

Department of Biology

Examination paper for BI3010 – Population genetics

Academic contact during examination:

Jarle Mork (phone 90973351)

Henrik Jensen (phone 91897064)

Examination date: December 18th 2014

Examination time (from-to): 09:00 – 13:00

Permitted examination support material: Calculators Casio fx-82ES PLUS, Citizen SR-270X, Citizen SR-270X College, Hewlett Packard HP30S.

Other information:

Language: English + Bokmål + Nynorsk

Number of pages (front page included): 16

Number of pages enclosed: Appendix: 1

Checked by:

Date

Signature

ENGLISH

Questions are differently weighted (1-3)

Question 1 (weight 3)

- a) What is the effect of each of the "four evolutionary forces" on genetic variability within and between natural populations?
- b) Explain the Hardy-Weinberg law (HW) with words, formulae and a numerical example.
- c) Perform a chi-squared Goodness-of-fit test of conformance to HW of the following genotypic distribution for a 2-allele (A, B) locus polymorphism in a diploid organism: **AA:30, AB:40, BB:30**. Use Appendix Table 1 to decide whether the deviation is statistically significant at the 0.05 level or not.
- d) Explain "the Wahlund effect" and its relation to Sewall Wright's F_{ST} .
- e) Calculate the F_{ST} for the two populations in *Table A* (below).
- f) Perform an RxC tests of the homogeneity of genotypic and allelic composition in the two populations (use Appendix Table 1 to find the statistical significance of any differences).

Table A. Genotypic composition in two populations

	AA	AB	BB
Population 1	25	50	25
Population 2	81	18	1

Question 2 (weight 2)

- a) Assume a single-locus, two-allelic blue/white skin-color polymorphism in a snail species, where the allele (B) for blue color is dominant over the allele (W) for white color. In a sample of 100 specimens we observe 91 blue and 9 white snails. Calculate the allele frequencies of B and W in the sample.
- b) What would be the expected proportions of color phenotypes among 100 offspring from a single cross between a BB male and a BW female?
- c) Table B (below) contains real data from Trondheimsfjord cod. 50 individuals were genotyped for two autosomal loci, and their combined genotypes were crosstabulated. Use an appropriate test to explore whether the combined genotype groups at the two loci occur in the expected proportions in Table B, or not. Use Appendix Table 1 to find and report the significance level (P-value) resulting from the test, and discuss briefly what circumstances

can cause deviations from expected combined genotype proportions in such a sample from a natural population.

Table B. Two-locus combined genotypes in a sample of cod.

Locus <i>Hbl</i>	Locus <i>LDH3</i>			Σ
	<i>CC</i>	<i>CD</i>	<i>DD</i>	
AA	4	13	7	
AB	5	14	5	
BB	1	1	0	
Σ				50

Question 3 (weight 2)

- Define the concept "effective population size" (N_e).
- A natural population of size $N_e=100$ individuals had a heterozygosity of $H=0.5$ in generation 1. In the five subsequent generations the N_e varied in the following way: 70-10-50-30-40. Calculate the overall N_e value for the entire period.
- Small populations lose genetic variation each generation by random genetic drift according to a well-known formula. Calculate the expected heterozygosity at the start of generation 7 in the population described in b) above.

Question 4 (weight 2)

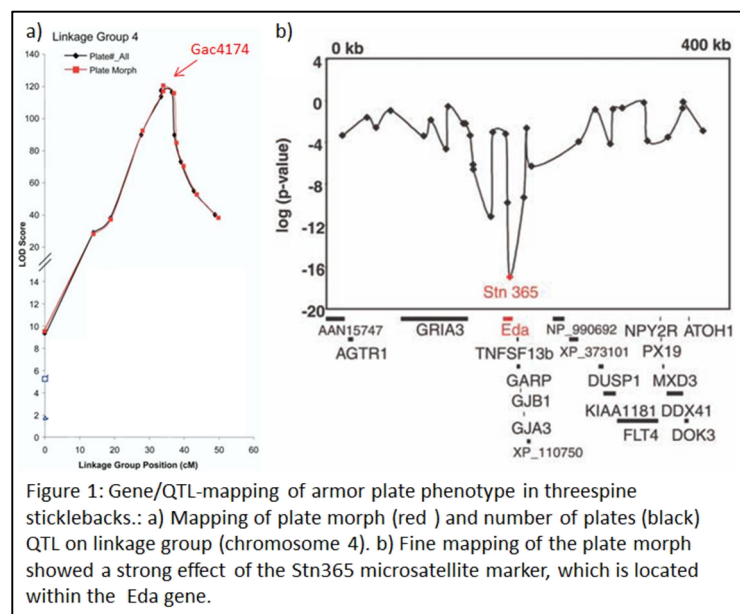
- Explain the following concepts from breeding genetics: parental population, offspring population (F1), phenotypic mean value, selection differential (S), heritability (H^2 and h^2), selection response (R).
- A captive population of milking cows produced an average of 3000 kg milk per year. Among them, a subgroup consisting of the best-yielding part of the population, with an average of 4000 kg per year, was selected to parent the next generation. Assuming that the narrow sense heritability for milk yield is 0.36 – what was the expected average milk yield per year of the offspring generation (F1)?
- Why is a strict control with and reduction of environmental variation so important in experiments designed to estimate the heritability of quantitative traits?

