

Evolutionsbiologi

Skriftlig eksamen, den 18. juni 2014

Med hjælpemidler
(Lærebøger, noter, computer)

For hver opgave oplyses dens vægtning.

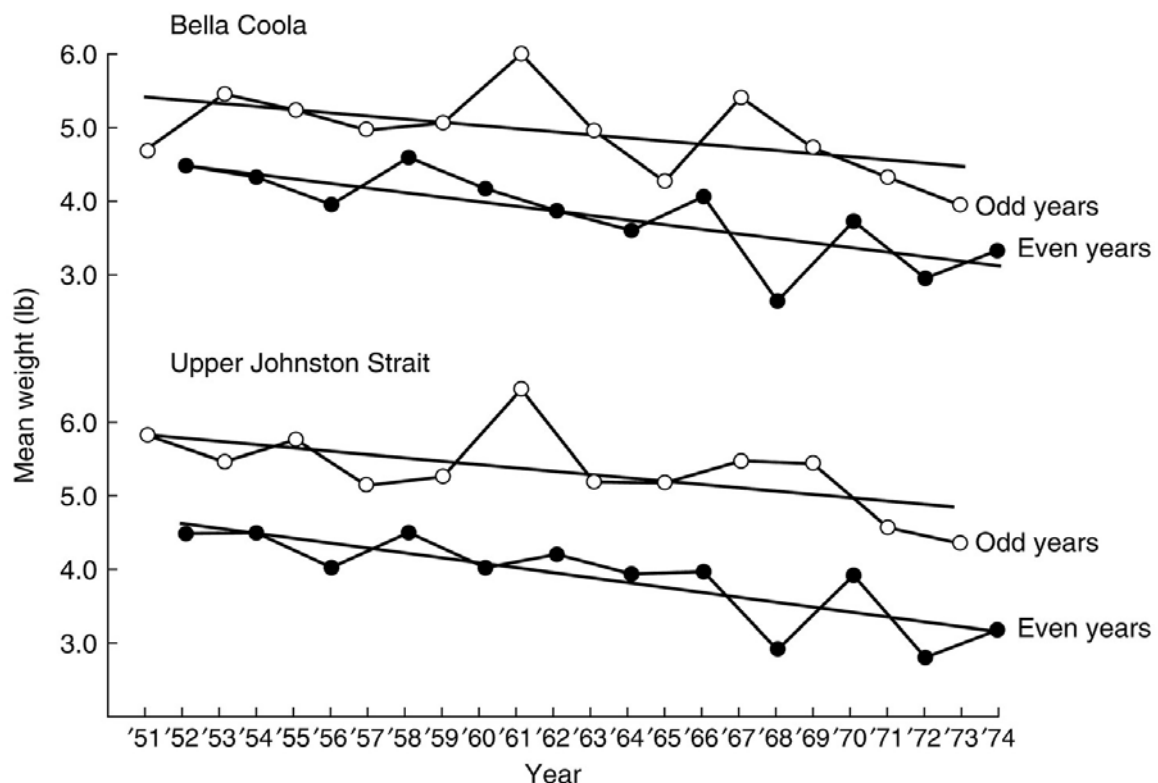
Dette opgavesæt indeholder 7 sider inklusive forsiden
*Der må **IKKE** benyttes blyant.*
Opgaverne skal afleveres i den rækkefølge, de er stillet i.

Jørgen Bundgaard Søren Rosendahl
Hans R. Siegismund Jesper Givskov Sørensen Jeppe Vinther

Opgave 1 (35%)

Pukkellaks (*Oncorhynchus gorbuscha*) er udbredt i det østlige Stillehav og yngler i ferskvand i Canada og USA. Den har en speciel livshistorie. Den lever i nøjagtigt to år. Det meste af livet tilbringes i havet, hvorfra den i to-årsalderen vandrer tilbage til ferskvand. (Den er anadrom.) De fleste vandrer tilbage til den flod, hvor de blev født. Her forplanter laksene sig og dør lige derefter. Vi skal her undersøge, hvorvidt laks, der lever i det samme flodsystem, og som yngler i henholdsvis lige og ulige år er genetisk adskilt fra hinanden. Pukkellaks er en eftertragtet spisefisk og har været fisket efter igennem mange år. Man bruger hovedsageligt net sat op i flodmundinger, hvor man fanger laks, der vender tilbage for at yngle. Figuren nedenunder viser gennemsnitsvægten (i pund, lb) for to floder i USA, Bella Coola og Upper Johnston Strait. Gennemsnitsvægten angives for såvel ulige år (åbne cirkler) som for lige år (fyldte cirkler) i de to floder.

