

UNIVERSITETET FOR MILJØ- OG BIOVITENSKAP



Forord

Etter å ha fullført en bachelor i Molekylær biologi ved Universitetet i Bergen hadde jeg et ønske om å fordype meg i genetikk. Valget falt derfor på Bioteknologi ved Universitetet i Ås. Jeg har hatt to fantastiske år ved UMB, deriblant et semester med utveksling til USA. Masteroppgaven har gitt meg nye kunnskaper innen genetikk, praktiske metoder og bioinformatikk.

I begynnelsen av første året på masterprogrammet hadde jeg en meget interessant forelesning med Professor Sigbjørn Lien. Forelesningen omhandlet "Single Nucleotide Polymorphisms" og bruken av disse innenfor avl. Det ble også gjort rede for arbeidet hos CIGENE og hva slags forskning som ble gjort der. Forelesningen og arbeidet ved CIGENE fremstod som særdeles interessant og utfordrende, derfor valgte jeg å ta masteroppgaven min hos CIGENE.

Oppgaven har gitt meg inspirasjon til å fortsette å jobbe innenfor genetikk. Prosessen har vært lærerik og utfordrende og jeg har lært mange nye metoder og teknikker. Neste steg er å finne en jobb som er like utfordrende og interessant som denne masteroppgaven.

Jeg vil takke Tomas Moen og Professor Sigbjørn Lien for god veiledning gjennom hele oppgaven. Tusen takk for all hjelp. Jeg vil også takke Kent Matthew som har gitt god veiledning i laboratoriet. I tillegg vil jeg takke Kristil Sundaasen, Helene Meaas Svendsen, Linda Ripel, Anne Guri Marøy og Hanne Hellerud Hansen som alle har vært svært hjelpsomme med mitt labarbeid.

Ås, mai 2013

Bente Kristin Velle

Sammendrag

IPNV infiserer Atlantisk laks i oppdrettsanlegg som skaper store økonomiske kostnader for oppdrettsnæringen og lidelse for laksen. Avlsselskap har tatt i bruk en ny metode for å avle frem laks som har forbedret resistens mot infeksjon av IPNV. Metoden benytter seg av genetiske markører som er koblet til en QTL som forklarer store deler av variasjonen i egenskapen. Det eksisterer to genotyper for dette området, der den ene er koblet til den resistente genotypen(QQ), og den andre er koblet til den ikke-resistente genotypen (qq). I QTL området er det identifisert to gener som kan være ansvarlige for resistens mot IPNV. Den beste kandidaten er E-cadherin, som fungerer som en transmembran reseptor i epithelia-vevet, men det har ikke blitt identifisert et kausativt gen og polymorfisme som forklarer denne variasjonen. Målet med denne avhandlingen var å analysere mRNA sekvensen til E-cadherin for å identifisere mulige variasjoner mellom de to genotypene.

Ved hjelp av genetisk bioteknologiske metoder har mRNA sekvensen blitt studert ved benyttelse av 31 prøver av Atlantisk laks i voksen- og yngelfase. Dette har ført til deteksjon av en mulig kausativ SNP som forårsaker skifte av aminosyre i posisjon 325 i proteinsekvensen. Variasjonen er en punktmutasjon som medfører en endring fra serine til proline. Disse aminosyrene har forskjellige egenskaper som kan forårsake endringer i proteinsekvensen. I tillegg er det mulighet for at punktmutasjonen er i et sensitivt område i enden av EC2-domenet og i begynnelsen av EC3-domenet. SNPen er en meget god kandidat, men videre arbeid er nødvendig for å bekrefte om mutasjonen er kausal i forhold til å gjøre laks resistent mot IPNV.

Abstract

IPNV infecting of Atlantic salmon creates substantial economic costs for the aquaculture and suffering for salmon. Breeding companies have adopted a new method to breed salmon that have a higher resistance to infection to IPNV. In this method they are using markers linked to a QTL that accounts for 83% of the variation. There are two genotypes for this area on chromosome 26; one connected to the resistance phenotype (QQ) and one connected to the non-resistant phenotype (qq). In the QTL area it has been identified two genes that may be responsible for the IPNV resistance. The main candidate is E-cadherin, which functions as a transmembrane receptor in epithelia tissue. The causative gene and underlying causative DNA variation explaining this variation has not yet been identified. The aim of this thesis is to analyze the mRNA sequence of the E-cadherin and to identify possible differences between the two genotypes.

Using genetic biotechnological methods, the mRNA has been characterized by sequencing 31 samples of Atlantic salmon. The animals were either adults or in the fry stage. This has led to the observation of a possible causative SNP result in a change from serine to proline at position 325 in the amino acid sequence. Such a structural change at the end of EC-2 domain can maybe affect how Ca^{2+} bind to the link between EC2 and EC-3. This may change the structure of the enzyme, which in turn may affect the function. Although these results are promising it is premature to conclude that this is the causative mutation given the limited number of salmon included in the study. Ongoing work is needed to verify that this is the causative SNP that make some individuals more resistance against IPNV than others.

Innholdsfortegnelse

1. INTRODUKSJON.....	8
1.1 Atlantisk laks	8
1.2 Betydningen av Atlantisk laks i akvakultur.....	10
1.3 Produksjonen av Atlantisk laks i norsk akvakultur	10
1.4 IPNV	12
1.5 Avl av IPN resistente laks basert på familie seleksjon	14
1.6 Markørassistert seleksjon (MAS) for økt IPN-resistens	15
1.7 Identifikasjon av enkel nukleotid polymorfisme (SNP) mellom ulike IPN-QTL genotyper	16
1.8 E-cadherin	17
1.9 Mulige kilder til variasjon.....	20
1.10 Mål for oppgaven	21
2. MATERIAL OG METODER.....	22
2.1 Dannelse av en referansesekvens av cDNA til E-cadherin og dens paraloge gen.....	22
Sekvens material	22
est2genom.....	23
UGENE	23
ExPasy Translate	24
Basic Local Alignment Search Tool (BLAST)	24
2.2 Primerdesign:	25
Primer3.....	25
”Mismatch” program	26
Sammenligning av ”mismach” fil og Primer3 fil.....	26
2.3 Laboratoriemetoder	34
Forberedelse til PCR, sekvensering og qPCR	34
Kvalitative metoder.....	36
- PCR.....	37
- Agarose gel elektroforese:.....	37
- Gradient PCR med endring i $MgCl_2$ konsentrasjonen	38
- DNA sekvensering.....	39
- Analyse av kvantitativ metode resultat	40
- Phred	40
- Phrap.....	41
- Consed	41

- PolyPhred.....	41
Kvantitative metoder	42
- qPCR.....	42
- Effektivitets test	43
- qPCR reaksjonen.....	45
- Delta Delta Ct	45
3. Resultat og Diskusjon	46
3.1 Sekvenser av mRNA til E-cadherin på kromosom 26 og kromosom 11.....	46
3.2 Variasjon mellom mRNA sekvensen til genotype qq og QQ til E-cadherin.....	47
3.3 Fra mRNA til aminosyre sekvens	50
Figur 16: Sammenligning av aminosyresekvensen til genotype QQ og genotype qq	52
3.4 Komparativ analyse mellom begge genotypene	53
3.5 Måle ekspresjons nivå av E-cadherin (qPCR)	54
4. Oppsummering	58
5. Framtidig perspektiv	59
6. Referanser	60
Appendix:	71
I. Primersett 1.	71
II. Primersett 2	73
III. Buffer og løsninger	73
IV. Referansesekvensen til cDNA av E-cadherin på kromosom 26.....	73
V. Referansesekvensen til cDNA av det paraloge genet på kromosom 11.....	74
VI. Protokoll til PCR.....	75
VII. Protokoll til gradient PCR:.....	76
VIII. Protokoll til gel elektroforese:.....	79
IX. Forberedelse til Sanger Sekvensering.....	80
X. Oversikt over konsentrasjonen over mRNA brukt i qPCR.....	82
XI. Optimaliserte primerpar	82
XII. Ladder som ble brukt i gel elektroforese	83
XIII. Sekvens til E-cadherin på kromosom 26 (genotype QQ)	83
XIV. Sekvensen til E-cadherin på kromosom 26 (genotype qq)	84
XV. Sekvensen til E-cadherin på kromosom 26 (genotype Qq).....	85
XVI. Sekvensen til det paraloge genet på kromosom 11	86

XVII. Sammenligning av konsensussekvensene til ulike genotypene til E-cadherin på kromosom 26 til konsensussekvensen til paralog på kromosom 11.....	87
XVIII. Komparativ sammenligning mellom ulike arter	100

Figurer

Figur 1: Kromosomene til laks	9
Figur 2: Livssyklusen til villaks	12
Figur 3: Genomet til IPN viruset	13
Figur 4: Kapsidet til IPN viruset	13
Figur 5: Klassisk Cadherin.....	18
Figur 6: E-Cadherin	19
Figur 7: Genet E-cadherin på kromosom 26 og dens mRNA sekvens	20
Figur 8: Primerpar i Primersett 1	28
Figur 9: Qiagen Revers Transkripsjon.....	36
Figur 10: Agarose gel med PCR produkt av Primersett 1	38
Figur 11: Innmerkingsreaksjon.....	40
Figur 12: Eksempel på et amplifikasjonsplott	44
Figur 13: Standardkurve fra samme eksperiment som amplifiseringsplottet	44
Figur 14: Alignment av sekvenser	49
Figur 15: Serine og Proline	51
Figur 16: Sammenligning av aminosyresekvensen til genotype QQ og genotype qq	52
Figur 17: E-cadherin sammenligning	53
Figur 18: Genotype qq mot genotype QQ	55
Figur 19: Genotype qq mot genotype QQ	56
Figur 20: Smittet mot ikke smittet i genotype QQ	57
Figur 21: Smittet mot ikke smittet i genotype qq	57
Figur 22: "Mass DNA ladder"	83

Tabeller

Tabell 1: Identifiserte SNPer	49
Tabell 2: Primersett 1	71
Tabell 3: Primersett 2	73
Tabell 4: Buffere	73
Tabell 5: Mastermiks for PCR reaksjon	76
Tabell 6: PCR program	76
Tabell 7: Mastermiks for gradient PCR.....	77
Tabell 8: Mengde av MgCl ₂ og H ₂ O.	78
Tabell 9: Program for PCR	78
Tabell 10: Temperatur under amplifisering.....	78
Tabell 11: Mastermiks for innmerkingsreaksjonen	81
Tabell 12: PCR programmet for innmerkingsreaksjonen	81
Tabell 13: Konsentrasjonen til RNA prøvene brukt i qPCR	82
Tabell 14: Forholdene for optimaliseringen av primerpar i Primersett 1	82

Forkortelse i tekst

Bp	Basepar
BLAST	Basic Local Alignment Search Tool
Ct	Threshold syklus
ddNTP	Dideoxysyribonucleotides
dNTP	Deoxyribonucleotides
E-cadherin	Epitelial-cadherin
EC-domenet	ekstracellulær cadherin domenet
HSMB	Hjerte- og skjelettmuskelbetennelse
ICSASG	International Collaboration to Sequence the Atlantic Salmon Genom
IgH	Immunoglobulin heavy chain
ILA	Infeksiøs lakseanemi
IPN	Infeksiøs pankreasnekrose
IPNV	Infeksiøs pankreasnekrose virus
MAS	Markør-assistert seleksjon
nr/nt	nucleotide collection
ORF	Open leseramme
PCR	Polymerase kjedereaksjon
PD	Pankreas sykdom
SNP	Single nucleotide polymorphism
QTL	Quantitative Trait Locus
qPCR	Real-Time PCR
RLT	Guanidine-thiocyanate

1. INTRODUKSJON

1.1 Atlantisk laks

Atlantisk laks (*Salmo salar*) har gjennom tidene vært en fisk som har fanget menneskets oppmerksomhet. Under Julius Cæsars tid ble fiskens evne til å hoppe opp Gallias elven lagt merke til av soldatene hans. Fisken ble gitt navnet ”*Salmo the leaper*” (Johansson 1997).

Denne fasinasjonen har vart fram til i dag, hvor den har gjort laksen til en sentral del, blant annet innen sportsfiske og matproduksjon i Norge og resten av verden.

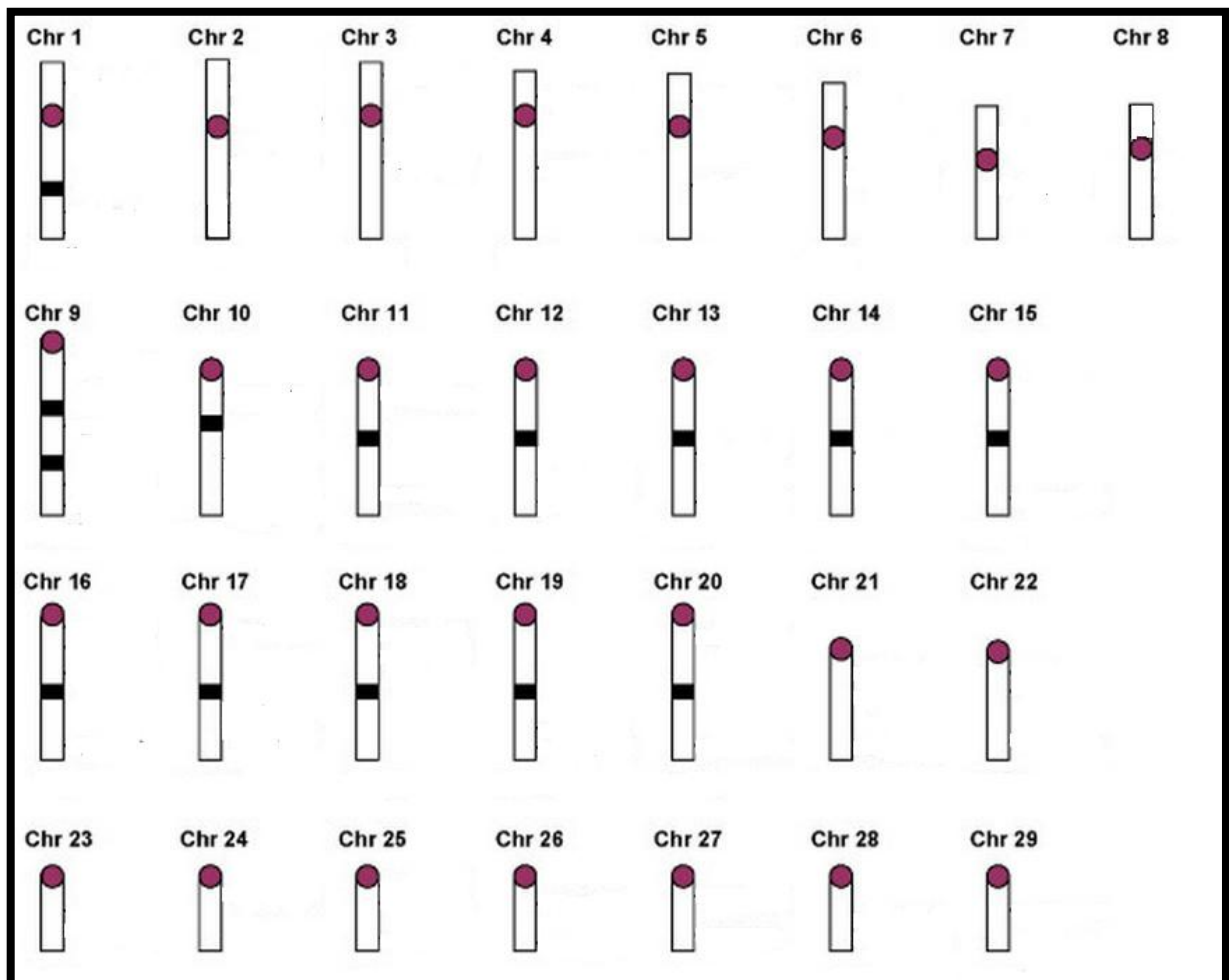
Avlselskaper, som Aqua Gen, har hatt stor framgang i å avle frem gunstige egenskaper i Atlantisk laks. Her har seleksjon på økonomiske viktige fenotyper, som blant annet resistens mot alvorlige sykdommer, filefarge og tilvekst stått i fokus. Her møter avlsselskapene mange utfordringer og det er stor interesse for å bruke genominformasjon for å øke presisjonen i avlarbeidet. Atlantisk Laks tilhører subfamilien Salmonidae i ordenen Salmoniformes sammen med arter som ørret, harr og regnbueørret. Forfedre til artene har trolig gjennomgått en hel genom duplisering for 25-100 millioner år siden. Data tyder på at stamfedre av subfamilien Salmonidae har eksistert som tetraploid og dette har gitt oss de ulike artene vi har i dag. Tetraploidseringen genererte homologe kromosom, og dermed dannet paraloge gener på ulike kromosom (Venkatesh 2003). I dag eksisterer det fortsatt stor sekvenslikhet mellom de dupliserte sekvensområdene noe som vanskeliggjør konstruksjonen av en god genomsesekvens for Atlantisk laks (Davidson 2010) og kompliserer arbeidet med å kartlegge gener som påvirker viktige egenskaper i oppdrettslaks og villaks. Eksempel på dette er to immunoglobulin ”heavy chain” (*IgH*) ”locus”, IGH-A og IGH-B, som har 81-85% likhet i deler av sekvensen (Yasuike, de Boer et al. 2010). Genomet til laks er fortsatt i endring og er trolig på vei tilbake til en stabil diploid tilstand (Lien, Gidskehaug et al. 2011).

En annen utfordring med genomet til laks er repeterende elementer. Disse finnes i store deler av genomet og har ofte en lengde på mer enn 1500 bp (Davidson, Koop et al. 2010 B).

Sammen med store sekvenslikheter mellom de dupliserte områdene, skaper dette problemer for dannelsen av en fullstendig referansesekvens. I dag eksisterer det kun en midlertidig referansesekvens av genomet til laks i databasen Genbank(AGKD00000000.1), men dette er ikke en fullstendig sekvens. Denne ble dannet av ”International Collaboration to Sequence the Atlantic Salmon Genom”(ICSASG). Referansesekvensen er basert på sekvensering av genomet til Europeisk Atlantisk laks som består av 29 par kromosomer, vist i figur 1

(Davidson, Koop et al. 2010 A). Målet for ICSASG er å danne en fullstendig og annotert genomsekvens (Davidson, Koop et al. 2010). Denne genomsekvensen kan blant annet benyttes i arbeidet mot sykdommer i oppdrettsnæringen. Det har blitt forsket på immunforsvaret i mange år og man har klart å utvikle vaksiner mot diverse bakterier. Problemet ligger i utvikling av vaksiner mot virus, da det viser seg at man ikke har tilstrekkelig med informasjon om immunresponsen til å danne en effektiv vaksine. Her kan genomsekvensen benyttes til å organisere kjente immunogener samt å identifisere nye. Dette vil være en begynnelse i arbeidet for å bygge opp kunnskapen om immunresponsen til laks ved infeksjon av virus (Fiskeri-og-havbruksnæringens-forskningsfond. 2012 s. 35).

Figur 1: Kromosomene til laks



Figur 1 viser oversikten over de 29 ulike kromosomene som danner genomet til Europeisk laks. (Phillips, Keatley et al. 2009)

1.2 Betydningen av Atlantisk laks i akvakultur

Akvakultur bidrar til å tilfredsstille den økende etterspørselen etter mat med produksjon av ulike fiskearter. Mens antall fisk fanget under fiske i hele verden har stagnert, har den totale produksjonen i akvakultur hatt en positiv økning fra 47,3 million tonn fisk i 2008 til hele 63,6 millioner tonn i 2011. Produksjonen i Norge følger den samme trenden, fra 151 000 tonn i 1990 til 1,46 millioner tonn i 2010. Dette gjør Norge til den ledende produsenten i Europa (39,95 %), og den sjuende største produsenten (1,68%) på verdensmarkedet i 2010. Den arten som har vært den viktigste bidragsyteren til denne produksjonsveksten i Norge er nettopp Atlantisk laks. Norge er en av de dominerende produsentene av laks (Food-and-Agriculture-Organization 2012) og i 2012 kom hele 60% av produksjonsmengden av laks fra Norge (Bessesen 08.04.2013). Selv om akvakulturen har en positiv økning i antall fisk produsert, er det fortsatt mange faktorer rundt produksjonen av laks som fører til store økonomiske tap. Her er det muligheter for optimalisering av disse faktorene og øke produksjonsmengden.

1.3 Produksjonen av Atlantisk laks i norsk akvakultur

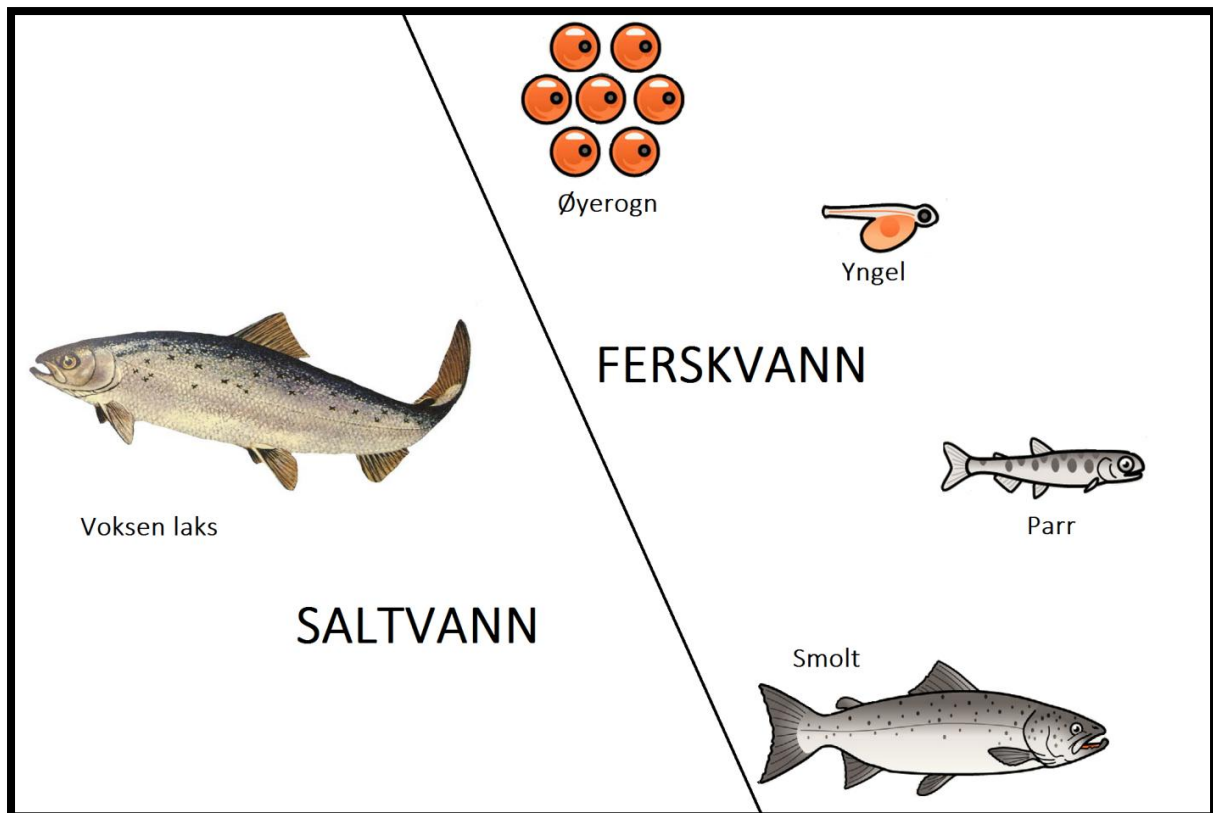
Produksjonen av laks i norske oppdrettsanlegg etterligner den anadrome livsstilen til villaksen. Dette innebærer at den lever de første to til fem årene av livssyklusen som plommeseekkyngel og yngel i ferskvann (Møller Christensen 1977). Neste fase starter med smoltifisering, der yngelen gjennomgår komplekse endringer i atferd, morfologi og fysiologi, for å kunne vandre ut i havet. Disse endringene inkluderer utvikling av hypoosmoregulatorisk evne og økning i veksthormon (Handeland, Imsland et al. 2013). Første fase i livssyklusen hos oppdrettslaks er mellom 8 til 18 måneder med foring i ferskvann. Den vil deretter gjennomgå smoltifisering, som starter den neste fasen i livssyklusen. Smolten blir plassert i merder i saltvann, der den vokser til 3-6 kg i løpet av 12-18 måneder (Havforskningsinstituttet 2009; http://www.imr.no/temasider/fisk/laks/laks_i_oppdrett/nb-no).

Det er mange utfordringer for å finne de optimale forholdene i produksjonen av laks. Foringstype, antall individer, merde størrelse og lokalitet er alle faktorer som påvirker lakseproduksjonen (Havforskningsinstituttet 2009; <http://www.imr.no/temasider/akvakultur/lakseoppdrett/>). Et stort antall individer samlet på et

lite område, medfører at fisken blir mer sårbar for infeksjon av patogener, ettersom patogenene da lettere kan formere seg og spre seg til andre verter. Resultatet er stor dødelighet som fører til store økonomiske kostnader og lidelse for fisken. Hovedutfordringen innenfor infeksjon av patogener, er for tiden virus og lakselus. Ved infeksjon av virus har særlig infeksjøs lakseanemi (ILA), Pankreas sykdom (PD), hjerte- og skjelettmuskelbetennelse (HSMB) og Infeksjøs pankreasnekrose (IPN), skapt store problemer (Veterinærinstituttet 2013 s. 8-14). I naturlige biologiske fauna har ikke disse virusene store skadevirkninger, trolig fordi fisken har et stort område å bevege seg på. I tillegg har villaksen hatt lengre tid til å tilpasse seg disse utfordringene, blant annet ved å få et motstandsdyktig immunforsvar. Derimot har oppdrettslaksen et genom og immunforsvar som trolig ikke er godt nok tilpasset det nye sykdomspresset i merdene. I oppdrettsanlegg er det også faktorer som kan ha medført økt stresspåkjenning for laksen. Stresset påvirker trolig immunforsvaret og gjør den svakere i kampen mot infeksjoner (Veterinærinstituttet 2013 s. 13). Kombinasjonen av disse faktorene har trolig ført til stor dødelighet av laks ved infeksjon av IPN viruset (IPNV).

IPNV har vært kjent helt siden 1940 tallet (M'Gonigle 1941), og ble isolert fra laks i Norge i 1975 (Hastein and Krogsrud 1976). Det var ikke før i 1985 det ble oppdaget stor mortalitet i Atlantisk fisk post smolts ved oppdrettsanlegg (Krogsrud, Håstein et al. 1989). Siden da har IPNV blitt en av de mest alvorlige virusinfeksjonene i Norge. I 2009 var det hele 223 utbrudds tilfeller i Norge ved ulike anlegg (Veterinærinstituttet 2013 s. 7). Dette har medført store produksjons- og økonomiske tap, både i ferskvannfasen og saltvannsfasen.

Figur 2: Livssyklusen til villaks

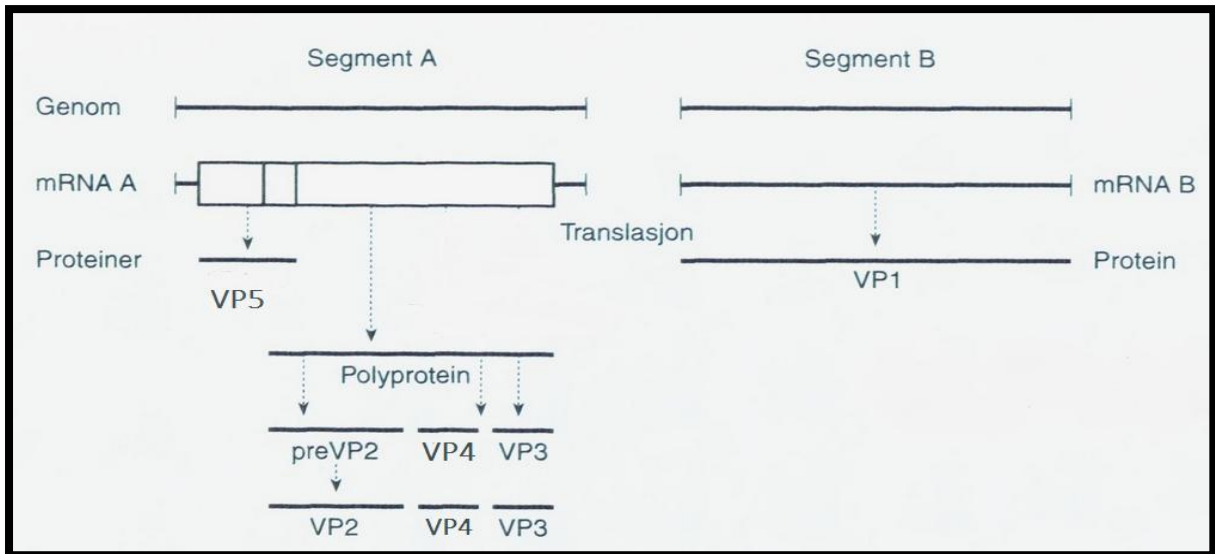


Figur 2 viser livssyklusen til villaks. Øyerogn blir klekt i ferskvann. I ferskvann er fasene yngel, Parr og smolt. Smolt gjennomgår smoltifisering som gjør at den kan vandre ut i havet. Der vokser den seg stor til en voksen laks som kan komme tilbake til ferskvann for å gyte.

1.4 IPNV

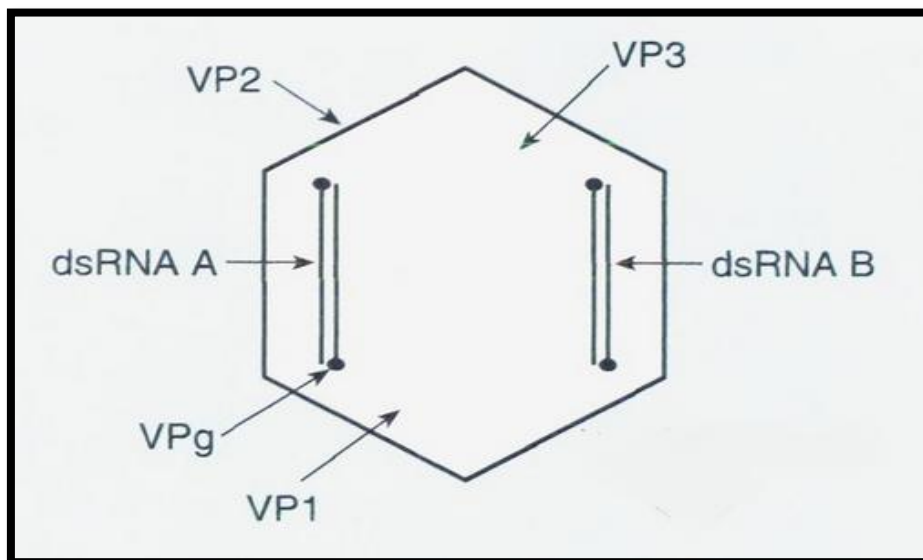
IPN viruset stammer fra slekta Aquabirnavirus som hører til familien Birnaviridae (Kelly and Loh 1972, Dobos 1995). Det finnes to serotyper av IPNV, disse heter Serogruppe A og Serogruppe B. Til dags dato er det ti ulike serotyper kjent, og 9 av disse hører til Serogruppe A (Hill and Way 1995). IPNV er et bi-segmentert dobbeltrådet RNA virus, som består av et 60 nm diameter ikosaedrisk symmetrisk kapsid (figur 4). Dette kapsidet beskytter RNA tråden med et enkelt membranløst skall. RNA trådene er dannet av segment A og segment B, som koder for fem ulike protein (VP1, VP2, VP3, VP4 og VP5) (Dobos 1995). Denne bearbeidelsen av RNA trådene, som danner proteinene, er illustrert i figur 3. Funksjonen til disse proteinene er kjent med unntatt av det ikke strukturelle argenin-rike proteinet VP5 (Heppell, Tarrab et al. 1995). VP1 er et RNA-avhengig RNA polymerase som replikerer genomet til viruset. VP2 og VP3 er strukturelle proteiner som bygger opp det ytre og det indre av kapsidet. Det siste proteinet, VP4, fungerer som en protease under RNA bearbeidelse (Dobos 1995). Lokaliteten til proteinene i kapsidet er vist i figur 4.

Figur 3: Genomet til IPN viruset



Figur 3 viser genomet til IPN viruset. Genomet består av segment A og segment B som koder for fem ulike proteiner. Segment A blir transkribert til mRNA som koder for fire protein (VP2, VP3, VP4 og VP5). Dette segmentet har to overlappende leserammer som koder for proteinet VP5 og et polyprotein (Heppell, Tarrab et al. 1995). Polyproteinet blir kløvet av proteinet VP4 som er en virale protease (Duncan, Nagy et al. 1987). Dette danner proteinene preVP2, VP4 og VP3, og deretter blir preVP2 bearbeidet av vertens protease til VP2. Segment B koder kun for en mRNA sekvens som blir translert til et RNA-avhengig RNA polymerase (VP1) (Duncan, Mason et al. 1991). Figuren er hentet fra boken "Fiskehelse og fiskesykdommer" (Bergh and Poppe 1999).

Figur 4: Kapsidet til IPN viruset



Figuren viser oppbyggingen av skallet til IPN viruset og hvor hvert enkelt protein er lokalisert. VP1 finnes i kapsidet, både som fri form og bundet form (VPg), til dsRNA A og dsRNA B. VP2 dekker den ytre siden av skallet til IPN viruset. Den indre strukturen til kapsidet er bygget opp av en "ribonucleoprotein" kjernestruktur, som kan bestå av VP3 bundet til dsRNA (Hjalmarsson, Carlalm et al. 1999). Figuren er hentet fra boken "Fiskehelse og fiskesykdommer" (Bergh and Poppe 1999).

Viruset er kjent for å ramme et mangfold av arter, både i saltvann og i ferskvann. IPNV har minst 32 ulike verter, blant annet bekkerøye, torsk, regnbueørret, kveite og laks (Bergh and Poppe 1999). I Norge er det hovedsaklig laks i fiskeoppdrett som er den meste kjente verten (Veterinærinstituttet 2013 s.13). Det har blitt gjort mye forskning på dette området, men fremdeles vet man ikke hvordan viruset infiserer verten. IPNV infeksjon forårsaker alt fra 0 til 100% mortalitet hos laks i oppdrettsanlegg (Bergh and Poppe 1999 s. 190). Infeksjonen skjer spesielt i yngelfasen i ferskvann, og i post-smoltfasen i saltvann. Individuer som overlever infeksjonen blir en smittebærer som varer livet ut. Viruset blir skilt ut i vannet og kan overleve en lengre periode (Veterinærinstituttet 2012 s.13). Symptomene for en infisert fisk er; sideveis svømming, mørkere farge, utstående øyner og utspilt buk. Infeksjonen forårsaker også indre symptomer, som indre blødninger i buken, blodfattige indre organer og en tom tarm. Årsaken er at IPN viruset angriper acinærcellene og fettvev i den eksokrine delen av bukspyttkjertelen, som forårsaker fokale nekrose (Bergh and Poppe 1999, Sandtrø 2011).

1.5 Avl av IPN resistente laks basert på familie seleksjon

Fra 2001 har Aqua Gen, i sitt avlsprogram på laks, selektert for familier som er resistente mot IPN (familie seleksjon). Denne klassiske seleksjonsprosessen har blitt basert på en smittetest for å plukke ut de dyra som er mest resistente mot IPN. Under testen blir flere søskengrupper utsatt for smitte fra viruset. Deretter registrerer man dødeligheten i de ulike søskengruppene og individer fra familiene med høyest overlevelse blir valgt som stamfisker for neste generasjon. Individuer som har vært i kontakt med IPN viruset kan ikke brukes som foreldre for neste generasjon. Med andre ord blir det et tilfeldig valg mellom ikke-testede individer fra samme familien. Denne typen seleksjon har vist seg å være effektiv (Storset, Strand et al. 2007), men man går glipp av muligheten til å velge ut de beste fiskene fra de beste familiene. Dette er ikke optimalt, fordi halvparten av den genetiske variasjonen er å finne mellom fullsøsken. Man ser derfor etter nye muligheter for å selektere for IPN resistens i laks.

1.6 Markørassistert seleksjon (MAS) for økt IPN-resistens

En Quantitative Trait Locus (QTL) er et område på genomet hvor det ligger en eller flere gener som virker inn på en egenskap. Den omtrentlige plasseringen til en QTL i genomet kan bestemmes ved hjelp av såkalt QTL-kartlegging. En identifisert QTL kan videre brukes i avl til å velge ut den beste stamfisk, ved hjelp av såkalt markør-assistert seleksjon (MAS). MAS er en metode som kan brukes til å selektere avlsdyr med de ønskede egenskapene, for å benytte disse som foreldre til neste generasjon. Metoden benytter DNA-markører som gjør den også gyldig i de tilfeller man ikke har direkte mulighet til å måle egenskapen man selekterer for. Ved bruk av metoden er man altså i stand til å teste individuelle fisker for den resistente genotypen, uten at fisken kommer i kontakt med viruset. Fisken kan derfor brukes videre som stamfisk til neste generasjon. Metoden sørger altså for at hele det genetiske potensialet blir fullt utnyttet.

Aqua Gen, CIGENE og Nofima har siden 2005 jobbet med identifisering av QTL for resistens mot IPN i laks, med formål å bruke QTL i MAS for økt IPN-resistens. Deler av disse arbeidene ble publisert i 2009 der man blant annet beskriver DNA markører koblet til en QTL for IPN resistens. Denne QTLEN forklarte 29% og 83% av de fenotypiske og genetiske variasjonene, og lokalisjonen ble identifisert til en 4 cM region på kromosom 26 (Moen, Baranski et al. 2009). Liknende resultater er også funnet i Skottland (Houston, Haley et al. 2008). Begge forskningsgruppene har konkludert med at markører koblet til QTLEN sammen med markørene, kan benyttes i MAS til å forbedre egenskapen.

Moen (2009) brukte i sitt opprinnelige "genom-scan" rundt 150 mikrosatelitter spredt utover genomet. Senere la de til enda flere mikrosatelitter i området rundt QTLEN, for å plassere QTLEN med større nøyaktighet (Moen, Baranski et al. 2009). Videre fant de ut at dersom man lager en haplotype av tre mikrosatelitt-markørene i QTL-området, vil allelene til denne haplotypen være sterkt korrelerte med graden av IPN-resistens, på tvers av familier. Dette var utgangspunktet for en DNA-test som Aqua Gen har anvendt til å bestemme genotypen til IPN-QTLEN i dyr fra sin populasjon. Testen kategoriserer dyra som qq, Qq eller QQ, avhengig av om de har 0,1 eller 2 kopier av allelet som gir høy resistens (Moen, upublisert).

I 2009 begynte Aqua Gen å bruke denne DNA-testen til å produsere lakserogn med økt resistens mot IPN. Siden den gang har andelen såkalte "QTL-rogn" vokst kraftig, og i dag selger Aqua Gen nesten bare rogn som er selektert på denne måten. Laboratorieforsøk og data fra oppdrettsnæringen har vist at QTL-rogn gir fisk med sterkt forbedret resistens mot IPN

(AquaGen 2012). I den norske fiskehelse rapporten fra 2012 ble det rapportert 119 tilfeller av IPN infeksjon i fisk i Norske oppdrett. 110 av disse tilfellene var infeksjon av laks, både i settefisk anlegg og i sjøfase anlegg. I tidligere år har det blitt diagnostisert høyere tilfeller av IPN utbrudd, fra 174 tilfeller i 2002 til hele 223 tilfeller i 2009 (Veterinærinstituttet 2013 s. 7). Deler av nedgangen fra 2009 til 2011 skyldes trolig benyttelsen av MAS i avl av laks.