Question 5 (weight 3):

Threespine sticklebacks (*Gasterosteus aculeatus*) offer a powerful system to study the genetic basis of adaptive evolution. After the last ice age, marine sticklebacks colonized and adapted to many recently formed freshwater habitats. Repeated evolution of similar phenotypes occurred, such as for example substantial loss of bony armor plates on the lateral side of the body (low-plated form). Distinctive marine and freshwater forms can still hybridize, and this makes it possible to map genes underlying e.g. armor plate evolution. In 2004 Colosimo et al. (PLOS Biology) collected individual sticklebacks from a marine population in Japan (all individuals were completely armor plated) and a freshwater population in western Canada (all individuals were low-plated) and brought these fish into the lab. At that time the best genetic resource available for threespine sticklebacks was a panel of 160 microsatellite markers distributed across the stickleback genome.

- What are the two main methods used for gene/QTL-mapping? Give a brief explanation of the kind of data you need in each of these approaches.
- Explain what linkage disequilibrium (LD) is, and why LD is important in most gene/QTL-mapping approaches. (Note that LD is defined as a special case of “gametic phase disequilibrium” in the book.)
- Explain how you would have carried out an experiment with the aim to map the gene(s) underlying armor plates in threespine sticklebacks if you had the same resources as Colosimo et al. (2004).

The results of Colosimo et al. (2004) are shown in the figure to the right. As you see from the figure they managed to map a gene/QTL underlying the plate morph and the number of armor plates to a relatively narrow region on linkage group (chromosome) number four. This region included a microsatellite marker called Gac4174. Fine-mapping based on a new set of 32 microsatellite markers within a 400kb



region close to Gac4174 was carried out by Colosimo et al. in a new study in 2005 (Science). The distribution of alleles at each of the 32 new microsatellite markers was then examined in a sample of

46 completely plated and 45 low-plated fish from a single interbreeding wild population in California, USA. The result from this fine mapping is presented in the figure below. As you can see from the figure this study identified the gene ectodysplasin (Eda) as the major locus controlling armor plate patterning in threespine sticklebacks.

Colosimo et al. (2005) then constructed a phylogenetic tree based on DNA sequences of the Eda gene in 15 low plated freshwater populations and 10 completely plated marine populations across the northern hemisphere.

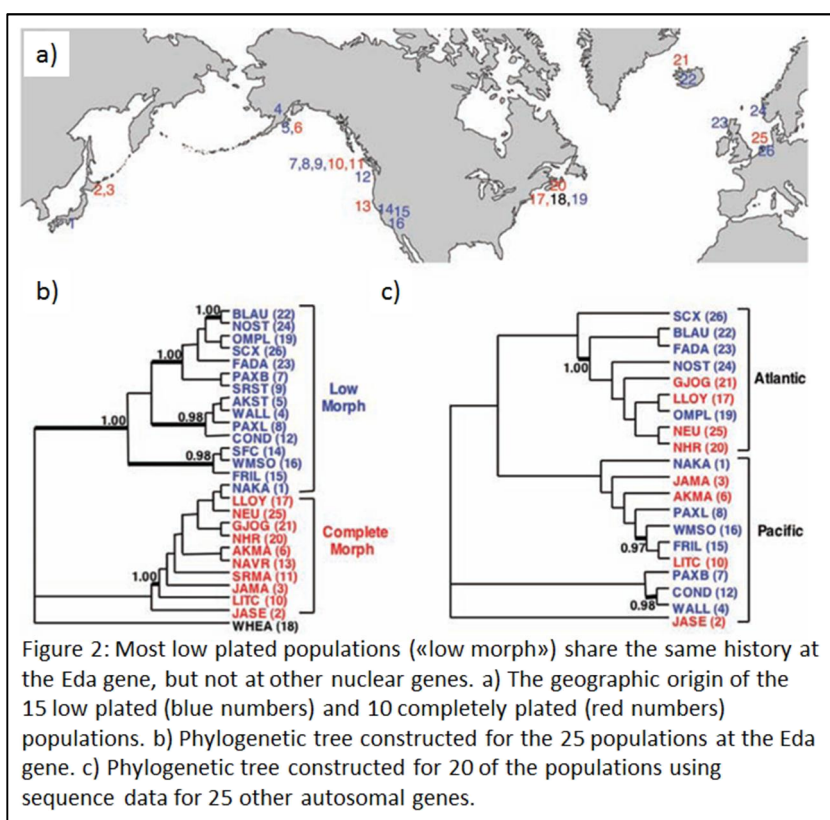
Individuals from a subset of these populations (12 low plated and 8 completely plated) were also sequenced at 25 randomly selected autosomal genes, and a phylogenetic tree was constructed based on 193 SNPs within these genes (note that Jones et al. (2012, Nature) estimated a phylogenetic tree with the same topography for these populations when whole genome sequence data was

used). The locations of the sampled populations and the two phylogenetic trees are shown in the figure above.