Pink Salmon Spawning Phase



a) Beskriv **kort** de to vigtigste observationer fra figuren.

I begge floder ser vi at laks, der yngler i ulige år, er større end laks, der yngler året før eller efter (i lige år). For begge floder ses desuden, at den gennemsnitlige størrelse aftager jævnt fra 1951 til 1974.

- b) Gennem fiskeriet har man tilsyneladende hovedsageligt fanget de store laks. Hvilken form for selektion foregår?

Retningsselektion mod store laks.

- c) Hvis denne form for selektion på en kvantitativ karakter fortsætter gennem mange generationer, observeres der oftest et selektionsplateau. Hvad er narrow-sense heritabiliteten, h^2 , i selektionsplateauet?

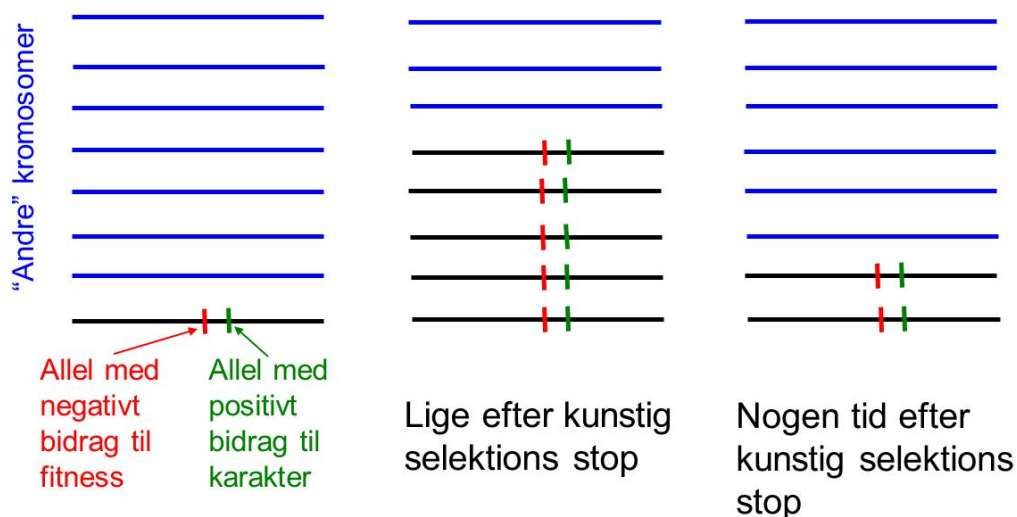
$$h^2 = 0$$

- d) Betyder det, at der ingen variation er for denne karakter i populationen?

Nej, det betyder kun, at den additive genetiske varians, V_A , er nul, da $h^2 = V_A/V_P = 0$. Der kan godt være andre komponenter, der er positive, f.eks. V_D .

- e) Når man standser selektionen på en kvantitativ karakter, ses ofte, at populationens middelværdi bevæger sig tilbage i retning mod den værdi, den havde før selektionen blev påbegyndt. Hvad skyldes dette fænomen?

