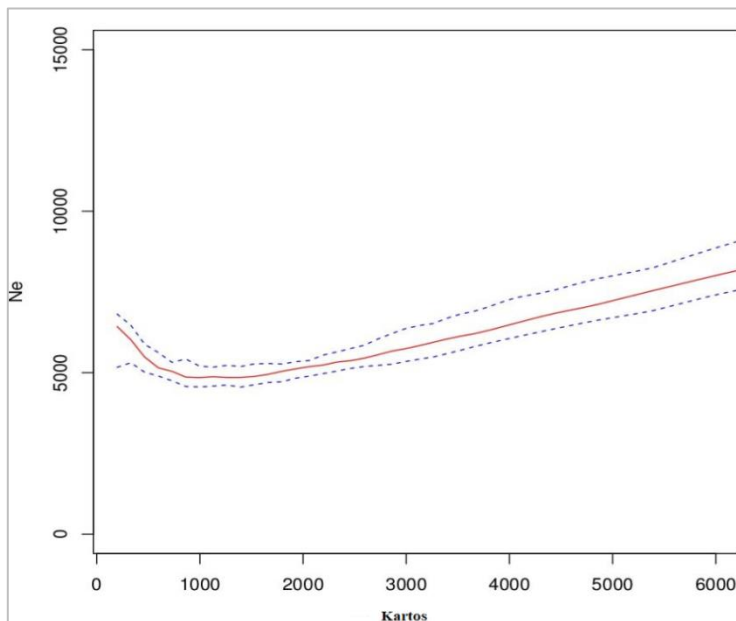
Lietuvos populiacijos  $N_e$  apskaičiuotas nuo 200 iki 6 000 kartų su prielaida, kad karta lygi 25 metams. Nustatytas Lietuvos populiacijos ilgalaikis  $N_e$  yra 5 404 su pasikliautiniu intervalu (PI) [4 910; 5 643].  $N_e$  reikšmių kitimas nuo 200 iki 6 000 kartų pateiktas 3.10 paveiksle.

**3.6 lentelė.** HGDP ir CEPH populiacijų *Ne* įverčiai, 2015 m. *NeON* paketu nustatyti M. Mezzavilla'os ir S. Ghirrotto'o

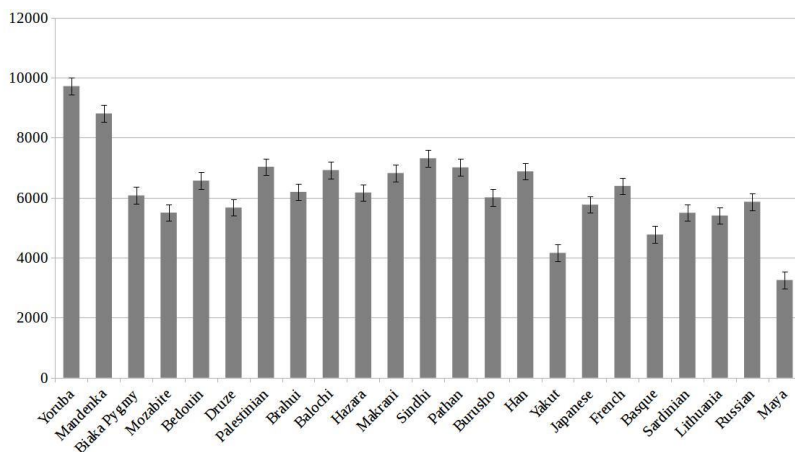
<b>Populiacija</b>	<b>Mėginių skaičius</b>	<b><i>Ne</i></b>	<b>95 proc. PI</b>
Yoruba	21	9 717	[9 066; 10 552]
Mandenka	22	8 806	[8 164; 9 501]
Sindhi	24	7 313	[6 824; 8 099]
Palestinian	45	7 030	[6 453; 7 466]
Pathan	22	7 009	[6 522; 7 501]
Balochi	24	6 920	[6 380; 7 354]
Han	34	6 877	[6 364; 7 192]
Makrani	25	6 821	[6 379; 7 555]
Bedouin	46	6 569	[6 070; 7 058]
French	27	6 392	[5 924; 6 704]
Brahui	24	6 193	[5 815; 6 583]
Hazara	22	6 172	[5 767; 6 841]
BiakaPygmy	21	6 074	[5 671; 6 439]
Burusho	25	6 010	[5 650; 6 614]
Russian	23	5 863	[5 308; 6 228]
Japanese	28	5 767	[5 193; 6 192]
Druze	42	5 670	[5 415; 6 108]
Mozabite	27	5 503	[5 034; 5 917]
Sardinian	28	5 497	[5 320; 6 024]
Basque	22	4 773	[4 396; 5 138]
Yakut	25	4 160	[3 859; 4 445]
Maya	21	3 255	[2 927; 3 600]

Lietuvos *Ne* nuo 150 000 (6 000 karta) iki 25 000 (1 000 karta) metų nuolatos mažėjo. *Ne* didėjimas stebimas apie 25 000 metus, kaip ir daugelio kitų ne Afrikos, ypač Europos ir Azijos, populiacijų [91].

Lyginant analizuotų populiacijų *Ne* reikšmes, stebima įvairovė nuo ~10 000 Afrikos populiacijose iki ~4 000 majų ir jakutų populiacijose (žr. 3.11 pav.). Lietuvos *Ne* yra ganėtinai mažas, kaip ir daugelio kitų Šiaurės Europos populiacijų, pvz., Suomijos *Ne* yra 5 200 galimai dėl butelio kaklelio efekto, asocijuoto su individų sumažėjimu ir rekolonizacija prieš paskutinį ledynmetį ir po jo, įvykusio prieš 13 000–18 000 m. [44]. Populiacijos didėjimas nuo 8 000 iki 5 000 metų, arba nuo 320 iki 200 kartų.

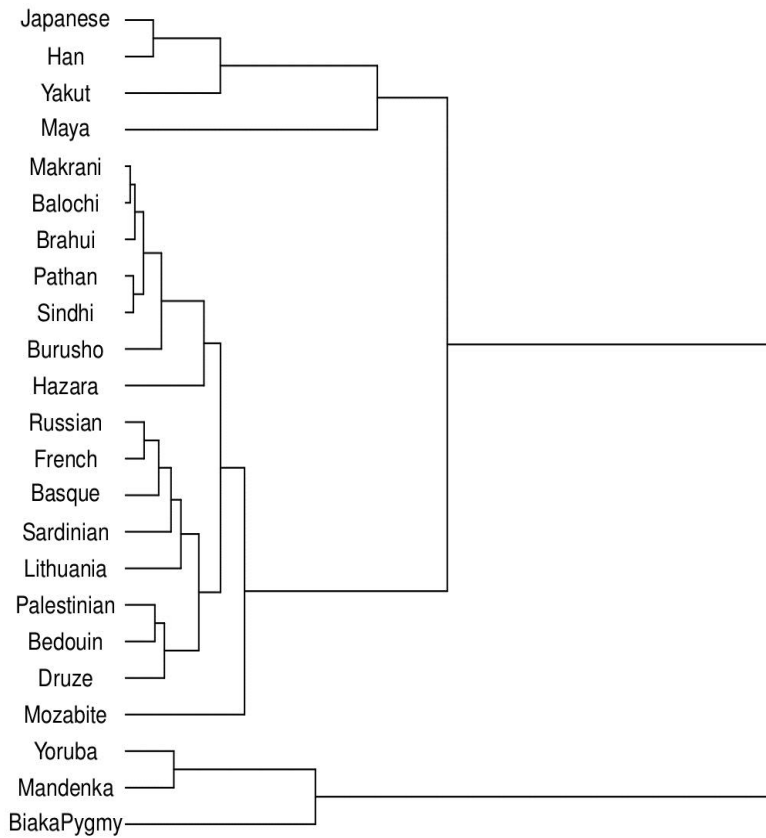


**3.10 pav.** Lietuvos efektyvaus populiacijos dydžio kitimas prieš 5 000 (200 karta)–150 000 (6 000 karta) metų. X ašyje pateikiamas laikas kartomis, y ašyje –  $N_e$  reikšmės su pasikliautiniais intervalais (5 ir 95 procentilės), punktyrinės linijos



**3.11 pav.** Analizuotų populiacijų  $N_e$  su 95 proc. pasikliautinoju intervalu

Panaudojus  $N_e$  reikšmes ir apskaičiuotų genetinių atstumų,  $F_{ST}$ , tarp tiriamų 23-ųjų populiacijų matricą, buvo rekonstruotas divergencijos laikas kartomis (žr. 3.7 lentelę) tarp populiacijų. Jis pavaizduotas filogenetiniame medyje, kai naudojama artimiausio kaimyno jungtis (žr. 3.12 pav.).



**3.12 pav.** Filogenetinis medis, paremtas divergencijos laiku, nustatytu kartomis tarp analizuotų populiacijų

**3.7 lentelė.** Tiriamų 23 populiacijų divergencijos laiko, pateikto metais, nustatyto R programinės įrangos *NeON* paketu, matrica

	Balochi	Brahui	Palestinian	Druze	Bedouin	Mozabite	Yoruba	Mandenka	BiakaPygmy	Lithuania	Populiacija	
										0	<b>Lithuania</b>	
									0	49 318	BiakaPygmy	Afrika
								0	17 043	51 510	Mandenka	
							0	4 182	15 826	55 652	Yoruba	
					0	0	36 304	33 473	36 822	11 173	Mozabite	Vidurio rytai
				0	3 805	8 078	49 979	46 468	47 829	9 122	Bedouin	
				0	2 924	6 768	55 746	51 677	50 879	7 376	Druze	
			0	5 926	7 089	10 763	52 920	49 262	50 634	7 848	Palestinian	
	0	609	5 711	5 756	6 988	10 909	53 919	50 096	50 537	7 721	Brahui	
	7 420	7 852	12 172	11 758	13 053	15 132	55 900	52 128	53 166	7 656	Balochi	
	402	1 006	5 191	5 502	6 358	9 982	57 798	53 801	53 667	10 008	Hazara	Centrinė pėtų Azija
	1 091	1 839	7 250	7 391	8 666	12 075	51 901	48 430	49 908	7 876	Makrani	
	1 126	1 902	6 401	6 322	7 897	11 764	55 853	52 176	53 618	8 260	Sindhi	
	3 324	3 905	8 642	8 344	9 930	12 953	58 356	54 378	55 270	6 911	Pathan	
	31 980	3 1392	38 550	36 719	38 485	37 781	56 003	52 145	51 797	8 140	Burusho	
	21 292	20 784	26 229	24 300	26 154	25 497	81 986	76 867	73 372	30 560	Han	Rytų Azija
	29 649	28 946	35 708	33 726	35 462	34 432	65 984	61 047	56 404	19 875	Yakut	
	6 391	6 600	4 511	4 233	6 017	9 147	77 289	72 125	68 239	28 166	Japanese	
	8 500	8 455	6 051	5 631	7 236	9 621	62 545	58 116	57 487	3 790	French	Europa
	9 639	9 601	5 130	4 850	6 314	8 923	59 900	55 276	53 072	5 447	Basque	
	6 235	6 515	6 628	6 180	8 030	10 708	62 123	57 521	55 906	7 131	Sardinian	
	27 786	26 601	32 061	28 945	31 561	30 395	60 643	56 317	55 255	2 814	Russian	
				28 945	31 561	30 395	74 270	68 433	60 760	23 674	Maya	





pietų ir Rytų Azijos populiacijų. Ilgos filogenetinio medžio šakos (senesnis atsiskyrimas) atskiria afrikiečius nuo ne Afrikos populiacijų. Pavyzdžiui, mozabite populiacija iš Šiaurės Afrikos klasterizuojasi šalia Vidurio rytų Azijos populiacijų (beduinai, palestiniečiai, druzai) ir taip parodomas šių populiacijų genetinis panašumas, nustatytas ir kitais tyrimais [100].

Seniausiai divergavusios yra Afrikos (jorubai, mandenkai, *biaka pygmy*) ir Rytų Azijos (haniai, jakutai, japonai) populiacijos – prieš 70 368 metus, PI [64 036; 76 700], taip pat Afrikos ir majų populiacijos – prieš 67 821 metus, PI [50 989; 84 653]. Vidutinis divergencijos laikas tarp Afrikos ir Europos populiacijų nustatytas prieš 56 536 metus, PI [54 000; 59 071], tarp Europos ir Rytų Azijos – prieš 31 261 metus, PI [27 089; 35 433]. Vėliausias (angl. *The Most recent*) atsiskyrimas nustatytas tarp Europos ir Vidurio rytų Azijos – prieš 7 074 metus, PI [5 973; 8 176], ir tarp Europos ir Centrinės pietų Azijos – prieš 8 970 metus, PI [8 191; 9 750].

Lietuviai pirmiausiai atsiskyrė nuo Afrikos – prieš 52 160 metų, PI [44 169; 60 151], ir daug vėliau nuo Rytų Azijos – prieš 26 201 metus, PI [12 272; 40 129]. Atsiskyrimas nuo Centrinės pietų ir Vidurio rytų Azijos įvyko panašiu laiku – prieš 8 082 metus, PI [7 198; 8 965], ir 8 880 metų, PI [6 179; 11 581]. Paskutinis genetinis atsiskyrimas nustatytas tarp Lietuvos ir Rusijos – prieš 2 814 metų, ir Prancūzijos – prieš 3 790 metų. Gauti rezultatai taip pat parodė, kad Lietuva buvo pirma iš visų tirtų Europos populiacijų (prancūzų, baskų, sardiniečių), atsiskyrusi nuo Vidurio rytų Azijos.

#### 3.4. Praeities įvykių tarp Lietuvos populiacijos etnolingvistinių grupių rekonstravimas, remiantis efektyviu populiacijos dydžiu ir divergencijos analize

Siekiant rekonstruoti praeities įvykius tarp Lietuvos populiacijos etnolingvistinių grupių, remiantis efektyviu populiacijos dydžiu ir divergencijos analize, analizuoti 295 tiriamieji iš šešių Lietuvos etnolingvistinių grupių: trijų aukštaičių grupių (Vakarų ( $n = 52$ ), Pietų ( $n = 51$ ), Rytų ( $n = 48$ )) ir trijų žemaičių grupių (Šiaurės ( $n = 61$ ), Pietų ( $n = 59$ ) ir Vakarų ( $n = 24$ )) genetiniai duomenys. Po plataus masto genotipavimo duomenų kokybės kontrolės analizei naudoti 294 tiriamųjų 568 040 autosomų VNP duomenys.

Atlikus skaičiavimus R programinės įrangos *NeON* paketu, nustatytas kiekvienos etnolingvistinės grupės harmoninis *Ne*, kurio pasiskirstymas pateiktas 4.13 paveiksle. Mažiausias *Ne* nustatytas Vakarų žemaičių grupėje – 4 940 [4 674; 5 304], o didžiausias – Vakarų aukštaičių grupėje – 5 314 [4 829; 5 490] (žr. 4.13 pav. ir 3.8 lentelę). Didesnės *Ne* reikšmės stebimos visose trijose aukštaičių grupėse, palyginti su žemaičių grupėmis.



**3.13 pav.** *Ne* su pasikliautiniais intervalais pasiskirstymas Lietuvos populiacijos šešiose etnolingvistinėse grupėse. RA – Rytų aukštaičiai, PA – Pietų aukštaičiai, VA – Vakarų aukštaičiai, ŠŽ – Šiaurės žemaičiai, PŽ – Pietų žemaičiai, VŽ – Vakarų žemaičiai

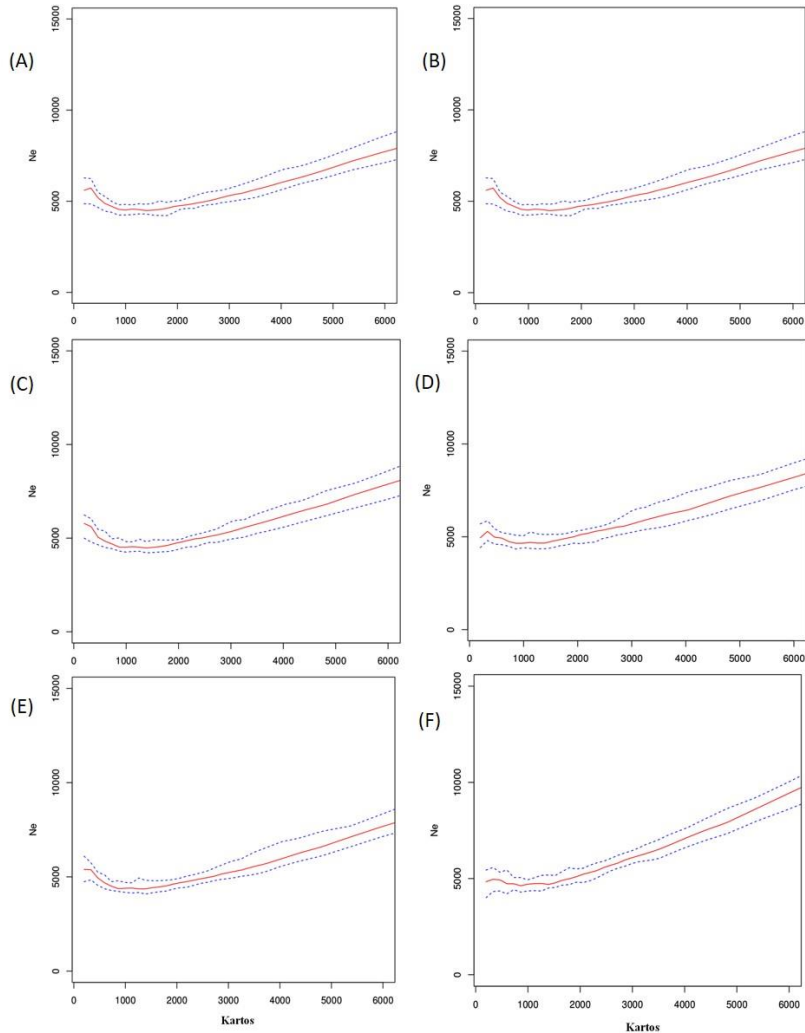
Atlikus *Ne* dydžių statistinę analizę tarp dviejų pagrindinių Lietuvos populiacijos etnolingvistinių grupių (aukštaičių ir žemaičių), nustatytas statistiškai reikšmingas skirtumas ( $p < \alpha$ ,  $\alpha = 0,001$ ). Naudotas Mann'o-Whitney'aus-Wilcoxon'o testas.

**3.8 lentelė.** Tirtų šešių Lietuvos etnolingvistinių grupių *Ne* įverčiai su pasikliautiniais intervalais. PA – Pietų aukštaičiai, PŽ – Pietų žemaičiai, RA – Rytų aukštaičiai, VA – Vakarų aukštaičiai, VŽ – Vakarų žemaičiai, ŠŽ – Šiaurės žemaičiai

Etnolingvistinė grupė	<i>Ne</i>	95 proc. PI
PA	5 308	[4 914; 5 626]
PŽ	5 079	[4 792; 5 365]
RA	5 298	[4 827; 5 537]
VA	5 314	[4 829; 5 490]
VŽ	4 940	[4 674; 5 304]
ŠŽ	5 081	[4 757; 5 393]

Analizuojant *Ne* kitimą laike, stebimi panašūs *Ne* svyravimai visose šešiose etnolingvistinėse grupėse (žr. 3.14 pav.). Aukštaičių ir žemaičių grupėse prieš 150 000–37 500 metų (prieš 6 000–1 500 kartų) stebimas pastovus *Ne*

mažėjimas, galimai dėl įkūrėjo efekto migruojant šiuolaikiniams žmonėms iš Afrikos. Neolito laikotarpiu visose Lietuvos populiacijos etnolingvistinėse grupėse stebimas  $N_e$  didėjimas. Neolito laikotarpiu  $N_e$  skirtumai stebimi tarp Vakarų aukštaičių ir Vakarų žemaičių grupių. Vakarų aukštaičių grupėje stebimas sparčiausias  $N_e$  didėjimas, Vakarų žemaičių grupėje – sparčiausias mažėjimas.



**3.14 pav.** Efektyvaus populiacijos dydžio kitimas laike prieš 5 000 (200 karta)–150 000 (6 000 karta) metų. X ašyje pateikiamas laikas kartomis, y ašyje –  $N_e$  reikšmės su pasikliautiniais intervalais (5 ir 95 procentilės), punktyrinės linijos. (A) Pietų aukštaičiai; (B) Rytų aukštaičiai; (C) Vakarų aukštaičiai; (D) Pietų žemaičiai; (E) Šiaurės žemaičiai; (F) Vakarų žemaičiai

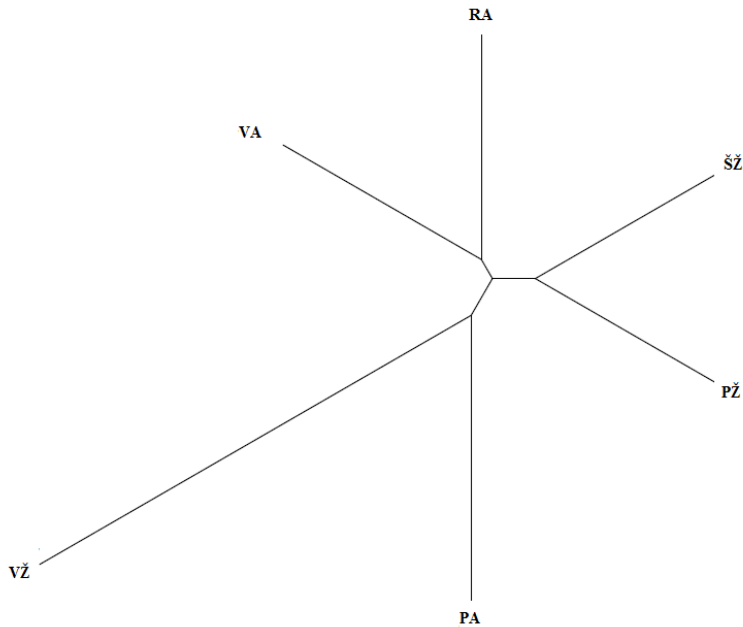
Panaudojus  $Ne$  reikšmes ir apskaičiuotų genetinių atstumų,  $F_{ST}$ , tarp tirtų šešių Lietuvos etnolingvistinių grupių matricą (žr. IV priedo 2 lentelę), buvo rekonstruotas divergencijos laikas kartomis (karta – 25 metai) (žr. 3.9 lentelę) ir atvaizduotas neorientuotame (angl. *unrooted*) UPGMA (angl. *Unweighted Pair Group Method with Arithmetic Mean, UPGMA*) filogenetiniame medyje (žr. 3.15 pav.).

**3.9 lentelė.** Divergencijos laikas metais tarp Lietuvos populiacijos šešių etnolingvistinių grupių. RA – Rytų aukštaičiai, PA – Pietų aukštaičiai, VA – Vakarų aukštaičiai, ŠŽ – Šiaurės žemaičiai, PŽ – Pietų žemaičiai, VŽ – Vakarų žemaičiai

	PA	PŽ	RA	ŠŽ	VA	VŽ
PA	0					
PŽ	6 725	0				
RA	7 325	6 450	0			
ŠŽ	7 125	4 775	5 800	0		
VA	5 850	5 650	5 275	5 225	0	
VŽ	9 975	8 350	9 650	8 250	9 350	0

Apskaičiuotas divergencijos laikas tarp etnolingvistinių grupių parodė, kad seniausiai nuo Pietų aukštaičių divergavo Vakarų žemaičiai – prieš 9 975 metus. Artimiausia divergencija stebima tarp etnolingvistinių grupių iš to paties ar artimo geografinio regiono: tarp Rytų ir Vakarų aukštaičių prieš 5 275 metus, tarp Pietų ir Šiaurės žemaičių prieš 4 775 metus ir tarp Šiaurės žemaičių ir Vakarų aukštaičių prieš 5 225 metus. Vidutinis divergencijos laikas tarp aukštaičių ir žemaičių grupių yra 7 327 metai.

Ilgesnės filogenetinio medžio šakos reiškia tolimesnį atsiskyrimo laiką. Vakarų žemaičių ir Pietų aukštaičių grupės atsiskiria nuo likusių keturių etnolingvistinių grupių – Vakarų ir Rytų aukštaičių bei Šiaurės ir Pietų žemaičių.



**3.15 pav.** Šešių Lietuvos etnolingvistinių grupių (RA – Rytų aukštaičiai, PA – Pietų aukštaičiai, VA – Vakarų aukštaičiai, ŠŽ – Šiaurės žemaičiai, PŽ – Pietų žemaičiai, VŽ – Vakarų žemaičiai) UPGMA filogenetinis medis, sudarytas pagal divergencijos laiką tarp grupių

### 3.5. Lietuvos trumpalaikio efektyvaus populiacijos dydžio nustatymas bei $N_e$ ir $N$ santykio įvertinimas

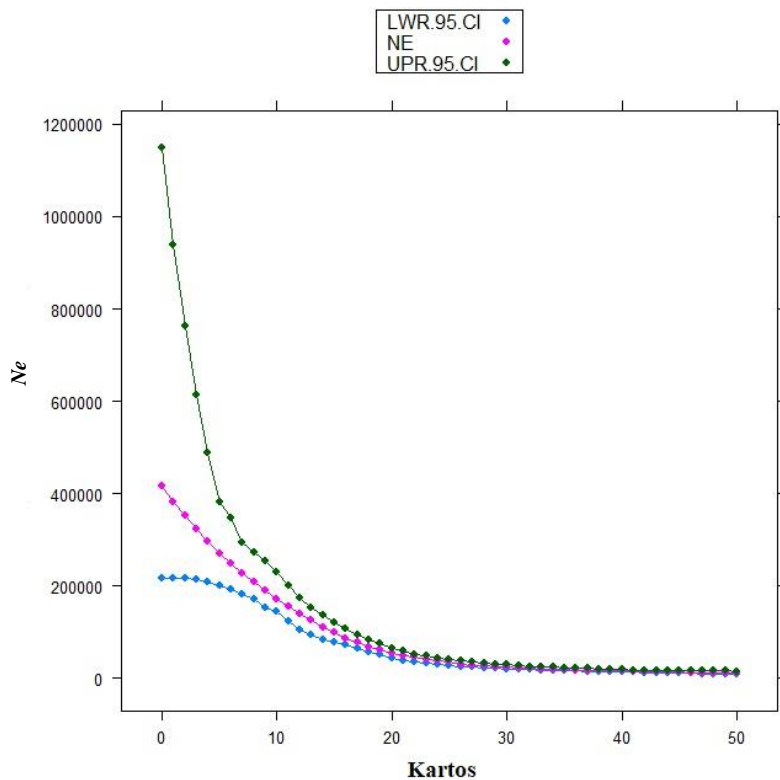
Trumpalaikis efektyvus populiacijos dydis nustatytas Lietuvos populiacijai ir dviem pagrindinėms etnolingvistinėms grupėms (aukštaičiams ir žemaičiams), remiantis identiškais pagal kilmę segmentais ir panaudojus neparametrinį skaičiavimo metodą [95]. Analizuoti 295 tiriamųjų iš šešių Lietuvos etnolingvistinių grupių genetiniai duomenys, t. y. trys aukštaičių grupės (Vakarų ( $n = 52$ ), Pietų ( $n = 51$ ), Rytų ( $n = 48$ )) ir trys žemaičių grupės (Šiaurės ( $n = 61$ ), Pietų ( $n = 59$ ) ir Vakarų ( $n = 24$ )). Po plataus masto VNP genotipavimo duomenų kokybės kontrolės analizei naudoti 294 tiriamųjų 568 040 autosomų VNP duomenys.

