THAÍS LIMA MARQUES

STRATEGIES AIMING GREATER EFFICIENCY IN THE TOBACCO BREEDING PROGRAM

LAVRAS - MG
2021

THAÍS LIMA MARQUES

STRATEGIES AMING GREATER EFFICIENCY IN THE TOBACCO BREEDING PROGRAM

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, área de concentração em Genética Quantitativa, para obtenção do título de Doutor.

Prof. Dr. Magno Antonio Patto Ramalho
Orientador

LAVRAS - MG
2021

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Marques, Thaís Lima.
Strategies aiming greater efficiency in the tobacco breeding program / Thaís Lima Marques. - 2021.
73 p.

Orientador(a): Magno Antonio Patto Ramalho.

Tese (doutorado) - Universidade Federal de Lavras, 2021.
Bibliografia.

1. plant breeding. 2. quantitative genetics. 3. phenotyping. I. Ramalho, Magno Antonio Patto. II. Título.

THAÍS LIMA MARQUES

STRATEGIES AMING GREATER EFFICIENCY IN THE TOBACCO BREEDING PROGRAM

ESTRATÉGIAS VISANDO AUMENTO NA EFICIÊNCIA NO PROGRAMA DE MELHORAMENTO DE TABACO

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, área de concentração em Genética Quantitativa, para obtenção do título de Doutor.

APROVADA em 09 de Novembro de 2021

Dr. Adriano Teodoro Bruzi, UFLA

Dr. José Maria Villela Padua, UFLA

Dr. Irving Joseph Berger, BAT

Dr. Ricardo Augusto Diniz Cabral Ferreira, SYNGENTA

Prof. Dr. Magno Antonio Patto Ramalho
Orientador

LAVRAS - MG
2021

À Deus,

pela Sua abundante graça.

Ofereço

Aos meus pais, Maria e Paulo,

com todo amor.

Dedico

ACKNOWLEDGMENTS

À Deus, sou grata por tudo.

À minha família, em que mesmo com a minha ausência, estão sempre presentes em toda a minha caminhada. Sou especialmente grata aos meus pais, Maria e Paulo, por todo amor. Ao meu irmão Guilherme pela grande amizade e incentivo a continuar. Ao meu esposo Gabriel pelo companheirismo e paciência. Aos meus tios e primos por me receberem e acolherem em Lavras, especialmente a minha avó Clarice por todo carinho. Muito obrigada por acreditarem em meus sonhos.

Ao meu grande exemplo de pessoa e profissional, professor Dr. Magno Ramalho, o qual tive a grande honra em tê-lo como meu orientador durante o doutorado. Agradeço por todos os ensinamentos, atenção, paciência e disponibilidade nos últimos anos. O entusiasmo que o professor Magno tem pela ciência é contagiante o qual admiro desde 2012 quando, durante a graduação, tive o privilégio em ser aluna. Meus sinceros agradecimentos pelos conselhos, levarei para sempre comigo e espero que o encerramento desse ciclo seja apenas o início da nossa amizade.

Professor Dr. Katy Rainey for welcoming me at Purdue University. I appreciate the opportunity and confidence. All the experience acquired was important and contributed to the execution of this work. Thank you so much to everyone who was with me in West Lafayette, especially to my great Chi Alpha friends who were always by my side. I treasure all the memories of the lived moments.

A todos profissionais que participaram da minha formação. À professora Dr. Édila Von Pinho por ter sido minha primeira orientadora, pelo incentivo inicial e co-orientação durante o mestrado juntamente com o professor Dr. Renzo Von Pinho. À pesquisadora Dr. Ângela Abreu pela amizade, conselhos, ajuda e apoio. Aos professores Dr. José Maria e Dr. Adriano Bruzi pelo apoio à pesquisa. A todos os professores durante a pós-graduação, pelos exemplos de dedicação e por contribuírem com ensinamentos para meu aprendizado.

Aos colegas e amigos do GEN e do feijão por compartilharem comigo as mesmas angustias e alegrias durante toda a pós-graduação, muito obrigada pela amizade. O apoio de vocês continuará sendo fundamental.

À Universidade Federal de Lavras, especialmente ao Departamento de Biologia e ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas por contribuíram para a minha formação profissional e desenvolvimento pessoal.

À British American Tobacco (BAT Brasil) por fornecer recursos para a realização dessa pesquisa. Agradeço a toda equipa do Centro de Melhoramento de Tabaco pela condução do experimento no campo, pela disponibilidade e interação entre a universidade e a empresa. A todos os envolvidos na idealização do trabalho e acompanhamento até a fase final.

O presente trabalho foi realizado com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes, Código de financiamento 001), pela concessão da bolsa de estudos.

Finalmente, a todos que de algum modo contribuíram para que fosse possível a realização desse trabalho e foram responsáveis para que essa etapa da minha vida fosse especial e inesquecível, meus sinceros agradecimentos.

“It is not the strength of your faith that saves you,
but the strength of Him upon whom you rely”
Charles H. Spurgeon

RESUMO

A cultura do tabaco apresenta grande importância socioeconômica para o Brasil, onde mais de 90% da produção brasileira é exportada e milhares de famílias estão envolvidas com o cultivo. Nesse contexto, é importante a obtenção de informações visando aumento na eficiência dos programas de melhoramento. No programa de seleção recorrente (SR) de tabaco diferentes estratégias podem ser empregadas na seleção e recombinação. Além disso, a escolha da melhor estratégia depende dos componentes da variância genética. Adicionalmente, visando aumentar a eficiência nos programas de melhoramento, o processo de colheita do tabaco demanda maior mão-de-obra e é também o processo com maior possibilidade em contribuir para o erro experimental. No entanto, uma das alternativas que vem sendo utilizada em outras espécies de plantas é a fenotipagem por imagens da cultura. Dessa forma, o objetivo desse estudo foi estimar os componentes da variância genética em uma população de tabaco e comparar as diferentes estratégias de condução da SR. Adicionalmente, verificar a viabilidade do emprego da fenotipagem da produção de folhas (produtividade) por meio de imagens coletadas bor veículos aéreos não tripuláveis e suas aplicações no programa de melhoramento de tabaco. No primeiro estudo, a variância de dominância foin nula para todos caracteres avaliados, indicando que o efeito de dominância (d) não é importante para a população de tabaco FCV empregada nesse estudo. Nessa condição, a estimative de D_1, a qual depende de d, não é significativa, indicando que a depressão por endogamia não influencia no ganho com a seleção (GS) no programa de SR; portanto, qualquer itpo de progênie pode ser utilizada. Por meio da equação do GS, foi constatado que o emprego de linhagens duplo-haploides (DH) na SR de tabaco representa a melhor estratégia, pois associa maior GS por unidade de tempo e gera a linhagem no final de cada ciclo, a qual é o objetivo final do programa de melhoramento genético. No entanto, no caso em que não for utillizado linhagens DHs, outras estratégias também apresentam expressivo GS, como o emprego alternativo de linhagens endogâmicas e irmãos germanos (IG) para avaliação e progênies de IG para recombinação. No segundo trabalho, constatou-se que: i) a fenotipagem por imagens apresenta boa acurácia em detectar diferenças entre os híbridos/linhagens de tabaco; ii) a cobertura do dossel (CC), obtida por meio da fenotipagem por imagens, correlacionou com a produtividade de folhas verdes/curadas; iii) o emprego da fenotipagem por imagens deve ser estimulada na cultura do tabaco visando a obteção da produtividade de folhas verdes/curadas; iv) conjecturou-se que a aplicação de imagens é viável em diversas situações dentro do programa de melhoramento visando a produtividade de folhas verdes/curadas.

Palavras-chave: melhoramento de plantas, genética quantitativa, fenotipagem, acurácia seletiva.

ABSTRACT

Tobacco crops have great socioeconomic importance in Brazil, where more than 90% of the Brazilian production is exported and thousands of families are involved with its cultivation. In this context, it is important to obtain information to improve the efficiency of the breeding programs. In tobacco recurrent selection (RS) different strategies can be employed in selection and recombination. Moreover, choosing the best strategy depends on the genetic variance components. Additionally, aiming to improve the efficiency of the breeding programs, the tobacco harvest is the process that demands the most labor and is also the process that has the highest possibility of contributing to experimental error. However, one of the alternatives that have been used in other species is phenotyping through images of the crop. Thus, the objective of this study was to estimate the genetic variance components in a tobacco population and compare the different strategies to conduct RS. Additionally, to verify the viability of employing the phenotyping of leaf mass (yield) through images collected by unmanned aerial vehicles (UAVs) and its likely applications in the tobacco breeding program. In the first study, it was found that the dominance variance was null for all evaluated traits, indicating that the dominance effect (d) is not important for the FCV tobacco population used in this study. In this condition, the estimate of D_1, which depends on d, is not significant, meaning that inbreeding does not influence the gain from selection (GS) in the RS program; therefore, any type of progeny may be used. From the GS equation, it was found that the use of doubled haploid (DH) lines in tobacco RS represents the best strategy, as it associates greater GS per time unit and generates the line at the end of each cycle, which is the ultimate goal of the genetic breeding program. However, in the case of not using DH lines, other strategies allow expressive GS, such as the alternate use of inbred and FS progenies for evaluation and FS progenies for recombination. In the second study, it was found that: i) image phenotyping presented good accuracy in detecting differences among the tobacco hybrids/lines, with the accuracy rising as the number of days after planting increased; ii) the canopy cover (CC), assessed through image phenotyping, correlated with the yield of green/cured leaves; iii) the employment of image phenotyping must be stimulated in tobacco crops aiming the obtainment of the green/cured mass yield; iv) it was conjectured that the application of images is viable in several situations within a breeding program aiming at the green/cured mass yield.

Keywords: plant breeding, quantitative genetics, phenotyping, selective accuracy.

LIST OF FIGURES

SECOND PART – ARTICLES

ARTICLE 1

ARTICLE 2

LIST OF TABLES

SECOND PART – ARTICLES

ARTICLE 1

APPENDIX A

APPENDIX B

ARTICLE 2

APPENDIX A

SUMMARY

FIRST PART

1 INTRODUCTION

Tobacco crops have great economic importance in Brazil, where more than 90% of the Brazilian production is exported, moving over 2.1 billion dollars per year (Tobacco Growers' Association of Brazil – AFUBRA, 2021). In addition to its great economic importance, tobacco also has high social importance for the country, because thousands of families (159 thousand families in the 2018/19 harvest) are involved with its cultivation in 14 Brazilian states, with the south region being responsible for over 97% of the production in the country. In this context, given the great socioeconomic relevance of the crop in Brazil, it is important to obtain information to improve the efficiency of the breeding programs and, consequently, to contribute to yield increase.

In the tobacco recurrent selection (RS) program, as in other autogamous species, it is possible to use different types of progenies and even doubled-haploid lines (DH). However, when using inbred progenies ($S_{0:1}$, $S_{0:2,}$ and DH) in the RS it is necessary to estimate de genetic variance component D_1 (genetic covariance between the additive effects of the alleles and the dominance effects of the homozygotes) where, as it is a genetic covariance, it can assume negative values, in addition to being present in the numerator of the gain from selection equation (SOUZA JÚNIOR, 1989). In light of the above, it appears that if D_1 is of great magnitude and negative, the use of inbred progenies in the RS would not be a good alternative.

As in tobacco and other autogamous species it is possible to use different types of progenies and even DH lines, the gain from selection is a function of the type of progeny evaluated and recombined (RAMALHO *et al.*, 2021). This aspect is particularly important in autogamous plants because from the estimates of the genetic variance components obtained from the RS population, it is possible to identify better alternatives of types of progenies to be evaluated and recombined.

Additionally, in the conduction of the genetic breeding program, besides being the process that demands the most labor, the harvest is also that which has the highest possibility of contributing to experimental error. Thus, alternatives that might reduce the experimental error and the cost of the harvest operation are evidently desired. One of the alternatives that have been employed in other plant species, in this case, to estimate grain production, is phenotyping through images of the crop (FERNANDEZ-GALLEGO *et al*., 2019; GALLI *et al*., 2020; HU; KNAPP; SCHMIDHALTER, 2020; KRAUSE *et al*., 2020; MOREIRA *et al*.,

2019; NATARAJAN *et al*., 2019; XAVIER *et al*., 2017). The perspectives for the employment of image phenotyping of tobacco aiming at predicting green mass yield are more promising than for grains because it depends only on plant canopy. However, no references of use with tobacco crops were found in the literature.

Thus, the objective of this study was to estimate the genetic variance components in a RS tobacco population and through these estimates, using the gain from selection equation, compare the different strategies to conduct RS. Additionally, to verify the viability of employing the phenotyping of leaf mass (yield) through images collected by unmanned aerial vehicles (UAVs) and its likely applications in the tobacco breeding program.

REFERENCES

FERNANDEZ-GALLEGO, J. A. et al. Low-cost assessment of grain yield in durum wheat using RGB images. **European Journal of Agronomy**, v. 105, p. 146-156, 2019.

GALLI, G. et al. Optimization of UAS-based high-throughput phenotyping to estimate plant health and grain yield in sorghum. **The Plant Phenome Journal**, v. 3, n. 1, p. 1-14, 2020.

HU, Y.; KNAPP, S.; SCHMIDHALTER, U. Advancing high-throughput phenotyping of wheat in early selection cycles. **Remote Sensing**, Basel, v. 12, n. 574, p. 1-10, 2020.

KRAUSE, M. R. et al. Aerial high-throughput phenotyping enables indirect selection for grain yield at the early generation, seed-limited stages in breeding programs. **Crop Science**, Madison, v. 60, p. 3096-3114, 2020.

MOREIRA, F. F. et al. Improving the efficiency of soybean breeding with high-throughput canopy phenotyping. **Plant Methods**, London, v. 15, n. 139, p. 1-9, 2019.

NATARAJAN, S. et al. High-throughput phenotyping of indirect traits for early-stage selection in sugarcane breeding. **Remote Sensing**, Basel, v. 11, n. 2952, p. 1-18, 2019.

RAMALHO, M. A. P. *et al*. **Aplicações da genética quantitativa no melhoramento de plantas autógamas.** 1 ed. Lavras: Ed. UFLA, 2012.

SOUZA JÚNIOR, Cláudio Lopes. **Componentes da variância genética e suas implicações no melhoramento vegetal.** 1 ed. Piracicaba: Ed. FEALQ, 1989.

TOBACCO GROWER'S ASSOCIATION OF BRAZIL (AFUBRA). Rio Grande do Sul, 2021. Available at: < https://afubra.com.br/fumicultura-brasil.html>. Accessed on: July 6, 2021.

XAVIER, A. et al. Genetic architecture of phenomic-enabled canopy coverage in *Glycine max*. **Genetics**, Austin, v. 206, p. 1081-1089, 2017.

SECOND PART - ARTICLES

ARTICLE 1 – STRATEGIES FOR THE RECURRENT SELECTION PROGRAM IN TOBACCO

Artigo redigido conforme a NBR 6022 (ABNT, 2018) e formatado de acordo com o Manual da UFLA de apresentação de teses e dissertações.

ABSTRACT

In tobacco crops, different strategies may be implemented to select and recombine the best progenies. However, choosing the best strategy depends on the genetic variance component. When using inbred progenies with an allelic frequency other than $^1/_2$, one of the genetic variance components is the genetic covariance between additive effects and homozygous dominance effects (D_1). This component may present negative values and is part of the numerator in the equation for gain from selection (GS). Consequently, the identification of the progeny type to be used in a recurrent selection (RS) program of the flue-cured Virginia (FCV) varietal group depends on knowing the magnitude of D_1. Thus, the objective of this study was to estimate the genetic variance components in a tobacco population and compare the different strategies to conduct RS. From an RS population, half-sib (HS), full-sib (FS), and $S_{1:2}$ progenies were evaluated in different experiments. The traits assessed were green leaf yield (GLY), total sugar content (TSC), total alkaloid content (TAC), and leaf width (LW). From the phenotypic and error variances, the additive variance, dominance variance, and D_1 were estimated for all traits. With the obtained estimates, the gain from selection was estimated for the seven different strategies for conducting the RS program. The dominance variance was null for all evaluated traits, indicating that the dominance effect (d) is not important for the FCV tobacco population used in this study. In this condition, the estimate of D_1, which depends on d, is not significant, meaning that inbreeding does not influence the GS in the RS program; therefore, any type of progeny may be used. From the GS equation, it was found that the use of doubled haploid (DH) lines in tobacco RS represents the best strategy, as it associates greater GS per time unit and generates the line at the end of each cycle, which is the ultimate goal of the genetic breeding program. However, in the case of not using DH lines, other strategies allow expressive GS, such as the alternate use of inbred and FS progenies for evaluation and FS progenies for recombination.

Keywords: Plant breeding, quantitative genetics, comparison of selection methods, inbreeding.

1 INTRODUCTION

In addition to having great economic importance, the tobacco plant has numerous biological advantages, being considered a model plant in several research areas. In genetics and plant breeding, tobacco brings some peculiar advantages: easy vegetative propagation, which increases the viability of using doubled haploid (DH) lines (LEMOS, 2021; SOOD et al., 2013); the presence of many flowers bearing floral biology that provides an enormous amount of seeds per plant, facilitates artificial pollination, increases viability in pollen storage, and enables the obtainment of several types of progeny to be evaluated; and, since the leaves are the commercial product of tobacco, selection may be carried out before flowering. Thus, although tobacco is an autogamous plant, it is possible to evaluate selection alternatives applicable to nearly all other cultivated plants. However, these possibilities have been little explored in plant breeding.

In tobacco, numerous traits are considered in the selection, including those of agronomic importance such as yield and those related to industrial aspects. It is noteworthy that most of these traits are quantitative, i.e., controlled by a large number of genes. In this situation, recurrent selection (RS) should be the main option to accumulate favorable alleles for the traits under selection (BERNARDO, 2020; RAMALHO et al., 2012). The use of RS was initially proposed for allogamous plants; however, in the last decades, it has been intensified in autogamous plants (LOPES et al., 2019; SOARES et al., 2020). In the case of tobacco, for example, no report was found in the literature on the use of RS, although it has recently been adopted in Brazil.

In the RS stages, evaluating the progenies is the most difficult and decisive step, as the selection will only be efficient, increasing the frequency of favorable alleles in the population, if solely the best progenies are intercrossed. In the case of autogamous plants, inbred progenies have been predominantly used since the evaluation of progenies happens concomitantly to the inbreeding, generating lines in the process, which is the ultimate goal of the program (LOPES et al., 2019; SOARES et al., 2020).

When inbred progenies are used in RS, in which the allelic frequency differs from $^1/_2$, a new component emerges in the genetic variance: D_1, which refers to the genetic covariance between the additive effects of the alleles and the dominance effects of the homozygotes (SOUZA JÚNIOR, 2001). As the component is a covariance, it may assume positive or negative values. In addition, the greater the inbreeding level of the progeny is, the larger the contribution of D_1, being able to reach $4D_1$ when evaluating the lines. Since this component is part of the numerator of the equation for gain from selection (GS), when the estimate of D_1 is negative,

the selection may reduce the average phenotypic expression of the population/line instead of improving it.

