FUNCTIONAL PLANT BREEDING JOURNAL

Scientific Journal. ISSN 2595-9433
Volume 5, Article n. 7, Jan/December
D.O.I. http://dx.doi.org/10.35418/2526-4117/v5a7
Received: 12/16/2022 - Accepted: 03/04/2022

INTRAPOPULATION RECURRENT SELECTION STRATEGIES IN PLANT BREEDING

Magno Antonio Patto Ramalho[1], Thaise da Silva Souza[2]*, Vitor Passos da Silva Junior[3]

[1] *Federal University of Lavras, Department of Biology, Lavras, Minas Gerais State, Brazil;*
[2] *Corteva Agriscience, Precision Phenotyping Research, Sorriso, Mato Grosso State, Brazil;*
[3] *Bracell, Eucalyptus Breeding Research, Lençóis Paulista, São Paulo State, Brazil.*
** Corresponding author: thaisefloresta@gmail.com*

Abstract: The conduction of each intrapopulation recurrent selection (IRS) cycle typically requires a few years. There are several factors, with and without the control of plant breeders, that affect the success obtained with the IRS. For it to be more efficient, such factors must be judiciously assessed. In this light, this study aimed to compare a large number of variables involved in the efficiency of the IRS. Selection gain (SG) estimates with the IRS were obtained. Considering different variables under the control of the breeders such as the following: two types of progenies, half-sib (HS) and full-sib (FS); three experiment evaluation conduction strategies with multiple tree plots (MTP), single tree plots (STP), and clonal progeny test (CPT), and other variables. Variables without the control of plant breeders were also considered: mean allele frequencies of the populations (p); level of dominance (d/a); broad-sense heritability at the individual level (h^2). Over 3000 SG estimates were obtained. When the CPT was used, the SG was always higher than that obtained with STP or MTP; the SG when employing HS was always higher than with FS. However, the employment of HS, especially when using clones for the recombination, has the serious restriction of not allowing higher selection intensities and maintaining the same number of progenies in the different IRS calculations. In this context, the employment of FS is more advisable since it allows applying a more significant selection intensity and enables more homogeneous recombination.

Keywords: plant breeding, intrapopulation recurrent selection, selection gain, clonal progeny test.

Introduction

The need for grains, fruits, and fiber to meet humanity's demands is growing. Increased productivity in agriculture has been achieved through new technologies for crop management and genetic improvement. One of the options to continue obtaining new cultivars that are better than the preexisting ones, through genetic improvement, is recurrent selection (RS). RS can be intrapopulational (IRS) or interpopulational or reciprocal (RRS). The SR was initially proposed for corn, as commented by Hallauer, Carena and Miranda Filho (2010), and was later used for several other species (Nelson et al. 2018; Rutkoski 2019).

The *Eucalyptus* genus is a group of plants that can be used as a reference in the search for strategies that increase the efficiency of RS programs. This is because it allows the use of any type of progeny in evaluations for selection and several recombination options. Thus, comparisons between RS strategies, obtained with eucalyptus, can be extrapolated, with small adjustments, to other cultivated plants. Both RS procedures (IRS or RRS) have been used in eucalyptus breeding in Brazil; however, the IRS has been used for a longer time (Pires et al. 2011). The IRS aims to improve the frequency of favorable alleles in the population through successive cycles of selection and recombination of the best individuals/progenies. The use of IRS intensified with the use of cloning, selecting the best individuals from the best progenies to be cloned.

However, it was found over time that the correlation between the individual in the progeny test and its clone in the clonal tests was low (Reis et al. 2011; Furtini et al. 2012), highlighting the need to use other strategies for the RS. One option that is currently being implemented is called clonal progeny test (CPT). Resende (2002) commented on the CPT, calling it progenies with repeated individuals through cloning. He mentioned that the use of experiments with repeat individuals by cloning was suggested by Burdon and Shelbourne in 1974 to improve the estimates of genetic parameters. Its use in genetic improvement occurred a few years later (Foster and Shaw 1988). Shelbourne (1992) showed that its efficiency was superior to other selection methods used until then. Resende (2002) presented the entire development of the CPT considering, however, the use of experimental plots with more than one plant.

Comparisons between RS conduction procedures were carried out both for autogamous plants (Cobb et al. 2019; Atlin and Econopouly 2022) and for allogamous plants (Hallauer, Carena and Miranda Filho 2010; Resende 2015), mainly through the expression of the expected gain with the selection. There are several alternatives for conducting the IRS that involve variables under the control of the breeders, such as: the type of progenies, which has been more used in allogamous plants, that is, progenies of half-siblings (HS) or Full-sib (FS); the number of progenies, replications, and plants per progeny. Numerous other variables also occur without the direct control of breeders, such as the average allele frequency of favorable alleles, the type of predominant allelic interaction, and the possible heritability to be obtained. All of these alternatives have been relatively underexplored in terms of research.

A strategy not yet explored in comparisons of RS procedures is to use as a reference the genetic variance of the population considered equal to 1.0 ($V_G = 1$) (Atlin and Econopouly 2022). This strategy has some advantages: a) the information obtained applies to any characteristic, regardless of the unit of measurement. The expected gain unit is the estimate of the of the population under selection; b) allows both the genetic variance components (V_A or V_D) and the environmental variance components to be transformed into units.

As the challenges in obtaining new cultivars are enormous, it is necessary to have maximum efficiency in the process. Based on the above, the present work was carried out with the objective of comparing IRS strategies involving the numerous variables under the control of the breeders, including the CPT, considering different allele frequencies in the population and type of allelic interactions and heritabilities.

Material and methods

Some variables were considered in the evaluations of non-inbred progenies, HS and FS. Three experimental strategies were implemented: evaluation of progenies in MTP, STP, and CPT experiments. In the different strategies, all analyses of variance will be considered at the individual level. Evidently, the analyses may be carried out using mixed models, obtaining the same variance estimates with the procedures proposed in this case. Obviously, the comparisons will be valid regardless of the procedure used in the analyses.