Lietuvos populiacijos trumpalaikis  $N_e$  apskaičiuotas nuo 50 iki nulinės kartos, su prielaida, kad kartą lygi 25 metams. Trumpalaikis  $N_e$  su pasikliautinaisiais intervalais 50-iai kartų pateiktas 3.10 lentelėje, o jo pasiskirstymas laike – 3.16 paveiksle.

**3.10 lentelė.** Lietuvos populiacijos 50 kartų trumpalaikis *Ne*

<b>Karta (k)</b>	<b><i>Ne</i></b>	<b>95 proc. PI</b>
0	417 000	[218 000; 1 150 000]
1	383 000	[217 000; 938 000]
2	352 000	[217 000; 763 000]
3	323 000	[214 000; 614 000]
4	296 000	[209 000; 488 000]
5	271 000	[200 000; 382 000]
6	248 000	[192 000; 346 000]
7	227 000	[181 000; 294 000]
8	208 000	[171 000; 273 000]
9	190 000	[153 000; 253 000]
10	172 000	[144 000; 230 000]
11	155 000	[124 000; 202 000]
12	140 000	[106 000; 173 000]
13	125 000	[93 400; 154 000.]
14	111 000	[84 800; 137 000]
15	98 400	[77 700; 122 000]
16	87 000	[71 500; 107 000]
17	77 100	[65 700; 94 600]
18	68 800	[58 100; 84 500]
19	61 500	[50 300; 74 900]
20	55 000	[44 100; 65 900]
21	49 300	[38 800; 58 600]
22	44 900	[35 300; 52 200]
23	40 900	[32 800; 47 800]
24	37 500	[30 200; 43 600]
25	34 300	[28 000; 40 300]
26	31 300	[26 000; 37 500]
27	28 700	[24 200; 34 900]
28	26 900	[23 000; 32 600]
29	25 300	[21 800; 30 500]
30	23 700	[20 600; 29 100]
31	22 300	[19 600; 27 500]
32	21 000	[18 800; 25 800]
33	20 000	[17 900; 24 400]
34	19 100	[17 200; 23 900]
35	18 500	[16 400; 23 000]
36	17 800	[15 700; 22 000]
37	17 100	[15 100; 21 200]
38	16 700	[14 600; 20 400]
39	16 100	[14 100; 19 400]
40	15 700	[13 500; 18 700]
41	15 400	[12 900; 18 000]
42	14 900	[12 300; 17 900]
43	14 300	[11 800; 18 000]

<b>Karta (k)</b>	<b><i>N<sub>e</sub></i></b>	<b>95 proc. PI</b>
44	13 600	[11 400; 17 400]
45	13 200	[11 000; 17 100]
46	12 700	[10 300; 16 600]
47	12 600	[9 960; 16 300]
48	12 200	[9 690; 16 000]
49	12 000	[9 430; 15 700]
50	11 900	[9 280; 15 300]



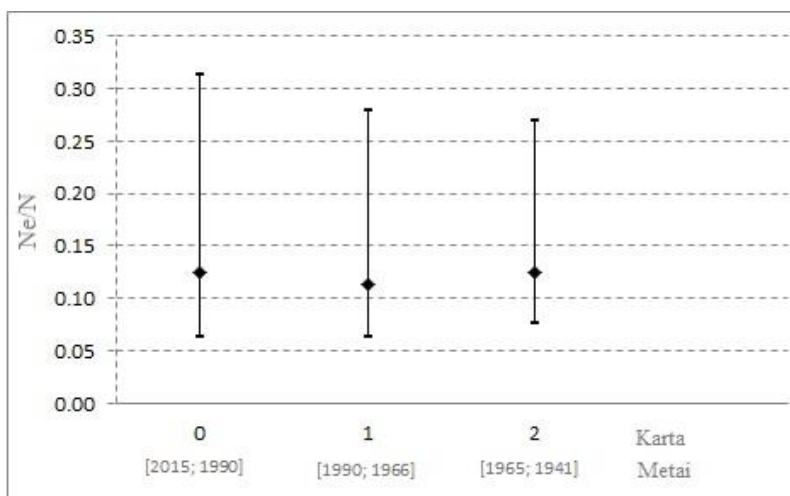
**3.16 pav.** Lietuvos populiacijos trumpalaikis  $N_e$ , nustatytas 50 kartų su 95 proc. pasikliautinaisiais intervalais

Prieš 50 kartų, arba prieš 1 250 metų, Lietuvos  $N_e$  buvo 11 900. Vidutinis  $N_e$  nuo 50 iki 30 kartos buvo 16 228. Ekspontinis didėjimas stebimas nuo 25 kartos.

Darbe vertintas efektyvaus populiacijos dydžio ir bendro individų skaičiaus Lietuvos populiacijoje konkrečių kartų santykis ( $N_e/N$ ). Patikimi bendro individų skaičiaus Lietuvos populiacijoje duomenys buvo gauti tik

pirmų trijų kartų iš *Eurostat* (IV priedo 5 lentelė), Lietuvos statistikos metraščio ([http://osp.stat.gov.lt/en/statistikos\\_leidiniu-katalogas?proc.20\\_publication=1673](http://osp.stat.gov.lt/en/statistikos_leidiniu-katalogas?proc.20_publication=1673)) bei A. Stulginskio ir E. Galvanausko knygos (1923) apie pirmąjį Lietuvos populiacijos surašymą.

Nulinės kartos (1991–2015 m.)  $N_e$  – 417 000 (95 proc. PI [218 000; 1 150 000]) ir vidutinis  $N$  – 3 373 154. Pirmos kartos (1966–1990 m.)  $N_e$  – 383 000 (95 proc. PI [217 000; 938 000]) ir vidutinis  $N$  – 3 351 015. Antros kartos (1941–1965 m.)  $N_e$  – 352 000 (95 proc. PI [217 000; 763 000]),  $N$  – 2 816 800. Apskaičiuotas antros kartos  $N_e$  ir  $N$  santykis lygus 0,125 (95 proc. PI [0,077; 0,271]), pirmos kartos  $N_e$  ir  $N$  santykis yra 0,114 (95 proc. PI [0,065; 0,280]) ir nulinės kartos  $N_e$  ir  $N$  santykis yra 0,124 (95 proc. PI [0,065; 0,341]) (žr. 3.17 pav.). Nustatytas Lietuvos efektyvus populiacijos dydis sudaro 1/10 dalį bendro individų skaičiaus Lietuvos populiacijoje.

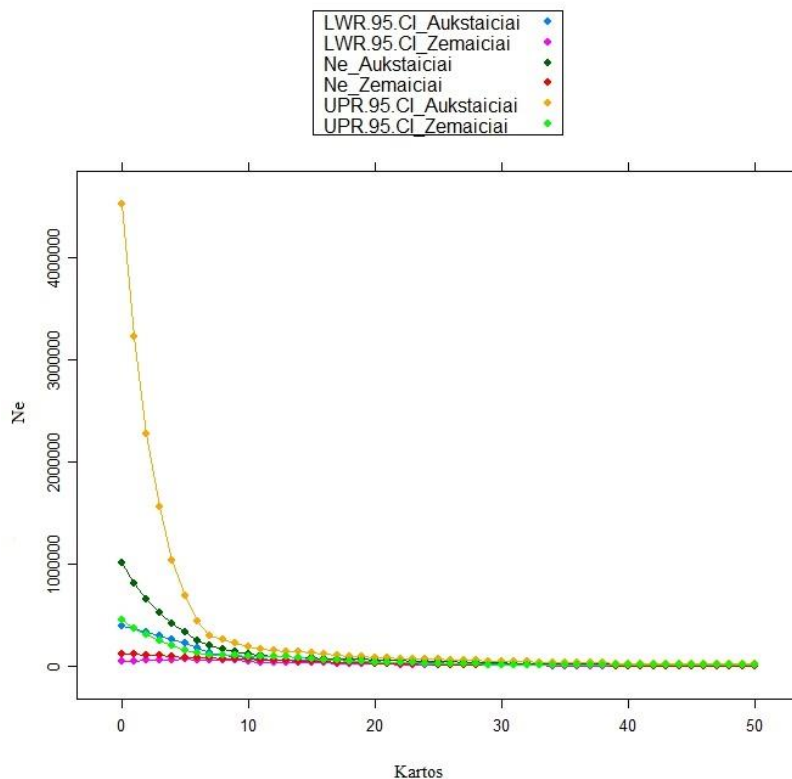


**3.17 pav.** Lietuvos populiacijos trijų kartų  $N_e$  ir  $N$  santykis su 95 proc. pasikliautiniu intervalu

Taip pat apskaičiuotas trumpalaikis  $N_e$  dviem pagrindinėms etnolingvistinėms grupėms – aukštaičiams ir žemaičiams (žr. IV priedo 6 ir 7 lenteles). Lyginant dvi pagrindines grupes, aukštaičių grupėje stebimos didesnės trumpalaikio  $N_e$  reikšmės negu žemaičių grupėje (žr. 3.18 pav.). Prieš 50 kartų, arba prieš 1 250 metų, aukštaičių grupės  $N_e$  buvo 16 900, o žemaičių grupės – 7 950. Aukštaičių grupės apskaičiuotas vidutinis  $N_e$  nuo 0 iki 50 kartos yra 127 088, o žemaičių grupės – 39 364, t. y. tris kartus mažesnis, palyginti su aukštaičiais. Spartesnis  $N_e$  didėjimas stebimas nuo 21 kartos aukštaičių grupėje ir nuo ~18 kartos žemaičių grupėje. Atlikus  $N_e$  dydžių statistinę analizę, nustatytas statistiškai reikšmingas skirtumas tarp



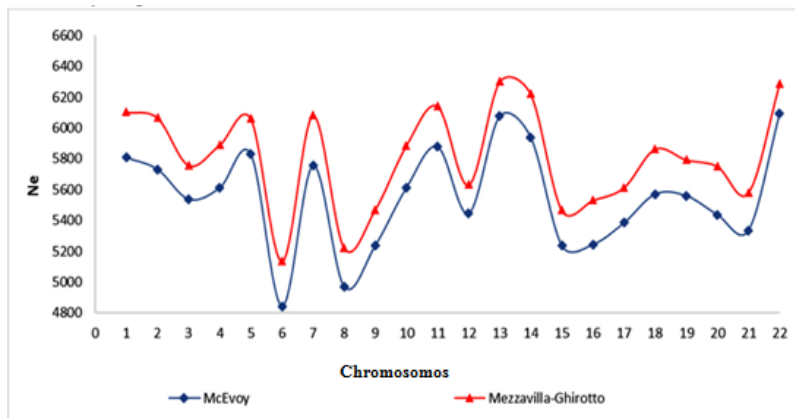
grupių ( $p < \alpha$ ,  $\alpha = 0,0002$ ), naudotas Mann'o-Whitney'aus-Wilcoxon'o testas.



**3.18 pav.** Lietuvos populiacijos dviejų pagrindinių etnolingvistinių grupių, aukštaičių ir žemaičių, trumpalaikis  $N_e$ , nustatytas 50 kartų su 95 proc. pasikliautiniu intervalu

### 3.6. Matematinė lyginamoji analizė efektyviam populiacijos dydžiui nustatyti

Darbe nagrinėjami du matematiniai metodai, t. y. McEvoy'aus ir Mezzavilla'os-Ghirotto'o, kurie skirti nustatyti efektyvųjį populiacijos dydį (žr. III priedo III dalį). Atlikta abiem metodais gautų  $N_e$  rezultatų lyginamoji analizė.  $N_e$  medianos buvo apskaičiuojamos kiekvienai autosomai imant nustatytas kiekvienos autosomos 250 taškų (Mezzavilla'os-Ghirotto'os metodas) arba 50 taškų (McEvoy'aus metodas) (žr. 3.19 pav.). Statistiškai reikšmingam skirtumui tarp  $N_e$  grupių nustatyti naudotas neparametrinis Mann'o-Whitney'aus U testas.



**3.19 pav.** Kiekvienos autosomos efektyvus populiacijos dydis

P reikšmė  $<0,05$  laikytina statistiškai reikšminga. Analizės rezultatai parodė, kad efektyvus populiacijos dydis, nustatytas McEvoy'aus metodu, yra 5 481, 95 proc. PI [5 384; 5 563]; Mezzavilla'os-Ghirotto'o metodu  $N_e$  nustatyta 5 722, 95 proc. PI [5 679; 5 755]. Šie skirtumai buvo statistiškai reikšmingi ( $p < 0,001$ ). Tačiau statistiškai reikšmingi skirtumai nebuvo nustatyti su visomis chromosomomis (žr. 3.11 lentelę). Efektyvaus populiacijos dydžio rangų vidurkis (angl. *mean rank*), apskaičiuotas McEvoy'aus metodu, yra 2 783,27, o Mezzavilla'os-Ghirotto'o metodu – 3 352,69. Didesnės  $N_e$  reikšmės siejamos su didesniu vidurkiu, pvz., naudojant Mezzavilla'os-Ghirotto'o metodą [91].

**3.11 lentelė.** Dviejų efektyvaus populiacijos dydžio nustatymo metodų (McEvoy'aus ir Mezzavilla'os-Ghirotto'o) apskaičiuotos  $N_e$  reikšmės autosomoms su p reikšmėmis

Chromosoma	McEvoy'aus metodas	Mezzavilla'os-Ghirotto'o metodas	p reikšmė
1	5 807,5	6 101,0	0,1075
2	5 731,5	6 063,0	0,1569
3	5 538,0	5 756,0	0,0414*
4	5 611,5	5 888,0	0,0745
5	5 829,0	6 057,0	0,0372*
6	4 843,5	5 132,5	2,1162-08**
7	5 756,0	6 079,0	0,0397*
8	4 970,5	5 219,5	0,0005**
9	5 237,0	5 467,0	0,0107*
10	5 611,0	5 882,5	0,0883
11	5 879,0	6 139,5	0,1484

Chromosoma	McEvoy'aus metodas	Mezzavilla'os-Ghirotto'o metodas	p reikšmė
12	5 444,0	5 628,0	0,0253*
13	6 076,0	6 297,0	0,1228
14	5 939,0	6 216,5	0,0263*
15	5 236,5	5 462,5	0,1656
16	5 245,0	5 529,0	0,0191*
17	5 385,0	5 611,5	0,0275*
18	5 571,0	5 861,5	0,0016*
19	5 560,5	5 791,5	0,0120*
20	5 438,0	5 749,0	0,0532
21	5 335,5	5 577,5	0,0679
22	6 092,0	6 281,5	0,0233*

Statistiškai reikšmingos p reikšmės \*p < 0,005, \*\*p < 0,001.

*Ne* nustatymas galimas abiem analizuotais metodais, tačiau reikėtų turėti omenyje skirtingas rekombinacijų atstumų kategorijas, kurias sukuria metodai.

Metodu, kuriuo sudaroma daugiau rekombinacijos genetinių atstumų, apskaičiuojamos didesnės *Ne* reikšmės, o metodas, sukuriantis mažiau genetinių atstumų kategorijų, apskaičiuoja mažesnes *Ne* reikšmes.

Siekiant turėti didesnės skiriamosios gebos rekombinacijos genetinių atstumų kategorijas, nustatant *Ne*, turėtų būti naudojamas Mezzavilla'os-Ghirotto'o metodas, kitu atveju – McEvoy'aus metodas.

#### 4. REZULTATŲ APIBENDRINIMAS

Šio darbo tikslas buvo atlikti Lietuvos populiacijos genetinės struktūros ir evoliucinių veiksnių – efektyvaus populiacijos dydžio ir gamtinės atrankos, nustatymą ir analizę pagal plataus masto vieno nukleotido polimorfizmo genotipavimo duomenis.

Ankstyvasis Baltijos šalių apgyvendinimo periodas yra mažai ištyrinėtas. Pirmieji žmogaus pėdsakai Pietų Pabaltijyje randami vėlyvojo paleolito periodu prieš 35–40 tūkstančių metų, tačiau jų archeologinius pėdsakus ištrynė ledynai. Pasibaigus paskutiniam ledynmečiui, prieš 13–12 tūkstančių metų, dabartinėje Lietuvos teritorijoje ėmė kurtis gyventojai. Pirmieji indoeuropiečiai Lietuvos vakarinėje teritorijoje pasirodė maždaug prieš 5–4 tūkstančius metų. Jie susimaišė su autochtonais ir suformavo baltus. Tačiau neaišku, kaip šie procesai atsispindi šiuolaikinių lietuvių genome.

Nors lietuvių tautą suformavo baltiškos gentys, tačiau pagal antropologinius duomenis šie skirtumai išnyko tik viduramžiais, ir nuo to laiko Lietuvos populiacija tiek Rytų Europos, tiek visos Europos genetinės įvairovės požiūriu yra labai homogeniška [101]. Nors kai kurių tyrimų rezultatai verčia abejoti šiuo teiginiu. Ištyrus 1 110 moksleivių iš septynių Lietuvos rajonų pirštų ir delnų dermatoglifinius požymius, nustatytas didžiausias fenotipinis skirtumas tarp Pietų aukštaičių ir žemaičių [102]. Taip pat skirtumai tarp etnolingvistinių grupių buvo nustatyti atlikus AB0 ir Rh(D) kraujo grupių tyrimą [3]. Pagal kraujo grupių pasiskirstymą Žemaitija yra homogeniškesnė negu Aukštaitija. Bendrai aukštaičių grupei nustatytas heterogeniškumas pagal Rh(D) faktorių [3]. Tai paaiškinama tuo, kad žemaičiai yra autochtonai, o aukštaičiai – vėliau atsikraustę rytiniai baltai [3]. Tiriant mtDNR genetinius žymenis, nustatyti nedideli skirtumai tarp aukštaičių ir žemaičių [3]. Detalesnis Lietuvos populiacijos genetinės struktūros įvertinimas ir padėtis kitų populiacijų atžvilgiu skirtingais žymenimis atliktas 2004 m. D. Kasperavičiūtės [1]. Ištyrus etnolingvistines grupes pagal mtDNR ir Y chromosomą, statistiškai reikšmingi skirtumai tarp grupių nenustatyti [1, 103]. I. Uktverytė (2014 m.), tirdama lietuvių populiaciją, remdamasi Y chromosoma, taip pat nustatė Lietuvos populiacijos homogeniškumą, tačiau, remdamasi tolygiai po genomą pasiskirsčiusių VNP alelių įvairove, nustatė, kad Žemaitija yra homogeniškesnė negu Aukštaitija [27].

Šiame darbe Lietuvos populiacijos vidinė genetinė struktūra bei padėtis kitų populiacijų atžvilgiu analizuota remiantis plataus masto VNP

genotipavimo duomenimis ir *SmartPCA* algoritmu. Atlikta šešių Lietuvos etnolingvistinių grupių pagrindinių komponentų analizė pagal plataus masto VNP žymenis patvirtino tirtos lietuvių populiacijos homogeniškumą. Lietuvos populiacijos vidinis homogeniškumas taip pat nustatytas modeliu, paremtu ADMIXTURE metodu, kuriame dominuoja vienas bendras genetinis komponentas tarp visų šešių etnolingvistinių grupių. Nustatytos mažos  $F_{ST}$  reikšmės taip pat parodė grupių genetinį ir geografinį artumą.

Lietuvos padėtis kitų pasaulio populiacijų atžvilgiu atitiko esamą geografinę padėtį. Atlikus Lietuvos populiacijos genetinės struktūros įvertinimą Europos populiacijų atžvilgiu ir atvaizdavus PK, pastebimas populiacijų išsidėstymas pagal pietų vakarų–šiaurės rytų ašį. Pirma PK atskyrė lietuvius nuo kitų analizuotų Europos populiacijų. Lietuvių populiacija yra artimiausia CEU populiacijai, nustatytas mažiausias  $F_{ST}$  (0,006), palyginti su kitomis tirtomis Europos populiacijomis. Panašūs rezultatai gauti M. Nelis ir kt. tyrime (2009) bei I. Uktverytės disertaciniame darbe (2014) [24, 27].

Remiantis šiais rezultatais, gamtinės atrankos tyrimui pasirinktos referentinės Europos populiacijos CEU ir FIN.

Išanalizavus globalų lietuvių populiacijos kilmės profilį ADMIXTURE programa, esant tinkamiausiam aštuonių kilmių komponentui, Lietuvos populiacijai nustatytas dominuojantis protėvinis genetinis komponentas (geltona spalva), kurio nedidelis santykis randamas ir gretimose analizuotose Europos populiacijose (CEU, FIN, GBR) (žr. 3.5 pav.). Remiantis ADMIXTURE ir PKA rezultatais, galima daryti išvadą, kad Lietuvos populiacija yra homogeniška ir pasižymi specifinėmis genominėmis ypatybėmis tirtų Europos populiacijų atžvilgiu. Tačiau reikalingi tolesni tyrimai, patvirtinantys šią išvadą. Dėl Baltijos šalių kaimynių genetinių duomenų neprieinamumo nebuvo galimybės atlikti Lietuvos populiacijos genetinės struktūros analizę jų atžvilgiu.

Atlikus teigiamos gamtinės atrankos analizę, panaudojant plataus masto VNP genotipavimo duomenis ir tris statistinius metodus, identifikuotos specifinės Lietuvos populiacijai kandidatines teigiamos gamtinės atrankos veikiamos genomo sritys. Stipriausias specifinis signalas Lietuvos populiacijai nustatytas 6 chromosomoje su dviem nesinoniminėmis pakaitomis uoslės receptorių gene, *OR2B6*. Kitų mokslininkų nustatyta uoslės receptorių didelė įvairovė tarp populiacijų galimai yra dėl vietinės adaptacijos [104]. Gamtinės atrankos poveikis nustatytas tik tam tikrų uoslės receptorių genų atveju [105–108]. Kodėl gamtinė atranka veikia kai kuriuos uoslės receptorius, iki galo nėra išaiškinta, tačiau siejama su mityba ir sveikata [109–112]. Darbe nustatyti kiti genai kandidatai, susiję su žmonių adaptacija: dieta, odos pigmentacija, ir

socialiniais įgūdžiais. Kaip ir kitų mokslininkų tyrimuose, nustatytas genas *SLC24A5*, kuris susijęs su europiečių šviesios odos pigmentacija [113]. *SLC24A5* genas koduoja baltymą nešiklį, kuris dalyvauja melanino sintezėje ir yra vienas svarbiausių genų, turinčių įtakos žmonių pigmentacijai [114]. Šis genas nustatytas visais trim metodais, o tai indikuoja senesnės gamtinės atrankos poveikį šiam genui. XP-EHH ir  $F_{ST}$  metodais nustatytas stiprus signalas 9 chromosomoje tarpgeninėje srityje (12352971–12537279), kurioje vienoje pusėje yra *PTPRD-AS2* genas, kitoje pusėje – *TYRPI* genas. Polimorfizmai *TYRPI* gene asocijuoti su plaukų ir akies rainelės spalva Europos populiacijose [98, 115]. Lietuvių populiacijoje nustatyta tarpgeninė sritis, esanti per ~156 kb nuo *TYRPI* geno, kuri gali dalyvauti šio geno reguliacijoje. Reguliaciniai elementai gali būti nutolę per tūkstančius bazių porų nuo geno, kurį reguliuoja [116]. *SIC24A5* ir *TYRPI* genai yra veikiami nesenos gamtinės atrankos, prasidėjusios prieš 11–19 tūkstančių metų, po šiuolaikinių europiečių protėvių atsiskyrimo nuo Rytų Azijos [117]. Kitas su pigmentacija susijęs genas (*HERC2*) nustatytas Tajima's D metodu. Jis asocijuotas su mėlynų akių spalva [118]. Tačiau šio geno sritis buvo per kelias pozicijas už 1 proc. Tajima's D p reikšmių išskirčių ribų. Šios srities (chr15:28437093-28627092) Tajima's D įvertis nustatytas –0,78 kai p reikšmė yra 0,001 bei 10 langų srityje. XP-EHH ir  $F_{ST}$  metodais LT ir CEU populiacijose nustatytas *SLC24A3* genas, kuris koduoja Na<sup>+</sup>/Ca<sup>2+</sup> mainų 3 tipo baltymą, atsakingą už kofaktorių transportą. Polimorfizmai šiame gene asocijuoti su vazokonstrikcija ir hipertenzija [116]. Šio geno variantų atranka galimai prasidėjo medžiotojų rinkėjų populiacijose, kuriose vyravo gyvūninės kilmės dieta. Šiai dietai būdingas žemas natrio chlorido, tačiau didelis kalio druskų suvartojimas, palyginti su pasikeitusia dieta po neolito [117]. Senesnės gamtinės atrankos poveikis nustatytas 6 chromosomos, chr6:35265879–5395878, geninėje srityje, kai p reikšmė yra 0,0005, su genu *PPARD* – peroksisomų proliferatoriaus–aktyvatoriaus receptorius delta. *PPARD* geno raiška vyksta melanocituose, smegenyse ir storojoje žarnoje, dalyvauja lipidų, cholesterolio metabolizme [119]. Šis genas asocijuotas su žmogaus odos pigmentacija ir atsakingas už riebalų įsisavinimą, būtent Europos populiacijose [119–121]. Taip pat Tajima's D metodu identifiкуotas *GPM6A* (glikoproteinas M6A) genas (chr4:176190684–176400683) LT ir FIN populiacijose, asocijuotas su nervų sistema, galbūt kalbos, socialiniais įgūdžiais [122]. Dauguma nustatytų su gamtinės atrankos poveikiu genetinių sričių yra tarpgeninėse srityse, kurių funkciją sudėtinga paaiškinti, tačiau kurios galimai dalyvauja genų raiškos reguliavime.

XP-EHH ir  $F_{ST}$  statistiniais metodais nustatyti dar trys genai (*PLA2G7*, *MINK1*, *ENO3*) su nesinoniminėmis pakaitomis, kurie asocijuoti

su ligomis. *PLA2G7* gene esanti pakaita rs1805018 (exon7:c.T593C) yra siejama su astma. *MINK1* geno pakaitos rs2302319 (exon21:c.G2476A) A alelio dažnis LT populiacijoje yra 0,05, nustatytas CADD 20,3 indikuoja esant šią pakaitą tarp 1 proc. labiausiai žalingų. *ENO3* geno pakaitos rs238238 (exon4:c.A212G) G alelio dažnis LT populiacijoje yra 0,28, o CADD 13,24 indikuoja esant šią pakaitą tarp 10 proc. labiausiai žalingų.

Šio tyrimo pagrindinis trūkumas – negalėjimas nustatyti retų variantų, remiantis plataus masto VNP genotipavimo duomenimis. Šį trūkumą išspręstų naujos kartos sekoskaitos duomenų panaudojimas. Taip pat reikalingas nustatytų variantų funkcinis patvirtinimas.

Atlikta Lietuvos efektyvaus populiacijos dydžio, divergencijos laiko ir evoliucinių ryšių su kitomis pasaulio populiacijomis analizė pagal plataus masto VNP genotipavimo duomenis suteikė išsamesnių žinių apie Lietuvos populiacijos demografiją praeityje ir dabartyje.