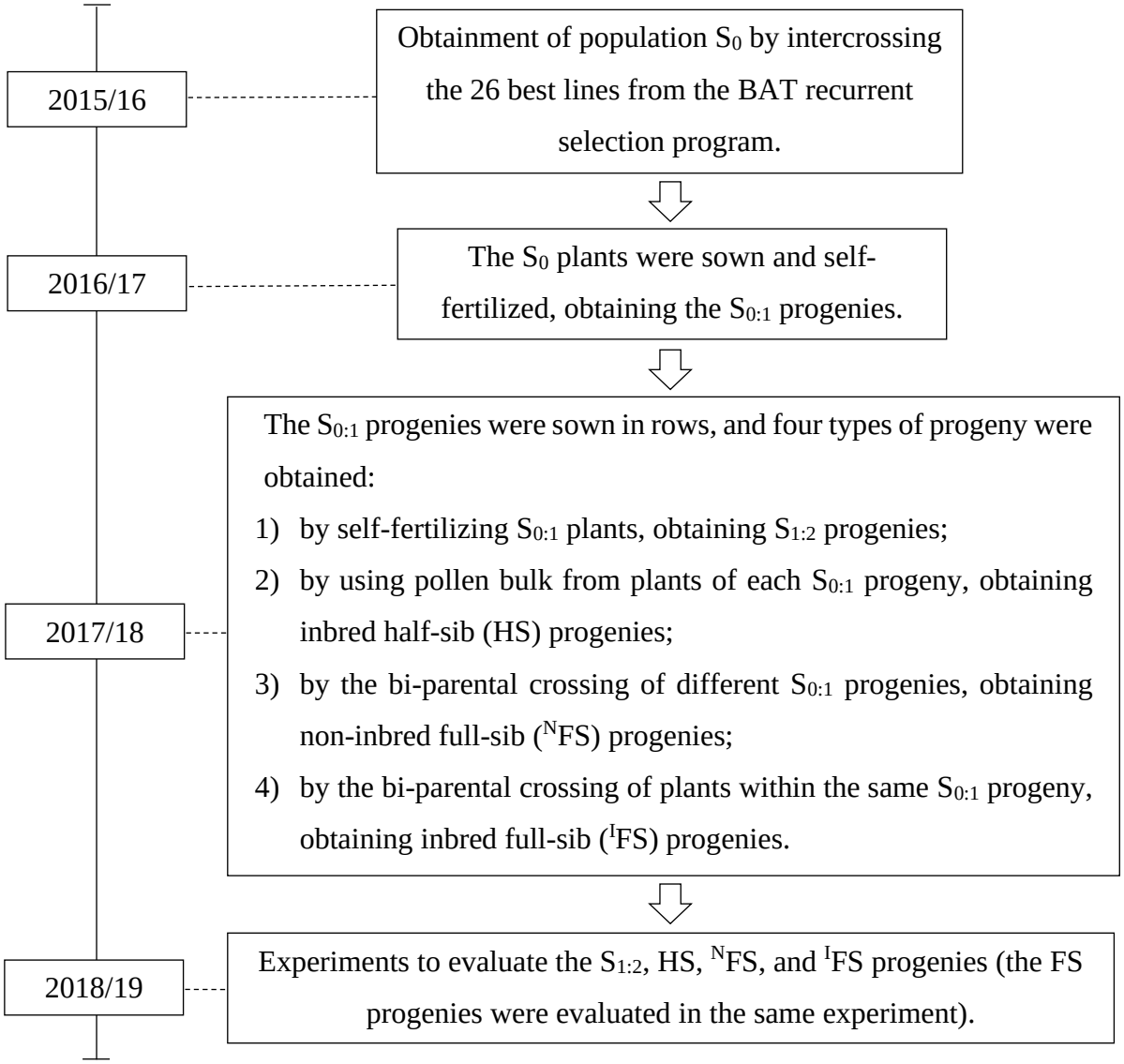
In light of the above, it is important to estimate the contribution of different genetic variance components to RS. Estimates of the additive genetic variance (σ_A^2) and the dominance variance (σ_D^2) are often reported in the literature for several species (HALLAUER; CARENA; MIRANDA FILHO, 2010; MORAIS et al., 1997; NOVOSELOVIC et al., 2004; SOUZA; RAMALHO, 1995). For tobacco, σ_A^2 and σ_D^2 estimates have been obtained over time, especially in Europe a few decades ago (MATHER, 1949; MATZINGER; MANN; COCKERHAM, 1962; ROBINSON; MANN; COMSTOCK, 1954). However, none of these studies were carried out in Brazil. Regarding the D_1 estimates, no reports were found for tobacco, although, for other plant species, some studies obtained negative estimates of D_1 (MORAIS et al., 1997; SOUZA; RAMALHO, 1995).

In tobacco and other autogamous species, it is possible to use different types of progeny and even to obtain DHs; thereby, it would be important to compare RS strategies involving progeny types in the evaluation and recombination stages. For this, the gain from selection (GS) should be estimated. This procedure has been mostly adopted in allogamous plants (HALLAUER; CARENA; MIRANDA FILHO, 2010). However, for autogamous plants, especially using inbred progenies in the evaluation, no reports were found. Thus, the present study was carried out to estimate the genetic variance components in an RS population of flue-cured Virginia (FCV) tobacco and compare different strategies for conducting RS applicable to most plant species.

2 MATERIAL AND METHODS

The experimental data used in this study were kindly provided by British American Tobacco (BAT) Brazil. The progenies were obtained and the experiments were conducted at the BAT experimental station located in Mafra (altitude 848 m, latitude 26˚10' S, and longitude 49˚48' W), in the state of Santa Catarina, southern Brazil. The steps of the research carried out in the field at the experimental station are shown in Figure 2.1.

Figure 2.1 – Workflow of the steps carried out in the research.



Source: from the author (2021).

As shown in Figure 2.1, the $S_{0:1}$ progenies were initially obtained from the RS program of flue-cured Virginia (FCV) tobacco. From the $S_{0:1}$ progenies, four types of progeny were obtained: $S_{1:2}$, by inbreeding $S_{0:1}$; inbred half-sibs (HS), using a pollen bulk collected from

plants of the same $S_{0:1}$ progeny; and full-sibs (FS), by crossing pairs of $S_{0:1}$ plants. When the crossing was carried out within the same progeny, inbred full-sibs (^{I}FS) were obtained, and when it was performed with pairs of plants from different $S_{0:1}$ progenies, non-inbred full-sibs (^{N}FS) were generated.

In the 2018/19 harvest, 126 progenies were evaluated, with 42 progenies of each type ($S_{1:2}$, HS, and FS). The FS progenies consisted of 22 ^{N}FS and 20 ^{I}FS. For each type of progeny, the experiments were conducted in a randomized block design with three replicates, totalizing 49 treatments, 42 progenies, and seven commercial controls of FCV tobacco. The controls were the same for the three experiments, which were contiguous, constituting three close areas in the field.

The plots were formed by a row with ten plants, with a spacing of 1.35 m between rows and 0.70 m between plants. The cultural practices were carried out according to the technical recommendations of the company for tobacco. At approximately 75 to 90 days after transplanting the seedlings to the experimental area, the plants were topped. Subsequently, the plants were harvested staggered, evaluating the green leaf yield (GLY) of the plots in kg. The total sugar content (TSC, in %), total alkaloid content (TAC, in %), and average leaf width (LW, in cm) were also evaluated.

Prior to analyzing the data, an adjustment was made by replicate of the phenotypic data in each type of progeny according to the performance of the seven common controls. Then, the data for each progeny type were subjected to an analysis of variance (TABLE 2.1) using the following model:

$$Y_{ij} = m + p_i + r_j + e_{ij},$$

where Y_{ij} refers to the observed value of the plot that received progeny *i* in replicate *j*, m is the overall mean, p_j is the effect of progeny *i*, with $p_i \sim N(0, \sigma_p^2)$, r_j is the effect of replicate *j*, with $r_j \sim N(0, \sigma_r^2)$, and e_{ij} is the experimental error related to the observation Y_{ij}, with $e_{ij} \sim N(0, \sigma_E^2)$.

Aiming to compare the means of the progeny types, a grouping was performed by employing the test by Scott and Knott (1974), using the estimate of the average error of the different experiments. The selective accuracy ($r_{g\hat{g}}$) was estimated by $r_{g\hat{g}} = \sqrt{1 - \frac{1}{F}}$.

The estimates of the coefficients of phenotypic and genotypic correlation among the traits, disregarding the type of progeny, were obtained from the following estimators (RAMALHO; FERREIRA; de OLIVEIRA, 2012):

$$r_{F_{XY}} = \frac{cov_{F_{XY}}}{\sqrt{\sigma_{F_X}^2 \sigma_{F_Y}^2}},$$

$$r_{G_{XY}} = \frac{COV_{G_{XY}}}{\sqrt{\sigma^2_{G_X}\sigma^2_{G_Y}}},$$

where $COV_{F_{XY}}$ and $COV_{G_{XY}}$ refer to the phenotypic and genotypic covariances, respectively, between traits *X* and *Y*, V_{F_X} and V_{F_Y} are the phenotypic variances for traits *X* and *Y*, and V_{G_X} and V_{G_Y} correspond to the genotypic variances for traits *X* and *Y*.

The correlation estimates involved data from all progenies. The $COV_{F_{XY}}$ was obtained from the covariance between the means of traits *X* and *Y*, whereas $COV_{G_{XY}}$ was obtained by the estimated covariance between the value of trait *X* in replicate *j* and trait *Y* in replicate *j'* or *j"* (RAMALHO et al., 2012). Student's t-test at the 5% significance level was used to verify whether the phenotypic correlations differed from zero.

Table 2.1 – Scheme of the analysis of variance and expected values of the mean squares E(MS).

SV	DF	MS	E(MS)	Calc F
HS	$i^{HS} - 1$	Q_1	$\sigma^2_{E_{HS}} + r\sigma^2_{HS}$	Q_1/Q_2
HS Error	$(i^{HS} - 1)(j - 1)$	Q_2	$\sigma^2_{E_{HS}}$	
$S_{1:2}$	$i^{S1:2} - 1$	Q_3	$\sigma^2_{E_{S1:2}} + r\sigma^2_{S_{1:2}}$	Q_3/Q_4
$S_{1:2}$ Error	$(i^{S1:2} - 1)(j - 1)$	Q_4	$\sigma^2_{E_{S1:2}}$	
FS	$i^{FS} - 1$	Q_5	$\sigma^2_{E_{FS}} + r\sigma^2_{FS}$	Q_5/Q_9
^{N}FS	$i^{N_{FS}} - 1$	Q_6	$\sigma^2_{E_{FS}} + r\sigma^2_{N_{FS}}$	Q_6/Q_9
^{I}FS	$i^{I_{FS}} - 1$	Q_7	$\sigma^2_{E_{FS}} + r\sigma^2_{I_{FS}}$	Q_7/Q_9
^{N}FS vs ^{I}FS	$(i^{FS} - 1) - \left[\left(i^{N_{FS}} - 1\right) + \left(i^{I_{FS}} - 1\right)\right]$	Q_8	-	Q_8/Q_9
FS Error	$(i^{FS} - 1)(j - 1)$	Q_9	$\sigma^2_{E_{FS}}$	

Number of progenies (*i*), number of replicates (*j*).

Source: from the author (2021).

Estimates of the phenotypic variance among means (σ^2_F) of progenies HS, $S_{1:2}$, ^{N}FS, and ^{I}FS were obtained by Q_1/j, Q_3/j, Q_6/j, and Q_7/j, respectively. The estimates of the confidence interval (CI) were calculated using the following equation:

$$CI_{1-\alpha}: \left[\frac{(n-1)\sigma^2_F}{x^2_{(\alpha/2;n-1)}}; \frac{(n-1)\sigma^2_F}{x^2_{(1-\alpha/2;n-1)}}\right],$$

where $x^2_{(\alpha/2;n-1)}$ and $x^2_{(1-\alpha/2;n-1)}$ are the values of the chi-square distribution with $n-1$ degrees of freedom.

With the estimates of phenotypic variance (σ^2_F) and errors (σ^2_E), the genetic variance components were estimated for all traits using two models. Model 1 contained the additive variance (σ^2_A), dominance variance (σ^2_D), D_1, and σ^2_E. In turn, only σ^2_A and σ^2_E were present in

Model 2. The demonstrations of the equations showing the genetic variance components for the different progeny types may be seen in Appendix B.

The genetic variance components were estimated using the iterative weighted least squares method, repeating the variance estimations until the estimated *Y* vector was better fitted to the observed value, detected by the stabilization of the coefficient of determination (R^2), using the following estimator:

$$\beta = (X'W^{-1}X)^{-1}X'W^{-1}Y,$$

where β is the vector of parameters, Y is the vector of observed variances, X refers to the matrix of known coefficients according to each model, and W is the diagonal weighting matrix obtained from the variance of one variance, as follows:

$$Y = \begin{bmatrix} \sigma^2_{F_{HS}} \\ \sigma^2_{F_{S1:2}} \\ \sigma^2_{F_{N_{FS}}} \\ \sigma^2_{F_{I_{FS}}} \\ \sigma^2_{E_{HS}} \\ \sigma^2_{E_{S1:2}} \\ \sigma^2_{E_{FS}} \end{bmatrix}$$

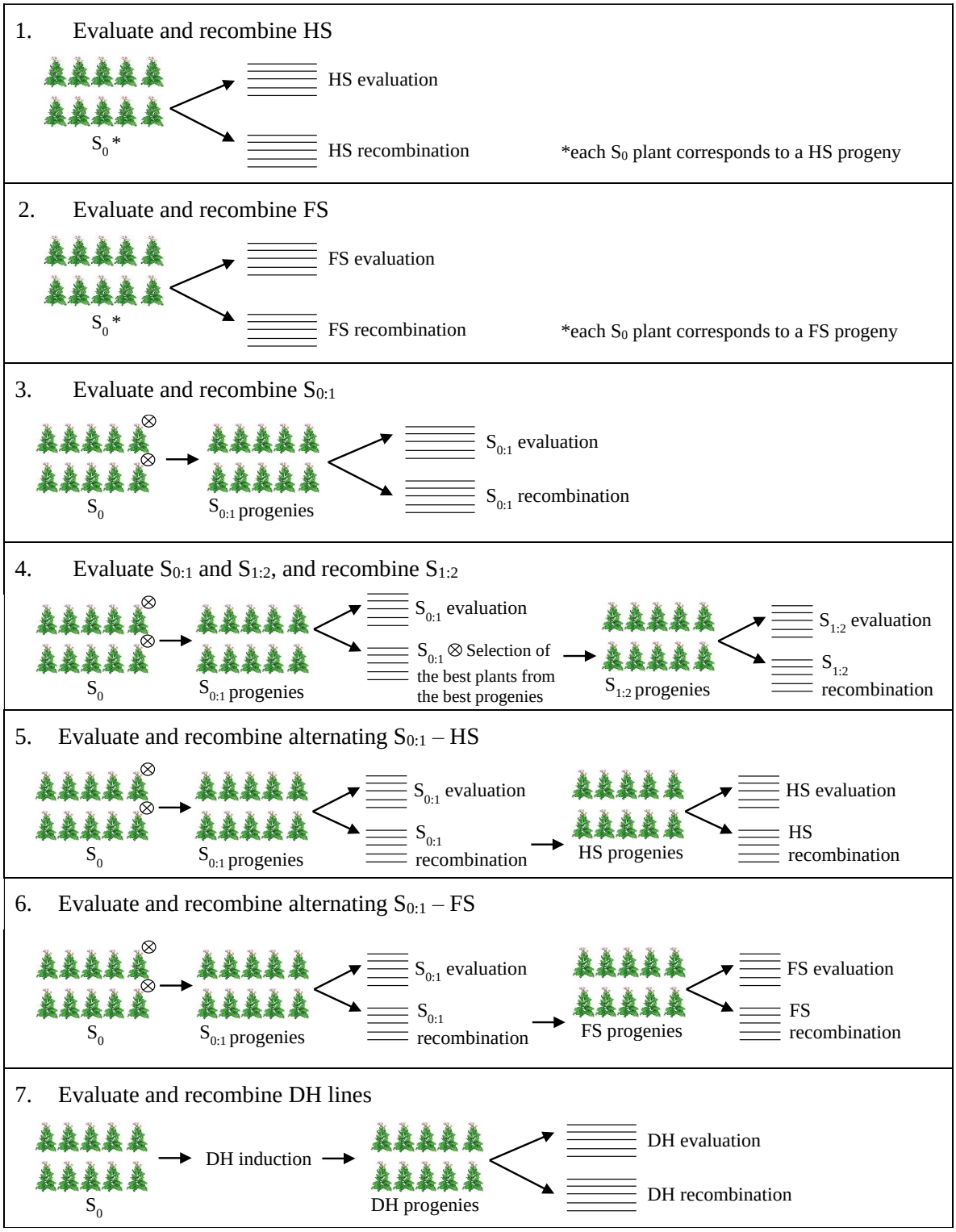
$$X = \begin{array}{c} \begin{matrix} \sigma^2_A & \sigma^2_D & D_1 & \sigma^2_{E_{HS}} & \sigma^2_{E_{S1:2}} & \sigma^2_{E_{FS}} \end{matrix} \\ \begin{bmatrix} 3/8 & 0 & 0 & 1/3 & 0 & 0 \\ 3/2 & 1/8 & 5/2 & 0 & 1/3 & 0 \\ 1/2 & 1/4 & 0 & 0 & 0 & 1/3 \\ 3/4 & 9/16 & 0 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

$$W = \begin{bmatrix} \frac{DF_{HS}+2}{2\sigma^2_{F_{HS}}} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{DF_{S_{1:2}}+2}{2\sigma^2_{F_{S1:2}}} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{DF_{N_{FS}}+2}{2\sigma^2_{F_{N_{FS}}}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{DF_{I_{FS}}+2}{2\sigma^2_{F_{I_{FS}}}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{DF_{E_{HS}}+2}{2\sigma^2_{E_{HS}}} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{DF_{E_{S1:2}}+2}{2\sigma^2_{E_{S1:2}}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{DF_{E_{FS}}+2}{2\sigma^2_{E_{FS}}} \end{bmatrix}$$

The heritability (h^2) estimates for each type of progeny ($S_{1:2}$, HS, and FS) were obtained by estimator $h^2 = {\sigma^2_G}/{\sigma^2_F}$. Lastly, the gain from selection (GS) per plant was estimated for the different strategies used in tobacco RS (FIGURE 2.2). As the GS was calculated per plant, the variance estimates were divided by the number of plants per plot (ten plants). The estimator used depends on the estimates of σ^2_A and D_1. The selection intensity applied was 10%, i.e., $i = 1.755$, and parental control was performed in both male and female plants, with $c = 1$ for all strategies.

All statistical analyses were performed in "R" (R CORE TEAM, 2020), and the genetic variance components were estimated using software "MAPGEN" (FERREIRA; ZAMBALDE, 1997).

Figure 2.2 – Schemes of strategies for conducting recurrent selection in a tobacco program.



Source: from the author (2021).

