

IMMUNOGLOBULIINIGEENIT
VERTAILEVASTA JA EVOLUTIIVISESTA NÄKÖKULMASTA

Eero Schroderus

Eläinlääketieteen lisensiaatin tutkielma

Eläinlääketieteellisten biotieteiden osasto

Eläinlääketieteellinen tiedekunta

Helsingin yliopisto

2012

TIIVISTELMÄ

Sisällysluettelo

LUETTELO LYHENTEISTÄ	i
1 JOHDANTO	1
2 KIRJALLISUUSKATSAUS	4
2.1 Immunoglobuliini	4
2.1.1 Immunoglobuliinin rakenne	4
2.1.2 Immunoglobuliinien monimuotoisuuden tuottaminen	5
2.1.3 Immunoglobuliini-isotyypit ja luokanvaihto	7
2.2 Immunoglobuliinigeenit eläinryhmittäin	8
2.2.1 Rustokalalut	8
2.2.2 Viuhkaeväiset luukalat	13
2.2.3 Varsieväiset luukalat	17
2.2.4 Tetrapodit	18
3 TAVOITTEET	33
4 AINEISTO JA MENETELMÄT	33
4.1 Geenien etsintä	33
4.2 Geenien annotointi ja toimivuuden arviointi	34
4.3 Fylogeneettinen analyysi	35
4.4 Geenien nimeäminen	35
5 TULOKSET	36
6 POHDINTA	40
7 LÄHDELUETTELO	44

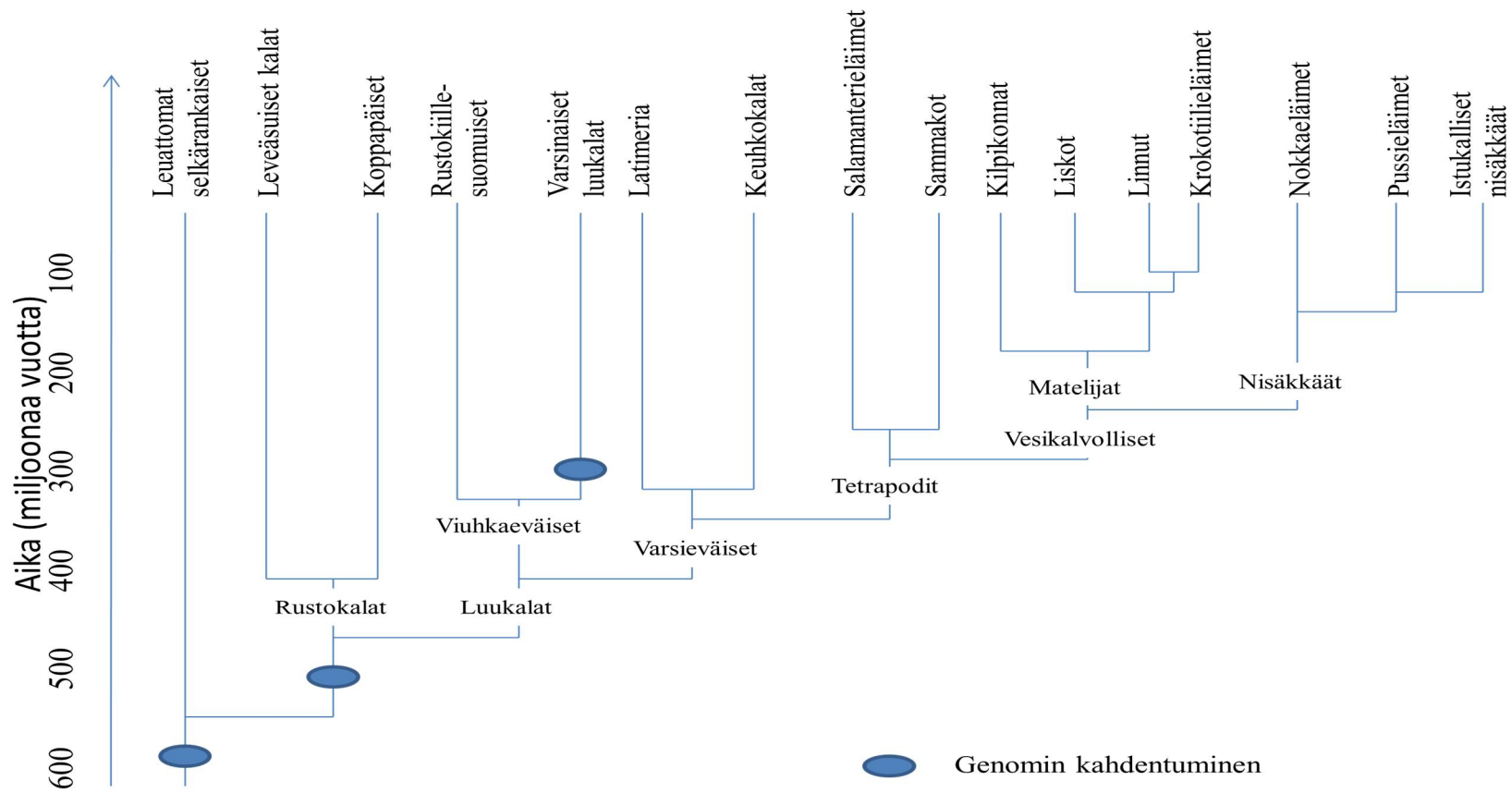
LUETTELO LYHENTEISTÄ

Lyhenne	Selitys
Ig	Immunoglobuliini
V _L	Kevytketjun vaihteleva domeeni
V _H	Raskasketjun vaihteleva domeeni
CDR	Complementary determining region
FR	Framework region
C _L	Kevytketjun vakiodomeeni
C _H	Raskasketjun vakiodomeeni
RSS	Recombination signal sequence
TdT	Terminal deoxynucleotidyl transferase
AID	Activation-induced (cytidine) deaminase
IGHV	Raskasketjun vaihtelevan domeenin V-segmenttiä koodaava geeni
IGHD	Raskasketjun vaihtelevan domeenin D-segmenttiä koodaava geeni
IGHJ	Raskasketjun vaihtelevan domeenin J-segmenttiä koodaava geeni
C _μ	Raskasketjun IgM isotyypin koodaava geeni
C _ω	Raskasketjun IgW isotyypin koodaava geeni
C _δ	Raskasketjun IgD isotyypin koodaava geeni
C _ζ	Raskasketjun IgZ isotyypin koodaava geeni
C _τ	Raskasketjun IgT isotyypin koodaava geeni
C _ν	Raskasketjun IgY isotyypin koodaava geeni
C _α	Raskasketjun IgA isotyypin koodaava geeni
C _ο	Raskasketjun IgO isotyypin koodaava geeni
C _γ	Raskasketjun IgG isotyypin koodaava geeni
C _ε	Raskasketjun IgE isotyypin koodaava geeni
IGLV	Kevytketjun λ isotyypin vaihtelevan domeenin V-segmenttiä koodaava geeni
IGLJ	Kevytketjun λ isotyypin vaihtelevan domeenin J-segmenttiä koodaava geeni
C _λ	Kevytketjun λ isotyypin vakiodomeenia koodaava geeni
IGKV	Kevytketjun κ isotyypin vaihtelevan domeenin V-segmenttiä koodaava geeni
IGKJ	Kevytketjun κ isotyypin vaihtelevan domeenin J-segmenttiä koodaava geeni
C _κ	Kevytketjun κ isotyypin vakiodomeenia koodaava geeni
IGRV	Kevytketjun ρ isotyypin vaihtelevan domeenin V-segmenttiä koodaava geeni
IGRJ	Kevytketjun ρ isotyypin vaihtelevan domeenin J-segmenttiä koodaava geeni
C _ρ	Kevytketjun ρ isotyypin vakiodomeenia koodaava geeni
IGSV	Kevytketjun σ isotyypin vaihtelevan domeenin V-segmenttiä koodaava geeni
IGSJ	Kevytketjun σ isotyypin vaihtelevan domeenin J-segmenttiä koodaava geeni
C _σ	Kevytketjun σ isotyypin vakiodomeenia koodaava geeni
BAC	Bacterial artificial chromosome
EST	Expressed sequence tag

1 JOHDANTO

Leuallisten selkärankaisten hankinnainen immuunijärjestelmä pystyy reagoimaan lähestulkoon mihin tahansa taudinaiheuttajaan (Murphy ym. 2008). Hankinnaisen immuunijärjestelmän tärkeimmät osat ovat kaksi lymfosyyttien linjaa, T- ja B-lymfosyytit. T-lymfosyyttien reseptori (TCR) tunnistaa muiden solujen pinnallaan MHC-molekyyleissä esittelemiä antigeenejä, kun taas B-lymfosyyttien reseptori (BCR) tunnistaa vapaita antigeenejä. Siten T-lymfosyytit osallistuvat ensisijaisesti solunsisäisten ja B-lymfosyytit vastaavasti solunulkoisten taudinaiheuttajien torjuntaan. Taudinaiheuttajien toistuvien rakenteiden tunnistamiseen perustuvassa luonnollisessa immuunijärjestelmässä jokaiselle reseptoriproteiinille on oma geeninsä. Hankinnainen immuunijärjestelmä edustaa täysin päinvastaista strategiaa: jokaisen lymfosyytin antigeenireseptorin ainutlaatuisuus varmistetaan kokoamalla reseptoria koodaava geeni lyhyistä segmenteistä. Hankinnaisessa immuunijärjestelmässä erilaisten antigeenireseptoreiden määrää rajoittaaakin ainoastaan niitä tuottavan solupopulaation koko. Immuunijärjestelmän arvioidaan pystyvän tuottamaan spesifisyydeltään jopa 10^{15} erilaista immunoglobuliinia (Tizard 2009). Osa lymfosyyteistä erilaistuu muistisoluiksi, jotka kykenevät käynnistämään immuunivasteen nopeasti kohdatessaan uudestaan omaan antigeenireseptoriinsa sitoutuvan antigeenin (Murphy ym. 2008). Plasmasoluiksi erilaistuneet B-lymfosyytit erittävät immunoglobuliinia (Ig), joka on BCR:n liukoinen muoto. Immunoglobuliinin tehtävä on neutraloida taudinaiheuttaja takertumalla siihen ja tehostaa sen tuhoutumista ohjaamalla paikalle muita immuunijärjestelmään kuuluvia soluja ja molekyylejä.

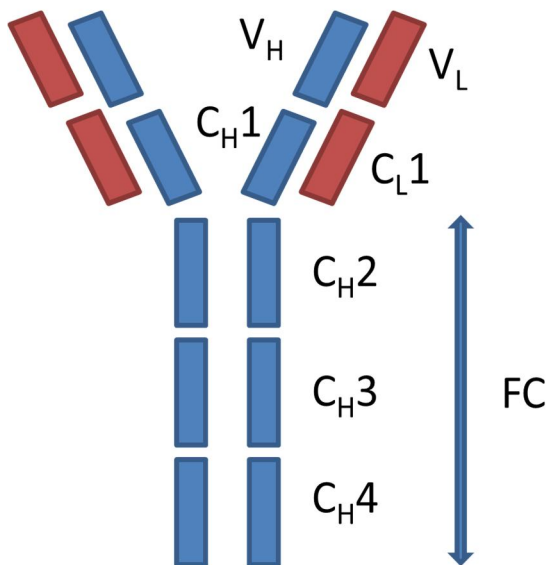
Adaptiivisen immuunijärjestelmän evoluutio ulottuu leuallisia selkärankaista kauemmaksi (Flajnik & Kasahara 2010). Leuattomien selkärankaisten VLR-reseptoreihin perustuva adaptiivinen immuunijärjestelmä hyödyntää osittain samoja entsyymejä kuin leuallisten selkärankaisten adaptiivinen immuunijärjestelmä. Rustokalat ovat kuitenkin vanhin selkärankaisten luokka, jonka adaptiivinen immuunijärjestelmä perustuu MHC-, TCR- ja BCR-reseptoreihin (Cooper & Alder 2006). Leuallisten selkärankaisten adaptiivinen immuunijärjestelmä on siis kehittynyt vajaan sadan miljoonan vuoden aikana, joka erotti leuattomien selkärankaisten ja rustokalojen eroamista omiksi kehityslinjoikseen (Kuva 1; Flajnik & Kasahara 2010). Adaptiivisen immuunijärjestelmän kehittymisen aikoihin selkärankaisten genomissa



Kuva 1. Selkärankaisten evolutiiviset suhteet.

tapahtui suuria muutoksia. Koko genomi kahdentui kahdesti, mikä on saattanut edistää adaptiivisen immuunijärjestelmän kehittymistä (Kuva 1; Kasahara ym. 1997).

Immunoglobuliini koostuu tavallisesti neljästä polypeptidiketjusta, kahdesta identtisestä raskasketjusta (IgH) ja kahdesta identtisestä kevytketjusta (IgL) (Kuva 2; Murphy ym. 2008). Immunoglobuliinien monimuotoisuus saadaan aikaiseksi enimmilläänkin vain muutamasta sadasta geenistä kokoamalla kevyt- ja raskasketjuja koodaavat geenit lyhyemmistä segmenteistä, jolloin rekombinaatio ja geenisegmenttien liitoskohtien epätarkkuus lisäävät monimuotoisuutta. Lisäksi valmiisiin immunoglobuliinigeeneihin kohdistuu somaattisia mutaatioita. Raskasketjun antigeenin tunnistavaa vaihtelevaa osaa koodaava geeni koostuu kolmesta (V, D ja J) ja kevytketjun kahdesta (V ja J) geenisegmentistä. Raskasketjun loppuosaa koodaava C-geeni määrää immunoglobuliinin biologisen aktiivisuuteen.



Kuva 2. Tyypillisen immunoglobuliinin rakenne.

Vaihtelevaa osaa koodaavat immunoglobuliinigeenit jaetaan sekvenssien samankaltaisuuden perusteella ryhmiin ja alaryhmiin. Yli 75 % identtiset geenit lasketaan samaan alaryhmään kuuluviksi (Giudicelli & Lefranc 1999). Vanhemmassa kirjallisuudessa on käytössä myös perheen käsite, joka tarkoittaa yli 80 % identtisiä genejä (Brodeur & Riblet 1984). Erityisesti V-geenien evoluutiota on tutkittu

alaryhmien/perheiden kautta. Alaryhmä voi säilyä satoja miljoonia vuosia, ja hyvinkin kaukaisten lajien alaryhmät voivat vastata toisiaan (Andersson & Matsunaga 1995). Alaryhmiä laajemmissa kokonaisuuksissa esimerkiksi raskasketjun V-geenit (IGHV) luokitellaan viiteen ryhmään (A-E) (Ota & Nei 1994).

Tämä tutkielma alkaa tiiviillä katsauksella immunoglobuliinien rakenteeseen ja mekanismeihin, joilla immunoglobuliinien monimuotoisuus saavutetaan. Huolimatta immunoglobuliinien samankaltaisuudesta kaikilla leuallisilla selkärankaisilla, immunoglobuliinigeenien järjestäytymisessä on suuria eroja (Hsu ym. 2006). Pääosan kirjallisuuskatsauksesta muodostaa immunoglobuliinigeenien vertailu leuallisten selkärankaisten luokkien välillä evolutiivisesta näkökulmasta. Lääketieteellisestä merkityksestä johtuen erityisesti ihmisen ja koe-eläimenä käytetyn hiiren immunoglobuliinigeenit tunnetaan hyvin (Lefranc 2001a; Johnston ym. 2006). Kuitenkin jo pelkästään nisäkkäiden välillä on suuria eroja immunoglobuliinien monimuotoisuuden tuottamisessa (Hsu ym. 2006). Immunoglobuliinigeenien tunteminen yhä useammilla eläinlajeilla auttaa ymmärtämään entistä paremmin hankinnaista immuunijärjestelmää ja sen evoluutiota. Erityisesti useiden eläinlajien perimän sekvensointi viime vuosina on avannut uusia ovia vertailevalle immunologialle.

2 KIRJALLISUUSKATSAUS

2.1 Immunoglobuliini

2.1.1 Immunoglobuliinin rakenne

Immunoglobuliini on suuri Y-muotoinen glykoproteiini, joka tyypillisimmin koostuu neljästä polypeptidiketjusta, kahdesta identtisestä raskas- ja kahdesta identtisestä kevytketjusta (Kuva 2; Murphy ym. 2008). Raskasketjut ovat liittyneet toisiinsa disulfidi-sidoksilla, samoin kevytketjut raskasketjuihin. Joidenkin lajien immunoglobuliineissa Y:n sakarit liittyvät varteen taipuisalla saranaosalla. Immunoglobuliinit koostuvat peräkkäisistä, noin 110 aminohapon mittaisista

laskostuneista proteiimirakenteista, domeeneista. Kevytketju koostuu aina kahdesta domeenista, raskasketjun domeenien lukumäärän vaihdeltaessa immunoglobuliinin isotyypistä riippuen kolmesta jopa seitsemääntoista (Danilova ym. 2005; Tizard 2009). Kummassakin ketjussa ensimmäisenä sijaitsee vaihteleva domeeni (V_L ja V_H) (Murphy ym. 2008). V_L - ja V_H -domeeneissa on kolme aminohappojärjestykseltään erityisen paljon vaihtelevaa aluetta (CDR1, CDR2 ja CDR3). Nämä hypervariaabelit alueet yhdistyvät muodostaen immunoglobuliinin antigeenin tunnistavat alueet määräten immunoglobuliinin sitoutumiskyvyn. CDR-alueiden välissä sijaitsevat vähemmän vaihtelevat FR- eli runkoalueet. V-domeeneja seuraavien vakiodomeenien (C_L ja C_H) aminohappojärjestyksessä on IgH-isotyyppien sisällä vain vähän vaihtelua. Kevytketju ja kaksi raskasketjun ensimmäistä domeenia (V_H ja C_{H1}) muodostavat Y:n sakarit (Kuva 2). Immunoglobuliinin varsi eli Fc-osa määrää immunoglobuliinin isotyypin ja vaikuttaa immunoglobuliinin biologisiin ominaisuuksiin.

2.1.2 Immunoglobuliinien monimuotoisuuden tuottaminen

Ensimmäinen vaihe immunoglobuliinien monimuotoisuuden luomisessa on rekombinaatio (Murphy ym. 2008). V_H -domeenia koodaava geeni koostuu kolmesta (V, D ja J) ja V_L -domeenia koodaava geeni kahdesta (V ja J) geenisegmentistä. Kunkin geenisegmentin lukumäärä perimässä vaihtelee muutamasta satoihin, joista V(D)J-rekombinaatioksi kutsutussa tapahtumasarjassa valikoituu satunnaisesti yksi. V-geeni koodaa noin sadan, ja J-geeni noin kymmenen aminohapon mittaista osaa immunoglobuliinista. D-geenin pituus vaihtelee, sen koodatessa lajista riippuen yhdestä kymmeneen aminohappoa. Lisäksi jotkut lajit käyttävät V(D)J-rekombinaatiossa useampaa D-geeniä. V_L - ja V_H -domeeneja koodaavien geenien järjestely genomissa poikkeaa suuresti eri selkärankaisluokkien välillä (Danilova & Amemiya 2009).

V(D)J-rekombinaatiota ohjaavat geenisegmenttejä reunustavat rekombinaatiosignaalisekvenssit (RSS) (Murphy ym. 2008). RSS koostuu kolmesta osasta: koodaavassa sekvenssissä kiinni olevasta seitsemän emäksen mittaisesta heptameerista, 12 tai 23 emäksen mittaisesta väliskeestä ja yhdeksän emäksen mittaisesta nonameerista. Heptameerin ja nonameerin emäsjärjestys on hyvin konservoitunut, mutta väliskeen emäsjärjestys vaihtelee. Väliskeen pituus kuitenkin vastaa aina yhtä tai kahta DNA:n kaksoiskierrteen kierrosta (12 tai 23 emästä). V(D)J-

rekombinaatiossa RAG1- ja RAG2-proteiinit sitoutuvat RSS:hin, tuovat ne lähelle toisiaan ja katkaisevat DNA-ketjun erottaen koodaavan osan RSS:stä. Pääsääntöisesti geenisegmentti, jota rajaa 12 emäksen mittaisen välikkeen sisältävä RSS, liittyy geenisegmenttiin, jota rajaa 23 emäksen mittaisen välikkeen sisältävä RSS. Tämä ns. 12/23 sääntö takaa rekombinaation tapahtumisen oikeiden geenisegmenttien välillä. Rekombinaatiossa tapahtuu joko DNA:n deleetio tai inversio. Deleetiossa RSS:t ja niiden väliin jäävä DNA poistuu. Jos yhdistettävät segmentit poikkeavat polaarisuudeltaan, tapahtuu inversio, jossa osa DNA:sta kääntyy. Geenisegmenttien käyttö V(D)J-rekombinaatiossa ei ole täysin sattumanvaraista, vaan esimerkiksi V-geeneistä käytetään eniten 3'-pään puoleisia geenejä (Tizard 2009).

