

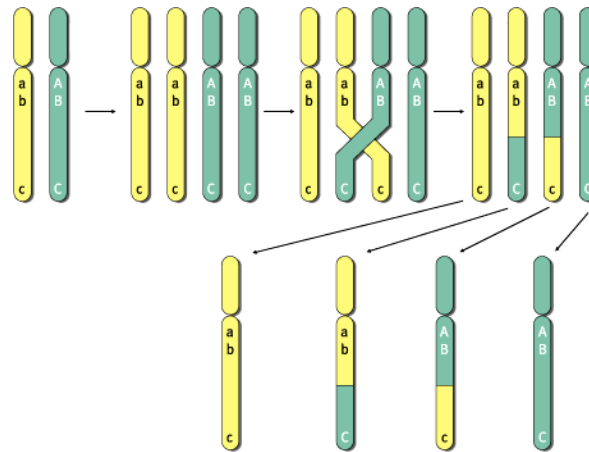
## III

## AHELDUSANALÜÜS

## 3.1 Sissejuhatus aheldusanalüüsi

## 3.1.1 Geenide aheldus, ristsiire, rekombinatsioon, lookuste vaheline kaugus, marker

Õeldakse, et kaks geeni on **aheldunud**, kui nad paiknevad ühes kromosoomis ja päranduvad seega ka peaaegu alati koos. Ometigi pole geenide aheldus absoluutne: meioosi algfaasis on homoloogiliste kromosoomide paardumisel ehk konjugeerumisel jälgitavad nende vahelised ühendused – **kiasmid**, kus võib aset leida homoloogiliste kromosoomide kromatiidiosade vahetus ehk ristsiire (vt joonis 14).



Joonis 14. Meioosi käigus toimub geneetilise materjali vahetus ehk rekombinatsioon sugurakkude samade kromosoomide isa- ja emapoolse koopia vahel, mille tõttu ei saa laps oma vanemate kromosoomide täpseid koopiaid. Selle aluseks on *crossing-over* ehk ristsiire. Joonisel on näidatud kuidas on toimunud kromosoomide pikema öla fragmendi vahetus.

Kiasmide paiknemine kromosoomil eeldatakse enamasti järgivat **Poissoni protsessi**<sup>1</sup>, kusjuures vastavad protsessid erinevate indiviidide ja ka sama indiviidi erinevate kromosoomide korral eeldatakse olevat sõltumatud.

Kui kahe homoloogses kromosoomis paikneva alleeli vahel on toimunud paaritu arv ristsiirdeid (vaid sellisel juhul saame me teavet krossingoveri toimumisest), õeldakse, et nende kahe alleeli vahel on aset leidnud **rekombinatsioon** (kaks alleeli on rekombineerunud). Sedasi, kromosoomide vastavate lõikude vahetumisel, kujunevad gameetides alleelide uued, vanematest erinevad kombinatsioonid. Erinevate alleelidega gameetide kombineerumisel võib populatsioonis kujuneda lõpmatu hulk erisuguseid genotüüpe ja järelikult ka fenotüüpe. Kahe lookuse vahelise kauguse ja nende vahel aset leidnud rekombinatsioonide arvu vahekorra kohta kehtib **Morgani seadus**:

<sup>1</sup> **Juhuslik protsess** kujutab enesest juhuslike suuruste  $N(t)$  peret (hulka)  $\{N(t), t \geq 0\}$ , parameeter  $t$  märgib aega, kohta ruumis, ...

Olgu  $N(t)$  mingite sündmuste (näiteks ristsiirete) toimumiste koguarv ajavahemikus/intervallis  $[0, t]$ . Juhuslikku protsessi  $\{N(t), t \geq 0\}$  nimetatakse **Poissoni protsessiks** intensiivsusega  $\lambda > 0$ , kui

a)  $N(0) = 0$ ;

b) protsessi juurdekasvud on sõltumatud (st iga  $t_1 < t_2 < t_3 < t_4$  korral  $N(t_4) - N(t_3)$  ja  $N(t_2) - N(t_1)$  on sõltumatud juhuslikud suurused);

c) sündmuste arv mistahes lõigul (st, et sõltub vaid lõigu pikkusest – tegu on stationaarse protsessiga) pikkusega  $t$  on Poissoni jaotusega juhuslik suurus keskväärtusega  $\lambda t$ , st suvalise  $s, t \geq 0$  korral:

$$P[N(t+s) - N(s) = n] = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad n = 0, 1, 2, \dots$$

**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.

Sellest seadusest lähtuvalt on defineeritud ka geenidevahelise kauguse mõõtmisel kasutatav ühik **morgan** – kahe geeni vaheline kaugus on üks morgan (1 M), kui nende geenide vaheline keskmine ristsiirete sagedus ühe meioosi jooksul on 1 (enam kasutatakse ühikut **sentimorgan** ehk morganiid – cM, mis vastab 1%-lisele ristsiirete sagedusele).

Kuna kaks erinevates kromosoomides paiknevat alleeli lahknevad meioosi käigus teineteisest sõltumatult, on nende vahelise rekombinatsiooni tõenäosus 1/2. Ühes kromosoomis paiknevad alleelid saavad lahkneda vaid nende vahel aset leidnud ristsiirde tagajärjel, kusjuures mida lähemal nad üksteisele paiknevad, seda suurem on tõenäosus, et järglasele päranduvad just kas kaks isalt saadud alleeli või siis kaks emalt saadud alleeli (Morgani seadus). Selline mitterekombinantsete gameetide suhteline ülekaal rekombinantsete gameetidega võrreldes tähendab, et vaadeldavate lookuste vaheline rekombineerumise tõenäosus on väiksem kui 1/2. Siit tuleneb ka teine, küllalt sageli kasutatav geenide (lookuste) ahelduse definitsioon, nimelt:

kui kahe lookuse vaheline rekombineerumise tõenäosus (**rekombinatsioonimäär**) on väiksem kui 1/2, öeldakse, et need geenid (lookused) on aheldunud.

Morgani seadusel ja sellest lähtuvalt defineeritud mõõtühikul baseerub geenide (lookuste) lineaarne järjestamine ja nende vaheliste kauguste määramine, so **geenide kaardistamine (kromosoomikaardid)**. Näiteks kui lookuste A ja B vaheline rekombineerumise tõenäosus  $r_{AB} = 0,09$ , lookuste A ja C vaheline rekombineerumise tõenäosus  $r_{AC} = 0,27$  ja lookuste B ja C vaheline rekombineerumise tõenäosus  $r_{BC} = 0,21$ , võime me need lookused järjestada kui ABC. Märkime veel, et  $r_{AC} < r_{AB} + r_{BC}$ , seda põhjusel, et osad B ja C vahelised rekombinatsioonid võivad “nullida” A ja B vahel aset leidnud rekombinatsioonid – vaadeldes vaid lookusi A ja C, jäävad sellised vahepeal aset leidnud kromosoomide osade vahetused märkamata.

