

Jalostusarvojen laskenta genomisella eläinmallilla

Minna Koivula, Esa Mäntysaari ja Ismo Strandén

MTT, Biotekniikka- ja elintarviketutkimus, Biometrinen genetiikka, 31600 Jokioinen
e-mail. [etunimi.sukunimi\(at\)mtt.fi](mailto:etunimi.sukunimi(at)mtt.fi)

TIIVISTELMÄ

Tällä hetkellä maailmalla käytössä olevat genomiset arvostelumallit ovat yleensä kaksi- tai kolmivaiheisia. Näissä eläimille lasketaan ensin perinteiset jalostusarvot, sitten yhdistetään jälkeläisten perusteella jalostusarvosteltujen sonnien jalostusindeksit ja SNP markkeritiedot, jolloin saadaan genomisen arvon ennustemalli. Tämän jälkeen voidaan ennustaa genomiset arvot niille eläimille, joilta on genomitiedot. Jos genomiseen arvoon yhdistetään eläimen polveutumistiedot, saadaan genomisen jalostusarvo-indeksi, GEBV. Genomisten jalostusarvojen luotettavuus on parempi, kun yhdistetään genomisen informaatio perinteisiin jalostusarvosteluihin käyttämällä ns. genomista eläinmallia eli single-step GBLUP:ia (ssGBLUP). Jalostusarvot lasketaan samanlaisella eläinmalli BLUP menettelyllä kuin aikaisemminkin, mutta nyt genotyypitettyjen eläinten sukulaisuudet perustuvat genotyypitietoihin ja muiden eläinten sukulaisuudet sukupuutitietoihin, kun aikaisemmin kaikkien eläinten sukulaisuudet perustuivat sukupuutitietoihin. Arvostelussa käytetään suoraan alkuperäisiä havaintoja, jolloin genomisen informaatio tulee huomioiduksi myös ympäristövaikutuksia ratkaistaessa. ssGBLUP-menetelmän haasteena on menetelmän laskennallinen vaativuus, sekä genomimatriisin, **G**, ja tavallisen sukulaisuusmatriisin, **A**, luotettava yhdistäminen ns. **H**-sukulaisuusmatriisiksi. Tässä tutkimuksen tavoitteena oli arvioida genomisen eläinmallin toimivuutta Pohjoismaisessa koelypsymallissa, käyttämällä oikeita koelypsuhavaintoja. Arvostelut laskettiin MiX99-ohjelmalla. Tulokset osoittivat, että ssGBLUP:n käyttö koelypsymallissa oli paitsi mahdollista, se myös tuotti paremmat jalostusarvojen ennusteet kuin ilman genomista tietoa. Menetelmällä lasketut luotettavuudet olivat korkeampia, samoin ennusteet olivat harhattomampia (korkeampi regressiokerroin) verrattuna perinteiseen polveutumisindeksiin. Seuraavana tavoitteena on laskentavuon parantaminen niin, että **G**:n käänteismatriisia ei tarvita, jolloin genotyypitettyjen eläinten määrä ei rajoita menetelmän käyttöä. Genomisissa jalostusarvosteluissa tullaan siirtymään ssGBLUP:n käyttöön, koska genotyypitettyjen eläinten lukumäärä kasvaa koko ajan ja yhä useammin genotyypitetään myös lemmiä. Tällöin nykyisin käytettävät arvostelumallit käyvät epäluotettavammiksi. ssGBLUP:n laskennallista tehokkuutta ja etenkin genomimatriisin muodostamista on kuitenkin vielä parannettava, etenkin, jos siirrytään käyttämään ns. HD eli High Density genotyyppiä.

Asiasanat: Genominen valinta, genomiset arvostelut, G-matriisi, jalostusarvostelu, eläinmalli