V- ja J-geenien liitoskohta muodostaa yhden immunoglobuliinin hypervariaabeleista alueista (Murphy ym. 2008). Raskasketjussa tämä CDR3-alue on osittain myös D-geenin koodaama. V(D)J-rekombinaatiossa yhdistettävien geenien liitoskohtiin satunnaisesti lisättävät ja poistettavat nukleotidit kasvattavat huomattavasti CDR3-alueen vaihtelua. RSS:en poistamisen jälkeen DNA:n toinen juoste katkeaa satunnaisesta kohdasta ja katkaistut nukleotidit liittyvät muodostuvan yksijuosteisen DNA-ketjun jatkoksi, jolloin syntyvää palindromimaista sekvenssiä kutsutaan P-nukleotideiksi. Terminaalinen deoksinukleotidyylitransferaasi (TdT) liittää yksijuosteisten segmenttien päihin satunnaisia nukleotideja, joita puolestaan kutsutaan N-nukleotideiksi. Juosteiden pariutumisen jälkeen eksonukleaasi poistaa parittomat nukleotidit, ja korjausentsyymit täydentävät DNA:n kaksijuosteiseksi. N-nukleotideja lisätään yleensä ainoastaan IgH-geeneihin, koska TdT:n ekspressio vähenee ennen IgL-geenien järjestelyä. Toisinaan liitoskohdasta myös poistetaan nukleotideja, jolloin lyhyestä D-segmentistä ei välttämättä jää mitään jäljelle.

V(D)J-rekombinaatiolla aikaansaatu primaarinen immunoglobuliinivalikoima laajentuu edelleen V-domeenia koodaavaan geeniin kohdistuvilla pistemutaatioilla (Murphy ym. 2008). Ihmisellä somaattinen hypermutaatio tapahtuu itukeskuksissa vasta B-solujen aktivoituttua ja vaatii signaalin myös aktivoituiltä T-soluilta, mutta esimerkiksi lampaalla somaattinen hypermutaatio alkaa jo ennen antigeenin kohtaamista (Reynaud ym. 1995). Somaattisessa hypermutaatiossa sytidiinideaminaasi (AID) muuttaa DNA:n sytidiinin urasiiliksi, joka korvautuu joko tymiinillä tai satunnaisella nukleotidilla

(Murphy ym. 2008). FR-alueita muuttavat mutaatiot tuottavat helposti viallisen BCR:n, mikä johtaa kyseisen B-solun kuolemaan. Toisaalta CDR-alueille keskittyvät mutaatiot parantavat joidenkin kloonien BCR:en sitoutumista antigeeniin. Tässä affiniteettimaturaatioksi kutsutussa ilmiössä immunoglobuliinien sitoutumiskyky antigeeniin paranee immuunivasteen kuluessa.

Ainakin linnuilla ja kaneilla immunoglobuliinien primäärivalikoima kasvaa geenikonversiolla (Parvari ym. 1988; Mage ym. 2006). Somaattisen hypermutaation tavoin myös geenikonversio perustuu AID-entsyymin toimintaan (Murphy ym. 2008). AID:n toiminnan aikaansaama katkos DNA:ssa täytetään käyttäen mallina jotain ylävirtaan sijaitsevaa V-geeniä tai pseudogeeniä.

Eri lajit hyödyntävät immunoglobuliinien monimuotoisuuden tuottamisessa eri mekanismeja eri suhteissa (Tizard 2009). B-soluja läpi elämänsä tuottavilla lajeilla on laajempi rekombinatorinen monimuotoisuus, kun taas lajit, joilla primäärinen immunoglobuliinivalikoima syntyy jo nuoruudessa, luottavat enemmän somaattiseen hypermutaatioon ja/tai geenikonversioon.

2.1.3 Immunoglobuliini-isotyypit ja luokanvaihto

Kaikilta selkärankaisten luokilta tunnetaan useita Ig-isotyyppejä, joiden tehtävä immuunipuolustuksessa vaihtelee (Tizard 2009). Ig-isotyypit luokitellaan toiminnan, sekvenssin ja IgH-lokuksen rakenteen perusteella. Erityisesti nisäkkäillä jotkin isotyypit jakaantuvat useisiin alatyyppeihin. Kaikille selkärankaisten luokille yhteiset IgM ja IgD ovat evolutiivisesti vanhimmat isotyypit. Tetrapodeilla sekä IgM että IgD toimivat aktivoimattomien B-solujen solukalvoreseptoreina, mutta IgM on immuunivasteen ensimmäisessä vaiheessa tuotettava isotyyppi. IgM:n rakenne ja toiminta ovat säilyneet eri selkärankaisten luokilla hyvin samankaltaisena, kun taas IgD on muuttunut huomattavasti evoluution kuluessa. Jotkut lajit ovat jopa menettäneet IgD:n (Edholm ym. 2011a). Rustokaloja lukuunottamatta IgM:ä koodaava C μ -geeni ja IgD:tä koodaava C δ -geeni sijaitsevat peräkkäin ja aktivoimattomassa B-solussa niiden mRNA:t syntyvät vaihtoehtoisella silmukoinnilla samasta transkriptistä (Murphy ym. 2008).

Sammakkoeläimistä lähtien B-lymfosyytit kykenevät vaihtamaan tuottamaansa Ig-isotyyppiä immuunivasteen kuluessa (Flajnik & Kasahara 2010). Luokanvaihto mahdollistaa tietyille antigeenille spesifisen V_H-domeenin liittämisen eri biologisen aktiivisuuden omaaviin C_H-domeeneihin, mutta se vaatii peruuttamattomia muutoksia DNA:ssa (Murphy ym. 2008). Luokanvaihtoa ohjaavat toistuvat DNA-sekvenssit (S, vaihtoalueet, switch regions) sijaitsevat C_H-geeniä edeltävässä intronissa. C δ -geenillä ei ole vaihtoaluetta, koska se ilmentyy C μ :n yhteydessä. B-solun vaihtaessa tuottamaansa Ig-isotyyppiä, tapahtuu rekombinaatio S μ :n ja seuraavaksi tuotettavaa Ig-isotyyppiä koodaavaa C_H-geeniä edeltävän vaihtoalueen välillä. Jokaisen C_H-geenin promoottori sijaitsee vaihtoalueen edellä, ja geenin transkriptio on edellytys luokanvaihdon tapahtumiselle. Transkription aikana AID saa aikaan katkoksia ei-koodaavaan ja välillisesti myös koodaavaan DNA-juosteeseen. Korjausentsyymit tunnistavat katkokset, siirtävät keskenään homologiset vaihtoalueet vierekkäin ja poistavat niiden väliin jäävän DNA:n.

Myös IgL-ketjut voidaan jaotella eri isotyyppisiin (σ -cart, σ , λ , κ). Kaikki neljä esiintyvät ainoastaan rustokaloilla (Criscitello & Flajnik 2007). Maaselkärangaisilta eli tetrapodeilta tunnetaan kolme IgL-isotyyppiä (σ , λ , κ), joista IgL λ on kaikilla lajeilla (Das ym. 2008). IgL-isotyyppien suhteellinen osuus vaihtelee paljon eläinlajeittain (Tizard 2009). IgL-isotyyppien väliltä ei ole löytynyt toiminnallisia eroja, vaikkakin eräissä sairauksissa jokin IgL-isotyyppi voi olla hallitseva (McGavin & Zachary 2007). Lisäksi ainakin kahdesti evoluutio on tuottanut pelkästään IgH-ketjun homodimeerinä esiintyvän immunoglobuliinin, joten IgL-ketju ei liene välttämätön immunoglobuliinin toiminnalle (Pilstrom 2002).

2.2 Immunoglobuliinigenit eläinryhmittäin

2.2.1 Rustokalat

Rustokalat (*Chondrichthyes*) ovat vanhin nykyisistä leuallisten selkärankaisten luokista ja samalla vanhin eläinryhmä, jolla on nisäkkäiden kanssa samoihin molekyyliin perustuva hankinnainen immuunijärjestelmä (Litman ym. 1999). Rustokalat jaetaan kahteen alaluokkaan, leveäsuiset kalat (*Elasmobranchii*) ja koppapäiset (*Holocephali*)

(Kuva 1). Ensin mainittuun kuuluvat rauskut ja hait, jälkimmäiseen sillikuningaskalat. Leveäsuisiin kaloihin ja koppapäisiin johtavat kehityslinjat ovat eronneet toisistaan jo yli 400 miljoonaa vuotta sitten pian sen jälkeen kun rustokalojen ja luukalojen linjat erkanivat (Inoue ym. 2010).

Rustokalojen IgH-geenit

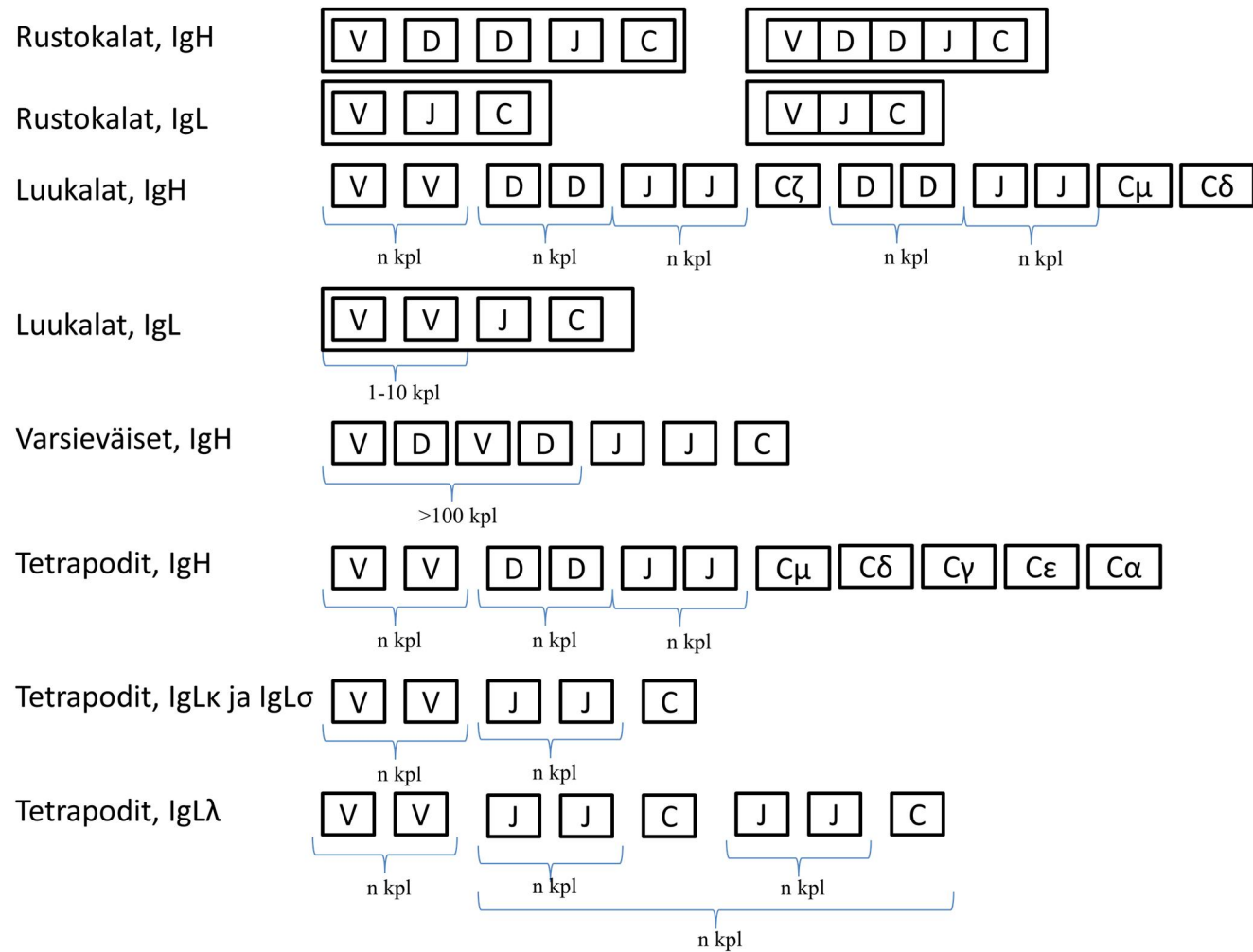
Kaikki rustokalojen IgH-geenit ovat järjestäytyneet yhden IGHV-, kahdesta kolmeen IGHD-, yhden IGHJ- ja yhden C-geenin sisältäviksi minilokuksiksi tai klustereiksi (Kuva 3). Rustokalojen IgH-geenit rekombinoituvat ainoastaan klusterin sisällä (Litman ym. 1999). Rustokaloilta tunnetaan neljä IgH-isotyyppiä (IgM, IgM_{1gj}, IgW ja IgNAR), jotka IgM_{1gj}:tä lukuunottamatta esiintyvät kaikki sekä solukalvoreseptoreina että eritettynä (Rumfelt ym. 2001; Dooley & Flajnik 2006). Eritettävässä IgM:ssä on neljä C_μ-domeenia ja yhtä domeenia pidempi solukalvoreseptori tuotetaan kahden viimeisen eksonin vaihtoehoisella silmukoinnilla (Hsu ym. 2006). IgW vastaa muiden selkärankaisten IgD:tä (Ohta & Flajnik 2006). IgW:n kaksi isotyyppiä (kaksi ja kuusi domeenia) tuotetaan mRNA:n vaihtoehoisella silmukoinnilla. Pitkä muoto on solukalvoreseptorina yhtä eksonia lyhyempi (Rumfelt ym. 2004). Koppapäisiltä ei vielä ole löytynyt IgD:tä vastaavaa immunoglobuliinia (Edholm ym. 2011a). Sekä IgM- että IgW-isotyyppejä koodaavien klustereiden määrä vaihtelee lajeittain hyvin paljon (Taulukko 1). Molempien isotyyppien klustereissa on kaksi D-geeniä. Joillakin rustokaloilla osassa IgM- ja IgW-klustereita V_H-domeenia koodaavat geenit on liitetty jo ituradassa. Liittäminen on voinut tapahtua joko kaikkien V_H-domeenia koodaavien geenien tai ainoastaan V- ja D-geenien (VD) tai edellisen lisäksi myös D- ja J-geenien välillä (VD-DJ) (Kuva 3; Litman ym. 1999). Ituradassa liitetyt geenit ovat kehittyneet tavanomaisista klustereista ja niiden oletetaan tarkoitetun jotain tiettyä taudinaiheuttajaa vastaan (Kokubu ym. 1988).

Kaikki rustokalojen IgM- ja IgW-klustereiden IGHV-geenit kuuluvat selkärankaisten E-ryhmään (Ota & Nei 1994). Kaikilla tutkituilla leveäsuuisilla kaloilla IgM-klustereiden IGHV-geenit muodostavat yhden perheen, jotka ovat syntyneet lajiutumisen jälkeen (Shen ym. 1996; Rumfelt ym. 2004). Atlantinpartahain IgM-klustereiden IGHV-geenit voidaan jakaa edelleen pienempiin ryhmiin, jotka vastaavat C_μ-geenien luokittelua

(Malecek ym. 2005). Vaikka C-geenien uskotaan muuttuvan hitaammin kuin IGHV-geenien, erityisesti C μ 2-domeeneissa on atlantinpartailla paljon vaihtelua (Lee ym. 2008). IgW-klustereiden IGHV-geenit jakaantuvat lajeittain useampiin perheisiin (Shen ym. 1996). Esimerkiksi atlantinpartailla IgW-klustereiden IGHV-geenit muodostavat vähintään kolme perhettä (Rumfelt ym. 2004). Fylogeneettisesti IgW-klustereiden IGHV-geenit vaikuttavat vanhemmilta kuin IgM-klustereiden IGHV-geenit (Shen ym. 1996).

IgM_{1gJ} ei oikeastaan ole oma isotyypinsä, vaan ainoastaan nuorilla atlantinpartailla ilmentyvä IgM:n alatyyppejä, josta puuttuu IgM:n C μ 2:sta vastaava domeeni (Rumfelt ym. 2001). Atlantinpartailla on yksi IgM_{1gJ}-klusteri, jonka V-, D- ja J-geenit on liitetty jo ituradassa (Rumfelt ym. 2001). IgM_{1gJ}:n IGHV-geeni on eronnut IgM:n IGHV-perheestä noin 100 miljoonaa vuotta sitten (Rumfelt ym. 2001).

Neljännessä IgH-isotyypissä, IgNAR:ssa (novel antigen receptor), ei ole lainkaan kevytketjuja, vaan se muodostuu muista Ig-isotyypeistä poiketen IgH-ketjun homodimeerinä (Greenberg ym. 1995). IgNAR ilmentyy IgW:n tavoin kahtena muotona (neljä tai kuusi domeenia). Neljä pidemmän muodon C_H-domeeneista on homologisia IgW:n pitkän ketjun kanssa (Dooley & Flajnik 2006). IgNAR:n V-geenit eivät muistuta sen enempää muita IGHV-geenejä, kuin kevytketjujen tai TCR:n V-geenejä. IgNAR saattaa muistuttaa alkuperäistä hankinnaisen immuunijärjestelmän antigeenireseptoria, joka edelsi nykyisiä TCR:a ja BCR:a, tai joka tapauksessa se ainakin on hyvin vanha reseptoriluokka. Oletettavasti IgNAR:n VDJ-klusteri rekombinoitui jossain vaiheessa IgW:n kanssa (Hsu ym. 2006). Kaikkiaan atlantinpartailla on löydetty neljä IgNAR-klusteria. IgNAR-klustereissa on kaksi tai kolme D-segmenttiä, joiden vaihtoehtoisella rekombinaatiolla saavutetaan laaja primaariantigeenireseptorivalikoima (Hsu ym. 2006). IgNAR-geeneihin kohdistuu myös voimakas somaattinen hypermutaatio immuunivasteen aikan (Dooley & Flajnik 2006).



Kuva 3. Immunoglobuliinigeenien järjestäminen eri eläinryhmillä. Rustokalojen ja luukalojen klusterit kehystetty.

Taulukko 1. Leveäsuisten kalojen IgH-klustereiden lukumääriä.

laji	IgM	IgM _{1gJ}	IgW	IgNAR
Atlantinpartahai ¹	15-25	1	4-7	4
Sarvihai ²	100-200	-	?	-
Ruskohai ³	16	-	?	-
Silonselkärausku ⁴	50	-	>100	-

? = tuntematon määrä; - = IgH-isotyyppiä ei ole löydetty kyseiseltä lajilta

¹ Rumfelt ym. 2001, ²Dooley & Flajnik 2006; ³Shen ym. 1996; ⁴Harding ym. 1990, Dooley & Flajnik 2006

Rustokalojen IgL-geenit

Leveäsuissa rustokaloilla on ainoana eläinryhmänä neljä IgL-isotyyppiä (Criscitiello & Flajnik 2007). Uusimpien tutkimusten mukaan aiemmin nimetyt rustokalojen IgL-isotyypit II ja III vastaavat muiden selkärankaisten IgL λ - ja IgL κ -isotyyppijä.

Vastaavasti rustokalojen (cartilaginous fish) isotyyppi I on nimetty uudelleen σ -cartiksi, koska se muistuttaa muiden selkärankaisten isotyypeistä eniten IgL σ :a. Viimeisenä rustokalojen IgL-isotyypeistä löydettiin luukalojen ja sammakkoeläinten IgL σ :n vastine. Kaikissa leveäsuisten kalojen IgL-klustereissa on yksi V-, yksi J- ja yksi C-geeni, jotka ovat samaa polariteettia (Kuva 3; Edholm ym. 2011b). Klustereiden lukumäärä isotyypeittäin vaihtelee hyvin paljon lajeittain (Taulukko 2). Esimerkiksi ruskohailla on noin 200 IgL λ -klusteria, kun taas sarvihailta on löydetty kymmenen ja atlantinpartahailta ainoastaan kuusi IgL λ -klusteria (Hohman ym. 1993; Rast ym. 1994; Lee ym. 2002). IgL σ -klustereiden lukumäärä on jokaisella tutkitulla lajilla alhaisin. Kaikilla tutkituilla lajeilla IgL λ -klustereiden geenit ovat ituradassa liitettyjä (Dooley & Flajnik 2006). Muissa isotyypeissä ituradassa liitettyjen geenien osuus vaihtelee lajeittain. Esimerkiksi atlantinpartahailla yksi IgL σ -cart-klusteista samoin kuin pieni osa IgL κ -klusteista on liitetty ituradassa (Fleurant ym. 2004; Dooley & Flajnik 2006). Edelleen, kaikki silonselkärauskun IgL σ -cart-klusterit ovat ituradassa liitettyjä, ja sarvihailta kaikki taas ovat liittämättömiä (Dooley & Flajnik 2006). Leveäsuista kaloista poiketen koppapäisiltä on löydetty ainoastaan yksi IgL-isotyyppi. Helmisillikuninkaalla (*Hydrolagus colliei*) on noin 30 IgL λ -klusteria (Rast ym. 1994).