Nustačius ilgalaikį Lietuvos populiacijos *Ne* nuo 6 000 iki 200 kartų, stebimas butelio kaklelio efektas iki 1 000 kartos galimai dėl šiuolaikinių žmonių kilimo iš Afrikos ir vėlesnio paplitimo po visą pasaulį. *Ne* didėjimas stebimas panašiu laiku ir daugelio kitų ne Afrikos populiacijų, ypač Europos ir Azijos [88]. Tačiau to ir tikimasi palaikant žmonių paplitimo iš Afrikos teoriją [120]. Dabartiniai europiečiai yra emigrantai iš Afrikos. Pirmą didžiąją migracijos bangą iš Afrikos įvyko prieš ~40–60 tūkstančių metų. Šios migracijos palikuonys apgyvendino Europos, Azijos ir kitus pasaulio regionus [58].

Tačiau nustatytas Lietuvos *Ne* yra ganėtinai mažas, kaip ir daugelio kitų Šiaurės Europos populiacijų, pvz., Suomijos *Ne* yra 5 200, o vidutinis *Ne* CEU, NET, SWE, DEN ir UK populiacijų yra 8 200 dėl galimai butelio kaklelio efekto, asocijuoto su individų skaičiaus sumažėjimu ir rekolonizacija prieš paskutinį ledynmetį ir po jo, maždaug prieš 13 000–18 000 metų [41, 52]. Populiacijos didėjimas prieš 8 000–5 000 metų (arba prieš 320–200 kartų) gali būti siejamas su žemdirbystės atsiradimu Neolito laikotarpiu. Nustačius divergencijos laiką tarp LT ir 22-iejų referentinių HGDP ir CEPH populiacijų, gauti panašūs rezultatai, palyginti su kitų mokslininkų rezultatais [55, 61, 62, 91]. Seniausiai atsiskyrė Afrika nuo Rytų Azijos. Tai įvyko prieš daugiau negu 70 000 metų. Divergencijos laikas tarp Afrikos ir Europos populiacijų nustatytas prieš 56 536 metus, tarp Europos ir Rytų Azijos – prieš 31 261 metus. Vėliausias atsiskyrimas nustatytas tarp Europos ir Vidurio rytų Azijos (prieš 7 074 metus) ir tarp Europos ir Centrinės pietų Azijos (prieš 8 970 metus).

Remiantis gautais rezultatais, galima daryti išvadą, kad pirminė žmonių migracija vyko iš Afrikos į Rytų Aziją prieš daugiau negu 70 000

metų, vėliau žmonės plito į Europą prieš ~56 000 metų ir galiausiai į Vidurio Rytus ir Centrinę Pietų ir Šiaurės Aziją prieš 52 000 metų. Taigi tarp Europos ir Afrikos egzistuoja artimesni ryšiai negu tarp Afrikos ir Rytų Azijos.

Divergencijos laiko analizė parodė, kad lietuvių populiacija buvo pirma, palyginti su kitomis Europos populiacijomis (Prancūzija, Baskai, Sardinija, Rusija), atsiskyrusi nuo Vidurio rytų Azijos prieš 8 800 metų.

Pagal R. Gimbutienę, baltų tautų susiformavimą lėmė ir indoeuropietiška kultūrinė sistema, išsirutuliojusi iš Eurazijos stepių (kurganų) kultūros [9]. Indoeuropiečių kalbą ir visuomeninę patriarchalinę, patrilinearinę ir trijų klasių struktūrą į Rytų Pabaltijį, Baltarusiją ir vidurio Rusiją apie 2 500 m. pr. m. e. atnešė virvelinės keramikos kultūros, suindoeuropietinti Vidurio Europos žmonės [9]. Tačiau anglų archeologo C. Renfrew iškelta Anatolijos teorija teigia, kad indoeuropiečių prokalbė gyvavo anksčiau, negu įprasta manyti, t. y. VII–VI tūkstantmetyje pr. m. e. Anatolijoje, o indoeuropiečių išplitimas Europoje siejamas su žemdirbių atsikėlimu iš Anatolijos į Pietryčių Europą [121].

Šio tyrimo rezultatai pateikia naujų duomenų diskusijai apie indoeuropiečių prokalbės kilmę ir paplitimą, palaikant hipotezę, kad indoeuropiečių kalbos į Europą atkeliavo iš Anatolijos prieš 8 000–9 500 metų [121]. Lietuvių kalba yra viena seniausių išlikusių indoeuropiečių kalbų [123]. Remdamiesi Anatolijos hipoteze, galime daryti prielaidą, kad lietuviai gali būti artimiausi protoindoeuropiečiams.

Ištyrus demografinius praeities įvykius tarp Lietuvos populiacijos etnolingvistinių grupių, nustatyta, kad aukštaičiai ir žemaičiai atsiskyrė prieš 9 000 metų. Tai gali būti siejama su tuo, kad prieš 11 000 metų vidurio Lietuva buvo netinkama sėsliam gyvenimui dėl klimato pokyčių [5]. Atlikti Aukštumalos aukštapelkės archeologiniai tyrimai 2004 m. byloja apie pirmųjų gyventojų Lietuvos teritorijoje pasirodymą ir gyvenseną Pabaltijyje vėlyvojo paleolito ir ankstyvojo mezolito laikotarpiu, prieš 14–10 tūkstančių metų, galimai dėl palankesnių gyvenimo sąlygų [6, 7]. Pagal J. Žilinską ir R. Rimantienę, pirmieji žmogaus pėdsakai Lietuvos teritorijoje randami Pietų Pabaltijyje vėlyvojo paleolito periodu [5, 6]. Tai buvo medžiotojai, atėję iš Vidurio ir Vakarų Europos, jie atsinešė dvi kultūras – Svidrų ir Pabaltijo Madeleino.

Lietuvos populiacijos etnolingvistinių grupių ilgalaikis *Ne* parodė statistiškai reikšmingą aukštaičių ir žemaičių skirtumą.

Remiantis šio tyrimo rezultatais, galima daryti išvadą, kad indoeuropiečiai galimai atkeliavo į Lietuvos teritoriją neolito periodu ir prisidėjo prie skirtingų baltų genčių formavimosi, taip pat kad aukštaičiai ir žemaičiai galimai istoriškai formavosi kaip dvi atskiros baltų gentys.



Nustačius Lietuvos trumpalaikį efektyvų populiacijos dydį 50-iai kartų nuo dabarties, buvo įvertintas efektyvaus populiacijos dydžio ir bendro individų skaičiaus populiacijoje santykis. Žinant  $Ne$  ir  $N$  santykį, galima tiesiogiai nustatyti  $Ne$ , remiantis bendru individų skaičiumi populiacijoje, bei įvertinti ekologinių veiksnių įtaką efektyviam populiacijos dydžiui [57]. Tyrimo rezultatai parodė eksponentinį populiacijos augimą nuo 25 kartos. Nustatytas  $Ne$  ir  $N$  santykis Lietuvos populiacijos pirmoms trimis kartoms (karta – 25 metai) buvo panašus – apie 0,1. Lietuvos populiacijos  $Ne$  ir  $N$  santykis yra mažas, palyginti su kitų tyrimų rezultatais, kurie svyruoja nuo 0,21 iki 0,65 [55]. Pagal L. Nunney'ų (1994; 1996),  $Ne$  ir  $N$  santykis paprastai yra artimas 0,5 ir retai būna mažesnis už 0,25 ar didesnis už 0,75 [46, 124]. Labai mažas  $Ne$  ir  $N$  santykis (0,1) parodo, kad populiacijoje egzistuoja veiksnių, pvz., moterų vaisingumo ir lyčių skirtumai, persidengiančios kartos ar individų skaičiaus svyravimai, kurie turi įtakos šiam santykiui ir mažina  $Ne$  [125]. Be to, natūralus populiacijos dydžio svyravimas, neveikiant kitiems veiksniams, yra dažnai pakankama mažo  $Ne$  ir  $N$  santykio priežastis [125].

Lietuvos populiacija yra nedidelė, istorijoje patyrusi butelio kaklelio efekto padarinius (stebimas staigus populiacijos dydžio sumažėjimas antroje kartoje) ir ekspansiją (staigus populiacijos augimas nulinėje kartoje), o tai galimai daro įtaką nustatytam mažam  $Ne$  ir  $N$  santykiui. Be to, ne itin reprezentatyvi šio tyrimo imtis ir Lietuvos populiacijos struktūra galėjo turėti įtakos netikslumams, atsiradusiems analizuojant artimiausias kartas nuo dabarties. Tačiau atsižvelgiant į gautus tyrimo rezultatus, tikrasis trumpalaikis efektyvus populiacijos dydis neperžengia 95 proc. pasikliautiną intervalo ribų. R. S. Browning'as ir B. L. Browning'as (2015) teigia, kad 200 individų analizės rezultatai yra pakankamai geri, tačiau tikslesni skaičiavimai gaunami analizuojant apie 1 000 individų [95].

Didesnė tiriamųjų imtis galbūt pateiktų tikslesnius trumpalaikio ir ilgalaikio  $Ne$  bei divergencijos laiko skaičiavimus, kadangi nereprezentatyvi imtis arba nevienoda tiriamų subpopuliacijų imtis gali turėti įtakos paklaidai, atsirandančiai analizuojant alelių dažnių pasiskirstymą. Šio tyrimo rezultatai ne tik papildo fundamentaliojo mokslo žinias apie Lietuvos populiacijos priešistorę, demografinius procesus, tačiau ir parodo, jog yra reikalingi tolesni tyrimai, kad būtų galima paremti šio tyrimo išvadas.

## IŠVADOS

1. Išanalizavus tirtos lietuvių populiacijos genetinę struktūrą pagal plataus masto vieno nukleotido polimorfizmo genotipavimo duomenis, nustatytas Lietuvos populiacijos homogeniškumas ir genetinis unikalumas kitų tirtų populiacijų atžvilgiu.
2. Remiantis plataus masto vieno nukleotido polimorfizmo genotipavimo duomenimis ir trimis statistiniais metodais, XP-EHH,  $F_{ST}$  ir Tajima's D, nustatytos specifinės lietuvių populiacijai kandidatinės teigiamos gamtinės atrankos veikiamos genomo sritys – identifikuoti genai kandidatai, susiję su žmonių adaptacija: dieta (*OR2B6*, *SLC24A3*, *PPARD*), pigmentacija (*SLC24A5*, *TYRP1*, *PPARD*), imunitetu (*IL26*, *IL22*, *HLA*, *BRD2*) ir ligomis (*PLA2G7*, *ENO3*, *COL8A2*).
3. Remiantis Lietuvos ilgalaikio efektyvaus populiacijos dydžio, divergencijos laiko ir evoliucinių ryšių su kitomis pasaulio populiacijomis analizės pagal plataus masto vieno nukleotido polimorfizmo genotipavimo duomenis rezultatais, nustatyta, kad pirminė žmonių migracija vyko iš Afrikos į Rytų Aziją, vėliau žmonės plito į Europą ir galiausiai – į Vidurio Rytus ir Centrinę pietų bei Šiaurės Aziją.
4. Lietuva buvo pirma populiacija, palyginti su kitomis Europos populiacijomis (Prancūzija, Baskai, Sardinija, Rusija), atsiskyrusi nuo Vidurio rytų Azijos prieš 8 800 metų.
5. Remiantis rekonstruotais praeities įvykiais tarp Lietuvos populiacijos etnolingvistinių grupių, nustatyta, kad:
  - 5.1. Ilgalaikis  $N_e$  statistiškai reikšmingai skiriasi tarp aukštaičių ir žemaičių grupių.
  - 5.2. Indoeuropiečiai, atkeliavę į Lietuvos teritoriją neolito periodu, prisidėjo prie skirtingų baltų genčių formavimosi, o tai galimai nulėmė ir lietuvių genetinę įvairovę bei skirtumus.
6. Nustačius Lietuvos trumpalaikį efektyvųjį populiacijos dydį 50-iai kartų nuo dabarties ir įvertinus efektyvaus populiacijos dydžio ir bendro individų skaičiaus populiacijoje santykį, nustatyta, kad:
  - 6.1. Lietuvos populiacijos  $N_e$  ir  $N$  santykis yra mažas (0,1).
  - 6.2. Lietuvos populiacija yra nedidelė, istorijoje patyrusi butelio kaklelio efekto padarinius ir ekspansiją, o tai galimai lemia nustatytą mažą  $N_e$  ir  $N$  santykį.

## LITERATŪROS SĄRAŠAS

1. Kasperaviciute D, Kucinskas V, Stoneking M. Y chromosome and mitochondrial DNA variation in Lithuanians. *Ann Hum Genet* 2004, 68(Pt 5): 438–452.
2. Kučinskas V. Population genetics of Lithuanians. *Ann Hum Biol* 2001, 28(1): 1–14.
3. Kucinskas V. Human mitochondrial DNA variation in Lithuania. *Anthropol Anz* 1994, 52(4): 289–295.
4. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR *et al.* A global reference for human genetic variation. *Nature* 2015, 526(7571): 68–74.
5. Rimantienė R. *Akmens amžius Lietuvoje*. Vilnius, Lithuania: Žiburys, 1966.
6. Žilinskas J: *Lietuvių protėviai (The ancestors of Lithuanians)*. Kaunas: Raidės, 1937.
7. Grigaliūnas M. The first palaeolithic and mesolithic settlements in Aukštumala, Šilutė district, the Nemunas river delta. *Archaeologia baltica 20* 2013: 175–189.
8. Mittnik A, Wang CC, Pfrenkle S, Daubaras M, Zarina G, Hallgren F, Allmae R, Khartanovich V, Moiseyev V, Torv M *et al.* The genetic prehistory of the Baltic Sea region. *Nat Commun* 2018, 9(1): 442.
9. Gimbutienė M. *Baltai priešistoriniais laikais: etnogenėzė, materialinė kultūra ir mitologija*. Vilnius: Mokslas, 1985.
10. Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A *et al.* A human genome diversity cell line panel. *Science* 2002, 296(5566): 261–262.
11. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL *et al.* A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 2001, 409(6822): 928–933.
12. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al.* Initial sequencing and analysis of the human genome. *Nature* 2001, 409(6822): 860–921.
13. Jobling M, Hollox E, Kivisild T, Tyler-Smith C. *Human Evolutionary Genetics*, 2 edn: Garland Science 2014.
14. Pugach I, Stoneking M. Genome-wide insights into the genetic history of human populations. *Investig Genet* 2015, 6: 6.
15. Kucinskas V. Population genetics of Lithuanians. *Ann Hum Biol* 2001, 28(1) :1–14.
16. Kučinskas V. *Genomo įvairovė: Lietuviai Europoje*. Vilnius: Spalvų šalis, 2004.

17. Karakachoff M, Duforet-Frebourg N, Simonet F, Le Scouarnec S, Pellen N, Lecointe S, Charpentier E, Gros F, Cauchi S, Froguel P *et al.* Fine-scale human genetic structure in Western France. *Eur J Hum Genet* 2015, 23(6): 831–836.
18. Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population structure on large genetic association studies. *Nature genetics* 2004, 36(5): 512–517.
19. Novembre J, Di Rienzo A. Spatial patterns of variation due to natural selection in humans. *Nature reviews Genetics* 2009, 10(11): 745–755.
20. Shriver MD, Kennedy GC, Parra EJ, Lawson HA, Sonpar V, Huang J, Akey JM, Jones KW. The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Hum Genomics* 2004, 1(4): 274–286.
21. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome research* 2009, 19(9): 1655–1664.
22. Heath SC, Gut IG, Brennan P, McKay JD, Bencko V, Fabianova E, Foretova L, Georges M, Janout V, Kabesch M *et al.* Investigation of the fine structure of European populations with applications to disease association studies. *Eur J Hum Genet* 2008, 16(12): 1413–1429.
23. Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR *et al.* Genes mirror geography within Europe. *Nature* 2008, 456(7218): 98–101.
24. Nelis M, Esko T, Mägi R, Zimprich F, Zimprich A, Toncheva D, Karachanak S, Piskácková T, Balascák I, Peltonen L *et al.* Genetic structure of Europeans: a view from the North-East. *PLoS one* 2009, 4(5): e5472.
25. Elhaik E, Tatarinova T, Chebotarev D, Piras IS, Maria Calò C, De Montis A, Atzori M, Marini M, Tofanelli S, Francalacci P *et al.* Geographic population structure analysis of worldwide human populations infers their biogeographical origins. *Nat Commun* 2014, 5: 3513.
26. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000, 155(2): 945–959.
27. Uktverytė I. *Lietuvos etnolingvistinių grupių genetinės struktūros analizė remiantis informatyviais genomo žymenimis*. Vilnius: Vilniaus universitetas, 2014.
28. Wright S. Classification of the factors of evolution. *Cold Spring Harb Symp Quant Biol* 1955, 20: 16–24D.
29. Wright S. *Evolution in Mendelian populations*, vol. 16. Print, 1931.
30. Crow JF, Kimura M. *An Introduction to Population Genetics Theory*. Blackburn Press, 2009.
31. Kimura M, Crow J. *The Measurement of Effective Population Number*, vol. 17; 1963.

32. Frankham R. Inbreeding in the wild really does matter. *Heredity (Edinb)* 2010, 104(2): 124.
33. Lynch M, Conery J, Burger R. Mutation Accumulation and the Extinction of Small Populations. *The American Naturalist* 1995, 146(4): 489–518.
34. Crow JF. General theory of population genetics: synthesis. *Cold Spring Harb Symp Quant Biol* 1955, 20: 54–59.
35. Eyre-Walker A, Keightley PD, Smith NG, Gaffney D. Quantifying the slightly deleterious mutation model of molecular evolution. *Mol Biol Evol* 2002, 19(12): 2142–2149.
36. Vicoso B, Charlesworth B. Evolution on the X chromosome: unusual patterns and processes. *Nature reviews Genetics* 2006, 7(8): 645–653.
37. Presgraves DC. Recombination enhances protein adaptation in *Drosophila melanogaster*. *Curr Biol* 2005, 15(18): 1651–1656.
38. Wang J. Estimation of effective population sizes from data on genetic markers. *Philos Trans R Soc Lond B Biol Sci* 2005, 360(1459): 1395–1409.
39. Anderson EC. An efficient Monte Carlo method for estimating  $N_e$  from temporally spaced samples using a coalescent-based likelihood. *Genetics* 2005, 170(2): 955–967.
40. Nei M, Tajima F. Genetic drift and estimation of effective population size. *Genetics* 1981, 98(3): 625–640.
41. Nomura T. Estimation of effective number of breeders from molecular coancestry of single cohort sample. *Evol Appl* 2008, 1(3): 462–474.
42. Waples RS, DO C. Idne: a program for estimating effective population size from data on linkage disequilibrium. *Mol Ecol Resour* 2008, 8(4): 753–756.
43. Wang J. A new method for estimating effective population sizes from a single sample of multilocus genotypes. *Mol Ecol* 2009, 18(10): 2148–2164.
44. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R *et al*. Linkage disequilibrium in the human genome. *Nature* 2001, 411(6834): 199–204.
45. Caballero A. Developments in the prediction of effective population size. *Heredity (Edinb)* 1994, 73 (Pt 6): 657–679.
46. Nunney L, Elam DR. Estimating the Effective Population Size of Conserved Populations. *Conservation Biology* 1994, 8(1): 175–184.
47. Wright S. *The theory of gene frequencies*. Chicago: University of Chicago Press, 1969.
48. Laporte V, Charlesworth B. Effective population size and population subdivision in demographically structured populations. *Genetics* 2002, 162(1): 501–519.

49. Gazal S, Génin E, Leutenegger AL. Relationship inference from the genetic data on parents or offspring: A comparative study. *Theor Popul Biol* 2016, 107: 31–38.
50. Nordborg M. Structured coalescent processes on different time scales. *Genetics* 1997, 146(4): 1501–1514.
51. Hedrick PW. Genetic variation in a heterogeneous environment. V. Spatial heterogeneity in finite populations. *Genetics* 1978, 89(2): 389–401.
52. Hill WG, Robertson A. Linkage disequilibrium in finite populations. *Theor Appl Genet* 1968, 38(6): 226–231.
53. Tenesa A, Navarro P, Hayes BJ, Duffy DL, Clarke GM, Goddard ME, Visscher PM. Recent human effective population size estimated from linkage disequilibrium. *Genome research* 2007, 17(4): 520–526.
54. Takahata N. Allelic genealogy and human evolution. *Mol Biol Evol* 1993, 10(1): 2–22.
55. McEvoy BP, Powell JE, Goddard ME, Visscher PM. Human population dispersal “Out of Africa” estimated from linkage disequilibrium and allele frequencies of SNPs. *Genome research* 2011, 21(6): 821–829.
56. Burgess DJ. Human evolution: multigenome coalescence for modern demographic insights. *Nature reviews Genetics* 2014, 15(8): 514.
57. Kalinowski ST, Waples RS. Relationship of Effective to Census Size in Fluctuating Populations. *Conservation Biology* 2002, 16(1): 129–136.
58. Engen S, Lande R, Saether BE, Gienapp P. Estimating the ratio of effective to actual size of an age-structured population from individual demographic data. *J Evol Biol* 2010, 23(6): 1148–1158.
59. Frankham R. Conservation genetics. *Annu Rev Genet* 1995, 29: 305–327.
60. Palstra FP, Ruzzante DE. Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Mol Ecol* 2008, 17(15): 3428–3447.
61. Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A. Bayesian inference of ancient human demography from individual genome sequences. *Nature genetics* 2011, 43(10): 1031–1034.
62. Tassi F, Ghirotto S, Mezzavilla M, Vilaça ST, De Santi L, Barbujani G. Early modern human dispersal from Africa: genomic evidence for multiple waves of migration. In: *Investig Genet*. vol. 6. London, 2015.
63. Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW. Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. *Proceedings of the National Academy of Sciences* 2000, 97(13): 7360–7365.
64. Ingman M, Kaessmann H, Pääbo S, Gyllensten U. Mitochondrial genome variation and the origin of modern humans. *Nature* 2000, 408(6813): 708–713.

65. Stringer C. Modern human origins: progress and prospects. *Philos Trans R Soc Lond B Biol Sci* 2002, 357(1420): 563–579.
66. Messer PW, Petrov DA. Population genomics of rapid adaptation by soft selective sweeps. *Trends Ecol Evol* 2013, 28(11): 659–669.
67. Berg JJ, Coop G. A population genetic signal of polygenic adaptation. *Plos Genet* 2014, 10(8): e1004412.
68. Ptak SE, Przeworski M. Evidence for population growth in humans is confounded by fine-scale population structure. *Trends Genet* 2002, 18(11): 559–563.
69. Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G, Drousiotou A, Dangerfield B, Lefranc G, Loiselet J *et al.* Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science* 2001, 293(5529): 455–462.
70. Hamblin MT, Di Rienzo A. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *American journal of human genetics* 2000, 66(5): 1669–1679.
71. Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* 2002, 419(6909): 832–837.
72. Balaesque PL, Ballereau SJ, Jobling MA. Challenges in human genetic diversity: demographic history and adaptation. *Human molecular genetics* 2007, 16 Spec No. 2: R134–139.
73. Slatkin M, Bertorelle G. The use of intraallelic variability for testing neutrality and estimating population growth rate. *Genetics* 2001, 158(2): 865–874.
74. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* 2007, 449(7164): 913–918.
75. Tajima F. DNA polymorphism in a subdivided population: the expected number of segregating sites in the two-subpopulation model. *Genetics* 1989, 123(1): 229–240.
76. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. *Evolution* 1984, 38(6): 1358–1370.
77. Wright S. Systems of Mating. I. the Biometric Relations between Parent and Offspring. *Genetics* 1921, 6(2): 111–123.
78. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. Interrogating a high-density SNP map for signatures of natural selection. *Genome research* 2002, 12(12): 1805–1814.
79. Beaumont MA, Balding DJ. Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol* 2004, 13(4): 969–980.

80. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. In: *American journal of human genetics* 2007, 81: 559–575.
81. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *Plos Genet* 2006, 2(12): e190.
82. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics* 2010, 26(22): 2867–2873.
83. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 2010, 10(3): 564–567.
84. Delaneau O, Zagury JF. Haplotype inference. *Methods Mol Biol* 2012, 888: 177–196.
85. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST *et al.* The variant call format and VCFtools. *Bioinformatics* 2011, 27(15): 2156–2158.
86. Szpiech ZA, Hernandez RD. Selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol Biol Evol* 2014, 31(10): 2824–2827.
87. Pybus M, Dall'Olio GM, Luisi P, Uzkudun M, Carreno-Torres A, Pavlidis P, Laayouni H, Bertranpetit J, Engelken J. 1000 Genomes Selection Browser 1.0: a genome browser dedicated to signatures of natural selection in modern humans. *Nucleic acids research* 2014, 42 (Database issue): D903–909.
88. Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol* 2014, 31(7): 1929–1936.
89. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* 2010, 38(16): e164.
90. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics* 2014, 46(3): 310–315.
91. Mezzavilla M, Ghirotto S. Neon: An R Package to Estimate Human Effective Population Size and Divergence Time from Patterns of Linkage Disequilibrium between SNPS. *J Comput Sci Syst Biol* 2015, 8: 037–044.
92. Hayes BJ, Visscher PM, McPartlan HC, Goddard ME. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome research* 2003, 13(4): 635–643.
93. Benazzo A, Panziera A, Bertorelle G. 4P: fast computing of population genetics statistics from large DNA polymorphism panels. *Ecol Evol* 2015, 5(1): 172–175.



94. Team RC. R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. 2014.
95. Browning SR, Browning BL. Accurate Non-parametric Estimation of Recent Effective Population Size from Segments of Identity by Descent. *American journal of human genetics* 2015, 97(3): 404–418.
96. Namjou B, Keddache M, Marsolo K, Wagner M, Lingren T, Cobb B, Perry C, Kennebeck S, Holm IA, Li R *et al.* EMR-linked GWAS study: investigation of variation landscape of loci for body mass index in children. *Front Genet* 2013, 4: 268.
97. Iliff BW, Riazuddin SA, Gottsch JD. The genetics of Fuchs' corneal dystrophy. *Expert Rev Ophthalmol* 2012, 7(4): 363–375.
98. Hider JL, Gittelman RM, Shah T, Edwards M, Rosenbloom A, Akey JM, Parra EJ. Exploring signatures of positive selection in pigmentation candidate genes in populations of East Asian ancestry. *BMC Evol Biol* 2013, 13: 150.
99. Blekhman R, Man O, Herrmann L, Boyko AR, Indap A, Kosiol C, Bustamante CD, Teshima KM, Przeworski M. Natural selection on genes that underlie human disease susceptibility. *Curr Biol* 2008, 18(12): 883–889.
100. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. Genetic structure of human populations. *Science* 2002, 298(5602): 2381–2385.
101. Česnys G. Anthropological roots of Lithuanians. *Science, Arts and Lithuanians* 1991, 8: 9–14.
102. Klevcova NI, Kučinskis V. Dermatoglifičeskije Ocobenosti Litovcev. *Voprosy Antropologii*, 81: 74–88.
103. Kasperaviciute D, Kucinskis V. Variability of the human mitochondrial DNA control region sequences in the Lithuanian population. *J Appl Genet* 2002, 43(2): 255–260.
104. Menashe I, Man O, Lancet D, Gilad Y. Population differences in haplotype structure within a human olfactory receptor gene cluster. *Human molecular genetics* 2002, 11(12): 1381–1390.
105. Gilad Y, Lancet D. Population differences in the human functional olfactory repertoire. *Mol Biol Evol* 2003, 20(3): 307–314.
106. Gilad Y, Man O, Glusman G. A comparison of the human and chimpanzee olfactory receptor gene repertoires. *Genome research* 2005, 15(2): 224–230.
107. Gimelbrant AA, Skaletsky H, Chess A. Selective pressures on the olfactory receptor repertoire since the human-chimpanzee divergence. *Proc Natl Acad Sci U S A* 2004, 101(24): 9019–9022.
108. Moreno-Estrada A, Casals F, Ramírez-Soriano A, Oliva B, Calafell F, Bertranpetit J, Bosch E. Signatures of selection in the human olfactory receptor OR511 gene. *Mol Biol Evol* 2008, 25(1): 144–154.