Finally, Colosimo et al. (2005) compared marine (completely plated) and freshwater (non-plated) individuals and estimated some molecular genetic parameters using sequence data for exons in the Eda gene. They found $d_N=0.0055$, $d_S=0.016$, and hence $d_N/d_S=0.344$ (note that d_N is sometimes called K_N or K_A , and that d_S is sometimes called K_S). One simple way to examine whether a particular gene shows signs of selection is to calculate the d_N/d_S -ratio for this gene when the gene's DNA sequence is compared in two species, populations or groups of individuals that differ for the phenotype encoded by the gene.

d) What is d_N and d_S ?

e) Explain what different values of the d_N/d_S -ratio indicate (i.e. $d_N/d_S < 1$, $d_N/d_S = 1$ and $d_N/d_S > 1$).



Incomplete lineage sorting is a concept which is sometimes relevant when constructing phylogenetic species trees. This phenomenon may also occur at the within-species level when populations are compared.

- f) Explain what incomplete lineage sorting is, and use this concept to give a likely explanation for the lack of correspondence between phylogenetic trees for Eda (Figure 2b) and other autosomal genes (Figure 2c).

BOKMÅL

Oppgavene er forskjellig vektet (1-3)

Oppgave 1 (vekt 3)

- Hva er effekten av hver av de "fire evolusjonære kreftene" på genetisk variabilitet innen og mellom naturlige populasjoner?
- Forklar Hardy-Weinbergs lov (HW) med ord, formler og numerisk eksempel.
- Utfør en chi-kvadrat Goodness-of-fit test av tilpasning til HW for følgende fordeling av av genotyper i en 2-allel (A, B) locus polymorfisme i en diploid organisme: **AA:30, AB:40, BB:30**. Benytt Appendix Table 1 for å avgjøre om avvik fra HW er statistisk signifikant på det vanlige 0,05 nivå.
- Forklar begrepet "Wahlund effekt" og dets sammenheng med Sewall Wrights F_{ST} .
- Beregn F_{ST} for populasjonene i Tabell A (nedenfor).
- Utfør en RxC test for homogenitet i genotyper og allel for de to populasjonene i Tabell A. Bruk Appendix Table 1 for å avgjøre om et eventuelt avvik er signifikant.

Tabell A: Genotypesammensetning i to populasjoner

	AA	AB	BB
Population 1	25	50	25
Population 2	81	18	1

Oppgave 2 (vekt 2)

- Anta en enkelt-locus, to-allel blå/hvit hudfarge polymorfisme hos en snegleart, der allel (B) for blå farge er dominant over allel (W) for hvit farge. I en stikkprøve på 100 individ observeres 91 blå individ og 9 hvite. Beregn allelfrekvensene for allel A og B i stikkprøven.
- Hva ville være de forventede proporsjonene av farge-fenotyper blant 100 avkom fra en krysning mellom en BB hann og en BW hunn?
- Tabell A (nedenfor) inneholder data fra torsk fra Trondheimsfjorden. 50 individ ble genotypet for to autosomale loci, og deres kombinerte genotype-grupper ble tallet og krysstabulert (Tabell A). Benytt en passende test for å undersøke om gruppene av kombinerte genotyper for de to loci opptrer i de forventede proporsjoner i Tabell B, eller ikke. Benytt Appendix Table 1 for å finne og rapportere signifikansnivået (P-verdien) fra testen, og drøft hvilke

omstendigheter som kan avstedkomme avvik fra forventede kombinerte genotype proporsjoner en slik stikkprøve fra en naturlig populasjon .

Tabell B. 2-locus kombinerte genotyper i en stikkprøve av torsk.

Locus <i>Hbl</i>	Locus <i>LDH3</i>			Σ
	<i>CC</i>	<i>CD</i>	<i>DD</i>	
AA	4	13	7	
AB	5	14	5	
BB	1	1	0	
Σ				50

Oppgave 3 (vekt 2)

- Definer begrepet "effektiv populasjonestørrelse" (N_e).
- En naturlig populasjon av størrelse $N_e=100$ individ hadde en heterozygositet $H=0.5$ i generasjon 1. I de fem påfølgende generasjoner varierte N_e med følgende størrelser: 70-10-50-30-40. Beregn den overordnede N_e value for hele perioden (generasjon 1- 6).
- Små populasjoner taper genetisk variabilitet hver generasjon pga tilfeldig genetisk drift. Størrelsen på tapet mellom hver generasjon følger en velkjent formel. Beregn den forventede heterozygositeten i starten av generasjon 7 i den populasjonen som er beskrevet i b) ovenfor.

Oppgave 4 (vekt 2)

- Forklar følgende begrep fra avlsgenetikken: parentalgenerasjon, avkom generasjon (F1), fenotypisk middelværdi (P), seleksjonsdifferensial (S), arvbarhet (H^2 and h^2), seleksjonsrespons (R).
- En besetning av melkekyr produserte gjennomsnittlig 3000 kg pr individ pr år. Fra denne besetningen ble en gruppe blant de mest høyt ytende, med et middel på 4000 kg pr individ pr år, valgt ut som stamdyr for neste generasjon. Anta at arvbarheten for melkeproduksjon er 0,36 – hva forventes gjennomsnittlig produksjon pr dyr pr år å være i avkomgenerasjonen (F1).
- Hvorfor er en streng kontroll med og reduksjon av miljøvariasjon så viktig i arvbarhetsstudier?

