

# Single-stage and two-stage approaches to genomic selection.

[Schulz-Streeck T](#), [Ogutu JO](#), [Piepho HP](#).

## Abstract

Genomic selection (GS) is a method for predicting breeding values of plants or animals using many molecular markers that is commonly implemented in two stages. In plant breeding the first stage usually involves computation of adjusted means for genotypes which are then used to predict genomic breeding values in the second stage. We compared two classical stage-wise approaches, which either ignore or approximate correlations among the means by a diagonal matrix, and a new method, to a single-stage analysis for GS using ridge regression best linear unbiased prediction (RR-BLUP). The new stage-wise method rotates (orthogonalizes) the adjusted means from the first stage before submitting them to the second stage. This makes the errors approximately independently and identically normally distributed, which is a prerequisite for many procedures that are potentially useful for GS such as machine learning methods (e.g. boosting) and regularized regression methods (e.g. lasso). This is illustrated in this paper using componentwise boosting. The componentwise boosting method minimizes squared error loss using least squares and iteratively and automatically selects markers that are most predictive of genomic breeding values. Results are compared with those of RR-BLUP using fivefold cross-validation. The new stage-wise approach with rotated means was slightly more similar to the single-stage analysis than the classical two-stage approaches based on non-rotated means for two unbalanced datasets. This suggests that rotation is a worthwhile pre-processing step in GS for the two-stage approaches for unbalanced datasets. Moreover, the predictive accuracy of stage-wise RR-BLUP was higher (5.0-6.1%) than that of componentwise boosting.