1.7 Identifikasjon av enkel nukleotid polymorfisme (SNP) mellom ulike IPN-QTL genotyper

MAS er ofte begrenset til seleksjon innenfor de familiene hvor man kjenner koblingsfasen mellom markørene og den kausale DNA-variasjonen. En kausativ DNA-variasjon, eller SNP i veldig sterk LD med denne, vil kunne brukes i seleksjon på tvers av familier. På denne måten har man en bedre metode til å selektere for IPN-resistens. I tillegg er det interessant å identifisere genet som blir påvirket av denne DNA-variasjonen for å få en forståelse av den biologiske prosessen bak. For å identifisere variasjonen mellom de to genotypene, ble 22 QQ og 23 qq dyr sekvensert med ny sekvenseringsteknologi (Illumina PE (Bentley, Balasubramanian et al. 2008)). Resultatet var sekvensmengde to ganger hele genomet per dyr, som ble sammenlignet med en referansesekvens. Variasjoner mellom QQ og qq gjorde det mulig å identifisere flere enkeltbasemutasjoner (single nucleotide polymorphisms = SNPs), som for det meste (men ikke alltid) nedarves sammen med QTLen. Det vil si at disse SNPene kan brukes til å beregne genotypen til QTLen, og den antatte genotypen er vanligvis, men altså ikke alltid rett. Disse SNPene ligger innenfor et område på omtrent 30 000 basepar på kromosom 26. Aqua Gen bruker nå disse SNPene i sin DNA-test for IPN-resistens, men for å øke nøyaktigheten bruker man en haplotype av to SNPer framfor en enkelt SNP (Thomas Moen, upublisert).

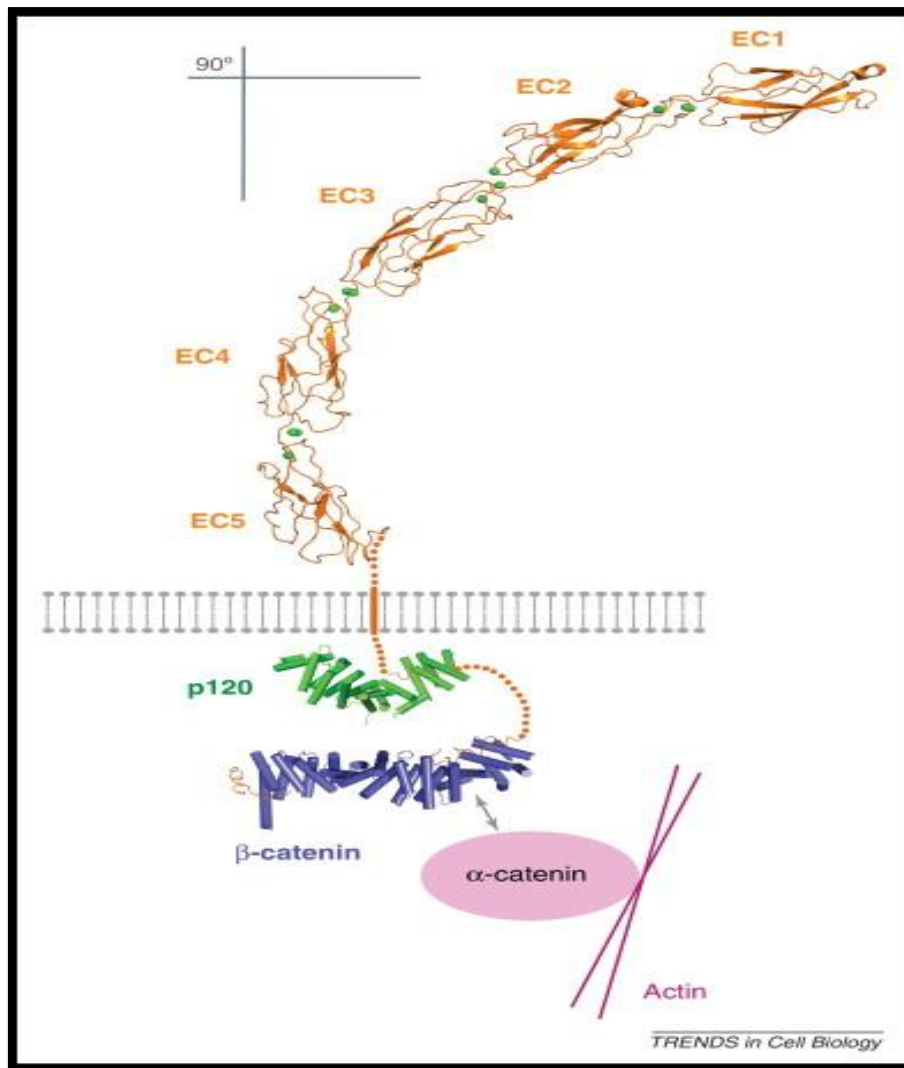
Ingen av de identifiserte SNPene gir en endring i et protein, og ingen av dem har en åpenbar funksjonell betydning, og som nevnt nedarves ingen av dem etter helt samme mønster som QTLen. Allikevel er de en viktig ledetråd i søket etter den kausale mutasjonen, fordi de peker mot et lite område på genomet. Dette området inneholder to gener, Epitelial(E)-cadherin og FAM96b. Det har foreløpig ikke lyktes forskerne i Aqua Gen og CIGENE å finne polymorfismer innenfor dette område som er enda sterkere korrelert med QTLen (Thomas Moen, upublisert).

1.8 E-cadherin

E-cadherin er en del av en stor superfamilie, kalt cadherin, med over 350 medlemmer (Hulpiau and van Roy 2009). Den består av glycoproteiner, som fungerer som celleoverflate reseptorer. Her bidrar de først og fremst i kalsium-avhengig celle-celle gjenkjennelser (Takeichi 1988) og til å binde cellene sammen. Dette påvirker hvordan cellene beveger seg og kommuniserer med hverandre. I tillegg kan de ha innvirkning på hvordan cellen gjennomgår morfogenetiske endringer (Gumbiner 2005, Brasch, Harrison et al. 2012, Saito, Tucker et al. 2012).

Superfamilien er delt inn i flere undergrupper. Noen av disse gruppene er klassiske cadherin, desmosomal cadherin og protocadherin. Denne inndelingen er basert på sekvenslikheter, domene arrangering og antall ekstracellulær cadherin (EC) domene. EC domene kan variere i antall domener og struktur. Klassisk Cadherin i vertebrater har fem EC domener og en enkel cytoplasmisk domene. Denne er nærmere beskrevet i figur 5. Klassisk Cadherin er delt inn i gruppene Type 1 og Type 2 (Brasch, Harrison et al. 2012). Type 1 består blant annet av E-cadherin, N-cadherin og P-cadherin (Saito, Tucker et al. 2012).

Figur 5: Klassisk Cadherin

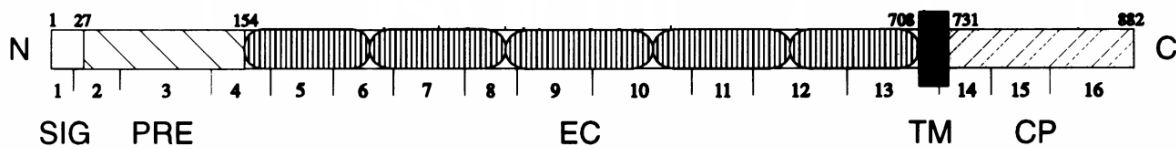


Figuren viser strukturen til klassisk cadherin. Cadherin består av tre hoveddeler; Ectodomene, transmembran domene og cytoplasmisk domene. I N-terminalen er ectodomene som består av fem EC-domener, som blir bundet av Ca²⁺ (grønne prikker). Disse stabiliserer EC-domenet, som får en vinkel på 90 grader. Noe av strukturen er ukjent, dette er vist som en prikkete linje. Det konserverte cytoplasmiske domenet binder til p120 og beta-catenin. For å bli linket til cytoskjeletet bindes Beta-catenin til alfa-catenin. Figuren er hentet fra: (Brasch, Harrison et al. 2012).

E-cadherin regner for å være et av de mest kjente proteinene i gruppen cadherin (Alberts B 2002). Det er et transemembran glycoprotein. Den er blant annet uttrykt i epitelcellevev, som danner og opprettholder strukturen av vevet. Dette skjer ved at det blir dannet en stor klynge av E-cadherin på celleoverflaten. Her vil E-cadherin danne en binding til nabo reseptoren på samme celle samtidig som den binder til en reseptor på motsatt celle. Dette styrker de intracellulære bindingene mellom cellene (Ozaki, Obata et al. 2010).

Forskning i menneske og andre vertebrater har gitt oss mye informasjon om E-cadherin, deriblant hvordan dens struktur er bygget opp. Før E-cadherin er et aktivt protein har den en aminosyresekvens rundt 850 aminosyrer (Figur 6). For å danne et aktivt protein blir et domene på rundt 129 aa på aminoenden kløyyet av (Ozawa and Kemler 1990). Proteinet er på rundt 700 aminosyrer og består av et ekstracellulært domene, transmembran domene og et cytoplasmisk domene. Det cytoplasmiske domenet er veldig konserververt og har en lengde på rundt 150 aminosyrer på den ”carboxy” terminalen (Nagafuchi and Takeichi 1988). Det ekstracellulære domenet har fem EC-domenene på 110 aminosyrer hver, disse finnes på aminoenden av sekvensen. EC-domenet blir stabilisert ved binding av CA^{2+} til linkene mellom EC-domenene (Brasch, Harrison et al. 2012).

Figur 6: E-Cadherin



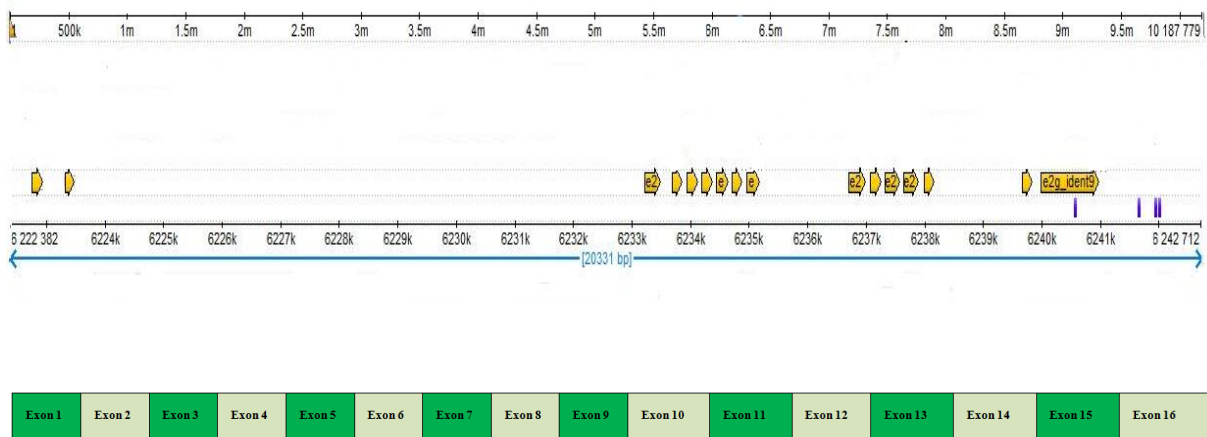
Figuren viser strukturen til E-cadherin. SIG=signal peptid. PRE= ende som blir kløyyet av. EC= Det ekstracellulære domenet med fem EC-domenene. TM=transmembran domenet. CP= Det cytoplasmiske domenet. Figuren er hentet fra (Berx, Cleton-Jansen et al. 1995).

I mennesket er tap av funksjon mutasjon i dette genet korrelert med ulike typer kreft. Det skjer ofte mutasjoner i exon 7, 8 og 9 som tilsvarer domenenene EC2 og EC3 (Endo, Ashida et al. 2001). Det er også kjent at mutasjoner i konsensus splise setet kan skape ulike typer mRNA som kan skape sykdommer (Li, Gao et al. 2013). Ved infeksjoner i pattedyr er E-cadherin kjent for å bli brukt av patogener. Bakterier bruker domenet på den ekstracellulære siden til å feste seg til celleoverflaten, og deretter starter en ”endocytosis” prosess for å komme inn i cellen (Mengaud, Ohayon et al. 1996, Pizarro-Cerdá, Kühbacher et al. 2012).

I genomet til laks eksisterer det to kopier av E-cadherin med full lengde. Disse er lokalisert til kromosom 11 og kromosom 26. Begge genene består av 16 exoner og er større en 3500 nukleotider (Figur 7). Kromosom 11 og 26 inneholder ”homeologe” kromosomområder som

oppstod fra samme kromosom etter den siste dupliseringen av laksegenomet. Det er fortsatt ukjent om begge genene er uttrykt i levra til laks, men det er kjent at proteinet er uttrykt i epitelieavv til mange andre arter, som menneske og zebrafisk. Hypotesen er at reseptoren E-cadherin på kromosom 26 blir brukt av IPN viruset i infeksjonsprosessen av laks. Variasjonen i sekvensen i genotype QQ kan forårsake en endring i reseptoren som gjør at IPN viruset ikke kan bruke den. Som en konsekvens kan ikke viruset komme seg inn i cellene for å få replikert genomet.

Figur 7: Genet E-cadherin på kromosom 26 og dens mRNA sekvens



Figuren viser en oversikt over de 16 ulike exon i genet E-cadherin på kromosom 26 som er ”alignet” mot genomsekvensen til laks. I tillegg er det en oversikt over mRNA sekvensen med exon spleiset sammen. Dette er en oversikt tatt fra programmet UGENE.

1.9 Mulige kilder til variasjon

Dersom E-cadherin er det genet som forårsaker forskjellen mellom de to allelene Q og q i QTLen, er det mulig at dette skjer ved at det finnes to varianter av genet med ulike aminosyresekvenser. I så fall finnes det en eller flere polymorfismer (for eksempel SNPer) i genet som gir endringen av aminosyresekvensen. Men det kan også finnes polymorfismer som gir andre typer endringer i genet. For eksempel kan det dannes alternative mRNA ved bruk av alternative promotorer, alternativ spleising og/eller polyadenylering. Alternative prosesser kan oppstå ved dannelsen av en mutasjon i et sensitivt område som vil gi et anerledes sluttresultat. Dette kan påvirke strukturen og funksjonen til proteinet. Ved transkripsjon av DNA kan det benyttes alternative promotorer som kan ligge utenfor eller i det tidligere transkriberte

området. Dette vil gi et mRNA produkt med ulik lengde som også vil inneholde ulike exoner. Alternativ polyadenylering er en annen bearbeidelses prosess som gir ulike ender på mRNAet. Når transkripsjonen er ferdig vil mRNAet bli kuttet ved polyadenyleringssetet og deretter få festet på en poly(A) hale. Et mRNA produkt kan inneholde flere slike seter, som kan gi variasjon i 3' ende av mRNAet. I noen sjeldne tilfeller kan også setet være i exon området, noe som vil påvirke det endelige protein produktet. Den siste av de tre nevnte kildene for variasjon i mRNA er alternativ spleising. Ved normal spleising blir alle intron mellom exon fjernet og alle exon blir spleiset sammen i samme rekkefølge. Denne prosessen blir gjennomført av spleisosomet som er et stort RNA-protein-kompleks. Alternativ spleising er en prosess som forårsaker ulike protein isoformer fra et enkelt gen. Dette ble oppdaget på 1980-tallet, og ga en forklaring til hvorfor man fant flere proteiner i proteomen enn eksisterende gen i genomet. Når et gen består av flere exon kan et eller flere exon bli eliminert fra det endelige produktet (Lodish 2008). I tillegg kan en mutasjon i konsensus spleisesete forårsake exon delesjon i mRNA fordi dette punktet på sekvensen er veldig sensitiv til endringer (Li, Gao et al. 2013). Ved å undersøke mRNA sekvensen til E-cadherin kan en bekrefte om det eksisterer variasjoner mellom de ulike genotypene.

1.10 Mål for oppgaven

- Undersøke om E-cadherin på kromosom 11 og kromosom 26 er uttrykt i leveren til laks.
- Finne sekvensene til "messenger" RNA (mRNA) av begge versjonene av E-cadherin.
- Bruke disse sekvensene til å undersøke om det eksisterer variasjon som SNP, exon insersjoner og delesjoner og spleisevarianter mellom genotypene QQ, Qq og qq til E-cadherin på Kromosom 26.
- Måle ekspresjonsnivået av E-cadherin (qPCR) på kromosom 11 og kromosom 26 ved prøver tatt fra
 - individ med genotype QQ og genotype qq
 - IPN smittet og ikke-smittede individer

2. MATERIAL OG METODER

For å analysere mRNA sekvensen og ekspresjonsnivået til genet E-cadherin på kromosom 26 og dens paraloge gen på kromosom 11 ble det i denne oppgaven brukt disse metodene; polymerase kjedereaksjon(PCR), sekvensering og Realt-time PCR (qPCR). Her ble det benyttet individer som hadde ulik genotype for QTLen for IPN-resistens. I første del blir det gjennomgått ulike forberedelser for de laboratoriske metodene. I den andre delen blir disse laboratoriske metodene beskrevet og det blir forklart hvordan hver metode bidrar til å besvare målene for oppgaven.

2.1 Dannelse av en referansesekvens av cDNA til E-cadherin og dens paraloge gen.

En referansesekvens kan brukes som grunnlag når en designer primere og skal analysere et eventuelt resultat. Derfor var det viktig å lage en referansesekvens som dekker hele genet. En slik cDNA sekvens var ikke tilgjengelig for E-cadherin i laks. Dersom primere blir basert på en dårlig referansesekvens kan dette lede til at primerne ikke binder til cDNA produktet, selv om det er tilstede i genomet. I tillegg vil dette gjøre det vanskeligere å analysere resultatet. For å danne en referansesekvens med god kvalitet ble det brukt ulike dataprogrammer. Denne prosessen blir beskrevet under, samtidig som det blir forklart hvorfor man bruker akkurat disse programmene.

Sekvens material

Dannelsen av en referansesekvens ble basert på flere cDNA sekvenser og en genomsekvens. Genomsekvensen ble utviklet av CIGENE (upublisert), mens cDNA sekvensene ble dannet i et upublisert arbeid ledet av Ben Koop. I dette arbeidet ble det benyttet ca. 500 000 EST som er tilgjengelig i databasen Genbank. EST sekvensene ble slått sammen for å danne flere contig av cDNA sekvenser fra laks. Et contig er et sett av overlappende DNA sekvenser som overlapper på et bestemt område på nukleotidsekvensen. Noen av contig bestod blant annet av cDNA sekvenser for E-cadherin og dens paralog. Hver contig dannet en konsensussekvens som kunne inneholde alt fra en liten del av et exon til mange exoner. Enkelte av delene i de ulike konsensussekvensene kunne i tillegg overlappe hverandre. I denne oppgaven ble det

benyttet konsensussekvenser basert på contig 34030 for cDNA sekvensen til E-cadherin på kromosom 26. For paralogen på kromosom 11 ble det benyttet to konsensussekvenser basert på contig 20606 og contig 6256. Dannelsen av disse referansesekvensene blir beskrevet under.

est2genom

Ved dannelse av en referansesekvens var det viktig å definere exon-intron overganger grensene. Primerepar som skulle binde til cDNA sekvensen til E-cadherin, under metodene PCR og sekvensering, kunne ikke binde i overgangen mellom exonene. I metoden qPCR var det derimot viktig at primerparet festet seg i overgangen mellom exonene. Det var derfor nødvendig å finne ut nøyaktig hvor overgangene mellom exonene i cDNA-sekvensen var. Programmet est2genom ble brukt til å definere exon-intron overgangene i genomsekvensen. Dette programmet er et gen prediksjonsprogram som er basert på sekvens homologi mellom en cDNA-sekvens og en ikke spleiset genom sekvens. Programmet bruker en modifisert "Smith-Waterman" algoritme for å detektere like områder mellom sekvensene (Mott 1997). est2genom "aligner" cDNA-sekvenser fra laks opp mot en genomsekvens. cDNA-sekvensene som passet til genomsekvensen var først blitt plukket ut fra et større sett med cDNA-sekvenser, ved hjelp av BLAST. Denne "alignment" var basert på "default" verdier. Resultatet ble en oversikt over alle exon for genet E-cadherin og dens paraloge gen i genomsekvensen. Oversikten ble tilslutt benyttet til å finne overgangene mellom exonene i mRNA-sekvensen.

UGENE

For å visualisere resultatet framstilt i est2genom, ble multiplattform programmet UGENE benyttet. UGENE er et program som forenkler arbeidet for forskere gjennom å samle diverse programmer innenfor bioinformatikk for bearbeiding, analyse og visualisering av informasjon. Programmet gjør det mulig å visualisere annoterte genomsekvenser, multiple sekvens "alignments" og diverse andre biologiske objekter (Okonechnikov, Golosova et al. 2012). UGENE er derfor et godt egnet verktøy for å visualisere "alignement" mellom genomsekvensen og cDNA sekvensene. Ved å kopiere cDNA sekvensen for hvert exon og søke etter exonet i de ulike konsensussekvensene, er man i stand til å definere overgangen mellom exonene. Ved å søke etter exonene ble det også mulig å finne konsensussekvenser som

inneholdt alle exonene. Det ble funnet en konsensussekvens for genet E-cadherin på kromosom 26 som inneholdt alle exonene som er basert på contig 34030. Derimot ble det ikke funnet en konsensussekvens som representerte hele cDNA-sekvensen til det paraloge genet på kromosom 11. I dette tilfellet ble det benyttet konsensussekvensen til contig 20606 og contig 6256, som representerte begynnelsen og slutten på cDNA sekvensen. I tillegg ble det benyttet deler av genomsekvensen for å fylle ut et område på rundt 600 nuklotider som manglet i referansesekvensen.

ExPasy Translate

Etter dannelsen av referansesekvensen (mRNA) var det behov for å kontrollere at den var satt sammen på riktig måte. Dette kunne kontrolleres ved å translaterer cDNA sekvensen til en aminosyresekvens for å se om den inneholdt minst en lang åpen leseramme (ORF). Et program som gjør akkurat dette er ExPasy Translate (Gasteiger, Gattiker et al. 2003). Programmet oversetter cDNA sekvensen til alle seks leserammer som gjør det mulig å lete etter en lang sekvens uten stopp kodon.

Basic Local Alignment Search Tool (BLAST)

En siste kontroll av referansesekvensen ble gjort for kontrollere likheten mellom sekvensen og andre cDNA sekvenser av E-cadherin i "nucleotide collection" (nr/nt) databasen. I denne prosessen ble det benyttet programmet "Basic Local Alignment Search Tool" (BLAST). Dette programmet er et lokalt "alignment" program som sammenligner nukleotid og aminosyre sekvenser opp mot en sekvensdatabase eller egne sekvenser. Søket er basert på en algoritme som går hurtigere enn andre "alignment" programmer, samtidig som den opprettholder sensitiviteten i søket. (Først lager den en oversikt med ordet som består av alle mulige sekvensmønster på 11 nukleotider. Når den har funnet sekvenser med stor signifikans, vil den starte "alignment" prosessen på begge sider av sekvensen (Altschul, Gish et al. 1990, Mount 2007)).

2.2 Primerdesign:

Referansesekvensen ble videre brukt som grunnlag for design av primere som skulle binde til mRNA for E-cadherin på kromosom 26 eller dens paralog på kromosom 11. I første del av dette kapittelet beskrives prosessen rundt dannelsen av primerparene i programmet Primer3. Det var behov for to ulike typer primere som skulle fungere i ulike metoder. Den ene primer typen skulle fungere under PCR og sekvensering, mens den andre primer typen skulle fungere under qPCR. Under design av primerparene i laks er det i tillegg viktig å ta hensyn til sekvenslikheten mellom de paraloge genene. Her er det viktig å velge primerpar som binder til det ene av de to paraloge genene, men ikke til begge. Derfor ble det laget nye programmer som gjør det lettere å ta hensyn til dette kriteriet. Disse programmene blir forklart i siste del av dette kapittelet.

Primer3

For å kunne analysere resultatet fra PCR, sekvensering og qPCR trenger man et produkt med høy konsentrasjon og god spesifisitet. For å produsere et bra produkt trenger man primere som fungerer optimalt under reaksjonen. Det er derfor viktig å designe primere etter forholdene til den spesifikke reaksjonen. Ved benyttelse av programmet Primer3 ved kommandolinjen kan man spesifisere akkurat disse verdiene. Primer3 er et program som designer og analyserer primere basert på "default" - og egenspesifiserte verdier. Disse egenspesifiserte verdiene kunne man bestemme før programmet ble kjørt (Untergasser, Cutcutache et al. 2012). Verdiene kan inkludere bindingsområde, primerlengde, produktlengde etc. De fleste verdier ble kjørt i en "default" modus, dvs standerverdier som er spesifisert i manualen til primer3. Her har smeltetemperaturen(T_m) blitt basert på "thermodynamic approach" (http://primer3.wi.mit.edu/primer3web_help.htm). De egenspesifiserte verdiene varierte ut fra hvilken metode de skulle benyttes i. Disse blir nå beskrevet hver for seg som Primersett 1 og Primersett 2.

Primersett 1 ble brukt i både PCR og sekvensering for å finne eventuelle forskjeller mellom de to genotypene QQ og qq. Dette kunne blant annet være variasjoner som alternativ spleising. Bindingsområdet til primerne ble derfor spesifisert, slik at primerne ikke festet seg ved eller i overgangene mellom exonene. I tillegg var primerlengden satt mellom 18 bp til 22 bp og produktlengde varierte mellom 500bp til 1000 bp.

Primersett 2 var designet for å finne forskjell i konsentrasjonen mellom mRNA produktet fra genotypene QQ og qq. I motsetning til Primersett 1 var det viktig at primerne festet seg til overgangene mellom exonene. I qPCR med CyberGreen er det ikke mulig å kontrollere om primerne har festet seg til flere produkter, som man kan gjøres ved PCR ved benyttelsen av gel elektroforese. RNA prøvene kan være forurenset av genomisk DNA som kan påvirke resultatet. For å være sikker på at ikke genomisk DNA ble registrert, måtte primerne binde i området i mRNA som var forskjellig fra det genomiske DNA. Andre forhold som ble spesifisert var primerlengde, som skulle være mellom 18bp og 22bp, og produktlengde som kunne variere mellom 400bp til 500bp.

Sluttresultatet fra Primer3 for hvert primersett var flere lister på 1000 mulige primerpar. Hver liste representerte et bindingsområde i cDNA-sekvensen (til E-cadherin på kromosom 26 eller dens paralog på kromosom 11) og hver cDNA-sekvens hadde flere bindingsområder. For å velge det best tilpassede primerparet innenfor hvert bindingsområde måtte man spesifisere forskjeller mellom mRNA sekvensen til E-cadherin og dens paralog.

”Mismatch” program

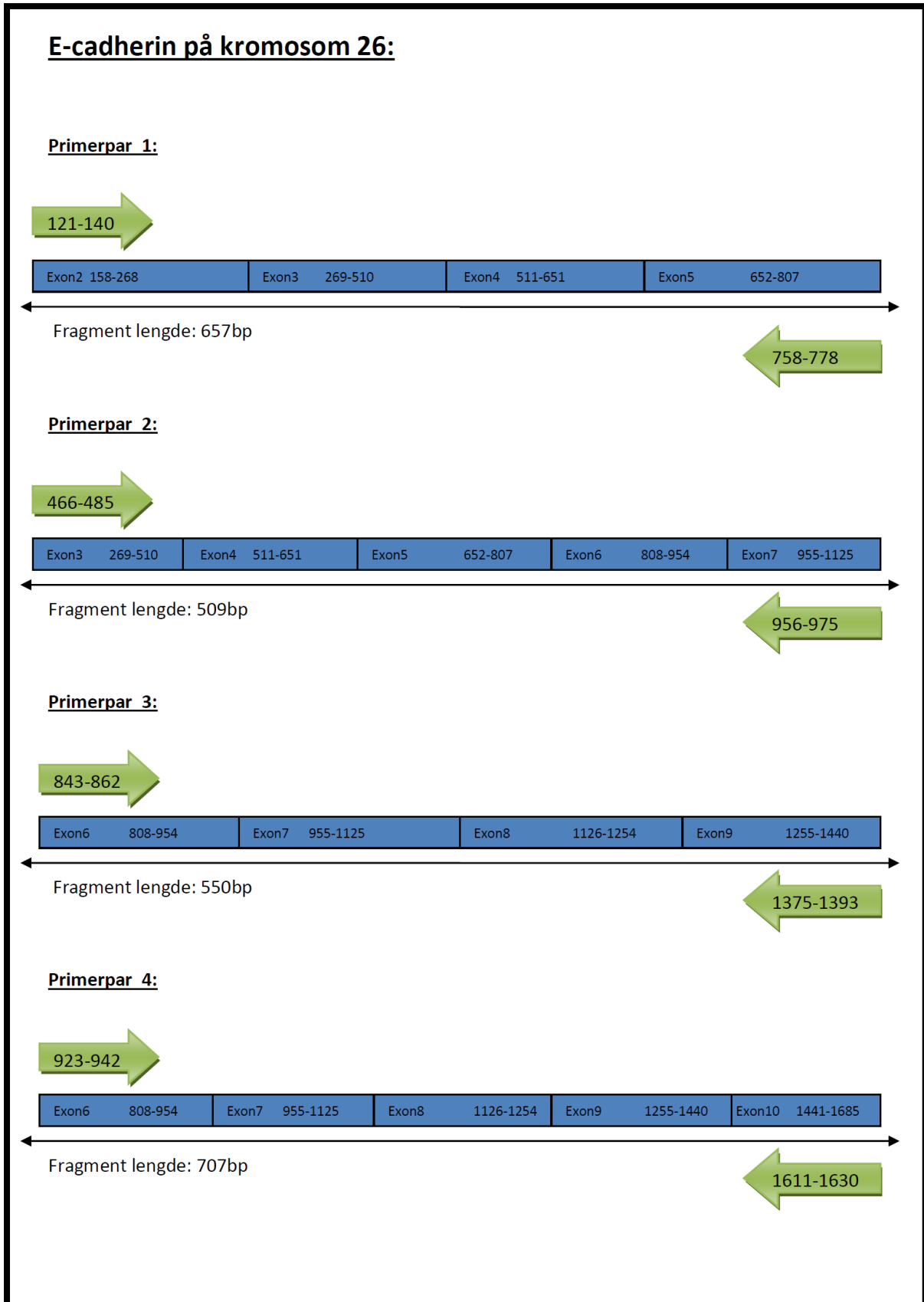
Et viktig kriterium ved designet av primere er at den kun binder seg til ett bestemt cDNA produkt. Den store likheten mellom de to paraloge E-cadherin genene kan føre til at primerparet binder seg til begge gen under PCR. For å finne primerpar som bare bindes til cDNA fra ett gen ble det brukt et Python-program av Thomas Moen (upublisert). Dette programmet kjøres ved kommandolinjen og det bruker BLAST til å ”aligne” referansesekvensen til de to genene. Programmet lager en ”mismatch” fil som viser antall ”mismatches” mellom de to genene for alle mulige primere med lengde mellom 17bp til 30bp innenfor den overlappende sekvensen. Hver base i den overlappende sekvensen vil altså fungere som et startpunkt for primere og hver base har fjorten primere som varierer i lengde. Denne listen kan kombineres med resultatet fra Primer3 for å velge ut de beste primerparene.

Sammenligning av ”mismatch” fil og Primer3 fil

For å finne det mest optimale primerparet fra listen på 1000 primerpar ble den kombinert med ”Mismatch” filen. For å kombinere disse listene ble det brukt et annet program. (Thomas Moen. upublisert). Dette python baserte programmet kjøres ved kommandolinjen for å

kombinere disse filene. Resultatet blir en liste med 1000 primerpar som også viste antall "mismatch" hver primer hadde. For å finne det mest optimale primerparet må det tilfredsstille et minimumskrav på minst to ulike baser mellom primeren og den paraloge sekvensen. I tillegg måtte primerparet ha en lav samlet "penalty" verdi. Dette er en verdi som oppsummerer hvor bra primersettet er, tatt i betraktning til viktige parametere, deriblant smeltetemperaturen. GC-prosenten beskriver sannsynligheten for at et primermolekyl skal "anneales" til et annet osv. Den endelige listen over primerpar finnes i tabell 2 og 3 i appendiks, og viser antall "mismatch" og den samlede "penalty" verdien. Figur 8 viser det hvor primerparene binder på E-cadherin og dens paraloge gen.

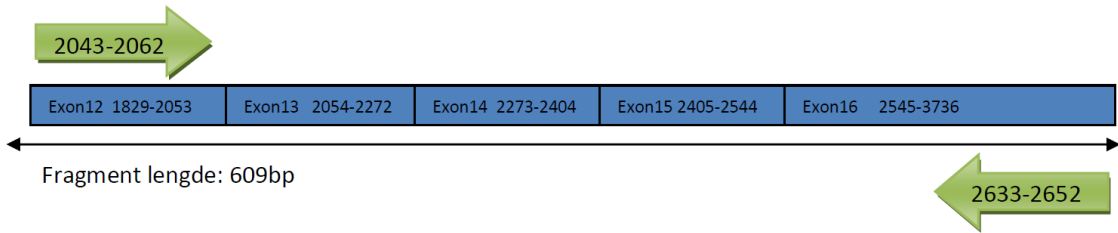
Figur 8: Primerpar i Primerset 1



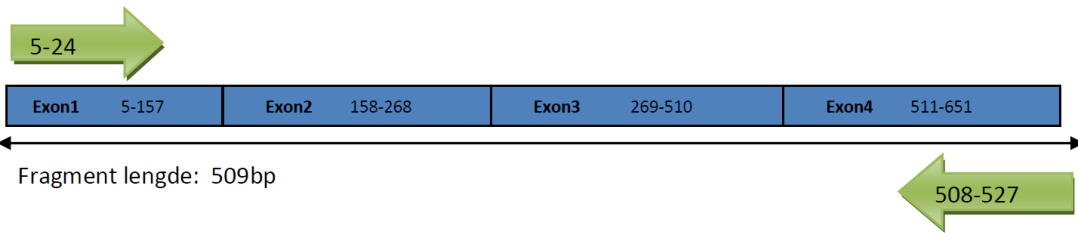
Primerpar 5:



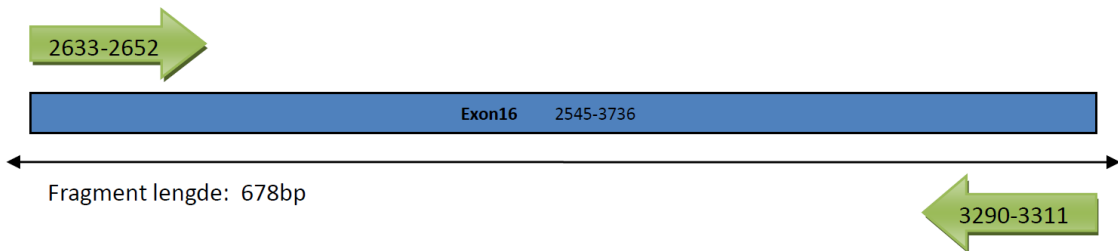
Primerpar 6:



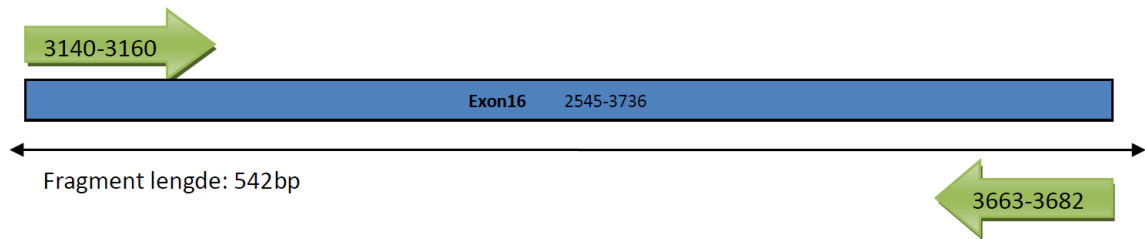
Primerpar 7:



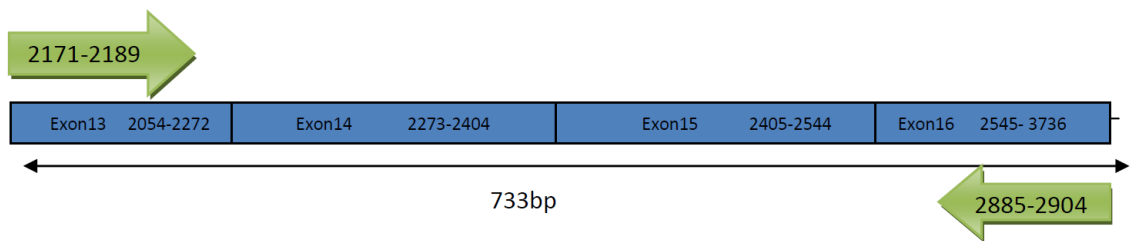
Primerpar 8:



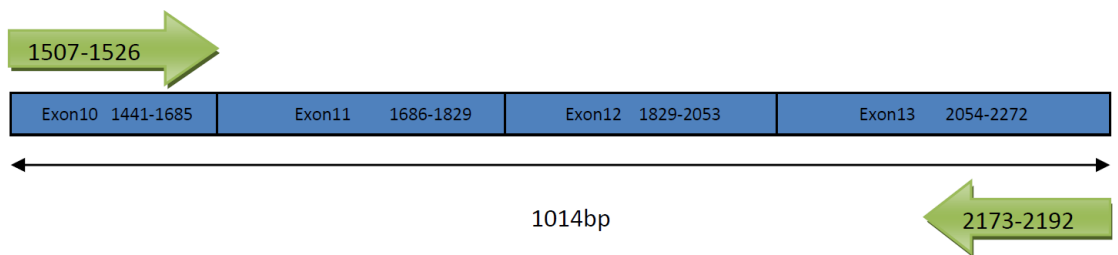
Primerpar 9:



Primerpar 10:

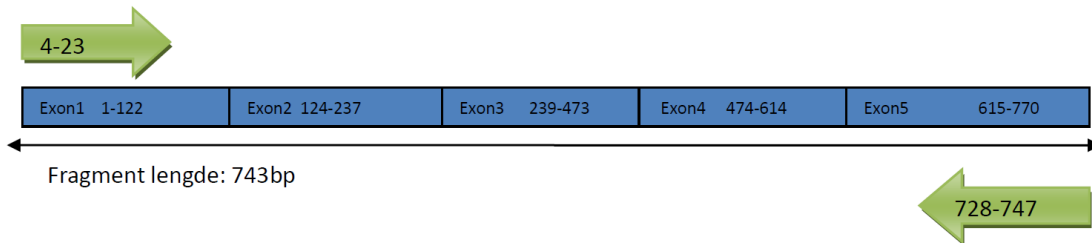


Primerpar 11:

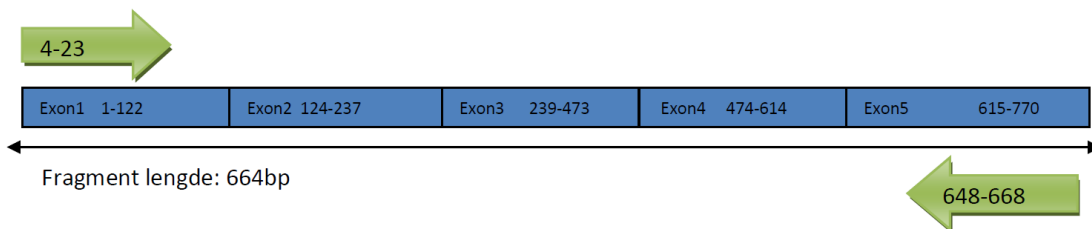


E-cadherin på kromosom 11:

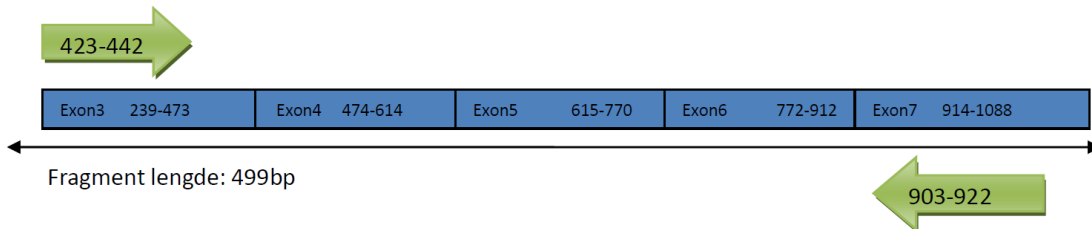
Primerpar 12:



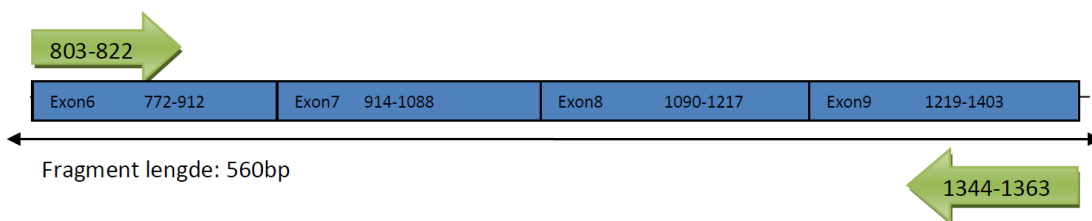
Primerpar 13:



