

I

POPULATSIOONIGENEETIKA GENOTÜÜPIDE TASEMEL

1.1 Geneetilise informatsiooni molekulaarne kodeerimine

1.1.1 Rakk, kromosoom, DNA

Räägitakse, et DNA kuju avastamine toimus tänu keerdrepi nägemisele hetkel, mil teadlase aju töötas just sellel õigel vastuvõtmistemperatuuril. Kogu geneetikateadus oleks võinud kujuneda hoopis teistsuguseks, kui ta oleks lifti kasutanud.*

Terry Pratchett
"Ürgsorts"

* Kuigi tõenäoliselt kiiremaks. Ja selle maksimaalne kandejõud oleks olnud kuni 14 inimest.

Rakk on väikseim elus ehitusühik, mis suudab kas üksi või paljurakulise organismi koostisosana kasvada, areneda ja paljuneda. Kõrvaloleval joonisel 1.1 on kujutatud loomaraku ehitus. Raku erinevate osade ehituse ja ülesannetega võib huviline tutvuda näiteks 12. klassi bioloogia õpiku või siis mõne teatmeteose vahendusel, lühike selgitus on toodud ka ajakirja "Eesti Loodus" 2000. a. aprillikuu numbris ja eestikeelseid üksikasjalikke õpetusi võib netist leida ülikoolide rakubioloogiakursuste loengute näol.

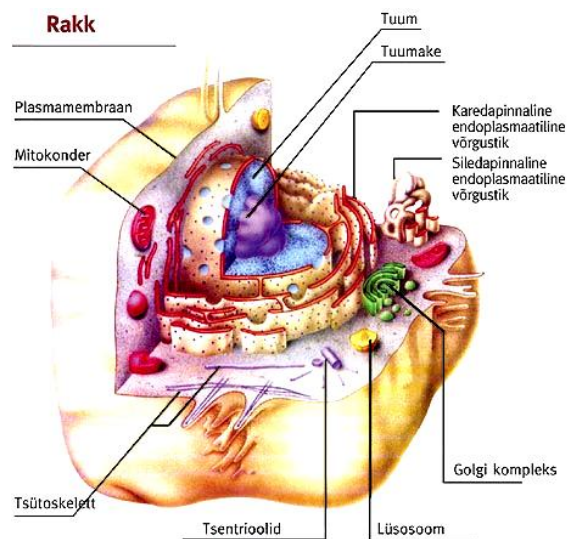
Iga rakk sisaldab kogu liigile omast **geneetilist materjali**, millest enamus on koondunud **raku tuuma** ja on organiseerunud struktuuridesse, mida nimetatakse **kromosoomideks**.

Kromosoom kujutab enesest biheeliksi (topeltspiraali) kujuliselt keerdunud **DNA** molekuli (joonis 1.2).

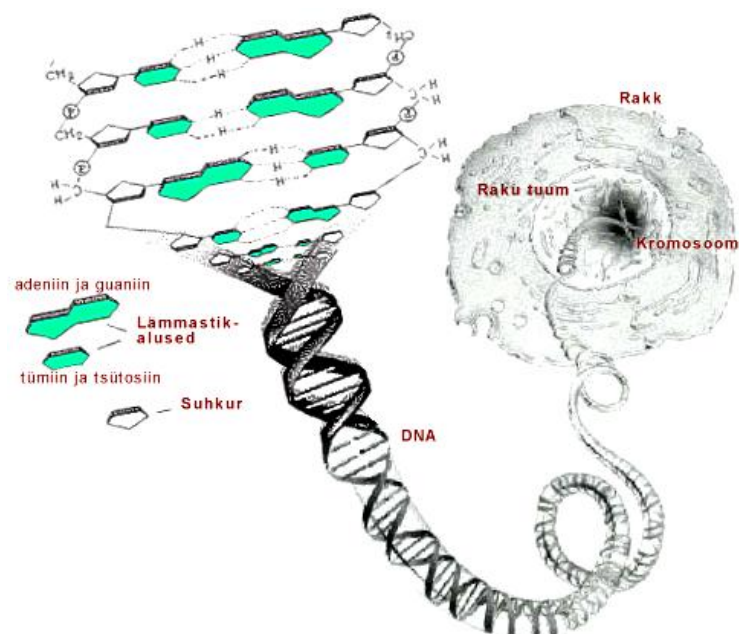
DNA (desoksüribonukleiinhape) on polinukleotiidne ahel, kus **nukleotiidid** on moodustunud kolme ühendi, lämmastikaluse, suhkru (desoksüriboosi) ja fosforhappe jäägi (fosfaatrühma), omavahelisel liitumisel.

DNA ehituses esineb neli erinevat **lämmastikalust**: A - adeniin (adenosiin), G - guaniin (guanosiin), T - tümiin (tümidiin) ja C - tsütosiin (tsütidiin).

Nukleotiidide omavahelise liitumise tulemusel tekib üks DNA ahel. Kahe omavahel koos püsiva ahela aluseks on **DNA komplementaarsusprintsip** – DNA-molekuli moodustavad nukleotiid-ahelad püsivad koos kindla reegli kohaselt – ühe ahela adeniini vastas on teises ahelas alati tümiin ja guaniini vastas tsütosiin (A-T, G-C, T-A, C-G). Järelikult, teades ühe ahela nukleotiidset järjestust ehk **primaarstruktuuri**, saame määrata ka teise ahela nukleotiidse järjestuse.



Joonis 1.1. Looma raku ehitus („Eesti Loodus”, 2000, 4).



Joonis 1.2. DNA paiknemine ja struktuur.

DNA molekuli **sekundaarstruktuur** moodustub vesiniksidemete ühendatud kahe ahela keerdumisel **biheeliksisse** (joonised 1.2 ja 1.3).

DNA on kromosoomide olulisim ehitusvahend. DNA tähtsus seisneb päriliku informatsiooni säilitamises ja selle täpses ülekandes raku jagunemisel.

Kuna geneetiline informatsioon on jagatud paljudesse DNA molekulidesse, siis tähendab see seda, et **kromosoomide arv** tuumas on suurem kui üks.

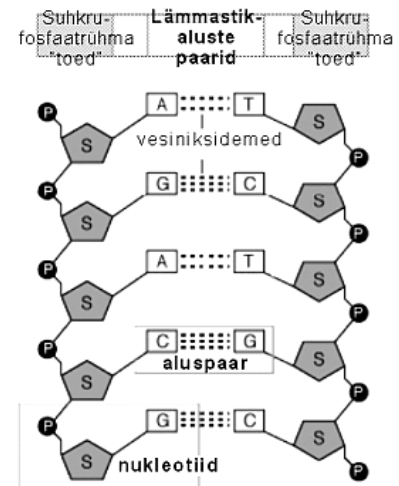
Kõrgematel organismidel (taimed, loomad) on kromosoomide arv **diploidne** (rakutuumas $2n$ kromosoomi), mis tähendab, et iga kromosoom on dubleeritud – esineb kahes eksemplaris.

Vaid **sugurakkudes** on igast **homoloogsete** kromosoomide (paardunud kromosoomid, mis sisaldavad samasid pärilikke tunnuseid määravaid gene, millest kumbki on pärit eri vanemalt) paarist ainult üks kromosoom (seega on suguraku tuumas kokku n kromosoomi, mida nimetatakse **haploidseks** kromosoomistikuks ja mis moodustavad organismi **genoomi**).

Diploidsetes rakkudes on kaks genoomi ($2n$), ühe annab sügoodi (viljastunud munaraku) moodustumisel isas- ja teise emarakk.

Kromosoomide arv varieerub ulatuslikult eri liikidel ja seda isegi ühe organismiderühma piires. Kromosoomiarvu varieerumises ei ole tendentsi olla suurem või väiksem evolutsiooniliselt arenenumatel liikidel, st. kromosoomide arv raku ei ole seoses organismi keerukusega (vt. näiteid tabelist 1.1).

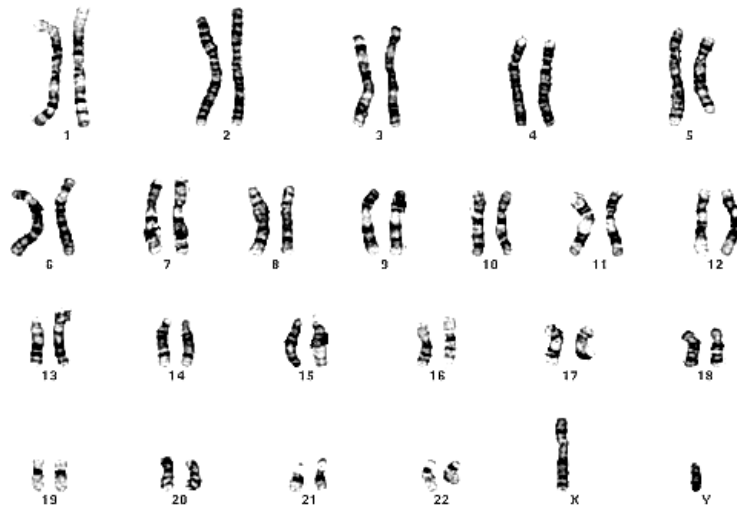
Inimesel on 46 kromosoomi (23 kromosoomipaari), kogupikkusega ~1,6 meetrit, mis koosnevad kokku 3,4 miljardist nukleotiidipaarist. Lahksoolistel liikidel on üks kromosoomipaar sugupooliti erinev ja seda nimetatakse **sugukromosoomide** (ehk gonosoomide) paariks, kõigis ülejäänud kromosoomipaarides on homoloogsed kromosoomid oma kujult ja suuruselt sarnased (joonis 4).



Joonis 1.3. DNA struktuur.

Tabel 1.1. Diploidne kromosoomiarv erinevatel liikidel

Liik	$2n$	Liik	$2n$		
<u>Loomad</u>		<u>Taimed ja seened</u>			
delfiin	<i>Stenella plagiodon</i>	44	kartul	<i>Solanum tuberosum</i>	48
inimene	<i>Homo sapiens</i>	46	mänd	<i>Pinus species</i>	24
hobune	<i>Equus caballus</i>	64	nisu	<i>Triticum monococcum</i>	24
kana	<i>Gallus domesticus</i>	*78	oder	<i>Hordeum vulgare</i>	14
karpkala	<i>Cyprinus carpio</i>	104	põlduba	<i>Vicia faba</i>	12
kass	<i>Felis domesticus</i>	38	pärm	<i>Saccharomyces cerevisiae</i>	32
kits	<i>Carpea hircus</i>	60	tomat	<i>Lycopersicon esculentum</i>	24
koer	<i>Canis familiaris</i>	78	tubakas	<i>Nicotiana tabacum</i>	48
koduhiir	<i>Mus musculus</i>	40	hallitusseened	<i>Aspargillus nidulans</i>	16
lammas	<i>Ovis aries</i>	54		<i>Neurospora crassa</i>	14
laulusääsk	<i>Culex pipiens</i>	6		<i>Penicillium species</i>	8
sig	<i>Sus domesticus</i>	38			
siidiliblikas	<i>Bombyx mori</i>	56	* kromosoomiarv varieerub või pole täpselt teada		
simpans	<i>Pan troglodytes</i>	48			
toakärbes	<i>Musca domestica</i>	12			
veis	<i>Bos taurus</i>	60			
äädikakärbes	<i>Drosophila melanogaster</i>	8			

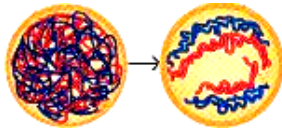


Joonis 1.4. Mehe kromosoomistik – 22 autosoomide paari (kromosoomi, mis esinevad paariliselt liigi kõigil indiviididel) ja sugukromosoomid X ja Y (naistel on kaks X-kromosoomi).

1.1.2 Rakkude paljunemine, mitoos, meioos, ristsiire

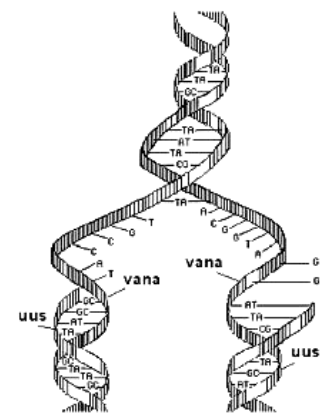
Rakud paljunevad jagunemise teel, mille käigus tekib ühest lähtrakust kaks identset tütarrakku. On olemas kahte tüüpi raku jagunemist. Ühte tüüpi jagunemise tulemusel moodustuvad uued rakud, mis on vajalikud organismi kasvamiseks ja igapäevaselt kahjustuste või haiguste tõttu surevate miljonite rakkude asendamiseks. Sellist raku jagunemist nimetatakse **mitoosiks**.

Teist tüüpi rakkude jagunemisel, mida nimetatakse **meioosiks**, tekivad sugurakud, mille ühinemisel saab alguse uus elusolend.

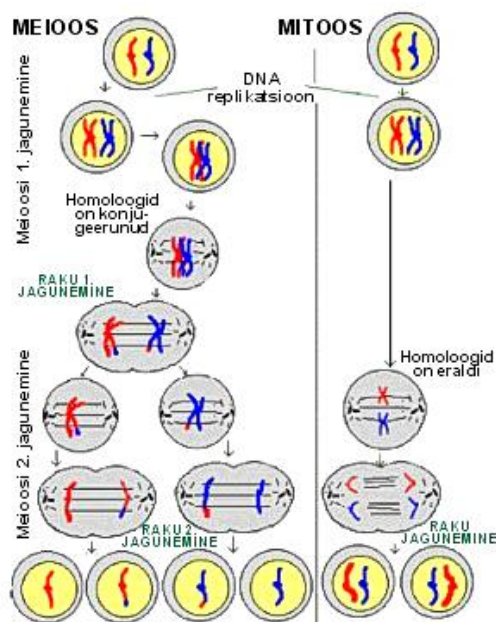


Joonis 1.5. Mitoosi- ja meiosisprotsessile eelnev kromosoomide kondenseerumine.

Mitoosile kui ka meioosile eelneb mingit nõõripundart meenutav DNA-ahelate kondenseerumine npaktseteks kromosoomideks (joonis 1.5) ja seejärel **replikatsioon**, mille käigus kromosoomi moodustav DNA-ahel kahestub ja vastavalt



Joonis 1.6. Kromosoomi replikatsioon.



Joonis 1.7. Meioosi ja mitoosi võrdlus.

DNA komplementaarsusprintsibiile genereeritakse mõlema vana nukleotiidi ahela kõrvale uus (joonis 1.6). Sellisel teel tekkinud **tütarkromatiidid** jäävad esialgu teineteisega seotuks spetsiaalse erinevatest kordus-DNA järjestustest ja valkudest koosneva kromosoomipiirkonna – tsentromeeri kaudu. See on ka ainus piisavalt tihedalt pakitud kromosoomi struktuur, mis võimaldab kromosoomide valgusmikroskoobiga vaadeldes eristada ja nende struktuursete iseärasustega tutvuda. Seetõttu kujutatakse tavaliselt ka joonistel kromosoomi koosnevana kahest tütarkromatiidist (joonised 1.2 ja 1.8).

Mitoosi korral järgneb kromosoomide replikatsioonile tütarkromatiidide lahkumise tütarkromosoomidena ja liikumine raku tuuma eri pooltesse (joonis 1.7), misjärel poolduvad ka tuum ja juba eelnevalt dublitseeritud organelle kahte ossa jagades rakk ise.

Meioos jagatakse kahte etappi. Esimesel meiotilisel jagunemisel replitseerunud homologsed kromosoomid konjugeeruvad, DNA-ahelad lõikuvad, katkevad, vahetavad osi ja ühinevad uuesti (joonis 1.7 – meioosi 1. jagunemine). Sellist homologsete kromosoomide ema- ja isapoolsete koopiade vahelist geneetilist materjali vahetust nimetatakse **ristsiirdeks** e **crossing-over'ks** ja see on ka põhjuseks, miks järglane ei saa oma vanemate kromosoomide täpseid koopiaid. Peale konjugeerumist kromosoomid lahknevad raku tuuma erinevatele poolustele ja rakk jaguneb kaheks. Mõlemad tütararakud sisaldavad ühte homologset kromosoomi, mis koosneb kahest tütarchromatiidist.

Teisel meiotilisel jagunemisel lahknevad tütarchromatiidid analoogselt mitoosile ja kummastki diploidsest tütararakust tekib kaks haploidset suguraku.

Peale viljastumist tungib seemnerakk munarakku, nende tuumad ühinevad, kahekordistades jällegi kromosoomide arvu. See üks viljastatud munarakk (sügoot) ongi uue elusolendi alguseks. Kõik teised rakud ja seeläbi ka kogu organism tekivad sellest raku jagunemise teel.

1.1.3 Valkude süntees, geneetiline kood

Kromosoomi moodustava DNA olulisim ülesanne on säilitada ja edasi anda infot **valkude** moodustamiseks. Valgud e proteiinid on raku olulisemad ehituskivid, samas sünteesitakse valkude kaasabil omakorda väga paljusid muid aineid. Valgud koosnevad omavahel ahelaks ühendatud sadadest või isegi tuhandetest **aminohapetest**. Valgu omadused ja selle paiknemiskoha raku määrab seda moodustavate aminohapete järjestus. Erinevaid aminohappeid on seejuures vaid 20. Selle, mis järjekorras ja kombinatsioonides aminohapped valgu moodustamiseks haakuvad, määrab ära raku tuumas paiknevate nukleotiidide lineaarne järjestus. Nukleotiidide järjestuse tõlkimine valgu aminohappeliseks järjestuseks toimub spetsiaalse võtme – **geneetilise koodi** – abil.

Geneetilise koodi kohaselt vastab ühele aminohappele kolme nukleotiidiline järjestus (**koodon**). Geneetilise koodi sõnastik koosneb siis 64-st kolmetähelisest sõnast (nelja nukleotiidi kolme kaupa kombineerides saame $4^3 = 64$), millele vastavad 20 erinevat aminohapet. See elu seisukohalt väga oluline sõnastik on enamuses organismides sama nn. universaalne geneetiline kood (UGK). Koodonite ja aminohapete vastavust kujutatakse tavaliselt tabelina (vt. tabel 1.2).

Sellist seost DNA ja valkude vahel, milles vahepealse etapina osaleb ka DNA-ga sarnane RNA-molekul (RNA-s on T - tümiini asemel U - uratsiil ning erinevalt DNA kaksikahelalisest struktuurist koosneb RNA vaid ühest nukleotiidide ahelast) nimetatakse **molekulaarbioloogia põhipostulaadiks**: DNA teeb RNA ja RNA teeb valgu (esineb ka vastupidine teisendus – RNA -> DNA, küll aga mitte valgust RNA-ks ja seeläbi DNA-ks).

Eristatakse kolme põhilist geneetilise informatsiooni ülekande protsessi:

1. eelmises peatükis juba käsitletud **replikatsioon** – päriliku materjali (mis võib olla nii DNA kui RNA) kahekordistumine;
2. **transkriptsioon**, mis tähendab mahakirjutamist ja kujutab enesest RNA sünteesi DNA matriitsi alusel;

Tabel 1.2. Universaalne geneetiline kood. Lisaks aminohappeid kodeerivatele koodonitele on UGK-s ka kolm stop-koodonit, millele ei vasta ühtegi aminohapet, aga mida kasutatakse valgusünteesi lõpetamisel nagu punkti lause lõpus (tabelis tähistatud 'Stop').

		2. nukleotiid				
		U	C	A	G	
1. nukleotiid	U	Phe Leu Leu	Ser	Tyr Stop Stop	Cys Stop Trp	U, C A G
	C	Leu	Pro	His Gln	Arg	U, C A, G
	A	Ile Ile Met	Thr	His Lys Lys	Ser Arg Arg	U, C A G
	G	Val	Ala	Asp Glu Glu	Gly	U, C A G

3. **translatsioon**, mis tähendab tõlkimist ja kujutab enesest RNA (seega ka DNA) nukleotiidses järjestuses tõlkimist valkude aminohappeliseks järjestuseks. Valkude sünteesiks vajalikku geneetilist informatsiooni kannab mRNA (matriits- ehk informatsiooniline-RNA). Valgu biosünteesi viib läbi ribosoom (RNA-st ja valkudest koosnev organoid). Aminohapped seatakse ribosoomi abil vastavusse mRNA-s sisalduva geneetilise informatsiooniga tRNA (transport-RNA) vahendusel.

Näide. Punaste vereliblede tähtsaima koostisosa, hemoglobiini, teisendid.

