



MAKERA TUTKIMUSHANKE 2012-2013

Teräväpiirto genomitiedon hyödyntäminen jalostusarvosteluissa

HD Genomiikka

High Definition genome wide information in genetic evaluations

LOPPURAPORTTI

Geneettinen Tutkimus

Biotekniikka- ja elintarviketutkimus

Maa- ja elintarviketalouden tutkimuskeskus MTT

Esa Mäntysaari¹, Johanna Vilkki², Minna Koivula¹, Timo Knürr¹ ja Ismo Strandén¹

¹Biometrisen genetiikan tutkimus,

²Eläingenomiikan tutkimus

E-mail: Etunimi.Sukunimi@mtt.fi

Sisältö

Hankkeen ohjausryhmä	2
1. Hankkeen tausta	3
2. Hankkeen tavoitteet.....	4
3. Hankkeen tulokset.....	4
3.1 Hankkeessa käytettävän datan hankinta, laatuanalyysit ja päivitys (WP0).....	4
3.2. Eläinten rotualkuperän huomioon ottava genominen malli (WP1).....	6
3.3. HD-aineistoa tehokkaimmin hyödyntävät laskennalliset mallit (WP2).....	8
3.4 HD-aineiston käyttö assosiaatioanalyysissä, ja monitekijäisiin ominaisuuksiin vaikuttavien geenien tehokartoittaminen (WP3).....	10
3.5 Genotyypitettyjen sonnien ja lehmien arvostelujen yhdistäminen ns. single-step analyysillä kaikkien tuotantotarkkailussa mukana olevien eläinten fenotyyppisiin tuloksiin (WP4)	12
3.3 Toteutusvaiheen arviointi	14
3.4 Julkaisut.....	14
4. Tulosten arviointi	15
4.1 Tulosten käytännön sovellutuskelpoisuus	15
4.2 Tulosten tieteellinen merkitys	16
LIITE 2 Julkaisut.....	20

Hankkeen ohjausryhmä

Puheenjohtaja, MMM, Neuvotteleva virkamies **Tuula Pehu**

Viking Genetics, Rotukoordinaattori **Auli Himanen**

Faba, Tuotekehitys, NAV arvostelut; Tutkimusagronomi **Jukka Pösö**

Valio Oy, Alkutuotanto, Kehityspäällikkö, **Jaana Kiljunen**

Helsingin Yliopisto, Maataloustieteen laitos, Lehtori **Jarmo Juga**

1. Hankkeen tausta

Genomisessa valintaohjelmassa valitaan vanhempaissukupolvi ns. genomisten jalostusarvojen perusteella. Genominen jalostusarvostelu perustuu koko genomien kattavaan tiheään genotyypitykseen, jonka avulla pyritään ennustamaan eläinten jalostusarvo (eli perinnöllinen arvo joka näkyy eläimen jälkeläisten poikkeamana populaation keskiarvosta). Genominen jalostusarvostelu pohjautuu Meuwissen ym. (2001) esittämiin keskeisiin menetelmiin, eli genomiseen BLUP ennusteeseen ja epälineaarisiin BayesA ja BayesB malleihin, joilla pyritään mallintamaan samanaikaisesti eläimen kaikissa kromosomeissa olevat yksittäiset additiiviset geenivaikutukset¹. Schaefferin (2006)² laskelmien perusteella oikein toteutettu genominen jalostusarvostelu voi nopeuttaa jalostuksellista edistymistä jopa kaksinkertaiseksi!

Genomisen valinnan toteuttaminen edellyttää kustannustehokasta genotyypitysalustaa. Tiheitä kytkentäkartoja koottaessa on hyödynnetty DNA:n emäsjärjestyksessä esiintyvää yksittäisten nukleotidien vaihtelua (single nucleotide polymorphism, SNP). Vuonna 2007 Illumina Inc. esitteli BovineSNP50-sirun, jossa SNP merkkien määrä oli 54 000. Vuonna 2010 Illumina³ toi markkinoille BovineHD-sirun, jossa oli jo 795 000 autosomaalista SNP-merkkiä. Lisäksi vuonna 2011 Illumina toi markkinoille BovineLD sirun (2012 BovineLD+), jossa merkkejä oli 4000 (8000), mutta joka laski genotyypityskustannukset alle 50€ hintaluokkaan. Meillä pohjoismaissa genominen valinta tai valintaohjelman kehittäminen käynnistyi vuonna 2008. Genomisten jalostusarvostelujen kehittämiseksi on kaikki Suomessa, Ruotsissa ja Tanskassa 2000-luvulla käytetyt keinosiemennyssonnit genotyypitetty BovineSNP50 tai BovineHD siruilla.

Ensimmäiset pohjoismaiset genomiset jalostusarvostelut kehitettiin yhteispohjoismaisissa tutkimushankkeissa (MTT ja HY, Suomi; Aarhus University, Tanska ja SLU, Ruotsi). Tässä raportoitavaa HD *Genomics* -hanketta edelsi MAKERAn rahoittama *Genominen valinta* 2009-2011. *Genominen valinta* -hankkeen tuloksena otettiin vuonna 2011 käyttöön keinosiemennyssonniin genomiset arvostelut. Niistä on tullut naudanjalostusohjelman tärkeimmät työkalut. Tällä hetkellä (2013) yli 75% holstein keinosiemennyksistä tehdään genomisesti valituilla nuorilla sonneilla. Punaisella rodulla genomisesti valittujen sonniin osuus on noussut 53%:iin. Genomisen valinnan hitaampi käyttöönotto punaisella rodulla johtuu genomisten arvostelujen alhaisemmasta luotettavuudesta verrattuna holstein rodulla saavutettuun luotettavuuteen.

HD *Genomiikka* -hankkeen keskeinen tavoite oli tutkia pohjoismaisen punaisen ja suomalaisen ayrshire rodun genomisten arvostelujen problematiikkaa, ja löytää ratkaisu rodun genomisten arvostelujen ongelmiin.

¹ Meuwissen, T.H.E., Hayes, B.J. and Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps, *Genetics* 157: 1819-1829.

² Schaeffer L.R. J. *Animal Breeding and Genetics*, 2006

³ ©Illumina, Inc. <http://www.illumina.com>

2. Hankkeen tavoitteet

Tutkimushankkeen tavoitteena oli kehittää uusia, koko genomin tuntemukseen perustuvia menetelmiä suomalaisen ayrshiren sekä yleisesti pohjoismaisen punaisen lypsykarjan (RDC) geenien kartoitukseen ja jalostusarvostelujen ja jalostusohjelmien kehittämiseen. Osatavoitteet olivat:

1. Kehittää uusi genominen malli, joka mahdollistaa rotualkuperältään heterogeenisen aineiston luotettavan analysoinnin.
2. Parantaa laskennallisia menetelmiä niin että HD-aineistoa voidaan hyödyntää tehokkaimmin.
3. Hyödyntää HD-aineistoa assosiaatioanalyseissä, ja tehokartoittaa monitekijäisiin ominaisuuksiin vaikuttavia genejä.
4. Yhdistää genotyypitettyjen sonnien ja lehmien arvostelut ns. single-step analyysillä kaikkien tuotannon tarkkailussa mukana olevien eläinten fenotyyppeihin tuloksiin.

3. Hankkeen tulokset

Hanke on jaettu neljään tavoitteiden mukaiseen työpakettiin WP1-WP4 sekä niitä tukevaan ja yhdistävään työpakettiin (WP0). Seuraavassa on esitelty työpaketeittain ensin alkuperäiset tavoitteet (sisennettynä) ja näiden jälkeen on käyty läpi tärkeimmät tulokset. Menetelmiin liittyvät yleiset viitteet on annettu alaviitteinä, omat tutkimukset löytyvät LIITTEEN 2 julkaisuluettelosta.

3.1 Hankkeessa käytettävän datan hankinta, laatuanalyysit ja päivitys (WP0)

Naudan uusi HD-siru (Illumina) sisältää noin 770 000 merkkiä. Hanke saa käyttöönsä Quantomics-hankkeessa Tanskassa HD-tyypitettyjen 250 suomalaisen ay-sonnin genotyypit. Ensin tälle aineistolle tehdään kattava laatutarkistus (tyypitysvirheet, id-virheet, mendelistinen periytyminen, informatiivisuus/MAF). Suomalaisten ay-genotyypitysten lisäksi hyödynnämme SRB (210) ja RDM (247 kpl) HD-genotyypit.

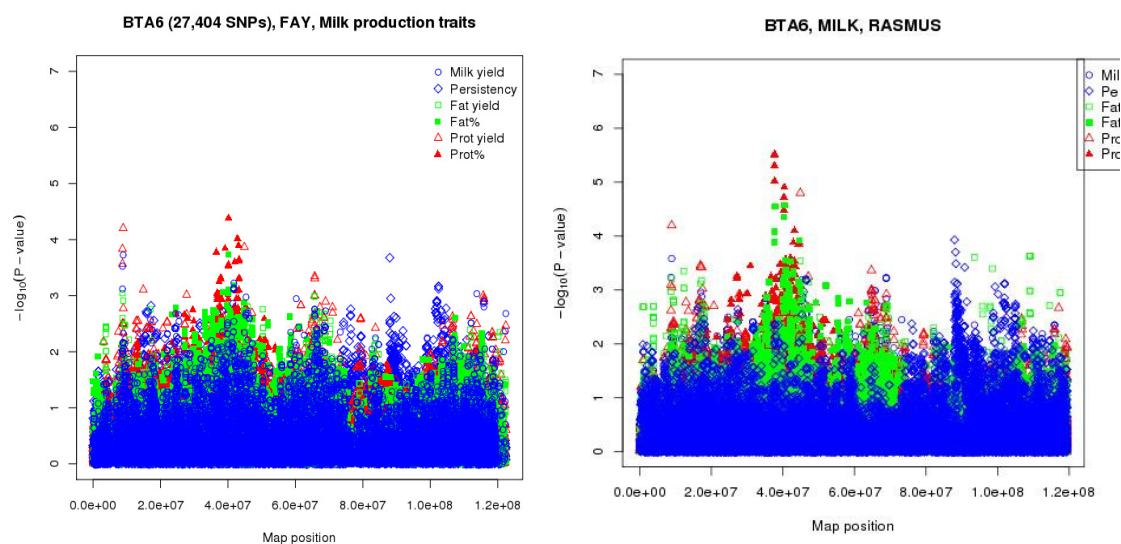
HD –sirulla määritettyjen eläinten määrä on pienempi kuin 54 000 merkkiä sisältävällä sirulla (54K) määritettyjen, joten on tarve imputoida (ennustaa tyypittämättömien lokusten genotyypit) kaikki 54K-sirulla määritetyt eläimet HD-tiheyteen. Imputointi tehdään 1) pelkästään suomalaisen ayrshiren sisällä, 2) yli koko punaisen populaation 3) punaiset ja holstein yhdessä. Näistä saatuja genotyyppejä verrataan keskenään ja alkuperäisiin todellisiin genotyyppeihin sen selvittämiseksi, millä tavalla imputointi antaa parhaiten todellista vastaavan tuloksen.

