

3.3 Aheldusanalüüs kvantitatiivsete tunnuste korral

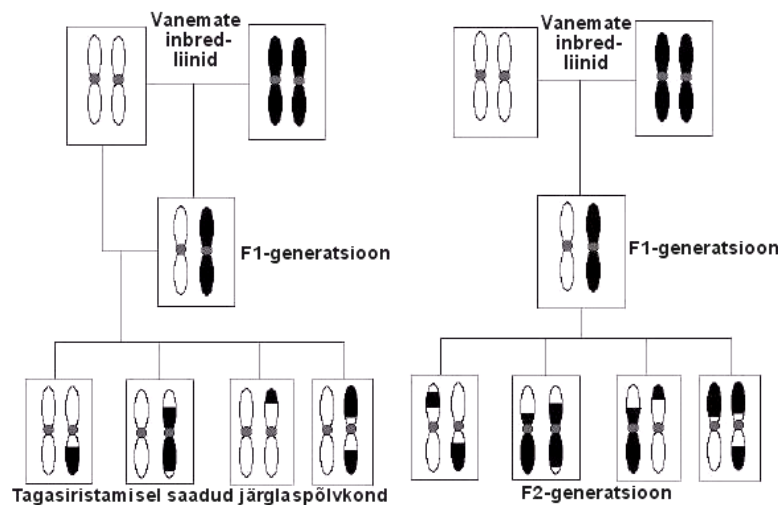
Kvantitatiivsete (arvuliste, mõõdetavate) tunnuste korral on pisut keeruline ette kujutada geeni, mis selle väärtused üksinda ära määraks. Õieti on see kvantitatiivse tunnuse korral võimatu, sest kvantitatiivsel tunnusel on teoreetiliselt lõpmatu hulk (reaalsuses vähemalt suur hulk) erinevaid väärtusi, millele peaks siis vastama ka lõpmatu hulk erinevaid allelele ühes geenilookuses. Pigem on mõistlik arvata, et suure hulga geneetiliste ja mittegeneetiliste mõjutegurite hulgas võivad leiduda mõned (geenid), mille mõju on pisut suurem ja silmatorkavam. Selliseid suhteliselt suurem mõjuga geene nimetatakse **põhigeenideks** (*major genes*) ja nende asukohta **QTL**-ks (kvantitatiivse tunnuse lookuseks, *quantitative trait loci*).

Sageli kasutatakse QTL-i kandidaatide (põhigeenide) leidmiseks ristamiseksperimente või valiktakse analüüsitavad isendid kindlate sugulussidemete alusel.

3.3.1 Marker-gruppide vahelise erinevuse testimine dispersioonanalüüsiga

Enamus põhigeeni leidmiseks korraldatud eksperimente võtavad aluseks kaks inbredliini (st, et vanemateks valitakse täielikult homosügootsed isendid). Selliste vanemate ristamisel saame geneetiliselt identsed heterosügootsed järglased (põlvkond F_1). Olgu oletatavaks uuritavat tunnust märkimisväärselt mõjutavaks põhigeeni lookuseks e QTL-ks dialleelne lookus Q (alleelid vastavalt Q_1 ja Q_2) ning vaadeldud indiviididel määratud markerlookuseks dialleelne lookus A (alleelid vastavalt A_1 ja A_2). Inbredliinidest pärit homosügootsed vanemad omavad siis vastavalt genotüüpe $Q_1A_1|Q_1A_1$ ja $Q_2A_2|Q_2A_2$ ning nende ristamisel saadud F_1 -põlvkonna järglased genotüüpe $Q_1A_1|Q_2A_2$.

Selgitamaks, kas lookused A ja Q on aheldunud, analüüsitakse andmeid fenotüübi ja markeri(te) kohta kas põlvkonnas F_2 või tagasiristamise teel saadud järglastel. Mõlemad need ristamisskeemid on kujutatud skemaatiliselt joonistel 16 ja 17.

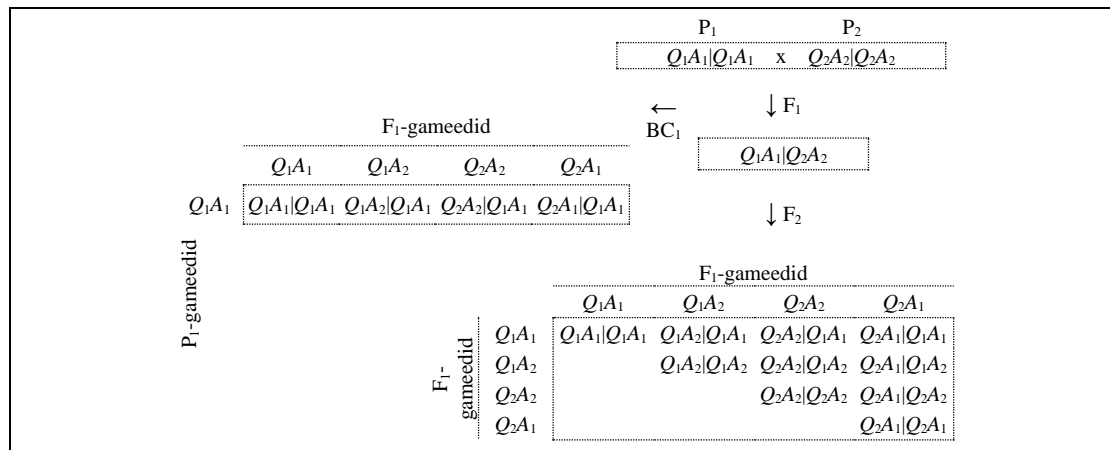


Joonis 16. Tagasiristamisega ja tavaline ristamisskeem (joonistel on kujutatud vaid üht homologsete kromosoomide paari ja 4 analüüsitavat järglast – reaalses uuringutes on vaatluse all kogu genotüüp ja üle 100 järglase).

Tagasiristamise (*backcrossing*) korral ristatakse F_1 -indiviidid ühega vanemliinidest. Selle tagajärjel on järglaspõlvkonna üks kromosoomidest identne ristamisel kasutatud inbredliinist pärit vanema kromosoomiga, teine kromosoom on aga tänu ristsirdele kahe inbredliini kombinatsioon.

Tavalise ristamisskeemi (*intercrossing*) korral ristatakse omavahel põlvkonna F_1 isendeid. Generatsiooni F_2 indiviidide genotüüp on kombinatsioon oma vanemate genotüüpidest, kusjuures iga üksiku genotüübi esinemise sagedus sõltub QTL-i ja markerlookuse vahelisest rekombinatsioonimäärast r .

Järgnevalt tuletame konkreetseid teste vaid tagasiristamisega eksperimendi tarvis, sest selle puhul on võrreldavaid kombinatsioone vähem ja valemite kirjanepanek lihtsam.



Joonis 17. Kahe, markerlookuse A ja QTL-i Q suhtes homosügootse, vanemliini ristamine – F₁-põlvkonnas on kõik individid geneetiliselt identsed heterosügootid genotüübiga Q₁A₁|Q₂A₂, F₂-põlvkonnas leidub 16 erinevat ja tagasiristamisel saadud BC₁-generatsioonis 4 erinevat genotüüpi.

