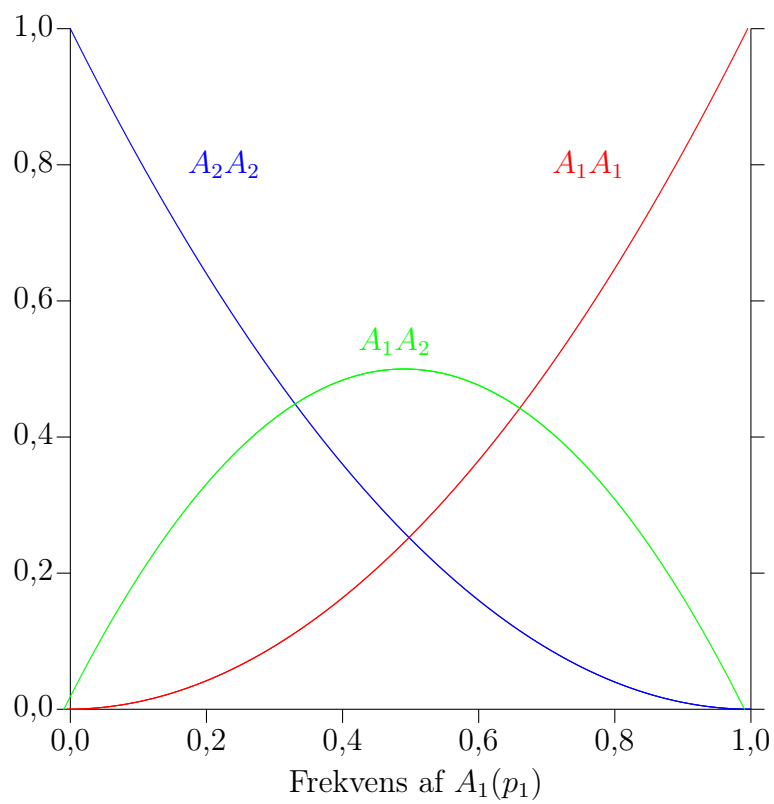


Introduktion til populationsgenetik

Hans R. Siegismund



Biologisk Institut
Københavns Universitet
2019

Indhold

1	Genetisk variation	1
1.1	Variation i enkelte loci	1
1.1.1	Genotype- og allelfrekvenser	2
1.1.2	Hardy-Weinberg-loven	3
1.1.3	Variation i loci med dominante alleler	7
1.1.4	Variation i kønsbundne loci	8
1.1.5	Variation i to loci	8
1.1.6	Genetisk diversitet	11
1.2	Variation på DNA-niveau	11
1.2.1	Nukleotiddiversitet	12
1.3	Fænotypisk variation	13
1.3.1	Komponenter af kvantitative karakterer	14
1.3.2	Heritabilitet	15
	Indeks	19

Kapitel 1

Genetisk variation

Genetisk variation er forudsætningen for evolution. Uden variation kan der ikke ske nogen evolution. Det er derfor vigtigt at kunne kvantificere genetisk variation. Den kan undersøges på flere niveauer. Starter vi på det nederste niveau, kan vi få information om selve **DNA-sekvensen** af det område af genomet, man er interesseret i. Variationen på et bestemt basepars plads vil så være en af de fire baser, A, C, G eller T. næste trin er at beskrive variationen i et **locus**, som betegner et bestemt område af genomet – ofte bruges det som reference til et gens placering. Her betegnes de forskellige former som alleler. Det kan for eksempel være proteinkodende gener, mikro- eller minisatelliter. Generne er lokaliseret på **kromosomer**. Disses antal og form kan være variabel og udgør således en del af en arts genetiske variation.

Den genetiske variation, der studeres med de ovennævnte metoder, består af diskrete enheder, hvis transmission fra generation til generation i princippet kan følges. Dette gælder ikke for **fænotypen**, der er under indflydelse af et større antal gener, samt af det omgivende miljø. Eksempler på sådanne karakterer er størrelse, farve og som det nok vigtigste individuel fitness, det vil sige overlevelse og reproduktiv succes.

I det følgende vil vi se på metoder til beskrivelse og måling af genetisk variation på de fire forskellige niveauer. Som udgangspunkt analyserer vi stikprøver taget fra populationer. Vi starter – ud fra historiske grunde – med variation i enkelte loci. Derefter følger variation på DNA- og fænotypisk niveau. Variation på kromosomalt niveau kan ofte beskrives ved hjælp af de samme metoder som for variation i enkelte loci.

1.1 Variation i enkelte loci

De næste afsnit gennemgår genetisk variation på enkelt-locus niveau. De tjener som et fundament for flere af de senere afsnit. Formålet er at se på simple love, der beskriver og forklarer fordelingen af genetisk materiale i naturlige populationer. Selv om disse love udledes for variationen i et enkelt locus, så kan de anvendes for de fleste metoder, man benytter for at beskrive genetisk variation.

1.1.1 Genotype- og allelfrekvenser

Lad os undersøge en population af en diploid dyreart for genetisk variation, som kodes af et enkelt locus. Vi vil sjældent kunne bestemme variationen i alle individer, men må nøjes med at tage en stikprøve. Vi antager, at det er en tilfældigt udtaget del af hele populationen og bekymrer os foreløbigt ikke om, hvordan man bestemmer et individs genotype i et givet locus. Endvidere antager vi, at den genetiske variation er codominant, det vil sige, at vi kan observere begge alleler i et heterozygot individ. Populationen er polymorf i et autosomalt locus med to alleler, A_1 og A_2 , (en egentlig definition og diskussion af begrebet polymorfi gives i et senere afsnit). Med to alleler findes derfor tre genotyper, A_1A_1 , A_1A_2 og A_2A_2 . **Genotypefrekvensen** er ens blandt hanner og hunner og beskrives med parametrene, P_{11} , P_{12} og P_{22} .