3 RESULTS

The accuracy estimates ($r_{g\hat{g}}$) ranged from 0.42 to 0.83 (TABLE 3.1). The evaluation of the $S_{1:2}$ progenies presented the highest estimates of $r_{g\hat{g}}$ for all evaluated traits except the GLY, which was greater in the FS progenies. As expected, the significance of the source of variation "progenies" is associated with the estimate of $r_{g\hat{g}}$. The significance varied among the progeny types and evaluated traits (TABLE 1, APPENDIX A). Only for the TAC were significant differences ($p \leq 0.05$) found for the three types of progeny studied. Concerning the GLY and LW, the F-test was not significant for the HS progenies. For the TSC, the source of variation "progenies" was only significant when the $S_{1:2}$ progenies were evaluated. These results reflect the frequency distribution of means for each progeny type. As expected, the $S_{1:2}$ progenies showed greater variation (FIGURE 3.1).

Table 3.1 – Accuracy ($r_{g\hat{g}}$) and mean estimates of the experiments evaluating the inbred half-sib (HS), $S_{1:2}$, and full-sib (FS) for green leaf yield (GLY), total sugar content (TSC), total alkaloid content (TAC), and leaf width (LW).

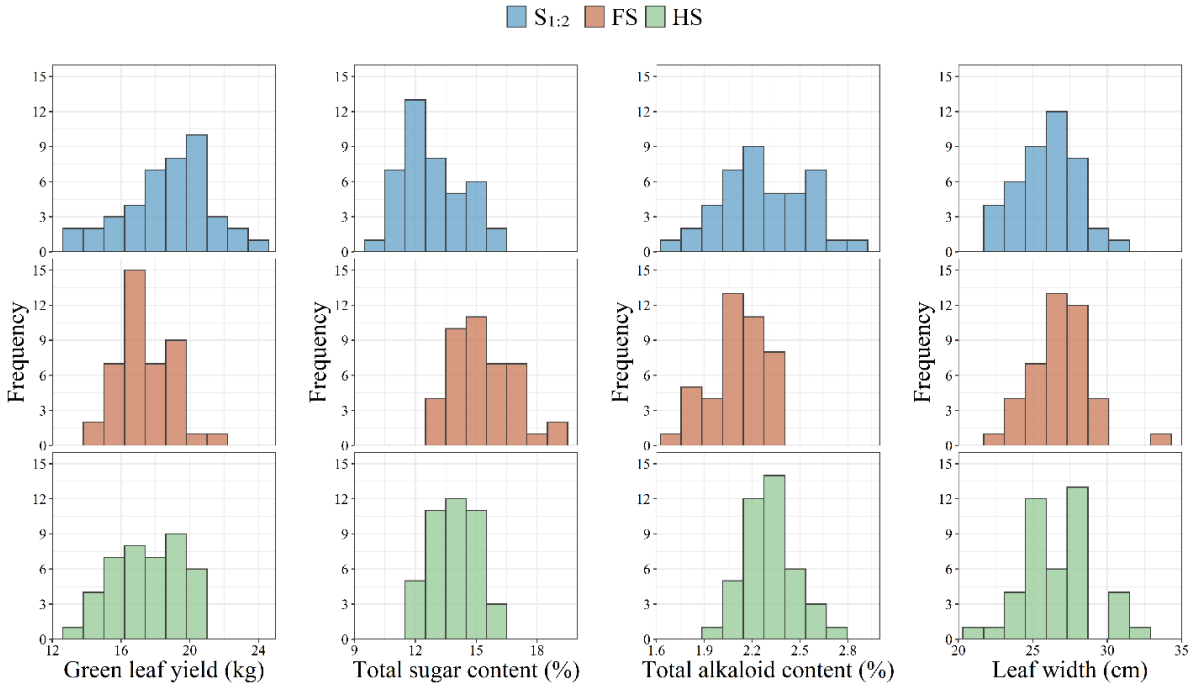
	Progenies	GLY	TSC	TAC	LW
$r_{g\hat{g}}$	**HS**	0.42	0.51	0.65	0.55
	$S_{1:2}$	0.61	0.82	0.83	0.74
	FS	0.71	0.48	0.80	0.61
Mean[1]	**HS**	17.43 b	13.91 b	2.31 a	26.65 a
	$S_{1:2}$	18.78 a	12.93 c	2.26 a	26.01 a
	FS	17.40 b	15.34 a	2.11 b	26.86 a

[1] Means followed by the same letter belong to the same group according to the test by Scott and Knott (1974) $\alpha = 5\%$.

Source: from the author (2021).

The means of the progeny types were grouped using the Scott-Knott test (TABLE 3.1). According to the test, the $S_{1:2}$ progenies presented a higher GLY average than the HS and FS progenies, which belong to the same group. Concerning the TSC, a greater value was found for the FS progenies, followed by the HS progenies. For the TAC, the HS and $S_{1:2}$ progenies showed the highest averages.

Figure 3.1 – Frequency distribution of means of the $S_{1:2}$, full-sib (FS), and half-sib (HS) for green leaf yield (kg plot^{-1}), total sugar content (%), total alkaloid content (%), and leaf width (cm).



Source: from the author (2021).

In general, significant phenotypic correlation estimates were obtained for most of the pairs of traits evaluated, except between the TSC and LW and the TAC and LW (TABLE 3.2). The GLY showed significant correlations with all traits, yet with small magnitudes. The estimates of genotypic correlation ranged from 0.13 to 0.47. Negative estimates were found between the GLY and TSC, the TSC and TAC, and the TAC and LW. The genotypic correlations involving the GLY were similar to the phenotypic correlation estimates.

Table 3.2 – Estimates of phenotypic (upper diagonal) and genotypic (lower diagonal) correlation coefficients among green leaf yield (GLY), total sugar content (TSC), total alkaloid content (TAC), and leaf width (LW).

	GLY	**TSC**	**TAC**	**LW**
GLY	-	-0.34 *	0.32 *	0.27 *
TSC	-0.35	-	-0.49 *	-0.05 ^{ns}
TAC	0.27	-0.47	-	-0.05 ^{ns}
LW	0.28	0.13	-0.33	-

* and ^{ns}: significant ($p \leq 0.05$) and non-significant ($p > 0.05$) according to Student's t-test, respectively.
Source: from the author (2021).

The estimates of phenotypic variance (σ_F^2) and error variance (σ_E^2) differed among the progeny types studied (TABLE 3.3). For the GLY and TAC, the most considerable σ_F^2 estimate

in absolute value was found in the $S_{1:2}$ progenies. However, the confidence intervals overlap among the types of progeny. In this situation, the absolute value does not necessarily indicate that the estimates differ. In turn, for the other traits, TSC and LW, a larger absolute value of σ_F^2 was revealed for the ^{N}FS progenies.

Table 3.3 – Estimates of phenotypic variance (σ_F^2) and error variance (σ_E^2) for green leaf yield (GLY), total sugar content (TSC), total alkaloid content (TAC), and leaf width (LW) in the inbred half-sib (HS), $S_{1:2}$, non-inbred full-sib (^{N}FS), and inbred full-sib (^{I}FS) progenies.

		GLY	TSC	TAC	LW
$\widehat{\sigma}_F^2$	**HS**	4.05 [2.80;6.80][1]	1.38 [0.96;2.32]	0.03 [0.02;0.05]	5.21 [3.60;8.74]
	$S_{1:2}$	6.33 [4.37;10.62]	2.19 [1.52;3.68]	0.07 [0.05;0.12]	3.69 [2.55;6.20]
	^{N}FS	2.83 [1.64;6.03]	2.48 [1.43;5.28]	0.03 [0.02;0.07]	5.38 [3.11;11.47]
	^{I}FS	2.29 [1.35;4.67]	1.94 [1.15;3.97]	0.03 [0.02;0.07]	2.33 [1.38;4.76]
$\widehat{\sigma}_E^2$	**HS**	9.96	3.09	0.05	10.81
	$S_{1:2}$	11.85	2.20	0.07	4.98
	FS	3.83	5.63	0.04	7.01

[1] Confidence interval of the means at a 5% probability.

Source: from the author (2021).

Regarding the error variance (σ_E^2), lower estimates for the GLY and TAC were found for the FS progenies, while, for the TSC and LW, the $S_{1:2}$ progenies presented the lowest estimates of σ_E^2. The estimates of σ_F^2 and σ_E^2 for the different progenies were used to estimate the genetic variance (σ_G^2) components for the traits under study.

Estimates of the genetic variance components using Model 1, including σ_A^2, σ_D^2, and D_1, presented high values for the coefficient of determination (R^2), indicating that this model was well-fitted to the results of the experimental tests (TABLE 3.4). However, the estimates of the dominance genetic variance (σ_D^2) were negative or null for all traits. Thus, the estimates of genetic variance using Model 2 were obtained without considering σ_D^2 and D_1. Even considering only the σ_A^2 as responsible for all the inheritable variation, the majority of the R^2 values were higher than 99%, indicating that this component explained most of the observed genetic variance. Estimates of σ_E^2 obtained for each type of progeny had different magnitudes, although the confidence intervals overlapped in some situations.

Table 3.4 – Estimates of genetic and environmental variance using Model 1 (considering all components of σ_G^2) and Model 2 (considering only σ_A^2 as a component of σ_G^2) for green leaf yield (GLY), total sugar content (TSC), total alkaloid content (TAC), and leaf width (LW).

		GLY		TSC		TAC		LW	
Model 1	$\widehat{\boldsymbol{\sigma}}_{\boldsymbol{A}}^{\boldsymbol{2}}$	3.15	[2.41;4.32][1]	1.05	[0.77;1.51]	0.03	[0.03;0.05]	7.25	[5.39;10.27]
	$\widehat{\boldsymbol{\sigma}}_{\boldsymbol{D}}^{\boldsymbol{2}}$	-2.06	[-1.59;-2.78]	-1.12	[-0.85;-1.54]	0.00	[0.00;0.00]	-9.25	[-7.04;-12.70]
	$\boldsymbol{D_1}$	-0.84	[-0.66;-1.10]	0.01	[0.01;0.02]	0.00	[0.00;0.00]	-3.07	[-2.31;-4.30]
	$\widehat{\boldsymbol{\sigma}}_{\boldsymbol{E_{HS}}}^{\boldsymbol{2}}$	9.67	[7.38;13.23]	3.07	[2.31;4.27]	0.05	[0.03;0.06]	10.35	[7.85;14.27]
	$\widehat{\boldsymbol{\sigma}}_{\boldsymbol{E_{S1:2}}}^{\boldsymbol{2}}$	11.85	[8.92;16.51]	2.20	[1.66;3.07]	0.07	[0.05;0.09]	4.98	[3.75;6.94]
	$\widehat{\boldsymbol{\sigma}}_{\boldsymbol{E_{FS}}}^{\boldsymbol{2}}$	3.89	[2.94;5.40]	5.73	[4.37;7.86]	0.04	[0.03;0.05]	7.30	[5.53;10.09]
	$\boldsymbol{R^2}$	99.64		99.84		99.74		98.53	
Model 2	$\widehat{\boldsymbol{\sigma}}_{\boldsymbol{A}}^{\boldsymbol{2}}$	1.82	[1.42;2.40]	0.92	[0.68;1.32]	0.03	[0.03;0.04]	1.53	[1.13;2.19]
	$\widehat{\boldsymbol{\sigma}}_{\boldsymbol{E_{HS}}}^{\boldsymbol{2}}$	10.00	[7.90;13.07]	3.10	[2.44;4.08]	0.05	[0.04;0.06]	11.68	[9.23;15.25]
	$\widehat{\boldsymbol{\sigma}}_{\boldsymbol{E_{S1:2}}}^{\boldsymbol{2}}$	11.69	[9.00;15.81]	2.22	[1.67;3.08]	0.07	[0.05;0.09]	4.92	[3.72;6.81]
	$\widehat{\boldsymbol{\sigma}}_{\boldsymbol{E_{FS}}}^{\boldsymbol{2}}$	3.92	[3.00;5.34]	5.47	[4.31;7.19]	0.04	[0.03;0.05]	7.45	[5.84;9.83]
	$\boldsymbol{R^2}$	99.42		99.62		99.73		96.45	

Estimates of additive genetic variance ($\hat{\sigma}_A^2$), dominance genetic variance ($\hat{\sigma}_D^2$), genetic covariance of the additive and homozygous dominance effects (D_1), variance of the mean error of HS progenies ($\hat{\sigma}_{E_{HS}}^2$), variance of the mean error of $S_{1:2}$ progenies ($\hat{\sigma}_{E_{S1:2}}^2$), variance of the mean error of FS progenies ($\hat{\sigma}_{E_{FS}}^2$), and coefficient of determination (R^2).

[1] Confidence interval of the means at a 5% probability.

Source: from the author (2021).

Estimates of h^2 were calculated for each type of progeny involving only the different σ_G^2 among progenies (TABLE 3.5). The estimates of h^2 differed among traits and progeny types, with a higher estimate being observed in the $S_{1:2}$ progenies for all traits except the GLY, as expected. These results were very consistent with the accuracy estimates.

Table 3.5 – Heritability estimates for selection among means of inbred half-sib (h_{HS}^2), $S_{1:2}$ ($h_{S1:2}^2$), and full-sib (h_{FS}^2) progenies for green leaf yield (GLY), total sugar content (TSC), total alkaloid content (TAC), and leaf width (LW).

Progenies	GLY	TSC	TAC	LW
$\boldsymbol{h_{HS}^2}$	0.18	0.26	0.43	0.31
$\boldsymbol{h_{S1:2}^2}$	0.38	0.67	0.69	0.55
$\boldsymbol{h_{FS}^2}$	0.50	0.23	0.64	0.37

Source: from the author (2021).

With the estimation of the variance components for the GLY, the gain from selection (GS) was estimated (TABLE 3.6) for the different strategies that may be applied to tobacco

(FIGURE 2.2). The values were obtained by the estimator $GS = ic \sigma_A^2 \Big/ \sqrt{\sigma_F^2}$, where i refers to the selection intensity and c to parental control. The GS estimates varied among the different strategies proposed for conducting an RS program. When considering the total GS in percentage (based on the overall mean), disregarding the duration of each cycle, the estimates ranged from 6.54% to 52.30%. Higher values were found for the use of doubled haploid (DH) lines, whereas the smallest values occurred for the HS progenies. As expected, when considering the GS per unit of time (year), the variation among the estimates decreased, with values ranging from 6.54% to 17.43%. Similarly, a high value of GS was estimated for the DH lines, and a lower number was found for the HS progenies in the evaluation and recombination.

Table 3.6 – Estimates of total gain from selection (GS) and gain from selection per year (GS year^{-1}) for the different types of progenies evaluated in the RS for green leaf yield of FCV tobacco.

Type of progeny evaluated	$\boldsymbol{\sigma_A^2}$	$\boldsymbol{t}$	$\boldsymbol{GS}$	$\boldsymbol{GS\ year^{-1}}$	$\boldsymbol{N_e}$
HS	$1/4$	1	0.12 (6.54)[1]	0.12 (6.54)	4
FS	$1/2$	1	0.23 (13.08)	0.23 (13.08)	2
$S_{0:1}$	1	2	0.47 (26.15)	0.23 (13.08)	1
$S_{1:2}$	$3/2$	3	0.70 (39.23)	0.23 (13.08)	0.67
$S_{0:1}$ – HS[2]	$1 - 1/4$	3	0.58 (32.69)	0.19 (10.90)	-
$S_{0:1}$ – FS[2]	$1 - 1/2$	3	0.70 (39.23)	0.23 (13.08)	-
DH	2	3	0.93 (52.30)	0.31 (17.43)	-

Proportion of what is exploited of additive genetic variance (σ_A^2), time (t) in years for an RS cycle, and effective size (N_e). Selection intensity (10%): $i = 1.755$; $\sigma_A^2 = 0.182$ (g/plant)^2; $\sigma_F^2 = 0.465$ (g/plant)^2.
[1] Estimated gain from selection in percentage based on the overall mean.
[2] For evaluation and recombination strategies concerning $S_{0:1}$ and HS or FS progenies, two RS cycles were performed in three years.

Source: from the author (2021).

4 DISCUSSION

The reference population employed in this study was obtained by the BAT company from the intercrossing of the 26 best lines in an FCV tobacco breeding program. In this context, in population S_0, the alleles responsible for the phenotypic expression of the different traits may assume any value between 0 and 1. Unlike what occurs when the population stems from the crossing between two lines and, hence, the loci that would be segregating have an allelic frequency of $^1/_2$. When the allelic frequencies differ from $^1/_2$, with inbreeding, the genetic variance of the population, in addition to the additive variance (σ_A^2) and dominance variance (σ_D^2), also contains the components D_1 (genetic covariance of the additive and homozygous dominance effects), D_2 (genetic variance of the homozygous dominance effects), and $\breve{H}$ (inbreeding depression) (COCKERHAM; WEIR, 1984). In this case, to estimate the genetic components, it is clearly necessary to have at least five equations to calculate the four genetic parameters and, at the same time, test them. It was only possible to obtain variances among the progenies $S_{1:2}$, HS, ^{N}FS, and ^{I}FS; thus, the four equations only allowed to estimate three components, namely σ_A^2, σ_D^2, and D_1. In this situation, the components D_2 and $\breve{H}$ are included in σ_D^2.

In order to estimate the genetic components, genetic variance among progenies (σ_G^2) or phenotypic variance based on progeny means (σ_F^2) may be used. The latter option is preferable because it is expected to input less error into the estimate; therefore, it was used to estimate the components in this study. Souza and Ramalho (1995) adopted the same approach to estimate the genetic variance components in the common bean crop.

The σ_F^2 on a progeny means basis, in addition to the genetic variance among progenies (σ_G^2), contains the error variance (σ_E^2) (RAMALHO et al., 2012). In turn, (σ_E^2) depends on the environmental variance among plots (σ_e^2) and the phenotypic variance within plots (σ_d^2). Since σ_d^2 is a phenotypic variance, it has two components: the environmental variance (σ_w^2) and the genetic variance $\left(\sigma_{G_d}^2\right)$ within plots (RAMALHO et al., 2012). In tobacco, no report was found concerning the magnitude of σ_e^2 in relation to σ_w^2. However, in maize (HALLAUER; CARENA; MIRANDA FILHO, 2010; RUSSELL et al., 1978) and common beans (SOUZA; RAMALHO, 1995), for example, σ_w^2 is always greater than σ_e^2.

This study did not aim at estimating these two components. However, $\sigma_{G_d}^2$ depends on the type of progeny, so $\sigma_{G_d}^2$ for the HS progenies will contain $^9/_8\,\sigma_A^2 + {}^3/_4\,\sigma_D^2$, whereas for the

$S_{1:2}$ progenies, $\sigma^2_{G_{d_{S1:2}}} = {}^1/_4\,\sigma^2_A + {}^5/_{16}\,\sigma^2_D$. When considering ^{N}FS, $\sigma^2_{G_{d_{N_{FS}}}} = \sigma^2_A + {}^1/_2\,\sigma^2_D$, and, for ^{I}FS, $\sigma^2_{G_{d\ I_{FS}}} = {}^3/_4\,\sigma^2_A + {}^3/_{16}\,\sigma^2_D$. Therefore, as σ^2_E has different genetic constitutions within the progenies, the estimates (σ^2_E) were obtained separately for each progeny type (TABELA 3.3).

