Determination of the optimal number of evaluations in half-sib progenies of kale by Bayesian approach

Alcinei M Azevedo¹; Derly José H da Silva²; Rogério Seus²; Elis Marina de Freitas²; Daniel F Afonso²; Cristiane L Gomes²; Ignácio Aspiazu³

¹Universidade Federal de Minas Gerais, Instituto de Ciências Agrárias (UFMG), Montes Claros-MG, Brasil; alcineimistico@hotmail.com; ²Universidade Federal de Viçosa (UFV), Viçosa-MG, Brasil; derly@ufv.br; rseusagro@hotmail.com; elis_stavano@hotmail.com; danielferreirafonso@gmail.com; cristiane.lourenco@ufv.br; ³Universidade Estadual de Montes Claros (UNIMONTES), Janaúba-MG, Brasil; aspiazu@gmail.com

ABSTRACT

Kale has a long vegetative cycle, requiring a lot of labor, due to the need for tutoring, thinning and multiple harvests, leading to difficulties in the maintenance and evaluation of experiments. Thus, the objective was to estimate the minimum number of evaluations for the assertive selection of half-sib progenies of kale by means of a repeatability study by Bayesian approach. Twenty four half-sib progenies were evaluated in a randomized block design with four replicates and five plants per plot. The number of shoots, number of marketable leaves, fresh mass of marketable leaves and fresh mass per leaf were measured throughout 15 harvests. All traits showed high estimates of the repeatability, indicating high regularity in the expression of the traits during the harvesting period. With eight harvests it is possible to evaluate all the traits with a coefficient of determination superior to 85% in half-sib progenies of kale.

Keywords: Brassica oleracea var. acephala, repeatability, genetic improvement, production of leaves.

Received on June 29, 2020; accepted on October 16, 2020

RESUMO

Determinação do número ótimo de avaliações em progênies de meios-irmãos de couve por abordagem Bayesiana

A couve possui ciclo vegetativo longo, requerendo muita mão-de-obra, pela necessidade de tutoramento, desbrotas e colheitas múltiplas, trazendo dificuldades na manutenção e avaliação de experimentos. Assim, objetivou-se estimar o número mínimo de avaliações para a seleção assertiva de famílias de meios-irmãos de couve por meio do estudo de repetibilidade, utilizando-se para isso a abordagem bayesiana. Foram avaliadas 24 famílias de meios-irmãos de couve no delineamento em blocos casualizados com quatro repetições e cinco plantas por parcela. O número de brotações, número de folhas comerciais, massa fresca de folha e massa fresca por folha foram avaliadas em 15 colheitas. Todas as características tiveram altas estimativas do coeficiente de repetibilidade, indicando elevada regularidade na expressão das características avaliadas ao longo do período de colheitas. Com 8 colheitas é possível avaliar todas as características com um coeficiente de determinação superior a 85% em progênies de meios-irmãos de couve.

Palavras-chave: Brassica oleracea var. acephala, repetibilidade, melhoramento genético, produção de folhas.
can be used advantageously, since it allows obtaining the a posteriori distribution and credibility intervals of the estimated parameters (Gonçalves-Vidigal et al., 2008; Mathew et al., 2012; Rodovalho et al., 2014). This makes the technique very informative (Mathew et al., 2012) and facilitates hypothesis testing. In addition, Bayesian inference enables the evaluation of experiments with unbalanced data and the study of complex statistical models (Waldmann & Ericsson, 2006; Bink et al., 2007). Consequently, its use is increasing among breeders, not only for the analysis of molecular data, but also for phenotypic data (Waldmann & Ericsson, 2006; Omer et al., 2016). In this sense, the objective was to estimate the minimum number of evaluations for the study of complex statistical models (Azevedo et al., 2017). Assuming that $e_{ij} \sim N(0, \sigma^2_e)$, the sampling distribution of the observed data (function of maximum likelihood) is:

$$y_{ij} \mid m, g_i, a_j, \sigma^2_e \sim N(m + g_i + a_j, \sigma^2_e).$$

For the location parameters, the following a priori distributions were considered:

$$m \mid u_m, \sigma^2_m \sim N(u_m, \sigma^2_m).$$
$$g_i \mid u_g, \sigma^2_g \sim N(u_g, \sigma^2_g)$$
$$a_j \mid u_a, \sigma^2_a \sim N(u_a, \sigma^2_a)$$