**Näide.** Olgu meil kahe lookuse vaheline kaugus  $x = 0,5$  M, st. keskmiselt üks ristsiire kahe meioosi kohta. Et ristsiire on küllaltki harva esinev sündmus, on loomulik selle esinemist vaadelda kui Poissoni jaotusega juhuslikku suurust. Tõenäosuste, et ühe meioosi jooksul toimub  $k = 0, 1, 2, \dots$  ristsiiret, arvutamiseks sobib siis valem:

$$p(k) = \frac{0,5^k}{k!} e^{-0,5}.$$

Seega

$$\begin{aligned} p(0) &= 0,6065 \\ p(1) &= 0,3033 \\ p(2) &= 0,0758 \\ p(3) &= 0,0126 \\ p(4) &= 0,0016 \\ p(5) &= 0,0002 \\ &\dots \end{aligned}$$

Rekombinatsiooni tõenäosus  $r$  on võrdne paarituarvuliste ristsiirete toimumise tõenäosusega:

$$r = p(1) + p(3) + \dots = 0,3161.$$

Ja tõenäosus, et meil ei õnnestu vaadeldud intervallis rekombinatsiooni tuvastada, on

$$1 - r = p(0) + p(2) + p(4) + \dots = 0,6839.$$

Kromatiidide vaheliste ristsiirete arv sõltub kromosoomi pikkusest, näiteks inimese 1. kromosoomi pikkuseks on ligikaudu 2 M ning 21. kromosoomi pikkuseks ligikaudu 0,5 M. Seejuures on geenide rekombineerumise sagedus naistel suurem kui meestel – naiste kromosoomikaardi ulatus on ligikaudu 39 M, samal ajal on meeste genoomi suuruseks hinnatud vaid ligikaudu 26,5 M.

Kromosoomi geneetilist kaarti võib kõrvutada ka **kromosoomi füüsilise kaardiga**, kus geenidevahelist kaugust mõõdetakse aluspaarides (või kilo-aluspaarides – kb). Inimesel vastab 1 cM-le

geneetilisel kaardil keskmiselt ligikaudu 1-2 Mb DNA-d (1-2 miljonit aluspaari), pärmil aga näiteks 30 miljonit aluspaari. Seejuures on sellise seose varieeruvus genoomi eri osade vahel küllaltki suur.

Kui nüüd võrrelda kahe indiviidi mingi kromosoomi mingit kindlat piirkonda moodustavate aluspaaride jada, siis enamus neist aluspaaridest osutub identseteks. Näiteks on hinnatud, et kahe inimese genoomi aluspaaride jadas leidub vaid 1-2 erinevust 1000 aluspaari kohta. Siiski ei oma kaks indiviidi (va. identsed kaksikud) oma miljonitest aluspaaridest koosnevas geneetilises materjalis täpselt samu aluspaare täpselt samades kohtades, mööda kromosoomi liikudes võime siin-seal avastada suuremaid või väiksemaid erinevusi ehk **polümorfisme**. Selliseid paiku, kus need erinevused esinevad, kutsutakse molekulaargeneetilisteks **markeriteks**. Lisaks suurele varieeruvusele (mittesugulased omavad suure tõenäosusega erinevaid markerallelele) eeldatakse markerite kohta ka, et nad on lihtsalt ja usaldusväärset määratavad. Markerid annavad ette orientiirid geneetiliste kaartide koostamiseks ja geenide ja fenotüübi vaheliste seoste tuvastamiseks.

Kui nõ klassikalisteks markeriteks on erinevused fenotüübi alusel määratavates tunnustes, mida loetakse indikaatoreiks erinevatele genotüüpidele (näiteks veregrupid, värvipimedus), siis uuema aja molekulaargeneetiliste markerite all mõistetakse juba konkreetseid DNA tasandil määratud erinevusi. Ühed tähelepanuväärsemad ja enamkasutatud molekulaargeneetilised markerid on kindlasti **mikro-** ja **minisatelliidid**, mis kujutavad enesest kas siis lühemaid (2-4 aluspaari) või pikemaid DNA kordusjärjestusi. Vastava markeri väärtuseks (alleeliks) loetakse lihtsalt konkreetse järjestuse (levinim on tsütosiini ja adenini paar, CA) korduste arvu. Tänapäevased markerid võimaldavad uuringuid teostada veelgi täpsemalt, fikseerides ka vaid ühenukleotiidilised erinevused indiviidide vahel – siit ka nende markerite nimetus: **SNP** (*single nucleotide polymorphism*). Pärts uus suund markerite kasutamisel on aga jälle vähe üldisema poole tagasi – nimelt on leitud, et üksikute SNP-de korral on nende varieeruvus liialt väike (erinevaid alleele saab maksimaalselt olla vaid 4 ja enamasti esinevad SNP-d üksnes 2-e erineva variandina) ning samas ei toimu ristisire sugugi juhuslikes kohtades vaid püüab jätta genoomi kodeerivad osad ühte tükki, siis võib analüüside täpsuse huvides olla sobilikum vaadelda markeritena hoopis SNP-de blokke (ehk **haplotüüpide blokke**).

### 3.1.2 Kaardistamisfunktsioonid (*map functions*)

Et lookustevaheliste rekombinatsioonide arvud ei ole üldjuhul aditiivsed, on peale iga uue geeni tuvastamist vajalik sentimorganites esitatud geneetilise kaardi ümberrekendamine. Selle vältimiseks määratakse geneetilisel kaardil kujutatavad kaugused nn **kaardistamisfunktsioonide** baasil. Kaardistamisfunktsioonid tõlgivad kahe lookuse vahelised rekombinatsioonide sagedused sentimorganites esitatud vahemaadeks geneetilisel kaardil. Viimased kujutavad enesest aga aheldusanalüüsil kasutatavate mudelite parameetreid.