Johdanto

Genomisessa valinnassa jalostusarvon ennustamisessa ja valintapäätöksen tekemisessä käytetään joko yksilön koko genomia tai siitä määriteltyjä geenimerkkejä. Periaatteena on, että DNA-testistä saatu genomitieto yhdistetään perinteiseen sukupuuhun ja ominaisuuksien mittaustietoon perustuvaan jalostusarvosteluun. Tällä hetkellä useimmat käytössä olevat genomiset arvostelumallit ovat kaksi- tai kolmivaiheisia (VanRaden 2008, Hayes ym. 2009, VanRaden ym. 2009). Tällöin eläimille lasketaan ensin perinteiset jalostusarvo -indeksit eli EBV:t. Tämän jälkeen yhdistetään jälkeläisten perusteella jalostusarvosteltujen sonnien jalostusindeksit ja SNP-markkeritiedot, jolloin saadaan genomisen arvon ennustemalli. Näin voidaan ennustaa genomiset arvot niille eläimille, joilta on genomitietoa, mutta joilla ei vielä ole muuta omaa informaatiota. Yleensä genomisen arvo lasketaan käyttäen vain jälkeläisarvostelu-informaatiota. Tällöin nuoren eläimen genomisen arvostelun luotettavuutta voidaan vielä parantaa yhdistämällä sen polveutumisindeksi ja genomisen arvo. Näin kolmannessa vaiheessa muodostuu genomien jalostusarvostelu-indeksi, GEBV. Meillä Pohjoismaisessa arvostelussa tämä tehdään lineaari-mallilla, jossa genomisen arvo ja perinteinen EBV käsitellään kahtena eri ominaisuutena, joiden välinen korrelaatio on yhtä suuri kuin genomisen arvostelun luotettavuus.

Toisaalta genomisen informaatio voidaan sisällyttää suoraan perinteiseen eläinmalli BLUP jalostusarvosteluun. Tässä ns. single-step- menetelmässä (ssGBLUP) genotyyppitettyjen eläinten sukulaisuudet perustuvat genotyyppi-tietoihin ja muiden eläinten sukulaisuudet sukupuutietoihin (Aguilar ym. 2010, Christensen ja Lund 2010). Single-step arvostelussa käytetään suoraan alkuperäisiä havaintoja, jolloin genomisen informaatio tulee huomioiduksi myös ympäristövaikutuksia ratkaistaessa ja tuottaa eläimille suoraan genomisen jalostusarvon. Menetelmä on laskennallisesti vaativa. Omat haasteensa luo myös genomisen sukulaisuusmatriisin, **G**, ja tavallisen sukulaisuusmatriisin, **A**, luotettava yhdistäminen ns. **H**-matriisiksi. Jottei sukulaisuudet tule huomioiduksi kahteen kertaan, uusi yhtenäistetty sukulaisuusmatriisi ei voi vain yksinkertaisesti sisältää kumpaakin informaatiota, vaan se sisältää sukulaisuusinformaation lisäksi ns. genomisen informaatiolisän (mm. Meuwissen ym. 2011). Ensimmäisten single-step kokeilujen jälkeen on havaittu, että genomisten arvostelujen luotettavuus paranee ja harha pienenee, jos **G**-matriisia korjataan kohti **A**-matriisia ennen yhdistämistä **H**-matriisiksi (Christensen ym. 2012, Misztal ym. 2013). Tässä tutkimuksen tavoitteena oli arvioida ssGBLUP:n toimivuutta Pohjoismaisessa koelypsymallissa, käytettäessä oikeita koelypsuhavainnoja.

Käytetyt lyhenteet:

EBV: Jalostusarvo-indeksi, engl. estimated breeding value; eläimen jalostusarvon ennuste, nykyisellään usein eläinmalli BLUP.

PA: Polveutumis-indeksi, engl. parent average; nuoren eläimen jalostusarvon ennuste, jossa ainoa informaatio on vanhempien keskiarvo.

R²: Jalostusarvostelujen arvosteluvarmuus, engl. evaluation accuracy; kertoo jalostusarvostelun (EBV tai GEBV) ja todellisen jalostusarvon välisen korrelaation. Yleensä esitetään luotettavuutena, eli arvosteluvarmuuden neliönä R², tällöin analoginen periytymisasteen määritelmän kanssa (kuinka monta prosenttia vaihtelusta havainnoissa johtuu jalostusarvojen vaihtelusta).

DGV: Genomien arvo, engl. direct genomic value, DGV; genotyyppitettyille eläimille laskettu geenimerkkivaikutusten summa.

GEBV: Genominen jalostusarvo -indeksi, engl. genomic enhanced breeding value, genotyyppitettyille eläimille laskettu indeksi, jossa on yhdistetty genomisen arvo ja perinteinen EBV (yleensä polveutumisindeksi).

ssGBLUP: Genominen eläinmalli jalostusarvostelu, engl. single-step GBLUP; jalostusarvostelu, jossa normaalissa eläinmallissa on mukana genotyyppitettyillä eläimillä genomisen informaatio.