Olgu meil ristatud individid inbredliinist genotüübiga Q₁A₁|Q₁A₁ ja F₁-generatsioonist genotüübiga Q₁A₁|Q₂A₂. Kõik järglaspõlvkonna isendid pärivad tõenäosusega üks gameedi Q₁A₁. Generatsioonist F₁ pärit vanemalt saadakse alleelid Q₁A₁ tõenäosusega 1/2(1 - r), alleelid Q₁A₂ tõenäosusega r/2, alleelid Q₂A₁ tõenäosusega r/2 ja alleelid Q₂A₂ tõenäosusega 1/2(1 - r). Tagasiristamisel saadud individid võib jagada kahte gruppi: markergenotüübiga A₁|A₁ ja markergenotüübiga A₁|A₂. Neist esimesed omavad QTL-genotüüpi Q₁|Q₁ tõenäosusega 1 - r ja QTL-genotüüpi Q₁|Q₂ tõenäosusega r, teised analoogselt genotüüpi Q₁|Q₁ tõenäosusega r ja genotüüpi Q₁|Q₂ tõenäosusega 1 - r. Marker-genotüübiga A₁|A₁ individide oodatav fenotüübiväärtus avaldub siis seosena

$$\mu_{A_1|A_1} = (1 - r)\mu_{Q_1|Q_1} + r\mu_{Q_1|Q_2}$$

kus $\mu_{Q_1|Q_1}$ märgib uuritava tunnuse keskvaartust individidel QTL-genotüübiga Q₁|Q₁ ja $\mu_{Q_1|Q_2}$ individidel QTL-genotüübiga Q₁|Q₂. Samaselt on markergenotüübiga A₁|A₂ individide oodatav fenotüübiväärtus esitatav seosena

$$\mu_{A_1|A_2} = r\mu_{Q_1|Q_1} + (1 - r)\mu_{Q_1|Q_2} .$$

Nende oodatavate fenotüübiväärtuste erinevus avaldub kujul

$$\mu_{A_1|A_1} - \mu_{A_1|A_2} = \underbrace{(\mu_{Q_1|Q_1} - \mu_{Q_1|Q_2})}_{\delta}(1 - 2r) = \delta(1 - 2r) . \tag{3.1}$$

Kui QTL ja marker ei ole aheldunud, siis $r = 1/2$ ja $\mu_{A_1|A_1} - \mu_{A_1|A_2} = 0$. Sellest lähtuvalt on loomulik marker-gruppide võrdlemisel kontrollida hüpoteeside paari H₀: $\mu_{A_1|A_1} - \mu_{A_1|A_2} = 0$, H₁: $\mu_{A_1|A_1} - \mu_{A_1|A_2} \neq 0$.

Ligikaudu sobib toodud hüpoteeside kontrollimiseks tavaline t-test (ligikaudu seepärast, et uuritava tunnuse jaotuseks võrreldavais gruppides on segujaotus). Et vähegi keerulisemate ristamissüsteemide ja/või enam kui kahe erineva alleeliga lookuste korral on võrreldavaid marker-gruppe rohkem kui kaks, kasutatakse keskmiste võrdlemiseks enamasti dispersioonanalüüsi (siit ka kirjeldatud meetodi populaarsem nimetus – **ANOVA-** e **dispersioonanalüüsi meetod**).

Kirjeldatud meetodi peamine probleem on, et statistikameetoditega tuvastatud erinevus ei ole geneetika mõistes üheselt tõlgendatav – mitmed $\delta = \mu_{Q_1|Q_1} - \mu_{Q_1|Q_2}$ ja r väärtused vastavad samale vahele $\mu_{A_1|A_1} - \mu_{A_1|A_2}$. Lisaks omistatakse kõigil juhtudel QTL-i asukoht markerlookuse kohta, mistõttu markerlookusest suhteliselt kaugel paiknev märkimisväärse mõjuga geen võib jääda tuvastamata või siis saada lokaliseeritud suure veaga.

□ Võimsuse ja valimi mahu arvutamisest marker-gruppide vahelise erinevuse dispersioonanalüüsiga testimisel

Leidmaks eelnevalt kirjeldatud viisil konstrueeritud testi võimsust, eeldame, et QTL-genotüübiga Q₁|Q₁ individide fenotüübijaotus on normaaljaotus parameetritega $\mu_{Q_1|Q_1}$ ja σ^2 ning genotüübiga Q₁|Q₂ individide fenotüübijaotus normaaljaotus parameetritega $\mu_{Q_1|Q_2}$ ja σ^2 .

Tähistame markergenotüübiga A₁|A₁ indiviidi j fenotüübiväärtust y_{1j} ja markergenotüübiga A₁|A₂ indiviidi j fenotüübiväärtust y_{2j}. Esimese marker-grupi sisest varieeruvust näitav dispersioon on siis

esitatav QTL-genotüüpidega $Q_1|Q_1$ ja $Q_1|Q_2$ isendite gruppide sisese ja vahelise varieeruvuse summana:

$$\text{Var}(y_{1j}) = E[\text{Var}(y_{1j} | \text{QTL-genotüüp})] + \text{Var}[E(y_{1j} | \text{QTL-genotüüp})]. \quad (3.2)$$

Selle summa esimene liidetav (QTL-gruppide sisene varieeruvus) on

$$E[\text{Var}(y_{1j} | \text{QTL-genotüüp})] = \sigma^2(1-r) + \sigma^2 r = \sigma^2$$

ja teine liidetav (QTL-gruppide vaheline varieeruvus)

$$\begin{aligned} \text{Var}[E(y_{1j} | \text{QTL-genotüüp})] &= E[E(y_{1j} | \text{QTL-genotüüp})^2] - \{E[E(y_{1j} | \text{QTL-genotüüp})]\}^2 \\ &= [(1-r)\mu_{Q_1|Q_1}^2 + r\mu_{Q_1|Q_2}^2] - [(1-r)^2\mu_{Q_1|Q_1}^2 + 2r(1-r)\mu_{Q_1|Q_1}\mu_{Q_1|Q_2} + r^2\mu_{Q_1|Q_2}^2] \\ &= (1-r)r\mu_{Q_1|Q_1}^2 + (1-r)r\mu_{Q_1|Q_2}^2 - (1-r)2r\mu_{Q_1|Q_1}\mu_{Q_1|Q_2} \\ &= (\mu_{Q_1|Q_1}^2 - 2\mu_{Q_1|Q_1}\mu_{Q_1|Q_2} + \mu_{Q_1|Q_2}^2)(1-r)r = (\mu_{Q_1|Q_1} - \mu_{Q_1|Q_2})^2(1-r)r = \delta^2(1-r)r. \end{aligned}$$

Seega on dispersioon (3.2) avaldatav kujul

$$\text{Var}(y_{1j}) = \sigma^2 + \delta^2(1-r)r. \quad (3.3)$$

Analoogselt on esitatav ka marker-grupi $A_1|A_2$ sisene dispersioon. Marker-gruppide keskmiste vahelise erinevuse jaotus on siis ligikaudu normaaljaotus:

$$\bar{y}_1 - \bar{y}_2 \sim N(\delta(1-2r), 2[\sigma^2 + \delta^2(1-r)r]/n),$$

kus n on indiviidide arv igas grupis. Juhul, kui n on suur, on t-jaotus lähendatav normaaljaotusega ja testi võimsuse leidmiseks võime rakendada normaaljaotusele vastavat teststatistikut

$$Z = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{2[\sigma^2 + \delta^2(1-r)r]/n}}.$$