Olgu meil vaatluse all  $k$  heterosügootset lookust. Võimalikud rekombinatsioonid suguraku moodustumisel neis lookustes asuvatest alleelidest on kirjeldatavad  $k-1 \times 1$  vektoriga, kus selle vektori  $j$ . element võrdub ühega, kui  $j$ . ja  $j+1$ . lookuse vahel leidis aset rekombinatsioon, ja nulliga rekombinatsiooni puudumisel. Kokku on  $k$ -lookuselisel analüüsil vaja läbiuurida  $2^{k-1}$  juhtu, sest just nii palju on  $k$  lookuse korral erinevaid rekombinatsioonide kombinatsioone (sündmusi). Arvestades, et nende sündmuste tõenäosuste summa võrdub ühega, on hinnatavaid parameetreid  $2^{k-1}-1$ . Rekombinatsioonide sageduse ja lookustevahelise kauguse vahelist seost modelleerivad kaardistamisfunktsioonid võimaldavad arvutada need  $2^{k-1}-1$  sagedust kromosoomi geneetilisel kaardil mõõdetud  $k-1$  kauguse abil. Ning seega kahaneb aheldusanalüüsi parameetrite hulk  $k-1$  Morgani skaalal esitatud lookustevahelise kauguseni.

Enne konkreetsete kaardistamisfunktsioonide juurde asumist veel mõned terminid ja tähelepanekud lookustevaheliste rekombinatsioonide kohta. Olgu meil kolm lookust järjestatuna  $ABC$ . Kui rekombinatsioonide toimumised intervallides  $A-B$  ja  $B-C$  on sõltumatud, avaldub kahekordse rekombinatsiooni esinemise tõenäosus  $g_{11}$  intervallis  $A-C$  korrutisena

$$g_{11} = r_{AB}r_{BC},$$

kus  $r_{ij}$  märgib lookuste  $i$  ja  $j$  vahelise rekombinatsiooni tõenäosust. Kui rekombinatsioonide toimumine kahes vaadeldavas intervallis ei ole sõltumatu, on kahekordse rekombinatsiooni esinemise tõenäosus leitav valemist

$$g_{11} = c \times r_{AB}r_{BC},$$

kus kordajat  $c$  nimetatakse **koosesinemise (kokkulangevuse) kordajaks** (*coefficient of coincidence*). Efekti, mille korral krossingoveri esinemine mingis kromosoomi piirkonnas pärsib selle toimumist naaberpiirkonnas, nimetatakse **interferentsuseks** (*interference*) ja arvutatakse seosest  $I = 1 - c$ . Sõltumatuse korral järelikult  $c = 1$  ja  $I = 0$ .

Lihtsaim kaardistamisfunktsioon (tuntud ka kui **Morgani kaardistamisfunktsioon**) baseerub eeldusel, et vaadeldavas intervallis saab toimuda maksimaalselt vaid 1 ristsiire ja et ristsiirde toimumise tõenäosus intervallis on proportsionaalne selle intervalli pikkusega. Sellisel juhul võib lühikestes kromosoomilõikudes lugeda ligikaudselt õigeks samasuse

$$r_x = x,$$

kus  $x$  märgib vaadeldava intervalli pikkust Morganites ja  $r_x$  on rekombinatsiooni toimumise tõenäosus selles intervallis.

Keerulisematel juhtudel kasutatakse kaardistamisfunktsioonide tuletamiseks enamasti ühte järgmisest kahest teest. Esimese korral eeldatakse mingi tõenäosusliku seose kehtimist ristsiirete arvu kohta intervallis  $x$  ning leitakse seejärel rekombinatsioonide sagedus kui paaritu arvu ristsiirete toimumise tõenäosus selles intervallis. Teisel juhul vaadeldakse rekombinatsioone kahes lähestikku paiknevas kromosoomi piirkonnas kirjeldavat mudelit, arvestades seejuures ka võimalikku interferentsust. Mudeli alusel püstitatakse diferentsiaalvõrrand, mille lahendid viivadki kaardistamisfunktsioonini. Järgnevalt on mõlemal lähenemisel peatunud vähe pikemalt.

Tähistagu  $P_t$  tõenäosust, et intervallis pikkusega  $x$  Morganit leiab aset  $t$  ristsiiret. Et rekombinatsioon intervalli otspunktide vahel on tuvastatav vaid paaritu arvu ristsiirete korral, avaldub rekombinatsiooni esinemise tõenäosus intervallis  $x$  seosena

$$r_x = P_1 + P_3 + P_5 + \dots = \frac{1}{2} \left[ 1 - \sum_{t=0}^{\infty} P_t \times (-1)^t \right] = \frac{1}{2} [1 - P(-1)] \quad (3.1)$$

kus  $P(s) = \sum_t P_t s^t$  on ristsiirete arvu jaotusele vastav **tõenäosusi genereeriv funktsioon**<sup>2</sup>.

J. B. S. Haldane kasutas 1919. aastal rekombinatsioonide arvu modelleerimiseks Poissoni protsessi, mille kohaselt on tõenäosused  $P_t$  arvutatavad Poissoni jaotuse tõenäosusfunktsioonist. Seejuures on vaja eeldada, et ristsiirded ühes intervallis toimuvad üksteisest sõltumatult ja et ristsiirete tõenäosus lühikesel intervallil on proportsionaalne selle intervalli pikkusega. Tõenäosus, et intervallis pikkusega  $x$  leiab aset  $t$  ristsiiret, on arvutatav valemist

$$P_t = \frac{(\lambda x)^t e^{-\lambda x}}{t!}.$$

Et Poissoni jaotuse parameeter  $\lambda$  on defineeritud kui ühikulise pikkusega intervallis toimunud sündmuste (ristsiirete) arv ja kahe lookuse vaheline kaugus geneetilisel kaardil võrdub nende lookuste vahel aset leidnud ristsiirete arvuga, siis  $\lambda = 1$  ja

$$P_t = \frac{(x)^t e^{-x}}{t!}.$$

Viimasele võrdusele vastav tõenäosusi genereeriv funktsioon on

$$P(s) = \sum_t \frac{x^t e^{-x} s^t}{t!} = \sum_t \underbrace{\frac{(xs)^t e^{-xs}}{t!}}_{e^{-xs}} = e^{x(s-1)} \quad (3.2)$$

Valemid (3.1) ja (3.2) annavad kokku **Haldane kaardistamisfunktsiooni**

$$r_x = \frac{1}{2} (1 - e^{-2x}). \quad (3.3)$$

Viimase pöördfunktsioon, mis võimaldab leida lookuste vahelise kauguse nende vahelisele rekombinatsioonimäärale tuginedes, on

$$x = \begin{cases} -\frac{1}{2} \ln(1 - 2r_x), & \text{kui } 0 \leq r_x < \frac{1}{2}; \\ \infty, & \text{kui } r_x = \frac{1}{2}. \end{cases} \quad (3.4)$$

Näiteks konverteerides aheldusanalüüsil hinnatud rekombinatsioonimäärat 22% kauguseks geneetilisel kaardil, saame vastavalt valemile (3.4), et  $x = -\frac{1}{2} \ln(1 - 2 \times 0,22) = 0,29$  M ehk 29 cM.