109. Matsui A, Go Y, Niimura Y. Degeneration of olfactory receptor gene repertoires in primates: no direct link to full trichromatic vision. *Mol Biol Evol* 2010, 27(5): 1192–1200.
110. Jaeger SR, McRae JF, Bava CM, Beresford MK, Hunter D, Jia Y, Chheang SL, Jin D, Peng M, Gamble JC *et al.* A Mendelian trait for olfactory sensitivity affects odor experience and food selection. *Curr Biol* 2013, 23(16): 1601–1605.
111. Pluznick JL, Zou DJ, Zhang X, Yan Q, Rodriguez-Gil DJ, Eisner C, Wells E, Greer CA, Wang T, Firestein S *et al.* Functional expression of the olfactory signaling system in the kidney. *Proc Natl Acad Sci U S A* 2009, 106(6): 2059–2064.
112. Busse D, Kudella P, Grüning NM, Gisselmann G, Ständer S, Luger T, Jacobsen F, Steinsträßer L, Paus R, Gkogkolou P *et al.* A synthetic sandalwood odorant induces wound-healing processes in human keratinocytes via the olfactory receptor OR2AT4. *J Invest Dermatol* 2014, 134(11): 2823–2832.
113. Basu Mallick C, Iliescu FM, Möls M, Hill S, Tamang R, Chaubey G, Goto R, Ho SY, Gallego Romero I, Crivellaro F *et al.* The light skin allele of SLC24A5 in South Asians and Europeans shares identity by descent. *Plos Genet* 2013, 9(11): e1003912.
114. Branicki W, Liu F, van Duijn K, Draus-Barini J, Pośpiech E, Walsh S, Kupiec T, Wojas-Pelc A, Kayser M. Model-based prediction of human hair color using DNA variants. *Hum Genet* 2011, 129(4): 443–454.
115. Sulem P, Gudbjartsson DF, Stacey SN, Helgason A, Rafnar T, Jakobsdottir M, Steinberg S, Gudjonsson SA, Palsson A, Thorleifsson G *et al.* Two newly identified genetic determinants of pigmentation in Europeans. *Nature genetics* 2008, 40(7): 835–837.
116. Kleinjan DA, van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. *American journal of human genetics* 2005, 76(1): 8–32.
117. Beleza S, Santos AM, McEvoy B, Alves I, Martinho C, Cameron E, Shriver MD, Parra EJ, Rocha J. The timing of pigmentation lightening in Europeans. *Mol Biol Evol* 2013, 30(1): 24–35.
118. Eiberg H, Troelsen J, Nielsen M, Mikkelsen A, Mengel-From J, Kjaer KW, Hansen L. Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. *Hum Genet* 2008, 123(2): 177–187.
119. Gattepaille L, Günther T, Jakobsson M. Inferring Past Effective Population Size from Distributions of Coalescent Times. *Genetics* 2016, 204(3): 1191–1206.
120. Templeton A. Out of Africa again and again. *Nature* 2002, 416(6876): 45–51.
121. Renfrew C. *Archaeology and Language*. The Puzzle of Indo-European Origins: Pimlico; 1987.

122. Nunney L. The influence of variation in female fecundity on effective population size. *Biological Journal of the Linnean Society* 1996, 59(4): 411–425.
123. Klimas A, Skrupskelis IK. English and Lithuanian: Two Candidates for the International Languages. *Lituanus* 1969, 15(3).
124. Excoffier L. Evolution of human mitochondrial DNA: evidence for departure from a pure neutral model of populations at equilibrium. *J Mol Evol* 1990, 30: 125–139.
125. Excoffier L, Langaney A. Origin and differentiation of human mitochondrial DNA. *Am J Human Genet* 1989, 44: 73–85.

## PADĖKA

Esu dėkinga savo mokslinio darbo vadovui **akad. prof. habil. dr. Vaidučiui Kučinskui** už darbo idėją ir suteiktą galimybę ją įgyvendinti. Gerbiamam Profesoriumi noriu padėkoti už pasitikėjimą, palaikymą, vertingus patarimus ir rūpestį vykdant mokslo tiriamąjį darbą Vilniaus universiteto Medicinos fakulteto Biomedicinos mokslų instituto Žmogaus ir medicininės genetikos katedroje.

Nuoširdžiausiai dėkoju **dr. Almai Molytei** už praktinę pagalbą, konsultavimą, vertingus patarimus ir visapusišką ne tik profesionalų, bet ir draugišką, nuoširdų palaikymą visos doktorantūros metu įgyvendinant užsibrėžtus tikslus. J. V. Gėtė yra pasakęs: „Geriausias būdas išlaikyti draugystę – dalytis mintimis apie savo darbus. Žmones labiausiai suartina veikla.“

Taip pat norėčiau padėkoti LITGEN projekto, kurio tiriamieji ir genetiniai duomenys buvo panaudoti šioje disertacijoje, vadovui **akad. prof. habil. dr. Vaidučiui Kučinskui** ir projekto vykdytojams, kurie prisidėjo prie mėginių surinkimo, DNR išskyrimo ir genotipavimo darbų.

Labai ačiū **dr. Laimai Ambrozaitytei** už pagalbą vykdant plataus masto VNP genotipavimo darbus disertacijos metu.

Nuoširdžiai dėkoju katedros **vyr. spec. Daliai Kuzminskienei** už pačius vertingiausius patarimus sprendžiant doktorantūros metu iškilusius įvairius klausimus. Ačiū, Dalia, už Jūsų begalinę kantrybę ir pagalbą.

Dėkoju Žmogaus ir medicininės genetikos katedros kolektyvui už suteiktą pagalbą ir patarimus.

Esu dėkinga šioms studijoms, kurių metu ne tik įgijau didžiulės patirties vykdama mokslo tiriamąjį darbą tiek Lietuvoje, tiek užsienyje, bet ir suradau naujų bendraminčių, draugų.

Ačiū mokslinių rezultatų viešinimo bendrakeivei **Birutei Burnytei** už sėkmingą ir malonų bendradarbiavimą ir suvokimą, kad gyvenimas yra kelionė. Kuo mažiau bagožo pasiimi, tuo daugiau įspūdžių gali parsivežti.

Galiausiai noriu padėkoti savo patiems artimiausiems ir mylimiausiems žmonėms – sūnui **Eduardui**, sesei **Laurai**, mamai **Stefai**, broliui **Mindaugui** ir jo šeimai, tetoms **Danai** ir **Genutei**, už moralinį palaikymą ir buvimą šalia, kai man to reikėjo. Ypač dėkoju sesei **Laurai**, be kurios pagalbos, nueitas mokslo kelias būtų buvęs sudėtingesnis.

## PRIEDAI

**I PRIEDAS.** Genominės DNR išskyrimo ir DNR koncentracijos bei švarumo nustatymo protokolai

Genominės DNR išskyrimas fenolio ir chloroformo ekstrakcijos metodu  
protokolas

Reagentai ir laboratorinė įranga

---

Lizuojantis buferis (0,32 M sacharozės; 5 mM MgCl<sub>2</sub>; 1 proc. Triton X – 100; 10 mM Tris-HCl, pH = 7,5)

---

Ląstelių branduolių pernešimo buferis (10 mM Tris-HCl, pH = 10,5; 1 mM EDTA; 0,15 mM NaCl)

---

10 proc. natrio dodecilsulfatas

---

2 proc. proteinazė K

---

Fenolis [C<sub>6</sub>H<sub>6</sub>O]

---

Chloroformas [CHCl<sub>3</sub>]

---

Chloroformo ir izoamilo alkoholio mišinys

---

3 M natrio acetato tirpalas

---

70 proc. etanolis [C<sub>2</sub>H<sub>6</sub>O]

---

96 proc. etanolis [C<sub>2</sub>H<sub>6</sub>O]

---

TE buferis (10 mM Tris-HCl; 1 mM EDTA, pH 8)

---

Didejonizuotas vanduo (ddH<sub>2</sub>O)

---

Vakuuminiai mėgintuvėliai su antikoaguliantu EDTA

---

Sterilūs mėgintuvėliai (50 ml, 2 ml, 1,5 ml, 0,5 ml)

---

Latekso pirštinės

---

Centrifuga

---

Mikrocentrifuga

---

Sterilūs antgaliai (1 000 µl, 100 µl, 20 µl)

---

Sterilios Pastero pipetės (3 ml)

---

Mikropipetės (100–1 000 µL, 10–100 µL, 2–20 µL)

---

Liofilizatorius (angl. *Eppendorf Concentrator 5301*, gamintojas *Eppendorf AG*, Vokietija)

---

Termostatas (angl. *Amersham Hybridization Oven/Shaker*, gamintojas *GE Healthcare, General Electric Corp.*, Jungtinė Karalystė)

---

## Darbo eiga

1. Veninis kraujas imamas į 5–10 ml vakuuminį mėgintuvėlį su antikoaguliantu EDTA.
2. Kraujas perpilamas į sterilų 45 ml mėgintuvėlį, iki 35 ml pripilama lizuojančio buferio. Centrifuguojama 10 min. 10 °C temperatūroje (šaldomojoje *Universal 32R*, *Hettich* centrifugoje) 3 000 aps./min. greičiu.
3. Viršnuosėdinis skystis nupilamas, nuosėdos sumaišomos pipetuoiant, pripilama lizuojančio buferio iki 10 ml ir centrifuguojama 10 min. 10 °C temperatūroje 3 000 aps./min. greičiu.
4. Viršnuosėdinis skystis nupilamas, nuosėdos sumaišomos pipete, pripilama lizuojančio buferio iki 5 ml. Centrifuguojama 10 min. 10 °C temperatūroje 3 000 aps./min. greičiu.
5. Viršnuosėdinis skystis nupilamas ir įpilama 400 µl ląstelių branduolių pernešimo buferio (10 mM Tris-HCl, pH = 10,5; 1 mM EDTA; 0,15 mM NaCl).
6. Visas turinys sterilia pipete perkliamas į 2 ml sterilų mėgintuvėlį, įpilama 20 µl 10 proc. natrio dodecilsulfato tirpalo ir 10 µl 2 proc. proteinazės K tirpalo.
7. Mišinys inkubuojamas 37 °C temperatūroje termostate 16 val.
8. Į mėgintuvėlį įpilama 400 µl fenolio ir mėgintuvėlis vartomas 10 min., po to centrifuguojama 2 min. 5 000 aps./min. greičiu.
9. Viršutinis sluoksnis nusiurbiamas ir perkliamas į 2 ml sterilų mėgintuvėlį, įpilama 200 µl fenolio ir 200 µl chloroformo. Mėgintuvėlis vartomas 5–10 min. Centrifuguojama 2 min. 5 000 aps./min. greičiu.
10. Viršutinis sluoksnis nusiurbiamas ir perkliamas į 2 ml sterilų mėgintuvėlį, įpilama 500 µl chloroformo ir izoamilo alkoholio mišinio (24:1). Mėgintuvėlis vartomas 5 min., po to centrifuguojama 2 min. 5 000 aps./min. greičiu.
11. Viršutinis sluoksnis nusiurbiamas ir perkliamas į 2 ml sterilų mėgintuvėlį, įpilama 400 µl chloroformo. Mėgintuvėlis vartomas 5 min., po to centrifuguojama 2 min. 5 000 aps./min. greičiu.
12. Viršutinis sluoksnis nusiurbiamas ir padalijus į 2 dalis perkliamas į 0,5 ml ir 1,5 ml sterilius mėgintuvėlius. Į kiekvieną mėgintuvėlį įpilama po 20 µl 3 M natrio acetato tirpalo ir po 500 µl 96 proc. šalto etanolio.
13. Mėgintuvėlius vartant, DNR iškrinta į nuosėdas. Centrifuguojama 5 min. 10 000–13 200 aps./min. greičiu.
14. Etanolis nupilamas ir įpilama 500 µl 70 proc. etanolio. Centrifuguojama 5 min. 10 000–13 200 aps./min. greičiu.

15. Etanolis nupilamas ir įpilama 500 µl 96 proc. etanolio. Centrifuguojama 5 min. 10 000–13 200 aps./min. greičiu.
16. Vienas (0,5 ml) mėgintuvėlis saugomas –20 °C temperatūroje. Kito mėgintuvėlio etanolis nupilamas ir mėgintuvėlyje likęs turinys liofilizuojamas 30 °C temperatūroje.
17. Išdžiovinta DNR užpilama 100 µl TE buferio (10 mM Tris-HCl; 1 mM EDTA, pH = 8) ir laikoma 4 °C temperatūroje, kol ištirpsta, mažiausiai 24 val.

Genominės DNR išskyrimas automatizuota *TECAN Freedom EVO® 200* sistema protokolas

96 proc. etanolis [C <sub>2</sub> H <sub>6</sub> O]
99 proc. izopropanolis (2-propanolis, CH <sub>3</sub> -CHOH-CH <sub>3</sub> )
Lizuojantis buferis (angl. <i>eLysis Buffer</i> )
Plovimo buferis etanolio ir izopropanolio pagrindu (angl. <i>Alcohol wash</i> )
Išplovimo buferis (angl. <i>Elution buffer</i> )
Ląstelių lizės tirpalas (angl. <i>Cell Lysis Solution</i> )
0,25 M NaOH
Magnetinės dalelės (angl. <i>MagneSil® Paramagnetic Particles</i> )
Didejonizuotas vanduo (ddH <sub>2</sub> O)
DNR išskyrimo rinkinys (angl. <i>MagneSil® Genomic, Large Volume System</i> , gamintojas <i>Promega Corp., US</i> )
DNR išskyrimo automatizuota sistema <i>TECAN Freedom EVO® 200</i>
Kaitinimo blokas su adapteriu (angl. <i>Digital Heatblock</i> )
Magnetinis stovas
Purtyklė su adapteriu
NALGENE indas
Sterilūs mėgintuvėliai (1,5 ml, 50 ml)
Latekso pirštinės

Reagentai ir laboratorinė įranga

Darbo eiga

1. Magnetines daleles, kurias laikomos 4 °C temperatūroje arba turi nuosėdų, prieš naudojimą inkubuoti 37 °C temperatūroje 30 min., po to purtyti 5 sekundes ir nusukti.
2. Patikrinti, ar *TECAN Freedom EVO® 200* sistemoje nėra nutekėjimų.
3. Jei yra skysčio, ištuštinti atliekų indą (angl. *Waste bottle*).
4. Skysčių sistemos (angl. *Liquid system*) indą pripildyti dejonizuotu vandeniu, sumaišytu su nedideliu kiekiu 1 proc. izopropanolio.

5. Nuvalyti beplauše servetėle prietaiso antgalius ir patikrinti, ar nėra vandens lašų.
6. Patikrinti, ar nėra oro burbulų prietaiso nutekėjimo vamzdžiuose.
7. Įjungti kaitinimo bloką su adapteriu ir nustatyti 95 °C temperatūrą.
8. Įjungti purtyklę su adapteriu.
9. Į prietaiso (1) poziciją įstatyti sterilius 1,5 ml mėgintuvėlius. Jų kiekis priklauso nuo kraujo ėminių skaičiaus (8, 16 arba 24).
10. Į prietaiso (2) pozicijas įstatyti atidarytus mėgintuvėlius su veniniu krauju (10 mm / 5 ml arba 16 mm / 10 ml su EDTA).
11. Į prietaiso stovelį įstatyti 50 ml tuščius mėgintuvėlius (8, 16 arba 24) ir pastatyti ant magnetinio paviršiaus (angl. *Magnet Position*).
12. Į prietaiso stovelį įstatyti aštuonis 50 ml mėgintuvėlius, pripildytus reikiamu kiekiu magnetinių dalelių.
13. Įpilti 900 ml lizuojančio buferio į L raide pažymėtą indą.
14. Įpilti 900 ml plovimo buferio etanolio ir izopropanolio pagrindu į A raide pažymėtą indą.
15. Įpilti 200 ml išplovimo buferio į E raide pažymėtą indą.
16. Įpilti 200 ml ląstelių lizės buferio į W raide pažymėtą indą.
17. Uždaryti prietaiso duris.
18. Įjungti prietaisą – 0,5 sek. laikyti nuspaudus mygtuką *ON / OFF* (žalia šviesa yra indikacija, kad prietaisas įjungtas).
19. Įjungti prietaiso PC ir programą *Freedom EVOware*® *Standard 2.3 (EVOware 2.3)*.
20. Pasižymėti *Edit an existing process* ir spausti *Start your selection* (žalia rodyklė).
21. Eiti į katalogą *Favorites* ir pasirinkti *Tecan\_gDNA\_XL\_10mm\_tubes\_V4\_0*, spausti *Run*.
22. Atlikti prietaiso plovimą – spausti dešinės pelės klavišą ir pasirinkti *Flush* ir *Run (Volume in waste = 30 ml; Volume in cleaner = 30 ml; pažymėti Use fast wash module)*.
23. Spausti *Run* ir pasirinkti *Start recovery mode: 0 (no)*, įvesti mėginių skaičių ir spausti *Yes*, jeigu perkeliame kraują iš 5 ml mėgintuvėlių į 50 ml mėgintuvėlius.
24. Pradėti procesą su vienu išplovimu.
25. Spausti *START the application*.
26. Pasibaigus procesui, spausti *Exit* ir *OK*.
27. Prietaisas išjungiamas spaudžiant mygtuką *ON / OFF*.
28. Patikrinti, ar nėra skysčių nutekėjimo. Antgalius nuvalyti išoriškai ir iš vidaus naudojant servetėles be plaušų ir 70 proc. etanolį.



29. Į prietaiso 42 poziciją įstatyti du indus: vieną su 96 proc. etanoliumi, kitą su 0,25 M NaOH.
30. Įjungti *EVOware 2.3* programą.
31. Pasirinkus *Run maintenance*, toliau pasirinkti *Clean\_Fix\_Tips\_Inside* ir spausti *Start your selection* (žalia rodyklė), spausti *Run*.

**1 lentelė.** DNR išskyrimui automatizuota TECAN sistema reikalingų reagentų tūriai

Ėminių skaičius	Lizuojantis buferis (ml)	Plovimo buferis (ml)	Išplovimo buferis (ml*)	Magnetinės dalelės (ml)
1	13,5 + V <sub>2lentelė</sub>	37,5	2,5	V <sub>2lentelė</sub>
8	108 + V <sub>2lentelė</sub>	300	20	V <sub>2lentelė</sub>
16	216 + V <sub>2lentelė</sub>	600	40	V <sub>2lentelė</sub>
24	324 + V <sub>2lentelė</sub>	900	60	V <sub>2lentelė</sub>

\*Išplovimo buferio kiekį galima keisti. Kuo mažesnis išplovimo buferio tūris, tuo didesnė išgrynintos DNR koncentracija (µg/ml), tačiau mažesnis bendras kiekis (µg). Jei norime didesnio DNR kiekio, išplovimas atliekamas du kartus, išplovimo buferio tūris dvigubas.

**2 lentelė.** DNR išskyrimui automatizuota TECAN sistema reikalingų reagentų tūriai

Ėminio tūris (ml)	Lizuojančio buferio tūris (ml)	Magnetinių dalelių tūris (ml)
1	0,9	0,35
2	1,8	0,7
3	2,7	1,05
4	3,6	1,4
5	4,5	1,75
6	5,4	2,1
7	6,3	2,45
8	7,2	2,8
9	8,1	3,15
10	9	3,5

## Genominės DNR koncentracijos ir švarumo matavimo protokolas

Išskirtos DNR koncentracija ir švarumas nustatomi spektrofotometru *NanoDropR ND-1000*. DNR tirpalo švarumas patikrinamas matuojant tirpalo optinį tankį 260 nm ir 280 nm ilgio bangomis. Optinių tankių santykis turi būti 1,7–1,9.

### Reagentai ir laboratorinė įranga

Didejonizuotas vanduo (ddH <sub>2</sub> O)
TE buferis (10 mM Tris-HCl, 1 mM EDTA, pH = 8)
gDNR tirpalas
Mikrocentrifuga (0,5 mL, 1,5 mL)
Mikropipetės (2–20 µl, 10–100 µl)
Sterilūs mėgintuvėliai (0,5 mL, 1,5 mL)
Sterilūs antgaliai (20 µL, 100 µL)
Purtyklė
<i>NanoDrop® ND-1000</i> spektrofotometras
Beplaušės servetėlės
Latekso pirštinės

### Darbo eiga

1. Įjungti *ND-1000 v3.5.2* programą.
2. Pasirinkti *Nucleic Acid*.
3. Pakelti spektrofotometro svirtelę ir beplauše servetėle, suvilgyta ddH<sub>2</sub>O, nuvalyti apatinį ir viršutinį DNR matavimo paviršius.
4. Aktyvuoti spektrofotometrą užlašinus mikropipete 2 µL ddH<sub>2</sub>O ir nuspaudus *OK*.
5. Pasibaigus prietaiso aktyvacijai, spausti *Blanc*.
6. Nuvalyti beplauše servetėle apatinį ir viršutinį DNR matavimo paviršius.
7. Supurtyti ir nucentrifuguoti gDNR tirpalo mėgintuvėlį.
8. Užlašinti mikropipete 2 µL gDNR tirpalo ant spektrofotometro matavimo paviršiaus ir spausti *Measure*. Spausti pakartotinai *Measure* ir išvesti gautų rezultatų vidurkį.
9. Išskirtos gDNR tirpalą, esant poreikiui, skiesti ddH<sub>2</sub>O arba TE buferiu iki reikiamos koncentracijos.
10. Baigus matavimą, spektrofotometro paviršius nuvalomas beplauše servetėle ir išplaunamas ddH<sub>2</sub>O.

## II PRIEDAS. Genominės DNR genotipavimo *Illumina* VNP lustais protokolai

Genotipavimo reakcija remiasi viso genomo amplifikacija, pagausinant DNR iki 1 000 kartų. DNR yra fragmentuojama, nusodinama, resuspenduojama ir hibridizuojama su VNP specifiniais pradmenimis, kurie yra imobilizuoti luste, bei atliekama vieno nukleotido pratęsimo reakcija, atliekamas pratęsto nukleotido žymėjimas ir žymės sustiprinimas dažant. Fluorochromais žymėti nukleotidai yra detektuojami *Illumina HiScanSQ* optine sistema, kuri identifikuoja spalvos ir signalo intensyvumą.

Genominės DNR genotipavimo *Illumina Infinium® HD* VNP lustais protokolas

Protokolas parengtas pagal *Illumina Infinium® HD Assay Ultra* naudotojo vadovą.

Genominės DNR gausinimas paruoštoje MSA3 plokštelėje

### Reagentai

Pavadinimas	Kiekis	Laikymas	Tiekėjas
MA1	1/96 mėginiams	Nuo –15 °C iki –25 °C	<i>Illumina</i>
MA2	1/96 mėginiams	Nuo –15 °C iki –25 °C	<i>Illumina</i>
MSM	1/96 mėginiams	Nuo –15 °C iki –25 °C	<i>Illumina</i>
0,1N NaOH	15 ml/96 mėginiams	Nuo 2 °C iki 8 °C	Naudotojas
96 šulinėlių 0,8 ml mikrotitro MIDI plokštelė	1 plokštelė		Naudotojas
DNR mėginiai (50 ng/μl)		Nuo –15 °C iki –25 °C	Naudotojas

### Pasiruošimas

- ✓ Įjungti *Illumina* hibridizacijos krosnį, nustatyti 37 °C temperatūrą.
- ✓ Ant naujos MIDI plokštelės užklijuoti MSA3 brūkšninį kodą.
- ✓ MA1, MA2 ir MSM reagentus atitirpinti kambario temperatūroje, centrifuguoti 280 xg.
- ✓ gDNR mėginius atitirpinti kambario temperatūroje.

## Darbo eiga

1. Perkelti po 20 µl MA1 reagento į MSA3 plokštelės šulinėlius, atsižvelgiant į gDNR mėginių skaičių.
2. Perkelti po 4 µl gDNR į MSA3 plokštelės šulinėlius.
3. Perkelti po 4 µl 0,1N NaOH į MSA3 plokštelės šulinėlius.
4. Uždengti MSA3 plokštelę dengiamąja guma.
5. Supurtyti plokštelę 1 600 aps./min. greičiu 1 min.
6. Centrifuguoti plokštelę 280 xg 1 min.
7. Inkubuoti plokštelę kambario temperatūroje 10 min.
8. Perkelti po 34 µl MA2 į kiekvieną MSA3 plokštelės šulinėlį su DNR mėginiais.
9. Perkelti po 38 µl MSM į kiekvieną MSA3 plokštelės šulinėlį su DNR mėginiais.
10. Uždengti MSA3 plokštelę dengiamąja guma.
11. Supurtyti plokštelę 1 600 aps./min. greičiu 1 min.
12. Centrifuguoti plokštelę 280 xg 1 min.
13. Inkubuoti plokštelę *Illumina* hibridizacijos krosnyje 20–24 val. 37 °C temperatūroje.

## Amplifikuotos gDNR fermentinė fragmentacija MSA3 plokštelėje

### Reagentai

Pavadinimas	Kiekis	Laikymas	Tiekėjas
FMS	1/96 mėginiams	Nuo –15 °C iki –25 °C	<i>Illumina</i>

### Pasiruošimas

- ✓ Įjungti mikroplokštelių inkubatorių *Hybex*<sup>®</sup> (angl. *Heat block*) su uždėta MIDI plokštele, nustatyti 37 °C temperatūrą.
- ✓ Atitirpinti FMS reagentą kambario temperatūroje. Ranka pavartyti, centrifuguoti 280 xg.
- ✓ Išimti MSA3 plokštelę iš *Illumina* hibridizacijos krosnies.
- ✓ Nuimti dengiamąją gumą.

## Darbo eiga

1. Centrifuguoti plokštelę 50 xg 1 min.
2. Perkelti po 25 µl FMS į kiekvieną MSA3 plokštelės šulinėlį su DNR mėginiais.
3. Uždengti MSA3 plokštelę dengiamąja guma.
4. Supurtyti plokštelę 1 600 aps./min. greičiu 1 min.

5. Centrifuguoti plokštelę 50 xg 1 min. 22 °C temperatūroje.
6. Inkubuoti plokštelę mikroplokštelių inkubatoriuje 37 °C temperatūroje 1 val.

gDNR nusodinimas MSA3 plokštelėje

#### Reagentai

Pavadinimas	Kiekis	Laikymas	Tiekėjas
PM1	1/96 mėginiams	Nuo 2 °C iki 8 °C	Illumina
100 proc. 2-propanolis	30 ml/96 mėginiams	Kambario temperatūra	Naudotojas

#### Pasiruošimas

- ✓ Įjungti, jei išjungtas, mikroplokštelių inkubatorių *Hybex*<sup>®</sup> (angl. *Heat block*) su uždėta MIDI plokštele, nustatyti 37 °C temperatūrą.
- ✓ Jei užšaldyta, atitirpinti MSA3 plokštelę kambario temperatūroje. Centrifuguoti 50 xg.
- ✓ Atitirpinti PM1 reagentą kambario temperatūroje. Centrifuguoti 280 xg 1 min.
- ✓ Nuimti dengiamąją gumą nuo MSA3 plokštelės.