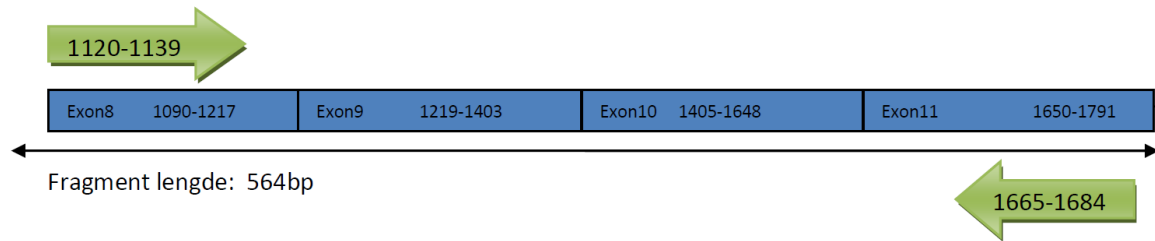
Primerpar 14:



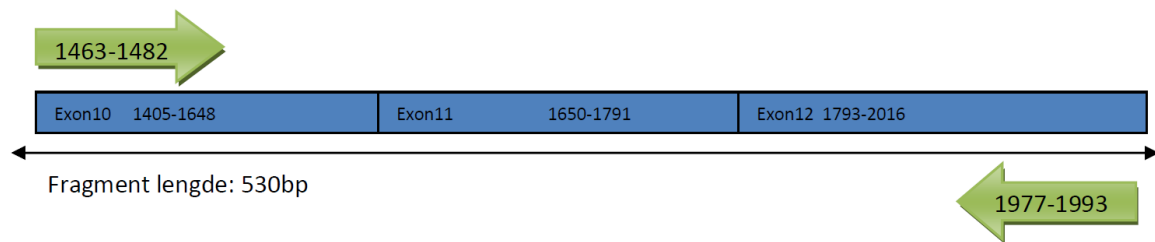
Primerpar 15:



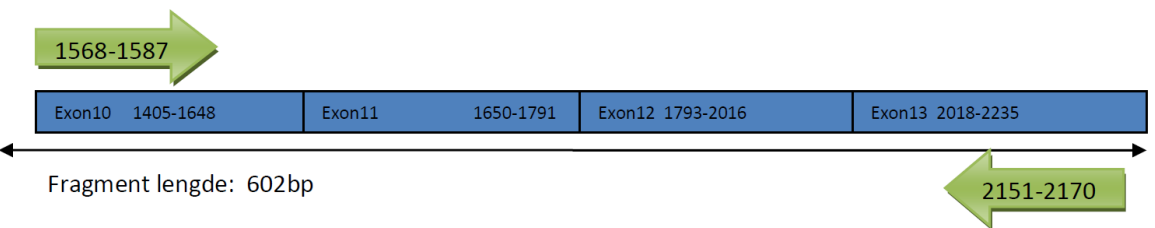
Primerpar 16:



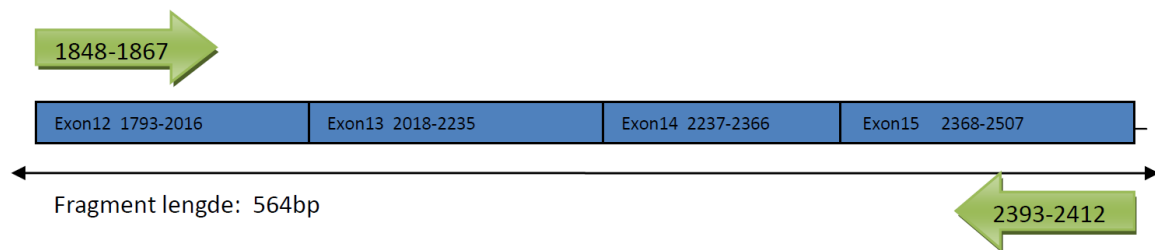
Primerpar 17:

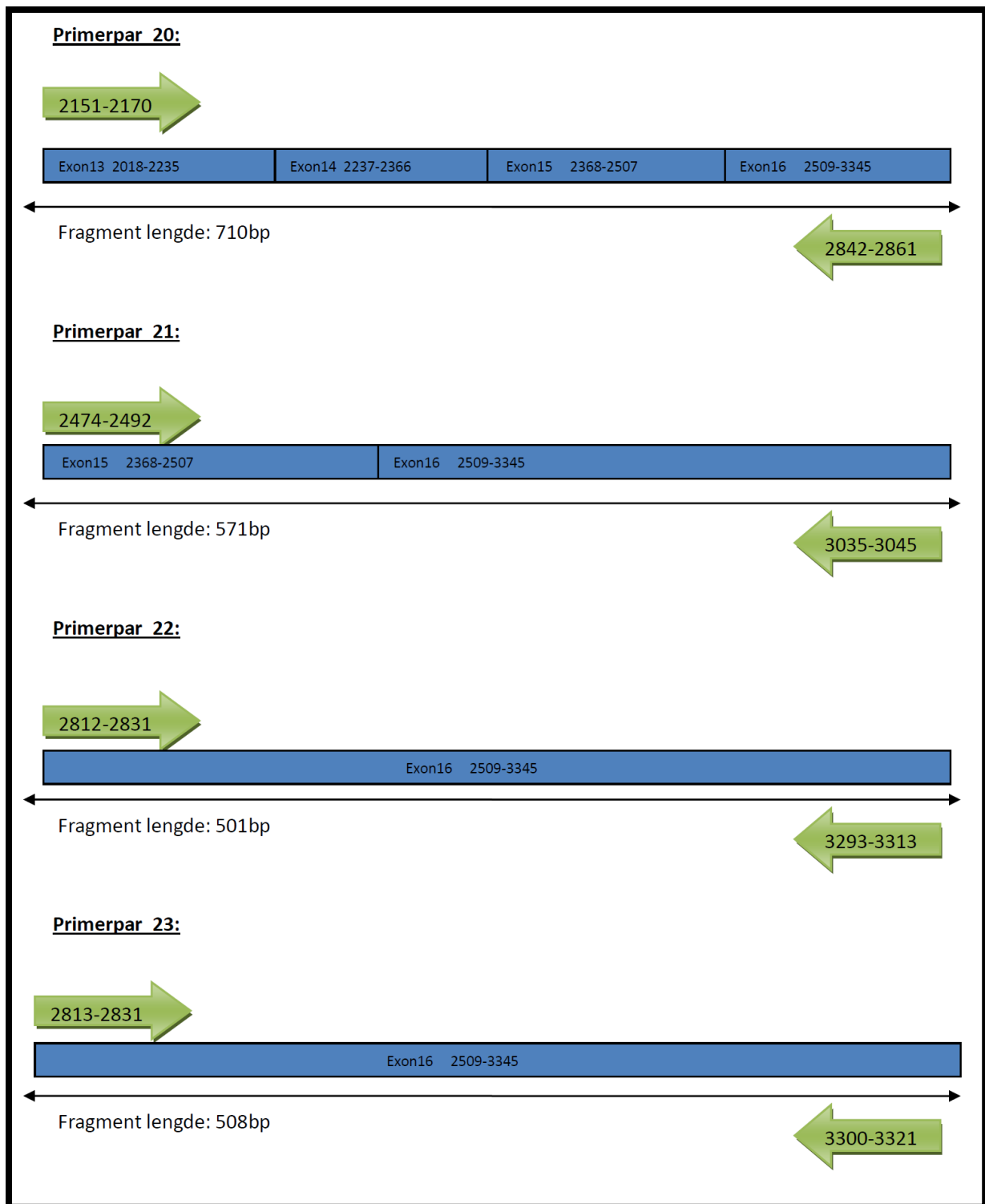


Primerpar 18:



Primerpar 19:





Figuren viser primerpar i Primersett 1. Hver enkel primer er vist med en grønn pil og området som den binder til er beskrevet i basepar (bp). Området som danner PCR produktet er vist i blå, som også beskriver hvilke exon som er del av dette produktet. Hele lengden av exonet er vist i den blå delen (Beskrevet i basepar). Fragmentlengden tilsvare lengden på PCR produktet.

2.3 Laboratoriemetoder

I dette avsnittet gis en oversikt over bioteknologiske metoder som er brukt i laboratoriet i dette eksperimentet. I første del beskrives hele prosessen fra nedbryting av vev til dannelsen av cDNA. cDNA blir videre brukt i kvalitative metoder for å avgjøre hvorvidt E-cadherin og dens paraloge gen er uttrykt i leveren til laks og hvorvidt det finnes sekvensvariasjon mellom de dyr som har de to ulike homozygote genotypene for QTLen for IPN-resistens. I siste del blir det beskrevet et mindre kvantitativt eksperiment basert på cDNA prøvene. Dette kan vise om genet E-cadherin og dens paraloge gen er høyere eller lavere uttrykt i individer med genotype QQ i forhold til individer med genotype qq. Effekten av IPN smitte ble også undersøkt, herunder om smitten forandrer ekspresjonsnivået av E-cadherin og dens paraloge gen.

Forberedelse til PCR, sekvensering og qPCR

Dyremateriale

Eksperimentet ble basert på prøver tatt fra leveren hos laks i enten voksen eller yngel fase. Fiskene ble delt opp i grupper basert på deres genotype (QQ, Qq og qq). Yngelen kom fra Aqua Gen sin stamme. Denne stammen er en oppdrettspopulasjon med opphav i fisk samlet fra 41 elever i Norge tidlig på 1970-tallet. Yngelen har gjennomgått et smitteforsøk hos Havbruksstasjonen i Tromsø i 2010. Smitteforsøket begynte med en startforing, seks dager før fisken ble utsatt for IPN gjennom metoden Badsmitte. Det ble samlet prøver fra smitteforsøket både før og etter smitte med IPN-viruset. Det ble også brukt prøver fra laks i voksen fase for å dekke ulike faser i syklusen. Disse voksne individene ble hentet fra en slaktetest hos Aqua Gen. Slaktetesten ble gjennomført når laksen var tilnærmet 4 kg, som er normal standardvekt for slakting. Laksen som ble brukt i denne oppgaven hadde alle fått bestemt en genotype for QTLen for IPN-resistens. Denne genotypen hadde laksen fått ved hjelp av DNA-testen som Aqua Gen bruker når de selekterer fisk for økt IPN-resistens. Testen bruker SNPer lokalisert i nærheten av E-cadherin genet, i tillegg til andre SNPer. Denne testen bestemmer genotypen til QTLen med stor sikkerhet, men 100 % sikker er den ikke.

RNA ekstrahering

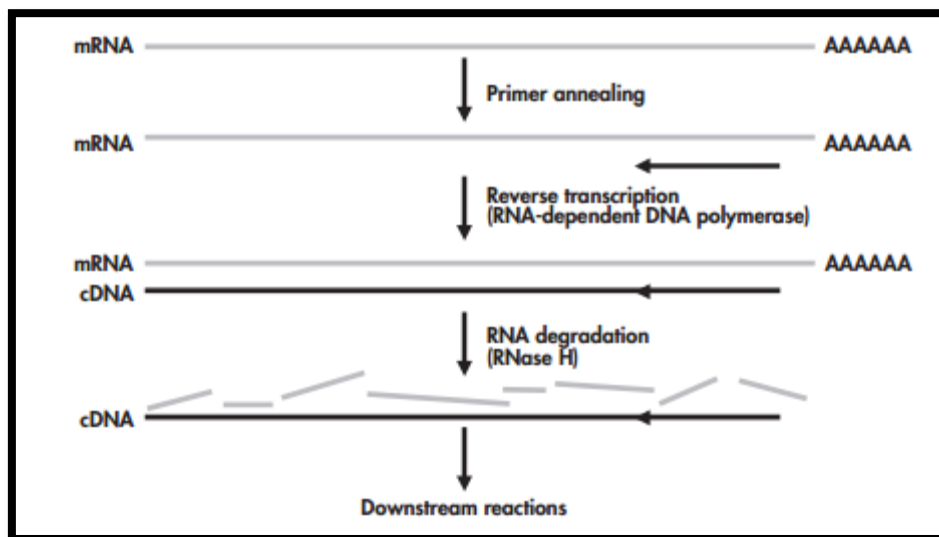
Prøvene tatt fra leveren til laksen ble brukt til å danne høy kvalitets RNA prøver som kunne brukes i PCR og qPCR. For å rense RNA fra 10mg vevsprøvene ble det benyttet RNeasy Mini Kit (Qiagen,Hilden,Tyskland). Denne prosessen ble fulgt i henhold til RNeasy Mini Handbook (Qiagen 2006) for rensing av total RNA dyre vev side 39. Prosessen består av tre hoveddeler. Først lysering og homogenisering, deretter fjerning av genomisk DNA og tilslutt rensing. Vevet ble lysert og homogenisert i en Tissuelyser utfra protokollen på side 16 i TissueLyser Handbook (Qiagen 2010). Her ble vevet lagt i en 350 ul guanidine-thiocyanate (RLT) løsning i en 2 ml microsentrifuge tube. Tuben inneholdt en liten metallkule som kolliderte med vevet under ristingen for å løse opp vevet. Her ble cellemembranen til cellen knust, og RNA og andre biologiske molekyler ble frigjort. Store proteiner og karbohydrater ble kuttet opp i biter som dannet en homogenisert væske. Dette er en viktig prosess fordi den bidrar til å fjerne RNase som kan skade det frie RNA i løsningen. Supernatanten fra RNA prøven ble deretter blandet med 50% ethanol som hjalp RNA i bindingen til en silica-basert membran filter i "spin column". Etter at RNA hadde festet seg til denne membranen ble DNase, fra DNaseFree DNase "kit" (Qiagen, Hilden,Tyskland), tilsatt for å fjerne alt genomisk DNA. Det er nødvendig å fjerne genomisk DNA fordi det kan medføre kontaminering og et falskt resultat under PCR, qPCR og sekvensering. Denne prosessen følger protokoll på side 69 i appendiks D i RNeasy Mini Handbook (Qiagen 2006). Under det siste steget, vasking, blir alle andre partikler enn RNA skylt bort. Etter tre vaskesteg ble RNA fortynnet i 50 ul RNase fritt vann. Denne fortynnede RNA prøven ble målt med Nanodrop, for å kontrollere at RNA ekstraheringer hadde fungert.

Revers Transkripsjon av RNA

RNA prøvene ble videre brukt til å danne cDNA som videre kan brukes i metoder som PCR og qPCR. Teknikken som ble brukt kalles Revers transkripsjon, og det ble benyttet QuantiTect® Revers Transkripsjons Kit (Qiagen,Hilden,Tyskland). Metoden ble fulgt i henhold til protokoll på side 12 i Quantitect Revers Transcription Handbook (Qiagen 2005). Denne reaksjonen er avhengig av et enzym, kalt Revers Transkriptase, som fester seg til et RNA templat og danner flere kopier av cDNA. Revers transkriptase er et multifunksjonelt enzym som til vanlig dannes i retrovirus, men som nå er mulig å produsere i Escherichia coli (E.coli). I dette "kit" brukes en optimalisert blanding av "rekombinante heterodimeric" enzym uttrykt i E.Coli, kalt Omniscript Revers Transcriptase og Sensiscript Revers Transcriptase.

Den første delen av protokollen består av en elimineringsprosess av genomisk DNA som innebar at 0.5 ul av RNA templatet ble rensset. Revers transkripsjon reaksjonen (Figur 9) starter med at forskjellige primere og oligo-dT binder seg til RNA molekylet. Ved binding i ulike områder, sikrer denne blandingen av primere og oligo-dT, at Revers transkriptase kan binde seg til ulike områder i RNA molekylet. Binding til RNA molekylet gjør at enzymet kan bruke sin RNA-avhengige DNA polymerase funksjon til å transkribere en komplementær cDNA-sekvens. Det blir da dannet et RNA:DNA hybrid som polymerasen bruker sin hybrid-avhengige exoribonuklease (RNase H) aktivitet på. RNA i komplekset blir da brytt ned og danner tilslutt et enkeltråd cDNA. Sluttproduktet ble en 20 ul cDNA prøve som ble fortynnet 1:1 med destillert H₂O. Disse cDNA prøvene ble videre benyttet i kvalitative og kvantitative metoder.

Figur 9: Qiagen Revers Transkripsjon



Figuren viser Qiagen Revers Transkripsjon. Først vil oligo(dT) og rimerene anneale til RNA sekvensen, som vil etterfølges av at RNA-avhengig DNA polymerasen binder seg til RNA. Den vil danne et RNA:DNA hybrid, som vil bli brutt opp av RNase H aktivitet av DNA polymersen som degraderer RNA. Sluttproduktet blir et enkeltrådet cDNA. Figuren er tatt fra Quantitect Revers Transcription Handbook (Qiagen 2005)

Kvalitative metoder

For å avgjøre om genet E-cadherin på kromosom 26 og dens paralog på kromosom 11 er uttrykt i leveren på laks, ble det benyttet metodene PCR og sekvensering av PCR-produkt. I dette kapittelet blir det i første del beskrevet hvordan cDNA blir amplifisert under metoden PCR ved bruk av Primersett 1. For å kontrollere at det ble dannet et produkt ved bruk av disse

primerne ble det gjennomført gel elektroforese. I neste steg beskrives prosessen før sekvenseringen og selve sekvenseringen av produktet. Her blir det dannet flere små sekvenser som, i siste del av kapittelet, blir samlet og analysert ved bruk av ulike bioinformatikk programmer. Konsensussekvensen ble benyttet til å vise om genet var aktivt og for finne mulige variasjoner mellom de ulike genotypene.

- **PCR**

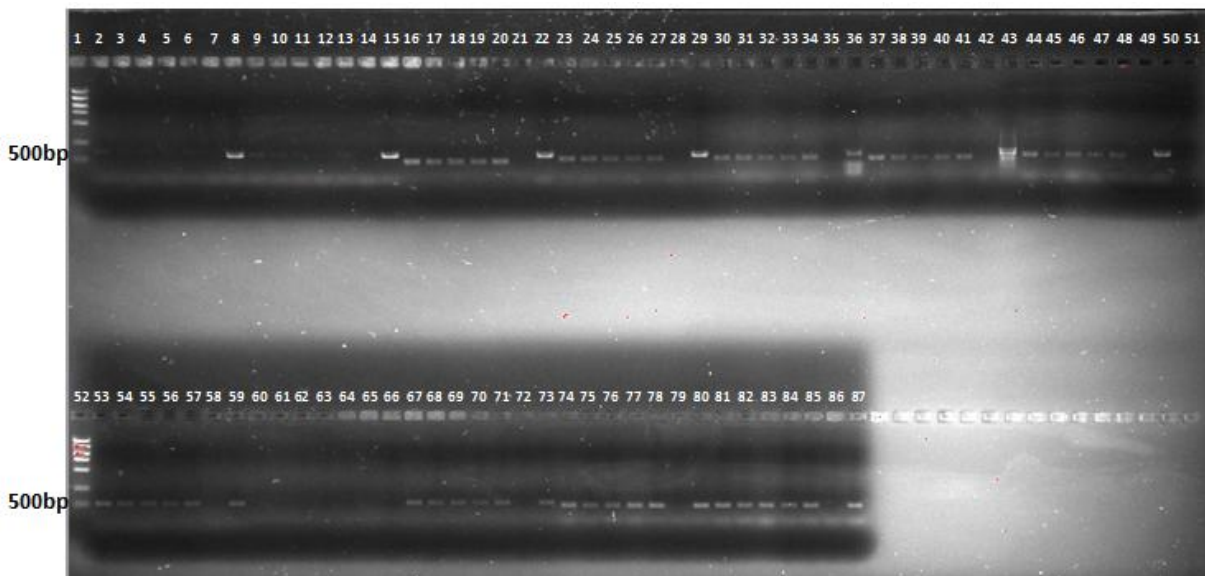
For å avgjøre om genet E-cadherin og dens paralog er uttrykt i leveren hos laks må man bevise at det eksisterer mRNA fra disse genene. Dette kan gjøres ved sekvensering av cDNA prøver som er tatt fra dette vevet. Prøvene må ha høy konsentrasjon av den spesifikke cDNA sekvensen for å oppnå et godt resultat under sekvensering. Ved å benytte teknikken PCR kan man få dannet prøver med høy konsentrasjon av den spesifikke cDNA. Metoden ble publisert i en artikkel i Science (1985) skrevet av Kary B. Mullis (Saiki, Scharf et al. 1985). Dette er en ”in-vitro” teknikk som, ved bruk av spesialdesignede primere, bruker en varmestabil DNA polymerase til å amplifisere en spesifikk DNA sekvens. PCR reaksjonen gjennomføres i tre trinn, disse består av denaturering, hybridisering og polymerisering. De tre stegene ble gjentatt 30 sykluser for å amplifisere opp store mengder av den utvalgte DNA sekvensen. For mer detaljer rundt PCR se (Sjøberg 2006). Under dette eksperimentet ble det benyttet 31 cDNA prøver fra individer som representerte alle genotyper (QQ, Qq, qq). Framgangsmåten er beskrevet i avsnitt VII i appendiks.

- **Agarose gel elektroforese:**

For å avgjøre hvorvidt PCR reaksjonen fungerte optimalt og om det ble produsert et produkt med høy konsentrasjon og med riktig lengde, ble prøvene kjørt i en gel elektroforese (Protokoll er i appendiks, avsnitt VIII). Gel elektroforese brukes til å separere de negativt ladde DNA molekylene basert på størrelse, ved bruk av elektrisk strøm. Under elektroforesen ble det benyttet agarose som gel material. 1% agarose gel kan separere DNA fra 500 bp til 6000 bp. Med mindre DNA molekyler må man øke til 2% agarose for å øke spesifisiteten (Sjøberg 2006 s. 174-175). Denne metoden kan brukes til å gi antydninger om lengden på produktet og om hvor høy konsentrasjonen er. Ved å sammenligne hvor langt cDNAet har vandret på gelen, med ”Mass DNA ladder” (N3237S, New England Biolabs, Ipswich, England)(Appendiks avsnitt XII), kan man beregne lengden på produktet. Dersom man har et

tydelig bånd, er det en indikasjon på at PCR reaksjonen mest sannsynlig har fungert og at man har en bra konsentrasjon av produktet. I figur 10 er det vist et eksempel på et typisk oppsett ved gel elektroforese. Under disse PCR forholdene fungerte ikke alle primerpar (1,7,8,9, 12,13 og 20) ettersom de ga svake eller ingen bånd. Se for eksempel Primerpar12 og Primerpar 13 i Figur 10. Primerparene ble derfor optimalisert ved å endre på ulike forhold i reaksjonen som ble benyttet i protokollen i avsnitt VII i appendiks.

Figur 10: Agarose gel med PCR produkt av Primersett 1



Figuren viser resultatet av agarose gel med PCR produkt av primersett 1 for E-cadherin på kromosom 11. Her ser man at alle primerpar fungerer med unntak av primerpar 12, primerpar13 og primerpar 20. Banen viser følgende prøver:(1) "Mass" ladder (2-8)Primerpar 12; (9-15) Primerpar 13; (16-22) Primerpar 14; (23-29) Primerpar 15; (30-36) Primerpar 16; (37-43) Primerpar 17; (44-50); Primerpar 18 (51) Kontroll uten primerpar (52) "Mass" ladder; (53-59) Primerpar 19; (60-66) Primerpar 20; (67-73); Primerpar 21; (74-80) Primerpar 22, (81-87) Primerpar 22. Hvert primerpar har sju brønner med ulikt cDNA og kontroller: (A) cDNA 1 (B) cDNA 3; (C) cDNA 5; (D) cDNA 7; (E) cDNA 11; (F) Negativ kontroll uten cDNA; (J) Positiv kontroll (gamle PCR prøver).

- Gradient PCR med endring i $MgCl_2$ konsentrasjonen

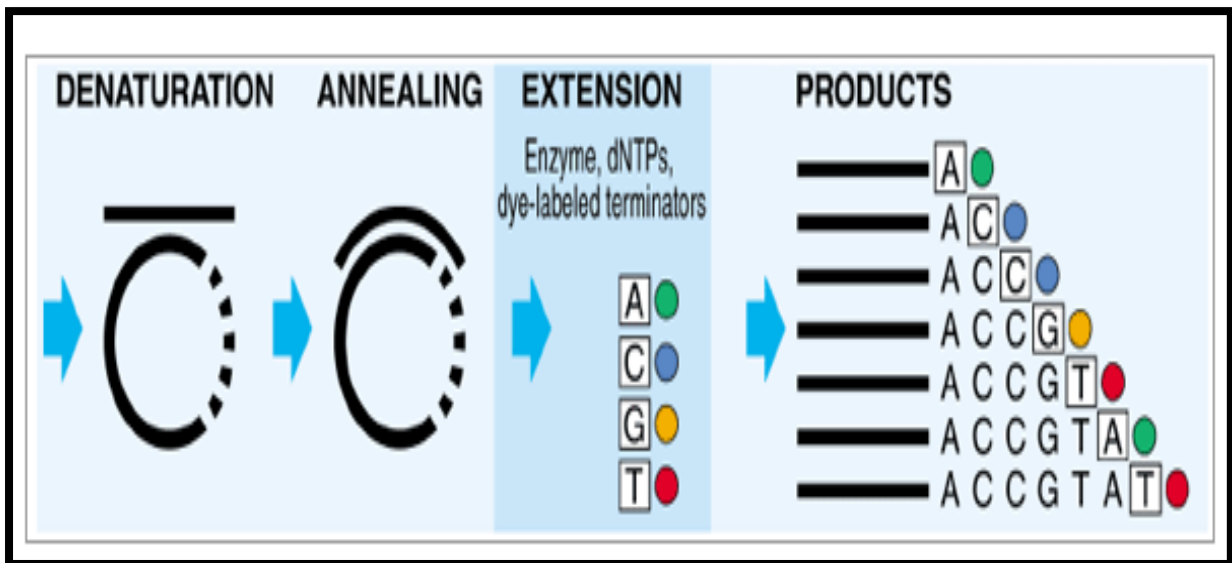
Under PCR er det ikke alltid slik at alle primerpar fungerer under samme forhold. Det er mange forhold som kan varieres for å få primerparene til å fungere optimalt og danne et produkt med høy konsentrasjon. De primerparene som ikke fungerte under normale forhold, ble testet i en gradient PCR med varierende $MgCl_2$ konsentrasjon. Denne protokollen er beskrevet i avsnitt VII i appendiks. Gradient PCR er en metode som optimaliserer "annealing" temperaturen for å sørge for optimal primerbinding til templat. Det var ni forskjellige temperaturer i "annealing" fasen under PCR reaksjonen, disse varierte i temperatur fra 46.3

grader Celsius til 59.7 grader Celsius. I tillegg ble $MgCl_2$ konsentrasjon endret, som påvirker bindingsforholdet mellom polymerasen og dNTP. $MgCl_2$ fungerer som en kofaktor for polymerasen ved å binde til dNTP. Ved å variere konsentrasjonen av $MgCl_2$ finner man det optimale forholdet mellom disse reagensene. I bufferen som ble brukt under normale forhold var det allerede tilsatt 1.5mM $MgCl_2$, med en "stock" konsentrasjon på 15mM. I protokollen fra Qiagen anbefales det å øke $MgCl_2$ gradvis til en endelig konsentrasjon rundt 5 mM i prøven (Qiagen 2010 s. 32). (Den ble økt til 4.5 i dette eksperimentet). I tillegg til å endre på disse forholdene, ble en ekstra reagent tilsatt i mastermiksen, Q-Solution. Denne reagensen kan forbedre oppløsningen av trådene i en dsDNA kompleks. Ofte eksisterer det sekundære strukturer og høy GC i cDNA templat som kan ha negativ påvirkning på smelteoppførselen av cDNA. Q-Solution reduserer effekten av disse faktorene (Qiagen 2010). Ved å endre alle disse forholdene ble det produsert produkt under PCR som ble bekreftet ved gel elektroforese. De endelige PCR forholdene for disse primerparene er beskrevet i avsnitt XII i appendiks.

- **DNA sekvensering**

For å bekrefte at PCR produktene var cDNA fra E-cadherin og dens paraloge, måtte produktet renses og sekvenseres. I 1977 ble det publisert en metode for DNA sekvensering av Sanger som var basert på gel elektroforese (Sanger, Nicklen et al. 1977). I dag har man videreutviklet Sanger metoden til en moderne automatisk sekvensering basert på kapillær elektroforese. Sanger-kjede termineringsmetoden er basert på bruk av DNA polymerase, fluorescerende "dideoxysyribonucleotides" (ddNTP) og "deoxyribonucleotider" (dNTP) i en innmerkingsreaksjon (Figur11). De fire ulike fluorescerende ddNTP er i begrenset mengde, noe som medfører tilfeldig terminering av sekvensen ved binding. De ulike fluorescerende ddNTP vil ha ulik farge, som gjør at vi kan identifisere ende nukleotid. Etter innmerkingsreaksjonen ble prøven kjørt i en kapillær elektroforese for å separere fragmentene. Lengden på produktet vil avgjøre rekkefølgen på fragmentene som kommer ut av kapillære kolumnen. De minste fragmentene kommer først, som tilsvarer den første nukleotid i sekvensen. Deretter bygger den sekvensen bortover (Sjøberg 2006 s. 187-189). Protokollen som ble benyttet er beskrevet i avsnitt IX i appendiks. Det endelige produktet etter sekvensering ble så analysert med ulike bioinformatikk verktøy.

Figur 11: Innmerkingsreaksjon



Figuren viser framgangsmåten i innmerkingsreaksjonen. Først vil det skje en denaturering av dobbeltrådet DNA template. Deretter vil primeren ”anneale” til en singelstrandet DNA sekvens. Polymerasen vil danne sekvenser med ulik lengde på grunn av ddNTP og dNTP. Bildet er hentet fra Life Technologies (Life-Technologies)

- **Analyse av kvantitativ metode resultat**

Under sekvenseringen dannes flere sekvenser fra samme cDNA. For å sette disse riktig sammen og danne en fullstendig sekvens, ble det benyttet flere bioinformatikk programmer. Ved å benytte programmene Blast og PolyPhrap ble den resulterende sekvensen analysert for å identifisere eventuell variasjon.

- **Phred**

For å kunne samle sekvensen til en lang konsensussekvens er det nødvendig å vite hvor sikker hver base i alle sekvensene var. Basert på basene som hadde best kvalitet kunne man danne den sekvensen som var sikrest. For å bestemme kvalitetsverdien på enkeltbasene ble kommandolinjeprogrammet Phred benyttet. Programmet brukte cDNA sekvensens ”trace” fil som har blitt dannet under sekvensering. Phred gir enkeltbasene en kvalitetsverdi basert på en log-transformed error sannsynlighet, som danner filer som kan leses av Phrap (Ewing and Green 1998, Ewing, Hillier et al. 1998).

- **Phrap**

For å samle disse sekvensene til en sekvens basert på den beste kvalitetsverdien, ble programmet Phrap benyttet. Dette er et kommandolinje program som bruker Phred kvalitetscore for å samle sekvenser og danne en eller flere contig. Resultatet vil se ut som en ”mosaic” av sekvenser med både dårlig og god kvalitetsverdi alignet sammen (Green 1994).

- **Consed**

For å vurdere sammensetningen av sekvensene, kan man visualisere resultatet fra Phrap i Consed (Gordon, Abajian et al. 1998). På denne måten er det mulig å redigere og kontrollere kvaliteten på filene. I tillegg ble det foreslått en mulig konsensussekvens basert på basene med den beste kvaliteten. Denne sekvensen ble blastet opp mot nr/nt databasen, for å bekrefte at sekvensen tilhørte E-cadherin eller dens paraloge gen. Sekvensen ble også blastet opp mot referansensekvensen for å se etter mulige variasjoner i sekvensen, som alternativ spleising.

- **PolyPhred**

Mulige sekvens produkter fra Phrap ble benyttet i en analyse for å identifisere mulige SNPer i hver genotype. Det ble benyttet et program kalt PolyPhred (Bhangale, Stephens et al. 2006, Stephens, Sloan et al. 2006). Programmet bruker filene fra Phrap til å finne mulige variasjoner mellom individene innenfor en genotype. Dette programmet tar hensyn til kvaliteten på hver base, og basert på dette produserer den en liste over mulige SNPer. Det endelige resultatet ble vist i programmet Consed for å kunne kontrollere resultatet manuelt. Her ble SNP mønsteret mellom genotype QQ og genotype qq sammenlignet for å finne mulige forskjeller og likheter. Nucleotidesekvensen ble tilslutt translatert til en aminosyresekvens ved bruk av programmet Expsy Translate. En åpen leseramme med en mulig proteinsekvens ble observert for begge genotype (QQ og qq). Disse aminosyresekvensene ble deretter sammenlignet ved bruk av BLAST for å se etter eventuelle forskjeller mellom genotypene. I tillegg ble aminosyresekvensen til E-cadherin sammenlignet ved bruk av ClustalW(Chenna, Sugawara et al. 2003).

Kvantitative metoder

For å se etter forskjell i uttrykkningen mellom genotypene (QQ,qq) ble det benyttet kvantitative metoder for å måle konsentrasjonen til cDNA. Det ble benyttet både dyr samlet før smitte og dyr samlet etter smitte, fordi man ønsket å se om en eventuell forskjell mellom genotypene var til stede både før og etter smitte. Det ble brukt åtte forskjellige dyr som hadde forskjellige genotyper (qq, QQ). For å kunne sammenligne de ulike cDNA prøvene var det viktig at de hadde lik konsentrasjon. Dette fjerner muligheten for at en eventuell forskjell mellom QQ og qq er et resultat av ulik startkonsentrasjon. Konsentrasjonen til RNA prøvene ble målt tre ganger ved bruk av Qubit. Her ble det benyttet Quant-IT RNA Assay Kit (Life Technologies, Carlsbad, California, USA) som ble utført nøyaktig etter protokoll som var vedlagt. Det ble brukt 1 ul av RNA i en 200 ul total blanding, som ble lest av Qubit måleren. Basert på den gjennomsnittlige verdien ble prøvene fortynnet til den laveste konsentrasjonen (Appendiks, avsnitt X). Disse prøvene ble videre benyttet i en kvantativ metode kalt Real-time PCR(qPCR).

- qPCR

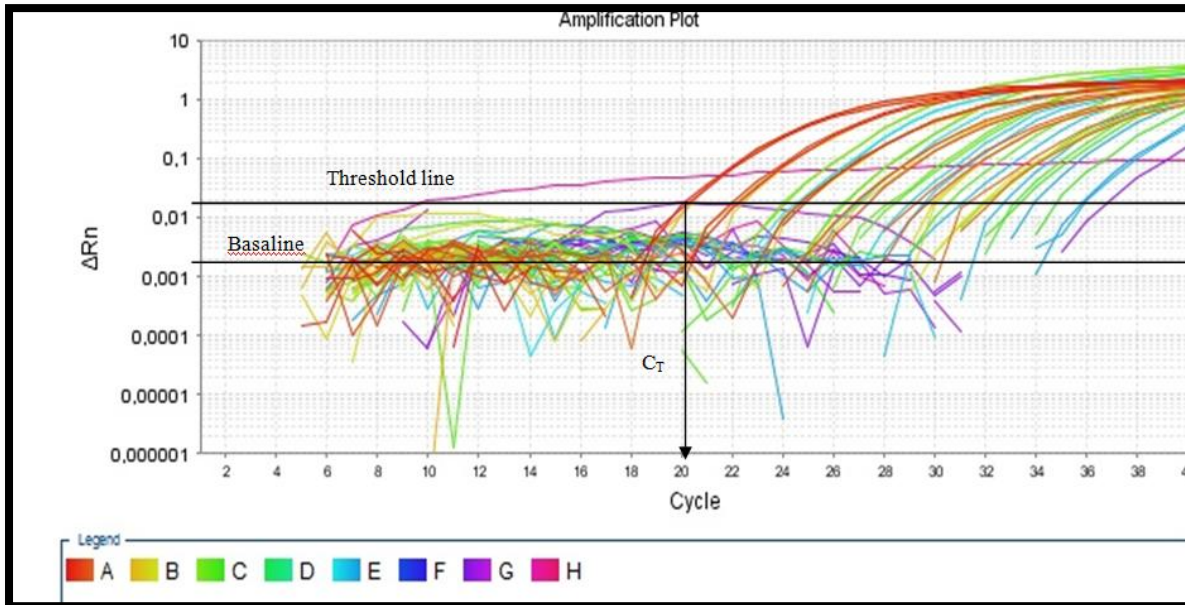
For å kunne måle konsentrasjonen i de ulike genotypene og deretter sammenligne de, ble det benyttet metoden qPCR. Under denne metoden blir en spesifikk cDNA sekvens amplifisert samtidig som cDNA mengden blir kvantifisert. Teknikken er basert på registrering av et signal fra en fluorescens farge bundet til nukleinsyren. Dette signalet vil være proporsjonalt med konsentrasjonen av PCR produktet i alle sykluser. Dette gjør det mulig å gjennomføre sammenligningsanalyse mellom ulike individer av et spesifikt gen, uttrykt i et bestemt vev (Invitrogen 2008 s. 4). Grunnen til at man kan sammenligne de ulike prøvene, er fordi resultatet fra hver primer blir normalisert ved bruk av en referanseprimer. Det ble benyttet to referanseprimerpar som heter EF1AB Olsvik(Olsvik, Lie et al. 2005) og EF1AB SMJ(Skjoesol, Skjaeveland et al. 2011) (Metabion, Martinsried, Tyskland). Benyttelsen av disse er nærmere forklart under metoden Delta Delta Ct. Først ble amplifiseringseffektiviteten testet for referanseprimerne og primerepar i Primersett 2.

- **Effektivitets test**

Under forberedelse for qPCR reaksjonen ble amplifiserings effektivitet til primerparene (Primersett 2) testet. I reaksjonen ble det benyttet SYBR Select Master Mix (Life Technologies, Carlsbad, California, USA) og vedlagt protokoll fra Life technologie (Life-technologies 2012). Dette er et viktig steg før en kjører qPCR, fordi ulik effektivitet ved amplifisering kan gi falskt resultat. Testen består av å kjøre vanlig qPCR med en fortynningsrekke på seks ulike konsentrasjoner av et cDNA med to replikater. Det ble benyttet fortyngningene 1:1, 1:5, 1:25, 1:125 og 1:625. Hver prøve i fortynningsrekken ble fortynnet i 20 ul dH₂O, med unntak av startkonsentrasjonen. Det ble tatt 2 ul av de fortynnede cDNA prøvene i en 20 ul total mastermiks. Denne mastermiksen inneholdt 0,5 av hver 10 uM primer og det fluorescerende kjemikaliet SYBR® GreenER™, som binder RNA (og DNA). Den endelige mastermiksen ble tatt i en 96-Brønn Plate, som ble tatt i en 7500 Real Time PCR maskin fra Applied Biosystems. Under reaksjonen vil SYBR Green binde seg direkte til alle dobbel strandet DNA og sende ut et fluorescens signal. På grunn av denne uspesifikke bindingen kan den også binde til uønsket dobbel strandet DNA og primer-dimere (Qiagen 2004 s. 5). For å unngå falskt resultat er det derfor nødvendig med høyt spesifikke primere i tillegg til å teste for primer-dimer med smeltekurve.