S. Karlin lähtus oma kaardistamisfunktsiooni tuletamisel 1984. aastal sellest, et  $t$  on ristsiirde esinemise arv  $N$  võimaliku katkemise koha kohta intervallis  $x$  ja  $p$  märgib ristsiirde esinemise tõenäosust igal üksikul juhul. Ristsiirete arv  $t$  on siis binoomijaotusega parameetritega  $N$  ja  $p$ . Geneetilisel kaardil

<sup>2</sup> Diskreetse juhusliku suuruse  $X = 0, 1, 2, \dots$  **tõenäosusi genereeriv funktsioon** on tõenäosusfunktsiooni

$p_k = P(X = k)$  esitus astmereana kujul  $\varphi(t) = E(t^X) = \sum_{k=0}^{\infty} p_k t^k$ .

kujutatud lookuste vahelise kauguse definitsioonist järeldub, et distants  $x = E(t) = Np$  ja seega  $p = x/N$ . Tõenäosus, et intervallis pikkusega  $x$  leiab aset  $t$  ristsiret, on arvutatav valemist

$$P_t = \binom{N}{t} (x/N)^t (1 - x/N)^{N-t}.$$

Vastav tõenäosusi genereeriv funktsioon on

$$\begin{aligned} P(S) &= \sum_{t=0}^N \binom{N}{t} (x/N)^t (1 - x/N)^{N-t} s^t \\ &= \sum_{t=0}^N \binom{N}{t} (xs/N)^t (1 - x/N)^{N-t} = [xs/N + (1 - x/N)]^N, \end{aligned} \quad (3.5)$$

sest  $\sum_{t=0}^N \binom{N}{t} a^t b^{N-t} = (a + b)^N$ . Valemeist (3.1) ja (3.5) saame **binomiaalse kaardistamisfunktsiooni**:

$$r_x = \begin{cases} \frac{1}{2} [1 - (1 - 2x/N)^N], & \text{kui } x < N/2 \\ \frac{1}{2}, & \text{kui } x \geq N/2 \end{cases}. \quad (3.6)$$

Funktsiooni (3.6) pöördfunktsioon on

$$x = \frac{1}{2} N [1 - (1 - 2r_x)^{1/N}].$$

Rekombinatsioonimäärale 22% vastav kaugus geneetilisel kaardil leituna binomiaalse kaardistamisfunktsiooni abil (eeldades, et kiasmide arv  $N = 2$ ), on  $x = 0,252$  M ehk 25,2 cM.

Teine võimalus kaardistamisfunktsioonide tuletamiseks on võtta vaatluse alla kromosoomi kaks naaberpiirkonda ja püüda kirjeldada neis toimuvaid rekombinatsioone. Olgu meil kolm lookust  $A$ ,  $B$  ja  $C$  järjestatud  $ABC$  ning olgu lookuste  $A$  ja  $B$  vaheline kaugus  $x$  ja lookuste  $B$  ja  $C$  vaheline kaugus  $h$ . Tähistame tuletatavat kaardistamisfunktsiooni, mis seab lookustevahelisele kaugusele vastavusse rekombinatsioonimäära,  $M(x)$ . Juhul, kui  $x$  on piisavalt väike, eeldatakse kehtivat samasus  $r_x = M(x) = x$ .

Lisaks tähistagu  $g_{\kappa_i}$  rekombinatsioonide kombinatsioonis  $\kappa_i$  esinemise tõenäosust (näiteks märgib  $g_{10}$  sündmuse, et esimeses intervallis toimus ja teises ei toimunud rekombinatsiooni, tõenäosust).

Sellise tähistuse abil on lookuste  $A$  ja  $C$  vahel aset leidnud rekombinatsiooni tõenäosus esitatav summana  $r_{AC} = g_{10} + g_{01}$ . Interferentsuse puudumisel saab selle avaldise arendada kujule:

$$r_{AC} = g_{10} + g_{01} = r_{AB}(1 - r_{BC}) + (1 - r_{AB})r_{BC} = r_{AB} + r_{BC} - 2r_{AB}r_{BC},$$

kus  $r_{AB}r_{BC} = g_{11}$  on kahekordse rekombinatsiooni tõenäosus. Kui interferentsust ei saa välistada, on kahekordse rekombinatsiooni tõenäosus esitatav valemiga  $g_{11} = c \times r_{AB}r_{BC}$  ja lookuste  $A$  ja  $C$  vaheline rekombinatsioonimäär avaldub kujul

$$r_{AC} = r_{AB} + r_{BC} - 2c r_{AB}r_{BC},$$

kus  $c$  on koosinemise kordaja. Asendades viimases valemis rekombinatsioonimäärad kaardistamisfunktsiooniga, saame võrduse

$$M(x + h) = M(x) + M(h) - 2c M(x)M(h),$$

mille võime esitada ka kujul

$$\frac{M(x + h) - M(x)}{h} = \frac{M(h) - 2c M(x)M(h)}{h}. \quad (3.7)$$

Minnes avaldises (3.7) piirile  $h \rightarrow 0$ , saame, et  $\frac{M(h)}{h} \rightarrow 1$  ja seega

$$\frac{dr_x}{dx} = 1 - 2c_0 r_x, \quad (3.8)$$

millest

$$x = \int \frac{1}{1 - 2c_0 r_x} dr_x = \frac{-1}{2c_0} \ln(1 - 2c_0 r_x). \quad (3.9)$$

Kordajat  $c_0$  nimetatakse **marginalseks koosinemise kordajaks** ja ta kujutab enesest kordaja  $c$  piirväärtust juhul, kui üks kahest kõrvuti olevast intervallist läheneb nullile.

Esitades võrrandi (3.9) kohal  $c_0 = 1$ , saame tulemuseks Haldane kaardistamisfunktsiooni, mis on ka loomulik, sest  $c_0 = 1$  korral toimuvad rekombinatsioonid kahes vaadeldavas piirkonnas sõltumatult ja seega on ka Poissoni jaotuse eeldused täidetud.

