

Should gene interactions be included in genomic evaluation in animal breeding?

(Pitääkö geenien yhdysvaikutus sisällyttää eläinjalostuksen genomiseen arvosteluun?)

Asko Mäki-Tanila¹, W.G. Hill²

¹MTT Agrifood Research Finland, Biotechnology and Food Research, Jokioinen

²University of Edinburgh, Institute of Evolutionary Biology, Edinburgh, UK

TIIVISTELMÄ

The genetic comparison of animals is based on their own performance and that of animals sharing genetic factors with them. Their expected genetic similarity is deduced from pedigree information and also now directly using a large number of molecular genetic markers over the genome (genomic breeding values). Quantitative trait analyses may also include gene interaction or epistatic effects. Additive x additive interaction effects have been found, particularly in crosses of inbred and widely diverse selected lines. These and gene functional studies have generated much interest in including the interaction effects in genome-wide analyses within populations, including animal breeding stocks. Several issues need consideration before incorporating them in genetic models: influence of gene interaction on the genetic evaluation and on the gains produced by selection, proportion of epistatic variance with multiple genes, expectations with common allele frequency distributions, and probability of finding interaction effects with the genomic tools. - The average effect of an allele already includes interaction effects with other loci, but with magnitude dependent on their frequencies. If a major epistatic effect is favourable, selection may fix the respective allele quickly. With milder effects the frequencies of interacting favourable alleles at both loci of pair will increase. - Even with additive effects in an underlying genotype, the relationship between phenotypes and genotypes may be non-linear and there is epistasis on the observed scale. An example is a categorical trait (diseased or not), where the analysis on the observed scale using an approximating model can be transformed to the underlying additive scale. In the multiplicative model the amount of epistasis increases with the coefficient of variation (CV), but the proportion never exceeds $1 - \ln(1+CV^2)/CV^2$, and most of the epistatic variance is due to two-locus interactions. - The additive variance is directly proportional to heterozygosity (H), with a maximum at allele frequency $1/2$ in a biallelic case. Additive x additive variance requires segregation in both the interacting loci A and B and is proportional to $H_A H_B$, and correspondingly for more loci. Hence epistatic variance can reach high values only when allele frequencies near $1/2$. - As the number of loci (n) is increased, average effects at individual loci decline with $1/\sqrt{n}$ (i.e. variance as $1/n$). Similarly additive x additive effects must decline as $1/n$. In genome-wide analyses, the number of effects to be estimated is the square of that for individual loci. With many thousands of markers very stringent test criteria have to be used so the power is very low. It has become obvious that the genomic tools cannot harvest all the existing genetic variation. In particular the variation due to rare alleles is often undetected. Such problems are even more likely in considering interaction effects. In summary, gene interaction effects are automatically utilized in selection using additive models while most epistatic effects are expected to be very small and difficult to detect in genome-wide analyses.

Asiasanat:

genominen arvostelu, epistasia, eläinjalostus

Johdanto

Eläimen geneettisestä tasosta saadaan käsitys käyttämällä tiedon lähteinä sen omaa tulosta sekä tietoja yksilöiltä, joilla on samoja geenejä. Sukulaisilla on samoja geenejä enemmän kuin populaation yksilöillä keskimäärin. Yhdenmukaisuuden aste päätellään sukupuusta.

Nykyään samojen geenien osuudesta saa suoran mitan tyypittämällä eläimet koko perimän eli genomin alueelta tiheällä DNA-merkkien joukolla. Eläimen ja sen toisen vanhemman – isän tai emän – geneistä puolet on samoja. Myös täyssisarten geenien odotetaan olevan 50-prosenttisesti samanlaisia. Tämän puolikkaan käyttäytyminen on kuitenkin erilainen näissä kahdessa sukulaisuustyyppissä (Hill 1993).

Kaikissa jälkeläisissä tarkalleen puolet perimästä on tullut isältä ja toinen puoli emältä. Yhteisissä osuuksissa ei ole mitään vaihtelua. Vaikka täyssisarten genomi on keskimäärin puoliksi yhteistä alkuperää, samankaltaisuuden osuudet vaihtelevat. Molekyyligenetiikka tulee tässä apuun. DNA-merkkien avulla saadaan tarkka arvio siitä, kuinka suuri osuus tarkasteltavan sisarusparin genomista on samanlaista.

