

Pohjoismaisten punaisten rotujen hedelmällisyysarvostelu genomisella eläinmallilla

Kaarina Matilainen¹⁾, Ismo Strandén¹⁾, Anna-Maria Tyrisevä¹⁾ ja Esa Mäntysaari¹⁾

¹⁾*Luomnonvarakeskus (Luke), Vihreä teknologia, Myllytie 1, 31600 Jokioinen, kaarina.matilainen@luke.fi*

TIIVISTELMÄ

Pohjoismaista lypsykarjan hedelmällisyysominaisuuksien jalostusarvostelua ollaan päivittämässä. Koska yhä useammasta eläimestä on genominen tieto, uutta mallia halutaan testata myös niin että genominen tieto on yhdistettynä arvosteluihin. Yleinen menetelmä genomisen ja perinteisen tiedon yhdistämiseen on ns. genominen eläinmalli eli single-step –menetelmä. Menetelmässä voidaan käyttää suoraan samoja havaintoja kuin perinteisessä arvostelussakin, mutta eläinten sukulaisuudet perustuvat sekä genotyyppi- että sukupuutietoihin. Tässä tutkimuksessa hedelmällisyyden genominen eläinravostelu toteutettiin pohjoismaiselle punaiselle lypsyrodulle. Arvostelussa käytetyt mallin tekijät ja varianssiparametrit olivat samat kuin perinteisellä eläinmallilla. Malli sisälsi yksitoista toisistaan riippuvaa ominaisuutta. Eri lypsykausien havainnot käsiteltiin eri ominaisuuksina. Uusimattomuusprosentti ja aika ensimmäisestä siemennyksestä viimeiseen siemennykseen määriteltiin sekä hiehoille että kaikille kolmelle ensimmäiselle lypsykaudelle erikseen. Aika poikimisesta ensimmäiseen siemennykseen mitataan vain lehmillä. Kaikilla ominaisuuksilla oli matala periytymisaste (0,01-0,04) ja lehmien ominaisuuksien välillä oli korkeat geneettiset korrelaatiot (0,60-0,88 lypsykausien välillä). Aineisto sisälsi havaintoja neljältä miljoonalta eläimeltä ja sukupuutiedot olivat yli viidelle miljoonalle eläimelle. Arvostelussa mukana olevista eläimistä genomitieto oli saatavilla 23520 eläimeltä. Genominen eläinmalliarvostelu, samoin kuin alkuperäinen perinteinen eläinmalliarvostelukin, tehtiin MiX99-ohjelmalla. Genominen eläinmalliarvostelu osoittautui hitaasti suppenevaksi. Tulosten ratkaisemiseen tarvittiin 7 päivää ja 29063 iteraatiota, kun perinteisen eläinmalliarvostelun tulokset ratkesivat viiden tunnin ja 3052 iteraation jälkeen. Genomisessa eläinmallissa kahdesta eri lähteestä olevien tietojen yhteensopivuus eläinten geneettisen samankaltaisuuden määrittämisessä on tärkeää. Siksi genomista arvostelua muutettiin niin, että geneettiset ryhmät otettiin huomioon sekä sukupuuhun että genotyyppitietoihin perustuvassa informaatiossa. Tämän jälkeen ratkaisut saatiin kolmentoista tunnin ja 2453 iteraation jälkeen. Genomisen ja perinteisen arvostelun geneettiset trendit seurasivat hyvin toisiaan ja syntymävuosittain lasketut jalostusarvojen korrelaatiot olivat pääasiassa yli 0,97. Ensimmäiset ajot hedelmällisyysominaisuuksien genomiselle arvostelulle osoittivat, että siirtyminen eläinmallista täyteen genomiseen eläinmalliin on toteutettavissa.