Selective accuracy $\left(r_{g\hat{g}}\right)$ depends on the existence of genetic variance among the progenies (σ^2_G) and the relative contribution of the experimental error (σ^2_E), σ^2_G relative to σ^2_E. To put it differently, $r_{g\hat{g}} = \sqrt{h^2} = \sqrt{1 - {}^1/_F}$ (de RESENDE; SILVA; AZEVEDO, 2014; de RESENDE; DUARTE, 2007), i.e., considering the same σ^2_e and σ^2_w, the accuracy will vary according to the type of progeny being evaluated. In the case of the HS progenies, as previously mentioned, considering only σ^2_A, $\sigma^2_{G_{HS}} = {}^3/_8\,\sigma^2_A$, whereas for the $S_{1:2}$ progenies, $\sigma^2_{G_{S1:2}} = {}^3/_2\,\sigma^2_A$, meaning that $r_{g\hat{g}_{S1:2}}$ is expected to be greater than $r_{g\hat{g}_{HS}}$, as observed for the four traits in this study (TABELA 3.1).

It could be argued that the σ^2_E for the HS progenies should be greater than for the $S_{1:2}$ progenies due to the difference in genetic variance within progenies. However, it should be noted that this component within the plot is divided by the number of plants that the plot contains (RAMALHO; FERREIRA, de OLIVEIRA, 2012). Thus, the contribution of this component is expected to be small. Estimates of σ^2_E (TABLE 3.3) were higher for the HS progenies than for the $S_{1:2}$ progenies except for the GLY, in which case the opposite occurred, and for the TAC, for which the estimates were the same. Therefore, it appears that it is not possible to infer which type of progeny was evaluated with greater precision.

The GLY is directly associated with the tobacco product that is commercialized: the mass of cured leaves. The GLY takes into account the mass of the freshly harvested leaves, which seemingly has greater experimental precision than the cured leaf mass. Even though no report was found to substantiate this observation, most tobacco studies use the GLY (REDDY; DWIVEDI; SHARMA, 2012; TANG et al., 2020; TOLEDO et al., 2013), including this research.

Surprisingly, a small magnitude of genetic correlation was estimated for the GLY and LW $(r_G = 0.28)$. In fact, the GLY depends on other traits such as the leaf number, width, and length. These two other traits (leaf number and length) likely explain the variation in the GLY more. No reports of estimates of the genetic correlations between leaf number and width and the GLY either. However, Robinson, Mann, and Comstock (1954) estimated the genetic correlation between leaf width and length with cured leaf yield in tobacco and found values of

0.49 and 0.44, respectively. Moreover, the authors obtained an even higher correlation estimate between leaf number and cured leaf yield ($r_G = 0.61$). Thereby, it appears that the leaf number is the trait that most explains the variation in tobacco yield. It should also be emphasized that, in this study, leaf number and length were also obtained, yet the data were not included in this article due to low accuracy estimates.

Progeny averages for the TSC ranged from 12.9% to 15.3%. Regarding the TAC, the variation was even smaller, from 2.1% to 2.3%. It is worth noting that, since the TSC and TAC were obtained as leaf dry matter percentages, it is harder to infer the magnitude of the value due to the dilution effect, given that all other treatment conditions were the same. The higher the leaf dry matter content is, the lower the percentage of the TSC and TAC, proportionally.

When estimating the genetic variance components, Model 1 (containing σ_A^2, σ_D^2, and D_1) presented an excellent fit, as demonstrated by the high estimates of R^2 (TABLE 3.4). However, the σ_D^2 values were null or negative. Thus, it was decided to estimate the σ_A^2 without considering σ_D^2 and D_1 (Model 2). Evidently, the σ_A^2 estimates were different from those found using Model 1 for all traits, while the estimates for R^2 were very similar between the models. The magnitude of σ_A^2 varied according to the trait. Nevertheless, it explained most of the genetic variance, which is often reported in the literature when the components of σ_G^2 are estimated in autogamous plants (BONATO; VELLO, 1999; MORAIS et al., 1997; MORETO; RAMALHO; ABREU, 2007; NOVOSELOVIC et al., 2004; SOUZA; RAMALHO, 1995). According to Bernardo (2020), regardless of the plant reproductive biology (autogamous or otherwise), the estimate of σ_A^2 is always greater than σ_D^2 depending on how it is estimated.

Tobacco was one of the pioneer species for which genetic variance component estimates were obtained. Especially in England, from 1949, tobacco was used as a model plant to estimate mean and variance components (MATHER, 1949). Since then, numerous other studies were published reporting the estimates of σ_A^2 and σ_D^2 (ROBINSON; MANN; COMSTOCK, 1954; MATZINGER; MANN; COCKERHAM, 1962). Moreover, several reports in the literature pointed out that, for some tobacco traits, the σ_D^2 explained very little of the variation observed among progenies (ROBINSON; MANN; COMSTOCK, 1954; MATZINGER; MANN; COCKERHAM, 1962). Yet, no reports were found involving the germplasm of tobacco cultivated in Brazil.

In tobacco crops, it is common to commercialize hybrids. In principle, one could think that heterosis (h) is expressive and that, in this condition, the contribution of loci in heterozygosis (d) would be fundamental for manifesting several traits and, consequently, the

σ_D^2 should differ from zero, contrary to what has been discussed so far. Nevertheless, h in tobacco is usually low. For the GLY, on the average of several crossings, h is less than 10% (CARVALHO, 2016; PSCHEIDT, 2020). However, even though h is a small-magnitude component, the use of hybrids is justified since it represents a way for the companies to protect their lines. In the male-sterility approach, a sterile male line is artificially crossed with its fertile male isogenic version. However, since an artificial cross is performed, it is preferable to use a sterile male line A and another fertile male line B. Thereby, the process exploits the hybrid vigor. Although the level of h is low, it brings an additional gain for the company along with germplasm protection. Additionally, it should be noted that some hypotheses explain the h without necessarily requiring a significant contribution from heterozygous loci (BARANWAL et al., 2012; JIANG et al., 2017; MELCHINGER et al., 2007; YAO et al., 2013).

As already mentioned, the D_1 component occurs when there is inbreeding and an allele frequency different from $^1/_2$ (SOUZA JÚNIOR, 1989). Considering an RS program, having information about the magnitude of D_1 is fundamental because it is related to the equation for gain from selection (GS). D_1 is part of the numerator of the GS equation, and its contribution tends to be greater as inbreeding increases, reaching $4D_1$ when the individuals are completely homozygous (SOUZA JÚNIOR, 1989). As component D_1 is a covariance, it may assume positive or negative values. When negative, instead of increasing the GS, it reduces it, which is obviously undesirable.

Considering only one locus, D_1 is given by $D_1 = -2pq(q-p)[a+(q-p)d]d$ (COCKERHAM; WEIR, 1984; SOUZA JÚNIOR, 1989). The component is dependent on the allele frequencies $(p \text{ and } q)$ and the contribution of the homozygous loci (a) and heterozygous loci (d). Regarding the allele frequency, in simulation studies, Crisóstomo (1989, cited by SOUZA; RAMALHO, 1995) and Fernandes (1990, cited by SOUZA; RAMALHO, 1995) found that, when the allele frequency of the population is lower than $^1/_2$, the estimate of D_1 is negative. However, as the reference population used in this study was obtained by crossing the best lines of an RS program, it is unlikely that the allele frequency of the population is less than $^1/_2$ and, thus, D_1 was expected to be positive. As already mentioned, the contribution of d for the traits evaluated should be null or very small in this study, as confirmed by the results.

In the literature, estimates of D_1 are uncommon regardless of the species. For tobacco, no reference was found, while estimates of D_1 were negative for rice and beans (MORAIS et al., 1997; SOUZA; RAMALHO, 1995). From the estimates obtained in the RS population of the varietal group Virginia evaluated in this study, one may infer that D_1 should not affect

significantly the RS program that has been carried out, even evaluating doubled haploid (DH) lines that better exploit the component D_1 ($4D_1$ in relation to the reference population S_0 or F_2) (SOUZA JÚNIOR, 1989, RAMALHO et al., 2012).

With these estimates, it is possible to assess which would be the best selection strategy for conducting the RS program. Although tobacco is an autogamous plant, there is no restriction on using any type of progeny, which is uncommon in other autogamous species. In tobacco, the artificial hybridization process is easy since the plant produces countless flowers with copious amounts of seeds, even when artificial pollination is performed. Another great advantage is that the pollen may be easily stored. Moreover, the pollen bulk may be obtained efficiently as it is easy to mix pollen stored in similar proportions, an ideal condition for obtaining HS progenies, for example, which is not possible for most cultivated species. It also merits emphasis that DH lines in tobacco crops may be obtained both by in vitro and in vivo processes (HANCOCK et al., 2015; SOOD et al., 2013).

RS consists of any cyclical breeding process that involves obtaining progenies, evaluating them, and recombining the best ones (RAMALHO et al., 2012). This breeding approach has been stimulated in tobacco for some reasons, including the ease of obtaining any type of progeny. RS makes the activities of breeders much more dynamic: for instance, it decreases the time needed annually to choose the bi-parental crossings to be performed. In addition, there is no need to obtain multiple segregating populations to be advanced. Thus, less work is needed to conduct segregating populations, and the number of notes and possible errors in genealogy reduces drastically. The RS process is as dynamic as any other; if a new line is generated, it may and should be incorporated into the recombination of the next cycle. The rules for the registration and protection of cultivars do not require specific inclusion regarding genealogy; one may simply inform that the new cultivar comes from an RS program. It is also useful for the company to use the RS because, in the case of a dismissed leader, the program may be easily conducted by another person since the only crucial aspect is the reference population, i.e., the last recombined generation. Other advantages from the genetical perspective have been discussed by Bernardo (2020), Fouilloux and Bannerot (1988), Geraldi (1997), and Ramalho et al. (2012).

Based on what has been reported, the most important step would be to define the best strategy for conducting an RS program. Some alternatives may be applied to compare the different selection methods. The first would be the direct process, conducting selective cycles under field conditions using different methods. This approach has been used little due to the difficulty of conducting several methods simultaneously, besides the restriction in the

generalization of results. There are very few reports in the literature on its employment, with one having been carried out with common beans (RAPOSO; RAMALHO; ABREU, 2000) and another with maize using S_1 progenies and DH lines (BORDES et al., 2007). Another alternative would be to use data from the literature in which the same population is submitted to RS cycles using different selection strategies. In this case, the drawbacks involve the differences in the selection intensities used in each situation and the experimental details in the evaluation of progenies (HALLAUER; CARENA; MIRANDA FILHO, 2010). The third alternative, more versatile than the others, would be to use the gain from selection (GS), also called the breeder's equation (COBB et al., 2019; XU et al., 2017), to simulate the use of different strategies for conducting RS. Thus, based on the GS estimate, predicting and choosing the best RS strategy for a given condition is possible. This was the alternative employed in this study since tobacco allows the use of any type of progeny with similar difficulty levels.

One requirement to apply the breeder's equation is to know the genetic variance for the different types of progeny. Considering that σ_D^2 and D_1 are null or not expressive in the genetic manifestation in tobacco, the variance analysis data were used to have the estimates for predicting GS. Moreover, to make the estimates comparable in any situation, GS was estimated per plant. An important variable in the breeder's equation is the time required to conduct each cycle. The time varies according to the type of progeny evaluated and/or recombined (TABLE 3.6). Thus, the focus of the discussion, from now on, will be directed to GS per year, rendering the different procedures directly comparable.

The lowest estimate of the annual GS, in percentage, was obtained for the HS progenies, being evaluated and recombined. It is worth noting that, for tobacco, it is possible to conduct one RS cycle per year. For instance, three replicates are used for evaluation, and a fourth replicate is implanted one month after the evaluation experiment for the recombination process (FIGURE 2.2). After harvesting the experiment, the data are analyzed, and the progenies to be recombined are identified. Note that, in the evaluation experiment, the plants are topped just as in a commercial field, whereas, in the recombination batch (fourth replicate), this process is not carried out. In the recombination batch, the unselected progenies are eliminated, and the recombinant seeds are obtained through a pollen bulk. Thus, the S_0 plants of the next cycle are generated, with the offspring of each plant constituting an HS progeny to be evaluated in the next cycle.

It is important to stress that, in none of the strategies, the GS was estimated within the progenies due to the lack of a phenotypic variance estimate within the plots. However, it may and must be performed visually in the recombination process. For example, in the case of HS

progenies, in the recombination batch, it is possible to choose the best plants from each HS progeny to donate the pollen and to be pollinated by the bulk. Even though it is a mass selection, it is carried out in both parents since the trait under selection, the GLY, is expressed before flowering. This is impossible for species the commercial products of which are expressed after flowering, such as maize.

For example, if the objective is to evaluate 200 HS progenies in each cycle and the selection intensity among progenies is 10%, a total of twenty progenies would be selected. Thus, in order to have the 200 progenies again, it would be necessary to select ten plants within each selected progeny. In this case, as mentioned above, the selection is carried out on both the male and female sides, and, since within HS progenies there is $^3/_4\ \sigma_A^2$, all the σ_A^2 available is exploited by the selection within. Thus, if the visual selection is efficient, it is expected that, for HS progenies, the gain from selection within may even be greater than between. It is worth emphasizing that, among the various strategies tested for conducting RS, the greatest gain within is expected for HS progenies (RAMALHO, 1977 cited by HALLAUER; CARENA; MIRANDA FILHO, 2010), and, evidently, the gain within is null for DH lines.

When using FS progenies, the GS per year is higher than for HS progenies. However, it may be harder to direct the crossing pairs to be recombined, which requires some time and many notes. Additionally, on average, the effective size of FS progenies is half that of HS progenies; thus, it would be necessary to evaluate twice the number of FS progenies than HS progenies to maintain the same size (SOUZA JÚNIOR, 2001). The issue of effective size has been widely discussed, and evidence shows that, even with a small effective size, it is possible to continue to have long-term gains with RS (HALLAUER; CARENA; MIRANDA FILHO, 2010). Furthermore, it must be emphasized that, in autogamous plants, the issue of effective size is not very limiting since, at any time of recombination, new lines that have stood out in other situations may be included in a new selective cycle (RAMALHO et al., 2012).

When evaluating self-fertilized progenies, i.e., $S_{0:1}$ progenies, two harvests are required for each cycle, two years in the case of tobacco: one to obtain progenies and the other to evaluate and recombine them. The recombination of $S_{0:1}$ progenies may be done using HS or FS progenies without affecting the result because, in both cases, an S_0 population of the next RS cycle is obtained. The second-year is necessary to self-fertilize the S_0 population and obtain $S_{0:1}$ progenies again. Another strategy for conducting RS would be to use $S_{1:2}$ progenies, exploring a greater proportion of σ_A^2 (TABLE 3.6). However, an additional generation is necessary, which, in terms of annual GS, in principal, did not present an advantage relative to $S_{0:1}$ (TABLE 3.6).

In this strategy, the $S_{0:1}$ progenies would be evaluated in three replicates, and the fourth replicate would be used to select the best plants within the best progenies selected. Since inbred progenies have a small effective size, it would be necessary to evaluate a higher number of $S_{1:2}$ progenies, for example, 300 progenies. Considering 200 $S_{0:1}$ progenies been evaluated, a small selection intensity would be applied, such as 15%, with a total of 30 progenies being selected. Thus, it would be necessary to select ten plants within each selected progeny. In this strategy, the GS would be higher due to the selection within plants, and due to the smaller phenotypic variance as it involves two generations of evaluation ($S_{0:1}$ and $S_{1:2}$). In addition, since the evaluation is carried out in two harvests, the aspect of the progenies versus harvests interaction, which is crucial in plant breeding, could be mitigated by evaluating the progenies in two different harvests.

Another viable possibility in tobacco is to carry out a combined process, evaluating $S_{0:1}$ progenies in one cycle and HS or FS progenies in another. At the moment of recombining the $S_{0:1}$ progenies, it is possible to obtain the HS or FS progenies. In the next year, these progenies are evaluated and recombined to obtain the $S_{0:1}$ progenies by self-fertilization in the next harvest and, thus, continue the RS process (FIGURE 2.2). Note that the annual GS alternating $S_{0:1}$ with FS progenies was about 20% higher than alternating $S_{0:1}$ with HS progenies (TABLE 3.6). However, there are two evaluations of the progenies every three years in both situations, which is a huge advantage.

Considering that the annual GS for progenies of FS, $S_{0:1}$, $S_{1:2}$, and alternating $S_{0:1}$ with FS were the same, a gain of 13.08%, it is questioned what would be the best strategy for conducting RS. In principle, one may say that the best strategy is to use $S_{1:2}$ progenies due to the faster obtainment of lines compared to the other strategies. Nonetheless, three harvests are needed to carry out each cycle, and this time is shorter for the other strategies.

Finally, with the viability of using DH lines, especially using anther cultures (LEMOS, 2021), no previous strategy obtained the same annual GS as with DH (TABLE 3.6). Using DH, the GS was 33% higher relative to the best strategies previously mentioned. However, two considerations are important for implementing this approach: obtaining DH lines are more expensive than other types of progeny, and it is necessary to evaluate a greater number of DH lines compared with other progeny types to succeed equally. This may be attributed to the mean of DH lines, which, due to wide segregation, is lower than when obtaining lines by the conventional process (LEMOS, 2021). It should also be emphasized that, unlike with the other strategies, it is impossible to obtain a gain from selection within DH lines. This may reduce its advantage in relation to other strategies that have the possibility of selection within. Finally, an

expressive point in favor of the use of DH, regardless of the GS, is that, in each cycle, the final product is the line and, therefore, these DH lines may already participate in the value for cultivation and use (VCU) experiments for a later recommendation.

5 CONCLUSIONS

The dominance variance was null for all evaluated traits, indicating that the dominance effect (d) is not important for the FCV tobacco population used in this study. In this condition, the estimate of D_1, which depends on d, is not expressive. Therefore, inbreeding in the RS program does not influence negatively the GS, and, consequently, any type of progeny may be used. The use of DH lines in the RS program in FCV tobacco represents the best strategy since it associates a greater GS per unit of time and generates a line at the end of each cycle, which is the ultimate goal of a genetic breeding program. However, in the case of not using DH lines, other strategies allow expressive gains, such as the alternate use of inbred progenies and FS progenies for evaluation and FS progenies for recombination.

REFERENCES

BARANWAL, V. K. *et al*. Heterosis: emerging ideas about hybrid vigour. **Journal of Experimental Botany**, v. 63, n. 18, p. 6309-6314, 2012.

BERNARDO, Rex. **Breeding for quantitative traits in plants.** 2. ed. Woodbury: Ed. Stemma Press, 2020.

BONATO, E. R.; VELLO, N. A. Aspectos genéticos do tempo para o florescimento em variantes naturais de soja. **Pesquisa Agropecuária Brasileira**, Brasília, v. 34, n. 6, p. 989-993, 1999.

BORDES, J. *et al*. Doubled-haploid versus single-seed descent and S1 family variation for testcross performance in a maize population. **Euphytica**, Wageningen, v. 154, n. 1, p. 41-51, 2007.