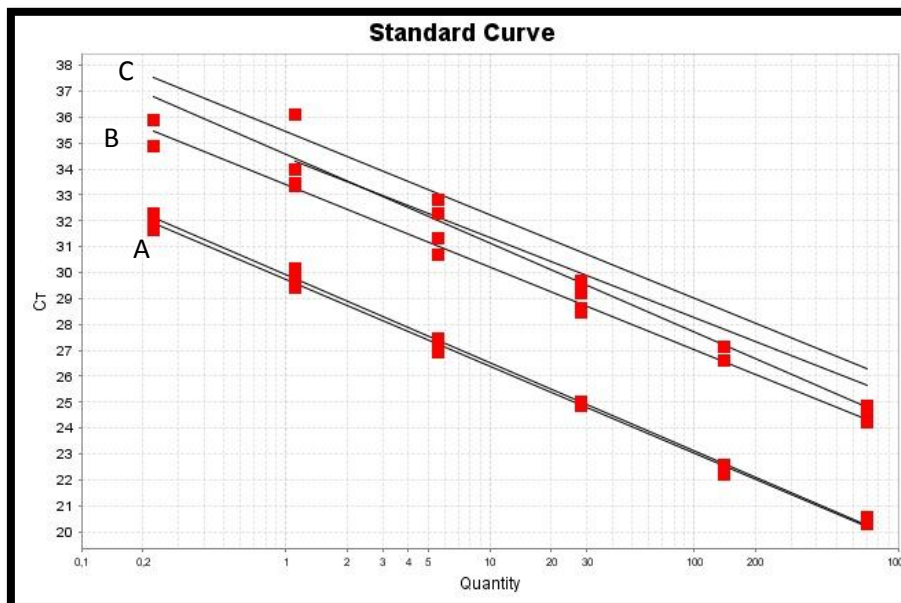
I resultatet ble det dannet et amplifiseringsplott og en standardkurve. Amplifiseringsplottet (Figur 12) viser mengde fluroscent signal på y akse mot antall sykluser på x-aksen. Dette gir en oversikt over dannelse av produkt over et bestemt tidsrom. Nå vil det være mulig å finne viktige verdier som "baseline", "Threshold line", og "Threshold syklus". "Baseline" er syklusene i begynnelsen av PCR reaksjonen, som har ingen endringer i det fluorescerende signalet. Når det skjer en signifikant økning i det fluorescerende signalet, vil dette nivået i fluorescerende signalet tilsvare "Threshold". Antall sykluser det tar er "Threshold" syklusen (Ct). Denne Ct verdien blir deretter plottet mot startkonsentrasjonen på de ulike cDNA prøvene. Dette danner standardkurven (Figur 13) som beskriver primer effektiviteten. Den gir en "slope" verdi som bør ligge rundt -3.364 for å ha en amplifiserings effektivitet tilsvarende 100 % (Invitrogen 2008 s. 4-12). For å kunne sammenligne resultatet mellom primerene bør alle primere ligge rundt 100 prosent med et mulig avvik på 5%.

Figur 12: Eksempel på et amplifikasjonsplott



Figuren viser et eksempel på et amplifikasjonsplott fra effektivitetstesten av primersett 2 og primereparene for referansegnet. (A)SMJ; (B)Olsvik; (C)Primerpar 1; (E)Primerpar 2. Her representerer delta R_n økningen i fluorescerende signalet på y-aksen mot antall sykluser på X-aksen. Her vises Baseline, Threshold line og C_T verdien. Det er viktig at replikatene fra en fortyning for et primerepar ligger veldig tett, i tillegg til å ha en fin stigende bue.

Figur 13: Standardkurve fra samme eksperiment som amplifiseringsplottet



Figuren viser et eksempel på en standardkurve fra samme eksperiment som amplifiseringsplottet i figur 12. A) SMJ og Olsvik; B) Primerpar 1; C) Primerpar 2. (Siste linje er fra en primer som ikke var med i qPCR). Her er C_T verdien på Y-aksen plottet mot kvantiteten på X-aksen. Linjene til replikatene bør ligge tett inntil hverandre. Hver linje som synker representerer en økning i startkonsentrasjon. "Slope" nummeret til denne linjen avgjør effektiviteten til primerene.

- **qPCR reaksjonen**

Selve qPCR reaksjonen er basert på samme protokoll som effektivitetstesten. Alle de åtte 40 ul cDNA prøvene ble fortynnet 1:1 med dH₂O. Under denne reaksjonen ble det benyttet tre replikater fra samme cDNA som alle har lik konsentrasjon. Basert på resultatet fra reaksjonen ble det dannet et amplifiseringsplott og en standerkurve. Her ble C_T verdien veldig sentral for å finne variasjonene mellom genotypene. For å kvantifisere konsentrasjonen av de ulike prøvene for et utvalgt gen, og deretter finne eventuelle forskjeller, kan en bruke en teknikk kalt Relativ kvantifisering (Qiagen 2004 s. 15). Teknikken består av å sammenligne uttrykket av et gen i to ulike prøver. Her kan man finne ut om genet er oppregulert eller nedregulert i forhold til den andre prøven. Utrekningen i denne metoden ble basert på Delta Delta Ct metoden.

- **Delta Delta Ct**

Metoden Delta Delta Ct er basert på normalisering av gen uttrykt i de eventuelle prøvene, ved bruk av et endogent referansegen til å korrigere for mulig variasjon mellom dem. Denne variasjonen kan innebære ulikt RNA innhold, ulik effektivitet i revers transkriptase, mulig RNA degresjon og hemmer i RNA prøven. Et referansegen må ha et stabilt uttrykk i alle prøver, uansett tilstand. I laks er genet EF1-AB antatt å være stabilt, derfor ble brukt to primerpar som binder til akkurat dette genet (Olsvik, Lie et al. 2005, Skjesol, Skjaeveland et al. 2011). For å normalisere prøvene ble den gjennomsnittlige Ct verdien for referanse genet i en bestemt cDNA prøve, trukket fra Ct verdien til primerparene i samme cDNA prøve. Deretter ble det dannet en gjennomsnittlig normalisert Ct verdi basert på prøvene med genotype qq. Den gjennomsnittlige Ct verdien ble trukket fra alle de fire normaliserte prøvene med genotype QQ individer. Ved å opphøye tallet 2 med disse verdiene(2^{Δ}), vet man hvor mye, mer eller mindre, hver enkel QQ prøve er uttrykt i forhold til den gjennomsnittlige verdien av prøvene med genotype qq (Qiagen 2004 s.20). Samme fremgangsmåte ble brukt for å finne forskjellen mellom smittet og ikke-smittet laks innenfor en genotype. Her ble den gjennomsnittlige normaliserte Ct verdien av ikke-smittet prøver trukket fra hvert enkelt smittet prøve.

3. Resultat og Diskusjon

3.1 Sekvenser av mRNA til E-cadherin på kromosom 26 og kromosom 11

I første del av oppgaven ble det benyttet cDNA fra leveren til 31 ulike laks (voksen og yngel), for å undersøke om genet E-cadherin og dens paralog på kromosom 11 var uttrykt i leveren til fisken. cDNA sekvensene ble kopiert opp ved bruk av PCR og deretter sekvensert.

Sekvensene var mellom 500 til 1000 basepar og ble tilslutt samlet i tre grupper basert på genotypene innenfor hvert gen. Det ble dannet tre ulike konsensussekvenser, QQ, Qq og qq, for E-cadherin på kromosom 26, og en konsensussekvens for paralogen på kromosom 11.

Sekvens materialet som konsensussekvensene var basert på hadde variabel kvalitet og lengde. Dette var et resultat av at flere primerpar ikke fungerte optimalt under reaksjonen i kombinasjon med lav konsentrasjon av E-cadherin i cDNA fra yngel. Under designet av primerparene kunne man ikke alltid velge det primerparet som var best tilpasset forholdene i PCR. Noen av de beste primerne kunne binde seg til mRNA fra det paraloge genet, noe som kunne ha gitt falskt resultat under PCR.

Flere forhold i reaksjonen ble optimalisert, i tillegg til at det ble inkludert flere cDNA prøver fra voksne individer. Dette ga flere og bedre produkter av E-cadherin på kromosom 26 selv om det fortsatt var variabel kvalitet og lengde på produktene. Variasjonen var spesielt gjeldende for produktene for genotype Qq, som dannet en kortere konsensus sekvens med stor variasjon i kvaliteten. Her ble det også benyttet et mindre antall individer i forhold til de andre genotypene. For de to andre konsensussekvensene hadde man en fullstendig sekvens med den forventede lengden. Noen deler av sekvensene hadde bra dekningsgrad og høy kvalitet mens andre områder hadde få sekvenser og dårlig kvalitet. Ved å sammenligne sekvensen opp mot referansesekvensen manglet det fire basepar på 5' enden og 53 basepar på 3' enden for genotype qq og QQ. Dette var forventet ettersom ingen av primerparene festet seg til den ytterste delen av sekvensen. Primerpar som ble foreslått i dette området hadde dårlig primerpar "penalty" verdi og for få "mismatch" til det paraloge genet. For E-cadherin på kromosom 11 ble det ikke dannet en fullstendig sekvens. Det var mangelen på sekvens fra basepar 2171 til basepar 2644 som tilsvarer et område på 473 basepar. Dette området var basert på to primerpar som ikke fungerte optimalt under PCR. PCR produktet hadde for lav

konsentrasjon til å gi et bra sekvenseringsresultat. Optimaliseringen av forholdene rundt primerparet fungerte delvis, men det var ikke nok tid til å teste ut andre forhold i reaksjonen.

Når man laget primerparene ble det tatt høyde for at primerene ikke skulle binde til begge paraloge gen samtidig. Dette var basert på forskjeller mellom referansesekvensen til de paraloge genene. For å kontrollere at det også eksisterte forskjell mellom konsensussekvensene til de paraloge genene ble sekvensene sammenlignet. Sammenligningen ble utført ved å benytte programmet BLAST, og resultatet er vist i avsnitt XVII i appendiks. Ved å sammenligne disse sekvensene ble det observert forskjell mellom konsensussekvensene.

Alle sekvensene ble sammenlignet med referansesekvensen for å kontrollere at sekvensen tilhørte genet E-cadherin og dens paralog. Det er kjent at det er store likheter mellom bestemte domener av cadherin familien. Derfor er det desto viktigere å kontrollere at vi har rett gen. Det ble observert stor likhet mellom sekvensene, noe som antyder at sekvensen tilhører E-cadherin. Konsensussekvensene ble også kontrollert ved å blaste de opp mot nr/nt databasen. Det ble kun observert sekvenser fra cDNA til E-cadherin fra laks og andre arter som hadde store likheter. Derfor konkluderes det med at E-cadherin på kromosom 26 og det paraloge genet på kromosom 11 er uttrykt i leveren til laks.

3.2 Variasjon mellom mRNA sekvensen til genotype qq og QQ til E-cadherin

På grunnlag av konsensussekvensene til genotypene qq og QQ kan man identifisere eventuelle variasjoner. Variasjonene mellom sekvensene kan være årsaken til at noen individer har større sannsynlighet for å bli smittet av IPN viruset. Ved å sammenligne begge sekvensene med referansesekvensen ble det tidligere observert stor likhet mellom de ulike genotypene og referansesekvensen. Alle exon ble observert i sekvensen. Dette betyr at mRNA ikke har gjennomgått alternativ spleising. Heller ingen andre deler av sekvensen manglet, noe som antyder at det ikke eksisterer andre prosesser som kan endre mRNA sekvensen. På en annen side ville man ikke observert om det ble benyttet en promotor som ligger "upstream" i samme leseramme, ettersom sekvensen da hadde ligget utenfor området som primerene binder til. Fremdeles er det altså en mulighet for at andre mekanismer skaper individer som varierer i overlevelsessevne ved infeksjon av IPNV.

Ved deteksjon av mulige SNPer i sekvensene, innenfor hver genotype, ble det observert tilsammen sju SNPer for begge genotyper (Tabell 1). To av disse forskjellene ble identifisert i begge genotyper, mens de resterende fem SNP kun var til stede i den ene av de to genotypene. Den ene SNP ved posisjon 1123 i genotype QQ hadde variasjon mellom CC,CT og TT. Det var 12 individer som hadde CC, mens det var kun et individ som hadde TT. Individet med TT kan ha feil genotype siden QTL genotypen har blitt beregnet på grunnlag av DNA-tester som ikke 100% sikre. I genotype qq eksisterte det ingen variasjon i denne posisjonen, den hadde kun varianten TT (i 17 individer). Det var tydelig forskjell mellom de homozygote genotypene, der den ene har C og den andre har T. (Figur 14). Sekvensområdet som SNPen var detektert i ble basert på flere sekvenser av god kvalitet.

I tidligere arbeider for å finne variasjon mellom genotypene har man ikke påvist denne forskjellen i nukleotidsekvensen. Gjennom å sekvensere hele genomet til dyr med kjent genotype qq eller QQ, identifiserte Thomas Moen et al. andre SNPer med stor forskjell mellom qq og QQ (upublisert). Sekvensene fra disse dyrene ble ”alignet” opp mot en helgenom referansesekvens. I området til den aktuelle SNPen (i posisjon 1123/1281 i nukleotidsekvensene) var det ingen sekvenser fra qq-eller QQ-dyrene ”alignet” til referansesekvensen. En mulig feil i referansesekvensen kan være årsaken til at sekvensene ikke ble ”alignet” til dette området, noe som kan forklare hvorfor denne SNPen ikke har blitt funnet før. Oppgaven viser en antydning til at den aktuelle SNPen kan være den ansvarlige for variasjonen mellom genotypene til E-cadherin. Resultatet indikerer dermed at individer med CC kan ha større sannsynlighet for å overleve en infeksjon av IPNV.

Forsøket er likevel basert på for få individer til å konkludere med at dette er den kausative mutasjonen. I tillegg kan det fortsatt eksistere andre mutasjoner som ikke ble detektert på grunn av den dårlige sekvenskvaliteten i noen områder av genet.

Tabell 1: Identifiserte SNPer

Genotype som SNP er detektert i	Posisjon	Homozygot (høyest antall individer)	Hetrozygot	Homozygot (med minst antall individer)	Den andre genotype	Samme posisjon i den andre genotype	Homozygot (høyest antall individer)	Hetrozygot	Homozygot (med minst antall individer)
qq	1351	CC (11)	AC (1)	AA(1)	QQ	1193	CC(17)	CA(0)	AA(1)
qq	1413	CC (13)	GC(1)	GG(1)	QQ	1255	CC(17)	CG(0)	GG(1)
qq	1710	AA (14)	AG(3)	GG(1)	QQ	1552	AA(17)	-	-
qq	1868	CC (11)	CT (5)	TT(1)	QQ	1709	CC(12)	-	-
qq	2039	CC (9)	CT (2)	TT (0)	QQ	1881	CC(15)	-	-
QQ	1089	TT(9)	CT(1)	CC(1)	Qq	1247	TT(8)	-	-
QQ	1123	CC(12)	CT(3)	TT(1)	Qq	1281	TT(12)	-	-
QQ	1193	CC(17)	CA(0)	AA(1)	Qq	1351	CC (11)	AC (1)	AA(1)
QQ	1255	CC(17)	CG(0)	GG(1)	Qq	1413	CC (13)	GC(1)	GG(1)

Tabellen viser identifiserte SNPer i begge genotyper. I tillegg viser den variasjonen i samme posisjon i den andre genotypen. Her ser man posisjonen SNPen ble identifisert i og hvilken genotype den hører til. Alle ulike variasjoner per SNP vises, med antall individer som har denne varianten i parentes. Rosa farge indikerer like SNPer, som og er identifisert i begge genotyper. Gul farge indikerer den mulige kausale SNPen.

Figur 14: Alignment av sekvenser

```

ContigAA      AAGCATCAAACCAGGTGACGAGGTCATGCAGGTAACGGCCACTGATGCTGATGAGGAGG 1200
ContigRR      AAGCATCAAACCAGGTGACGAGGTCATGCAGGTAACGGCCACTGATGCTGATGAGGAGG 1042
*****

ContigAA      GCTCTGCCAATTCTGATGTCAGATACACCATTCTCAGTCAGGAGCCTCCACTACCAAGCC 1260
ContigRR      GCTCTGCCAATTCTGATGTCAGATACACCATTCTCAGTCAGGAGCCTCCACTACCAAGCC 1102
*****

ContigAA      CCAACATGTTTGTTCATCAACTCTGTGACTGGAGGGATTTCGGGTTAATGCACCTGGGTTGG 1320
ContigRR      CCAACATGTTTGTTCATCAACCTGTGACTGGAGGGATTTCGGGTTAATGCACCTGGGTTGG 1162
*****

ContigAA      ACAGAGAGAAAATTCCCAAATATACTTTGGCAATCCAAGCTGCCGATATGGAGGGAAAATG 1380
ContigRR      ACAGAGAGAAAATTCCCAAATATACTTTGGCAATCCAAGCTGCCGATATGGAGGGAAAATG 1222
*****

ContigAA      GCCTTACCAGCTTTGGCAAAGCCATCATTACACTAACGGACAGCAATGACAACGCGCCAC 1440
ContigRR      GCCTTACCAGCTTTGGCAAAGCCATCATTACACTAACGGACAGCAATGACAACGCGCCAC 1282
*****

ContigAA      AGTTTGTGACGCCTTCGTACACCGTGTTCAGTCCCAGAGAATAAAGTGGATGCCTTGGTGG 1500
ContigRR      AGTTTGTGACGCCTTCGTACACCGTGTTCAGTCCCAGAGAATAAAGTGGATGCCTTGGTGG 1342
*****

ContigAA      TGAAAATGCCAGTGACTGATGGAGATGAGCCTCACTCCTCTGCCTGGGCCACCACCTACA 1560
ContigRR      TGAAAATGCCAGTGACTGATGGAGATGAGCCTCACTCCTCTGCCTGGGCCACCACCTACA 1402
*****

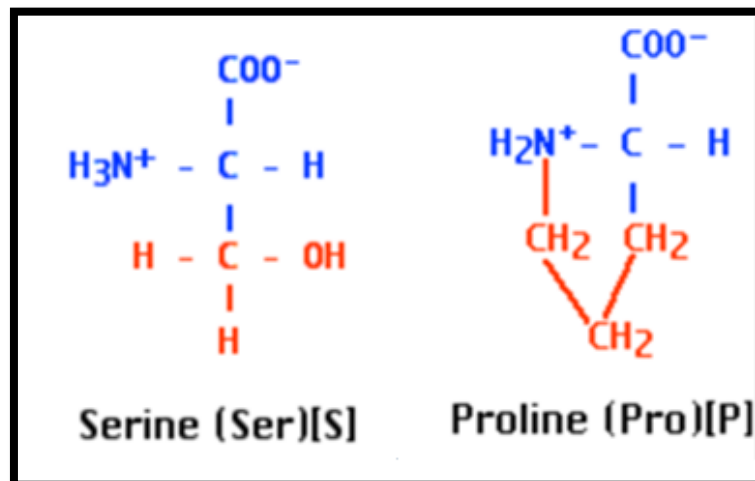
```

Figuren viser "alignment" av sekvenser fra genotype qq og genotype QQ. Her er det en mulig variasjon mellom sekvensene, i genotype qq er det basen T posisjon 1181, mens i genotype QQ er det basen C i posisjon 1123.

3.3 Fra mRNA til aminosyre sekvens

Endringer i mRNA sekvens trenger nødvendigvis ikke ha store konsekvenser for proteinet. Denne endringen kan lede til samme aminosyre, eller en kjemisk lik aminosyre, som ikke vil forandre strukturen til proteinet. Ved kjemiske ulike aminosyrer kan dette resultere i endring i struktur og funksjon. Ved å sammenligne aminosyresekvensen for mulige endringer vil man kunne få en indikasjon om hvorvidt SNPen, i posisjon 1123/1281 i nukleotidsekvensen, kan endre proteinstrukturen. Nukleotidsekvensen ble translert til en aminosyresekvens ved bruk av Expasy Translate (Gasteiger, Gattiker et al. 2003). Aminosyren i posisjon 325 tilsvarer samme posisjon i nukleotidsekvensen som den SNPen som ble detektert mellom genotypene. For mRNA i genotype QQ ble det observert aminosyre prolin, mens for genotype qq ble det observert aminosyren serine (se figur 16). Dette viser at endringen i nukleotidsekvensen har ledet til endringer i aminosyresekvensen. Aminosyrene har forskjellige egenskaper som kan påvirke strukturen i proteinet. Proline er en hydrofobisk aminosyre som har en veldig stiv struktur på grunn av sidekjeden sin. Den består av en 5-sidet ring hvor amin-nitrogenet er bundet til to karbonmolekyler (se figur 15). Med denne aminosyren i proteinkjeden vil den skape en kink som fører til begrensinger i folde muligheter. I tillegg har den hydrofobiske sidegruppen dårlig evne til å binde til vann og vil vende inn mot aminosyrekjeden. Serine derimot er en hydrofil aminosyre med en polar uladd sidegruppe (se figur 15). Den hydrofile sidegruppen kan danne hydrogenbindinger med H₂O, som vil gjøre at denne sidegruppen vender ut fra aminosyrekjeden. Den danner ingen kink i proteinsekvensen som gjør at den ikke lager samme type begrensing i foldemulighetene slik som sidegruppen i Proline (Lodish 2008 s. 41-43 og s. 1161). Skifte av aminosyre i proteinsekvensen kan endre strukturen til proteinet, som igjen kan påvirke egenskapene til reseptoren. Dette kan videre endre hvordan reseptoren kommuniserer med andre molekyler. Dersom dette er reseptoren som IPNV bruker for å binde seg til cellen, kan denne prosessen lede til at IPNV viruset ikke kommer seg inn i cellen. Dette er i samsvar med andre resultater som viser at QQ-fisken i liten grad er smittebærere etter en IPN-infeksjon, mens qq-fisken vanligvis er smittebærer selv om den overlever en IPN-infeksjon (Aqua Gen, upubliserte data). At fisken ikke er smittebærer indikerer at viruset ikke klarer å komme seg inn i cellene.

Figur 15: Serine og Proline



Figuren viser strukturen av serine og proline.
Figuren er hentet fra "Kimball's Biology Pages" (Kimball 1994)

Figur 16: Sammenligning av aminosyresekvensen til genotype QQ og genotype qq

Query	1	MGAFWFVELGVLILFLQAFKPGSSSESTCLPGFNSEIYIFKVERNHLQSGRRLGKVVFDCC	60
Sbjct	1	MGAFWFVELGVLILFLQAFKPGSSSESTCLPGFNSEIYIFKVERNHLQSGRRLGKVVFDCC	60
Query	61	TSRTSFLFHSEDSHFVKVDGDTLKLERGLTLHNGHKEVYVSTQSKGKKITVPVRVLHEAR	120
Sbjct	61	TSRTSFLFHSEDSHFVKVDGDTLKLERGLTLHNGHKEVYVSTQSKGKKITVPVRVLHEAR	120
Query	121	HGHHHNNHEMTTQPKPGASLSLPVLFNFKSSGGLKRRKRDWVIPPINFPENDRGPFPPKIM	180
Sbjct	121	HGHHHNNHEMTTQPKPGASLSLPVLFNFKSSGGLKRRKRDWVIPPINFPENDRGPFPPKIM	180
Query	181	VQIRSNNDKEVKIQYSITGTGADLPPVGI FTVDKNSGNLYVTEPLDREKKDKYILLAHAV	240
Sbjct	181	VQIRSNNDKEVKIQYSITGTGADLPPVGI FTVDKNSGNLYVTEPLDREKKDKYILLAHAV	240
Query	241	AVGAGIAEDPMEIIVKVIDMNDNKPVFTQDPFMGTVP EASKPGDEVMQVTATDADEEGSA	300
Sbjct	241	AVGAGIAEDPMEIIVKVIDMNDNKPVFTQDPFMGTVP EASKPGDEVMQVTATDADEEGSA	300
Query	301	NSDVRVTILSQEPPLPSPNMFVIN E VTGGIRVNAPGLDREKIPKYTLAIQAADMEGNGLT	360
Sbjct	301	NSDVRVTILSQEPPLPSPNMFVIN I VTGGIRVNAPGLDREKIPKYTLAIQAADMEGNGLT	360
Query	361	SFGKAIITLTDSNDNAPQFVTPSYTVSVPENKVDALVVKMPVTDGDEPHSSAWATTYKIV	420
Sbjct	361	SFGKAIITLTDSNDNAPQFVTPSYTVSVPENKVDALVVKMPVTDGDEPHSSAWATTYKIV	420
Query	421	DGDPQGLFNVSTGPSKLEGIITTA KPLDFEKNKNTLLVTVQNEVPFTISMPTSTATVVV	480
Sbjct	421	DGDPQGLFNVSTGPSKLEGIITTA KPLDFEKNKNTLLVTVQNEVPFTISMPTSTATVVV	480
Query	481	NVEDVNEAPVFTPVEKIIIRKPEDLPVDSDLVLYTATDPDTARNQKVTYKIRNDNAGWLSI	540
Sbjct	481	NVEDVNEAPVFTPVEKIIIRKPEDLPVDSDLVLYTATDPDTARNQKVTYKIRNDNAGWLSI	540
Query	541	NKDTGLIKVKSLMDRESTFVQDNKYSVIVL GIDNDEIPATGTGLIIIELEDVNDNAPTID	600
Sbjct	541	NKDTGLIKVKSLMDRESTFVQDNKYSVIVL GIDNDEIPATGTGLIIIELEDVNDNAPTID	600
Query	601	ESVIRVCNKESSPQLLSVTDKDGAGFTAPYTVQLQGSSHSNWTARMNDTKTGI ILLTKTM	660
Sbjct	601	ESVIRVCNKESSPQLLSVTDKDGAGFTAPYTVQLQGSSHSNWTARMNDTKTGI ILLTKTM	660
Query	661	LDSGDYTVVLRVSDNQGLHHDSTIQASVCDCKGADVQCSDKAVAGFGISSILGILGAVLL	720
Sbjct	661	LDSGDYTVVLRVSDNQGLHHDSTIQASVCDCKGADVQCSDKAVAGFGISSILGILGAVLL	720
Query	721	LLLLSLLLLMFLRKRGGEGKEPLLQEDDVRDNIYYYDEEGGGEDDQDFDLSVLHRGLDNR	780
Sbjct	721	LLLLSLLLLMFLRKRGGEGKEPLLQEDDVRDNIYYYDEEGGGEDDQDFDLSVLHRGLDNR	780
Query	781	PDVFRNDIAPT MARPEYRPRPANPADIGNFIDN LKAADNDPTAPPYD SLLVFDYEGGGS	840
Sbjct	781	PDVFRNDIAPT MARPEYRPRPANPADIGNFIDN LKAADNDPTAPPYD SLLVFDYEGGGS	840
Query	841	EAGSLSSLN SSSSGDDQDYDLLQEWGPRFKKLSDMYGGGED 881	
Sbjct	841	EAGSLSSLN SSSSGDDQDYDLLQEWGPRFKKLSDMYGGGED 881	

Figuren viser en sammenligning av aminosyresekvensen, til E-cadherin på kromosom 26, mellom genotype QQ og qq. Her er det en forskjell i posisjon 325

3.4 Komparativ analyse mellom begge genotypene

Komparativ analyse mellom ulike arter kan bidra til å få en bedre forståelse av gen i arter som nettopp har blitt identifisert. Ved å sammenligne aminosyresekvensen for begge homozygote genotyper for E-cadherin, med andre aminosyresekvenser av E-cadherin, kan man få en indikasjon av hva som er den opprinnelige aminosyren i posisjon 325. Det ble brukt aminosyresekvenser av E-cadherin fra andre fiskearter og paralogen på kromosom 11 i laks. Fiskearter som ble inkludert i analysen var zebrafisk (*Danio rerio*), blåsefisk (*Tetraodon nigroviridis*), Japansk Risfisk (*Oryzias latipes*, Medaka). Etersom det ikke var tilgjengelige sekvenser i de forskjellige databasene (nr/nt og ensembl(Flicek, Amode et al. 2010)) ble ikke sekvensene fra E-cadherin i Torsk (*Gadus morhua*) og Stingsild (*Gasterosteidae*) benyttet. I figur 17 ser man at Japansk risfisk og laks(qq) har aminosyren serine, mens de resterende artene har aminosyren proline. Dette gir ingen indikasjon på hva som kan være den opprinnelige aminosyren i E-cadherin. Dersom man kun sammenligner aminosyresekvensen til paralogene av E-cadherin i laks fremstår proline som den beste kandidaten. Proline finnes både i proteinet til paralogen på kromosom 11 og i E-cadherin med genotype QQ. Basert på denne informasjonen er det derimot ikke mulig å konkludere med hva som er den opprinnelige aminosyren i posisjon 325 i E-cadherin.

Figur 17: E-cadherin sammenligning

TetraodonNigroviridis	NFEVIQVNATDLDE-PNSYNSDIRYKILSQEPKLPDNLFTINP	194
OryziasLatipes	GTTVIQVKATDADE-PGNDNSIIKYTILSQEPKLPDMSFAIN	343
SalmoSalarQQ	GDEVMQVTATDADE-EGSANSDVRYTILSQEPPLPSPNMFVIN	331
SalmoSalarqq	GDEVMQVTATDADE-EGSANSDVRYTILSQEPPLPSPNMFVIN	331
SalmoSalar11	GDEFMQVTATDADE-EGSANSDVRYTILSQEPPLPSPNMFVIN	202
DanioRerio	GEVFMTVTATDADDKENTDNADISYVILSQDPPSPKPNMFAIN	313
	. . : * .*** * : . . * : : * * :*** * * . . :* .** . . * *	

Figuren viser sammenligning av E-cadherin mellom *Danio rerio*, *Oryzias latipes*(Medaka), *Tetraodon Nigroviridis* og laks. *Salmo Salar QQ* og *Salmo Salar qq* representerer de ulike homozygote genotypene for E-cadherin. Zebrafisk, *Tetraodon Nigroviridis* og laks(QQ) har aminosyren Proline. laks(qq) og *Oryzias Latipes* har Serine(S) i posisjon 325 i E-cadherin (merket rødt i figuren).

Det har blitt forsket på genene og genomsekvensen til mennesker og zebrafisk i mange år. Ved å sammenligne genet E-cadherin fra disse (og andre arter) med E-cadherin fra laks kan man få en forståelse av oppbyggingen av ulike domener og funksjon. Sammenligningen av

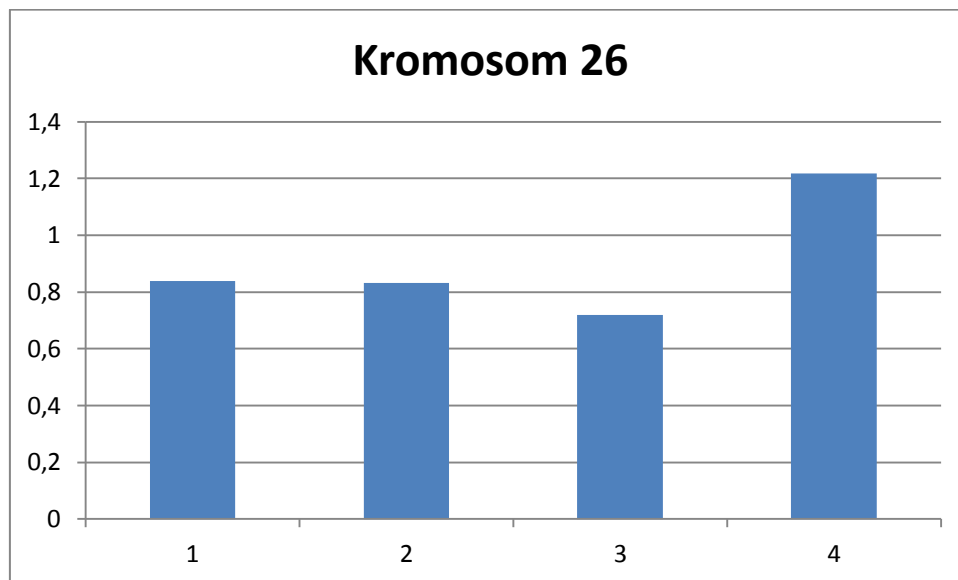
aminosyresekvensen viste stor likhet i de siste 150 aminosyrene i "carboxy" terminalen (sekvens sammenligningen vises i kapittel XVIII i Appendiks). Domenet er i den cytoplasmiske siden av cellen og er veldig konservert i de fleste arter. I aminoenden var det mindre likheter mellom sekvensen fra de ulike artene. Posisjon 325 i aminosyresekvensen ligger trolig i enden av EC-2 domenet mot EC-3 domenet (Figur 6), som er rundt samme området hvor Ca^{2+} binder. Endringer i området til EC-2 og EC-3 i E-cadherin er ofte involvert i utvikling av kreft i mennesker. Dette kan antyde at området er viktig for funksjonen til reseptoren. Alle disse antydningene er basert på informasjon om E-cadherin fra andre organismer, som kan variere fra den virkelige strukturen til E-cadherin i laks. Derfor er det viktig å studere strukturen til proteinet i laks for å få en bedre forståelse om konsekvensen av denne variasjonen.

3.5 Måle ekspresjons nivå av E-cadherin (qPCR)

Den totale mengden av mRNA i leveren på laks kan muligens påvirke evnen til å motstå infeksjon av IPN viruset. Evne til å motstå infeksjon kan gjenspeiles i forskjeller i uttrykkningen av genet. Her kan en mulig mutasjon påvirke aktivitetsnivået (genekspresjon) til E-cadherin genet på kromosom 26. Derfor var det interessant å undersøke om genotype QQ har lavere eller høyere konsentrasjon av mRNA i forhold til genotype qq. Ved å benytte åtte cDNA prøver fra leveren til laks ble qPCR metoden benyttet til å kvantifisere konsentrasjonen i hver prøve. Dette viste at individer med genotype qq og individer med genotype QQ hadde veldig liten forskjell i ekspresjon av E-cadherin genet. I figur 18 ser man at E-cadherin på kromosom 26 er ca. 0,9 ganger høyere uttrykt i individer med genotype QQ i forhold til individer med genotype qq. I forsøket ble det kun benyttet fire individer innenfor hver genotype, som fører til at hvert individ kan ha stor innflytelse. Antallet dyr som er undersøkt er for lite til at man kan si noe sikkert om forskjellene i uttrykk mellom qq og QQ. Samtidig er det slik at man, på grunn av den store forskjellen i IPN-resistens mellom qq-dyr og QQ-dyr, forventer en stor forskjell i uttrykk av E-cadherin mellom qq-dyr og QQ-dyr, dersom forskjell i uttrykk av dette genet henger sammen med IPN-resistensen. Derfor tyder resultatene på at forskjell i uttrykk av E-cadherin ikke henger sammen med IPN-resistensen. For å få et sikkert svar på dette spørsmålet bør eksperimentet gjentas på et større antall dyr fra populasjonen..

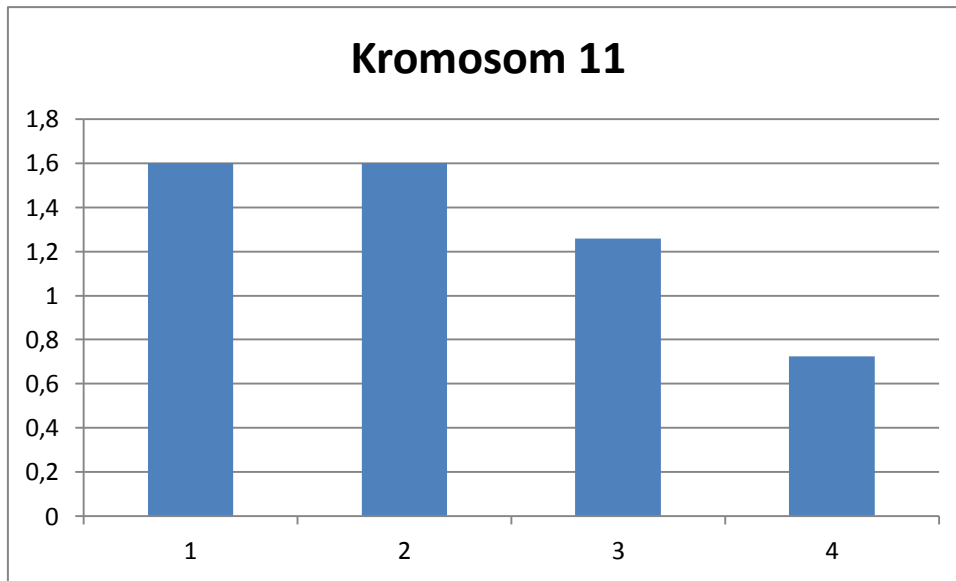
Heller ikke E-cadherin på kromosom 11 ser ut til å være høyere uttrykt i individer med genotype QQ i forhold til individer med genotype qq (Figur 19). Dette var imidlertid heller ikke forventet, ettersom QTLen for IPN-resistens ligger på kromosom 26 og ikke på kromosom 11.

Figur 18: Genotype qq mot genotype QQ



Figuren viser forskjellen i genuttrykk mellom genotype qq og genotype QQ av E-cadherin på kromosom 26. Genotype qq er satt til 0. Høyden på radene tilsvarer hvor mange ganger mer E-cadherin på QQ er uttrykket i forhold til qq.

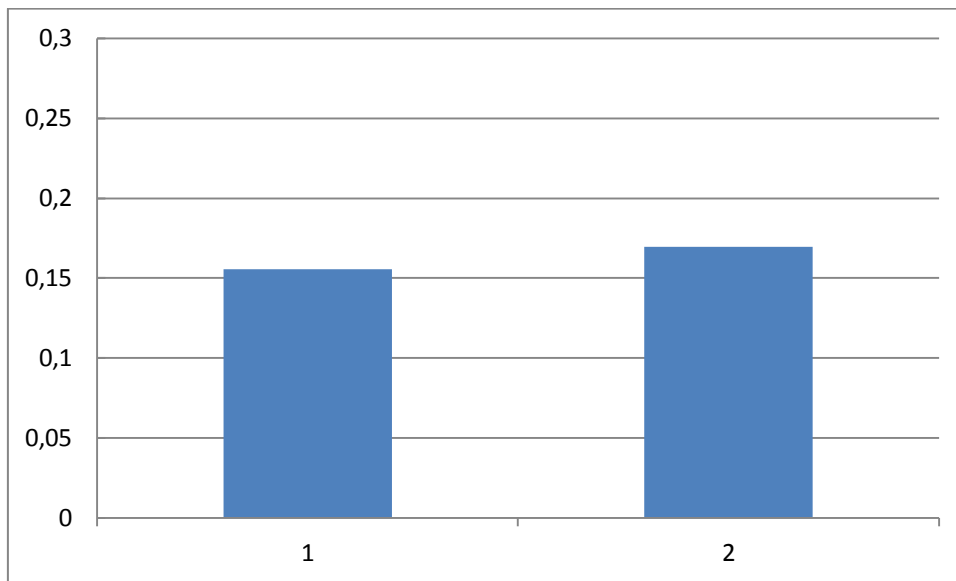
Figur 19: Genotype qq mot genotype QQ



Figuren viser forskjellen i genuttrykk mellom genotype qq og genotype QQ av E-cadherin på kromosom 11. Genotype qq er satt til 0. Høyden på radene tilsvarer hvor mange ganger mer E-cadherin på QQ er uttrykket i forhold til qq.

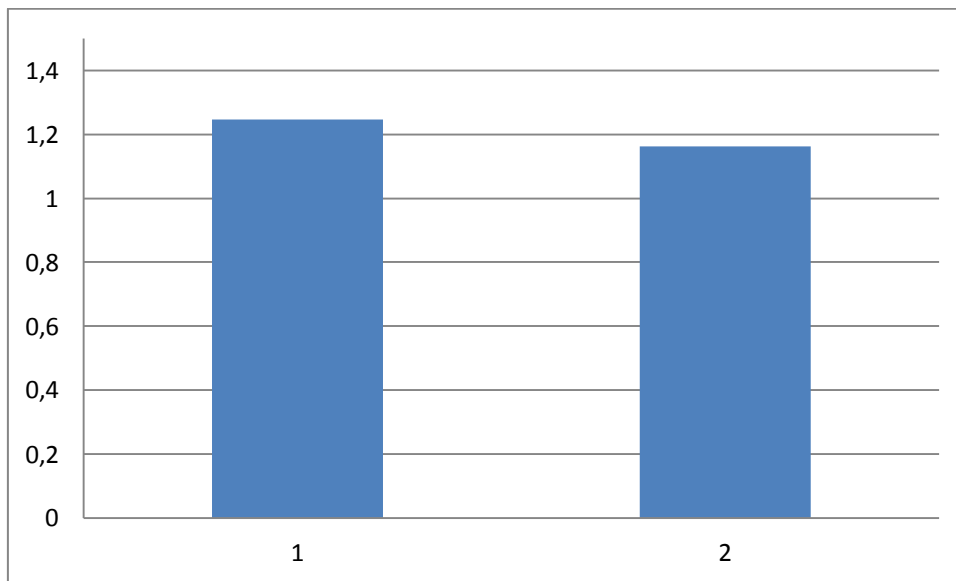
For å se om smitte av IPN vil endre aktivitetsnivået av genet E-cadherin på kromosom 26 ble individene delt opp i enda mindre grupper innenfor hver genotype. Det var to individer som hadde blitt smittet av IPN, mens to individer ikke hadde vært i kontakt med IPN. I figur 20 og figur 21 vises forskjellen mellom usmitta og smitta individer på kromosom 26. Her ligger de fleste prøvene rundt samme verdi, som viser at IPN mest sannsynlig ikke påvirker uttrykkelsen av genet E-cadherin. Dette resultatet er også basert på veldig få individer. Forsøket bør derfor gjentas med et større antall individer innenfor populasjon for å oppnå et mer signifikant resultat.