Oppgave 5 (vekt 3)

Trepigget stingsild (*Gasterosteus aculeatus*) utgjør et meget godt system for å studere det genetiske grunnlaget for adaptiv evolusjon. Etter siste istid koloniserte og tilpasset

marine stingsild seg til mange ferskvannshabitater som nylig var dannet. Gjennom evolusjon av lignende fenotyper skjedde, slik som for eksempel betydelig tap av beinaktige «panserplater» på sidene av kroppen (få-plate-type). Atskilte typer i marine miljøer og i ferskvann kan fortsatt hybridisere, og dette gjør det mulig å kartlegge gener som koder for feks «panserplate»-evolusjon. I 2004 samla Colosimo et al. (PLOS Biology) inn individer av stingsild fra en marin populasjon i Japan (alle individer hadde fullstendige «panserplater») og en ferskvannspopulasjon i vest-Canada (alle individer var av få-plate-typen) og tok dem inn i laboratoriet sitt. På den tiden var den beste genetiske ressursen som var tilgjengelig for trepigget stingsild et panel bestående av 160 mikrosatelitt-markører som var spredt utover i stingsild-genomet.

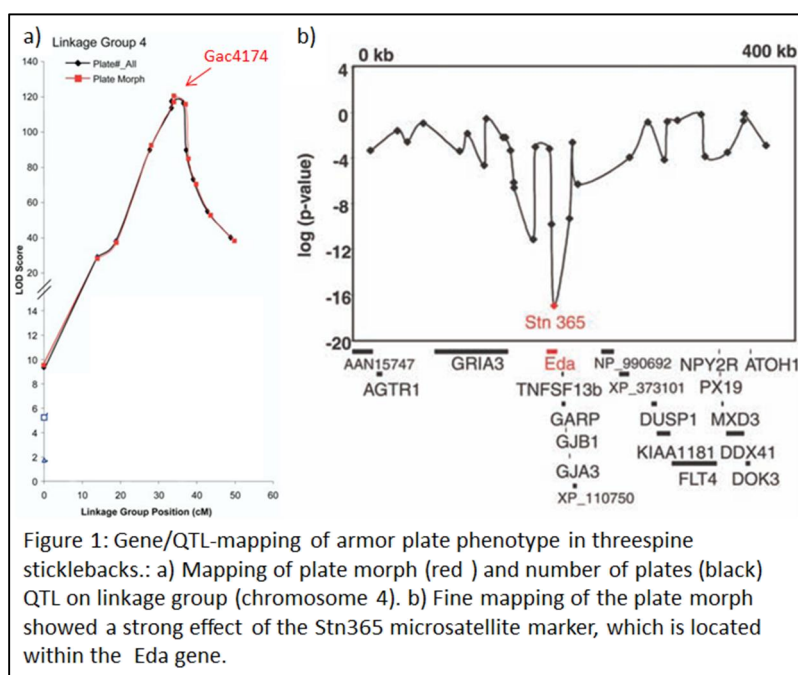
a) Hvilke er de to hovedmetodene som brukes for gen-/QTL-kartlegging? Gi en kort forklaring av hva slags data du trenger i hver av disse fremgangsmåtene.

b) Forklar hva koblings ulikevekt (linkage disequilibrium, LD) er, og hvorfor LD er viktig i de fleste metoder for gen-/QTL-kartlegging. (Merk at LD er definert som et spesielt tilfelle av «gametic phase disequilibrium» i boka.)

c) Forklar hvordan du ville ha utført et eksperiment der målet er å kartlegge genet/genene som koder for «panserplater» i trepigget stingsild hvis du hadde de samme ressursene som Colosimo et al. (2004).

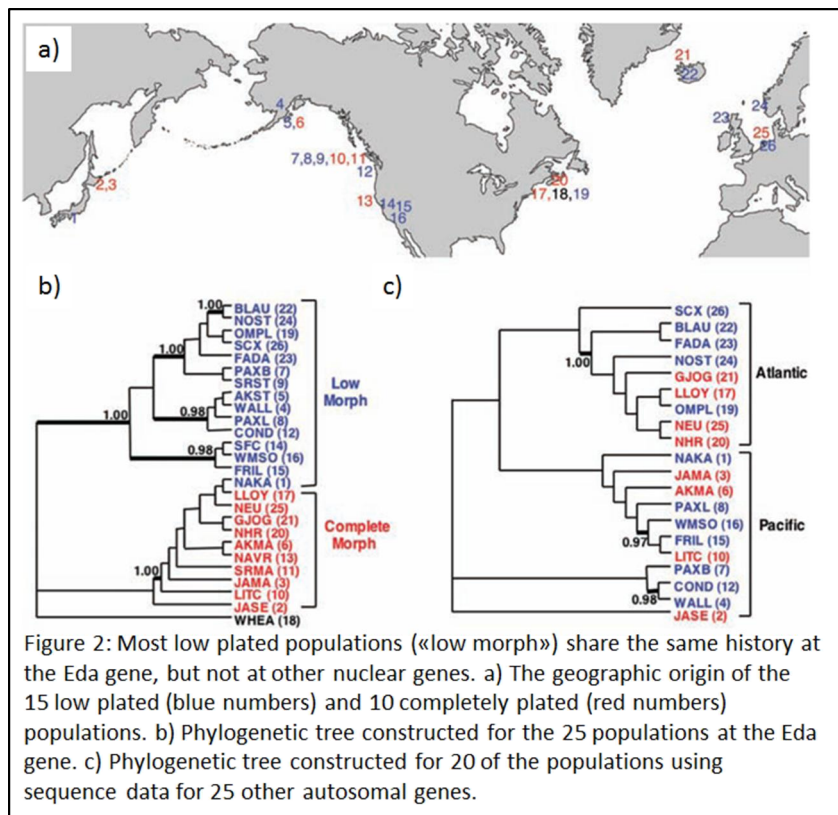
Resultatene til Colosimo et al. (2004) vises i Figur 1a. Som du ser på figuren klarte de å kartlegge