Käytössämme oli 54K-sirun genotyypitystiedot 3387 suomalaisesta ayrshire-naudasta. Näistä 238 yksilölle saatiin Quantomics-projektin kautta genotyypit Illuminan Bovine HD BeadChipillä, joka sisältää yhteensä 777 962 SNP-merkkiä. HD- aineistolle tehtyjen laaduntarkistusten jälkeen käytettävissä oli 565 360 SNP-merkin genotyypit. SNP-merkkien karttana kaikissa analyyseissä käytettiin naudan genomin versiota UMD3.1. Näiden perusteella imputoitiin (Beagle⁴) koko eläinjoukolla HD-genotyypit. Imputoinnin keskimääräinen tarkkuus oli $r^2 = 0.93$, vain alle 3 %:lla markkereista tarkkuudeksi tuli alle 0.8 eli $r^2 < 0.8$. Huonosti ennustettavat markkerigenotyypit näyttivät esiintyvän tietyillä alueilla genomissa, yleensä kromosomien päissä sekä myös kromosomeissa 12 ja 26. Imputointitapojen vertailua varten saimme Tanskasta HD-aineiston, jonka imputoinnissa oli käytetty koko punaista populaatiota joko pelkästään tai yhdessä holsteinin kanssa. Punaisen rodun

⁴ Browning, S.R., and B.L. Browning. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. The American Journal of Human Genetics 81.5 (2007): 1084-1097.

eläimistä oli HD-sirulla genotyypitettyjä eläimiä 723, joita käytettiin 8085 eläimen genotyyppien imputointiin 54K sirulta HD sirun tiheyteen. Imputaatioon testattiin muutamia ohjelmia ja ohjelmissa eri vaihtoehtoja, mm. sukupuun huomiointia. Imputaation testauksen ja imputoinnin teki Rasmus Brøndum Aarhusin yliopistossa (Brøndum ym., 2012⁵). Parhaimmaksi imputointiohjelmaksi osoittautui Beagle, joka oli nopea ja melko luotettava.

Testasimme imputoitujen genotyyppien käyttökelpoisuutta maidontuotanto-ominaisuuksien assosiaatiotesteissä. Totesimme, että suuremman aineiston avulla imputoidut genotyypit antoivat myös assosiaatiokartoituksessa tarkemman tuloksen.



Kuva 3.1. Suomalaisesta ayrshire-aineistosta imputoitujen HD-genotyyppien (vasen kuva) ja koko pohjoismaisen punaisen rotuaineistolla imputoitujen HD-genotyyppien (oikeanpuoleinen kuva, Rasmus Brøndum (2012)) perusteella tehdyt maito-ominaisuuksien assosiaatioanalyysit (GenAbel) naudan kromosomissa 6. X-akselilla on assosiaation merkitsevyys, y-akselilla paikka kromosomissa. Assosiaatioiden merkitsevyys kasvoi ja paikka tarkentui käytettäessä isomman aineiston perusteella imputoituja genotyyppejä.

Genomisessa arvostelussa ja assosiaatioanalyseissä käytetään ominaisuuksina maidontuotanto- (maito-, valkuais- ja rasvatuotokset sekä valkuais- ja rasvapitoisuudet), utareterveysominaisuuksia (kliininen utaretulehdus, somaattinen soluluku) ja naarashedelmällisyysominaisuuksia. Näiden ominaisuuksien fenotyyppiaineisto päivitetään kattamaan nuorimmat eläimet. Utareterveysominaisuuksille tuotetaan koko punaista populaatiota käyttäen uudet DRP (deregressed breeding values) tai DYD-arvot (daughter yield deviation). Maidontuotanto-ominaisuuksien osalta arvioidaan ensin mahdolliset hyödyt siirtyä deregressoiduista jalostusarvoista DYD-arvoihin.

Assosiaatioanalyysijä varten päivitettiin fenotyyppiaineisto kattamaan nuorimmatkin eläimet. Näiden eläinten spermanäytteet haettiin Hollolasta ja eristettiin DNA jatkotutkimuksia varten (WP3). Utareterveysominaisuuksille laskettiin koko pohjoismaisen punaisen populaation pohjalta uudet DYD-arvot. Naarashedelmällisyyden ja maidontuotannon osalta käytettiin DRP-arvoja.

Genomisten arvostelujen testauksessa ja kehittämisessä käytettiin suoraan pohjoismaisen jalostusarvosteluyhtiön (NAV) arvosteluista laskettuja DRP-arvoja. Isämalliin perustuvaa GBLUP mallia varten deregressiot tehtiin vain sonnien jalostusarvoille, mutta kun siirryttiin ns. single-step analyysiin (WP4) deregressio tehtiin kaikille pohjoismaisille punaisille lehmillä. Mallin kehityksessä päähuomio kiinnitettiin tuotanto-ominaisuuksiin (maito, rasva ja valkuainen). Osassa

⁵ Brøndum, R. F., Ma, P., Lund, M. S., & Su, G. 2012. Short communication: Genotype imputation within and across Nordic cattle breeds." *Journal of dairy science* 95.11 (2012): 6795-6800.

tutkimuksista arvostelumallia testattiin myös alhaisemman periytymisasteen ominaisuuksilla, joina käytettiin sekä hedelmällisyysarvosteluja että utareterveysarvosteluja. Haploblokkimallin kehityksessä (WP1) haluttiin tuloksien olevan vertailukelpoisia NAVn laskemiin genomisiin arvosteluihin ja tämän vuoksi käytettiin suoraan NAVn laskemia DRP arvoja.

Genomisten arvostelujen validointi on perustunut ns. ”Interbull GEBV validation” testiin. Se edellyttää, että kokonaisaineistosta muodostetaan tutkimusreferenssi-aineisto katkaisemalla pois noin 4 viimeisimmän vuoden fenotyypiset havainnot. Kun kehitettyä genomista mallia sovelletaan katkaistuun aineistoon, saadaan nuorille genotyypitetyille eläimille GEBVt ilman tytärinformaatiota. Validaatioissa tarkastellaan näiden GEBV arvojen ja tulevien tyttärien perusteella arvioitavien jalostusarvojen yhteyttä. Yhdenmukaisuuden vuoksi on yksinkertaisinta käyttää kaikissa tutkimustyöpaketeissa samaa GEBV-validaatio -aineistorakennetta.

Single-step (WP4) mallissa genomisen informaatio yhdistetään suoraan valtakunnallisiin jalostusarvosteluihin. Eli tällöin havainnot eivät ole sonnikohtaisia DRP-arvoja, vaan havaintoja tyttäristä. Tämän jälkeen mallin validointia ei voitaisi enää tehdä nuorilla sonneilla. Tätä varten omaksuimme validaatioprotokollan, jossa genomisen single-step malli ratkaistiin lehmiin perustuvasta katkaistusta havaintoaineistosta. Tämän jälkeen koko havaintoaineistoa käytetään valtakunnallisiin arvosteluihin, joista nuorien kandidaattisonniin arvostelut deregressoidaan validaatiotarkoituksiin.

3.2. Eläinten rotualkuperän huomioon ottava genomisen malli (WP1)

Päätyneen Genominen valinta –hankkeen aikana kehitettiin malli, jossa joka eläimelle määriteltiin genomisen jalostusarvo jokaisen alkuperäisen perusrodun suhteen. Tämä perustui satunnaisregressiomalliin ja genotyyppien perustella laskettuun genomiseen sukulaisuusmatriisiin.

Genomisessa sukulaisuusmatriisissa käytettiin keskimääräisiä pohjoismaisen populaation alleelifrekvenssejä. Nyt Genominen sukulaisuusmatriisiin yksilöiden (odotettavissa olevat) alleelifrekvenssit korvataan rotualkuperän huomiovilla ennustetuilla frekvensseillä. Tätä mallia testataan sekä GBLUP, että single-step (WP4) arvosteluissa.

Uutena lähestymistapana käytetään ns. MAS arvostelumallia. Aluksi kartoitetaan noin 200-400 ominaisuuteen eniten vaikuttavaa genomi-segmenttiä. Seuraavaksi etsitään segmenttien yli määritellyt haplotyyppit, jotka lisätään arvostelumalliin. Geenialueet tunnistetaan sekä WP2:n muuttujavalinta malleja hyödyntäen että WP3:n assosiaatiokartoitusmalleja hyödyntäen. Haplotyyppiblokkien perusteella voidaan rakentaa keskimääräinen genotyyppiin perustuva geenialkuperämatriisi (IBD relationship matrix), joka hyödyntäisi paremmin alkuperäisissä peruspopulaatioissa olleet SNP-merkki – geenivaikutus assosiaatiot. Haplotyyppiblokkien määrittelyssä voidaan myös hyödyntää suoraan rotualkuperän mukaista eläinten klusterointia, ja ns. ancestral origin segmenttejä

Tällä hetkellä pohjoismaisessa genomisessa arvostelussa käytetään ns. GBLUP menetelmää, jossa eläinten genomiset arvot (DGV) lasketaan hyödyntäen genomista sukulaisuusmatriisia, ja lopuksi DGV arvot yhdistetään olemassa oleviin arvosteluihin genomisiksi jalostusarvoiksi (GEBV).

Monirotuisen populaation genomisia jalostusarvoja voidaan mallintaa satunnaisregressiomallilla. Tällöin oletetaan, että genomisen sukulaisuusmatriisi kuvaa eläinten väliset sukulaisuudet kuten homogeenisessä populaatiossakin, mutta että eläimen sukutaulun mukaan laskettu rotuosuus kertoo paljonko yksilössä on kunkin rodun peruspopulaation ”jalostusarvoa”. Mallinnettaessa pohjoismaisen punaisen rodun genomisia jalostusarvoja satunnaisregressiomallilla (Makgahlela ym. 2013a) laskettiin joka eläimelle rotuosuus neljästä peruspopulaatiosta (suomalainen ay, ruotsalainen SRB, norjalainen NRF, ja muut rodut yhteensä, ml. Tanskan RDM). Esimerkiksi Suomessa vuonna 2006 kantakirjatuissa punaisen rodun sonneissa oli keskimäärin 57% alkuperäistä ay, 24% SRB, 17% NRF ja 2% muita punaisia rotuja. Satunnaisregressiomallilla laskettujen DGVn validointiluotettavuus oli

0.32 (0.30 ja 0.29) maidolle ja valkuaiselle ja 0.42 (0.43) rasvatuotokselle (suluissa vertailun vuoksi normaalin yhdenrodun genomisen mallin luotettavuus samassa aineistossa).