DRP: Deregressoitu jalostusarvo, engl. deregressed genetic merit; jalostusarvon ennusteista eli EBV:stä takaisin ratkaistu eläinkohtainen tulos. Sonnilla DRP vastaa tyttärien tulosten keskiarvon ja tyttärien emien keskiarvojen välistä erotusta.

Aineisto ja menetelmät

Aineisto

Tutkimuksessa käytettiin Pohjoismaista punaisen rodun (RDC) koelypsyaineistoa 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 lemmiä. Genomiaineistossa oli jokaisella genotyyppitettyllä eläimellä 46,914 SNP- markkeria. Tutkimuksessa analysoitiin NAV:n käyttämällä koelypsymallilla (Lidauer ym. 2006) sekä (I) koko koelypsyaineisto että (II) katkaistu aineisto, josta oli poistettu neljän viimeisen vuoden havainnot. Saatujen indeksien luotettavuus arvioitiin ns. validointimenetelmällä, jossa arvioidaan markkerivaikutukset vanhemmalla referenssijoukolla (aineisto II) ja ennustetaan nuorempien kandidaattieläimien tuloksia koko aineistossa I.

Tilastolliset menetelmät

Single-step GBLUP arvosteluissa on genotyyppitetuille eläimille ja niiden sukulaisille oletettu samankaltainen yhtenäistetty jalostusarvo. Jalostusarvoilla on erilainen sukulaisuusmatriisi eli varianssirakenne (\mathbf{H} -sukulaisuusmatriisi) riippuen siitä onko eläin genotyyppitetty vai ei.

Jalostusarvot ratkaistaan BLUP-menetelmässä ns. sekamalliyhtälöistä. Sekamallin yhtälöissä tarvitaan jalostusarvojen varianssirakenteen kääntematriisi. Perinteisessä eläinmallissa tarvitaan sukulaisuusmatriisin kääntematriisi \mathbf{A}^{-1} . Genomisessa eläinmallissa käytetään sen sijasta \mathbf{H}^{-1} -matriisia, joka voidaan kirjoittaa sukulaisuusmatriisin ja genomisen lisäinformaation summana (Aguilar ym. 2010; Christensen ja Lund 2010):

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

jossa \mathbf{A}_{22} on genotyyppitettyjen eläinten sukupuuhun perustuva sukulaisuusmatriisi ja $\mathbf{G}_w = w\mathbf{G}_0 + (1-w)\mathbf{A}_{22}$ on matriisi jossa on yhdistetty genomisen informaatio ja sukupuuinformaatio; w on kerroin, joka määrittää miten suuri osa geneettisestä vaihtelusta johtuu polygeenisistä tekijöistä, joita markkerit eivät selitä. Tärkein ero tavalliseen eläinmalliin on matriisi $\mathbf{B} = \mathbf{G}_w^{-1} - \mathbf{A}_{22}^{-1}$. Tässä tutkimuksessa käytimme \mathbf{H}^{-1} -matriisin rakentamisessa ns. Misztalin esittämää painotettua menetelmää (Misztal ym. 2013), jossa $\mathbf{B} = \tau\mathbf{G}_w^{-1} - \omega\mathbf{A}_{22}^{-1}$ ja oletusarvoinamme oli $w=0.10$, $\tau=1.5$ ja $\omega=0.5$.

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 1999). MiX99 -ohjelma lukee erikseen tavallisen sukupuutiedoston ja \mathbf{B} -matriisin. ssGBLUP arvostelujen keskimääräisen arvosteluvarmuuden validointi tehtiin Interbull GEBV validointitestillä (Mäntysaari ym. 2010), joka perustuu aikaisemmasta vanhasta aineistosta laskettujen jalostusarvojen kykyyn ennustaa tuotoksia uudemmassa aineistossa. Ensimmäin laskettiin jalostusarvot koko aineistosta I. Sitten laskettiin GEBV:t katkaistulla aineistolla II. Koko aineistolla lasketuista jalostusarvojen ennusteista (EBV_I) deregressoitiin standardoidut tuotospoikkeamat kaikille eläimille. Nuorten sonnien deregressoitu jalostusarvo (DRP) vastaa karkeasti sonnien tyttäryhmän keskiarvoa. Tämän jälkeen kandidaattisonnien ja nuorten lehmien DRP:tä mallinnettiin aineistosta II laskettuihin indekseihin (PA tai ssGBLUP)