For the variance components, the inverse chi-squared distribution was considered as the a priori distribution:

$$\sigma^2_e \mid \nu_g, s_g \sim \nu_g s_g X^{-2}_{\nu_g},$$
$$\sigma^2_a \mid \nu_a, s_a \sim \nu_a s_a X^{-2}_{\nu_a}$$
$$\sigma^2_m \mid \nu_e, s_e \sim \nu_e s_e X^{-2}_{\nu_e}$$

Therefore, the joint a posteriori distribution can be represented by:

$$P(\theta \mid y_{ij}) \propto \prod_{i=1}^{24} \prod_{j=1}^{15} \exp \left( \frac{y_{ij} - (m + g_i + a_j)^2}{2\sigma^2_e} \right) (\sigma^2_e)^{-0.5} \exp \left( \frac{(m - u_m)^2}{2\sigma^2_m} \right) (\sigma^2_m)^{-0.5} \exp \left( \frac{(g_i - u_g)^2}{2\sigma^2_g} \right) (\sigma^2_g)^{-0.5} \exp \left( \frac{(a_j - u_a)^2}{2\sigma^2_a} \right) (\sigma^2_a)^{-0.5} \exp \left( \frac{\nu_g s_g}{2\nu_{\nu_g}} \right) \left(\frac{\nu_g}{\nu_g + \nu_g} \right) (\sigma^2_g)^{-0.5} \exp \left( \frac{\nu_a s_a}{2\nu_{\nu_a}} \right) \left(\frac{\nu_a}{\nu_a + \nu_a} \right) (\sigma^2_a)^{-0.5} \exp \left( \frac{\nu_e s_e}{2\nu_{\nu_e}} \right) \left(\frac{\nu_e}{\nu_e + \nu_e} \right) (\sigma^2_e)^{-0.5} \exp \left( \frac{\nu_{\nu_g} s_g}{2\nu_{\nu_g}} \right) \left(\frac{\nu_{\nu_g}}{\nu_{\nu_g} + \nu_{\nu_g}} \right) (\sigma^2_g)^{-0.5} \exp \left( \frac{\nu_{\nu_a} s_a}{2\nu_{\nu_a}} \right) \left(\frac{\nu_{\nu_a}}{\nu_{\nu_a} + \nu_{\nu_a}} \right) (\sigma^2_a)^{-0.5} \exp \left( \frac{\nu_{\nu_e} s_e}{2\nu_{\nu_e}} \right) \left(\frac{\nu_{\nu_e}}{\nu_{\nu_e} + \nu_{\nu_e}} \right) (\sigma^2_e)$$

The MCMC chains, 1,000,000 iterations per characteristic were established. Burnin of 100,000 iterations and thin of 500 iterations was used, resulting in a total sample of 1,800 iterations for each characteristic. After confirming the convergence by the Geweke test ($p > 0.05$) and the absence of autocorrelation, the residual coefficient of variation $CV_e = \frac{100}{m} \sqrt{\sigma^2_e}$, coefficient of repeatability $r = \frac{\sigma^2_e}{(\sigma^2_m + \sigma^2_g + \sigma^2_a)}$, coefficient of determination $R^2 = 100 \times \frac{r}{1 + r (15 - 1)}$ and the optimal number of harvests $n_o = R^2_0 (1 - r)/[1 + r (15 - 1)]$ were estimated. For all the parameters $(\sigma^2_a, \sigma^2_g, \sigma^2_m, m, CV_e, r, R^2)$ the HPD 95% (high probability density) interval and mode were estimated with the aid of the Bayesian Output Analysis (BOA) package of the software R.