Taulukko 2. Rustokalojen IgL-klustereiden lukumääriä.

laji	λ	κ	σ	σ -cart
Atlantinpartahai ¹	6	noin 60	3	3
Sarvihai ²	10	>3	3	noin 40
Ruskohai ³	noin 200	?	?	?
Silonselkärausku ⁴	noin 15	?	2	89
Helmisillikuningas ⁵	noin 30	-	-	-

? = tuntematon määrä; - = IgL-isotyyppejä ei ole löydetty kyseiseltä lajilta

¹Lee ym. 2000, Fleurant ym. 2004, Criscitiello & Flajnik 2007; ²Rast ym. 1994, Criscitiello & Flajnik 2007; ³Hohman ym. 1993; ⁴Rast ym. 1994, Anderson ym. 1995, Criscitiello & Flajnik 2007; ⁵Rast ym. 1994

2.2.2 Viuhkaeväiset luukalat

Luukalat (*Osteichthyes*) on suuri selkärankaisten yläluokka, joka jaetaan yleensä kahteen luokkaan: viuhkaeväiset (*Actinopterygii*) ja varsieväiset (*Sarcopterygii*) (Benton 2005). Ensimmäiset nykyisen kaltaiset viuhkaeväiset ilmestyivät noin 400 miljoonaa vuotta sitten. Valtaosa viuhkaeväisten nykyisistä noin 25 000 lajista kuuluu *Neopterygii*-alaluokkaan ja edelleen varsinaisten luukalojen (*Teleostei*) osaluokkaan. Viuhkaeväisten toiseen, primitiivisempään alaluokkaan, rustokiillesuomuiset (*Chondrostei*), kuuluu ainoastaan 52 tunnettua lajia, esimerkiksi sampikalat. Varsinaisten luukalojen genomi kahdentui sen jälkeen kun ne olivat eriytyneet omaksi kehityshaarakseen, mikä on saattanut vaikuttaa niiden immunoglobuliinigeenien monipuolisuuteen (Hsu ym. 2006; Steinke ym. 2006).

Viuhkaeväisten luukalojen IgH-geenit

Varsinaisten luukalojen IgH-lokus muistuttaa rakenteeltaan ja geenien ilmentymisen ajoittumisen perusteella T-solujen TCR $\delta\alpha$ -reseptorilokusta (Danilova & Amemiya 2009). Yhden teorian mukaan se voisi edustaa alkuperäistä adaptiivisen immuunijärjestelmän antigeenireseptoria koodaavan lokuksen rakennetta, jonka rustokaloihin ja maaselkärankaisiin johtavat kehityslinjat olisivat menettäneet. Poikkeuksellisen varsinaisten luukalojen IgH-lokuksesta verrattuna muihin selkärankaisten luokkiin tekee ainutlaatuinen IgH-isotyyppi, IgZ, jota koodaavat C ζ -

geenit sijaitsevat ylävirtaan C μ - ja C δ -geeneistä (Kuva 3; Danilova & Amemiya 2009). IgZ löydettiin seeprakalalta (*Danio rerio*) samaan aikaan kun kirjolohelta (*Oncorhynchus mykiss*) löydettiin niin ikään uusi Ig-isotyyppi, joka sai nimen IgT (Danilova ym. 2005; Hansen ym. 2005). IgH-lokuksen rakenteen ja fylogeneettisten analyysien perusteella kirjolohen IgT kuitenkin vastaa IgZ:a (Gambon-Deza ym. 2010). IgZ koostuu valtaosalla lajeista neljästä domeenista, mutta esimerkiksi kolmipiikin (*Gasterosteus aculeatus*) vastaava immunoglobuliini koostuu ainoastaan kolmesta domeenista (Bao ym. 2010a; Gambon-Deza ym. 2010). C ζ - ja C μ -geenit rekombinoituvat eri IGHD- ja IGHJ-geenien kanssa. Seeprakalan C ζ -geeni ja sen kanssa rekombinoituvat IGHD- ja IGHJ-geenit sijaitsevat ylävirtaan C μ - ja C δ -geenien kanssa rekombinoituvista IGHD- ja IGHJ-geeneistä (Kuva 3). IGHD- ja IGHJ-geenit eivät rekombinoitu näiden klustereiden välillä, mutta IgZ- ja IgM-isotyypeissä käytetään samoja IGHV-geenejä (Danilova ym. 2005). Varsinaisten luukalojen IgH-lokuksen rakenteesta voi päätellä, että IgZ ilmentyy ennen kuin IgM, koska jälkimmäisen ilmentymiseen liittyvä DNA:n järjestely poistaa C ζ -geenit. Mahdollisesti IgZ ja IgM ilmentyvät kokonaan eri B-solupopulaatioissa, koska IgZ voi toimia myös BCR:na (Danilova ym. 2005).

IgD on tähän mennessä löytynyt kaikilta tutkituilta luukaloilta (Edholm ym. 2011a). IgD:n mRNA muodostuu silmukoimalla C μ 1-eksoniin lajista riippuen vaihteleva määrä C δ -eksoneita. Kun eri lajien C δ -geenien syntymiseen vaikuttaneet eksonien kahdentumiset otetaan huomioon, varsinaisilla luukaloilla on kaikkiaan seitsemän ainutlaatuista C δ -eksonia (Edholm ym. 2011a). Esimerkiksi seeprakalan C δ -geeni on syntynyt C δ 2-, C δ 3- ja C δ 4-eksonien kahdentuessa useaan kertaan (Danilova ym. 2005). Joillakin lajeilla on lisäksi tapahtunut deleetioita. Esimerkiksi turska (*Gadus Morhua*) on menettänyt eksonit C δ 3-6 (Stenvik & Jorgensen 2000). Monnilla eri IgH-lokus koodaa solukalvoreseptorina toimivaa ja eritettävää IgD:tä (Bengten ym. 2002).

Varsinaisten luukalojen IgH-geenit ovat tyypillisesti jakautuneet useaan lokukseen. Esimerkiksi lohella (*Salmo salar*) koko IgH-lokus on kahdentunut (Yasuike ym. 2010). Toisessa lohen IgH-lokuksista sijaitsee C μ - ja C δ -geenien 5'-puolella viisi ja toisessa kolme C τ -geeniä. Kolme yhteensä kahdeksasta C τ -geenistä on toimivia, joten lohella on kaikkiaan seitsemän toimivaa C H -geeniä. Useimmat lohen C H -geeneistä

rekombinoituvat omien IGHV ja IGHD -geeniensä kanssa, joten eri Ig-luokkien antigeenispesifisyys poikkeaa toisistaan. Monnin (*Ictalurus punctatus*) IgH-lokus koostuu kolmesta C μ - ja C δ -geenin muodostamasta klusterista, joiden syntyyn vaikuttaneista transpositioista ja kahdentumisista johtuen monni on mahdollisesti menettänyt C ζ -geenin (Bengten ym. 2002; Bengten ym. 2006). Kolmipiikin IgH-geenit ovat jakautuneet neljään lokukseen (Bao ym. 2010a; Gambon-Deza ym. 2010). Kolme C ζ -, C μ - ja C δ -geenit sisältävää lokusta vastaavat kukin esimerkiksi seeprakalan IgH-lokusta. Neljäs lokus koodaa ainoastaan IgZ:a.

IGHV-geenien määrä varsinaisilla luukaloilla vaihtelee paljon lajista riippuen (Taulukko 3). Eniten IGHV-geenejä on löydetty monnilta ja lohelta, mutta kummallakin lajilla jopa puolet saattaa olla pseudogeenejä (Yang ym. 2003; Yasuie ym. 2010). Lohella luonnonvalinta on saattanut suosia erityisen laajaa ituradan immunoglobuliinivalikoimaa, koska lohi vaelluskalana altistuu usean eri ympäristön taudinaiheuttajille (Yasuie ym. 2010). Vastaavasti turskalla, jolla on erityisen heikko humoraalinen immuunivaste, ituradan immunoglobuliinivalikoimaa on luukalojen suppeimmasta päästä (Stenvik ym. 2000). Kaikki luukalojen IGHV-geenit kuuluvat selkärankaisten IGHV-geenien ryhmiin C ja D (Ota & Nei 1994).

Taulukko 3. V_H-domeenia koodaavien geenien lukumäärät varsinaisilla luukaloilla. Rekombinatorista monimuotoisuutta ei voi suoraan päätellä IGHV-, IGHD- ja IGHJ-geenien tulona, koska lajista riippuen geenit voivat olla jakautuneena useaan lokukseen, ja C τ / ζ -geenit käyttävät eri IGHD- ja IGHJ-geenejä kuin C μ -geenit.

	IGHV	IGHV- perheitä	IGHD	IGHJ
Monni ¹	200	13	3	9
Seeprakala ²	47	14	7	7
Kirjolohi ³	>130	13	9	7
Lohi ⁴	>300	18	21	25
Turska ⁵	>50	4	5-7	2-3
Kolmipiikki ⁶	49	4	13	16

Jatkoa Taulukko 3. ¹Bengten ym. 2006; ²Danilova ym. 2005; ³Hansen ym. 2005, Brown ym. 2006; ⁴Yasuike ym. 2010; ⁵Solem & Stenvik 2006; ⁶Bao ym. 2010a

Viuhkaeväisten luukalojen IgL-geenit

Kuten rustokaloilla, IgL-geenit ovat varsinaisilla luukaloilla järjestäytyneet klustereiksi (Kuva 3; Litman ym. 1999). Klusterityyppinen IgL-geenien järjestäytyminen on kuitenkin näillä kahdella ryhmällä todennäköisesti kehittynyt erikseen, koska evolutiivisesti varsinaisia luukaloja vanhemmilla rustokiillesuomuisilla IgL-geenit ovat järjestäytyneet samoin kuin tetrapodeilla (Lundqvist ym. 1996). Aivan lopullista tietoa rustokiillesuomuisten IgL-geeneistä ei tosin vielä välttämättä ole, koska ainoastaan yhden isotyypin (IgL κ) geenit on tunnistettu vasta yhdellä lajilla (sampi, *Acipenser baeri*) (Lundqvist ym. 1996). Toisaalta varsinaisten luukalojen IgL-klusterit myös poikkeavat monella tavalla rustokalojen IgL-klustereista. Rustokaloista poiketen varsinaisten luukalojen IgL-klusterit sijaitsevat varsin lähekkäin, ja V-geenien määrä klusteria kohden vaihtelee yhdestä kymmeneen (Edholm ym. 2011b). Lisäksi varsinaisten luukalojen IgL κ -klustereissa V-geenit poikkeavat polaarisuudeltaan J- ja C-geeneistä (Edholm ym. 2011b). Ainakin seeprakalalla IgL κ -isotyypissä voi tapahtua myös klustereiden välisiä rekombinaatioita (Zimmerman ym. 2008). Joidenkin lajien IgL-klustereissa V-geenejä sijaitsee myös C-geenin 3'-puolella (Zimmerman ym. 2008; Bao ym. 2010a). Pääsääntöisesti IgL κ -klustereita vaikuttaa kaikilla tutkituilla lajeilla olevan enemmän kuin IgL λ - tai IgL σ -klustereita (Taulukko 4). Osa lajeista on jopa menettänyt IgL λ :n (Criscitiello & Flajnik 2007).

Taulukko 4. Varsinaisten luukalojen IgL-klusterit.

laji	IgL λ	IgL κ	IgL σ
Monni¹	2	~65	2
Seeprakala²	-	14	2
Kirjolohi³	?	>10	10
Turska⁴	4	20	?
Kolmipiikki⁵	-	29	8

? = tuntematon määrä; - = IgL-isotyyppejä ei ole löydetty kyseiseltä lajilta

Jatkoa Taulukko 4. ¹Edholm ym. 2009, Ghaffari & Lobb 1993; ²Hsu & Criscitiello 2006; ³Daggfeldt ym. 1993; ⁴Daggfeldt ym. 1993, Edholm ym. 2009; ⁵Bao ym. 2010a

Varsinaisten luukalojen IgL-isotyyppien luokittelu oli pitkään ongelmallista, koska niiden ei katsottu vastaavan nisäkkäiden kahta isotyyppiä. Nimeäminen tapahtui lajien ja löytöjärjestyksen perusteella, ja seurauksena olikin varsin sekava termistö (Edholm ym. 2011b). Fylogeneettisten analyysien, geenien järjestymisen ja RSS:en perusteella on kuitenkin osoitettu, että aiemmin löydetyt varsinaisten luukalojen IgL-isotyypit vastaavat muiden selkärankaisten IgL κ - ja IgL σ -isotyyppijä (Taulukko 5; Criscitiello & Flajnik 2007; Edholm ym. 2011b). Lopulta varsinaisilta luukaloilta tunnistettiin myös IgL λ :n vastine (Edholm ym. 2009). Varsinaisten luukalojen C_L-geenit jakautuvat fylogeneettisesti neljään ryhmään siten, että C κ -geenit jakautuvat kahteen ryhmään noudattaen aiemmin käytettyä jakoa L1/G- ja L3/F-isotyyppihin. IGKV-geenit kuitenkin sekoittuvat L1/G-L3/F-jaosta riippumatta, ja niinpä IgL-lokusten V-geenit muodostavat C_L-geeneistä poiketen ainoastaan kolme ryhmää IgL κ -, IgL λ - ja IgL σ -isotyyppien mukaan (Edholm ym. 2009; Edholm ym. 2011b).

Taulukko 5. Varsinaisten luukalojen IgL κ - ja IgL σ -isotyyppien vastaavuus aiemman nimeämiskäytännön kanssa.

laji	IgLκ	IgLσ
Monni	G(L1), F(L3)	
Seeprakala	L1, L3	L2
Kirjolohi	L1, L3	L2
Turska	L1	
Kolmipiikki	L1A, L1B	L2

2.2.3 Varsieväiset luukalat

Varsieväiset luukalat ovat maaselkärankaisten lähimpiä eläviä sukulaisia (Benton 2005). Ensimmäiset varsieväiset ilmestyivät noin 418 miljoonaa vuotta sitten. Varsieväisten luukalojen *Coelacanth*-alaluokan viimeinen elossa oleva edustaja on *Latimeria* (*Latimeria*). Toiseen alaluokkaan (*Rhipidistia*) kuuluvat keuhkokalat (*Dipnoi*). *Coelacanth* ja *Rhipidistia* alaluokat polveutuvat ilmeisesti yhteisestä kantamuodosta,

joka on maaselkärankaisiin johtavan linjan sisarryhmä (Kuva 1; Shan & Gras 2011). Varsieväisten luukalojen geenitutkimusta hidastaa niiden perimän suuri koko, joka voi olla jopa 20-kertainen ihmiseen verrattuna (Stavnezer & Amemiya 2004). Esimerkiksi varsieväisten luukalojen IgL-geenejä ei vielä tunneta lainkaan (Edholm ym. 2011b).

Varsieväisten luukalojen IgH-geenit

Huolimatta evolutiivisesta asemastaan viuhkaeväisten luukalojen ja maaselkärankaisten välillä, varsieväisten luukalojen IgH-geenien järjestäytyminen vaikuttaa olevan rustokalojen ja viuhkaeväisten luukalojen välistä (Stavnezer & Amemiya 2004). Varsieväisiltä luukaloilta tunnetaan kaksi IgH-isotyyppiä, jotka molemmat ilmeisesti koodataan omassa lokuksessaan, koska kukin IGHV-geeni ilmentyy ainoastaan yhtä isotyyppiä edustavissa immunoglobuliineissa (Ota ym. 2003; Stavnezer & Amemiya 2004). Varsieväisten luukalojen IgM ei poikkea muiden selkärankaisten vastaavasta immunoglobuliinista. Toinen isotyyppi taas vastaa muiden selkärankaisten IgW/IgD:tä (Edholm ym. 2011a; Sun ym. 2011). IgW ilmentyy etiopian keuhkokalalla (*Protopterus aethiopicus*) kahtena muotona. IgW:n lyhyt kolmesta domeenista ja pitkä kahdeksasta domeenista koostuva isotyyppi tuotetaan samasta geenistä vaihtoehtoisella silmukoinnilla (Ota ym. 2003). Keuhkokalalta on tunnistettu 13 IGHV-geeniä, jotka jakautuvat kahdeksaan perheeseen (Ota ym. 2003). Latimerian IGHV- ja IGHD-geenit sijaitsevat poikkeuksellisesti parettain (Kuva 3). Näitä VD-klustereita on jopa yli sata kumpaakin IgH-isotyyppiä koodaavissa lokuksissa (Amemiya ym. 1993; Stavnezer & Amemiya 2004).

2.2.4 Tetrapodit

Ensimmäiset tetrapodit (*Tetrapoda*), sammakkoeläimet (*Amphibia*), kehittyivät noin 390 miljoonaa vuotta sitten (Benton 2005). Nykyisten sammakkoeläinten alaluokkien, sammakoiden (*Anura*) ja salamanterieläinten (*Urodela*), viimeinen yhteinen esi-isä eli noin 290 miljoonaa vuotta sitten. Matelijat (*Sauropsida*) kehittyivät kivihiihlikaudella noin 300 miljoonaa vuotta sitten matelijamaisista labyrodonteista, jotka luultavasti olivat ensimmäisiä vesikalvullisia, toisin sanoen kokonaan maaelämään sopeutuneita eläimiä. Matelijoiden kehityshaara jakautui pian kehittymisensä jälkeen nisäkäsmäisiin ja toisaalta edelleen nykyisiin matelijoihin johtaviin linjoihin. Dinosaurukset ja lintujen lähimmät nykyisin elävät sukulaiset, krokotiilieläimet, kehittyivät triaskaudella noin

250-200 miljoonaa vuotta sitten. Linnut puolestaan kehittyivät pienistä teropodidinosauruksista jurakaudella (noin 200-146 miljoonaa vuotta sitten). Nuorimmatkin lintujen nykyisistä lahoista olivat kehittyneet jo noin 65 miljoonaa vuotta sitten (van Tuinen & Hedges 2001). Esimerkiksi immunologiassa paljon tutkittujen kana- ja sorsalintujen viimeinen yhteinen edeltäjä eli 90 miljoonaa vuotta sitten.

Nokkaeläinten (*Monotremata*) esi-isät erosivat pussieläimiin (*Metatheria*) ja istukallisiin nisäkkäisiin (*Eutheria*) johtavasta kehityslinjasta noin 166 miljoonaa vuotta sitten (Bininda-Emonds ym. 2007). Nykyisten nokkaeläinten, vesinokkaeläimen (*Ornithorhynchus anatinus*) ja nokkasiilien lahkoon (*Tachyglossidae*) kehityshaarat erosivat puolestaan noin 64 miljoonaa vuotta sitten. Pussieläinten ja istukallisten nisäkkäiden kehityslinjat erosivat noin 148 miljoonaa vuotta sitten. Nykyisistä pussieläimistä vanhin laho on amerikanpussieläimet (*Didelphimorphia*), joka erosi omaksi linjakseen noin 83 miljoonaa vuotta. Istukallisten nisäkkäiden lahkot kehittyivät noin 85-75 miljoonaa vuotta sitten.

Tetrapodien IgH-geenit

Sammakkoeläimet ovat evolutiivisesti vanhin selkärankaisten luokka, jolla tapahtuu immunoglobuliinin luokanvaihto. Siinä missä kaloilla B-solu tuottaa samaa Ig-isotyyppiä koko ikänsä, tetrapodit voivat muokata immuunivastettaan vaihtamalla B-solun tuottaman immunoglobuliinin biologista aktiivisuutta (Flajnik & Kasahara 2010). Kaikkiaan tetrapodeilla on kolmesta kuuteen immunoglobuliiniluokkaa (Taulukko 6). Pääsääntöisesti tetrapodien IgH-geenit sijaitsevat yhdessä lokuksessa (Stavnezer & Amemiya 2004).

IgM toimii kaikilla tetrapodeilla aktivoimattomien B-solujen solukalvoreseptorina ja on immuunivasteen ensimmäisessä vaiheessa tuotettava immunoglobuliini. Eritetty IgM koostuu kaikilla tetrapodeilla viidestä domeenista, ja plasmassa se esiintyy polymeerinä (Flajnik & Kasahara 2010). Karolinan anolislisko (*Anolis carolinensis*) tuottaa normaalin IgM:n lisäksi lyhyttä solukalvoreseptorina toimivaa IgM:ä, josta puuttuvat

C μ 1- ja C μ 3-domeenit. Molemmat muodot tuotetaan yhden C μ -geenin vaihtoehtoisella silmukoinnilla (Gambon Deza ym. 2009; Wei ym. 2009). IgM on todennäköisesti IgD:tä lukuunottamatta kaikkein muiden IgH-isotyyppien edeltäjä.

Taulukko 6. Immunoglobuliiniluokat tetrapodeilla.

Eläinryhmä	Immunoglobuliiniluokat
Sammakoeläimet	
Sammakot	IgM, IgD, IgY, IgF, IgX
Salamanterieläimet	IgM, IgD, IgY, IgX
Matelijat	
Liskot	IgM, IgD, IgY, IgA
Kilpikonnat	IgM, IgD, IgY
Linnut	IgM, IgY, IgA
Nisäkkäät	
Nokkaeläimet	IgM, IgD, IgG, IgE, IgA, IgO
Pussieläimet	IgM, IgG, IgE, IgA
Istukalliset nisäkkäät	IgM, IgD, IgG, IgE, IgA

IgM:stä poiketen IgD on muuttunut evoluution kuluessa hyvin paljon. Tetrapodeista IgD jopa puuttuu kokonaan linnuilta, pussieläimiltä ja eräitä istukallisilta nisäkkäiltä (Edholm ym. 2011a). Kynsisammakon C δ -geenissä on kahdeksan eksonia, joista C δ 7-8 vaikuttavat syntyneen C δ 5-6-eksoneiden kahdentuessa. Kynsisammakko ei myöskään silmukoi C μ 1-eksonia C δ -geeniin IgD:n ilmentyessä (Zhao ym. 2006a).