The basic equation used to obtain the expected gains from recurrent selection per cycle, between and within non-inbred progenies and considering that recombination is always carried out using the selected individuals (clones) was the following:

$$SG_{Total} = SG_{Between} + SG_{Within} = \quad \text{(Eq.1)}$$
$$= \left\{\left[i_B\left(\frac{1/2\,V_A}{\sqrt{V_{\bar{F}}}}\right)\right] + \left[i_W\left(\frac{1/2\,V_A}{\sqrt{V_d}}\right)\right]\right\}$$

Where i_B and i_W are the standardized selection intensities between and within the progenies, respectively, V_A is the additive genetic variance, $V_{\bar{F}}$ is the phenotypic variance between progeny means, and V_d is the phenotypic variance between individuals within the evaluated progenies.

Based on the expected mean squares E(MS), one may infer that the phenotypic variance between progeny means ($V_{\bar{F}}$) for the MTP strategy is obtained by $V_{\bar{F}}=V_d/kr+V_e/r+V_p$, with V_d and V_p varying with the type of progeny used. If $HS=V_{d_{HS}}=V_w+V_{Gd}=V_w+¾V_A+V_D$, where V_w is the environmental variance within the plots and V_{Gd} is the genetic variance between plants from different progenies. In other words, for $V_{GdHS}=¾V_A+V_D$, where V_D is the dominance variance and $V_{GdFS}=½V_A+¾V_D$. When a single-tree plot is used, $V_{\bar{F}}=V_{e^*}/r+V_p$, with the equation components already having been defined.

To obtain V_A and V_D, it was assumed that, for a locus (Bernardo, 2020), $V_A=2p(1-p)[a+(1-2p)d]^2$ and $V_D=[2p(1-2p)d]^2$, where p is the frequency of the favorable allele in the population and a and d are the contributions of the homozygous and heterozygous loci, respectively, relative to the mean of the two homozygotes in the manifestation of the trait.

To allow obtaining information at the population level, a procedure similar to that adopted by Vencovsky et al. (2010). It was considered that the difference in the allele frequency in a given population fits a Beta distribution. A random variable p with a Beta distribution has the following as its density function, where $0 < p < 1$, x and z are parameters, with $x > -1$ and $z > -1$, and Γ is the Gamma Function, with $\Gamma(x+1) = x\Gamma(x) = X!$.

(Eq.2)

$$t(p) = \frac{\Gamma(x+z+2)}{\Gamma(x+1)\,\Gamma(z+1)}\,px\,(1-p)^z$$

The mean value of p in the distribution is given by:

$$\text{(Eq.3)} \quad \bar{p} = \frac{x+1}{x+z+2}$$

Three distribution functions were obtained from the Beta distribution, simulating a little-improved population (x = 1 and z = 3), an intermediary population (x = z = 1), and an improved population (x = 3 and z = 1). From these distribution functions, $\bar{V}_D$ and $\bar{V}_A$ were obtained (mean dominance variance and mean additive variance, respectively). Hence, when $\bar{V}_D$ and $\bar{V}_A$ are mentioned in this work, the mean value of the estimate as a function of the frequency distribution is considered.

The estimates were obtained through the following estimators:

(Eq.4)

$$\bar{V}_A = \int_0^1 2p\,(1-p)[a+(1-2p)d]^2\, f(p)\,dp$$

(Eq.5)

$$\bar{V}_D = \int_0^1 [2p\,(1-p)d]^2\, f(p)\,dp$$

For each allele frequency distribution, the allelic interactions of complete dominance d = a, the partial dominance d/a = 1/2, and the absence of dominance d = 0 were considered. The proportions of the mean estimates $\bar{V}_D$ and $\bar{V}_A$ for the three populations were obtained using the same procedure adopted by Atlin and Econopouly (2022), considering the total genetic variance (V_G) equal to 1.0 (Table1).

The environmental variance was obtained from the broad-sense heritabilities (h^2) at the individual level. The h^2 values considered were 0.2, 0.4, and 0.6. Hence, considering $h^2 = 0.2$, we have the following:

$$\text{(Eq.6)} \quad h^2 = V_G/(V_E+V_G) = (V_A+V_D)/(V_E+V_A+V_D)$$

Table 1. Estimates (proportions) of mean additive genetic variance ($\bar{V}_A$) and mean dominance variance ($\bar{V}_D$) as a function of population frequency distribution. Values obtained for three populations, differing in mean allele frequencies and level of dominance (d/a). Estimates considering the population having the total genetic variance equal to 1 (V_G).

$\bar{p}$	(d/a)	$\bar{V}_A/\bar{V}_D$	$\bar{V}_D$	$\bar{V}_A$
		$2.5d^2/(6a^2 + 3ad + d^2)$*		
0.333	0.0	0	-	1.000
	0.5	5/62	0.0746	0.9264
	1.0	1/4	0.2	0.8
		$3d^2/(7a^2 + d^2)$		
0.500	0.0	0	-	1.000
	0.5	3/29	0.0937	0.9063
	1.0	3/8	0.2727	0.7272
		$2.5d^2/(6a^2 + d^2 - 3ad)$		
0.667	0.0	0	-	1.000
	0.5	5/38	0.1163	0.8837
	1.0	2.5/4	0.3846	0.6154

*Mean additive genetic variance ($\bar{V}_A$) and dominance ($\bar{V}_D$) after integrating the respective distribution function used in the estimates of the $\bar{V}_A/\bar{V}_D$.

Using the procedure proposed by Atlin and Econopouly (2022), the V_E values for the different h^2 were expressed as units of the genetic variance of the population (V_G), always considered to be equal to 1.0. For example, with $V_A + V_D = V_G = 1.0$, we have $h^2 = 0.2 = V_G/(V_G + V_E)$, hence $V_E = 4.0V_G$. In this situation, the environmental variance will be four times the genetic variance of the population under selection, with V_E being the total environmental variance of the area where the experiments will be conducted, disregarding the isolated variation among replicates, $V_e + V_W$. Since the variance within the plot (V_W) is typically far superior to V_e, three situations will be considered in obtaining the SG estimates: $V_W = 3,10$, or 20 times V_e.