CARVALHO, Bruna Line. **Breeding for multiple traits using phenotypic and genotypic information in tobacco for tropical conditions.** 2016. Tese (Doutorado em Genética e Melhoramento de Plantas) – Universidade Federal de Lavras, Lavras, 2016.

COBB, J. N. *et al*. Enhancing the rate of genetic gain in public-sector plant breeding programs: lessons from the breeder's equation. **Theoretical and applied genetics**, Berlim, v. 132, n. 3, p. 627-645, 2019.

COCKERHAM, E.; WEIR, B. S. Covariances of relatives stemming from a population undergoing mixed self and random mating. **Biometrics**, Washington, v. 40, p. 157-164, 1984.

CRISÓSTOMO, J. R. 1989. In: SOUZA, G. A.; RAMALHO, M. A. P. Estimates of genetic and phenotypic variance of some traits of dry bean using a segregant population from the cross Jalo x Small White. **Brazilian Journal of Genetics**, Ribeirão Preto, v. 18, n.1, p. 87-91, 1995.

FERNANDES, J. S. C. 1990. In: SOUZA, G. A.; RAMALHO, M. A. P. Estimates of genetic and phenotypic variance of some traits of dry bean using a segregant population from the cross Jalo x Small White. **Brazilian Journal of Genetics**, Ribeirão Preto, v. 18, n.1, p. 87-91, 1995.

FERREIRA, D.F.; ZAMBALDE, A.L. Simplificação de algumas técnicas especiais da experimentação agropecuária no MAPGEN e Softwares correlatos. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE INFORMÁTICA APLICADA A AGROPECUÁRIA E A AGROINDÚSTRIA, SOFTEX 2000, Belo Horizonte, 1997. *Anais*...Belo Horizonte, p.285-291, 1997.

FOUILLOUX, G.; BANNEROT, H. Selection methods in the common bean (*Phaseolus vulgaris* L.). In: GEPTS, P. **Genetic resources of *Phaseolus* beans.** Dordrecht: Kluwer Academic, 1988.

GERALDI, I. O. O Sellección recurrente em el mejormiento de plantas. In: GUIMARÃES, E. P. (Ed.). **Seleccion recurrente em arroz.** CIAT: Cali, 1997.

HALLAUER, A. R.; CANENA, M. J.; MIRANDA FILHO, J. B. **Quantitative genetics in maize breeding**. 3 ed. New York: Ed. Springer, 2010.

HANCOCK, W. G. *et al*. Identification of maternal haploids of Nicotiana tabacum aided by transgenic expression of green fluorescent protein: evidence for chromosome elimination in the *N. tabacum*× *N. africana* interspecific cross. **Molecular breeding**, v. 35, n. 9, p. 1-14, 2015.

JIANG, Y. *et al*. A quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. **Nature genetics**, New York, v. 49, n. 12, p. 1741-1746, 2017.

LEMOS, Roxane do Carmo. **Performance of doubled haploid lines in relation to those obtained by the conventional breeding method in tobacco**. Lavras: Ed. UFLA, 2021.

LOPES, F. S. *et al.* Recurrent selection in common bean aiming at resistance to white mold in a greenhouse. **Crop Breeding and Applied Biotechnology**, Viçosa, v. 19, n. 1, p. 95-101, 2019.

MATHER, K. The genetical theory of continuous variation. **Hereditas**, Landskrona, v. 35, n. S1, p. 376-401, 1949.

MATZINGER, D. F.; MANN, T. J.; COCKERHAM, C. C. Diallel crosses in *Nicotiana tabacum*. **Crop Science**, Madison, v. 2, p. 383-386, 1962.

MELCHINGER, A. E. *et al*. The role of epistasis in the manifestation of heterosis: a systems-oriented approach, **Genetics**, Austin, v. 177, n. 3, p. 1815-1825, 2007.

MORAIS, O. P. *et al*. Estimação dos parâmetros genéticos da população de arroz irrigado CNA-IRAT 4/0/3. **Pesquisa Agropecuária Brasileira**, Brasília, v. 32, n. 4, p. 421-433, 1997.

MORETO, A. L.; RAMALHO, M. A. P.; ABREU, A. F. B. Estimação dos componentes da variância fenotípica em feijoeiro utilizando o método genealógico. **Ciência e Agrotecnologia**, Lavras, v. 31, n. 4, p. 1035-1042, 2007.

NOVOSELOVIC, D. *et al*. Quantitative inheritance of some wheat plant traits. **Genetics and Molecular Biology**, Ribeirão Preto, v. 27, n. 1, p. 92-98, 2004.

PSCHEIDT, Anderson. **Capacidade de combinação entre linhagens de tabaco dark.** 2020. Dissertação (Mestrado Profissional em Genética e Melhoramento de Plantas) – Universidade Federal de Lavras, Lavras, 2020.

R CORE TEAM. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2020.

RAMALHO, M. A. P. Eficiência relativa de alguns processos de seleção intrapopulacional no milho baseados em famílias não endógamas. 1977. In: HALLAUER, A. R.; CANENA, M. J.; MIRANDA FILHO, J. B. **Quantitative genetics in maize breeding**. 3 ed. New York: Ed. Springer, 2010.

RAMALHO, M. A. P.; FERREIRA, D. F.; de OLIVEIRA, A. C. **Experimentação em Genética e Melhoramento de Plantas.** 3 ed. Lavras: Ed. UFLA, 2012.

RAMALHO, M. A. P. *et al*. **Aplicações da genética quantitativa no melhoramento de plantas autógamas.** 1 ed. Lavras: Ed. UFLA, 2012.

RAPOSO, F. V. ; RAMALHO, M. A. P. ; ABREU, A. F. B. . Comparação de métodos de condução de populações segregantes do feijoeiro. **Pesquisa Agropecuária Brasileira**, Brasília, v. 35, n. 10, p. 1991-1997, 2000.

REDDY, T. V.; DWIVEDI, S.; SHARMA, N. K. Development of tilling by sequencing platform towards enhanced leaf yield in tobacco. **Industrial Crops and Products**, v. 40, p. 324-335, 2012.

RESENDE, M. D. V.; DUARTE, J. B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical**, Goiânia, v. 37, n. 3, p. 182-194, 2007.

RESENDE, M. D. V.; SILVA, F. F. E.; AZEVEDO, C. F. **Estatística matemática, biométrica e computacional.** 1 ed. Viçosa: Ed. Viçosa, 2014.

ROBINSON, H. F.; MANN, T. J.; COMSTOCK, R. E. An analysis of quantitative variability in *Nicotiana tabacum*. **Heredity**, Edinburgh, v. 8, n. 3, p. 365, 1954.

RUSSELL, W. K. *et al*. Allocation of resources in breeding for resistance in maize to second brood of the European corn borer. **Maydica**, Italy, v. 23, n. 1, p. 11–20, 1978.

SCOTT, A. J.; KNOTT, M. A cluster analysis method for grouping means in the analysis of variance. **Biometrics**, Washington, v. 30, n. 3, p. 507-512, 1974.

SOARES, I. O. *et al.* Genetic and phenotypic parameters associated with soybean progenies in a recurrent selection program. **Crop Breeding and Applied Biotechnology**, Viçosa, v. 20, n. 4, 2020.

SOOD, S. *et al.* Improving androgenesis-mediated doubled haploid production efficiency of FCV tobacco (Nicotiana tabacum L.) through in vitro colchicine application. **Plant Breeding**, Berlin, v. 132, n. 6, p. 764-771, 2013.

SOUZA JÚNIOR, Cláudio Lopes. **Componentes da variância genética e suas implicações no melhoramento vegetal.** 1 ed. Piracicaba: Ed. FEALQ, 1989.

SOUZA JÚNIOR, C. L. In: NASS, Luciano Lourenço. **Recursos genéticos e melhoramento-Plantas**. Rondonópolis: Ed. Fundação MT, 2001.

SOUZA, G. A.; RAMALHO, M. A. P. Estimates of genetic and phenotypic variance of some traits of dry bean using a segregant population from the cross Jalo x Small White. **Brazilian Journal of Genetics**, Ribeirão Preto, v. 18, n.1, p. 87-91, 1995.

TANG, Z. *et al*. Climatic factors determine the yield and quality of Honghe flue-cured tobacco. **Scientific Reports**, v. 10, n. 1, p. 1-12, 2020.

TOLEDO, F. H. R. B. *et al*. Experimental strategies in carrying out VCU for tobacco crop I: plot design and size. **Genetics and Molecular Research**, Ribeirão Preto, v. 12, n. 3, p. 3766-3774, 2013.

XU, Y. *et al.* Enhancing genetic gain in the era of molecular breeding. **Journal of Experimental Botany**, v. 68, n. 11, p. 2641-2666, 2017.

YAO, H. *et al*. Genomic dosage effects on heterosis in triploid maize. **Procedings of the National Academy of Sciences**, v. 110, n. 7, p. 2665-2669, 2013.

APPENDIX A

Table 1 – Summary of the variance analysis for green leaf yield (GLY), total sugar content (TSC), total alkaloid content (TAC), and leaf width (LW).

SV	DF	MS			
		GLY	TSC	TAC	LW
HS	41	12.15 ^{ns}	4.15 ^{ns}	0.08 *	15.62 ^{ns}
HS Error	82	9.96	3.09	0.05	10.81
$S_{1:2}$	41	18.98 *	6.58 **	0.21 **	11.08 **
$S_{1:2}$ Error	82	11.85	2.20	0.07	4.98
FS	41	7.66 **	7.29 ^{ns}	0.10 **	11.13 *
^{N}FS	19	8.49 **	7.43 ^{ns}	0.10 **	16.13 **
^{I}FS	21	6.87 *	5.83 ^{ns}	0.10 **	7.00 ^{ns}
^{N}FS vs ^{I}FS	1	8.42 ^{ns}	35.16 *	0.04 ^{ns}	2.92 ^{ns}
FS Error	82	3.83	5.63	0.04	7.01
Mean		17.87	14.06	2.23	26.51

Experiments for evaluating progenies of inbred half-sibs (HS), $S_{1:2}$, and full-sibs (FS), containing non-inbred (^{N}FS) and inbred full-sibs (^{I}FS).
**, *, and ^{ns}: significant ($p \leq 0.01$), significant ($p \leq 0.05$), and non-significant ($p > 0.05$) by the F-test, respectively.

Source: from the author (2021).

APPENDIX B

i) Genetic variance for $S_{1:2}$ progenies

According to SOUZA JÚNIOR (1989), the genetic variance for inbred progenies is estimated by the following equation:

$$C_{tgg'} = (1 + F_t)\sigma_A^2 + \left[\frac{(1-F_g)(1-F_{g'})}{1-F_t}\right]\sigma_D^2 + \left(F_g + F_{g'} + 2F_t\right)D_1 + \left[F_t + \frac{(F_g - F_t)(F_{g'} - F_t)}{2(1-F_t)}\right]D_2 + \left[\frac{F_t(1-F_g)(1-F_{g'})}{1-F_t}\right]\breve{H},$$

where

$F_t = 1 - \left(\frac{1}{2}\right)^t$;

$F_g = 1 - \left(\frac{1}{2}\right)^g$;

$F_{g'} = 1 - \left(\frac{1}{2}\right)^{g'}$.

Thus, the $S_{1:2}$ genetic variance $\left(\sigma^2_{G_{S1:2}}\right)$ is:

$F_t = F_1 = 1 - \left(\frac{1}{2}\right)^1 = \frac{1}{2}$;

$F_g = F_{g'} = F_2 = 1 - \left(\frac{1}{2}\right)^2 = \frac{3}{4}$;

$\sigma^2_{G_{S1:2}} = \left(1 + \frac{1}{2}\right)\sigma_A^2 + \left[\frac{\left(1-\frac{3}{4}\right)\left(1-\frac{3}{4}\right)}{1-\frac{1}{2}}\right]\sigma_D^2 + \left(\frac{3}{4} + \frac{3}{4} + 2\frac{1}{2}\right)D_1 + \left[\frac{1}{2} + \frac{\left(\frac{3}{4}-\frac{1}{2}\right)\left(\frac{3}{4}-\frac{1}{2}\right)}{2\left(1-\frac{1}{2}\right)}\right]D_2 + \left[\frac{\frac{1}{2}\left(1-\frac{3}{4}\right)\left(1-\frac{3}{4}\right)}{1-\frac{1}{2}}\right]\breve{H}$;

$\sigma^2_{G_{S1:2}} = \frac{3}{2}\sigma_A^2 + \left[\frac{\left(\frac{1}{4}\right)\left(\frac{1}{4}\right)}{\frac{1}{2}}\right]\sigma_D^2 + \frac{5}{2}D_1 + \left[\frac{1}{2} + \frac{\left(\frac{1}{4}\right)\left(\frac{1}{4}\right)}{2\left(\frac{1}{2}\right)}\right]D_2 + \left[\frac{\frac{1}{2}\left(\frac{1}{4}\right)\left(\frac{1}{4}\right)}{\frac{1}{2}}\right]\breve{H}$;

$\sigma^2_{G_{S1:2}} = \frac{3}{2}\sigma_A^2 + \frac{1}{8}\sigma_D^2 + \frac{5}{2}D_1 + \frac{9}{16}D_2 + \frac{1}{16}\breve{H}$.

ii) Genetic variance for inbred half-sib (HS) progenies

Table 1 – Genotypic values with *p* and *q* allele frequencies.

Genotypes	Frequency	Genotypes in progenies			Genotypic values
		BB	Bb	bb	
BB	$p\left(p+\frac{1}{2}q\right)$	p	q	-	$pa+qd$
Bb	pq	$\frac{1}{2}p$	$\frac{1}{2}$	$\frac{1}{2}q$	$\frac{1}{2}pa+\frac{1}{2}d-\frac{1}{2}qa$
bb	$q\left(q+\frac{1}{2}p\right)$	-	p	q	$pd-qa$

$$mean = [p^2(pa+qd)] + \left[2pq\left(\tfrac{1}{2}pa+\tfrac{1}{2}d-\tfrac{1}{2}qa\right)\right] + [q^2(pd-qa)]$$

$$mean = p^3a + p^2qd + p^2qa + pqd - pq^2a + pq^2d - q^3a$$

$$mean = a(p^3 + p^2q - pq^2 - q^3) + pqd(p+1+q)$$

$$mean = a[(p^2-q^2)(p+q)] + 2pqd$$

$$mean = ap^2 - aq^2 + 2pqd$$

$$\sigma^2_{G_{HS}} = \left[p\left(p+\tfrac{1}{2}q\right)\right](pa+qd)^2 + \left[pq\left(\tfrac{1}{2}pa+\tfrac{1}{2}d-\tfrac{1}{2}qa\right)^2\right] + \left[q\left(q+\tfrac{1}{2}p\right)\right](pd-qa)^2 - (ap^2 - aq^2 + 2pqd)^2$$

$$\sigma^2_{G_{HS}} = \left(p^2+\tfrac{1}{2}pq\right)(p^2a^2 + 2pqad + q^2d^2) + pq\left(\tfrac{1}{4}p^2a^2 + \tfrac{1}{4}pad - \tfrac{1}{4}pqa^2 + \tfrac{1}{4}pad + \tfrac{1}{4}d^2 - \tfrac{1}{4}qad - \tfrac{1}{4}pqa^2 - \tfrac{1}{4}qad + \tfrac{1}{4}q^2a^2\right) + \left(q^2+\tfrac{1}{2}pq\right)(p^2d^2 - 2pqad + q^2a^2) - (p^4a^2 - p^2q^2a^2 + 2p^3qad - p^2q^2a^2 + q^4a^2 - 2pq^3ad + 2p^3qad - 2pq^3ad + 4p^2q^2d^2)$$

$$\sigma^2_{G_{HS}} = p^4a^2 + 2p^3qad + p^2q^2d^2 + \tfrac{1}{2}p^3qa^2 + p^2q^2ad + \tfrac{1}{2}pq^3d^2 + \tfrac{1}{4}p^3qa^2 + \tfrac{1}{4}p^2qad - \tfrac{1}{4}p^2q^2a^2 + \tfrac{1}{4}p^2qad + \tfrac{1}{4}pqd^2 - \tfrac{1}{4}pq^2ad - \tfrac{1}{4}p^2q^2a^2 - \tfrac{1}{4}pq^2ad + \tfrac{1}{4}pq^3a^2 + p^2q^2d^2 - 2pq^3ad + q^4a^2 + \tfrac{1}{2}p^3qd^2 - p^2q^2ad + \tfrac{1}{2}pq^3a^2 - p^4a^2 + p^2q^2a^2 - 2p^3qad + p^2q^2a^2 - q^4a^2 + 2pq^3ad - 2p^3qad + 2pq^3ad - 4p^2q^2d^2$$

$$\sigma^2_{G_{HS}} = a^2\left(p^4 + \tfrac{1}{2}p^3q + \tfrac{1}{4}p^3q - \tfrac{1}{4}p^2q^2 - \tfrac{1}{4}p^2q^2 + \tfrac{1}{4}pq^3 + q^4 + \tfrac{1}{2}pq^3 - p^4 + p^2q^2 + p^2q^2 - q^4\right) + ad\left(2p^3q + p^2q^2 + \tfrac{1}{4}p^2q + \tfrac{1}{4}p^2q - \tfrac{1}{4}pq^2 - \tfrac{1}{4}pq^2 - 2pq^3 - p^2q^2 - 2p^3q + 2pq^3 - 2p^3q + 2pq^3\right) + d^2\left(p^2q^2 + \tfrac{1}{2}pq^3 + \tfrac{1}{4}pq + p^2q^2 + \tfrac{1}{2}p^3q - 4p^2q^2\right)$$

$$\sigma^2_{G_{HS}} = a^2\left(\tfrac{3}{4}p^3q + \tfrac{3}{2}p^2q^2 + \tfrac{3}{4}pq^3\right) + ad\left(-2p^3q + \tfrac{1}{2}p^2q - \tfrac{1}{2}pq^2 + 2pq^3\right) + d^2\left(p^2q^2 + \tfrac{1}{2}pq^3 + \tfrac{1}{2}p^3q\right) + d^2\left(\tfrac{1}{4}pq + p^2q^2 - 4p^2q^2\right)$$

$$\sigma^2_{G_{HS}} = \tfrac{3}{4}pqa^2(p^2 + 2pq + q^2) + pqad\left(\tfrac{1}{2}p - \tfrac{1}{2}q\right)(p + q) + pqad(-2p^2 + 2q^2) + \tfrac{1}{2}pqd^2(p^2 + 2pq + q^2) + d^2\left(\tfrac{1}{4}pq - 3p^2q^2\right)$$

$$\sigma^2_{G_{HS}} = \tfrac{3}{4}pqa^2 + pqad\left(\tfrac{1}{2}p^2 - \tfrac{1}{2}q^2 - 2p^2 + 2q^2\right) + pqd^2\left(\tfrac{1}{2} + \tfrac{1}{4} - 3pq\right)$$

$$\sigma^2_{G_{HS}} = \tfrac{3}{4}pqa^2 + pqad\left(\tfrac{3}{2}q^2 - \tfrac{3}{2}p^2\right) + pqd^2\left(\tfrac{3}{4} - 3pq\right)$$

$$\sigma^2_{G_{HS}} = \tfrac{3}{4}pqa^2 + \tfrac{3}{2}pqad(q^2 - p^2) + \tfrac{3}{4}pqd^2(1 - 4pq)$$

$$\sigma^2_{G_{HS}} = \tfrac{3}{4}pqa^2 + \tfrac{3}{2}pqad[(q - p)(q + p)] + \tfrac{3}{4}pqd^2(p^2 + 2pq + q^2 - 4pq)$$

$$\sigma^2_{G_{HS}} = \tfrac{3}{4}pqa^2 + \tfrac{3}{2}pqad(q - p) + \tfrac{3}{4}pqd^2(p^2 - 2pq + q^2)$$

$$\sigma^2_{G_{HS}} = \tfrac{3}{4}pqa^2 + \tfrac{3}{2}pqad(q - p) + \tfrac{3}{4}pqd^2(q - p)^2$$

$$\sigma^2_{G_{HS}} = \tfrac{3}{4}pq[a^2 + 2(q - p)ad + (q - p)^2d^2]$$

$$\sigma^2_{G_{HS}} = \tfrac{3}{4}pq\alpha^2$$

$$\sigma^2_{G_{HS}} = \tfrac{3}{8}\sigma_A^2$$

iii) Genetic variance for non-inbred full-sib (^{N}FS) progenies

Table 2 – Genotypic values and frequency for each bi-parental cross, with allele frequencies *p* and *q*.