Yli roturyhmien laskettu genominen sukulaisuusmatriisi olettaa, että genotyypitettyjen eläinten keskinäinen genominen sukulaisuus on keskimäärin nolla, samoin kuin kaikkien eläinten genominen sukusiitosaste. Tämä oletus poikkeaa sukupuun perusteella lasketuista arvoista, joiden mukaan vanhimpien, perussukupolven eläinten sukusiitos keskinäinen sukulaisuus ovat nolla. Genotyyppeihin ja sukupuuhun perustuvat sukulaisuuskertoimet saatiin samalle tasolle, kun SNP-merkkien alleelifrekvenssit arvioitiin ns. geenikylläisyys mallilla (Gengler ym., 2007)⁶ ja käyttäen rotuosuuksia kovariaabeleina (Makgahlela ym. 2013b). Alleelifrekvenssejä käytetään laskettaessa sukulaisuuskertoimia, ja näin ollen populaation keskiarvosta poikkeavat alleelifrekvenssit johtavat suuriin eroihin genomisissa sukulaisuuksissa. Tämä osoittautui todeksi erityisesti Tanskassa rekisteröityjen sonnien kohdalla. Alkuperäisten (ja virallisissa genomisissa arvosteluissa käytettyjen) alleelifrekvenssien perusteella tanskalaisten sonnien genomisen sukulaisuusmatriisin diagonaalelementit olivat keskimäärin 1.136 (kaikkien keskiarvo 1.019). Kun malli perustui neljän eri perussukupolven alleelifrekvensseihin, laski tämä keskiarvo tanskalaisten sonnien kohdalla 0.976 (kaikkien keskiarvon ollessa 0.986). Luku ei ole täysin vertailukelpoinen sukupuusta laskettavaan a_{ii} (eli $1 + \text{sukusiitosaste}$) termiin, mutta on silti yhtenevä polveutumisen perusteella laskettujen sukusiitosasteiden kanssa. Rodun historian perusteella tanskalaisten sonnien sukusiitosaste on alhaisempi kuin Suomessa ja Ruotsissa syntyneillä, ja Tanskan sonnit ovat geneettisesti etäällä muista maista. Vaikka eri alleelifrekvensseillä lasketut genomiset sukulaisuusmatriisit poikkeavat toisistaan paljonkin, tällä ei juuri ollut vaikutusta genomisten arvostelujen luotettavuuteen. Single-step arvostelussa (kts. WP4) käytettävän genomisen sukulaisuusmatriisin tulee olla vertailukelpoinen normaalin sukulaisuusmatriisin kanssa. Normaalisti genominen sukulaisuusmatriisi skaalataan niin, että sen diagonaalelementtien keskiarvo (minus 1)⁷ on sama kuin genotyypitettyjen eläinten sukupuun perusteella laskettujen sukusiitosasteiden keskiarvo. Kun käytettiin rotukohtaisin alleelifrekvenssein laskettua genomista sukulaisuusmatriisia, voitiin ko. skaalaus jättää tekemättä (Makgahlela ym. 2013c). Samalla genomisten jalostusarvojen jakauma normalisoitui, mutta arvostelujen luotettavuus oli sama kuin aikaisemminkin.

Nyt käytössä oleva GBLUP perustuu oletukseen että jokaiseen ominaisuuteen vaikuttaa monia (satoja tai tuhansia) pienivaikutteisia geenejä (QTL'ia), ja a priori jokainen SNP geenimerkki saattaa liittyä johonkin QTL'ään. Ns. bayesiläisillä muuttujaosajoukon valintamenetelmillä (variable selection, ks. WP2) pyritään valitsemaan ennusteyhtälöön vain ne SNP merkit joilla on aito yhteys (linkage) geenivaikutusten kanssa. Tämä yhteys on ilmeisen usein populaatio- ja ominaisuuskohtainen, ja kun pohjoismainen punainen rotu sisältää useita eri peruspopulaatioita, saattavat nämä yhteydet olla eri perheissä erilaisia, ja lisäksi risteytyksissä syntyy virheellisiä yhteyksiä. Testaamassamme haplotyyppiblokkien mallissa oletetaan, että QTL on paremmin tunnistettavissa, kun seurataan useiden peräkkäisten SNP merkkien periytymistä "blokkina" perussukupolvesta lähtien. Ensinnäkin käytettiin BayesB menetelmää suurimpien SNP-merkkivaikutusten tunnistamiseen. Näiden perusteella valittiin yksittäisiä merkkejä (750 tai 1500) joiden ympäriltä tarkasteltiin viiden SNP merkin haploblokkeja. Haploblokkien rakentaminen edellyttää genotyyppien faasaamista haplotyypeiksi, joissa kaikilla eläimillä perimä on jaettu eri vanhemmilta saatujen kromosomien tai kromosominosien ketjuiksi. Faasaminen tehtiin ensin Findhap -ohjelmalla (VanRaden ym., 2011)⁸ ja myöhemmin Beagle ohjelmalla. Vertasimme 750 ja 1500 haploblokki -tekijää sisältäviä jalostusarvostelumalleja GBLUP menetelmään (Knurr ym. 2013). Samalla testattiin viiden, kolmen tai yhden SNP-merkin haploblokki-pituuksia. Maidon arvosteluissa paras haploblokki-malli oli parempi ($R^2=0.48$) kuin GBLUP ($R^2=0.46$), rasvan ja tytärhedelmällisyyden kohdalla molemmat mallit olivat yhtä hyviä, ja

⁶ Gengler, N., P. Mayeres, and M. Szydlowski. 2007. A simple method to approximate gene content in large pedigree populations: Application to the myostatin gene in dual-purpose Belgian blue cattle. *Animal* 1:21-28.

⁷ Sukusiitoskerroin = (sukulaisuusmatriisin diagonaalelementti - 1)

⁸ VanRaden, P. M., O'Connell, J. R., Wiggans, G. R., & Weigel, K. A. 2011. Genomic evaluations with many more genotypes. *Genet. Sel. Evol.* 43(10).

valkuaisarvosteluissa GBLUP oli parempi ($R^2=0.41$ vs. 0.43). Haploblokkimallin teoreettista potentiaalia arvioitiin valitsemalla parhaat SNP-merkit käyttäen koko aineistoa. Kun näin valittujen haploblokkien ratkaisut laskettiin referenssipopulaatiosta, saatiin testiaineistossa kolmen SNP-merkin haploblokki mallille luotettavuudet $R^2_{\text{maito}}=0.67$, $R^2_{\text{valkuainen}}=0.58$, $R^2_{\text{rasva}}=0.57$ ja $R^2_{\text{hedelm.}}=0.54$. Näin ollen haploblokkimallin ongelmaksi pääteltiin väärin blokkien valinta. Koska blokkien valinta ei perustunut varsinaisesti itse blokkeihin vaan yksittäisiin SNP merkkeihin, analyysit on päätetty tehdä uudestaan valitsemalla blokit suoraan ilman välivaihetta. Tähän antaa mahdollisuuden uusi analysointityökalu Bayz (<http://www.bayz.biz/>).

3.3. HD-aineistoa tehokkaimmin hyödyntävät laskennalliset mallit (WP2)

Genomisissa malleissa yleisimmin oletetaan että kaikilla SNP-merkeillä on yhtä suuri mahdollisuus kuvata jalostusarvoa (GBLUP). Epälineaarissa tai ns. bayes-malleissa oletetaan, että kaikilla vaikutuksilla on oma a'prior jakauma, jolloin perusoletuksen mukaan osa SNP-merkeistä ei ole yhteydessä QTL-vaikutuksiin. Perusaineiston genotyypitystiheyttä käytettäessä GBLUP ja bayes mallien välillä ei ole nähty merkittävää eroa mallien antamissa arvosteluvarmuuksissa. Toisaalta monissa tutkimuksissa on havaittu, että GBLUP menetelmää ei juurikaan voida parantaa käyttämällä HD-genotyyppisiä.

Tämän vuoksi HD-tuloksien mallintamiseen käytetään bayes'laista muuttuja valintamenetelmää. Markkerien valinta voi perustua joko osana mallia tai kaksivaiheisena. Osana mallia perustuvia menetelmiä ovat mm. BayesB ja sen yleistykset sekä muut mixture-tyyppiset mallit kuten stochastic search variable selection. Kaksivaiheisessa menettelyssä tavoitteena on löytää noin 10 000 – 20 000 SNP –merkkiä, joiden yhteydessä on jalostusarvoa ennustavia geenimuotoja. Tällöin ensimmäisessä vaiheessa siis hyväksytään jonkun kriteerin mukaisesti tilastollisesta mallista markkereita. Esimerkiksi ensimmäisessä vaiheessa voidaan käyttää BayesB mallia, jonka perusteella hyväksytään suurimman markkerivaikutuksen SNP-merkit. Mallinnuksessa voidaan jakaa SNP-merkkejä esim. kromosomeittain blokkeihin joille voidaan muodostaa tarkempi malli kuin perinteisellä kaikki markkerit yhtenäisesti käsittelevälle mallilla. Toinen vaihtoehto HD-tulosten käytölle on fyysisesti lyhempien ja yhdenmukaisemmin käyttäytyvien haplotyyppiblokkien määrittely, jossa vierekkäiset SNP-merkit muodostavat blokkeja.

Pohjoismaisen RDC rodun genomista aineistoa analysoitiin käyttämällä joko 54K tai HD sirujen SNP-geenotyyppisiä. Valitsimme 54K sirulta 36 379 SNP-markkeria niiden laadun perusteella. Samojen laatu-kriteerien perusteella valittiin HD sirulta 410 559 SNP-markkeria. Käytimme pohjoismaisesta lypsykarja-arvostelusta saatuja tuotanto-ominaisuuksien tietoja maito-, rasva- ja valkuaiskiloille. Analyysi perustui perinteiseen ennustevalidointimalliin, jossa 1470 sonnia muodostivat referenssipopulaation ja 710 sonnia kandidaattipopulaation. Kaikki sonnit olivat syntyneet Suomessa ja kandidaattisonnit olivat syntyneet myöhemmin kuin referenssipopulaation sonnit.

Tarkastelimme kolmea tilastollista mallia. Ne olivat tavallinen normaalijakauman malli (Gc), yleistetty BayesA (GtA) ja yleistetty BayesB (GtB). Koska tässä yhtenä päämääränä oli bayes'laisten mallin valintakriteerien tarkastelu, laskenta perustui Markov chain Monte Carlo menetelmiin, jotka ovat laskennallisesti melko intensiivisiä. Käytännön jalostusarvostelu perustuu usein laskennallisesti helpompiin BLUP-malleihin. Bayeslaisten mallien voidaan kuitenkin odottaa toimivan paremmin, kun käytettyjen SNP markkerien määrä on suuri, koska bayeslaiset mallit pystyvät paremmin erottamaan yksittäisten SNP markkerien vaikutuksia kuin esimerkiksi SNP-BLUP. Tämän takia bayes'laisia malleja käytetään usein genomisen aineiston assosiaatioanalyysissä (GWAS, genomwide association study). BayesA ja BayesB mallit ovat normaalijakauman mallia parempina erityisesti GWAS tutkimuksissa, mutta ne voivat toimia hyvin myös tavallisessa jalostusarvon ennustamisessa.