In the case of the expected SG with the selection of cloned progenies ,the evaluations involve clones of the plants of each progeny to be evaluated. Hence, the simultaneous evaluation of the HS or FS and clones was considered for each individual from the progenies. In this case, the experiments were always in STP. The number of individuals evaluated was the product **n·q·r**, where **n** is the number of progenies, **q** is the number of individuals per progeny, and **r** is the number of replicates, i.e., clones of each individual.

Numerous variables were considered in the SG estimates, that is the three selection strategies MTP, STP and CPT, the two types of progenies, FS and HS, three mean allele frequencies of the populations, three types of d/a,three h^2 magnitudes, and two selection intensities between and within progenies. The approximate number of individuals being evaluated was always considered in obtaining the SG estimates. This number is the result of the product, i.e., the numbers of progenies, replicates, and plants per plot. In the case of the STP, k = 1. The total selection intensity considered was 0.01.

From the expected mean squares, considering the different estimates of the additive genetic variance ($\bar{V}_A$), dominance variance ($\bar{V}_D$), and heritabilities (h^2), the phenotypic variances between progeny means (V_F) and the phenotypic variances between individuals within the plots (V_d) were estimated for the two types of progeny used and the three selection strategies. Recalling that, in the equation for the SG, i is the tabulated value of the selection intensity (Ramalho et al. 2012). If the number of progenies or plants to be submitted to selection is lower than 50, the value of used was estimated from the following equation by Wricke and Weber (1986), where is the value of the standardized selection intensity for large populations, f is the selected proportion, and k is the number of

individuals being selected within the plot or the number of replicates in the case of STP.

$$i^* = i - (1 - f)/[2if(k + 1)] \quad \text{(Eq.7)}$$

Results

The number of estimates obtained from the expected selection gains (SG) was over 3000. In this condition, it is infeasible to present all results obtained. Due to this, we sought to choose variables that allow inferring the trend of what occurs with the others. In most situations, two conditions of the allele frequencies of the population ($\overline{p}$ = 0.333 and 0.667), two proportions of $\overline{V}_D/\overline{V}_A$ (d/a = 0 and 1), and two heritabilities (h^2 = 0.2 and 0.6) were considered.

The proportions between estimates V_A and V_D, considering different mean allele frequencies of the populations and the average level of dominance (d/a), show how it was expected that the sum of V_A and V_D is always one, as considered (Table1). Hence, in all results presented, the unit used will always be the genetic variance (V_G) of the population under selection, regardless of the trait. Evidently, when d/a = 0, V_D is null, and all the V_G is due to V_A, i.e., V_A = 1. When the d/a is not zero, the proportion of V_D relative to V_A grows with the increase in the frequency of the favorable allele (p). However, the maximum occurred with the highest value of p used, i.e., 0.667, and with d/a = 1. In this condition, V_D corresponded to 38.46% of V_G.

As already mentioned, it was also possible to express the environmental variance (V_E) as a function of V_G, having the broad-sense h^2 as a reference at the individual level. Hence, for h^2 = 0.2, V_E would be 4 V_G. This proportion decreases with the increase in h^2, as expected. Considering h^2 = 0.6, the environmental variance becomes $V_E = 0.667V_G$. When the multiple-tree plot (MTP) strategy is used, there is the proportion of environmental variance within (V_W) and between (V_e) progenies as a function of V_G. For example, when V_w/V_e = 3 and h^2 = 0.2, the sum of $V_w + V_e = V_E = 4V_G$. In this condition, $V_e = 1{,}0V_G$ and $V_w = 3{,}0V_G$. When, V_w/V_e = 3 and h^2 = 0.6, there is: $V_E = 0{,}667V_G$ and $V_e = 0{,}5V_G$ and $V_w = 0{,}167V_G$.

When using experiments with multiple-tree plots, (results not shown), it was found that for a population with $\overline{p}$ = 0.333, d/a = 0, and h^2 = 0.2, when V_w/V_e changes from 3 to 20, the value of $V_{d_{HS}}$ increases by 21.6%. In turn, with h^2 = 0.6, in the same condition, the increase was only 10.8%. Similar results occurred with $V_{d_{FS}}$, albeit with slightly higher proportions. When the square root of the phenotypic variance between the mean of the progenies $\sqrt{V_F}$ is considered, which is the denominator of the equation of $SG_{between}$, the opposite of that reported earlier for V_d occurs, i.e., the reduction of $\sqrt{V_F}$ with the increase of the V_w/V_e ratio. It was observed that, for $\overline{p}$ = 0.333, d/a = 0, and h^2 = 0.2 and $\sqrt{V_F}$ with HS progenies, the effect V_w/V_e = 3 decreased relative to V_w/V_e = 20 by 18.9%. For $\sqrt{V_{F_{FS}}}$, the reduction was smaller, yet still expressive: 13.7%. These results show that special attention must be given to reducing the environmental variation within the plots, especially when the trait has a lower h^2.

Other variables were also involved in the MTP strategy, such as the numbers of plants per plot (k) and replicates (r). The values of **k** and **r** do not affect the estimate for V_d; however, the product **k×r** evidently has implications in the selection intensity (i) that may be applied within the progenies. The smaller **k×r** is, the greater the proportion of individuals that may be selected, i.e., the lower the value of i and, thus, the lower the SG. However, using HS or FS, $\sqrt{V_F}$ decreases with the increase in **k×r**, allowing a more considerable gain from selection between progenies. Fixing r = 3, h^2 = 0.2, and d/a = 1, $\sqrt{V_{F_{HS}}}$ went from 0.785 to 0.685 with k = 5 or 10, i.e., a 12.8% decrease. Using FS, the reduction was less significant in the same condition: 8.6%. When the effect of going from three to five replicates is compared, always considering ten plants per plot, d/a = 1, $\overline{p}$ = 0.333, and h^2 = 0.2, there is a decrease in $\sqrt{V_{F_{HS}}}$ of 12.3%, and the reduction in $\sqrt{V_{F_{FS}}}$ is of 7.6%. This effect of the number of replicates and plants per plot decrease with the increase of h^2. One may infer that the increase in **k×r** is only justifiable for the trait and conditions in which the history of h^2 is low, regardless of the type of allelic interaction present in the trait control.