Figur 20: Smittet mot ikke smittet i genotype QQ



Figuren viser smittet mot ikke smittet i QQ for E-cadherin på kromosom 26. Usmittet er satt til 0. Høyden på radene tilsvarer hvor mange ganger mer E-cadherin i smittet er uttrykket i forhold til ikke-smittet individer.

Figur 21: Smittet mot ikke smittet i genotype qq



Figuren viser smittet mot ikke smittet i qq for E-cadherin på kromosom 26. Usmittet er satt til 0. Høyden på radene tilsvarer hvor mange ganger mer E-cadherin i smittet indivd er uttrykket i forhold E-cadherin til ikke-smittet individer.

4. Oppsummering

I denne oppgaven ble det bekreftet av E-cadherin og dens paraloge gen på kromosom 11 er aktive. Det ble dannet en sekvens for hver genotype til E-cadherin på kromosom 26 og en samla sekvens for det paraloge genet på kromosom 11. Det ble identifisert en mulig funksjonell variasjon mellom genotype QQ og genotype qq som kan forårsake endringer i proteinstrukturen. Det ble ikke observert noen form for forskjeller i ekspresjonen av E-cadherin mellom de to genotypene. Det er fortsatt mulighet for at det eksisterer andre variasjoner mellom disse genotypene ettersom konsensussekvensen for hver av dem var basert på sekvensmaterial som varierte i kvalitet. I tillegg ble det benyttet et mindre antall individer per genotype. Derfor er det mulighet at denne SNP'en mellom de to genotypene ikke er ekte og at man ved en tilfeldighet har valgt prøver med samme variasjon. Eksperimentet bør derfor gjentas for å bekrefte resultatet ved bruk av et større antall individer.

Denne SNP'en er en lovende kandidat som en mulig kausal variasjon for genetiske forskjeller i resistens mot IPN i laks. Punktmutasjonen er i posisjon 325 i proteinsekvensen. Denne forårsaker et skifte fra serine til proline i aminosyresekvensen. Disse aminosyrene har forskjellige egenskaper som kan forårsake endringer i proteinsekvensen. I tillegg er det mulighet for at punktmutasjonen er i et sensitivt område i enden av EC2-domenet og i begynnelsen av EC3-domenet. Videre studie av strukturen til proteinet og strukturen for de ulike genotypene, kan bidra til å finne ut om denne mutasjonen forårsaker endringer i reseptor egenskapene.

5. Framtidig perspektiv

Denne masteroppgaven har beskrevet variasjonen mellom genotype QQ og genotype qq i laks. Det ble detektert en SNP i posisjon 325 i aminosyresekvensen som er en punktmutasjon og som kan være ansvarlig for variasjonen mellom genotypene. Dette må bekreftes med et større antall individer med bruk av andre funksjonelle og bioinformatiske metoder.

Første fase vil være å sammenligne nedarvingsmønsteret av SNPen og QTLen. Om de følger det samme nedarvingsmønsteret kan dette gi en indikasjon på at dette er en mulig kausativ mutasjon. Eksperimentet kan bestå av å genotype SNPen i et stort materiale som består av dyr fra en smittetest for IPN. I tillegg vil man genotype både foreldre og avkom for hvert enkelt dyr. Disse dyra blir delt inn i søskengrupper og en kan dermed følge nedarvingen av både QTLen og SNPen for tre generasjoner. Her kan man bestemme om nedarvingen av SNPen følger samme mønster som nedarvingen av QTLen.

Neste steg blir å bevise at dette er den SNPen som er ansvarlig for QTLen. Her kan man teste ut den mest sannsynlige hypotesen, at E-cadherin fungerer som reseptor for viruset. Man kan bruke ulike funksjonelle analyser for å se på hvordan E-cadherin interagerer med overflateproteinet til viruset. For eksempel kan man se om de to ulike formene av E-cadherin (med serin eller prolin) virker som reseptor for viruset, ved hjelp av såkalt fluorescence resonance energy transfer (FRET).

Dersom dette er den kausative mutasjonen kan man bruke denne SNPen for å direkte avle for denne resistente egenskapen hos laks. Dette vil i så fall være et mer presist verktøy i arbeidet mot IPNV. Resultatet blir trolig store økonomiske fordeler samt at man sparer laksen for store lidelser forårsaket av dette viruset.

6. Referanser

Alberts B, J. A., Lewis J, et al. (2002). Molecular Biology of the Cell. 4th edition. . New York, Garland Science.

Altschul, S. F., et al. (1990). "Basic local alignment search tool." J Mol Biol 215(3): 403-410.

Aquagen (2012). "Feltresultater av QTL-innOva IPN - fra rogn til slakt." Retrieved 1, from <http://aquagen.no/filestore/01-2012FeltresultateravQTL-innOvalPN-frarogntilslakt.pdf> (Lest: 09.05.2013).

Bentley, D. R., et al. (2008). "Accurate whole human genome sequencing using reversible terminator chemistry." Nature 456(7218): 53-59.

Bergh, Ø. and T. Poppe (1999). Fiskehelse og fiskesykdommer. Oslo, Universitetsforl.

Berx, G., et al. (1995). "E-cadherin is a tumour/invasion suppressor gene mutated in human lobular breast cancers." EMBO J 14(24): 6107-6115.

Bessesen, T. (08.04.2013). "NORGE – VERDENSLEDENDE PÅ LAKS." from <http://laksefakta.no/nokkelinfo.html> (lest 29.04.2013).

Bhangale, T. R., et al. (2006). "Automating resequencing-based detection of insertion-deletion polymorphisms." Nat Genet 38(12): 1457-1462.

Brasch, J., et al. (2012). "Thinking outside the cell: how cadherins drive adhesion." Trends Cell Biol 22(6): 299-310.

Chenna, R., et al. (2003). "Multiple sequence alignment with the Clustal series of programs." Nucleic Acids Res 31(13): 3497-3500.

Davidson, W. S., et al. (2010). "Sequencing the genome of the Atlantic salmon (*Salmo salar*)."
Genome Biol 11(9): 403.

Davidson, W. S., et al. (2010 A). "Sequencing the genome of the Atlantic salmon (*Salmo salar*)."
Genome Biol 11(9): 403. Som har sitert til: (Hartley SE. The chromosomes of salmonid fishes. Biol Rev. 1987;1962:1197–1214. doi: 1910.1111/j.1469-1185X.1987.tb00663.x.) (King T, Verspoor E, Spidle AP, Gross R, Phillips RB, Koljonen M-L, Sanchez JA, Morrison CL. In: *The Atlantic Salmon: Genetics, Conservation and Management*. Verspoor E, Stradmeyer L, Nielsen JL, editor. Oxford: Blackwell; 02007. Biodiversity and population structure. pp. 00117–00166.)

Davidson, W. S., et al. (2010 B). "Sequencing the genome of the Atlantic salmon (*Salmo salar*)."
Genome Biol 11(9): 403. Som har sitert til: de Boer JG, Yazawa R, Davidson WS, Koop BF. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. BMC Genomics. 2007;2008:2422. doi: 2010.1186/1471-2164-2008-2422.

Dobos, P. (1995). "The molecular biology of infectious pancreatic necrosis virus (IPNV)."
Annual Review of Fish Diseases 5(0): 25-54.

Duncan, R., et al. (1991). "Sequence analysis of infectious pancreatic necrosis virus genome segment B and its encoded VP1 protein: a putative RNA-dependent RNA polymerase lacking the Gly-Asp-Asp motif." Virology 181(2): 541-552.

Duncan, R., et al. (1987). "Synthesis of the infectious pancreatic necrosis virus polyprotein, detection of a virus-encoded protease, and fine structure mapping of genome segment A coding regions." J Virol 61(12): 3655-3664.

Endo, K., et al. (2001). "E-cadherin gene mutations in human intrahepatic cholangiocarcinoma." The Journal of Pathology 193(3): 310-317.

Ewing, B. and P. Green (1998). "Base-calling of automated sequencer traces using phred. II. Error probabilities." Genome Res 8(3): 186-194.

Ewing, B., et al. (1998). "Base-calling of automated sequencer traces using phred. I. Accuracy assessment." Genome Res 8(3): 175-185.

Fiskeri-og-havbruksnæringens-forskningsfond. (2012). Utredning: Hvordan kan kartleggingen av laksens genom bidra til å løse utfordringene i norsk havbruksnæring? Oslo. s. 35, Fiskeri- og havbruksnæringens forskningsfond.

Flicek, P., et al. (2010). "Ensembl 2011." Nucleic Acids Res.

Food-and-Agriculture-Organization (2012). The state of world fisheries and aquaculture. O. Barbaroux, G. Bizzarri, M.R. Hasan, L. Miuccio, J. Saha, J. Sanders, J. Spaul and J. Van Acker. Rome, FAO Fisheries and Aquaculture Department. <http://www.fao.org/docrep/016/i2727e/i2727e.pdf> (lest 29.04.2013).

Gasteiger, E., et al. (2003). "ExPASy: The proteomics server for in-depth protein knowledge and analysis." Nucleic Acids Res 31(13): 3784-3788.

Gordon, D., et al. (1998). "Consed: a graphical tool for sequence finishing." Genome Res 8(3): 195-202.

Green, P. (1994). "DOCUMENTATION FOR PHRAP AND CROSS_MATCH." from <http://www.phrap.org/phredphrap/phrap.html> (Lest: 09.02.13).

Gumbiner, B. M. (2005). "Regulation of cadherin-mediated adhesion in morphogenesis." Nat Rev Mol Cell Biol 6(8): 622-634.

Handeland, S. O., et al. (2013). "Physiology during smoltification in Atlantic salmon: effect of melatonin implants." Fish Physiol Biochem.

Hastein, T. and J. Krogsrud (1976). "Infectious pancreatic necrosis: first isolation of virus from fish in Norway." Acta Vet Scand 17(1): 109-111.

Havforskningsinstituttet (2009). "Biologi hos laks i oppdrett." (http://www.imr.no/temasider/fisk/laks/laks_i_oppdrett/nb-no (Lest: 04.05.2013)).

Havforskningsinstituttet (2009). "Lakseoppdrett." from <http://www.imr.no/temasider/akvakultur/lakseoppdrett/> (lest: 29.04.2013).

Heppell, J., et al. (1995). "Characterization of the small open reading frame on genome segment A of infectious pancreatic necrosis virus." J Gen Virol 76 (Pt 8): 2091-2096.

Hill, B. J. and K. Way (1995). "Serological classification of infectious pancreatic necrosis (IPN) virus and other aquatic birnaviruses." Annual Review of Fish Diseases 5(0): 55-77.

Hjalmarsson, A., et al. (1999). "Infectious pancreatic necrosis virus: identification of a VP3-containing ribonucleoprotein core structure and evidence for O-linked glycosylation of the capsid protein VP2." J Virol 73(4): 3484-3490.

Houston, R. D., et al. (2008). "Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*)." Genetics 178(2): 1109-1115.

Hulpiau, P. and F. van Roy (2009). "Molecular evolution of the cadherin superfamily." Int J Biochem Cell Biol 41(2): 349-369.

Invitrogen (2008). Real-Time PCR: From theory to practice.

Johansson, J. (1997). Laks. [Oslo], Landbruksforl.

Kelly, R. K. and P. C. Loh (1972). "Electron microscopical and biochemical characterization of infectious pancreatic necrosis virus." J Virol 10(4): 824-834.

Kimball, J. W. (1994). "Kimball's Biology Pages." from <http://users.rcn.com/jkimball.ma.ultranet/BiologyPages/> (Lest: 12.05.2013).

Krogsrud, J., et al. (1989). Infectious Pancreatic Necrosis Virus in Norwegian Fish Farms. Viruses of Lower Vertebrates. W. Ahne and E. Kurstak, Springer Berlin Heidelberg: 284-291.

Li, X., et al. (2013). "Mutation screen and RNA analysis disclose the changed splicing of the E-cadherin transcription in gastric cancer." Familial Cancer: 1-8.

Lien, S., et al. (2011). "A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns." BMC Genomics 12: 615.

Life-Technologies. "Immakulerinsreaksjon." from <http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/dna-sequencing-fragment-analysis/overview-of-dna-sequencing/sequencing-chemistries.html> (lest: 30.04.13).

Life-technologies (2012). "SYBR® Select Master Mix." from http://tools.invitrogen.com/content/sfs/manuals/cms_103305.pdf (lest: 01.03.2013).

Lodish, H. (2008). Molecular cell biology. New York, Freeman.

M'Gonigle, R. H. (1941). "Acute Catarrhal Enteritis of Salmonid Fingerlings." Transactions of the American Fisheries Society 70(1): 297-303.

Mengaud, J., et al. (1996). "E-Cadherin Is the Receptor for Internalin, a Surface Protein Required for Entry of *L. monocytogenes* into Epithelial Cells." Cell 84(6): 923-932.

Moen, T., et al. (2009). "Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): population-level associations between markers and trait." BMC Genomics 10: 368.

Mott, R. (1997). "EST_GENOME: a program to align spliced DNA sequences to unspliced genomic DNA." Comput Appl Biosci 13(4): 477-478.

Mount, D. W. (2007). "Using the Basic Local Alignment Search Tool (BLAST)." Cold Spring Harbor Protocols 2007(7): pdb.top17.

Møller Christensen, J. (1977). Fiskeliv i Nordsjøen. [Oslo], Cappelen.

Nagafuchi, A. and M. Takeichi (1988). "Cell binding function of E-cadherin is regulated by the cytoplasmic domain." EMBO J 7(12): 3679-3684.

Okonechnikov, K., et al. (2012). "Unipro UGENE: a unified bioinformatics toolkit." Bioinformatics 28(8): 1166-1167.

Olsvik, P. A., et al. (2005). "Evaluation of potential reference genes in real-time RT-PCR studies of Atlantic salmon." BMC Mol Biol 6: 21.

Ozaki, C., et al. (2010). "The extracellular domains of E- and N-cadherin determine the scattered punctate localization in epithelial cells and the cytoplasmic domains modulate the localization." J Biochem 147(3): 415-425.

Ozawa, M. and R. Kemler (1990). "Correct proteolytic cleavage is required for the cell adhesive function of uvomorulin." The Journal of Cell Biology 111(4): 1645-1650.

Phillips, R. B., et al. (2009). Assignment of Atlantic salmon (*Salmo salar*) linkage groups to specific chromosomes: conservation of large syntenic blocks corresponding to whole chromosome arms in rainbow trout (*Oncorhynchus mykiss*). BMC Genet. 10: 46.

Pizarro-Cerdá, J., et al. (2012). "Entry of *Listeria monocytogenes* in Mammalian Epithelial Cells: An Updated View." Cold Spring Harbor Perspectives in Medicine 2(11).

Qiagen (2004). "Critical Factors for Successful Real-Time PCR." from <http://www.icmb.utexas.edu/core/DNA/qPCR/QiagenRT-PCR.pdf> (Lest: 09.05.2013).

Qiagen (2005). "Quantitect Reverse Transcription Handbook." from http://aquacul4.fish.washington.edu/Protocols:Information%20Sheets/Commercial%20Protocols:Manuals/Qiagen%20-%20QuantiTect_Reverse_Transcription_Handbook.pdf (lest:20.02.2013).

Qiagen (2006). "RNeasy® Mini Handbook." from http://labs.fhcrc.org/fero/Protocols/RNeasy_Mini_Handbook.pdf. (Lest 30.04.2013).

Qiagen (2010). "HotStarTaq® Plus PCR Handbook ". from <http://www.qiagen.com/resources/Download.aspx?id=%7BC505B538-7399-43B7-AD10-D27643013D10%7D&lang=en&ver=1>. (Lest: 09.05.2013).

Qiagen (2010). "TissueLyser Handbook." from http://www.nhm.ac.uk/resources-rx/files/qiagen-tissuelyser_handbook-95094.pdf (lest: 30.04.2013).

Saiki, R. K., et al. (1985). "Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia." Science 230(4732): 1350-1354.

Saito, M., et al. (2012). "Classical and desmosomal cadherins at a glance." J Cell Sci 125(Pt 11): 2547-2552.

Sandtrø, A. (2011) IPN-viruset og virulensfaktorer. Tekna Fiskehelse

Sanger, F., et al. (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A 74(12): 5463-5467.

Sjøberg, N. O. (2006). Molekylær genetik: genteknologi, humant DNA. Nesbru, Vett & viten.

Skjesol, A., et al. (2011). "IPNV with high and low virulence: host immune responses and viral mutations during infection." Virologia 8: 396.

Stephens, M., et al. (2006). "Automating sequence-based detection and genotyping of SNPs from diploid samples." Nat Genet 38(3): 375-381.

Storset, A., et al. (2007). "Response to selection for resistance against infectious pancreatic necrosis in Atlantic salmon (*Salmo salar* L.)." Aquaculture 272, Supplement 1(0): S62-S68.

Takeichi, M. (1988). "The cadherins: cell-cell adhesion molecules controlling animal morphogenesis." Development 102(4): 639-655. Som har sitert til : (YOSHIDA, C. & TAKEICHI, M. (1982). Teratocarcinoma cell adhesion: Identification of a cell surface protein involved in calcium-dependent cell aggregation. *Cell* 1928, 1217-1224.) (YOSHIDA-NORO, C, SUZUKI, N. & TAKEICHI, M. (1984). Molecular nature of the calcium- dependent cell-cell adhesion system in mouse teratocarcinoma and embryonic cells studied with a monoclonal antibody. *Devi Biol.* 1101, 1919-1927).

Untergasser, A., et al. (2012). "Primer3--new capabilities and interfaces." Nucleic Acids Res 40(15): e115.

Venkatesh, B. (2003). "Evolution and diversity of fish genomes." Current Opinion in Genetics & Development 13(6): 588-592.

Veterinærinstituttet (2012). Fiskehelserapporten 2011. Oslo s. 13.

Veterinærinstituttet (2013). Fiskehelserapporten 2012. Oslo s. 7-14, Veterinærinstituttet.

Yasuike, M., et al. (2010). "Evolution of duplicated IgH loci in Atlantic salmon, *Salmo salar*." BMC Genomics 11: 486.

Appendix:

I. Primersett 1.

Tabell 2: Primersett 1

Navn	Retning	Sekvens	Bindingspunkt	Lengde	TM	GC Prosent	"Mismatch"	Primerpar "Penalty"
Primer 1F	Framover	TTCGTGGAGTTGGGCGTATT	121	20	59,678	50	2	1,46
Primer 1R	Revers	AGGGCTCGGTCACATAGAGAT	778	21	60,133	52	3	
Primer 2F	Framover	CATGGCCACCACCACAATCA	466	20	60,899	55	7	1,02
Primer 2R	Revers	CGTTACCTGCATGACCTCGT	975	20	60,109	55	3	
Primer 3F	Framover	AGCTGAGGATCCCATGGAGA	843	20	59,733	55	3	0,86
Primer 3R	Revers	CGTTGAAAAGGCCTTGAGGG	1393	20	59,402	55	3	
Primer 4F	Framover	TGGGAACAGTTCCTGAAGCA	923	20	59,157	50	2	0,88
Primer 4R	Revers	CACTGTCAACAGGGAGGTCC	1630	20	59,965	60	3	
Primer 5F	Framover	CCCTCAAGGCCTTTTCAACG	1374	20	59,402	55	3	1,21
Primer 5R	Revers	TGCAGACCCTGATCACACTC	1927	20	59,39	55	4	
Primer 6F	Framover	GGACGACACAAAGACTGGCA	2043	20	60,532	55	4	0,57
Primer 6R	Revers	GTTGAGGGAAGTGGAGGAGC	2652	20	60,036	60	2	
Primer 7F	Framover	CCAGAAAGCACACGTTTGAA	5	20	57,158	45	-1	2,88
Primer 7R	Revers	GACAGACTTGCTCCTGGCTT	527	20	59,94	55	3	
Primer 8F	Framover	GCTCCCTCAGTTCCTCAAC	2633	20	60,036	60	2	2,14
Primer 8R	Revers	AGCAGAACTGTGAGCTTCTTCA	3311	20	59,896	45	-1	
Primer 9F	Framover	GCCAAATCAGCCTTACGATGT	3140	21	58,983	47	3	2,28
Primer 9R	Revers	AAGGCACCTCAGGGTTTTGT	3682	20	59,738	50	-1	
Primer 10F	Framover	GTGACTGCAAAGGAGCGGA	2169	19	60,302	58	2	1,49
Primer 10R	Revers	TGGTCCCATGTCTGCACAAA	2902	20	59,815	50	4	
Primer 11F	Framover	ACAATCAGCATGCCACCTC	1505	20	60,682	55	7	1,36
Primer 11R	Revers	ACATCCGCTCCTTTGCAGTC	2190	20	60,675	55	3	
Primer 12F	Framover	CCCGTCTACGCGTAAACTGA	4	20	59,831	55	-1	0,42
Primer 12R	Revers	GTCCAAAGGCTGGGTCACAT	747	20	60,251	55	2	
Primer 13F	Framover	CCCGTCTACGCGTAAACTGA	4	20	59,831	55	-1	2,25
Primer 13R	Revers	TTGCCAGTGATGCTGTACATG	668	21	58,914	47	3	

Primer 14F	Framover	GCATGAGGCCAGACACTACC	423	20	60,463	60	4	
Primer 14R	Revers	CATCACCTGGTTTTGCTGCC	922	20	60,038	55	3	0,5
Primer 15F	Framover	GTAAGGCTGAGGCACCCATG	803	20	60,749	60	4	
Primer 15R	Revers	TGCTCACACTGAAAAGGCCT	1363	20	59,817	50	2	0,93
Primer 16F	Framover	GCTGCTGATTTGGAGGGAGA	1120	20	59,746	55	3	
Primer 16R	Revers	TGAGCCATCCAGCAACATCA	1684	20	59,671	50	2	0,58
Primer 17F	Framover	TCCATTTGCCATCCCTCTG	1463	20	59,47	55	7	
Primer 17R	Revers	TCCAGTTGGAAAGGGACGAC	1996	20	59,605	55	2	0,65
Primer 18F	Framover	GACCTGAGGACCTCCATGTG	1568	20	59,461	60	3	
Primer 18R	Revers	TATTGGCGCACTGGACATCA	2170	20	59,748	50	3	0,79
Primer 19F	Framover	TGACAACCTCCAACCATCG	1848	20	59,964	55	4	
Primer 19R	Revers	CGGACGGTTATCCAGACCTC	2412	20	59,615	60	2	0,42
Primer 20F	Framover	TGATGTCCAGTGCCCAATA	2151	20	59,748	50	3	
Primer 20R	Revers	CCACGGCTGCACAAATCAAA	2861	20	59,968	50	8	0,28
Primer 21F	Framover	CCAACCCAGCCGACATAGG	2474	19	60,153	63	2	
Primer 21R	Revers	AGAAGCAAGAACCAGCAAACA	3045	21	58,622	43	9	3,53
Primer 22F	Framover	CCATCTTGTGTGTTTCGGGC	2812	20	59,759	55	5	
Primer 22R	Revers	GGCCATCAATGTACCAAAGGG	3313	21	59,243	52	-1	2
Primer 23F	Framover	CCATCTTGTGTGTTTCGGGC	2812	20	59,759	55	5	
Primer 23R	Revers	ATTCACATGGCCATCAATGTAC	3321	22	57,076	41	-1	5,16

Tabellen viser en oversikt over Primersett 1 med primer "stock" konsentrasjon på 200 pmol/µl (Metabion, Martinsried, Tyskland). Bindingspunktet viser første base som primeren binder til. Lengden representerer hvor mange nukleotider primeren består av. T_M representerer smeltetemperaturen til primerene og GC viser antall g og c i sekvensen. "Mismatch" viser antall baser som er forskjellige mellom de paraloge sekvensene. Primerpar "penalty" går fra 0 og opp. 0 er den optimale primeren og desto høyere den blir detso dårligere tilpasset er den til forholdene i reaksjonen.

II. Primersett 2

Tabell 3: Primersett 2

Navn	Kromo- som	Retning	Sekvens	Binding s punkt	Lengd e	TM	G C %	Mismatch	Primerpa r Penalty	Produkt lengde
Primer 1	26	Framove r	TGGAGATGAGCCTCACTCC T	1317	20	59,66 2	55	2	0,59	139
Primer 2	26	Revers	CTCAAAGTCAAGCGGCTTG G	1455	20	59,76 1	55	3		
Primer 3	11	Framove r	TTGCTGTGGGTGCAGGTAA G	788	20	60,53 7	55	2	0,59	135
Primer 4	11	Revers	CATCACCTGGTTTTGCTGCC	922	20	60,03 8	55	3		

Tabellen viser en oversikt over Primersett 2 med primer "stock" konsentrasjon på 200 pmol/μl (Metabion, Martinsried, Tyskland). Bindingspunktet viser første base som primeren binder til. Lengden representerer hvor mange nukleotider primeren består av. T_M representerer smeltetemperaturen til primerene og GC viser antall g og c i sekvensen. "Mismatch" viser antall baser som er forskjellige mellom de paraloge sekvensene. Primerpar "penalty" går fra 0 og opp. 0 er den optimale primeren og desto høyere den blir detso dårligere tilpasset er den til forholdene i reaksjonen.

III. Buffer og løsninger

Tabell 4: Buffere

Reagent	Innhold
400 ul 1X TE-Buffer	400 ul Tris-EDTA Buffer 100x Concentrate 40 mL destilert H ₂ O
TAE buffer, 50X	160 mL TAE buffer, 50x Resten dH ₂ O Endelig volum 8L

Tabellen viser buffere brukt i ulike reaksjoner.

IV. Referansesekvensen til cDNA av E-cadherin på kromosom 26

>E-cadherin kromosom 26

TTGGCCAGAAAGCACACGTTTGAAAAATAGACTTTTCTCATCTACACGTAAGGGATTACTTGTACTATTGGTGGAACGACT
ATCCAAAAATTATTCAAAAACAAATGGGGGCATTTGGTTCGTGGAGTTGGGCGTATTAATATTATTTCTCCAGGCTTTCAAAC
CAGGGTCGTCGGAGTCAACATGTCTACACAGGTTTCAATTCAGAAATATACATTTTCAAAGTGGAAAGAAATCACTTACAAAGT
GGCCGGAGACTAGGAAAAGTGGTTTTTGGATGACTGCACCAGCCGACCAGCTTTCTTTTCACTCCGAGGATTCACACTTCAA
AGTAGATGGCGACGGGACGCTGAAACTGGAGAGGGGGGCTGACTCTGCATAATGGACATAAGGAGGTCTATGTCTTACCCAG
TCCAAGGGCAAGAAGATCACGGTTCCAGTCAGAGTGTGCATGAGGCCAGACATGGCCACCACCACAATCACCATGAGATGA
CCACCAGCCCAAGCCAGGAGCAAGTCTGTCTCTACCTGTCTGAACTTCCCAAGTCTTCAGGAGGTCTGAAAAGAAGGAA
GAGGGACTGGGTCAATCTCCCATCAACTTCCAGAGAATGACCCGAGGCCCTTCCCAAGATTATGGTGCAGATCAGGTCCA
ACAATGATAAAGAGGTGAAGATCCAGTACAGCATCACTGGCACTGGGGCAGACCTACCTCCTGTGGGAATCTTCACTGTGGA
CAAAAACCTCTGGGAATCTCTATGTGACCGAGCCCTGGACAGGGAGAAAAAAGACAAATACATTCTCCTAGCCCATGCTGTT

GCAGTGGGTGCAGGTATAGCTGAGGATCCCATGGAGATCATTGTGAAAAGTCATCGATATGAATGACAACAAACCTGTATTTC
CCCAAGATCCATTTATGGGAACAGTTCTGAAGCATCAAAACCAGGTGACGAGGTCATGCAGGTAACGGCCACTGATGCTGA
TGAGGAGGCTCACTCCCTGCTTGGCCACCACCTACAAGATAGTTGACGGGGACCCCTCAAGGCCTTTTCAACGTGAGCACAG
TCAACTCTGTGACTGGAGGGATTGGGTTAATGCACCTGGGTTGGACAGAGAGAAAATTCCAAATATACTTTGGCAATCCA
AGCTGCCGATATGGAGGGAAATGGCCTTACCAGCTTTGGCAAAGCCATCATTACACTAACGGACAGCAATGACAACGCGCCA
CAGTTTGTGACGCCTTCGTACACCGTGTGAGTCCAGAGAATAAAGTGGATGCCTTGGTGGTAAAAATGCCAGTGACTGATGG
AGATGAGCCTCACTCCCTGCTTGGCCACCACCTACAAGATAGTTGACGGGGACCCCTCAAGGCCTTTTCAACGTGAGCACAG
GCCCTAGCAAGCTGGAGGGCATCATTACAACAGCCAAGCCGCTTGACTTTGAGAAGAACAACAAGTACTACTCTGCTGGTCCAC
TGTGACAGAAATGAAGTCCATTACAATCAGCATGCCACCTCCACTGCTACTGTTGTAGTGAATGTGGAGGATGTGAATGAAG
CTCCAGTCTTACCCCCAGTGGAGAAGATTATCAGGAAACCTGAGGACCTCCCTGTTGACAGTGAACCTGGTTCTGTACACAGCC
ACAGACCCAGACACCGAAGGAATCAGAAAAGTCACATACAAGATACGCAATGATAATGCTGGATGGTCCAGTATCAACAAA
GACACTGGGCTGATCAAAAGTCAAGAGCCTCATGGACAGAGAATCCACTTTTGTCCAAGACAACAATACTGTTATTGTTCT
GGGCATCGACAACGATGAAATTCAGCAACTGGCACTGGCACCTTATCATAGAGCTGGAGGATGTGAATGATAATGCTCCA
ACCATTGACGAGAGTGTGATCAGGGTCTGCAACAAGGAGTCCCTCCACAGTTGTTGTCAGTCACTGATAAAGATGGTGGC
GCTTCACTGCTCCATACACCGTACAGCTTCAGGGTCTGCCATTCTAACTGGACTGCCAGAATGGACGACACAAAAGACTGGC
ATTACTGACTCTGAAGACTATGTTGGACAGTGAATATACACGGTTGCTGAGAGTGTCTGAGAGTGTCTGACAAACCAGGGCTGACCA
TGACAGCACCATCCAGGCCTCCGTCTGTGACTGCAAAGGAGCGGATGTCCAGTGTCCGATAAAGCTGTAGCAGGCTTCGGC
ATCTCTAGCATTCTGGGAATTCTAGGAGCAGTCTTACTACTCCTATTGTTGTCTTCTGCTGCTCATGTTCTGAGGAAGAGA
GGTGGCGAGAAGAAGGAGCCTCTGCTGCAGGAGGACGATGTGAGGGACAACATCTACTACTATGACGAGGAGGGAGGTGGC
GAGGATGACCAGATTTGACTTGGAGCCTGTCACAGAGGTTGGATAAACCCTGCTGATGTTTCCGTAATGACATCGCTCC
AACCATGGCCCGGCCAGAGTATCGTCCACGACCCGCCAACCCAGCCGACATTGGCAACTTCATTGATGATAACCTGAAGGCA
GCTGACAACGACCCCACTGCTCCTCCCTACGACTCCCTCCTGGTGTGTTGACTATGAAGGAGGTGGCTCTGAGGCTGGCTCCCT
CAGTTCCTCAACTCTCCAGCTCAGGAGACGACAGGACTACGACCTCCTTCAAGAGTGGGGCCCGCTTCAAGAAGCTGT
CTGACATGTACGGAGGAGGAGATTGAGAAGTCAGTCAGCCACTAAACCTTTTTGCTTGGAACTGTCGGTTCACAC
TTTGGTCTGTAACCTGAAGACCTTTTCTTTATTTCCCTTTTTATGCGTTCGGGCACATTTACAGTTTGGCAATTTGATTTGTG
CAGACATGGGACCATTTAGAAAAAATGTGTTCAATGCTCGTCTCAGCATGGGGAGGGTGGCATACTCTGGCTCTGACTA
GCAGGATGTTGACATGGATTTTACACAATTCAGCAGTCTTTCTAATGGACTCTTAAACAACCCAGATTGTAATTTAAT
ATGCTCTTTTGGCGTGTATCGAAGTCTTTTTTTTTATATATGACTATCCTTTGGCATGAAATGAAGTACTGAAAGCC
AAATCAGCCTTACGATGTTTTGTTAACTGAGATTAATATACAGCTCAATAAATGCTTTTTAAAAATGTTTGATAAAACA
AGTTTGGGTTGTTTTATTCCTTTTCAATTTAATCATATTAATCACTAATTATATTTTTTTCTGAAGAAGCTCACAGTCTGC
TTTTATTTTATTACAGTCTGCTTTTTTAAATTAAGATATCCTTTTGGTACAATGATGGATGTGAATATTTGTATTAAATACA
TTGTAATTTGCTTATTTGAGCTCAATTAATCAAATTCACAGCTGAGACTAGATGGCGACAATTGAATAAAACAGGTCTT
GGTCAATGAATGAGATTTTCTTGTCCCGCAGAACAGATTGTACAAATTAATTTAAGTAATAAGGCTGGAGGGGTGTGG
TGTATGTGGCAACATATCTCGGCTAGGGGCTGACTTATGCACGATGCAACACGGAGTGCCTGGACAGAGCCCTTAGCCGTG
GCTTATTGGCCATATACCACAAAACCTGAGGTGCCTACTGCTATTATAAACTGGTTACCAATGTAATTGGAGCAATAAAAA
TGTGCTTGGC

V. Referansesekvensen til cDNA av det paraloge genet på kromosom 11.

>E-cadherin kromosom 11

CCTCCCGTCTACGCGTAAACTGATTACTTTTTACTGTTGGTGGAAACGACTATTCTGAAATTATTCAAAACAAATGGGGCCCTTT
TGGTTCGTGGAGTTGGGGTTATTAACATTGTTTCTCCAGGCTTCAAGCCAGGGTCTGCGGAGGAGTCAAAATGTATACCAGG
TTTCAATCCGAAATGTACATTTTCAAAGTGGAAAGAAATCACTTACAAGTGGCCGGAGACTAGGAAAAGTGGTTTTTGTGAT
GACTCACCGCCGACCACTTTCTCTTCACTCCGAGGATTAACGCTTCAAAGTAGATGGCGACGGGACGCTGAAACTGAA
GAGGGGGGTGACTCTGCATAATGGACATAAGGAGTCTATGTCTTACCCAGTCCATGGGCAAGAAGATCACTGTTCCAGTC
AGAGTGTTCATGAGGCCAGACACTACCACATCACCATGAGATGACCACCCAGCCAGCCTGGATCAAGTCTGTCTTACC
ATGAGCGAGGCCCCTTCCCAAGATTATGGTGCAGATCAGGTCCAGCAATGATAAAAGAGGTGAAGATCATGTACAGCATCAC
TGGCAATGGGGCAGACCAACCTCCTGTGGGACTTTCAGTGTGGACAAAACCTCTGGGATTCTATGTGACCCAGCCTTTGG
ACAGGGAGAAAAAGACAAATACATTCTCCTAGCCATGCTGTTGCTGTGGGTGCAAGGTAAGGCTGAGGCACCCATGGAGGT
CATTGTGAAAGTCACTGACCAAAAATGACAACAAACCTGTATTTACCAAAATCCATTTATGGGAACAGTCCCTGAGGCAGCA
AAACCAGGTGATGAGTTCATGCAGGTAACAGCCACTGATGCTGATGAGGAGGGCTCTGCAATTTGATGTACAGATACACCA
TTCTCAGTCAGGAGCCTCCACTACCAAGCCCCAACATGTTTGTATCAACCCTGTGACTGGAGGGATTCCGGTTAACGCACCT
GGGTTGGACAGAGAGAAAATCTTAAATATACTTTGGAAATCCAAGCTGCTGATTTGGAGGGAGATGGCCTTACCAGCTTTG
GCAAAGCCATCATTACAGTAACGGACAGCAATGACAACGCGCCACAGTTTGTGACGCTTCTGTACTGCTGCTGCTCCAGA
GAATAAAGTATGATGCTTGGTGGTAAAAATGCCAGTACTGATGGGATGATGCTACTCCTCTGCTGGCCACCACCTTCA
AGATAGTTGACGGGGACCTTGAAGGCTTTTCACTGTGAGCAGGGCCTAGCAAGCTGGAAGGCTCATTACGACAGAAAA
GCCCTTACTTTGAGAAGAACAACAAGTACACTCTGCTGGTCACTGTGGTGAATGAAGTCCCATTGGCCATCCCTCTGCCCCA
CCTCCACTGCTACTGTTATAGTGAATGTGGAGGATGTGAATGAAGTCCAGTCTTCAAACCAAGTGGAGAAGATGATCAGAAG
ACCTGAGGACCTCATGTGGACAGCGACCTGGTTCTGTACACAGCCACAGACCCAGACTGCAAGGAATCAGAAAGTCA
TACAAGATTAAAGAAATGATGTTGCTGGTCACTGATCAACAAGACACTGGGCTGATAAAAAGTCAAGACTCTCATGGACA
GAGAATCCACTTTTGTCCAAGACAACAATACTCTGTTGTTGTTCTGGGCACTGACAACGATGAAATCCCAGCAACGGGCACT
GGCACCTTATCATAGAGCTGGAGGATGTGAATGACAACCCTCAACCATCGACGAGGAGGATGATTAAGGTCTGCAACAAGG
AGTCTCCCCACAGCTGCTGTCAATCACTGATAAAGACGGAGCGGGCTTCGCTGCTCCATACACTGTACAGCTTACGGGGTCC

TCCCTTTCCAACCTGGACTGCCAGAATGAATGACTCAAAAACCTGGCATTATCCTGACACTGAAGACTATGTTGGACAGTGGAGA
TTACACGGTTGTCTGAGAGTGTCTGACAACCAGGGCCTGCACCAGGAGAGCACCATCTGGCCTCCGTGTGTGACTGCAAA
GGAGCTGATGTCCAGTGCGCCAATAAAGTTGTAGCAAGCTTCGGCCTCTTAGCATTCTGGGAATTCTAGGAGCCATTTACT
ACTCTATTGTTGTCTCTTCTGCTGCTCATGTTCTGAGGAAGAGAGGTGGTGAGAAGAAGGAGCCTCTGCTGCAGGAGGACG
ATGTCAGGGACAACATCTACTACTACGACGAGGAGGGAGGTGGCGAGGATGACCAGGATTTGACTTGAGTGTCTGCACAG
AGGTCTGGATAACCGTCCGGATGTTTTCCGTAATGACATCGCTCCAACCATGGCCCGCCAGAGTATCGCCCACGACCCGCCA
ACCCAGCCGACATAGGCAACTTCATTGATGATAACCTGAAGGCAGCTGACAACGACCCCACTGCTCCTCCCTACGACTCCCTC
CTGGTGTGTTGACTATGAAGGAGGTGGCTCTGGGGCCGGCTCTCTCAGCTCCCTCAACTCCTCCAGCTCAGGAGACGACCAGGA
CTACGACCTCCTTCAACAGTGGGGCCCGCTTCAAGAAGCTGTCTGACATGTATGGAGGAGGAGGATTGAGAAGTCAGT
CAGCCTCTCTACCTTTTTGCTTGGAAAAGCGTCCAGTTCTCCACACTTGGGTCTGTAGGCTGAAGACCCTTTTTTTAACTTTATT
TCCCATCTTGTGTGTTTCGGGCACATTTACAGTTTGATTTGTGCAGCCGTGGGAACATTTTAGAGAATTGTGTTCAATGCTCGT
CTTCAGCATGGGGAGGGTGGCACACTCTGGCTCTGTACTAGCAGGATGTTGACATAGATTTTACACAGTTCAGCAGTGCTTT
CTAATGGACTCTTAATAACCCAGATTGTAATTTAATATGCATTGTTTGTGGTCTGCTTCTATGTTGGTGCATATC
ATAAAGTTTCTATAAAGACTATACATTGGCATGTAATGTAAGTTACTTCAAATCCAAATCAGCCTTAGGATTTTTTTGTTAACT
GAGATGAATTATCAAGCTCAAACGCTTTTTAAAAATGTTTGATAAAAAATATAAGTTTGGGTCTTTTGTGTTCCTTTTTCATTTA
AATGTTTTTTGTTTCTACTTGCTTCTGAAGAAGCACTTGTGAATTTTTTTTTATAAAGGATATCCCTTTGGTACATTGATGGCCA
TGTGAATAGTTTTAATAAATAAATTACAACAT

VI. Protokoll til PCR

Material:

- cDNA
- 200 pmol/ul Primere (se tabell i 2)
- Hot StarTaq®Plus DNA Polymerase, 5 units/ul (Qiagen,Hilden,Tyskland).
- PCR Buffer, 10x (15mM MgCl₂) (Qiagen,Hilden,Tyskland).
- dNTP (200 µM)
- Destillert H₂O
- Plast tube eller 96-brønn plate (Avhengig av antall prøver)
- Is

Protokoll:

1. Et oppsett for en typisk PCR reaksjon finner man i tabell 5. Den ble satt opp på is for å forhindre at primerne festet seg til templat, og degradering av cDNA.
2. Reagensene ble blandet og sentrifugert før brettet ble plassert i en PCR maskin. Reaksjonen fulgte det termiske syklusoppsettet fra tabell 6.
3. Etter prøvene hadde gjennomgått PCR reaksjonen, ble de lagret ved -18 grader Celsius for å forhindre degradering av PCR produktet.