ASIASANAT

eläinjalostus, genominen eläinmalli, hedelmällisyys

Johdanto

Lisääntymiskyky on välttämätön ja tärkeä taloudellinen tekijä maidontuotannossa. Tuotos- ja hedelmällisyysominaisuuksien välillä esiintyvä epäsuotuisa geneettinen yhteys on kuitenkin johtanut monissa maissa lehmien hedelmällisyyden heikkenemiseen. Tätä on lähdetty torjumaan laskemalla jalostusarvostelut myös lisääntymisominaisuuksille. Yhteispohjoinen jalostusarvostelu on perustunut monenominaisuuden malliin, jossa viisi ominaisuutta arvostellaan kahdessa ryhmässä. Nyt pohjoismaista lypsykarjan jalostusarvostelua hedelmällisyysominaisuuksille on päivitetty (Tyrisevä ym., 2016). Uusi jalostusarvostelu perustuu eläinmalliin, jossa lehmän hedelmällisyys arvostellaan eri ominaisuuksina kullakin lypsykaudella. Koska genotyyppitettyjen eläinten määrä kasvaa nykypäivänä nopeasti ja sen perusteella halutaan tehdä valintapäätöksiä, kaikki mahdollinen fenotyyppinen ja genotyyppinen tieto on tärkeä sisällyttää jalostusarvosteluihin. Myös hedelmällisyysominaisuuksien uutta mallia halutaan testata niin että genominen tieto on mukana arvostelussa.

Aguilar ym. (2010) sekä Christensen & Lund (2010) ovat esittäneet yleisen menetelmän genomisen ja perinteisen tiedon yhdistämiseen. Sitä kutsutaan single-step –menetelmäksi eli ns. genomiseksi eläinmalliksi. Menetelmässä voidaan käyttää suoraan samoja havaintoja kuin perinteisessä arvostelussakin, mutta eläinten geneettisen samankaltaisuuden määrittäminen perustuu sekä genotyyppi- että sukupuutietoihin. Menetelmän etuja ovat mm. genotyyppiin perustuvan valinnan huomioon ottaminen jalostusarvostelussa sekä korkeampi validaatioarvosteluvarmuus tavalliseen eläinmalliin verrattuna (esim. Tsuruta ym., 2011). Genomista eläinmallia on jo testattu pohjoismaisen lypsykarjan tuotosominaisuuksille lupaavin tulokset (Koivula ym., 2015). Tämän tutkimuksen tavoitteena oli laskea genomiset eläinmallin jalostusarvot hedelmällisyysominaisuuksille. Jalostusarvostelu toteutettiin pohjoismaiselle punaiselle lypsyrodulle.

Aineisto ja menetelmät

Aineistona oli Faban virallisissa pohjoismaisissa hedelmällisyysarvosteluissa vuonna 2014 käytetty tieto. Aineisto sisälsi havaintoja 4046233 eläimeltä ja sukupuutiedot olivat 5257722 eläimelle. Arvostelussa mukana olevista eläimistä genomitieto oli saatavilla 23520 eläimelle. Markkerien määrä oli 46914.

Arvostelussa käytettävä malli ja parametrit olivat samat kuin perinteisellä eläinmallilla (Tyrisevä ym., 2016). Malli sisälsi yksitoista toisistaan riippuvaa ominaisuutta. Eri lypsykausien havainnot käsiteltiin eri ominaisuuksina. Uusimattomuusprosentti (nrr) ja aika ensimmäisestä siemennyksestä viimeiseen siemennykseen (ifl) määriteltiin sekä hiehoille, että kaikille kolmelle ensimmäiselle lypsykaudelle (nrr0-nrr3 ja ifl0-ifl3). Lepokausi, eli aika poikimisesta ensimmäiseen siemennykseen (icf) mitataan vain lehmiltä (icf1-icf3). Kaikkiaan aineistossa oli 27591628 havaintoa yhdelletoista ominaisuudelle.