Vi tager en stikprøve på N individer. Disse N fordeler sig på de tre genotyper som følger

Genotype	A_1A_1	A_1A_2	A_2A_2	Sum
Antal	N_{11}	N_{12}	N_{22}	N
Frekvens	P_{11}	P_{12}	P_{22}	1

Indet Frekvensen P_{ij} af genotypen A_iA_j i stikprøven beregnes som brøkdelen N_{ij}/N . **Allelfrekvenserne** af A_1 og A_2 bestemmes ved simpel optælling af de to alleler i stikprøven, og ved at huske at homozygoter bærer to alleler af samme slags. Vi har derfor

$$\begin{aligned}
 p_1 &= \frac{2N_{11} + N_{12}}{2N} = \frac{N_1}{2N} \\
 &= P_{11} + \frac{1}{2}P_{12} \\
 p_2 &= \frac{N_{12} + 2N_{22}}{2N} = \frac{N_2}{2N} \\
 &= \frac{1}{2}P_{12} + P_{22}
 \end{aligned} \tag{1.1}$$

Summen af allelfrekvenserne er selvfølgelig 1

$$p_1 + p_2 = P_{11} + \frac{1}{2}P_{12} + \frac{1}{2}P_{12} + P_{22} = 1.$$

Et eksempel fra naturen

Bestemmelse af allelfrekvenser og deres varianter vil her blive illustreret med et eksempel fra den marine isopod *Jaera ischiosetosa* (Siegismund & Christensen 1992). En stikprøve indsamlet i en population ved Løgstør i Limfjorden blev undersøgt med enzymelektroforese for fire polymorfe loci. Et af disse loci (*Gpi*, glukose fosfat isomerase) gav følgende resultat hos hannerne

Genotype	A_1A_1	A_1A_2	A_2A_2	Sum
Antal	10	40	49	99
Frekvens	0,101	0,404	0,495	1,000

Frekvenserne af de to alleler bestemmes til

$$p_1 = \frac{2 \times 10 + 40}{2 \times 99} = \frac{60}{198} = 0,303$$

$$p_2 = \frac{40 + 2 \times 49}{2 \times 99} = \frac{138}{198} = 0,697,$$

hvor vi som kontrol undersøger, at

$$p_1 + p_2 = 0,303 + 0,697 \\ = 1,000.$$

1.1.2 Hardy-Weinberg-loven

Hidtil har vi beskrevet genetisk variation med genotypefrekvenser og har på grundlag af dem beregnet allelfrekvenser. Vi vil nu se på, under hvilke forhold, vi kan nøjes med at benytte allelfrekvenser, og ud fra disse eventuelt kan beregne frekvenserne af genotyperne. Dette gør det nemmere at beskrive den genetiske variation. I stedet for frekvensen af tre genotyper (hvoraf dog kun de to er uafhængige, da deres sum er 1) kan vi benytte to allelfrekvenser (hvor kun den ene er uafhængig).

Vi betragter en population af diploide individer, der er kontinuert udbredte i tid og rum. Vi ser her bort fra besværlige økologiske parametre som tæthed, ubegrænset vækstrate og bærekapacitet. Populationen antages at bestå af hanner og hunner, som udelukkende forplanter sig seksuelt. Populationens parametre antages at være kendte: hyppighederne af genotyperne A_1A_1 , A_1A_2 og A_2A_2 er henholdsvis P_{11} , P_{12} og P_{22} hos begge køn. Hvad er hyppigheden i den næste generation? For at besvare dette spørgsmål må vi gøre følgende antagelser:

- ens genotypefrekvens hos hanner og hunner
- ikke overlappende generationer
- tilfældig parring
- Mendelsk udspaltning
- uendelig stor population
- ingen mutation
- ingen migøration
- ingen selektion

Med tilfældig parring i populationen findes hyppigheden af forældre-kombinationen $A_1A_1 \times A_1A_1$ som P_{11}^2 . Dette giver udelukkende A_1A_1 afkom. Kombinationen af $A_1A_1 \times A_1A_2$ har hyppigheden $2P_{11}P_{12}$, fordi hunnen kan være homozygot og hannen heterozygot eller omvendt. En sådan parring vil med Mendelsk udspaltning resultere i lige mange A_1A_1 og A_1A_2 afkom. I Tabel 1.1 er samtlige seks kombinationer af forældre opført sammen med hyppigheden af deres afkom.

Populationen har nu efter én generation under de betingelser, der er angivet foroven, opnået en stabil fordeling af genotypernes hyppigheder. Fordelingen af genotyperne blandt afkommet bliver

$$\begin{array}{ccc} A_1A_1 & A_1A_2 & A_2A_2 \\ p_1^2 & 2p_1p_2 & p_2^2. \end{array} \quad (1.2)$$

Tabel 1.1: Kombinationer af hanner og hunner ved tilfældig parring samt Mendelsk udspaltning i afkommet.

Forældre-kombination	Frekvens	Afkom		
		A_1A_1	A_1A_2	A_2A_2
$A_1A_1 \times A_1A_1$	P_{11}^2	P_{11}^2		
$A_1A_1 \times A_1A_2$	$2P_{11}P_{12}$	$P_{11}P_{12}$	$P_{11}P_{12}$	
$A_1A_1 \times A_2A_2$	$2P_{11}P_{22}$		$2P_{11}P_{22}$	
$A_1A_2 \times A_1A_2$	P_{12}^2	$\frac{1}{4}P_{12}^2$	$\frac{1}{2}P_{12}^2$	$\frac{1}{4}P_{12}^2$
$A_1A_2 \times A_2A_2$	$2P_{12}P_{22}$		$P_{12}P_{22}$	$P_{12}P_{22}$
$A_2A_2 \times A_2A_2$	P_{22}^2			P_{22}^2
Sum	1	$(P_{11} + \frac{1}{2}P_{12})^2$	$2(P_{11} + \frac{1}{2}P_{12}) \times (P_{22} + \frac{1}{2}P_{12})$	$(P_{22} + \frac{1}{2}P_{12})^2$
		$=$ p_1^2	$=$ $2p_1p_2$	$=$ p_2^2

Dette stationære punkt betegnes **Hardy-Weinberg-ligevægt** efter den engelske matematiker G. H. Hardy og den tyske læge W. Weinberg, som begge i 1908 formulerede denne generelle lov. En særdeles vigtig egenskab er, at allelfrekvensen forbliver den samme. Endvidere gælder der, at hvis populationen én gang er kommet i Hardy-Weinberg-ligevægt, forbliver den der.

Hardy-Weinberg-ligevægten kan findes lidt nemmere end vist i Tabel 1.1, idet tilfældig parring mellem hanner og hunner samt Mendelsk udspaltning faktisk svarer til tilfældig sammensmeltning af gameter. Dette er vist i Tabel 1.2. Det ses, at genotypefordelingen er den samme som i ligning (1.2).

Man bør vogte sig for at drage slutningen, at hvis en polymorfi udviser Hardy-Weinberg-proportioner, så opfylder den alle betingelser for ligevægt. *Dette er ikke tilfældet!* Et simpelt eksempel kan tjene som illustration: Antag, at begge homozygoter er sterile, så kun heterozygoter før afkom. Hvis vi betragter populationen af voksne, har vi et perfekt

Tabel 1.2: Tilfældig sammensmeltning af gameter resulterer i Hardy-Weinberg-ligevægt blandt afkommet.