### Results and Discussion

Only the characteristic number of shoots showed overlapping of the HPD intervals for the variance components due to the effects of family and evaluation (Table 1). For the other traits, there were higher estimates of the variance components due to the effects of the evaluations in relation to the family effects, without overlapping the HPD intervals. The obtained asymmetric HPD for the variance components and genetic parameters are a peculiarity
of Bayesian inference and facilitate hypothesis testing (Azevedo et al., 2017).

The higher magnitudes of the variance components of the effects of the evaluations compared to the genetic effects (progenies), without overlapping the credibility interval, indicates the predominance of the evaluation effects when compared to the genetic effects. This happened for number of leaves, fresh leaf mass and fresh mass per leaf. The higher magnitudes of the variance components of the effects of the evaluations compared to the genetic effects were also verified by Brito et al. (2019) when evaluating half-sib kale. The higher magnitudes of the residual coefficient of variation for number of shoots and fresh leaf mass shows that these traits are more influenced by random effects of the environment (experimental error).

The highest mode for the coefficient of residual variation was found for the number of shoots (19.10%), and its HPD interval only did not overlap those values found for the number of leaves and fresh mass per leaf, which presented the lower modes (8.17 and 10.48%, respectively). There were overlapping HPD intervals between the coefficients of determination of all the traits. The

| Characteristics          | Parameters   | HPD          | Means   | Median | Mode |
|--------------------------|--------------|--------------|---------|--------|------|
|                          |              | 2.50% | 97.50% |        |      |      |
| Number of shoots         | $\sigma^2_e$ | 0.30  | 1.53   | 0.70   | 0.63 | 0.55 |
|                          | $\sigma^2_g$ | 0.37  | 1.28   | 0.70   | 0.66 | 0.57 |
|                          | $\sigma^2_a$ | 0.47  | 0.64   | 0.55   | 0.55 | 0.54 |
|                          | $m$          | 3.26  | 4.35   | 3.80   | 3.80 | 3.88 |
|                          | $CV_e(\%)$   | 16.73 | 23.32  | 19.60  | 19.43| 19.10|
|                          | $R^2(\%)$    | 90.89 | 97.28  | 94.58  | 94.74| 95.31|
| Number of leaves         | $\sigma^2_e$ | 0.58  | 2.67   | 1.25   | 1.13 | 0.93 |
|                          | $\sigma^2_g$ | 0.17  | 0.56   | 0.31   | 0.29 | 0.26 |
|                          | $\sigma^2_a$ | 0.14  | 0.18   | 0.16   | 0.16 | 0.16 |
|                          | $m$          | 4.19  | 5.42   | 4.82   | 4.83 | 4.88 |
|                          | $CV_e(\%)$   | 7.15  | 9.68   | 8.27   | 8.22 | 8.17 |
|                          | $R^2(\%)$    | 93.97 | 98.18  | 96.42  | 96.54| 96.67|
| Fresh mass of leaves     | $\sigma^2_e$ | 1672.24| 2292.04| 1950.32| 1939.04| 1922.51|
|                          | $\sigma^2_g$ | 955.14| 3387.68| 1852.92| 1738.61| 1616.06|
|                          | $m$          | 225.76| 373.76 | 308.16 | 310.71| 314.87|
|                          | $CV_e(\%)$   | 11.59 | 19.84  | 14.56  | 14.19| 13.75|
|                          | $R^2(\%)$    | 87.66 | 96.44  | 92.82  | 92.07| 93.28|
| Fresh mass per leaf      | $\sigma^2_e$ | 198.53| 950.31 | 431.65 | 384.70| 345.01|
|                          | $\sigma^2_g$ | 20.29 | 73.49  | 39.46  | 36.91| 31.77|
|                          | $m$          | 61.39 | 83.11  | 71.71  | 71.49| 71.12 |
|                          | $CV_e(\%)$   | 8.86  | 12.65  | 10.59  | 10.53| 10.48|
|                          | $R^2(\%)$    | 83.82 | 95.10  | 90.39  | 90.72| 91.37|

