

Resum

La selecció genòmica està canviant profundament el mercat del boví de llet. Actualment, és possible obtenir valoracions genètiques fiables d'animals molt joves sense necessitat de disposar del fenotip propi o el de les seves filles. Per tant, la resposta genètica d'un programa genòmic ben dissenyat supera netament a la selecció tradicional.

Aquesta tesi es compon de sis capítols en els que s'estudia l'establiment de les bases per a implementar un programa de selecció genòmica en el boví de llet espanyol. Amb aquesta finalitat, s'han realitzat estudis de simulació i valoracions genòmiques amb dades reals de la primera població de referència nacional.

L'objectiu principal d'aquesta tesi és contribuir a la implementació de la selecció genòmica en el boví de llet espanyol. Els objectius específics són: (1) Estudiar alternatives de genotipat en poblacions reduïdes de boví lleter. (2) Desenvolupar i validar metodologia per a l'avaluació de grans quantitats de genotips. (3) Estudiar l'efecte dels processos d'imputació de genotips en l'habilitat predictiva dels genotips resultants.

Les principals qüestions relacionades amb la selecció genòmica en boví lleter van ser discutides el capítol 1 incloent: aspectes estadístics i genètics en què es basa la selecció genòmica, disseny de poblacions de referència adequades, revisió de la metodologia desenvolupada per a l'avaluació, disseny i metodologia de programes d'imputació i implementació de la selecció genòmica en boví de llet a nivell de programa de selecció, centre d'inseminació i granja comercial. La selecció genòmica està revolucionant el mercat del boví de llet, ja que és possible aconseguir valors genètics molt més precisos d'animals joves, en comparació amb els obtinguts mitjançant índexs de pedigrí tradicionals. Aquesta millora està modificant un dels principis tradicionals del mercat de boví de llet com era la preferència d'ús de toros amb altes fiabilitats respecte animals amb valors genètics *a priori* superiors.

En el capítol 2 es va realitzar un estudi de simulació comparant estratègies de genotipat selectiu en poblacions de femelles enfront de l'ús de selecció tradicional o selecció genòmica amb una població de referència de mascles. La població espanyola estava formada per una mica més de 1,600 toros amb prova de progènie. Aquest mida no és, en principi, suficient per obtenir prediccions genòmiques d'alta fiabilitat. Per tant, calia avaluar diferents alternatives per incrementar l'habilitat predictiva de les avaluacions. Les estratègies que inclouen el genotipat com a població de referència dels animals en ambdós extrems de la distribució permetien millorar la precisió de l'avaluació. Els resultats usant 1,000 genotips van ser 0.50 per al caràcter de baixa heretabilitat i 0.63 per al d'heretabilitat mitjana quan la variable dependent fou el fenotip ajustat. Quan varen usar-se valors genètics com a variable dependent, les correlacions van ser 0.48 i 0.63, respectivament. Per als mateixos caràcters, una població de 996 mascles va obtenir correlacions de 0.48 i 0.55 en les prediccions posteriors. L'estudi conclou que l'estratègia

de genotipat que proporciona la major correlació és la que inclou les femelles de les dues cues de la distribució de fenotips. D'altra banda es fa evident que la mera inclusió de les femelles d'èlit, que són les habitualment genotipades, produeix resultats molt pobres en la predicció de valors genòmics.

En el capítol 3, el Random Boosting és comparat amb altres mètodes d'avaluació genòmica utilitzant metodologia Bayessiana (Bayes-A i LASSO Bayessià) i amb un G-BLUP usant la matriu genòmica. La població de referència espanyola va ser utilitzada per comparar aquests mètodes en termes de precisió i biaix. Les prediccions genòmiques van ser més precises que l'índex de pedigrí tradicional a l'hora de predir els resultats de futurs test de progènie. Els guanys obtinguts en precisió derivats de l'ús de la selecció genòmica depenen del caràcter avaluat i varien entre 0.04 i 0.42 unitats de correlació de Pearson. Els resultats promig entre caràcters demostraren que el LASSO Bayessià va obtenir majors correlacions superant al Random Boosting, Bayes-A i BLUP genòmic en 0.01, 0.03 i 0.03 unitats, respectivament. Les prediccions obtingudes amb el LASSO també van mostrar menys desviacions respecte la mitja, 0.02, 0.03 i 0.10 menys que Bayes-A, R Boost i G-BLUP, respectivament. Les prediccions usant Random Boosting van obtenir coeficients de regressió més propers a la unitat que la resta de mètodes i els errors mitjans quadràtics van ser un 2%, 10% i 12% inferiors als obtinguts a partir del B-LASSO, Bayes-A i G-BLUP, respectivament. L'estudi conclou que el Random Boosting és una metodologia aplicable en selecció genòmica i competitiva en termes d'habilitat predictiva.