		Hun-gameter	
		A_1	A_2
Han-gameter		p_1	p_2
	$A_1 p_1$	p_1^2	p_1p_2
	$A_2 p_2$	p_2p_1	p_2^2

fit til de forventede Hardy-Weinberg-proportioner, nemlig

$$\begin{array}{ccc} A_1A_1 & A_1A_2 & A_2A_2 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{array}$$

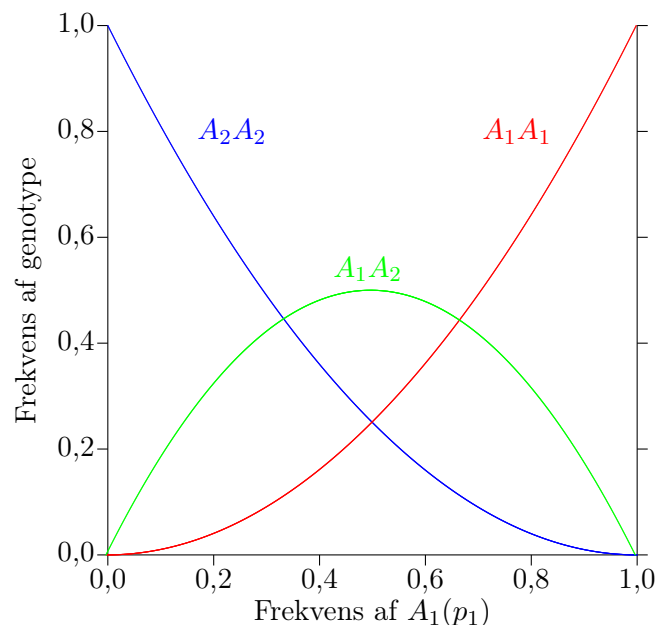
som er resultatet af udspaltningen af den eneste fertile kombination, $A_1A_2 \times A_1A_2$ i generationen før. Vi ser, at selv med en særdeles hård selektion kan man have en perfekt overensstemmelse med Hardy-Weinberg-frekvenser i en population.

Egenskaber af Hardy-Weinberg-fordelingen

I et autosomalt locus med to alleler er hyppigheden af de tre genotyper givet ved formel 1.2. Dette er illustreret i Figur 1.1. Det ses, at hvis allel A_1 har en meget lav frekvens, vil den hovedsageligt forekomme hos heterozygoter, da frekvensen af p_1^2 vil være meget lav. Dette illustreres tydeligt hos recessive arvelige sygdomme, hvor langt de fleste skadelige alleler findes hos raske heterozygote bærere i populationen. Figuren viser endvidere, at når der er to alleler i populationen vil hyppigheden af heterozygoter aldrig overstige $1/2$, som er den værdi hvor begge alleler er lige hyppige i populationen, d.v.s. $p_1 = p_2 = 1/2$.

Test af Hardy-Weinberg-proportioner

Når man har indsamlet en stikprøve fra en population og bestemt variationen i et autosomalt locus vil man ofte være interesseret i, hvorvidt genotypefordelingen stemmer overens med Hardy-Weinberg-proportioner. Betragt en stikprøve på N individer i et locus med



Figur 1.1: Frekvens af de tre genotyper som funktion af allelfrekvensen i et autosomalt locus med to alleler.

Tabel 1.3: Observerede genotyper og forventet Hardy-Weinberg-fordeling i et autosomalt locus med to alleler.

	A_1A_1	A_1A_2	A_2A_2	Sum
Observeret	N_{11}	N_{12}	N_{22}	N
Forventet	E_{11}	E_{12}	E_{22}	N
	$=$	$=$	$=$	
	p_1^2N	$2p_1p_2N$	p_2^2N	

to allele gener A_1 og A_2 . Allelfrekvenserne findes ved simpel optælling som angivet i ligning (1.1). Det er en god idé både at bestemme p_1 og p_2 ; hvis deres sum er lig med 1, har man regnet rigtigt. En test gennemføres ved at sammenligne de forventede med de observerede antal, se Tabel 1.3. Den observerede og den forventede fordeling sammenlignes med en χ^2 -test. Teststørrelsen

$$\chi^2 = \frac{(N_{11} - E_{11})^2}{E_{11}} + \frac{(N_{12} - E_{12})^2}{E_{12}} + \frac{(N_{22} - E_{22})^2}{E_{22}} \quad (1.3)$$

er tilnærmelsesvist χ^2 -fordelt med én frihedsgørad. Antallet af frihedsgrader fremkommer på følgende måde: Med en given stikprøvestørrelse på N har vi to frihedsgørader til at beskrive de tre genotyper med, da deres sum er N . Vi bruger én frihedsgørad til at estimere $p_1 (= 1 - p_2)$, så der bliver én frihedsgørad tilovers til testen. Er teststørrelsen større end 3,84 forkaster vi hypotesen på 5% niveauet; det vil sige, den genetiske variation i det betragtede locus anses for ikke at være i overensstemmelse med Hardy-Weinberg-proportioner.

Hvis vi accepterer hypotesen, kan vi kun udtale, at genotypfordelingen stemmer overens med Hardy-Weinberg-fordelingen; *vi kan ikke sige noget om, hvorvidt der er Hardy-Weinberg-ligevægt i populationen*. Når man laver en test for Hardy-Weinberg-proportioner, er det vigtigt at overveje om den genotypiske sammensætning eventuelt ændrer sig i løbet af artens livscyklus. Hos delvist selvbestøvende planter ses ofte, at en del af de indavlede individer elimineres i løbet af opvæksten. Her før man en forkastelse i en test for Hardy-Weinberg-proportioner hos tidlige udviklingsstadier, medens afvigelserne senere aftager og måske ikke vil kunne opdages ved en test. Hvis naturlig selektion virker på polymorfien, er det heller ikke ligegyldigt, hvornår man laver en test for Hardy-Weinberg-proportioner.

Vi kan nu bruge denne metode til at teste, om der er Hardy-Weinberg-proportioner i eksemplet med den marine isopod *Jaera ischiosetosa*. Sammenlignes den observerede med den forventede genotypfordeling

Genotype	A_1A_1	A_1A_2	A_2A_2	Sum
Observeret	10	40	49	99
Forventet	9,09	41,82	48,09	99,01

ved hjælp af udregningerne i ligning (1.3) fås

$$\begin{aligned}\chi^2 &= \frac{(10 - 9,09)^2}{9,09} + \frac{(40 - 41,82)^2}{41,82} + \frac{(49 - 48,09)^2}{48,09} \\ &= 0,19,\end{aligned}$$

hvilket ikke er signifikant på 5% niveauet, da et opslag i en tabel over χ^2 -fordelingen viser, at $P = 0,665$. Vi kan derfor sige, at den genotypiske fordeling i det undersøgte *Gpi* locus stemmer overens med den forventede Hardy-Weinberg-fordeling.