Table 1. HPD (high probability density) interval, mean, median and mode for the variance components of the evaluations ($\sigma^2_e$), progenies ($\sigma^2_g$), error ($\sigma^2_e$), for the general means ($m$), residual coefficient of variation ($CV_e$) and coefficient of determination ($R^2$) in half-sib progenies of kale. Viçosa, UFV, 2020.
traits with higher mode values for this parameter were number of leaves and number of shoots (96.67 and 95.31%, respectively).

The lowest mode for repeatability was found for fresh matter per leaf, followed by fresh matter of leaves, with mode values of 0.40 and 0.47 (Figure 1). However, the number of leaves and number of shoots showed the highest estimates of repeatability, with mode values of 0.65 and 0.55, respectively. However, there was an overlap of the HPD interval of repeatability in all traits. The repeatability estimation can vary from 0 to 1, and high coefficients allow to predict the real value for a given characteristic with few measurements (Oliveira & Moura, 2010). The highest estimates for the repeatability coefficient of the number of marketable leaves (Figure 1) were also found by Brito et al. (2019). This indicates for this trait, a smaller increase in the experimental accuracy due to the increase in the number of evaluations (Della Bruna et al., 2012). On the other hand, the trait fresh mass per leaf, with smaller estimates, requires a greater number of harvests for a selection with greater efficiency and reliability.

The number of wished harvests according to the used coefficient of determination indicates that the traits mass of fresh matter per leaf and mass of fresh leaves require a greater number of evaluations for the efficient selection of progenies (Figure 2). This can be justified by the percentage of water in the leaves that can vary at the time of harvest, due to variations in soil moisture, relative humidity or temperature. For this, 13 harvests are required to guarantee the coefficient of determination of 90%, and eight harvests to reach a coefficient of determination of 85%. This information is important, and indicates that in future experiments with half-sib progenies of kale, it is possible to have considerable precision with only eight harvests. This number of harvests is much higher than that found by Azevedo et al. (2012), which suggest only three harvests to obtain a coefficient of variation higher than 95%. Among the justifications for the need of a smaller number of harvests found by these authors stands out the fact that they evaluated kale clones. In the present work, the genetic variability within each treatment (half-sib progenies) may have contributed to the lower repeatability coefficients. This justification agrees with the work

Figure 1. Graphical representation of HPD 95% (high probability density) interval, mode and density of the a posteriori distribution of coefficients of repeatability for the number of shoots (A), number of leaves (B), fresh mass of leaves (C) and fresh mass per leaf (D) in half-sib progenies of kale. Viçosa, UFV, 2020.
done by Cruz et al. (2012), in which the authors state that the repeatability coefficient may vary according to the genetic structure of the studied population (clones, half-sib progenies, complete-sib progenies). The higher number of measurements required for the evaluation of fresh mass per leaf may be due to a higher interaction between genotypes and temporary environment in these traits. A possible cause for this interaction may be the regulation of the character by different gene sets, which may be more or less active, depending on the developmental state of the individual (Cruz et al., 2012). More than 8 harvests were also necessary to obtain the determination coefficient greater than 85% by Brito et al. (2019) to half-sib progeny.

On the other hand, the number of leaves is the characteristic that requires a smaller number of evaluations, followed by the number of shoots. From the mode of the a posteriori distribution, it is estimated that three, two, four and six evaluations of the traits number of shoots, number of marketable leaves, fresh matter of leaves and fresh matter per leaf are required, respectively, if a coefficient of determination of 80% is desired (Figure 2). To obtain a coefficient of determination of 85%, four, three, six and eight evaluations are required for the number of shoots, number of marketable leaves, fresh matter of leaves and fresh matter per leaf, respectively. To achieve the 90% coefficient of determination, seven, five, ten and 12 evaluations are required, respectively. To obtain the coefficient of 95%, 14, 10, 20 and 26 evaluations are necessary, respectively.