genet/QTLen som koder for «panserplate»-morfen og antall «panserplater» til en relativt begrenset region på koblingsgruppe (kromosom) nummer fire.



Denne regionen inneholdt en mikrosatelitt-markør kalt Gac4174. Finkartlegging basert på et nytt sett med 32 mikrosatelitt-markører i en 400kb region nær Gac4174 ble utført av Colosimo et al. i en ny studie i 2005 (Science). Fordelingen av alleler på hver av de 32 nye mikrosatelitt-markørene ble da undersøkt i et sampel bestående av 46 individer med fullstendige «panserplater» og 45 fisk av «få-plate-typen» fra en populasjon der disse typene krysser seg med hverandre i vill tilstand i California, USA. Resultatene fra denne finkartleggingen er presentert i Figur 1b. Som du kan se fra figuren identifiserte denne studien genet ectodysplasin (Eda) som det viktigste locuset som kontrollerer «panserplate»-utvikling hos trepigget stingsild.

Colosimo et al. (2005) konstruerte så et fylogenetisk tre basert på DNA-sekvensene til Eda genet i 15 ferskvannspopulasjoner med få «panserplater» og 10 marine populasjoner med full «panserplate»-morfologi over



hele den nordlige jord-halvkule. Individer fra et utvalg av disse populasjonene (12 av få-plate-typen og 8 av typen med fullstendige «panserplater») ble også sekvensert på 25 tilfeldig valgte autosomale gener, og et fylogenetisk tre ble konstruert basert på 193 SNPs i disse genene (merk at Jones et al. (2012, Nature) estimerte et fylogenetisk tre med den samme topografien for disse populasjonene ved å bruke helgenom sekvenseringsdata). Plasseringen av de ulike populasjonene og de to fylogenetiske trærne er vist i Figur 2.

Til slutt sammenlignet Colosimo et al. (2005) stingsild av den marine typen (med fullstendige «panserplater») og av ferskvannstypen (uten «panserplater»), og estimerte

forskjellige molekylærgenetiske parametere ved å bruke sekvensdata for eksoner i Eda-genet. De fant at $d_N=0.0055$, $d_S=0.016$, og dermed at $d_N/d_S=0.344$ (merk at d_N enkelte ganger kalles K_N eller K_A , og at d_S enkelte ganger kalles K_S). En enkel måte å undersøke om et bestemt gen viser tegn til seleksjon er å beregne d_N/d_S -ratioen for dette genet når genets DNA-sekvenser sammenlignes i to arter, populasjoner, eller grupper av individer som er forskjellige når det gjelder fenotypen som genet koder for.

d) Hva er d_N og d_S ?

e) Forklar hva ulike verdier for d_N/d_S -ratioen indikerer (dvs. $d_N/d_S < 1$, $d_N/d_S = 1$ og $d_N/d_S > 1$).

Ufullstendig avstammings-sortering (incomplete lineage sorting) er et konsept som enkelte ganger er relevant når en konstruerer fylogenetiske artstrær. Dette fenomenet kan også forekomme på innen-artsnivå når en sammenligner populasjoner.

f) Forklar hva ufullstendig avstammings-sortering (incomplete lineage sorting) er, og bruk dette konseptet til å gi en sannsynlig forklaring for mangelen på samsvar mellom de fylogenetiske trærne for Eda (Figur 2b) og andre autosomale gener (Figur 2c).

NYNORSK

Oppgåvene er ulikt vekta (1-3)

Oppgåve 1 (vekt 3)

- a) Kva er effekten av kvar av dei «fire evolusjonære kreftene» på genetisk variasjon innan og mellom naturlege populasjonar?
- b) Forklar Hardy-Weinberg loven (HW) med ord, formlar og numerisk døme.
- c) Utfør ein chi-kvadrat Goodness-of-fit test av tilpassing til HW for fylgjande fordeling av genotypar i ein 2-allel (A, B) locus polymorfisme i ein diploid organisme: **$AA:30, AB:40, BB:30$** .
Bruk Appendix Table 1 for å avgjere om avvik frå HW er statistisk signifikant på det vanlege 0,05 nivået.
- d) Forklar omgrepet "Wahlund effekt" og samanhangen det har med Sewall Wright sin F_{ST} .
- e) Rekn ut F_{ST} for populasjonene i Tabell A (nedanfor)
- f) Utfør ein RxC test av homogenitet for genotypar og allel i dei to populasjonane i Tabell A, og bruk Appendix Table 1 for å avgjere resultatet er statistisk signifikant.

Tabell A. Genotypesamansetning i to populasjonar.