1.1.3 Variation i loci med dominante alleler

Bestemmelsen af allelfrekvenser i en population er forholdsvis enkel for codominante alleler, hvor vi for hvert individ kan observere, hvilke alleler det bærer. Situationen bliver lidt mere vanskelig, når der er dominans mellem allelerne. Vi betragter et autosomalt locus med to alleler, D og d , hvor D dominerer over d . Der findes derfor to fænotyper i populationen, nemlig $D-$ og dd , hvor vi kun hos den sidste entydigt er i stand til at bestemme dens genotype. fænotypen $D-$ er enten homozygot DD eller heterozygot Dd . Hvis vi antager, at betingelserne for Hardy-Weinberg-ligevægten er opfyldt i populationen, og at allelerne D og d optræder med hyppighederne p_D og p_d , kan populationens sammensætning beskrives på følgende måde

Genotype	Fænotype	Frekvens
DD	$D-$	p_D^2
Dd	$D-$	$2p_D p_d$
dd	dd	p_d^2
		$\left. \begin{array}{l} p_D^2 \\ 2p_D p_d \\ p_d^2 \end{array} \right\} = P_{D-}$
		$\quad \quad \quad = P_d$

Allelfrekvenserne beregnes ud fra relationen mellem frekvensen af den recessive allel og frekvensen af den recessive homozygot. Da

$$p_d^2 = P_{dd}$$

får vi, at

$$p_d = \sqrt{P_{dd}}$$

og dermed

$$p_D = 1 - p_d.$$

forudsætningen herfor er som sagt, at der er Hardy-Weinberg-proportioner for dette locus i populationen. Såfremt dette ikke er opfyldt, vil denne fremgangsmåde ikke resultere i et meningsfyldt resultat. Vi har ingen mulighed for at teste for Hardy-Weinberg-proportioner, da fænotypefordelingen beskrives med én enkelt frihedsgørad. De beregnede allelfrekvenser beskrives ligeledes med én enkelt frihedsgørad, så der er ingen frihedsgørader til en test.

1.1.4 Variation i kønsbundne loci

Hardy-Weinberg loven blev udledt for variation i et autosomt locus. Kønsbundne loci opfører sig lidt anderledes. Vi vil nu se på et simpelt system, hvor kønnet bestemmes af to forskellige kønskromosomer, X og Y , som det blandt andet kendes fra pattedyr og *Drosophila* og hos få særkønnede planter som humle og hamp. Her hører bærere af to X -kromosomer til det homogametiske køn (hunner), medens bærere af et X - og et Y -kromosom tilhører det heterogametiske køn (hanner). Hos fugle og sommerfugle er kønsbestemmelsen omvendt. Hanner er homogametiske, medens hunner er heterogametiske. Man ser ofte, at der i sådanne systemer benyttes betegnelserne Z og W for de kromosomer, der svarer til henholdsvis X - og Y -kromosomer.

Kønskromosomer kan opdeles to regioner: én, hvor der sker en parring af homologe dele af X - og Y -kromosomerne i meiosen, og én, der omfatter de ikke-homologe dele af kønskromosomerne, hvor der ikke sker nogen parring i meiosen. Vi ser på variation, der bestemmes af gener, som sidder i denne region på X -kromosomet. Hunnerne bærer to kopier af sådanne gener, medens hannerne kun har én kopi, de er hemizygote. Betragt en population, hvor der i et locus \mathcal{A} er observeret to alleler, A_1 og A_2 . Hos hunnerne findes derfor tre genotyper, A_1A_1 , A_1A_2 og A_2A_2 , medens der hos hannerne kun forekommer to genotyper, A_1Y og A_2Y . Frekvenserne af de to alleler bestemmes hos hunnerne som i et autosomt locus, se afsnit 1.1.1. Hos hannerne er det lidt nemmere at bestemme frekvenserne af A_1 og A_2 , de er lig med frekvenserne af hanner med genotyperne A_1Y og A_2Y .

Med tilfældig gametsammensmeltning og ens allelfrekvenser hos hanner og hunner forekommer der Hardy-Weinberg-proportioner hos hunnerne, medens hannerne genotypfrekvenser er lig med allelfrekvenserne. Her er genotypefordelingen i de to køn så givet ved

Hunner			Hanner	
A_1A_1	A_1A_2	A_2A_2	A_1Y	A_2Y
\bar{p}_1^2	$2\bar{p}_1\bar{p}_2$	\bar{p}_2^2	\bar{p}_1	\bar{p}_2

Forskellen i genotypefordelingen mellem hanner og hunner bliver specielt tydeligt i tilfælde, hvor den ene allel dominerer over den anden. Her vil den recessive fænotype være langt hyppigere hos hannerne. Hvis vi antager, at A_2 er recessiv over for A_1 , vil den forekomme med en frekvens, der er $1/\bar{p}_2$ større end A_2A_2 homozygoten hos hunnerne. Hos sjældne arvelige sygdomme som hæmofili (blødersygdommen) hos mennesket er det derfor meget sjældent at observere syge kvinder. Det er stort set kun mænd, der lider af denne sygdom. Det samme gælder for rød-grøn farveblindhed.

1.1.5 Variation i to loci

Hardy-Weinberg-loven viste, at en enkel generation med tilfældig parring er tilstrækkelig til at genotyperne i et autosomt locus vil fordeles efter de forventede Hardy-Weinberg-proportioner. Dette gælder for alle loci i genomet, men betragtes variationen i to eller flere

loci samtidigt, fordeles den ikke nødvendigvis uafhængigt af hinanden i de forskellige loci. Vi nøjes her med at se på to loci, men selv for dette simple system bliver dynamikken betydelig mere kompleks end for variation i et enkelt locus, fordi der skal tages hensyn til mulige associationer mellem alleler i de to loci.

Betragt to autosmale loci A og B , der begge er polymorfe med to alleler. Frekvensen af allelerne i populationen er givet ved

$$\begin{aligned} A_1: p_1 \\ A_2: p_2 \\ B_1: q_1 \\ B_2: q_2 \end{aligned}$$

hvor der som sædvanligt gælder, at

$$p_1 + p_2 = q_1 + q_2 = 1.$$

Med to loci, der hver har to alleler, vil der i alt kunne være fire forskellige gameter. Lad f_{11} være frekvensen af gameter, der bærer allel A_1 i locus A og allel B_1 i locus B , og lad f_{12} være frekvensen af gameter, der bærer allel A_1 i locus A og allel B_2 i locus B , og så videre. Der gælder selvfølgelig, at

$$f_{11} + f_{12} + f_{21} + f_{22} = 1.$$

Det er let at beregne allelfrekvenserne ud fra gametfrekvenserne. De findes som

$$p_1 = f_{11} + f_{12} \tag{1.4}$$

$$p_2 = f_{21} + f_{22} \tag{1.5}$$

$$q_1 = f_{11} + f_{21} \tag{1.6}$$

$$q_2 = f_{12} + f_{22}, \tag{1.7}$$

hvor vi adderer frekvenserne af de par af gameter, der hver bærer den allel, hvis frekvens vi ønsker at beregne.