Samas on loomulik, et väiksemale rekombinatsioonimäärale vastab ka väiksem koosinemise kordaja. Teoreetiliselt võib nii saada lõpmatu arvu kaardistamisfunktsioone. D. D. Kosambi märkis 1944. aastal, et vaadates kordajat  $c_0$  rekombinatsioonimäärana  $r_x$  funktsioonina, saame ühe konkreetse kaardistamisfunktsiooni. Lihtsaim selline funktsioon, mis kasvab vahemikus  $0 < r_x < 1/2$ , võrdub nulliga kui  $r_x = 0$  ja ühega kui  $r_x = 1/2$ , on  $c_0 = 2r_x$ . Diferentsiaalvõrrand (3.8) saab siis kuju

$$\frac{dr_x}{dx} = 1 - 2c_0r_x = 1 - 4r_x^2.$$

Selle lahendamisel saame **Kosambi kaardistamisfunktsiooni**

$$r_x = \frac{1}{2} \frac{e^{4x} - 1}{e^{4x} + 1}, \tag{3.10}$$

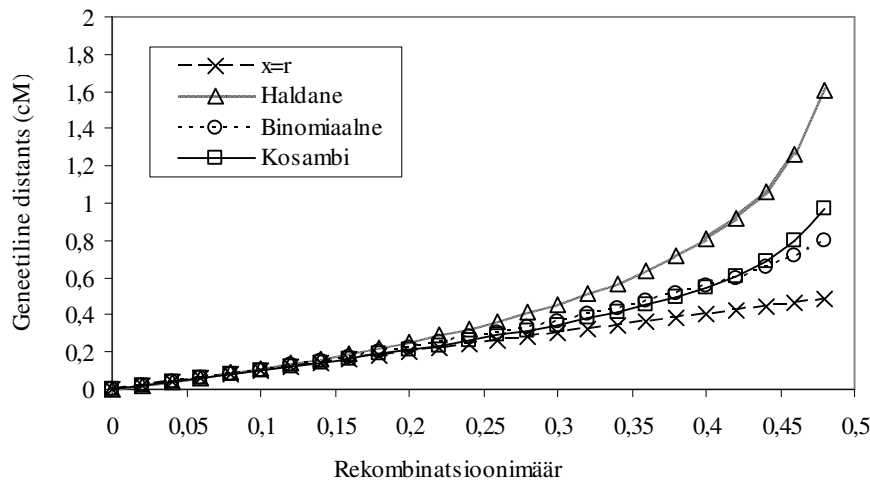
pöördfunktsiooniga

$$x = \frac{1}{4} \ln \frac{1 + 2r_x}{1 - 2r_x}. \tag{3.11}$$

Leides rekombinatsioonimäärale 22% vastava kauguse geneetilisel kaardil Kosambi kaardistamisfunktsiooni abil, saame valemist (3.11), et  $x = 0,236$  M ehk 23,6 cM.

Joonisel 15 on kujutatud mõningate kaardistamisfunktsioonide graafikud.

Lisaks siin kirjeldatutele on välja töötatud teisi kaardistamisfunktsioone. Enamus neist on tuletatavad valemist (3.8), andes vaid koosinemise kordajale uue kuju. Näiteks **Carter-Falconeri kaardistamisfunktsiooni** saame, lahendades diferentsiaalvõrrandi juhul  $c_0 = (2r_x)^3$ .



Joonis 15. Kaardistamisfunktsioonid (binomiaalse kaardistamisfunktsiooni korral  $N = 2$ ).

### 3.1.3 Rekombinatsiooni(de) tõenäosuste arvutamine paljude lookuste korral

Järgnevalt tuletame seosed, mis võimaldavad leida paljudes intervallides toimuda võivate rekombinatsioonide nn **ühistõenäosust**, arvestades kõiki teadaolevaid andmeid markerite ja nende rekombineerumissageduste kohta. Olgu meil vaatluse all  $k$  lookust. Võimalikud rekombinatsioonid neis lookustes asuvate alleelide sugurakkudesse jagunemisel on kirjeldatavad  $2^{k-1} \times k - 1$ -maatriksiga, mille  $i$ . reavektori  $\kappa_i$   $j$ . element võrdub ühega, kui  $j$ . ja  $j+1$ . lookuse vahel leidis aset rekombinatsioon, ja nulliga rekombinatsiooni puudumisel.

Kõigi  $2^{k-1}$  rekombinatsioonide kombinatsiooni ühistõenäosuste leidmiseks toome sisse hüpoteetilise kromosoomi piirkonna  $W_{\delta_i}$  ja sellele vastava  $k - 1 \times 1$ -vektori  $\delta_i$ , mille  $j$ . element võrdub ühega, kui  $j$ . ja  $j+1$ . lookuse vaheline intervall kuulub piirkonda  $W_{\delta_i}$ , ja nulliga vastupidisel juhul. Selliseid vektoreid  $\delta_i$  on kokku  $2^{k-1}$ , sest just nii palju on erinevaid võimalusi  $k - 1$  lookuste vahelise intervalli kuulumiseks või mittekuulumiseks mingisse piirkonda. Piirkonna  $W_{\delta_i}$  ulatus on defineeritud seosega

$$x(\delta_i) = \sum_{j=1}^{k-1} \delta_{ji} x_j,$$

kus  $x_j$  on  $j$ . ja  $j+1$ . lookuse vaheline kaugus geneetilisel kaardil.

Paaritu arvu rekombinatsioonide aset leidmise tõenäosust piirkonnas  $W_{\delta_i}$  tähistatakse  $R(\delta_i)$  ja nimetatakse vektorile  $\delta_i$  vastavaks **rekombinatsiooniväärtuseks**. Kaardistamisfunktsiooni  $r(x) = M(x)$  korral on rekombinatsiooniväärtus esitatav võrdusena

$$R(\delta_i) = M[x(\delta_i)].$$

Rekombinatsiooniväärtus  $\delta_i$  suhtes on leitav ka nende sündmuste, mille korral toimub piirkonnas  $W_{\delta_i}$  paaritu arv rekombinatsioone, tõenäosuste  $g_{\kappa_i}$  summa. Näiteks kui  $k=4$  ja  $\delta_i = [1 \ 0 \ 1]^T$ , siis  $R(\delta_i) = g_{001} + g_{011} + g_{100} + g_{110}$ .