Salamanterieläimiin kuuluvalta kylkiluuvesiliskolta (*Pleurodeles waltl*) löytynyt IgP vastanee IgD:tä. Fylogeneettisesti IgP ryhmittyy kynsisammakon IgD:n seuraan (Sun ym. 2011). IgP:tä tuotetaan erityisesti toukkavaiheessa, ja siinä hyödynnetään suhteellisen vähän rekombinatorista kapasiteettia. IgP:tä tuottavat B-solut mahdollisesti vastaavat nisäkkäiden B-1-soluja, jotka toimivat lähinnä luonnollisen immunitetin kaltaisesti (Schaerlinger ym. 2008). Karolinan anoliskolla C δ -geenissä on yksitoista eksonia, mutta vain neljä niistä ilmentyy (Gambon Deza ym. 2009; Wei ym. 2009). Leopardigekolla (*Eublepharis macularius*) on kaksi C δ -geeniä. Pitkä 11 eksonista koostuva geeni vastaa karolinan anoliskon C δ -geeniä. Lyhyemmässä muodossa on seitsemän eksonia, joista osa on ilmeisesti peräisin C α -geenistä (Gambon-Deza &

Sanchez Espinel 2008). Pehmeäkilpikongan (*Pelodiscus sinensis*) IgD:ssä on kuusi C δ -domeenia (Xu ym. 2009).

Vesinokkaeläimen IgD muistuttaa enemmän matelijoiden kuin muiden nisäkkäiden vastaavaa isotyyppiä. Vesinokkaeläimen C δ -geeni on esimerkiksi kymmenellä eksonillaan selvästi pidempi kuin istukallisten nisäkkäiden C δ -geeni (Zhao ym. 2009). Yhdeltäkään pussieläimeltä ei vielä ole löytynyt IgD:tä, joten ne ovat todennäköisesti menettäneet C δ -geenin ennen nykyisten pussieläinlahkojen kehittymistä. Päästäisopossumilla (*Monodelphis domestica*) alue C μ - ja C γ -geenien välissä, jossa C δ -geeni sijaitsee muilla nisäkkäillä, sisältää runsaasti endogeenisiä retroviruselementtejä (Wang ym. 2009). Istukallisten nisäkkäiden C δ -geeni on lyhyt verrattuna muihin selkärangkaisiin, ja sen koodaamassa proteiinissa on saranaosa. Istukallisten nisäkkäiden C δ -geenin eksonit vastaavat vesinokkaeläimen eksoneita C δ 1, C δ 6 ja C δ 7 (Zhao ym. 2009). Kuitenkin myös istukallisilla nisäkkäillä C δ -geenin rakenne vaihtelee hyvin paljon lajien välillä (Sun ym. 2011). Hiiren (*Mus musculus*) ja rotan (*Rattus norvegicus*) IgD:ssä on C δ 2:n puuttuessa vain kaksi C δ -domeenia (Zhao & Hammarstrom 2003). Siialla (*Sus scrofa*) ja märehijöillä C δ 1 on puolestaan korvautunut IgM:n ensimmäisellä domeenilla (C μ 1). Korvautuminen on kuitenkin ilmeisesti tapahtunut erikseen märehijöihin ja sikaan johtavissa linjoissa (Zhao ym. 2002; Zhao ym. 2003). Siialla IgD:n mRNA:han voidaan silmukoida vaihtoehtoisesti C μ 1 tai C δ 1. (Zhao ym. 2003). Lisäksi sika ja märehijät ovat ihmisen ohella ainoita nisäkkäitä, joilla voi tapahtua varsinainen luokanvaihto IgM:stä IgD:hen (Zhao ym. 2002; Chen ym. 2009). Kanilta (*Oryctolagus cuniculus*), afrikannorsulta (*Loxodonta africana*) ja alpakalta (*Lama pacos*) ei ole löytynyt toimivaa C δ -geeniä (Achour ym. 2008; Butler ym. 2009; Guo ym. 2011). Afrikannorsulta on kuitenkin löytynyt C μ -geenin vierestä muiden nisäkkäiden C δ 3-eksonia vastaava sekvenssi (Guo ym. 2011). Alpakalla on kokonainen C δ -geeni, mutta toimimattomaksi mutatoituneena (Achour ym. 2008).

Sammakkoeläimiltä, matelijoilta ja linnuilta löytyvä IgY muistuttaa toiminnaltaan ja rakenteeltaan nisäkkäiden IgG:tä, jonka edeltäjä se todennäköisesti onkin (Warr ym. 1995). Salamanterialäimiin kuuluvalla aksolotlilla (*Ambystoma mexicanum*) IgY toimii IgA:n kaltaisesti suolen limakalvojen suojana (Fellah ym. 1992). IgY on todennäköisesti myös ainoastaan kynsisammakolta löydetyn IgF:n edeltäjä. IgF:n kaksi

C_H -domeenia muistuttavat eniten kynsisammakon IgY:n tai erikseen vertailtuna nisäkkäiden IgG:n domeeneja. IgF:ssä on ainoana sammakkoeläinten Ig-isotyypeistä saranaosa (Zhao ym. 2006a). Joillakin matelijoilla ja linnuilla IgY esiintyy kahtena muotona (Warr ym. 1995). IgY:n lyhyessä muodossa (IgY(Δ Fc)) on tavanomaisen neljän sijaan vain kaksi ensimmäistä C_H -domeenia. Koska IgY(Δ Fc):stä puuttuu Fc-osa, se ei pysty sitoutumaan Fc-reseptoreihin. IgY(Δ Fc) siis kykenee ainoastaan neutraloimaan taudinaiheuttajia ilman immunoglobuliinille tavallista sekundaarista biologista aktiivisuutta. Vielä tuntemattomasta syystä kuitenkin esimerkiksi sorsalintujen IgY:stä valtaosa on lyhyttä muotoa, ja sen määrä lisääntyy immuunivasteen kuluessa (Lundqvist ym. 2006). Sekä karoliinan anoliskolla että ankalla (*Anas platyrhynchos*) IgY ja IgY(Δ Fc) tuotetaan samasta C_V -geenistä vaihtoehtoisella silmukoinnilla (Lundqvist ym. 2006; Wei ym. 2009). Erot mRNA:n silmukoinnissa kuitenkin paljastavat, että IgY(Δ Fc) on syntynyt liskoihin ja lintuihin johtavissa kehityslinjoissa erikseen (Gambon Deza ym. 2009; Wei ym. 2009).

Sammakkoeläinten IgX muodostaa IgM:n tavoin polymeerejä, mutta osallistuu ensisijaisesti limakalvojen immuunipuolustukseen ja siten vastaa toiminnaltaan lähinnä kehittyneempien selkärankaisten IgA:ta (Flajnik 2002). Salamantierieläimistä IgX on löydetty aksolotlilta, mutta ei kylkiluuvesiliskolta (Schaerlinger ym. 2008; Schaerlinger & Frippiat 2008). Vastaavasti matelijoissa karoliinan anoliskoon johtanut kehityslinja on menettänyt C_α -geenin (Gambon Deza ym. 2009). Sekä kynsisammakon ja aksolotlin IgX:ssä että leopardigekon ja lintujen IgA:ssa on neljä C_H -domeenia. Molemmissa isotyypeissä kaksi ensimmäistä domeenia muistuttavat IgY:n C_V1 - ja C_V2 -domeeneja, ja kaksi seuraavaa IgM:n $C_\mu3$ - ja $C_\mu4$ -domeeneja (Gambon Deza ym. 2007; Schaerlinger & Frippiat 2008). Selvästikin IgX/A on syntynyt IgM:n ja IgY:n rekombinaation tuloksena, mutta toistaiseksi ei voida sanoa, ovatko IgX ja IgA homologisia vai konvergentin evoluution tulosta (Gambon Deza ym. 2007; Gambon-Deza & Sanchez Espinel 2008). Lintujen IgA on kuitenkin homologinen nisäkkäiden IgA:n kanssa (Lundqvist ym. 2001; Lundqvist ym. 2006). Lintujen IgH-lokus on järjestynyt poikkeavalla tavalla, siten että C_α -geeni on siirtynyt C_μ - ja C_V -geenien väliin ja sen polaarisuus on vaihtunut. Immunoglobuliinin luokanvaihdon yhteydessä on siis C_μ -geenin deleetion sijaan käännettävä C_μ - ja C_α -geenit sisältävä blokki, mikä voi selittää IgA:n viivästyneen ilmentymisen ankalla. C_α -geenin insertio C_μ -geenin viereen

voi myös selittää C δ -geenin puuttumisen linnuilta (Zhao ym. 2000; Lundqvist ym. 2001).

Nisäkkäiden IgA on menettänyt alempien selkärankaisten C α 2-domeenin. Lisäksi nisäkkäiden IgA:han on kehittynyt toisen eksonin jatkeeksi saranaosa (Tucker ym. 1981). Jo nokkaeläinten IgA edustaa tätä rakennetta, joten IgA:n saranaosa on kehittynyt hyvin varhain nisäkkäiden kehityksen aikana (Vernersson ym. 2010). Nokkaeläimillä on kaksi C α -geeniä, jotka poikkeavat toisistaan hyvin paljon. Todennäköisesti C α -geenin kahdentuminen nokkaeläimillä tapahtui varhaisessa vaiheessa (Vernersson ym. 2010). Istukallisista nisäkkäistä kanilla on peräti kolmetoista IgA-isotyyppiä, jotka poikkevat toisistaan erityisesti C α 1-domeenin ja saranaosan perusteella (Burnett ym. 1989). Kanin IgA-isotyypeillä saattaa olla toiminnallisia eroja, koska niiden ilmentyminen on kudosspesifistä (Spiekerpolet ym. 1993).

IgG ja IgE ovat ainoastaan nisäkkäillä tavattavia Ig-isotyyppiä (Flajnik 2002). IgG on nisäkkäiden yleisin immunoglobuliini ja osallistuu muun muassa komplementin aktivoimiseen. IgE puolestaan osallistuu loisinfektioiden torjuntaan ja toimii pääasiassa basofiilien ja syöttösolujen aktivoijana (Murphy ym. 2008). IgG ja IgE lienevät kehittyneet alempien selkärankaisten IgY:stä (Warr ym. 1995). IgE:ssä on vielä IgY:n tavoin neljä C H -domeenia, mutta IgG on menettänyt yhden C H -domeenin, ja lisäksi IgG:hen on kehittynyt saranaosa (Murphy ym. 2008). Ainoastaan vesinokkaeläimeltä löydetty IgO saattaa olla IgY:n ja IgG:n välimuoto. IgO:ssa on IgY:n tavoin neljä C H -domeenia, mutta siinä on jo saranaosa kuten IgG:ssä. Lisäksi C δ -geeni sijaitsee genomissa välittömästi C δ -geenin jälkeen ja on fylogeneettisesti lähellä nisäkkäiden C γ -geenejä (Zhao ym. 2009). Istukallisilla nisäkkäillä C γ -geenien lukumäärä vaihtelee lajeittain huomattavasti (Taulukko 7). Eri IgG-alytyypeille on kehittynyt toisistaan poikkeavia tehtäviä (Stavnezer & Amemiya 2004).

Ihmisen IgH-lokuksen rakenne C H -geenien osalta poikkeaa muista nisäkkäistä (Flanagan & Rabbitts 1982). Ihmisellä on ensin kahdentunut C γ -geeni, ja sen jälkeen (C γ -C γ -C ϵ -C α)-blokki, johtaen varsin monimutkaiseen rakenteeseen (C μ -C δ -C γ -C γ -C ϵ -C α -C γ -C γ -C ϵ -C α). Ensimmäinen C ϵ -geeneistä on pseudogeeni. Ihmisen C H -geenien

lukumäärissä on paljon polymorfismia (Rabbani ym. 1996; Brusco ym. 1997). Tältä osin vertailu muihin lajeihin on vaikeaa, koska muilla lajeilla IgH-lokusta ei ole tutkittu yhtä intensiivisesti.

Kameleilla (*Camelidae*) on rustokalojen lisäksi ainoana eläinryhmänä Ig-isotyyppi, josta puuttuvat kevytketjut (Hamerscasterman ym. 1993). Kameleiden HCAb on IgG:n alatyyppejä, joka kevytketjujen lisäksi on menettänyt C γ 1-domeenin. Kameleiden HCAb-geenissä on C γ 1-domeenia vastaava eksoni, mutta se poistetaan mRNA:n silmunoinnin yhteydessä (Nguyen ym. 1999). Esimerkiksi dromedaarilla (*Camelus dromedarius*) viisi kaikkiaan yhdeksästä C γ -geenistä on toimivia, ja niistä kaksi ilmentyy perinteisinä immunoglobuliineina ja kolme IgH:n homodimeerina (De Genst ym. 2006). Alpakkan neljästä C γ -geenistä kaksi ilmentyy IgH:n homodimeerina ja kaksi perinteisessä immunoglobuliinissa (Achour ym. 2008).

Taulukko 7. Immunoglobuliiniluokkien alatyyppeiden lukumääriä istukallisilla nisäkkäillä.

laji	IgG	IgE	IgA
Ihminen ¹	4	1	2
Sika ²	11	1	1
Nauta ³	3	1	1
Lammas ⁴	2	1	1
Hiiri ⁵	4	1	1
Kani ⁶	1	1	13
Dromedaari ⁷	5	1	1
Alpakka ⁸	4	1	1
Afrikannorsu ⁹	8	?	?

? = ei tietoa

¹Rabbani ym. 1996; ²Butler ym. 2009; ³Zhao ym. 2006b; ⁴Jenne ym. 2006; ⁵Shimizu ym. 1982; ⁶Mage ym. 2006; ⁷de Genst 2006; ⁸Achour ym. 2008; ⁹Guo ym. 2011

Pääsääntöisesti tetrapodien IgH-lokuksessa IGHV-, IGHD- ja IGHJ-geenit ovat järjestäytyneet omiksi ryhmikseen (Kuva 3; Flajnik & Kasahara 2010). Tetrapodien

IGHV-geenit kuuluvat selkärankaisten A-, B- tai C-ryhmään, joista kahteen ensin mainittuun kuuluu ainoastaan tetrapodien geenejä (Ota & Nei 1994). Nisäkkäiden IGHV-geenit jaetaan perinteisesti kolmeen klaaniin, siten että I klaani sisältyy selkärankaisten ryhmään A, II klaani ryhmään B ja III klaani ryhmään C (Ota & Nei 1994).

Sammakkoeläinten ja matelijoiden IGHV-geeneille ominaista on suuri määrä pieniä perheitä, mikä kertoo hitaasta erilaistumisesta. Kynsisammakolla on vähintään 40 IGHV-geeniä jakautuneena yhteentoista perheeseen (Haire ym. 1990). IGHD-geenejä on ainakin 17, ja IGHJ-geenejä ainakin 10. Kynsisammakko saattaa käyttää osassa rekombinaatioita kahta IGHD-geeniä, mikä lisää CDR3-alueen monimuotoisuutta. Aksolotlilla on jopa 70 IGHV-geeniä, vähintään kolme IGHD-geeniä ja yhdeksän IGHJ-geeniä (Golub & Charlemagne 1998). Aksolotlin IGHV-geenit jakaantuvat 11 perheeseen, joista löytyy yhteneväisyyksiä esimerkiksi kirjolohen, kynsisammakon ja ihmisen IGHV-perheiden kanssa. Eri IGHV-perheisiin kuuluvia geenejä käytetään epätasaisesti, aksolotlille ominaisten geenien käytön ollessa vähäisintä. Karolinan anolisiskolta on löydetty 71 IGHV-, 13 IGHD- ja kahdeksan IGHJ-geeniä (Gambon Deza ym. 2009). Sen IGHV-geenit jakautuvat 28 perheeseen. Kahdella vähiten vaihtelevalla karolinan anoliskon IGHV-perheellä on selvä vastine kynsisammakon IGHV-perheissä.

Linnut poikkeavat eniten muista tetrapodeista immunoglobuliinien monimuotoisuuden tuottamisessa (Flajnik 2002). Kombinatorisesti saavutettava monimuotoisuus on lintujen immunoglobuliineissa erittäin alhainen. Kana (*Gallus gallus domesticus*) käyttää V(D)J-rekombinaatiossa ainoastaan yhtä IGHV-geeniä (Parvari ym. 1988). Ankalla toimivia IGHV-geenejä saattaa olla useampia (Lundqvist ym. 2006). Kanalla on 15 keskenään hyvin samankaltaista IGHD-geeniä, joista kana käyttää V(D)J-rekombinaatiossa säännöllisesti vain kahta (Reynaud ym. 1991). Ankalla on yksi tai kaksi IGHD-geeniä (Lundqvist ym. 2001). Sekä kanalla että ankalla on ainoastaan yksi IGHJ-geeni (Lundqvist ym. 2001). V(D)J-rekombinaatiossa muodostettava geeni on linnuille ainoastaan lähtökohta, josta geenikonversiolla luodaan varsinainen immunoglobuliinien monimuotoisuus. Geenikonversio tapahtuu linnuilla alaperärauhasessa (Fabriciuksen bursa) varsin lyhyenä ajanjaksona kuoriutumisen

jälkeen, jolloin muodostuvalla immunoglobuliinivalikoimalla linnun on tultava toimeen koko ikänsä (McCormack ym. 1991). Kanalla sijaitsee V(D)J-rekombinaatiossa käytetyn IGHV-geenin 5'-puolella suuri määrä geenikonversiossa luovuttajina toimivia pseudogeenejä (Reynaud ym. 1994). Pseudogeeneit vaikuttavat itse asiassa fuusioituneilta VD-geeneiltä (Ratcliffe 2006). Kanan toimiva IGHV-geeni ja pseudogeeneit kuuluvat yhteen tiiviiseen perheeseen (Ota & Nei 1995; Lundqvist ym. 2006). Kanan ja ankan IGHV-geeneit kuuluvat selkärankaisten IGHV-geenein C-ryhmään, johon kuuluu nisäkkäiden III klaanin lisäksi esimerkiksi latimerian, kynsisammakon ja kaimaanin geenejä (Ota & Nei 1994; Sitnikova & Su 1998).

Nisäkkäistä primitiivisimmällä, vesinokkaeläimellä, ilmentyy ainakin 25 IGHV-geeniä (Johansson ym. 2002). Ilmentyvät vesinokkaeläimen IGHV-geeneit ryhtymät nisäkkäiden III klaanin sisällä täysin erilleen muiden lajien geeneistä ja ovat mahdollisesti lähtöisin yhdestä geenistä vesinokkaeläimen omaksi kehityslinjaksi eriytymisen jälkeen. Tutkitut geeneit kuuluvat ilmeisesti muutamaan toisiaan lähellä olevaan perheeseen. Vesinokkaeläimen IGHV-perheiden sisällä on tavanomaista enemmän vaihtelua CDR1- ja CDR2-alueilla. Lisäksi vesinokkaeläimen immunoglobuliinien CDR3-alueet ovat hyvin pitkät ja rekombinaatiossa käytetään mahdollisesti useampia IGHD-geenejä. IGHD-geenejä vesinokkaeläimellä on useita, mutta tarkka lukumäärä ei ole selvillä. Genomisesta DNA:sta vesinokkaeläimeltä on löydetty 11 IGHJ-geeniä, joista yksi on pseudogeeni (Zhao ym. 2009). Vesinokkaeläimestä poiketen nokkasiilin (*Tachyglossus aculeatus*) IGHV-geeneit jakautuvat kaikkiin kolmeen nisäkkäiden klaaniin (Belov & Hellman 2003).

Nykyisten pussieläinten yhteinen esi-isä on jossain vaiheessa menettänyt valtaosan IGHV-lokuksen geneettisestä vaihtelusta, joka on alkanut rakentua uudestaan yhden perheen tai ehkäpä vain yhden geenin kautta. Kaikki tutkitut pussieläinten IGHV-geeneit kuuluvat nisäkkäiden III klaaniin, jonka sisälläkin ne muodostavat oman ryhmänsä (Baker ym. 2005). Amerikanpussieläimiin kuuluvalla päästäisopossumilla on 25 IGHV-geeniä, joista 23 kuuluu yhteen perheeseen (Wang ym. 2009). Kahteen muuhun päästäisopossumin IGHV-perheeseen kuuluu kumpaankin yksin geeni. Niistä toinen on fylogeneettisesti etäisin muista tutkituista pussieläinten IGHV-geeneistä ja sijaitsee myös kromosomissa kauimpana IGHD-geeneistä, ja toinen on ainoa nisäkkäiltä löydetty

ituradassa VD-liitetty immunoglobuliinigeeni. Kettukusun (*Trichosurus vulpecula*) IGHV-valikoima on muistuttaa päästäisopossumia (Baker ym. 2005). Kettukusulla ilmentyy 16 IGHV-geeniä, jotka kahta lukuun ottamatta kuuluvat yhteen perheeseen. IGHD-geenejä päästäisopossumilla on yhdeksän, joista seitsemän on toimivia (Wang ym. 2009). Poikkeuksellisesti yksi IGHD-geeneistä sijaitsee IGHV-geenien seassa, ja sitä käytetään rekombinaatiossa ainoastaan siitä 5' suuntaan sijaitsevien IGHV-geenien kanssa.