Koblingsuligevægt

Et centralt spørgsmål er nu, om variationen i de to loci *statistisk* set fordeler sig uafhængigt af hinanden. Hvis dette er tilfældet, vil frekvenserne af de fire gameter være lig med produkterne af frekvenserne af de alleler, de bærer. Det vil sige

$$\begin{aligned} f_{11} &= p_1 q_1 \\ f_{12} &= p_1 q_2 \\ f_{21} &= p_2 q_1 \\ f_{22} &= p_2 q_2. \end{aligned}$$

I dette tilfælde taler man om **koblingsligevægt**, i modsat fald betegnes det **koblingsuligevægt**.¹ Man støder ofte på betegnelserne gametfaseligevægt og gametfaseuligevægt

¹på engelsk henholdsvis "linkage equilibrium" og "linkage disequilibrium" (Lewontin & Kojima 1960).

for disse to situationer.² Betegnelsen koblingsuligevægt er måske knap så heldig, da man måske kunne ledes til at tro, at lociene er koblede. Det behøver de ikke være, for at der kan opstå koblingsuligevægt mellem dem. Koblingsligevægt refererer udelukkende til, hvorvidt der er en statistisk uafhængighed mellem variationen i de to loci i gameterne. Er denne uafhængighed ikke til stede, må man inddrage en parameter ud over allelfrekvenserne i de to loci for at beskrive gametfrekvenserne. Dette gøres ved

$$f_{11} = p_1 q_1 + D \quad (1.8)$$

$$f_{12} = p_1 q_2 - D \quad (1.9)$$

$$f_{21} = p_2 q_1 - D \quad (1.10)$$

$$f_{22} = p_2 q_2 + D, \quad (1.11)$$

hvor parameteren D passende betegnes som koblingsuligevægten. Den beskriver afvigelsen fra koblingsligevægten

$$D = f_{11} - p_1 q_1. \quad (1.12)$$

Evolutionær ændring af koblingsuligevægten

Såfremt forudsætningerne for Hardy-Weinberg-ligevægten er opfyldte, vil der efter en tid indstille sig koblingsligevægt i populationen. Raten er afhængig af, hvor tæt de to loci er koblede til hinanden. Hvis vi måler afstanden mellem dem med **rekombinationsraten** r , kan det vises, at gametfrekvenserne ændrer sig fra generation til generation med

$$f'_{11} = f_{11} - rD \quad (1.13)$$

$$f'_{12} = f_{12} + rD \quad (1.14)$$

$$f'_{21} = f_{21} + rD \quad (1.15)$$

$$f'_{22} = f_{22} - rD. \quad (1.16)$$

Bemærk, at det kun er gametfrekvenserne, der ændrer sig. Allelfrekvenserne er konstante i de to loci, da forudsætningerne for Hardy-Weinberg-ligevægten jo er opfyldte.

Man kan vise, at værdien af D_t efter t generationer er givet ved

$$D_t = (1 - r)^t D_0,$$

hvor D_0 er koblingsuligevægten i generation 0. Vi ser derfor, at jo tættere koblede de to loci er, desto længere tid tager det at komme i koblingsligevægt. Selv for ukoblede loci med $r = 1/2$ ses, at det tager et vist antal generationer for at komme i denne ligevægt. Afstanden til den vil for ukoblede loci halveres i hver generation.

²på engelsk henholdsvis “genetic phase equilibrium” og “genetic phase disequilibrium” (Crow & Kimura 1970).

1.1.6 Genetisk diversitet

Polymorfi

I beskrivelsen af genetisk variation har vi benyttet eksempler med mere end én allel, det vil sige, at de benyttede loci var polymorfe. Spørgsmålet er nu, hvornår man betragter et locus som polymorft? Den mest simple definition er, at hvis bare der findes mere end én allel i populationen, kaldes den for polymorf. Nu stammer vores kendskab til populationen udelukkende fra stikprøven, der jo har en begrænset størrelse. Derfor kan vi ikke bruge en definition på polymorfi, at der skal optræde mindst to alleler i en stikprøve. Jo mindre stikprøven er, des større er sandsynligheden for at sjældne alleler ikke kommer med i stikprøven. Man benytter derfor en pragmatisk **definition på polymorfi**. Et locus kaldes polymorft efter 0,95-kriteriet, hvis den hyppigste allel har en frekvens, der er mindre end 0,95. Tilsvarende har man også et 0,99-kriterium (Nei 1987).

Heterozygoti og gendiversitet

Heterozygoti er bedre egnet til at beskrive genetisk variation end graden af polymorfi er. Her bestemmes brøkdelen af alle observerede genotyper, der er heterozygote. Det er et godt mål for genetisk diversitet i et locus, såfremt genotypfordelingen stemmer nogenlunde overens med Hardy-Weinberg-proportioner. Er dette ikke tilfældet, er det et misvisende mål. Dette gælder specielt for planter, som er selvbestøvende. Jo højere selvbestøvningsraten er, desto større er hyppigheden af homozygoter i populationen. Dette problem løses ved at beregne den **forventede heterozygoti**, H_e , efter Hardy-Weinberg-loven. Denne beregnes lettest ved at indse, at hyppigheden af de forventede heterozygoter er lig med én minus hyppigheden af alle homozygoter

$$H_e = 1 - \sum_{i=1}^a p_i^2. \quad (1.17)$$

Denne størrelse betegnes som **gendiversitet** (Nei 1987). Man kan betragte den som sandsynligheden for, at to tilfældigt udtrukne gener i et locus er forskellige alleler. H_e har derfor også mening i haploide organismer eller for organeller som mitokondrier eller kloroplaster, der jo også er haploide. Her er H_e et mål for, hvor sandsynligt det er, at to tilfældigt udtrukne individer (eller organeller) bærer forskellige allele former af et gen.

1.2 Variation på DNA-niveau

Den størst mulige information om genetisk variation fås selvfølgelig ved kendskabet til polymorfier i selve DNA'et. Det benyttes i stigende omfang til karakterisering af genetisk variation i naturlige populationer og til undersøgelse af deres struktur.

1.2.1 Nukleotiddiversitet

Variationen på selve nukleotid-niveauet kan beskrives ved at bestemme antallet af **segregerende sites**, hvor der observeres mere end én base. I Tabel 1.4 er vist et hypotetisk eksempel med seks DNA-sekvenser. De er alle forskellige og en optælling viser, at de segregerer i syv ud af ti sites. Denne metode kan anvendes, selv om alle sekvenserne er forskellige. Metoden er imidlertid afhængig af stikprøvens størrelse, hvilket kan gøre det vanskeligt at sammenligne populationer, hvor der er taget stikprøver af forskellig størrelse.