Parents ♀		Parents ♂		Frequency	Genotypic values
		BB	p^2	p^4	a
BB	p^2	**Bb**	$2pq$	$2p^3q$	$\frac{1}{2}a+\frac{1}{2}d$
		bb	q^2	p^2q^2	d
		BB	p^2	$2p^3q$	$\frac{1}{2}a+\frac{1}{2}d$
Bb	$2pq$	**Bb**	$2pq$	$4p^2q^2$	$\frac{1}{2}d$
		bb	q^2	$2pq^3$	$\frac{1}{2}d-\frac{1}{2}a$
		BB	p^2	p^2q^2	d
bb	q^2	**Bb**	$2pq$	$2pq^3$	$\frac{1}{2}d-\frac{1}{2}a$
		bb	q^2	q^4	$-a$

$$mean = [(p^4)a] + \left[(2p^3q)\left(\tfrac{1}{2}a+\tfrac{1}{2}d\right)\right] + [(p^2q^2)d] + \left[(2p^3q)\left(\tfrac{1}{2}a+\tfrac{1}{2}d\right)\right] + \left[(4p^2q^2)\left(\tfrac{1}{2}d\right)\right] + \left[(2pq^3)\left(\tfrac{1}{2}d-\tfrac{1}{2}a\right)\right] + [(p^2q^2)d] + \left[(2pq^3)\left(\tfrac{1}{2}d-\tfrac{1}{2}a\right)\right] + [(q^4)(-a)]$$

$$mean = p^4a + p^3qa + p^3qd + p^2q^2d + p^3qa + p^3qd + 2p^2q^2d + pq^3d - pq^3a + p^2q^2d + pq^3d - pq^3a - q^4a$$

$$mean = a(p^4 + p^3q + p^3q - pq^3 - pq^3 - q^4) + d(p^3q + p^2q^2 + p^3q + 2p^2q^2 + pq^3 + p^2q^2 + pq^3)$$

$$mean = a(p^4 + 2p^3q - 2pq^3 - q^4) + d(2p^3q + 4p^2q^2 + 2pq^3)$$

$$mean = a[(p^2 - q^2)(p^2 + 2pq + q^2)] + pqd(2p^2 + 4pq + 2q^2)$$

$$mean = ap^2 - aq^2 + pqd[2(p+q)^2]$$

$$mean = ap^2 - aq^2 + 2pqd$$

Table 3 – Summary of genotypic values and frequency for each bi-parental cross, with allele frequencies *p* and *q*.

Parents		Frequency	Genotypic values
♀	♂		
BB	**BB**	p^4	a
BB	**Bb**	$4p^3q$	$\frac{1}{2}a + \frac{1}{2}d$
BB	**bb**	$2p^2q^2$	d
Bb	**Bb**	$4p^2q^2$	$\frac{1}{2}d$
Bb	**bb**	$4pq^3$	$\frac{1}{2}d - \frac{1}{2}a$
bb	**bb**	q^4	$-a$

$$\sigma^2_{G_{N_{FS}}} = [(p^4)a^2] + \left[(4p^3q)\left(\tfrac{1}{2}a + \tfrac{1}{2}d\right)^2\right] + [(2p^2q^2)d^2] + \left[4p^2q^2\left(\tfrac{1}{2}d\right)^2\right]$$
$$+ \left[(4pq^3)\left(\tfrac{1}{2}d - \tfrac{1}{2}a\right)^2\right] + [(q^4)(-a)^2] - (ap^2 - aq^2 + 2pqd)^2$$

$$\sigma^2_{G_{N_{FS}}} = p^4a^2 + \left[(4p^3q)\left(\tfrac{1}{4}a^2 + \tfrac{1}{2}ad + \tfrac{1}{2}d^2\right)\right] + 2p^2q^2d^2 + p^2q^2d^2$$
$$+ \left[(4pq^3)\left(\tfrac{1}{4}d^2 - \tfrac{1}{2}ad + \tfrac{1}{4}a^2\right)\right] + q^4a^2$$
$$- (p^4a^2 - p^2q^2a^2 + 2p^3qad - p^2q^2a^2 + q^4a^2 - 2pq^3ad + 2p^3qad$$
$$- 2pq^3ad + 4p^2q^2d^2)$$

$$\sigma^2_{G_{N_{FS}}} = p^4a^2 + p^3qa^2 + 2p^3qad + p^3qd^2 + 2p^2q^2d^2 + p^2q^2d^2 + pq^3d^2 - 2pq^3ad$$
$$+ pq^3a^2 + q^4a^2 - p^4a^2 + p^2q^2a^2 - 2p^3qad + p^2q^2a^2 - q^4a^2 + 2pq^3ad$$
$$- 2p^3qad + 2pq^3ad - 4p^2q^2d^2$$

$$\sigma^2_{G_{N_{FS}}} = a^2(p^4 + p^3q + pq^3 + q^4 - p^4 + p^2q^2 + p^2q^2 - q^4)$$
$$+ ad(2p^3q - 2pq^3 - 2p^3q + 2pq^3 - 2p^3q + 2pq^3)$$
$$+ d^2(p^3q + 2p^2q^2 + p^2q^2 + pq^3 - 4p^2q^2)$$

$$\sigma^2_{G_{N_{FS}}} = pqa^2(p^2 + 2pq + q^2) + 2pqad(q^2 - p^2) + pqd^2(p^2 - 2pq + q^2) + p^2q^2d^2$$

$$\sigma^2_{G_{N_{FS}}} = pqa^2 + 2pqad[(q-p)(p+q)] + pqd^2(q-p)^2 + p^2q^2d^2$$

$$\sigma^2_{G_{N_{FS}}} = pq[a^2 + 2(q-p)ad + (q-p)^2d^2] + p^2q^2d^2$$

$$\sigma^2_{G_{N_{FS}}} = pq[a + (q-p)d]^2 + p^2q^2d^2$$

$$\sigma^2_{G_{N_{FS}}} = pq\alpha^2 + p^2q^2d^2$$

$$\sigma^2_{G_{N_{FS}}} = \tfrac{1}{2}\sigma^2_A + \tfrac{1}{4}\sigma^2_D$$

iv) Genetic variance for inbred full-sib (^{I}FS) progenies

Table 4 – Genotypic values and frequency for each bi-parental cross, with allele frequencies *p* and *q*.

Parents ♀		Parents ♂		Frequency	Genotypic values
BB	$p\left(p+\frac{1}{2}q\right)$	**BB**	$p\left(p+\frac{1}{2}q\right)$	$p^4+p^3q+\frac{1}{4}p^2q^2$	a
		Bb	pq	$p^3q+\frac{1}{2}p^2q^2$	$\frac{1}{2}a+\frac{1}{2}d$
		bb	$q\left(q+\frac{1}{2}p\right)$	$p^2q^2+\frac{1}{2}pq^3+\frac{1}{2}p^3q+\frac{1}{4}p^2q^2$	d
Bb	pq	**BB**	$p\left(p+\frac{1}{2}q\right)$	$p^3q+\frac{1}{2}p^2q^2$	$\frac{1}{2}a+\frac{1}{2}d$
		Bb	pq	p^2q^2	$\frac{1}{2}d$
		bb	$q\left(q+\frac{1}{2}p\right)$	$pq^3+\frac{1}{2}p^2q^2$	$\frac{1}{2}d-\frac{1}{2}a$
bb	$q\left(q+\frac{1}{2}p\right)$	**BB**	$p\left(p+\frac{1}{2}q\right)$	$p^2q^2+\frac{1}{2}pq^3+\frac{1}{2}p^3q+\frac{1}{4}p^2q^2$	d
		Bb	pq	$pq^3+\frac{1}{2}p^2q^2$	$\frac{1}{2}d-\frac{1}{2}a$
		bb	$q\left(q+\frac{1}{2}p\right)$	$q^4+pq^3+\frac{1}{4}p^2q^2$	$-a$

$$mean = \left[\left(p^4+p^3q+\tfrac{1}{4}p^2q^2\right)a\right]+\left[\left(p^3q+\tfrac{1}{2}p^2q^2\right)\left(\tfrac{1}{2}a+\tfrac{1}{2}d\right)\right]$$
$$+\left[\left(p^2q^2+\tfrac{1}{2}pq^3+\tfrac{1}{2}p^3q+\tfrac{1}{4}p^2q^2\right)d\right]+\left[\left(p^3q+\tfrac{1}{2}p^2q^2\right)\left(\tfrac{1}{2}a+\tfrac{1}{2}d\right)\right]$$
$$+\left[p^2q^2\left(\tfrac{1}{2}d\right)\right]+\left[\left(pq^3+\tfrac{1}{2}p^2q^2\right)\left(\tfrac{1}{2}d-\tfrac{1}{2}a\right)\right]$$
$$+\left[\left(p^2q^2+\tfrac{1}{2}pq^3+\tfrac{1}{2}p^3q+\tfrac{1}{4}p^2q^2\right)d\right]+\left[\left(pq^3+\tfrac{1}{2}p^2q^2\right)\left(\tfrac{1}{2}d-\tfrac{1}{2}a\right)\right]$$
$$+\left[\left(q^4+pq^3+\tfrac{1}{4}p^2q^2\right)(-a)\right]$$

$$mean = p^4a+p^3qa+\tfrac{1}{4}p^2q^2a+\tfrac{1}{2}p^3qa+\tfrac{1}{2}p^3qd+\tfrac{1}{4}p^2q^2a+\tfrac{1}{4}p^2q^2d+p^2q^2d+\tfrac{1}{2}pq^3d$$
$$+\tfrac{1}{2}p^3qd+\tfrac{1}{4}p^2q^2d+\tfrac{1}{2}p^3qa+\tfrac{1}{2}p^3qd+\tfrac{1}{4}p^2q^2a+\tfrac{1}{4}p^2q^2d+\tfrac{1}{2}p^2q^2d$$
$$+\tfrac{1}{2}pq^3d-\tfrac{1}{2}pq^3a+\tfrac{1}{4}p^2q^2d-\tfrac{1}{4}p^2q^2a+p^2q^2d+\tfrac{1}{2}pq^3d+\tfrac{1}{2}p^3qd+\tfrac{1}{4}p^2q^2d$$
$$+\tfrac{1}{2}pq^3d-\tfrac{1}{2}pq^3a+\tfrac{1}{4}p^2q^2d-\tfrac{1}{4}p^2q^2a-q^4a-pq^3a-\tfrac{1}{4}p^2q^2a$$

$$mean = a\left(p^4+p^3q+\tfrac{1}{4}p^2q^2+\tfrac{1}{2}p^3q+\tfrac{1}{4}p^2q^2+\tfrac{1}{2}p^3q+\tfrac{1}{4}p^2q^2-\tfrac{1}{2}pq^3-\tfrac{1}{4}p^2q^2-\tfrac{1}{2}pq^3\right.$$
$$\left.-\tfrac{1}{4}p^2q^2-q^4-pq^3-\tfrac{1}{4}p^2q^2\right)$$
$$+d\left(\tfrac{1}{2}p^3q+\tfrac{1}{4}p^2q^2+p^2q^2+\tfrac{1}{2}pq^3+\tfrac{1}{2}p^3q+\tfrac{1}{4}p^2q^2+\tfrac{1}{2}p^3q+\tfrac{1}{4}p^2q^2+\tfrac{1}{2}p^2q^2\right.$$
$$\left.+\tfrac{1}{2}pq^3+\tfrac{1}{4}p^2q^2+p^2q^2+\tfrac{1}{2}pq^3+\tfrac{1}{2}p^3q+\tfrac{1}{4}p^2q^2+\tfrac{1}{2}pq^3+\tfrac{1}{4}p^2q^2\right)$$

$$mean = a(p^4+2p^3q-2pq^3-q^4)+d(2p^3q+4p^2q^2+2pq^3)$$

$$mean = a[(p^2 - q^2)(p^2 + 2pq + q^2)] + pqd(2p^2 + 4pq + 2q^2)$$

$$mean = ap^2 - aq^2 + pqd[2(p + q)^2]$$

$$mean = ap^2 - aq^2 + 2pqd$$

Table 5 – Summary of genotypic values and frequency for each bi-parental cross, with allele frequencies *p* and *q*.

Parents		Frequency	Genotypic values
♀	♂		
BB	**BB**	$p^4 + p^3q + \frac{1}{4}p^2q^2$	a
BB	**Bb**	$2p^3q + p^2q^2$	$\frac{1}{2}a + \frac{1}{2}d$
BB	**bb**	$2p^2q^2 + p^3q + pq^3 + \frac{1}{2}p^2q^2$	d
Bb	**Bb**	p^2q^2	$\frac{1}{2}d$
Bb	**bb**	$2pq^3 + p^2q^2$	$\frac{1}{2}d - \frac{1}{2}a$
bb	**bb**	$q^4 + pq^3 + \frac{1}{4}p^2q^2$	$-a$

$$\sigma^2_{G\ I_{FS}} = \left[\left(p^4 + p^3q + \tfrac{1}{4}p^2q^2\right)a^2\right] + \left[(2p^3q + p^2q^2)\left(\tfrac{1}{2}a + \tfrac{1}{2}d\right)^2\right]$$
$$+ \left[\left(2p^2q^2 + pq^3 + p^3q + \tfrac{1}{2}p^2q^2\right)d^2\right] + \left[p^2q^2\left(\tfrac{1}{2}d\right)^2\right]$$
$$+ \left[(2pq^3 + p^2q^2)\left(\tfrac{1}{2}d - \tfrac{1}{2}a\right)^2\right] + \left[\left(q^4 + pq^3 + \tfrac{1}{4}p^2q^2\right)(-a)^2\right]$$
$$- (ap^2 - aq^2 + 2pqd)^2$$

$$\sigma^2_{G\ I_{FS}} = p^4a^2 + p^3qa^2 + \tfrac{1}{4}p^2q^2a^2 + \left[(2p^3q + p^2q^2)\left(\tfrac{1}{4}a^2 + \tfrac{1}{2}ad + \tfrac{1}{2}d^2\right)\right] + 2p^2q^2d^2$$
$$+ p^3qd^2 + pq^3d^2 + \tfrac{1}{2}p^2q^2d^2 + \tfrac{1}{4}p^2q^2d^2$$
$$+ \left[(2pq^3 + p^2q^2)\left(\tfrac{1}{4}d^2 - \tfrac{1}{2}ad + \tfrac{1}{4}a^2\right)\right] + q^4a^2 + pq^3a^2 + \tfrac{1}{4}p^2q^2a^2$$
$$- (p^4a^2 - p^2q^2a^2 + 2p^3qad - p^2q^2a^2 + q^4a^2 - 2pq^3ad + 2p^3qad$$
$$- 2pq^3ad + 4p^2q^2d^2)$$

$$\sigma^2_{G\ I_{FS}} = p^4a^2 + p^3qa^2 + \tfrac{1}{4}p^2q^2a^2 + \tfrac{1}{2}p^3qa^2 + p^3qad + \tfrac{1}{2}p^3qd^2 + \tfrac{1}{4}p^2q^2a^2 + \tfrac{1}{2}p^2q^2ad$$
$$+ \tfrac{1}{4}p^2q^2d^2 + 2p^2q^2d^2 + p^3qd^2 + pq^3d^2 + \tfrac{1}{2}p^2q^2d^2 + \tfrac{1}{4}p^2q^2d^2 + \tfrac{1}{2}pq^3d^2$$
$$- pq^3ad + \tfrac{1}{2}pq^3a^2 + \tfrac{1}{4}p^2q^2d^2 - \tfrac{1}{2}p^2q^2ad + \tfrac{1}{4}p^2q^2a^2 + q^4a^2 + pq^3a^2$$
$$+ \tfrac{1}{4}p^2q^2a^2 - p^4a^2 + p^2q^2a^2 - 2p^3qad + p^2q^2a^2 - q^4a^2 + 2pq^3ad$$
$$- 2p^3qad + 2pq^3ad - 4p^2q^2d^2$$

$$\sigma^2_{G\ I_{FS}} = a^2\left(p^4 + p^3q + \frac{1}{4}p^2q^2 + \frac{1}{2}p^3q + \frac{1}{4}p^2q^2 + \frac{1}{2}pq^3 + \frac{1}{4}p^2q^2 + q^4 + pq^3 + \frac{1}{4}p^2q^2 - p^4 + p^2q^2 + p^2q^2 - q^4\right) + ad\left(p^3q + \frac{1}{2}p^2q^2 - pq^3 - \frac{1}{2}p^2q^2 - 2p^3q + 2pq^3 - 2p^3q + 2pq^3\right) + d^2\left(\frac{1}{2}p^3q + \frac{1}{4}p^2q^2 + 2p^2q^2 + p^3q + pq^3 + \frac{1}{2}p^2q^2 + \frac{1}{4}p^2q^2 + \frac{1}{2}pq^3 + \frac{1}{4}p^2q^2 - 4p^2q^2\right)$$

$$\sigma^2_{G\ I_{FS}} = a^2\left(\frac{3}{2}p^3q + 3p^2q^2 + \frac{3}{2}pq^3\right) + ad(-3p^3q + 3pq^3) + d^2\left(\frac{3}{2}p^3q - 3p^2q^2 + \frac{3}{2}pq^3\right) + \frac{9}{4}p^2q^2d^2$$

$$\sigma^2_{G\ I_{FS}} = \frac{3}{2}pqa^2(p^2 + 2pq + q^2) + 3pqad(q^2 - p^2) + \frac{3}{2}pqd^2(p^2 - 2pq + q^2) + \frac{9}{4}p^2q^2d^2$$

$$\sigma^2_{G\ I_{FS}} = \frac{3}{2}pqa^2 + 3pqad[(q-p)(p+q)] + \frac{3}{2}pqd^2(q-p)^2 + \frac{9}{4}p^2q^2d^2$$

$$\sigma^2_{G\ I_{FS}} = \frac{3}{2}pq[a^2 + 2(q-p)ad + (q-p)^2d^2] + \frac{9}{4}p^2q^2d^2$$

$$\sigma^2_{G\ I_{FS}} = \frac{3}{2}pq[a + (q-p)d]^2 + \frac{9}{4}p^2q^2d^2$$

$$\sigma^2_{G\ I_{FS}} = \frac{3}{2}pq\alpha^2 + \frac{9}{4}p^2q^2d^2$$

$$\sigma^2_{G\ I_{FS}} = \frac{3}{4}\sigma^2_A + \frac{9}{16}\sigma^2_D$$

ARTICLE 2 – IS EMPLOYING IMAGE PHENOTYPING VIABLE TO PREDICT TOBACCO LEAF YIELD?