Testi võimsus on leitav kui tõenäosus $P(Z > z_{\alpha/2})$, kus z_{α} on normaaljaotuse α -täiendkvantiil, st $P(Z > z_{\alpha}) = \alpha$. Sisuka hüpoteesi kehtimise korral on statistiku keskväärtus esitatav seosena

$$E(Z) = \frac{\delta(1-2r)}{\sqrt{2[\sigma^2 + \delta^2(1-r)r]/n}}. \quad (3.4)$$

Testi võimsus on siis esitatav tõenäosusena

$$P[Z - E(Z) > z_{\alpha/2} - E(Z)].$$

Kuna vahe $Z - E(Z)$ on standardse normaaljaotusega, siis peab võimsuse $1 - \beta$ saavutamiseks kehtima võrdus $z_{\alpha/2} - E(Z) = z_{1-\beta} = -z_{\beta}$. Asendades viimases võrduses $E(Z)$ suhtega (3.4) ja lahendades saadud võrrandi n suhtes, saame võimsuse $1 - \beta$ jaoks vajalikud gruppide suurused:

$$n = (z_{\alpha/2} + z_{\beta})^2 \frac{2[\sigma^2 + \delta^2(1-r)r]}{\delta^2(1-2r)^2}. \quad (3.5)$$

Näide. Olgu meil eksperiment läbi viidud sarnaselt joonisel 17 kujutatud skeemile. Geenide aditiivne efekt põlvkonnas F_2 on kirjeldatav aditiivdispersiooni σ_a^2 abil. Tähistagu p QTL-i osa generatsiooni F_2 aditiivgeneetilises varieeruvuses. Märkus: vanempõlvkonnad (P_1 ja P_2) on täielikult homosügootsed, põlvkonnas F_2 on aga juba 50% isendeid heterosügootid – seega on QTL-st põhjustatud varieeruvus põlvkonnas F_2 poole suurem võrreldes vanempõlvkondadega. QTL-st tingitud osa aditiivgeneetilisest varieeruvusest avaldub siis suhtena

$$p = \frac{1}{2} \delta^2 / \sigma_a^2,$$

kus teatavasti $\delta = \mu_{Q_1|Q_1} - \mu_{Q_1|Q_2}$. Siis

$$\sigma_a^2 = \frac{1}{2} \delta^2 / p.$$

Uuritava tunnuse päritavus h^2 generatsioonis F_2 on

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2},$$

kus $\sigma_e^2 = \sigma_a^2(1-h^2)/h^2$ on keskkonna mõjudest tingitud varieeruvus. Aditiivgeneetiline varieeruvus tagasiristamisel saadud isendite vahel on poole väiksem võrreldes F_2 -generatsiooniga. Seega, kuna geenide aditiivsest efektist tingitud varieeruvus on QTL-i olemas-

olul vastavalt väiksem: $\sigma_a^2 - p \times \sigma_a^2 = \sigma_a^2 - \frac{1}{2}\delta^2$, on koguvarieeruvus põlvkonnas BC₁ esitatav summana

$$\sigma^2 = \sigma_e^2 + \frac{1}{2}(\sigma_a^2 - \frac{1}{2}\delta^2).$$

Asendades σ_e^2 avaldisega $\sigma_e^2 = \sigma_a^2(1-h^2)/h^2$, saame viimatisele dispersioonile kuju

$$\sigma^2 = \frac{1}{2}\delta^2 \frac{(1-h^2)}{ph^2} + \frac{1}{4}\delta^2 \left(\frac{1}{p} - 1\right) = \delta^2 \left[\frac{1}{2} \frac{(1-h^2)}{ph^2} + \frac{1}{4} \left(\frac{1}{p} - 1\right) \right].$$

Võrduse (3.5) saame nüüd oma näite jaoks esitada seosena

$$n = 2 \frac{(z_{\alpha/2} + z_{\beta})^2}{(1-2r)^2} \left[\frac{1}{2} \frac{(1-h^2)}{ph^2} + \frac{1}{4} \left(\frac{1}{p} - 1\right) + r(1-r) \right].$$

Tagasiristamisega eksperimendis põhigeeni tuvastamise võimsuse 0,9 saavutamiseks vajalikud valimi mahud on toodud järgnevas tabelis (tunnuse päritavuseks on võetud $h^2 = 0,25$).

<i>r</i>	<i>P</i>		
	0,04	0,08	0,16
0	1828	909	449
0,05	2260	1125	557
0,1	2863	1426	708
0,2	5097	2543	1266
0,4	45959	22974	11482

3.3.2 Suurima tõepära meetod

Nagu oli näha valemist (3.1), pole marker-gruppide fenotüübierinevuste alusel QTL-efekti võimalik üheselt hinnata. Üks võimalus probleemi lahendamiseks on kirjutada välja andmetele vastav tõepära-funktsioon ja maksimeerida see paralleelselt nii oletatavale QTL-genotüübile vastavate jaotusparameetrite kui ka rekombinatsioonimäära suhtes.

Vaatleme lihtsuse mõttes jällegi tagasiristamisega skeemi ning otsime dialleelset QTL-i. Eeldame, et QTL-genotüübiga $Q_1|Q_1$ indiviidide fenotüübijaotus on normaaljaotus parameetritega $\mu_{Q_1Q_1}$ ja σ^2 ning QTL-genotüübiga $Q_1|Q_2$ indiviidide fenotüübijaotus normaaljaotus parameetritega $\mu_{Q_1Q_2}$ ja σ^2 . Siis on nende indiviidide, kes on markerlookuse suhtes homosügooidid $A_1|A_1$, fenotüübijaotus esitatav kahe normaaljaotuse seguna:

$$f_1(y | \mu_{Q_1Q_1}, \mu_{Q_1Q_2}, \sigma, r) = (1-r) \times \phi\left(\frac{y - \mu_{Q_1Q_1}}{\sigma}\right) + r \times \phi\left(\frac{y - \mu_{Q_1Q_2}}{\sigma}\right),$$

kus r märgib markerlookuse ja QTL-i vahelist rekombinatsioonimäära ja ϕ tähistab standardse normaaljaotuse tihedusfunktsiooni. Analoogselt on markerlookuse suhtes heterosügootsete $A_1|A_2$ indiviidide fenotüübijaotus esitatav seguna:

$$f_2(y | \mu_{Q_1Q_1}, \mu_{Q_1Q_2}, \sigma, r) = (1-r) \times \phi\left(\frac{y - \mu_{Q_1Q_1}}{\sigma}\right) + r \times \phi\left(\frac{y - \mu_{Q_1Q_2}}{\sigma}\right).$$

Tähistades indiviidi i fenotüübi y_i ja tuues sisse muutuja x_i , mis võrdub 1-ga, kui i indiviid on markerlookuse suhtes homosügootne, ja 0-ga markerlookuse suhtes heterosügootse indiviidi korral, on andmetele vastav tõepärafunktsioon avaldatav korrutisena

$$L(\theta, r | y, x) = \prod_i [f_1(y_i | \theta, r)]^{x_i} [f_2(y_i | \theta, r)]^{1-x_i},$$

kus $\theta = (\mu_{Q_1Q_1}, \mu_{Q_1Q_2}, \sigma)$.