Genomisen arvostelun teho

C R Henderson'in kehittämässä BLUPissa eli sukutietoihin perustuvassa jalostusarvostelussa käytetään sukupuuta ja oletetaan, että yhden sukulaistyyppin yksilöt ovat kaikki sukulaisuuskertoimen osoittaman määrän verran samanlaisia. Genominen arvostelu (Meuwissen et al. 2001) taas pääsee käsiksi sukulaisten todellisiin samankaltaisuuksiin ja antaa sen takia tarkemman arvion esimerkiksi täyssisarten jalostusarvoista.

Kaukaisemmissa sukulaisissa sukupuusta laskettu samankaltaisuuden odotusarvo on alhainen. DNA-merkeistä saatava tieto ei tätä paranna. Genomisen arvostelun antama lisä on arvokkainta lähisukulaisilla (Habier et al 2013). Näin on erityisesti täyssisarilla. Heti tulevat mieleen lypsykarjan alkionsiirtojälkeläiset, tai jos ei olla fakkiutuneita nautaan, sikapahnueet tai kanojen ja kalojen isot täyssisarperheet.

Genomisessa arvostelussa on vielä toinenkin edullinen piirre. Kun genomin eri osien vaikutukset ovat erikokoisia, tätäkin vaihtelua voidaan jäljittää ja hyödyntää taas DNA-merkkien ja genomisten jalostusarvojen avulla.

Tässä muutamia syitä miksi genotyyppitys ja genominen jalostusarvostelu ovat hyödyllisiä ja suosittuja. Lypsykarjalla ylivoimaisina etuina ovat suppeampi sonnien jälkeläisarvostelu ja valinnan aikaistuminen (Schaeffer 2006). Samankaltaisuuden ja geenivaikutusten vaihtelun jäljille pääseminen vaatii mahdollisimman suuren merkkijoukon käyttöä, paljon tyypitettyjä yksilöitä ja luotettavat arviot näiden yksilöiden geneettisistä tasoista eli jalostusarvoista (VanRaden 2008).

Ovatko yhdysvaikutukset tärkeitä?

Genomin kattavia analyysejä on käytetty tuotokseen ja kestävyteen vaikuttavien geenien (QTL) etsimiseen. Naudalla tämän työn lähtökohtana ovat tarkat arviot yksilön (keinosiemennyssonnin) geenien vaikutusten summasta (jalostusarvosta) (e.g. Weller et al. 1990).

QTL-analyysejä on eri populaatioista tehty tuhansia. Lopputulos on ollut aika laiha. Ensiksikin, sopivalta tuntuvat fysiologisesti tärkeät geenit selittävät hyvin vähän tai tuskin lainkaan perinnöllistä vaihtelua. Toiseksi, uusia merkittäviä geenejä ei ole löytynyt oikein nimeksikään (e.g. Mäki-Tanila 2012).

Näyttää siltä, että eläinten väliset erot johtuvat sadoista pienivaikutteisista geeneistä tai näiden geenien kirjavasta säätelytekijöiden joukosta (Hill et al 2008). Perimää kauttaaltaan haravoiva genominen arvostelu osaa jäljittää tätä kuvaa tehokkaasti, kun suuresta merkkijoukosta riittää viestintuoja jokaisen geenin tai säätelytekijän vaihtoehdosta.

Jalostajat ovat tyytyväisiä jalostusarvojen parempaan laskentaan, itse geenien löytyminen olisi tietysti raflaava sivujuoni. Ne muutamat geenit jotka ovat löytyneet QTL-analyysien jatkotarkasteluissa, näyttävät vaikuttavan kukin omalla tavallaan ja enimmäkseen hyvin monimutkaisesti (Andersson 2009).

On melkein säännönmukaista, että geenin vaikutus riippuu siitä, minkä muiden geenien kanssa se toimii. Puhutaan geenien yhdysvaikutuksesta eli epistasiasta. Geenien epistaattiset vaikutukset voivat olla positiivisia tai negatiivisia.