Artigo redigido conforme a NBR 6022 (ABNT, 2018) e formatado de acordo com o Manual da UFLA de apresentação de teses e dissertações.

ABSTRACT

The employment of image phenotyping has been stimulated in several cultivated species. However, the efficiency of its use in tobacco breeding has not been seen in the literature. Hence, this work aimed to verify the viability of employing the phenotyping of leaf mass (yield) through images collected by unmanned aerial vehicles (UAVs) and its likely applications in the tobacco breeding program. Experiments involving the two most important varietal groups for tobacco crops in Brazil were used. Three tobacco experiments were assessed, two from the varietal group flue-cured Virginia (FCV) and one from air-cured Burley (ACB). The images for assessing the canopy coverage (CC) and the three vegetative indices, namely the brightness index (BI), normalized green-red difference index (NGRDI), and visible atmospherically resistant index (VARI), were obtained at different times after planting, while leaf yield was later evaluated through the green/cured leaf mass. Analyses of variance were performed for each obtained variable and also for yield. The Spearman correlations between the image variables and the yield and a multiple regression involving all pieces of information (CC, BI, NGRDI, and BI) from the images as independent variables and the yield as the dependent variable were estimated. The simultaneous employment of an index involving the CC and yield was assessed. Additionally, the viability of employing images as an alternative to reduce the number of replicates was verified. It was found that: i) image phenotyping presented good accuracy in detecting differences among the tobacco hybrids/lines; ii) the canopy coverage (CC), assessed through image phenotyping, correlated with the yield of green/cured leaves; iii) the employment of image phenotyping must be stimulated in tobacco crops aiming the obtainment of the green/cured mass yield; iv) it was conjectured that the application of images is viable in several situations within a breeding program aiming at the green/cured mass yield.

Keywords: plant breeding, phenotyping, canopy coverage, drones, selective accuracy.

1 INTRODUCTION

Tobacco crops have great socioeconomic importance in Brazil. Cultivation is predominantly carried out by farmers who have small areas, typically family farmers. According to data from the Tobacco Growers' Association of Brazil - AFUBRA (2021), 159 thousand families were involved with tobacco crops in fourteen Brazilian states in the 2018/19 harvest, with the south region being responsible for over 97% of the production in the country. It must be highlighted that the production of tobacco contributes to farmers who dedicate themselves to the crop having an income likely superior to that obtained by any other farming operation possible in the region. According to AFUBRA (2021), over 90% of Brazilian production is exported, moving over 2.1 billion dollars per year.

From the planting of seedlings in the field until the harvest process, tobacco cultivation is predominantly carried out manually. There are some varietal groups in the crop, with the main one being the flue-cured Virginia (FCV) and the second in importance being the air-cured Burley (ACB). In the case of the FCV, the harvest is the process that demands the most time because the leaves are harvested in several steps according to maturation. In turn, a single harvest is carried out for the ACB, yet it is also considered an arduous process. After each harvest, the leaves are cured, i.e., submitted to a drying process that differs depending on the varietal group. In the conduction of the genetic breeding program, besides being the process that demands the most labor, the harvest is also that which has the highest possibility of contributing to experimental error.

The obtainment of the production of the experimental plots is through the mass of the cured leaves. However, since hundreds of plots are normally assessed in breeding programs, the green mass of the leaves has mainly been used, aiming at reducing the experimental error. Alternatives that might reduce the experimental error and the cost of the harvest operation are evidently desired. One of the alternatives that have been employed in other plant species, in this case, to estimate grain production, is phenotyping through images of the crop (FERNANDEZ-GALLEGO et al., 2019; GALLI et al., 2020; HU; KNAPP; SCHMIDHALTER, 2020; KRAUSE et al., 2020; MOREIRA et al., 2019; NATARAJAN et al., 2019; XAVIER et al., 2017).

The perspectives for the employment of image phenotyping of tobacco aiming at predicting green mass yield are more promising than for grains because it depends only on plant canopy. However, no references of use with tobacco crops were found in the literature. Considering the breeding program especially, in which there is a trend of increasing

employment of progenies to be assessed and part of the experiments are conducted at farmer properties, the proof of the efficiency of image phenotyping is evidently crucial. The obtainment of images of the experimental areas is currently facilitated by the employment of unmanned aerial vehicles (UAVs). The drones or UAVs enable obtaining data from large areas and high-resolution images (YANG et al., 2020).

From the exposed, this work aimed to verify the viability of employing the phenotyping of leaf yield through images collected by UAVs and its likely applications in the tobacco breeding program.

2 MATERIAL AND METHODS

The data used in this research were provided by British American Tobacco (BAT) Brazil. Data from three experiments assessed in the 2020/2021 harvest at the BAT experimental station located in Mafra (altitude 848 m, latitude 26°10' S, longitude 49°48' W), in the state of Santa Catarina, in the south region of Brazil, were used. Information on the planting dates of the experiments is presented in Table 2.1. In Experiment 1, fifteen tobacco hybrids of varietal group flue-cured Virginia (FCV) were assessed. A randomized block design with four replicates was used, with each plot consisting of two 4.5 m rows, a spacing of 1.35 m between rows, and ten plants per row. Given that the harvest of the leaves is carried out manually in steps for varietal group FCV, the green mass of the leaves before being cured was obtained in the last two harvests of Experiment 1. In this case, the yield was assessed through the cured leaf yield (kg $plot^{-1}$).

Experiment 2 consisted of the assessment of sixty lines of FCV tobacco in a randomized block design with three replicates. The plots consisted of one 4.5 m row with ten plants and a spacing of 1.35 m between rows. The harvest was carried out in steps in a manner similar to that of Experiment 1; however, in this case, the green leaf yield was assessed (kg $plot^{-1}$).

In Experiment 3, 104 lines of tobacco of varietal group air-cured Burley (ACB) were assessed in a randomized block design with three replicates. The plots consisted of one 4.0 m row with ten plants and a spacing of 1.30 m between rows. For this varietal group, the harvest was carried out by cutting the plant base and posteriorly removing the leaves. In this experiment, the trait assessed was green leaf yield (kg $plot^{-1}$).

The red-green-blue (RGB) aerial images of the experiments were obtained using UAV DJI Phantom 4 with camera model FC6310 (4864 x 3648 pixels). The images had a 70% lateral and forward overlap that allowed image stitching. Details on the flight heights, the image resolutions, and the tobacco phenology stages that the images were collected according Orlando et al. (2011), for each experiment are presented in Table 2.1.

The image analysis, plot extraction, and classification were performed through a multilayer mosaic methodology using the software “Progeny”, as described by Hearst (2019). The data on the canopy coverage (CC - %) medians of the images of each plot was obtained through the percentage of pixels of the image classified as canopy pixels. Besides the canopy coverage, three vegetative indices (VIs) were estimated, namely the brightness index (BI), with $BI = \sqrt{(Red^2 + Green^2 + Blue^2)/3}$ (RICHARDSON; WIEGAND, 1977), the normalized

green-red difference index (NGRDI), $NGRDI = {(Green - Red)}/{(Green + Red)}$ (TUCKER, 1979), and the visible atmospherically resistant index (VARI), $VARI = {(Green - Red)}/{(Green + Red - Blue)}$ (GITELSON et al., 2002). The VIs were estimated in "R" (R CORE TEAM, 2020) using package "FIELDImageR" (MATIAS; CARAZA-HARTER; ENDELMAN, 2020).

Table 2.1 – Specifications of the flights carried out for the different phenology stages (ORLANDO et al., 2011) in the three experiments.

Experiment / Planting Date	Phenology stages (Code)	Flight height (m)	Resolution (cm pixel^{-1})
1 Sep. 10, 2020	Initial II (L)	65.5	1.75
	Crop development II (G)	35.4	0.94
	Pre-flowering III (CF)	34.4	0.92
	Flowering I (OF)	33.5	0.89
	Flowering II (AF)	34.1	0.91
2 Sep. 14, 2020	Initial II (L)	50.0	1.34
	Crop development II (G)	50.9	1.36
	Pre-flowering III (CF)	50.9	1.36
	Flowering II (AF)	23.5	0.63
3 Sep. 17, 2020	Initial II (L)	55.8	1.49
	Crop development II (G)	54.3	1.45
	Pre-flowering III (CF)	53.3	1.42
	Flowering II (AF)	55.2	1.47

Source: from the author (2021).

All data were submitted to an analysis of variance for each experiment, employing the following model:

$$Y_{ij} = m + l_i + r_j + e_{ij},$$

where Y_{ij} refers to the value observed in the plot that received hybrid/line *i* in replicate *j*, m is the overall mean, l_i is the effect of hybrid/line *i*, with $\sum_{i=0}^{n} \frac{l_i^2}{n-1}$, r_j is the effect of replicate *j*, with $r_j \sim N(0, \sigma_r^2)$, and e_{ij} is the experimental error associated with observation Y_{ij}, with $e_{ij} \sim N(0, \sigma_e^2)$.

Through the results of the analyses of variance, accuracy $(r_{gg'})$ was estimated using the estimator $r_{gg'} = \sqrt{1 - \frac{1}{F}}$, where F corresponds to the mean square of the hybrid/line divided by the mean square of the error. Moreover, the mean of the hybrids/lines was classified using the grouping test by Scott and Knott (1974).

Later, the Spearman correlations (STEEL; TORRIE; DICKEY, 1997) between the leaf yield and the four variables obtained through the image analyses (CC, BI, NGRDI, and VARI) were estimated. The multiple regression coefficients were also estimated by employing the leaf yield as the dependent variable (Y) and CC, BI, NGRDI, and VARI as the independent variables (X). The selection of the variables was carried out through the stepwise procedure based on the Akaike information criterion (AIC) (AKAIKE, 1974). Both the correlation and multiple regression estimates were made for each experiment separately.

To obtain the joint information between the leaf yield and the image, using the information from the image that most explained the variation in yield, the index was calculated from the sum of the standardized variables through the following equation:

$$Z_{ijk} = \frac{Y_{ijk} - \bar{Y}_{.jk}}{s_{.jk}} + 4,$$

where Z_{ijk} refers to the standardized variable corresponding to hybrid/line *i* in replicate *j* for variable *k* (yield or image), Y_{ijk} is the mean of progeny *i* in replicate *j* for variable *k*, $\bar{Y}_{.jk}$ is the overall mean of replicate *j* for variable *k*, and $s_{.jk}$ is the standard deviation of the means of the lines in replicate *j* for variable *k*. As the Z_{ijk} can assume negative and positive values, the constant 4 was added in all estimates, also enabling the analysis of the variable Z_{ijk}.

Later, the possibility of using the index aiming at reducing the number of replicates of the experiments was assessed. For such, all combinations of the different replicates for each experiment were considered. The index was submitted to an analysis of variance by employing the previously mentioned model for each of the possible situations.

All analyses were carried out in “R” (R CORE TEAM, 2020), and package “MASS” (VENABLES; RIPLEY, 2002) was used for the multiple regression analysis.

3 RESULTS

In Experiment 1, a significant difference ($p \leq 0.01$) was detected among the assessed hybrids for cured leaf yield (kg plot^{-1}) (TABLE 1, APPENDIX A), a fundamental condition to achieve the research objectives. It was also verified that the estimate of accuracy $(r_{gg'})$ was high (TABLE 3.1), indicating that the assessment of the trait was carried out with good precision. In the case of Experiment 2, in which sixty tobacco lines of the same varietal group as Experiment 1 were evaluated, i.e., varietal group flue-cured Virginia (FCV), no significant difference ($p > 0.05$) was detected among the lines for green leaf yield (kg plot^{-1}) in this case (TABLE 1, APPENDIX A). This fact contributed to the low estimate of $r_{gg'}$ (TABLE 3.1). In Experiment 3, now with lines of varietal group air-cured Burley (ACB), the results for yield were quite similar to those of Experiment 1 (TABLE 1, APPENDIX A, and TABLE 3.1).

Relative to the canopy coverage (CC) and the vegetative indices (VIs) of brightness index (BI), normalized green-red difference index (NGRDI), and visible atmospherically resistant index (VARI), obtained through the analysis of the images collected via unmanned air vehicle (UAV), significant differences among the hybrids/lines varied among the experiments and among the phenology stages. In Experiment 1, a difference ($p \leq 0.05$) was found among the assessed hybrids from pre-flowering III for the CC and the three VIs (TABLE 2, APPENDIX A).

In Experiment 2, significant differences were found among the lines ($p \leq 0.05$) for NGRDI and VARI in the assessments carried out from pre-flowering III (TABLE 3, APPENDIX A). For the BI, a significant effect ($p \leq 0.05$) occurred at initial II, pre-flowering III, and flowering II. In turn, for the CC, a significant difference ($p \leq 0.01$) was found only in the assessment carried out at pre-flowering III. In Experiment 3, the lines presented significant differences ($p \leq 0.05$) except for the CC evaluated at initial II ($p < 0.05$) (TABLE 4, APPENDIX A).

It must be emphasized that, in general, the estimates for $r_{gg'}$ were of greater magnitude with the advance of the crop stage (TABLE 3.1). The estimates of $r_{gg'}$ for CC and the three VIs (BI, NGRDI, and VARI) were similar and many times superior to the estimates of $r_{gg'}$ for the tobacco leaf yield. However, in Experiment 2 and for the CC, $r_{gg'}$ was almost always of smaller magnitude, likely due to the inexistence of an expressive variation among the lines, as has also already been emphasized for yield.

Table 3.1 – Estimates of the mean selective accuracy $(r_{gg'})$ of Experiments 1, 2, and 3 for canopy coverage (CC), brightness index (BI), normalized green-red difference index (NGRDI), visible atmospherically resistant index (VARI), and leaf yield assessed at different phenology stages.

	Phenology stages	$r_{gg'}$		
		Experiment 1	**Experiment 2**	**Experiment 3**
CC	Initial II	0.66	0.44	0.27
	Crop development II	0.62	0.50	0.64
	Pre-flowering III	0.88	0.66	0.77
	Flowering I	0.90	NA	NA
	Flowering II	0.90	0.32	0.76
BI	Initial II	0.35	0.57	0.59
	Crop development II	0.38	0.51	0.75
	Pre-flowering III	0.90	0.88	0.80
	Flowering I	0.88	NA	NA
	Flowering II	0.82	0.68	0.80
NGRDI	Initial II	-	-	0.62
	Crop development II	0.42	0.43	0.86
	Pre-flowering III	0.73	0.74	0.83
	Flowering I	0.90	NA	NA
	Flowering II	0.80	0.61	0.79
VARI	Initial II	-	-	0.59
	Crop development II	0.43	0.45	0.85
	Pre-flowering III	0.71	0.75	0.83
	Flowering I	0.86	NA	NA
	Flowering II	0.77	0.57	0.79
Yield		0.87	0.46	0.63

Source: from the author (2021).

The estimates of the classifying correlation coefficient between the phenotyping obtained by the traditional method, i.e., by weighing the cured or green leaves, and those obtained via images (CC, BI, NGRDI, and VARI) corroborate practically all previous comments (TABLE 3.2). Observe, for example, that, in general, the correlation rose with the increase of the phenology stage of the crop. This occurred for practically all assessments performed and, surprisingly, also occurred in the case of Experiment 2, in which the significant difference for yield among the lines assessed at flowering II was only detected at the probability level over 0.14 ($p \leq 0.14$). Considering that, in general, higher correlation estimates were found for the assessments carried out with the advance of the crop, the CC presented the greatest estimates relative to the VIs for the three evaluated experiments, with correlation estimates over 56%.

Table 3.2 – Estimates of the Spearman correlation coefficients between yield and the canopy coverage (CC), brightness index (BI), normalized green-red difference index (NGRDI), and visible atmospherically resistant index (VARI) assessed at different phenology stages for the three experiments.

	Phenology stages	Yield		
		Experiment 1	Experiment 2	Experiment 3
CC	Initial II	-0.10^{ns}	-0.11^{ns}	0.01^{ns}
	Crop development II	0.52*	0.11^{ns}	0.13^{ns}
	Pre-flowering III	0.83*	0.41*	0.30*
	Flowering I	0.72*		
	Flowering II	0.83*	0.77*	0.56*
BI	Initial II	0.38^{ns}	-0.15^{ns}	-0.14^{ns}
	Crop development II	0.30^{ns}	0.06^{ns}	-0.36*
	Pre-flowering III	0.29^{ns}	0.09^{ns}	-0.28*
	Flowering I	-0.35^{ns}		
	Flowering II	-0.57*	-0.27*	-0.10^{ns}
NGRDI	Initial II	0.08^{ns}	0.22^{ns}	0.10^{ns}
	Crop development II	-0.09^{ns}	-0.03^{ns}	0.29*
	Pre-flowering III	0.14^{ns}	0.08^{ns}	0.54*
	Flowering I	0.65*		
	Flowering II	0.83*	0.61*	0.38*
VARI	Initial II	0.10^{ns}	0.23^{ns}	0.09^{ns}
	Crop development II	-0.08^{ns}	0.06^{ns}	0.32*
	Pre-flowering III	0.17^{ns}	0.19^{ns}	0.54*
	Flowering I	0.52*		
	Flowering II	0.80*	0.60*	0.31*

* and ^{ns}: significant ($p \leq 0.05$) and non-significant ($p > 0.05$) according to the t-Student test, respectively.
Source: from the author (2021).

Aiming at answering the inquiry of what would occur if the four-information obtained via images were considered simultaneously, the multiple regressions were estimated between the dependent variable (Y) of tobacco leaf yield and the independent variables (X) of CC, BI, NGRI, and VARI considering only the flowering II stage (TABLE 3.3). It was found, for all the experiments, that the CC was the first variable to enter the model, i.e., it presented the smallest AIC value and, thus, the greatest influence on tobacco leaf productivity. It is also pertinent to emphasize that the VARI was present in the models selected for all three experiments.

In general, the models selected for each experiment presented good fitting according to the obtained coefficient of determination (R^2) estimates. The highest and lowest estimates for R^2 were found for Experiment 1 (76.3%) and Experiment 3 (56.4%), respectively. From the exposed, since the CC was present in all situations and with the greatest influence for all

experiments, the other results presented in this work focused only on the CC among the other three-information obtained through the analysis of the images.

Table 3.3 – Estimates of the coefficients of determination (R^2) through the stepwise analysis based on the Akaike Hoking information criterion (AIC) obtained in the multiple regression equations involving tobacco leaf yield and the canopy coverage (CC), brightness index (BI), normalized green-red difference index (NGRDI), and visible atmospherically resistant index (VARI) assessed at the phenology stage flowering II for Experiment 1, 2 and 3.