Man kan beskrive variation i DNA-sekvenser med en metode, som er analog til gendiversitet i et enkelt locus. Metoden er lige som for variation i ét locus forholdsvis uafhængig af stikprøvestørrelsen. Vi definerer **nukleotiddiversitet** som sandsynligheden for, at to tilfældigt udtrukne sekvenser fra populationen i et site bærer to forskellige baser. Hvis vi med π_{ij} betegner brøkdelen af sites i hvilken sekvenserne i og j bærer forskellige baser, så kan nukleotiddiversiteten beregnes som

$$\bar{\pi} = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^N \pi_{ij}}{N(N-1)/2}, \quad (1.18)$$

hvor N er antallet af sekvenser i stikprøven, og hvor summationen går over alle mulige par af forskellige sekvenser i stikprøven. Antallet af sammenligninger er givet ved $N(N-1)/2$. Fremgangsmåden kan illustreres med eksemplet i Tabel 1.4. Højre side af tabellen viser antallet af nukleotidforskelle mellem de seks sekvenser. Deres sum er 45, hvilket skal deles med $6(6-1)/2 = 15$ for at få den gennemsnitlige nukleotiddiversitet, som i dette tilfælde er 3. Sidstnævnte divideres med sekvensens længde (10) for at finde nukleotiddiversiteten per site på 0,3.

Tabel 1.4: Nukleotidvariation og antal nukleotidforskelle mellem seks DNA-sekvenser, m og s angiver sites, der er henholdsvis monomorfe og segregerende.

		Nukleotidforskelle					
Sekvens		1	2	3	4	5	6
aatgccgatt	1		2	2	2	2	3
aaAgGcgatt	2			2	4	4	1
aatgGcCatt	3				2	4	3
aatCccCatt	4					4	5
aatgccgaAA	5						5
aaAgGcgTtt	6						
mmsssmssss							

1.3 Fænotypisk variation

De metoder, der er blevet præsenteret i de forrige afsnit, benyttes hovedsageligt til at beskrive molekylær variation i naturlige populationer. De kan kun sjældent benyttes til at analysere variation i karakterer som størrelse, farve og bladareal. Disse karakterer udviser fænotypisk variation, hvor forskellene mellem individer ofte er så små, at de kan være svære at måle, og hvor det er vanskeligt at beskrive den genetiske baggrund, der bestemmer dem.

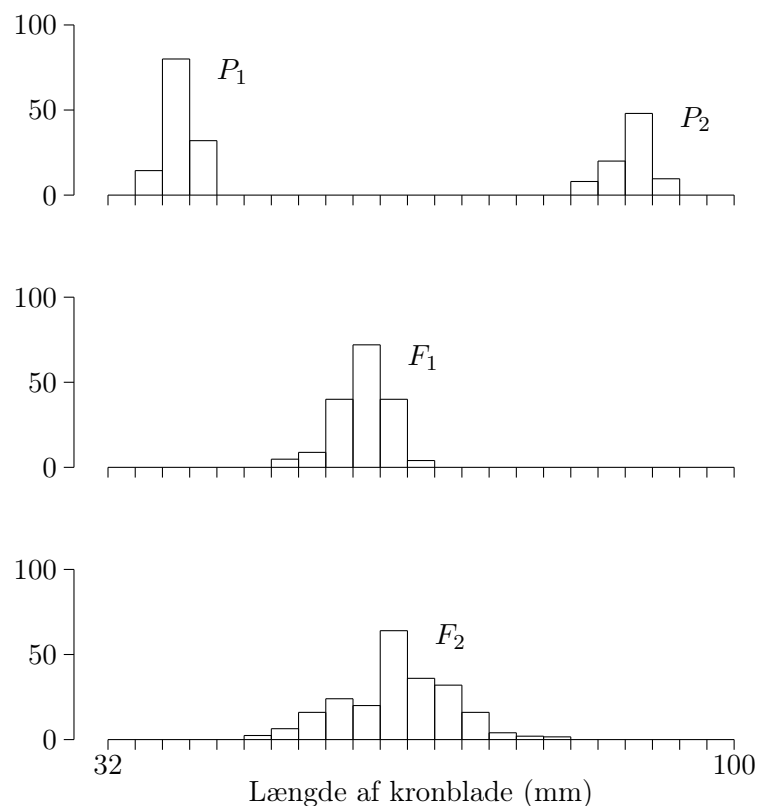
Til denne kategori hører også antal afkom, overlevelse og andre egenskaber, der er knyttet til et dyrs eller en plantes evne til at producere afkom til næste generation. Disse sidstnævnte karakterer er alle komponenter af fitness, der i sig selv kan betragtes som en kvantitativ karakter. Et kendskab til den genetiske baggrund af disse karakterer og til deres genetiske diversitet i naturlige populationer er derfor nødvendig for at forstå arters muligheder for at tilpasse sig til ændringer i det fysiske eller biologiske miljø.

For at beskrive en population må vi ty til statistiske metoder. Det mest simple består i at afbilde fordelingen på en passende skala. Ud fra denne beregnes så størrelser som middelværdi og varians. Kvantitative karakterer deles ofte op i tre grupper: kontinuerte, meristiske og tærskel-karakterer. Som navnet antyder, så vil fordelingen af **kontinuerte karakterer** kunne beskrives langs en akse, der er kontinuert. Hertil hører for eksempel vægt, længde eller diameter af stammen på træer. Disse karakterer kan måles med forskellige fysiske eller kemiske metoder, og deres størrelse vil være en funktion af den anvendte skala og af "måleinstrumentets" nøjagtighed. I princippet vil disse karakterer være kontinuert fordelt, men for det meste grupperes de i klasser. Fordelingen vil ofte – såfremt man har valgt en passende skala – være normalfordelt.

Tællelige karakterer optræder som diskrete enheder, det vil sige, at de kan tælles med hele tal. Hertil hører eksempler som antallet af hvirvler, afkom, børster og hår. Skalaen er givet på forhånd. Når antallet af observerede kasser er forholdsvis stort, vil deres fordeling også kunne tilnærmes med en normalfordeling.

Den sidste gruppe er **tærskel-karakterer**. Disse optræder oftest kun i to former, hvor tilstanden kan beskrives som tilstedeværelse eller mangel på en egenskab. Som eksempel kan nævnes, hvorvidt en enkelt fugl er trækfugl eller standfugl. At benytte betegnelsen kontinuert karakter for en sådan egenskab synes måske lidet intuitivt, men årsagen er, at dens genetik kan beskrives på samme måde som for de andre to grupper.

Fælles for de tre kategorier er, at den observerede fænotypiske fordeling er betinget af såvel en genetisk som en miljø-påvirkning. En given karakter vil ofte være under indflydelse af en lang række loci, hvis enkelte bidrag det oftest er umuligt at måle. De betegnes derfor ofte som **polygene** eller **kvantitative karakterer**. Den engelske betegnelse for gener, der bidrager til karakteren, er "minor genes" eller "polygenes" i modsætning til "major genes", hvis effekt med en passende metode direkte kan måles. Rent genetisk er der ingen forskel på disse to klasser af loci. For de fleste karakterer er det ukendt, hvorvidt det drejer sig om få loci (5–20) eller mange (100 eller flere) (Barton & Turelli 1989). Ved krydsninger mellem individer fra forskellige populationer kan man dog med passende statistiske redskaber få et skøn over antallet. Med molekylære teknikker er det i dag muligt at kortlægge disse såkaldte "**quantitative trait loci**", **QTL**.