Saadud tõepärafunktsioon maksimeeritakse esmalt θ suhtes, andes selleks rekombinatsioonimäärale r mingi esialgse väärtuse; seejärel, kasutades leitud suurima tõepära hinnangut $\hat{\theta}$, avaldatakse tõepärafunktsiooni maksimum parameetri r suhtes ja lahendatakse, saades tulemuseks suurima tõepära hinnangu \hat{r} ; leitud rekombinatsioonimäära hinnangut kasutatakse omakorda täpsustamiseks parameetri θ hinnangut, jne kuni protseduur koondub.

Hüpoteesi markerlookuse ja QTL-i ahelduse kohta kontrollitakse tõepärasuhte testi abil, võrreldes ülaltoodud ühe oletatava põhigeeni ja markeri ahelduse eeldusel konstrueeritud mudelit nullhüpoteesile vastava mudeliga, kus kõigi indiviidide fenotüübiväärtused eeldatakse jaotuvat vastavalt normaaljaotusele parameetritega μ ja σ ning tõepärafunktsioon avaldub korrutisena

$$L_0(\theta_0 | y) = \prod_i \phi((y_i - \mu)/\sigma),$$

kus $\theta_0 = (\mu, \sigma)$. *LOD*-skoor, mille täisarvulised väärtused 1, 2, 3 jne vastavad “ümmargustele” tõepärasuhetele 10, 100, 1000 jne, ning mille väärtust 3 loetakse geneetikas enamasti piirväärtuseks, millest suurema *LOD*-skoori väärtuse korral loetakse QTL leituks, avaldub kujul

$$LOD(r) = \log_{10} \left[\frac{\max L(\theta, r | y, x)}{L_0(\theta_0 | y)} \right]. \quad (3.6)$$

3.3.3 Regressioonanalüüs

QTL-i lokaliseerimiseks eksperimentaalsel ristamisel vahest kõige enam kasutatav meetod seisneb lühidalt järgnevas: fikseeritakse oletatava põhigeeni asukoht markerite vahel ning püütakse seejärel lähendada fenotüübil mõõdetud tunnuse väärtusi oletatavatesse QTL-i gruppidesse kuulumise tõenäosustega, hinnates selleks regressioonanalüüsi abil erinevatele QTL-genotüüpidele vastavaid oodatavaid fenotüübiväärtuseid. Tehes sama läbi kõigi markerite vaheliste piirkondade kohta, saame tulemuseks kõige tõenäolisema põhigeeni asukoha – intervalli, kus regressioonanalüüsi jääkide ruutude summa on minimaalne.

Võtame vaatluse alla kaks dialleelset markerlookust *A* ja *B* ning oletatava dialleelse QTL-i *Q* nende vahel (selliste oletatavat QTL-i sisaldava kromosoomiintervalli otstes paiknevate markerite tarvis on ingliskeelses kirjanduses kasutusel spetsiaalne nimetus – *flanking markers* – **piiravad markerid**).

Indiviidid inbrediinidest P_1 ja P_2 on vastavalt genotüüpidega $A_1Q_1B_1|A_1Q_1B_1$ ja $A_2Q_2B_2|A_2Q_2B_2$. Tähistame rekombinatsioonimäära lookuste *A* ja *Q* vahel r_A , lookuste *Q* ja *B* vahel r_B ning lookuste *A* ja *B* vahel r_{AB} . Seejuures eeldatakse viimane teadaolevaks, st et toetutakse juba teadaolevale geneetiliselt kaardile. Et rekombinatsioon intervallis *A*-st *B*-ni on fikseeritav vaid juhul, kui see leidis aset kas *A*-st *Q*-ni või *Q*-st *B*-ni, aga mitte mõlemal juhul korraga, siis kehtib järgmine rekombinatsioonimäärade vaheline seos: $r_{AB} = r_A + r_B - 2 r_A r_B$.

Tähistame QTL-genotüüpidele vastavad oodatavad fenotüübiväärtused $E(y | Q_1Q_1) = \mu_{Q_1Q_1}$ ja $E(y | Q_1Q_2) = \mu_{Q_1Q_2}$ ning eeldame ka uuritava tunnuse võrdset varieeruvust mõlema genotüübi korral.

Generatsioonis F_1 on kõik indiviidid genotüübiga $A_1Q_1B_1|A_2Q_2B_2$, nende ristamisel vanemliiniga P_1 saadavas generatsioonis BC_1 on võimalik leida neli erinevat marker- ja kaks QTL-genotüüpi. Iga järglane omab ühes kromosoomis P_1 -vanemalt saadud alleelikomplekti $A_1Q_1B_1$, teises kromosoomis paiknev F_1 -vanemalt päritud alleelikomplekt sõltub juhusest ja lookustevahelistest võimalikest ristsiiretest. Oletades, et järglane pärrib teisena markeralleelid A_2 ja B_1 , saab ta tõenäosusega $r_A(1-r_B)$ ka alleeli Q_1 – selleks peab toimuma rekombinatsioon markerlookuse *A* ja QTL-i *Q* vahel (tõenäosus r_A) ning ei tohi toimuda rekombinatsiooni lõigus *Q*-st *B*-ni (tõenäosus $1-r_B$). Tingliku tõenäosusena esitatuna on seos kujul

$$P \left(\frac{A_1Q_1B_1}{A_2Q_1B_1} \mid \frac{A_1Q_1B_1}{A_2B_1} \right) = \frac{r_A(1-r_B)}{r_A(1-r_B) + (1-r_A)r_B} = \frac{r_A(1-r_B)}{r_{AB}}.$$

Analoogselt on leitav ka tõenäosus, et järglane pärrib lisaks generatsiooni P_1 -vanemalt saadud QTL-alleelile Q_1 generatsiooni F_1 -vanemalt koos markeralleelidega A_2 ja B_1 QTL-alleeli Q_2 :

$$P \left(\frac{A_1Q_1B_1}{A_2Q_2B_1} \mid \frac{A_1Q_1B_1}{A_2B_1} \right) = \frac{(1-r_A)r_B}{r_A(1-r_B) + (1-r_A)r_B} = \frac{(1-r_A)r_B}{r_{AB}}.$$

Samamoodi on tinglikud tõenäosused leitavad kõigi nelja võimaliku marker-genotüübi ja kahe QTL-genotüübi korral (seda muidugi eeldades interferentsuse puudumist – vastasel juhul ei saaks markerite *A* ja *B* vahelist rekombinatsioonimäära esitada markerlookuste ja QTL-i vaheliste rekombinatsioonide toimumise ja/või mittetoimumise tõenäosuste korrutisena) – vt tabel 2.

Baseerudes leitud tõenäosustel püütakse uuritava tunnuse fenotüübiväärtusi lähendada mudeliga

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e},$$

kus \mathbf{y} on fenotüübiväärtuste $n \times 1$ -vektor, \mathbf{X} on tabelis 2 toodud tõenäosuste $n \times 2$ -maatriks, $\boldsymbol{\beta} = (\mu_{Q_1Q_1}, \mu_{Q_1Q_2})^T$ on tundmatute QTL-genotüüpide ooteväärtuste 2×1 -vektor, \mathbf{e} on juhuslike vigade $n \times 1$ -vektor.

