

Tilføjelse af SNP fra MD-chippen til markørsættet anvendt i den genomiske evaluering

Af Anders Glasius

Formålet med aktiviteten er at udnytte viden om stykker af genomet med stor effekt. Et nyligt studie på AU har fundet, at det ikke giver værdi at lægge større vægt på bestemte områder af genomet. Det betyder stadig, at det er vigtigt at inddrage SNP, der ikke tidligere har indgået i avlsarbejdet. Der er derfor udviklet procedurer for udtræk og klargøring af data. De hidtidige rutiner for SNP-håndtering er således blevet tilpasset og de nye data vil blive inddraget i avlsværdiurderingen af NAV i 2021.

Udtræk af data

Først tages markørerne fra den fælles del af MD_v2-chippen og summeres med markørsættene for de 3 racer, RDC, Holstein og Jersey i den genomiske evaluering. For dette nye summerede sæt af markører udtrækkes genotyper fra BC|SNPmax og på baggrund af dette datasæt og det nye markørsæt dannes nye filer til brug i FImpute for at beregne Minor Allele Frequencies (MAF). Så køres FImpute for de 3 racer og markører der ikke opfylder kravet på $MAF \geq 0,1\%$ fjernes. Dernæst dannes nye inputfiler til FImpute og der køres igen med det reducerede markørsæt og output datasæt gøres klar til brug i den genomiske evaluering for at undersøge hvilke markører der rent faktisk bidrager med information ud fra det nye markørsæt udvidet med markører fra MD-v2-chippen med $MAF \geq 0,1\%$.

Resultater for Jersey

Der er 41.897 markører i det nuværende markørsæt i den genomiske evaluering og plus markørerne fra MD_v2-chippen giver det 58.695 markører i første omgang FImpute. Herefter sættes krav til MAF på 0,1% og det fjerner 263 markører fra det nuværende markørsæt og der kan tilføjes 6.104 af de nye markører fra MD_v2 chippen, hvilket giver 47.738 markører der skal tjekkes om de bidrager med information.

MAF 0,1 pct

	MD_v2 chip	Nuværende markørsæt	Antal
MAF $\geq 0,1$ pct	6.104	41.634	47.738
MAF < 0,1 pct	10.694	263	10.957
Antal	16.798	41.897	58.695

Resultater for Holstein

Der er 46.342 markører i det nuværende markørsæt i den genomiske evaluering og plus markørerne fra MD_v2-chippen giver det 58.597 markører i første omgang FImpute. Herefter sættes krav til MAF på 0,1% og det fjerner 311 markører fra det nuværende markørsæt og der kan tilføjes 4.832 af de nye markører fra MD_v2 chippen, hvilket giver 50.863 markører der skal tjekkes, om de bidrager med information.

MAF 0,1 pct

	MD_v2 chip	Nuværende markørsæt	Antal
MAF $\geq 0,1$ pct	4.832	46.031	50.863
MAF < 0,1 pct	7.423	311	7.734
Antal	12.255	46.342	58.597

Resultater for RDC

Der er 4.6914 markører i det nuværende markørsæt i den genomiske evaluering og plus markørerne fra MD_v2-chippen giver det 58.722 markører i første omgang Fimpute. Herefter sættes krav til MAF på 0,1% og det fjerner 162 markører fra det nuværende markørsæt og der kan tilføjes 5.200 af de nye markører fra MD_v2 chippen, hvilket giver 51.952 markører, der skal tjekkes for hvorvidt de bidrager med information.

MAF 0,1 pct

	MD_v2 chip	Nuværende markørsæt	Antal
MAF $\geq$ 0,1 pct	5.200	46.752	51.952
MAF < 0,1 pct	6.608	162	6.770
Antal	11.808	46.914	58.722

Afslutning

Resultaterne viser, at inddragelse af ekstra SNP vil øge datagrundlaget med 10-15%. Dette er en væsentlig forøgelse, som kan give mere information til beregningen af genomiske avlsværdital