Uppsalassa työskentelevän Örjan Carlborg'in (Wahlberg et al 2009) ryhmä on löytänyt epistasiaa kanojen risteytysjälkeläisissä, joiden vanhempaislinjoissa painavimpien keskiarvo (8 viikon paino 1522 g) on melkein kymmenen kertaa kevyempien keskiarvo (181 g). Populaation sisäisestä vaihtelusta on Trudy Mackay'n (Huang et al. 2009) ryhmä Pohjois-Karoliinassa esittänyt arvioita QTL:ien epistasiasta banaanikärpäsellä - todeten kuitenkin perään että tulokset voitaisiin yhtä hyvin esittää yksinkertaisesti geenivaikutusten summina. Geenien yhdysvaikutukset ovat ehkä biokemiallisten ja fysiologisten prosessien hierarkian luonnollinen seuraus (Phillips 2008).

On siis olemassa tuloksia ja väitteitä yksittäisten geenien yhdysvaikutuksista. Epistasia on tällä hetkellä suosittu tutkimuskohde. Sen takia on hyvä muodostaa käsitys siitä, mikä geenien yhdysvaikutusten merkitys on vaihtelulle ja jalostukselle ja mitkä tekijät vaikuttavat yhdysvaikutusten löytymiseen. Ajankohtaisempi kysymys on, pitäisikö genomisessa arvostelussa lisätä geenivaikutusten ohkeen myös niiden yhdysvaikutukset?

Heterotsygotia vaihtelun määrän mittari

Valinnan vaikutus riippuu perinnöllisen vaihtelun määrästä, erityisesti geenien yksinkertaisten vaikutusten summan vaihtelusta (additiivisesta vaihtelusta).

Epistasia on monimutkainen tutkimuskohde. Sen tutkiminen kannattaa aloittaa kahden geenin tapauksesta. Ominaisuuteen voi vaikuttaa geenit A ja B. Yksilön kussakin geenissä on samojen (homotsygootti) tai eri (heterotsygootti) alleelien pari – toinen on tullut isältä ja toinen emältä.

	Bb	Bb	BB
aa	0	<i>b</i>	<i>2b</i>
Aa	<i>a</i>	<i>a + b + [ab]</i>	<i>a + 2b + 2[ab]</i>
AA	<i>2a</i>	<i>2a + b + 2[ab]</i>	<i>2a + 2b + 4[ab]</i>

Oletetaan että A-alleelin frekvenssi on p ja B-alleelin q . Silloin keskiarvo $\mu = 2 p a + 2 q b + 4 p q [ab]$. Alleelien keskimääräinen vaikutus voidaan arvioida perinteisen monimutkaisesti pienimmän neliösumman menetelmän mukaan (Cockerham 1954). Kojima (1959) oivalsi laskea alleelin keskimääräisen vaikutuksen johtamalla pienen alleelifrekvenssin vaikutuksen keskiarvoon eli

derivoimalla keskiarvon alleelifrekvenssin suhteen. Silloin esim. A-geenissä alleelimuutoksesta johtuvaksi keskimääräiseksi vaikutukseksi tulee

$$\alpha = \frac{1}{2} \frac{\partial \mu}{\partial p} = a + 2q [ab] \quad (1)$$

Samalla tavalla saadaan B-geenin keskimääräinen vaikutus β .

Yksilön jalostusarvo koostuu siinä olevien eri geenien alleelien keskimääräisistä vaikutuksista populaatioissa. Eri populaatioissa geenien vaihtelu ja alleelifrekvenssit ovat erilaisia, sen takia jalostusarvo riippuu vertailtavista yksilöistä ja epistasian yhteydessä myös niiden perimän yhdysvaikutuksista. Geenien interaktion voi olla positiivista tai negatiivista. Yksi mielenkiintoinen malli on sellainen jossa vain interaktiovaikutukset ovat nolasta poikkeavia (duplikaattimalli). Esimerkiksi usean tekijän säätelemässä biokemiallisessa prosessissa voi esiintyä tämäntyppisiä yhdysvaikutuksia (e.g. Keightley 1989).