Tilastollisten mallien vertailussa ja eri SNP sirujen käytössä käytettiin neljää kriteeriä. Referenssipopulaation perusteella laskettiin ennusteyhtälöt kandidaateille. Näiden arvojen perusteella laskettiin arvostelujen luotettavuus (R^2) eli korrelaation neliö ennusteiden ja oikeiden jalostusarvojen välille. Mitä suurempi R^2 , sitä parempi ennustemalli. Tämän lisäksi laskettiin kolme vaihtoehtoista kriteeriä. MSEP eli mean squared error of prediction käyttää samoja arvoja kuin R^2 , mutta arvojen välille lasketaan keskimääräinen neliöity erotus. Tämä on bayeslaisen tilastotieteen usein käyttämä ennustamisen kriteeri. Mitä pienempi MSEP sitä parempi ennustemalli. DICia (deviance information criterion) ja logaritminen mallin todennäköisyys (logPD) ovat bayeslaisessa tilastotieteessä käytettyjä mallien vertailukriteerejä. Ne eivät ole ennustepohjaisia eivätkä siis käytä kandidaattisonnien tietoja vaan perustuvat mallin toimivuuteen referenssipopulaatiossa. Molempien kriteerien osalta, mitä pienempi luku sitä parempi malli. Esitetty kahden mallin todennäköisyys logPD on -2 kertaa logaritmi mallin todennäköisyydestä, mikä mahdollistaa vertailun DIC lukuun. DIC approksimoi logPD arvoa, mutta laskennallisesti DIC on helpompi ja logPD luvun lasku on numeerisesti epästabiili. Kahden mallin todennäköisyyttä (logPD) voidaan käyttää laskemaan mallien vertailua kuvaava Bayesin osamäärä. Kun mallien välinen erotus on suurempi kuin 10, niin mallien katsotaan poikkeavan merkittävästi toisistaan.

Tulosten perusteella (Taulukot 3.3.1 ja 3.3.2) HD sirulla saadaan hiukan parempia ennusteita kuin 54K sirun avulla. Erot ovat kuitenkin ennustemalleissa melko pieniä. HD siru voi olla parempi kuin 54K siru, kun päämääränä on GWAS. Tätä ei kuitenkaan tässä vertailussa tutkittu.

Luotettavuus R^2 ja MSEP johtivat samoihin tulkintoihin malleista: suuri luotettavuus johti pieneen MSEP:hen. Samoin usein alhainen DIC vastasi suurta luotettavuutta. Sekä MSEP ja DIC näyttivät mallien välillä selvempiä eroja kuin luotettavuus. Mallin todennäköisyys logPD ei aina ollut yhteneväinen muiden kriteerien kanssa. Tämä johtunee siitä että tämän suuren laskeminen on numeerisesti epästabiilia käytetyllä menetelmällä, jolloin mallien väliset erot jäävät numeeristen ongelmien varjoon.

Taulukko 3.3.1. Maidontuotanto-ominaisuuksien tilastollisia tuloksia 54K genomisirulla (36379 SNP markkeria) kolmella eri mallilla (Gc, GtA, GtB) neljällä eri mallin hyvyttä kuvaavalla mitalla (r , MSEP, DIC, logBF). Mallit ovat Gc= normaalijakauman malli, GtA= yleistetty BayesA, GtB= yleistetty BayesB. Tilastolliset suureet ovat: R^2 = luotettavuus, MSEP= keskimääräinen neliöity erotus ennusteen ja oikean arvon välillä, DIC= deviance information criterion, joka approksimoi arvoa logPD, logPD= logaritminen mallin todennäköisyys.

		R^2	MSEP	DIC	logPD
Maito	Gc	0.29	74.6	9404	9103
	GtA	0.30	74.3	9307	9008
	GtB	0.31	74.1	9297	8983
Rasva	Gc	0.40	69.0	9325	9110
	GtA	0.45	64.2	9216	8987
	GtB	0.45	63.8	9207	8995
Valkuainen	Gc	0.27	95.5	9531	9216
	GtA	0.28	94.5	9510	9206
	GtB	0.28	94.5	9517	9240

Taulukko 3.3.2. Maidontuotanto-ominaisuuksien tilastollisia tuloksia HD genomisirulla (410 559 SNP markkeria) kolmella eri mallilla (Gc, GtA, GtB) neljällä eri mallin hyvyttä kuvaavalla mitalla (r , MSEP, DIC, logBF). Mallit ovat Gc= normaalijakauman malli, GtA= yleistetty BayesA, GtB= yleistetty BayesB. Tilastolliset suureet ovat: R^2 = luotettavuus, MSEP= keskimääräinen neliöity erotus ennusteen ja oikean arvon välillä, DIC= deviance information criterion, joka approksimoi arvoa logPD, logPD= logaritminen mallin todennäköisyys.

		R^2	MSEP	DIC	logPD
Maito	Gc	0.30	71.4	9231	9182
	GtA	0.33	68.6	9117	8769
	GtB	0.33	67.3	9070	8817
Rasva	Gc	0.42	64.8	9121	9292
	GtA	0.45	59.4	8933	8689
	GtB	0.47	59.0	8816	8530
Valkuainen	Gc	0.28	91.3	9347	9119
	GtA	0.28	90.9	9345	8956
	GtB	0.29	90.5	9340	8914

3.4 HD-aineiston käyttö assosiaatioanalyyseissä, ja monitekijäisiin ominaisuuksiin vaikuttavien geenien tehokartoittaminen (WP3)

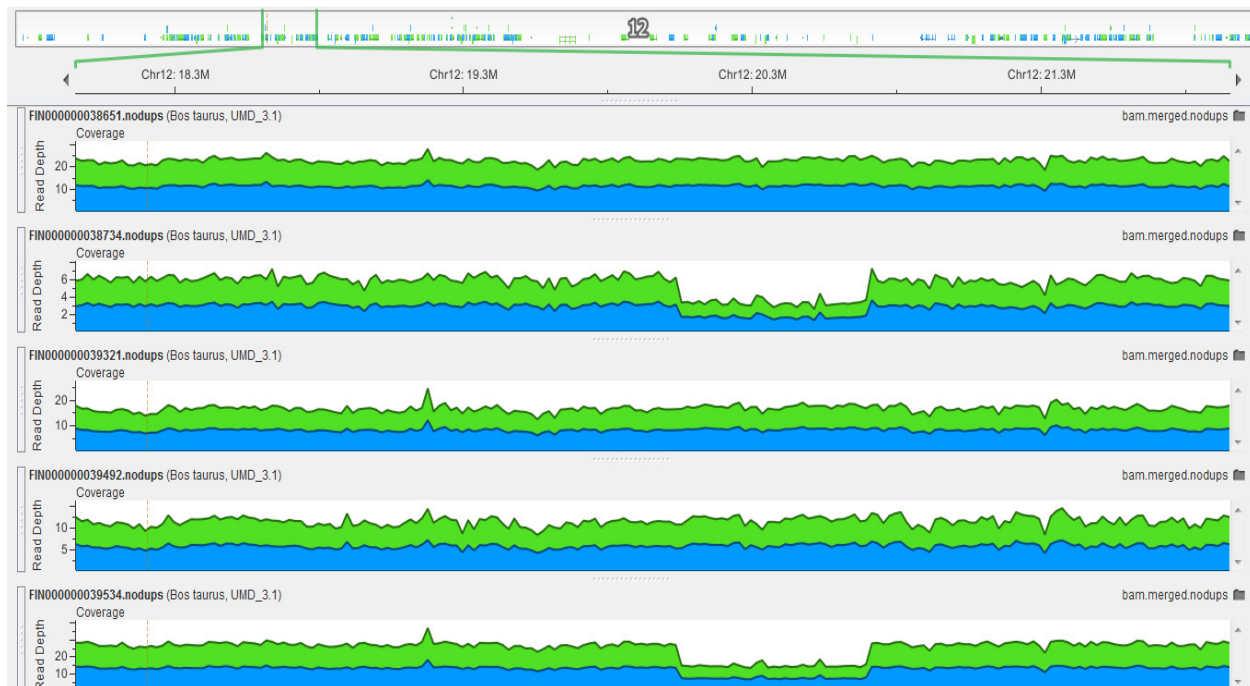
Kartoitetaan tärkeimpiin maidontuotanto-, utareterveys- ja hedelmällisyysominaisuuksiin vaikuttavat QTL-vaikutukset käyttäen HD-tiheyteen imputoitua aineistoa. Aineistona käytetään sekä suomalaisia Ayrshirejä yksinään että yhdistettyä punaista aineistoa. Assosiaatiokartoituksessa käytetään yksittäisiä SNP-vaikutuksia käyttäviä malleja ja haplotyyppisiä hyödyntäviä malleja, joiden tuloksia verrataan keskenään. Assosiaatiokartoituksesta saatuja SNP-vaikutusten ratkaisuja verrataan genomiarvostelussa käytettävillä Bayesiläisillä malleilla saatuihin ratkaisuihin. Tulokset hyödynnetään jalostusarvostelumallissa, sekä mahdollisesti sairausgeeneille kohdistetuissa geenitesteissä. Yhdistämällä 2. ja 3. kohdan kokemukset voidaan menetelmiä testata myös sekvenssidatalle.

Koko genomien kattavat HD-assosiaatioanalyytit tehtiin suomalaiselle ayrshirelle 1786 jälkeläisarvostellun sonnin avulla, jotka oli ensin genotyyppitetty naudan 54K-sirulla, ja joille oli laskettu DYD- (utareturehdus: mastiitti 1, mastiitti 2, SCS1, SCS2, SCS3) tai DRP-arvot (maidontuotanto-ominaisuudet, hedelmällisyysominaisuudet) (WP0). Sonneille imputoitiin 565 360 SNP-merkin genotyypit (WP0). Markkeri kerrallaan tehdyissä assosiaatioanalyyseissä (GenAbel; SVS7, Golden Helix), korjattiin mahdollinen populaatorakenteen vaikutus käyttämällä 54K-genotyyppien perustuvaa sukulaisuusmatriisia. Yleisesti HD-datalla tehdyt analyytit tarkensivat 54K-datalla saatuja assosiaatiokartoituksen tuloksia ja tunnistivat myös uusia alueita.