In the STP and CPT strategies, the number of plants evaluated per progeny has an effect, especially on the phenotypic variance between progeny means. In the case of the STP, the number of individuals per progeny is equal to **r**. In turn, in the CPT, it corresponds to **q**. With the increase in the number of plants of each progeny, $\sqrt{V_F}$ is reduced in both the evaluations with HS and FS (results not shown). This effect is more considerable the smaller h^2. For example, in STP with d/a = 0, $\bar{p}$ = 0.333, and of h^2 = 0.2, from 20 to 40 plants per progeny, $\sqrt{V_{F_{HS}}}$ decreased by 14.99% (0.698/0.607); with FS, the decrease was less significant: 8.82% (0.851/0.782). However, with CPT, in the same conditions, $\sqrt{V_{F_{FS}}}$ went from 0.595 with 20 plants to 0.550 with 40 plants, a decrease of 8.18%. In the case of , the effect of duplicating the number of plants was even less significant, only 4.20%. These results are repeated with the increase in individuals per area, albeit with an increment in the proportional difference. However, an additional advantage of using a more considerable number of plants per progeny is the possibility of applying a higher selection intensity within the progenies. As already commented for the MTP, the selection of one plant among 20 corresponds to a selection intensity of 5% (i = 1.844), while, among 40 plants, it becomes 2.5% (i = 2.118), i.e., it enables increasing the SG estimate within the progeny by 14.87%.

As expected, the phenotypic variances within and between progeny means decrease with the increase in h^2 regardless of the allele frequencies, the average degree of dominance, and the selection strategy. For the same heritability and allele frequency, V_d always increases with the increment in dominance (Table 2). The opposite occurs in the case of the root square of the phenotypic variances among progeny means. With the increase in the proportional contribution of V_D relative to V_A, $\sqrt{V_{\bar{F}}}$ the values decrease. This increase is more pronounced with higher values of h^2. For example, with $\bar{p}$ = 0.333 and h^2 = 0.2 in MTP, $\sqrt{V_{F_{HS}}}$ with d/a = 1 decreases by 4.7% [1 - (0.685/0.719)] relative to that obtained

Table 2. Estimates of phenotypic variances, in units of population genetic variance (V_G), within progenies (V_d) and the square root of phenotypic variance ($\sqrt{V_{\bar{F}}}$), between half sib (HS) or full sib (FS) progenies in different selection strategies (SS): Multiple tree plots (MTP), single tree plot (STP) and cloned progeny test (CPT). Results obtained considering different allele frequencies ($\bar{p}$), level of dominance (d/a) and heritability (h^2). In MTP, V_w/V_e = 10, r = 3 and k = 10, in TPC, number of plants/progenies of 30, r = 3 and in STP, r = 30.

SS	$\bar{p}$	h^2	d/a	$\sqrt{V_{d\ HS}}$	$\sqrt{V_{F_{HS}}}$	$\sqrt{V_{d\ FS}}$	$\sqrt{V_{F_{FS}}}$
MTP	0.333	0.2	0	4.390	0.719	4.136	0.871
			1	4.436	0.685	4.186	0.843
		0.6	0	1.356	0.562	1.106	0.746
			1	1.406	0.518	1.156	0.713
	0.667	0.2	1	4.482	0.652	4.232	0.816
CPT	0.333	0.2	0	2.083	0.565	1.833	0.749
			1	2.130	0.520	1.883	0.716
		0.6	0	0.972	0.531	0.722	0.724
			1	1.022	0.484	0.772	0.690
	0.667	0.2	1	2.179	0.476	1.929	0.684
		0.6	1	1.068	0.435	0.818	0.657
STP	0.333	0.2	0	4.750	0.639	4.500	0.806
			1	4.800	0.600	4.550	0.776
		0.6	0	1.417	0.545	1.167	0.734
			1	1.467	0.499	1.217	0.700
	0.667	0.2	1	4.846	0.562	4.596	0.746
		0.6	1	1.513	0.452	1.263	0.668

with d/a = 0. With the same conditions except that h^2 is 0.6, the decrease becomes 7.8% [1 - (0.518/0.562)]. The same observation is valid with FS progenies and in STP or CPT. Since the reference is $V_G = 1$, it must be emphasized that when d/a = 0, the absence of dominance, whatever occurs with $\overline{p} = 0.333$ and a given h^2 is repeated for any allele frequency, which is why the values are not shown.

Regardless of the type of conduction strategy, the expected gain from recurrent selection recombining clones of the best individuals was always superior with HS progenies relative to FS ones (Table 3). The superiority of the total gain from selection ($SG_T = SG_{between} + SG_{within}$) with HS on the average of the three strategies and considering the six situations presented was 19.8%. As estimated, the expected SG values are higher when a more considerable heritability of the trait is considered. The effect of h^2 was more pronounced with the employment of FS. For example, with $\overline{p} = 0.333$, on average, if $h^2 = 0.6$ is considered, the SG with HS was 28% of that obtained with $h^2 = 0.2$ and 30% with FS.