#### Darbo eiga

1. Perkelti po 25 µl PM1 į kiekvieną MSA3 plokštelės šulinėlį su DNR mėginiais.
2. Uždengti MSA3 plokštelę dengiamąja guma.
3. Supurtyti plokštelę 1 600 aps./min. greičiu 1 min.
4. Inkubuoti plokštelę mikroplokštelių inkubatoriuje 37 °C temperatūroje 5 min.
5. Centrifuguoti plokštelę 50 xg 1 min. 22 °C temperatūroje.
6. Perkelti po 155 µl 100 proc. 2-propanolio į kiekvieną MSA3 plokštelės šulinėlį su DNR mėginiais.
7. Uždengti MSA3 plokštelę su nauja dengiamąja guma.
8. Vartyti plokštelę ranka 10 kartų.
9. Inkubuoti plokštelę 4 °C temperatūroje 30 min.
10. Centrifuguoti plokštelę 3 000 xg 4 °C temperatūroje 20 min. Pasibaigus centrifugavimui, kitas etapas pradedamas nedelsiant. Uždelsus, centrifugavimas kartojamas.
11. Nuimti dengiamąją gumą nuo MSA3 plokštelės.

12. Apversti plokštelę staigiu judesiu ant beplaušių servetėlių, kad būtų pašalintas viršnuosėdinis sluoksnis.
13. Daužyti plokštelę stipriai apie 1 min., neleidžiant viršnuosėdiniam sluoksniui pakliūti į kitus šulinėlius.
14. Džiovinti neuždengtą apverstą plokštelę ant mėgintuvėlių stovo kambario temperatūroje 1 val. Išdžiuvus nuosėdoms, šulinėlių dugne turi matytis melsvos spalvos nuosėdos.

Nusodintos gDNR resuspendacija

Reagentai

Pavadinimas	Kiekis	Laikymas	Tiekėjas
RA1	7 ml/96 mėginiams	Nuo –15 °C iki –25 °C	<i>llumina</i>

Pasiruošimas

- ✓ Jei užšaldyta, atitirpinti MSA3 plokštelę kambario temperatūroje. Nuimti dengiamąją gumą nuo MSA3 plokštelės ir išmesti.
- ✓ Įjungti *llumina* hibridizacijos krosnį, nustatyti 48 °C temperatūrą.
- ✓ Įjungti folijos užlydytuvą (angl. *Heat Sealer*), kaitinti 20 min.
- ✓ Atitirpinti RA1 reagentą kambario temperatūroje. Keletą kartų invertuoti.

Darbo eiga

1. Perkelti po 23 µl RA1 į kiekvieną MSA3 plokštelės šulinėlį su DNR nuosėdomis.
2. Uždengti folijos lapą ant plokštelės ir užlydyti folijos užlydytuvu, laikant nuleistą ir prispaustą 5 s (jei reikia, pakartoti).
3. Inkubuoti plokštelę *llumina* hibridizacijos krosnyje 48 °C temperatūroje 1 val.
4. Supurtyti plokštelę 1 800 aps./min. greičiu 1 min.
5. Trumpai centrifuguoti plokštelę 280 xg.

## gDNR hibridizacija ant VNP lustų (12x1 VNP lustai)

### Reagentai

Pavadinimas	Kiekis (96 mėginiams)	Laikymas	Tiekėjas
PB2	2	Kambario temperatūra	<i>Illumina</i>
Lustai	8		<i>Illumina</i>
Hibridizacijos kameros (angl. <i>Hybridization Chambers</i> )	2		<i>Illumina</i>
Hibridizacijos kameros guminės tarpinės (angl. <i>Hybridization Chamber gaskets</i> )	2		<i>Illumina</i>
Hibridizacijos kameros įdėklai (angl. <i>Hybridization Chamber inserts</i> )	8		<i>Illumina</i>

### Pasiruošimas

- ✓ Įjungti mikroplokštelių inkubatorių *Hybex*<sup>®</sup> (angl. *Heat block*), nustatyti 95 °C temperatūrą.
- ✓ Įjungti *Illumina* hibridizacijos krosnį, nustatyti 48 °C temperatūrą, nustatyti 5 padėklo siūbavimo greitį.

### Darbo eiga

#### Hibridizacijos kameros paruošimas

1. Įdėti į hibridizacijos kamerą guminę tarpinę.
2. Perkelti po 400 µl PB2 į kiekvieną iš aštuonių hibridizacijos kameros buferio rezervuarą.
3. Uždengti hibridizacijos kamerą dangčiu ir palikti kambario temperatūroje.

#### gDNR hibridizacija ant VNP lustų

1. Denatūruoti MSA3 plokštelėje esančius mėginius mikroplokštelių inkubatoriuje 95 °C temperatūroje 20 min.
2. Išimti VNP lustus iš šaldytuvo (2–8 °C), neišpakuoti, kol nebus perkelti DNR mėginiai.
3. Po denatūravimo MSA3 plokštelę atvėsinti kambario temperatūroje 30 min.
4. Trumpai centrifuguoti MSA3 plokštelę 280 xg.

## Mėginių perkėlimas ant VNP lustų

1. Išpakuoti VNP lustus.
2. Įdėti VNP lustus į hibridizacijos kameros įdėklus taip, kad brūkšninis kodas sutaptų su ant įdėklo esančiu brūkšninio kodo simboliu.
3. Perkelti po 15 µl DNR mėginio į kiekvieną VNP lusto laukelį pagal mėginių vietą VNP luste, pažymėtą darbo protokole, ir užrašyti lusto brūkšninį kodą.
4. Nedelsiant įdėti hibridizacijos kameros įdėklus su VNP lustais į hibridizacijos kamerą ir uždengti dangčiu.
5. Užspausti gnybtus abiejose hibridizacijos kameros pusėse, po vieną iš abiejų pusių, įstrižai.
6. Įdėti hibridizacijos kamerą į iki 48 °C temperatūros įkaitusią *Illumina* hibridizacijos krosnį taip, kad hibridizacijos kameros gnybtai būtų nukreipti į kairę ir į dešinę krosnies sienas. Hibridizacijos kameros dangčio *Illumina* logotipas turi būti nukreiptas į tyrėją.
7. Nustatyti 5 padėklo siūbavimo greitį.
8. Inkubuoti 48 °C temperatūroje 16–24 val.

## XC4 reagento paruošimas VNP lustų dažymui

1. Įpilti 330 ml 100 proc. etanolio į XC4 butelį.
2. 15 s stipriai purtyti.
3. Palikti butelį nusistovėti kambario temperatūroje per naktį.

## VNP lustų plovimas

### Reagentai

Pavadinimas	Kiekis	Laikymas	Tiekėjas
PB1	550 ml	Kambario temperatūra	<i>Illumina</i>
Daugiavietis VNP lustų talpiklis (angl. <i>Multi-Sample BeadChip Alignment Fixture</i> )	1		<i>Illumina</i>
Pratekamosios kameros, <i>Te-Flow</i> pratekamosios kameros su juodais rėmeliais, tarpinėmis, dengiamaisiais stiklais ir gnybtais (angl. <i>Te-Flow Flow-Through Chambers with black frames, spacers, glass backplates, and clamps</i> )	1		<i>Illumina</i>
Stikliniai plovimo indai	2		<i>Illumina</i>
Plovimo laikiklis	1		<i>Illumina</i>



## Pasiruošimas

- ✓ Išimti hibridizacijos kamerą iš *Illumina* hibridizacijos krosnies. Prieš atidengiant, atvėsinti kambario temperatūroje 25 min.
- ✓ Pripildyti du stiklinius plovimo indus po 200 ml PB1 reagento.
- ✓ Įpilti į daugiavietį lustų talpiklį 150 ml PB1 reagento.
- ✓ Atskirti permatomas plastikines tarpines nuo baltų.
- ✓ Nuvalyti dengiamuosius stiklus. Pirma nuplauti 1 proc. detergentu (*Alconox*) ir nuvalyti 70 proc. etanolium. Prieš naudojant, nuo sausų stiklų nupūsti dulkes suspausti oru.

## Darbo eiga

1. Įstatyti metalinį lustų plovimo laikiklį į pirmą indą su PB1 reagentu.
2. Išimti iš hibridizacijos kameros hibridizacijos kameros įdėklus su VNP lustais.
3. Išimti po vieną VNP lustus iš hibridizacijos kameros įdėklų.
4. Laikant lustą vienoje rankoje, lusto brūkšninis kodas turi būti arčiau tyrėjo, kita ranka vienu lėtu įstrižu judesiu nuplėšti VNP lusto plėvelę (angl. *IntelliHyb Seal*), neliečiant lusto paviršiaus.
5. Nedelsiant ir atsargiai įstatyti VNP lustą į metalinį lustų plovimo laikiklį, kuris yra panardintas pirmame inde su PB1. Įsitikinti, kad lustas yra visiškai apsemtas.
6. Kartoti 4–5 etapus kiekvienam VNP lustui.
7. Kilnoti metalinį lustų plovimo laikiklį į viršų ir į apačią, kertant reagento paviršių, 1 min.
8. Perkelti metalinį lustų plovimo laikiklį į antrą indą su PB1.
9. Kilnoti metalinį lustų plovimo laikiklį į viršų ir į apačią, kertant reagento paviršių, 1 min.

## *Te-Flow*<sup>TM</sup> pratekamųjų kamerų surinkimas

1. Įdėti juodus rėmelius kiekvienam lustui į daugiavietį lustų talpiklį.
2. Įstatyti lustus į juodus rėmelius taip, kad brūkšninis kodas būtų nukreiptas į tyrėją. Įsitikinti, kad lustai apsemti PB1 reagento.
3. Uždėti permatomą plastikinę tarpinę ant panardintų VNP lustų.
4. Uždėti talpiklio laikiklį (angl. *Alignment Bar*) ant lustų talpiklio.
5. Nupūsti dulkes suspaustu sausu oru nuo dengiamųjų stiklų.
6. Uždėti stiklą ant VNP lustų su plastikinėmis tarpinėmis. Stiklo nuožulnioji vieta, skirta reagentų rezervuarui, turi būti nukreipta į lusto apačią su brūkšniniu kodu.
7. Užspausti gnybtus ant kiekvieno surinktos pratekamosios kameros galo

5 mm atstumu nuo viršutinio krašto ir 5 mm atstumu nuo reagentų rezervuaro pradžios.

8. Nukirpti išsikišusius plastikinių tarpinių galus.
9. Nedelsiant išplauti hibridizacijos kamerą su ddH<sub>2</sub>O.

VNP lustų dažymas

Reagentai

Pavadinimas	Kiekis	Laikymas	Tiekėjas
RA1	10 ml 1–8 lustams 20 ml 9–16 lustų 30 ml 17–24 lustams	Nuo –15 °C iki –25 °C	<i>llumina</i>
XC1	1/4 lustams	Nuo –15 °C iki –25 °C	<i>llumina</i>
XC2	1/4 lustams	Nuo –15 °C iki –25 °C	<i>llumina</i>
TEM	1/4 lustams	Nuo –15 °C iki –25 °C	<i>llumina</i>
XC3	50 ml 1–8 lustams 100 ml 9–16 lustų 150 ml 17–24 lustams	Kambario temperatūra	<i>llumina</i>
STM	1/4 lustams	Nuo –15 °C iki –25 °C	<i>llumina</i>
ATM	1/4 lustams	Nuo –15 °C iki –25 °C	<i>llumina</i>
PB1	310 ml 1–8 lustams 285 ml 9–24 lustams	Kambario temperatūra	<i>llumina</i>
XC4	310 ml 1–8 lustams 285 ml 9–24 lustams	Kambario temperatūra	<i>llumina</i>
Alkanoksas	Pagal poreikį	Kambario temperatūra	Naudotojas
EtOH	Pagal poreikį	Kambario temperatūra	Naudotojas
95 proc. formamidas/1 mM EDTA	15 ml 1–8 lustams 17 ml 9–16 lustų 25 ml 17–24 lustams	Nuo –15 °C iki –25 °C	Naudotojas

Pasiruošimas

- ✓ Įsitikinti, kad termostatas su vandens cirkuliacijos sistema pripildytas reikiamu kiekiu vandens.
- ✓ Įjungti termostatą su vandens cirkuliacijos sistema, nustatyti tokią temperatūrą, kad kamerų laikiklio sistemos temperatūra būtų 44 °C.
- ✓ Pašalinti burbulus iš kamerų laikiklio sistemos.
- ✓ Patikrinti kamerų laikiklio sistemos temperatūrą skirtingose vietose, naudojant *llumina* termometrą.
- ✓ Atšildyti reagentus iki kambario temperatūros, centrifuguoti 3 000 xg 3 min.

- ✓ Supurtyti butelį su XC4 reagentu.

#### Vienos bazės pratęsimo reakcija

1. Pasiekus kamerų laikiklio sistemai  $44\text{ }^{\circ}\text{C}\pm 0,5\text{ }^{\circ}\text{C}$  temperatūrą, nedelsiant įstatyti pratekamasias kameras.
2. Pakartotinai supurtyti XC4 butelį.
3. Perkelti į kiekvieną pratekamosios kameros reagentų rezervuarą:
  - a) 150  $\mu\text{l}$  RA1. Inkubuoti 30 s. Kartoti 5 kartus;
  - b) 450  $\mu\text{l}$  XC1. Inkubuoti 10 min.;
  - c) 450  $\mu\text{l}$  XC2. Inkubuoti 10 min.;
  - d) 200  $\mu\text{l}$  TEM. Inkubuoti 15 min.;
  - e) 450  $\mu\text{l}$  95 proc. formamido/mM EDTA. Inkubuoti 1 min. Pakartoti 1 kartą;
  - f) inkubuoti 5 min.;
  - g) pradėti mažinti kamerų laikiklio sistemos temperatūrą iki temperatūros, nurodytos ant STM buteliuko, jei nenurodyta, kad reikia sumažinti iki  $37\text{ }^{\circ}\text{C}$  temperatūros;
  - h) 450  $\mu\text{l}$  XC3. Inkubuoti 1 min. Pakartoti 1 kartą.
4. Laukti, kol kamerų laikiklio sistema pasieks reikiamą temperatūrą.

#### VNP lustų dažymas

1. Įjungti kompiuterį ir *Illumina HiScan<sup>TM</sup>SQ* prietaisą, jei lustų skenavimas planuojamas iš karto po dažymo.
2. Pridėti į kiekvieną pratekamosios kameros reagentų rezervuarą:
  - a) 250  $\mu\text{l}$  STM. Inkubuoti 10 min.;
  - b) 450  $\mu\text{l}$  XC3. Inkubuoti 1 min. Pakartoti 1 kartą ir inkubuoti 5 min.;
  - c) 250  $\mu\text{l}$  ATM. Inkubuoti 10 min.;
  - d) 450  $\mu\text{l}$  XC3. Inkubuoti 1 min. Pakartoti 1 kartą ir inkubuoti 5 min.;
  - e) 250  $\mu\text{l}$  STM. Inkubuoti 10 min.;
  - f) 450  $\mu\text{l}$  XC3. Inkubuoti 1 min. Pakartoti 1 kartą ir inkubuoti 5 min.;
  - g) 250  $\mu\text{l}$  ATM. Inkubuoti 10 min.;
  - h) 450  $\mu\text{l}$  XC3. Inkubuoti 1 min. Pakartoti 1 kartą ir inkubuoti 5 min.;
  - i) 250  $\mu\text{l}$  STM. Inkubuoti 10 min.;
  - j) 450  $\mu\text{l}$  XC3. Inkubuoti 1 min. Pakartoti 1 kartą ir inkubuoti 5 min.
3. Nedelsiant išimti pratekamasias kameras iš kamerų laikiklio sistemos ir padėti horizontaliai ant darbostalio.

## VNP lustų plovimas ir padengimas

1. Įpilti 310 ml PB1 reagento į plovimo indą.
2. Įstatyti lustų dažymo laikiklį į plovimo indą.
3. Išardyti pratekamašias kameras. Atsargiai mentele atkabinti gnybtus, nuimti dengiamuosius stiklus ir plastikines tarpines.
4. Nedelsiant įstatyti lustus į dažymo laikiklį, esantį dažymo inde, lustų brūkšninis kodas turi būti tyrėjo pusėje. Įsitikinti, kad lustai apsemti reagento.
5. Lėtai kilnoti dažymo laikiklį aukštyn ir žemyn 10 kartų, kertant reagento paviršių.
6. Pimirkyti 5 min.
7. Supurtyti XC4 butelį.
8. Įpilti 310 ml XC4 į plovimo indą, neleidžiant stovėti inde ilgiau nei 10 min.
9. Perkelti dažymo laikiklį į plovimo indą su XC4. Lustų brūkšninis kodas turi būti tyrėjo pusėje.
10. Lėtai kilnoti dažymo laikiklį aukštyn ir žemyn 10 kartų, kertant reagento paviršių.
11. Pimirkyti 5 min.
12. Išimti dažymo laikiklį iš plovimo indo ir padėti ant mėgintuvėlių stovo su į viršų nukreiptais lustų brūkšniniais kodais.
13. Ištraukti lustus iš dažymo laikiklio ir padėti ant mėgintuvėlių stovo su į viršų nukreiptais lustų brūkšniniais kodais.
14. Džiovinti lustus vakuuminėje džiovyklėje 50–55 min., esant 675 mmHg (0,9 bar).
15. Nuvalyti išdžiuvusių lustų apačią beplauše servetėle, suvilgyta 70 proc. etanoliu, pašalinant pridžiuvusio XC4 reagento likučius.
16. Išplauti naudotus laboratorinius reikmenis ddH<sub>2</sub>O.

## VNP lustų skenavimas *Illumina iScan* skeneriu *HiScan<sup>TM</sup>SQ* sistema

### Darbo eiga

1. Įjungti *iScan Control Software* programą. Suvesti prisijungimo duomenis.
2. Įsikelti iš anksto *DMAP Client* programa pagal kiekvieno VNP lusto unikalų brūkšninį kodą atsisiųstus kiekvieno VNP lusto unikalius dekodavimo (angl. *Decode*) .dmap failus.

3. Spausti *iScan Control Software* programos ikoną.
4. Spausti *Start*.
5. Atsidarius prietaiso durims, įdėti paruoštus VNP lustus į *Illumina* nuskaitymo takelius. Spausti *Next*.
6. Užsidarius prietaiso durims, laukti lustų brūkšninių kodų nuskaitymo.
7. Spausti *Scan*.

Genominės DNR genotipavimo *Illumina Infinium HTS* VNP lustais protokolas

Protokolas parengtas pagal *Illumina Infinium HTS Assay* naudotojo vadovą.

Genominės DNR gausinimas paruoštoje MSA3 plokštelėje

Reagentai

Pavadinimas	Kiekis	Laikymas	Tiekėjas
MA1	1/96 mėginiams	Nuo –15 °C iki –25 °C	<i>Illumina</i>
MA2	1/96 mėginiams	Nuo –15 °C iki –25 °C	<i>Illumina</i>
MSM	1/96 mėginiams	Nuo –15 °C iki –25 °C	<i>Illumina</i>
0,1N NaOH	15 ml/96 mėginiams	Nuo 2 °C iki 8 °C	Naudotojas
96 šulinėlių 0,8 ml mikrotitro MIDI plokštelė	1 plokštelė		Naudotojas
DNR mėginiai (50 ng/μl)		Nuo –15 °C iki –25 °C	Naudotojas

Pasiruošimas

- ✓ Įjungti *Illumina* hibridizacijos krosnį, nustatyti 37 °C temperatūrą.
- ✓ Ant naujos MIDI plokštelės užklijuoti MSA3 brūkšninį kodą.
- ✓ MA1, MA2 ir MSM reagentus atitirpinti kambario temperatūroje, centrifuguoti 280 xg.
- ✓ gDNR mėginius atitirpinti kambario temperatūroje.

Darbo eiga

1. Perkelti po 20 μl MA1 reagento į MSA3 plokštelės šulinėlius, atsižvelgiant į gDNR mėginių skaičių.
2. Perkelti po 4 μl gDNR į MSA3 plokštelės šulinėlius.
3. Perkelti po 4 μl 0,1N NaOH į MSA3 plokštelės šulinėlius.

4. Uždengti MSA3 plokštelę dengiamąja guma.
5. Supurtyti plokštelę 1 600 aps./min. 1 min.
6. Centrifuguoti plokštelę 280 xg 1 min.
7. Inkubuoti plokštelę kambario temperatūroje 10 min.
8. Perkelti po 34 µl MA2 į kiekvieną MSA3 plokštelės šulinėlį su DNR mėginiais.
9. Perkelti po 38 µl MSM į kiekvieną MSA3 plokštelės šulinėlį su DNR mėginiais.
10. Uždengti MSA3 plokštelę dengiamąja guma.
11. Supurtyti plokštelę 1 600 aps./min. 1 min.
12. Centrifuguoti plokštelę 280 xg 1 min.
13. Inkubuoti plokštelę *Illumina* hibridizacijos krosnyje 20–24 val. 37 °C temperatūroje.

Amplifikuotos gDNR fermentinė fragmentacija MSA3 plokštelėje

Reagentai

Pavadinimas	Kiekis	Laikymas	Tiekėjas
FMS	1/96 mėginiams	Nuo –15 °C iki –25 °C	<i>Illumina</i>

Pasiruošimas

- ✓ Įjungti mikroplokštelių inkubatorių *Hybex*<sup>®</sup> (angl. *Heat block*) su uždėta MIDI plokštele, nustatyti 37 °C temperatūrą.
- ✓ Atitirpinti FMS reagentą kambario temperatūroje. Ranka pavartyti, centrifuguoti 280 xg.
- ✓ Išimti MSA3 plokštelę iš *Illumina* hibridizacijos krosnies.
- ✓ Nuimti dengiamąją gumą.

Darbo eiga

1. Centrifuguoti plokštelę 50 xg 1 min.
2. Perkelti po 25 µl FMS į kiekvieną MSA3 plokštelės šulinėlį su DNR mėginiais.
3. Uždengti MSA3 plokštelę dengiamąja guma.
4. Supurtyti plokštelę 1 600 aps./min. 1 min.
5. Centrifuguoti plokštelę 50 xg 1 min. 22 °C temperatūroje.
6. Inkubuoti plokštelę mikroplokštelių inkubatoriuje 37 °C temperatūroje 1 val.

## gDNR nusodinimas MSA3 plokštelėje

### Reagentai

Pavadinimas	Kiekis	Laikymas	Tiekėjas
PM1	1/96 mėginiams	Nuo 2 °C iki 8 °C	<i>Illumina</i>
100 proc. 2-propanolis	30 ml/96 mėginiams	Kambario temperatūra	Naudotojas

### Pasiruošimas

- ✓ Įjungti, jei išjungtas, mikroplokštelių inkubatorių *Hybex*<sup>®</sup> (angl. *Heat block*) su uždėta MIDI plokštele, nustatyti 37 °C temperatūrą.
- ✓ Jei užšaldyta, atitirpinti MSA3 plokštelę kambario temperatūroje. Centrifuguoti 50 xg.
- ✓ Atitirpinti PM1 reagentą kambario temperatūroje. Centrifuguoti 280 xg 1 min.
- ✓ Nuimti dengiamąją gumą nuo MSA3 plokštelės.

### Darbo eiga

1. Perkelti po 50 µl PM1 į kiekvieną MSA3 plokštelės šulinėlį su DNR mėginiais.
2. Uždengti MSA3 plokštelę dengiamąja guma.
3. Supurtyti plokštelę 1 600 aps./min. 1 min.
4. Inkubuoti plokštelę mikroplokštelių inkubatoriuje 37 °C temperatūroje 5 min.
5. Centrifuguoti plokštelę 280 xg 1 min.
6. Atsargiai nuimti dengiamąją gumą nuo plokštelės ir išmesti.
7. Perkelti po 155 µl 100 proc. 2-propanolio į kiekvieną MSA3 plokštelės šulinėlį su DNR mėginiais.
8. Uždengti MSA3 plokštelę su nauja, sausa dengiamąja guma.
9. Vartyti plokštelę ranka 10 kartų.
10. Inkubuoti plokštelę 4 °C temperatūroje 30 min.
11. Centrifuguoti plokštelę 3 000 xg 4 °C temperatūroje 20 min. Pasibaigus centrifugavimui, kitas etapas pradedamas nedelsiant. Uždelsus, centrifugavimas kartojamas.
12. Nedelsiant išimti MSA3 plokštelę iš centrifugos.
13. Nuimti dengiamąją gumą nuo MSA3 plokštelės ir išmesti.
14. Apversti plokštelę staigiu judesiu ant beplaušių servetėlių, kad galima

- būtų pašalinti viršnuosėdinį sluoksnį.
15. Daužyti plokštelę stipriai apie 1 min., neleidžiant viršnuosėdiniam sluoksniui pakliūti į kitus šulinėlius.
  16. Džiovinti neuždengtą, apverstą plokštelę ant mėgintuvėlių stovo kambario temperatūroje 1 val. Išdžiuvus nuosėdoms, šulinėlių dugne turi matytis melsvos spalvos nuosėdos.

#### Nusodintos gDNR resuspendacija

##### Reagentai

Pavadinimas	Kiekis	Laikymas	Tiekėjas
RA1	7 ml/96 mėginiams	Nuo -15 °C iki -25 °C	<i>Illumina</i>

##### Pasiruošimas

- ✓ Jei užšaldyta, atitirpinti MSA3 plokštelę kambario temperatūroje. Nuimti dengiamąją gumą nuo MSA3 plokštelės ir išmesti.
- ✓ Įjungti *Illumina* hibridizacijos krosnį, nustatyti 48 °C temperatūrą.
- ✓ Įjungti folijos užlydytuvą (angl. *Heat Sealer*), kaitinti 20 min.
- ✓ Atitirpinti RA1 reagentą kambario temperatūroje. Keletą kartų invertuoti.

##### Darbo eiga

1. Perkelti po 23 µl RA1 į kiekvieną MSA3 plokštelės šulinėlį su DNR nuosėdomis.
2. Uždengti folijos lapą ant plokštelės ir užlydyti folijos užlydytuvu, laikant nuleistą ir prispaustą 5 s (jei reikia, pakartoti).
3. Inkubuoti plokštelę *Illumina* hibridizacijos krosnyje 48 °C temperatūroje 1 val.
4. Supurtyti plokštelę 1 800 aps./min. 1 min.
5. Trumpai centrifuguoti plokštelę 280 xg.



## gDNR hibridizacija ant HTS VNP lustu (24x1 VNP lustai)

### Reagentai

Pavadinimas	Kiekis (96 mėginiams)	Laikymas	Tiekėjas
PB2	1	Kambario temperatūra	<i>Illumina</i>
Lustai	4		<i>Illumina</i>
Hibridizacijos kameros (angl. <i>Hybridization Chambers</i> )	1		<i>Illumina</i>
Hibridizacijos kameros guminės tarpinės (angl. <i>Hybridization Chamber gaskets</i> )	1		<i>Illumina</i>
Hibridizacijos kameros įdėklai (angl. <i>Hybridization Chamber inserts</i> )	4		<i>Illumina</i>

### Pasiruošimas

- ✓ Įjungti mikroplokštelių inkubatorių *Hybex*<sup>®</sup> (angl. *Heat block*), nustatyti 95 °C temperatūrą.
- ✓ Įjungti *Illumina* hibridizacijos krosnį, nustatyti 48 °C temperatūrą, nustatyti 5 padėklo siūbavimo greitį.