Ominaisuuksiin vaikuttavia tekijöitä olivat: yleinen regressiotekijä, kolme kiinteää tekijää sekä satunnainen additiivinen geneettinen eläintekijä. Yleinen regressiotekijä kuvasi kokonaisheteroosivaikutusta. Kiinteät tekijät olivat ominaisuudesta riippuen joko karja×syntymävuosi- tai karja×ensimmäinen poikimavuosi-yhdysvaikutus, siemennysvuosi×kuukausi- tai poikimavuosi×kuukausi-yhdysvaikutus, sekä ikä ensimmäisen siemennyksen aikaan. Sukupuu sisälsi myös 315 geneettistä ryhmää, jotka mallinnettiin satunnaisina.

Perinteisellä eläinmallilla laskettuja geneettisiä parametrejä käytettiin myös genomisessa eläinmallissa. Kaikilla ominaisuuksilla oli matala periytymisaste. Alhaisin periytymisaste (0,015) oli kaikilla hiehojen ominaisuuksilla sekä lehmien uusimattomuusprosentilla. Lähes yhtä matalat periytymisasteet (0,03 ja 0,04) olivat lehmille määritetyillä ominaisuuksilla aika ensimmäisestä viimeiseen siemennykseen ja aika poikimisesta ensimmäiseen siemennykseen. Lisäksi lehmien ominaisuuksien väliset geneettiset yhteydet olivat korkeat (0,60-0,88 lypsykausien välillä).

Genomisessa eläinmallissa tarvitaan sukupuuhun perustuva sukulaisuusmatriisi A_{22} genotyyppitetuille eläimille. Sen jälkeen laskettiin genomitietojen perusteella genomisen sukulaisuusmatriisi G perustuen VanRadenin (2008) esittämään menetelmään 1:

$$G = \frac{ZZ'}{2 \sum p_i(1 - p_i)}$$

missä Z on markkereittain keskitetty genotyyppimatriisi ja p_i on alleelifrekvenssi lokuksessa i .

Yhdistämällä kahden sukulaisuusmatriisin \mathbf{A}_{22} ja \mathbf{G} tiedot saadaan genomisessa eläinmalliarvostelussa käytetty matriisiin \mathbf{H} käänteismatriisi (Misztal ym., 2009):

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_w^{-1} - \mathbf{A}_{22}^{-1} \end{bmatrix},$$

missä \mathbf{A}^{-1} on sukupuuhun perustuvan sukulaisuusmatriisin käänteismatriisi ja $\mathbf{G}_w = (1 - w)\mathbf{G} + w\mathbf{A}_{22}$. Tässä tutkimuksessa polygeenisen informaation painoksi w asetettiin 10%, eli genomisella tiedolla oletettiin selitettävän 90% additiivisesta geneettisestä vaihtelusta. Tämän jälkeen laskettiin MiX99-ohjelmalla (MiX99 Development Team, 2015) genominen eläinmalliarvostelu käyttäen matriisia \mathbf{H}^{-1} , samoin kuin alkuperäinen perinteinen eläinmalliarvostelukin käyttäen matriisia \mathbf{A}^{-1} .

Geneettisten ryhmien mukanaolo mallissa saattaa aiheuttaa ongelmia genomisen eläinmallin ratkaisemisessa, mikäli käytetään yllä esitettyä käänteismatriisia \mathbf{H}^{-1} (Misztal ym., 2013). Tämä johtuu siitä, että geneettiset ryhmät on otettu huomioon sukupuuhun perustuvassa informaatiomatriisissa \mathbf{A}^{-1} ns. QT-muunnoksen avulla (Quaas ja Pollak, 1981), eli:

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}^{*-1} & -\mathbf{A}^{*-1}\mathbf{Q} \\ -\mathbf{Q}'\mathbf{A}^{*-1} & \mathbf{Q}'\mathbf{A}^{*-1}\mathbf{Q} \end{bmatrix},$$

missä \mathbf{A}^{*-1} on sukulaisuusmatriisin käänteismatriisi kaikille eläimille ja matriisi \mathbf{Q} sisältää geneettisten ryhmien kontribuutiot kullekin eläimelle. Geneettisten ryhmien kontribuutioita ei kuitenkaan ole otettu huomioon genomiseen tietoon perustuvassa matriisissa. Ratkaisuna tähän Misztal ym. (2013) esitti geneettisten ryhmien kontribuutioiden lisäämistä seuraavasti. Määritellään $\mathbf{B} = [\mathbf{G}_w^{-1} - \mathbf{A}_{22}^{-1}]$. Tällöin genomisessa arvostelussa käytetty uusi käänteismatriisi on:

$$\mathbf{H}^{-1} = \begin{bmatrix} \mathbf{A}^{*-1} & -\mathbf{A}^{*-1}\mathbf{Q} \\ -\mathbf{Q}'\mathbf{A}^{*-1} & \mathbf{Q}'\mathbf{A}^{*-1}\mathbf{Q} \end{bmatrix} + \begin{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} & -\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \mathbf{Q} \\ -\mathbf{Q}' \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} & \mathbf{Q}' \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \mathbf{Q} \end{bmatrix}.$$

Tätä kutsutaan jatkossa genomisen arvostelun QT-muunnokseksi. Analyysseissä käytetty MiX99-ohjelmisto muodostaa sukulaisuusmatriisin käänteismatriisin \mathbf{A}^{-1} itse, mutta lukee harvamatriisimuodossa tiedostoon talletetun matriisin \mathbf{B} . Geneettisten ryhmien kontribuutioiden lisääminen genomiseen informaatioon ei siis aiheuttanut muutoksia MiX99-ohjelmistoon. Sen sijaan geneettisten ryhmien kontribuutioiden kertomisesta matriisiin \mathbf{B} kanssa johtuvia nollasta poikkeavia lukuja lisättiin alkuperäiseen matriisiin \mathbf{B} sisältävään tiedostoon.

MiX99-ohjelmisto on kehitetty suurille aineistoille ja sen ratkaisualgoritmeilla pystytään laskemaan jalostusarvostelumallin ratkaisut muodostamatta varsinaisesti jalostusarvostelujen sekamalliryhmää. Ratkaisut saadaan iteratiivisesti lukien sekä havaintoaineistoa, arvostelumallin tietoja että eläinten sukupuutiedostoa ja muodostamalla yhtälöryhmät uudelleen joka iteraatiokierroksella. Kierroksia toistetaan, kunnes ratkaisut eivät kierrokselta seuraavalle siirryttäessä enää muutu. Tätä kutsutaan ratkaisujen suppenemiseksi. MiX99-ohjelmisto käyttää tehokasta pohjustettua liittogradienttimenetelmää (Preconditioned Conjugate Gradient, PCG) sekamalliyhtälöryhmän ratkaisemiseksi. PCG-menetelmän oletettiin päätyneen lopullisiin ratkaisuihin kun peräkkäisten PCG-kierrosten ratkaisujen välinen suhteellinen ero oli pienempi kuin 10^{-10} . Samaa kriteeriä käytettiin myös genomiselle eläinmalliarvostelulle.