Fikseeritud markerlookuste *A* ja *B* ning nendevahelise oletatava QTL-i tarvis leitakse parameetervektori $\boldsymbol{\beta}$ vähimruutude hinnang

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

ning jääkide ruutude summa $SSE = \mathbf{y}^T \mathbf{y} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{y}$.

LOD-skoor avaldatakse regressioonanalüüsi tulemustest kujul

$$LOD(r) = \frac{n}{2} \log_{10} \left[\frac{SSE_0}{SSE} \right], \quad (3.7)$$

kus n on analüüsitud indiviidide arv, SSE_0 tähistab jääkide ruutude summat ilma QTL-efektita mudelist ja SSE sama regressiooniseose täpsust iseloomustavat suurust põhigeeni eeldusel konstrueeritud mudelist.

Marker-genotüüp	QTL-genotüüp	
	$Q_1 Q_1$	$Q_1 Q_2$
$A_1 B_1 A_1 B_1$	$\frac{(1-r_A)(1-r_B)}{(1-r_{AB})}$	$\frac{r_A r_B}{(1-r_{AB})}$
$A_1 B_1 A_1 B_2$	$\frac{(1-r_A) r_B}{r_{AB}}$	$\frac{r_A (1-r_B)}{r_{AB}}$
$A_1 B_1 A_2 B_1$	$\frac{r_A (1-r_B)}{r_{AB}}$	$\frac{(1-r_A) r_B}{r_{AB}}$
$A_1 B_1 A_2 B_2$	$\frac{r_A r_B}{(1-r_{AB})}$	$\frac{(1-r_A)(1-r_B)}{(1-r_{AB})}$

Tabel 2. QTL-genotüüpide tinglikud tõenäosused marker-genotüüpide suhtes generatsioonis BC₁.

Näide. Alljärgnev näide demonstreerib, et kirjeldatud regressioonanalüüsimeetod on lisaks eksperimentaalsetele ristamisskeemidele rakendatav ka, tuvastamaks põhigeeni heterosügootsete teada oleva meioosi faasiga vanemate ja nende järglaste alusel.

Olgu kahe piirava markeri ja nende vahele jääva oletatava dialleelse põhigeeni osas heterosügootne teadaoleva genotüübiga $A_1 Q_1 B_1 | A_2 Q_2 B_2$ isane indiviid ristatud populatsioonist juhuslikult valitud mitteteadaoleva genotüübiga emase indiviidiga. Markerite A ja B vaheline kaugus on teadaolevalt 10 cM.

Konstrueerime regressioonivõrrandi modelleerimaks genotüübiga $A_1 B_1 | A_1 B_1$ järglase fenotüübiväärtust markerist A 2 cM võrra markeri B pool asuva QTL-i eeldusel.

Sentimorganites esitatud geneetiliste distantside aheldusanalüüsil kasutatavateks rekombinatsioonimääradeks teisendamiseks kasutame Haldane kaardistamisfunktsiooni. Arvestades, et markerite A ja B vaheline kaugus on 0,1 M, avaldub nende vaheline rekombinatsioonimäär kujul

$$r_{AB} = \frac{1}{2} (1 - e^{-2 \times 0,1}) = 0,0906;$$

analoogselt $r_A = \frac{1}{2} (1 - e^{-2 \times 0,02}) = 0,0196$ ja $r_B = \frac{1}{2} (1 - e^{-2 \times 0,08}) = 0,0739$.

Tinglikud tõenäosused, et genotüübiga $A_1 B_1 | A_1 B_1$ järglane sai isalt koos haplotüübiga $A_1 B_1$ ka QTL-alleeli Q_1 või Q_2 , on järgmised:

$$P\left(\frac{A_1 Q_1 B_1}{A_1 B_1} \mid A_1 B_1\right) = \frac{(1-r_A)(1-r_B)}{(1-r_{AB})} = 0,9984,$$

$$P\left(\frac{A_1 Q_2 B_1}{A_1 B_1} \mid A_1 B_1\right) = \frac{r_A \times r_B}{(1-r_{AB})} = 0,0016.$$

Eeldades järgnevalt, et QTL-alleeli Q_1 sagedus populatsioonis $P(Q_1) = 0,2$ ning et QTL-alleelide ja markeralleelide kombinatsioonid on aegade jooksul tekkinud juhuslikult, saame leida ka tõenäosused järglase mistahes genotüübi tarvis:

$$P\left(\frac{A_1 Q_1 B_1}{A_1 Q_1 B_1} \mid A_1 B_1\right) = \frac{(1-r_A)(1-r_B)}{(1-r_{AB})} \times P(Q_1) = 0,9984 \times 0,2 = 0,19968,$$

$$P\left(\frac{A_1 Q_1 B_1}{A_1 Q_2 B_1} \mid A_1 B_1\right) = \frac{(1-r_A)(1-r_B)}{(1-r_{AB})} \times P(Q_2) + \frac{r_A \times r_B}{(1-r_{AB})} \times P(Q_1) = 0,79904,$$

$$P\left(\frac{A_1 Q_2 B_1}{A_1 Q_2 B_1} \mid A_1 B_1\right) = \frac{r_A \times r_B}{(1-r_{AB})} \times P(Q_2) = 0,00128.$$

Tähistades oletatavatele QTL-genotüüpidele vastavad keskmised fenotüübiväärtused $\mu_{Q_1Q_1} = \mu + a$, $\mu_{Q_1Q_2} = \mu + d$ ja $\mu_{Q_2Q_2} = \mu - a$, saame markergenotüübiga $A_1B_1|A_1B_1$ järglase keskmise fenotüübiväärtuse tarvis võrrandi

$$\begin{aligned} E\left(y \mid \frac{A_1B_1}{A_1B_1}\right) &= 0,19968 \times (\mu + a) + 0,79904 \times (\mu + d) + 0,00128 \times (\mu - a) \\ &= \mu + 0,19552 \times a + 0,79904 \times d. \end{aligned}$$

Ülesanne 8.

Analoogselt eelnevale näitele konstrueerida regressioonivõrrand modelleerimaks genotüübiga $A_1B_2|A_1B_2$ järglase fenotüübiväärtust markerist A 2 cM võrra markeri B pool asuva dialleelse QTL-i eeldusel. Rekombinatsioonimäärad leidke binomiaalse kaardistamisfunktsiooni alusel (võttes potentsiaalsete ristsiirete asukohtade arvuks $N = 2$). QTL-alleeli Q_1 sagedus populatsioonis on $P(Q_1) = 0,2$ ning QTL-alleelide ja markeralleelide kombinatsioonid on aegade jooksul tekkinud juhuslikult.

3.3.4 Intervallkaardistamine

Intervallkaardistamine (*interval mapping*) on üldine nimetus meetoditele, mida kasutatakse märkimisväärse mõjuga geenide lokaliseerimiseks mingi statistilise mudeli abil piiravate markerite alusel. Meetodi idee on triviaalne – markerite vaheline DNA-piirkond jagatakse erinevate potentsiaalsete QTL-i asukohtade näol lühikesteks vahemikeks e intervallideks.