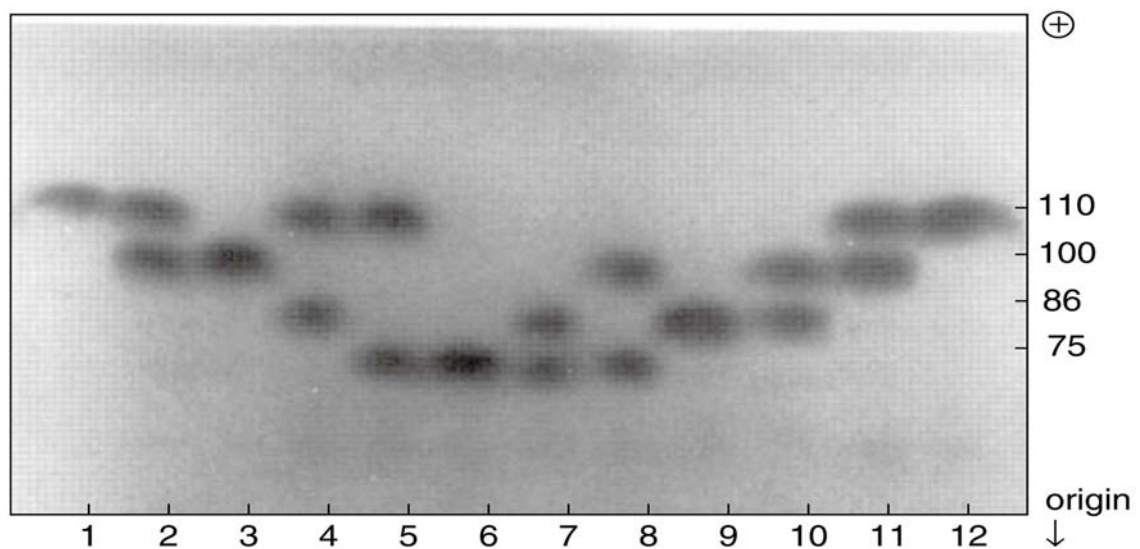
Det skyldes pleiotropi og hitchhiking. De gener, der bidrager til karakteren har for nogles vedkommende en negativ pleiotrop effekt på fitnessrelaterede karakterer. Naturlig selektion vil efter den kunstige selektions ophold fjerne disse igen og indirekte ændre karakterens middelværdi. Hitchhiking skyldes at skadelige gener, der er tæt koblet til gener med bidrag til karakteren, bliver selekteret fordi de er i koblingsligevægt med de selekterede gener. Efter kunstig selektions stop, bliver disse fjernet gennem naturlig selektion.



- f) I alle floder, hvor man har undersøgt pukkellaksen, viser det sig, at laks, der gyder i ulige år, er større end laks, der yngler i lige år. Hvad fortæller dette om slægtskabet mellem laks i de to årgangsgrupper og specielt om migration af laks mellem årgangene?

Laksene i lige og ulige år er to bestande, der har udviklet sig uafhængigt af hinanden. Genflowet mellem dem må være begrænset, ellers ville disse to grupper, der er fænotypisk forskellige med hensyn til størrelse, være smeltet sammen.

- g) Laksefiskeriet har fjernet en stor del af bestanden, så man frygter, den har gennemgået en flaskehals og derved mistet genetisk diversitet og måske udsættes for indavlsdepression. Man har for at undersøge om dette er tilfældet analyseret genetisk variation i en lang række loci. Figuren nedenunder viser resultatet af gel-elektroforese af et protein som bestemmes af fire alleler. De fire typer kan adskilles, da deres vandring i gelen er forskellig. Homozygoter giver ét bånd mens heterozygoter giver to bånd. Til højre er angivet relative mobiliteter af de forskellige alleler: 75, 86, 100, 110. I alt er der undersøgt variation hos 12 individer. Beregn frekvensen for de fire alleler.



$$p(75) = (3 + 1 \times 2) / (2 \times 12) = 0,208$$

$$p(86) = (3 + 1 \times 2) / (2 \times 12) = 0,208$$

$$p(100) = (4 + 1 \times 2) / (2 \times 12) = 0,250$$

$$p(110) = (4 + 2 \times 2) / (2 \times 12) = 0,333$$

- h) Hvad er gendiversiteten i dette locus?

$$\begin{aligned} H &= 1 - \sum_i p_i^2 \\ &= 1 - 0,208^2 - 0,208^2 - 0,250^2 - 0,333^2 \\ &= 0,740 \end{aligned}$$

- i) Gendiversiteten i dette locus er repræsentativt for den gennemsnitlige gendiversitet i denne art. Det svarer til det, der ses i gennemsnit hos laksefisk. Er der noget tegn på, at den har været igennem en flaskehals og har mistet en væsentlig del af sin variation?

Nej, det ser ud til, at der er meget variation i arten.

- j) Man har undersøgt, hvor meget af gendiversiteten, der skyldes differentiering mellem populationerne. Dette er beregnet for såvel gruppen, der yngler i lige år, og gruppen, der yngler i ulige år. Man benyttede Wright's F-statistik for at kvantificere dette. For de to grupper fandt man en F_{ST} på 0,005 for gruppen, der ynglede i ulige år, og en F_{ST} på 0,002 for gruppen, der ynglede i lige år. Til sammenligning kan nævnes, at laksen i Europa (*Salmo salar*) har en F_{ST} værdi for differentiering mellem dens populationer på 0,16. Den europæiske laks er kendt for at vende tilbage til den flod, hvor den stammer fra, når den skal yngle.