Assosiaatioanalyyseistä saatuja SNP-vaikutusten ratkaisuja maidontuotanto-ominaisuuksissa verrattiin genomiarvosteluissa saaduilla malleilla saatuihin ratkaisuihin. Verrattaessa genomiarvostelusta saatujen suurimpien vaikutusten sijaintia assosiaatioanalyyseiden merkitsevimpien alueiden sijaintiin, todettiin, että nämä eivät aina osuneet samoihin kohtiin (tai ainakaan samoihin SNP-merkkeihin). Toisin sanoen, suurimmiksi arvioidut vaikutukset eivät aina ole niitä merkitsevimpiä. Genominen arvostelu tehdään yleensä pelkästään arvioitujen vaikutusten, ei niiden merkitsevyyden perusteella.

Painottamalla varvimpia suurivaikutteisia SNP-merkkejä (ääriesimerkkinä tunnistetut kausaaliset mutaatiot) genomisessa arvostelussa voitaisiin parantaa arvosteluvarmuutta⁹.

Hedelmällisyysominaisuuksien HD-kartoituksella tunnistettiin assosioituneita alueita lähes kaikista kromosomeista; merkittävimmät olivat kromosomeissa 1, 2, 6, 8, 10, 12, 17, 19, 24 ja 27. Kaikkein merkittävin löydös oli kromosomissa 12. Muutamia alueita tarkasteltiin tarkemmin DNA-sekvenssin tasolla. Käytössämme oli 18 suomalaisen Ayrshire populaation esi-isäsonnin koko genomisen sekvenssi (Quantomics-hankkeesta sekvensoitu). Kromosomista 12 löytyi assosioituneen alueen yhteydestä yli 600 kiloemäksen mittainen deleetio kuudelta sekvensoiduista sonneista.



Kuva 3.4. Kromosomin 12 deleetioalue sekvensointiaineistosta nähtynä. Deleetion kantajilla (toiseksi ylin ja alin rivi) sekvenssin peittoaste deleetion alueella on puolet normaalista.

Deleetion vaikutus hedelmällisyyteen ja sen esiintyminen muissa populaatioissa tutkittiin yhteistyönä Aarhusin yliopiston ja Liegen yliopiston kanssa (Kadri ym. 2014). Deleetio todennäköisimmin vaikuttaa hedelmällisyyteen aiheuttamalla varhaisen alkion kuoleman, mikä pidentää tyhjääkautta ja lisää siemennysten määrää. Deleetio on yleisin suomalaisessa Ay-populaatioissa (32 % eläimistä on kantajia). Samalla alueella on todettu myös maidontuotantoon vaikuttava QTL (mm. Schulman ym. 2008), johon kohdistunut valinta on voinut johtaa deleetion yleistymiseen.

Utaretulehdukseen assosioituneista alueista valittiin 12 lähempään tarkasteluun. Näiltä alueilta annotoitiin kaikki sekvenssivariaatiot (Ensembl v71 gene set, UMD3.1; GATK). Laatatarkistukset läpäisseistä 240 051 SNP-merkistä 1834 sijaitsi koodaavilla alueilla (2 lisää ennen aikaisen lopetuskodonin, 5 muuttaa lähetti-RNAn silmukointikohdan ja 454 proteiinin aminohapon). Jatkoa varten nämä SNP-merkit asetettiin tärkeysjärjestykseen niiden ennustetun vaikutuksen avulla. SIFT-analyysien (VEP, Ensembl) perusteella 93 havaituista muutoksista osoittautui todennäköisesti haitallisiksi (deleterious). Lisäksi jatkoanalyysiin valittiin SNP:t, jotka sijaitsivat evolutiivisesti konservoituneilla alueilla (GERP-elementit) tai miRNA-alueilla. Valitut 384 SNP-merkkiä genotyyppitettiin uudessa otoksessa ayrshirepopulaatiosta (537 nuorempaa sonnia; WPO).

⁹ Lourenco ym., 2014

Uusista SNP-merkeistä 261 osoittautui informatiivisiksi uudessa otoksessa. Niiden assosiaatio utaretulehdusominaisuuksiin tutkittiin yhden lokuksen mixed model analyysillä (EMMAX, SVS7). Yhteensä 29 SNP-merkkiä kromosomeissa 2, 3, 5, 6, 8, ja 18 oli assosioitunut utaretulehdusominaisuuksiin, monet useampaan alaominaisuuteen. Erityisen mielenkiintoiseksi osoittautui kromosomi 6, jossa näyttää olevan samalla alueella useampia utaretulehdusominaisuuksiin vaikuttavia lokuksia. Tätä aluetta analysoitiin tarkemmin monen lokuksen mixed model analyysillä (MLMM, SVS7), jolla sekä utaretulehdusalttiudelle ensimmäisen laktaation alkupuolella (CM11) että somaattiselle soluluvulle ensimmäisessä laktaatiossa (SCS1) löydettiin kaksi todennäköistä vaikuttavaa SNP-merkkiä (Taulukko 3.4). Nämä sijaitsevat oletetuissa kandidaattigeeneissä, joko niiden evolutiivisesti konservoituneilla, siis oletettavasti merkityksellisillä alueilla (GERP) tai ne voivat vaikuttaa geenituotteen laatuun (RNAn rakenne, silmukointi) (Taulukko 3.4).

Taulukko 3.4. Monen lokuksen mixed model analyysissä löytyneet utaretulehdusominaisuuksiin (CM11, utaretulehduksen esiintyminen ensimmäisen laktaation 150 päivän aikana; SCS1, somaattinen soluluku ensimmäisessä laktaatiossa) vaikuttavat SNP-merkit kromosomeissa 6. Taulukon FDR (false discovery rate, väärän löydöksen todennäköisyys) arvot ilmaisevat assosiaation merkitsevyyttä mallilla, jossa kaksi merkitsevintä vaikutusta on mukana.

SNP	BTA6 bp	Vaikutus	Geeni	Ominaisuus	FDR
RS110211379_C_T	88179632	GERP-alue	SLC4A4	CM11	$< 10^{-5}$
RS133405583_A_G	90829396	RNA rakenne	ncRNA	CM11	< 0.25
RS136929700_C_A	87702291	silmukointi	AMBN	SCS1	< 0.0045
RS110291328_G_A	90559772	GERP-alue	IL8	SCS1	< 0.09

3.5 Genotyypitettyjen sonnien ja lehmien arvostelujen yhdistäminen ns. single-step analyysillä kaikkien tuotantotarkkailussa mukana olevien eläinten fenotyyppeihin tuloksiin (WP4)

Aikaisempien tutkimusten perusteella on näytetty, että genomiset arvostelut voidaan toteuttaa koko pohjoismaisessa jalostusarvostelussa mukana oleville 4.6 miljoonalle VG Red lehmälle ja sonnille. Tällöin käytettiin ns. single-step mallia deregressoiduille lehmäarvosteluille (ks. Genominen valinta -hankkeen väliraportti). Seuraavassa vaiheessa menettely tehdään suoraan fenotyypistä esim. koelypsymalli aineistoa käyttäen. Tarkastellaan mahdollisuutta single-step mallin rakentamiseksi niin, että ratkaisut perustuvat HD-aineiston SNP-merkkeihin tai haplotyyppiblokkeihin ilman GBLUP perusoletusta että kaikilla genomisen mallin tekijöillä on sama varianssi. Lopullisessa sovelluksessa tulee olla mahdollisuus käyttää HD-aineistoa tai haplotyyppiblokkeja. Lisäksi kehitetään yksittäisten eläinten genomisen yhdistelmäindeksin (GEBV) luotettavuuden arviointimenetelmä (ns. model based individual animal reliability) ja single-step mallin käytännön toteutus.

Niin sanotussa kaksivaiheisessa genomisessa arvostelussa havainnoista saatava informaatio ja genomisen informaatio yhdistetään ensin käyttäen ns. referenssipopulaatiota. Toisessa vaiheessa genotyypitettyille valintakandidaateille ratkaistaan genomisen jalostusarvo hyödyntämällä joko niiden sukulaisuuksia referenssipopulaatioon, tai referenssipopulaatiosta laskettuja SNP ratkaisuja. Koska kaikkia eläimiä ei ole genotyypitetty, saattaa osa informaatiosta puuttua referenssiaineistosta. Esim. sonnien emien hyvyys/huonous ei tule huomioiduksi genomisessa indeksissä. Tämän vuoksi genomiset indeksit on vielä yhdistetty olemassa oleviin jalostusarvosteluihin erilaisilla ns. blending –

menetelmillä (VanRaden, 2009¹⁰). Pohjoismaisissa jalostusarvosteluissa on vuodesta 2011 käytetty bivariate –blending mallia (Mäntysaari ja Strandén, 2010¹¹). Tätä menetelmää kehitettiin paremmaksi vuonna 2013, koska haluttiin lisätä myös genotyypitettyjen lehmien havainnot referenssipopulaatioon (Taskinen ym., 2013).

Single-step menetelmässä (ssGBLUP) genominen informaatio voidaan sisällyttää suoraan perinteiseen eläinmalli BLUP jalostusarvosteluun. Tällöin genotyypitettyjen eläinten sukulaisuudet perustuvat genotyypitietoihin ja muiden eläinten sukulaisuudet sukupuutietoihin (Aguilar ym. 2010¹², Christensen ja Lund, 2010¹³). Jalostusarvoilla on erilainen sukulaisuusmatriisi eli varianssirakenne (H-sukulaisuusmatriisi) riippuen siitä onko eläin genotyypitetty vai ei. Single-step arvostelussa käytetään suoraan alkuperäisiä havaintoja, jolloin genominen informaatio tulee huomioiduksi myös ympäristövaikutuksia ratkaistaessa ja tuottaa eläimille suoraan genomisen jalostusarvon. Lopulliset jalostusarvot ratkaistaan BLUP menetelmällä sekamalliyhtälöistä. Sekamallin yhtälöissä tarvitaan jalostusarvojen varianssirakenteen käänteismatriisi. Perinteisessä eläinmallissa tarvitaan sukulaisuusmatriisin käänteismatriisi A^{-1} . Genomisessa eläinmallissa käytetään sen sijasta H^{-1} -matriisia, joka voidaan kirjoittaa sukulaisuusmatriisin ja genomisen lisäinformaation summana.

Genomisten sukulaisuusmatriisien laskemista varten on MTT:llä kehitetty HGINV –ohjelma. Genomisissa arvosteluissa tarvittavat ohjelmistot on liitetty osaksi MTT:n omaa MiX99 jalostusarvosteluohjelmistoa (Strandén ja Lidauer 1999¹⁴). MiX99 –ohjelma lukee erikseen tavallisen sukupuutiedoston ja H-matriisin. ssGBLUP arvostelujen keskimääräisen arvosteluvarmuuden validointi tehtiin Interbull GEBV validointitestillä (Mäntysaari ym. 2010¹⁵). Validointi testissä arvioidaan sekä mallin ennustamiskyky ((arvosteluvarmuus)²), että harhattomuus. Harhattomuus esitetään GEBV arvojen regressiokertoimena b_1 niiden kuvatessa tytärkeskiarvoja. Kerroin 1.0 kertoo että 1.0 kg GEBV arvossa ilmenee 1.0 kg tuotosrona jälkeläisissä.