Tabell 5: Mastermiks for PCR reaksjon

Reagenser	"Stock" Konsentrasjon	Endelig konsentrasjon/mengde	Mengde (μL)
H ₂ O	-	-	13.7
Buffer	10X	1X	2.0
dNTP	2.5mM	200 μM	1.6
Primer R	10 μM	40pmol/100 μl	0.8
Primer F	10 μM	40pmol/100 μl	0.8
Polymerase	5U/ μl	2.5U/100 μl	0.1
Templat	1 μl	Ukjent	1.0
Total	-	-	20 μL

Tabellen viser mastermiks for PCR reaksjon.

Tabell 6: PCR program

Temperatur	Tid	Sykluser
94,0 C	5,0 min	1 syklus
94,0 C	0,30 min	30 sykluser- 35 sykluser
55,0 C	0,30 min	
72,0	0,45 min	
72,0	7,0 min	1 syklus
10	Uendelig	

Tabellen viser PCR program.

VII. Protokoll til gradient PCR:

Material:

- cDNA
- 200 pmol/ul Primere (se tabell 2)
- Hot StarTaq®Plus DNA Polymerase, 5 units/ul (Qiagen,Hilden,Tyskland).
- PCR Buffer, 10x (15mM MgCl₂) (Qiagen,Hilden,Tyskland).
- 200 μM dNTP (Qiagen,Hilden,Tyskland).
- Destillert H₂O

- Plast tube eller 96-brønn plate (Avhengig av antall prøver)
- Is
- 25mM MgCl₂ (Qiagen,Hilden,Tyskland).
- Q-Solution (Qiagen,Hilden,Tyskland).

Protokoll:

1. Et oppsett for en typisk gradient PCR reaksjon finnes i tabell 7. Den ble satt opp på is for å forhindre at primerne festet seg til templatene, og degradering av cDNA. Oppsett for variabel konsentrasjon til MgCl₂ står i tabell 8. Konsentrasjonen MgCl₂ varierer mellom 15mM til 19,5 mM. Endelig mengde H₂O vil variere avhengig av mengde MgCl₂ tilsatt, se oversikt i tabell 8.
2. Reagensene ble blandet og sentrifugert før brettet ble plassert i en PCR maskin. Reaksjonen fulgte det termiske syklus oppsettet fra tabell 9, med variasjon i temperatur ut fra tabell 10.
3. Etter prøvene hadde gjennomgått PCR reaksjonen, ble prøvene lagret ved -18 grader Celsius for å forhindre degradering av PCR produktet.

Tabell 7: Mastermiks for gradient PCR

Reagenser	"Stock" konsentrasjon	Endelig konsentrasjon/mengde	Volum (µL)
H ₂ O			Variabel
Q-Solution	5x	1x	4
MgCl ₂	25mM	Variabel	Variabel
Buffer	10X	1X	2
dNTP	2.5mM	200 µM	1.6
F-Primer	10uM	40pmol/100 µL	0.8
R-Primer	10uM	40pmol/100 µL	0.8
Polymerase	5U/ µL	2.5U/100 µL	0.1
Templat	1 µL	Unknown	1
Total	-	-	20

Tabellen viser mastermiks for gradient PCR.

Tabell 8: Mengde av MgCl₂ og H₂O.

Prøve	Endelig konsentrasjon av MgCl ₂ (mM)	Mengde MgCl ₂ (μL)	Mengde dH ₂ O (μL)
1	1.5	0	9.7
2	2.0	1.6	8.1
3	2.5	2.0	7.7
4	3.0	2.5	7.2
5	3.5	2.8	6.9
6	4.5	3.6	6.1

Tabellen viser mengde av MgCl₂ og H₂O tilsatt til en temperatur.

Tabell 9: Program for PCR

Temperatur (grader Celsius)	Tid (Minuter)	Sykluser
94,0	5,0	1 syklus
94,0	0,30	35 sykluser
46.3 -59,6	0,30	
72	0,45	
72	7,0	1 syklus
10	Uendelig	

Tabellen viser program for PCR

Tabell 10: Temperatur under amplifisering

Rad:	Temperatur (grader Celsius)
1	Blank
2	Blank
3	46.3
4	47.5
5	49.2
6	51.4
7	53.9

8	56.0
9	57.7
10	58.8
11	59.7
12	Blank

Tabellen viser en oversikt over temperaturer under amplifisering.

VIII. Protokoll til gel elektroforese:

Forberedelse av 1% /2% agarose gel:

Material:

- Tubeglass
- Gelformings sett
- Agarose (Sigma, Missouri, USA)
- TEA-buffer (Tabell 4)
- RedSafe™ Nucleic Acid Staining Solution (20,000x) (iNtRON Biotechnology, Boca Raton, USA)
- 50 µg/ml Mass DNA Ladder (New England Biolabs, Ipswich, England)
- 6x Gel Loading Dye Orange (New England Biolabs, Ipswich, England)

Protokoll:

1. Ved forberedelsen av en 100 ul stock løsning for 1 % agarose gel, ble 1 gram Agarose pulver løst opp i 100 ul 50x TAE buffer. For 2 % agarose gel innholdet av Agarose pulver dobbelt.
2. Løsning varmes deretter opp i mikrobølgeovn til den har nådd kokepunktet, dette for å løse opp Agarose pulvere i TAE bufferen. Deretter kjøles løsningen ned i romtemperatur til den har nådd ca. 60 grader Celsius.
3. Det ble tilsatt 1 ul Red Safe per 100 ml gel, blandet godt, også tilsatt i en gelform. Den ble liggende i ca. 20-30 min slik at gelen stivnet.

4. Agarose gelen ble lagt i elektroforese kammeret med tilsatt TAE buffer.
5. 1 ul av 6x Gel Loading Dye Orange Dye ble blandet med 5 ul prøve (6x). Etter prøvene ble satt i brønnene, ble prøvene satt på 70 volt i 30-40 min. Gelen ble studert i BioRad Molecular Imager GelDocTM XR+.

IX. Forberedelse til Sanger Sekvensering.

Material:

- Montage-PCR_{μ96} Renseplate
- Montage-SEQ_{μ96} Renseplate
- Vakuumpumpe
- PCR produkt
- 200 pmol/ul Primere (Se tabell 2)
- BigDye® Terminator v3.1 Cycle Sequencing Kit (Life Technologies, Carlsbad, California, USA)
- dH₂O

Protokoll:

Rensing av PCR produkt:

1. Startet med å tilsette 75 ul 1X TE-buffer til PCR produktet som ble blandet ved å pipettere 10 ganger opp og ned. Løsningen ble deretter overført til en Montage-PCR_{μ96} Renseplate, og satt på vakuum med et trykk på 14-15, til brønnene var tomme for væske.
2. 50 ul "sequencing wash solution" ble tilsatt til hver brønn før renseplaten ble satt på vakuum igjen. For å vaske PCR produktet ble samme prosedyre ble gjentatt en gang.
3. Etter vasking av PCR produktet ble det tilsatt 20 ul 1 xTE buffer til renseplaten. Deretter ble platen satt på risting i 10 min for å løsne DNA fra membranen i renseplaten. Væske ble overført fra renseplate til ny AB plate, som var klar for innmerkingsreaksjon.

Innmerkingsreaksjon:

1. En 96 brønn AB plate med rensset PCR produkt gir to innmerkingsreaksjonsplater (en ”forward” primer og en Revers primer). Et eksempel på oppsettet av mastermiksen vises i tabell 11. Reaksjonen fulgte PCR programmet satt opp i tabell 12.

Tabell 11: Mastermiks for innmerkingsreaksjonen

Reagenser	”Stock” konsentrasjon	Volum(μ L)
Primer ”Forward” eller ”Reverse”	10 pmol/ul	0.5
Seq BigDye Terminator buffer	5x	0.5
Templat	Ukjent	2
dH ₂ O	-	6
BigDye Terminator	-	1
Total		10

Tabellen viser mastermiks for innmerkingsreaksjonen.

Tabell 12: PCR programmet for innmerkingsreaksjonen

Temperatur (grader Celsius)	Tid (25 sykluser)
96	10sec
50	5 sek
60	4min
10	uendelig.

Tabellen viser PCR programmet for innmerkingsreaksjonen

Rensing av innmerkingsprodukt:

1. Innmerkingsproduktet er nå lyssensitivt og ble derfor dekt med aluminiumsfolie for å forhindre degradering av den fluorescense fargen. Det ble tilsatt 25ul ”injection solution” til innmerkings-PCR platen og deretter mikset ved pipettering.

2. Løsningen ble overført til en blå renseplate for å fjerne reagenser fra innmerkingsprosessen. Platen ble satt på vakuum med et trykk på 14-15 til brønnene var tomme for væske.
3. I Montage-SEQ_{μ96} Renseplate ble det tilsatt 25 ul "injection solution" i hver brønn, deretter ble den satt på vakuum med likt trykk. Prosessen ble gjentatt en gang.
4. Tilslutt ble det tilsatt 25 ul "injection solution" til hver prøve som deretter ble satt på en raster i 10 min. Løsningen med innmerkingsproduktet ble overført fra renseplaten til en ny AB 96-brønn. AB 96-brønnplaten ble spunnet ned og satt i en sekvenseringsbeholder klar for sekvensering.

X. Oversikt over konsentrasjonen over mRNA brukt i qPCR

Tabell 13: Konsentrasjonen til RNA prøvene brukt i qPCR

Prøver	Første måling (ng/ul)	Andre måling (ng/ul)	Tredje måling (ng/ul)	Gjennomsnittlig konsentrasjon (ng/ul)
4	48,1	26,7	35,8	36,87
5	42,7	40,5	38	40,40
9	81,6	60,9	58,1	66,87
19	100	90,1	83,4	91,17
20	33,8	31,8	32,2	32,60
26	32,5	24	39,7	32,07
29	36,7	31	42,7	36,80
40	36,2	22,6	37,2	32,00

Tabellen viser konsentrasjonen til RNA prøvene brukt i qPCR.

XI. Optimaliserte primerpar

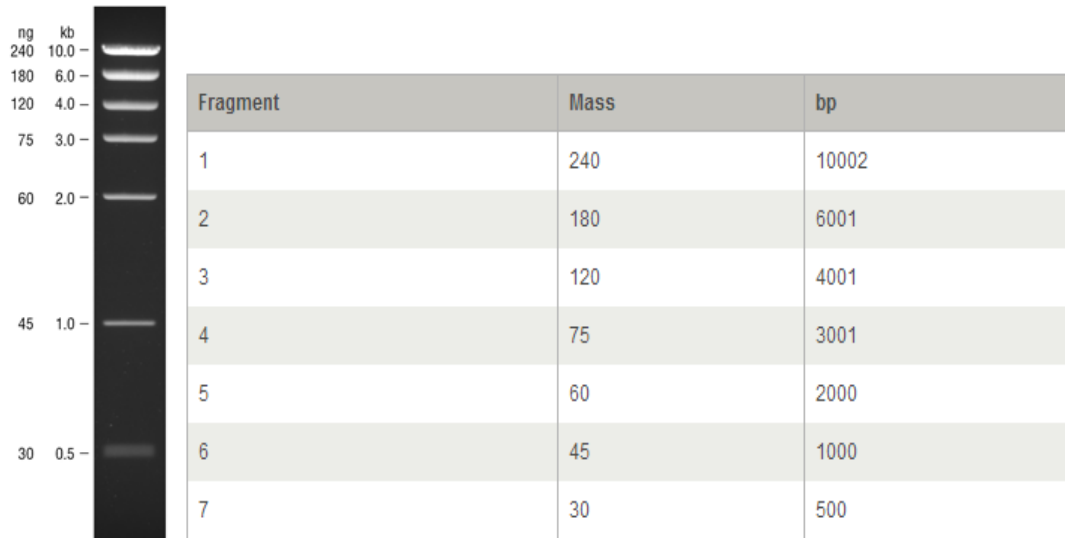
Tabell 14: Forholdene for optimaliseringen av primerpar i Primersett 1

Primerpar	Økning i MgCl ₂ konsentrasjon (μL)	Optimal Temperatur (grader celsius)
Primerpar 1	0,5	59.7
Primerpar 7	2	53.9
Primerpar 8	2.5	57.7
Primerpar 9	2	58.8
Primerpar 13	3.5	57.7

Tabellen viser forholdene for optimaliseringen av primerepar i Primersett 1.

XII. Ladder som ble brukt i gel elektroforese

Figur 22: ”Mass DNA ladder”



Figuren viser ”Mass DNA ladder” (N3237S, New England Biolabs, Ipswich, England).
 Bildet er hentet fra; <https://www.neb.com/products/n3237-mass-dna-ladder>

XIII. Sekvens til E-cadherin på kromosom 26 (genotype QQ)

>Contig QQ

```

AAGATAATAGGAGGAGGGTAGAGATAAAAAAGATAAAGTTAATTAGTTCAGAAAAGCACACGTTTGAAAAATAGACTTTT
CTCATCTACACGTAAGGGATTACTTGTTACTATTGGTGGAACGACTATCCAAAAATTATTCAAACAAATGGGGGCATTTTG
GTTTCGTGGAGTTGGGGCGTATTAATATTATTTCTCCAGGCTTCAAACCAGGGTCGTCGGAGTCAACATGTCTACCAGGTTTCA
ATTCAGAAAATATACATTTTCAAAGTGGAAGAAATCACTTACAAAGTGCCCGGAGACTAGGAAAAAGTGGTTTTTGTGACTG
CACCAGCCGACCAGCTTTCTTTTACTCCGAGGATTCACACTTCAAAGTAGATGGCGACGGGACGCTGAAACTGGAGAGG
GGGCTGACTCTGCATAATGGACATAAGGAGGTCTATGTCTTACCCAGTCCAAGGGCAAGAAGATCACGGTTCAGTCAGAG
TGCTGCATGAGGCCAGACATGGCCACCACCACAATCACCATGAGATGACCACCCAGCCCAAGCCAGGAGCAAGTCTGTCTCT
ACCTGTTCTGAACTCCCAAGTCTTCAGGAGGTCTGAAAAGAAGGAAGAGGGACTGGGTCATTCCTCCCATCAACTCCCAAG
AGAATGACCGAGGCCCCCTCCCAAGATTATGGTGCAGATCAGGTCCAACAATGATAAAGAGGTGAAGATCCAGTACAGCAT
CACTGGCACTGGGGCAGACCTACCTCCTGTGGGAATCTTCACTGTGGACAAAACTCTGGGAATCTCTATGTGACCGAGCCCT
TGGACAGGGAGAAAAAGACAAATACATTCTCCTAGCCCATGCTGTTGCAAGTGGGTGCAGGTATAGCTGAGGATCCCATGGA
GATCATTGTGAAAAGTCATCGATATGAATGACAACAAACAGTATTTTACCCAAGATCCATTTATGGGAACAGTTCCTGAAGCAT
CAAAACCAGGTGACGAGGTCATGCAGGTAACGGCCACTGATGCTGATGAGGAGGGCTCTGCCAATTCTGATGTCAGATACAC
CATTCTCAGTCAGGAGCCTCCACTACCAAGCCCCAACATGTTTGTCAATCAACCCTGTGACTGGAGGGATTTCGGGTTAATGCAC
CTGGGTTGGACAGAGAGAAAAATCCCAATATACTTTGGCAATCCAAGCTGCCGATATGGAGGGAAATGGCCTTACCAGCTT
TGGCAAAGCCATCATTACTAACGGACAGCAATGACAACGCGCCACAGTTTGTGACGCTTCGTACACCGTGTGACGTCCTCA
GAGAATAAAGTGGATGCCTTGGTGGTGAATGCCAGTACTGATGGAGATGAGCCTCACTCCTCTGCTGGGCCACCACCT
ACAAGATAGTTGACGGGGACCCTCAAGGCCTTTTCAACGTGAGCACAGGCCCTAGCAAGCTGGAGGGCATCATTACAACAGC
CAAGCCGCTTGACTTTGAGAAGAACAACAAGTACACTCTGCTGGTCACTGTGCAGAATGAAGTCCCATTCACAATCAGCATG
CCCACCTCCACTGCTACTGTTGTAGTGAATGTGGAGGATGTGAATGAAGTCCAGTCTTACCCCAGTGGAGAAGATTATCAG
GAAACCTGAGGACCTCCCTGTTGACAGTGACCTGGTTCTGTACACGGCCACAGACCCAGACACCCGAAGGAATCAGAAAAGTC
ACATACAAGATACGCAATGATAATGCTGGATGGCTCAGTATCAACAAAGACACTGGGCTGATCAAAGTCAAGAGCCTCATGG
ACAGAGAATCCACTTTTGTCCAAGACAACAATACTCTGTTATTGTTCTGGGCATCGACAACGATGAAATCCCGGCAACTGGC
ACTGGCACCCCTTATCATAGAGCTGGAGGATGTGAATGATAATGCTCCAACATTGACGAGAGTGTGATCAGGGTCTGCAACA
AAGAGTCTCCCAAGTGTGTCAGTCACTGATAAAGATGGTGCGGGCTTCACTGCTCCATACACCGTACAGCTTACAGGGG
TCGTCCCATTTAACTGGACTGCCAGAATGAACGACACAAAGACTGGCATTATCCTGACTCTGAAGACTATGTTGGACAGTGG
AGATTACACGGTGTCTGAGAGTGTCTGACAACCAGGGCCTGCACCATGACAGCACCATCCAGGCCTCCGTCTGTGACTGCA
AAGGAGCGGATGTCCAGTGTCCGATAAAGCTGTAGCAGGCTTCGGCATCTCTAGCATTCTGGGAATTCTAGGAGCAGTCTTA
    
```

CTACTCCTATTGTTGTCTCTTCTGCTCATGTTCTGAGGAAGAGAGGTTGGCGAGAAGAAGGAGCCTCTGCTGCAGGAGGA
CGATGTCAGGGACAACATCTACTACTATGACGAGGAGGGAGGTGGCGAGGATGACCAGGATTTGACTTGAGCGTTCTGCAC
AGAGGTCGGATAAACCGTCTGTATGTTTCCGTAATGACATCGTCCAACCATGGCCCGGACAGTATCGTCCACGACCCCGC
CAACCCAGCCGACATTTGCAACTTCATTGATGATAACCTGAAGGCAGCTGACAACGACCCCACTGCTCCTCCCTACGACTCCC
TCCTGGTGGTTGACTATGAAGGAGGTGGCTCTGAGGCTGGCTCCCTCAGTTCCTCACTCCTCCAGCTCAGGAGACGACCAG
GACTACGACCTCCTCAAGAGTGGGGCCCGCCTTCAAGAAGCTGTCTGACATGTACGGAGGAGGAGGATTGAGAAGTCA
GTCAGCCACTAAACCTTTTGGCTGGAAACTGTCCGGTTCTCCACACTTTGGTCTGTAAACTGAAGACCTTTTCTTATTTCCCC
TTTTATGCGTTTCGGGCACATTTACAGTTTGGCAATTTGATTTGTGCAGACATGGGACCATTTAGAAAAAATGTGTTCAATG
CTCGTCTTACAGCATGGGGAGGGTGGCATACTCTGGCTCTGACTAGCAGGATGTTGACATGGATTTTACACAATTCAGCAGT
TCTTTCTAATGGACTCTTAAACACCCAGATTGTACTTGAATTTAATATGCTTCTTTTGGCGTCTGATCGAAGTTCTTTTTTTT
ATATATGACTATCTTTGGCATGAAATTTGAAGTTACTGAAAGCCAATCAGCCTTACGATGTTTTTGTAACTGAGATTAATTA
TCAAGCTCAATAAATGCTTTTAAAAATGTTTGATAAAACAATAAGTTTGGGTTGTTTTATTCCTTTTCATTTTAACTCATATTAAT
CACTAATTATATATTTTTTCTGAAGAAGCTCACAGTTCTGCTTTTTATTTATTACAGTTCTGCTTTTTAAATTAAGATAT
CCTTTTGGTACAATGATGGATGTGAATATTTGTATTAATACATTTGAATTTGTCTTATTTTGTGCTCAATATTTCAAAATTCAC
AGCTGAGACACTAGATGGCGACAATTTGAATAAAACAGGTCTGGTCAATGAATGATGGCATTCCCTGTCCCGCAGAACAGA
TTGTACAAATTAATTAAGTAATAAGGCTGGAGGGGTTGGTGTATATGGCCAACATATCTCGGCTAAGGGCTGACTTAT
GCACGATGCAACACGGAGTGCTGGACAGAGCCCTGGCCGTGGCTTATTGGCCATATACCACAAAACCTGAGGTGCCTTA
TTNCTTTNNTTTNNTTTGTTCTTTCCGGTTGTTCTNNTTNTCTTTTTTTGCCTTCTTCTTGCTCTTCTGTTTTTCTCT
TTCTCCCTCCCTTGTCTTCTGCTTTTTCTCTCTTTCTTTCCCTTTCTTTCTTTCTTTCTGTGTTTTCTCTTTTTCTCTCT
TTTGGTTCGTTGGCGCTTTTTCTTTCCGCTTCTCTCTCTCCCTTCCNGTTTGGCGGTGTGTTCTTTCTCTTTTTCTCTCT
TCCTTTTTGTCTGCTTGGTCTCTTTTGGGGTGGTTTTTTCCCTCTTTTTTTCTTCTCTTTTCTGTGTTTTTTTTTTCTTT
GTTTTGTTCTCTTTGCTCCTTTTTCTTTTTTTTTTTTTCTTTCCGGCGTCTGTCTGCTGTTTTTCTCCCTTCTGCTTCTC
CCTTTTTTTCTTTCTCTTCTTTCTGCTCTCTCTTTTTTCGCGCTCTTTCCCTGTGCTTCTGCTGCTCTCTCTCTCTCT
TTCTTGCTCCTGTCCGTCCTTTCTGCTGTTCTTCTCTCTTCTGNTTCTCTCTGTGCTTGTCTCTTGTCTTGTCTT

XIV. Sekvensen til E-cadherin på kromosom 26 (genotype qq)

>Contig qq

AAAAAAAAAAAAATGAAAGAAAAAAATAAATTTAATAGAAATATAATAGAAATAGAGAGATTGAGTAATTGAAAGAAAA
AGGAAAAAAGAAAGAAAAGATAATAAAGAATATAAGAGAGAATAAGAAGATGATTAATTAATAGAGAAGTAAAAGTGA
AGGGAGGGGGGAGGGGGGGGAGAGGTGATAAAGTGAAGTTGGTTATTTCCAGAAAAGCACAGTTTTGAAAAATAGACTTTTTT
ATCTACACGTAAGAGATTAATCTTTACTATTGGTGGAAAGCTACTCAAAAAATTATTTCAAAACAATGGGGCATTTTGGTT
CGTGGAGTTGGGCGTATTAATATTATTTCTCCAGCTTCAAACAGGCTCGGAGTCAACATGTTACTACAGGTTCAATTC
AGAAATATACATTTTCAAAGTGGAAAGAAATCACTTCAAAGTGGCCGGAGACTAGGAAAAGTGGTTTTTGTGACTGCACC
AGCCGACCCAGCTTTCTTTTACTCCGAGGATTCACACTTCAAAGTAGATGGCGACGGGACGCTGAAACTGGAGAGGGGGC
TGACTCTGCATAATGGACATAAGGAGGTCTATGTCTTACCCAGTCCAAGGGCAAGAAAGATCACGGTTCCAGTCCAGTGTCT
GCATGAGGCCAGACATGGCCACCACCAATCACCATGAGATGACCACCCAGCCAAAGCCAGGAGCAAGTCTGCTTCACT
GTTCTGAACCTCCCCAAGTCTTCAGGAGGTCTGAAAAGAAAGAAAGAGGGACTGGGTCATTCTCCCATCAACTTCCCAGAGA
ATGACCGAGGCCCTTCCCCAAGATTATGGTGCAGATCAGGTCCAACAATGATAAAGAGGTGAAGATCCAGTACAGCATCAC
TGGCACTGGGGCAGACCTACCTCTGTGGGAATCTTACTGTGGACAAAACCTCTGGAATCTCTATGTGACCCGAGCCCTTGG
ACAGGGAGAAAAAGACAAAATACATTTCTTAGCCATGCTGTGCAAGTGGTGCAGGTATAGCTGAGGATCCCATGGAGT
CATTGTGAAAGTCAATCGATATGAATGACAACAACCTGTATTACCCAAGATCCATTTATGGGAAACAGTTCTGAAGCATCAA
AACCAGGTGACGAGGTGATGACAGGTAACGGCCACTGATGCTGATGAGGAGGGCTCTGCCAATCTGATGTCAGATACCCAT
TCTCAGTCCAGGAGCTCCACTACCAAGCCCAACATGTTTGTATCAACTCTGTGACTGGAGGGATTGGGTTAATGCACCTG
GGTTGGACAGAGAAAAATCCAAATATACTTTGGCAATCCAAGTGCCTGATATGGAGGGAATGGCCTTACCAGCTTTGG
CAAAGCCATCAATTACATAACGGACAGCAATGACAAACGGCCAGTTTGTGACGCTTCTGACAGCTTCTGACCCGACAG
AATAAAGTGGATGCCTTGGTGGTGAATAAGCCAGTGACTGATGGAGATGAGCCTCACTCTCTGCTGGGCCACCACCTACA
AGATAGTTGACGGGGACCCTCAAGGCCTTTCAACGTGAGCACAGGCCCTAGCAAGCTGGAGGGCATCTACAACAGCCAA
GCCGCTGACTTTGAGAAGAACAACAAGTACACTCTGCTGGTCACTGTGCAGAATGAAGTCCCATTCACAATCAGCATGCCCA
CCTCCACTGCTACTGTTGATGAAATGTGGAGGATGTGAATGAAGCTCCAGTCTTCAACCCAGTGGAGAAGATTATCAGGAAA
CCTGAGGACCTCCCTGTTGACAGTGTGACCTGGTTCTGTACACAGCCACAGACCCAGACACCCGCAAGGAATCAGAAAGTCACAT
ACAAGATACGCAATGAT AATGCTGGATGGCTCAGTATCAACAAAAGACACTGGGCTGATCAAAGTCAAGAGCCTCATGGACAG
AGAATCCACTTTTGTCCAAGACAACAATACTCTGTTATTGTTCTGGGCATCGACAACGATGAAATCCCGGCAACTGGCCTG
GCACCCATTATCATAGAGCTGGAGGATGTGAATGATAATGCTCCAACATTGACGAGAGTGTGATCAGGGTCTGCAAC AAGGA
GTCCTCCCCACAGTTGTTGTCAGTCACTGAT AAAGATGGTGGGGCTTCACTGCTCCATACACCGTACAGCTTTCAGGGGTCGT
CCCATTCTAAGTCCACTGCCAGAATGAACGACACAAAAGACTGGCATTATCCTGACTCTGAAGACTATGTTGGACAGTGGAGA
TTACACGGTTGCTCAGAGTGTCTGACAACCAGGGCTGCACCATGACAGCACCATCCAGGCCCTCCGTTCTGTGACTGCAAAG
GAGCGGATGTCCAGTGTCCGATAAAGCTGTAGCAGGCTTCCGATCTTAGCATTCTGGAATTTAGGAGCAGTCTACTACTA
CTCCTATTGTTGTTCTTTCTGCTGCTCATGTTCTGAGGAAGAGAGGTTGGCGAGAAGAAGGAGCCTCTGCTGCAGGAGGACGA
TGTACAGGGACAACATCTACTACTATGACGAGGAGGGAGGTTGGCGAGGATGACCAGGATTTGACTTGAGCGTTCTGCACAGA
GGTCTGGATAAACCCTGCTGATGTTTCCGTAATGACATCGCTCCAACCATGGCCCGGACAGATATCGTCCACGACCCGCCAA
CCCAGCCGACATTTGGCAACTTCATTGATGATAACCTGAAGCAGCTGACAACGACCCCACTGCTCTCCTACGACTCCCTC
TGGTTTTGACTATGAAGAGGTTGGCTCTGAGGCTGGCTCCCTCAGTTCCCTCAACTCTCAGCTCCAGTACGACGACCAGGAC
TACGACCTCCTTCAAAGTGGGGCCCGCCTTCAAGAAGCTGTCTGACATGTACGGAGGAGGAGGATTGAGAAGTCAAGTC
AGCCACTAAACCTTTTGGCTGGAAACTGTCCGGTTCTCCACACTTTGGTCTGTAAACTGAAGACCTTTTCTTATTTCCTCTT
TTATGCGTTTTCGGGCACATTTACAGTTTGGCAATTTGATTTGTGCAGACATGGGACCATTTTAGAAAAAATGTGTTCAATGCTC

GTCTTCAGCATGGGGAGGGTGGCATACTCTTGGCTCTGCACTAGCAGGATGTTGACATGGATTTTACACAATTCAGCAGTTCT
TTCTAATGGACTCTTAACAACCCAGATTGACTTGAATTTAATATGCTTCTTTTGGCGTCGATCGAAGTTCTTTTTTTTATA
TATGACTATCCTTTGGCATGAAATTTGAAGTTACTGAAAGCCAAATCAACCTTACGATGTTTTGTTAACTGAGATTAATTATCA
AGCTCAATAAATGCTTTTAAATGTTTGATAAACAACCTAAGTTTGGGTGTTTTTATTCCTTTTCATTTTAAATCATATTAATCACT
AATTATATATTTTTTCCCTGAAGAAGCTCACAGTTCTGCTTTTTATTTTATTACAGTTCTGCTTTTTTAAATTAAGATATCCTTT
TGGTACAATGATGGATGTGAATATTTGTATTAATACATTGTAATTTGTCTATTTTGGAGCTCAATTATTCAAATTCACAGCTG
AGACACTAGATGGCGACAATTTGAATAAAAACAGGTCTTGGTCAATGAATGATGGCATTTCCTTGTCCCGCAGAACAGATTGTA
CAAATTATATTTAAGTAATAAGGCTGGAGGGGTGTGGTGTATATGGCCAACATATCTCGGCTAAGGGCTGTACTTATGCACG
ATGCAACACGGAGTGCCTGGACAGAGCCCTTGGCCGTGGCTTATTGGCCATATACCACAAAACCTGAGGTGCCTTATTCTT
TNCTTTTTTTTGTTC

XV. Sekvensen til E-cadherin på kromosom 26 (genotype Qq)

>ContigQq

TACTTGTGACTATGTGGTGGAACTATCCAAAAATTATTCAAACCCCTTACTATGTTGGTTCGTGGAGTTGGGCGTATT
AATATTATTTCCAGGCTTCAAACCCAGGGTCTCGGAGTCAACATGTCTACCAGGTTCAATTCAGAAATATACATTTCAA
AGTGGAAAAGAAATCACTTACAAAAGTGGCCGGAGACTAGGAAAAAGTGGTTTTTGTGACTGCACCAGCCGACCAGCTTTCTC
TTTACTCCGAGGATTCACACTTCAAAGTAGATGGCGACGGACGCTGAAACTGGAGAGGGGGCTGACTTGCATAATGGAC
ATAAGGAGGTCTATGTCTTACCCAGTCCAAGGGCAAGAAGATCAGGTTCCAGTCAGAGTGTGCATGAGGCCAGACATGG
CCACCACCACAATCACCATGAGATGACCACCAGCCAAAGCCAGGAGCAAGTCTGTCTTACCTGTTCTGAACTTCCCAAGT
CTTCAGGAGGTCTGAAAAGAAAGGAAGAGGGACTGGGTCTTCTCCATCAACTTCCAGAGAATGACCGAGGCCCTTCCC
CAAGATTATGGTGCAGATCAGGTCCAACAATGATAAAGAGGTGAAGATCCAGTACAGCATCACTGGCACTGGGGCAGACCTA
CCTCCTGTGGGAATCTTACTGTGGACAAAACTCTGGGAATCTCTATGTGACCGAGCCCTTGGACAGGGAGAAAAAGACA
AATACATTCTCCTAGCCATGCTGTTGAGTGGGTGCAGGTATAGCTGAGGATCCCATGGAGATCATTGTGAAAGTCATCGAT
ATGAATGACAACAACCTGTATTTACCCAAGATCCATTTATGGGAACAGTTCTTGAAGCATCAAAACCCAGGTGACGAGGTCA
TGCAGGTAACGGCCACTGATGCTGATGAGGAGGGCTCTGCCAATTCTGATGTCAGATACACCATTCTCAGTCAGGAGCCTCCA
CTACCAAGCCCCAACATGTTTGTCTCAACTCTGTGACTGGAGGGATTCCGGTTAATGCACCTGGGTTGGACAGAGAGAAAA
TTCCCAAATATACTTTGGCAATCCAAGCTGCCGATATGGAGGGAAATGGCCTTACCAGCTTTGGCAAAGCCATCATTACACTA
ACGGACAGCAATGACAACGCGCCACAGTTTGTGACGCTTTCGTACACCGTGTCCAGAGAATAAAGTGGATGCCCTGG
TGGTGAATAATGCCAGTACTGATGAGAGATGAGCCTCACTCTCTGCCTGGGCCACCACCTACAAGATAGTTGACGGGGACCC
TCAAGGCTTTTCAACGTGAGCACAGGCCCTAGCAAGCTGGAGGGCATCATTACAACAGCCAAGCCGTTGACTTTGAGAAG
AACAACAAAGTACACTCTGCTGGTCACTGTGCAGAAATGAAGTCCCATTCACAATCAGCATGCCACCTCCACTGCTACTGTTGT
AGTGAATGTGGAGGATGTAATGAAGTCCAGTCTTACCCAGTGGAGAAGATTATCAGGAAACCTGAGGACCTCCCTGTT
GACAGTCACTGATAAAGATGGTGCAGCCACAGACCCAGACCCGCAAGGAATCAGAAAGTCACATAACAAGTACCAAGTACCA
ATGCTGGATGGCTCAGTATCAACAAGACACTGGGCTGATCAAAGTCAAGAGCCTCATGGACAGAGAATCCACTTTTGTCCA
AGACAACAAATACTCTGTTATTGTTCTGGGCATCGACAACGATGAAATCCCGCAACTGGCACTGGCACCCCTTATCATAGAGC
TGGAGGATGTGAATGATAATGCTCCAACCTGACGAGAGTGTGATCAGGGTCTGCAACAAGGAGTCCCTCCACAGTTGTT
GTCAGTCACTGATAAAGATGGTGCAGCCACAGACCCAGACCCGCAAGGAATCAGAAAGTCACATAACAAGTACCAAGTACCA
CCAGAATGAACGACACAAGACTGGCATTATCCTGACTCTGAAGACTATGTTGGACAGTGGAGATTACACGGTTGTCCTGAG
AGTGTCTGACAACCAGGGCTGCACCATGACAGCACCATCCAGGCTCCGTCTGTGACTGCAAAGGAGCGGATGTCCAGTGC
TCCGATAAAGCTGTAGCAGGCTTCCGCATCTTAGCATTCTGGGAATTCTAGGAGCAGTCTTACTACTCCTATTGTTGTCTCTT
CTGCTGCTCATGTTTCTGAGGAAGAGAGGTGGCGAGAAGAAGGAGCCCTGCTGCAGGAGGACGATGTCAGGGACAAACATCT
ACTACTATGACGAGGAGGGAGGTGGCGAGGATGACCAGGATTTGACTTTGAGCGTTCTGCACAGAGGTCTGGATAACCGTCC
TGATGTTTTCCGTAATGACATCGTCCAACCATGGCCCGCCAGAGTATCGTCCACGACCCGCCAACCCAGCCGACATTGGCA
ACTTCATTGATGATAACCTGAAGGCAGCTGACAACGACCCCACTGCTCCTCCCTACGACTCCCTCCTGGTGTGTTGACTATGAA
GGAGGTGGCTCTGAGGCTGGTCCCTCAGTTCCTCAACTCCTCAGCTCAGGAGACGACCAGGACTACGACCTCCTTCAAGA
GTGGGGCCCGCTTCAAGAAGCTGTCTGACATGTACGGAGGAGGAGGATTGAGAAGTCAGTCAGCCACTAAACCTTTTT
GCTTGGAAACTGTCCGTTTCTCCACTTTGGTCTGTAAACTGAAGACCTTTTCTTATTTCCCTTTTTATGCGTTTCGGGCAC
ATTTACAGTTTGGCAATTTGATTTGTGCAGACATGGGACCATTTAGAAAAAATGTGTTCAATGCTCGTCTTACGATGGAGA
GGTGGCATACTCTTGGCTCTGCACTAGCAGGATGTTGACATGCAATTTACACAATTCACAGTTCTTTCTAATGTTTTCTTAA
CAACGCCCCATTTACTTGAATTTAATATGCTTCTTTTACGCTGATGATGAAAGTTCTTTTTTTTTTAAATGAAAAACCTTTGGTC
AGAAATTGAAACTAATGGAAGCCAAACCCACCTTACCAAGGTTTTGGTAATTTGGGTTAAATAACCTCCCAACCAATGCCTT
TAAAAAAGTTTGA AAAACAATTAAGTTTGGGGGGCTTAATCCTTTTCATTTTAAACAAATAGTCACTAATTATATATTTTT
TCCGGAAGACGCTCAAGTTCTGCTTTTTATTTTACAGTTCTGCTTTTTAAATTAAGATATCCTTTTGGTACAATGATGGA
TGTGAATATTTGTATTAATAACATTTGTAATTTGTCTTATTTTGGAGTCAATTATTCAAATTCACAGCTGAGACACTAGATGGCG
ACAATTGAATAAAAACAGGTCTTGGTCAATGAATGATGGCATTTCCTTGTCCCGCAGAACAGATTGTACAAATTTATTTAAGT
AATAAGGCTGGAGGGGTGTGGTGTATATGGCCAACATATCTCGGCTAAGGGCTGTACTTATGCACGATGCATTTCCGAGTG
CCTGGACAGAGCCCTTAGCCGTGGCTTATTGGCCATATACCACAAAACCTGAAGGTGCCTTA

XVI. Sekvensen til det paraloge genet på kromosom 11

>Kromosom11

ATCCCTGTTCTCCCGAGTACGGAATGATCTTCACGTTGGTGGAAACGACTAGTCTGAAATTATTCAAAACAAATGGGGACCGTC
TGGTTCGCGGAGTTGGGGTTATTAACAGAGGATATCCAGGCTTTCAAGCCAGGGTTGTCGGAGGAGTCAAAATGTATACAG
GTTTCAATTCGGAAATGTACATTTTCAAAGTGGAGAGAAATCACTTACAAAAGTGGGCGGAGACTAGGAAAAGTGGTTTTGA
TGACTGCACCAGCCGCACCAGCTTTCTTTTCGCTCCGAGGATTCACGCTTCAAAGTAAATGGCGACGGGACGCTGAAACTGA
AGAGGGGGTACTCTGATAAATGGACATAAGGAGTTCTATGTCTTACCCAGTCCATGAGCAAGAAGATCAGGTGTTCCAG
TCAGAGTGAGTGCATGAGGCCAGACACTACCACCATCACCATGAGATGACCACCCAGCCCCAGCCTGGATCAAGTCTGTCTC
TACCTGTTCTGAACTTCCCCAAGTCTTCAGGAGGTCTGAAAAGAAGGAAGAGGGACTGGGTCATTCTCCCGTCAACTTCCCA
GAGAATGAGCGAGGCCCTTCCCCAAGATTATGGTGCAGATCAGGTCCAGCAATGATAAAGAGGTGAAGATCATGTACAGCA
TCACTGGCAATGGGGCAGACCAACCTCCTGTGGGACTTCACTGTGGACAAAACCTCTGGGATTCTCTATGTGACCCAGCCT
TTGGACAGGGGAGAAAAAAGACAAATACATTCTCCTAGCCCATGCTGTTGCTGTGGGTGCAGGTAAGGCTGAGGCACCCATGG
AGGTCATTGTGAAAGTCATCGACCAAAAATGACAACAAACCTGTATTTACCCAAAATCCATTTATGGGAACAGTCCCTGAGGC
AGCAAAACCAGGTGATGAGTTCATGCAGGTAACAGCCACTGATGCTGATGAGGAGGGCTCTGCCAATTCTGATGTACAGATAC
ACCATTCTCAGTCAGGAGCCTCCACTACCAAGCCCCAACATGTTTGTATCAAACCTGTGACTGGAGGGATTCCGGGTTAACGC
ACCTGGGTTGGACAGAGAGAAAATTCTTAAATATACTTTGAAAATCCAAGCTGCTGATTTGGAGGGAGATGGCCTTACCAGC
TTTGGCAAAGCCATCATTACAGTAACGGACAGCAATGACAACGCGCCACAGTTTGTGACGCCTTCGTACACTGTGTACAGTCCC
AGAGAATAAAGTAGATGCCTTGGTGGTGAATAATGCCAGTACTGATGGGGATGATCCTCACTCCTCTGCTGGGCCACCCACC
TTCAGATAGTTGACGGGGACCTGAAGGCCTTTTCAGTGTGAGCACAGGCCCTAGCAAGCTGGAAGGCATCATTACGACAG
AAAAGCCCTTGACTTTGAGAAGAACAACAAGTACACTCTGCTGGTCACTGTGGTGAATGAAGTCCATTTGCCATCCCTCTG
CCCACCTCCACTGCTACTGTTATAGTGAATGTGGAGGATGTGAATGAAGTCCAGTCTTCAAACCAGTGGAGAAGATGATCA
GAAGACCTGAGGACCTCCATGTGGACAGCGACCTGGTTCTGTACACAGCCACAGACCCAGACACTGCAAGGAATCAGAAAAGT
CACATAAAGATTAAGAATGATGTTGCTGGATGGCTCAGTATCAACAAAGACACTGGGCTGATAAAAAGTCAAGACTCTCATG
GACAGAGAATCCCTTTTGTCCAAGACAACAAATACTCTGTTGTTGTTCTGGGCACTGACAACGATGAAATCCCAGCAACGG
GCACTGGCACCCCTTATCATAGAGCTGGAGGATGTGAATGACAACCTCCAACCATCGACGAGAGGATGATTAAGGTCTGCAA
CAAGGAGTCCCTCCCCACAGCTGCTGCAATCACTGATAAAGACGGAGCAGGCTTCGCTGCTCCATACACTGTACAGCTTCAGG
GGTCGTCCCTTTCCAACCTGGACTGCCAGAATGAATGACTCAAAAACCTGGCATTATCCTGACACTGAGTACTATGTTGGACAGT
GGAGATTACACGGTTGTCTGAGAGTGTCTGACAACCCAGGGCTGCACCAGGAGAGCACCATCCTGGCCTCCGTTGTGACT
GCAAAGGAGCTGATGTCCAGTGCGCCAATAATTATTTCTTCTTTCTATCTTTTTCTCTCTCTTCTCTAATATATTTTTATTCTA
TACTTCTATACTATCTATTCATGATTCTTTAGATTTCACTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTCT
CTTTCTATCTTTCTTTACTTCTTCTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT
TATTCTCTACTCCTTATTTTATACTACAACCTTTTTTATCTTATACTTTACATCTATTACTGCTTTCTTTTCTTTTCTTTTCT
TTATATATCCTTCTTTAACTCTTATGTATTTTCTTTCTATCTGCTTTTACATTTCTTCTGTCATCTTTTCTACTTTTTACT
CTAATTTACTACTACTTTTCTTATTTTATCATTCTATTGTTCTATACTATATCTGTATACAACCTTTTATTACTATCTACTACT
ATTATCCTTCTATTCTTACTTAATCTTCGTTAGACTTGTATATCTTTTTATAGTCTCGCATACATCCTACGACCTCCTTCAAGAG
TGGGCCCCGCGCTTCAAGAAGCTGTCTGACATGTACGGAGGAGGAGGATTGAGAAGTCAGTCAGCCACTAAACCTTTTTG
CTTGAAAACCTGTCGGTTCTCCACACTTTGGTCTGTAAACTGAAGACCTTTTTCTTTATTTCCCTTTTTATGCGTTTCGGGCACA
TTTACAGTTTATTTGTGAACCGTGGGAACATTTTAGAGAATTGTGTTCAATGCTCGTCTTCAGCATGGGGAGGGTGGCACAC
TCTTGGCTCTGCACTAGCAGGATGTTGACATAGATTTTACACAGTTCAGCAGTGCTTCTAATGGACTCTTAATAACCCAGAT
TGTACTTGAATTTAATATGCATTGTTTGGTCTGCTTCTATGTTGGTGTATATCATAAAGTTTCTATAAAGACTATACAT
TGGCATGTAATGTAAGTTACTTCAAATCCAAATCAGCCTTAGGATTTTTTTGTTAACTGAGATGAATATCAAGCTCAAACGCT
TTTTAAAAATGTTTGATAAAAAATATAAGTTTGGGTCTTTTGTGTTCTTTTCTTTTAAATGTTTTTGTCTACTTGTCTGTA
AGAAGCACTGTGAATTTTTTTTTATAAAGGATATCCCTTTGGTACATTGATGGCCATGTGAATAAAA

XVII. Sammenligning av konsensussekvensene til ulike genotypene til E-cadherin på kromosom 26 til konsensussekvensen til paralog på kromosom 11.