Tulokset ja tulosten tarkastelu

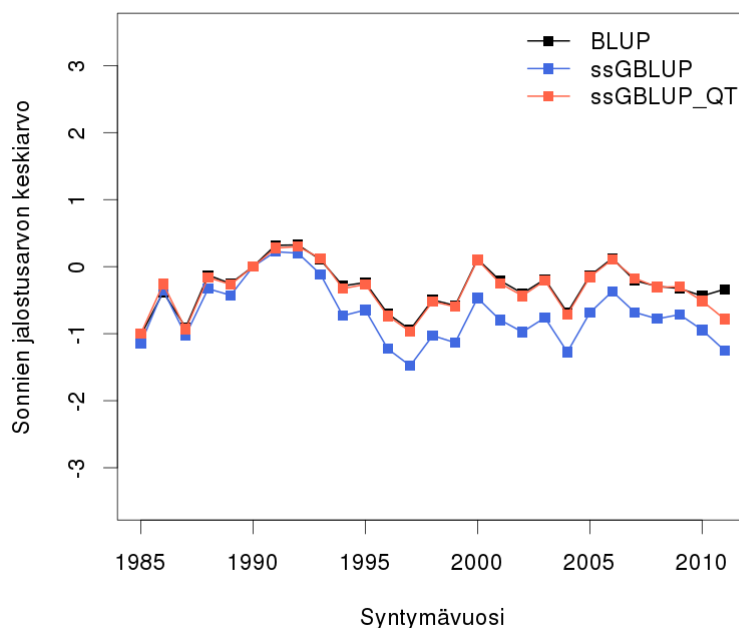
Genomisen eläinmalliarvostelun ratkaisut osoittautuivat hitaasti suppeneviksi. Käytettyyn ratkaisukriteeriin pääseminen vaati 7 päivää ja 29063 PCG-kierrosta, kun perinteisen eläinmalliarvostelun tulokset ratkesivat viiden tunnin ja 3052 PCG-kierroksen jälkeen. Huomattavaa genomisen eläinmallin ratkaisemisessa oli PCG-menetelmän epätasainen suppeneminen, mikä herättää epäilyksiä ratkaistujen jalostusarvojen luotettavuudesta. Nämä erot suppenemisessä voivat aiheutua eroista geneettisten ryhmien huomioon ottamisessa sukulaisuusmatriisissa \mathbf{A} ja genomisessa sukulaisuusmatriisissa \mathbf{G} , jotka muodostavat matriisin \mathbf{H} . Kun genomisessa arvostelussa käytettyä matriisia \mathbf{H}^{-1} päivitettiin ottamaan huomioon geneettiset ryhmät, PCG-menetelmän suppeneminen parani selvästi. Ratkaisut saatiin 2453 PCG-kierroksen jälkeen ja epämääräisiä heilahduksia

suppenemisessa ei enää näkynyt. Vaikka suppenemista saatiin huomattavasti nopeutettua PCG-kierrosten määrän perusteella, laskenta-aika kolmetoista tuntia oli silti huomattavasti pidempi kuin tavallisen eläinmallin viisi tuntia.

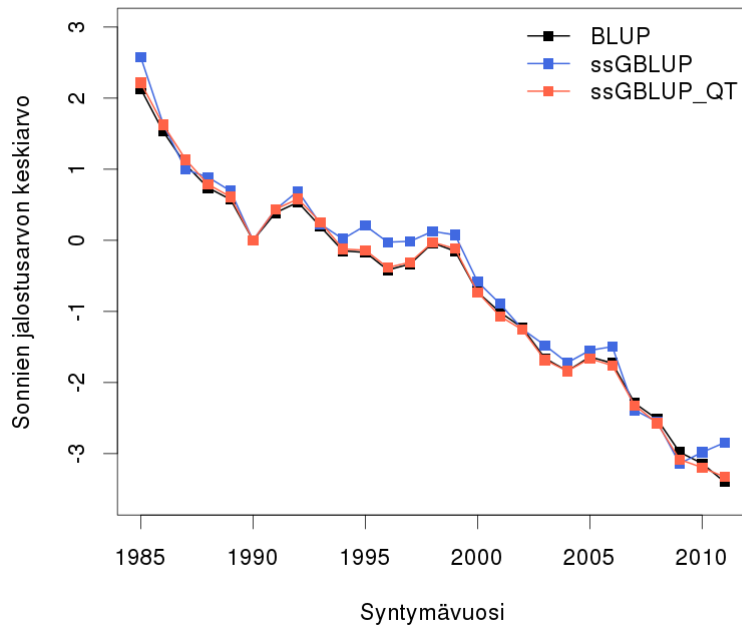
Tarkastellaan aluksi genomisen arvostelun tuloksia, jossa on käytetty perinteistä tapaa muodostaa matriisi \mathbf{H}^{-1} . Ominaisuudesta riippuen, genomisen ja perinteisen arvostelun geneettiset trendit seurasivat hyvin tai melko hyvin toisiaan. Esimerkkeinä kuvissa 1 ja 2 on tavallisella ja genomisella eläinmallilla saadut sonnien jalostusarvoihin perustuvat geneettiset trendit lepokauden pituudelle toisella lypsykaudella ja uusimattomuusprosentille kolmannella lypsykaudella. Vaikka geneettiset trendit muistuttavatkin toisiaan, eläinten järjestys syntymävuosittain saattaa vaihdella eri arvostelujen välillä. Siksi laskimme syntymävuosittain jalostusarvojen väliset korrelaatiot. Eri malleilla laskettujen arvostelujen välillä korrelaatiot olivat pääasiassa yli 0,90. Kahden arvostelun väliset järjestyserot olivat suurimmat uusimattomuusprosentin jalostusarvoissa, erityisesti kolmannella lypsykaudella (Kuva 3). Aivan kuten PCG-menetelmän suppeneminen parani QT-muunnoksen jälkeen, myös saadut jalostusarvot vastasivat paremmin perinteisen jalostusarvostelun tuloksia. Nyt kaikkien ominaisuuksien geneettiset trendit seurasivat perinteisten jalostusarvojen trendejä, lukuunottamatta mahdollisia pieniä eroja muutaman nuorimman sonniluokan tuloksissa (Kuvat 1 ja 2). Samoin syntymävuosittain lasketut jalostusarvojen korrelaatiot nousivat. Nyt ne olivat pääasiassa yli 0.97, myös uusimattomuusprosentille kolmannella lypsykaudella (Kuva 3).