Näiteks teades, et markerite A ja B vaheline kaugus geneetilisel kaardil on 10 cM, võime selle vahemiku jagada 10-ks võrdse ulatusega intervalliks. Seejärel fikseeritakse oletatava QTL-i asukohana esimese intervalli otspunkt, st et põhigeen paigutatakse 1 cM kaugusele markerist A .

Järgnevalt leitakse aheldusanalüüsi mudeli tarvis nii markerite omavaheline kui ka markerite ja oletatava QTL-i vahelised rekombinatsioonimäärad. Näiteks markerist A 1 cM kaugusel ja seega markerist B 9 cM kaugusel paikneva oletatava QTL-i korral saame Haldane kaardistamisfunktsioonist markeri A ja QTL-i vaheliseks rekombinatsioonimääraks

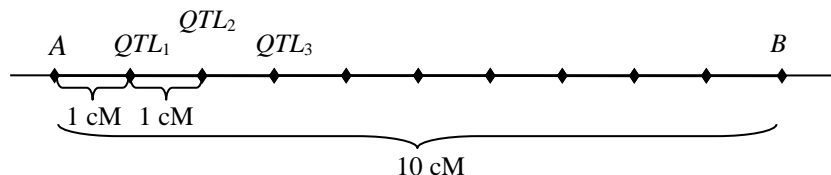
$$r_{A,QTL_1} = \frac{1}{2}(1 - e^{-2 \times 0,01}) = 0,009901$$

ning markeri B ja QTL-i vaheliseks rekombinatsioonimääraks

$$r_{QTL_1,B} = \frac{1}{2}(1 - e^{-2 \times 0,99}) = 0,430965.$$

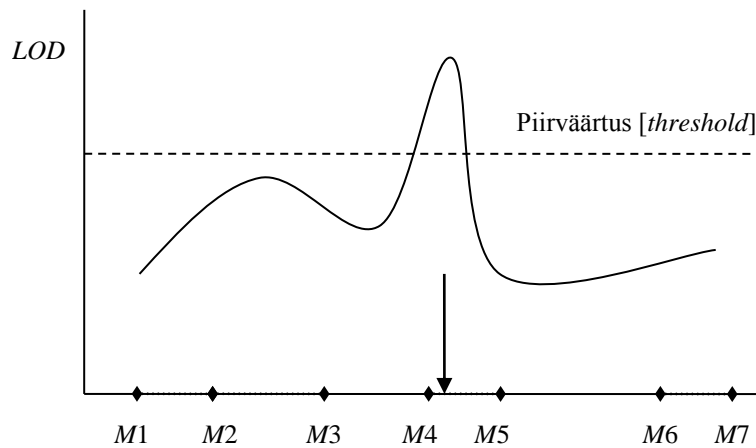
Leitud rekombinatsioonimäärade ja kasutatava statistilise mudeli alusel arvutatakse tõepärasuhte statistiku või LOD-skoori väärtus.

Kirjeldatud tegevusi korratakse järgmise potentsiaalse QTL-i asukoha tarvis (2 cM markerist A) jne (joonis 18).



Joonis 18. Markerite vahelise piirkonna intervallideks jagamine intervallkaardistamise korral.

Suurima tõepärasuhte statistiku või LOD-skoori väärtusega mudel vastab suurima mõjuga põhigeenile. Et oleks lihtsam välja valida intervalli, kus otsitav põhigeen tõenäoliselt paikneb, esitatakse erinevate intervallide korral leitud tõepärasuhte statistikute väärtused graafikul – seetõttu nimetataksegi vastavat meetodit intervallkaardistamiseks (joonis 19).



Joonis 19. Intervallkaardistamine.

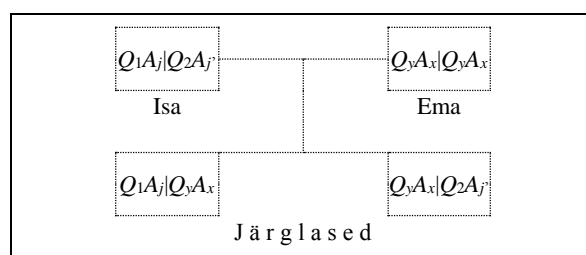
3.3.5 Poolõvede dispersioonanalüüs

Alternatiivne analüüs eelnevalt vaadeldud markergruppide vahelisele võrdlusele on uurida nn perekondade vahelist ja sisest varieeruvust, kasutades selleks dispersioonanalüüsi. Meetodi taust on lihtne – kui kaks poolõde või -venda pärivad ühiselt vanemalt sama markeralleeli, on markeriga aheldunud põhigeeni korral üsna tõenäoline, et nad pärivad sama alleeli ka QTL-s.

Järgnevalt vaatleme lihtsuse mõttes vaid poolõvede (ühe ühise vanemana vaatleme, nagu näiteks loomakasvatuses tavaks, isa) andmeid, perekonna moodustavad siis ühe isa järglased. Eeldame, et uuritavad isad pole omavahel sugulased ja et nad on määratava markerlookuse A suhtes heterosügootsed genotüübiga $A_jA_{j'}$. Emade genotüüp eeldatakse mitteteadaolevaks (vt joonis 20). Juhul, kui järglase genotüübi alusel on võimalik üheselt otsustada, kumma markeralleeli ta isalt päris, nimetatakse seda järglast **informatiivseks** (informatiivne ei ole järglane, kes on sarnaselt isale heterosügootne genotüübiga $A_jA_{j'}$). Informatiivsete järglaste andmete analüüsiks on kasutatav mudel

$$y_{ijk} = \mu + s_i + m_{ij} + e_{ijk} ,$$

kus y_{ijk} on isalt i markeralleeli j saanud järglase k fenotüübiväärtus, $\mu = E(y_{ijk})$, s_i on isa i juhuslik mõju keskvaartusega null ja dispersiooniga σ_s^2 , m_{ij} on isalt i päritud markeralleeli j juhuslik mõju keskvaartusega null ja dispersiooniga σ_m^2 , e_{ijk} on juhuslik viga dispersiooniga σ_e^2 . Kõik juhuslikud efektid eeldatakse olevat sõltumatud.



Joonis 20. Poolõvede skeem, heterosügootsetelt isadelt järglastele päranduvad markeralleelid on tähistatud A_j ja $A_{j'}$, emadelt päranduvad mitteteadaolevad markeralleelid A_x . Isadelt järglastele koos markeralleelidega edasi antavad QTL-alleelid on vastavalt Q_1 ja Q_2 ning emadelt saadud QTL-alleelid Q_y .

Kahe isalt i sama markeralleeli j saanud poolõde k ja k' vaheline kovariatsioon on siis esitatav summana

$$\text{Cov}(y_{ijk}, y_{ijk'}) = \sigma_s^2 + \sigma_m^2 .$$

Eeldades, et marker A ja otsitav QTL, millest põhjustatud varieeruvus moodustab kogu aditiivgeneetilise varieeruvusest σ_A^2 osa σ_Q^2 , on aheldunud ning arvestades, et kogu aditiivgeneetilise varieeruvusest on isalt päritud vaid $1/4$ ja et kummagi markeralleeli saamise tõenäosus järglastel on $1/2$, saame eelnevalt kirja pandud kovariatsiooni esitada ka kujul

$$\text{Cov}(y_{ijk}, y_{ijk'}) = [(1-r)^2 + r^2] \sigma_Q^2 / 2 + \sigma_A^2 / 4,$$

kus r on markerlookuse ja QTL-i vahelise rekombinatsiooni tõenäosus.