Ihmisellä seitsemään perheeseen jakautuneet IGHV-geenit edustavat kaikkia kolmea nisäkkäiden IGHV-klaania (Taulukko 8; Matsuda ym. 1998). Myös muilla laajan IGHV-valikoiman lajeilla kuten hiiri, rotta ja afrikannorsu on IGHV-geenejä kaikista kolmesta klaanista (Johnston ym. 2006; Hendricks ym. 2010; Guo ym. 2011). Rotalla (*Rattus norvegicus*) on kaikista tutkituista lajeista eniten IGHV-geenejä (Taulukko 8; Hendricks ym. 2010). Vain noin puolet hiiren ja kolmannes ihmisen ja rotan IGHV-geeneistä on toimivia (Johnston ym. 2006; Hendricks ym. 2010; Guo ym. 2011). Kanilla on 100-200 IGHV-geeniä, joista valtaosa kuitenkin lienee pseudogeenejä (Gallarda ym. 1985; Ros ym. 2004). Suuresta lukumäärästä huolimatta kaikki tähän mennessä tutkitut kanin IGHV-geenit kuuluvat samaan perheeseen (Ros ym. 2004). 70-90 %:ssa kanin immunoglobuliineja on käytetään proksimaalisinta IGHV-geeniä, josta tosin löytyy kolme alleelivarianttia (Mage ym. 2006). Kani käyttää immunoglobuliinivalikoiman laajentamiseen lintujen tapaan geenikonversiota (Mage ym. 2006).

Kaikilla tutkituilla kärkiastujiin (*Ungulata*) kuuluvilla lajeilla on ihmistä tai jyräjöitä suppeampi IGHV-valikoima (Taulukko 8). Naudalla (*Bos taurus*) ja sialla ilmentyy ainoastaan yksi IGHV-perhe, joskin varsinkin naudan kohdalla tutkimukset ovat toistaiseksi olleet varsin suppeita (Jenne ym. 2006; Zhao ym. 2006b; Butler ym. 2009). Lampaalla ilmentyviä IGHV-perheitä saattaa olla jopa kahdeksan (Berens ym. 1997; Charlton ym. 2000). Naudan ainoa tunnettu ilmentyvä IGHV-perhe kuuluu nisäkkäiden II klaaniin, kun taas sian ainoa IGHV-perhe kuuluu III klaaniin (Berens ym. 1997; Butler ym. 2009). Hevosella (*Equus caballus*) IGHV-valikoima on hieman laajempi käsittäen 50 geeniä, joista tosin vain 12 vaikuttaa toimivilta. Hevosien IGHV-geenit jakautuvat seitsemään perheeseen ja kaikkiin kolmeen klaaniin, mutta yksikään klaanin

III kuuluvista geneistä ei vaikuta toimivalta. Hevosella on myös paljon IGHD-genejä verrattuna muihin kärkiastujiin (Sun ym. 2010). Koiran (*Canis familiaris*) IGHV-valikoima vaikuttaa kärkiastujia laajemmalla (Bao ym. 2010b). Koiralla on noin 80 IGHV-geeniä, joista noin puolet on toimivia. Koiran IGHV-geenit jakautuvat kolmeen perheeseen ja kaikkiin kolmeen nisäkkäiden klaaniin, mutta neljää geeniä lukuun ottamatta kuuluvat yhteen III klaanin perheeseen (Bao ym. 2010b).

Kameleilla on tavanomaisissa immunoglobuliineissa käytettävien IGHV-geenien lisäksi erikseen HCAb-immunoglobuliineissa käytettävät IGHVH-geenit (Nguyen ym. 2002). Esimerkiksi dromedaarilla on noin 40 IGHVH- ja noin 50 IGHV-geeniä (Nguyen ym. 2000). Sekä IGHV- että IGHVH-geenit rekombinoituvat samojen IGHD- ja IGHJ-geenien kanssa. Kevytketjun mukana HCAb on menettänyt osan kombinatorisesta monimuotoisuudesta, mutta kompensatioksi IGHVH-geenien CDR-alueet ovat pidentyneet. Kaikki kamalien IGHV- ja IGHVH-geenit kuuluvat III klaaniin, jonka sisällä ne muodostavat oman tiiviin ryhmänsä.

Taulukko 8. Raskasketjun vaihtelevaa osaa koodaavat geenit erällä istukallisilla nisäkkäillä.

	IGHV	IGHV-perheet	IGHD	IGHJ
Ihminen ¹	123	7	27	10
Hiiri ²	195	16	10	4
Rotta ³	353	13	21	5
Afrikannorsu ⁴	112	7	87	6
Lammas ⁵	45	9	?	6
Nauta ⁶	>20	1-?	7	6
Sika ⁷	~20	1	1	1
Hevonen ⁸	50	7	40	8
Koira ⁹	80	3	6	3
Alpakka ¹⁰	88	3	7	7

¹Matsuda ym. 1998; ²Johnston ym. 2006, Ye 2004, Solin & Kaartinen 1992; ³Hendricks ym. 2010; ⁴Guo ym. 2011; ⁵Jenne ym. 2006, Dufour & Nau 1997; ⁶Zhao ym. 2006b, ⁷Butler ym. 2009; ⁸Sun ym. 2010; ⁹Bao ym. 2010b; ¹⁰Achour ym. 2008

Kaikkien istukallisten nisäkkäiden IGHJ-lokus on varsin samankaltainen (Dufour & Nau 1997). Fylogeneettisesti istukallisten nisäkkäiden IGHJ-geenit voidaan jakaa seitsemään ryhmään. Esimerkiksi ihmisen, hiiren, lampaan ja kanin IGHJ-geenit jakautuvat tasaisesti näihin ryhmiin. Mahdollisesti ryhmät vastaavat nykyisten istukallisten nisäkkäiden yhteisen esi-isän IGHJ-lokuksen geenejä.

Tetrapodien IgL-geenit

Kaikkien tetrapodien IgL-lokusten rakenne on hyvin yhdenmukainen (Das ym. 2008). IgL κ -lokuksessa, samoin kuin ainoastaan sammakoilta löytyvässä IgL σ -lokuksessa, on yksi C_L-geeni, josta ylävirtaan sijaitsee muutamia J- ja useita V-geenejä. Tetrapodien IgL λ -lokuksessa sen sijaan voi olla useita C_L-geenejä, jotka kukin rekombinoituvat yhden tai useamman J-geenin kanssa. Lajista riippuen nämä JC-blokit saattavat edelleen rekombinoitua kukin omien tai yhteisten V-geenien kanssa (Kuva 3).

Sammakoilla on ainoana tetrapodien ryhmänä kolme IgL-isotyyppiä, kun taas salamanterieläimiltä on löytynyt toistaiseksi ainoastaan IgL λ (Das ym. 2008). IgL κ -lokusta vastaavassa IgL ρ -lokuksessa on troppiikkikynsisammakolla (*Xenopus tropicalis*) vähintään kymmenen IGRV-, yhdeksän IGRJ- ja yksi C ρ -geeni (Qin ym. 2008). IgL σ -lokuksessa on vähintään kahdeksan IGSV-, neljä IGSJ- ja yksi C σ -geeni. IgL λ :a vastaavassa troppiikkikynsisammakon tyyppin III IgL-lokuksessa on vähintään 17 IGLV-, viisi IGLJ- ja kolme C λ -geeniä. Kahta C λ -geeniä edeltää yksi ja yhtä kolme IGLJ-geeniä. Troppiikkikynsisammakon IgL λ -lokuksessa on enemmän vaihtelua kuin IgL κ - tai IgL ρ -lokuksissa. IGLV-geenit muodostavat kuusi, ja IGRV- ja IGSV-geenit kummatkin yhden perheen. Aksolotlin IgL λ -lokuksesta on löytynyt vain yksi C λ -geeni (Andre ym. 2000). IGLV-geenejä aksolotlilta on löydetty kahdeksan, ja ne jakautuvat neljään perheeseen. Aksolotlin IgL λ -geenit tunnetaan ainoastaan cDNA:sta, joten tulokset saattavat muuttua genomi tasolle siirryttäessä. Sekä troppiikkikynsisammakon että aksolotlin IgL-lokusten V-geenit muistuttavat enemmän muiden tetrapodien kuin rusto- tai luukalojen geenejä (Das ym. 2008).

Karolinan anolisliskon IgL-geenit vastaavat järjestäytymiseltään nisäkkäiden geenejä (Wu ym. 2010). IgL κ -lokuksessa karolinan anolisliskolla on yksi C κ -geeni, kaksi IGKJ-geeniä 14 IGKV-geeniä. Kaikki IGKV-geenit kuuluvat yhteen perheeseen. IgL λ -lokuksessa löytyy enemmän vaihtelua. C λ -geenejä on kolme, ja kuhunkin niistä liittyy yksi IGLJ-geeni. 37 IGLV-geeniä jakautuvat vähintään 10 perheeseen. Osa karolinan anolisliskon IGLV perheistä on yhteisiä kynsisammakon kanssa. Kaikki karolinan anolisliskon IgL-geenit ilmentyvät, joten liskoilla VJ-rekombinaatio saattaa olla tärkeä osa immunoglobuliinien monimuotoisuuden luomisessa.

IgL λ on lintujen ainoa IgL-isotyyppi (Das ym. 2008). Linnut käyttävät IgL-ketjun monimuotoisuuden saavuttamiseen samoja mekanismeja kuin IgH-ketjun tapauksessa (McCormack ym. 1989). Kana, ankka ja varpuslintuihin kuuluva seeprapeippo (*Taeniopygia guttata*) käyttävät VJ-rekombinaatiossa yhtä IGLV- ja yhtä IGLJ-geeniä (Parvari ym. 1987; Lundqvist ym. 2006; Das ym. 2010). Myskisorsalla (*Cairina moschata*) toimivia IGLV-geenejä on muista tutkituista linnuista poiketen kaksi (McCormack ym. 1989). VJ-rekombinaatiossa käytettävää IGLV-geeniä täydentää suuri määrä geenikonversiossa käytettäviä pseudogeenejä (McCormack ym. 1989). Kanalla IGLV-pseudogeenejä on noin 25 (Ratcliffe 2006). Ainakin ankalla toimiva IGLV-geeni ja pseudogeeni kuuluvat samaan perheeseen (Lundqvist ym. 2006). Lintujen toimiva IGLV-geeni on läheisintä sukua liskojen ja sammakkoeläinten 3' pään puoleisimmille IGLV-perheille (Wu ym. 2010).

Vesinokkaeläimellä IGKV-geenit jakautuvat neljään ja nokkasiilillä yhdeksään perheeseen, mutta geenien kokonaismäärää ei tiedetä kummallakaan lajilla (Nowak ym. 2004; Johansson ym. 2005). Fylogeneettisesti nokkaeläinten IGKV-geenit ryhmittyvät samaan haaraan muiden nisäkkäiden IGKV-geenien kanssa (Nowak ym. 2004). Vesinokkaeläimen IgL λ -lokuksessa neljä C λ -geeniä assosioituvat jokainen yhdestä neljään IGLJ-geenin kanssa (Johansson ym. 2005). 29 IGLV-geeniä muodostavat kaksi perhettä, joista toiseen kuuluu 25 ja toiseen neljä geeniä. Molemmat perheet ryhmittyvät fylogeneettisesti erilleen muiden nisäkkäiden IGLV-geeneistä, joten vesinokkaeläimen IGLV-perheet ovat syntyneet lajin erottua nisäkkäiden yhteisestä kehityslinjasta. Huolimatta alhaisemmasta geeniperheiden määrästä IgL λ -lokuksessa verrattuna IgL κ -lokukseseen, λ -ketjut ovat κ -ketjuja vaihtelevampien CDR-alueiden ansiosta

monimuotoisempia. λ -ketjun käyttö on myös kymmenen kertaa κ -ketjua yleisempää vesinokkaeläimen immunoglobuliineissa.

Raskasketjun geneistä poiketen, pussieläinten kevytketjun V-geenit ryhmittyvät fylogeneettisesti muiden nisäkkäiden geenien sekaan (Baker ym. 2005; Wang ym. 2009). Pussieläinten IgL-geeneissä on siis enemmän vaihtelua kuin IgH-geeneissä, joten pussieläimillä kevytketju saattaa vastata suuremmasta osasta immunoglobuliinien monimuotoisuudesta kuin raskasketju (Baker ym. 2005). Päästäisopossumin IgL κ -lokuksessa on yksi C κ -geeni ja kaksi IGKJ-geeniä, mutta peräti 122 IGKV-geeniä (Wang ym. 2009). Päästäisopossumin IGKV-geenit jakautuvat seitsemään perheeseen, joista neljä suurinta käsittää valtasosan geneistä. Kettukusulla on noin neljäkymmentä IGKV-geeniä, jotka jakautuvat viiteen perheeseen (Baker ym. 2005). Päästäisopossumin IGLJ- ja C λ -geenit ovat järjestäytyneet kahdeksaksi J-C-pariksi (Wang ym. 2009). 64 IGLV-geeniä jakautuvat neljään perheeseen, joista suurimpaan kuuluu 54 geeniä ja pienimpään ainoastaan yksi. Kettukusulla on noin kolmekymmentä IGLV-geeniä neljään perheeseen jakautuneena (Baker ym. 2005).

Monilla istukallisilla nisäkkäillä joko IgL λ - tai IgL κ -lokus sisältää huomattavasti toista laajemman ituradan geenivalikoiman (Taulukko 9). Erot ituradan geenivalikoimassa saattavat heijastua myös λ - ja κ -ketjujen käytön yleisyyteen immunoglobuliineissa (Tizard 2009). Ihmisen genomissa IGKV-geenit ovat järjestäytyneet kahteen erilliseen ryhmään, joista toinen puuttuu osasta haplotyyppistä (Lefranc 2001b). Kanilla on tetrapodien yleisestä IgL κ -lokuksesta poiketen kaksi C κ -geeniä, joista kuitenkin pääasiassa käytetään vain toista (Benammar & Cazenave 1982). Lisäksi molemmat kanin C κ -geenit rekombinoituvat omien IGKJ-geeninsä kanssa (V_n-(J)₅-C-(J)₃-C) (Emorine & Max 1983; Ayadi ym. 1991). IGKJ-lokus on hyvin samanlainen kaikilla istukallisilla nisäkkäillä, ja IGKJ-geenit eri lajeilla vastaavat pitkälti toisiaan jopa järjestystä myöten (Butler ym. 2006). Istukallisten nisäkkäiden IGKV-geenit ryhmittyvät fylogeneettisesti erilleen alempien selkärankaisten geneistä, mutta istukallisten nisäkkäiden sisällä kaikkien lajien IGKV-perheet ryhmittyvät sekaisin (Sitnikova & Nei 1998; Sun ym. 2010).

Istukallisten nisäkkäiden IGLV-geenit jakautuvat useampiin ryhmiin kuin IGKV-geenit, mutta IGLV-geeneissäkin eri lajien perheet ryhmittyvät fylogeneettisesti sekaisin (Sitnikova & Su 1998; Sun ym. 2010). Hiiren IgLλ-lokus on istukallisten nisäkkäiden yksinkertaisin (Storb ym. 1989). Hiiren IgLλ-geenit ovat järjestäytyneet kahdeksi klusteriksi, joihin kuuluu yksi ja kaksi IGLV-geeniä (V2-Vx-(J2-C2)-(J4-C4)-Vl-(J3-C3)-(J1-C1)) (Hayzer 1990). Hiiren poikkeuksellinen IgLλ-lokuksen rakenne on syntynyt hiiren ja rotan kehityslinjojen eroamisen jälkeen. Rotan IgLλ-lokuksessa on 10-15 IGLV-geeniä ja yksi ilmentyvä Cλ-geeni (Aguilar & Gutman 1992). Myös hevosen IgLλ-lokuksen rakenne mahdollisesti poikkeaa istukallisten nisäkkäiden enemmistöstä. Hevoselta on löytynyt Cλ-geeneistä alavirtaan niiden kanssa vastakkaista polariteettia olevia IGLV-geenejä, mikä tosin saattaa selittyä myös virheellä genomikokoonpanossa (Sun ym. 2010).

Taulukko 9. Kevytketjun vaihtelevaa domeenia koodaavat geenit erällä istukallisilla nisäkkäillä.

	IGKV	IGKV- perheet	IGKJ	IGLV	IGLV- perheet	IGLJ-Cλ
Ihminen ¹	46-82	7	5	96	11	(J-C) ₇
Hiiri ²	164	18	5	3	2	2x(J-C) ₂
Afrikannorsu ³	153	8	3	15	6	(J-C) ₃
Lammas ⁴	6	4	3	>100	6	(J-C) ₂
Nauta ⁵	22	4	3	63	8	J-(J-C) ₂ -C ₃
Sika ⁶	>60	2	5	?	?	(J-C) _n
Hevonen ⁷	60	7	5	144	11	(J-C) ₇

? = ei tietoa

¹Lefranc 2001b,c; ²Martinez-Jean 2001; ³Guo ym. 2011; ⁴Jenne ym. 2006; ⁵Ekman ym. 2009; ⁶Butler ym. 2009; ⁷Sun. ym 2010;

3 TAVOITTEET

Naudan IgH-lokus, varsinkin vaihtelevaa domeenia koodaavien geenien osalta, on vielä huonosti tunnettu (Zhao ym. 2006b). Tähän mennessä tutkijat ovat olleet yksimielisiä ainoastaan yhden IGHV-alaryhmän ilmentymisestä, mutta ituradan IGHV-alaryhmien tai -geenien lukumäärästä ei ole saatu varmuutta. Naudan IGHV-geenejä on toistaiseksi tutkittu pääasiassa cDNA-tasolla tai hybridisaatiolla ihmisen ja hiiren sekvenssien kanssa (Berens ym. 1997; Saini ym. 1997). Naudan perimän selvittäminen (Elsik ym. 2009) mahdollisti vihdoin immunoglobuliinigeenien ituradan vaihtelun tutkimisen. Julkaistuun genomikokoonpanoon pohjautuen naudän immunoglobuliinin kevytketjua koodaavat geenit on jo raportoitu (Ekman ym. 2009). Tämän tutkielman kokeellisessa osassa on tarkoituksena ottaa tarkasteluun naudän IGHV-geenit. Tarkemmin ilmaistuna tavoitteena on:

- tunnistaa IGHV-geenit naudän genomikokoonpanosta
- arvioida naudän IGHV-geenien toimivuutta ja rakennetta
- suorittaa naudän IGHV-geenien fylogeneettinen analyysi

4 AINEISTO JA MENETELMÄT

4.1 Geenien etsintä

Tutkimuksen aineistona käytettiin kahta julkaistua versiota naudän genomista (Btau_4.2 ja UMD_3.1) ja genomisekvensoinnin raakadataa (Elsik ym. 2009; Zimin ym. 2009). Naudän genomiprojekti perustuu kahteen hereford-rotuiseen yksilöön (L1 Domino 99375 ja L1 Dominette 01449). Pääosin sekvenssidata on Dominetesta, mutta sitä on täydennetty Domineten isään Dominoon perustuvalla BAC-kirjastolla (Bovine Genome Project 2011). Genomikokoonpanoista Btau_4.2 on genomiprojektin virallinen ja UMD_3.1 Marylandin yliopiston samasta sekvensointiaineistosta kokoama kilpaileva versio. Tässä tutkimuksessa päädyttiin käyttämään molempia genomikokoonpanoja, koska varsinaisesta naudän genomiprojektin julkaisemasta versiosta löytyi alustavissa analyyseissä todella vähän IGHV-geenejä.

Molemmista genomikokoonpanoista suoritettiin iteratiivinen blastn-haku Geneious-ohjelman kautta (expect threshold-arvo $< 10^{-5}$) (Altschul ym. 1990; Drummond ym. 2011). Kyselysekvensseinä käytettiin aikaisemmin julkaistuja naudän ja lampaan IGHV-geenejä ja -mRNA-sekvenssejä, sekä kaikkia IMGT:n listaamia ihmisen ja hiiren IGHV-geenejä (Lefranc ym. 2011). Lisäksi haku suoritettiin genomisekvensoinnin raakadatasta, high throughput -sekvensseistä ja NCBI:n trace-sekvensseistä. Lopuksi IGHV-geenejä haettiin vielä UMD_3.1-genomikokoonpanon täydellisistä kromosomeista aiemmissä hauissa löytyneiden naudän IGHV-geenien RSS:llä. Viimeisessä vaiheessa oli tarkoitus löytää ne geenit, jotka poikkeavat aikaisemmin julkaistusta naudän ja lampaan sekä ihmisen ja hiiren geeneistä niin paljon, että jäävät löytymättä blastn-haussa.

4.2 Geenien annotointi ja toimivuuden arviointi

Löydettyjen geenien toimivuutta arvioitiin neljällä eri kriteerillä. Geenin promoottorin oktameerin tuli olla konservoitunut, lukukehyksen ehjä ja RSS:en IGHV-geeneille tyypilliset (Taulukko 10). Lisäksi geenin tuli koodata proteiiniin tietyt konservoituneet aminohapot omille paikoillaan.

Taulukko 10. IGHV-geenien konservoituneet promoottorisekvenssit ja RSS:t.

	Sekvenssi
Promoottorin oktameeri	5'ATGCAAAT3'
RSS (Heptameeri)	5'CACAGTG3'
RSS (Nonameeri)	5'ACAAAAACC'3

IMGT:n unique numbering -järjestelmän mukaan immunoglobuliini-molekyylillä on kysteini paikalla 23, tryptofaani paikalla 41, leusiini tai metioniini paikalla 89 ja kysteini paikalla 104 (Lefranc 1997). Löytyneistä IGHV-geeneistä tehtiin rinnastus, jonka perusteella selvitettiin, onko kullakin geenillä konservoituneita aminohappoja koodaavat kodonit oikeilla paikoilla. Geenien FR- ja CDR-alueet annotoitiin konservoituneiden aminohappojen perusteella Geneious-ohjelmassa.

Geenien ilmentyminen analysoitiin käyttäen IGHV-sekvensseillä haettua suurta EST-sekvenssikokoelmaa. Kutakin EST-sekvenssiä sovitettiin IGHV-geeneihin Geneioksen Assembly-toiminnolla.