Sammenligning av konsensussekvensen for genotype QQ med konsensussekvensen for paralog på kromosom 11.

```

Query   36   TTGGTGGAAACGACTAGTCTGAAATTATTCAAAACAAATGGGGACCGTCTGGTTCGCGGAG   95
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  115   TTGGTGGAAACGACTATCCAAAAATTATTCAAAACAAATGGGGGCATTTTGGTTCGTGGAG   174

Query   96   TTGGGGTTATTAACAGAGGATATCCAGGCTTTCAAGCCAGGGTTGTCGGAGGAGTCAAAA   155
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  175   TTGGGCGTATTAATATTATTTCTCCAGGCTTTCAAACCAGGGTCGTC---GGAGTCAACA   231

Query  156   TGTATACCAGGTTTCAATTCGGAAATGTACATTTTCAAAGTGGAGAGAAATCACTTACAA   215
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  232   TGTCTACCAGGTTTCAATTCAGAAATATACATTTTCAAAGTGGAAAGAAATCACTTACAA   291

Query  216   AGTGGGCGGAGACTAGGAAAAGTGGTTTTTGTGACTGCACCAGCCGACCAGCTTTCTC   275
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  292   AGTGGCCGGAGACTAGGAAAAGTGGTTTTTGTGACTGCACCAGCCGACCAGCTTTCTC   351

Query  276   TTTCGCTCCGAGGATTCACGCTTCAAAGTAAATGGCGACGGGACGCTGAAACTGAAGAGG   335
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  352   TTTCACTCCGAGGATTCACACTTCAAAGTAGATGGCGACGGGACGCTGAAACTGGAGAGG   411

Query  336   GGGGTGACTCTGCATAATGGACATAAGGAGTTCTATGTCTCTACCCAGTCCATGAGCAAG   395
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  412   GGGCTGACTCTGCATAATGGACATAAGGAGGTCTATGTCTCTACCCAGTCCAAGGGCAAG   471

Query  396   AAGATCAGCTGTTCCAGTCAGAGTGAGTGCATGAGGCCAGACA---CTACCA-C-C-ATC   449
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  472   AAGATCA-CGGTTCAGTCAGAGTG-CTGCATGAGGCCAGACATGGCCACCACCACAATC   529

Query  450   ACCATGAGATGACCACCCAGCCCAGCCTGGATCAAGTCTGTCTCTACCTGTTCTGAACT   509
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  530   ACCATGAGATGACCACCCAGCCCAAGCCAGGAGCAAGTCTGTCTCTACCTGTTCTGAACT   589

Query  510   TCCCCAAGTCTTCAGGAGGTCTGAAAAGAAGGAAGGGACTGGGTCATTCTCCCGTCA   569
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```


Sbjct 590 TCCCCAAGTCTTCAGGAGGTCTGAAAAGAAGGAAGAGGGACTGGGTCATTCCCTCCCATCA 649

Query 570 ACTTCCCAGAGAATGAGCGAGGCCCTTCCCCAAGATTATGGTGCAGATCAGGTCCAGCA 629
 |||

Sbjct 650 ACTTCCCAGAGAATGACCGAGGCCCTTCCCCAAGATTATGGTGCAGATCAGGTCCAACA 709

Query 630 ATGATAAAGAGGTGAAGATCATGTACAGCATCACTGGCAATGGGGCAGACCAACCTCCTG 689
 |||

Sbjct 710 ATGATAAAGAGGTGAAGATCCAGTACAGCATCACTGGCACTGGGGCAGACCTACCTCCTG 769

Query 690 TGGGACTCTTCACTGTGGACAAAACTCTGGGATCTCTATGTGACCCAGCCTTTGGACA 749
 |||

Sbjct 770 TGGGAATCTTCACTGTGGACAAAACTCTGGGAATCTCTATGTGACCGAGCCCTTGGACA 829

Query 750 GGGAGAAAAAGACAAATACATTCTCTAGCCCATGCTGTTGCTGTGGGTGCAGGTAAGG 809
 |||

Sbjct 830 GGGAGAAAAAGACAAATACATTCTCTAGCCCATGCTGTTGCAGTGGGTGCAGGTATAG 889

Query 810 CTGAGGCACCCATGGAGGTCATTGTGAAAGTCATCGACAAAATGACAACAAACCTGTAT 869
 |||

Sbjct 890 CTGAGGATCCCATGGAGATCATTGTGAAAGTCATCGATATGAATGACAACAAACCTGTAT 949

Query 870 TTACCCAAAATCCATTTATGGGAACAGTCCCTGAGGCAGCAAACCAGGTGATGAGTTCA 929
 |||

Sbjct 950 TTACCCAAGATCCATTTATGGGAACAGTTCCTGAAGCATCAAACCAGGTGACGAGGTCA 1009

Query 930 TGCAGGTAACAGCCACTGATGCTGATGAGGAGGGCTCTGCCAATTCTGATGTCAGATACA 989
 |||

Sbjct 1010 TGCAGGTAACGGCCACTGATGCTGATGAGGAGGGCTCTGCCAATTCTGATGTCAGATACA 1069

Query 990 CCATTCTCAGTCAGGAGCCTCCACTACCAAGCCCCAACATGTTTGTCAATCAACCCTGTGA 1049
 |||

Sbjct 1070 CCATTCTCAGTCAGGAGCCTCCACTACCAAGCCCCAACATGTTTGTCAATCAACCCTGTGA 1129

Query 1050 CTGGAGGGATTCGGGTTAACGCACCTGGGTTGGACAGAGAGAAAATTCTTAAATATACTT 1109
 |||

Sbjct 1130 CTGGAGGGATTCGGGTTAATGCACCTGGGTTGGACAGAGAGAAAATTCCCAAATATACTT 1189

Query 1110 TGGAAATCCAAGCTGCTGATTTGGAGGGAGATGGCCTTACCAGCTTTGGCAAAGCCATCA 1169
 |||

Sbjct 1190 TGGCAATCCAAGCTGCCGATATGGAGGGAAATGGCCTTACCAGCTTTGGCAAAGCCATCA 1249

Query 1170 TTACAGTAACGGACAGCAATGACAACGCGCCACAGTTTGTGACGCCTTCGTACTACTGTGT 1229

Sbjct	1250	TTACTACTAACGGACAGCAATGACAACGCGCCACAGTTTGTGACGCCTTCGTACACCGTGT	1309
Query	1230	CAGTCCCAGAGAATAAAGTAGATGCCTTGGTGGTGAAAATGCCAGTGACTGATGGGGATG	1289
Sbjct	1310	CAGTCCCAGAGAATAAAGTGGATGCCTTGGTGGTGAAAATGCCAGTGACTGATGGAGATG	1369
Query	1290	ATCCTCACTCCTCTGCCTGGGCCACCACCTTCAAGATAGTTGACGGGGACCCTGAAGGCC	1349
Sbjct	1370	AGCCTCACTCCTCTGCCTGGGCCACCACCTACAAGATAGTTGACGGGGACCCTCAAGGCC	1429
Query	1350	TTTTCAGTGTGAGCACAGGCCCTAGCAAGCTGGAAGGCATCATTACGACAGAAAAGCCCC	1409
Sbjct	1430	TTTTCAACGTGAGCACAGGCCCTAGCAAGCTGGAGGCATCATTACAACAGCCAAGCCGC	1489
Query	1410	TTGACTTTGAGAAGAACAACAAGTACTCTGCTGGTCACTGTGGTGAATGAAGTCCCAT	1469
Sbjct	1490	TTGACTTTGAGAAGAACAACAAGTACTCTGCTGGTCACTGTGCAGAATGAAGTCCCAT	1549
Query	1470	TTGCCATCCCTCTGCCCACCTCCACTGCTACTGTTATAGTGAATGTGGAGGATGTGAATG	1529
Sbjct	1550	TCACAATCAGCATGCCACCTCCACTGCTACTGTTGTAGTGAATGTGGAGGATGTGAATG	1609
Query	1530	AAGCTCCAGTCTTCAAACCAGTGGAGAAGATGATCAGAAGACCTGAGGACCTCCATGTGG	1589
Sbjct	1610	AAGCTCCAGTCTTCAACCCAGTGGAGAAGATTATCAGGAAACCTGAGGACCTCCCTGTTG	1669
Query	1590	ACAGCGACCTGGTTCTGTACACAGCCACAGACCCAGACACTGCAAGGAATCAGAAAGTCA	1649
Sbjct	1670	ACAGTGACCTGGTTCTGTACACAGCCACAGACCCAGACACCGCAAGGAATCAGAAAGTCA	1729
Query	1650	CATACAAGATTAAG-AATGATGTTGCTGGATGGCTCAGTATCAACAAAGACACTGGGCTG	1708
Sbjct	1730	CATACAAGAT-ACGCAATGATAATGCTGGATGGCTCAGTATCAACAAAGACACTGGGCTG	1788
Query	1709	ATAAAAGTCAAGACTCTCATGGACAGAGAATCCACTTTTGTCCAAGACAACAAATACTCT	1768
Sbjct	1789	ATCAAAGTCAAGAGCCTCATGGACAGAGAATCCACTTTTGTCCAAGACAACAAATACTCT	1848
Query	1769	GTTGTTGTTCTGGGCACTGACAACGATGAAATCCCAGCAACGGGCACTGGCACCCCTTATC	1828
Sbjct	1849	GTTATTGTTCTGGGCATCGACAACGATGAAATCCCAGCAACTGGCACTGGCACCCCTTATC	1908

```

Query 1829 ATAGAGCTGGAGGATGTGAATGACAACCCTCCAACCATCGACGAGAGGATGATTAAGGTC 1888
          |||
Sbjct 1909 ATAGAGCTGGAGGATGTGAATGATAATGCTCCAACCATGACGAGAGTGTGATCAGGGTC 1968

Query 1889 TGCAACAAGGAGTCCTCCCCACAGCTGCTGTCAATCACTGATAAAGACGGAGCAGGCTTC 1948
          |||
Sbjct 1969 TGCAACAAGGAGTCCTCCCCACAGTTGTTGTCTGACTGATAAAGATGGTGCAGGCTTC 2028

Query 1949 GCTGCTCCATACACTGTACAGCTTCAGGGGTCGTCCTTTCCAACCTGGACTGCCAGAATG 2008
          |||
Sbjct 2029 ACTGCTCCATACACCGTACAGCTTCAGGGGTCGTCCTTCTAACTGGACTGCCAGAATG 2088

Query 2009 AATGACTCAAAAACCTGGCATTATCCTGACACTGACTACTATGTTGGACAGTGGAGATTAC 2068
          |||
Sbjct 2089 AACGACACAAAGACTGGCATTATCCTGACTCTGAAGACTATGTTGGACAGTGGAGATTAC 2148

Query 2069 ACGGTTGTCCTGAGAGTGTCTGACAACCAGGGCCTGCACCAGGAGAGCACCATCCTGGCC 2128
          |||
Sbjct 2149 ACGGTTGTCCTGAGAGTGTCTGACAACCAGGGCCTGCACCATGACAGCACCATCCAGGCC 2208

Query 2129 TCCGTGTGTGACTGCAAAGGAGCTGATGTCCAGTGCGCCAATAA 2172
          |||
Sbjct 2209 TCCGTCTGTGACTGCAAAGGAGCGGATGTCCAGTGTCCGATAA 2252

HULL

Query 2730 CTACGACCTCCTTCAAGAGTGGGGCCCGCGCTTCAAGAAGCTGTCTGACATGTACGGAGG 2789
          |||
Sbjct 2724 CTACGACCTCCTTCAAGAGTGGGGCCCGCGCTTCAAGAAGCTGTCTGACATGTACGGAGG 2783

Query 2790 AGGAGAGGATTGAGAAGTCAGTCAGCCACTAAACCTTTTTGCTTGAAAACCTGTCCGGTTC 2849
          |||
Sbjct 2784 AGGAGAGGATTGAGAAGTCAGTCAGCCACTAAACCTTTTTGCTTGAAAACCTGTCCGGTTC 2843

Query 2850 TCCACACTTTGGTCTGTAAACTGAAGACCTTTTCTTTATTTCCCCTTTTATGCGTTTCG 2909
          |||
Sbjct 2844 TCCACACTTTGGTCTGTAAACTGAAGACCTTTTCTTTATTTCCCCTTTTATGCGTTTCG 2903

Query 2910 GGCACATTTACAGTTT---A-TTTG-T--G--CAACCGTGGGAACATTTTAGAGAAT-TG 2959
          |||
Sbjct 2904 GGCACATTTACAGTTTGGCAATTTGATTTGTGCAGACATGGGACCATTTTAGAAAAATG 2963

Query 2960 TGTTCATGCTCGTCTTCAGCATGGGGAGGGTGGCACACTCTGGCTCTGCACTAGCAGG 3019

```

```

|||||
Sbjct  2964  TGTTC AATGCTCGTCTTCAGCATGGGGAGGGTGGCATACTCTTGGCTCTGCACTAGCAGG  3023

Query  3020  ATGTTGACATAGATTTTACACAGTTCAGCAGTGCCTTCTAATGGACTCTTAATAACCCCA  3079
|||||
Sbjct  3024  ATGTTGACATGGATTTTACACAATTCAGCAGTCTTCTAATGGACTCTTAACAACCCCA  3083

Query  3080  GATTGTACTTGAATTTAATATGCATTGTTTGCTGGT-TCTTGCTTCTATGTTGGTGTCA  3138
|||||
Sbjct  3084  GATTGTACTTGAATTTAATATGC-TTCTTT--TGGCGTCGTA-T-CGAAGTTCTT-T--T  3135

Query  3139  ATCATAAAGTTTCTATAAAGACTATACATTGGCATGTAATGTAAGTACTTCAAATCCAA  3198
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Sbjct  3136  -T--T----TTTATATAT-GACTATCCTTTGGCATGAAATTGAAGTACTG-AAAGCCAA  3186

Query  3199  ATCAGCCTTAGGAatTTTTTGTAACTGAGATGAATTATCAAGCTCAA-AC--GCTTTTT  3255
|||||
Sbjct  3187  ATCAGCCTTACGATGTTTTTGTAACTGAGATTAATTATCAAGCTCAATAAATGCTTTT-  3245

Query  3256  AAAAAATGTTTGATAAAAAATATAAGTTTGGGTCTTTTGTGTTCCCTTTTCATTTTAA  3310
|||||
Sbjct  3246  AAAAAATGTTTGATAAACAACTAAGTTTGGGTGTTT-TATTCCTTTTCATTTTAA  3299

```

Sammenligning av konsensussekvensen for genotype qq med paraloge på kromosom 11

```

Query  36  TTGGTGG AACGACTAGTCTGAAATTATTCAAACAAATGGGGACCGTCTGGTTCGCGGAG  95
|||||
Sbjct  273  TTGGTGG AACGACTATCCAAAAATTATTCAAACAAATGGGGGCATTTGGTTCGTGGAG  332

Query  96  TTGGGGTTATTAACAGAGGATATCCAGGCTTTCAAGCCAGGGTTGTCGGAGGAGTCAAAA  155
|||||
Sbjct  333  TTGGGCGTATTAATATTATTTCTCCAGGCTTTCAAACCAGGTCGTC---GGAGTCAACA  389

Query  156  TGTATACCAGGTTTCAATTCGGAAATGTACATTTTCAAAGTGGAGAGAAATCACTTACAA  215
|||
Sbjct  390  TGTCTACCAGGTTTCAATTCAGAAATATACATTTTCAAAGTGGAAAGAAATCACTTACAA  449

Query  216  AGTGGGCGGAGACTAGGAAAAGTGGTTTTTGATGACTGCACCAGCCGCACCAGCTTTCTC  275
|||||
Sbjct  450  AGTGGCCGGAGACTAGGAAAAGTGGTTTTTGATGACTGCACCAGCCGCACCAGCTTTCTC  509

Query  276  TTTGCTCCGAGGATTCACGCTTCAAAGTAAATGGCGACGGGACGCTGAAACTGAAGAGG  335

```

```

||||| ||||||||||||||| ||||||||||| ||||||||||||||||||| |||||||
Sbjct 510 TTTCACTCCGAGGATTACACTTCAAAGTAGATGGCGACGGGACGCTGAAACTGGAGAGG 569

Query 336 GGGGTGACTCTGCATAATGGACATAAGGAGTTCTATGTCTCTACCCAGTCCATGAGCAAG 395
||| ||||||||||||||||||| ||||||||||||||||||| | |||||||
Sbjct 570 GGGCTGACTCTGCATAATGGACATAAGGAGGTCTATGTCTCTACCCAGTCCAAGGGCAAG 629

Query 396 AAGATCAGCTGTTCCAGTCAGAGTGAGTGCATGAGGCCAGACA---CTACCA-C-C-ATC 449
||||||| | ||||||||||||||| ||||||||||||||| | |||| | |||
Sbjct 630 AAGATCA-CGGTTCAGTCAGAGTG-CTGCATGAGGCCAGACATGGCCACCACCACAATC 687

Query 450 ACCATGAGATGACCACCCAGCCCCAGCCTGGATCAAGTCTGTCTCTACCTGTTCTGAACT 509
||||||||||||||||| |||| ||| |||||||||||||||||||
Sbjct 688 ACCATGAGATGACCACCCAGCCCAAGCCAGGAGCAAGTCTGTCTCTACCTGTTCTGAACT 747

Query 510 TCCCAAGTCTTCAGGAGGTCTGAAAAGAAGGAAGAGGGACTGGGTCATTCTCCCGTCA 569
||||||||||||||||| ||||||||||||||||||| ||||
Sbjct 748 TCCCAAGTCTTCAGGAGGTCTGAAAAGAAGGAAGAGGGACTGGGTCATTCTCCCATCA 807

Query 570 ACTTCCCAGAGAATGAGCGAGGCCCTTCCCAAGATTATGGTGCAGATCAGGTCCAGCA 629
||||||||||||| ||||||||||||||||||| ||||
Sbjct 808 ACTTCCCAGAGAATGACCGAGGCCCTTCCCAAGATTATGGTGCAGATCAGGTCCAACA 867

Query 630 ATGATAAAGAGGTGAAGATCATGTACAGCATCACTGGCAATGGGGCAGACCAACCTCTG 689
||||||||||||| ||||||||||||||| ||||||||| |||||||
Sbjct 868 ATGATAAAGAGGTGAAGATCCAGTACAGCATCACTGGCACTGGGGCAGACCTACCTCTG 927

Query 690 TGGGACTCTTCACTGTGGACAAAACTCTGGGATTCTCTATGTGACCCAGCCTTTGGACA 749
||||| ||||||||||||||| ||||||||||| ||||| |||||
Sbjct 928 TGGGAATCTTCACTGTGGACAAAACTCTGGGAATCTCTATGTGACCGAGCCCTTGGACA 987

Query 750 GGGAGAAAAAAGACAAATACATTCTCCTAGCCCATGCTGTTGCTGTGGGTGCAGGTAAGG 809
||||||||||||| ||||||||||| ||||| |||||
Sbjct 988 GGGAGAAAAAAGACAAATACATTCTCCTAGCCCATGCTGTTGCTGTGGGTGCAGGTAAGG 1047

Query 810 CTGAGGCACCCATGGAGGTCATTGTGAAAGTCATCGACCAAAATGACAACAAACCTGTAT 869
|||||| | ||||| ||||||||||| |||||||||||
Sbjct 1048 CTGAGGATCCCATGGAGATCATTGTGAAAGTCATCGATATGAATGACAACAAACCTGTAT 1107

Query 870 TTACCCAAAATCCATTTATGGGAACAGTCCCTGAGGCAGCAAACCAGGTGATGAGTTCA 929
||||||| ||||||||||| |||| | ||||||||| ||| |||
Sbjct 1108 TTACCCAAGATCCATTTATGGGAACAGTTCCTGAAGCATCAAACCAGGTGACGAGGTCA 1167

```

Query 930 TGCAGGTAACAGCCACTGATGCTGATGAGGAGGGCTCTGCCAATTCTGATGTCAGATACA 989
 |||
 Sbjct 1168 TGCAGGTAACAGCCACTGATGCTGATGAGGAGGGCTCTGCCAATTCTGATGTCAGATACA 1227

Query 990 CCATTCTCAGTCAGGAGCCTCCACTACCAAGCCCCAACATGTTTGTTCATCAACCCTGTGA 1049
 |||
 Sbjct 1228 CCATTCTCAGTCAGGAGCCTCCACTACCAAGCCCCAACATGTTTGTTCATCAACTCTGTGA 1287

Query 1050 CTGGAGGGATTTCGGGTTAACGCACCTGGGTTGGACAGAGAGAAAATTCCTAAATATACTT 1109
 |||
 Sbjct 1288 CTGGAGGGATTTCGGGTTAATGCACCTGGGTTGGACAGAGAGAAAATTCCTAAATATACTT 1347

Query 1110 TGGAAATCCAAGCTGCTGATTTGGAGGGAGATGGCCTTACCAGCTTTGGCAAAGCCATCA 1169
 |||
 Sbjct 1348 TGGCAATCCAAGCTGCCGATATGGAGGGAAATGGCCTTACCAGCTTTGGCAAAGCCATCA 1407

Query 1170 TTACAGTAACGGACAGCAATGACAACGCGCCACAGTTTGTGACGCCTTCGTACACTGTGT 1229
 |||
 Sbjct 1408 TTACACTAACGGACAGCAATGACAACGCGCCACAGTTTGTGACGCCTTCGTACACCGTGT 1467

Query 1230 CAGTCCCAGAGAATAAAGTAGATGCCTTGGTGGTAAAATGCCAGTGACTGATGGGGATG 1289
 |||
 Sbjct 1468 CAGTCCCAGAGAATAAAGTGGATGCCTTGGTGGTAAAATGCCAGTGACTGATGGAGATG 1527

Query 1290 ATCCTCACTCCTCTGCCTGGGCCACCACCTTCAAGATAGTTGACGGGGACCCTGAAGGCC 1349
 |
 Sbjct 1528 AGCCTCACTCCTCTGCCTGGGCCACCACCTACAAGATAGTTGACGGGGACCCTCAAGGCC 1587

Query 1350 TTTTCAGTGTGAGCACAGGCCCTAGCAAGCTGGAAGGCATCATTACGACAGAAAAGCCCC 1409
 |||
 Sbjct 1588 TTTTCAACGTGAGCACAGGCCCTAGCAAGCTGGAGGGCATCATTACAACAGCCAAGCCGC 1647

Query 1410 TTGACTTTGAGAAGAACAACAAGTACACTCTGCTGGTCACTGTGGTGAATGAAGTCCCAT 1469
 |||
 Sbjct 1648 TTGACTTTGAGAAGAACAACAAGTACACTCTGCTGGTCACTGTGCAGAATGAAGTCCCAT 1707

Query 1470 TTGCCATCCCTCTGCCCACCTCCACTGCTACTGTTATAGTGAATGTGGAGGATGTGAATG 1529
 |
 Sbjct 1708 TCACAATCAGCATGCCACCTCCACTGCTACTGTTGATAGTGAATGTGGAGGATGTGAATG 1767

Query 1530 AAGCTCCAGTCTTCAAACCAGTGGAGAAGATGATCAGAAGACCTGAGGACCTCCATGTGG 1589
 |||
 Sbjct 1768 AAGCTCCAGTCTTCAACCCAGTGGAGAAGATTATCAGGAAACCTGAGGACCTCCCTGTTG 1827

Query 1590 ACAGCGACCTGGTTCTGTACACAGCCACAGACCCAGACACTGCAAGGAATCAGAAAAGTCA 1649
 ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Sbjct 1828 ACAGTGACCTGGTTCTGTACACAGCCACAGACCCAGACACCGCAAGGAATCAGAAAAGTCA 1887

Query 1650 CATAACAAGATTAAG-AATGATGTTGCTGGATGGCTCAGTATCAACAAAGACACTGGGCTG 1708
 ||||| ||| ||||| ||||||||||||||||||||||||||||||||||||||||||||
 Sbjct 1888 CATAACAAGAT-ACGCAATGATAATGCTGGATGGCTCAGTATCAACAAAGACACTGGGCTG 1946

Query 1709 ATAAAAGTCAAGACTCTCATGGACAGAGAATCCACTTTTGTCCAAGACAACAAATACTCT 1768
 || ||||||||| ||||||||||||||||||||||||||||||||||||||||||||
 Sbjct 1947 ATCAAAGTCAAGAGCCTCATGGACAGAGAATCCACTTTTGTCCAAGACAACAAATACTCT 2006

Query 1769 GTTGTGTTCTGGGCACTGACAACGATGAAATCCCAGCAACGGGCACTGGCACCCCTTATC 1828
 ||| ||||||||| |||||||||||||||||| ||||| ||||||||||||||||
 Sbjct 2007 GTTATTGTTCTGGGCATCGACAACGATGAAATCCCGGCAACTGGCACTGGCACCCCTTATC 2066

Query 1829 ATAGAGCTGGAGGATGTGAATGACAACCTCCAACCATCGACGAGAGGATGATTAAGGTC 1888
 |||||||||||||||||| || ||||||||| ||||| ||||| |||||
 Sbjct 2067 ATAGAGCTGGAGGATGTGAATGATAATGCTCCAACCATGACGAGAGTGTGATCAGGGTC 2126

Query 1889 TGCAACAAGGAGTCCTCCCCACAGTGTGTCAATCACTGATAAAGACGGAGCAGGCTTC 1948
 |||||||||||||||||| || ||||| ||||||||| || || |||||
 Sbjct 2127 TGCAACAAGGAGTCCTCCCCACAGTTGTTGTGTCAGTCACTGATAAAGATGGTGCGGGCTTC 2186

Query 1949 GCTGCTCCATACACTGTACAGCTTCAGGGGTCGTCCTTTCCAACCTGGACTGCCAGAATG 2008
 ||||||||| |||||||||||||||||| ||| ||||||||| |||||
 Sbjct 2187 ACTGCTCCATACACCGTACAGCTTCAGGGGTCGTCCTTCTAACTGGACTGCCAGAATG 2246

Query 2009 AATGACTCAAAAAGTGGCATTATCCTGACACTGAGTACTATGTTGGACAGTGGAGATTAC 2068
 || ||| ||| |||||||||||||||||| ||||| ||||||||| |||||
 Sbjct 2247 AACGACACAAAGACTGGCATTATCCTGACTCTGAAGACTATGTTGGACAGTGGAGATTAC 2306

Query 2069 ACGGTTGTCCTGAGAGTGTCTGACAACCAGGGCCTGCACCAGGAGAGCACCATCCTGGCC 2128
 |||||||||||||||||| || ||||| ||||| |||||
 Sbjct 2307 ACGGTTGTCCTGAGAGTGTCTGACAACCAGGGCCTGCACCATGACAGCACCATCCAGGCC 2366

Query 2129 TCCGTGTGTGACTGCAAAGGAGCTGATGTCCAGTGCGCCAATAA 2172
 ||||| |||||||||||||| ||||||||| || |||||
 Sbjct 2367 TCCGTCTGTGACTGCAAAGGAGCGGATGTCCAGTGTCCGATAA 2410

HULL

Query	2730	CTACGACCTCCTTCAAGAGTGGGGCCCGCCTTCAAGAAGCTGTCTGACATGTACGGAGG	2789
Sbjct	2882	CTACGACCTCCTTCAAGAGTGGGGCCCGCCTTCAAGAAGCTGTCTGACATGTACGGAGG	2941
Query	2790	AGGAGAGGATTGAGAAGTCAGTCAGCCACTAAACCTTTTTGCTTGGAACCTGTCCGGTTC	2849
Sbjct	2942	AGGAGAGGATTGAGAAGTCAGTCAGCCACTAAACCTTTTTGCTTGGAACCTGTCCGGTTC	3001
Query	2850	TCCACACTTTGGTCTGTAAACTGAAGACCTTTTCTTTATTTCCCTTTTATGCGTTTCG	2909
Sbjct	3002	TCCACACTTTGGTCTGTAAACTGAAGACCTTTTCTTTATTTCCCTTTTATGCGTTTCG	3061
Query	2910	GGCACATTTACAGTTT---A-TTTG-T--G--CAACCGTGGGAACATTTAGAGAAT-TG	2959
Sbjct	3062	GGCACATTTACAGTTTGGCAATTTGATTTGTGCAGACATGGGACCATTTAGAAAAAATG	3121
Query	2960	TGTTCAATGCTCGTCTTCAGCATGGGGAGGGTGGCACACTCTTGGCTCTGCACTAGCAGG	3019
Sbjct	3122	TGTTCAATGCTCGTCTTCAGCATGGGGAGGGTGGCATACTCTTGGCTCTGCACTAGCAGG	3181
Query	3020	ATGTTGACATAGATTTTACACAGTTCAGCAGTGCCTTCTAATGGACTCTTAATAACCCCA	3079
Sbjct	3182	ATGTTGACATGGATTTTACACAATTCAGCAGTCTTCTAATGGACTCTTAACAACCCCA	3241
Query	3080	GATTGTACTTGAATTTAATATGCATTGTTTGCTGGT-TCTTGCTTCTATGTTGGTGTCAT	3138
Sbjct	3242	GATTGTACTTGAATTTAATATGC-TTCTTT--TGCGTCGTA-T-CGAAGTTCTT-T--T	3293
Query	3139	ATCATAAAGTTTCTATAAAGACTATACATTGGCATGTAATGTAAGTTACTTCAAATCCAA	3198
Sbjct	3294	-T--T----TTTATATAT-GACTATCCTTTGGCATGAAATTGAAGTTACTG-AAAGCCAA	3344
Query	3199	ATCAGCCTTAGGAatTTTTtGTAACTGAGATGAATTATCAAGCTCAA-AC--GCTTTTT	3255
Sbjct	3345	ATCAACCTTACGATGTTTTTGTTAACTGAGATTAATTATCAAGCTCAATAAATGCTTTT-	3403
Query	3256	AAAAATGTTTGATAAAAATATAAGTTTGGGTCTTTTGTGTTTCCTTTTCATTTTAA	3310
Sbjct	3404	AAAA-TGTTTGATAAAACAACTAAGTTTGGGTGTTT-TATTCCTTTTCATTTTAA	3456

Sammenligning av konsensussekvensen for genotype Qq med konsensussekvensen til paraloge på kromosom 11

```

Query 84   TGGTTCGCGGAGTTGGGGTTATTAACAGAGGATATCCAGGCTTTCAAGCCAGGGTTGTGCG 143
          ||||| ||||| ||||| | | ||||| ||||| |||||
Sbjct 62   TGGTTCGTGGAGTTGGGCGTATTAATATTATTTCTCCAGGCTTTCAAACCAGGGTCGTC- 120

Query 144  GAGGAGTCAAAATGTATACCAGGTTTCAATTCGGAAATGTACATTTTCAAAGTGGAGAGA 203
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 121  --GGAGTCAACATGTCTACCAGGTTTCAATTCAGAAATATACATTTTCAAAGTGGAAAGA 178

Query 204  AATCACTTACAAAGTGGGCGGAGACTAGGAAAAGTGGTTTTTGTGACTGCACCAGCCGC 263
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 179  AATCACTTACAAAGTGGCCGGAGACTAGGAAAAGTGGTTTTTGTGACTGCACCAGCCGC 238

Query 264  ACCAGCTTTCTCTTTCGCTCCGAGGATTCACGCTTCAAAGTAAATGGCGACGGGACGCTG 323
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 239  ACCAGCTTTCTCTTTCCTCCGAGGATTCACACTTCAAAGTAGATGGCGACGGGACGCTG 298

Query 324  AACTGAAGAGGGGGGTGACTCTGCATAATGGACATAAGGAGTCTATGTCTCTACCCAG 383
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 299  AACTGGAGAGGGGGCTGACTCTGCATAATGGACATAAGGAGGTCTATGTCTCTACCCAG 358

Query 384  TCCATGAGCAAGAAGATCAGCTGTTCCAGTCAGAGTGAGTGCATGAGGCCAGACA---CT 440
          |||| | ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 359  TCCAAGGGCAAGAAGATCA-CGGTTCAGTCAGAGTG-CTGCATGAGGCCAGACATGGCC 416

Query 441  ACCA-C-C-ATCACCATGAGATGACCACCCAGCCCCAGCCTGGATCAAGTCTGTCTCTAC 497
          |||| | | ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 417  ACCACCACAATCACCATGAGATGACCACCCAGCCAAGCCAGGAGCAAGTCTGTCTCTAC 476

Query 498  CTGTTCTGAACTTCCCCAAGTCTTCAGGAGGTCTGAAAAGAAGGAAGAGGGACTGGGTCA 557
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 477  CTGTTCTGAACTTCCCCAAGTCTTCAGGAGGTCTGAAAAGAAGGAAGAGGGACTGGGTCA 536

Query 558  TTCCTCCCGTCAACTTCCCAGAGAATGAGCGAGGCCCTTCCCCAAGATTATGGTGCAGA 617
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 537  TTCCTCCCATCAACTTCCCAGAGAATGACCGAGGCCCTTCCCCAAGATTATGGTGCAGA 596