Ühiselt isalt i erinevad markeralleelid j ja j' pärinud järglaste k ja k' vaheline kovariatsioon on

$$\text{Cov}(y_{ijk}, y_{ij'k'}) = \sigma_s^2,$$

ehk eeldades jällegi QTL-i ja markerlookuse aheldust

$$\text{Cov}(y_{ijk}, y_{ij'k'}) = 2r(1-r) \sigma_Q^2 / 2 + \sigma_A^2 / 4.$$

Neljast viimasest võrdusest kokku järeldub, et

$$\sigma_m^2 = [(1-r)^2 + r^2 - 2r(1-r)] \sigma_Q^2 / 2 = (1-2r)^2 \sigma_Q^2 / 2$$

ja et $\sigma_m^2 \neq 0$, kui $r \neq 1/2$. Testimaks, kas $\sigma_m^2 > 0$ (kas QTL-i mõju on statistiliselt oluline), arvutame teststatistiku

$$F = \frac{MS_{m(s)}}{MS_e},$$

kus $MS_{m(s)}$ on isade sisesele markeralleelide vahelisele erinevusele vastav keskruudu väärtus ja MS_e on juhuslikule veale vastav keskruut. Nullhüpoteesi $\sigma_m^2 = 0$ kehtides on leitud statistik F -jaotusega vabadusastmete arvuga ν_1 ja ν_2 , kus $\nu_1 = n_s$ ja $\nu_2 = 2n_s \sum_{i,j} (n_{ij} - 1)$, n_s – isade arv, n_{ij} – isalt i markeralleeli j pärinud järglaste arv.

3.3.6 Aheldusanalüüs kompleksse sugupuu korral – QTL-efektiga üldine lineaarne segamudel

Eeldatakse, et fenotüüp P on kujunenud järgmise mudeli alusel:

$$P = Q_m + Q_f + A + E,$$

kus Q_m ja Q_f on vastavalt isalt ja emalt päritud põhigeeni efektid, A on polügeenne efekt ja E väljendab keskkonna mõjusid.

Põhigeeni mõju ja asukoha, polügeense efekti (aretusväärtuse) ning vajadusel ka fikseeritud efektide hindamiseks vaatleme segamudelit

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{T}\mathbf{q} + \mathbf{Z}\mathbf{a} + \mathbf{e}, \quad (3.8)$$

kus \mathbf{y} on vaatluste $N \times 1$ -vektor, $\boldsymbol{\beta}$ on fikseeritud efektide $k \times 1$ -vektor, \mathbf{X} fikseeritud efektidele vastav $N \times k$ -plaanimaatriks, \mathbf{a} on aditiivsete polügeensete juhuslike efektide $n \times 1$ -vektor, \mathbf{Z} on juhuslikele geneetilistele efektidele vastav $N \times n$ -plaanimaatriks, \mathbf{T} on $n \times 2n$ -maatriks, mis seab igale indiviidile vastavusse tema kaks QTL-alleeli, \mathbf{q} on indiviidide kahekaupa esitatud juhuslike QTL-efektide $2n \times 1$ -vektor ja \mathbf{e} on juhuslike vigade $N \times 1$ -vektor.

Maatriksid \mathbf{X} , \mathbf{Z} ja \mathbf{T} eeldatakse teadaolevaiks. Juhuslikud efektid eeldatakse olevat normaaljaotustega vastavalt

$$\mathbf{a} \sim N(\mathbf{0}, \mathbf{A}\sigma_a^2), \quad \mathbf{q} | M, p \sim N(\mathbf{0}, \mathbf{Q}_{p,M}\sigma_q^2), \quad \mathbf{e} \sim N(\mathbf{0}, \mathbf{R}\sigma_e^2),$$

kus σ_a^2 , σ_q^2 ja σ_e^2 on vastavalt polügeensetest efektist ja QTL-alleeli mõjust tingitud dispersioonid ning jääkvarieeruvus. Maatriks \mathbf{A} on $n \times n$ **aditiivgeneetilise suguluse maatriks** ja $\mathbf{Q}_{p,M}$ on markergenotüüpidest ja nende aheldusest M ning QTL-i asukohast p sõltuv QTL-s paiknevate alleelide $2n \times 2n$ dispersioonimaatriks (nn **gameedilise suguluse maatriks**). Markerite kohta teadaolevat informatsiooni kasutatakse leidmaks tõenäosust, et kaks alleeli on identsed päritolult (need tõenäosused kujutavadki enesest maatriksi \mathbf{Q} elemente).

Lisaks eeldatakse, et

$$\text{Cov}(\mathbf{a}, \mathbf{q}) = \mathbf{0}, \quad \text{Cov}(\mathbf{a}, \mathbf{e}) = \mathbf{0} \quad \text{ja} \quad \text{Cov}(\mathbf{e}, \mathbf{q}) = \mathbf{0}.$$

Mudelile (3.8) vastav segamudeli võrrand on

$$\begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z} \mathbf{T} & \mathbf{X}^T \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{T}^T \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{T}^T \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{X} + \mathbf{Q}_{p,M}^{-1} \frac{\sigma_e^2}{\sigma_q^2} & \mathbf{T}^T \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} + \mathbf{A}^{-1} \frac{\sigma_e^2}{\sigma_a^2} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{q}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{T}^T \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{y} \\ \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{y} \end{bmatrix}.$$

Eeldades, et fenotüübiväärtused on mitmemõõtmelise normaaljaotusega $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$, kus

$$\text{Var}(\mathbf{y}) = \mathbf{V} = \mathbf{T}\mathbf{Z}\mathbf{Q}\mathbf{Z}^T\mathbf{T}^T\sigma_q^2 + \mathbf{Z}\mathbf{A}\mathbf{Z}^T\sigma_a^2 + \mathbf{R}\sigma_e^2,$$

saame tõepärafunktsiooni täielike markerandmete korral esitada seosena

$$L(\boldsymbol{\beta}, p, \theta | \mathbf{y}, M) \propto f(\mathbf{y} | \mathbf{X}\boldsymbol{\beta}, \mathbf{V}),$$

kus $f(\mathbf{y} | \mathbf{X}\boldsymbol{\beta}, \mathbf{V}) = 2\pi^{-0.5N} |\mathbf{V}|^{-0.5} \exp[-0.5(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]$ ja $\theta = \{\sigma_a^2, \sigma_q^2, \sigma_e^2\}$.