4.3 Fylogeneettinen analyysi

Fylogeneettisessä analyysissä käytettiin IGHV-geenien FR1-FR3 -alueiden sekvenssejä. Fylogeneettisen puun muodostamiseen vaadittu rinnastus tehtiin MAFFT-ohjelmalla (Kato ym. 2002), ja puu rakennettiin bayesilaisin menetelmin MrBayes-ohjelmalla (Huelsenbeck & Ronquist 2001). Analyysissä käytettiin HKY85-substituutiomallia ja 100 000 toistoa. Tässä tutkimuksessa löydettyjen naudan IGHV-geenien lisäksi fylogeneettiseen analyysiin otettiin mukaan kaikki aikaisemmin julkaistut naudan (GenBank access number: BTU55164-BTU55175) ja lampaan (GenBank access number: Z49180-Z49188) genomiset IGHV-sekvenssit, satunnaisia muiden nautaa lähellä olevien lajien IGHV-sekvenssejä ja geenejä kaikista ihmisen ja hiiren IGHV-alaryhmistä. Ulkoryhmänä käytettiin kanan IGHV-sekvenssiä (GenBank access number: Gallus_CHKIGHVAA, M30319).

Sekvenssi-identtisyysmatriisi laskettiin MAFFT-ohjelmalla tehdystä rinnastuksesta BioEdit-ohjelmalla (versio 7.0.9). Sekvenssiltään yli 75 % identtiset geenit laskettiin samaan alaryhmään kuuluvaksi.

4.4 Geenien nimeäminen

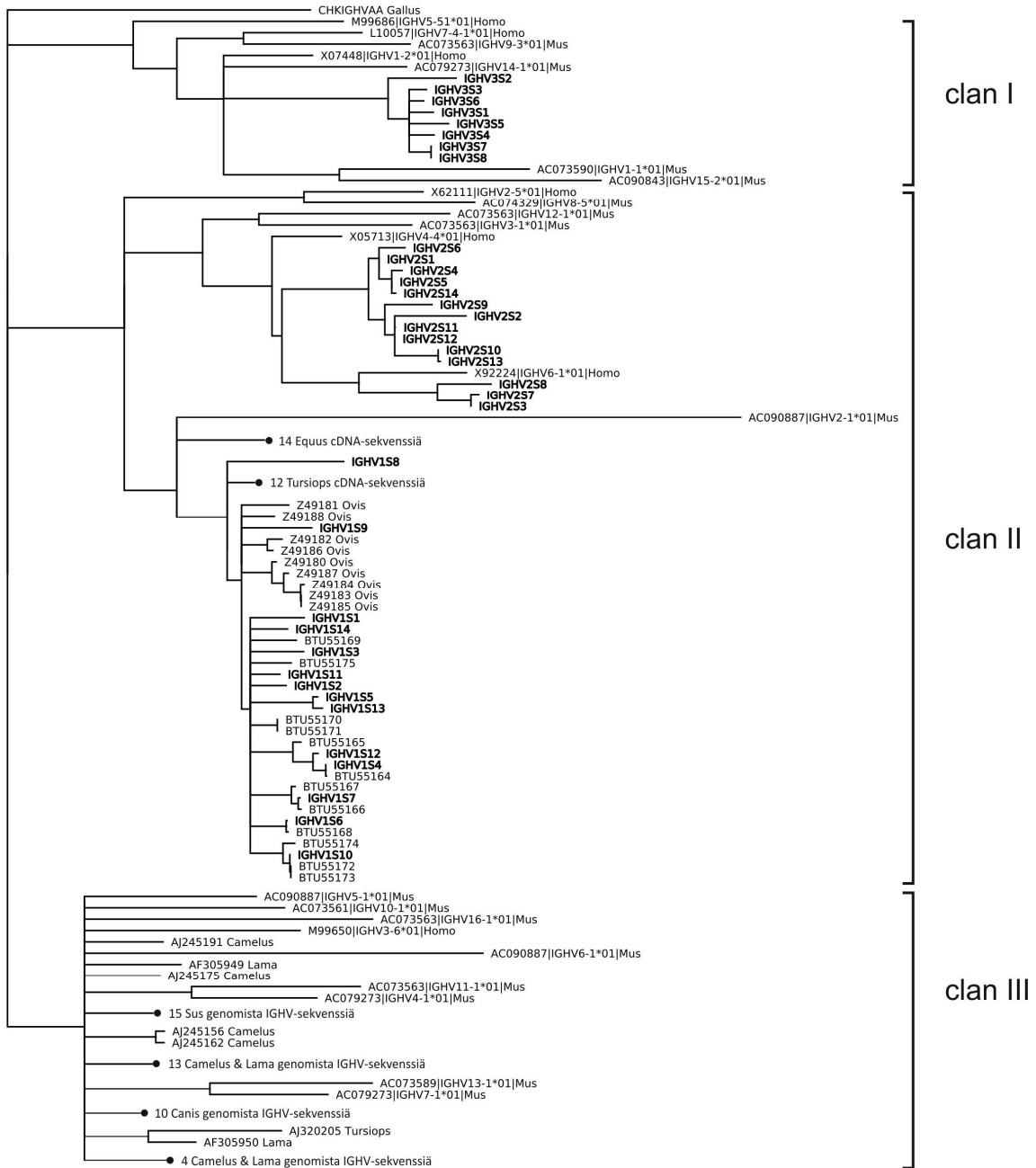
Löydetyt IGHV-geenit nimettiin IMGT:n nimeämiskäytännön mukaan (Giudicelli & Lefranc 1999). Siinä geenin tyyppiä (IGHV) seuraava numero tarkoittaa alaryhmää ja viimeinen numero geeniä alaryhmän sisällä. Koska geenien sijainti kromosomissa on vielä epävarmaa genomikokoonpanon virheiden vuoksi, nimeen lisättiin alaryhmän ja geenin numeron väliin väliaikaista nimeämistä merkitsevä S-kirjain.

5 TULOKSET

Naudan IGHV-geenejä löytyi UMD3.1-genomikokoonpanon kromosomeista 7 ja 21. Yhteensä löydettiin 36 IGHV-geeniä. 31 geeniä löytyi UMD3.1-genomikokoonpanosta ja viisi geeniä trace-sekvensseistä. Koska Btau_4.2-genomikokoonpanosta löytyi ainoastaan kaksi geeniä, jotka olivat vielä identtisiä UMD3.1-kokoonpanosta löytyneiden geenien kanssa, työ päätettiin perustaa ainoastaan UMD3.1-genomikokoonpanoon. Kromosomista 7 löytyi 17 geeniä, ja kromosomista 21 löytyi 12 geeniä. Kahta UMD3.1-genomikokoonpanosta löytynyttä geeniä (IGHV1S10, IGHV2S13) ei ollut sijoitettu mihinkään kromosomiin. Myöskään viiden ainoastaan trace-sekvensseistä löytyneen geenin (IGHV1S11-IGHV1S14 ja IGHV2S14) kromosomisijaintia ei voi päätellä. Kaksi UMD3.1-genomikokoonpanosta löydetyistä geeneistä olivat sekvensseiltään identtisiä (IGHV3S7 ja IGHV3S8), mutta sijoitettuna eri paikkaan kromosomissa, joten ne laskettiin eri geeneiksi. (Taulukko 11).

Kymmenen geeniä täytti kaikki tässä tutkimuksessa asetetut toimivan IGHV-geenin kriteerit, mutta EST-tietokannan mukaan vain kahdeksan ilmentyy. Kaikkiaan EST-osumia löytyi 12 geenillä, mutta neljän kohdalla niitä oli niin vähän, ettei geeni laskettu tässä tutkimuksessa toimivaksi. (Taulukko 11).

Löydetyt IGHV-geenit jaettiin fylogeneettisen analyysin ja sekvenssi-identtisyyksien perusteella kolmeen alaryhmään. IGHV1:ksi nimeämäämme alaryhmään kuului 14 tässä tutkimuksessa löydetyt geenin lisäksi ainoa aiemmin raportoitu ilmentyvä naudan IGHV-perhe (BovV_{H1}) (Saini ym. 1997). Fylogeneettisen analyysin perusteella IGHV1 ja 14 geenistä koostuva IGHV2 kuuluivat nisäkkäiden II klaaniin. Kahdeksasta geenistä koostuva IGHV3 kuului nisäkkäiden I klaaniin. IGHV3-alaryhmässä leusiini paikassa 89 on korvautunut metioniinilla. Kaikki ilmentyvät geenit kuuluivat IGHV1-alaryhmään ja niiden FR- ja CDR -alueet olivat yhtä pitkät. (Taulukot 11 ja 12; Kuva 4).



Kuva 4. Fylogeneettinen puu naudan ja lähisukuisten lajien IGHV-geeneistä.

Taulukko 11. Naudan IGHV-geenit ominaisuuksineen.

Nimi	Ch	Oct	Exon1	Exon2	Cys23	Trp41	Leu89	Cys104	Hepta	Nona	Func	EST	FR1	CDR1	FR2	CDR2	FR3
IGHV1S1	21	OK	OK	OK	OK	OK	OK	OK	OK	OK	Kyllä	277	75	24	51	21	114
IGHV1S2	21	OK	OK	OK	OK	OK	OK	OK	OK	OK	Kyllä	195	75	24	51	21	114
IGHV1S3	21	OK	OK	OK	OK	OK	Met	OK	OK	OK	Kyllä	417	75	24	51	21	114
IGHV1S4	21	OK	OK	OK	OK	OK	OK	OK	OK	OK	Kyllä	2	75	24	51	21	114
IGHV1S5	7	OK	OK	STOP	OK	OK	-	OK	OK	OK	Ei	15	75	24	51	21	114
IGHV1S6	7	OK	OK	STOPS	OK	OK	OK	OK	OK	OK	Kyllä	88	75	24	51	21	114
IGHV1S7	7	(OK)	OK	OK	OK	OK	OK	OK	OK	(OK)	Kyllä	67	75	24	51	21	114
IGHV1S8	7	OK	OK	STOPS	OK	OK	-	OK	OK	OK	Ei	0	75	24	51	-	-
IGHV1S9	7	-	OK	STOPS	-	OK	-	OK	OK	(OK)	Ei	0	-	24	51	21	-
IGHV1S10	-	OK	OK	OK	OK	OK	OK	OK	OK	(OK)	Kyllä	242	75	24	51	21	114
IGHV1S11	-	OK	OK	OK	OK	OK	OK	OK	OK	OK	Kyllä	342	75	24	51	21	114
IGHV1S12	-	OK	OK	OK	OK	OK	OK	OK	OK	OK	Kyllä	5	75	24	51	21	114
IGHV1S13	-	OK	OK	STOP	OK	OK	-	OK	OK	(OK)	Ei	3	75	24	51	21	114
IGHV1S14	-	OK	OK	OK	OK	OK	OK	OK	OK	OK	Kyllä	461	75	24	51	21	114
IGHV2S1	21	OK	OK	STOP	OK	(OK)	OK	OK	(OK)	(OK)	Ei	0	75	25	51	21	114
IGHV2S2	21	(OK)	(OK)	STOPS	OK	OK	OK	-	OK	(OK)	Ei	0	71	25	51	21	111
IGHV2S3	21	(OK)	STOP	STOPS	OK	-	OK	OK	OK	(OK)	Ei	0	58	23	-	-	-
IGHV2S4	21	OK	(OK)	STOP	OK	(OK)	OK	OK	OK	(OK)	Ei	0	77	25	51	21	114
IGHV2S5	21	(OK)	(OK)	STOPS	OK	(OK)	OK	OK	OK	(OK)	Ei	0	75	25	51	21	114
IGHV2S6	7	OK	(OK)	STOP	OK	(OK)	OK	OK	OK	(OK)	Ei	0	75	25	51	21	112
IGHV2S7	7	(OK)	STOP	STOPS	OK	-	OK	OK	OK	(OK)	Ei	0	58	23	-	-	-
IGHV2S8	7	OK	OK	STOPS	OK	-	OK	OK	OK	(OK)	Ei	0	74	23	-	-	-
IGHV2S9	7	OK	(OK)	OK	OK	OK	OK	-	OK	(OK)	Ei	0	75	27	51	21	114
IGHV2S10	7	(OK)	OK	STOPS	(OK)	(OK)	(OK)	(OK)	OK	(OK)	Ei	0	-	25	51	21	111
IGHV2S11	7	(OK)	OK	STOPS	OK	OK	OK	OK	OK	(OK)	Ei	0	72	25	51	21	112
IGHV2S12	7	(OK)	OK	STOPS	OK	OK	OK	OK	OK	(OK)	Ei	0	72	25	51	21	112
IGHV2S13	-	(OK)	OK	STOPS	OK	OK	OK	OK	OK	(OK)	Ei	0	-	25	51	21	112
IGHV2S14	-	-	(OK)	STOPS	OK	(OK)	OK	OK	(OK)	-	Ei	0	75	25	51	21	114
IGHV3S1	21	(OK)	(OK)	STOPS	OK	-	Met89	(OK)	(OK)	(OK)	Ei	0	74	21	51	23	114
IGHV3S2	21	(OK)	(OK)	STOPS	OK	-	Met89	(OK)	(OK)	-	Ei	0	74	21	51	23	114

Taulukko 11. Jatkoa

Nimi	Ch	Oct	Exon1	Exon2	Cys23	Trp41	Leu89	Cys104	Hepta	Nona	Func	EST	FR1	CDR1	FR2	CDR2	FR3
IGHV3S3	21	(OK)	(OK)	STOPS	OK	OK	Met89	(OK)	(OK)	-	Ei	0	74	21	51	23	114
IGHV3S4	7	(OK)	(OK)	STOPS	OK	-	Met89	(OK)	(OK)	(OK)	Ei	0	74	21	51	23	113
IGHV3S5	7	(OK)	-	STOPS	OK	OK	Met89	(OK)	(OK)	(OK)	Ei	0	65	24	51	23	113
IGHV3S6	7	(OK)	(OK)	STOPS	OK	-	Met89	(OK)	(OK)	-	Ei	0	74	21	51	23	114
IGHV3S7	7	(OK)	(OK)	STOPS	OK	OK	-	(OK)	(OK)	(OK)	Ei	0	74	21	51	23	114
IGHV3S8	7	(OK)	(OK)	STOPS	OK	OK	-	(OK)	(OK)	(OK)	Ei	0	74	21	51	23	114

Ch = kromosomi; Oct = promoottori oktameeri; Exon1 = 1. eksoni; Exon2 = 2. eksoni; Cys23 = kysteini 23. aminohappona; Leu89 = leusiini 89. aminohappona; Cys104 = kysteini 104. aminohappona; Hepta = rekombinaatiosignaalin heptameeri; Nona = rekombinaatiosignaalin nonameeri; Func = vaikuttaako geeni toimivalta; EST = EST osuimien lkm; OK = kyseinen ehto täyttyy/promoottori tai RSS on konsensussekvenssi; (OK) = promoottori tai RSS löytyy, mutta poikkeaa konsensussekvenssistä; STOPS = STOP-kodoni keskellä lukukehystä; Met89 = Leu89:n korvaa metioniini; - = ei tietoa/ehto ei täyty

Taulukko 12. FR1-FR3-alueiden sekvenssi-identtisyyksien vaihteluväli ja keskiarvo (suluissa) naudän IGHV-alaryhmien sisällä ja välillä (%).

	IGHV1	IGHV2	IGHV3
IGHV1	75 – 99 (89)		
IGHV2	56 – 68 (64)	78 – 99 (86)	
IGHV3	48 – 54 (52)	47 – 55 (51)	84 – 100 (94)

6 POHDINTA

Tämä tutkimus vahvistaa käsityksen naudän suppeasta ituradan IGHV-geenivalikoimasta (Berens ym. 1997; Saini ym. 1997; Sinclair ym. 1997; Lopez ym. 1998). Kaikki naudän ilmentyvät IGHV-geenit kuuluvat yhteen alaryhmään (IGHV1), johon pseudogeenit ja aiemmin julkaistut sekvenssit mukaanlukien kuuluu 26 geeniä. Samoin tutkimuksemme vahvistaa aiemmat epäilyt ilmentymättömistä IGHV-alaryhmistä (Berens ym. 1997). Naudalla on kaksi kokonaan pseudogeenistä koostuvaa IGHV-alaryhmää, joista toinen kuuluu nisäkkäiden I klaaniin, ja toinen IGHV1-alaryhmän tavoin II klaaniin. Naudalta siis puuttuvat kokonaan nisäkkäiden III klaaniin kuuluvat IGHV-geenit.

Tässä tutkimuksessa naudalta löydettiin 36 uutta IGHV-geeniä, joista kuitenkin todennäköisesti ainoastaan kahdeksan ilmentyy. IGHV-geenien evoluutiossa tasapainottelevat uusien geenien synty kahdentumalla ja vanhojen rappeutuminen mutaatioiden kertyessä pseudogeeniksi (Ota & Nei 1994; Nei ym. 1997). IGHV-geenien suuresta lukumäärästä johtuen yhteen geeniin kohdistuva valintapaine on alhainen, joten pseudogeenitkin voivat säilyä genomissa pitkään ja muodostaa merkittävän osan IGHV-geenien kokonaismäärästä. Kaikilla selkärangkaisilla merkittävä osa IGHV-geeneistä on pseudogeeninä. Esimerkiksi ihmisellä ja rotalla vain yksi kolmasosa IGHV-geeneistä on toimivia. Toisaalta näillä lajeilla IGHV-geenien kokonaismääräkin on huomattavasti suurempi kuin naudalla (Matsuda ym. 1998; Hendricks ym. 2010).

Yhteensä tässä tutkimuksessa löydettyjä ja aiemmin julkaistuja naudan IGHV-geenejä on 48, joka saattaa olla yliarvio naudan ituradan IGHV-geenivalikoimasta. Yleisesti ottaen IGHV-geenit ovat hyvin polymorfisia. Sekvenssi-identtisyyksien perusteella ei voi todeta, ovatko kaksi geeniä toistensa alleeleja vai paralogeja (Pramanik ym. 2011). Toisin sanoen onko kyse lokuksen sisäisestä vai kahden kahden lokuksen välisestä vaihtelusta. Täysin homotsygootin yksilön perimässä alleeleista johtuva vaihtelu olisi eliminoitu, mutta naudan genomiprojekti perustuu kahteen yksilöön, jolloin aineistossa on saattanut näiden yksilöiden korkeasta sukusiitosasteesta (30 %) huolimatta olla useita alleeleita samasta geenistä. Lisäksi kahden tässä tutkimuksessa löydetyn täysin identtisen geenin kohdalla voimme vain todeta, että käytössä olevan tiedon perusteella kyseessä ovat eri geenit. Tässä tutkimuksessa ei myöskään löydetty aiemmin julkaistujen IGHV-geenien kanssa täysin identtisiä sekvenssejä, mikä sekkin antaa ymmärtää, että 48 geenistä osa on ainoastaan saman geenin eri alleelleita. Toisaalta esimerkiksi ihmisellä IGHV-geenien lukumäärässä esiintyy myös haplotyyppikohtaista vaihtelua (Lefranc 2001a). Vastaava ilmiö myös naudalla selittäisi aiemmin raportoitujen geenien puuttumisen tässä tutkimuksessa löydettyjen joukosta.

Naudan IGHV-geenejä löytyi UMD_3.1-genomikokoonpanon kromosomeista 7 ja 21. Kahden täydellisen toimivan IgH-lokukseen löytyminen nisäkkäältä olisi todella poikkeuksellista (Hsu ym. 2006). Todennäköisesti löydöksessä on sittenkin kyse virheestä genomikokoonpanossa. Ihmisen IgH-locus sijaitsee kromosomeissa 14 (Cook ym. 1994). Hybridisaatioon perustuvan kromosomien homologian perusteella naudan kromosomi 7 vastaa ihmisen ihmisen kromosomeja 5 ja 19, ja naudan kromosomi 21 vastaa ihmisen kromosomeja 14 ja 15 (Chowdhary ym. 1996). Myös naudan C_H-geenit sisältävä locus on paikallistettu kromosomiin 21 (Gu ym. 1992). Toisaalta naudalla on löydetty μ -geenin kaltainen IgML1-geeni myös kromosomista 11 (Gu ym. 1992; Tobinjanzen & Womack 1992). Edelleen, kromosomista 11 on löydetty myös IGHJ-geenejä (Hosseini ym. 2004). Poistogeenisiä nautoja tutkimalla on vahvistettu, että molemmat naudan IgH-lokukset ovat toimivia (Kuroiwa ym. 2009). Jos kromosomissa 11 ei ole IGHV- ja IGHD-geenejä, toimivan IgH-geenin kokoaminen vaatii geenien järjestelyä kromosomien välillä. Luukaloilla, joilla voi olla useita IgH-lokuksia, IGHV- ja C_H-geenit rekombinoituvat myös lokuksien välillä (Yasuike ym. 2010). Kanilla ja hiirellä puolestaan tapahtuu immunoglobuliinin luokanvaihtoa kromosomien välillä (Knight ym. 1995; Reynaud ym. 2005). Naudan IGHV-geenien sijainnille saadaan

lopullinen varmistus vasta kun ne fyysisesti paikannettu esimerkiksi *in situ* -hybridisaatiolla.