Tutkimuksessa käytettiin pohjoismaista RDC rodun koelypsyaaineistoa heinäkuulta 2013. Koko aineisto käsitti n.82 miljoonaa havaintoa 3.7 miljoonalta lehmältä. Sukupuussa oli kaikkiaan 4.9 miljoonaa eläintä. Genomiaineistossa oli 9107 RDC eläintä, joista 5315 oli sonneja ja 3792 lemiä. Genomiaineistossa oli jokaisella genotyypitettyllä eläimellä 46 914 SNP- markkeria. Tutkimuksessa analysoitiin NAV:n käyttämällä koelypsymallilla (Lidauer ym. 2006¹⁶) sekä (I) koko koelypsyaaineisto että (II) katkaistu aineisto, josta oli poistettu neljän viimeisen vuoden havainnot. Saatujen indeksiennusteiden luotettavuus arvioitiin validointimenetelmällä, jossa arvioidaan markkerivaikutukset vanhemmalla referenssijoukolla (aineisto II) ja ennustetaan nuorempien kandidaattieläimien tuloksia koko aineistossa I.

Tulokset osoittivat, että genomisen eläinmallin käyttö koelypsymallissa oli paitsi mahdollista, se myös tuotti luotettavampia jalostusarvojen ennusteet. Menetelmällä laskettujen genomisten jalostusarvojen luotettavuudet olivat korkeampia, samoin ennusteet olivat harhattomampia (korkeampi regressio-

¹⁰ VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. Journal of dairy science, 91(11), 4414-4423.

¹¹ Mäntysaari, E.A. & Strandén, I. 2010. Use of bivariate EBV-DGV model to combine genomic and conventional breeding value evaluation. In: Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Aug. 1-6, 2010, Leipzig, Germany. 4 p. (CD-ROM file:///D:/assets/pdf/0353.pdf).

¹² Aguilar, I., Misztal, I., Johnson, D.L., Legarra, A. & Tsuruta, S. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743-752.

¹³ Christensen, O.F. & Lund, M.S. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42:2.

¹⁴ Strandén, I. & Lidauer, M. 1999. Solving large mixed models using preconditioned conjugate gradient iteration. J. Dairy Sci. 82: 2779-2787.

¹⁵ Mäntysaari, E.A., Liu, Z. & VanRaden, P. 2010. Interbull validation test for genomic evaluations. Interbull Bulletin. 40: 1-5.

¹⁶ Lidauer, M., Pedersen, J., Pösö, J., Mäntysaari, E.A. Strandén, I., Madsen, P., Nielsen, U.S., Eriksson, J.-Å., Johansson, K. & Aamand, G.P. 2006. Joint Nordic Test Day Model: Evaluation Model. Interbull Open Meeting, Kuopio, Finland, June 4th - 6th, 2006.

kerroin) verrattuna polveutumisindeksiin sekä sonneilla että lehmillä (taulukko 3.5). Aikaisempiin genomisiin jalostusindekseihin verrattuna sonnien ssGBLUP arvosteluilla oli 5 - 18 prosenttiyksikköä korkeammat luotettavuudet (Koivula ym. 2012). Taulukosta 3.5 voi havaita, että genomisella eläinmallilla harhattomuutta kuvaavat b_1 regressiokertoimet olivat 0.97 – 1.15 maitotuotoksella, valkuaisuotoksella 0.84 - 1.01 ja rasvatuotoksella 0.80 – 1.01. Tulokset tukevat ajatusta, että ssGBLUP parantaa genomisten arvostelujen luotettavuutta.

Taulukko 3.5. Regressiomallin regressiokerroin (b_1) sekä luotettavuudet (R^2) ssGBLUP malleista ja polveutumisindeksistä (PA) eri tuotosominaisuuksille a) RDC sonneilla ja b) RDC lehmillä. Vertailun vuoksi taulukossa tulokset aiemmista isämälli DGV:stä sekä isämälli GEBV:stä *)Koivula ym. 2012.

Sonnit	Maito		Valkuainen		Rasva	
	b_1	R^2	b_1	R^2	b_1	R^2
PA	0.91	0.37	0.75	0.28	0.65	0.30
ssGBLUP	0.97	0.49	0.84	0.42	0.80	0.47
DGV*)	0.77	0.30	0.77	0.31	0.86	0.40
GEBV _{SM} *)	0.69	0.32	0.74	0.35	0.90	0.44
Lehmät	Maito		Valkuainen		Rasva	
	b_1	R^2	b_1	R^2	b_1	R^2
PA	1.08	0.28	0.90	0.17	0.92	0.22
ssGBLUP	1.15	0.40	1.01	0.28	1.01	0.33

3.3 Toteutusvaiheen arviointi

Koko genomien kattava genotyyppitys on vielä uutta ja kehittyvää teknologiaa. Tämä vaikeuttaa tutkimuksen suunnittelua, koska eteen tulevia ongelmia, tai mahdollisuuksia, ei voida tietää edeltä käsin. Tässä hankkeessa päädyttiin joustavaan mukautumiseen ja pyrittiin aina kohdistumaan uusiin tarpeisiin ja mahdollisuuksiin. Tämä ei muuttanut varsinaisia tutkimuksen painopisteitä (WP1,...,WP4), mutta antoi painopisteiden sisällä liikkumavaraa. Hankkeessa yhteistyökumppanina olivat paitsi tutkimuslaitokset (lähinnä Aarhus University) myös genomista valintaa toteuttava Viking Genetics ja genomisista arvosteluista vastaava NAV. Genomisen valinnan käyttöönotossa ilmenneet uudet tutkimustarpeet pyrittiin ottamaan huomioon, jopa niin että hankkeen pääpaino osittain siirtyi akateemisesta julkaisusta käytännön vaikuttavuuden varmistamiseen.

3.4 Julkaisut

Katso liite 1.

4. Tulosten arviointi

4.1 Tulosten käytännön sovelluskelpoisuus

Projektin aikana on nähty, että genotyypit määritetään yhä suuremmalle määrälle eläimiä. Genotyypitys tehdään eri tiheyksillä siruilla. Perustiheys on 54K, mutta lehmille käytetään usein alhaisen tiheyden sirua kuten LD+ (8K), ja joillekin sonneille HD sirua tai valikoiduilta määritetään jopa koko sekvenssi. Alhaisemman tiheyden genomitieto imputoidaan suurempaan asteittain, eli 8K imputoidaan 54K tiheyteen, sitten 54K imputoiduilla vahvistettuna HD –tiheyteen jne. Tämän projektin ja muiden tutkimusten mukaan, käytännön jalostusarvosteluissa 54K tiheyden genotyypitys on usein riittävä ja lisähyöty HD sirusta on vähäinen. Tarkempi HD-tiheyden genotyypitystieto hyödyttää GWAS tutkimuksia. Käytännön jalostusarvosteluille ylimääräisestä genomitiedosta voi olla jopa haittaa, koska jalostusarvostelujen laskennasta voi tulla laskennallisesti liian raskas genotyypitettyjen eläinmäärän kasvaessa suureksi.

Genomisen tiedon käyttöön jalostusarvostelussa on useita tapoja, joista käytetyimmät ovat ns. single-step sekä bivariate blending. Tässä hankkeessa käytettiin single-step jalostusarvosteluja ja osoitettiin niiden toteuttamisen olevan mahdollista suurellakin aineistolla. Genotyypitettyjen määrän lisääntyessä single-step jalostusarvostelussa törmätään kuitenkin laskennallisiin ongelmiin. Kun single-step ei ole käytettävissä, kaksivaiheinen bivariate-blending menetelmä on hyvä vaihtoehto. Nykyinen Pohjoismainen lypsykarjan jalostusarvostelu käyttää bivariate-blending menetelmää. Bivariate blending menetelmä perustuu kuitenkin teoreettisiin yksinkertaistuksiin verrattuna single-step menetelmään, jossa kaikki tieto, sekä genomisen että perinteisen, huomioidaan samanaikaisesti.

Kokonaisuudessaan hanke toteutettiin kiinteässä yhteydessä käytännön jalostustoimijoiden kanssa. Hankkeen tulokset raportoitiin Pohjoismaisessa Genomisen valinta –johtoryhmässä, ja kuukausittaisissa tutkimuksen ja toteutuksen alaryhmäkokouksissa. Hankkeen keskeisinä tavoitteina oli löytää genomisen jalostusarvostelumalli, joka toimisi hyvin Pohjoismaisilla lypsykarjaroduilla, sekä tarjota uusia mahdollisuuksia genomisen informaation käyttöön jalostusohjelmassa. Suuri osa tuloksista liittyy genomisen jalostusarvostelun tai genomisen valinnan yksityiskohtien hiomiseen.

Esimerkkinä suuremmista käyttöön otetuista, tai käyttöönottoa odottavista tuloksista:

- 1) NAVn genomiset jalostusarvostelut on ennen julkaisua yhdistetty perinteisiin jalostusarvosteluihin ns. bivariate blending menetelmällä. WP2 tuloksissa esitetty uudistettu bivariate blending mahdollistaa myös lehmien sisällyttämisen referenssipopulaatioon. Käyttöön vuonna 2014.
- 2) MiX99 ohjelmistossa toteutetut genomiset jalostusarvostelumallit ovat käytössä Iso Britanniassa (UK), Irlannissa, Saksassa ja Itävallassa. Meillä pohjoismaisissa on toistaiseksi käytössä GBLUP arvostelu, mutta WP4 kehitetyt single-step arvostelut otettaneen käyttöön 2014/15.
- 3) Informaatiota löytyneistä haitallisista mutaatioista, esimerkiksi kromosomin 12 deleetiosta voidaan käyttää myös yksittäisten paritusten suunnittelussa tilatasolla välttämään kantaja-kantaja yhdistelmiä. Käyttöön 2014.

4.2 Tulosten tieteellinen merkitys

Genomisen jalostusarvostelumallin rakentaminen pienille ja perimältään heterogeenisille populaatioille on tarpeellista myös muiden maiden jalostusohjelmille. Meillä testatut kolme tilastollista malli, eli rotusuudet sisältävä satunnaisregressiomalli, rotukohtaisiin alleelifrekvensseihin perustuva genomisen sukulaisuusmatriisi, ja haplotyyppi blokkeihin perustuva genomisen malli, ovat kaikki herättäneet kiinnostusta myös muissa maissa. Jälkimmäinen malli yhdistää myös kaksi eri lähestymistapaa, koska geenikartoituksissa saatujen assosiaatioanalyysien tuloksia voidaan käyttää tehostamaan genomista jalostusarvostelua (merkitsevimmille SNP-merkeille voidaan antaa arvostelussa suurempi painoarvo). Tärkeänä havaintona voidaan myös pitää, että pelkällä tiheämmällä genotyyppityksellä, eli HD tiheydellä, ei näytä olevan ratkaisevaa hyötyä perimältään heterogeenisten populaatioiden genomisessa arvostelussa.