$$\text{DRP} = b_0 + b_1 * \text{GEBV} + e,$$

jossa regressiokerroin b_1 kuvaa arvostelujen harhattomuutta. Jotta saadut ennusteet olisivat harhattomia, regressiokerroin (b_1) pitäisi olla lähellä yhtä, jolloin 1 indeksipiste genomisissa arvosteluissa vastaisi 1 indeksipistettä jälkeläisarvosteluissa. Validointi regressioyhtälön mallinsetelyaste R^2 puolestaan kuvaa genomisten arvostelujen luotettavuutta.

Tulokset ja tulosten tarkastelu

Tulokset osoittivat, että genominen eläinmallin käyttö koelypsymallissa oli paitsi mahdollista, se myös tuotti luotettavimmat jalostusarvojen ennusteet. Menetelmällä laskettujen genomisten jalostusarvojen luotettavuudet olivat korkeampia, samoin ennusteet olivat harhattomampia (korkeampi regressiokerroin) verrattuna polveutumisindeksiin sekä sonneilla että lehmillä (taulukko 1 ja taulukko 2). Aikaisempiin genomisiin jalostus-indekseihin verrattuna olivat sonnien ssGBLUP arvosteluilla 5 - 18 prosenttiyksikköä korkeammat luotettavuudet (Koivula ym. 2012).

Taulukoista 1 ja 2 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. Alhaiset regressiokertoimet sonneilla osoittavat että ssG-BLUP:n antamat GEBV:t yliarvioivat jalostusarvojen välistä vaihtelua. Toisaalta myös polveutumisindeksin (PA) antamat b_1 kertoimet osoittavat, että nuorten sonnien saamat polveutumisindeksit ovat odotettua suurempia. Genomisella eläinmallilla regressiokertoimet ovat maidolla ja valkuaisella vähän lähempänä yhtä verrattuna aikaisempiin genomiarvosteluihin (Koivula ym. 2012), mikä tukee ajatusta, että ssGBLUP parantaa genomisten arvostelujen luotettavuutta. Jostain syystä rasvalla ssGBLUP tuottaa hivenen alhaisemman regressiokertoimen kuin aikaisemmissa genomiarvosteluissa.

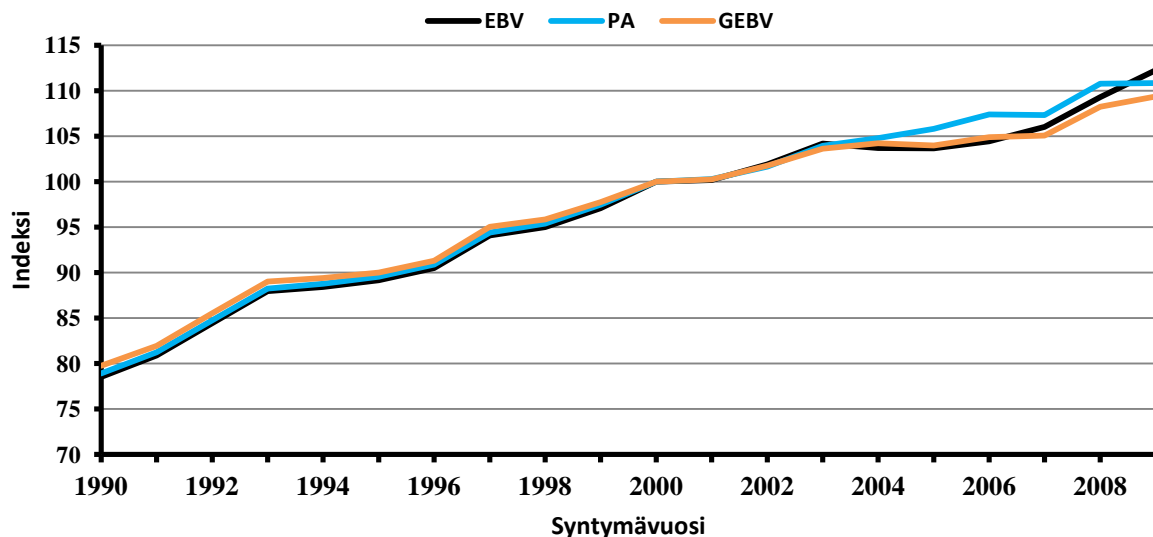
Kuvasta 1 voidaan havaita, että sonnien GEBV:stä laskettu maidon genominen jalostusarvojen trendi (keskimääräinen jalostusarvo/syntymävuosi) noudattaa paremmin sonnien jalostusarvojen trendiä kuin polveutumisindeksien perusteella lasketut trendit.