Et fylogenetisk træ for alle populationer estimeret ud fra allelfrekvenser i 16 loci blev delt op i to monofyletiske grupper: én med populationer, der yngler i ulige år og én med populationer, der yngler i lige år. Grupperne blev understøttet af en bootstrap-værdi på 1000 (ud af 1000 tilfældige subsamples).

Benyt disse oplysninger til at give et kvalitativt bud på, hvor stor migrationen er mellem forskellige floder og for genflow mellem lige og ulige årgange af laksen.

Den meget lave F_{ST} værdi for begge grupper tyder på, at laksene ikke er så stedfaste og nogle gange ikke vender tilbage til, hvor de blev født. Der synes at være betydeligt mere migration mellem floder end det er tilfældet med den europæiske laks. Til gengæld viser det fylogenetiske træ, der stort set ikke at være gen-flow mellem årgangene. Dette støttes også af, hvad der blev nævnt under spørgsmål f).

Opgave 2 (25%)

Den europæiske musvåge (*Buteo buteo*) findes i tre forskellige morfer, en mørk (M), en intermediær (I) og en lys (L) morf. De er afbildet i figuren til højre i samme nævnte rækkefølge. Krüger et al. (2001) undersøgte ud fra feltobservationer, hvordan farvemorferne blev nedarvet.



De undersøgte udspaltningen i forskellige fænotyper blandt afkom for alle mulige kombinationer af parringer mellem de tre morfer. Resultatet er givet i nedenstående tabel, hvor den procentvise udspaltning er angivet for afkomsgenerationen (N er antallet af par):

Forældre	N	Afkom		
		M	I	L
M × M	2	100	0	0
M × I	22	36	64	0
M × L	4	0	100	0
I × I	90	23	64	13
I × L	41	2	49	49
L × L	3	0	0	100

- a) Forklar, hvordan farvemorferne nedarves hos musvågen. Afkommets udspaltning er angivet i procenter, og der er derfor ingen mulighed for at lave statistiske tests.

Kommenter, hvor der optræder en observation, der er problematisk og giv et bud på, hvordan den kan forklares.

Det tyder på, at farverne bestemmes af to allele gener, M og m , i et autosomalt locus. Genotypen MM resulterer i den mørke fase, Mm i den intermediære fase og mm i den lyse fase. Udspaltningerne—på nær én eller nogen få, se nedenfor—passer fint med denne hypotese.

Der synes dog at være et lille problem med kombinationen $I \times L$. Her forventer man ikke, at der optræder et afkom, der er mørkt, M . Det kan enten skyldes, at det kan være svært at score den rigtige farvemorf, eller det kan være resultatet af en parring, hvor en heterozygot hun (I, Mm) i krydset $I \times L$ har parret sig med en mørk eller intermediær han og fået et mørkt afkom (MM).

- b) Tæl antallet af de tre morfer i forældregenerationen op (Beregn den observerede genotypefordeling blandt voksne)

Morf	Mørk	Intermediær	Lys	Sum
Antal	$2 \times 2 + 22 + 4$ = 30	$22 + 2 \times 90 + 41$ = 243	$4 + 41 + 2 \times 3$ = 51	324

- c) Beregn frekvenserne for de to alleler.

$$p(\text{mørk}) = (2 \times 30 + 243) / (2 \times 324) = 0,468$$

$$p(\text{lys}) = (2 \times 51 + 243) / (2 \times 324) = 0,532$$

- d) Hvad er de forventede genotypefrekvenser ifølge Hardy-Weinberg loven? Karakteriser en eventuel afvigelse og lav en test for afvigelse fra Hardy-Weinberg loven.

