

¹Appendix: Cross Validation Procedure for Genomic Selection

²The overall 5-fold cross validation procedure proceeds as follows.

```

3 Randomly split the data into 5 groups of equal size.
4 for  $i = 1, \dots, 5$  do
5   Set the  $i$ th group as test set (20% of the full dataset) and the remaining
6   groups as training set (80% of the full data).
7   if the model is a multi-trait regularized model then
8     Hold out 20% of the training set.
9     Fit the model on the rest of the training set for various regularization
10     parameters choices.
11     Pick the "best" regularization parameters, i.e., the choice with highest
12     prediction accuracy on the holdout set, where the prediction accuracy of
13     a given choice of regularization parameters is the Pearson correlation
14     coefficient between the observed phenotypes in the holdout set and the
15     GEBVs obtained using the estimated  $B$ ,  $C_g$  and  $C_e$  for that choice .
16     Refit the model with the "best" choice on the full training set to obtain
17     estimates of  $B$ ,  $C_g$  and  $C_e$ .
18   end
19 else
20   | Fit the model on the training set to obtain estimates of  $B$ ,  $C_g$  and  $C_e$ .
21 end
22 Use the estimates of  $B, C_g$  and  $C_e$  to obtain GEBVs in the test set.
23 Compute the prediction accuracy by calculating the Pearson correlation
24 coefficient between the observed phenotypes in the test set and the GEBVs.
25 end

```

Algorithm 1: Description of the 5-Fold Cross-Validation Procedure

²⁶In our experiments, the above procedure is repeated 50 times.

²⁷**Author details**

²⁸**References**