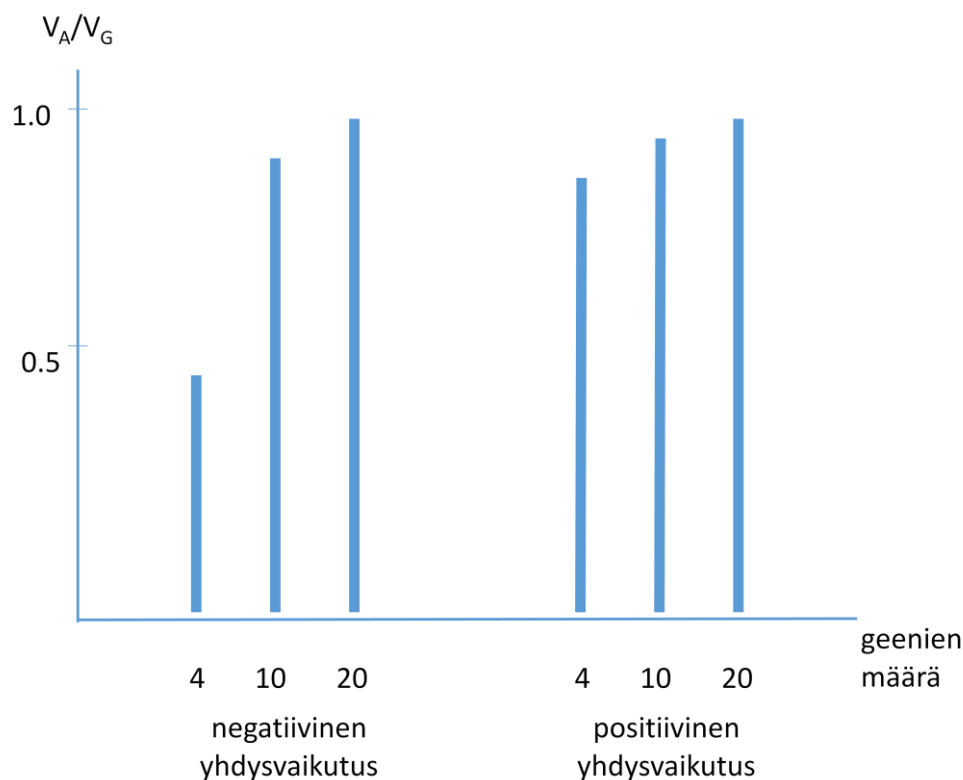
Jonkin ominaisuuden additiivinen vaihtelu (ja periytyvyysaste) riippuu ominaisuuteen vaikuttavissa geneeissä olevien vaihtoehtojen määrästä ja yleisyyksistä, lyhyesti sanottuna heterotsygoottisten genotyyppien osuudesta (Falconer & Mackay 1996, Hill et al. 2008). Esimerkiksi A-geenin heterotsygotia on $H_A = 2p(1-p)$. Kahdesta geenistä johtuva additiivinen varianssi on $V_A = H_A \alpha^2 + H_B \beta^2$. Lausekkeesta (1) huomataan että geeni-interaktion vaikutus sisältyy additiiviseen vaihteluun, sen kontribuutio riippuu geeniparissa olevien alleelien frekvensseistä. Heterotsygotia on maksimissaan, kun eri geneeissä alleelit ovat yhtä yleisiä, eli kun kahden alleelin tapauksessa kummankin osuus on puolet. Jos kahdella geenillä on myös yhdysvaikutus, additiivisen vaihtelun lisäksi on olemassa epistaattista vaihtelua. Kojiman mukaan keskiarvon toinen derivaatta toisen lokuksen suhteen antaa additiivinen x additiivinen -yhdysvaikutuksen keskimääräisen vaikutuksen $[\alpha\beta] = \frac{1}{4} \frac{\partial^2 \mu}{\partial p \partial q} = [ab]$. Epistaattisen vaihtelun esiintyminen edellyttää, että molemmista geneeistä on populaatioissa useita alleeleja eli molemmat geenit ovat heterotsygoottisia. Kahden geenin yhdysvaikutuksen varianssi on $V_{AA} = H_A H_B [\alpha\beta]^2$.