### Darbo eiga

#### Hibridizacijos kameros paruošimas

1. Įdėti į hibridizacijos kamerą guminę tarpinę.
2. Perkelti po 400 μl PB2 į kiekvieną iš aštuonių hibridizacijos kameros buferio rezervuarą.
3. Uždengti hibridizacijos kamerą dangčiu ir palikti kambario temperatūroje.

#### gDNR hibridizacija ant VNP lustu

1. Denatūruoti MSA3 plokštelėje esančius mėginius mikroplokštelių inkubatoriuje 95 °C temperatūroje 20 min.
2. Išimti VNP lustus iš šaldytuvo (2 °C–8 °C), neišpakuoti, kol nebus perkelti DNR mėginiai.
3. Po denatūravimo MSA3 plokštelę atvėsinti kambario temperatūroje 30 min.
4. Trumpai centrifuguoti MSA3 plokštelę 280 xg. Nuimti folijos lapą.

## Mėginių perkėlimas ant VNP lustų

1. Išpakuoti VNP lustus.
2. Įdėti VNP lustus į hibridizacijos kameros įdėklus taip, kad brūkšninis kodas sutaptų su ant įdėklo esančiu brūkšninio kodo simboliu.
3. Perkelti po 14 µl DNR mėginio į kiekvieną VNP lusto laukelį pagal mėginių vietą VNP luste, pažymėtą darbo protokole, ir užrašyti lusto brūkšninį kodą.
4. Nedelsiant įdėti hibridizacijos kameros įdėklus su VNP lustais į hibridizacijos kamerą ir uždengti dangčiu.
5. Užspausti gnybtus abiejose hibridizacijos kameros pusėse, po vieną iš abiejų pusių, įstrižai.
6. Įdėti hibridizacijos kamerą į iki 48 °C temperatūros įkaitusią *Illumina* hibridizacijos krosnį taip, kad hibridizacijos kameros gnybtai būtų nukreipti į kairę ir dešinę krosnies sienas. Hibridizacijos kameros dangčio *Illumina* logotipas turi būti nukreiptas į tyrėją.
7. Nustatyti 5 padėklo siūbavimo greitį.
8. Inkubuoti 48 °C temperatūroje 16–24 val.

## XC4 reagento paruošimas VNP lustų dažymui

1. Įpilti 330 ml 100 proc. etanolio į XC4 butelį.
2. 15 s stipriai purtyti.
3. Palikti butelį nusistovėti kambario temperatūroje per naktį.

## VNP lustų plovimas

### Reagentai

Pavadinimas	Kiekis	Laikymas	Tiekėjas
PB1	550 ml 1–8 lustų 700 ml 9–16 lustų 850 ml 17–24 lustų	Kambario temperatūra	<i>Illumina</i>
Daugiavietis VNP lustų talpiklis (angl. <i>Multi-Sample BeadChip Alignment Fixture</i> )	1/8 lustams		<i>Illumina</i>
Pratekamosios kameros, <i>Te-Flow</i> pratekamosios kameros su juodais rėmeliais, LCG tarpinėmis, LCG dengiamaisiais stiklais ir gnybtais (angl. <i>Te-Flow LCG Flow-through chambers, with black frames, LCG spacers, LCG glass back plates, and clamps</i> )	1/lustui		<i>Illumina</i>
Stikliniai plovimo indai	2/8 lustams		<i>Illumina</i>
Plovimo laikiklis	1/8 lustams		<i>Illumina</i>

## Pasiruošimas

- ✓ Išimti hibridizacijos kamerą iš *Illumina* hibridizacijos krosnies. Prieš atidengiant, atvėsinti kambario temperatūroje 30 min.
- ✓ Pripildyti du stiklinius plovimo indus po 200 ml PB1 reagento.
- ✓ Įpilti į daugiavietį lustų talpiklį 150 ml PB1 reagento.
- ✓ Atskirti permatomas LCG plastikines tarpines nuo baltų.
- ✓ Nuvalyti LCG dengiamuosius stiklus. Pirmiau nuplauti 1 proc. detergentu (*Alconox*), po to nuvalyti 70 proc. etanolium. Prieš naudojant, nuo sausų stiklų nupūsti dulkes suspaustu oru.

## Darbo eiga

1. Įstatyti metalinį lustų plovimo laikiklį į pirmą indą su PB1 reagentu.
2. Išimti iš hibridizacijos kameros įdėklus su VNP lustais.
3. Išimti po vieną VNP lustus iš hibridizacijos kameros įdėklų.
4. Laikant lustą vienoje rankoje, lusto brūkšninis kodas turi būti arčiau tyrėjo, kita ranka vienu lėtu įstrižu judesiu nuplėšti VNP lusto plėvelę (angl. *IntelliHyb Seal*), neliečiant lusto paviršiaus.
5. Nedelsiant ir atsargiai įstatyti VNP lustą į metalinį lustų plovimo laikiklį, kuris yra panardintas pirmame inde su PB1. Įsitikinti, kad lustas yra visiškai apsemtas.
6. Kartoti 4–5 etapus kiekvienam VNP lustui.
7. Kilnoti metalinį lustų plovimo laikiklį į viršų ir į apačią, kertant reagento paviršių, 1 min.
8. Perkelti metalinį lustų plovimo laikiklį į antrą indą su PB1.
9. Kilnoti metalinį lustų plovimo laikiklį į viršų ir į apačią, kertant reagento paviršių, 1 min.

## *Te-Flow*<sup>TM</sup> pratekamųjų kamerų surinkimas

1. Įdėti juodus rėmelius kiekvienam lustui į daugiavietį lustų talpiklį.
2. Įstatyti lustus į juodus rėmelius taip, kad brūkšninis kodas būtų nukreiptas į tyrėją. Įsitikinti, kad lustai apsemti PB1 reagento.
3. Uždėti permatomą LCG plastikinę tarpinę ant panardintų VNP lustų.
4. Uždėti talpiklio laikiklį (angl. *Alignment Bar*) ant lustų talpiklio.
5. Nupūsti dulkes suspaustu sausu oru nuo LCG dengiamųjų stiklų.
6. Uždėti LCG dengiamuosius stiklus ant VNP lustų su plastikinėmis tarpinėmis. Stiklo nuožulnioji vieta, skirta reagentų rezervuarui, turi būti nukreipta į lusto apačią su brūkšniniu kodu.

7. Užspausti gnybtus ant kiekvieno surinktos pratekamosios kameros galo 5 mm atstumu nuo viršutinio krašto ir 5 mm atstumu nuo reagentų rezervuaro pradžios.
8. Nukirpti išsikišusius plastikinių tarpinių galus.
9. Nedelsiant išplauti hibridizacijos kamerą su ddH<sub>2</sub>O.

VNP lustų dažymas

Reagentai

Pavadinimas	Kiekis	Laikymas	Tiekėjas
RA1	10 ml 1–8 lustams 20 ml 9–16 lustų 30 ml 17–24 lustams	Nuo –15 °C iki –25 °C	<i>illumina</i>
LX1	2/8 lustams	Nuo –15 °C iki –25 °C	<i>illumina</i>
LX2	2/8 lustams	Nuo –15 °C iki –25 °C	<i>illumina</i>
EML	2/8 lustams	Nuo –15 °C iki –25 °C	<i>illumina</i>
XC3	50 ml 1–8 lustams 100 ml 9–16 lustų 150 ml 17–24 lustams	Kambario temperatūra	<i>illumina</i>
SML	2/8 lustams	Nuo –15 °C iki –25 °C	<i>illumina</i>
ATM	2/8 lustams	Nuo –15 °C iki –25 °C	<i>illumina</i>
PB1	310 ml 1–8 lustams 285 ml 9–24 lustams	Kambario temperatūra	<i>illumina</i>
XC4	310 ml 1–8 lustams 285 ml 9–24 lustams	Kambario temperatūra	<i>illumina</i>
Alkanoksas	Pagal poreikį	Kambario temperatūra	Naudotojas
EtOH	Pagal poreikį	Kambario temperatūra	Naudotojas
95proc. formamidas/1 mM EDTA	15 ml 1–8 lustams 17 ml 9–16 lustų 25 ml 17–24 lustams	Nuo –15 °C iki –25 °C	Naudotoja

Pasiruošimas

- ✓ Įsitikinti, kad termostatas su vandens cirkuliacijos sistema pripildytas reikiamu kiekiu vandens.
- ✓ Įjungti termostatą su vandens cirkuliacijos sistema, nustatyti tokią temperatūrą, kad kamerų laikiklio sistemos temperatūra būtų 44 °C ± 0,5 °C.
- ✓ Pašalinti burbulus iš kamerų laikiklio sistemos.
- ✓ Patikrinti kamerų laikiklio sistemos temperatūrą skirtingose vietose, naudojant *illumina* termometrą.

- ✓ Atšildyti reagentus iki kambario temperatūros, centrifuguoti 3 000 xg 3 min.
- ✓ Supurtyti butelį su XC4 reagentu.

#### Vienos bazės pratęsimo reakcija

1. Pasiekus kamerų laikiklio sistemai 44 °C temperatūrą, nedelsiant įstatyti pratekamasias kameras.
2. Pakartotinai supurtyti XC4 butelį.
3. Perkelti į kiekvieną pratekamosios kameros reagentų rezervuarą:
  - a) 150 µl RA1. Inkubuoti 30 s. Kartoti 5 kartus;
  - b) 225 µl LX1. Pakartoti 1 kartą\*. Inkubuoti 10 min.;
  - c) 225 µl LX2. Pakartoti 1 kartą\*. Inkubuoti 10 min.;
  - d) 300 µl EML. Inkubuoti 15 min.;
  - e) 250 µl 95 proc. formamido/mM EDTA. Inkubuoti 1 min. Pakartoti 2 kartus\*;
  - f) inkubuoti 5 min.;
  - g) pradėti mažinti kamerų laikiklio sistemos temperatūrą iki temperatūros, nurodytos ant SML buteliuko.
  - h) 250 µl XC3. Inkubuoti 1 min. Pakartoti 2 kartus\*.
4. Laukti, kol kamerų laikiklio sistema pasieks reikiamą temperatūrą.  
\*Reagentų išpilstymo (angl. *Dispensing*) metodas.

#### VNP lustų dažymas

1. Įjungti kompiuterį ir *Illumina HiScan<sup>TM</sup>SQ* prietaisą, jei lustų skenavimas planuojamas iš karto po dažymo.
2. Perkelti į kiekvieną pratekamosios kameros reagentų rezervuarą:
  - a) 250 µl SML. Inkubuoti 10 min.;
  - b) 250 µl XC3. Inkubuoti 1 min. Pakartoti 2 kartus\* ir inkubuoti 5 min.;
  - c) 250 µl ATM. Inkubuoti 10 min.;
  - d) 250 µl XC3. Inkubuoti 1 min. Pakartoti 2 kartus\* ir inkubuoti 5 min.;
  - e) 250 µl SML. Inkubuoti 10 min.;
  - f) 250 µl XC3. Inkubuoti 1 min. Pakartoti 2 kartus\* ir inkubuoti 5 min.;
  - g) 250 µl ATM. Inkubuoti 10 min.;
  - h) 250 µl XC3. Inkubuoti 1 min. Pakartoti 2 kartus\* ir inkubuoti 5 min.;

- i) 250 µl SML. Inkubuoti 10 min.;
  - j) 250 µl XC3. Inkubuoti 1 min. Pakartoti 2 kartus\* ir inkubuoti 5 min.
3. Nedelsiant išimti pratekamašias kameras iš kamerų laikiklio sistemos ir padėti horizontaliai ant darbatalio.

#### VNP lustų plovimas ir padengimas

1. Įpilti 310 ml PB1 reagento į plovimo indą.
2. Įstatyti lustų dažymo laikiklį į plovimo indą.
3. Išardyti pratekamašias kameras. Atsargiai mentele atkabinti gnybtus, nuimti dengiamuosius stiklus ir plastikines tarpines.
4. Nedelsiant įstatyti lustus į dažymo laikiklį, esantį dažymo inde, lustų brūkšninis kodas turi būti nukreiptas nuo tyrėjo. Įsitikinti, kad lustai apsemti reagento.
5. Lėtai kilnoti dažymo laikiklį aukštyn ir žemyn, kertant reagento paviršių, 10 kartų.
6. Pamirkyti 5 min.
7. Supurtyti XC4 butelį.
8. Įpilti 310 ml XC4 į plovimo indą, neleidžiant stovėti inde ilgiau nei 10 min.
9. Perkelti dažymo laikiklį į plovimo indą su XC4. Lustų brūkšninis kodas turi būti nukreiptas nuo tyrėjo.
10. Lėtai kilnoti dažymo laikiklį aukštyn ir žemyn, kertant reagento paviršių, 10 kartų.
11. Pamirkyti 5 min.
12. Išimti dažymo laikiklį iš plovimo indo ir padėti ant mėgintuvėlių stovo su į viršų nukreiptais lustų brūkšniniais kodais.
13. Ištraukti lustus iš dažymo laikiklio ir padėti ant mėgintuvėlių stovo su į viršų nukreiptais lustų brūkšniniais kodais.
14. Džiovinti lustus vakuuminėje džiovyklėje 50–55 min., esant 508 mmHg (0,68 bar).
15. Nuvalyti išdžiuvusių lustų apačią beplauše servetėle, suvilgyta 70 proc. etanoliu, pašalinant pridžiuvusio XC4 reagento likučius.
16. Išplauti naudotus laboratorinius reikmenis ddH<sub>2</sub>O.

## VNP lustų skenavimas *Illumina iScan* skeneriu *HiScan<sup>TM</sup>SQ* sistema

### Darbo eiga

1. Įjungti *iScan Control Software* programą. Suvesti prisijungimo duomenis.
2. Įsikelti iš anksto *DMAP Client* programa pagal kiekvieno VNP lusto unikalų brūkšninį kodą atsisiųstus kiekvieno VNP lusto unikalius dekodavimo (angl. *Decode*) .dmap failus.
3. Spausti *iScan Control Software* programos ikoną.
4. Spausti *Start*.
5. Atsidarius prietaiso durims, įdėti paruoštus VNP lustus į *Illumina* nuskaitymo takelius. Spausti *Next*.
6. Užsidarius prietaiso durims, laukti lustų brūkšninių kodų nuskaitymo.
7. Spausti *Scan*.

### III PRIEDAS. Analizės metodai

#### I. Projekto sukūrimas *GenomeStudio* v2011.1 programa

Parengta pagal *GenomeStudio Genotyping Module v1.0* naudotojo vadovą.

1. Atidaryti *GenomeStudio* programą.
2. *New Project* lange spausti *Genotyping* arba pasirinkti *File / New Project / Genotyping*.
3. Spausti *Next*.
4. Atsidariusiame lange *Project Repository* laukelyje nurodyti direktoriją, kur bus išsaugotas sukurtas projektas, ir žemiau esančiame laukelyje *Project Name* sukurti projekto pavadinimą.
5. Spausti *Next*.
6. Pasirinkti *Use sample sheet to load sample intensities*.
7. Spausti *Next*.
8. Atsidariusiame lange ieškoti ir pasirinkti projektui reikalingus failus: (1) *Sample Sheet*, .csv failas, informacija apie mėginius ir jų vietą luste; (2) *Data Repository*, \*.idat failai, VNP intensyvumo informacija; (3) *Manifest Repository*, .bpm failai, VNP identifikavimas luste.
9. Spausti *Next*.
10. Atsidariusiame lange pažymėti *Import cluster positions from a cluster file*, .egt failai, reikalingos VNP klasterizavimui. Pažymėti langelius *Pre-Calculate* ir *Cluster SNPs*. *Gen Call Threshold* parametro vertė paliekama 0,15.
11. Spausti *Finish*.

#### II. Įvesties bylų sukūrimas *GenomeStudio* programa

1. *GenomeStudio* programoje pasirinkti *Analysis | Reports | Report Wizard / Custom Report / PLINK Input Report 2.1.3 by Illumina, Inc. From Illumina, Inc.*
2. Pasirinkti įvesties failų parametrus.
3. Spausti *Next*.
4. Atsidariusiame lange pažymėti *Selected Samples*.
5. Spausti *Next*.



6. Atsidariusiame lange pažymėti *Remove zeroed SNPs from the report.*
7. Spausti *Next.*
8. Atsidariusio lango *Output Path* langelyje nurodyti sukurtų įvesties bylų išsaugojimo direktoriją. *Output Path* laukelyje nurodyti įvesties failo pavadinimą.
9. Spausti *Finish.*

III. Efektyvaus populiacijos dydžio ir divergencijos laiko nustatymo R programinės įrangos *NeON* v. 1.0 paketu protokolas

Parengta pagal *NeON* 2013 m. spalio 24 d. naudotojo vadovą.

1. Nepusiausvyros sankibos nustatymas *NeLD* funkcija

Komandinė eilutė: *NeLD(plink.file, geno=0.02, mind=0.9, ld.window.kb=500, ld.window=9999, outfile="output.ld")*

*plink.file* – genotipuotų duomenų *plink* programos .bed .bim .fam failai

*geno* – VNP, kuriems trūksta >2 proc. genotipo, pašalinimas

*mind* – individų, kuriems trūksta >10 proc. genotipo, pašalinimas

*ld.window.kb* – nepusiausvyros sankibos nustatymo apimtis chromosomoje, 500 kb

*ld.window* – maksimalus VNP skaičius vienam skaičiavimui, 9 999 VNP

*outfile* – sukurtos išvesties failo pavadinimas (informacija apie  $r^2_{LD}$  tarp žymenų)

2. Efektyvaus populiacijos dydžio nustatymas *Nestimate* funkcija

Komandinė eilutė: *Nestimate(file.ld, sample.size, min.R2=0.001, max.R2=0.999, method='MG', min.cfr=5)*

*file.ld* – *NeLD* funkcijos išvesties failas .ld

*sample.size* – individų skaičius, naudotas  $r^2$  nustatyti populiacijoje

*min.R2* – mažiausia leistina  $r^2$  reikšmė

*max.R2* – didžiausia leistina  $r^2$  reikšmė

*method* – galimi du skirtingi metodai *Ne* apskaičiuoti: 1) \*'McEvoy'; 2) 'MG', pagal nutylėjimą

*min.cfr* – mažiausias leistinas kiekvienoje genetinių atstumų kategorijoje palyginimų skaičius *Ne* vidurkiui apskaičiuoti, pagal nutylėjimą 5

\**Nestimate*, remiantis McEvoy'aus metodu, sukuria 50 nepersidengiančių genetinių atstumų kategorijų nuo 0,005 iki 0,25 cM. Mezzavilla'os-Ghirotto'o metodas, kuris buvo naudotas šiame darbe, sukuria 250

persidengiančių genetinių atstumų kategorijų kas 0,001 cM nuo 0,005 iki 0,25 cM. Efektyvus populiacijos dydis apskaičiuojamas kiekvienai genetinių atstumų kategorijai (50 arba 250).

3. \*Ilgalaikio efektyvaus populiacijos dydžio nustatymas *Ne\_CI* funkcija

Komandinė eilutė: *Ne\_CI(Nestimate.output, ci=c(0.05, 0.5, 0.95))*

*Nestimate.output* – *Nestimate* funkcijos išvesties failas

*ci* – vektorius, pagal kurį apskaičiuojami ilgalaikio *Ne* pasikliautiniai intervalai, pagal nutylėjimą 5, 50 ir 95 procentilės

\*Ilgalaikis *Ne* apskaičiuojamas kaip harmoninis vidurkis.

4. Efektyvaus populiacijos dydžio nustatymas laike su pasikliautiniais intervalais *Ne\_Med* funkcija

Komandinė eilutė: *Ne\_Med(Nestimate.output, method='MG', ci=FALSE, ci.int=c(0.05, 0.5, 0.95))*

*Nestimate.output* – *Nestimate* funkcijos išvesties failas

*method* – du galimi metodai *Ne* apskaičiuoti: 1) 'McEvoy'; 2) 'MG'

*ci* – jeigu reikia apskaičiuoti pasikliautinuosius intervalus (pagal nutylėjimą FALSE)

*ci.int* – jeigu *ci=TRUE*, nurodyti kvartilius (pagal nutylėjimą 5, 50 ir 95 procentilės)

5. Atvaizdavimas *Neplot* funkcija

Komandinė eilutė: *Neplot(Ne.file, approx=TRUE, ylim=c(0, 15000), xlim=c(200, 6000), main="Ne from linkage disequilibrium", xlab="Generation ago", ylab="Ne", ci=TRUE)*

*Ne.file* – *Ne\_Med* funkcijos išvesties failas

*approx* – taškų interpoliacijai, pagal nutylėjimą TRUE

*ylim* – vektorius su viršutinėmis ir apatinėmis *Ne* reikšmėmis y ašyje, pagal nutylėjimą (0, 15000)

*xlim* – vektorius su viršutinėmis ir apatinėmis kartų reikšmėmis x ašyje, pagal nutylėjimą (200, 6 000)

*main* – grafikos pavadinimas

*xlab* – x ašies pavadinimas

*ylab* – y ašies pavadinimas

*ci* – pasikliautinių intervalų atvaizdavimui, pagal nutylėjimą (TRUE)

6. Divergencijos laiko tarp populiacijų nustatymas *Tdverg* funkcija

Komandinė eilutė: *Tdverg(Fst, All\_H)*

*F<sub>ST</sub>* – genetinių atstumų matrica

*All\_H* – .txt failas su kiekvienos populiacijos ilgalaikiu *Ne*

Išvesties failas matrica su divergencijos laiku, kartomis, tarp tiriamų populiacijų. Atvaizdavimui naudojama R programinės įrangos *phangorn* biblioteka.

#### IV. Trumpalaikio efektyvaus populiacijos dydžio analizės protokolas

Parengta pagal *IBDseq* ir *IBDNe* 2013 m. lapkričio 7 d. naudotojo vadovą.

##### 1. Identiškų pagal kilmę segmentų nustatymas *IBDseq* v.r1206 programiniu paketu

Komandinė eilutė: `java -jar ibdseq.r1206.jar gt=[46] chrom=[number] errormax=0.0025 out=[prefix]`

*gt* – VCF formato įvesties failas

*chrom* – nurodoma analizuojama chromosoma

*errormax* – maksimali alelio dažnio paklaida

*out* – išvesties failo pavadinimas

Programa sukuria tris išvesties failus – .log, .ibd, .hbd. *R2-filtered.Log* byloje įrašoma informacija apie analizuojamų mėginių, žymenų skaičių, naudojamus parametrus. *R2-filtered* byloje įrašoma pašalintų žymenų informacija. .ibd byloje pateikiama informacija apie nustatytus identiškus pagal kilmę segmentus.

##### 2. Trumpalaikio efektyvaus populiacijos dydžio nustatymas *IBDNe* v. 04Sep15.e78 programiniu paketu

Komandinė eilutė: `cat *.ibd | java -jar ibdne.04Sep15.e78.jar map=plink.GRCh37 minibd=7 out=[prefix]`

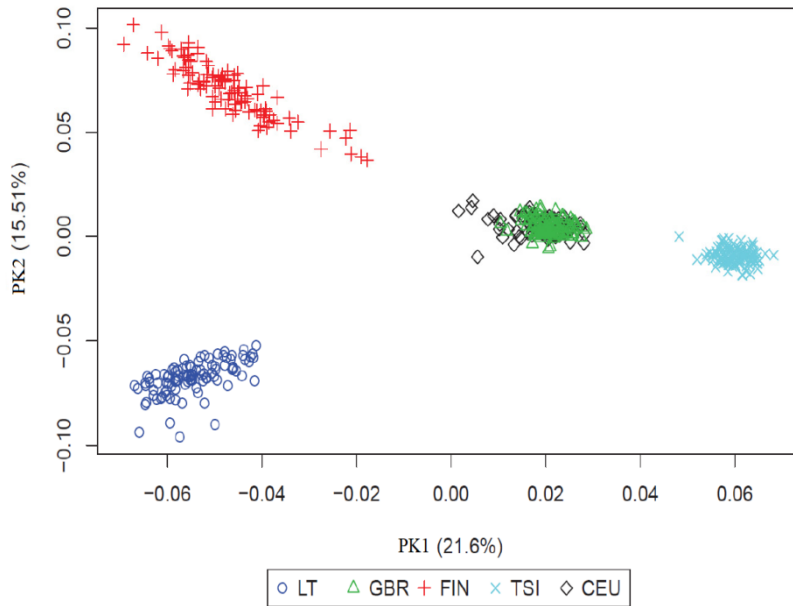
IBD – *IBDseq* programos sukurti .ibd failai, kurie sujungiami *cat* parametru

*map* – genolapis

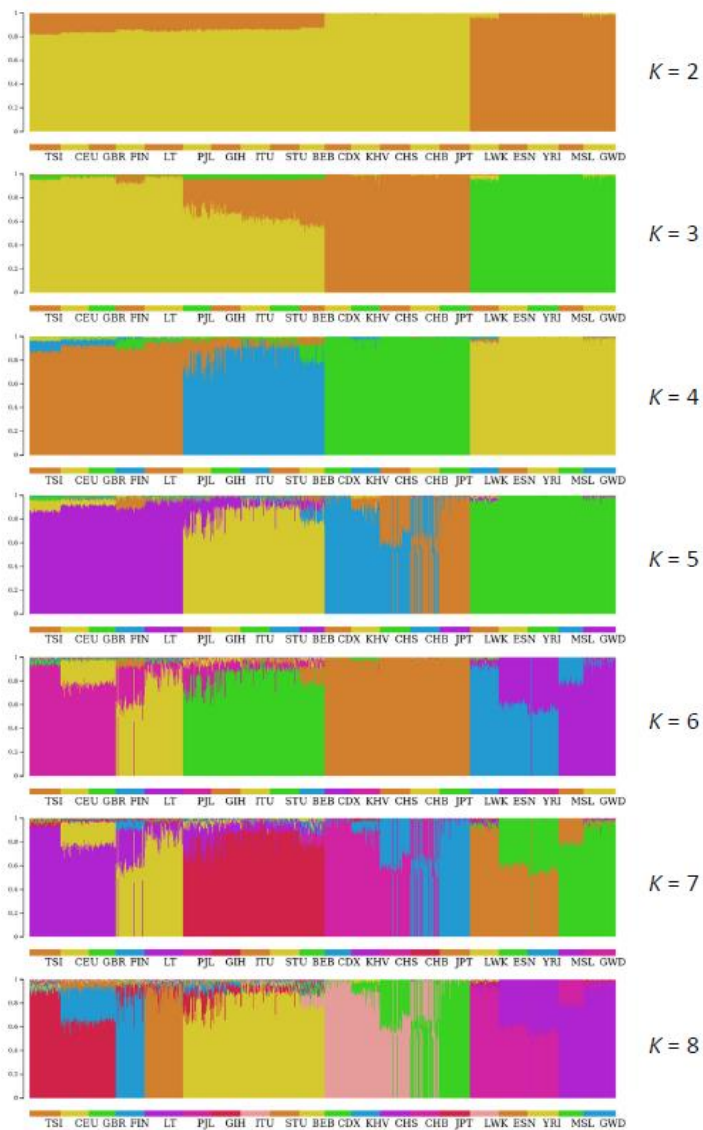
*minibd* – IBD segmentų, trumpesnių nei *minibd* (mūsų atveju 7), pašalinimas

*out* – išvesties failo pavadinimas

#### IV PRIEDAS. Papildoma rezultatų medžiaga



**1 pav.** PKA, panaudojant 100 atsitiktiniu būdu atrinktų Lietuvos populiacijos individų ir keturių 1 000 genomų populiacijų (GBR, FIN, TSI ir CEU) VNP duomenis