Hemoglobiini molekul (*Hb*-molekul) koosneb neljast kahte tüüpi polüpeptiidahelast (α - ja β -globiinist). Haigused, nagu sirpaneemia ja aneemia (kehvveresus), on tingitud muutustest 146 aminohappe pikkuses β -globiini ahelas, ja täpsemalt, muutustest selle ahela 6.-s positsioonis:

Normaalsetel inimestel (*HbA*) on β -globiini ahela 6.-l positsioonil glutamiinhape (Glu), sirpaneemia-haigetel (*HbS*) aga valiin (Val) ning aneemia-haigetel (*HbC*) lüsiin (Lys).

1.2 Klassikalise geneetika seadused

Klassikaline geneetika hõlmab 1) geeniteooriat e **mendelismi**, mis postuleerib, et organismide pärilikud tunnused on määratud diskreetsete ja püsivate geneetiliste elementide – geenide – poolt, mille pärandumine vanemalt järglastele baseerub sugurakkude moodustumise ning viljastumise käigus toimuva geenide lahknemise ja kombineerumise seaduspäradel ning on kirjeldatav tõenäosusteooria abil (Mendeli seadused); 2) **kromosoomiteooriat**, mis annab reeglid geenide paiknemise ja koospärandumise tarvis (Morgani seadus); 3) **mutatsiooniteooriat**, mis kirjeldab geen-, kromosoom- ja genoommutatsioonide olemust; 4) **populatsioonigeneetikat**, mis formuleerib alleeli- ja genotüübisageduste muutused populatsioonis baseeruvana mendelismil ning kromosoomi- ja mutatsiooniteoorial (oluliseim seadus – Hardy-Weinbergi e populatsiooni geneetilise tasakaalu seadus).

Järgnev osa peatükist 1 käsitleb valdavalt alleeli- ja genotüübisageduste muutusi populatsioonis. Lisaks on antud lühiülevaade klassikalist tõenäosusteooriat illustreerivaist Mendeli seadustest. Ristsiirete toimumise ja seeläbi geenide paiknemise modelleerimisel (millele aluse annab Morgani seadus¹) kasutatavate juhuslike protsesside ja diferentsiaalvõrrandite käsitlemine ei mahu käesoleva kursuse raamidesse.

1.2.1 Klassikalise geneetika põhimõisted ja terminoloogia

Geen – lõik DNA-st, mis sisaldab infot ühe valgu genereerimiseks e valgu kirjeldus DNA tasandil (joonis 1.8). Geen võib olla määratud vaid 1000 aluspaariga, aga võib sisaldada ka mitut miljonit aluspaari, seejuures ei ole geen enamasti üks katkematu jupp, vaid asub DNA-l mitmeks osaks tükeldatult. Inimesel on ~30 tuhat geeni. Võrdluseks: ümarussil on genee ~19000, äädikakärbsel ~13600, geneetikute poolt enimuuritud taimel, müürlaual ~25500, pärmiseenel 6000 – seega ei anna geenide koguarv veel aimu kogu elusorganismist. Miljoni aluspaari kohta on inimesel 12 geeni, äädikakärbsel aga 117. Kogu inimese DNA-st teatakse vaid 30-40%-l olevat kindel funktsioon, kogu ülejäänud osa on lihtsalt ballast (et äkki läheb kunagi vaja). Erinevus kahe inimese DNA vahel on ~0,2% – see annab ettekujutuse mänguruumist geenide funktsioonide selgitamisel inimeste omavahelise võrdlemise teel (võrdluseks: inimese DNA erinevus oma kõige lähemast sugulasest šimpansist on 2%).

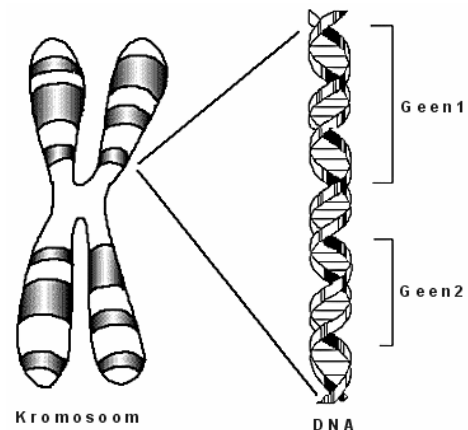
Populatsioonigeneetikas tähistatakse genee tavaliselt ladina tähestiku tähtedega (*A, a, B, C* jne); molekulaarbioloogias ja/või konkreetse funktsiooniga geenist rääkides on aga kasutusel liigispetsiifilised nimed/koodid tähtede-numbrite kombinatsiooni kujul.

Lookus – geeni asukoht DNA-s (kromosoomi geneetiliselt kaardil).

Alleel – geeniteisend, geeni esinemisvorm; üks kahest või mitmest alternatiivsest geenivariandist, mis asuvad populatsiooni isendite homoloogiliste kromosoomide samas lookuses ja toimivad sama tunnuse kujunemisele (tekitades selle eri vorme või avaldumisastmeid). Diploidsel isendil ei saa kunagi olla üle kahe erineva alleeli, ehkki alleeliseerias (erinevate geeniteisendite jadas) võib neid olla kümneid.

Kui üks geen esineb vaid kahel erineval kujul (2 alleeli) on tegemist **dialleelsusega**, kui aga alleele on rohkem, siis **polüalleelsusega**. Dialleelsuse korral tähistatakse üht alleeli suure ja teist väikese tähega (*A* ja *a*), polüalleelsuse korral kasutatakse enamasti indekseid (A_1, A_2, \dots). Sõltuvalt alleelide avaldumisest ja nende funktsioonist on kasutusel hulk erinevaid ja spetsiifilisi tähistusi.

Genotüüp – indiviidi geneetiliste lookuste alleelne koosseis. Diploidsete organismide genotüübi igas geenipaaris on üks alleel saadud isalt, teine emalt.



Joonis 1.8. Geeni paiknemine kromosoomil.

¹ **Morgani seadus**

aheldusreegel: ühes kromosoomis lähestikku paiknevad geenid on lineaarses ahelduses ja päranduvad enamasti koos, sealjuures seda sagedamini, mida väiksem on nende vahemaa piki kromosoomi;

ristsiirdereegel: aheldunud geenid rekombineeruvad meiotilise ristsiirde kaudu, kusjuures nende ümberkombineerumise sagedus on seda suurem, mida suurem on geenide vahemaa piki kromosoomi.

Isendeid, kellel on mõlemas homologses kromosoomis teatud geenilookuses sama alleel (näiteks A ja A , B ja B , q ja q jne.), nimetatakse **homosügootseteks** ja nende genotüüpi vastavate lookuste suhtes märgitakse vastavalt AA , BB või qq .

Kui homologsete kromosoomide samas lookuses on erinevad alleelid (genotüüp näiteks Aa), nimetatakse isendit **heterosügootseks**.

Enam, kui üht lookust hõlmav genotüüp esitatakse lookuste alleelse koosseisu jadana – a'la $AaBBC_1C_3$.

Haplotüüp – haploidne genotüüp, haploidse raku (gameedi) või organismi paaritu genotüüp – üksnes ühelt vanemalt päritud alleelide kogum; kitsamas tähenduses ka kui mingi kromosoomi(lõigu) tihedalt aheldunud lookuste alleelne koosseis. Viimasel juhul tähistatakse haplotüüpi sinna kuuluvate alleelide tähistega (nt vererühmade Rhesus-süsteemi, mis on määratud kolme dialleelse geeni – C , c ; D , d ; E , e – poolt, alleelsed haplotüübid on CDe , cdE , CdE jne).

Kui genotüüpi jagunemine haplotüüpideks on teada, eristatakse erinevad haplotüübid genotüüpi tähistuses kald- või püstjoonega – a'la ABC_1/aBC_3 .

Fenotüüp – indiviidi avaldunud genotüüp (enamasti genotüüpi ja keskkonna koosmõju tulemusel kujunenud tunnused) – näiteks füüsilised tunnused (pikkus, kaal, juuste värvus), biokeemilised näitajad (metaboliitide kontsentratsioon, ensüümide aktiivsus) ja tervisliku seisundi hinnangud (hüpertensioon, diabeet).

Juhul, kui heterosügootne genotüüp (Aa) avaldub samamoodi kui homosügootne genotüüp (AA), nimetatakse selliselt prevaleerivat alleeli (tunnust) **dominantseks**. Dialleelsel juhul tähistatakse dominantset alleeli suure tähega (A). Polüalleelsuse korral võib esineda ka astmeline dominantsus ($A^1 > A^2 > A^3$).

Alleeli (tunnust), mis avaldub vaid homosügootse genotüüpi (aa) korral, nimetatakse **retsessiivseks** ja tähistatakse väikese tähega (a).

Heterosügootidel võib peale dominantse ja retsessiivsuse esineda ka veel **kodominantsus** – mõlema alleeli (tunnuse) avaldumine, või **intermediaarsus** – tunnuse vahepealne avaldumine võrreldes homosügootidega.

Näited.

Inimese vererühmade ABO-süsteem (dominantsus ja kodominantsus).

Meditiinilises käsitluses eristatakse 4 ABO-rühma (nn. veregruppi): 0, A, B ja AB. Need on määratud ühe geeni 3 alleeliga, millest 2 (I^A ja I^B) determineerivad erinevaid glükosiidseid antigeene erütrotsüütide (punaste vereliblede) pinnal, 3. alleel (i ehk I^0) on aga funktsioonita.

Alleelid I^A ja I^B on teineteise suhtes kodominantsed, alleeli i suhtes aga dominantset.

Kuus genotüüpi määravad vererühmi (fenotüüpe) järgmiselt: $I^A I^A$ ja $I^A i$ – grupp A, $I^B I^B$ ja $I^B i$ – grupp B, $I^A I^B$ – grupp AB ning ii – grupp 0.

Inimese vererühmade MN-süsteem (kodominantsus).

Inimese vererakud võivad toota 2 erinevat produkti – N- ja M-antigeeni. Neid antigeene toodavad sama geeni 2 alleelset varianti. Alleeli M suhtes homosügootid toodavad ainult M-antigeeni, alleeli N suhtes homosügootid aga ainult N-antigeeni.

Heterosügootides üks alleel teist maha ei suru, vaid avalduvad mõlemad, ning seetõttu on verest testitavad nii M- kui ka N-antigeen. Sel juhul on alleelid kodominantsed. Kuna kodominantsuse puhul avalduvad alleelid teineteisest sõltumatult, märgitakse mõlemad alleelid suurte tähtedega ja üleval indeksina. Seega on M ja N alleelide tähistused L^M ja L^N . Täht L tuleneb konkreetset juhul erinevate vererühmade avastaja Karl Landsteineri nimest.

Inimeste lainesjuukselisus ja hääletüüp (intermediaarsus).

LL – lokkis, Ll – lainjas, ll – sirge.

BB – bass/alt, Bb – bariton/metsosopran, bb – tenor/sopran.

Küülikute karvavärvus (astmeline dominantsus).

Küülikute karvavärvust määrab üks konkreetne geen tähistusega c . Sellel geenil on 4 erinevat alleeli: c – albiino (c tuleneb inglisekeelsest sõnast “colorness”, värvusetu), c^h – himaalaja e hermeliin, c^{ch} – tšintšilja ja c^+ – metsiktüüp.

Homosügootses olekus on igal alleelil kindel toime karva värvusele. cc küülikud on üleni valge karvaga (albiinod, punaste silmadega), $c^h c^h$ küülikud on valged mustade kõrvade, käppade ja ninaotsaga ning punaste silmadega (nn hermeliinvärvus), $c^{ch} c^{ch}$ küülikud on valgete karvadega, millel on mustad otsad (nö hõbehallid, kusjuures värvus varieerub heledast tumedani) ja $c^+ c^+$ küülikud on tumedakarvalised. Kuna enamus looduslikus populatsioonis elavaid küülikuid on tumedakarvalised, siis kutsutakse c^+ alleeli metsiktüüpi alleeliks. + märk on geneetikutel metsiktüübi tähiseks. Geenid nimetatakse sageli mutantse alleeli järgi ja enamasti just selle alleeli järgi, mille efekt on kõige markantsem (antud juhul valge karvavärvus).

Küüliku karvavärvust määravate alleelide vahel valitseb domineerumises hierarhia: $c^+ > c^{ch} > c^h > c$. Lahtiseletatult tähendab see seda, et metsiktüüpi alleel on täielikult funktsionaalne, tšintšilja ja himaalaja alleelid võimaldavad produtseerida pigmenti vaid osaliselt ning albiino üldsegi mitte. Erinevad alleelide kombinatsioonid heterosügootidel viivad erinevatele fenotüüptidele. Kõik metsiktüüpi alleeli omavad isendid on fenotüübilt tumedakarvalised, $c^{ch} c$ heterosügoot hele tšintšilja, $c^{ch} c^h$ alleelidega küülik hele tšintšilja mustade kõrvade, käppade ja ninaga ning $c^h c$ heterosügoot on fenotüübilt hermeliin.



Tšintšilja



Hermeliin

1.2.2 Mendeli seadused²

Mendeli seadused, avastatud esmakordselt G. Mendeli poolt (1865) ja üleavastatud H. de Vries'i, C. Correns'i ja E. v. Tschermak'i poolt (1900), võtavad kokku klassikalises geneetikas käsitletavate geenide pärandumise ja kombineerumise peamised seaduspärasused.

Mendeli I seadus e ühetaolisuse seadus. Erinevate homosügootsete isendite ristamisel on esimese põlvkonna järglased (F_1)³ kõik ühetaolised heterosügootid ($AA \times aa \rightarrow Aa$ või $aaBB \times AAbb \rightarrow AaBb$).

Mendeli II seadus e lahknemisseadus. Heterosügootide järglaskonnas toimub geneetiline lahknemine, nii et kindlates sagedussuhetes tekivad nii homosügootsed kui ka heterosügootsed isendid.

Algselt oli seadus formuleeritud dialleelse lookuse tarvis: $Aa \times Aa \rightarrow 1/4AA + 2/4Aa + 1/4aa$.

Saadava kolme genotüübi sageduste 1:2:1-suhe tuleneb alleelide lahknemisest erinevatesse gameetidesse meioosis ja nende vabast paardumisest viljastumisel, mis on väljendatav järgmise avaldisega: $Aa \times Aa \rightarrow (1/2A + 1/2a) \times (1/2A + 1/2a) \rightarrow 1/4AA + 2/4Aa + 1/4aa$.

Kirjeldatud genotüübiline 1:2:1-lahknemine võib avalduda erinevate fenotüübiliste lahknemistena (kas 3:1- või 1:2:1-suhteis), olenevalt alleelide interaktsioonist (kas 3/4 dominantseid + 1/4 retsessiivseid või 1/4 tunnusega 1 + 2/4 vahepealseid + 1/4 tunnusega 2).

² Tekst suure osas pärit Mart Viikmaa „Klassikalise geneetika leksikonist” (<http://biomedicum.ut.ee/~martv/genolex.html>)

³ Populatsioonigeneetikas kasutatakse erinevate generatsioonide e põlvkondade tarvis järgmist tähistust: P = vanemate põlvkond (lad. k. „parentes” – vanemad); järglaskonna tähis on F (lad. k. „filia” – tütar; „filialis” – poeg, poja-), esimene järglaspõlvkond on F_1 , teine F_2 jne.

Lahknemisseadus toimib aga ka siis, kui ainult üks vanem on heterosügootne, ja ka siis, kui vanemad on erinevad heterosügootid: $Aa \times aa \rightarrow 1/2Aa + 1/2aa$ või (trialeelsuse korral) $A_1a \times A_2a \rightarrow 1/4A_1A_2 + 1/4A_1a + 1/4A_2a + 1/4aa$.

Näide.

Mõlemad vanemad on blondid. Neil sündis punapäine tütar. Milline on tõenäosus, et järgmisena sünnib: a) punapäine poiss; b) blond tütar?

Vanemad peavad olema dominantsete tunnustega heterosügootid (muidu ei saa teistsuguse tunnusega laps sündida): $Rr \times Rr = \frac{3}{4}R-$ (blond) + $\frac{1}{4}rr$ (punapäine).

a) $\frac{1}{4}$ (punapäine) $\times \frac{1}{2}$ (poiss) = $\frac{1}{8}$;

b) $\frac{3}{4}$ (blond) $\times \frac{1}{2}$ (tüdruk) = $\frac{3}{8}$.

Ülesanne 1. Arvutage, millised on tõenäosused, et heterosügootsete pruunisilmsete vanemate kahest lapsest

a) mõlemad on pruunisilmsed; b) mõlemad on sinisilmsed; c) üks on pruuni-, teine sinisilmne.

Mendeli III seadus e sõltumatu lahknemise seadus e geenide vaba kombineerumise seadus. Polühübriidide⁴ aheldumata alleelipaarid lahknevad üksteisest sõltumatult, mistõttu järglaskonnas toimub kõikide võimalike alleelipaaride vaba kombineerumine.

Selle seaduse kohaselt moodustab polüheterosügoot võrdse sagedusega 2^n erineva haplotüübiga gameeti (kus n on heterosügootsete geenipaaride arv). Nii moodustab kahe lookuse suhtes heterosügootne organism (dihübriid) 4 tüüpi gameete: $AaBb \rightarrow 1/4AB + 1/4Ab + 1/4aB + 1/4ab$; trihübriid 8 tüüpi: $AaBbDd \rightarrow 1/8ABD + 1/8ABd + 1/8AbD + 1/8Abd + 1/8aBD + 1/8aBd + 1/8abd + 1/8abd$. Sõltumatu lahknemise seadus ei kehti aheldunud lookuste korral (sellest tuleb juttu pisut edaspidi).

Näide.

Lainesjuukseline bariton on abielus lainesjuukselise metsosopraniga. Vanemate mõlemad tunnused on intermediaarsed. Missugune on tõenäosus, et nende laps on: a) lainesjuukseline bass; b) lokkisjuukseline sopran?

Tunnuste intermediaarsuse tõttu peavad vanemad olema diheterosügootid. Analüüsime oodatavat lahknemist kummagi tunnuse suhtes eraldi.

Lainesjuukselistel vanematel ($Ss \times Ss$) võib sündida kolme juuksejuga lapsi selliste tõenäosustega: $1/4$ lokkisjuukselisi (S/S), $2/4$ lainesjuukselisi (S/s) ja $1/4$ sirgejuukselisi (s/s).

Hääletüüp on soost sõltuv tunnus, mis on määratud ühe alleelipaariga: mehed vastavalt bassid (Hb/Hb), baritonid (Hb/Ht) ja tenorid (Ht/Ht) ning naised aldid (Hb/Hb) metsosopranid (Hb/Ht) ja sopranid (Ht/Ht). Seega tuleb oodatava lapse hääletüübi ennustamisel arvestada ka sugu.

Oodatavad tõenäosused on:

$1/2 (S/s) \times 1/4 (Hb/Hb) \times 1/2$ (poiss) = $1/16$ – lainesjuukseline bass;

$1/4 (S/S) \times 1/4 (Ht/Ht) \times 1/2$ (tüdruk) = $1/32$ – lokkisjuukseline sopran.

1.2.3 Alleeli- ja genotüübisagedused

Populatsiooni geneetilist seisundit iseloomustab **genotüüpide jaotus**, st populatsiooni koostis genotüüpide järgi. Dialleelse lookuse (alleelidega A ja a) korral on genotüüpide jaotuseks vektor $P(AA), P(Aa), P(aa)$, kus

$$P(AA) + P(Aa) + P(aa) = 1.$$

⁴ Kui jälgitakse ainult ühe geeni (ühe alleelipaari) pärandumist (seega ühe tunnuse avaldumist järglastel) siis on tegemist **monohübriidse ristamisega**. Kui samaaegselt on vaatluse all kaks või rohkem geeni (uuritakse samaaegselt kahe või rohkema tunnuse avaldumist järglastel põlvest põlve), siis on tegemist **polühübriidse** (di-, tri-, tetra-, pentahübriidse jne) **ristamisega**.

Genotüüpide jaotus ei ole püsiv – järglas põlvkonnas toimub alleelide ümberjaotumine üldiselt teistsugusteks genotüüpideks. Seepärast on populatsiooni pikemal jälgimisel otstarbekam kasutada **alleelijaotust**. Vaadeldud dialleelse lookuse korral on alleelijaotuseks vektor $P(A), P(a)$, kus

$$P(A) + P(a) = 1.$$

Kui genotüübijaotus $P(AA), P(Aa), P(aa)$ on teada, siis alleelijaotused saab alati leida valemitega

$$P(A) = P(AA) + \frac{1}{2}P(Aa) \text{ ja } P(a) = 1 - P(A).$$

Järgnevalt uurime genotüübi- ja alleelijaotuse leidmist üldjuhul.