Piirkonnas  $W_{\delta_i}$  aset leidvate rekombinatsioonide arv  $s_{ij}$  on rekombinatsioonide kombinatsiooni  $\kappa_j$  korral leitav korrutisest

$$s_{ij} = \delta_i^T \kappa_j^T \quad (i, j = 1, \dots, 2^{k-1}).$$

Seega on vektorile  $\delta_i$  vastav rekombinatsiooniväärtus esitatav ka seosena

$$R(\delta_i) = \sum_{\substack{j, \text{ kui } s_{ij} \\ \text{on paaritu}}} g_{\kappa_j} = \frac{1}{2} \left[ 1 - \sum_{j=1}^{2^{k-1}} (-1)^{s_{ij}} g_{\kappa_j} \right].$$

Meie näite korral, kus  $k=4$  ja  $\delta_i = [1 \ 0 \ 1]^T$ , saame viimase valemi rakendamisel juba eelnevalt tuletatud seose:

$$\begin{aligned} R(\delta_i) &= \frac{1}{2} \left[ 1 - \sum_{j=1}^{2^{k-1}} (-1)^{s_{ij}} g_{\kappa_j} \right] \\ &= \frac{1}{2} \{ 1 - [(-1)^0 g_{000} + (-1)^1 g_{001} + (-1)^0 g_{010} + (-1)^1 g_{100} \\ &\quad + (-1)^1 g_{011} + (-1)^2 g_{101} + (-1)^1 g_{110} + (-1)^2 g_{111}] \} \\ &= \frac{1}{2} \{ 1 - (-g_{001} - g_{100} - g_{011} - g_{110} + g_{000} + g_{010} + g_{101} + g_{111}) \} \\ &= \frac{1}{2} \{ g_{001} + g_{100} + g_{011} + g_{110} + [1 - (g_{000} + g_{010} + g_{101} + g_{111})] \} \\ &= \frac{1}{2} [2 \times (g_{001} + g_{100} + g_{011} + g_{110})] \\ &= g_{001} + g_{100} + g_{011} + g_{110}. \end{aligned}$$

Maatriksvõrdusena on seos rekombinatsiooniväärtuste  $R(\delta_i)$  ja tõenäosuste  $g_{\kappa_i}$  vahel esitatav kujul

$$\mathbf{r} = \frac{1}{2} (\mathbf{1} - \mathbf{A}\mathbf{g}), \quad (3.12)$$

kus  $\mathbf{r}$  on rekombinatsioonimäärade  $2^{k-1} \times 1$ -vektor,  $\mathbf{1}$  on ühtedest koosnev  $2^{k-1} \times 1$ -vektor, maatriks  $\mathbf{A}$  on  $2^{k-1} \times 2^{k-1}$ -maatriks, mille  $ij$ . element  $a_{ij} = (-1)^{s_{ij}}$ , ja  $\mathbf{g}$  on rekombinatsioonisündmuste ühistõenäosuste  $2^{k-1} \times 1$ -vektor. Valemist (3.12) saame, et

$$\mathbf{A}\mathbf{g} = \mathbf{1} - 2\mathbf{r}.$$

Maatriksi  $\mathbf{A}$  definitsioonist on selge et  $\mathbf{A} = \mathbf{A}^T$  (maatriks  $\mathbf{A}$  on sümmeetriline) ja et  $\mathbf{a}_i^T \mathbf{a}_j = 2^{k-1}$ , kui  $i = j = 1, \dots, 2^{k-1}$  ning  $\mathbf{a}_i^T \mathbf{a}_j = 0$ , kui  $i \neq j$ . Siis  $\mathbf{A}\mathbf{A} = 2^{k-1} \mathbf{I}$  ja  $\mathbf{A}^{-1} = \frac{1}{2^{k-1}} \mathbf{A}$ . Ühistõenäosuste vektor  $\mathbf{g}$  on nüüd avaldatav seosest

$$\mathbf{g} = \mathbf{A}^{-1}(\mathbf{1} - 2\mathbf{r}) = \frac{\mathbf{A}(\mathbf{1} - 2\mathbf{r})}{2^{k-1}}. \quad (3.13)$$

Iga üksiku rekombinatsioonide kombinatsiooni tõenäosus on siis leitav valemist

$$g_{\kappa_i} = \sum_{j=1}^{2^{k-1}} (-1)^{s_{ij}} \frac{1 - 2R(\delta_j)}{2^{k-1}}.$$

**Näide.** Olgu meil vaatluse all neli lookust ( $k=4$ ), rekombinatsioonimääradega  $r_j$  kolmes intervallis vastavalt 0,1, 0,05 ja 0,2. Nendele rekombinatsioonimääradele vastavad Haldane, Kosambi ja binomiaalse ( $N=2$ ) kaardistamisfunktsiooni abil arvutatud vahemaad geneetilisel kaardil  $x_j$  on toodud tabelis. Vektorid  $\kappa_i$  ja  $\delta_i$  on kirjas maatriksi  $\mathbf{U}$  ridadena ja rekombinatsioonide arvud  $s_{ij}$  maatriksina  $\mathbf{S}$ .

Intervall	$r_j$	$x_j$		
$j$		Haldane	Binomiaalne	Kosambi
1	0,1	0,1116	0,1056	0,1014
2	0,05	0,0527	0,0513	0,0502
3	0,2	0,2554	0,2254	0,2118

$$U = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}, S = UU^T = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 2 & 1 & 2 \\ 0 & 0 & 1 & 1 & 1 & 1 & 2 & 2 \\ 0 & 1 & 1 & 2 & 1 & 2 & 2 & 3 \end{bmatrix}.$$

Maatriks **A** on siis

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}.$$

Piirkonnale  $W_{\delta_i}$  vastavad vahemaad geneetilisel kaardil  $x(\delta_i)$  ja vastavad Haldane, Kosambi ja binomiaalse ( $N = 2$ ) kaardistamisfunktsiooni abil arvutatud rekombinatsiooniväärtused  $R(\delta_i)$  on kirjas järgnevas tabelis.

$\delta_i^T$	Haldane		Binomiaalne		Kosambi	
	$x(\delta_i)$	$R(\delta_i)$	$x(\delta_i)$	$R(\delta_i)$	$x(\delta_i)$	$R(\delta_i)$
0 0 0	0	0	0	0	0	0
0 0 1	0,2554	0,2	0,2254	0,2	0,2118	0,2
0 1 0	0,0527	0,05	0,0513	0,05	0,0502	0,05
0 1 1	0,3081	0,23	0,2767	0,2384	0,2620	0,2404
1 0 0	0,1116	0,1	0,1056	0,1	0,1014	0,1
1 0 1	0,3670	0,26	0,3310	0,2762	0,3132	0,2778
1 1 0	0,1642	0,14	0,1569	0,1446	0,1516	0,1471
1 1 1	0,4197	0,284	0,3823	0,3092	0,3634	0,3105

Pannes arvutatud rekombinatsiooniväärtused ja leitud maatriksi **A** valemisse (3.13), saamegi otsitud rekombinatsioonide ühistõenäosused  $g_{\kappa_i}$ :

$\kappa_i$	$g_{\kappa_i}$		
	Haldane	Binomiaalne	Kosambi
0 0 0	0,684	0,6704	0,6687
0 0 1	0,171	0,1823	0,1829
0 1 0	0,036	0,0415	0,0425
0 1 1	0,009	0,0058	0,0060
1 0 0	0,076	0,0854	0,0862
1 0 1	0,019	0,0119	0,0123
1 1 0	0,004	0,0027	0,0026
1 1 1	0,001	0	-0,0012

Märkus1. Interferentsuse puudumisel on rekombinatsioonide tõenäosused Haldane kaardistamisfunktsiooni korral leitavad ka lihtsamalt. Näiteks

$$g_{110} = nr_2(1 - r_3) = 0,1 \times 0,05 \times (1 - 0,2) = 0,004.$$

Märkus2. Kosambi kaardistamisfunktsioon ei ole kasutatav enam kui 3 lookuse korral. Sellest annab märku ka negatiivne ühistõenäosus  $g_{111}$ .