In the selection condition considering the selection intensity always the same, 10% between and 10% within, regardless of the selection strategy considered, the SG between progenies explained most of the SG_T obtained. On the average of the six estimates presented for each selection alternative, the gain from selection within progenies was 27.1% of the $SG_T$when using HS and 35.3% for FS. It was also observed that the d/a affected the gain estimates. Considering the three selection strategies, $\overline{p} = 0.333$, the decrease with the increase of the d/a from 0 to 1 was 19% with $h^2 = 0.2$, and 17% for $h^2 = 0.6$. One may surmise that, since the intrapopulation recurrent selection explores only the additive variance, the SG_T decreases with the

Table 3. Estimates of selection gain in units of population genetic variance, within (SG_W), between (SG_B) and total (SG_T). Using half sib (HS) or full sib (FS) progenies in different selection strategies (SS): Multiple tree plots (MTP), single tree plot (STP) and clonal progeny test (CPT). Results obtained considering different allele frequencies ($\overline{p}$), level of dominance (d/a) and heritability (h^2). In MTP, $V_w/V_e = 10$, r = 3 and k = 10; in CPT, number of plants/progenies = 30, r = 3; in STP, r = 30. Selection intensity of 10% between, and 10% within.

SS	$\overline{p}$	h^2	d/a	SG_{WHS}	SG_{BHS}	SG_{THS}	SG_{WFS}	SG_{BFS}	SG_{TFS}	% (SG_{TSH}/SG_{TFS})
MTP	0.333	0.2	0	0.40	1.226	1.626	0.410	1.011	1.421	14.4
			1	0.32	1.028	1.348	0.326	0.835	1.161	16.1
		0.6	0	0.72	1.568	2.288	0.792	1.179	1.971	16.1
			1	0.56	1.362	1.922	0.607	0.987	1.594	20.6
	0.667	0.2	1	0.24	0.832	1.072	0.250	0.664	0.914	17.3
		0.6	1	0.43	1.149	1.579	0.469	0.795	1.264	24.9
Mean				**0.44**	**1.194**	**1.639**	**0.476**	**0.912**	**1.388**	**18.2**
CPT	0.333	0.2	0	0.578	1.557	2.135	0.617	1.175	1.792	19.1
			1	0.458	1.354	1.812	0.487	0.983	1.470	23.3
		0.6	0	0.847	1.572	2.419	0.982	1.215	2.197	10.1
			1	0.661	1.454	2.115	0.760	1.020	1.780	18.8
	0.667	0.2	1	0.348	1.138	1.486	0.370	0.792	1.162	27.9
		0.6	1	0.497	1.181	1.678	0.568	0.825	1.393	20.5
Mean				**0.565**	**1.376**	**1.941**	**0.631**	**1.001**	**1.632**	**20.0**
STP	0.333	0.2	0	0.383	1.377	1.760	0.394	1.092	1.486	18.4
			1	0.305	1.173	1.478	0.313	0.907	1.220	21.2
		0.6	0	0.701	1.615	2.316	0.772	1.198	1.970	11.8
			1	0.552	1.411	1.963	0.605	1.006	1.611	21.8
	0.667	0.2	1	0.233	0.964	1.197	0.240	0.726	0.966	23.9
		0.6	1	0.418	1.198	1.616	0.457	0.811	1.268	27.4
Mean				**0.432**	**1.290**	**1.722**	**0.463**	**0.957**	**1.420**	**21.3**

increment in the proportion of V_D relative to V_A, as expected.

In the three selection strategies, we started from the assumption that the number of progenies evaluated (n) was always the same, with the same occurring with the number of plants per progeny. However, in the case of the CPT, since the evaluation of the clones of each individual is considered, the number of plants was greater that of the MTP and STP. In all six situations presented (Table3), the SG estimates with the employment of CPT were superior to those with STP and MTP. On the average of the six alternatives, with HS, the CPT presented selection gains superior to the MTP and STP strategies by 12.7% and 18.4%. With FS, in the same conditions, the CPT surpassed the MTP and the STP by 17.6% and 14.9%, respectively. In turn, the $SG_T$of the STP was superior to that of the MTP. However, the differences in the mean estimates of gain were lower: 5.1% with HS progenies and only 2.3% with FS progenies (Table3).

Discussion

We sought to perform the comparisons among the improvement strategies by using the most variables involved in the process. With adjustments, the so-called breeder's equation applies to most situations (Cobb et al. 2019). It is important to stress that the comparisons were focused on the conduction of selection in one generation (selective cycle). However, the two main aspects of any recurrent selection program were considered: the strategies in the evaluations for selecting the progenies/individuals and in recombination.

It should be noted that all comparisons were performed considering the use of the least squares method .However, the employment of mixed models is growing in agricultural data analysis. One may question the implications of this fact in the comparisons made. The first is that if balanced experiments are considered, without losing plants or plots, the results will be the same (Bernardo 2020). However, even with a more considerable imbalance of the experiments, although the ways of estimating the variance components are different, proportionally, the comparisons should not be affected significantly since it is expected that the considerations will be the same in the different selection strategies used.

The fact that a genetic variance equal to one ($V_G = 1$) was used as a reference, as done by Atlin and Econopouly (2022) to compare some selection strategies in autogamous plants, has some advantages, as already mentioned: a) information applies to any trait, regardless of the measurement unit. The unit of the expected gain is the estimate of the V_G of the population under selection; b) it allows both the genetic variance components (V_A or V_D) and the environmental variance to be transformed into units of V_G. This alternative allows comparing countless possibilities of selection strategy using the breeder's equation.

The plant populations of any cultivated species are typically at different improvement stages. In other words, they have different frequencies of the favorable alleles of the different genes, which control the expression of the trait under selection. In this condition, for the comparisons among improvement strategies to be more reliable, one must make the comparisons for populations with different allele frequencies. At first, one may imagine that they must be low in recently introduced natural populations. In turn, for synthetic populations produced by companies and obtained by crossing the best clones, they must be medium to high. In this context, what has been done is to consider a single allele frequency (Sherboune 1992; Hallauer, Carena and Miranda Filho 2010). In this work, we sought to compare little improved ($\overline{p} = 0.333$), averagely improved ($\overline{p} = 0.555$), and well improved ($\overline{p} = 0.667$) populations. However, certainly, having populations with low means of favorable alleles, such as an average of $\overline{p} = 0.333$, does not indicate that all alleles have the same frequency.They may vary from 0 to 1. Using the beta distribution, with a mean of $\overline{p} = 0.333$, one may infer that approximately 66.3% of the alleles of the different loci should individually have an allele frequency under 0.4. With

$\overline{p}$ = 0.667, this value changes to 8.7%. Thus, the employment of mean allele frequencies of 0.333 and 0.667 should represent most of the expected situations.