1.2.3.1 Alleelijaotuse leidmine genotüübijaotusest

Juht 1. Eeldame esmalt, et genotüüpide arvud (või nende esinemistõenäosused) populatsioonis on teada, ning tähistame genotüübi $A_i A_j$ esinemiste arvu populatsioonis N_{ij} . Populatsiooni suurus on siis $N = \sum_{ij} N_{ij}$, alleelide arv populatsioonis on $2N$ ja alleel A_i esineb $N_{ii} + \sum_{j \neq i} N_{ij}$ korda. Seega

$$p_i = P(A_i) = \frac{N_{ii} + \sum_{j \neq i} N_{ij}}{2N} = P(A_i A_i) + \frac{1}{2} \sum_{j \neq i} P(A_i A_j). \quad (1.1)$$

Viimane valem näitab ühtlasi, kuidas arvutada teadaoleva genotüübijaotuse alusel alleelijaotust.

Juht 2. Olgu meil järgnevalt genotüüpide arvud N_{ij} teada populatsioonist juhuslikult võetud valimis. Alleelijaotuse hindamiseks kasutame suurima tõepära meetodit (STM), valides esialgu hinnatavateks parameetriteks genotüüpide tõenäosused $p_{ij} = P(A_i A_j)$, mis on kirjeldatavad multinoomijaotuse⁵ abil ja millelele vastav tõepärafunktsioon esitub kujul

$$L(p_{11}, \dots, p_{ij}, \dots, p_{kk}) = \frac{N!}{N_{11}! \dots N_{ij}! \dots N_{kk}!} p_{11}^{N_{11}} \dots p_{ij}^{N_{ij}} \dots p_{kk}^{N_{kk}},$$

kus k on erinevate alleelide arv. Logaritmiline tõepärafunktsioon, mis tuleb maksimeerida koos lisa-kitsendust $\sum p_{ij} = 1$ arvestava Lagrange'i liikmega, on

$$l(p_{11}, \dots, p_{kk}) \propto \sum_{ij} N_{ij} \ln(p_{ij}) - \lambda (\sum_{ij} p_{ij} - 1).$$

Võrdsustades osatuletised $\partial l / \partial p_{ij}$ nulliga, saame võrdused $N_{ij} = \lambda p_{ij}$, mille summeerimisel üle i ja j saame Lagrange'i kordaja hinnanguks $\lambda = N$. Seega on genotüüpide tõenäosuste STM hinnanguteks $\hat{p}_{ij} = N_{ij} / N$.

Edasi kasutame STM omadust, mille kohaselt saab jaotuse parameetrite sileda funktsiooni STM hinnangu siis, kui võrdsustada selle funktsiooni argumendid (jaotuse parameetrid) nende STM hinnangutega. Vaadeldes alleelide tõenäosusi genotüüpide tõenäosuste (e siis jaotuse parameetrite) funktsioonidena vastavalt valemile (1.1), saame alleelide tõenäosuste STM hinnangud

$$\hat{p}_i = \hat{p}_{ii} + \frac{1}{2} \sum_{j \neq i} \hat{p}_{ij} = \frac{N_{ii} + \sum_{j \neq i} N_{ij}}{2N}. \quad (1.2)$$

Ülesanne 2. Tuletage suurima tõepära hinnangud vahetult p_1, \dots, p_k jaoks.

1.2.3.2 Genotüübijaotuse leidmine alleelijaotusest

Valemeist (1.1) ja (1.2) on näha, et erinevatele genotüübijaotustele võivad vastata samad alleelijaotused. Näiteks nii genotüübitõenäosuste $p_{11} = \frac{1}{2}, p_{12} = 0, p_{22} = \frac{1}{2}$ kui ka $p_{11} = \frac{1}{4}, p_{12} = \frac{1}{2}, p_{22} = \frac{1}{4}$ puhul $p_1 = p_2 = \frac{1}{2}$. Seega ei ole üldjuhul võimalik leida alleelijaotusest genotüübijaotust.

Üheks oluliseks erijuhuks, kus alleelijaotus määrab genotüübijaotuse üheselt, on Hardy-Weinbergi tasakaal populatsioonis. Sellisel juhul kehtib valem

$$p_{ij} = \begin{cases} 2p_i p_j, & \text{kui } i \neq j; \\ p_i^2, & \text{kui } i = j. \end{cases} \quad (1.3)$$

⁵ Diskreetse juhusliku vektori (X_1, X_2, \dots, X_m) jaotust nimetatakse multinoomijaotuseks, kui

$$P(X_1 = \kappa_1, \dots, X_m = \kappa_m) = \frac{n!}{\kappa_1! \dots \kappa_m!} p_1^{\kappa_1} \dots p_m^{\kappa_m}, \text{ kus } n \in \mathbb{N}, p_i > 0, \sum_{i=1}^m p_i = 1, \sum_{i=1}^m \kappa_i = n.$$

Tõlgendus: katseseerias, mis koosneb n katses, võivad esineda sündmused A_1, \dots, A_m tõenäosustega p_1, \dots, p_m ; tõenäosusfunktsioon väljendab nende sündmuste toimumiste arvude kombinatsiooni $X_1 = \kappa_1, \dots, X_m = \kappa_m$ tõenäosust.

1.3 Geneetiline tasakaal populatsioonis

1.3.1 Panmiktiline populatsioon. Hardy-Weinbergi seadus

Panmiktiline populatsioon on defineeritud kui populatsioon, milles

- 1) puuduvad mutatsioonid;
- 2) kõigi genotüüpidega isendid on võrdse sigivusega (võrdse eluvõime ja viljakusega), st. puudub valik;
- 3) ristumine indiviidide vahel on täiesti juhuslik;
- 4) ei toimu isendite vahetust teiste populatsioonidega (migratsiooni), st. populatsioon on isoleeritud;
- 5) isendite arvukus on püsivalt väga suur, statistilises mõttes lõpmatu.

Inglise matemaatik G. H. Hardy ja saksa arst W. Weinberg formuleerisid 1908. aastal teineteisest sõltumatult printsiibi, mis käsitleb genotüübi- ja geenisagedusi panmiktilises populatsioonis, postuleerides, et taoline populatsioon on geneetilise tasakaalu seisundis.

Hardy-Weinbergi (HW) seadus e populatsiooni geneetilise tasakaalu seadus e populatsiooni-geneetika põhiseadus: panmiktilises populatsioonis, mis on geneetilise tasakaalu seisundis, püsivad alleeli- ja genotüübisagedused põlvkonniti muutumatutena.

Populatsioon on **geneetilise tasakaalu** (Hardy-Weinbergi tasakaalu) seisundis, kui genotüübisagedused on määratud alleelisagedustest ruutseose kaudu. Näiteks dialleelse lookuse korral, alleelidega A ja a ning alleelisagedustega $P(A) = p$ ja $P(a) = q$, väljendub populatsiooni tasakaalulisus seoses $(p + q)^2 = 1$, kusjuures genotüübisagedused on määratud selle ruutpolünoomi liikmete kaudu: $P(AA) = p^2$, $P(Aa) = 2pq$ ja $P(aa) = q^2$ (vt (1.3)).

Teoreem (tasakaalu tekkimisest). Panmiktilise populatsiooni järglaspõlvkond on iga lookuse osas Hardy-Weinbergi tasakaalus (e panmiktilises populatsioonis tekib tasakaal ühe põlvkonna jooksul).

Tõestus. Vaatame lihtsuse mõttes dialleelset lookust alleelidega A ja a ning alleelisagedustega (tõenäosustega) $P(A) = p$ ja $P(a) = q$. Olgu genotüüpide AA , Aa ja aa tõenäosused lähtepõlvkonnas vastavalt P , Q ja R ning järglaspõlvkonnas P' , Q' ja R' . Tabel 1.3 võtab kokku ristumistüübid, nende tõenäosused ja järglaste genotüüpide jaotuse.

Tabel 1.3. Ristumistüübid, nende tõenäosused ja järglaste genotüüpide jaotus dialleelse lookuse korral.

Ristumistüüp	Ristumise tõenäosus	Järglaste jaotus		
		AA	Aa	aa
$AA \times AA$	P^2	1		
$AA \times Aa$	$2PQ$	$\frac{1}{2}$	$\frac{1}{2}$	
$AA \times aa$	$2PR$		1	
$Aa \times Aa$	Q^2	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
$Aa \times aa$	$2QR$		$\frac{1}{2}$	$\frac{1}{2}$
$aa \times aa$	R^2			1

Täistõenäosuse valemist saame genotüüpide tõenäosused järglaspõlvkonnas

$$P' = P(AA) = 1 \cdot P^2 + \frac{1}{2} \cdot 2PQ + \frac{1}{4} \cdot Q^2 = (P + \frac{1}{2}Q)^2,$$

$$Q' = P(Aa) = \frac{1}{2} \cdot 2PQ + 1 \cdot 2PR + \frac{1}{2} \cdot Q^2 + \frac{1}{2} \cdot 2QR$$

$$= 2(P + \frac{1}{2}Q)(R + \frac{1}{2}Q),$$

$$R' = P(aa) = \frac{1}{4} \cdot Q^2 + \frac{1}{2} \cdot 2QR + 1 \cdot R^2 = (R + \frac{1}{2}Q)^2.$$

Samal ajal tänu valemile (1.2)

$$p = P + \frac{1}{2}Q \text{ ja } q = R + \frac{1}{2}Q.$$

Kokkuvõttes saame seosed $P' = p^2$, $Q' = 2pq$ ja $R' = q^2$, mistõttu vastavalt definitsioonile (1.3) on järglaspõlvkond Hardy-Weinbergi tasakaalus. ■

Näide (Teinberg, 1978). Olgu populatsioonis mittetasakaalulised genotüübisagedused $P(AA) = 0,42$, $P(Aa) = 0,36$, $P(aa) = 0,22$ ja neile vastavad alleelisagedused $P(A) = 0,6$ ja $P(a) = 0,4$.

Arvutame genotüübi- ja alleelisagedused järgmises põlvkonnas:

♀ \ ♂	$P(A) = 0,6$	$P(a) = 0,4$
$P(A) = 0,6$	$P(AA) = 0,36$	$P(Aa) = 0,24$
$P(a) = 0,4$	$P(Aa) = 0,24$	$P(aa) = 0,16$

Alleelisagedused: $P(A) = 0,36 + 0,48/2 = 0,60$; $P(a) = 0,16 + 0,48/2 = 0,40$.

Tulemusest nähtub, et alleelisagedused on täpselt samad, kui eelnevas põlvkonnas. Kujunenud tasakaalulised genotüübisagedused säilivad edaspidi samuti muutumatutena põlvkonnast põlvkonda, seni, kuni ei toimi tasakaalu häirivad tegurid.

Panmiktilist populatsiooni ei ole muidugi olemas. Väga paljudel juhtudel on aga suhteliselt lühikeste ajavahemike kestel (mõni põlvkond) kõrvalekalded neist tingimustest niivõrd väikesed, et populatsiooni võib mitmete geenide suhtes vaadelda praktiliselt tasakaalulisena. Mõnikord võivad nende tegurite vastassuunalised toimed teineteist neutraliseerida, nii et populatsioon on geneetiliselt tasakaalustatud, kuigi talle toimivad võrdlemisi tugevad dünaamikategurid. Kuidas mingite panmiksise eelduste paikapidamatus populatsiooni geneetilist struktuuri mõjutab, seda uurime järgnevais peatükkides.

Populatsiooni geneetilise tasakaalu tingimuse võib esitada ka alleeli- ja genotüübijaotuste kaudu: genotüübijaotus J_g on alleelijaotuste J_a otsekorrutis,

$$J_g = J_a \otimes J_a.$$

Siinjuures tuleks märkida, et viimasest korrutisest avalduvad kõikvõimalikud alleelide kombinatsioonide sagedused eraldi, traditsiooniliste genotüübisageduste saamiseks tuleb osad neist summeerida.

Näide. Olgu meil dialleelne lookus, alleelijaotusega $J_a = P(A), P(a)^T$. Genotüübijaotus $J_g = P(AA), \underbrace{P(Aa), P(aA)}_{P(Aa)}, P(aa)^T$ avaldub siis alleelijaotuste otsekorrutisena:

$$J_g = \begin{bmatrix} P(A) \\ P(a) \end{bmatrix} \otimes \begin{bmatrix} P(A) \\ P(a) \end{bmatrix} = \begin{bmatrix} P(A)P(A) \\ P(A)P(a) \\ P(a)P(A) \\ P(a)P(a) \end{bmatrix}.$$

Siit $P(AA) = [P(A)]^2$, $P(aa) = [P(a)]^2$ ja $P(Aa) = P(A)P(a) + P(a)P(A) = 2P(A)P(a)$, sest fenotüübis avalduvad genotüübid Aa ja aA ühtviisi ning traditsiooniliselt ei eristata neid teineteisest. Saadud tulemused vastavad ka valemile (1.3).

1.3.2 Alleelijaotuse leidmine fenotüübijaotusest. Hardy-Weinbergi tasakaalu kontrollimine

Olgu teada valimisse sattunud fenotüüpide kordsused N_i . Kui genotüüpide ja fenotüüpide vahel on üksühene vastavus, siis on teada ka genotüüpide kordsused ja saab hinnata alleelisagedusi. Taoline situatsioon leiab aset näiteks kodominantsuse ja intermediaarsuse korral.

Dominantsuse korral ei määra aga fenotüüp genotüüpi üheselt. HW seaduse eeldusel saab mõningail lihtsamail juhtudel kõik alleeli- ja genotüübisagedused siiski arvutada.

Näiteks dominantse dialleelse lookuse korral alleelidega A ja a ning alleelisagedustega $P(A) = p$ ja $P(a) = q$ moodustavad genotüübilt puhta rühma vaid retsessiivse tunnusega isendid (aa). Retsessiivsete indiviidide sagedus (genotüübi aa sagedus) tasakaalulises populatsioonis on aga, vastavalt HW seadusele, võrdne retsessiivse geeni sageduse ruuduga (q^2). Siit saame leida $q = \sqrt{q^2}$ ja $p = 1 - q$.

Näide. Holstein-friisi tõugu lehmadel USA-s sünnib keskmiselt 1 punase-valgekirju vasikas iga 200 vasika kohta (Rice jt, 1957). On teada, et musta värvust määrav alleel (B) domineerib punast värvust määrava alleeli (b) üle. Arvutada geenisagedused.

Genotüübi bb sagedus: $R = 1/200 = 0,005$.

Alleeli b sagedus: $q = \sqrt{0,005} = 0,0707$.

Alleeli B sagedus: $p = 1 - 0,0707 = 0,9293$.

Genotüübisagedused populatsioonis arvutame Hardy-Weinbergi seaduse põhjal:

$P(BB) = p^2 = 0,8636$, $P(Bb) = 2pq = 0,1314$ ja $P(bb) = q^2 = 0,005$.

Fenotüübisagedused: $0,9950$ ($BB + Bb$) ja $0,0050$ bb .

Kontrollimaks, kas populatsioon on uuritava tunnuse osas Hardy-Weinbergi tasakaalus, kasutatakse χ^2 -kriteeriumit. Selleks hinnatakse kõigepealt alleelitõenäosused, millest HW tasakaalu eeldusel arvutatakse valemitega (1.3) fenotüüpide teoreetilised tõenäosused. Korrutades viimased valimi mahuga N , saadakse fenotüüpide oodatavad kordsused \hat{N}_i HW tasakaalu korral. Viimaseid

võrreldakse vastavate empiiriliste kordsustega, kasutades χ^2 -statistikut. Kui hinnata tuleb k sõltumatut alleelitõenäosust, on vabadusastmeid $df = \text{fenotüüpide arv} - k - 1$.

Näide. Islandil juhuslikult valitud inimeste veregruppide MM , MN ja NN sagedused saadi vastavalt 233, 385 ja 129 (Mourant, 1954). On need andmed vastavuses Hardy-Weinbergi seadusega?

Leiame vaadeldavate veregruppide oodatavad sagedused eeldusel, et nende kohta kehtib HW seadus.

Kokku osales uuringus 747 inimest ja vaatluse all oli seega $2 \times 747 = 1494$ alleeli. Alleel M esines $(2 \times 233) + 385 = 851$ korda \Rightarrow alleeli M empiiriline sagedus on $851/1494 = 0,5696$. Analoogselt saame alleeli N sageduseks 0,4304.

Genotüüpide MM , MN ja NN HW seadusele vastavad sagedused on siis $(0,5696)^2 = 0,3245$, $2(0,5696)(0,4304) = 0,4903$ ja $(0,4304)^2 = 0,1852$. Korrutades saadud sagedused inimeste arvuga, saame HW seadusele vastavad erinevate veregruppidega inimeste arvud: 242,37 (veregrupiga MM), 366,26 (MN) ja 138,37 (NN).

Empiiriliste andmete võrdlemiseks teoreetiliselt tuletatutega sobiva χ^2 -statistiku leidmiseks vajalikud arvutused on koondatud tabelisse.

	MM	MN	NN	Kokku
Empiiriline (E)	233	385	129	747
Teoreetiline (T)	242,37	366,26	138,37	747
$(E - T)^2 / T$	0,3622	0,9588	0,6345	1,9555

Leitud statistiku väärtusele 1,9555 vastav olulisuse tõenäosus ühe vabadusastme korral (liisaks tavapärasele ühe vabadusastme kaole, mis kaasneb teoreetiliste absoluutsete sageduste arvutamisel kitsendava tingimusena nende summa fikseerimisega, on alleelisageduste hindamisel tarvilik fikseerida ka üks neist – et alleelisageduste summa võrduks ühega – põhjustades sellega ka teise vabadusastme kao) on 0,1620. Seega ei ole alust lükata ümber hüpoteesi veregruppide MN -süsteemi geneetilisest tasakaalust Islandil.

Ülesanne 3. Populatsioonis on 120 punasekirjut ja 1225 mustakirjut veist, viimastest omakorda 890 on punase geeni kandjad.

a) Milline on erinevate genotüüpide ja alleelide sagedus selles populatsioonis?

b) Mitu punasekirjut vasikat peaks sündima järgmises põlvkonnas, kui populatsioon on geneetiliselt tasakaalus ja populatsiooni maht jääb samaks?

Ülesanne 4. Veiste veregruppide EAF-süsteem.

Veiste veregruppide EAF-lookuses on 2 alleeli – '01' ja '02', mis defineerivad 3 erinevat geno- ja fenotüüpi. Eesti punast tõugu veiste populatsioonist valiti 40 looma ja määrati nende veregrupid nimetatud lookuse osas. Fenotüüpide kordsused valimis on $N_{01/01} = 13$, $N_{01/02} = 23$, $N_{02/02} = 4$.

Kas populatsioon on HW tasakaalus?

1.4 Alleelide assotsiatsioon ja ahelduse mittetasakaalus

1.4.1 Alleelide assotsiatsioon

Vaatleme lookuseid (geene) A ja B , alleelidega ning nende esinemissagedustega populatsioonis vastavalt A_1, A_2, \dots, A_m ja B_1, B_2, \dots, B_n ning p_1, p_2, \dots, p_m ja q_1, q_2, \dots, q_n . Iga indiviidi genotüüp koosneb teatavasti kahest haplotüübist – üks saadakse emalt ja teine isalt. Iga haplotüüp sisaldab mõlemat uuritavat lookust ning mingit alleeli kummaski neist. Et geenidel A ja B on vastavalt m ja n erinevat esinemisvormi, on erinevaid haplotüüpe (alleelide kombinatsioone) kokku mn : $A_1B_1, A_1B_2, \dots, A_mB_n$. Viimaste sagedust tähistame $h_{11}, h_{12}, \dots, h_{mn}$.

Kui sündmused, et lookuses A paikneb alleel A_i ja lookuses B alleel B_j , on sõltumatud, on nende alleelide koosinemise sagedus võrdne alleelide A_i ja B_j sageduste korrutisega: $h_{ij} = p_i q_j$.

Viimase võrduse mittekehtimisel loetakse alleelid **assotsieerunuiks** (nad ei esine teineteisest sõltumatult). Kui $h_{ij} > p_i q_j$, on alleelide A_i ja B_j vahel positiivne assotsiatsioon, kui aga $h_{ij} < p_i q_j$, siis negatiivne assotsiatsioon.