**Ülesanne 3.**

Näidake, et kui ristsiirde võimalikud toimumiskohad järgivad statsionaarset Poissoni protsessi, siis järgivad Poissoni protsessi ka juhuslikult valitud meioosil toimuvate ristsiirete asukohad (eeldame interferentsuse puudumist).

**Vihjed.** Jagame kromosoomi mittelõikuvateks lõplikuks arvuks intervallideks  $I_1, I_2, \dots, I_k$ . Tähistagu  $N(i)$  ristsiirde võimalike toimumiskohtade (kiasmide) arvu intervallis  $I_i$  ja  $M(i)$  juhuslikult valitud meioosil toimunud ristsiirete arvu intervallis  $I_i$ .

Vaja on näidata, et ristsiirete arvud  $M(i)$  on sõltumatud ja  $M(i) \sim P(d_i)$ , st et  $M(i)$  on Poissoni protsess ( $d_i$  – intervalli  $I_i$  pikkus).

Ja veel – et eeldasime interferentsuse puudumist, siis igas potentsiaalses kohas toimub ristsiire tõenäosusega  $1/2$ .

**Ülesanne 4.**

Tuletage Carter-Falconeri kaardistamisfunktsiooni avaldis geneetilisele kaardile vastava kauguse  $x$  leidmiseks rekombinatsioonimäärast  $r_x$ , võttes diferentsiaalvõrrandis (3.8) marginaalseks koosinemise kordajaks  $c_0 = (2r_x)^3$ . Kasutades tuletatud avaldist, leidke rekombinatsioonimäärale 22% vastav kaugus geneetilisel kaardil.

**Ülesanne 5.**

Olgu meil kolm lookust  $A, B$  ja  $C$  rekombinatsioonimääradega vastavalt  $r_{AB} = 0,3$  ja  $r_{BC} = 0,1$ .

- Arvutage lookuste  $A$  ja  $B$  ning  $B$  ja  $C$  vahelised geneetilised distantid cM-tes ning teiseks kogu intervalli ( $A$  kuni  $C$ ) pikkus tagasi rekombinatsiooni sagedusele. Kasutage paralleelselt nii Haldane kui ka Kosambi kaardistamisfunktsioone. Miks on erinevad nõ otse arvatud ja kaardistamisfunktsioonidest kogu cM-tes mõõdetud geneetilise distantsi alusel leitud rekombinatsioonimäärad?
- Leidke lookuste  $A$  ja  $C$  vaheliste rekombinatsioonide ühistõenäosused.
- Võtame arvesse veel ka markeri  $D$ , kus  $r_{CD} = 0,05$ . Leidke  $g_{011}$ , so sündmuse, et rekombinatsioonid toimuvad intervallides  $B - C$  ja  $C - D$ , aga ei toimu intervallis  $A - B$ , ühistõenäosus.

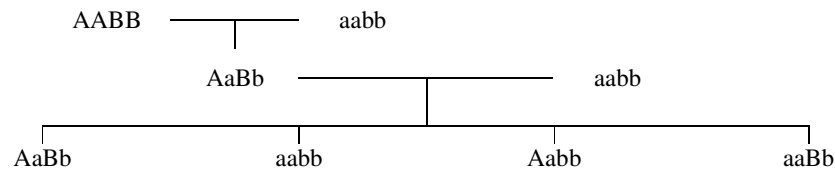
**3.1.4 Aheldusanalüüsi olemus, markerite informatiivsus ja identsus päritolult**

Oma esialgselt definitsioonilt on aheldusanalüüs meetod, mis baseeruvana Morgani seadustel, püüab teha järeldusi uuritud lookuste suhtelise paiknemise kohta genoomis. Lihtsaimal, vaid kahe lookuse analüüsil, taandub probleem nende lookuste aheldumuse testimisele ja nende vahelise rekombineerumise tõenäosuse hindamisele. Juhul, kui juhuslikult valitud indiviidide genoomi moodustanud sugurakkude kohta on täpselt kokku loetavad vaadeldavate lookuste suhtes rekombinantsed gameetid, seisneb ahelduse test kontrollimises, kas rekombinantsete gameetide proportsioon võrdub  $1/2$ -ga (ahelduse puudumisele vastav nullhüpotees) või on sellest väiksem (lookuste aheldust kinnitav sisukas hüpotees).

Et igal indiviidil on võimalik määrata vaid tema genotüüp uuritavate markerlookuste osas, siis iga genotüübi kahe haplotüübina esitamiseks on lisaks vaja teada ka vanemate genotüüpe (et tuvastada, kummalt vanemalt konkreetne alleel pärandus, ehk millised alleelid moodustavad ühe haplotüübi). Selleks, et osata öelda, kas uuritava indiviidi ühe haplotüübina identifitseeritud alleelid paiknevad samas haplotüübis ka selle indiviidi vanemal, ehk kas sugurakkude moodustumisel vanema organismis on nende alleelide vahel aset leidnud rekombinatsioon või mitte, on vaja teada ka vanavanemate genotüüpe.

Täpsustamaks, kas genotüübi  $A_1A_2B_1B_2$  moodustavad haplotüübid  $A_1B_1$  ja  $A_2B_2$  või  $A_1B_2$  ja  $A_2B_1$ , kasutatakse tähistusi  $A_1B_1|A_2B_2$  või  $A_1B_2|A_2B_1$ , mis aheldusanalüüsi seisukohast märgivad erinevaid genotüüpe. Juhul, kui võtta arvesse ka see, kummalt vanemalt mingi haplotüüp pärandus, võib esialgne genotüübi kirjeldus vastata neljale erinevale genotüübile:  $A_1B_1|A_2B_2$ ,  $A_2B_2|A_1B_1$ ,  $A_1B_2|A_2B_1$  või  $A_2B_1|A_1B_2$  (kus traditsiooniliselt on esimesena märgitud haplotüüp pärit isalt). Viimati defineeritud genotüüpe nimetatakse ka **järjestatud genotüüpideks**.