	AA	AB	BB
Population 1	25	50	25
Population 2	81	18	1

Oppgåve 2 (vekt 2)

- a) Anta ein enkelt-locus, to-allel blå/kvit hudfarge polymorfisme hjå ein snegleart, der allel (B) for blå farge er dominant over allel (W) for kvit farge. I ei stikkprøve på 100 individ vert det observert 91 blå individ og 9 kvite. Rekn ut allelfrekvensane for allel B og W i stikkprøva.
- b) Kva ville vere dei forventa roporsjonane av farge-fenotypar blant 100 avkom fra ei kryssing mellom ein BB han og ei BW hoe?
- c) Tabell A (nedanfor) inneheld data frå torsk fra Trondheimsfjorden. 50 individ blei genotypa for to autosomale loci, og deira kombinerte genotype-grupper blei talde opp og krysstabulert (Tabell B). Bruk ein passende test for å undersøkje om gruppene av kombinerte genotypar for dei to loci opptrer i dei forventa proporsjonane i Tabell B, eller ikkje. Bruk Appendix Table 1 for å finne og rapportere signifikansnivået (P -verdien) frå testen, og drøft kva for omstende som kan leie til

avvik frå forventa kombinerte genotype proporsjoner ein slik stikkprøve frå ein naturleg populasjon.

Tabell B. 2-locus kombinerte genotyper i en stikkprøve av torsk.

Locus <i>Hbl</i>	Locus <i>LDH3</i>			Σ
	<i>CC</i>	<i>CD</i>	<i>DD</i>	
↕				
<i>AA</i>	4	13	7	
<i>AB</i>	5	14	5	
<i>BB</i>	1	1	0	
Σ				50

Oppgåve 3 (vekt 2)

- Definer omgrepet "effektiv populasjonsstorleik" (N_e).
- Ein naturleg populasjon av storleik $N_e=100$ individ hadde ein heterozygositet $H=0.5$ i generasjon 1. I dei fem påfølgjande generasjonane variera N_e med følgjande st
- orleikar: 70-10-50-30-40. Rekn ut det overordna N_e verdet for heile perioden (generasjon 1- 6).
- Små populasjonar taper genetisk variasjon i kvar generasjon pga tilfeldeg genetisk drift. Storleiken på tapet mellom kvar generasjon fylgjer ein velkjend formel. Rekn ut den forventa heterozygositeten i starten av generasjon 7 i den populasjonen som er beskrevet i b) ovanfor.

Oppgåve 4 (vekt 2)

- Forklar fylgjande omgrep frå avlsgenetikken: parentalgenerasjon, avkomgenerasjon (F1), fenotypisk medelverde (P), seleksjonsdifferensial (S), arvbarheit (H^2 and h^2), seleksjonsrespons (R).
- Ein bestand av mjølkekyr produserte årleg gjennomsnittleg 3000 kg pr individ pr år. Frå denne bestanden vart ei gruppe frå dei mest høgtytande, med eit medelverde på 4000 kg pr individ pr år, velde ut som stamdyr for neste generasjon. Dersom arvbarheita for mjølkeproduksjon er 0,36 – kva vert den forventa gjennomsnittlege produksjon pr dyr pr år i avkomgenerasjonen (F1)?
- Kvifor er ein streng kontroll med og reduksjon av miljøvariasjon så viktig i arvbarheitsstudier?

Oppg ve 5 (vekt 3)

Trepigget stingsild (*Gasterosteus aculeatus*) utgj r eit sv rt godt system for   studere det genetiske grunnlaget for adaptiv evolusjon. Etter siste istid koloniserte og tilpassa marine stingsild seg til mange ferskvasshabitat som nylig var danna. Gjentatt evolusjon av liknande fenotypar skjedde, slik som for eksempel betydeleg tap av beinaktige «panserplater» p  sidene av kroppen (f -plate-type).  tskilte typar i marine milj  og i ferskvatn kan fortsatt danne hybridar, og dette gj r det mulig   kartlegge gen som koder for feks «panserplate»-evolusjon. I 2004 samla Colosimo et al. (PLOS Biology) inn individ av stingsild fr  ein marin populasjon i Japan (alle individ hadde fullstendige «panserplater») og ein ferskvasspopulasjon i vest-Canada (alle individ var av f -plate-typen) og tok dei inn i laboratoriet sitt. P  den tida var den beste genetiske ressursen som var tilgjengeleg for trepigget stingsild eit panel som bestod av 160 mikrosatelitt-mark rar som var spreidd utover i stingsild-genomet.