Näide. Vaatleme dialleelseid lookuseid A ja B alleelidega vastavalt A_1 ja A_2 ning B_1 ja B_2 . Erinevaid haplotüüpe on neli: A_1B_1, A_1B_2, A_2B_1 ja A_2B_2 , mille sagedused leituna mingis suures populatsioonis on vastavalt 0,4, 0,1, 0,2 ja 0,3.

Kas mõned neist vaadeldavais lookustes paiknevaist alleelidest on antud populatsioonis assotsieerunud?

Lookuses A paiknevate alleelide sagedused on $p_1 = 0,4 + 0,1 = 0,5$ ja $p_2 = 0,2 + 0,3 = 0,5$ ning lookuses B paiknevate alleelide sagedused on $q_1 = 0,4 + 0,2 = 0,6$ ja $q_2 = 0,1 + 0,3 = 0,4$.

Alleelidevahelise assotsiatsiooni puudumisel oleks nende koosinemise sagedused vastavalt

$A_1B_1: 0,5 \cdot 0,6 = 0,3$, $A_1B_2: 0,5 \cdot 0,4 = 0,2$, $A_2B_1: 0,5 \cdot 0,6 = 0,3$ ja $A_2B_2: 0,5 \cdot 0,4 = 0,2$.

Et tegelikud haplotüüpide esinemissagedused (vastavalt 0,4, 0,1, 0,2 ja 0,3) on erinevad viimati arvatud sagedustest, eksisteerib uuritud alleelide vahel ilmne assotsieeritus – alleelid A_1 ja B_1 ning A_2 ja B_2 esinevad koos märksa sagedamini kui oleks omane nende sõltumatuse juhule.

1.4.2 Ahelduse mittetasakaalus

Edasi uurime, kuidas muutub alleelide assotsiatsioon panmiktilises populatsioonis generatsioonist generatsioonini. Vaatleme jällegi lookuseid A ja B , alleelidega ning nende esinemissagedustega populatsioonis vastavalt A_1, A_2, \dots, A_m ja B_1, B_2, \dots, B_n ning p_1, p_2, \dots, p_m ja q_1, q_2, \dots, q_n . Lisaks eeldame, et nende lookuste vaheline rekombinatsioonimäär⁶ on θ .

Olgu haplotüübi $A_i B_j$ esinemissagedus vaadeldavas generatsioonis h_{ij0} . Sama haplotüüp järgmises generatsioonis võib olla kas rekombinantne (alleelid A_i ja B_j on ühte kromosoomi kokku sattunud lookuste A ja B vahel leidnud rekombinatsiooni tulemusel) või mitterekombinantne (vastav alleelide kombinatsioon eksisteeris vaadeldava vanema genotüübis juba enne sugurakkude moodustumist). Esimene juht esineb tõenäosusega θ ja teine tõenäosusega $1 - \theta$. Kokku avaldub tõenäosus, et järglasele antakse edasi haplotüüp $A_i B_j$, summana:

$$h_{ij1} = (1 - \theta)h_{ij0} + \theta p_i q_j \quad (1.4)$$

Haplotüübisageduse muutus generatsioonist 1 võrreldes generatsiooniga 0 on

$$h_{ij1} - h_{ij0} = \theta(p_i q_j - h_{ij0}).$$

Siit on näha, et haplotüübisagedus ei muutu, kui $h_{ij0} = p_i q_j$ (st, et puudub alleelide assotsiatsioon), ning et kõik erinevused h_{ij0} ja $p_i q_j$ vahel on proportsionaalsed rekombinatsioonimääraga θ . Kui $h_{ij0} = p_i q_j$ iga i ja j korral, on haplotüübisagedused generatsioonide viisi konstantsed ja kaks vaadeldud lookust öel-

⁶ Kahe lookuse vaheline **rekombinatsioon** = paaritu arv ristisirded (vaid sellisel juhul on võimalik tuvastada ristisirdede toimumist).

Rekombinatsioonimäär – kahe lookuse vaheline rekombineerumise tõenäosus ühel meiosisiprotsessil.

dakse olevat **tasakaalulises ahelduses**. Vastasel juhul loetakse lookused olevaks **mittetasakaalulises ahelduses** (LD – *linkage disequilibrium*).

Kiirus, millega populatsioon läheneb tasakaalulise ahelduse seisundisse, sõltub eelkõige rekombinatsioonimäärast θ . Kirjutades seose (1.4) ümber kujul

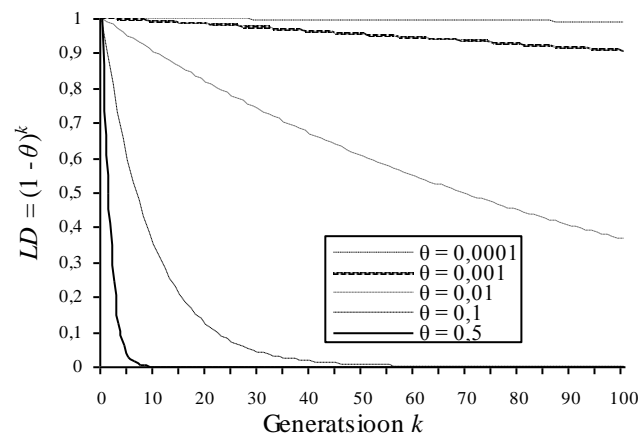
$$h_{ij1} - p_i q_j = (1 - \theta)(h_{ij0} - p_i q_j),$$

näeme, et erinevus haplotüübi sageduse ja selle tasakaalu juhule vastava väärtuse vahel kahaneb $(1 - \theta)$ korda generatsiooni kohta. Seega on peale k generatsiooni populatsiooni erinevus tasakaalust leitav seosest

$$h_{ijk} - p_i q_j = (1 - \theta)^k (h_{ij0} - p_i q_j).$$

Et viimast vahet kasutatakse mõõtmaks alleelidevahelise assotsiatsiooni suurust, loetakse ahelduse mittetasakaalulisust alleelide assotsiatsiooni sünonüümiks.

Joonisel 1.9 on kujutatud ahelduse mittetasakaalulisuse muutus generatsioonide viisi panmiktilises populatsioonis erinevate rekombinatsioonimäärade korral.



Joonis 1.9. Ahelduse mittetasakaalulisuse vähenemine generatsioonide viisi panmiktilises populatsioonis sõltuvalt lookustevahelisest rekombinatsioonimäärast.

Näide. Olgu kahe lookuse vaheline rekombinatsioonimäär 0,01. Eeldades suurt populatsiooni ja juhuslikku paaritumist selles, kui mitme generatsiooniga väheneb alleelide assotsiatsioon (mõõdetuna haplotüübisageduste erinevusena nende tasakaalu juhule vastavaist väärtustest) poole võrra?

Et alleelide assotsiatsioon väheneb $(1 - \theta)$ korda generatsiooni kohta ja $(1 - \theta)^k$ korda k generatsiooni kohta, tuleb ahelduse mittetasakaalulisuse poole võrra vähenemiseks kuluvate generatsioonide arvu k leidmiseks võrdsustada kordaja $(1 - \theta)^k$ $1/2$ -ga.

Võttes võrduse mõlemaist pooltest logaritmi, saame seose

$$k = \ln(0,5)/\ln(1 - \theta),$$

milles asendades $\theta = 0,01$, saame vastuseks $k = 69$ generatsiooni.

1.5 Populatsiooni geneetiline struktuur panmiksise eelduste mittekehtimise korral

1.5.1 Mutatsiooniprotsess

Mutatsioonid on DNA või kromosoomide spontaansed muutused. Et kromosoom- ja genoommutatsioonide toimet populatsiooni geneetilisele struktuurile on suhteliselt raske üldistada nende tüüpide mitmekesisuse tõttu, vaatleme siinkohal ainult geen- ehk punktmutatsioonide mõju geenisagedustele. Nendel mutatsioonidel on arvatavasti ka evolutsioonis kõige olulisem tähtsus.

Mutatsioonid võib jagada laias laastus kaheks: **harvaesinevad mutatsioonid**, mis tekivad ainult kord pikema ajavahemiku jooksul, ei põhjusta alleelisageduses märkimisväärseid muutusi ja mille korral üksiku mutantse alleeli säilimise tõenäosus evolutsioonis läheneb praktiliselt nullile (sest selle alleeli sagedus väheneb iga põlvkonnaga); **korduvad mutatsioonid**, mis tekivad teatud ajavahemike järel uuesti ja produtseerivad küllaldaselt mutantseid gene populatsiooni geneetilise tasakaalu muutmiseks.

Korduvad mutatsioonid on tavaliselt pöörduvad, s. t. et normaalne („uluk-“) alleel võib muutuda mutantseks ja vastupidi. Normaalse alleeli muutumist mutantseks nimetatakse **otsemutatsiooniks**, mutantse muutumist normaalseks aga **pöördmutatsiooniks**.

Tähistatakse geenmutatsioone enamasti kujul $A_i \leftrightarrow A_j$.

Mutatsioonide tõenäosus on väga väike, $u \cdot 10^{-5} \dots 10^{-8}$ raku ühe jagunemise kohta.

Mutatsioonide dünaamikat on sobilik modelleerida **Markovi ahelate**⁷ abil, kus ahela olekujaotuse mingil sammul (ajal, põlvkonnas) moodustab alleelijaotus populatsioonis, näiteks $\mathbf{p} = P(A_1) P(A_2) P(A_3)^T$, ja alleelide muteerumistõenäosused, $P(A_1 \rightarrow A_2), P(A_1 \rightarrow A_3), P(A_2 \rightarrow A_1)$ jne, on antud üleminekumatriksi \mathbf{M} abil (eeldades mutatsioonide sageduse konstantsust ajas e statsionaarset Markovi ahelat).

Juhul, kui populatsiooni panmiktilisust rikuvad üksnes mutatsioonid, on **statsionaarne** olek, millest edasi alleelide jaotus hoolimata võimalikest mutatsioonidest enam ei muutu, leitav võrrandist $\mathbf{M}\mathbf{p} = \mathbf{p}$.

Näide. Vaatleme trialleelset lookust alleelidega A_1, A_2 ja A_3 ning olgu üleminekumatriks antud järgmise tabelina (tumedamas raamis):

Statsionaarsete alleelisageduste leidmiseks lahendame järgneva võrrandisüsteemi, kus kaks võrrandit saame statsionaarsele olekule vastavast võrrandist $\mathbf{M}\mathbf{p} = \mathbf{p}$ (kolmas on neist lineaarselt sõltuv):

		Algolekud		
		A_1	A_2	A_3
Lõpp-olekud	A_1	$1 - 4 \cdot 10^{-6}$	$3 \cdot 10^{-7}$	0
	A_2	10^{-6}	$1 - 3 \cdot 10^{-7}$	$3 \cdot 10^{-6}$
	A_3	$3 \cdot 10^{-6}$		$1 - 3 \cdot 10^{-6}$
Kokku		1	1	1

$$(1 - 4 \cdot 10^{-6}) P(A_1) + 3 \cdot 10^{-7} P(A_2) + 0 \cdot P(A_3) = P(A_1),$$

$$10^{-6} P(A_1) + (1 - 3 \cdot 10^{-7}) P(A_2) + 3 \cdot 10^{-6} P(A_3) = P(A_2);$$

kolmandaks võrrandiks võtame $P(A_1) + P(A_2) + P(A_3) = 1$.

Süsteemi lahendiks on $P(A_1) = 3/46$, $P(A_2) = 40/46$ ja $P(A_3) = 3/46$.

Seega, kui populatsiooni panmiktilisust rikuvad ainult mutatsioonid, siis ajapikku kujunebki niisugune alleelide jaotus.

⁷ Juhuslike suuruste jada $\{X_t\}$, kus t võib omada lõpliku (või loenduva) arvu väärtusi, nimetatakse **Markovi ahelaks**, kui tõenäosus

$$P(X_t = j \mid \underbrace{X_0 = k_0, X_1 = k_1, \dots, X_{t-1} = k_{t-1}}_{\text{minevik}}) = P(X_t = j \mid \underbrace{X_{t-1} = k_{t-1}}_{\text{olevik}}),$$

st. järgmise seisundi prognoosimiseks on vaja teada ainult praegust seisundit.

Võimalikke X_t väärtusi nimetatakse Markovi ahela **seisunditeks** ehk **olekuteks**.

Olekujaotus $\mathbf{p}(t)$ on Markovi ahela võimalikes olekuis viibimise tõenäosusi sammul t kajastav vektor, $\mathbf{p}(t) = p_1(t) p_2(t) \dots^T$, kus $p_i(t)$ on sammul t olekus i viibimise tõenäosus; olekujaotus $\mathbf{p}(0)$ tähistab algolekujaotust.

Tõenäosust $P(X_t = j \mid X_{t-1} = i) = p_{ij}^{(t)}$ nimetatakse **üleminekutõenäosuseks** – seisundist i seisundisse j ülemineku tõenäosus sammul (ajahetkel, põlvkonnaks) t . Kui üleminekutõenäosused $p_{ij}^{(t)}$ ei sõltu t väärtusest, st. iga t korral $p_{ij}^{(t)} = p_{ij}$, siis Markovi ahelat nimetatakse **homogeenseks** e **statsionaarseks**.

Maatriksit $\mathbf{M} = p_{ij}$ nimetatakse Markovi ahela **ülemineku(tõenäosuste)maatriksiks**.

Olekujaotus sammul t on avaldatav kujul $\mathbf{p}(t) = \mathbf{M}\mathbf{p}(t-1)$ ja statsionaarse ahela korral ka kujul $\mathbf{p}(t) = \mathbf{M}^t\mathbf{p}(0)$.

Ülesanne 5. Alleel a_1 muteerub alleeliks a_2 tõenäosusega 10^{-6} ja alleeliks a_3 tõenäosusega 10^{-7} , a_3 muteerub alleeliks a_1 aga tõenäosusega 5×10^{-8} . Teised üleminekud ei ole praktiliselt võimalikud. Leidke populatsiooni statsionaarne seisund.

1.5.2 Valik

Vaatame lihtsuse mõttes dialleelset lookust alleelidega A ja a , mille tõenäosused i -ndas põlvkonnas olgu $P(A)_i = p_i$ ja $P(a)_i = q_i$ ($p_i + q_i = 1$). Oletame, et Hardy-Weinbergi tasakaalus populatsioonis hakkab toimima valik, mida iseloomustavad **selektiooniindeksid** s_1, s_2 ja s_3 , st et genotüüpe AA, Aa ja aa kõrvaldatakse vastavalt proportsioonides s_1, s_2 ja s_3 . Näiteks, kui $s_1 = 0$, jäävad kõik genotüübega AA indiviidid alles, kui aga $s_1 = 1$, siis kõrvaldab valik kõik AA genotüübid. Eeldame ka, et populatsiooni suurus on konstantne, st et matemaatilises esituses järglaspõlvkonna indiviidide arv normeeritakse (jagatakse mingi konstandiga T) peale valikut nii, et populatsiooni suurus ei muutuks. Valiku toime võtab kokku tabel 1.4.

Suurus T tuleb valida nii, et alleelitõenäosuste summa pärast valikut võrduks ühega. Tabelist 1.4 näeme (vt ka valem (1.1)), et A tõenäosus peale valikut, p_i , avaldub A valikueelse tõenäosuse p_{i-1} kaudu valemiga (arvestades, et $q_{i-1} = 1 - p_{i-1}$)

$$p_i = \left[\frac{(1-s_1)p_{i-1}^2}{T} + \frac{1}{2}(1-s_2)2p_{i-1}(1-p_{i-1})/T \right] \\ = \left[\frac{(s_2-s_1)p_{i-1}^2 + (1-s_2)p_{i-1}}{T} \right].$$

Tabel 1.4. Valiku toime genotüüpide tõenäosustele

Genotüüp	Tõenäosus enne valiku toimimist	Tõenäosus pärast valiku toimimist
AA	p^2	$(1-s_1)p^2/T$
Aa	$2pq$	$(1-s_2)2pq/T$
aa	q^2	$(1-s_3)q^2/T$
Kokku	1	1

Tabeli 1.4 viimasest veerust saame avaldada normeeriva konstandi T kujul

$$T = p_{i-1}^2(2s_2 - s_1 - s_3) + 2p_{i-1}(s_3 - s_2) + 1 - s_3,$$

millest järeldub, et

$$p_i = \frac{(s_2 - s_1)p_{i-1}^2 + (1 - s_2)p_{i-1}}{p_{i-1}^2(2s_2 - s_1 - s_3) + 2p_{i-1}(s_3 - s_2) + 1 - s_3}. \quad (1.5)$$

Uurime viimast valemit erijuhul, kui elimineeritakse ainult homosügoote aa (st $s_1 = s_2 = 0, s_3 > 0$). Siis

$$p_i = p_{i-1} / (-s_3 p_{i-1}^2 + 2s_3 p_{i-1} + 1 - s_3) = p_{i-1} / [-s_3(1 - p_{i-1})^2 + 1]$$

ja alleeli A sageduse muutus $\Delta p = p_i - p_{i-1}$ avaldub kujul

$$\Delta p = \frac{p_{i-1} + s_3(1 - p_{i-1})^2 p_{i-1} - p_{i-1}}{-s_3(1 - p_{i-1})^2 + 1} = \frac{s_3 q_{i-1}^2 (1 - q_{i-1})}{1 - s_3 q_{i-1}^2},$$

kus $q_{i-1} = 1 - p_{i-1}$ on alleeli a tõenäosus lähtepõlvkonnas. Saadud valem näitab, et haruldaste retsessiivsete (sageli letaalsete) alleelide protsent kahaneb valiku toimel aeglaselt.

Tõepoolest, kui $s_3 = 1$, siis $\Delta p = q_{i-1}^2 / (1 + q_{i-1})$, ning kui alleeli a tõenäosus q_{i-1} on väike, siis $\Delta p \approx q_{i-1}^2$. Näiteks kui haruldase letaalse alleeli korral $q_{i-1} = 0,001$, siis järglaspõlvkonnas on selle alleeli tõenäosus ikka veel $q_i = q_{i-1} - \Delta p = 0,000999$. Selle efekti tõttu on igas populatsioonis palju latentseid retsessiivseid „defektseid“ (normist hälbivaid) allelele, mis valikule on „kättesaamatud“.

Ülesanne 6. Uurime letaalset retsessiivset alleeli, st oletame, et $s_1 = s_2 = 0$ ja $s_3 = 1$, ning modelleerime selle sageduse q vähenemist populatsioonis.

- a) Võttes vaatluse alla generatsioonid $0, \dots, k$, tuletage valem leidamaks, kui mitu generatsiooni läheb vaja, et retsessiivse alleeli sagedus väheneks q_0 -st q_k -ni.

Märkus. Esmalt avaldage alleelisagedus q_k algse alleelisageduse q_0 ja generatsioonide arvu k kaudu ning sealt juba generatsioonide arv k alleelisageduste q_0 ja q_k kaudu.

- b) Oletame, et albinism on määratud ühe retsessiivse alleeli poolt (tegelikult on neid allelele vähemalt kaks). Euroopas on albinismi sagedus umbes $1/20000$ (so albinismi alleeli suhtes homosügootsete inimeste sagedus!).

Kui mitu generatsiooni ja kui mitu aastat (võttes ühe generatsiooni pikkuseks 25 aastat) kulub selleks, et albinismi sagedust kaks korda vähendada, rakendades valikut maksimaalselt (st võttes $s_1 = s_2 = 0$ ja $s_3 = 1$)?

Teoreem (populatsiooni statsionaarsetest seisunditest). Tähistame

$$p_{\text{lim}} = \frac{s_3 - s_2}{s_1 + s_3 - 2s_2}, \quad (1.6)$$

kus s_1 , s_2 ja s_3 on genotüüpide AA , Aa ja aa selektsiooniindeksid. Siis p_{lim} on $p = P(A)$ jaoks tõmbav (tõukav) asümptoot vastavalt sellele, kas $s_1 + s_3 - 2s_2 > 0$ (< 0). Statsionaarseteks seisunditeks on $p = 0$, $p = 1$ ja kui $0 < p_{\text{lim}} < 1$, siis ka p_{lim} .

Kui $s_1 + s_3 - 2s_2 = 0$, siis $s_3 < s_2$ korral on piirjuhaks $p = 0$, $s_3 > s_2$ korral $p = 1$ ja $s_3 = s_2$ korral (st $s_1 = s_3 = s_2$ korral) populatsiooni struktuur ei muutu.

Tõestus.