Using these frequency distributions, it was also possible to estimate the proportions of the mean additive variance ($\overline{V}_A$) and mean dominance ($\overline{V}_D$), considering different estimates of the average degree of dominance. As expected, when d = 0, V_G was entirely additive. However, with d different from zero, the proportion of V_D increased with the rise in the allele frequency of the population (Table 2). However, even in the situation of d = a, the highest allele frequency of the population V_D corresponded to 38.5% of V_G.

As expected, the heritability estimates among the traits presented in the literature with eucalyptus breeding vary greatly (Sumardi, Kurniawan and Prastyono, 2016; Tambarussi et al. 2017; Chen et al. 2018; Ekomono, Rambolarimana and Bouvet, 2022). In this work, three broad-sense h^2 estimates were considered at the individual level (h^2 = 0.2, 0.4, and 0.6), expecting them to represent most situations that breeders encounter. Highlighting once again that we considered V_G = 1, one may infer the importance of the environmental variance (V_E) relative to the genetic variance of the population. Hence, for h^2 of 0.2, V_E will correspond to 4 V_G. With h^2 = 0.6, this value becomes V_E, only 0.67 of V_G. Hence, in the comparisons presented, it is easier to comment on the effect of h^2 on the different improvement strategies. As expected, the lower h^2 is, the lower the expected gain from selection (Table 3). However, some selection strategies respond more to a decrease in heritability. For example, in the comparisons involving the use of more plants per plot and the STP, considering the same number of plants in the experiment, the advantage of the STP is more significant the lower h^2 is (Table 3).

The phenotypic variance among HS progeny means was consistently lower than that of the FS progenies regardless of the allele frequencies of the population, the interaction type, and the numbers of replicates or plants per plot. In turn, the opposite occurred for the phenotypic variance estimates within the progenies (V_D), i.e., $V_{d_{FS}}$ was always of a lower magnitude than $V_{d_{HS}}$. However, the magnitude of the difference $V_{d_{HS}}$ - $V_{d_{FS}}$ was always more significant than the difference. Considering that the numerator of the equation for the expected gain from selection is the same for HS and FS when the recombination is carried out using clones, i.e., ½ for both the selection between and within progenies, one may infer that the total gain (SG_T) from the selection (between + within) using the same selection intensity will always be more considerable using HS than FS (Table3). It must be emphasized that these results differ from most selection method comparison situations involving eucalyptus in the literature, for which the employment of FS is superior to HS (Shelbourne, 1992). The difference is that, in these comparisons, clones of the individuals selected for the progeny recombinations were not used, and, in this condition, the numerator of the equation for $SG_{between}$ is ¼ for HS and ½ for FS.

Results in the literature on eucalyptus breeding show that the behavior of the clones in the progeny tests and their performance in the clonal tests have little repeatability (Furtini et al. 2012; Reis et al. 2011). These results have stimulated some companies to adopt cloned progeny test (CPT). In the context of comparing selection strategies, all observations made previously are valid for the CPT. Additionally, CPT enables more significant accuracy in the evaluations of the progenies and the individuals within each progeny. A great additional advantage is that it decreases the time to obtain a new clone by at least three to five years, depending on the time used to produce a reasonable number of cloned seedlings of all individuals from different progenies to be evaluated from a single seedling.

There are two alternatives when conducting CPT: selecting the progenies/individuals for the continuity of the RS and/or selecting the best clones for forest exploration. In the latter case, one may or may not use the information from the progeny test. When the information from the progeny test is not used, the individuals are selected while having as a reference only their “per se” performance,

provided by the clones derived from it. In this case, all the available genetic variance is used in the selection, not just V_A, as occurs in RS. No reference was found in the literature relative to the employment of the information on the progenies in CPT to select the best clones. However, in a study comparing selection gains, Shelbourne (1992) showed that CPT (denominated by the author as open-pollinated breeding population cloning – OP BP cloning) with four to ten ramets per clone and ten per progeny resulted in gains expressively superior to those obtained with conventional progeny test with the same number of plants per progeny.

In the context of RS, there are several options for recombination. However, through simulations, Abreu et al. (2013) showed that the use of the best individuals/clones from the best progenies allows obtaining more significant gain with RS than the employment of remnant seeds of the best progenies or, in the case of perennial plants, the plants that originated the progenies. However, HS progenies are obtained when clones are used for the recombination and the crossbreeding occurs randomly. One of the disadvantages of this recombination is that the flowering of the trees from different clones may not coincide. Hence, the pollination may not be carried out well. In this condition, the success of the RS is harmed despite all efforts to evaluate the progenies. To avoid this problem, the recombination could be done through artificial crossbreeding of the selected plants/clones. The information available for eucalyptus show that artificial hybridization is practically viable (Assis and Mafia 2007). In this case, to obtain HS progenies, each plant would be pollinated by a pollen mixture of the other selected individuals. This procedure is feasible because, in the case of eucalyptus, the pollen grains may be stored without losing viability. No reports of the employment of this procedure for eucalyptus were found, but it is routinely used in other species, such as tobacco (Marques et al. 2022).

The employment of clones for the recombination, as has been used for eucalyptus, has the additional advantage that the individual selected in the progeny testing has its performance confirmed in the clonal test performed right after. However, this procedure requires additional time to complete the RS cycle, which may not compensate for the additional gain obtained. This has been one of the reasons for stimulating the employment of cloned progeny test. In CPT, the clones of each individual of a given progeny are simultaneously evaluated in the progeny tests. As already commented, the SG_T with CPT is superior to the traditional process (Table3).