Assosiaatioanalyysillä saatiin lisää tietoa hedelmällisyyteen ja utaretulehdusalttiuteen vaikuttavista kromosomialueista, ja niissä olevista mahdollisista kandidaatti-SNP:istä. Jatkotutkimuksilla voidaan tutkia näiden kandidaatti-SNP:ien vaikutusta kohdegeenien ekspressioon. Assosiaatiotuloksia voidaan myös jatkossa hyödyntää meta-analyyseissä, joissa yhdistetään tuloksia eri pohjoismaisista tutkimuksista; jolloin odotetaan saatavan entistä tarkempia tuloksia. Jatkotutkimuksissa (uusi pohjoismainen yhteistyöhanke Genomics in Herds) on tarkoitus käyttää sekä HD-genotyyppidataa että koko genomien sekvenssidataa hedelmällisyyteen vaikuttavien letaalimutaatioiden tunnistamiseen ja tämän tiedon hyödyntämiseen jalostusohjelmissa.

Genomisen tiedon käyttö on tähän saakka yhdistänyt perinteisen jalostusarvostelujen tarpeet ja GWAS tutkimusten tarpeet. Jatkossakin molemmat hyötyvät genomisesta tiedosta, mutta aineiston käyttöön liittyvät ongelmat ja siis tutkimustarpeet voivat eriytyä. Nykyisille jalostusarvosteluille riittää genotyyppitys vähäisemmän tiheyden sirulla (esim. 54K), kun taas GWAS tutkimukset hyötyvät tiheämmän sirun ja sekvenssitiedon käytöstä. Genomisen tiedon käyttö jalostuksessa tulee jatkossa painottumaan yhä enemmän suuremman eläinmäärän genotyyppitykseen ja siitä seuraavan kaiken suuren genotyyppitietomäärän käytöstä. Tärkeätä on ennustaa jalostusarvoa mahdollisimman hyvin, jolloin tilastollisten menetelmien ei tarvitse ole tarkkoja yksittäisten geenien tarkasta mallintamisesta. GWAS hyötyy genotyyppitettyjen määrän kasvusta, jolloin kartoituksessa käytetyt tilastolliset menetelmät tulevat tarkemmiksi. Mutta toisin kuin jalostusarvosteluissa, imputaatiomenetelmien merkitys korostuu kun suuri osa genotyyppitetystä tiedosta perustuu imputoituihin tietoihin eikä havaittuihin genotyypeihin. Genotyyppitystiedon määrän lisääntyessä tulee yhä tärkeämmäksi pystyä karsimaan tarpeetonta genomista tietoa ja löytää merkityksellisiä genomialueita, jotta käytännön jalostusarvojen laskenta olisi toteutettavissa nopeasti.

5. Hankeosapuolet ja yhteistyö

Tutkimus tehtiin MTT:n Biotekniikka- ja elintarviketutkimuksen geneettisen tutkimuksen ryhmässä. Työpakettien 1, 2 ja 4 tutkimuksesta vastasi lähinnä Biometrisen genetiikan tiimi (BGE), ja työpaketista 3 Eläingenomiikan tiimi (EGE). Alla olevassa luettelossa on tarkemmin määritelty eri tutkijoiden vastualueet:

Professori Esa Mäntysaari, BGE, vastuullinen johtaja genomiset jalostusarvostelut

Professori Johanna Vilkki, EGE, yhteyshenkilö, vastuullinen johtaja genotyyppitykset

Vanhempi tutkija MMT Nina Schulman, EGE, assosiaatioanalyysit

Tutkija FM Terhi Iso-Touru, EGE, SNP-datan käsittely, assosiaatioanalyysit ja haplotyyppianalyysit

Tutkija FT Miika Tapio, GDI, SNP-datan käsittely, imputointi, haplotyyppianalyysit

Erikoistutkija PhD Ismo Strandén, BGE, genomiset arvostelumallit, bayesian analyysit.

Tutkija FT Minna Koivula, BGE, single-step analyysit

Tutkija FT Matti Taskinen, BGE, bi-variate blending

Tutkija FM Timo Knürr, BGE, haplotyyppi avusteinen genominen arvostelu

Jatko-opiskelija Mahlako. Makgahlela, väitteli 7.2.2014, BGE, genomiset jalostusarvot

MTTn ulkopuolisia yhteistyökumppaneita ovat olleet *University of Aarhus, Department of Genetics and Biotechnology, Tanska*, jossa tutkimuksesta on vastannut Dept. Head *Mogens Sandø Lund*, sekä *Pohjoismainen jalostusarvosteluyhtiö NAV Tanska*, jossa yhteisestä koordinoinnista on vastannut *MD Dr. Gert Aamand*.

NAV on myös antanut hankkeen käyttöön eläinarvosteluissa käytettävän fenotyyppisen aineiston.

Hankkeessa käytetty genotyyppiaineisto on pääosin Aarhus Universityn; Viking Genetics'in; NAVn; *Faba, Suomi*; ja *Växa, Ruotsi*, perustaman "Genotyypitys" –projektin omaisuutta. *Viking Genetics*'in yhteyshenkilö on ollut *tutkimusjohtaja Søren Borchersen*, Tanska sekä Rotukoordinaattori *Auli Himanen*, Suomi. *Faba osk* yhteyshenkilö on ollut kehityspäällikkö *Minna Toivonen*.

Teräväpiirto genomitiedon hyödyntäminen jalostusarvosteluissa *HD Genomiikka*

High Definition genome wide information in genetic evaluations

Vastuuorganisaatio	Maa- ja elintarviketalouden tutkimuskeskus MTT Biotekniikka- ja elintarviketutkimus prof. Esa Mäntysaari ja prof. Johanna Vilkki Alimentum, Myllytie 1 31600 Jokioinen puh. 040-5106013
Kesto	2012 – 2013 (Loppuraportti 1.4.2014)
Rahoitus	Kokonaiskustannukset 814 920 euroa MMM:ltä saatu kokonaisrahoitus 150 000 euroa Hankkeen toteuttajan oma rahoitus 522 470 euroa Muista julkisista lähteistä saatu rahoitus 0 euroa Muu rahoitus 142 450 euroa
Avainsanat	Genominen valinta, Genominen jalostusarvostelu, SNP, lypsykarja

Tiivistelmä

TAVOITTEET

Tutkimushankkeen tavoitteena oli kehittää uusia, koko genomien tuntemukseen perustuvia menetelmiä Suomalaisen ayrshiren sekä yleisesti Pohjoismaisen punaisen lypsykarjan geenien kartoitukseen ja jalostusarvostelujen ja jalostusohjelmien kehittämiseen. Osatavoitteet olivat: i) Kehittää uusi genominen malli, joka mahdollistaa rotualkuperältään heterogeenisen aineiston luotettavan analysoinnin; ii) Parantaa laskennallisia menetelmiä niin että HD-aineistoa voidaan hyödyntää tehokkaimmin; iii) Hyödyntää HD-aineistoa assosiaatioanalyseissä, ja tehokartoittaa monitekijäisiin ominaisuuksiin vaikuttavia geenejä; iv) Yhdistää genotyyppitettyjen sonnien ja lehmien arvostelut ns. single-step analyysillä kaikkien tuotannon tarkkailussa mukana olevien eläinten fenotyyppiin tuloksiin.

TULOKSET

Tutkimuksissa käytetyt genomiset aineistot hankittiin yhteistyössä pohjoismaisten jalostusjärjestöjen ja tutkimuslaitosten sekä jalostusarvostelu-yhtiö NAVin kanssa. Tutkimuksissa käytettiin 9107 pohjoismaisen punaisen rodun (RDC) eläimen genotyyppityksiä Illuminan Bovine54k genotyyppitysalustalla (54 000 SNP merkkiä), 723 RDC eläimen genotyyppityksiä BovineHD (770k SNP merkkiä) ja 18 Suomen RDC jalostusohjelmalle keskeisen sonnien koko sekvenssiä. Enimmillään fenotyyppi-aineistona käytettiin 3.7 miljoonan RDC lehmän koelypsyhavaintoja.

Testasimme kolme vaihtoehtoista genomista mallia: Rotukohtainen satunnaisregressiomalli; rotukohtaisiin alleelifrekvensseihin perustuvaan genomiseen sukulaisuusmatriisiin perustuva BLUP malli ja useita SNP merkkejä sisältäviä kromosomisegmenttejä hyödyntävä haplotyyppimalli. Näistä haplotyyppimalli osoittautui lupaavaksi ja sen jatkokehitystä jatketaan.

HD aineistolla saatiin vähän parempi genomisten arvostelujen luotettavuus kuin perus merkkitiheydellä, mutta saavutettu arvosteluvarmuus ei ollut silti lähellä muilla roduilla saavutettavia. Toisaalta assosiaatio-analyyseissä HD tiheydellä pystytään paikallistamaan enemmän ja tarkemmin ominaisuuksiin vaikuttavia genomialueita. Löydettiin mm. hedelmällisyyteen voimakkaasti vaikuttava alue kromosomista 12, ja käyttäen kokosekvenssidataa alueelta löydettiin pitkä noin 600 kiloemäksen deletio. Sen yleisyys ayrshire-populaatiossa oli 32%. Deletio ilmentyy kuten resessiivinen letaalitekijä, eli paritettaessa kaksi deletion kantajaa, neljäsosa alkioista kuolee varhaisessa vaiheessa.

Nyt käytössä olevaa vain sonnien genotyypityksiä hyödyntävää GBLUP mallia verrattiin ns. single-step genomiarvosteluun, jossa ei-genotyypitetyillä eläimillä ovat normaalit sukupuuhun perustuvat sukulaisuudet, ja genotyypitetyillä eläimillä genotyypityksiin perustuvat genomiset sukulaisuudet. Single-step arvosteluilla saavutettiin systemaattisesti parempi luotettavuus kuin GBLUP mallilla. Tulosten perusteella single-step mallin käyttöön siirtyminen kannattaa tehdä niin pian kuin mahdollista. Single-step arvostelujen ongelma on lievä yliennustaminen, joka johtaa siihen että genomisilla indekseillä sonnien välinen vaihtelu on 3-20% suurempaa kuin näkyy tyttäristä laskettuna.

TULOSTEN ARVIOINTI

Hanke toteutettiin läheisessä yhteistyössä käytännön jalostustoimijoiden kanssa ja monet hankkeen tulokset on jo otettu käyttöön pohjoismaisessa jalostusohjelmassa, tai tullaan ottamaan hankkeen päätyttyä. Erityisenä vahvuutena voidaan pitää yhteistyöverkoston syntymistä sekä pohjoismaisten tutkimuslaitosten välillä että kansainvälisesti laajemminkin. Julkaisutuotannoltaan hanke oli hyvä, joskin osa tehdystä työstä on vielä kirjoitusvaiheessa.