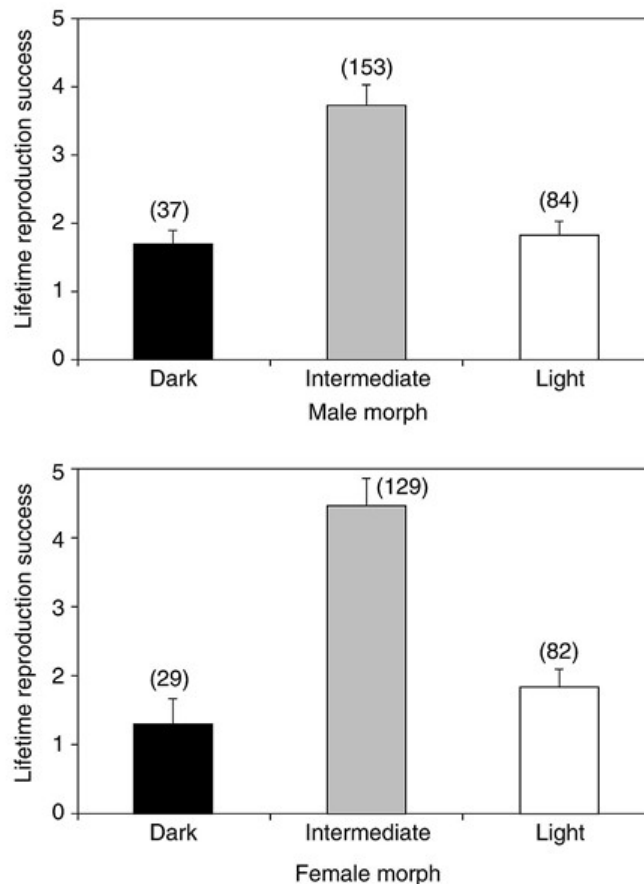
Morf	Mørk	Intermediær	Lys	Sum
Obs	30	243	51	324
Exp	70,84	161,32	91,84	

$$\chi^2 = (30 - 70,84)^2 / 70,84 + (243 - 161,32)^2 / 161,32 + (51 - 91,84)^2 / 91,84$$

$$= 83,06$$

Der er et stort overskud af heterozygoter og et tilsvarende underskud af homozygoter i forhold til Hardy-Weinberg proportioner.

Feltstudiet af musvågen omfattede også en opgørelse over, hvor mange afkom de tre forskellige morfer fik igennem et helt liv (lifetime reproductive output). Dette er vist i den følgende figur.



- e) Når lifetime reproductive output benyttes som et bud på fitness for de tre genotyper i dette locus, besvar da hvilken form for selektion, der foregår i dette locus.

Overdominant selektion, heterozygoten har den største fitness. Faktisk mere end dobbelt så stor som homozygoternes.

- f) Hvad vil allelfrekvenserne være i et sådant system (cirka)? Stemmer de overens med de observerede allelfrekvenser (fundet i c)?

Fitness af de to homozygoter er næsten ens, med den lyse havende en lidt større fitness. Derfor forventer vi en stabil ligevægt med en allelfrekvens på ca. 0,5 for hver af dem, med en lidt højere frekvens for den lyse allel. Dette stemmer fint overens med de fundne værdier.

Opgave 3 (20 %)

Betrakt de følgende DNA sekvenser:

Art1	actagcagat	cgacaatggt	actgagatag	agctcctaga
Art2C...
Art3	...a.....C.C..	...a..C...
Art4	...a.....C.C..	...g..C...

Her er baser, som i Art 2, 3 og 4 er identiske med basen i den homologe position i Art 1, vist som et punktum.

- a) Udregn antallet af mutationer der adskiller disse sekvenser, ved at færdiggøre følgende tabel af afstande:

	ART1	ART 2	ART 3	ART 4
Art 1	-			
Art 2		-		
Art 3			-	1
Art 4			1	-

	ART1	ART 2	ART 3	ART 4
Art 1	-	1	5	5
Art 2	1	-	4	4
Art 3	5	4	-	1
Art 4	5	4	1	-

- b) Vis, hvilket træ (uden rod) brug af UPGMA metoden ville føre til (tip: du behøver kun at finde de første to arter som skal grupperes sammen for at finde træet uden rod).

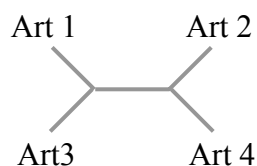
Art 1 og 2 samt/eller art 3 og art 4 har den mindste distance så de grupperes først sammen. Dette leder til træet ((1, 2), (3, 4)).

- c) Hvad er nukleotiddiversiteten for disse sekvenser.