Relative to recombination, it is worth noting that, when using clones aiming at obtaining HS progenies, a problem occurs in the proportion of plants that must be selected. For example, imagine that 200 progenies were evaluated with 20 plants per progeny, and, in the end, 20 individuals were selected. The proportion selected was 20 out of 4000 individuals (selection intensity of 0.5%). If ten plants/clones are used from each individual to recombine the recombination batch, there would be 200 plants. They would be randomized to allow all 20 clones to have similar chances of receiving pollen from the others. In this condition, 20 new progenies will be obtained to give continuity to the RS. This occurs because the seeds of each plant from the same clone will be mixed. At first, one could think that, as the gametes of each plant from the same clone will hardly be genotypically equal, the descendants of each plant could be a new HS progeny. However, on average, it was expected that the performances of plants from the same clone, HS, would be similar. Therefore, regardless of the number of plants, 20 clones will produce 20 progenies. This is a serious restriction of the employment of clones or the individuals of the progeny tests in recombination.

Compared with the recombination aiming to obtain FS, since the crossings is performed in pairs, it would be possible to obtain 190 FS progenies with 20 individuals/clones [(20 × 19)/2]. For example, if there were 190 HS or FS progenies with 30 plants per progeny, i.e., 5700 plants. By applying the selection intensities of 10.53% between progenies and 3.33% within progenies, there would be, in the end, 20 plants for the recombination. If they are HS, the maximum number of prog-

enies, due to the reason already commented on, would be 20 and no longer 190. In turn, with FS, as commented, it would again be 190. What selection intensity would need to be applied with HS to obtain the same number of progenies as the FS (190)? Since the population has 5700 individuals, the proportion to be selected would be 190 out of 5700, i.e., 0.0333. With the obtainment of FS, the selected proportion would be 20 out of 5700 (0.0035). Hence, the proportion selected with HS would be 9.5 times that with FS. Evidently, this is a considerable advantage of FS relative to HS.

It is assumed that each HS progeny corresponds to the effective size (Ne) of 4 and each FS progeny to Ne = 2 (Hallauer, Carena, & Miranda Filho, 2010). To exemplify the consequence of using the same Ne, 190 FS progenies and 95 HS progenies and the continuity of the process will be considered, yet maintaining the same number of progenies for the evaluation of the following cycle. For such, the beginning of the process will be considered with 95 HS progenies and 30 plants per plot. With FS, we would have 190 progenies and 15 plants per plot, i.e., the same number of plants in the experiment. One possibility to maintain the same number of progenies in the successive cycles with HS would be to apply a selection intensity between progenies of 25% (95 × 0.25 = 24) and to select, within progenies, four plants per progeny, i.e., a selected proportion of 13.33% (4/30 = 0.133). The total selected proportion with HS should be 0.033 (95/2850). To maintain the same number of 190 progenies with FS, it would be necessary to select 20 progenies with one plant per progeny, with a selected proportion of 0.007 (20/2850), i.e., a selected proportion between progenies of 20/190 = 0.1053 and within progenies of 1/15 = 0.067. In this condition, considering CPT, r = 3 and the same Ne, the SG_T estimate is very similarwith FS relative to HS when the population is little improved, d/a = 0, and h^2 = 0,2 (SG_{TFS} = 1.37 and SG_{THS} = 1.35). This same observation is valid for other allele frequencies, *gmd* and h^2.

Not considering the Ne, in the same condition, 190 progenies and 30 plants per plot would be evaluated in both conditions. To have once again 190 progenies aiming at the continuity of the RS, the selected proportion with HS would be 33.3% (190 × 0.333 = 64) between progenies and 10% (30 × 0.10 = 3) within progenies. In other words, 192 (64 × 3) plants that, once interbred, will maintain the 190 progenies. With FS, the selected proportions would be 10.53% (190 × 0.1053 = 20) between progenies, and 3.33% (30 × 0.033 = 1) within progenies. Hence, in the end, we would have 20 plants that, once crossed in pairs, would maintain the 190 progenies. In this situation,with the SG_T estimate with FS, regardless of h^2, will always be higher than with HS. For example, considering CTP with r = 3, p = 0.33, d/a = 0 and h^2 = 0.2, the estimate would be SG_{TFS} = 1.55, ie 26% higher than that obtained with HS (SG_{THS} = 1.23). The advantage of FS in relation to HS reduces with the increase of h^2 and gmd, but even so it is still expressive.

The matter of the effective size may be viewed in two aspects: the first would be in terms of the conservation/preservation of the genetic properties of the populations, and the second, of the breeding program. The Ne value has more meaning in the first case. When one is performing the selection, especially considering a perennial plant, for which the duration of the selective cycles requires some time, maintaining a large Ne in detriment to the proportion that may be selected is not justified. There are results in the literature that show that the Ne does not need to be large for there to be a success in long-term selection (Backer and Curnow 1969; Rawlings1970). Through simulations, Backer and Curnow (1969) showed that the probability of fixing a given favorable allele is dependent on the Ne and the initial frequency of the allele ($\overline{p}$). They showed, for example, that the probability of changing the frequency $\overline{p}$ = 0.3 to $\overline{p}$ = 1 with RS is high, above 93%, with Ne = 30. Additionally, it should be noted that it is possible, throughout the selective cycles, to add other clones during the recombination. However, the inclusion must only be of clones proven superior and, evidently, at a very small proportion. Adding clones with low averages decreases

or even eliminates the progress with the selection of the previous cycles.

From the exposed, in recombination using clones, which is the best option companies have, the limitation of applying higher selection intensities when using HS is a severe restriction. Considering this fact associated with the difficulty of having efficient recombination when using HS, the employment of FS progenies must be stimulated.