Naudan IGHV1- ja IGHV2-alaryhmät kuuluvat nisäkkäiden II klaaniin, kun taas IGHV3 kuuluu I klaaniin. Ilmentymätön IGHV2 on todennäköisesti havaittu aikaisemmin hybridisaatiolla hiiren IGHV3-alaryhmään kuuluvien geenien kanssa (Berens ym. 1997). Samalla menetelmällä naudalta on löydetty myös hiiren IGHV5- ja IGHV7-alaryhmiä vastaavat, ja siten III klaaniin kuuluvat alaryhmät (Tutter & Riblet 1989; Berens ym. 1997). IGHV-alaryhmien tunnistaminen hybridisaatiolla käyttäen koettimina toisen eläinlajin sekvenssejä ei luultavasti ole täysin luotettava menetelmä, vaikka toisaalta nimenomaan III klaaniin kuuluvien geenien oletetaan muuttuvan muita klaaneja hitaammin (Tutter & Riblet 1989). Tässä tutkimuksessa ei kuitenkaan saatu vahvistusta naudnan III klaaniin kuuluville IGHV-alaryhmille. Koska tässä tutkimuksessa käytettiin hakusekvensseinä kaikkia ihmisen ja hiiren IGHV-geenejä, ja lisäksi geenejä haettiin suoraan rekombinaatiosignaalisekvensseillä, on melko epätodennäköistä, että kokonaisia alaryhmiä olisi jäänyt löytymättä. Geenejä on voinut jäädä löytymättä ainoastaan, jos ne puuttuisivat naudnan genomiprojektin sekvenssintidatasta, mikä sekin tuntuu melko epätodennäköiseltä, koska naudnan perimä on genomiprojektissa luettu keskimäärin seitsemään kertaan (Elsik ym. 2009).

III klaania pidetään nisäkkäiden IGHV-klaaneista evolutiivisesti vanhimpana (Tutter & Riblet 1989). Naudnan kaukaisemmista sukulaisista sian ainoa IGHV-perhe kuuluu III klaaniin (Butler ym. 2009). Myös kaikki kamelieläinten IGHV-perheet kuuluvat III klaaniin (Nguyen ym. 2002). Lampaalla toistaiseksi ainoa genomisesta DNA:sta tunnettu IGHV-perhe kuuluu naudnan IGHV1-alaryhmän tavoin II klaaniin (Dufour ym. 1996). Lampaan IGHV-valikoima saattaa kuitenkin olla laajempi. Lampaan mRNA-sekvensseistä on tunnistettu jopa yhdeksän IGHV-perhettä, joissa on edustettuina kaikki nisäkkäiden IGHV-klaanit (Charlton ym. 2000). mRNA-tutkimuksessa on immunoglobuliinigeenien kohdalla aina virhelähteenä V(D)J-rekombinaation jälkeen tapahtuneet somaattiset mutaatiot, eivätkä muut tutkijat ole onnistuneet toistamaan Charltonin ym. (2000) tuloksia (Jenne ym. 2006).

Miksi nauta ilmentää ainoastaan yhden alaryhmän geenejä? IGHV-geenien sijoittumisella ei pitäisi olla vaikutusta ilmentymiseen, koska kaikkiin kolmeen alaryhmään kuuluvat geenit sijaitsevat lomittain naudan perimässä. Eri IGHV-alaryhmien koodaamat immunoglobuliinit saattavat poiketa ominaisuuksiltaan. Esimerkiksi I klaaniin kuuluvassa naudan IGHV3-perheessä leusiini paikassa 89 on korvautunut metioniinilla, mikä näiden geenien ilmentyessä on saattanut vaikuttaa antigeeniin sitoutuvan alueen laskostumiseen. Muutenkin IGHV3-alaryhmässä on vain vähän vaihtelua (sekvenssit keskimäärin 94 % samoja), joten sen oltava verraten nuori alaryhmä, tai siihen on kohdistunut menneisyydessä stabiloiva luonnonvalinta. Ensimmäinen vaihtoehto ei todennäköisesti pidä paikkaansa, koska IGHV3-alaryhmän geenit sijaitsevat genomissa muiden alaryhmien geenien kanssa sekaisin. Jos IGHV3-perhe olisi ollut tärkeä naudan edeltäjien selviytymiselle, luonnonvalinta olisi karsinut siihen kohdistuneet mutaatiot tehokkaasti. Edelleen, IGHV3-alaryhmän muuttuminen naudalle vahingolliseksi jossain vaiheessa voisi selittää sen rappeutumisen pseudogeneiksi STOP-kodoneita aikaansaavilla pistemutaatioilla. Märehtijät ovat yksimahaisiin verrattuna hyvin riippuvaisia ruoansulatuskanavan symbioottisista mikrobeista. Ehkä I ja III klaaniin kuuluvien IGHV-geenien koodaamalla immunoglobuliineilla on ominaisuuksia, joiden vuoksi märehtijöiden ja suolistomikrobien yhteiselo häiriintyisi. Ajatus voi tuntua kaukaa haetulta, mutta esimerkiksi hiirillä tietyt antigeenit voivat stimuloida eri IGHV-perheisiin kuuluvien geenien ilmentymistä (Jeong & Teale 1988). Tietysti on mahdollista, että naudan suppea IGHV-valikoima on ainoastaan sattuman tulosta. Kaikista selkärankaisten luokista löytyy lajeja, joiden IGHV-valikoima on menneisyydessä supistunut ja alkanut sitten rakentumaan uudelleen. Esimerkiksi kaikki vesinokkaeläimen IGHV-geenit vaikuttavat syntyneen yhden ainoan geenin monistuessa vesinokkaeläimen ja nokkasiilien kehityslinjojen eroamisen jälkeen (Johansson ym. 2002). Naudan IGHV-alaryhmät ovat suhteellisen suuria, mikä sopisi teoriaan menetetyistä ja uudestaan rakentuvasta IGHV-valikoimasta.

Moneen muuhun lajiin verrattuna naudan IGHV-valikoima on hyvin suppea. Ituradan kolmesta IGHV-alaryhmästä ainoastaan yksi ilmentyy, ja siihenkin kuuluvat geenit ovat keskenään hyvin samankaltaisia. Esimerkiksi kaikkien ilmentyvien geenien FR- ja CDR-alueet ovat yhtä pitkät. Kun nauta lisäksi käyttää pääasiassa vain yhtä IGHJ-geeniä (Hosseini ym. 2004), naudan IgH-ketjun rekombinatorinen monimuotoisuus on

väistämättä alhainen (Zhao ym. 2006b). Nautaan johtava kehityslinja on menettänyt III klaaniin kuuluvat IGHV-geenit aikaisintaan sikoihin johtavan linjan eroamisen jälkeen. Tarkempi ajoitus vaatisi naudalle läheisempää sukua olevien lajien, kuten valaiden ja hirvieläinten perimän selvittämistä. Samalla saataisiin ehkä valaistusta kysymykseen miksi I klaaniin kuuluvat IGHV-geenit ovat naudalla rappeutuneet pseudogeeneiksi. Naudan IGHV-alaryhmien tunteminen auttaa ymmärtämään immunoglobuliinien monimuotoisuuden tuottamista suppean ituradan IGHV-valikoiman lajeilla.

7 LÄHDELUETTELO

- Achour I, Cavelier P, Tichit M, Bouchier C, Lafaye P, Rougeon F. Tetrameric and homodimeric camelid IgGs originate from the same IgH locus. *J. Immunol.* 2008, 181:2001-2009.
- Aguilar B, Gutman G. Transcription and Diversity of Immunoglobulin-Lambda Chain Variable Genes in the Rat. *Immunogenetics* 1992, 37:39-48.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. *J. Mol. Biol.* 1990, 215:403-410.
- Amemiya CT, Ohta Y, Litman RT, Rast JP, Haire RN, Litman GW. V(h)gene Organization in a Relict Species, the Coelacanth *Latimeria-Chalumnae* - Evolutionary Implications. *Proc. Natl. Acad. Sci. U. S. A.* 1993, 90:6661-6665.
- Andersson E, Matsunaga T. Evolution of Immunoglobulin Heavy-Chain Variable Region Genes - a Vh Family can Last for 150-200 Million Years Or Longer. *Immunogenetics* 1995, 41:18-28.
- Andre S, Guillet F, Charlemagne J, Fellah JS. Structure and diversity of Mexican axolotl lambda light chains. *Immunogenetics* 2000, 52:137-144.
- Ayadi H, Marche P, Cazenave P. Evolution of the Rabbit Immunoglobulin Kappa-Chain Genes. *Immunogenetics* 1991, 34:201-207.

- Baker ML, Belov K, Miller RD. Unusually similar patterns of antibody V segment diversity in distantly related marsupials. *J. Immunol.* 2005, 174:5665-5671.
- Bao Y, Wang T, Guo Y, Zhao Z, Li N, Zhao Y. The immunoglobulin gene loci in the teleost *Gasterosteus aculeatus*. *Fish Shellfish Immun.* 2010a, 28:40-48.
- Bao Y, Guo Y, Xiao S, Zhao Z. Molecular characterization of the VH repertoire in *Canis familiaris*. *Vet. Immunol. Immunopathol.* 2010b, 137:64-75.
- Belov K, Hellman L. Immunoglobulin genetics of *Ornithorhynchus anatinus* (platypus) and *Tachyglossus aculeatus* (short-beaked echidna). *Comp. Biochem. Phys. A* 2003, 136:811-819.
- Benammar A, Cazenave P. A 2nd Rabbit Kappa-Isotype. *J. Exp. Med.* 1982, 156:585-595.
- Bengten E, Quiniou SMA, Stuge TB, Katagiri T, Miller NW, Clem LW, Warr GW, Wilson M. The IgH locus of the channel catfish, *Ictalurus punctatus*, contains multiple constant region gene sequences: Different genes encode heavy chains of membrane and secreted IgD. *J. Immunol.* 2002, 169:2488-2497.
- Bengten E, Quiniou S, Hikima J, Waldbieser G, Warr GW, Miller NW, Wilson M. Structure of the catfish IGH locus: analysis of the region including the single functional IGHM gene. *Immunogenetics* 2006, 58:831-844.
- Benton MJ. *Vertebrate Palaeontology*. Blackwell Science, Oxford 2005.
- Berens SJ, Wylie DE, Lopez OJ. Use of a single V-H family and long CDR3s in the variable region of cattle Ig heavy chains. *Int. Immunol.* 1997, 9:189-199.
- Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD, Grenyer R, ym. The delayed rise of present-day mammals. *Nature* 2007, 446:507-512.
- Bovine Genome Project. <http://www.hgsc.bcm.tmc.edu/project-species-m-Bovine.hgsc?pageLocation=Bovine>, haettu 20.6.2011.

- Brodeur P, Riblet R. The Immunoglobulin Heavy-Chain Variable Region (Igh-V) Locus in the Mouse .1. One Hundred Igh-V Genes Comprise 7 Families of Homologous Genes. *Eur. J. Immunol.* 1984, 14:922-930.
- Brown GD, Kaattari IM, Kaattari SL. Two new Ig VH gene families in *Oncorhynchus mykiss*. *Immunogenetics* 2006, 58:933-936.
- Brusco A, Cinque F, Saviozzi S, Boccazzi C, DeMarchi M, Carbonara A. The G4 gene is duplicated in 44% of human immunoglobulin heavy chain constant region haplotypes. *Hum. Genet.* 1997, 100:84-89.
- Burnett RC, Hanly WC, Zhai SK, Knight KL. The Iga Heavy-Chain Gene Family in Rabbit - Cloning and Sequence-Analysis of 13 C-Alpha-Genes. *EMBO J.* 1989, 8:4041-4047.
- Butler J, Sun J, Wertz N, Sinkora M. Antibody repertoire development in swine. *Dev. Comp. Immunol.* 2006, 30:199-221.
- Butler JE, Zhao Y, Sinkora M, Wertz N, Kacsokovics I. Immunoglobulins, antibody repertoire and B cell development. *Dev. Comp. Immunol.* 2009, 33:321-333.
- Charlton KA, Moyle S, Porter AJR, Harris WJ. Analysis of the diversity of a sheep antibody repertoire as revealed from a bacteriophage display library. *J. Immunol.* 2000, 164:6221-6229.
- Chen K, Xu W, Wilson M, He B, Miller NW, Bengten E, ym. Immunoglobulin D enhances immune surveillance by activating antimicrobial, proinflammatory and B cell-stimulating programs in basophils. *Nat. Immunol.* 2009, 10:889-U121.
- Chowdhary B, Fronicke L, Gustavsson I, Scherthan H. Comparative analysis of the cattle and human genomes: Detection of ZOO-FISH and gene mapping-based chromosomal homologies. *Mamm. Genome* 1996, 7:297-302.
- Cook G, Tomlinson I, Walter G, Riethman H, Carter N, Buluwela L, Winter G, Rabbitts T. A Map of the Human-Immunoglobulin V-H Locus Completed by Analysis of the Telomeric Region of Chromosome 14q. *Nat. Genet.* 1994, 7:162-168.

- Cooper M, Alder M. The evolution of adaptive immune systems. *Cell* 2006, 124:815-822.
- Criscitiello MF, Flajnik MF. Four primordial immunoglobulin light chain isotypes, including lambda and kappa, identified in the most primitive living jawed vertebrates. *Eur. J. Immunol.* 2007, 37:2683-2694.
- Daggfeldt A, Bengtén E, Pilström L. A cluster type organization of the loci of the immunoglobulin light chain in Atlantic cod (*Gadus morhua* L) and rainbow trout (*Oncorhynchus mykiss* Walbaum) indicated by nucleotide sequences of cDNA and hybridization analysis. *Immunogenetics* 1993, 38:199-209.
- Danilova N, Bussmann J, Jekosch K, Steiner LA. The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. *Nat. Immunol.* 2005, 6:295-302.
- Danilova N, Amemiya CT. Going Adaptive The Saga of Antibodies. Year in Evolutionary Biology 2009 2009, 1168:130-155.
- Das S, Nikolaidis N, Klein J, Nei M. Evolutionary redefinition of immunoglobulin light chain isotypes in tetrapods using molecular markers. *Proc. Natl. Acad. Sci. U. S. A.* 2008, 105:16647-16652.
- Das S, Mohamedy U, Hirano M, Nei M, Nikolaidis N. Analysis of the Immunoglobulin Light Chain Genes in Zebra Finch: Evolutionary Implications. *Mol. Biol. Evol.* 2010, 27:113-120.
- de Genst E, Saerens D, Muyldermans S, Conrath K. Antibody repertoire development in camelids. *Dev. Comp. Immunol.* 2006, 30:187-198.
- Dooley H, Flajnik MF. Antibody repertoire development in cartilaginous fish. *Dev. Comp. Immunol.* 2006, 30:43-56.
- Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Duran C ym. Geneious v5.4, Available from <http://www.geneious.com/>, haettu 1.4.2011
- Dufour V, Malinge S, Nau F. The sheep Ig variable region repertoire consists of a single V-H family. *J. Immunol.* 1996, 156:2163-2170.

- Dufour V, Nau F. Genomic organization of the sheep immunoglobulin JH segments and their contribution to heavy chain variable region diversity. *Immunogenetics* 1997, 46:283-292.
- Edholm E, Wilson M, Sahoo M, Miller NW, Pilstrom L, Wermenstam NE, Bengten E. Identification of Ig sigma and Ig lambda in channel catfish, *Ictalurus punctatus*, and Ig lambda in Atlantic cod, *Gadus morhua*. *Immunogenetics* 2009, 61:353-370.
- Edholm E, Bengten E, Wilson M. Insights into the function of IgD. *Dev. Comp. Immunol.* 2011a, 35:1309-1316.
- Edholm E, Wilson M, Bengten E. Immunoglobulin light (IgL) chains in ectothermic vertebrates. *Dev. Comp. Immunol.* 2011b, 35:906-915.
- Ekman A, Niku M, Liljavirta J, Iivanainen A. Bos taurus genome sequence reveals the assortment of immunoglobulin and surrogate light chain genes in domestic cattle. *Bmc Immunology* 2009, 10:22.
- Elsik CG, Tellam RL, Worley KC, Gibbs RA, Abatepaulo ARR, Abbey CA, ym. The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. *Science* 2009, 324:522-528.
- Emorine L, Max E. Structural-Analysis of a Rabbit Immunoglobulin Chi-2 J-C Locus Reveals Multiple Deletions. *Nucleic Acids Res.* 1983, 11:8877-8890.
- Fellah JS, Iscaki S, Vaerman JP, and Charlemagne J. Transient Developmental Expression of IgY Secretory Component like Protein in the Gut of the Axolotl (*Ambystoma mexicanum*). *Dev. Immunol.* 1992, 2:181-190.
- Flajnik MF. Comparative analyses of immunoglobulin genes: Surprises and portents. *Nature Rev. Immunol.* 2002, 2:688-698.
- Flajnik MF, Kasahara M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nature Rev. Genet.* 2010, 11:47-59.
- Flanagan J, Rabbitts T. Arrangement of Human-Immunoglobulin Heavy-Chain Constant Region Genes Implies Evolutionary Duplication of a Segment Containing Gamma-Genes, Epsilon-Genes and Alpha-Genes. *Nature* 1982, 300:709-713.

- Fleurant M, Changchien L, Chen CT, Flajnik MF, Hsu E. Shark Ig light chain junctions are as diverse as in heavy chains. *J. Immunol.* 2004, 173:5574-5582.
- Gallarda JL, Gleason KS, Knight KL. Organization of Rabbit Immunoglobulin Genes .1. Structure and Multiplicity of Germ-Line Vh Genes. *J. Immunol.* 1985, 135:4222-4228.
- Gambon Deza F, Sanchez Espinel C, Valdueza Beneitez J. A novel IgA-like immunoglobulin in the reptile *Eublepharis macularius*. *Dev. Comp. Immunol.* 2007, 31:596-605.
- Gambon Deza F, Sanchez Espinel C, Mompo SM. The immunoglobulin heavy chain locus in the reptile *Anolis carolinensis*. *Mol. Immunol.* 2009, 46:1679-1687.
- Gambon-Deza F, Sanchez Espinel C. IgD in the reptile leopard gecko. *Mol. Immunol.* 2008, 45:3470-3476.
- Gambon-Deza F, Sanchez-Espinel C, Magadan-Mompo S. Presence of an unique IgT on the IGH locus in three-spined stickleback fish (*Gasterosteus aculeatus*) and the very recent generation of a repertoire of VH genes. *Dev. Comp. Immunol.* 2010, 34:114-122.
- Ghaffari SH, Lobb CJ. Immunoglobulin Light Chain in the Channel catfish An Unusual Genomic Organizational Pattern of Segmental Genes. *J. Immunol.* 1993, 151:6900-6912.
- Giudicelli V, Lefranc MP. Ontology for immunogenetics: the IMGT-ONTOLOGY. *Bioinformatics* 1999, 15:1047-1054.
- Golub R, Charlemagne J. Structure, diversity, and repertoire of V-H families in the Mexican axolotl. *J. Immunol.* 1998, 160:1233-1239.
- Greenberg AS, Avila D, Hughes M, Hughes A, Mckinney EC, Flajnik MF. A New Antigen Receptor Gene Family that Undergoes Rearrangement and Extensive Somatic Diversification in Sharks. *Nature* 1995, 374:168-173.

- Gu F, Chowdhary BP, Andersson L, Harbitz I, Gustavsson I. Assignment of the Bovine Immunoglobulin Gamma Heavy-Chain (Ighg) Gene to Chromosome-21q24 by Insitu Hybridization. *Hereditas* 1992, 117:237-240.
- Guo Y, Bao Y, Wang H, Hu X, Zhao Z, Li N, Zhao Y. A Preliminary Analysis of the Immunoglobulin Genes in the African Elephant (*Loxodonta africana*). *Plos One* 2011, 6:e16889.
- Haire RN, Amemiya CT, Suzuki D, Litman GW. 11 Distinct Vh-Gene Families and Additional Patterns of Sequence Variation Suggest a High Degree of Immunoglobulin Gene Complexity in a Lower Vertebrate, *Xenopus-Laevis*. *J. Exp. Med.* 1990, 171:1721-1737.
- Hamerscasterman C, Atarhouch T, Muyldermans S, Robinson G, Hamers C, Songa E, Bendahman N, Hamers R. Naturally-Occurring Antibodies Devoid of Light-Chains. *Nature* 1993, 363:446-448.
- Harding FA, Cohen N, Litman GW. Immunoglobulin heavy chain gene organization and complexity in the skate, *Raja erinacea*. *Nucleic Acids Res* 1990, 18:1015-2010.
- Hansen JD, Landis ED, Phillips RB. Discovery of a unique Ig heavy-chain isotype (IgT) in rainbow trout: Implications for a distinctive B cell developmental pathway in teleost fish. *Proc. Natl. Acad. Sci. U. S. A.* 2005, 102:6919-6924.
- Hayzer D. Immunoglobulin-Lambda Light Chain Evolution - Igl and Igl-Like Sequences Form 3 Major Groups. *Immunogenetics* 1990, 32:157-174.
- Hendricks J, Terpstra P, Dammers PM, Somasundaram R, Visser A, Stoel M, Bos NA, Kroese FGM. Organization of the variable region of the immunoglobulin heavy-chain gene locus of the rat. *Immunogenetics* 2010, 62:479-486.
- Hohman VS, Schuchman DB, Schluter SF, Marchalonis JJ. Genomic Clone for Sandbar Shark Lambda-Light Chain - Generation of Diversity in the Absence of Gene Rearrangement. *Proc. Natl. Acad. Sci. U. S. A.* 1993, 90:9882-9886.
- Hosseini A, Campbell G, Prorocic M, Aitken R. Duplicated copies of the bovine J(H) locus contribute to the Ig repertoire. *Int. Immunol.* 2004, 16:843-852.