Tämän takia epistaattista vaihtelua havaitaan parhaiten risteytyspopulaatioissa, joissa alleelien osuudet ovat vaihtelevissa geneeissä lähes yhtä suuria, eli 1/2. Silloinkin vaihtelevien geeniparien osuus on paljon pienempi, eli puolikkaiden tulo 1/4. Yleensä alleelifrekvenssit ovat mieluummin lähellä nolaa tai ykköstä kuin puolikasta, ja sen takia myös heterotsygotian taso on äskeistä 25 prosenttia alhaisempi (e.g. Crow & Kimura 1970) ja edelleen epistasian osuus vaihtelusta pienempi.

Yhdysvaikutus voi olla monimutkaisempaa. Kolmen geenin interaktion tapauksessa kahden geenin on yhdysvaikutusten vaihtelu verrannollinen heterotsygotian kuutioon. Korkeamman asteen yhdysvaikutusten antama lisä epistaattiseen vaihteluun on tämän ajattelun mukaan entistä pienempää. Toisin sanoen, jos arvioimme epistasian merkitystä, kahden-kolmen geenin yhdysvaikutusten tarkastelu riittää eikä kannata mennä monimutkaisempiin interaktioihin.

Mitä enemmän geenejä, sitä pienempiä vaikutuksia

Kun vaihteluun vaikuttaa paljon geenejä, on yksittäisen geenin vaikutus pieni. Ja aina vain pienempi, mitä enemmän geenejä on. Kun vaikuttavia geenejä on 50, on mahdollisia erilaisia geenipareja 1250. Kun geenejä on useampia tuhansia, on potentiaalisia pareja miljoonia. Jotta geenivaikutusten kokonaisuus pysyy äärellisenä mitta-asteikon sisällä, on geenipareille spesifisten yhdysvaikutusten suuruuden oltava olemattoman pieni.



Kuva. Kahden geenin interaktion vaikutus kvantitatiivisen ominaisuuden vaihteluun (V_A additiivinen geneettinen vaihtelu, V_G on koko geneettinen vaihtelu) kun geenien (alleelifrekvenssi kaikissa geneeissä sama 0.5) määrä kasvaa. Geenin ja interaktion vaikutus ovat yhtä suuria.

Johdonmukaisesti QTL-analyysien tulosten mukaan voi ennustaa että yhdysvaikutusten löytyminen on hyvin epätodennäköistä. Genomin kattavissa analyyseissä on tällä hetkellä käytössä satojen tuhansien DNA-merkkien joukko. Geeniparien vaikutusten arviointiin olisi näin käytettävissä miljardien joukko. Jotta tällaisesta analyysistä löytyisi vaihteluun vaikuttavia geenejä ja vielä lisäksi tilastollisen merkitsevyysrahan ylittäviä yhdysvaikutuksia, tarvitaan miljoonia genotyyppitettyjä eläimiä tuotostietoineen. Periaatteessa genomisessa arvostelussa voidaan tilastolliseen analyysiin laittaa geenien vaikutuksia jäljittävien merkkien jälkeen merkkien muodostamat parit.

Aina kun genotyypin ja fenotyypin arvon välillä on ei-lineaarinen suhde, esiintyy havaintoasteikolla epistaattista vaihtelua. Tästä ovat esimerkkejä fitness- tai kynnysominaisuuksien vaihtelu. Multiplikatiivinen malli on esimerkki tapauksesta jossa asteikon muutoksella päästään takaisin puhtaaseen additiiviseen vaihteluun. Multiplikatiivinen malli soveltuu matemaattisten ominaisuuksiensa takia epistasiavaihtelun tutkimiseen. Havaintojen vaihtelukerroin (CV , ominaisuudesta riippuen 0.1 – 0.3) on hyvä mittari ei-lineaarisuuden määrälle. Kun ominaisuuteen vaikuttavien geenien määrä on hyvin suuri, havainnot ovat log-normaalisti jakautuneita ja geneettisen vaihtelun additiivinen osa on $\ln(1+CV^2)/CV^2$. Toisin sanoen multiplikatiivisen interaktion yhteydessä äärettömön suuri geenijoukko tuottaa korkeallakin CV :llä (0.3) enimmäkseen (97%) additiivista vaihtelua. Epistaattinen vaihtelu on pääosin lokusparien tuottamaa.