Leidmaks populatsiooni statsionaarseid seisundeid konstantsete selektsiooniindeksitega määratud valiku suhtes, so siis olekujaotusi, mis hoolimata jätkuvalt mõjuvast valikust enam ei muutu, tuleb lahendada võrrand $\Delta p = p_i - p_{i-1} = 0$ p_{i-1} suhtes. Valemist (1.5) järeldub, et

$$\begin{aligned} \Delta p &= \frac{p_{i-1}^2(s_2 - s_1) + p_{i-1}(1 - s_2) - p_{i-1}^3(2s_2 - s_1 - s_3) - 2p_{i-1}^2(s_3 - s_2) - p_{i-1}(1 - s_3)}{p_{i-1}^2(2s_2 - s_1 - s_3) + 2p_{i-1}(s_3 - s_2) + 1 - s_3} \\ &= \frac{p_{i-1}^3(s_1 + s_3 - 2s_2) + p_{i-1}^2(3s_2 - s_1 - 2s_3) + p_{i-1}(s_3 - s_2)}{p_{i-1}^2(2s_2 - s_1 - s_3) + 2p_{i-1}(s_3 - s_2) + 1 - s_3}. \end{aligned} \quad (1.7)$$

Võrdsustades viimase avaldise nulliga, saame kuupvõrrandi p_{i-1} suhtes:

$$p_{i-1}^3(s_1 + s_3 - 2s_2) + p_{i-1}^2(3s_2 - s_1 - 2s_3) + p_{i-1}(s_3 - s_2) = 0. \quad (1.8)$$

Selle lahendamisel saame esmalt, et $p_{i-1} = 0$, sest võrrand (1.8) on ümber kirjutatav kujul

$$p_{i-1}[p_{i-1}^2(s_1 + s_3 - 2s_2) + p_{i-1}(3s_2 - s_1 - 2s_3) + (s_3 - s_2)] = 0.$$

Grupeerides järgnevalt avaldise $p_{i-1}^2(s_1 + s_3 - 2s_2) + p_{i-1}(3s_2 - s_1 - 2s_3) + (s_3 - s_2)$ keskmise liidetava sulgudesisesed liikmed sobivalt ja tuues ühised tegurid sulgude ette, saame seose kujul

$$(1 - p_{i-1})[p_{i-1}(2s_2 - s_1 - s_3) + s_3 - s_2] = 0,$$

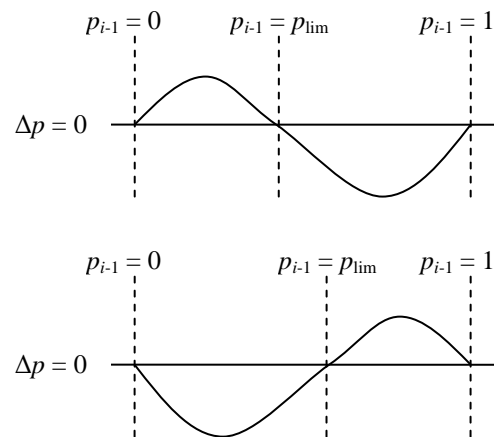
millest $p_{i-1} = 1$ või $p_{i-1} = (s_3 - s_2)/(s_1 + s_3 - 2s_2)$.

Seega alleelisageduse $p = P(A)$ edaspidist muutust fikseeritud valikustrateegia korral ei toimu (st $\Delta p = p_i - p_{i-1} = 0$), kui $p = 0$, $p = 1$ või kui $0 < p = p_{\text{lim}} < 1$.

Kui alleeli A tõenäosus populatsioonis on valiku toimima hakkamise hetkel statsionaarsete seisundite vahel, siis peab aja jooksul $P(A)$ kas suurenema ($\Delta p > 0$) või vähenema ($\Delta p < 0$).

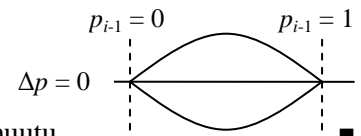
Valemeist (1.7) ja (1.8) nähtub, et alleelitõenäosuse muutus Δp avaldub valikueelse alleelitõenäosuse p_{i-1} kuupfunktsioonina, kusjuures kuupliikme kordaja märk on määratud avaldisega $s_1 + s_3 - 2s_2$. Et kuupliikme kordaja märk näitab omakorda, kuhu suunda kuupfunktsiooni graafiku „saba” pöörab – kui kordaja on nullist suurem, siis pöörab saba üles ja vastupidi – siis kehtivad järgmised seaduspärad:

- kui $s_1 + s_3 - 2s_2 > 0$, siis
 - ✓ $0 < p_{i-1} < p_{\text{lim}}$ korral $\Delta p > 0$, mis tähendab, et alleeli A tõenäosus järglaspõlvkonnas suureneb ja p_{lim} on $p = P(A)$ jaoks tõmbav asümptoot,
 - ✓ $p_{\text{lim}} < p_{i-1} < 1$ korral $\Delta p < 0$, mis tähendab, et alleeli A tõenäosus järglaspõlvkonnas väheneb ja p_{lim} on $p = P(A)$ jaoks jällegi tõmbav asümptoot;
- kui $s_1 + s_3 - 2s_2 < 0$, siis
 - ✓ $0 < p_{i-1} < p_{\text{lim}}$ korral $\Delta p < 0$, mis tähendab, et alleeli A tõenäosus järglaspõlvkonnas väheneb ja p_{lim} on $p = P(A)$ jaoks tõukav asümptoot,
 - ✓ $p_{\text{lim}} < p_{i-1} < 1$ korral $\Delta p > 0$, mis tähendab, et alleeli A tõenäosus järglaspõlvkonnas suureneb ja p_{lim} on $p = P(A)$ jaoks samuti tõukav asümptoot.



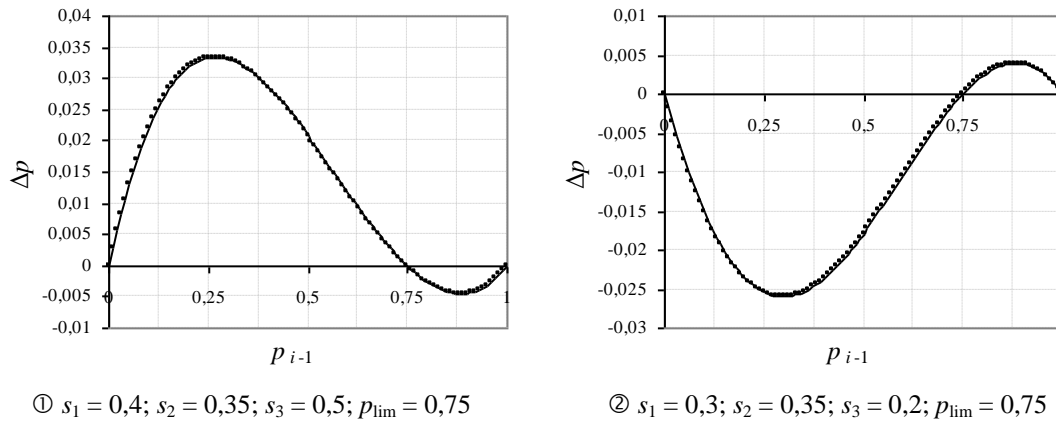
Kui $s_1 + s_3 - 2s_2 = 0$, siis avaldub Δp alleelitõenäosuse p_{i-1} ruutfunktsioonina, kusjuures ruutliikme kordaja märk on määratud avaldisega $3s_2 - s_1 - 2s_3 = s_2 - s_3$. Seega,

- kui $s_3 < s_2$, siis $\Delta p < 0$, mistõttu alleeli A tõenäosus $p \rightarrow 0$,
- kui $s_3 > s_2$, siis $\Delta p > 0$, mistõttu alleeli A tõenäosus $p \rightarrow 1$,
- kui $s_3 = s_2$, siis ka $s_1 = s_3 = s_2$, $\Delta p = 0$ ja populatsiooni struktuur ei muutu.

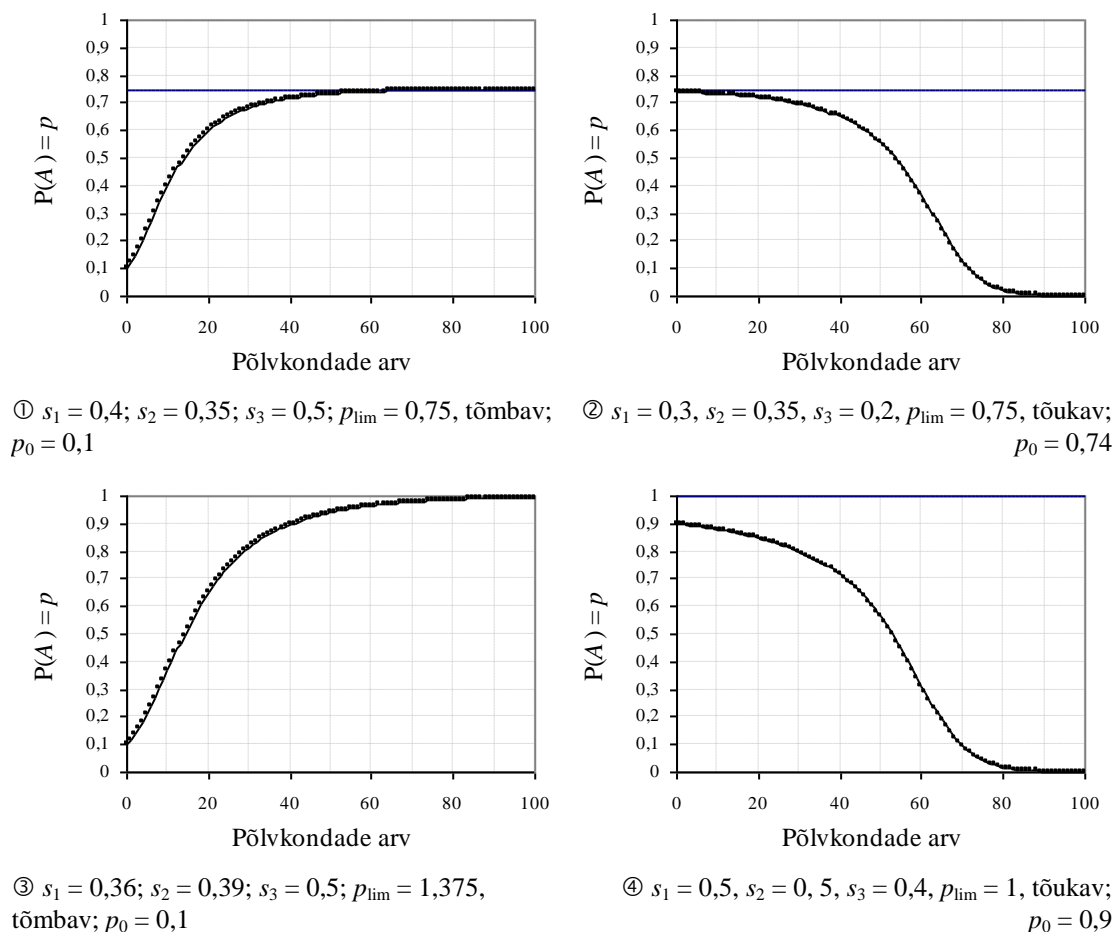


Teoreemis väidetu paremaks mõistmiseks on joonisel 1.10 on kujutatud alleeli A sageduse muutus Δp , arvutatuna valemi (1.7) järgi, erinevate valikustrateegiatega korral. Vasakpoolsel joonisel $s_1 + s_3 - 2s_2 > 0$, millele vastavalt $\Delta p > 0$ (alleeli A sagedus suureneb), kui $p_{i-1} < p_{\text{lim}}$, ja $\Delta p < 0$ (alleeli A sagedus väheneb), kui $p_{i-1} > p_{\text{lim}}$. Parempoolsel joonisel $s_1 + s_3 - 2s_2 < 0$ (millele vastavalt $\Delta p < 0$, kui $p_{i-1} < p_{\text{lim}}$, ja $\Delta p > 0$, kui $p_{i-1} > p_{\text{lim}}$).

Joonisel 1.11 on kujutatud alleeli A sageduse $P(A) = p$ (valem (1.5)) muutumine populatsioonis põlvkondade jooksul erinevate valikustrateegiatega korral (seejuures on graafikutel ① ja ② võetud aluseks samad valikustrateegiad, mis joonise 1.10 graafikuil). Tõmbava asümptoodi p_{lim} korral (graafikud ① ja ③) liigub graafik piirides $0 \dots 1$ asümptoodi poole, tõukava asümptoodi korral (graafikud ② ja ④) aga sellest eemale. Kui alustada A tõenäosusega $p_0 = 0$, $p_0 = 1$ või $p_0 = p_{\text{lim}}$, ei toimuks valiku toimel mingeid muutusi.



Joonis 1.10. Alleeli A sageduse $P(A) = p$ muutus Δp sõltuvalt p väärtustest erinevate valikustrateegiatega korral.



Joonis 1.11. Alleeli A sageduse $P(A) = p$ muutus populatsioonis erinevate valikustrateegiatega korral.

Näide. Thalasseemia (*thalassaemia*) on verehaigus, mida põhjustab letaalne retsessiivne alleel t , kusjuures heterosügootidel Tt esineb küll tervisehäireid, kuid üldiselt on nad elujõulised. Ühes Lõuna-Malta külas osutusid uuritud 740-st koolilapsest 163 heterosügootseks. Oletame, et loodete populatsioon, mis eksisteeris enne, kui valik kõrvaldas genotüübiga tt lapsed, oli Hardy-Weinbergi tasakaalus. Oletame ka, et valik on toimunud piisavalt kaua, nii et uuritav populatsioon on saavutanud statsionaarse seisundi. Hindame nendel eeldustel, kumb genotüüpidest TT või Tt on sealsetes tingimustes elujõulisem.

Antud juhul $s_3 = 1$ ning küsimus on s_1 ja s_2 võrdlemises. Kuna populatsioonis esines mõlemat alleeli, peab $P(T) = p_{\text{lim}}$, sest vastasel korral oleksid statsionaarseteks seisunditeks $P(T) = 0$ või $P(T) = 1$, samuti peab väärtus $P(T)$ olema tõmbav.

Seega $P(T) = p_{\text{lim}} = (1 - s_2)/(s_1 + 1 - 2s_2)$, millest $s_2[2P(T) - 1] + 1 - P(T) = P(T)s_1$ ja

$$s_1 = \frac{P(t)}{P(T)} + \frac{2P(T) - 1}{P(T)} s_2 > \frac{P(t)}{P(T)} s_2 + \frac{2P(T) - 1}{P(T)} s_2 = s_2,$$

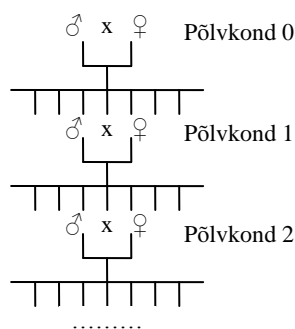
sest $s_2 < 1$ ja $P(T) + P(t) = 1$. Seega on genotüüp TT keskmiselt väiksema elujõulisusega kui Tt – kuigi tavalistes elutingimustes on Tt nõrgem (toob kaasa tervisehäireid), võib aeg-ajalt esineda olukordi, kus Tt on oluliselt parem (näiteks võib Tt olla mõne hävitava nakkushaiguse suhtes vähem vastuvõtlik kui TT).

Ülesanne 7. Homosügoot AA ja aa kõrvaldatakse vastavalt intensiivsusega 0,2 ja 0,5, heterosügoot aga intensiivsusega 0,3. Missuguse statsionaarse seisundi poole populatsioon liigub, kui algselt on populatsioonis mõlemad alleelid?

1.5.3 Regulaarsed ristamissüsteemid, inbriiding

Panmiktilise populatsiooni põhieeldus – juhuslik vaba ristumine – on sageli rikutud seetõttu, et mingis mõttes lähedased indiviidid ristuvad suurema tõenäosusega kui erinevad. Kui suurema tõenäosusega ristuvad geneetiliselt sarnased (suguluses olevad) indiviidid, on tegemist **inbriidinguga** e **sisearetusega**. Inbriidingut kasutatakse aretustöös suurema geneetilise homogeensuse (erijuhul puhta liini) saamiseks. **Regulaarses ristamissüsteemis** ristatakse suguluses olevaid indiviide kindla skeemi järgi, kus iga teatava põlvkondade arvu järel korratakse tsüklikult üht ja sama paaritamisvarianti. Tutvume järgnevalt ühe lihtsama ja hästi uuritud regulaarse ristamissüsteemiga, kasutades Markovi ahelaid.

Õde-vend ristamissüsteemi modelleerimisel vaadeldakse süsteemina (= Markovi ahelana) vanematepaari, mille olekuteks on genotüüpide kõikvõimalikud kombinatsioonid ja diskreetset aega arvestatakse põlvkondade järgi. Järgmise põlvkonna vanematepaariks (st süsteemi olekuks järgmisel ajamomendil) valitakse eelmise paari järglaste seast juhuslikult isane ja emane (vt



Joonis 1.12. Õde-vend ristamissüsteem

joonis 1.12). Vaatame lihtsaimat juhtu, kus genotüüp on määratud ühe dialleelse lookusega (alleelid A ja a). Võimalikeks genotüüpideks on homosügootid AA ja aa ning heterosügoot Aa . Vanematepaari kui süsteemi olekud 1, ..., 6 olgu vastavalt $AA \times AA$, $aaxaa$, $AA \times Aa$, $AA \times aa$, $Aa \times Aa$ ja $Aa \times aa$.

Olekuid $AA \times AA$ ja $aaxaa$ nimetatakse **puhasliinideks**. Puhasliin on neelav olek⁸: kui vanempaar on puhasliinist, on ka kõik järglased samast puhasliinist. Õde-vend ristamissüsteem viib tõenäosusega 1 puhasliini tekkimiseni – tavaliselt just see ongi vastava ristamissüsteemi kasutamise eesmärk.

⁸ Markovi ahela olekut i nimetatakse **neelavaks olekuks**, kui $p_{ii} = 1$. Kui kõik neelavate olekute klassid on ühelelendilised, on tegu **neelava Markovi ahelaga**, kusjuures viimase korral jõuab süsteem tõenäosusega 1 mingisse neelavasse olekusse.

Markovi ahela olekute **kanooniline numeratsioon** tähendab, et esmalt nummerdatakse kõik neelavad olekud ja seejärel ülejäänud. Üleminekumaatriksi kanooniline kuju neelava Markovi ahela korral on $\mathbf{M} = \begin{pmatrix} \mathbf{I} & \mathbf{B} \\ \mathbf{0} & \mathbf{Q} \end{pmatrix}$.

Maatriksit $\mathbf{F} = (\mathbf{I} - \mathbf{Q})^{-1}$ nimetatakse **fundamentaalmatriksiks**. Fundamentaalmatriksi element F_{ji} näitab, kui mitu korda keskmiselt süsteem külastab mitteneelavat olekut j alustades sammul 0 mitteneelavast olekust i .

Olekute nummerdamisel on kasutatud nn kanoonilist numeratsiooni – kõigepealt nummerdatakse neelavad olekud $AAxAA$ ja $aaxaa$ ning siis ülejäänud.