- Hsu E, Criscitiello MF. Diverse Immunoglobulin Light Chain Organizations in Fish Retain Potential to Revise B Cell Receptor Specificities. *J. Immunol.* 2006, 177:2452-2462.
- Hsu E, Pulham N, Rumpf LL, Flajnik MF. The plasticity of immunoglobulin gene systems in evolution. *Immunol. Rev.* 2006, 210:8-26.
- Huelsenbeck J, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 2001, 17:754-755.
- Inoue JG, Miya M, Lam K, Tay B, Danks JA, Bell J, Walker TI, Venkatesh B. Evolutionary Origin and Phylogeny of the Modern Holocephalans (Chondrichthyes: Chimaeriformes): A Mitogenomic Perspective. *Mol. Biol. Evol.* 2010, 27:2576-2586.
- Jenne CN, Kennedy LJ, Reynolds JD. Antibody repertoire development in the sheep. *Dev. Comp. Immunol.* 2006, 30:165-174.
- Jeong H, Teale J. Comparison of the Fetal and Adult Functional B-Cell Repertoires by Analysis of Vh Gene Family Expression. *J. Exp. Med.* 1988, 168:589-603.
- Johansson J, Aveskogh M, Munday B, Hellman L. Heavy chain V region diversity in the duck-billed platypus (*Ornithorhynchus anatinus*): Long and highly variable complementarity-determining region 3 compensates for limited germline diversity. *J. Immunol.* 2002, 168:5155-5162.
- Johansson J, Salazar JN, Aveskogh M, Munday B, Miller RD, Hellman L. High variability in complementarity-determining regions compensates for a low number of V gene families in the lambda light chain locus of the platypus. *Eur. J. Immunol.* 2005, 35:3008-3019.
- Johnston CM, Wood AL, Bolland DJ, Corcoran AE. Complete sequence assembly and characterization of the C57BL/6 mouse Ig heavy chain V region. *J. Immunol.* 2006, 176:4221-4234.
- Kasahara M, Nakaya J, Satta Y, Takahata N. Chromosomal duplication and the emergence of the adaptive immune system. *Trends Genet.* 1997, 13:90-92.

- Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002, 30:3059-3066.
- Knight KL, Kingzette M, Crane MA, Zhai SK. Transchromosomally Derived Ig Heavy-Chains. *J. Immunol.* 1995, 155:684-691.
- Kokubu F, Litman R, Shamblott MJ, Hinds K, Litman GW. Diverse Organization of Immunoglobulin Vh-Gene Loci in a Primitive Vertebrate. *EMBO J.* 1988, 7:3413-3422.
- Kuroiwa Y, Kasinathan P, Sathiyaseelan T, Jiao J, Matsushita H, Sathiyaseelan J, ym. Antigen-specific human polyclonal antibodies from hyperimmunized cattle. *Nat. Biotechnol.* 2009, 27:173-181.
- Lee SS, Fitch D, Flajnik MF, Hsu E. Rearrangement of Immunoglobulin Genes in Shark Germ Cells. *J. Exp. Med.* 2000, 191:1637-1647.
- Lee SS, Tranchina D, Ohta Y, Flajnik MF, Hsu E. Hypermutation in shark immunoglobulin light chain genes results in contiguous substitutions. *Immunity* 2002, 16:571-582.
- Lee V, Huang JL, Lui MF, Malecek K, Ohta Y, Mooers A, Hsu E. The evolution of multiple isotypic IgM heavy chain genes in the shark. *J. Immunol.* 2008, 180:7461-7470.
- Lefranc MP. Unique database numbering system for immunogenetic analysis; Current literature. *Immunol. Today* 1997, 18:509-509.
- Lefranc MP. Nomenclature of the human immunoglobulin heavy (IGH) genes. *Exp. Clin. Immunogenet.* 2001a, 18:100-116.
- Lefranc MP. Nomenclature of the human immunoglobulin kappa (IGK) genes. *Exp. Clin. Immunogenet.* 2001b, 18:161-174.
- Lefranc MP. Nomenclature of the human immunoglobulin lambda (IGL) genes. *Exp. Clin. Immunogenet.* 2001c, 18:242-254.

- Lefranc MP, Giudicelli V, Ginestoux C, Jabado-Michaloud J, Folch G, Bellahcene F
ym. IMGT(R), the international ImMunoGeneTics information system(R).
www.imgt.org, haettu 1.12.2011
- Litman GF, Anderson MK, Rast J. Evolution of antigen binding receptors. *Annu. Rev. Immunol.* 1999, 17:109-147.
- Lopez O, Perez C, Wylie D. A single V-H family and long CDR3s are the targets for hypermutation in bovine immunoglobulin heavy chains. *Immunol. Rev.* 1998, 162:55-66.
- Lundqvist M, Bengten E, Stromberg S, Pilstrom L. Ig light chain gene in the Siberian sturgeon (*Acipenser baeri*) - Implications for the evolution of the immune system. *J. Immunol.* 1996, 157:2031-2038.
- Lundqvist ML, Middleton DL, Hazard S, Warr GW. The immunoglobulin heavy chain locus of the duck - Genomic organization and expression of D, J, and C region genes. *J. Biol. Chem.* 2001, 276:46729-46736.
- Lundqvist ML, Middleton DL, Radford C, Warr GW, Magor KE. Immunoglobulins of the non-galliform birds: Antibody expression and repertoire in the duck. *Dev. Comp. Immunol.* 2006, 30:93-100.
- Mage RG, Lanning D, Knight KL. B cell and antibody repertoire development in rabbits: The requirement of gut-associated lymphoid tissues. *Dev. Comp. Immunol.* 2006, 30:137-153.
- Malecek K, Brandman J, Brodsky J, Ohta Y, Flajnik M, Hsu E. Somatic hypermutation and junctional diversification at Ig heavy chain loci in the nurse shark. *J. Immunol.* 2005, 175:8105-8115.
- Martinez-Jean C, Folch G, Lefranc MP. Nomenclature and Overview of the Mouse (*Mus musculus* and *Mus sp*) Immunoglobulin Kappa (IGK) Genes. *Exp. Clin. Immunogenet.* 2001, 18:255-279.
- Matsuda F, Ishii K, Bourvagnet P, Kuma K, Hayashida H, Miyata T, Honjo T. The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. *J. Exp. Med.* 1998, 188:2151-2162.

- McCormack W, Carlson L, Tjoelker L, Thompson C. Evolutionary comparison of the avian Ig sub(L) locus: Combinatorial diversity plays a role in the generation of the antibody repertoire in some avian species. *Int. Immunol.* 1989, 1:332-341.
- McCormack W, Tjoelker L, Thompson C. Avian B-Cell Development - Generation of an Immunoglobulin Repertoire by Gene Conversion. *Annu. Rev. Immunol.* 1991, 9:219-241.
- McGavin MD, Zachary JD. Pathologic basis of veterinary disease. Mosby Elsevier, St. Louis, Missouri 2007.
- Murphy K, Travers P, Walport M. Immunobiology. Garland Science, New York 2008.
- Nei M, Gu X, Sitnikova T. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc. Natl. Acad. Sci. U. S. A.* 1997, 94:7799-7806.
- Nguyen VK, Hamers R, Wyns L, Muyldermans S. Loss of splice consensus signal is responsible for the removal of the entire C(H)1 domain of the functional camel IGG2A heavy-chain antibodies. *Mol. Immunol.* 1999, 36:515-524.
- Nguyen VK, Hamers R, Wyns L, Muyldermans S. Camel heavy-chain antibodies: diverse germline VHH and specific mechanisms enlarge the antigen-binding repertoire. *EMBO J.* 2000, 19:921-930.
- Nguyen VK, Su C, Muyldermans S, van der Loo W. Heavy-chain antibodies in Camelidae; a case of evolutionary innovation. *Immunogenetics* 2002, 54:39-47.
- Nowak MA, Parra ZE, Hellman L, Miller RD. The complexity of expressed kappa light chains in egg-laying mammals. *Immunogenetics* 2004, 56:555-563.
- Ohta Y, Flajnik M. IgD, like IgM, is a primordial immunoglobulin class perpetuated in most jawed vertebrates. *Proc. Natl. Acad. Sci. U. S. A.* 2006, 103:10723-10728.
- Ota T, Nei M. Divergent Evolution and Evolution by the Birth-And-Death Process in the Immunoglobulin V-H Gene Family. *Mol. Biol. Evol.* 1994, 11:469-482.

- Ota T, Nei M. Evolution of Immunoglobulin V-H Pseudogenes in Chickens. *Mol. Biol. Evol.* 1995, 12:94-102.
- Ota T, Rast JP, Litman GW, Amemiya CT. Lineage-restricted retention of a primitive immunoglobulin heavy chain isotype within the Dipnoi reveals an evolutionary paradox. *Proc. Natl. Acad. Sci. U. S. A.* 2003, 100:2501-2506.
- Parvari R, Ziv E, Lentner F, Telor S, Burstein Y, Schechter I. Analyses of Chicken Immunoglobulin Light Chain Cdna Clones Indicate a Few Germline V-Lambda-Genes and Allotypes of the C-Lambda-Locus. *EMBO J.* 1987, 6:97-102.
- Parvari R, Avivi A, Lentner F, Ziv E, Telor S, Burstein Y, Schechter I. Chicken Immunoglobulin Gamma-Heavy Chains - Limited Vh Gene Repertoire, Combinatorial Diversification by D-Gene Segments and Evolution of the Heavy-Chain Locus. *EMBO J.* 1988, 7:739-744.
- Pilstrom L. The mysterious immunoglobulin light chain. *Dev. Comp. Immunol.* 2002, 26:207-215.
- Pramanik S, Cui X, Wang H, Ching N, Hu G, Shen L, ym. Segmental duplication as one of the driving forces underlying the diversity of the human immunoglobulin heavy chain variable gene region. *BMC Genomics* 2011, 12:78.
- Qin T, Ren L, Hu X, Guo Y, Fei J, Zhu Q, ym. Genomic organization of the immunoglobulin light chain gene loci in *Xenopus tropicalis*: Evolutionary implications. *Dev. Comp. Immunol.* 2008, 32:156-165.
- Rabbani H, Pan Q, Kondo N, Smith CIE, Hammarstrom L. Duplications and deletions of the human IGHC locus: Evolutionary implications. *Immunogenetics* 1996, 45:136-141.
- Rast JP, Anderson MK, Ota T, Litman RT, Margittai M, Shablott MJ, ym. Immunoglobulin Light-Chain Class Multiplicity and Alternative Organizational Forms in Early Vertebrate Phylogeny. *Immunogenetics* 1994, 40:83-99.
- Ratcliffe MJH. Antibodies, immunoglobulin genes and the bursa of Fabricius in chicken B cell development. *Dev. Comp. Immunol.* 2006, 30:101-118.

- Reynaud CA, Anquez V, Weill J. The Chicken D-Locus and its Contribution to the Immunoglobulin Heavy-Chain Repertoire. *Eur. J. Immunol.* 1991, 21:2661-2670.
- Reynaud CA, Bertocci B, Dahan A, Weill JC. Formation of the Chicken B-Cell Repertoire - Ontogeny, Regulation of Ig Gene Rearrangement, and Diversification by Gene Conversion. *Adv. Immunol.* 1994, 57:353-378.
- Reynaud CA, Garcia C, Hein W, Weill J. Hypermutation Generating the Sheep Immunoglobulin Repertoire is an Antigen-Independent Process. *Cell* 1995, 80:115-125.
- Reynaud S, Delpy L, Fleury L, Dougier HL, Sirac C, Cogne M. Interallelic class switch recombination contributes significantly to class switching in mouse B cells. *J. Immunol.* 2005, 174:6176-6183.
- Ros F, Puels J, Reichenberger N, van Schooten W, Buelow R, Platzer J. Sequence analysis of 0.5 Mb of the rabbit germline immunoglobulin heavy chain locus. *Gene* 2004, 330:49-59.
- Rumfelt LL, Avila D, Diaz M, Bartl S, McKinney EC, Flajnik MF. A shark antibody heavy chain encoded by a nonsomatically rearranged VDJ is preferentially expressed in early development and is convergent with mammalian IgG. *Proc. Natl. Acad. Sci. U. S. A.* 2001, 98:1775-1780.
- Rumfelt LL, Lohr RL, Dooley H, Flajnik MF. Diversity and repertoire of IgW and IgM VH families in the newborn nurse shark. *BMC immunology* 2004, PMC420240:8.
- Saini SS, Hein WR, Kaushik A. A single predominantly expressed polymorphic immunoglobulin V-H gene family, related to mammalian group, I, clan, II, is identified in cattle. *Mol. Immunol.* 1997, 34:641-651.
- Schaerlinger B, Fripiat J. IgX antibodies in the urodele amphibian *Ambystoma mexicanum*. *Dev. Comp. Immunol.* 2008, 32:908-915.
- Schaerlinger B, Bascove M, Fripiat J. A new isotype of immunoglobulin heavy chain in the urodele amphibian *Pleurodeles waltl* predominantly expressed in larvae. *Mol. Immunol.* 2008, 45:776-786.

- Shan Y, Gras R. 43 genes support the lungfish-coelacanth grouping related to the closest living relative of tetrapods with the Bayesian method under the coalescence model. *BMC Research Notes* 2011, 4:49.
- Shen SX, Bernstein RM, Schluter SF, Marchalonis JJ. Heavy-chain variable regions in carcharhine sharks: Development of a comprehensive model for the evolution of V-H domains among the gnathanstomes. *Immunol. Cell Biol.* 1996, 74:357-364.
- Shimizu A, Takahashi N, Yaoita Y, Honjo T. Organization of the constant-region gene family of the mouse immunoglobulin heavy chain. *Cell.* 1982, 28:499-506.
- Sinclair MC, Gilchrist J, Aitken R. Bovine IgG repertoire is dominated by a single diversified V-H gene family. *J. Immunol.* 1997, 159:3883-3889.
- Sitnikova T, Nei M. Evolution of immunoglobulin kappa chain variable region genes in vertebrates. *Mol. Biol. Evol.* 1998, 15:50-60.
- Sitnikova T, Su C. Coevolution of immunoglobulin heavy- and light-chain variable-region gene families. *Mol. Biol. Evol.* 1998, 15:617-625.
- Solem ST, Stenvik J. Antibody repertoire development in teleosts-a review with emphasis on salmonids and *Gadus morhus* L. *Dev. Comp. Immunol.* 2006, 30:57-76.
- Solin MJ, Kaartinen M. Allelic polymorphism of mouse *Igh-J* locus, which encodes immunoglobulin heavy chain joining (J_H) segments. *Immunogenetics* 1992, 36:306-313.
- Spiekerpolet H, Yam P, Knight K. Differential Expression of 13 Iga-Heavy Chain Genes in Rabbit Lymphoid-Tissues. *J. Immunol.* 1993, 150:5457-5465.
- Stavnezer J, Amemiya CT. Evolution of isotype switching. *Semin. Immunol.* 2004, 16:257-275.
- Steinke D, Hoegg S, Brinkmann H, Meyer A. Three rounds (1R/2R/3R) of genome duplications and the evolution of the glycolytic pathway in vertebrates RID D-4700-2009 RID C-9826-2009. *Bmc Biology* 2006, 4:16.

- Stenvik J, Jorgensen TO. Immunoglobulin D (IgD) of Atlantic cod has a unique structure. *Immunogenetics* 2000, 51:452-461.
- Stenvik J, Lundback AS, Jorgensen TO, Pilstrom L. Variable region diversity of the Atlantic cod (*Gadus morhua* L.) immunoglobulin heavy chain. *Immunogenetics* 2000, 51:670-680.
- Storb U, Haasch D, Arp B, Sanchez P, Cazenave P, Miller J. Physical Linkage of Mouse Lambda-Genes by Pulsed-Field Gel-Electrophoresis Suggests that the Rearrangement Process Favors Proximate Target Sequences. *Mol. Cell. Biol.* 1989, 9:711-718.
- Sun Y, Wang C, Wang Y, Zhang T, Ren L, Hu X, ym. A comprehensive analysis of germline and expressed immunoglobulin repertoire in the horse. *Dev. Comp. Immunol.* 2010, 34:1009-1020.
- Sun Y, Wei Z, Hammarstrom L, Zhao Y. The immunoglobulin δ gene in jawed vertebrates: A comparative overview. *Dev. Comp. Immunol.* 2011, 35:975-981.
- Tizard IR. *Veterinary Immunology*. Saunders, St. Louis, Missouri 2009.
- Tobinjanzen TC, Womack JE. Comparative Mapping of *Ighg1*, *Ighm*, *Fes*, and *Fos* in Domestic Cattle. *Immunogenetics* 1992, 36:157-165.
- Tucker P, Slightom J, Blattner F. Mouse Iga Heavy-Chain Gene Sequence - Implications for Evolution of Immunoglobulin Hinge Exons. *Proc. Natl. Acad. Sci. U. S. A.* 1981, 78:7684-7688.
- Tutter A, Riblet R. Conservation of an Immunoglobulin Variable-Region Gene Family Indicates a Specific, Noncoding Function. *Proc. Natl. Acad. Sci. U. S. A.* 1989, 86:7460-7464.
- van Tuinen M, Hedges SB. Calibration of avian molecular clocks. *Mol. Biol. Evol.* 2001, 18:206-213.
- Vernersson M, Belov K, Aveskogh M, Hellman L. Cloning and structural analysis of two highly divergent IgA isotypes, IgA1 and IgA2 from the duck billed platypus, *Ornithorhynchus anatinus*. *Mol. Immunol.* 2010, 47:785-791.

- Wang X, Olp JJ, Miller RD. On the genomics of immunoglobulins in the gray, short-tailed opossum *Monodelphis domestica*. *Immunogenetics* 2009, 61:581-596.
- Warr GW, Magor KE, Higgins DA. IgY - Clues to the Origins of Modern Antibodies. *Immunol. Today* 1995, 16:392-398.
- Wei Z, Wu Q, Ren L, Hu X, Guo Y, Warr GW, ym. Expression of IgM, IgD, and IgY in a Reptile, *Anolis carolinensis*. *J. Immunol.* 2009, 183:3858-3864.
- Wu Q, Wei Z, Yang Z, Wang T, Ren L, Hu X, ym. Phylogeny, genomic organization and expression of lambda and kappa immunoglobulin light chain genes in a reptile, *Anolis carolinensis*. *Dev. Comp. Immunol.* 2010, 34:579-589.
- Xu Z, Wang GL, Nie P. IgM, IgD and IgY and their expression pattern in the Chinese soft-shelled turtle *Pelodiscus sinensis*. *Mol. Immunol.* 2009, 46:2124-2132.
- Yang FX, Ventura-Holman T, Waldbieser GC, Lobb CJ. Structure, genomic organization, and phylogenetic implications of six new VH families in the channel catfish. *Mol. Immunol.* 2003, 40:247-260.
- Yasuike M, de Boer J, von Schalburg KR, Cooper GA, McKinnel L, Messmer A, ym. Evolution of duplicated IgH loci in Atlantic salmon, *Salmo salar*. *BMC Genomics* 2010, 11:486.
- Ye J. The immunoglobulin IGHD gene locus in C57BL/6 mice. *Immunogenetics* 2004, 56:399-404.
- Zhao Y, Rabbani H, Shimizu A, Hammarstrom L. Mapping of the chicken immunoglobulin heavy-chain constant region gene locus reveals an inverted alpha gene upstream of a condensed epsilon gene. *Immunology* 2000, 101:348-353.
- Zhao Y, Kacsokovics I, Pan Q, Liberles DA, Geli J, Davis SK, ym. Artiodactyl IgD: The missing link. *J. Immunol.* 2002, 169:4408-4416.
- Zhao Y, Pan-Hammarstrom Q, Kacsokovics I, Hammarstrom L. The porcine Ig delta gene: Unique constant region domain in its chimeric splicing of the first heavy chain transcripts(1,2). *J. Immunol.* 2003, 171:1312-1318.

- Zhao Y, Hammarstrom L. Cloning of the complete rat immunoglobulin delta gene: evolutionary implications. *Immunology* 2003, 108:288-295.
- Zhao Y, Pan-Hammarstrom Q, Yu S, Wertz N, Zhang X, Li N, ym. Identification of IgF, a hinge-region-containing Ig class, and IgD in *Xenopus tropicalis*. *Proc. Natl. Acad. Sci. U. S. A.* 2006a, 103:12087-12092.
- Zhao Y, Jackson SM, Aitken R. The bovine antibody repertoire. *Dev. Comp. Immunol.* 2006b, 30:175-186.
- Zhao Y, Cui H, Whittington CM, Wei Z, Zhang X, Zhang Z, ym. Ornithorhynchus anatinus (Platypus) Links the Evolution of Immunoglobulin Genes in Eutherian Mammals and Nonmammalian Tetrapods. *J. Immunol.* 2009, 183:3285-3293.
- Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, ym. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 2009, 10:R42.
- Zimmerman AM, Yeo G, Howe K, Maddox BJ, Steiner LA. Immunoglobulin light chain (IgL) genes in zebrafish: Genomic configurations and inversional rearrangements between (V-L-J(L)-C-L) gene clusters. *Dev. Comp. Immunol.* 2008, 32:421-434.