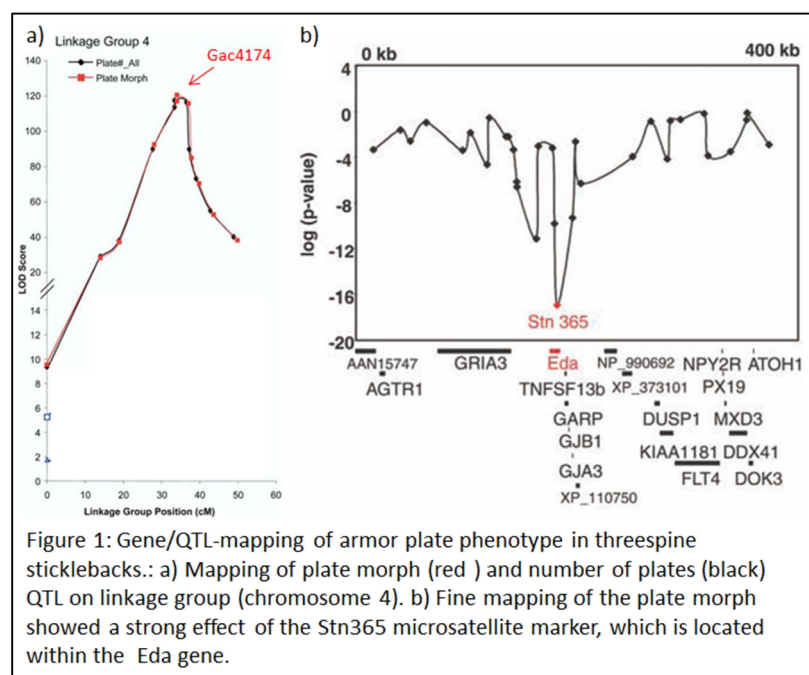
a) Kva for nokre er dei to hovudmetodane som blir brukt for gen-/QTL-kartlegging? Gi ei kort forklaring av kva slags data du treng i kvar av desse framgangsm tane.

b) Forklar kva koplings ulikevekt (linkage disequilibrium, LD) er, og kvifor LD er viktig i dei fleste metodar for gen-/QTL-kartlegging. (Merk at LD er definert som eit spesielt tilfelle av «gametic phase disequilibrium» i boka.)

c) Forklar korleis du ville ha utf rt eit eksperiment der m let er   kartlegge gen/gena som koder for «panserplater» i trepigget stingsild om du hadde dei same ressursane som Colosimo et al. (2004).

Resultata til Colosimo et al. (2004) vises i Figur 1a. Som du ser p  figuren klarte dei   kartlegge gen/QTLen som koder for «panserplate»-morfen og talet p  «panserplater» til ein relativt avgrensa region p 

koplingsgruppe (kromosom) nummer fire. Denne regionen inneheldt ein mikrosatelitt-mark r kalla Gac4174. Finkartlegging basert p  eit nytt sett med 32 mikrosatelitt-mark rar i ein 400kb region n r Gac4174 vart utf rt av Colosimo et al. i ein ny studie i 2005 (Science). Fordelinga av allel p  kvar av dei 32 nye mikrosatelitt-mark rane vart da unders kt i et sampel som bestod av 46 individ med



fullstendige «panserplater» og 45 fisk av «få-plate-typen» frå ein populasjon der desse typene kryssar seg med kvarandre i vill tilstand i California, USA. Resultata frå denne finkartlegginga er presentert i Figur 1b. Som du kan sjå frå figuren identifiserte denne studien genet ectodysplasin (Eda) som det viktigaste locuset som kontrollerer «panserplate»-utvikling hos trepigget stingsild.

Colosimo et al. (2005) konstruerte så eit fylogenetisk tre basert på DNA-sekvensane til Eda genet i 15 ferskvasspopulasjonar med få «panserplater» og 10 marine populasjonar med full «panserplate»-morfologi over heile den nordlige jordhalvkule. Individ frå eit utval av desse populasjonane (12 av få-plate-typen og 8 av typen med fullstendige «panserplater») vart også sekvensert på 25 tilfeldig valde autosomale gen, og eit fylogenetisk tre vart konstruert basert på 193 SNPs i desse gena (merk at

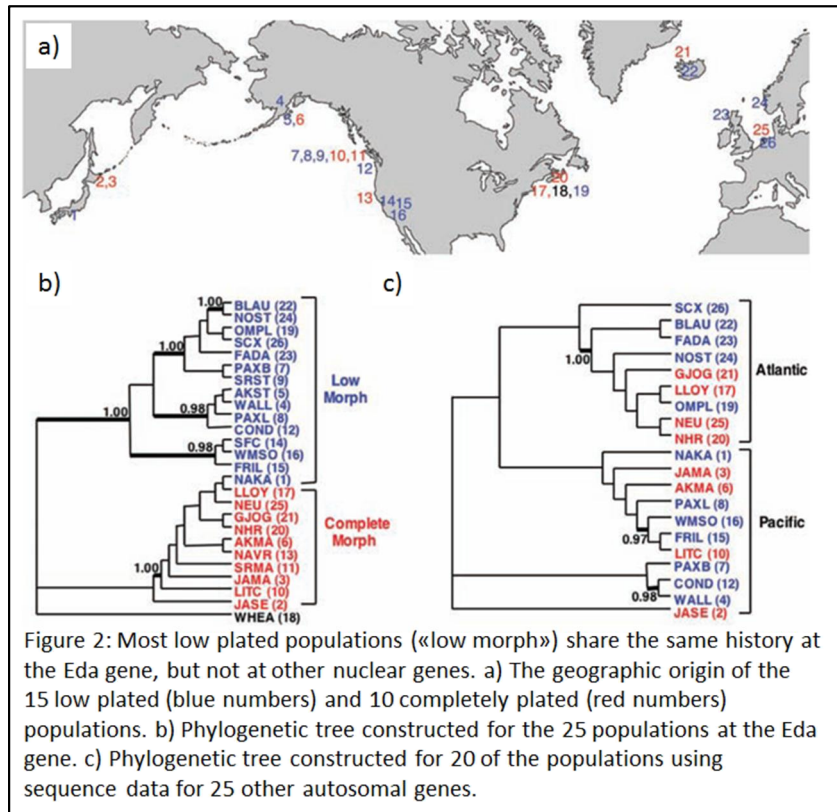
Jones et al. (2012, Nature) estimerte eit fylogenetisk tre med den same topografien for desse populasjonane ved å bruke heilgenom sekvenseringsdata). Plasseringa av dei ulike populasjonane og dei to fylogenetiske trea er vist i Figur 2.