Genomista monenominaisuuden jalostusarvostelua on tutkittu lehmien tiinehtymisprosentille holstein-rodulla (Aguilar ym., 2011). Samoin kuin hedelmällisyysominaisuuksilla tässä tutkimuksessa, myös tiinehtymisprosentilla on alhainen periytymisaste, mutta korkea geneettinen korrelaatio lypsykausien välillä. Tiinehtymisprosentille käytetyssä mallissa (Aguilar ym., 2011) ei kuitenkaan ollut geneettisiä ryhmiä ja teknisesti genomisen arvostelun tekeminen onnistui käyttämällä perinteistä \mathbf{H}^{-1} -matriisia. Genomisen tiedon käyttö jalostusarvostelussa johti validaatioarvosteluvarmuuden paranemiseen tiinehtymisprosentissa (Aguilar ym., 2011). Tätä odotamme myös jatkotutkimuksissa pohjoismaisen lypsykarjan hedelmällisyysominaisuuksissa.

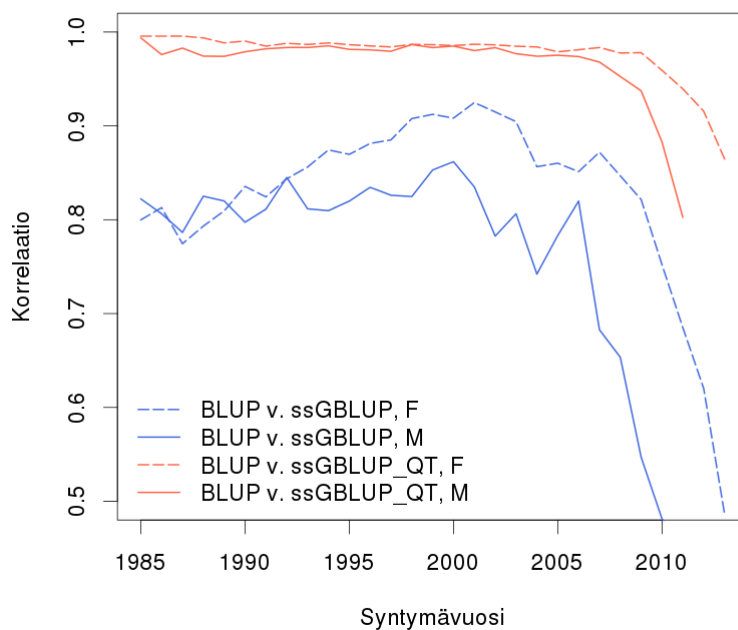
Kuva 1. Geneettinen trendi lepokauden pituudelle toisella lypsykaudella perinteisellä eläinmallilla (BLUP) sekä genomisella eläinmallilla kun QT-muunnosta ei ole tehty (ssGBLUP) tai on tehty (ssGBLUP_QT).