Olekute tinglikud tõenäosused (tinglikud, kuna sõltuvad lähteolekust) järgmises põlvkonnas saame leida vastavalt Mendeli I ja II seadusele. Näiteks, kui lähteolekuks on vanematepaar $AaxAa$, siis genereerib kumbki vanem võrdseis proportsioonides sugurakke A ja a , viimaste juhusliku kombineerumise puhul on $\frac{1}{4}$ järglastest genotüübiga AA , $\frac{1}{2}$ Aa ja $\frac{1}{4}$ aa ; järglasi omakorda juhuslikult paaritades võib saada kõikvõimalikke vanematepaare vastavalt tõenäosustega $\frac{1}{16}, \frac{1}{16}, \frac{1}{4}, \frac{1}{8}, \frac{1}{4}$ ja $\frac{1}{4}$. Analoogselt saame leida uue vanematepaari tõenäosused teistsuguste kui $AaxAa$ lähte paaride korral. Kokkuvõttes saame üleminekutõenäosuste maatriksi \mathbf{M} järgmisel kanoonilisel kujul:

$$\mathbf{M} = \begin{array}{c} \text{l} \\ \text{õ} \\ \text{p} \\ \text{o} \\ \text{l} \\ \text{e} \\ \text{k} \\ \text{u} \\ \text{d} \end{array} \begin{array}{c} \text{AAxAA} \\ \text{aaxaa} \\ \text{AAxAa} \\ \text{AAxaa} \\ \text{AaxAa} \\ \text{Aaxaa} \end{array} \begin{array}{c} \text{algolekud} \\ \text{AAxAA} \quad \text{aaxaa} \quad \text{AAxAa} \quad \text{AAxaa} \quad \text{AaxAa} \quad \text{Aaxaa} \end{array} \begin{pmatrix} 1 & 0 & \frac{1}{4} & 0 & \frac{1}{16} & 0 \\ 0 & 1 & 0 & 0 & \frac{1}{16} & \frac{1}{4} \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{8} & 0 \\ 0 & 0 & \frac{1}{4} & 1 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{1}{2} \end{pmatrix}. \quad (1.9)$$

Oodatava põlvkondade arvu (kaasa arvatud lähtepõlvkond), mis kulub puhasliini kujunemiseks, saab leida fundamentaalmaatriksi

$$\mathbf{F} = (\mathbf{I} - \mathbf{Q})^{-1} = \begin{bmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{8} & 0 \\ \frac{1}{4} & 1 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} \end{pmatrix} \\ \hline \end{bmatrix}^{-1} = \frac{1}{6} \begin{pmatrix} 16 & 8 & 8 & 4 \\ 1 & 8 & 2 & 1 \\ 8 & 16 & 16 & 8 \\ 4 & 8 & 8 & 16 \end{pmatrix}$$

abil. Fundamentaalmatriksi read ja veerud vastavad mitteneelavatele olekutele ja maatriksi elemendid näitavad, kui mitu korda süsteem keskmiselt läbib reale vastavat mitteneelavat olekut alustades veerule vastavast mitteneelavast olekust. Kui lähte paar on $AaxAa$, kulub keskmiselt $(16+1+8+4)/6 = 4,8333$ põlvkonda puhasliini tekkimiseni. Ülejäänud puhasliinile mittevastavaist olekuist jõuab puhasliinini keskmiselt vastavalt 6,6666, 5,6666 ja 4,8333 põlvkonna pärast. Paneme tähele, et olekust $AAxaa$ alustades tekib puhasliin keskmiselt 1 põlvkond hiljem kui olekust $AaxAa$ alustades, sest olekust $AAxaa$ läheb süsteem kõigepealt olekusse $AaxAa$.

Puhasliini tõenäosus $p_{\text{puhas}}(n)$ pärast n põlvkonda kestnud aretustööd avaldub valemiga

$$p_{\text{puhas}}(n) = (1 \ 1 \ 0 \ 0 \ 0 \ 0) \mathbf{p}(n) = (1 \ 1 \ 0 \ 0 \ 0 \ 0) \mathbf{M}^n \mathbf{p}(0), \quad (1.10)$$

kus n . põlvkonna olekujaotuse $\mathbf{p}(n)$ vasakult maatriksiga $(1 \ 1 \ 0 \ 0 \ 0 \ 0)$ korrutamine annab tulemuseks homosügootsete vanemate paaride esinemise tõenäosuse (eeldusel, et olekud on nummerdatud kanooniliselt – st neelavaist olekuist $AAxAA$ ja $aaxaa$ alustades).

Funktsiooni $f(\mathbf{M}) = \mathbf{M}^n$ väärtuse valemis (1.10) võime leida nii „jõuga” ja arvuti abil kui ka avaldada algebraliselt kasutades üldist maatriksi funktsiooni arvutamise võtet⁹. Viimase puhul leiame esmalt

⁹ Mingi maatriksfunktsioon $f(\mathbf{A})$ on kõige üldisemalt esitatav kujul

$$f(\mathbf{A}) = f(\lambda_{11})\mathbf{Z}_{11} + f(\lambda_{12})\mathbf{Z}_{12} + \dots + f^{(n-1)}(\lambda_{1n})\mathbf{Z}_{1n} + \dots + f(\lambda_{k1})\mathbf{Z}_{k1} + f(\lambda_{k2})\mathbf{Z}_{k2} + \dots + f^{(m-1)}(\lambda_{km})\mathbf{Z}_{km} = \sum_{i=1}^k \sum_{j=1}^{n_i} f^{(j-1)}(\lambda_{ij})\mathbf{Z}_{ij},$$

kus λ_{ij} on maatriksi \mathbf{A} i . omaväärtus ja indeks j märgib selle kordsust, $f^{(j-1)}(\lambda_{ij})$ on funktsiooni f ($j-1$). tuletise väärtus kohal λ_{ij} ning \mathbf{Z}_{ij} on üksnes maatriksist \mathbf{A} (ja mitte funktsioonist f) sõltuv konstantne maatriks.

Maatriksi \mathbf{A} omaväärtused leitakse võrrandist $|\mathbf{A} - \lambda\mathbf{I}| = 0$ ja konstantsete maatriksite \mathbf{Z}_{ij} leidmiseks konstrueeritakse võimalikult lihtsatest maatriksi \mathbf{A} funktsioonidest (a’la $f(\mathbf{A}) = \mathbf{I}$, $f(\mathbf{A}) = \mathbf{A}$ jne) võrrandisüsteem ülaltoodud kujul. Seejuures tuleb astmefunktsioonide $f(\mathbf{A}) = \mathbf{A}^r$ korral 0-maatriksist erinev vaid esimene kordseile omaväärtustele $\lambda_{i1}, \dots, \lambda_{in}$ vastavaist maatrikseist \mathbf{Z}_{ij} ja seega $\mathbf{A}^r = \sum_{i=1}^k \lambda_i^r \mathbf{Z}_i$.

Näide. Vaatame 2-olekulist Markovi ahelat, kus üleminekud ühest olekust teise on antud üleminekumaatriksiga

$$\mathbf{P} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{pmatrix} \text{ ja meid huvitab olekujaotus hetkel } t = 1000, \mathbf{p}(1000) = \mathbf{P}^{1000} \mathbf{p}(0). \text{ Maatriksfunktsiooni } \mathbf{P}^{1000} \text{ leidmiseks}$$

arvutame esmalt maatriksi \mathbf{P} omaväärtused, lahendades järgmise võrrandi:

$$|\mathbf{P} - \lambda\mathbf{I}| = \begin{vmatrix} 1-\lambda & \frac{1}{2} \\ 0 & \frac{1}{2}-\lambda \end{vmatrix} = 0 \Rightarrow (1-\lambda)(\frac{1}{2}-\lambda) = 0 \Rightarrow \lambda_1 = 1, \lambda_2 = \frac{1}{2}.$$

maatriksi \mathbf{M} omaväärtused, mis tulevad $\lambda_{1,1} = \lambda_{1,2} = 1$, $\lambda_2 = 1/2$, $\lambda_3 = 1/4$, $\lambda_4 = (1 + \sqrt{5})/4$ ja $\lambda_5 = (1 - \sqrt{5})/4$.

Konstantsete maatriksite $\mathbf{Z}_1, \dots, \mathbf{Z}_5$ ($\mathbf{Z}_{1,1} =: \mathbf{Z}_1, \mathbf{Z}_{1,2} = \mathbf{0}$) leidmiseks konstrueerime järgmise võrrandisüsteemi:

$$\begin{cases} \mathbf{I} = \mathbf{Z}_1 + \mathbf{Z}_2 + \mathbf{Z}_3 + \mathbf{Z}_4 + \mathbf{Z}_5 \\ \mathbf{M} = \mathbf{Z}_1 + \frac{1}{2} \mathbf{Z}_2 + \frac{1}{4} \mathbf{Z}_3 + \frac{1+\sqrt{5}}{4} \mathbf{Z}_4 + \frac{1-\sqrt{5}}{4} \mathbf{Z}_5 \\ \mathbf{M}^2 = \mathbf{Z}_1 + \frac{1}{2} \mathbf{Z}_2 + \frac{1}{4} \mathbf{Z}_3 + \frac{1+\sqrt{5}}{4} \mathbf{Z}_4 + \frac{1-\sqrt{5}}{4} \mathbf{Z}_5 \\ \mathbf{M}^3 = \mathbf{Z}_1 + \frac{1}{2} \mathbf{Z}_2 + \frac{1}{4} \mathbf{Z}_3 + \frac{1+\sqrt{5}}{4} \mathbf{Z}_4 + \frac{1-\sqrt{5}}{4} \mathbf{Z}_5 \\ \mathbf{M}^4 = \mathbf{Z}_1 + \frac{1}{2} \mathbf{Z}_2 + \frac{1}{4} \mathbf{Z}_3 + \frac{1+\sqrt{5}}{4} \mathbf{Z}_4 + \frac{1-\sqrt{5}}{4} \mathbf{Z}_5 \end{cases}$$

ja lahendame selle komponentide $\mathbf{Z}_1, \dots, \mathbf{Z}_5$ suhtes. Tulemuseks saame

$$\mathbf{Z}_1 = \begin{pmatrix} 1 & 0 & 0,75 & 0,5 & 0,5 & 0,25 \\ 0 & 1 & 0,25 & 0,5 & 0,5 & 0,75 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \mathbf{Z}_2 = \begin{pmatrix} 0 & 0 & -0,25 & 0 & 0 & 0,25 \\ 0 & 0 & 0,25 & 0 & 0 & -0,25 \\ 0 & 0 & 0,5 & 0 & 0 & -0,5 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0,5 & 0 & 0 & 0,5 \end{pmatrix}, \mathbf{Z}_3 = \begin{pmatrix} 0 & 0 & -0,05 & 0,2 & 0,05 & -0,05 \\ 0 & 0 & -0,05 & 0,2 & 0,05 & -0,05 \\ 0 & 0 & 0,2 & -0,8 & -0,2 & 0,2 \\ 0 & 0 & -0,1 & 0,4 & 0,1 & -0,1 \\ 0 & 0 & -0,2 & 0,8 & 0,2 & -0,2 \\ 0 & 0 & 0,2 & -0,8 & -0,2 & 0,2 \end{pmatrix},$$

$$\mathbf{Z}_4 = \begin{pmatrix} 0 & 0 & -0,44861 & -0,68541 & -0,55451 & -0,44861 \\ 0 & 0 & -0,44861 & -0,68541 & -0,55451 & -0,44861 \\ 0 & 0 & 0,26180 & 0,4 & 0,32361 & 0,26180 \\ 0 & 0 & 0,05 & 0,07639 & 0,06180 & 0,05 \\ 0 & 0 & 0,32361 & 0,49443 & 0,4 & 0,32361 \\ 0 & 0 & 0,26180 & 0,4 & 0,32361 & 0,26180 \end{pmatrix},$$

$$\mathbf{Z}_5 = \begin{pmatrix} 0 & 0 & -0,00139 & -0,01459 & 0,00451 & -0,00139 \\ 0 & 0 & -0,00139 & -0,01459 & 0,00451 & -0,00139 \\ 0 & 0 & 0,03820 & 0,4 & -0,12361 & 0,03820 \\ 0 & 0 & 0,05 & 0,52361 & -0,16180 & 0,05 \\ 0 & 0 & -0,12361 & -1,29443 & 0,4 & -0,12361 \\ 0 & 0 & 0,03820 & 0,4 & -0,12361 & 0,03820 \end{pmatrix}.$$

Järelikult, kui alustada kõige ebasoodsama paariga $AAXaa$ (algolekujaotus $\mathbf{p}(0) = (0 \ 0 \ 0 \ 1 \ 0 \ 0)^T$), siis puhasliini tõenäosus n -ndas järglaspõlvkonnas on

$$\begin{aligned} p_{puhas}(n) &= 1 \ 1 \ 0 \ 0 \ 0 \ 0 \left[\mathbf{Z}_1 + \left(\frac{1}{2}\right)^n \mathbf{Z}_2 + \left(\frac{1}{4}\right)^n \mathbf{Z}_3 + \left(\frac{1+\sqrt{5}}{4}\right)^n \mathbf{Z}_4 + \left(\frac{1-\sqrt{5}}{4}\right)^n \mathbf{Z}_5 \right] \mathbf{p}(0) \\ &= 1 + \left(\frac{1}{4}\right)^n 0,4 - \left(\frac{1+\sqrt{5}}{4}\right)^n 1,37082 - \left(\frac{1-\sqrt{5}}{4}\right)^n 0,02918. \end{aligned}$$

Vastavad tõenäosused mõnede n väärtuste korral on toodud tabelis 1.5 ja puhasliinini jõudmise dünaamika on kujutatud joonisel 1.13.

Konstantsete maatriksite \mathbf{Z}_1 ja \mathbf{Z}_2 leidmiseks konstrueerime vajaliku võrrandisüsteemi ja lahendame selle:

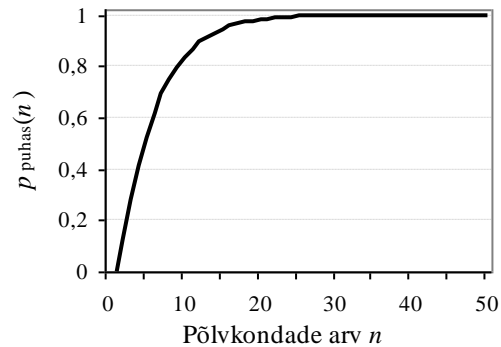
$$\begin{aligned} f_1(\mathbf{P}) = f_1(\lambda_1)\mathbf{Z}_1 + f_1(\lambda_2)\mathbf{Z}_2 &=> \mathbf{I} = 1 \cdot \mathbf{Z}_1 + 1 \cdot \mathbf{Z}_2 &=> \frac{1}{2}\mathbf{Z}_2 = \mathbf{I} - \mathbf{P} &=> \mathbf{Z}_2 = \begin{pmatrix} 0 & -1 \\ 0 & 1 \end{pmatrix}, \mathbf{Z}_1 = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}. \\ f_2(\mathbf{P}) = f_2(\lambda_1)\mathbf{Z}_1 + f_2(\lambda_2)\mathbf{Z}_2 &=> \mathbf{P} = 1 \cdot \mathbf{Z}_1 + \frac{1}{2} \cdot \mathbf{Z}_2 &=> \mathbf{Z}_1 = \mathbf{I} - \mathbf{Z}_2 &=> \mathbf{Z}_2 = \begin{pmatrix} 0 & -1 \\ 0 & 1 \end{pmatrix}, \mathbf{Z}_1 = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

Edasi saame arvutada \mathbf{P}^{1000} ja ahela olekujaotuse $\mathbf{p}(1000)$, võttes näiteks $\mathbf{p}(0) = (0 \ 1)^T$:

$$\mathbf{P}^{1000} = \lambda_1^{1000} \mathbf{Z}_1 + \lambda_2^{1000} \mathbf{Z}_2 = 1 \cdot \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} + \frac{1}{2^{1000}} \cdot \begin{pmatrix} 0 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 - \frac{1}{2^{1000}} \\ 0 & \frac{1}{2^{1000}} \end{pmatrix}, \mathbf{p}(1000) = \begin{pmatrix} 1 & 1 - \frac{1}{2^{1000}} \\ 0 & \frac{1}{2^{1000}} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 - \frac{1}{2^{1000}} \\ \frac{1}{2^{1000}} \end{pmatrix}.$$

n	$p_{\text{puhas}}(n)$
1	0,000
2	0,125
3	0,28125
5	0,52539
10	0,83535
15	0,94294
20	0,98022
50	0,99997

Tabel 1.5. Puhasliini tekkimine öde-vend ristamis-süsteemis algolekujaoituse $\mathbf{p}(0) = (0 \ 0 \ 0 \ 1 \ 0 \ 0)^T$ korral.



Joonis 1.13. Puhasliini tekkimine öde-vend ristamis-süsteemis algolekujaoituse $\mathbf{p}(0) = (0 \ 0 \ 0 \ 1 \ 0 \ 0)^T$ korral.

Ülesanne 8. Kirjutada välja Markovi ahela üleminekutõenäosuste maatriks dialleelisel juhul isetolmlemise korral ning leida, mitmendas põlvkonnas (sh lähtepõlvkonnas) tekib puhasliin.

Ülesanne 9. Tõuisast (homostügootne soovitud geenivariandi suhtes) ristatakse algul juhusliku emasega (mängige läbi kõikvõimalikud emase indiviidi genotüübid – 3 erinevat varianti), siis sellest ristamisest sündinud tütreaga jne, iga kord viimasest ristamisest sündinud tütreaga. Eeldame dialleelset lookust.

a) Kas tekib puhasliin?

b) Kui tekib, siis kui mitme põlvkonna pärast on puhasliini tõenäosus $\geq 0,8$? Lahendused leidke algebraliselt (lahendage vastavad võrratused).

1.5.4 Mitteregulaarne sisearetus, inbriidingukoefitsient

Kaks alleeli on **päritolult identsed** (*identical by descent, IBD*), kui nad on sama eellasalleeli keemilised koopiad (replikatsioonid samast DNA molekulist, võibolla läbi vahepealsete DNA koopiate). Alleelid võivad olla võrdsed ka muul viisil, näiteks kui üks alleel on muteerunud teiseks, kuid see ei ole päritolult identsus.

Indiviidi **inbriidingukoefitsiendiks** (tähist. F) nimetatakse tõenäosust, et tema suvalises fikseeritud lookuses on alleelid päritolult identsed. See tõenäosus ei olene alleelidest ja lookusest, kuid oleneb sellest, missugustes sugulussidemetes on indiviidi eellased omavahel (sest sama alleel saab päranduda nii emalt kui ka isalt üksnes siis, kui leidub vähemalt üks ühine esivanem – ema ja isa on omavahel suguluses). Panmiktilise populatsiooni korral on kõigi indiviidide inbriidingukoefitsiendid 0-d, sest lõpmatu suure populatsiooni ja juhusliku ristumise korral on tõenäosus, et mingi indiviidi vanemad on sugulased, 0 (lõpmata kauget sugulust ei loeta suguluseks – vastasel korral oleksid kõik isendid sugulased).

Üldise seaduspära inbriidingukoefitsiendi leidmiseks, mis baseerub kõigi ühiste eellasteni viivate teede (radade, inglise keeles *path*) läbikäimisel sugupuus, tuletas esmakordselt USA geneetik Sewall Wright 1921. aastal.

Teoreem. Indiviidi O inbriidingukoefitsient avaldub valemiga

$$F_O = \sum_{i=1}^l \sum_{j=1}^{k_i} \frac{1}{2}^{n_{ij}+m_{ij}+1} (1 + F_{W_i}), \quad (1.11)$$

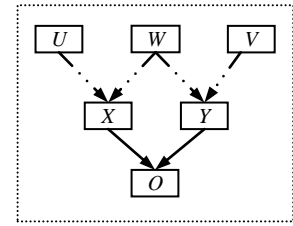
kus l tähistab indiviidi O vanemate X ja Y ühiste eellaste arvu, k_i erinevate radade arvu indiviididest X ja Y ühise i . eellaseni W_i , n_{ij} ja m_{ij} tähistavad lõikude arvu vastavalt indiviidide X ja W_i ning Y ja W_i vahelisel j -l rajal ning F_{W_i} on i . ühise eellase W_i inbriidingukoefitsient.

Tõestus. Vaatame indiviidi O kahe alleeli kõikvõimalikke põlvnemisteid, liikudes sugupuus O juurest ülespoole, ühe alleeli tee läbi ühe vanema, teise alleeli tee läbi teise vanema. Igale põlvnemisteede variantile vastab tõenäosus, et alleelid on põlvnenud just niisuguse variandi kohaselt, kusjuures kõigi põlvnemisvariantide tõenäosuste summa on 1 (sest üks neist variantidest realiseerus).