Taulukko 1. Regressiomallin regressiokerroin (b_1) sekä luotettavuudet (R^2) ssGBLUP malleista ja polveutumisindeksistä (PA) eri tuotosominaisuuksille RDC sonneilla. Vertailun vuoksi taulukossa tulokset aiemmista isämalli DGV:stä sekä isämalli GEBV:stä *)Koivula ym 2012.

	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

Taulukko 2. Regressiomallin regressiokerroin (b_1) sekä luotettavuudet (R^2) ssGBLUP malleista ja polveutumisindeksistä (PA) eri tuotosominaisuuksille RDC lehmillä.

	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



Kuva 1. Sonniin jalostusarvojen trendit maidolla laskettuna tavallisista jalostusarvoista (EBV), genomisista jalostusarvoista (GEBV) ja polveutumisindekseistä (PA). GEBV ja PA perustuvat katkaistusta datasta saatuihin jalostusarvojen ennusteisiin.

Johtopäätökset

Genomitiedolla lisätyn eläinmalli ssGBLUP:n luotettavuudet validaatiotestissä olivat korkeampia, samoin ennusteet olivat harhattomampia (korkeampi regressioeroin) verrattuna perinteiseen polveutumisindeksiin. Genomisissa arvosteluissa tultaneen siirtymään ssGBLUP:n käyttöön, koska genotyyppitettyjen eläinten lukumäärä kasvaa koko ajan ja yhä useammin genotyyppitetään myös lehmiä. Tällöin nykyisin käytettävät arvostelumallit käyvät epäluotettavammiksi. ssGBLUP:n laskennallista tehokkuutta ja etenkin genomimatriisin muodostamista on kuitenkin vielä parannettava, etenkin, jos siirrytään käyttämään ns. HD eli High Density genotyyppijä.

Kirjallisuus

- 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.
- Christensen, O.F. & Lund, M.S.** 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42:2.
- Christensen, O.F.** 2012. Compatibility of pedigree-based and marker-based relationship matrices for single-step genetic evaluation. *Genet. Sel. Evol.* 44:37.
- Hayes, B.J., Bowman, P.J., Chamberlain, A.J. & Goddard, M.E.** 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92: 433-443.
- Koivula, M., Strandén, I., Su, G. & Mäntysaari, E.A.** 2012. Different methods to calculate genomic predictions – comparisons of SNP-BLUP, G-BLUP and H-BLUP. *J. Dairy Sci.*, 95:4065-4073.
- 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.
- Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E.** 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819-1829.

- Meuwissen T.H.E., Luan, T. & Woolliams, J.A.** 2011. The unified approach to the use of genomic and pedigree information in genomic evaluations revisited. *J Anim. Breed. Genet.* 128: 429-439
- Misztal, I., Aggrey, S.E. & Muir, W.M.** 2013. Experiences with a single-step genome evaluation. *Poultry Science* 92 :2530–2534.
- Misztal, I., Tsuruta, S. Aguilar, I., Legarra, A., VanRaden, P. M. & Lawlor, T.J.** 2013. Methods to approximate reliabilities in single-step genomic evaluation *J. Dairy Sci.* 96: 647-654..
- Mäntysaari, E.A., Liu, Z. & VanRaden, P.** 2010. Interbull validation test for genomic evaluations. *Interbull Bull.* 40: 1-5.
- Strandén, I. & Lidauer, M.** 1999. Solving large mixed models using preconditioned conjugate gradient iteration. *J. Dairy Sci.* 82: 2779-2787.
- VanRaden, P.M.** 2008. Efficient Methods to Compute Genomic Predictions. *J. Dairy Sci.* 91: 4414-4423.
- VanRaden, P.M., Van Tassell, C.P., Wiggans, G.R., Sonstegard, T.S. & Schnabel, R.D.** 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci.* 92: 16-24.