Vaikka tuloksia saavutettiin tuli yhä selvemäksi, että genomisen jalostusohjelman tehokkaaksi toteuttamiseksi on tehtävä vielä paljon lisäkehitystä. Samaan aikaan kun tässä hankkeessa tarkasteltiin HD-genotyyppien hyödyntämistä, on käytännön jalostusohjelmissa yhä enemmän siirrytty käyttämään matalan tiheyden ns. LD -genotyypityksiä.

Julkaisut

Vertaisarvioituja 10 kpl, muita tieteellisiä ja kongressijulkaisuja 17 kpl.

LIITE 2 Julkaisut

Asiantuntijatarkastetut julkaisut:

- KADRI NK, SAHANA G, CHARLIER C, ISO-TOURU T, GULDBRANDTSEN B, KARIM L, NIELSEN US, PANITZ F, AAMAND GP, SCHULMAN N, GEORGES M, VILKKI J, LUND MS, DRUET T. 2014. A 660-Kb Deletion with Antagonistic Effects on Fertility and Milk Production Segregates at High Frequency in Nordic Red Cattle: Additional Evidence for the Common Occurrence of Balancing Selection in Livestock. *PLoS Genet* 10(1): e1004049. doi:10.1371/journal.pgen.1004049
- KNÜRR, T., STRANDÉN, I., KOIVULA, M., AAMAND, G.P., MÄNTYSAARI E. A. 2014. Haplotype-assisted genomic evaluations in Nordic Red dairy cattle. (to be submitted).
- KOIVULA, M., STRANDÉN, I, SU, G., MÄNTYSAARI, E A. 2012. Different methods to calculate genomic predictions : comparisons of BLUP at the single nucleotide polymorphism level (SNP-BLUP), BLUP at the individual level (G-BLUP), and the one-step approach (H-BLUP). *Journal of Dairy Science* 95 7: 4065-4073. [doi]
- KOIVULA, M., STRANDÉN, I, PÖSÖ J., AAMAND, G.P., MÄNTYSAARI, E. A 2014. Single step genomic evaluation using multi-trait random regression models and test-day data (to be submitted).
- MAKGAHLELA, M. LINAH, MÄNTYSAARI, E A., STRANDÉN, I, KOIVULA, M., NIELSEN, U. S., SILLANPÄÄ, M. J., JUGA, J. 2013a. Across breed multi-trait random regression genomic predictions in the Nordic Red dairy cattle. *Journal of Animal Breeding and Genetics* 130 1: 10-19. [doi]
- MAKGAHLELA, M. L., STRANDÉN, I, NIELSEN, U. S., SILLANPÄÄ, M. J., MÄNTYSAARI, E. A. 2013b. The estimation of genomic relationships using breedwise allele frequencies among animals in multibreed populations. *Journal of Dairy Science* 96 8: 5364-5375. [doi]
- MAKGAHLELA, M. L., STRANDÉN, I, NIELSEN, U. S., SILLANPÄÄ, M. J., MÄNTYSAARI, E. A. 2014c. Using the unified relationship matrix adjusted by breed-wise allele frequencies in genomic evaluation of a multibreed population. *Journal of Dairy Science* 97 2: 1117-1127. [doi]
- RIUS-VILARRASA, E., BRØNDUM, R.F., STRANDÉN, I., GULDBRANDTSEN, B., STRANDBERG, E., LUND, M.S., FIKSE, W.F. 2012. Influence of model specifications on the reliabilities of genomic prediction in a Swedish-Finnish Red Breed Cattle population. *Journal of Animal Breeding and Genetics*, 129:369-379.
- STRANDÉN, I, MÄNTYSAARI, E. 2013. Use of random regression model as an alternative for multibreed relationship matrix. *Journal of Animal Breeding and Genetics* 130 1: 4-9. [doi]
- SU, G., MADSEN, P., NIELSEN, U. S., MÄNTYSAARI, E A., AAMAND, G. P., CHRISTENSEN, O. F., LUND, M. S. 2012. Genomic prediction for Nordic Red Cattle using one-step and selection index blending. *Journal of Dairy Science* 95 2: 909-917. [doi]

Muut tieteelliset julkaisut:

- GULDBRANDTSEN B., G.SU., AAMAND, G.P., STRANDÉN, I, LUND, M.S. 2013. Should the markers on X chromosome be used for genomic prediction? In: Proceedings of the 2013 Interbull meeting, Nantes, France, August 23-25, 2013. *Interbull Bulletin* 47: 125-130. [url]
- JUGA, J., SILLANPÄÄ, M. J., MÄNTYSAARI, E. 2012. Lypsykarjan genominen valinta. In: *Maailma muuttuu : muuttuuko maatalous? / Mervi Seppänen (toim.).* p. 165-172.

- JUGA, J., SILLANPÄÄ, M., MÄNTYSAARI, E. 2012. Genomisk selektion av mjölkboskap. In: *Världen förändras : förändras jordbruket?* / Mervi Seppänen (red.). p. 165-172.
- KOIVULA, M., MÄNTYSAARI, E, STRANDÉN, I. 2012. SNP-BLUP, G-BLUP ja H-BLUP : johdanto genomisiin arvosteluihin. In: *Maataloustieteen Päivät 2012, 10.-11.1.2012 Viikki, Helsinki : esitelmä- ja posteritiivistelmät* / Toim. Nina Schulman. Suomen maataloustieteellisen seuran tiedote 29: p. 27. [url]
- KOIVULA, M., MÄNTYSAARI, E, STRANDÉN, I. 2012. SNP-BLUP, G-BLUP ja H-BLUP : johdanto genomisiin arvosteluihin. In: *Maataloustieteen Päivät 2012, 10.-11.1.2012 Viikki, Helsinki : esitelmät, posterit* / Toim. Nina Schulman ja Heini Kauppinen. Suomen maataloustieteellisen seuran tiedote 28: 5 p. [url]
- KNÜRR, T., STRANDÉN, I., KOIVULA, M., AAMAND, G. P., MÄNTYSAARI, E. A. 2013. Haplotype-assisted genomic evaluations in Nordic Red dairy cattle. In: *Book of abstracts of the 64th annual Meeting of the European federation of animal science, Nantes, France, 26-30 August 2013.* EAAP book of abstracts 19: 454. [doi]
- KOIVULA, M., STRANDÉN, I, PÖSÖ, J, PEDERSEN J., AAMAND, G., MÄNTYSAARI, E A. 2012. Single step genomic evaluations for the Nordic Red Dairy Cattle test day data. In: *Proceedings of the 2012 Interbull meeting, Cork, Ireland, May 28-31, 2012.* Interbull Bulletin 46: 115-120. [url]
- MAKGAHLELA, M., STRANDÉN, I, NIELSEN, U.S., SILLANPÄÄ, M.J., MÄNTYSAARI, E A. 2013. The adjusted genomic relationships by allele frequencies within breeds and use in single-step GBLUP. In: *Book of abstracts of the 64th annual meeting of the European Federation of Animal Science, Nantes, France, 26-30 August 2013.* EAAP Book of abstracts 19: 237. [doi]
- MAKGAHLELA, M., KNÜRR, T., AAMAND, G.P., STRANDÉN, I, MÄNTYSAARI, E A. 2013. Single step evaluations using haplotype segments. In: *Proceedings of the 2013 Interbull meeting, Nantes, France, August 23-25 2013.* Interbull Bulletin 47: 217-221. [url]
- MAKGAHLELA, M. L., MÄNTYSAARI, E A., STRANDÉN, I, KOIVULA, M., NIELSEN, U. S., SILLANPÄÄ, M. J., JUGA, J. 2012. Investigation of reliability of genomic predictions in the admixed Nordic Red dairy cattle. In: *Maataloustieteen Päivät 2012, 10.-11.1.2012 Viikki, Helsinki : esitelmä- ja posteritiivistelmät* / Toim. Nina Schulman. Suomen maataloustieteellisen seuran tiedote 29: p. 25. [url]
- MAKGAHLELA, M. L., MÄNTYSAARI, E A., STRANDÉN, I, KOIVULA, M., NIELSEN, U. S., SILLANPÄÄ, M. J., JUGA, J. 2012. Investigation of reliability of genomic predictions in the admixed Nordic Red dairy cattle. In: *Maataloustieteen Päivät 2012, 10.-11.1.2012 Viikki, Helsinki : esitelmät, posterit* / Toim. Nina Schulman ja Heini Kauppinen. Suomen maataloustieteellisen seuran tiedote 28: 5 p. [url]
- MAKGAHLELA, M. L., STRANDÉN, I, NIELSEN, U. S., SILLANPÄÄ, M., J., JUGA, J., MÄNTYSAARI, E A. 2012. Comparative assessment of methods for estimating genomic relationships and their use in predictions in an admixed population. In: *Proceedings of the 2012 Interbull meeting, Cork, Ireland, May 28-31, 2012.* Interbull Bulletin 46: 92-96. [url].
- MAKGAHLELA, M. L. 2014. Accounting for population admixture in genomic evaluations. Ph.D. Thesis, Helsinki University. Dept of Agricultural Sciences, Publication 30. Helsinki.
- MÄNTYSAARI, E A., KOIVULA, M.. 2012. GEBV validation test revisited. In: *Proceedings of the 2012 Interbull technical workshop, Verona, Italy, February 2-3, 2012.* Interbull Bulletin 45: 5 p. [url]
- TASKINEN, M., MÄNTYSAARI, E. A., LIDAUER, M., KNÜRR, T., PÖSÖ, J., SU, G., AAMAND, G.P., STRANDÉN, I. 2013. Comparison of model reliabilities from single-step and bivariate

blending methods. In: Proceedings of the 2013 Interbull meeting, August 23-25, Nantes, France, 2013. Interbull Bulletin 47: 246-251. [url]

VILKKI, J., ISO-TOURU, T., SCHULMAN, N.F., DOLEZAL, M.A., BAGNATO, A., SOLLER, M., PANITZ, F., HOLM, L-E. 2013. [Revisiting QTL affecting clinical mastitis by high-density GWAS and resequencing in the Finnish Ayrshire dairy cattle.](#) International Plant & Animal Genome XXI Conference, San Diego, CA, January 12-16, 2013.

VILKKI, J., DOLEZAL, M., SAHANA, G., BAGNATO, A., TAFER, H., ISO_TOURU, T., PANITZ, F., SANTUS, E., SOLLER, M. 2013 [Validation of QTL affecting mastitis in dairy cattle.](#) In: Book of abstracts of the 64th annual meeting of the European Federation of Animal Science, Nantes, France, 26-30 August, 2013. EAAP book of abstracts 19: 538.