Kui indiviidi O vanemal X ja Y on üks ühine vanem W , siis tõenäosusega $\frac{1}{4}$ on mõlemad indiviidi O alleelid valitud lookuses pärandunud W -lt, kusjuures omakorda pooltel juhtudel neist on nii läbi X -i

kui ka Y -i pärandunud sama alleel (sellisel juhul on indiviidi O alleelid päritolult identsed tõenäosusega 1) ja pooltel juhtudel erinev alleel (sellisel juhul on indiviidi O alleelid päritolult identsed tõenäosusega F_W). Kokku on indiviidi O alleelide põlvnemisteed ja nende tõenäosused järgmised:

Tee läbi X	Tee läbi Y	Variandi tõenäosus	Tinglik tõenäosus, et alleelid on päritolult identsed
OXU	OYV	$\frac{1}{4}$	0
OXU	OYW	$\frac{1}{4}$	0
OXW	OYV	$\frac{1}{4}$	0
OXW	OYW	$\frac{1}{4}$	$\frac{1}{2} + \frac{1}{2}F_W$



Kasutades täistõenäosuse valemite, saame siit tõenäosuse, et O alleelid on identsed päritolult:

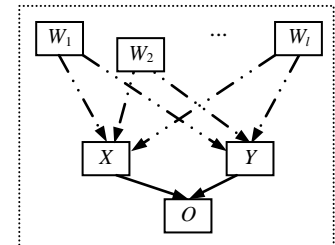
$$F_O = \frac{1}{4} \cdot (\frac{1}{2} + \frac{1}{2}F_W) = \frac{1}{8}(1 + F_W).$$

Kui indiviidi W näol on tegu varasema eellasega, siis tuleb inbriidingukoefitsienti korrutada $\frac{1}{2}$ -ga iga täiendava indiviidide X ja W või Y ja W vahelise generatsiooni kohta, sest iga generatsiooniga väheneb sama alleeli pärimise tõenäosus $\frac{1}{2}$ korda. Valemina on inbriidingukoefitsient esitatav kujul

$$F_O = \frac{1}{2} [(\frac{1}{2})^{n+m}(1 + F_W)] = (\frac{1}{2})^{n+m+1}(1 + F_W)$$

kus n ja m tähistavad vastavalt indiviidide X ja W ning Y ja W vahelist generatsioonide arvu.

Kuna erinevaid radu indiviididest X ja Y ühise eellaseeni W võib olla mitmeid, tuleb kõik need sama alleeli edasikandumisvõimalused inbriidingukoefitsiendi arvutamisel arvesse võtta. Tulemuseks on summa üle kõigi radade ühise eellaseeni:

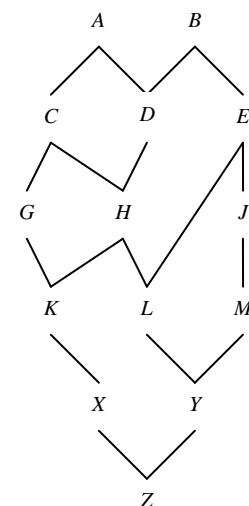


$$F_O = \sum_{j=1}^k \frac{1}{2}^{n_j+m_j+1} (1 + F_W),$$

kus k tähistab erinevate radade arvu ning n_j ja m_j tähistavad vastavalt indiviidide X ja W ning Y ja W vahelist generatsioonide arvu läbi j . raja.

Kui ühiseid eellasi on mitu (näiteks l), on vajalik summeerimine üle nende kõigi, mis annabki tulemuseks valemi (1.11). Seejuures tuleb silmas pidada, et kui alleelide pärimisteed lõikuvad (lõikepunktis on ühine eellane), siis tuleb trajektoor sellest ühisest eellasest ettepoole (ülespoole) arvestusest välja jätta. ■

Näide 1. Analüüsime põlvnemispuud kõrvaloleval joonisel, leides indiviidi Z inbriidingukoefitsiendi. Indiviidi Z vanemate, X ja Y , ühised eellased on A, B, C ja H (on ka D , aga põlvnemistee temani kulgeb nii X -i kui ka Y -i puhul läbi ühise eellase H , mistõttu alleelide põlvnemine seda teed mööda on juba arvesse võetud). Koondame arvutused alljärgnevasse tabelisse.

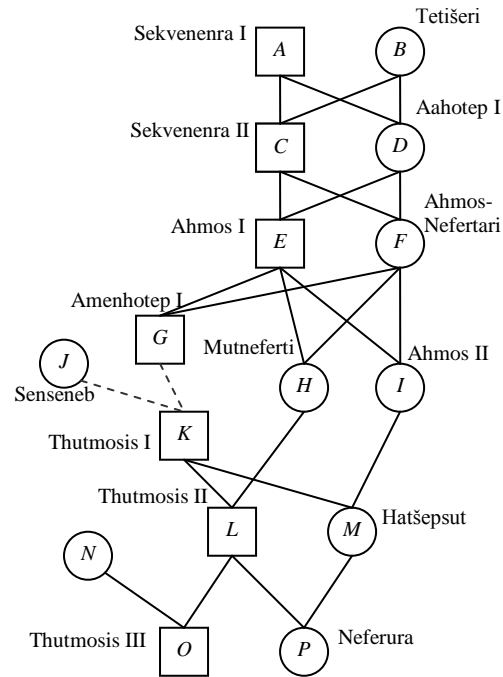


Tee	$n_{ij} + m_{ij}$	F_i	F_Z liidetav
$XKHLY$	4	$F_H = \frac{1}{8}$	$(\frac{1}{2})^5(1 + \frac{1}{8})$
$XKGCHLY$	6	0	$(\frac{1}{2})^7$
$XKHDBELY$	7	0	$(\frac{1}{2})^8$
$XKHDBEJMY$	8	0	$(\frac{1}{2})^9$
$XKGCADHLY$	8	0	$(\frac{1}{2})^9$

$$F_Z = (\frac{1}{2})^5 \cdot (1 + \frac{1}{8}) + (\frac{1}{2})^7 + (\frac{1}{2})^8 + 2 \cdot (\frac{1}{2})^9 = \frac{13}{256} = 0,05078.$$

Näide 2. Joonisel 1.14 on kujutatud osa Egiptuse Uuele riigile aluse pannud vaaraode 18. dünastia (1552 e.m.a.) sugupuust. Leiame kuninganna Hatšepsuti ja vaarao Thutmosis II tütre Neferura inbriidingukoefitsiendi, jättes arvesse võtmata vaarao Thutmosis I spekulatiivse pärinemise (joonisel katkendlikud jooned).

Tee	$n_{ij} + m_{ij}$	F_i	F_P liidetav
<i>LKM</i>	2	0	$(\frac{1}{2})^3$
<i>LHEIM</i>	4	$F_E = \frac{1}{4}$	$(\frac{1}{2})^5(1 + \frac{1}{4})$
<i>LHFIM</i>	4	$F_F = \frac{1}{4}$	$(\frac{1}{2})^5(1 + \frac{1}{4})$
<i>LHECFIM</i>	6	0	$(\frac{1}{2})^7$
<i>LHFCEIM</i>	6	0	$(\frac{1}{2})^7$
<i>LHFDEIM</i>	6	0	$(\frac{1}{2})^7$
<i>LHEDFIM</i>	6	0	$(\frac{1}{2})^7$
<i>LHECADFIM</i>	8	0	$(\frac{1}{2})^9$
<i>LHEDACFIM</i>	8	0	$(\frac{1}{2})^9$
<i>LHFCADEIM</i>	8	0	$(\frac{1}{2})^9$
<i>LHFDACEIM</i>	8	0	$(\frac{1}{2})^9$
<i>LHECBDFIM</i>	8	0	$(\frac{1}{2})^9$
<i>LHEDBCFIM</i>	8	0	$(\frac{1}{2})^9$
<i>LHFCBDEIM</i>	8	0	$(\frac{1}{2})^9$
<i>LHFDBCEIM</i>	8	0	$(\frac{1}{2})^9$



Joonis 1.14. Osa Egiptuse vaaraode 18. dünastia sugupuu algusest.

$$F_P = (\frac{1}{2})^3 + 2 \cdot (\frac{1}{2})^5 \cdot (1 + \frac{1}{4}) + 4 \cdot (\frac{1}{2})^7 + 8 \cdot (\frac{1}{2})^9$$

$$= \frac{1}{8} + \frac{5}{64} + \frac{1}{32} + \frac{1}{64} = \frac{16}{64} = 0,25.$$

Ülesanne 10. Leida joonisel 1.14 kujutatud sugupuu põhjal vaaraode Thutmosis III ja Amenhotep I inbriidingukoefitsiendid.

1.5.5 Inbriidingukoefitsient regulaarses ristamissüsteemis

Tuletame inbriidingukoefitsiendi väärtuse õde-vend ristamissüsteemis suvalises põlvkonnas kahel erineval viisil – esmalt vahetult inbriidingukoefitsiendi arvutamise valemist lähtudes ja teiseks Markovi ahela abil.

Eeldame, et õde-vend ristamiskeemi aluseks olevad vanemad (põlvkond 0) ei ole sugulased. Sellisel juhul $F_0 = 0$ ja samuti $F_1 = 0$. Inbriidingukoefitsiendi väärtus põlvkonnas 2 tuleb vastavalt valemile (1.11) $F_2 = (\frac{1}{2})^{2+1}(1 + F_0) + (\frac{1}{2})^{2+1}(1 + F_0) = \frac{1}{4}$, põlvkonnas 3 (vt ka joonis 1.15)

$$F_3 = \underbrace{2 \cdot (\frac{1}{2})^{2+1}(1 + F_1)}_{\text{rajad ECF ja EDF}} + \underbrace{4 \cdot (\frac{1}{2})^{4+1}(1 + F_0)}_{\text{ECADF, ECBDF, EDACF, EDBCF}} = (\frac{1}{2})^2(1 + F_1) + (\frac{1}{2})^3(1 + F_0)$$

jne. Kokkuvõttes avaldub inbriidingukoefitsient põlvkonnas t seosena

$$F_t = (\frac{1}{2})^2(1 + F_{t-2}) + (\frac{1}{2})^3(1 + F_{t-3}) + \dots + (\frac{1}{2})^t(1 + F_0)$$

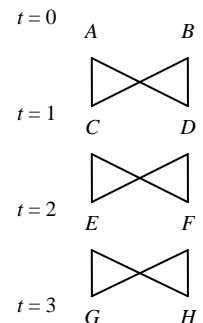
kust $2F_t - F_{t-1} = \frac{1}{2}(1 + F_{t-2})$ ja $F_t = \frac{1}{2}F_{t-1} + \frac{1}{4}F_{t-2} + \frac{1}{4}$. Viimasest esitusest saame omakorda matriksvõrduse

$$\begin{pmatrix} F_{t+1} \\ F_t \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ 1 & 0 \end{pmatrix} \begin{pmatrix} F_t \\ F_{t-1} \end{pmatrix} + \begin{pmatrix} \frac{1}{4} \\ 0 \end{pmatrix},$$

mis on matrikseid sobivalt tähistades esitatav ka kujul

$$\mathbf{U}_{t+1} = \mathbf{A}\mathbf{U}_t + \mathbf{C},$$

kus \mathbf{A} ja \mathbf{C} on konstantsed matriksid ja $\mathbf{U}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}^T$. Saadud valemist korduvalt rakendades leiame, et $\mathbf{U}_2 = \mathbf{C}$, $\mathbf{U}_3 = (\mathbf{A} + \mathbf{I})\mathbf{C}$, $\mathbf{U}_4 = (\mathbf{A}^2 + \mathbf{A} + \mathbf{I})\mathbf{C}$, ..., $\mathbf{U}_t = (\mathbf{A}^{t-2} + \mathbf{A}^{t-3} + \dots + \mathbf{I})\mathbf{C}$. Järelikult on \mathbf{U}_t avaldatav matriksi \mathbf{A} funktsiooni $f_t(\mathbf{A})$ abil: $\mathbf{U}_t = f_t(\mathbf{A})\mathbf{C}$.



Joonis 1.15. Õde-vend ristamissüsteemi skeem.

Et skalaaride korral kehtib vastavalt geomeetrilise progressiooni summa valemile¹⁰ seos

$$f_i(x) = x^{t-2} + x^{t-3} + \dots + 1 = (x-1)^{-1}(x^{t-1} - 1),$$

siis kandes selle üle maatriksfunktsioonile, saame $f_i(\mathbf{A}) = (\mathbf{A} - \mathbf{I})^{-1}(\mathbf{A}^{t-1} - \mathbf{I})$.

Funktsiooni $f_i(\mathbf{A})$ leidmiseks kasutame maatriksfunktsiooni leidmise üldist skeemi, leides esmalt maatriksi \mathbf{A} omaväärtused λ_1 ja λ_2 ning seejärel maatriksid \mathbf{Z}_1 ja \mathbf{Z}_2 , mis rahuldavad võrrandeid $f(\mathbf{A}) = f(\lambda_1)\mathbf{Z}_1 + f(\lambda_2)\mathbf{Z}_2$. Maatriksi \mathbf{A} omaväärtused saame võrrandist

$$\begin{vmatrix} \frac{1}{2} - \lambda & \frac{1}{4} \\ 1 & -\lambda \end{vmatrix} = 0 \Rightarrow \lambda^2 - \frac{1}{2}\lambda - \frac{1}{4} = 0 \Rightarrow \lambda_1, \lambda_2 = \frac{1}{4}(1 \pm \sqrt{5})$$

ja maatriksid \mathbf{Z}_1 ja \mathbf{Z}_2 võrrandisüsteemist

$$\begin{cases} \mathbf{Z}_1 + \mathbf{Z}_2 = \mathbf{I} \\ \frac{1}{4}(1 + \sqrt{5})\mathbf{Z}_1 + \frac{1}{4}(1 - \sqrt{5})\mathbf{Z}_2 = \mathbf{A} \end{cases} \Rightarrow \mathbf{Z}_1 = \frac{1}{2\sqrt{5}} \begin{pmatrix} \sqrt{5} + 1 & 1 \\ 4 & \sqrt{5} - 1 \end{pmatrix}, \mathbf{Z}_2 = \frac{1}{2\sqrt{5}} \begin{pmatrix} \sqrt{5} - 1 & -1 \\ -4 & \sqrt{5} + 1 \end{pmatrix}.$$

Maatriksfunktsioon $f_i(\mathbf{A})$ avaldub leitud omaväärtuste ja konstantsete maatriksite kaudu kujul

$$f_i(\mathbf{A}) = \frac{\lambda_1^{t-1} - 1}{\lambda_1 - 1} \mathbf{Z}_1 + \frac{\lambda_2^{t-1} - 1}{\lambda_2 - 1} \mathbf{Z}_2.$$

Maatriksi \mathbf{U}_t esitusest järeldub, et

$$F_t = \mathbf{1} \mathbf{0} \mathbf{U}_t = \mathbf{1} \mathbf{0} f_i(\mathbf{A}) \mathbf{C},$$

millest omakorda tänu maatriksi \mathbf{C} esitusele järeldub, et F_t võrdub $\frac{1}{4}$ -kordse maatriksfunktsiooni $f_i(\mathbf{A})$ vasaku ülemise nurga elemendiga: $F_t = \frac{1}{4}\{f_i(\mathbf{A})\}_{11}$.

Rakendades viimast tõdemust maatriksi \mathbf{A} omaväärtuste λ_1 ja λ_2 ning konstantsete maatriksite \mathbf{Z}_1 ja \mathbf{Z}_2 kaudu avaldatud maatriksfunktsioonile $f_i(\mathbf{A})$, saame inbriidingukoefitsiendi väärtuseks põlvkonnas t

$$F_t = \frac{1}{4} \left(\frac{\lambda_1^{t-1} - 1}{\lambda_1 - 1} \times \frac{\sqrt{5} + 1}{2\sqrt{5}} + \frac{\lambda_2^{t-1} - 1}{\lambda_2 - 1} \times \frac{\sqrt{5} - 1}{2\sqrt{5}} \right),$$

mis peale mõningaid teisendusi on viidav kujule

$$F_t = 1 - \frac{4}{\sqrt{5}} \left[\left(\frac{1 + \sqrt{5}}{4} \right)^{t+2} - \left(\frac{1 - \sqrt{5}}{4} \right)^{t+2} \right]$$

või siis ka kujule

$$\begin{aligned} F_t &= 1 - \frac{2}{5 - 3\sqrt{5}} \left(\frac{1 + \sqrt{5}}{4} \right)^t + \frac{2}{5 + 3\sqrt{5}} \left(\frac{1 - \sqrt{5}}{4} \right)^t \\ &= 1 - 1,17082 \left(\frac{1 + \sqrt{5}}{4} \right)^t + 0,17082 \left(\frac{1 - \sqrt{5}}{4} \right)^t. \end{aligned} \quad (1.12)$$

Järgnevalt püüame jõuda samaväärses esituseni õde-vend ristamissüsteemi Markovi ahela abil modelleerides. Märgime esmalt, et kui p_1, \dots, p_k on alleelide tõenäosused mingis lookuses ja populatsioon on Hardy-Weinbergi tasakaalus, siis on see lookus põlvkonnas t homosügootne tõenäosusega

$$Q_t = F_t + (1 - F_t)(p_1^2 + \dots + p_k^2), \quad (1.13)$$

kus F on inbriidingukoefitsient. Siin $1 - F$ on tõenäosus, et alleelid lookuses ei ole päritolult identsed, ja see tõenäosus on korrutatud tingliku tõenäosusega, et alleelid on juhuslikult identsed. Dialleelsel juhul, kui alleelid on võrdtõenäosed, valem lihtsustub:

$$Q_t = \frac{1}{2}(1 + F_t). \quad (1.14)$$

Analoogselt peatüki 1.5.3 käsitlusele võtame vaatluse alla Markovi ahela kanooniliselt nummerdatud olekud (vanematepaarid) $AAxAA$, $aaxaa$, $AxAa$, $AAxaa$, $AaxAa$ ja $Aaxaa$, mille tõenäosused põlvkonnas t moodustavad olekujaotuse $\mathbf{p}(t)$. Viimane avaldub statsionaarse Markovi ahela eeldusel kujul

$$\mathbf{p}(t) = \mathbf{M}^t \mathbf{p}(0),$$

¹⁰ Geomeetrilise jada, mille üldliige avaldub kujul $a_0 q^k$, esimese n liikme summa avaldub valemiga $a_0 + a_0 q + a_0 q^2 + \dots + a_0 q^{n-1} = (a_0 - a_0 q^n)/(1 - q)$. Tuleneb see geomeetrilise jada lõpmatu summa valemist $\sum_{k=0}^{\infty} a_0 q^k = a_0/(1 - q)$, sest $\sum_{k=0}^{n-1} a_0 q^k = \sum_{k=0}^{\infty} a_0 q^k - \sum_{k=n}^{\infty} a_0 q^k = [a_0/(1 - q)] - [a_0 q^n/(1 - q)] = (a_0 - a_0 q^n)/(1 - q)$.

kus \mathbf{M} on õde-vend ristamissüsteemile vastava Markovi ahela üleminekutõenäosuste maatriks kujul (1.9) ja $\mathbf{p}(0)$ on algolekujaotus. Eeldades, et esimese vanempaari valik toimub juhuslikult ning mõlemad alleelid A ja a on võrdtõenäolised, on algolekujaotus kujul $\mathbf{p}(0) = (\frac{1}{16} \ \frac{1}{16} \ \frac{1}{4} \ \frac{1}{8} \ \frac{1}{4} \ \frac{1}{4})^T$. Järgnevalt toome sisse maatriksi

$$\mathbf{B} = \begin{pmatrix} 1 & 1 & \frac{1}{2} & 1 & 0 & \frac{1}{2} \end{pmatrix},$$

mille elemendid näitavad, kui suure tõenäosusega on mingist vanempaarist juhuslikult valitud indiviid homosügootne – näiteks vanematepaari $AAXaAa$ korral on homosügootse indiviidi tõenäosus $\frac{1}{2}$. On ilmne, et

$$\mathbf{Q}_t = \mathbf{B}\mathbf{p}(t) = \mathbf{B}\mathbf{M}^t\mathbf{p}(0). \quad (1.15)$$

Üleminekutõenäosuste maatriksi \mathbf{M} funktsiooni $f(\mathbf{M}) = \mathbf{M}^t$ leiame üldist maatriksfunktsioonide arvutamise võtet kasutades kujul

$$\mathbf{M}^t = \sum_{i=1}^k (\lambda_i^t \mathbf{Z}_i), \quad (1.16)$$

kus λ_i on maatriksi \mathbf{M} i . mittekordne omaväärtus, \mathbf{Z}_i on sobivalt valitud konstantne maatriks ja k on erinevate omaväärtuste arv. Avaldise (1.16) ja (1.15) kokku võttes saame, et

$$\mathbf{Q}_t = \mathbf{B} \sum_{i=1}^k \lambda_i^t \mathbf{Z}_i \mathbf{p}(0) = \sum_{i=1}^k \lambda_i^t \mathbf{B}\mathbf{Z}_i\mathbf{p}(0). \quad (1.17)$$

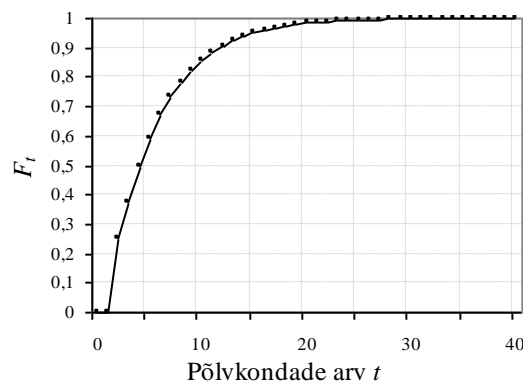
Peatükis 1.5.3 on välja arvatud maatriksi \mathbf{M} viis mittekordset omaväärtust $\lambda_1 = 1$, $\lambda_2 = 1/2$, $\lambda_3 = 1/4$, $\lambda_4 = (1 + \sqrt{5})/4$ ja $\lambda_5 = (1 - \sqrt{5})/4$ ning neile vastavad konstantsed maatriksid \mathbf{Z}_i . Pannes leitud suurused avaldise (1.17) ja viies läbi maatrikskorrutised $\mathbf{B}\mathbf{Z}_i\mathbf{p}(0)$, saame

$$\begin{aligned} \mathbf{Q}_t &= \sum_{i=1}^5 \left[\lambda_i^t \times \mathbf{B}\mathbf{Z}_i\mathbf{p}(0) \right] = 1^t \times 1 + \frac{1}{2}^t \times 0 + \frac{1}{4}^t \times 0 + \left(\frac{1 + \sqrt{5}}{4} \right)^t \times (-0,585541) + \left(\frac{1 - \sqrt{5}}{4} \right)^t \times (0,08541) \\ &= 1 - 0,585541 \left(\frac{1 + \sqrt{5}}{4} \right)^t + 0,08541 \left(\frac{1 - \sqrt{5}}{4} \right)^t. \end{aligned}$$

Seosest (1.14) järeldeb, et $F_t = 2\mathbf{Q}_t - 1$, millest saame seosega (1.12) analoogse valemi:

$$F_t = 1 - 1,17082 \left(\frac{1 + \sqrt{5}}{4} \right)^t + 0,17082 \left(\frac{1 - \sqrt{5}}{4} \right)^t.$$

Joonisel 1.16 on kujutatud inbriidingukoefitsiendi muutus õde-vend ristamissüsteemis põlvkondade viisi – nagu oligi eeldada, läheneb päritolult identsete alleelide tõenäosus kiirelt ühele.