Johtopäätökset

Geenien yhdysvaikutuksia on löytynyt useampia kertaluokkia toisistaan poikkeavien linjojen risteytyksistä. Epistasian vaikutuksesta jalostuspopulaation sisäiseen vaihteluun ei ole toistaiseksi varteenotettavaa näyttöä. Teoriankaan perusteella ei epistaattisen vaihtelun odoteta olevan merkittävää.

Epistasiaa on hankala hyödyntää valinnassa koska edulliset geeniyhdistelmät purkautuvat sukupolvesta toiseen mentäessä. Yhdysvaikutukset on kyllä mahdollista sisällyttää tilastollisiin analyyseihin parantamaan arvostelulaskennan tarkkuutta. Toisaalta geenien yhdysvaikutukset sisältyvät jo yksittäisten geenien keskimääräisiin vaikutuksiin ja niiden kerryttämiin jalostusarvoihin. Sen takia genomisen arvostelun voi näillä näkymin tehdä huoletta ilman yhdysvaikutusten tuomaa monimutkaisuutta.

Kiitokset: Tutkin tätä aluetta työskennellessäni W G (Bill) Hillin kanssa Edinburghin yliopistossa OECD-apurahalla.

Kirjallisuus

- Andersson L** 2009 Genome-wide association analysis in domestic animals: a powerful approach for genetic dissection of trait loci. *Genetica* 136:341-349
- Cockerham CC** 1954 An Extension of the Concept of Partitioning Hereditary Variance for Analysis of Covariances among Relatives When Epistasis Is Present. *Genetics* 39: 859–882.
- Crow JF, Kimura M** 1970 An Introduction to Population Genetics Theory. New York: Harper & Row.
- Falconer DS, Mackay TFC** 1996 Introduction to quantitative genetics, 4th edition. Harlow, Essex: Longmans Green.
- Habier D, Fernando RL, Garrick DJ** 2013 Genomic BLUP decoded: a look into the black box of genomic prediction. *Genetics*. 194: 597-607
- Hill WG** 1993 Variation in genetic identity within kinships. *Heredity* 71:652–653.
- Hill WG, Goddard ME, Visscher PM** 2008 Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet* 4: e1000008.doi:10.1371/journal.pgen.1000008
- Huang W, Richards S, Carbone MA, Zhu D, Anholt RR, Ayroles JF, Duncan L, Jordan KW, Lawrence F, Magwire MM, Warner CB, Blankenburg K, Han Y, Javaid M, Jayaseelan J, Jhangiani SN, Muzny D, Onger F, Perales L, Wu YQ, Zhang Y, Zou X, Stone EA, Gibbs RA, Mackay TF** 2012 Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. *Proc Natl Acad Sci U S A*. 109: 15553-15559.
- Kojima K** 1959 Role of Epistasis and Overdominance in Stability of Equilibria with Selection. *Proc Natl Acad Sci U S A* 45: 984–989.
- Mäki-Tanila, A** 2012 Animal molecular genetics: from major genes to genomics. In: Encyclopedia of sustainability science and technology. Volume 1 / Ed. Meyers, Robert A.. Springer Science+Business Media. p. 473-497.
- Phillips PC** 2008 Epistasis – the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* 9: 855–867.
- Schaeffer LR** 2006 Strategy for applying genome-wide selection in dairy cattle. *J Anim Breed Genet*. 123: 218-223.
- VanRaden P** 2008 Efficient methods to compute genomic predictions. *J Dairy Sci* 91:4414-4423
- Wahlberg P, Carlborg O, Foglio M, Tordoir X, Syvänen AC, Lathrop M, Gut IG, Siegel PB, Andersson L.** 2009 Genetic analysis of an F(2) intercross between two chicken lines divergently selected for body-weight. *BMC Genomics* 10:248. doi: 10.1186/1471-2164-10-248.
- Weller JI, Kashi Y, Soller M.** 1990 Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *J Dairy Sci*. 73: 2525-2537.