**Näide.** Vaatleme kolme generatsiooni ühes perekonnas, kus genotüüp on määratud 2 dialleelse markeri osas.



Toodud põlvnemisskeem hõlmab 8 indiviidi ja seega potentsiaalselt 16 erinevat haplotüüpi. Esimeses nn vanavanemate generatsioonis on indiviidide haplotüübid tänu homosügootsusele tuvastatavad (AB|AB vanaisal ja ab|ab vanaemal), kuid võimalikke rekombinatsioone, mille tulemusel sellised haplotüübid võisid tekkida, pole eelneva generatsiooni mitteteadmise tõttu võimalik leida. Teises nn vanemate generatsioonis on haplotüübid samuti leitavad (AB|ab isal ja ab|ab emal), kuid jällegi pole võimalik tuvastada rekombinatsioone. Isa puhul on see tingitud tema vanemate homosügootsusest mõlema lookuse suhtes – ükskõik, kas rekombinatsioon toimus või mitte, meioosi lõpuks moodustunud sugurakud on uuritud lookuste suhtes ikka ühesugused; ema kohta pole aga jällegi teada tema vanemate genotüüpi.

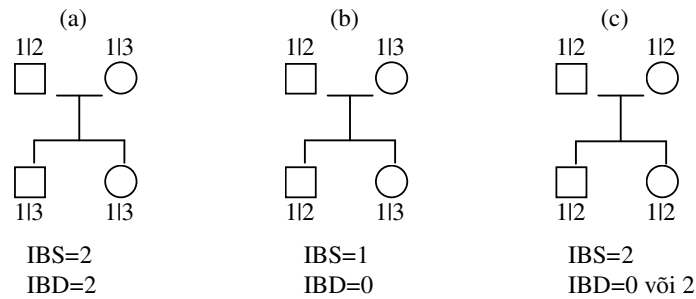
Kolmanda generatsiooni indiviidid pärivad kõik emalt haplotüübi ab ega anna seega mingit informatsiooni võimaliku rekombinatsiooni kohta selle haplotüübi moodustumisel. Seega peavad nelja järglase teised haplotüübid pärandununa isalt sisaldama alleele AB, ab, Ab ja aB. Et isa genotüüp on juba jagatud haplotüüpideks kujul AB|ab, siis on selge, et kahe esimese järglase isalt päritud haplotüübid on mitterekombinantid, aga kahe ülejäänud järglase haplotüüpe moodustavate sugurakkude kujunemisel on uuritavate lookuste vahel toimunud rekombinatsioon.

Kokkuvõttes ei olnud vaatluse all olnud 8 indiviidi 16 haplotüübist tervelt 12 korral mingit võimalust tuvastada kas rekombinatsioon toimus või mitte ning seega on need 12 haplotüüpi aheldusanalüüsi seisukohast ebainformatiivsed. Ülejäänud 4 haplotüüpi öeldakse olevat **täielikult informatiivsed** (*fully informative*), sest nende rekombineeritus (*recombination status*) on täpselt tuvastatav (analoogne põhimõtte alusel defineeritakse mõnes kirjutises ka informatiivsed meioosid ja/või gameetid).

Kuigi sarnaselt kompleksele segregatsioonanalüüsile on ka aheldusanalüüsi puhul välja kirjutatavad üldised tõepärafunktsioonid arvestamiseks mitteteadaolevate haplotüüpide puhul kõigi võimalike alleelikombinatsioonidega, püütakse enamasti uuritavaiks valida vaid informatiivseid indiviide (so siis neid, kelle puhul on täpselt teada, millise alleeli/haplotüübi ta milliselt vanemalt päris). Informatsiooni sellest, millistest haplotüüpidest vanemate sugurakud (ehk siis uuritava järglase haplotüübid) meioosi protsessil kokku pannakse, nimetatakse **meioosi faasiks** (ka ahelduse faasiks – *phase of meiosis, linkage phase*).

Enamasti on isalt ja emalt päritud haplotüübi (markeralleeli) eristamine indiviidide genotüüpides vajalik selgitamiseks, kas kahe indiviidi samas lookuses paiknevad oma **olemuselt identsed** alleelid (*identical-by-state* – IBS) on **identsed** ka **päritolult** (*identical-by-descent* – IBD), st et kas nad on ühise eellase vastava alleeli füüsilised koopiad.

**Näide.** Olgu kolmes perekonnas (isa, ema ja 2 last), kus eelaste kohta midagi teada ei ole, genotüüp määratud ühe markerlookuse osas. Alleelid vaadeldavas lookuses on tähistatud numbritega 1, 2 ja 3 ning perekonnad (a), (b) ja (c). Päritolult identsete (IBD) ja olemuselt identsete (IBS) alleelide arvud on toodud iga perekonna all.



Perekonna (a) puhul on mõlema haplotüübi pärandumine üheselt tuvastatav – alleel 3 järglaste genotüübis saab olla pärit vaid emalt, mistõttu peab alleel 1 mõlemal järglasel olema saadud isalt, ning kokkuvõttes seega  $IBS=IBD=2$ .

Perekonna (b) puhul on samuti haplotüüpide pärandumine üheselt tuvastatav, Mõlemal järglasel on üks olemuselt identne alleel (1), aga et see alleel on saadud erinevatel vanematelt, kes eeldatakse suguluses mitteolevaiks, siis  $IBD=0$ .

Perekonnas (c) on mõlemad järglased uuritud lookuse suhtes geneetiliselt identsed, st  $IBS=2$ , samas ei ole aga võimalik tuvastada alleelide täpset pärandumist, st et võrdse tõenäosusega (Mendeli seaduste kohaselt) jagavad kaks järglast kas kahte või mitte ühtegi päritolult identset alleeli.

Põhigeenide ehk QTL-ide otsingul aheldusanalüüsi abil on kaks põhilist lähenemist. Nõ **parameetrilise aheldusanalüüsi** korral püstitatakse kõigepealt mudel kirjeldamaks fenotüübiväärtusi oletatavatele geneetilistele efektidele tuginedes, hinnatakse mudeli parameetrid arvestades, et kahe lookuse (seega ka tunnust määrava geeni lookuse ja markerlookuse) vaheline kaugus on proportsionaalne nende vahel aset leidnud rekombinatsioonide arvuga, ning püütakse välja selgitada, kas konstrueeritud mudel peegeldab fenotüübiväärtusi paremini võrreldes mingi lihtsama mudeliga. Nõ **mitteparameetriline** e **mudelivaba aheldusanalüüs** püüab indiviidide suuremat fenotüübilist sarnasust põhjendada nende suurema päritolult identsete alleelide arvu abil, fikseerimata mingit matemaatilist feno- ja genotüübi vahelist seost.