Joonis 1.16. Inbriidingukoefitsiendi suurenemine õde-vend ristamissüsteemis

1.5.6 Geneetilise defekti tõenäosus inbriidisel indiviidil

Inbriiding suurendab oluliselt geneetiliste defektide tõenäosust. Oletame, et genoomis on N lookust, kus igäühes võib esineda väikese tõenäosusega ($q = 0,001$) retsessiivne alleel, mis homosügootsuse korral avaldub geneetilise defektina (st, et ükskõik milline lookus N -st võimalikust on homosügootne, on indiviid geneetiliselt defektne). Lookuste arv N on tõenäoliselt väga suur, näiteks $N = 10000$. Arvutame tõenäosuse, et vaatlusalune indiviid on geneetilise defektiga.

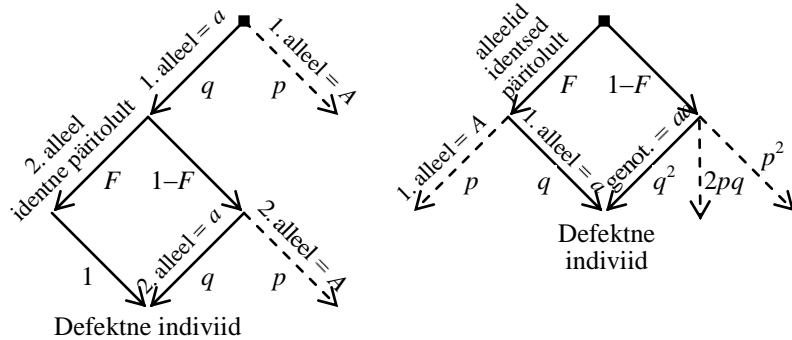
Esmalt eeldame, et populatsioon on Hardy-Weinbergi tasakaalus ja indiviidi inbriidingukoefitsient on null: $F = 0$. Tõenäosus, et mingi konkreetne lookus põhjustab defekti, on Hardy-Weinbergi tasakaalu

tõttu $q^2 = 10^{-6}$. Tõenäosus, et lookus ei põhjusta defekti, on $1 - 10^{-6}$ ja tõenäosus, et ükski lookus ei põhjusta defekti, on lookuste sõltumatuse korral $(1 - 10^{-6})^N$. Seega on defekti tõenäosus indiviidil

$$P_{\text{defekt}} = 1 - (1 - q^2)^N \approx 1 - e^{-Nq^2}, \quad (1.18)$$

antud juhul $P_{\text{defekt}} \approx 0,00995$.

Leiame nüüd defekti tõenäosuse indiviidil, kelle inbriidingukoefitsient $F \neq 0$. Defektne lookus kujuneb teatavate sõltumatute sündmuste samaaegsel toimumisel (vastavaid võimalusi on kujutatud kõrvalolevatel skeemidel). Tõenäosus, et parajasti vaadeldav lookus põhjustab defekti, on kirjutatav kujul $Fq + (1 - F)q^2$ (kas alleelid on päritolult identsed ning samaaegselt on üks juhuslikult valitud alleel geneetilist defekti põhjustav – tõenäosus selliseks sündmuseks on Fq , või ei ole alleelid vaatlusaluses lookuses päritolult identsed, aga on mõlemad geneetilist defekti põhjustavad – tõenäosus selliseks sündmuseks on $(1 - F)q^2$). Analoogselt eespool vaadelduga saame defektse indiviidi tõenäosuseks



Defektne indiviid

$$P_{\text{defekt}} = 1 - [1 - Fq - (1 - F)q^2]^N \approx 1 - e^{-NF(1-q)q - Nq^2}. \quad (1.19)$$

Võttes inbriidingukoefitsiendiks üsna mõõduka $F = 0,1$, saame defektse indiviidi tõenäosuseks $P_{\text{defekt}} \approx 0,635$. Juhul, kui võimalikke geneetilisi defekte põhjustavaid lookusi on rohkem, näiteks 50000, on ka mingi lookuse suhtes defektse indiviidi sündimise tõenäosus märgatavalt suurem, $F = 0,1$ korral $P_{\text{defekt}} \approx 0,9936$ – sellisel juhul on normaalse indiviidi sündimine praktiliselt võimatu.

Valemite (1.18) ja (1.19) võrdlusest tuleneb, et normaalse inbriidse indiviidi sündimise tõenäosus moodustab vaid $e^{-NF(1-q)q}$ osa vastavast tõenäosusest mitteinbriidse indiviidi korral. Meie näites ($q = 0,001$; $N = 10000$; $F = 0,1$) on see suhe $e^{-1} = 0,368$; $N = 50000$ korral aga vaid $e^{-5} = 0,0067$.

Tabelis 1.6 on toodud mõned geneetilise defekti tõenäosused, mis on arvutatud saadud valemitega.

Tabel 1.6. Geneetilise defekti tõenäosus (ümardatud 4. kohani peale koma) sõltuvalt defektse alleeli tõenäosusest lookuses (q) ja indiviidi inbriidingukoefitsiendist (F). Eeldatud on, et lookuste arv on 10000 ($N = 10000$).

Defektse alleeli tõenäosus q	Inbriidingukoefitsient F				
	0	0,01	0,05	0,1	0,25
0,01	0,6321	0,8633	0,9974	1,0000	1,0000
0,005	0,2212	0,5265	0,9353	0,9946	1,0000
0,001	0,0100	0,1041	0,3992	0,6354	0,9185
0,0001	0,0001	0,0100	0,0489	0,0952	0,2213

Ülesanne 11. Genotüübi olulises osas on 1000 lookust. Oletame, et igas lookuses võib tõenäosusega q esineda retsessiivne geneetilist defekti põhjustav alleel. Panmiktilises populatsioonis on defektse indiviidi sündimise tõenäosus 0,01.

Kui suur on selles populatsioonis defekti tõenäosus indiviidil, kelle vanemad on õde ja vend?

1.5.7 Migratsioon

Mingi genotüübi väljaränne (emigratsioon) on samaväärne selle genotüübi (osalise) elimineerimisega populatsioonist. Seega võib väljarännet, mille intensiivsus on võrdeline genotüübi osakaaluga populatsioonis, käsitleda valikuna ja kirjeldada selektsiooniindeksite abil.

Sisserände (immigratsiooni) kirjeldamiseks on üldjuhul tarvis omaette metoodikat. Üks lähenemisviis on järgmine. Olgu alleelide jaotus põhipopulatsioonis $\mathbf{p}_{(1)}$ ja immigrantide hulgas $\mathbf{p}_{(2)}$. Pärast immigratsiooni toimumist on alleelide jaotuseks populatsioonis segu $(1 - \tau)\mathbf{p}_{(1)} + \tau\mathbf{p}_{(2)}$, kus τ ($0 \leq \tau \leq 1$) on sisserände intensiivsus. Käesolevas kursuses sisserände modelleerimist pikemalt ei käsitleta.

¹¹ $\lim_{n \rightarrow \infty} (1 + \frac{x}{n})^n = e^x$

1.5.8 Lõplikud populatsioonid. Alleeli kadumistõenäosus lõplikus populatsioonis

Panmiktiline populatsioon on lõpmata suur, mistõttu sellises populatsioonis on tõenäosused ja suhtelised sagedused ning samuti keskvärtus ja aritmeetiline keskmine võrdsed. Väikestes populatsioonides ilmneb aga juhuslik fluktueerimine¹² (ebaregulaarne võnkumine, **juhuslik triiv**), mis mängib populatsioonigeneetikas olulist rolli. Juhusliku triivi tõttu võib nendel perioodidel, kus populatsioon on väikesearvuline, rida alleele kaduma minna, kuid võib ka juhtuda, et haruldased alleelid muutuvad tavaliseks (nn **pudelikaela efekt**).

Alleeli kadumine võib üldjuhul tähendada kas alleeli täielikku füüsilist kadumist populatsioonist või siis alleeli esinemistõenäosuse lähenemist nullile. Järgnevalt uurime kadumist esimeses tähenduses. Jälgime mingi alleeli a paljunemist (antud juhul siis mitte lõpmatu hulga alleelide hulgast juhuslikult valitud alleeli paljunemist, vaid ühe konkreetse alleeli, mida on populatsioonis lõplik hulk – võib olla ka ainueksplar – paljunemist) ehk siis seda, kui mitu alleeli a tekib järglaspõlvkonda ühest konkreetsest vanempõlvkonna alleelist a . Olgu X_0, X_1, X_2, \dots alleeli a arvud (kordsused) järjestikustes põlvkondades (X_i on täisarvulised mittenegatiivsed juhuslikud suurused, mis esitavad Markovi ahela järjestikuseid seisundeid ehk olekuid; $X_0 = 1$, sest vt ühte konkreetset alleeli). Seejuures on alleeli a kordsus põlvkonnas $n + 1$ esitatav kujul

$$X_{n+1} = \sum_{r=1}^{X_n} \xi_r,$$

kus sõltumatud juhuslikud suurused ξ_r tähistavad vanempõlvkonda (antud juhul siis põlvkonda n) kuuluva r -nda alleeli a järglaste arvu järglaspõlvkonnas (antud juhul siis põlvkonnas $n + 1$). Tähistame $P(\xi_r = k) = p_k$ ($k = 0, 1, 2, \dots$) ja oletame, et alleelil a on positiivse tõenäosusega järglasi, samas kui positiivse tõenäosusega võivad järglased ka puududa, st

$$0 < p_0 < 1 \quad (1.20)$$

(ehk teisisõnu, mingi väljasuremistõenäosus eksisteerib – $p_0 > 0$, aga samas on šanss ka mitte välja surra – $p_0 < 1$).

Defineerime tõenäosusi genereerivad funktsioonid¹³ ξ_r jaoks:

$$\varphi(s) := \varphi_{\xi_r}(s) = \sum_{k=0}^{\infty} P(\xi_r = k) s^k, \quad (1.21)$$

ja X_n ($n = 0, 1, 2, \dots$) jaoks:

$$\varphi_n(s) = \sum_{k=0}^{\infty} P(X_n = k) s^k. \quad (1.22)$$

Seose nende kahe tõenäosusi genereeriva funktsiooni vahel saame järgmiselt (põlvkonna $n + 1$ jaoks):

$$\begin{aligned} \varphi_{n+1}(s) &= \sum_{k=0}^{\infty} P(X_{n+1} = k) s^k = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} P(X_{n+1} = k \mid X_n = j) P(X_n = j) s^k \\ &= \sum_{k=0}^{\infty} s^k \sum_{j=0}^{\infty} P(X_n = j) P(\xi_1 + \dots + \xi_j = k) \\ &= \sum_{j=0}^{\infty} P(X_n = j) \sum_{k=0}^{\infty} P(\xi_1 + \dots + \xi_j = k) s^k \\ &= \sum_{j=0}^{\infty} P(X_n = j) [\varphi(s)]^j = \varphi_n[\varphi(s)] = \varphi_n[\varphi(s)], \end{aligned}$$

kus viimase rea kirjutamisel on arvestatud, et sõltumatute juhuslike suuruste summa tõenäosusi genereeriv funktsioon on liidetavate tõenäosusi genereerivate funktsioonide korrutis:

$$\sum_{k=0}^{\infty} P(\xi_1 + \dots + \xi_j = k) s^k = \varphi_{\xi_1 + \dots + \xi_j}(s) = \varphi_{\xi_1}(s) \cdot \dots \cdot \varphi_{\xi_j}(s) = [\varphi(s)]^j,$$

seejärel on kasutatud definitsiooni (1.22) ning lõpuks ilmset võrdust $\varphi_1(s) = \varphi(s)$. Leitud seos võimaldab alandada $\varphi_n(s)$ indeksit, samal ajal suurendades argumendis oleva $\varphi_n(s)$ indeksit. Seost korduvalt rakendades jõuame lõpuks ilusa valemieni

¹² fluktueerima <-eerida, -eerin> voogama, kõikumama, vahelduma; med võdisema. ▪ Fluktueeriv kiirgus. Fluktueeriv kõht (nt vesitõve puhul) [ÕS 2006].

¹³ Juhusliku suuruse $X = 0, 1, 2, \dots$ **tõenäosusi genereeriv funktsioon** on tõenäosusfunktsiooni $p_k = P(X = k)$ esitus astmereana kujul $\varphi(t) = E(t^X) = \sum_{k=0}^{\infty} p_k t^k$.

Mõningad tõenäosusi genereeriva funktsiooni omadused.

1) Juhusliku suuruse X tõenäosusfunktsioon avaldub kujul $p_k = \varphi^{(k)}(0)/k!$.

2) Juhusliku suuruse X keskvärtus avaldub kujul $E[X] = \varphi^{(1)}(1)$ ja k . järku faktoriaalne moment kujul $E[X!/(X - k)!] = E[X(X - 1)\dots(X - k + 1)] = \varphi^{(k)}(1)$.

3) Kui X_1, X_2, \dots, X_n on sõltumatud (ja mitte ilmtingimata sama jaotusega) juhuslikud suurused ja $S_n = \sum_{i=1}^n a_i X_i$, kus a_i -d on konstandid, siis $E(t^{S_n}) = E(t^{\sum_{i=1}^n a_i X_i}) = \varphi_{S_n}(t) = \varphi_{X_1}(t^{a_1}) \varphi_{X_2}(t^{a_2}) \dots \varphi_{X_n}(t^{a_n})$.

$$\varphi_{n+1}(s) = \varphi[\varphi_n(s)], \quad (1.23)$$

millest omakorda järeldub, et $\varphi_n(s)$ on teisenduse φ n -kordne rakendus argumentidele s .

Tähistame nüüd lühiduseks $q_n = P(X_n = 0)$ (q_n on tõenäosus, et n -ndas põlvkonnas alleeli a enam ei ole). Tõenäosusi genereerivate funktsioonide omadustest tulenevalt saame valemist (1.22), et

$$q_n = \lim_{s \rightarrow 0} \varphi_n(s) = \varphi_n(0).$$

Selle alusel

$$q_{n+1} = \varphi_{n+1}(0) = \varphi(\varphi_n(0)) = \varphi(q_n). \quad (1.24)$$

Et $\varphi(s)$ on rangelt kasvav funktsioon (sest (1.20) tõttu $p_0 < 1$, mistõttu $p_1 + p_2 + \dots = \sum_{i=0}^{\infty} p_i - p_0 > 0$, kusjuures kordajad valemis (1.21) on mittenegatiivsed) ja $q_1 = \varphi(0) = p_0 > 0$, siis

$$q_2 = \varphi(q_1) > \varphi(0) = q_1 > 0.$$

Siit induktsiooniga jätkates saame (1.24) tõttu rangete võrratuste ahela $0 < q_1 < q_2 < \dots < 1$. Seega peab leiduma piirväärtus $\pi = \lim_{n \rightarrow \infty} q_n$ ($0 < \pi \leq 1$), mis ongi tõenäosus, et alleel a millalgi kaob populatsioonist.

Et tõenäosusi genereeriv funktsioon $\varphi(s)$ on lõigul $[0, 1]$ pidev, siis saame valemi (1.24) alusel

$$\pi = \lim_{n \rightarrow \infty} \varphi(q_{n-1}) = \varphi(\lim_{n \rightarrow \infty} q_n) = \varphi(\pi).$$

Siit näeme, et π on võrrandi

$$\varphi(s) = s \quad (1.25)$$

lahend. Osutub, et π on selle võrrandi minimaalne positiivne lahend. Tõepoolest, olgu s_0 suvaline positiivne lahend. Siis $q_1 = \varphi(0) < \varphi(s_0) = s_0$. Induktiivselt, olgu juba tõestatud, et $q_n < s_0$. Siis (1.24) tõttu $q_{n+1} = \varphi(q_n) < \varphi(s_0) = s_0$ ja järelikult iga n korral $q_n < s_0$. Kuid $\pi = \lim_{n \rightarrow \infty} q_n \leq s_0$, π on võrrandi (1.25) vähim positiivne lahend.

Valemi alleeli a keskmise järglaste arvu leidmiseks saame järgmise arutelu tulemusena: tõenäosusi genereeriva funktsiooni (1.21) esimene tuletis avaldub kujul $\varphi'(s) = \sum_{k=0}^{\infty} k p_k s^{k-1}$, samas on üldteada, et diskreetse juhusliku suuruse ξ keskvärtus avaldub summamana $E(\xi) = \sum_{k=0}^{\infty} k p_k$, järelikult $E(\xi) = \varphi'(1)$.

Funktsiooni $y = \varphi(s)$ tuletis kohal $s=1$ määrab ka selle, kas $\varphi(s)$ graafik lõikab $s < 1$ korral diagonaali $\varphi(s) = s$ või mitte. Siit saame järelduse: alleeli allesjäämise tõenäosus on nullist suurem ainult siis, kui alleeli keskmine järglaste arv on suurem kui 1.

Näide. Uurime olukorda, kus mingi alleel võib järgmises põlvkonnas esineda sagedusega 0, 1, 2 või 3 võrdse sagedusega $\frac{1}{4}$, st alleeli a järglaste arvu ξ jaotustabel on kujul

k	0	1	2	3
$P(\xi = k)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Leiame, kui suure tõenäosusega kaovad selle alleeli järglased.

Järglaste arvu tõenäosusi genereeriv funktsioon on kujul

$$\varphi(s) = \frac{1}{4}s^0 + \frac{1}{4}s^1 + \frac{1}{4}s^2 + \frac{1}{4}s^3,$$

mis on vastavalt geomeetrilise jada liikmete summa valemile ümber kirjutatava ka kujul

$$\varphi(s) = \frac{1}{4}(1 - s^4)/(1 - s).$$

Väljasuremistõenäosus π on võrrandi $\frac{1}{4}(1 - s^4)/(1 - s) = s$ vähim positiivne lahend, milleks tuleb $\pi \approx 0,4$.

Keskmise järglaste arvu võime leida nii tõenäosusi genereeriva funktsiooni tuletisest kui ka otse jaotustabelist. Esimesel juhul saame, et

$$E(\xi) = \varphi'(1) = \frac{1}{4} + \frac{1}{2} \cdot 1 + \frac{3}{4} \cdot 1^2 = \frac{9}{4} = 1,5;$$

teisel juhul aga

$$E(\xi) = \frac{1}{4}(0 + 1 + 2 + 3) = \frac{6}{4} = 1,5.$$