**2 pav.** 120 atsitiktiniu būdu atrinktų Lietuvos populiacijos asmenų ir 1 000 genomų projekto 19 populiacijų priemaišos analizė, kai  $K$  yra nuo 2 iki 8

**1 lentelė.** Kandidatiniai netolimos teigiamos gamtinės atrankos veikiami VNP, nustatyti XP-EHH ir  $F_{ST}$  metodais Lietuvos populiacijoje

CHR	POS	SNP	POP	Func, refGene	Gene, refGene	GeneDetail,refGene	EgzoninėFunc, refGene	$F_{ST}$	CADD13_Raw Score	CADD13_PHRKD
1	35484861	rs11263938	LT-YRI	introninė	ZMYM6	NA	NA	0,66	0,062288	3 210
1	35628789	rs6425939	LT-YRI	intergeninė	ZMYM1;SFPQ	dist=47330;dist=20412	NA	0,27	0,705777	8 865
1	35635453	rs12064550	LT-YRI	intergeninė	ZMYM1;SFPQ	dist=53994;dist=13748	NA	0,33	-0,045246	2 172
1	36549157	rs3754076	LT-YRI	upstream	TEKT2	dist=519	NA	0,51	0,233075	5 024
1	36561956	rs274750	LT-YRI	UTR3	COL8A2	NM_005202;c.*1214G>T;NM_001294347;c.*1214G>T	NA	0,67	0,571959	7 960
1	36562551	rs274752	LT-YRI	UTR3	COL8A2	NM_005202;c.*619T>C;NM_001294347;c.*619T>C	NA	0,67	-0,139848	1 452
1	231866461	rs2793102	LT-CEU	ncRNA_introninė	TSN4X-DISCI	,	,	0,09	0,154628	4 195
1	231873198	rs12027635	LT-CEU	ncRNA_introninė	TSN4X-DISCI	,	,	0,00	0,324494	5 928
1	231908347	rs7534681	LT-CEU	ncRNA_introninė	TSN4X-DISCI	,	,	0,00	0,082856	3 425
1	245363423	rs6428904	LT-CEU	introninė	KIF26B	NA	NA	0,00	-0,009549	2 495
1	246303128	rs7551290	LT-CEU	introninė	SMYD3	NA	NA	0,01	0,123673	3 862
1	247400259	rs10924968	LT-CEU	intergeninė	MIR3916;NIR5	dist=34897;dist=19115	NA	0,02	0,641084	8 437
1	247401645	rs2793289	LT-CEU	intergeninė	MIR3916;NIR5	dist=36283;dist=17729	NA	0,01	0,302216	5 715
2	56070352	rs934278	LT-FIN	intergeninė	PNPT1;EFEMP1	dist=149307;dist=22745	,	,	0,188001	4 552
2	56118046	rs10865291	LT-FIN	introninė	EFEMP1	,	,	0,00	-0,396913	393
2	56130640	rs1430195	LT-FIN	introninė	EFEMP1	,	,	0,00	0,101362	3 622
2	56161538	rs7608756	LT-FIN	intergeninė	EFEMP1;MIR217HG	dist=10240;dist=29003	,	0,00	0,137864	4 015
3	2347409	rs12632736	LT-FIN	introninė	CNTN4	NA	NA	0,00	-0,140683	1 447
3	2347522	rs1626056	LT-FIN	introninė	CNTN4	NA	NA	0,07	0,349619	6 163

CHR	POS	SNP	POP	Func.refGene	Gene.refGene	GeneDetail.refGene	EggnineFunc.refGene	F <sub>ST</sub>	CADD13_Raw Score	CADD13_PHRD
3	2351848	rs1148053	LT-FIN	introninè	CNTN4	NA	NA	0,08	0,194951	4 626
3	2354035	rs17194266	LT-FIN	introninè	CNTN4	NA	NA	0,00	0,486751	7 329
3	2355167	rs11129096	LT-FIN	introninè	CNTN4	NA	NA	0,13	0,029484	2 875
3	112511598	rs9862107	LT-FIN	intergeninè	LOC101929694;CD200R1L	dist=43432;dist=22958	NA	0,01	-0,269640	777
3	112796030	rs1795172	LT-FIN	intergeninè	C3orf17;LOC101929717	dist=57475;dist=65167	NA	0,06	1,194553	11,72
3	112797260	rs6784662	LT-FIN	intergeninè	C3orf17;LOC101929717	dist=58705;dist=63937	NA	0,06	-0,105878	1 687
3	112811428	rs7625411	LT-FIN	intergeninè	C3orf17;LOC101929717	dist=72873;dist=49769	NA	0,13	0,025052	2 830
3	112816192	rs9816697	LT-FIN	intergeninè	C3orf17;LOC101929717	dist=77637;dist=45005	NA	0,00	0,173676	4 400
3	130243249	rs4682614	LT-FIN	intergeninè	COL6A5;COL6A6	dist=39559;dist=35929	NA	0,03	-0,439187	310
3	131445951	rs17294121	LT-FIN	introninè	CPNE4	NA	NA	0,09	-0,064756	2 007
4	106552414	rs2044536	LT-YRI/FIN-YRI	UTR3	ARHGEF38	NM_017700:c.*208A>G	NA	0,56/0,50	0,904780	10,10
4	106593850	rs4235415	LT-YRI/FIN-YRI	introninè	ARHGEF38	NA	NA	0,68/0,66	-0,159845	1 326
4	106594625	rs11930965	LT-YRI/FIN-YRI	introninè	ARHGEF38	NA	NA	0,45/0,28	-0,048408	2 145
4	89629000	rs1063143	LT-FIN	UTR3	HERC3	NM_001271602:c.*889A>G;NM_014606:c.*889A>G	NA	0,01	-0,112530	1 639
4	89653257	rs1801915	LT-FIN	eggninè	FAM134	NA	synonymous SNV	0,09	-0,023946	2 362
4	89665193	rs3775391	LT-FIN	introninè	FAM134	NA	NA	0	-0,451950	289
4	89669852	rs11944668	LT-FIN	introninè	FAM134	NA	NA	0,05	0,807047	9 507
4	181637915	rs2545308	LT-FIN	intergeninè	NONE;LINC00290	dist=NONE;dist=347328	NA	0,04	0,124919	3 875

CHR	POS	SNP	POP	Func_refGene	Gene_refGene	GeneDetail_refGene	EggnineFunc_refGene	F <sub>ST</sub>	CADD13_Raw Score	CADD13_PHRD
4	181638467	rs6552438	LT-FIN	intergenine	NONE;LINC00290	dist=NONE;dist=346776	NA	0,04	1,753089	14,71
4	181682943	rs2613007	LT-FIN	intergenine	NONE;LINC00290	dist=NONE;dist=302300	NA	0,18	0,074804	3 341
4	181691738	rs918402	LT-FIN	intergenine	NONE;LINC00290	dist=NONE;dist=293505	NA	0,03	-0,128885	1 525
5	10073956	rs11742768	LT-CEU	intergenine	LOC285692;FAM173B	dist=170020;dist=151664	,	0,08	-0,176106	1 230
5	10085900	rs924968	LT-CEU	intergenine	LOC285692;FAM173B	dist=181964;dist=139720	,	0,03	0,023370	2 814
5	10108500	rs6886410	LT-CEU	intergenine	LOC285692;FAM173B	dist=204564;dist=117120	,	0,05	0,136078	3 996
5	10109791	rs884355	LT-CEU	intergenine	LOC285692;FAM173B	dist=205855;dist=115829	,	0,05	-0,017553	2 421
5	133177588	rs2457175	LT-FIN	intergenine	FSTL4;WSPAR	dist=229365;dist=71780	NA	0,13	0,136829	4 004
5	133180735	rs1703425	LT-FIN	intergenine	FSTL4;WSPAR	dist=232512;dist=68633	NA	0,11	0,498067	7 416
5	133183441	rs1681127	LT-FIN	intergenine	FSTL4;WSPAR	dist=235218;dist=65927	NA	0,12	1,141420	11,44
5	133638034	rs17165388	LT-FIN	intronin	CDKL3	NA	NA	0,05	0,076163	3 355
6	11710405	rs9348953	LT-FIN	intergenine	TMEM170B;ADTRP	dist=126648;dist=3483	NA	0,06	0,839141	9 703
6	11711968	rs210917	LT-FIN	intergenine	TMEM170B;ADTRP	dist=128211;dist=1920	NA	0,02	0,137101	4 007
6	11716759	rs1413346	LT-FIN	intronin	ADTRP	NA	NA	0,00	-0,034199	2 270
6	11717934	rs6933892	LT-FIN	intronin	ADTRP	NA	NA	0,10	0,134851	3 982
6	27811815	rs200997	LT-FIN	intergenine	HIST1H2BN;HIST1H2AL	dist=4927;dist=21292	NA	0,00	0,078818	3 383
6	27815494	rs200991	LT-FIN	intergenine	HIST1H2BN;HIST1H2AL	dist=8606;dist=17613	NA	0,02	0,542995	7 752
6	27815639	rs13194781	LT-FIN	intergenine	HIST1H2BN;HIST1H2AL	dist=8751;dist=17468	NA	0,00	0,203412	4 715
6	27816442	rs200989	LT-FIN	intergenine	HIST1H2BN;HIST1H2AL	dist=9554;dist=16665	NA	0,00	-0,131371	1 508



CHR	POS	SNP	POP	Func.refGene	Gene.refGene	GeneDetail.refGene	EggonineFunc.refGene	F <sub>ST</sub>	CADD13_Raw Score	CADD13_PHRD
					2AL					
6	27828667	rs9295751	LT-FIN	intergenic	HIST1H2BN; HIST1H2AL	dist=21779;dist=4440	NA	0,01	-0,067516	1 984
6	27834085	rs13199772	LT-FIN	downstream	HIST1H1B; HIST1H2AL	dist=485	NA	0,01	0,013734	2 719
6	27835930	rs200951	LT-FIN	upstream	HIST1H1B	dist=571	NA	0,04	0,264216	5 341
6	27849676	rs573179	LT-FIN	intergenic	HIST1H4L; HIST1H3J	dist=8387;dist=8417	NA	0,03	-0,143285	1 430
6	27852357	rs200979	LT-FIN	intergenic	HIST1H4L; HIST1H3J	dist=11068;dist=5736	NA	0,01	0,205590	4 738
6	27870772	rs9468252	LT-FIN	intergenic	HIST1H2BO; OR2B2	dist=9103;dist=8191	NA	0,01	-0,439180	310
6	27883201	rs1497526	LT-FIN	intergenic	OR2B2; OR2B6	dist=3027;dist=41818	NA	0,03	-0,081076	1 875
6	27887275	rs1015075	LT-FIN	intergenic	OR2B2; OR2B6	dist=7101;dist=37744	NA	0,01	-0,53209	180
6	27890631	rs6933825	LT-FIN	intergenic	OR2B2; OR2B6	dist=10457;dist=34388	NA	0,01	0,180975	4 478
6	27895213	rs156737	LT-FIN	intergenic	OR2B2; OR2B6	dist=15039;dist=29806	NA	0,04	0,595882	8 128
6	27895504	rs9468256	LT-FIN	intergenic	OR2B2; OR2B6	dist=15330;dist=29515	NA	0,01	0,595832	8 127
6	27901041	rs7760871	LT-FIN	intergenic	OR2B2; OR2B6	dist=20867;dist=23978	NA	0,02	1,756572	14,73
6	27903467	rs10456357	LT-FIN	intergenic	OR2B2; OR2B6	dist=23293;dist=21552	NA	0,14	-0,030068	2 307
6	27919688	rs9461425	LT-FIN	intergenic	OR2B2; OR2B6	dist=39514;dist=5331	NA	0,04	-0,343913	526
6	27925367	rs7767176	LT-FIN	eggonine	OR2B6	NA	nonsynonymous SNV	0; (0,12 CEU-YRI; 0,13 FIN-YRI; 0,0 CEU-FIN)	0,650666	8 501
6	27925827	rs9380030	LT-FIN	eggonine	OR2B6	NA	nonsynonymous SNV	0,00591924; (0,02 CEU-YRI; (0,03 FIN-YRI; (0,00010 CEU-FIN)	-0,893801	27

CHR	POS	SNP	POP	Func.refGene	Gene.refGene	GeneDetail.refGene	EggnineFunc.refGene	F <sub>ST</sub>	CADD13_Raw Score	CADD13_PhiRED
6	27928445	rs9295753	L-T-FIN	intergenine	OR2B6;ZNF165	dist=2485;dist=120037	NA	0,01	-0,151352	1 378
6	27936049	rs6922530	L-T-FIN	intergenine	OR2B6;ZNF165	dist=10089;dist=112433	NA	0,00	0,056192	3 146
6	27979625	rs10484402	L-T-FIN	intergenine	OR2B6;ZNF165	dist=53665;dist=68857	NA	0,02	0,309259	5 783
6	27981673	rs9368540	L-T-FIN	intergenine	OR2B6;ZNF165	dist=55713;dist=66809	NA	0,01	-0,618892	112
6	27986529	rs9368544	L-T-FIN	intergenine	OR2B6;ZNF165	dist=60569;dist=61953	NA	0,01	0,043620	3 017
6	27987776	rs149976	L-T-FIN	intergenine	OR2B6;ZNF165	dist=61816;dist=60706	NA	0,04	-0,146543	1 409
6	27998258	rs149990	L-T-FIN	intergenine	OR2B6;ZNF165	dist=72298;dist=50224	NA	0,00	-0,320872	595
6	27999044	rs9368545	L-T-FIN	intergenine	OR2B6;ZNF165	dist=73084;dist=49438	NA	0,83	0,931008	10,26
6	28002388	rs149943	L-T-FIN	intergenine	OR2B6;ZNF165	dist=76428;dist=46094	NA	0,00	0,171634	4 378
6	28018944	rs9393879	L-T-FIN	intergenine	OR2B6;ZNF165	dist=92984;dist=29538	NA	0,01	0,164865	4 305
6	28021589	rs203888	L-T-FIN	intergenine	OR2B6;ZNF165	dist=95629;dist=26893	NA	0,02	2,125758	17,02
6	28028390	rs9468277	L-T-FIN	intergenine	OR2B6;ZNF165	dist=102430;dist=20092	NA	0,03	0,250236	5 199
6	28033505	rs7741570	L-T-FIN	intergenine	OR2B6;ZNF165	dist=107545;dist=14977	NA	0,00	0,497524	7 412
6	28033523	rs10484403	L-T-FIN	intergenine	OR2B6;ZNF165	dist=107563;dist=14959	NA	0,03	0,695393	8 797
6	28039586	rs4713135	L-T-FIN	intergenine	OR2B6;ZNF165	dist=113626;dist=8896	NA	0,03	-0,263323	802
6	28046478	rs172164	L-T-FIN	intergenine	OR2B6;ZNF165	dist=120518;dist=2004	NA	0,00	-0,184676	1 181
6	28058635	rs1853097	L-T-FIN	ncRNA_egzomine	ZSCAN12P1	NA	NA	0,03	0,281140	5 509
6	28062066	rs203893	L-T-FIN	ncRNA_egzomine	ZSCAN12P1	NA	NA	0,04	0,007773	2 661
6	28064736	rs169433	L-T-FIN	intergenine	ZSCAN12P1;ZSCAN16-AS1	dist=1243;dist=24837	NA	0,00	1,583581	13,76
6	28077374	rs203884	L-T-FIN	intergenine	ZSCAN12P1;ZSCAN16-AS1	dist=13881;dist=12199	NA	0,04	0,048213	3 064
6	28083521	rs4713137	L-T-FIN	intergenine	ZSCAN12P1;ZSCAN16-AS1	dist=20028;dist=6052	NA	0,69	-0,439930	309

CHR	POS	SNP	POP	Func.refGene	Gene.refGene	GeneDetail.refGene	EggonineFunc.refGene	F <sub>ST</sub>	CADD13_Raw Score	CADD13_PHRED
6	28096280	rs1654774	LT-FIN	ncRNA_intronic	ZSCAN16-AS1	NA	NA	0,15	0,018535	2 766
6	46596286	rs3799866	LT-CEU	introninè	CYP39A1	,	,	0,04	0,913557	10,16
6	46599187	rs9472810	LT-CEU	introninè	CYP39A1	,	,	0,00	-0,260305	815
6	46603687	rs6924295	LT-CEU	introninè	CYP39A1	,	,	0,02	0,107860	3 692
6	46703513	rs1421378	LT-CEU	upstream	PLA2G7	dist=83	,	0,01	0,687211	8 743
6	46767645	rs3799859	LT-CEU	introninè	MEP1A	,	,	0,08	-0,178329	1 217
6	46767688	rs3799858	LT-CEU	introninè	MEP1A	,	,	0,01	0,187896	4 551
6	46832813	rs614826	LT-CEU	eggoninè	ADGRF5	,	synonymous SNV	0,04	1,035610	10,86
6	46834685	rs586024	LT-CEU	eggoninè	ADGRF5	,	nonsynonymous SNV	0; (0,3 CEU-YRI; 0,32 FIN-YRI; 0 CEU-FIN)	0,413035	6 727
7	30276161	rs6462196	LT-YRI	intergeninè	MTURN:ZNRIF2	dist=73780;dist=47762	NA	0,63	-0,034517	2 267
7	30299645	rs6462197	LT-YRI	intergeninè	MTURN:ZNRIF2	dist=97264;dist=24278	NA	0,66	0,395334	6 573
7	30370786	rs11974360	LT-YRI	introninè	ZNRIF2	NA	NA	0,53	-0,209407	1 049
7	30407445	rs42597	LT-YRI	downstream	ZNRIF2	dist=137	NA	0,08	-0,620893	111
7	30415102	rs17158939	LT-YRI	intergeninè	DKFZP586I1420;LINCOL1176	dist=2692;dist=15399	NA	0,40	0,065273	3 241
7	112994204	rs2692198	LT-FIN	intergeninè	LINC00998;PPP1R3	dist=235567;dist=52267	NA	0,18	-0,257565	826
7	113008823	rs1701843	LT-FIN	intergeninè	LINC00998;PPP1R3	dist=250186;dist=50805	NA	0,11	0,007092	2 654
7	113017399	rs7784560	LT-FIN	intergeninè	LINC00998;PPP1R3	dist=258762;dist=49948	NA	0,00	0,121539	3 839
7	113020044	rs11977351	LT-FIN	intergeninè	LINC00998;PPP1R3	dist=261407;dist=49683	NA	0,18	0,687103	8 742

CHR	POS	SNP	POP	Func.refGene	Gene.refGene	GeneDetail.refGene	EggnineFunc.refGene	F <sub>ST</sub>	CADD13_Raw Score	CADD13_PhiRED
7	129187430	rs2598179	LT-CEU	intergeninè	SMKRI;NRF1	dist=34657;dist=64113	,	0,00	1,126521	11,36
7	129196930	rs959444	LT-CEU	intergeninè	SMKRI;NRF1	dist=44157;dist=54613	,	0,00	-0,386348	417
7	129199328	rs10245560	LT-CEU	intergeninè	SMKRI;NRF1	dist=46555;dist=52215	,	0,06	-0,147858	1 400
7	129199487	rs11765308	LT-CEU	intergeninè	SMKRI;NRF1	dist=46714;dist=52056	,	0,08	0,088395	3 484
7	129210535	rs692489	LT-CEU	intergeninè	SMKRI;NRF1	dist=57762;dist=41008	,	0,04	0,800801	9 469
7	129214174	rs754386	LT-CEU	intergeninè	SMKRI;NRF1	dist=61401;dist=37369	,	0,00	-0,856385	32
7	129214387	rs754385	LT-CEU	intergeninè	SMKRI;NRF1	dist=61614;dist=37156	,	0,04	0,051625	3 099
7	129333905	rs9656383	LT-CEU	introninè	NRF1	,	,	0,04	0,323737	5 921
7	129342404	rs2402976	LT-CEU	introninè	NRF1	,	,	0,04	0,207353	4 756
7	129343967	rs11773023	LT-CEU	introninè	NRF1	,	,	0,07	0,221382	4 903
7	129350170	rs3736626	LT-CEU	introninè	NRF1	,	,	0,07	0,054016	3 124
9	12352971	rs10960616	LT-CEU	intergeninè	PTPRD-AS2;TYRPI	dist=1732551;dist=340415	,	0,02	-0,308530	635
9	12356017	rs1331350	LT-CEU	intergeninè	PTPRD-AS2;TYRPI	dist=1735597;dist=337369	,	0,02	-0,075974	1 916
9	12359563	rs4628342	LT-CEU	intergeninè	PTPRD-AS2;TYRPI	dist=1739143;dist=333823	,	0,08	0,039525	2 976
9	12382539	rs1952692	LT-CEU	intergeninè	PTPRD-AS2;TYRPI	dist=1762119;dist=310847	,	0,00	-0,175380	1 234
9	12387932	rs7469693	LT-CEU	intergeninè	PTPRD-AS2;TYRPI	dist=1767512;dist=305454	,	0,02	0,012560	2 707
9	12401666	rs10756369	LT-CEU	intergeninè	PTPRD-AS2;TYRPI	dist=1781246;dist=291720	,	0,02	-0,232582	936
9	12421929	rs12683218	LT-CEU	intergeninè	PTPRD-AS2;TYRPI	dist=1801509;dist=271457	,	0,00	-0,108137	1 670
9	12425767	rs10491747	LT-CEU	intergeninè	PTPRD-AS2;TYRPI	dist=1805347;dist=267619	,	0,00	-0,312610	622

CHR	POS	SNP	POP	Func.refGene	Gene.refGene	GeneDetail.refGene	EggnineFunc.refGene	F <sub>ST</sub>	CADD13_Raw Score	CADD13_PHRED
9	12443586	rs7860071	LT-CEU	intergenine	<i>PTPRD-AS2;TYRPI</i>	dist=1823166;dist=249800	,	0,01	-0,014792	2 446
9	12445143	rs1537181	LT-CEU	intergenine	<i>PTPRD-AS2;TYRPI</i>	dist=1824723;dist=248243	,	0,06	0,091601	3 518
9	12453240	rs4740523	LT-CEU	intergenine	<i>PTPRD-AS2;TYRPI</i>	dist=1832820;dist=240146	,	0,00	0,108702	3 701
9	12460256	rs4008031	LT-CEU	intergenine	<i>PTPRD-AS2;TYRPI</i>	dist=1839836;dist=233130	,	0,07	-0,221460	989
9	12463411	rs2026326	LT-CEU	intergenine	<i>PTPRD-AS2;TYRPI</i>	dist=1842991;dist=229975	,	0,00	-0,264885	796
9	12464876	rs2150696	LT-CEU	intergenine	<i>PTPRD-AS2;TYRPI</i>	dist=1844456;dist=228510	,	0,03	0,204795	4 730
9	12483221	rs1576659	LT-CEU	intergenine	<i>PTPRD-AS2;TYRPI</i>	dist=1862801;dist=210165	,	0,01	1,743048	14,65
9	12490247	rs791654	LT-CEU	intergenine	<i>PTPRD-AS2;TYRPI</i>	dist=1869827;dist=203139	,	0,00	-0,011758	2 475
9	12496752	rs2022314	LT-CEU	intergenine	<i>PTPRD-AS2;TYRPI</i>	dist=1876332;dist=196634	,	0,00	0,385907	6 490
9	12497036	rs7467390	LT-CEU	intergenine	<i>PTPRD-AS2;TYRPI</i>	dist=1876616;dist=196350	,	0,00	0,332357	6 002
9	12501498	rs791662	LT-CEU	intergenine	<i>PTPRD-AS2;TYRPI</i>	dist=1881078;dist=191888	,	0,03	-0,013432	2 459
9	12509667	rs16929075	LT-CEU	intergenine	<i>PTPRD-AS2;TYRPI</i>	dist=1889247;dist=183719	,	0,02	-0,210935	1 041
9	12510515	rs791667	LT-CEU	intergenine	<i>PTPRD-AS2;TYRPI</i>	dist=1890095;dist=182871	,	0,01	-0,108306	1 669
9	12513587	rs1041102	LT-CEU	intergenine	<i>PTPRD-AS2;TYRPI</i>	dist=1893167;dist=179799	,	0,02	0,412045	6 718
9	12515876	rs791669	LT-CEU	intergenine	<i>PTPRD-AS2;TYRPI</i>	dist=1895456;dist=177510	,	0,00	-0,458486	278
9	12522037	rs791681	LT-CEU	intergenine	<i>PTPRD-AS2;TYRPI</i>	dist=1901617;dist=171349	,	0,03	-0,100320	1 728
9	12525999	rs1325132	LT-CEU	intergenine	<i>PTPRD-AS2;TYRPI</i>	dist=1905579;dist=167387	,	0,00	-0,032816	2 282

CHR	POS	SNP	POP	Func.refGene	Gene.refGene	GeneDetail.refGene	EggnineFunc.refGene	F <sub>ST</sub>	CADD13_Raw Score	CADD13_PhiRED
9	12526192	rs702131	LT-CEU	intergeninè	<i>PTPRD-AS2;TYRPI</i>	dist=1905772;dist=167194	,	0,04	-0,103920	1 701
9	12529952	rs791695	LT-CEU	intergeninè	<i>PTPRD-AS2;TYRPI</i>	dist=1909532;dist=163434	,	0,00	-0,251223	853
9	12532584	rs791700	LT-CEU	intergeninè	<i>PTPRD-AS2;TYRPI</i>	dist=1912164;dist=160802	,	0,02	-0,023906	2 362
9	12537279	rs1570781	LT-CEU	intergeninè	<i>PTPRD-AS2;TYRPI</i>	dist=1916859;dist=156107	,	0,03	0,045809	3 040
10	116362835	rs7094673	LT-FIN	introninè	<i>ABLIMI</i>	NA	NA	0,06	-0,421969	342
10	116364311	rs10787539	LT-FIN	introninè	<i>ABLIMI</i>	NA	NA	0,04	-0,527961	187
10	116364424	rs10787540	LT-FIN	introninè	<i>ABLIMI</i>	NA	NA	0,04	-0,075708	1 918
10	116377962	rs2483603	LT-FIN	introninè	<i>ABLIMI</i>	NA	NA	0,10	-0,074559	1 927
11	129832270	rs7949651	LT-YRI/CEU-YRI	introninè	<i>PRDM10</i>	NA	NA	0,76/0,69	-0,309363	633
11	129843578	rs4937474	LT-YRI/CEU-YRI	introninè	<i>PRDM10</i>	NA	NA	0,51/0,36	0,215892	4 846
11	129860701	rs3740879	LT-YRI/CEU-YRI	introninè	<i>PRDM10</i>	NA	NA	0,41/0,28	-0,260621	814
11	129885191	rs2875185	LT-YRI/CEU-YRI	intergeninè	<i>LINC00167;APLP2</i>	dist=9810;dist=54525	NA	0,51/0,37	-0,179179	1 212
11	129910584	rs3808986	LT-YRI/CEU-YRI	intergeninè	<i>LINC00167;APLP2</i>	dist=35203;dist=29132	NA	0,60/0,38	1,042148	10,90
11	129979745	rs3819106	LT-YRI/CEU-YRI	introninè	<i>APLP2</i>	NA	NA	0,54/0,47	0,135147	3 986
11	129983282	rs11221970	LT-YRI/CEU-YRI	introninè	<i>APLP2</i>	NA	NA	0,30/0,16	0,481601	7 289
11	130639997	rs12287063	LT-YRI/CEU-YRI	intergeninè	<i>C1orf44;LOC100507431</i>	dist=52750;dist=74091	NA	0,29/0,17	-0,027046	2 334
11	130642904	rs4937565	LT-YRI/CEU-YRI	intergeninè	<i>C1orf44;LOC100507431</i>	dist=55657;dist=71184	NA	0,52/0,37	0,412222	6 720
11	130645506	rs6590507	LT-YRI/CEU-YRI	intergeninè	<i>C1orf44;LOC100507431</i>	dist=58259;dist=68582	NA	0,44/0,26	-0,206613	1 063