Query 618  TCAGGTCCAGCAATGATAAAGAGGTGAAGATCATGTACAGCATCACTGGCAATGGGCGAG 677
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Sbjct 597  TCAGGTCCAACAATGATAAAGAGGTGAAGATCCAGTACAGCATCACTGGCACTGGGCGAG 656

```

Query 678 ACCAACCTCCTGTGGGACTCTTCACTGTGGACAAAACTCTGGGATTCTCTATGTGACCC 737
 ||| ||||||||||| ||||||||||||||||||||||| |||||||||||
 Sbjct 657 ACCTACCTCCTGTGGGAATCTTCACTGTGGACAAAACTCTGGGAATCTCTATGTGACCG 716

Query 738 AGCCTTTGGACAGGGAGAAAAAGACAAATACATTCTCCTAGCCCATGCTGTTGCTGTGG 797
 |||| ||||||||||||||||||||||||||||||||||||||| ||||
 Sbjct 717 AGCCCTTGGACAGGGAGAAAAAGACAAATACATTCTCCTAGCCCATGCTGTTGCAGTGG 776

Query 798 GTGCAGGTAAGGCTGAGGCACCCATGGAGGTCATTGTGAAAGTCATCGACCAAAATGACA 857
 ||||||| |||||| | ||||||| ||||||||||||||||| |||||||
 Sbjct 777 GTGCAGGTATAGCTGAGGATCCCATGGAGATCATTGTGAAAGTCATCGATATGAATGACA 836

Query 858 ACAAACCTGTATTTACCCAAAATCCATTTATGGGAACAGTCCCTGAGGCAGCAAACCAG 917
 ||||||||||||||||||| ||||||||||||||||||| ||||| ||| |||||||||
 Sbjct 837 ACAAACCTGTATTTACCCAAGATCCATTTATGGGAACAGTTCCTGAAGCATCAAACCAG 896

Query 918 GTGATGAGTTCATGCAGGTAACAGCCACTGATGCTGATGAGGAGGGCTCTGCCAATCTG 977
 |||| ||| ||||||||||||| |||||||||||||||||||||||||||||
 Sbjct 897 GTGACGAGGTCATGCAGGTAACGGCCACTGATGCTGATGAGGAGGGCTCTGCCAATCTG 956

Query 978 ATGTCAGATACACCATTCTCAGTCAGGAGCCTCCACTACCAAGCCCCAACATGTTTGTC 1037
 |||||||||||||||||||||||||||||||||||||||||||||||||||
 Sbjct 957 ATGTCAGATACACCATTCTCAGTCAGGAGCCTCCACTACCAAGCCCCAACATGTTTGTC 1016

Query 1038 TCAACCCTGTGACTGGAGGGATTGGGTTAACGCACCTGGGTTGGACAGAGAGAAAATTC 1097
 ||||| ||||||||||||||||||||||| |||||||||||||||||||||||
 Sbjct 1017 TCAACTCTGTGACTGGAGGGATTGGGTTAATGCACCTGGGTTGGACAGAGAGAAAATTC 1076

Query 1098 TTAAATATACTTTGGAAATCCAAGCTGCTGATTTGGAGGGAGATGGCCTTACCAGCTTTG 1157
 ||||||||||||| ||||||||||||| ||| ||||||||| |||||||||||||||||
 Sbjct 1077 CCAAATATACTTTGGCAATCCAAGCTGCCGATATGGAGGGAAATGGCCTTACCAGCTTTG 1136

Query 1158 GCAAAGCCATCATTACAGTAACGGACAGCAATGACAACGCGCCACAGTTTGTGACGCCTT 1217
 ||||||||||||||||| |||||||||||||||||||||||||||||||||
 Sbjct 1137 GCAAAGCCATCATTACACTAACGGACAGCAATGACAACGCGCCACAGTTTGTGACGCCTT 1196

Query 1218 CGTACACTGTGTCAGTCCCAGAGAATAAAGTAGATGCCTTGGTGGTGAAAATGCCAGTGA 1277
 ||||||| ||||||||||||||||||||||| |||||||||||||||||||||
 Sbjct 1197 CGTACACCGTGTGTCAGTCCCAGAGAATAAAGTGGATGCCTTGGTGGTGAAAATGCCAGTGA 1256

Query 1278 CTGATGGGGATGATCCTCACTCCTCTGCCTGGGCCACCACCTTCAAGATAGTTGACGGGG 1337
 ||||||| ||||| ||||||||||||||||||||||| |||||||||||||||||
 Sbjct 1257 CTGATGGAGATGAGCCTCACTCCTCTGCCTGGGCCACCACCTACAAGATAGTTGACGGGG 1316


```

Sbjct  1916  GGTGCGGGCTTCACTGCTCCATACACCGTACAGCTTCAGGGGTCGTCCCATTCTAACTGG  1975

Query  1997  ACTGCCAGAATGAATGACTCAAAAAGTGGCATTATCCTGACACTGAGTACTATGTTGGAC  2056
|||||||
Sbjct  1976  ACTGCCAGAATGAACGACACAAAGACTGGCATTATCCTGACTCTGAAGACTATGTTGGAC  2035

Query  2057  AGTGGAGATTACACGGTGTCTCCTGAGAGTGTCTGACAACCAGGGCCTGCACCAGGAGAGC  2116
|||||||
Sbjct  2036  AGTGGAGATTACACGGTGTCTCCTGAGAGTGTCTGACAACCAGGGCCTGCACCATGACAGC  2095

Query  2117  ACCATCCTGGCCTCCGTGTGTGACTGCAAAGGAGCTGATGTCCAGTGCGCCAATAA  2172
|||||||
Sbjct  2096  ACCATCCAGGCCTCCGTCTGTGACTGCAAAGGAGCGGATGTCCAGTGCTCCGATAA  2151

HULL

Query  2730  CTACGACCTCCTTCAAGAGTGGGGCCCGCCTTCAAGAAGCTGTCTGACATGTACGGAGG  2789
|||||||
Sbjct  2623  CTACGACCTCCTTCAAGAGTGGGGCCCGCCTTCAAGAAGCTGTCTGACATGTACGGAGG  2682

Query  2790  AGGAGAGGATTGAGAAGTCAGTCAGCCACTAAACCTTTTTGCTTGAAACTGTCCGGTTC  2849
|||||||
Sbjct  2683  AGGAGAGGATTGAGAAGTCAGTCAGCCACTAAACCTTTTTGCTTGAAACTGTCCGGTTC  2742

Query  2850  TCCACACTTTGGTCTGTAAACTGAAGACCTTTTCTTTATTTCCCCTTTTATGCGTTTCG  2909
|||||||
Sbjct  2743  TCCACACTTTGGTCTGTAAACTGAAGACCTTTTCTTTATTTCCCCTTTTATGCGTTTCG  2802

Query  2910  GGCACATTTACAGTTT---A-TTG-T--G--CAACCGTGGAACATTTTAGAGAAT-TG  2959
|||||||
Sbjct  2803  GGCACATTTACAGTTTGGAATTTGATTTGTGCAGACATGGGACCATTTTAGAAAAAATG  2862

Query  2960  TGTTC AATGCTCGTCTT CAGCATGGGAGGGTGGCACACTCTGGCTCTGCACTAGCAGG  3019
|||||||
Sbjct  2863  TGTTC AATGCTCGTCTT CAGCATGGGAGGGTGGCATACTCTGGCTCTGCACTAGCAGG  2922

Query  3020  ATGTTGACATAGATTTTACACAGTTCAGCAGTGCTTTCTAATGGACTCTTAATAACCCCA  3079
|||||||
Sbjct  2923  ATGTTGACATGCATTTTACACAATTCAACAGTTCTTTCTAATGTTTTCTTAACAACGCCC  2982

Query  3080  GATTGACTTGAATTTAATATGCATTGTTT  3109
|||||||
Sbjct  2983  CATG TACTTGAATTTAATATGC-TTCTTT  3011

```

XVIII. Komparativ sammenligning mellom ulike arter

HomoSapiens	MGPWSR-----SLSALLLLLQVSSWLCQ--EPEP--CHPGFDAESYTF	39
PanTroglodytes	MGPWSR-----SLSALLLLLQVSSWLCQ--EPEP--CHPGFDAESYTF	39
MacacaMulatta	MGPWSR-----SLSALLLLLQVSSWLCQ--EPEP--CHPGFDAESYTF	39
CanisLupusFamiliaris	MGPARYGG---APALLLPLLLLLQVSSGLCQ--EPEP--CRPGFGADSYTF	43
BosTaurus	MGPWSR-----SLSALCCCCRCNPWLCR--EPEP--CIPGFGAESYTF	39
MusMusculus	MGARCR-----SFSALLLLLQVSSWLCQELEPEP--CSPGFSSSEVYTF	41
RattusNorvegicus	MGARCR-----SFSALLLLLQVSSWLCQPESESDSCRPGFSSEVYTF	43
GallusGallus	MGRRWGSPALQRFPVLVLLLLQVCGRRCD--EAAP--CQPGFAAETFSF	46
SalmoSalarQQ	MGAFWFV-----ELGVLIIFLQAFKPGSS-----ESTCLPGFNSEIYIF	39
SalmoSalarqq	MGAFWFV-----ELGVLIIFLQAFKPGSS-----ESTCLPGFNSEIYIF	39
TetraodonNigroviridis	-----	
OryziasLatipes	MGTAWIA-----VFGVFFFIFLEVSVVSTTQ---ESPCKPGFESDLLIF	40
DanioRerio	MACVTTV-----GLGVIFLFRVFFSSGYTH----MSICTPGFELEEFVF	40

HomoSapiens	TVPRRHLEGRVLRVNFEDCTGRQRTAYFSLDTRFKVGTGVIIVTKRPL	89
PanTroglodytes	TVPRHLLEGRVLRVNFEDCTGRQRTAYFSLDTRFKVGTGVIIVTKRPL	89
MacacaMulatta	TVPRRHLEGRVLRVNFEDCTGRQRTAYFSLDTRFKVGPDPGVIIVTKRPL	89
CanisLupusFamiliaris	TVPRRHLEGRVLRVNFEGCTGLPRTAYVSDTRFKVGTGVIIVTKRPL	93
BosTaurus	TVPRRNLEGRVLRVNFEGCAGLPRTVYVSDTRFKVHTDGVLTVRRPV	89
MusMusculus	PVPERHLEGRHVLRVRFEGCTGRPRTAFFSEDSRFKVDGDTITVKRHL	91
RattusNorvegicus	LVPERHLEGRHILGRVRFEGCTGRPRTAFFSEDSRFKVDGVIIVTKRHL	93
GallusGallus	SVPQDSVAAGRELGRVFAACSGRPWAVYVPTDTRFKVNGDGVVSTKRPL	96
SalmoSalarQQ	KVERNHLQSGRRLGKVVFDCTSRTSFLFHSEDSHFKVDGDGTLKLERGL	89
SalmoSalarqq	KVERNHLQSGRRLGKVVFDCTSRTSFLFHSEDSHFKVDGDGTLKLERGL	89
TetraodonNigroviridis	-----	
OryziasLatipes	KVNRKHLKAGTRLGKVGFTDCTDRIRFLFSTDDSRFVFTTDGVLTVKRAV	90
DanioRerio	KVHRNHLHSGKRLGKVTFNCDGRTRTLFQSIDKRFEINTDGTVTLKRQV	90

HomoSapiens	RFHNPQIHFLVYAWDSTYRKFS TKVTLNTVGHHR-----P	125
PanTroglodytes	RFHNPQIHFLVYAWDSTYRKFS TKVTLNTVGHHR-----P	125
MacacaMulatta	QFHNQPIHFLVYAWDSTYRKFS TKVTLNTVGHHSR-----T	125
CanisLupusFamiliaris	QLHKPEISFLVHAWDSSRRKLS TRVRLKAATHHHH-----H	129
BosTaurus	HLHRPELSFLVHAWDSTHRKLS TKVTLNLSAHHHH-----H	125
MusMusculus	KLHKLETSFLVHAWDSSHRELSTKVTLKSMGHHHH-----R	127
RattusNorvegicus	KLHKLETSFLVHAWDSSYRKLSTKVTLKSLGHHHH-----R	129
GallusGallus	TLYGRKISFTIYAQDAMGRHSARVTVG--RHRHR-----R	130
SalmoSalarQQ	TLHNGHKEVYVSTQ--SKGKKITVPVRLHEARH-----G	122
SalmoSalarqq	TLHNGHKEVYVSTQ--SKGKKITVPVRLHEARH-----G	122
TetraodonNigroviridis	-----	
OryziasLatipes	VLHEGHLDFVHSDSQGNKMTVPVMLRHGHHVSRDAERHHGGHRHQ	140
DanioRerio	TLHEGHKVFVSHAWDSSGMKHTASVRVERVPAQ-----	123

HomoSapiens	PPHQASVSGIQ-----AELLTFPNSSPG-LRRQKRDWVIPPISCPEN	166
PanTroglodytes	PPHQASVSGIQ-----AELLTFPNSSPG-LRRQKRDWVIPPISCPEN	166
MacacaMulatta	PPLHASVSGVQ-----AELLTFPNSSPG-LRRWKRWDVIPPISCPEN	166
CanisLupusFamiliaris	H--HDAPSKTQ-----TEVLTFPSSQHG-LRRQKRDWVIPPISCPEN	168
BosTaurus	HSHHDSPSGTQ-----TEVLTFPGPHHG-LRRQKRDWVIPPISCPEN	166
MusMusculus	HHHRDPASEN-----PELLMFPSVYPG-LRRQKRDWVIPPISCPEN	168
RattusNorvegicus	HHHRDPVSESN-----PELLTFPSFHQG-LRRQKRDWVIPPINCPEN	170
GallusGallus	HHHNHHLQDFT-----PAVLTFPKHDPGFLRRQKRDWVIPPISCPEN	172
SalmoSalarQQ	HHHNHHEMTTQPKPGASLSLPVLFNFKSSGG-LKRRKRWDVIPPINFPEN	171
SalmoSalarqq	HHHNHHEMTTQPKPGASLSLPVLFNFKSSGG-LKRRKRWDVIPPINFPEN	171
TetraodonNigroviridis	-----MEAQ-----VPVVFHKSSEG-LKRRKRDMVIPPINFPEN	34
OryziasLatipes	HHNTEVDSADTKETAGHPDVAVLYFPKSSKG-LRRRKRWDVIPPINIPEN	189
DanioRerio	-----VESSSD-----VDLTKN-----KRVKRGWIIPPISVSEN	152

: : : * * . : * * . * *

HomoSapiens	EKGFPKPNLVQIKSNKDKGKVFYSITGQGADTPPVGVFIIERETGWLKV	216
PanTroglodytes	EKGFPKPNLVQIKSNKDKGKVFYSITGQGADTPPVGVFIIERETGWLKV	216

MacacaMulatta	EKGPFPPKNLVQIKSNKDKEKGVFYSITGQGADTPPVGVFIERETGWLKV	216
CanisLupusFamiliaris	EKGPFPPKNLVQIKSNRDKEIKVFYSITGQGADAPPVGVFIERETGWLKV	218
BosTaurus	EKGPFPPKSLVQIKSNKEKETQVFYSITGQRADTPPVGVFIERETGWLKV	216
MusMusculus	EKGFEPPKLVQIKSNRDKEIKVFYSITGQGADKPPVGVFIERETGWLKV	218
RattusNorvegicus	QKGEFPQRLVQIKSNRDKETTVFYSITGPGADKPPVGVFIERETGWLKV	220
GallusGallus	HRGPYPMLVQIKSNKDKEKVVYSITGQGADSPVGVFIERETGWLEV	222
SalmoSalarQQ	DRGPFPPKIMVQIRSNNDKEVKIQYSITGTGADLPPVGI FTVDKNSGNLYV	221
SalmoSalarqq	DRGPFPPKIMVQIRSNNDKEVKIQYSITGTGADLPPVGI FTVDKNSGNLYV	221
TetraodonNigroviridis	HRGPYPLKISQIRSNKDDTRIYYSISGPGANQPPVNLFTMDRVSGTLFV	84
OryziasLatipes	SNGPFPLKMTVIRSSEDIKIKKIYYSITGPGANEDPVGLFTMDKDTGDLV	239
DanioRerio	SKGPFPPMLVQIKSDYAIETRLAYKITGEGADLDPKGIFTIDRLSGWVS	202
	. * : * : * : * : * : * : * : * : * : * : * : * : * : * : *	
HomoSapiens	TEPLDRERIATYTLFHAVSSNGNAVEDPMEILITVTDQNDNKPEFTQEV	266
PanTroglodytes	TEPLDRERIATYTLFHAVSSNGNAVEDPMEILITVTDQNDNKPEFTQEV	266
MacacaMulatta	TEPLDRENIATYTLFHAVSSNGNAVEDPMEILITVTDQNDNKPFVFTQEV	266
CanisLupusFamiliaris	TEPLDREQIAKYILYSHAVSSNGNAVEDPMEIVITVTDQNDNKPEFTQAV	268
BosTaurus	TQPLDREQIAKYILFHAVSSNGQAI EEPMEIVITVTDQNDNKPFQEQEV	266
MusMusculus	TQPLDREAIAKYILYSHAVSSNGEAVEDPMEIVITVTDQNDNRPEFTQPV	268
RattusNorvegicus	TQPLDREAIDKYLILYSHAVSSNGEAVEDPMEIVVTVTDQNDNRPEFIQEV	270
GallusGallus	TEQLDREKIDRYTLFHAVSSAGQPVEDPMEI IITVMDQNDNKPVFIKEV	272
SalmoSalarQQ	TEPLDREKKDKYILLAHAVAVGAGIAEDPMEI IVKVIDMNDNKPVFTQDP	271
SalmoSalarqq	TEPLDREKKDKYILLAHAVAVGAGIAEDPMEI IVKVIDMNDNKPVFTQDP	271
TetraodonNigroviridis	TQQLDREQQASYMLKAYAVAE GSGTAE EPMDI IVNVIDQNDNKPAFVKDT	134
OryziasLatipes	HQRLDREKQAHYTLFHAVDV-----EDPMEI IINVIDMNDNKPIFEQTS	283
DanioRerio	TQLLDREKKASYKLRHANGVDADVTEKPMDI IVTVTDQNDNKPVFTQNP	252
	: **** * * : : * * : : * * : : * * : : * * : :	
HomoSapiens	FKGSVMEGALPGTSMVEVTATDADDDVNTYNAAIAYTILSQPELDPKMN	316
PanTroglodytes	FKGSVMEGALPGTSMVEVTATDADDDVNTYNAAIAYTILSQPELDPKMN	316
MacacaMulatta	FKGSVMEGALPGTSMVEVTATDADDDVNTYNAAIAYSILSQPELDPKMN	316
CanisLupusFamiliaris	FQGSVTEGALPGTSMVQVTATDADDDVNTYNAAIAYSILTQDPLPSSMM	318
BosTaurus	FKASALEGALPGTSMVQVTATDIDDEVNTYTAAGYTI PAQDPMLPHNKM	316
MusMusculus	FEGFVAEGAVPGTSMVKVATDADDDVNTYNAAIAYTIVSQPELPHKMN	318
RattusNorvegicus	FEGSVAEGALPGTSMVQVATDADDDINTYNAAIAYTILSQPELPHKMN	320
GallusGallus	FVGYIEENAKPGTSMVTVNATDADDAVNTDNGIVSYSIVSQPPRPHQPM	322
SalmoSalarQQ	FMGTVPEASKPGDEVMTATDADE-EGSANS DVRYTILSQPELPSNM	320
SalmoSalarqq	FMGTVPEASKPGDEVMTATDADE-EGSANS DVRYTILSQPELPSNM	320
TetraodonNigroviridis	FLGEVPEASPTNFEVIQVNATDLDE-PNSYNSDIRYKILSQPEKLPDNL	183
OryziasLatipes	YVAEVAESSPKGTTVIQVKATDADE-PGNDNS I KYTILSQPEKLPDSDM	332
DanioRerio	FNGNVPEALEKGEVFMVTATDADDKENTDNADISYV IISQDPPSPKPNM	302
	: . * . . : * . * * * : . . . : * * : : * * : :	
HomoSapiens	FTINRNTGVISVVTGLDRESFPTYTLVVQAADLQGEGLSTTATAVITVT	366
PanTroglodytes	FTINRNTGVISVVTGLDRESFPTYTLVVQAADLQGEGLSTTATAVITVT	366
MacacaMulatta	FTINKNTGVISVVTGLDRESFPMYTLVVQAADLQGEGLSTTATAVITVT	366
CanisLupusFamiliaris	FTINKDTGVISVLTGLDREGVPMYTLVVQAADLQGEGLTTTATAVITVT	368
BosTaurus	FTINKETGVISVLTGLDRESFPTYTLMVQAADLNGEGLSTTATAVITVL	366
MusMusculus	FTVNRDTGVISVLTGLDRESYPTYTLVVQAADLQGEGLSTTAKAVITVK	368
RattusNorvegicus	FTVNRDTGVISVLTGLDRESYPTYTLVVQAADLQGEGLSTTAKAVITVK	370
GallusGallus	FTIDPAKGI ISVLTGLDRETTPNYTLIVQATDQEGKGLSNTATAI IEVT	372
SalmoSalarQQ	FVINPVTGGIRVNAPGLDREKIPKYTLAIQAADMEGNGLTSFGKAIITLT	370
SalmoSalarqq	FVINPVTGGIRVNAPGLDREKIPKYTLAIQAADMEGNGLTSFGKAIITLT	370
TetraodonNigroviridis	FTINPTTGGIRVNAGGLDREKYPEYTLLEVQAADLEGIGLTVGAKVILT	233
OryziasLatipes	FAINSFNGAIMVEGVGLDREKYPEYTLLEIQAADTKGEGLTGKTKVTLKVT	382
DanioRerio	FAINPVS GGISVLEKGLDREQWFRYTLVITATDMNGEGLSTGTAVITVT	352
	* . : : * :	
HomoSapiens	DTNDNPPIFNPPTYKQVPENEANVITTLKVTADAPNTPAWEAVYTIL	416
PanTroglodytes	DTNDNPPIFNPPTYKQVPENEANVITTLKVTADAPNTPAWEAVYTIL	416
MacacaMulatta	DTNDNPPVFNPTTYKQVPENEANFVITTLKVTADAPNTPAWEAVYTIL	416
CanisLupusFamiliaris	DINDNPPIFNPPTYQGRVPENEANVEIAVLKVTADAVPDPAPWAVYTIL	418
BosTaurus	DTNDNAPRFPNPTTYVGSVPENEANVAITTLVTDADDPNTPAWEAVYTVL	416
MusMusculus	DINDNAPVFNPTTYQGVPENEVNIATLKVTDADDPNTPAWKAVYTVV	418
RattusNorvegicus	DINDNAPVFNPTTYQGVLENEVGARIATLKVTDADDPNTPAWNAVYTVV	420
GallusGallus	DANDNIPVFNPTTYEGVVEENKPGTEVARLTVTDQDAPGSPAWQAVYHIK	422
SalmoSalarQQ	DSNDNAPQFVTPSYTVSVPEKVDALVVKMPVTDGDEPHSSAWATTYKIV	420
SalmoSalarqq	DSNDNAPQFVTPSYTVSVPEKVDALVVKMPVTDGDEPHSSAWATTYKIV	420

TetraodonNigroviridis	DSNDNAPAFQTQASYETSVAENKADSQILRMLVTDGDEPHSAAWNAKFTIV	283
OryziasLatipes	DSNDNPPVFTASTYQGSVDENAVGVLVVKMLVTDEDEPNTPAWNAKFKIV	432
DanioRerio	DSNDNAPLFEQSSYTASVPENQVGVEVAKLPVTDGDEPESTAWSTKYQII	402
	* * * * * . * * * . : : * * * * * : . * * : : :	
HomoSapiens	N-DDGGQFVVTNPNVNDGILKTAKG-----	441
PanTroglodytes	N-DDGGQFVVTNPNVNDGILKTAKG-----	441
MacacaMulatta	N-DNDGQFVVTNPNVNDGILKTAKG-----	441
CanisLupusFamiliaris	N-NNNDQFVVTNPNVNDGILKTAKG-----	443
BosTaurus	N-DNEKQFIVVTDVPTNEGTLKTAKG-----	441
MusMusculus	N-DPDQQFVVTNPNVNDGILKTAKG-----	443
RattusNorvegicus	N-DPDHQFTVITDPKTNEGILKTAKG-----	445
GallusGallus	SGNLDGAFSIIITDPSTNNGILKTAKG-----	448
SalmoSalarQQ	DGDPQGLFNVSTGSPKLEGIITAKP-----	446
SalmoSalarqq	DGDPQGLFNVSTGSPKLEGIITAKP-----	446
TetraodonNigroviridis	SGDPGGFFSVKGTGNKQEGILSTAKVHFHRKFGTLM SRLKKIRTL LLLC	333
OryziasLatipes	GGDPDKFFSIEGTGNKQEGIKTIKG-----	458
DanioRerio	AGDKGGFFNISTGSRLEGIITTVKP-----	428
	: * : * . : * : . * *	
HomoSapiens	-----LDFEAKQQYILHVAVTNVVPEVSLTSTATVTVDVLDVNEAP	484
PanTroglodytes	-----LDFEAKQQYILHVAVTNVVPEVSLTSTATVTVDVLDVNEAP	484
MacacaMulatta	-----LDFEAKQQYILHVAVTNVVPEVSLTSTATVTVDVLDVNEAP	484
CanisLupusFamiliaris	-----LDFEAKQQYILHVAVTNVVPEVSLTSTATVTVDVLDVNEAP	486
BosTaurus	-----LDFEAKQQYILYVAVTNVVPEVSLTSTATVTVDVLDVNEAP	484
MusMusculus	-----LDFEAKQQYILHVRVENEPEFEGSLVPSTATVTVDVLDVNEAP	486
RattusNorvegicus	-----LDFEAKQQYILHVTVENEPEFEGSLVPSTATVTVDVLDVNEAP	488
GallusGallus	-----LDYETKSRDYLVVTVENKVPLSVPITLSTASVLTVDVNEPP	491
SalmoSalarQQ	-----LDFEKNNKYTLTVQNEVPFTISMPSTATVTVNVEDVNEAP	489
SalmoSalarqq	-----LDFEKNNKYTLTVQNEVPFTISMPSTATVTVNVEDVNEAP	489
TetraodonNigroviridis	CPLPTKGLDFEKNKHTLLI AVENEV PFAVALPTATATVTVQDVNEAP	383
OryziasLatipes	-----LDFETNRKHTLLVRVENDVPPFAIPVVTSTATVTVTKDVNEPP	501
DanioRerio	-----LDYEKTKQYILSVIVVNDKDFVGPLPTSTATVTVNVDVNEPP	471
	** : * . : : * : * * : : : * * * * * * * *	
HomoSapiens	IFVPPEKRVEVSEDFGVGQEITSYTAQEPDTFMEQKITYRIWRDTANWLE	534
PanTroglodytes	IFVPPEKRVEVSEDFGVGQEITSYTAQEPDTFMEQKITYRIWRDTANWLE	534
MacacaMulatta	IFVPPEKRVEVSEDFGVGQEITSYTAREPDTFMEQKITYRIWRDAANWLE	534
CanisLupusFamiliaris	IFIPC PKVVSIPEDFGVGQEITSYTAEDPDTYMEQRITYRIWRDAAGWLE	536
BosTaurus	IFVPPQKRVEVPEDFGVGLEITSYTAREPDTFMEQKITYRIWRDTANWLE	534
MusMusculus	IFMPAERRVEVPEDFGVGQEITSYTAREPDTFMDQKITYRIWRDTANWLE	536
RattusNorvegicus	IFVPAEKRRVEVPEDFGVGLEIASYTAREPDTFMEQKITYRIWRDTANWLE	538
GallusGallus	VFVPPIKRVGVPEDLPVGGQVTSYTAQDPDRDMRQKITYRMGSDPAGWLY	541
SalmoSalarQQ	VFTPVEKII RKPEDLPVDSDLVLYTATDPDTARNQKVTYKIRNDNAGWLS	539
SalmoSalarqq	VFTPVEKII RKPEDLPVDSDLVLYTATDPDTARNQKVTYKIRNDNAGWLS	539
TetraodonNigroviridis	IFDPPEKQVSKGEDLPVGTDVVQYASDPDTARKQKVMYRII LNDPAGWLN	433
OryziasLatipes	IFKEKEMTVQKRELD SVESVVIKSEAEDPDPFARENIVKYKI I NDPDNWLK	551
DanioRerio	EFIPKEKFI SRPENLPVGSNLIPFTAIDPDETEKKQNITYRIGNDPSPDWLN	521
	* : * : * : * : * * : : * : * * . * *	
HomoSapiens	INPDTGAISTRAELDREDFEHVKNSTYTALIIATDNG-SPVATGTGTLLL	583
PanTroglodytes	INPDTGAISTRAELDREDFEHVKNSTYTALIIATDNG-SPVATGTGTLLL	583
MacacaMulatta	INPDTGAISTRAELDREDFEHVKNSTYTALIIATDND-HHLC-----	576
CanisLupusFamiliaris	VNPESGAIFTRAELDREDFEHVKNSTYEALIIAIDNG-SPVATGTGTLLL	585
BosTaurus	INPETGAISTRAELDREDDVHVKNSTYTALIIATDNG-SPPATGTGTLLL	583
MusMusculus	INPETGAIFTRAEMDREDAEHVKNSTYVALIIATDDG-SPIATGTGTLLL	585
RattusNorvegicus	INPETGVISTRAEMDREDEHVKNSTYTALIIATDDG-SPIATGTGTLLL	587
GallusGallus	IHPENGIVTATQPLDRES-VHAINSTYKAIILAVDNG-IPDTTGTGTLLL	589
SalmoSalarQQ	INKDTGLIKVKSLMDRES-TFVQDNKYSVIVLGI DND-EIPATGTGTLLI	587
SalmoSalarqq	INKDTGLIKVKSLMDRES-TFVQDNKYSVIVLGI DND-EIPATGTGTLLI	587
TetraodonNigroviridis	VDKETGVIVKTSMDRES-LFVKDNKYTALIGAYDDD-EVPATGTGTLMV	481
OryziasLatipes	VDEDSGLVKVKSMDRES-PNVKDNKYTVLLGAYDND-PVPATGTGTLLI	599
DanioRerio	ITG-SGQIQVKNALDRES-SNVKDGKYKALILALDNDVESPATGTGTLLVI	569
	: . * : . : * * * . : : . * . :	
HomoSapiens	ILSDVNDNAPIPEPTIIFCERNPKPQVINIIDADLPNTPSPFTAELTHG	633
PanTroglodytes	ILSDVNDNAPIPEPTIIFCERNPKPQVINIIDADLPNTPSPFTAELTHG	633
MacacaMulatta	-----FLESHFLDEEVKLIKK-----	592

CanisLupusFamiliaris VLSDVNDNGPIPEPRNMFDFCQKNPQPHVINIIDPDLPPNTSPFTAELTHG 635
 BosTaurus FLDDVNDNGPVPEPRMTDFCQRNPEPHIININDPDLPPNTSPFTAELTHG 633
 MusMusculus VLLDVNDNAPIPEPRNMQFCQRNPPHIIITILDPLPPNTSPFTAELTHG 635
 RattusNorvegicus VLSDVNDNAPIPEPRNMQFCQRNPKPHVITILDPLPPNTSPFTAELTHG 637
 GallusGallus LLQDVNDNGPTPEPRSFIEICSRQPEKQILSVDKDLPPHTYFFKAALEHG 639
 SalmoSalarQQ ELEDVNDNAPTIDESVIRVCNKESSPQLLSVTDKDGAGFTAPYTVQLQGS 637
 SalmoSalarqq ELEDVNDNAPTIDESVIRVCNKESSPQLLSVTDKDGAGFTAPYTVQLQGS 637
 TetraodonNigroviridis LLEDVNDNAPVIEERAFQVCRNKPAPQLLSVTDKDGPGNGAPFSVSLQDP 531
 OryziasLatipes ILEDVNDNAPAIVERETQICKHQRTPALLSVTDKDEAAHSAPYSVAVHEK 649
 DanioRerio ELQDVNDNAPVINERTIKLCNRESAPVLLSITDKDLPPFAGPFKVEPQGD 619

. . . : : .

HomoSapiens ASANWTIQYNDPTQESIILKPKMALEVGDYKINLKLMDNQNKDQVTTLEV 683
 PanTroglodytes ASANWTIQYNDPTQESIILKPKMALEVGDYKINLKLMDNQNKDQVTTLEV 683
 MacacaMulatta -----IAQESIILKPKIALEVGDYKINLKLMDNQQKDQVTTLEV 631
 CanisLupusFamiliaris ASVNWTIEYNDPARESLILKPKKTTLELGDYKINLKLTDNQNKDQVTTLDV 685
 BosTaurus ASVNWTIEYNDQERESLILKPKKTTLELGDHKINLKLIDNQNKDQVTTLDV 683
 MusMusculus ASVNWTIEYNDAAQESLILQPRKDLEIGEYKIHKLADNQNKDQVTTLDV 685
 RattusNorvegicus ASVNWTIEYNDAAQESLILQPRKDLEIGEYKINLKLSDNQNKDQVTTLEV 687
 GallusGallus SSNNWTVEIR--GQDELAMGLKKELEPGEYNI FVKLTDSQGAQVTVQVKA 687
 SalmoSalarQQ SHSNWTARMND-TKTGIILTLKTMLDSDYTVVLRVSDNQLHHDSTIQA 686
 SalmoSalarqq SHSNWTARMND-TKTGIILTLKTMLDSDYTVVLRVSDNQLHHDSTIQA 686
 TetraodonNigroviridis SRSNWTARMND-SKTGIILQLTKALESQYTVVLRVADNRGRDQDNTVQA 580
 OryziasLatipes YSANWTAKMNE-SKTGIEFLKSDIPEGKYKVVLRVADNRGLDQESSINV 698
 DanioRerio TSKNWSVFFNE-TGHFLNIKPSQSQLEQGEYKVVLRVADREGESQENIIQA 668

: : : * . . . : : * : . . .

HomoSapiens SVCDCEGAAGVCRKAQP-VEAGLQIP-AILGILGGILALLILLLLLLFL 731
 PanTroglodytes SVCDCEGAAGVCRKAQP-VEAGLQIP-AILGILGGILALLILLLLLLFL 731
 MacacaMulatta SVCDCEGAAGVCKKAP-VEAGMQIP-AILGILGGILALLILLLLLLFL 679
 CanisLupusFamiliaris FVCDCEGVNSCKRTAPYAEAGLQVP-AILGILGGILALLILLLLLLFLV 734
 BosTaurus HVCDCDGIVSNCRKARP-AEAGLQVP-AILGILGGILAFLLILLLLLLFLV 731
 MusMusculus HVCDCEGTVNNCMKAGI-VAAGLQVP-AILGILGGILALLILLLLLLFL 733
 RattusNorvegicus HVCDCEGTVNNCMKAIS-LEAGLQVP-AILGILGGILALLILLLLLLFL 735
 GallusGallus QVCECEGTAKNCERRSY-IVGGLGVP-AILGILGGILALLILLLLLLFLFA 735
 SalmoSalarQQ SVCDCKGADVQCSKAV---AGFGIS-SILGILGAVLLLLLSSLLLLMFL 732
 SalmoSalarqq SVCDCKGADVQCSKAV---AGFGIS-SILGILGAVLLLLLSSLLLLMFL 732
 TetraodonNigroviridis TVCECTGEDNQCQERVA---AGTNLP-VILGVLGAILLLLLMLVLLLLMFA 626
 OryziasLatipes EVCTCLENGLCSLARIG---GQEEMSPYIWGILGALLLLVIAVLGLYAY 745
 DanioRerio SVCDCKGEAFQCTDKQV---AGIPLF-GVLGVLGGIILLLLLLALLLLMFL 714

** * . . : : * . . . * : : : * *

HomoSapiens RRR--AVVKEPLLPPEDDTRDNVYYYDEEGGGEEDQDFDLSQLHRGLDAR 779
 PanTroglodytes RRR--AVVKEPLLPPEDDTRDNVYYYDEEGGGEEDQDFDLSQLHRGLDAR 779
 MacacaMulatta RRR--AVVKEPLLPPEDDTRDNVYYYDEEGGGEEDQDFDLSQLHRGLDAR 727
 CanisLupusFamiliaris RRR--RVVKEPLLPPEDDTRDNVYYYDEEGGGEEDQDFDLSQLHRGLDAR 782
 BosTaurus RRR--RVVKEPLLPPEDDTRDNVYYYDEEGGGEEDQDFDLSQLHRGLDAR 779
 MusMusculus RRR--TVVKEPLLPPDDTDRDNVYYYDEEGGGEEDQDFDLSQLHRGLDAR 781
 RattusNorvegicus RRR--TVVKEPLLPPDDTDRDNVYYYDEEGGGEEDQDFDLSQLHRGLDAR 783
 GallusGallus RRR--KVEKEPLLPPEDDMRDNVYNYDEEGGGEEDQDYDLSQLHRGLDAR 783
 SalmoSalarQQ RKR--GGEKPELLQEDDVRDNIYYYDEEGGGEEDDQDFDLSVLHRGLDNR 780
 SalmoSalarqq RKR--GGEKPELLQEDDVRDNIYYYDEEGGGEEDDQDFDLSVLHRGLDNR 780
 TetraodonNigroviridis RRR--RGPQKPELLQDTRDNIFFYDEEGGGEEDDQDYDLSVLHRGLDNR 674
 OryziasLatipes NKRQKKGEIGEPFEKD-IIRDNIYHYAEEGGGEDDQNFDSLVLHYGLDNK 794
 DanioRerio RRR--SNSKPELLPDDVDRDNIYYYDEEGGGEEDDQDFDLSVLHRGLDNR 762

: : : : * * : : * * * * * : : * * * * *

HomoSapiens PEVTRNDVAPTLMSPVRYLPRPANPDEIGNFIDENLKAADTDPTAPPYDS 829
 PanTroglodytes PEVTRNDVAPTLMSPVRYLPRPANPDEIGNFIDENLKAADTDPTAPPYDS 829
 MacacaMulatta PEVTRNDVAPTLLSVPQYRPRPANPDEIGNFIDENLKAADSPTAPPYDS 777
 CanisLupusFamiliaris PEVTRNDVAPTLLSVPQYRPRPANPDEIGNFIDENLKAADTDPTAPPYDS 832
 BosTaurus PEVTRNDVAPTLMSPVQYRPRPANPDEIGNFIDENLKAADSPTAPPYDS 829
 MusMusculus PEVTRNDVAPTLMSPVQYRPRPANPDEIGNFIDENLKAADSPTAPPYDS 831
 RattusNorvegicus PEVTRNDVAPTLMSPQYRPRPANPDEIGNFIDENLKAADSPTAPPYDS 833
 GallusGallus PEVTRNDVAPPLMAAPQYRPRPANPDEIGNFIDENLKAADTDPTAPPYDS 833
 SalmoSalarQQ PDVFRNDIAP-TMARPEYRPRPANPADIGNFIDENLKAADNDPTAPPYDS 829
 SalmoSalarqq PDVFRNDIAP-TMARPEYRPRPANPADIGNFIDENLKAADNDPTAPPYDS 829
 TetraodonNigroviridis PEVTRNDVMPNFMFAPQYRPRPANPEEIGNFIDENLKAADNDPTAPPYDS 724


```

OryziasLatipes      PDVLRSDMMPVTMSRPTYRTL PANPDEIGNFIEDNLKAADEDPTAPPYDS 844
DanioRerio          PEVFRNDVAPT FMPAPQYRPR PANPEEIGTFIDDNLKAAADNDPTAPPYDS 812
                    *:* *.*: * :. * * . ***** :*.**.:***:** *****

HomoSapiens         LLVFDYEGSGSEAASSLNSES DDKDQDYDYLNEWGNRFKKLADMYGGG 879
PanTroglodytes     LLVFDYEGSGSEAASSLNSES DDKDQDYDYLNEWGNRFKKLADMYGGG 879
MacacaMulatta      LLVFDYEGSGSEAASSLNSES DDKDQDYDYLNEWGNRFKKLADMYGGG 827
CanisLupusFamiliaris LLVFDYEGSGSEAASSLNSES DQDQDYDYLNEWGNRFKKLADMYGGG 882
BosTaurus          LLVFDYEGSGSEAA TLSSLNSES DQDQDYDYLNEWGNRFKKLADMYGGG 879
MusMusculus        LLVFDYEGSGSEAASSLNSES DQDQDYDYLNEWGNRFKKLADMYGGG 881
RattusNorvegicus   LLVFDYEGSGSEAASSLNSES DQDQDYDYLNEWGNRFKKLADMYGGG 883
GallusGallus       LLVFDYEGSGSEATSLSSLN SSSASDQDQDYDYLNEWGNRFKKLAELYGGG 883
SalmoSalarQQ       LLVFDYEGSGSEAGSLSSLN SSSSGDDQDYDLLQEWGPRFKKLSDMYGGG 879
SalmoSalarqq       LLVFDYEGSGSEAGSLSSLN SSSSGDDQDYDLLQEWGPRFKKLSDMYGGG 879
TetraodonNigroviridis LLVFDYEGSGSDAGSLSSLN SSSSGD-QDYDCLSQWGPRFKKLADMYGGG 773
OryziasLatipes     LLVFDYEGSGSDAGSLSSLN SSSNSGD-QDYNCLTEWGPRFKKLADMYGGG 893
DanioRerio         LLVFDYEGSGSDAGSLSSLN TSSSGNDQDYDFLNEWGNRFKKLADMYGGG 862
                    *****.**: :*****.* *.. ***: * :** *****.:**

HomoSapiens         EDD---- 882
PanTroglodytes     EDD---- 882
MacacaMulatta      EDD---- 830
CanisLupusFamiliaris EDD---- 885
BosTaurus          EDD---- 882
MusMusculus        EDD---- 884
RattusNorvegicus   EED---- 886
GallusGallus       EDDE--- 887
SalmoSalarQQ       ED----- 881
SalmoSalarqq       ED----- 881
TetraodonNigroviridis EDD---- 776
OryziasLatipes     EDDDDML 900
DanioRerio         ED----- 864
                    *:

```