In breeding programs, this information is important, as it permits knowing the minimum number of evaluations to compare genotypes (Patcharin et al., 2013). This allows avoiding the loss of time with evaluations beyond necessary, also avoiding evaluation for a very short period, which can lead to errors in the identification of the superior genotypes (Neves et al., 2010).

Therefore, it can be concluded that the number of leaves is the characteristic with higher repeatability, as opposed to the fresh mass per leaf, which requires a higher number of harvests for the selection of better half-sib progenies of kale. With eight harvests it is possible to evaluate all the traits with a coefficient of determination superior to 85% in

Figure 2. Estimation of mode, lower (IL) and upper (US) limits of the HPD 95% (high probability density) interval of the a posteriori distribution of the ideal number of harvests as a function of the different coefficients of determination required for number of shoots (A), number of leaves (B), fresh mass of leaf (C) and fresh mass per leaf (D) in half-sib progenies of kale. Viçosa, UFV, 2020.
Determination of the optimal number of evaluations in half-sib progenies of kale by Bayesian approach

half-sib progenies of kale.

ACKNOWLEDGEMENTS

Our thanks to CAPES (Council for Improvement of Personnel in Higher Education - Finance code 001), to FAPEMIG (Minas Gerais State Research Support Foundation) and CNPq (National Council for Scientific and Technological Development) for their support for this study.

REFERENCES

AGBAJE, EO; OKPARA, CS. 2013. Antitumor activity of aqueous extract of fresh leaf of Brassica oleracea linn. Var. Acephala (d.c) alef (Brassicaceae). International Research Journal of Pharmacy 4: 107-111.

AYAZ, FA; HAYIRLIOGLU-AYAZ, S; ALPAY-AYAZ, FA; HAYIRLIOGLU-AYAZ, S; GRÚZ, J; V ALENTOVÁ, K; ULRICHOVÁ, J; STRNAD, M. 2008. Phenolic acid contents of kale (Brassica oleracea L. var. acephala DC.) extracts and their antioxidant and antibacterial activities. Food Chemistry 107: 19-25.

AZEVEDO, AM; ANDRADE JÚNIOR, VC; PEDROSA, CC; FERNANDES, JSC; VALADARES, NR; FERREIRA, MRA; MARTINS, RA V . 2012. Desempenho de materiais e variabilidade genética em Panicum maximum l. var. acephala DC. Revista Brasileira de Zootecnia 36: 240-250.

AZEVEDO, AM; ANDRADE JÚNIOR, VC; SANTOS, AA; SOUSA JÚNIOR, AS; FERREIRA, MAM. 2017. Population parameters and selection of kale genotypes using Bayesian inference in a multi-trait linear model. Acta Scientiarum. Agromon 39: 25-31.

BINK, CAM; BOER, MP; TER BRAAK, CJF; CANSEN, J; VOORRIPS, RE; VAN DE WEGWE. 2007. Bayesian analysis of complex traits in pedigreed plant populations. Euphytica 161: 85-96.

BRITO, GO; ANDRADE JÚNIOR, VC; AZEVEDO, AM; DONATO, LMS; SILVA, LR; FERREIRA, MAM. 2019. Study of repeatability and phenotypical stabilization in kale using frequentist, Bayesian and bootstrap resampling approaches. Acta Scientiarum. Agronomy 41: e42606.

CRUZ, CD; REGAZZI, AJ; CARNEIRO, PCS. 2012. Modelos biométricos aplicados ao melhoramento genético. Viçosa: UFV. 514p.

DELLA BRUNA, E; MORETO, AL; DALBÓ, MA. 2012. Uso do coeficiente de repetibilidade na seleção de clones de peruaquê para o litoral sul de Santa Catarina. Revista Brasileira de Fruticultura 34: 206-215.