CHR	POS	SNP	POP	Func.refGene	Gene.refGene	GeneDetail.refGene	EggnineFunc.refGene	F <sub>ST</sub>	CADD13_Raw Score	CADD13_PHRD
			YRI		7431					
11	41897826	rs4756680	L-T-YRI	intergenine	LINC01499;LOC100507205	dist=39834;dist=311467	NA	0,56	-0,531021	184
11	41898126	rs12793795	L-T-YRI	intergenine	LINC01499;LOC100507205	dist=40134;dist=311167	NA	0,00	0,302380	5 717
11	41899112	rs7122740	L-T-YRI	intergenine	LINC01499;LOC100507205	dist=41120;dist=310181	NA	0,38	-0,116740	1 609
11	41915366	rs9300039	L-T-YRI	intergenine	LINC01499;LOC100507205	dist=57374;dist=293927	NA	0,09	-0,048499	2 144
11	41926369	rs2019626	L-T-YRI	intergenine	LINC01499;LOC100507205	dist=68377;dist=282924	NA	0,70	-0,019665	2 401
11	41968107	rs6485298	L-T-YRI	intergenine	LINC01499;LOC100507205	dist=110115;dist=241186	NA	0,37	0,070111	3 291
11	41978556	rs2672546	L-T-YRI	intergenine	LINC01499;LOC100507205	dist=120564;dist=230737	NA	0,11	-0,665929	87
11	42008948	rs11036649	L-T-YRI	intergenine	LINC01499;LOC100507205	dist=150956;dist=200345	NA	0,20	-0,078942	1 892
11	42012945	rs7127839	L-T-YRI	intergenine	LINC01499;LOC100507205	dist=154953;dist=196348	NA	0,33	0,311960	5 809
11	42307166	rs11036800	L-T-YRI	intergenine	LOC100507205;HNRNP3	dist=31926;dist=975888	NA	0,69	2,089550	16,79
12	68640583	rs1179246	L-T-CEU	intergenine	IL26;IL22	dist=21012;dist=1442		0,02	-0,368815	459
12	68641532	rs1182844	L-T-CEU	downstream	IL22	dist=493		0,33	0,200088	4 680
12	68647713	rs2227485	L-T-CEU	upstream	IL22	dist=432		0,02	1,458819	13,10
12	68654392	rs11177140	L-T-CEU	intergenine	IL22;MDMI	dist=7111;dist=33954		0	-0,175520	1 233
12	68654642	rs12100005	L-T-CEU	intergenine	IL22;MDMI	dist=7361;dist=33704		0,01	-0,412378	361
12	68654780	rs7954576	L-T-CEU	intergenine	IL22;MDMI	dist=7499;dist=33566		0,03	-0,219439	999
12	103326468	rs7973675	L-T-CEU	intergenine	PAH;ASCL1	dist=15087;dist=24984		0,02	-0,361867	477
12	103361379	rs17450273	L-T-CEU	intergenine	ASCL1;LOC1019290	dist=7085;dist=184241		0,02	0,031209	2 892

CHR	POS	SNP	POP	Func.refGene	Gene.refGene	GeneDetail.refGene	EggnineFunc.refGene	F <sub>ST</sub>	CADD13_Raw Score	CADD13_PHRD
					58					
12	103363796	rs4764926	LT-CEU	intergenine	ASCL1;LOC101929058	dist=9502;dist=181824	,	0,06	0,013212	2 714
12	103370585	rs4764719	LT-CEU	intergenine	ASCL1;LOC101929058	dist=16291;dist=175035	,	0,09	2,408538	18,88
12	103394540	rs2373952	LT-CEU	intergenine	ASCL1;LOC101929058	dist=40246;dist=151080	,	0,12	-0,008197	2 508
15	48389924	rs12440301	LT-YRI/CEU-YRI	intergenine	LINC01491;SLC24A5	dist=251491;dist=23245	NA	0,66/0,47	-0,123330	1 563
15	48523284	rs16960698	LT-YRI/CEU-YRI	intronine	SLC12A1	NA	NA	0,20/0,22	-0,175413	1 234
15	48526271	rs1531916	LT-YRI/CEU-YRI	intronine	SLC12A1	NA	NA	0,25/0,23	0,117648	3 797
15	48542027	rs2279366	LT-YRI/CEU-YRI	intronine	SLC12A1	NA	NA	0,39/0,35	0,577566	7 999
15	48670904	rs959324	LT-YRI/CEU-YRI	intergenine	DUT;FBN1	dist=35334;dist=29599	NA	0,37/0,25	0,085642	3 455
15	48675227	rs919129	LT-YRI/CEU-YRI	intergenine	DUT;FBN1	dist=39657;dist=25276	NA	0,32/0,32	0,270552	5 404
15	48683235	rs8025596	LT-YRI/CEU-YRI	intergenine	DUT;FBN1	dist=47665;dist=17268	NA	0,20/0,23	0,071525	3 306
15	48686175	rs10519170	LT-YRI/CEU-YRI	intergenine	DUT;FBN1	dist=50605;dist=14328	NA	0,15/0,18	-0,283583	723
15	51639821	rs2124872	LT-CEU	intronine	GLDN	,	,	0,04	0,081273	3 409
15	51640359	rs2445771	LT-CEU	intronine	GLDN	,	,	0,00	-0,125152	1 550
15	51644775	rs7163193	LT-CEU	intronine	GLDN	,	,	0,01	0,200521	4 685
15	51648534	rs2124873	LT-CEU	intronine	GLDN	,	,	0,06	-0,127994	1 531
15	51658293	rs4774588	LT-CEU	intronine	GLDN	,	,	0,05	-0,116042	1 614
15	51663386	rs17524318	LT-CEU	intronine	GLDN	,	,	0,02	-0,195115	1 124



CHR	POS	SNP	POP	Func, refGene	Gene, refGene	GeneDetail, refGene	EgzonInFunc, refGene	F <sub>ST</sub>	CADD13_Raw Score	CADD13_PHRED
15	51663597	rs4775942	LT-CEU	introninè	GLDN			0,07	0,062680	3 214
15	51669741	rs16964316	LT-CEU	introninè	GLDN			0,00	0,278370	5 482
15	51673630	rs10163098	LT-CEU	introninè	GLDN			0,00	-0,079687	1 886
15	58579956	rs2899624	LT-CEU	intergeninè	AQP9;LIPC	dist=101846;dist=144219		0,08	-0,130975	1 511
15	58580525	rs12910051	LT-CEU	intergeninè	AQP9;LIPC	dist=102415;dist=143650		0,03	-0,066702	1 991
15	58587409	rs16940021	LT-CEU	intergeninè	AQP9;LIPC	dist=109299;dist=136766		0,06	0,093928	3 543
15	58588116	rs4775028	LT-CEU	intergeninè	AQP9;LIPC	dist=110006;dist=136059		0,07	-0,624892	109
16	66962411	rs7194513	LT-FIN	intergeninè	RRAD;FAM96B	dist=2972;dist=3547	NA	0,08	-0,466429	266
16	66970561	rs11568314	LT-FIN	introninè	CES2	NA	NA	0,06	0,294789	5 643
16	66973381	rs8045523	LT-FIN	introninè	CES2	NA	NA	0,08	-0,246258	875
16	66977094	rs2241409	LT-FIN	introninè	CES2	NA	NA	0,07	0,165079	4 308
16	66977997	rs8192925	LT-FIN	UTR3	CES2	NM_003869.c.*69A>G; NM_198061.c.*69A>G	NA	0,00	0,217207	4 860
16	66999370	rs13336470	LT-FIN	introninè	CES3	NA	NA	0,09	0,119375	3 816
16	67028008	rs28571100	LT-FIN	introninè	CES4A	NA	NA	0,07	-0,423554	339
17	4722606	rs3764899	LT-FIN	introninè	PLD2	NA	NA	0,03	-0,064294	2 011
17	4742112	rs12450045	LT-FIN	introninè	MINK1	NA	NA	0,10	0,149280	4 138
17	4752260	rs8072531	LT-FIN	introninè	MINK1	NA	NA	0,08	0,329136	5 972
17	4755424	rs9904177	LT-FIN	introninè	MINK1	NA	NA	0,01	1,385614	12,71
17	4797305	rs2302319	LT-FIN	egzoninè	MINK1	NA	nonsyno- nymous SNV	0,0339696 (0,08 CEU- YRI; 0,12 FIN- YRI; 0,0009	2,623577	20,3

CHR	POS	SNP	POP	Func.refGene	Gene.refGene	GeneDetail.refGene	EggnineFunc.refGene	F <sub>ST</sub>	CADD13_Raw Score	CADD13_PHRD
17	4801288	rs1053754	L-T-FIN	UTR3	<i>CHRNAE;MINK1</i>	NM_000080:c.*743T>G;NM_015716:c.*706A>C;NM_153827:c.*706A>C;NM_170663:c.*706A>C;NM_001024937:c.*706A>C	NA	0,04	0,023832	2 818
17	4807926	rs3760490	L-T-FIN	intergeninè	<i>CHRNAE;GPI1B4</i>	dist=1557;dist=27644	NA	0,01	-0,518858	197
17	4812470	rs9905341	L-T-FIN	intergeninè	<i>CHRNAE;GPI1B4</i>	dist=6101;dist=23100	NA	0,12	0,379463	6 433
17	4831790	rs9914087	L-T-FIN	intergeninè	<i>CHRNAE;GPI1B4</i>	dist=25421;dist=3780	NA	0,06	0,132379	3 956
17	4839930	rs400688	L-T-FIN	downstream	<i>SLC25A11</i>	dist=496	NA	0,03	-0,543903	171
17	4845768	rs238247	L-T-FIN	introninè	<i>RNF167</i>	NA	NA	0,09	1,317549	12,36
17	4846045	rs11871232	L-T-FIN	introninè	<i>RNF167</i>	NA	NA	0,00	-0,479924	246
17	4856376	rs238238	L-T-FIN	egzoninè	<i>ENO3</i>	NA	non-synonymous SNV	0,0506347 (0,06 CEU-YRI; 0 FIN-YRI; 0,04 CEU-FIN)	1,486882	13,24
17	4863410	rs238224	L-T-FIN	introninè	<i>SPAG7</i>	NA	NA	0,03	-0,381122	429
17	4885982	rs238233	L-T-FIN	introninè	<i>CAMTA2</i>	NA	NA	0,10	-0,727456	62
20	19702049	rs3828002	L-T-FIN/CEU-FIN	UTR3	<i>SLC24A3</i>	NA_020689:c.*265G>A	NA	0,03/0,10	-0,468579	263
20	19702171	rs6046271	L-T-FIN/CEU-FIN	UTR3	<i>SLC24A3</i>	NA_020689:c.*387G>A	NA	0,00/0,04	0,860746	9 833
20	19702513	rs16981035	L-T-FIN/CEU-FIN	UTR3	<i>SLC24A3</i>	NA_020689:c.*729G>A	NA	0,12/0,044	-0,958205	19
20	19703335	rs8994	L-T-FIN/CEU-FIN	UTR3	<i>SLC24A3</i>	NA_020689:c.*1551G>A	NA	0,12/0,044	0,834476	9 675
20	19705073	rs1884769	L-T-FIN/CEU-FIN	intergeninè	<i>SLC24A3;RIN2</i>	dist=1532;dist=162092	NA	0	-0,386344	417

CHR	POS	SNP	POP	Func. refGene	Gene, refGene	GeneDetail, refGene	EggnineFunc, refGene	$F_{ST}$	CADD13_Raw Score	CADD13_PHRD
			FIN							
20	19706617	rs7272615	LT-FIN/CEU-FIN	intergenine	SLC24A3;RN2	dist=3076;dist=160548	N/A	0,04	1,770653	14,81
21	21389233	rs2825948	LT-CEU	intergenine	MIR548XHIG;LINC00320	dist=1257103;dist=725675	,	0,05	-0,261817	809
21	21504834	rs2826027	LT-CEU	intergenine	MIR548XHIG;LINC00320	dist=1372704;dist=610074	,	0,07	-0,050714	2 125
21	21526261	rs9982120	LT-CEU	intergenine	MIR548XHIG;LINC00320	dist=1394131;dist=588647	,	0,01	-0,013183	2 461
21	21535472	rs1034348	LT-CEU	intergenine	MIR548XHIG;LINC00320	dist=1403342;dist=579436	,	0,01	-0,079052	1 891
21	21535568	rs1034349	LT-CEU	intergenine	MIR548XHIG;LINC00320	dist=1403438;dist=579340	,	0,02	-0,027294	2 332
21	21543098	rs2826058	LT-CEU	intergenine	MIR548XHIG;LINC00320	dist=1410968;dist=571810	,	0,08	-0,254562	839
21	21561683	rs1980961	LT-CEU	intergenine	MIR548XHIG;LINC00320	dist=1429553;dist=553225	,	0,04	0,178514	4 451
21	21568418	rs2826076	LT-CEU	intergenine	MIR548XHIG;LINC00320	dist=1436288;dist=546490	,	0,04	-0,076038	1 915

**2 lentelė.** 23 analizuojamų populiacijų Weir'o ir Cockheram'o  $F_{ST}$  genetinių atstumų matrica pagal autosomų VNP žymenų pasiskirstymą

	Sindhi	Makra	Hazara	Baloch	Brahui	Palestini	Druze	Bedoui	Mozab	Yorub	Mandenk	BiakaPy	Lithuani
0,0256	0,0254	0,0339	0,0245	0,0262	0,02493	0,0262	0,0300	0,0401	0,1369	0,13498	0,15792	0	Lithuania
0,1480	0,1434	0,1608	0,1509	0,1519	0,14321	0,1591	0,1404	0,1194	0,0393	0,04478	0		BiakaPygmy
0,1214	0,1165	0,1338	0,1241	0,1250	0,11700	0,1330	0,1138	0,0893	0,0089	0			Mandinka
0,1229	0,1179	0,1354	0,1257	0,1267	0,11874	0,1349	0,1155	0,0910	0				Yoruba
0,0369	0,0318	0,0505	0,0345	0,0361	0,02137	0,0285	0,0214	0					Mozabite
0,0246	0,0188	0,0401	0,0205	0,0219	0,00745	0,0123	0						Bedouin
0,0225	0,0174	0,0389	0,0181	0,0196	0,00917	0							Druze
0,0200	0,0148	0,0362	0,0162	0,0177	0								Palestinian
0,0054	0,0030	0,0250	0,0018	0									Brahui
0,0030	0,0011	0,0224	0										Balochi
0,0189	0,0241	0											Hazara
0,0044	0												Makrani
0													Sindhi
													Pathan
													Burusho
													Han
													Yakut
													Japanese
													French
													Basque
													Sardinian
													Russian
													Maya

Maya	Russia	Sardini	Basqu	French	Japane	Yakut	Han	Burush	Pathan
0,1036	0,0099	0,0258	0,0211	0,0127	0,0959	0,0797	0,0947	0,0281	0,0220
0,2293	0,1690	0,1757	0,1777	0,1684	0,2058	0,1978	0,2027	0,1575	0,1554
0,2030	0,1423	0,1486	0,1502	0,1418	0,1796	0,1716	0,1780	0,1313	0,1285
0,2046	0,1441	0,1506	0,1524	0,1438	0,1810	0,1732	0,1793	0,1327	0,1302
0,1296	0,0369	0,0319	0,0367	0,0302	0,1150	0,1001	0,1149	0,0440	0,0369
0,1206	0,0255	0,0207	0,0252	0,0184	0,1086	0,0929	0,1081	0,0310	0,0230
0,1216	0,0212	0,0172	0,0213	0,0139	0,1112	0,0941	0,1104	0,0281	0,0197
0,1172	0,0203	0,0162	0,0203	0,0133	0,1056	0,0895	0,1049	0,0261	0,0180
0,1065	0,0213	0,0323	0,0303	0,0207	0,0922	0,0771	0,0916	0,0127	0,0057
0,1034	0,0193	0,0305	0,0286	0,0190	0,0892	0,0739	0,0885	0,0102	0,0032
0,0770	0,0285	0,0485	0,0453	0,0355	0,0368	0,0286	0,0359	0,0168	0,0177
0,1056	0,0204	0,0301	0,0287	0,0193	0,0918	0,0765	0,0912	0,0120	0,0051
0,0981	0,0205	0,0347	0,0317	0,0218	0,0808	0,0675	0,0802	0,0073	0,0018
0,0972	0,0162	0,0309	0,0275	0,0177	0,0817	0,0669	0,0809	0,0060	0
0,0929	0,0227	0,0398	0,0362	0,0259	0,0723	0,0602	0,0712	0	
0,0960	0,0965	0,1218	0,1193	0,1088	0,0075	0,0317	0		
0,0874	0,0763	0,1042	0,1006	0,0902	0,0298	0			
0,0964	0,0972	0,1227	0,1200	0,1096	0				
0,1160	0,0052	0,0090	0,0068	0					
0,1273	0,0143	0,0131	0						
0,1326	0,0198	0							
0,1046	0								
0									

**3 lentelė.** Divergencijos laikas metais taip tirtų žemynų, VR – Vidurio Rytai; CPA – centrinė pietų Azija; RA – Rytų Azija; Eur – Europa

	Afrika	VR	CPA	RA	Eur	Rusija	Maja	Lietuva
Afrika	0	46 833	53 374	70 368	56 536	57 405	67 821	52 160
VR		0	8 779	32 754	7 074	7 887	30 740	8 880
CPA			0	23 549	8 970	6 706	25 362	8 082
RA				0	31 261	27 322	21 795	26 201
Eur					0	3 490	27 972	5 456
Rusija						0	25 190	2 814
Maja							0	23 674
Lietuva								0

**4 lentelė.** Šešių Lietuvos populiacijos etnolingvistinių grupių Weir'o ir Cockheram'o  $F_{ST}$  genetinių atstumų matrica pagal autosomų VNP žymenų pasiskirstymą. RA – Rytų aukštaičiai, PA – Pietų aukštaičiai, VA – Vakarų aukštaičiai, ŠŽ – Šiaurės žemaičiai, PŽ – Pietų žemaičiai, VŽ – Vakarų žemaičiai

	PA	PŽ	RA	ŠŽ	VA	VŽ
PA	0	0,00258592	0,00276547	0,00273779	0,00220085	0,00388391
PŽ		0	0,00248541	0,00188559	0,00217026	0,00333001
RA			0	0,00223648	0,00198616	0,00376448
ŠŽ				0	0,00200576	0,00329323
VA					0	0,00364053
VŽ						

**5 lentelė.** Bendro individų skaičiaus Lietuvos populiacijoje duomenys iš *Eurostat*

Metai	Lietuva
1960	2 755 600
1961	2 801 500
1962	2 845 600
1963	2 881 100
1964	2 916 800
1965	2 953 600
1966	2 989 300
1967	3 026 800
1968	3 062 000
1969	3 095 700

<b>Metai</b>	<b>Lietuva</b>
1970	3 118 941
1971	3 160 437
1972	3 197 645
1973	3 229 598
1974	3 259 277
1975	3 288 510
1976	3 314 794
1977	3 342 533
1978	3 367 538
1979	3 391 490
1980	3 404 194
1981	3 422 210
1982	3 443 684
1983	3 470 673
1984	3 499 711
1985	3 528 698
1986	3 560 388
1987	3 597 439
1988	3 635 295
1989	3 674 802
1990	3 693 708
1991	3 701 968
1992	3 706 299
1993	3 693 929
1994	3 671 296
1995	3 642 991
1996	3 615 212
1997	3 588 013
1998	3 562 261
1999	3 536 401
2000	3 512 074
2001	3 486 998
2002	3 454 637
2003	3 431 497
2004	3 398 929
2005	3 355 220
2006	3 289 835
2007	3 249 983
2008	3 212 605
2009	3 183 856

<b>Metai</b>	<b>Lietuva</b>
2010	3 141 976
2011	3 052 588
2012	3 003 641
2013	2 971 905
2014	2 943 472
2015	2 921 262

**6 lentelė.** Trumpalaikio efektyvaus populiacijos dydžio nustatytos reikšmės Lietuvos populiacijos aukštaičių grupėje su pasikliautinaisiais intervalais 50-iai kartų

<b>GEN</b>	<b>NE</b>	<b>LWR-95 proc., PI*</b>	<b>UPR-95 proc., PI*</b>
0	1,01E+06	3,98E+05	4,52E+06
1	8,16E+05	3,69E+05	3,22E+06
2	6,58E+05	3,42E+05	2,27E+06
3	5,27E+05	3,02E+05	1,56E+06
4	4,20E+05	2,69E+05	1,04E+06
5	3,32E+05	2,33E+05	6,90E+05
6	2,60E+05	1,86E+05	4,44E+05
7	2,02E+05	1,38E+05	3,02E+05
8	1,67E+05	1,15E+05	2,62E+05
9	1,43E+05	9,72E+04	2,26E+05
10	1,24E+05	8,57E+04	1,94E+05
11	1,11E+05	7,75E+04	1,71E+05
12	1,02E+05	6,89E+04	1,58E+05
13	9,58E+04	6,85E+04	1,48E+05
14	9,02E+04	6,64E+04	1,42E+05
15	8,61E+04	6,69E+04	1,37E+05
16	8,22E+04	6,63E+04	1,29E+05
17	7,83E+04	6,33E+04	1,16E+05
18	7,50E+04	5,90E+04	1,04E+05
19	7,13E+04	5,37E+04	9,60E+04
20	6,74E+04	4,98E+04	8,84E+04
21	6,34E+04	4,62E+04	8,48E+04
22	6,02E+04	4,34E+04	8,17E+04
23	5,67E+04	4,03E+04	8,04E+04
24	5,35E+04	3,77E+04	7,83E+04



<b>GEN</b>	<b>NE</b>	<b>LWR-95 proc., PI*</b>	<b>UPR-95 proc., PI*</b>
25	5,01E+04	3,49E+04	7,32E+04
26	4,67E+04	3,34E+04	6,77E+04
27	4,35E+04	3,21E+04	6,48E+04
28	4,11E+04	3,07E+04	6,17E+04
29	3,89E+04	2,85E+04	5,81E+04
30	3,65E+04	2,66E+04	5,47E+04
31	3,43E+04	2,62E+04	5,20E+04
32	3,24E+04	2,55E+04	4,84E+04
33	3,07E+04	2,39E+04	4,52E+04
34	2,93E+04	2,23E+04	4,30E+04
35	2,82E+04	2,07E+04	4,12E+04
36	2,70E+04	1,93E+04	3,93E+04
37	2,59E+04	1,81E+04	3,77E+04
38	2,50E+04	1,73E+04	3,62E+04
39	2,41E+04	1,66E+04	3,45E+04
40	2,34E+04	1,59E+04	3,34E+04
41	2,27E+04	1,52E+04	3,23E+04
42	2,19E+04	1,47E+04	3,13E+04
43	2,08E+04	1,38E+04	3,06E+04
44	1,98E+04	1,29E+04	3,02E+04
45	1,92E+04	1,21E+04	2,92E+04
46	1,84E+04	1,16E+04	2,88E+04
47	1,80E+04	1,11E+04	2,82E+04
48	1,75E+04	1,09E+04	2,83E+04
49	1,71E+04	1,08E+04	2,84E+04
50	1,69E+04	1,06E+04	2,84E+04

\*Pasikliautinasis intervalas.

**7 lentelė.** Trumpalaikio efektyvaus populiacijos dydžio nustatytos reikšmės Lietuvos populiacijos žemaičių grupėje su pasikliautiniais intervalais 50-iai kartų

<b>GEN</b>	<b>NE</b>	<b>LWR-95 proc., PI*</b>	<b>UPR-95 proc., PI*</b>
0	1,27E+05	5,20E+04	4,58E+05
1	1,20E+05	5,59E+04	3,78E+05
2	1,13E+05	6,00E+04	3,11E+05
3	1,07E+05	6,43E+04	2,56E+05
4	1,00E+05	6,91E+04	2,10E+05
5	9,40E+04	7,21E+04	1,65E+05
6	8,81E+04	7,03E+04	1,39E+05
7	8,24E+04	6,80E+04	1,14E+05
8	7,83E+04	6,25E+04	1,14E+05
9	7,50E+04	5,89E+04	1,10E+05
10	7,16E+04	5,20E+04	1,01E+05
11	6,83E+04	4,59E+04	1,02E+05
12	6,50E+04	4,16E+04	1,02E+05
13	6,16E+04	4,00E+04	9,87E+04
14	5,77E+04	3,90E+04	8,45E+04
15	5,38E+04	3,89E+04	7,71E+04
16	4,96E+04	3,81E+04	6,90E+04
17	4,55E+04	3,43E+04	6,44E+04
18	4,18E+04	3,03E+04	5,83E+04
19	3,82E+04	2,91E+04	5,10E+04
20	3,47E+04	2,59E+04	4,61E+04
21	3,15E+04	2,36E+04	4,21E+04
22	2,90E+04	2,20E+04	3,81E+04
23	2,66E+04	2,00E+04	3,50E+04
24	2,46E+04	1,89E+04	3,28E+04
25	2,26E+04	1,81E+04	3,08E+04
26	2,07E+04	1,70E+04	2,73E+04
27	1,90E+04	1,56E+04	2,50E+04
28	1,79E+04	1,42E+04	2,36E+04
29	1,68E+04	1,31E+04	2,27E+04
30	1,57E+04	1,25E+04	2,10E+04
31	1,48E+04	1,20E+04	1,96E+04
32	1,40E+04	1,16E+04	1,87E+04
33	1,33E+04	1,12E+04	1,78E+04
34	1,27E+04	1,07E+04	1,75E+04

<b>GEN</b>	<b>NE</b>	<b>LWR-95 proc., PI*</b>	<b>UPR-95 proc., PI*</b>
35	1,22E+04	1,01E+04	1,80E+04
36	1,18E+04	9,35E+03	1,86E+04
37	1,13E+04	8,80E+03	1,90E+04
38	1,11E+04	8,34E+03	1,91E+04
39	1,07E+04	7,84E+03	1,85E+04
40	1,04E+04	7,52E+03	1,82E+04
41	1,02E+04	7,22E+03	1,79E+04
42	9,89E+03	6,97E+03	1,76E+04
43	9,46E+03	6,80E+03	1,73E+04
44	9,05E+03	6,49E+03	1,68E+04
45	8,79E+03	6,16E+03	1,64E+04
46	8,47E+03	5,83E+03	1,59E+04
47	8,35E+03	5,56E+03	1,56E+04
48	8,15E+03	5,36E+03	1,52E+04
49	7,99E+03	5,17E+03	1,54E+04
50	7,95E+03	4,96E+03	1,54E+04

\*Pasikliautinis intervalas.

Vilniaus universiteto leidykla  
Universiteto g. 1, LT-01513 Vilnius  
El. p. [info@leidykla.vu.lt](mailto:info@leidykla.vu.lt)  
[www.leidykla.vu.lt](http://www.leidykla.vu.lt)  
Tiražas \_\_ egz.