En el capítol 4 l'algoritme de machine learning Random Boosting avaluat en el capítol 3 és descrit i implementat per a selecció genòmica i adaptat a l'avaluació eficient de grans bases de dades. Després de la incorporació al consorci Eurogenomics, el programa genòmic espanyol va passar a disposar de més de 22,000 toros provats com a població de referència. Es va fer necessària doncs, l'implementació d'un mètode capaç d'avaluar aquest gran conjunt de dades en un temps raonable. El nou algoritme anomenat Random Boosting realitza de forma seqüencial una selecció aleatòria d'SNPs a cada iteració sobre els quals s'aplica un predictor feble. L'algoritme va ser avaluat sobre les dades reals de boví de llet emprades en el capítol 3 i van estudiar-se més en profunditat el comportament dels paràmetres de sintonització. Aquesta proposta de modificació del Boosting permet obtenir prediccions sense pèrdua de precisió ni increments de biaix emprant només un 1% del temps de computació original.

En el capítol 5 s'avalua l'efecte d'usar genotips de baixa densitat imputats mitjançant el programari Beagle pel que fa a la seva posterior habilitat predictiva quan aquests són incorporats a la població de referència. Amb aquesta finalitat, es varen utilitzar dos mètodes d'avaluació: Random Boosting i un BLUP amb matriu genòmica. Animals dels que s'en

coneixia els SNPs inclosos en els xips GoldenGate Bovine 3K i BovineLD BeadChip varen ser imputats fins a conèixer els SNP's inclosos en el BovineSNP50v2 BeadChip. Posteriorment, un segon procés d'imputació va permetre obtenir els SNP's inclosos en el BovineHD BeadChip. Els genotipats a baixa densitat després de ser imputats, van obtenir similar capacitat predictiva que els originals en densitat 50K. Tanmateix, només es va obtenir una petita millora (en 0.002 unitats de Pearson) a l'imputar HD. El major increment es va obtenir per a dies oberts on les correlacions en el grup de validació varen augmentar en 0.06 unitats de Pearson quan es van emprar els genotips imputats a HD. En funció de la densitat de genotipat, l'algoritme Random Boosting mostra més diferències que el BLUP genòmic. Ambdós mètodes varen obtenir resultats similars tret del cas d' percentatge de greix, on les prediccions obtingudes amb el Random Boosting varen ser superiors a les del G-BLUP en 0.20 unitats de correlació de Pearson. L'estudi conclou que la capacitat predictiva d'alguns caràcters pot millorar imputant la població de referència a HD i usant mètodes d'avaluació que siguin capaços d'adaptar-se a les diferents arquitectures genètiques possibles.

Finalment, en el capítol 6 es duu a terme una discussió general dels estudis presentats en els capítols anteriors que s'enllacen amb la implementació de la selecció genòmica en el boví lleter espanyol, desenvolupada paral·lelament a aquesta tesi doctoral. La primera població de referència, amb uns 1,600 toros, va ser avaluada en el capítol 4 i va ser usada per comparar els diferents mètodes i escenaris proposats en els capítols 3, 4 i 5. La primera avaluació genòmica obtinguda per als caràcters inclosos en el capítol 4 d'aquesta tesi va estar disponible per als centres d'inseminació inclosos en el programa al mes de setembre del 2011. La població d'Eurogenomics es va incorporar al novembre del mateix any, completant la primera avaluació per als caràcters inclosos en l'índex de selecció ICO al febrer de 2012 emprant el Random Boosting descrit en el capítol 3. El maig de 2012 les avaluacions del caràcter Proteïna van ser validades per INTERBULL i finalment el 30 novembre 2012 les primeres avaluacions genòmiques oficials van ser publicades on-line per la federació de ramaders CONAFE (<http://www.conafe.com/noticias/20121130a.htm>).