Conflict of Interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by all of them. The first draft of the manuscript was written by Magno Antonio Patto Ramalho and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

ABREU, G.B.; FERREIRA, D.F.; RAMALHO, M.A.P.; TOLEDO, F.H.R.B.;BUENO FILHO, J.S.S. 2013. Computer simulation for the evaluation of recombination strategies in Intrapopulation Recurrent Selection in Eucalyptus. **Silvae Genetica**,62:68–79.

ASSIS, T.F.; MAFIA, R.G. 2007.Hibridação e clonagem de Eucalyptus. In: Borém A (ed) **Biotecnologia Florestal**. Editora UFV, p.93-121,Viçosa.

ASSIS, T.F.; ABAD, J.I.M.; AGUIAR, A.M. 2015.Melhoramento Genético do Eucalipto. In: SCHUMACHER, M.V.; VIEIRA, M. (ed) **Silvicultura do Eucalipto no Brasil**. Editora UFSM, p. 217-244, Santa Maria.

ATLIN, G.N.; ECONOPOULY, B.F. 2022.Simple deterministic modeling can guide the design of breeding pipelines for self-pollinated. **Crop Science** ,62:661–678.

BAKER, L.H.; CURNOW, R.N. 1969. Choice of population size and use of variation between replicate populations in plant breeding selection programs. **Crop Science**, 19 (4): 566-566.

BERNARDO, R .2020.**Breeding for quantitative traits in plants**. Stemma Press, Woodbury.

BURDON, R.D.; SHELBOURNE, C.J.A. 1974. The use of vegetative propagules for obtaining genetic information. **New Zealand Journal of Forestry Science**, 4: 418-425.

CHEN, S.; WENG, Q.; LI, F; LI, M.; ZHOU, C.; GAN, S. 2018.Genetic parameters for growth and wood chemical properties in *Eucalyptus urophylla* × *E. tereticornis* hybrids. **Annals of Forest Science**, 75: 16.

COBB,J.N.;JUMA, R.U.; BISWAS, P.S. et al. 2019.Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation.**Theoretical and Applied Genetics**, 132: 627–645.

EKOMONO, C.G.M.; RAMBOLARIMANANA, T.; BOUVET, J.M. 2022. Preponderance of additive and non-additive variances for growth, ecophysiological and wood traits in Eucalyptus hybrid genotype-by-spacing interaction. **Tree Genetics and Genomes**, 18: 32.

FOSTER, G.S.; SHAW, D.V. 1988. Using clonal replicates to explore genetic variation in a perennial plant species. **Theoretical and Applied Genetics**,76:788–794.

FURTINI, I.V.; RAMALHO, M.A.P.; ABAD, J.I.M.; AGUIAR, A.M. 2012.Effect of different progeny test strategies in the performance of eucalypt clones. **Silvae Genetica**, 61 (3): 116-120.

HALLAUER, A.R.; CARENA, M.J.; MIRANDA FILHO, J.B. 2010.**Quantitative genetics in maize breeding**. Springer, New York.

MARQUES, T.L.; PADUA, J.M.V.; BERGER, I.J.; RAMALHO, M.A.P. 2022. Strategies for the recurrent selection program in tobacco breeding for green leaf yield. **Crop Science**, 62 (6): 2212-2221.

NELSON, R.; WIESNER-HANKS, T.; WISSER, R.2018. Navigating complexity to breed disease-resistant crops. **Nature Reviews Genetics**, 19:21–33.

PIRES, I.E.; RESENDE, M.D.V.; SILVA, R.L.; RESENDE JUNIOR, F.R. 2011. **Genética Florestal**. Arka, Viçosa.

RAMALHO, M.A.P.; ABREU, A.F.B.; SANTOS, J.B.; NUNES, J.A.R. 2012. **Aplicações da genética quantitativa no melhoramento de plantas autógamas**. Editora UFLA, Lavras.

RAWLINGS, J.O. 1970. **Presents status of research on long and short term recurrent select in finite population-choise of populaterm recurrent selection in finite population-choice of population size.** In Paper presented at 2nd Meeting of Working Group on Quantitative genetics USDA SFES, New Orleans, p. 1-15.

REIS, C.A.; GONÇALVES, F.M.A.; ROSSE, L.N.; COSTA, R.R.; RAMALHO, M.A.P. 2011. Correspondence between performance of Eucalyptus spp trees selected from family and clonal tests. **Genetics and Molecular Research**, 10 (2): 1172-1179.

RESENDE, M.D.V. 2002.**Genética biométrica e estatística no melhoramento de plantas perenes**. EMBRAPA Informação Tecnológica, Brasília.

RESENDE, M.D.V. 2015.**Genética quantitativa e de populações**. Suprema, Viçosa.

RUTKOSKI,J.E. 2019. A practical guide to genetic gain.**Advances inAgronomy**, 157: 217-249.

SANTOS, H.G.; GONÇALVES, F.M.A.; LIMA, J.L. 2021. Strategies for the analysis of single-tree plot experiments in *Eucalyptus* plantations. **Journal of Forestry Research**, 32: 2437–2445.

SHELBOURNE, C.J.A. 1992. Geneticas gains from different kinds of breeding population and seed or plant product population. Paper presented at the IUFRO symposium "Intensive Forestry: The Role of Eucalyptus" held in Durban, South Africa, p. 49-65.

TAMBARUSSI, E.V.; LIMA, B.M.; COSTA QUEIROZ, R. et al. 2017. Estimate of genetic parameters for early selection in clones of Eucalyptus spp.Scientia Forestalis, 45 (115): 507-517.

WRICKE, G.; WEBER, W.E. 1986. **Quantitative genetics and selection in plant breeding.** Walter de Gruyter, Berlin.

VENCOVSKY, R.; PEREIRA, M.B.; CRISOSTOMO, J.R.; FERREIRA, M.A.J. 2001. Genética e Melhoramento de Populações Mistas. In: Nass, L.L et al. **Recursos Genéticos e Melhoramento**. Editora Fundação Mato Grosso, p. 231-282.