Dispersioonikomponendid σ_a^2 , σ_q^2 ja σ_e^2 hinnatakse tavaliselt **REML-meetodil** (*REstricted* või *REsidual Maximum Likelihood method* – kitsendatud suurima tõepära meetod). REML-meetodi korral ei lähtuta mitte andmevektorist \mathbf{y} , vaid selle elementide lineaarkombinatsioonidest $\mathbf{K}^T \mathbf{y}$, kus maatriks \mathbf{K} on valitud nii, et $\mathbf{K}^T \mathbf{X} = \mathbf{0}$, st et $E(\mathbf{K}^T \mathbf{y}) = \mathbf{K}^T \mathbf{X}\boldsymbol{\beta} = \mathbf{0}$, $\text{Var}(\mathbf{K}^T \mathbf{y}) = \mathbf{K}^T \text{Var}(\mathbf{y})\mathbf{K} = \mathbf{K}^T \mathbf{V}\mathbf{K}$ ja tihe-
dusfunktsioon on kujul

$$f(\mathbf{K}^T \mathbf{y} | \mathbf{K}^T \mathbf{V}\mathbf{K}) = 2\pi^{-0.5N} |\mathbf{K}^T \mathbf{V}\mathbf{K}|^{-0.5} \exp[-0.5 \mathbf{y}^T \mathbf{K} (\mathbf{K}^T \mathbf{V}\mathbf{K})^{-1} \mathbf{K}^T \mathbf{y}].$$

Maksimeeritav logaritmiline tõepärafunktsioon on siis kujul

$$\ln L(p, \theta | \mathbf{K}^T \mathbf{y}, M) \propto \text{const} - 0.5 \ln |\mathbf{K}^T \mathbf{V}\mathbf{K}| - 0.5 \mathbf{y}^T \mathbf{K} (\mathbf{K}^T \mathbf{V}\mathbf{K})^{-1} \mathbf{K}^T \mathbf{y}.$$

Põhigeeni(de) olemasolu ja asukoha selgitamiseks võrreldakse tõepärasuhte statistiku abil omavahel QTL-i olemasolu eeldusel ja ilma QTL-efektita konstrueeritud mudeleid:

$$\Lambda(\mathbf{y}) = -2 \ln \left[\frac{L(\text{QTL-efektiga mudel})}{L(\text{QTL-efektita mudel})} \right].$$

Põhigeeni(de) tõenäoline asukoht leitakse, teostades kirjeldatud segamudelite analüüsi üle kogu markeritega kaetud kromosoomi(de) osa(de) (kõigis markeritevahelistes intervallides), arvutades iga kord uuesti gameedilise suguluse maatriksi \mathbf{Q} ja selle alusel kõik ülejäänud parameetrid – st, et teostatakse geneetiliste analüüsides puhul traditsiooniline intervallkaardistamine. Ja muidugi esitatakse tulemused mitte tõepärasuhte statistiku vaid hoopis *LOD*-skoori kujul.

Näide. Näitena juhusliku QTL-efektiga mudeli konstrueerimisest vaatleme väikest kolme indiviidiga andmestikku, kus kõigil indiviididel on määratud kaks markerit (m_{11} märgib esimese markeri, A , ühes kromosoomis ja m_{12} teises kromosoomis paiknevat alleeli; m_{21} ja m_{22} on siis vastavalt teise markeri, B , kaks alleeli):

i	isa	ema	μ	m_{11}	m_{12}	m_{21}	m_{22}
1	-	-	1	5	3	4	2
2	1	-	1	5	2	4	1
3	1	-	1	3	1	4	5

Andmete kirjeldamiseks kasutame mudelit (3.8),

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{T}\mathbf{q} + \mathbf{Z}\mathbf{a} + \mathbf{e},$$

kus $\boldsymbol{\beta} = \mu$, $\mathbf{q} = [q_i^p \ q_i^m \ q_2^p \ q_2^m \ q_3^p \ q_3^m]^T$, $\mathbf{a} = [a_1 \ a_2 \ a_3]^T$ ja plaanimaatriksid on kujul

$$\mathbf{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{Z} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{T} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

Indiviidi i . jaoks võib mudeli esitada kujul

$$y_i = \mu + q_i^p + q_i^m + a_i + e_i,$$

kus q_i^p ja q_i^m on vastavalt isalt ja emalt päritud põhigeeni mõjud.

Aditiivse geneetilise suguluse maatriksi \mathbf{A} elemendid kirjeldavad indiviidide vahelist polügeenset seost ja on võrdsed kahekordse tõenäosusega, et juhuslik alleel indiviidi i genotüübist on päritolult identne (\equiv) vastava alleeliga indiviidi j genotüübist, st. et peadiagonaalil paiknevate elementide (isendi sugulus iseendaga) väärtuseks on 1 pluss inbriidingukoefitsient, väljaspool peadiagonaali paiknevad aga indiviididevahelised sugu-

luskoefitsiendid (mis on $\frac{1}{2}$ täisõvede või vanema-järglase vahel, $\frac{1}{4}$ poolõvede või indiviidi ja tema vanavanemate vahel jne). Selle näite põlvnemisskeemile vastav aditiivse geneetilise suguluse maatriks on

$$\mathbf{A} = \begin{bmatrix} 1 & 0,5 & 0,5 \\ 0,5 & 1 & 0,25 \\ 0,5 & 0,25 & 1 \end{bmatrix}.$$

Gameedilise suguluse maatriks \mathbf{Q} on analoogne maatriksiga \mathbf{A} , ainult et kaks korda suurema dimensiooniga (sisaldab päritolult identne olemise tõenäosusi eraldi nii isalt kui ka emalt päritud alleelide kohta) ja konstrueeritakse ta enam mitte mingi juhusliku alleeli vaid kindla otsitava QTL-alleeli suhtes. Leitavad on need tõenäosused teadaolevate markeralleelide ja nende ahelduse andmetele tuginedes. Antud näite korral avaldub gameedilise suguluse maatriks kujul

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & p_{13} & 0 & p_{15} & 0 \\ 0 & 1 & p_{23} & 0 & p_{25} & 0 \\ p_{13} & p_{23} & 1 & 0 & p_{35} & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ p_{15} & p_{25} & p_{35} & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Indiviidil 1 on ühes kromosoomis koos (aheldunud) markeralleelid 5 ja 4, teises kromosoomis markeralleelid 3 ja 2, indiviid 2 pärib isalt haploidse kromosoomistiku markeralleelidega 5 ja 4, indiviid 3 aga markeralleelid 3 ja 4. Tõenäosused, et järglane sai isalt kas paternaalse (isapoolse) või siis maternaalse (emapoolse) QTL-alleeli on järgmised:

$$p_{13} = P(q_2^p \equiv q_{isa}^p | L, H_2) = \frac{(1 - r_{Aq})(1 - r_{qB})}{(1 - r_{AB})};$$

$$p_{23} = P(q_2^p \equiv q_{isa}^m | L, H_2) = \frac{r_{Aq}r_{qB}}{(1 - r_{AB})};$$

$$p_{15} = P(q_1^p \equiv q_{isa}^p | L, H_3) = \frac{r_{Aq}(1 - r_{qB})}{r_{AB}};$$

$$p_{25} = P(q_3^p \equiv q_{isa}^m | L, H_3) = \frac{(1 - r_{Aq})r_{qB}}{r_{AB}};$$

$$p_{35} = P(q_2^p \equiv q_3^p | L, H_2, H_3) = P(q_2^p \equiv q_{isa}^p | L, H_2)P(q_1^p \equiv q_{isa}^p | L, H_3) \\ + P(q_2^p \equiv q_{isa}^m | L, H_2)P(q_3^p \equiv q_{isa}^m | L, H_3).$$