GONÇALVES-VIDIGAL, MC; MORA, F; BIGNOTTO, TS; MUNHOZ, REF; SOUZA, LD. 2008. Heritability of quantitative traits in segregating common bean progenies using a Bayesian approach. Euphytica 164: 551-560.

KAULMANN, A; JONVILLE, MC; SCHNEIDER, YJ; HOFFMANN, L; BOHN, T. 2014. Carotenoids, polyphenols and micronutrient profiles of Brassica oleracea and plum varieties and their contribution to measures of total antioxidant capacity. Food chemistry 155: 240-250.

KORUS, A. 2011. Level of vitamin C, polyphenols, and antioxidant and enzymatic activity in three varieties of kale (Brassica oleracea l. var. acephala) at different stages of maturity. International Journal of Food Properties 14: 1069-1080.

MARTUSCELLO, JA; JANK, L; FONSECA, DM; CRUZ, CD; CUNHA, DNFV. 2007. Repetibilidade de caracteres agronômicos em Panicum maximum Jacq. Revista Brasileira de Zootecnia 36: 1975-1981.

MATHEW, B; BAUER, AM; KOISTINEN, P; REETZ, TC; LÉON, J; SILLANPAA, MJ. 2012. Bayesian adaptive Markov chain Monte Carlo estimation of genetic parameters. Heredity 109: 235-245.

NEVES, LG; BRUCKNER, CH; CRUZ, CD; BARELLI, MAA. 2010. Avaliação da repetibilidade e melhoramento de famílias de maracujazeiro. Revista Ceres 57: 480-485.

OLIVEIRA, MSP; MOURA, EF. 2010. Repetibilidade e número mínimo de medições para caracteres de cacho de babacu (Oenocarpus mapora). Revista Brasileira de Fruticultura 32: 1173-1179.

OMER, OS; ABDALLA, AWH; MOHAMMED, MH; SINGH, M. 2016. Bayesian estimation of genotypic and phenotypic correlations from crop variety trials. Crop Breeding and Applied Biotechnology 16: 14-21.

PATCHARIN, T; YAOANWAT, H; PUNTAREE, T; PEERASAK, S. 2013. Estimates of repeatability and path coefficient of bunch and fruit traits in bang boet Dura oil palm. Journal of Oil Palm Research 25: 108-115.

R DEVELOPMENT CORE TEAM. 2020. R: A Language and Environment for Statistical Computing. Áustria. http://www.R-project.org/

RODOVALHO, M; MORA, F; ARRIAGADA, O; MALDONADO, C; ARNOLD, E; SCAPIM, CA. 2014. Genetic evaluation of popcorn progenies using a Bayesian approach via the independence chain algorithm. Crop Breeding and Applied Biotechnology 14: 261-265.

SIKORA, E; CIEŚLIK, E; LESZCZYŃSKA, T; FILIPIAK-FLOKIEWICZ, A; PISULEWSKI, PM. 2008. The antioxidant activity of selected cruciferous vegetables subjected to aquathermal processing. Food Chemistry 107: 55-59.

SOENGAS, P; SOTELO, T; VELASCO, P; CARTEA, ME. 2011. Antioxidant properties of Brassica vegetables. Functional Plant Science and Biotechnology 5: 43-55.

TENKOUANO, A; ORTIZ, R; NOKOE, KS. 2012. Repeatability and optimum trial configuration for field-testing of banana and plantain. Scientia Horticulturae 140: 39-44.

VILAR, M; CARTEA, ME; PADILLA, G. 2008. The potential of kales as a promising vegetable crop. Euphytica 159: 153-165.

WALDMANN, P; ERICSSON, T. 2006. Comparison of REML and Gibbs sampling estimates of multi-trait genetic parameters in Scots pine. Theoretical and Applied Genetics 112: 1441-1451.