Til slutt samanlikna Colosimo et al. (2005) stingsild av den marine typen (med fullstendige «panserplater») og av ferskvasstypen (utan «panserplater»), og estimerte forskjellige molekylærgenetiske parametarar ved å bruke sekvensdata for ekson i Eda-genet. Dei fann at $d_N=0.0055$, $d_S=0.016$, og dermed at $d_N/d_S=0.344$ (merk at d_N enkelte gonger blir kalla K_N eller K_A , og at d_S enkelte gonger blir kalla K_S). Ein enkel måte for å undersøkje om eit bestemt gen viser teikn til seleksjon er å regne ut d_N/d_S -raten for dette genet når genet sine DNA-sekvensar blir samanlikna i to arter, populasjonar, eller grupper av individ som er forskjellige når det gjeld fenotypen som genet koder for.

d) Kva er d_N og d_S ?

e) Forklar kva ulike verdiar for d_N/d_S -raten indikerer (dvs. $d_N/d_S < 1$, $d_N/d_S = 1$ og $d_N/d_S > 1$).

Ufullstendig avstammings-sortering (incomplete lineage sorting) er eit konsept som enkelte gonger er relevant når ein konstruerer fylogenetiske artstre. Dette fenomenet kan også førekomme på innan-artsnivå når ein samanliknar populasjonar.



f) Forklar kva ufullstendig avstammings-sortering (incomplete lineage sorting) er, og bruk dette konseptet til å gi ei sannsynlig forklaring for mangelen på samsvar mellom dei fylogenetiske trea for Eda (Figur 2b) og andre autosomale gen (Figur 2c).

APPENDIX

Table 1. Critical values of the chi-squared distribution

Percentage Points of the Chi-Square Distribution									
Degrees of Freedom	Probability of a larger value of χ^2								
	0.99	0.95	0.90	0.75	0.50	0.25	0.10	0.05	0.01
1	0.000	0.004	0.016	0.102	0.455	1.32	2.71	3.84	6.63
2	0.020	0.103	0.211	0.575	1.386	2.77	4.61	5.99	9.21
3	0.115	0.352	0.584	1.212	2.366	4.11	6.25	7.81	11.34
4	0.297	0.711	1.064	1.923	3.357	5.39	7.78	9.49	13.28
5	0.554	1.145	1.610	2.675	4.351	6.63	9.24	11.07	15.09
6	0.872	1.635	2.204	3.455	5.348	7.84	10.64	12.59	16.81
7	1.239	2.167	2.833	4.255	6.346	9.04	12.02	14.07	18.48
8	1.647	2.733	3.490	5.071	7.344	10.22	13.36	15.51	20.09
9	2.088	3.325	4.168	5.899	8.343	11.39	14.68	16.92	21.67
10	2.558	3.940	4.865	6.737	9.342	12.55	15.99	18.31	23.21
11	3.053	4.575	5.578	7.584	10.341	13.70	17.28	19.68	24.72
12	3.571	5.226	6.304	8.438	11.340	14.85	18.55	21.03	26.22
13	4.107	5.892	7.042	9.299	12.340	15.98	19.81	22.36	27.69
14	4.660	6.571	7.790	10.165	13.339	17.12	21.06	23.68	29.14
15	5.229	7.261	8.547	11.037	14.339	18.25	22.31	25.00	30.58
16	5.812	7.962	9.312	11.912	15.338	19.37	23.54	26.30	32.00
17	6.408	8.672	10.085	12.792	16.338	20.49	24.77	27.59	33.41
18	7.015	9.390	10.865	13.675	17.338	21.60	25.99	28.87	34.80
19	7.633	10.117	11.651	14.562	18.338	22.72	27.20	30.14	36.19
20	8.260	10.851	12.443	15.452	19.337	23.83	28.41	31.41	37.57
22	9.542	12.338	14.041	17.240	21.337	26.04	30.81	33.92	40.29
24	10.856	13.848	15.659	19.037	23.337	28.24	33.20	36.42	42.98
26	12.198	15.379	17.292	20.843	25.336	30.43	35.56	38.89	45.64
28	13.565	16.928	18.939	22.657	27.336	32.62	37.92	41.34	48.28
30	14.953	18.493	20.599	24.478	29.336	34.80	40.26	43.77	50.89
40	22.164	26.509	29.051	33.660	39.335	45.62	51.80	55.76	63.69
50	27.707	34.764	37.689	42.942	49.335	56.33	63.17	67.50	76.15
60	37.485	43.188	46.459	52.294	59.335	66.98	74.40	79.08	88.38