Figur 1.2: Fordeling af antal planter med forskellig længde af kronblade i to rene linier P_1 og P_2 af tobaksplanten (*Nicotiana longiflora*), deres F_1 og F_2 . Data af East (1916), fra Grant (1975).

1.3.1 Komponenter af kvantitative karakterer

For et givet *individ* kan en karakter beskrives ved dens **fænotype** P . Denne kan betragtes som summen af bidrag fra **genotype**, G , og **miljø**, E ,³

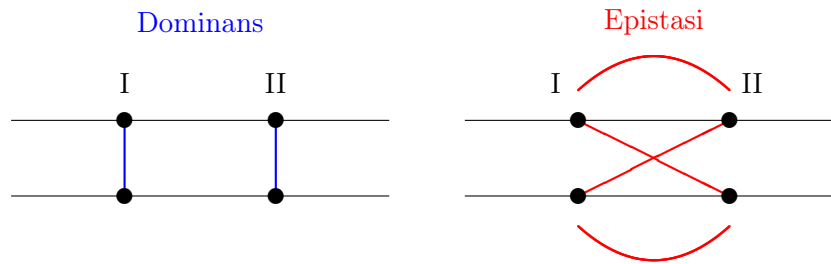
$$P = G + E.$$

Skal en karakter beskrives for en given *population*, afbildes fordelingen af fænotypen, og dens middelværdi og varians beregnes. Spørgsmålet er nu, hvor stor en del af den fænotypiske varians er genetisk betinget, og hvor stor en del skyldes miljøet? Konkret kan det formuleres som

$$V_P = V_G + V_E,$$

hvor V_P er den **fænotypiske varians**, V_G er den **genetiske varians** og V_E er **miljøvariansen**. (Vi ser bort fra eventuel interaktion mellem arv og miljø.) Et simpelt eksempel kan illustrere dette. I Figur 1.2 er længden af kronbladene afbildet for to indavlede sorter af tobaksplanter. Den observerede varians må derfor udelukkende bestå af miljøvariens,

³Forkortelserne er engelske, de står for henholdsvis “phenotype”, “genotype” og “environment”.



Figur 1.3: De to horizontale linier symboliserer to stykker homologe kromosomer hos et individ. På kromosomerne er to loci, I og II, angivet. Der er mulighed for interaktion mellem gener inden for samme locus, dominans, og mellem loci, epistasi.

det vil sige $V_P = V_E$. Midterst i billedet ses deres F_1 -afkom. Disses varians består igen udelukkende af miljøvariens, da de er genetisk identiske. Sammenlignes forældrenes fordeling med F_1 's fordeling ses, at deres varianser er af samme størrelse. Krydses F_1 -individer indbyrdes fås en F_2 -generation, der har en betydelig højere varians end nogen af de øvrige fordelinger. Denne øgning skyldes udspaltningen af alleler i de loci, der bidrager til forskellen mellem P_1 og P_2 . I dette tilfælde kunne man groft bestemme V_G i F_2 som differencen mellem dens V_P og gennemsnittet af variansen i de to første generationer, der giver et skøn over V_E .

Vi vil nu opsplitte den genetiske varians i dens komponenter og se på, hvorledes disse kan estimeres. Den genetiske varians består af en **additiv varians** V_A , en **dominans-variens** V_D og en **interaktionsvariens** V_I ,

$$V_G = V_A + V_D + V_I. \quad (1.19)$$

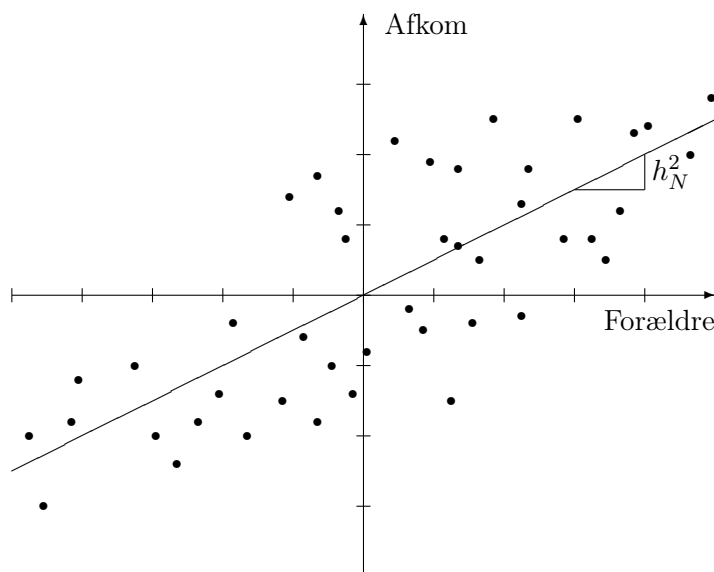
Dominans- og interaktionsdelen skyldes interaktion mellem gener inden for de enkelte loci, medens interaktionsvariensen er forårsaget af epistatisk interaktion mellem gener i de forskellige loci, se Fig. 1.3. For V_D og V_I gælder at de er positive, når der henholdsvis er afvigelse fra additive effekter inden for de enkelte loci eller mellem de forskellige loci. I Fig. 1.3 er locus I og II illustreret som siddende på samme kromosom. Dette er dog ingen forudsætning for epistatisk interaktion. Den forekommer ligesåvel blandt ikke-koblede loci.

1.3.2 Heritabilitet

Brøkdelen af den fænotypiske varians, der skyldes genetiske bidrag betegnes “**broad sense**” **heritabilitet** og beregnes som

$$h_B^2 = \frac{V_G}{V_P}.$$

Det er imidlertid en anden komponent, der er mere interessant. Den additive genetiske varians V_A er den del af den genetiske varians, V_G , der skyldes korrelationen mellem beslægtede individer. V_A er i princippet den variation, som kunstig eller naturlig selektion



Figur 1.4: Regression af afkommets værdi for en morfologisk karakter mod middelværdien af forældrenes middelværdi. Koordinatsystemet er blevet lagt så nulpunkterne for x og y -akse befinder sig i forældrenes og afkommets middelværdi.

har at arbejde med. Brøkdelen af den totale fænotypiske varians, som skyldes den additive genetiske varians, betegnes som **“narrow sense” heritabilitet**

