



Eläinjalostus: eläimiä, geenejä ja matemaattista tilastotiedettä

*Ismo Strandén*¹

Eläinjalostuksella muutetaan kotieläimiä haluttuun suuntaan. Eläinjalostuksesta voi tulla mieleen kloonaus tai ehkä geenimanipulaatio. Kotieläintuotannon eläinjalostus on kuitenkin aivan muuta, eikä tällaiselle bioteknologialle ole siinä tarvetta. Sen sijaan tärkeitä työkaluja ovat matemaattisen tilastotieteen tietokonesovellukset. Nämä sovellukset yhdistävät useiden alojen tietoa: genetiikkaa, biologiaa, tilastotiedettä, numeerisia menetelmiä ja tietojenkäsittelyä. Eläinjalostuksen matemaattisia menetelmiä käytetään tuotantoeläimille kuten lypsykarjan, sikojen ja broilerin jalostuksessa mutta myös lähes kaikille ihmisen kasvattamille eläimille.

Eläinjalostus toimii kuten luonnonvalinta eli evoluutio: parhaat yksilöt saavat eniten jälkeläisiä. Parhailla yksilöillä on parhaat geenit tai ainakin muita vähemmän huonoja geenejä. Valitettavasti emme tunne kaikkia geenejä, jotka vaikuttavat esim. eläinten terveyteen tai tuotantokykyyn. Voimme kuitenkin mitata eläimistä ominaisuuksia kuten maidon tuotantomäärä, mutta mittauksiin vaikuttavat geenien lisäksi mm. ravinnon ja ympäristön laatu. Käytännössä arvioimme eläimen geneettisen arvon matemaattisilla malleilla. Kyse on samasta asiasta kuin esimerkiksi tarkennettaessa suttuista kuvaa kuvankäsittelyohjelmalla. Itse asiassa samankaltaisia tilastotieteen malleja voidaan käyttää niin kuvankäsittelyssä kuin eläinjalostuksessa.

Eläimen geneettinen arvo pyritään arvioimaan kaikille tärkeille ominaisuuksille. Käytännössä geneettinen arvo lasketaan samanaikaisesti kaikille eläimille: sekä eläville, että jo kuolleille. Tällä tavalla kaikki ke-

rätty tieto tulee käytettyä hyväksi. Geneettisen arvon arvioimista kutsutaan jalostusarvosteluksi ja geneettistä arvoa jalostusarvoksi. Otan yksinkertaistaen esimerkkinä Pohjoismaiden (jossa ovat mukana Suomi, Ruotsi ja Tanska) jalostusarvostelun lypsykarjan maidontuotanto-ominaisuuksille. Ominaisuuksia ovat maidontuotanto sekä sen komponentit, joita ovat valkuainen ja rasva. Jalostusarvostelussa on useita vaiheita ja se sisältää useita ongelmia, joihin on omat ratkaisunsa. Yksi vaihe on ratkaista lineaarinen yhtälöryhmä, jossa on jopa 150 miljoonaa tuntematonta. Tällainen yhtälöryhmä saadaan tilastollisesta mallista, jossa huomioidaan kaikki tunnetut sukulaisuudet eläinpopulaatiossa. Ratkaisemisessa tarvitaan kehittyneitä numeerisia menetelmiä ja raskasta rinnakkaislaskentaa. Tällainen yhtälöryhmä on ratkaistava useita kertoja vuodessa. Tarkasteltavia ominaisuuksia lypsykarjan jalostuksessa on yhteensä 14, joista tässä siis oli kolme. Jokaiselle ominaisuudelle on oma tilastollinen malli ja ai-neisto.

Eläinjalostuksen kannalta ensimmäinen kysymys on: vaikuttavatko ominaisuuteen geenit ja kuinka paljon. Eläinjalostuksen matemaattiset mallit ovat tilastotieteen nk. sekamalleja, joissa on ns. kiinteitä ja satunnaisia tekijöitä. Eläinjalostuksen sekamalleissa kiinteillä tekijöillä otetaan huomioon havaintoon vaikuttavia ympäristötekijöitä, kuten eläimen ikä, sukupuoli ja mittauksen vuodenaika. Satunnaistekijöillä huomioidaan eläinten välisiä samankaltaisuuksia sukulaistiedon eli geenien takia. Jalostusarvo arvioidaan useimmiten tällaisella lineaarisella sekamallilla, koska sen rat-

¹Kirjoittaja on suorittanut filosofian maisterin ja lisensiaatin tutkinnot sovelletusta matematiikasta Turun yliopistossa sekä filosofian tohtorin tutkinnon (PhD) kotieläintieteestä Wisconsin-Madisonin yliopistossa. Nykyisin hän työskentelee erikoistutkijana Biotekniikka ja elintarvikkeet -yksikössä Maa- ja elintarviketalouden tutkimuskeskus MTT:ssä sekä dosenttina Helsingin yliopistossa.

kaiseminen on laskennallisesti yksinkertaista ja nopeaa.