Kuva 2. Geneettinen trendi kolmannen lypsykauden uusimattomuusprosentille perinteisellä eläinmallilla (BLUP) sekä genomisella eläinmallilla kun QT-muunnosta ei ole tehty (ssGBLUP) tai on tehty (ssGBLUP_QT).



Kuva 3. Kolmannen lypsykauden uusimattomuusprosentin jalostusarvojen korrelaatiot syntymävuosittain lehmillä (F) ja sonneille (M) sekä perinteisen eläinmallin ja genomisen eläinmallin välillä (BLUP v. ssGBLUP) että perinteisen eläinmallin ja QT-muunnosta käyttävän genomisen eläinmallin välillä (BLUP v. ssGBLUP_QT).



Johtopäätökset

Ensimmäiset tulokset hedelmällisyysominaisuuksien genomiselle arvostelulle osoittivat, että siirtyminen eläinmallista genomiseen eläinmalliin on teknisesti mahdollista. Jalostusarvojen ratkaisemiseen käytetyn PCG-menetelmän suppenemisen äärimmäinen hitaus ei kuitenkaan vakuuttanut tulosten oikeellisuudesta. Geneettisten ryhmien huomioiminen QT-muunnoksen avulla osoittautui tarpeelliseksi. Sen lisäksi että tällä saavutettiin nopeampi PCG-menetelmän suppeneminen, saatiin myös luotettavimmat jalostusarvot perinteisiin jalostusarvosteluihin tehtyjen vertailujen perusteella. Tulosten perusteella genomisen eläinmalliarvostelun tekeminen pohjoismaisen lypsykarjan hedelmällisyysominaisuuksille on mahdollista.

Kirjallisuus

Aguilar, I., Misztal, I., Johnson, D. L., Legarra, A., Tsuruta, S. & Lawlor, T. J. 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.

Aguilar, I., Misztal, I., Tsuruta, S., Wiggans, G. R. & Lawlor T. J. 2011. Multiple trait genomic evaluation of conception rate in Holsteins. J. Dairy Sci. 94:2621–2624.

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

Koivula, M., Strandén, I., Pösö, J., Aamand, G. P. & Mäntysaari, E. A. 2015. Single-step genomic evaluation using multitrait random regression model and test-day data. J. Dairy Sci. 98:2775-2784.

Misztal, I., Legarra, A. & Aguilar, I. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92:4648-4655.

Misztal, I., Vitezica, Z. G., Legarra, A., Aguilar, I. & Swan, A. A. 2013. Unknown-parent groups in single-step genomic evaluation. J. Anim. Breed. Genet. 130:252-258.

MiX99 Development Team 2015. MiX99: A software package for solving large mixed model equations. Release VIII/2015. Natural Resources Institute Finland (Luke). Jokioinen, Finland. URL: <http://www.luke.fi/mix99>.

Quaas, R. L. & Pollak, E. J. 1981. Modified equations for sire models with groups. J. Dairy Sci. 64:1868-1872.

Tsuruta, S., Misztal, I., Aguilar, I. & Lawlor, T. J. 2011. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. J. Dairy Sci. 94:4198-4204.

Tyrisevä, A.-M., Muuttoranta, K., Mäntysaari, E. A., Pösö, J., Aamand, G. P., Eriksson, J.-Å. & Lidauer, M. 2016. Lypsykarjan hedelmällisyyden jalostusarvostelu päivitetty eläinmalliksi. Maataloustieteen Päivät, 12.-13.1.2016.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414-4423.