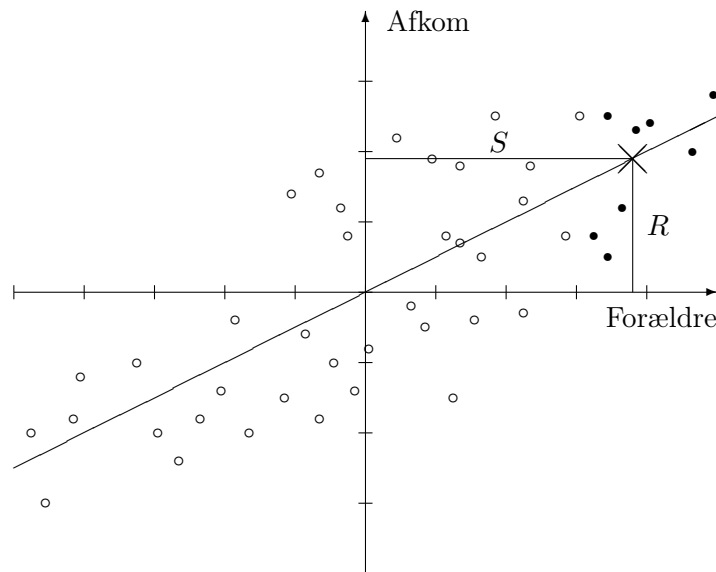
$$h_N^2 = \frac{V_A}{V_P}.$$

En given karakters heritabilitet kan bestemmes ud fra regressionen mellem beslægtede individer. Afbildes et enkelt afkoms værdi som funktion af forældrenes middelværdi for karakteren, kan man vise at heritabiliteten kan bestemmes som regressionskoefficienten, se Figur 1.4. Afbildes afkommets værdi som funktion af en enkelt forælders værdi, er heritabiliteten lig med to gange regressionskoefficienten. Der findes herudover andre metoder til bestemmelse af komponenter af den genetiske varians, se Falconer (1989).

Har man bestemt en karakters heritabilitet i en population, kan man benytte den til at forudsige virkningen af selektion på karakteren. Antag, at vi selekterer de forældre, som har de højeste værdier for karakteren i populationen. Middelværdien af deres afvigelser fra hele populationens gennemsnit betegner vi som **selektionsdifferentialet** S . Ser vi på deres afkom, vil de have en gennemsnitlig værdi, som ligger over populationens middelværdi, se Figur 1.5. Forskellen betegnes som **selektionsrespons**. Sammenhængen mellem heritabilitet, selektionsdifferentiale og respons er givet ved

$$R = h_N^2 S. \quad (1.20)$$

Vi ser, at jo højere den additive varians og dermed h_N^2 for en karakter er, desto kraftigere vil en population reagere på selektion, det være sig kunstig eller naturlig selektion.



Figur 1.5: Selektionsforsøg, hvor middelværdien af forældrenes afvigelser fra hele populationens middelværdi er angivet med S (selektionsdifferentialet). Afkommets gennemsnitlige afvigelse fra populationens middelværdi er vist med R (responsen).

Heritabiliteten for en karakter i en population kan beregnes ud fra korrelationen mellem beslægtede individer. Dens værdi gælder for populationen i den generation og i det miljø, den er blevet målt i. Selektion ændrer den genetiske sammensætning og dermed også heritabiliteten. I praksis har det dog vist sig, at et skøn over heritabiliteten også vil kunne bruges til at forudsige responsen gennem flere generationer (op til 10 eller flere), såfremt selektionen ikke er for hård (Falconer 1989).

Hvis kunstig eller naturlig selektion virker på en karakter igennem længere tid, forventer vi, at selektionen efterhånden “opbruger” den additive genetiske varians. Responsen aftager og standser til sidst. Raten er proportional med den additive varians. Fisher (1930) har formuleret det som et fundamentalt teorem for naturlig selektion: “*The rate of increase in fitness of any organism at any time is equal to its genetic variance⁴ in fitness at that time.*” Hvis vi ville være i stand til at observere selve fitness som karakter, ville dens additive varians være nul for populationer, som er kommet i ligevægt i forhold til naturlig selektion. Nu er det vanskeligt at måle fitness som sådan, men vi kan undersøge karakterer, der er relateret til fitness. Hertil hører de såkaldte “life-history”-karakterer. Det er blandt andet antal afkom, vækstrate eller overlevelse. Disse karakterer forventes at have lave heritabiliteter sammenlignet med mindre “livsnødvendige” karakterer. Dette er også fundet hos en lang række husdyr (Falconer 1989). Som et enkelt eksempel kan nævnes at heritabiliteten af svinets spæklag er 0,70, medens heritabiliteten af kuldstørrelsen er tæt på 0.

⁴Fisher’s “genetic variance” svarer til additiv varians i ligning 1.19.

Indeks

- Additiv varians, 15
- Allelfrekvens, 2
 - beregning, 2
 - dominans, 7
- Autosomal, 2
- Blødersygdom, 8
- Codominans, 2
- Dominans, 7
- Dominans-variens, 15
- Fænotype, 14
- Fænotypisk varians, 14
- Fænotypisk variation, 13–17
- Farveblindhed, 8
- Fisher's fundamentale teorem for naturlig selektion, 17
- Fugle
 - Kønsbestemmelse, 8
- Gametfaselige vægt, *Se* Koblinglige vægt
- Gametfrekvens, 9
- Gendiversitet, 11
 - definition, 11
- Genetisk varians, 14
- Genotypefrekvens, 2
 - beregning, 2
- Hamp, 8
- Hardy-Weinberg
 - dominans, 7
 - egenskaber, 5
 - lige vægt, 4
 - lov, 3–7
 - proportioner, 4
 - test, 5–7
 - frihedsgørader, 6
- Hemizygoti, 8
- Heritabilitet, 15–17
 - “broad sense”, 15
 - “narrow sense”, 16
 - beregning, 16
- Heterogametisk, 8
- Heterozygoti, 11
 - forventet, 11
- Homogametisk, 8
- Humle, 8
- Hæmofili, 8
- Interaktionsvariens, 15
- Jaera ischiosetosa*
 - allelfrekvenser, 2
 - Hardy-Weinberg proportioner, 6
- Koblingsulige vægt, 9–10
 - evolution, 10
- Kvantitativ karakter, 13, 14
 - kontinuert, 13
 - tællelig, 13
 - tærskel, 13
- Kønsbunden variation, 8
 - allelfrekvens, 8
- Kønskromosomer, 8
- Major genes, 13
- Miljø-variens, 14
- Minor genes, 13
- Nicotiana longiflora*, *Se* Tobaksplante
- Nukleotiddiversitet, 12
 - definition, 12
- Pattedyr
 - Kønsbestemmelse, 8

Polygen karakter, *Se* Kvantitativ karakter

Polygenes, 13

Polymorfi, 11

QTL, quantitative trait loci, 13

Rekombinationsrate, 10

Segregerende sites, 12

Selektionsdifferentiale, 16

Selektionsrespons, 16

Sommerfugle

Kønsbestemmelse, 8

Tilfældig parring, 3

Tobaksplante

variation i kvantitativ karakter, 14

W-kromosom, 8

X-kromosom, 8

Y-kromosom, 8

Z-kromosom, 8