Viimeisen kymmenen vuoden aikana on aloitettu genomisen tiedon käyttö jalostusarvosteluissa. Genomisella tiedolla tarkoitetaan genotyyppien tietämistä useista tuhansista geenipaikoista. Nämä geenipaikat eivät useinkaan ole varsinaisia ominaisuuteen vaikuttavia geenejä vaan niiden lähellä olevia ns. merkkigeenejä. Genomitiedon myötä analysoitavien aineistojen koko kasvaa edelleen. Kun aikaisemmin eläimen havaintoa kohden oli selittäviä tekijöitä, kuten ympäristötekijöitä, tavallisesti alle 10, genomitiedon myötä jokaiselta genotyyppitetyltä eläimeltä käytetään potentiaalisina selittävinä tekijöinä noin 400 000 merkkipaikkaa. Tällä hetkellä genotyyppityksen kalleuden takia vain osalta eläimistä tiedetään genotyyppit, mutta tulevaisuudessa ehkä paljon suuremmalta osalta. Tämä johtaa aikaisempaa laskentaintensiivisempään jalostusarvojen arviointiin ja mahdollisesti laajempaan bayeslaisen tilastotieteen käyttöön. Bayeslaisen tilastotieteen mallien ratkaiseminen vaatii usein simulaatiopohjaisten ratkaisumenetelmien käyttöä, koska ratkaiseminen vaatii moniulotteista integrointia.

Matemaattisen tilastotieteen menetelmiä käytetään siis mm. geneettisen arvon arviointiin eläinjalostuksessa. Samaa menetelmää voidaan käyttää myös kasvien ja metsien jalostuksessa. Genomitiedon matemaattiset mallit ovat samoja kaikilla geneettistä tietoa käsittelevillä aloilla, niin eläin-, kasvi- kuin ihmisgenetiikassa. Kaikilla näillä aloilla genomitieto on johtanut tarpeeseen analysoida massiivisen suuria aineistoja. Olemme tällä hetkellä keskellä genomisen aineiston vallankumousta, jossa tarve niiden analysoijista ja parempien matemaattisten mallien kehittäjistä vain lisääntyy. Vaikka siis kerron tässä artikkelissa eläinjalostuksesta, paljon kertomastani pätee myös kasvien jalostukseen sekä osittain ihmisgenetiikkaan.

Eläinjalostus on hyvin kansainvälistä ja avointa. Aivan kuten voit ostaa koiran ulkomailta, niin voit ostaa esim. hevosen tai tuotantoeläimen. Aina kun ostat eläimen,

ostat samalla sen jalostushistorian geenien muodossa. Eläinjalostuksen matemaattisista malleista julkaistaan kansainvälisissä kokouksissa. Tällä tavalla matemaattiset mallit ovat kaikkien arvioitavissa, sillä geneettisten arvojen uskottavuus on erittäin tärkeää. Ethän ostaisi koiraa omistajalta, joka ei halua näyttää lääkärintodistusta koiran terveydentilasta. Eläimelle laskettu geneettinen arvo vaikuttaa eläimestä maksettavaan hintaan, joten arvojen laskutavan on oltava luotettava.

Eläimet ovat monimutkaisia biologisia kokonaisuuksia. Eläinjalostuksessa pyritään matemaattisten mallien avulla ymmärtämään tätä kokonaisuutta ja parantamaan sitä hallitusti. Uuden genomisen tiedon myötä analysoitavien aineistojen koko on kasvanut ja kasvaa edelleen. Tämä ei kuitenkaan ole uutta, sillä eläinjalostuksen matemaattiset mallit ovat olleet jatkuvassa muutoksessa ja aineistot ovat suurentuneet viimeiset 50 vuotta. Tämä tulee jatkumaan. Ehkä sinä muutat eläinjalostuksen matemaattisia malleja, sillä varmaa on, että tämän päivän malleja ei käytetä enää 2020-luvulla.

Yhteenveto. Matemaattiset ja tietotekniikan taidot eläinjalostuksessa:

- tilastotiede, erityisesti lineaariset sekamallit
- bayeslainen tilastotiede ja todennäköisyyslaskenta
- lineaarinen algebra: matriisilaskenta
- iteratiiviset menetelmät: lineaariset ja epälineaariset yhtälöryhmät
- ohjelmointitaito: esim. Matlab/octave, R, Fortran

Muu tarpeellinen tieto:

- kvantitatiivinen genetiikka
- englannin kieli
- kiinnostus biologisten ilmiöiden mallintamiseen