Her benytter vi matricen med parvise antal forskelle mellem sekvenserne, hvor vi dividerer summen med antallet af sammenligninger og med sekvenslængden.

$$\pi = (1 + 5 + 4 + 5 + 4 + 1)/(6 \times 40) = 0,083$$

- d) Hvor mange mutationer skal der introduceres for at forklare forskellen i disse sekvenser ifølge parsimoniprincippet på nedenstående fylogeni:



9

- e) Hvilket træ uden rod er maksimum parsimoni træet, d.v.s. træet med færrest mulige mutationer ifølge parsimoniprincippet?

((1, 2), (3, 4)): 6

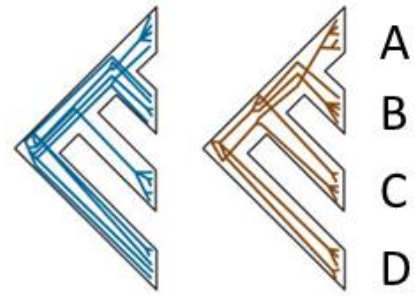
((1, 3), (2, 4)): 9 (fra 3d)

((1, 4), (2, 3)): 9

Træet ((1, 2), (3, 4)) er mest parsimonisk.

Opgave 4 (15 %)

Man kan studere tæt beslægtede arters stamtræer ved at estimere fylogener baseret på flere forskellige gener. Hvis man har en velunderbygget artsfylogeni, ser man nogle gange, at forskellige geners fylogenetiske træer ser anderledes ud. Gentræerne afspejler ikke artstræerne. Et eksempel er vist i figuren til højre, der viser fire arters fylogeni (A, B, C, D) samt to geners fylogeni.



- a) Hvad kaldes dette fænomen? Forklar, hvilke faktorer, der kan have haft indflydelse på dette. (Tip: inddrag populationsstørrelserne og tiden).

Incomplete lineage sorting. Når tiden, en fælles forfader har eksisteret (for A og B eller A, B og C) har været kort, vil der være en større sandsynlighed for, at forskellige haplotypers fylogeni afviger fra artstræet. Det samme er tilfældet, hvis populationsstørrelsen af en fælles stamform har været stor. Så er der også en stor sandsynlighed for, at forskellige gener i stamformen ikke har én fælles forfader i stamformen, men at forskellige haplotyper ender op i de afledte arter.

Ser man på kromosomal variation hos nært beslægtede taxa, ser man et andet billede. Her ser man nogle gange, at populationer, der er parapatrisk udbredte, er fikserede for forskellige kromosom-mutationer, som inversioner eller translokationer. Det ses specielt hos populationer, der ikke migrerer langt og som har små populationsstørrelser.

- b) Hvilken form for naturlig selektion er virksom, og hvorledes er de forskellige populationer blevet fikseret for kromosomvarianterne?

Heterozygoter for kromosomale varianter danner ofte ubalancerede gameter og har derfor en nedsat fitness. Det er underdominant selektion. Årsagen til, at de forskellige populationer er fikserede for forskellige kromosommutationer, skyldes formentlig genetisk drift, hvor en kromosommutation, der har en nedsat fitness hos heterozygoter, kan fikseres i populationer med en lille størrelse.

En endnu mere ekstrem form for kromosomal ændring ses ved artsdannelse, der omfatter hele genomet fra to forskellige arter.

- c) Hvilken form for artsdannelse drejer det sig om? Hos hvilke organismer forekommer den hyppigt.

Allopolyploidi. Den er almindelig hos planter.

Opgave 5 (5 %)

Figuren viser et UCSC genome browser screenshot af en del af det humane chromosome 3. Området indeholder et proteinkodende gen: ADP-ribosyl-arginine hydrolase.

- a) Ifølge RefSeq annoteringen som er vist på figuren, hvor mange mRNA isoformer er der af dette gen?

3 mRNA isoformer (de øverste tre rækker, UTR er lave kasser, exons høje kasser, introns er resten)

- b) Hvilken form for proces er ansvarlig for fremstilling af de forskellige isoformer?

Alternativ splejsning

- c) Ifølge RefSeq annoteringen som er vist på figuren, hvor mange forskellige protein-isoformer vil mRNA isoformerne kode for?

2 protein isoformer

- d) Ud fra konserveringsannoteringen vist på figuren, hvilke dele af genet er højest konserveret og hvad er en sandsynlig forklaring herpå?

De proteinkodende exons er højest konserverede. Proteinet er en funktionel del af genet, og mutationer i den proteinkodende del vil derfor være udsat for stærkere negative selektion end introns og UTR regioner.