	R^2 (%)	**Model Traits**
Experiment 1	76.3	CC + VARI + BI
Experiment 2	63.7	CC + VARI
Experiment 3	56.4	CC + NGRDI + VARI + BI

Source: from the author (2021).

Corroborating the previous results, the classifications of the hybrids/lines as for the CC and leaf yield tend to be similar. This observation may be confirmed with the data in Table 3.4, which compares the number of treatments belonging to the same group through the test by Scott and Knott (1974) for Experiments 1 and 3, i.e., experiments that presented significant differences among the hybrids/lines. A good agreement was found, considering that the test managed to classify the means in more than one group. It was observed that, in Experiment 1, there was a coincidence in the classification of nine hybrids, i.e., 60% of the assessed treatments. In Experiment 3, this percentage was even higher (75%).

At first, the most likely use of image phenotyping would be associating it with the data obtained manually, aiming to improve the selective efficiency. For such, the sum of the two standardized variables was estimated, i.e., the CC with at flowering II and the leaf yield, obtaining index Z. The analyses of variance involving Z were significant for Experiments 1 and 3, as occurred for the variables isolatedly (TABLE 5, APPENDIX A). Having the leaf yield as a reference, the inclusion of the datum collected via images, i.e., the canopy coverage, provided a slight increase in the estimates of $r_{g\hat{g}}$ (TABLE 3.5).

Table 3.4 – Number of hybrids/lines for the groups of means obtained via the test by Scott and Knott (1974) for the CC and tobacco leaf yield, and the number of hybrids/lines coinciding with the classification of the means for Experiments 1 and 3.

Scott-Knott	Experiment 1			Experiment 3		
	CC Flowering II	Yield	Coincident	CC Flowering II	Yield	Coincident
Group A	7	11	7	50	46	35
Group B	8	2	2	54	58	43
Group C	-	2	0	-	-	-

Source: from the author (2021).

Another aspect that could be explored is whether the use of image phenotyping could contribute to a possible reduction of the number of replicates used in the experiments. In the simulations of the use of different numbers of replicates, it was found that, for Experiment 1, the reduction of one replicate, i.e., from four to three replicates, using the index presented a relatively small accuracy decrease, lower than 2.5%. It draws attention that the estimate of $r_{gg'}$ for the yield with four replicates was the same as that using the Z index considering three replicates. In the case of Experiment 3, in which the number of replicates used was only three, the reduction to two replicates provided a 10% decrease in the estimate of $r_{gg'}$. However, the $r_{gg'}$ for index Z for the two replicates was, once again, equal to that obtained for yield with three replicates.

Table 3.5 – Estimates of the mean selective accuracy $(r_{gg'})$ for selection index Z for Experiments 1 and 3 considering all the replicates in the experiments and reducing the number of replicates.

Experiments	Number of Replicates	$r_{gg'}$	Confidence Interval
1	**4**	0.89	
	3	0.87	(0.84 – 0.90)
	2	0.81	(0.77 – 0.89)
3	**3**	0.70	
	2	0.63	(0.59 – 0.65)

Source: from the author (2021).

4 DISCUSSION

In tobacco crops, although there are several important traits, leaf yield is the most relevant, especially for producers, because the profit is directly related to the weight of commercialized cured leaves. However, besides being difficult, obtaining cured leaf mass in the breeding program has an associated error relative to the green leaf mass for operational reasons. For this reason, the assessment of cured leaf yield has only been performed in experiments that compose the Value for Cultivation and Use (VCU).

Evidence points to a high estimate of the correlation between green and cured mass of tobacco leaves; however, no reports were found in the literature. However, from the data obtained in Experiment 1, it was possible to verify a high correlation between the two variables $(r = 0.97)$, indicating, as expected, that the green mass represents the cured product well. Hence, in the other experiments in this work, the information obtained was only of the green leaf mass, as has commonly been used in the breeding program. Moreover, the term yield will be used in the discussion to identify the assessed commercial product, i.e., green mass and cured mass.

The estimates of accuracy $(r_{gg'})$ of the field experiments for yield presented an expressive magnitude, except in Experiment 2. When the $r_{gg'}$ is of high magnitude, one may infer that the experimental precision was good (RESENDE; DUARTE, 2007). Under this condition, the reliability of the results is evidently greater. However, as occurred in Experiment 2, with an estimate of $r_{gg'}$ of lower magnitude, one cannot infer that the precision was not good. This is because $r_{gg'}$ depends on the existence of a genetic variation among the hybrids/lines and also on the experimental error variance. Hence, if genetic variation does not occur, even if the experimental error is not of high magnitude, the accuracy tends to zero. Since, in the case of Experiment 2, the behavior of the lines was quite uniform, including the F test not being significant, in this condition, as has already been commented, one cannot state that the experimental precision was low. This type of behavior of the assessed lines hampered the inferences to be obtained from the association between image and yield in Experiment 2, and, thus, the discussion will be directed toward Experiments 1 and 3.

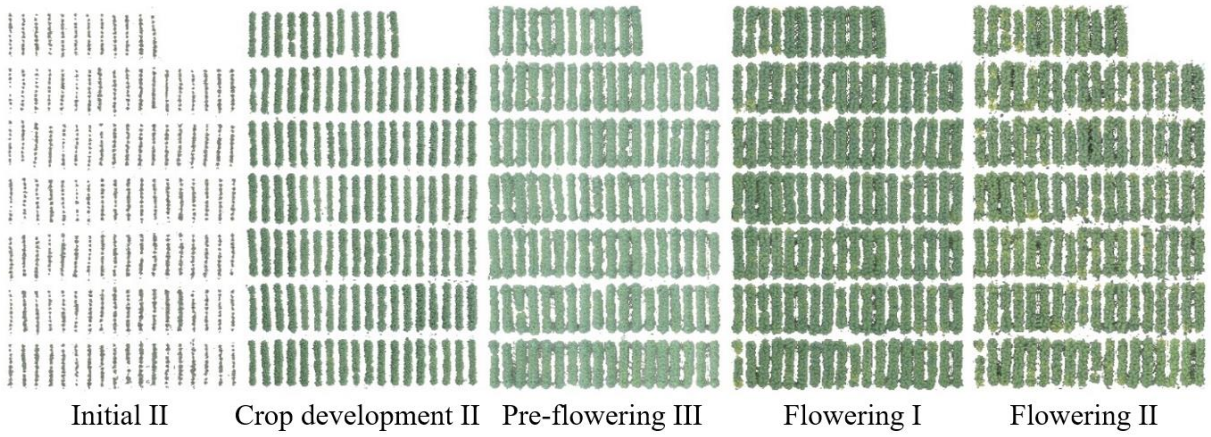
In general, the estimates of $r_{gg'}$ involving the information obtained via images, i.e., canopy coverage (CC) and the vegetative indices (VIs) of brightness index (BI), normalized green-red difference index (NGRDI), and visible atmospherically resistant index (VARI) may be considered of large magnitude. Reports in the literature also point to moderate to high

accuracy of phenotypic information obtained via image phenotyping for other crops. Relative to the CC, Makanza et al. (2018) found an estimate for r_{gg}, of 0.77 for maize crops. In soybean crops, the estimate was even higher (XAVIER et al., 2017). In work conducted in Brazil with bean crops, the r_{gg}, for the canopy coverage performed in the first days after the emergence of the plantlets was also high (NALIN; RAMALHO; CARVALHO, 2017). It must be highlighted that, in this case, no aerial images were used, and, besides, the research objective was to verify the difference in the speed of ground cover by the bean crop, aiming at reducing the occurrence of weed.

In the present research, r_{gg}, presented a difference relative to the time of obtainment of the images, i.e., the phenology stages. The estimates were of lower magnitude when obtained in the initial stages of the crop. It is likely that, in the first images, since the tobacco plants were still very small, the discrimination of the contrast between the leaf and the ground cover was not so expressive, as shown in Experiment 1 (FIGURE 4.1), which must have contributed to the greater experimental error estimate. Additionally, it was not possible to detect the difference among the treatments at the beginning because, as found in the analyses of variance, the source of variation "hybrids/lines" relative to the error variance increased with the advance of the crop. The opposite was reported for wheat crops by Fernandez-Gallego et al. (2019), who found estimates of r_{gg}, for the green area (GA), greener area (GGA), normalized green-red difference (NGRDI), and triangular greenness (TGI) indices were more significant at the beginning of the crop development, more specifically during the elongation and tillering of the plants.

In the case of tobacco, seemingly, the improvement of r_{gg}, with the advance of the crop stages, is a favorable condition. This is because, with the growth of the plant, one could assume that the self-shading of the leaves would reduce the possibility that image phenotyping would discriminate the assessed hybrids/lines. However, with the growth of the canopy, the tobacco plant typically resembles a cone and, thus, the larger the base of the cone is, the greater the yield should be and, consequently, the better the discrimination through image phenotyping will be.

Figure 4.1 – Images of the plots of Experiment 1 for assessing tobacco hybrids of varietal group flue-cured Virginia (FCV) collected via unmanned aerial vehicle (UAV) at different phenology stages.



Source: from the author (2021).

The proof of the previous observation, which is the focus of this work, was carried out through the verification of whether the differences found in image phenotyping reflect the tobacco plant yield. Some strategies were used to obtain this type of information. The first was through obtaining the estimate of Spearman's classifying correlation (STEEL; TORRIE; DICKEY, 1997). The estimate was particularly expressive for the CC, being superior to those obtained with the different VIs. Similar results were found in the literature for other crops such as cotton, maize, and soybean (FENG et al., 2020; GARCÍA-MARTÍNEZ et al., 2020; MAIMAITIJIANG et al., 2019).

It was expressive the fact that the estimates, for any of the employed methodologies, increased with the advance of the crop stage, as had already occurred with the $r_{gg'}$. Observe, for example, that, for Experiment 1, the estimate of the correlation between the CC and yield was 0.83 at flowering II, i.e., over 69% of the variation in yield was explained by the image. It must be emphasized that, by using the yield itself, it was verified through the analysis of variance (TABLE 1, APPENDIX A) that the source of variation hybrids explained 58% of the total variation of the experiment, disregarding the source of variation of blocks $\left({SQ_{Híbridos}}/{(SQ_{Híbridos} + SQ_{Erro})}\right)$. It is surmised that the CC explained the variation among the hybrids relative to yield with more intensity than the assessment of the cured leaf mass itself. Unfortunately, no reports were found in the literature regarding the correlation between yield and the phenotypic information obtained via images for tobacco crops. However, for other species such as soybean, maize, cotton, and rice, for example, the correlation estimates obtained were lower than those reported for tobacco in this research (FENG et al., 2020; GARCÍA-

MARTÍNEZ et al., 2020; MOREIRA et al., 2019; JARQUIN et al., 2018; MAKANZA et al., 2019; ZHOU et al., 2017).

However, in Experiment 3, the correlation estimates were of lower magnitude, with a correlation of 0.56 between yield and the CC at flowering II stage. Hence, the image explained only 31% of the variation, i.e., less than half of what occurred for varietal group FCV. The most likely hypothesis is that the difference among the lines of varietal group ACB in Experiment 3 was smaller than among the FCV hybrids, as the estimate of $r_{gg'}$ itself evinced. This observation reinforces the fact that the correlation between the image and the yield is always highly dependent on the occurrence of variation among the lines/hybrids. Still considering Experiment 3, the explanation of variation among the lines using green leaf mass and disregarding the block effect was greater ($R^2 = 45\%$) than the coefficient of determination obtained via the images, which, as commented, was 31%. This fact also reinforces another hypothesis to explain the lower estimate of the correlation between the CC and yield: the differences in the canopies of the plants in the ACB group were harder to detect through images. The differences in the plant bases are less pronounced than for the FCV group.

Another strategy to assess the efficiency of image phenotyping is the employment of multiple regression involving the dependent variable (yield) and the independent variables (the CC and the VIs) obtained in the flowering II stage for the reasons already commented. As expected, the results were coherent with the correlation estimates. However, observe that, in Experiment 1, the explanation of the variation in yield among hybrids was greater than when Spearman's classifying correlation of each piece of information obtained via images was estimated separately. It was found, however, that the CC was the variable that most explained the variation in yield and was present in all the models, as reported previously for other species (FENG et al., 2020; GARCÍA-MARTÍNEZ et al., 2020; MAIMAITIJIANG et al., 2019). It must be emphasized that the inclusion of the VIs contributed little to the fitting of the model.

The third strategy, of greater interest under the viewpoint of breeders, is how the grouping of the lines in the test by Scott and Knott (1974) behaves through image phenotyping (CC) and yield. It was found that, although the divergence among the hybrids/lines present in the experiments did not allow the formation of many groups, at most three, the correspondence in the grouping was, in general, high (TABLE 3.4). Hence, it is once again inferred that the employment of image phenotyping is feasible in the assessment of tobacco yield.

Finally, the possible applications of the employment of images in tobacco breeding remain to be commented on. At first, what the implication of the simultaneous employment of the information on CC and yield would be in the context of reducing the work in conducting

experiments. In other words, would the simultaneous use of the two variables enable reducing the number of replicates of the experiments? Although the number of replicates used in the experiments was not large, the analyses performed showed that it is viable to reduce replicates (TABLE 3.5) without compromising the experimental accuracy.

Another situation for applying image phenotyping is in the assessment of a large number of progenies/lines in experiments with replicates. Through the information obtained via the images, such as the CC, it would be possible to indirectly select only the progenies/lines with the worst performance to be discarded before the harvest of the leaves, reducing the operational costs of the assessments and, hence, allowing to evaluate a larger number of progenies. The continuity of the increment in tobacco yield in the future will certainly require the growing employment of progenies/lines to be assessed in the breeding program. For example, for employing doubled haploid (DH) lines in a recurrent selection (RS) program, the performance of a preliminary screening of the DH before they have been assessed with greater precision has been suggested, aiming at the identification of those that will be used in the recombination (LEMOS, 2021). Under this condition, the images from before the beginning of flowering should be collected quickly, acting as an early indirect selection, evidently without the need for obtaining the leaf mass. Hence, the DH with the worst performance are eliminated, and only those to be assessed later during the RS would be self-fertilized. For wheat crops, Fernandez-Gallego et al. (2019) stated that the early indirect selection allows for lower cost and less time spent, allowing breeders to increase the intensity of selection and, consequently, the gain. The authors recommend the use of image phenotyping aiming to obtain information on the VIs for the early prediction of wheat grain yield. Other work found in the literature also reinforces the possibility of performing indirect selection for yield through phenotypic information obtained via images, especially in the initial selection stages of the breeding program (GALLI et al., 2020; HU; KNAPP; SCHMIDHALTER, 2020; KRAUSE et al., 2020; MOREIRA et al., 2019; NATARAJAN et al., 2019; XAVIER et al., 2017).

5 CONCLUSIONS

i) Image phenotyping presented good accuracy (on average greater than 0.70) for detecting differences among the tobacco hybrids/lines. The accuracy increased with the advance of the crop stage.

ii) The canopy coverage (CC), assessed through image phenotyping, correlated with the green/cured leaf yield. The correlation ranged from 0.56 to 0.83% at the flowering II stage.

iii) The employment of image phenotyping must be stimulated in tobacco crops aiming the obtainment of the green/cured leaf yield.

iv) It was conjectured that the application of images is viable in several situations within a breeding program aiming at green/cured leaf yield.

REFERENCES

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716-723, 1974.

FENG, A. et al. Yield estimation in cotton using UAV-based multi-sensor imagery. **Biosystems Engineering**, v. 193, p. 101-114, 2020.

FERNANDEZ-GALLEGO, J. A. et al. Low-cost assessment of grain yield in durum wheat using RGB images. **European Journal of Agronomy**, v. 105, p. 146-156, 2019.

GALLI, G. et al. Optimization of UAS-based high-throughput phenotyping to estimate plant health and grain yield in sorghum. **The Plant Phenome Journal**, v. 3, n. 1, p. 1-14, 2020.

GARCÍA-MARTÍNEZ, H. et al. Corn grain yield estimation from vegetation indices, canopy cover, plant density, and a neural network using multispectral and RBG images acquired with unmanned aerial vehicles. **Agriculture**, v. 10, n. 277, p. 1-24, 2020.

GITELSON, A. A. et al. Novel algorithms for remote estimation of vegetation fraction. **Remote Sensing of Environment**, v. 80, n. 1, p. 76-87, 2002.

HEARST, Anthony Ahau. **Remote sensing of soybean canopy cover, color, and visible indicators of moisture stress using imagery from unmanned aircraft systems**. West Lafayette: Ed. Purdue University, 2019.

HU, Y.; KNAPP, S.; SCHMIDHALTER, U. Advancing high-throughput phenotyping of wheat in early selection cycles. **Remote Sensing**, Basel, v. 12, n. 574, p. 1-10, 2020.

JARQUIN, D et al. Increasing predictive ability by modeling interactions between environments, genotype and canopy coverage image data for soybeans. **Agronomy**, v. 8, n. 51, p. 1-15, 2018.

KRAUSE, M. R. et al. Aerial high-throughput phenotyping enables indirect selection for grain yield at the early generation, seed-limited stages in breeding programs. **Crop Science**, Madison, v. 60, p. 3096-3114, 2020.

LEMOS, Roxane do Carmo. **Performance of doubled haploid lines in relation to those obtained by the conventional breeding method in tobacco**. Lavras: Ed. UFLA, 2021.

MAIMAITIJIANG, M. et al. Vegetation index weighted canopy volume model (CVMVI) for soybean biomass estimation from unmanned aerial system-based RGB imagery. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 151, p. 27-41, 2019.

MAKANZA, R. et al. High-throughput phenotyping of canopy cover and senescence in maize field trials using aerial digital canopy imaging. **Remote Sensing**, Basel, v. 10, n. 330, p. 1-13, 2018.

MATIAS, F. I.; CARAZA-HARTER, M. V.; ENDELMAN, J. B. FIELDImageR: An R package to analyze orthomosaic images from agricultural field trials. **The Plant Phenome Journal**, v. 3, n.1, p. 1-6, 2020.

MOREIRA, F. F. et al. Improving the efficiency of soybean breeding with high-throughput canopy phenotyping. **Plant Methods**, London, v. 15, n. 139, p. 1-9, 2019.

NALIN, R. S.; RAMALHO, M. A. P.; CARVALHO, B. L. A genetic study of traits related to canopy coverage in common bean. **Crop Breeding and Applied Biotechnology**, Viçosa, v. 17, p. 313-319, 2017.

NATARAJAN, S. et al. High-throughput phenotyping of indirect traits for early-stage selection in sugarcane breeding. **Remote Sensing**, Basel, v. 11, n. 2952, p. 1-18, 2019.

ORLANDO, F. et al. Growth and development responses of tobacco (Nicotiana tabacum L.) to changes in physical and hydrological soil properties due to minimum tillage. **American Journal of Plant Sciences**, United States, v. 2, p. 334-344, 2011.

R CORE TEAM. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2020.

RESENDE, M. D. V.; DUARTE, J. B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical**, Goiânia, v. 37, n. 3, p. 182-194, 2007.

RICHARDSON, A. J.; WIEGAND, C. L. Distinguishing vegetation from soil background information. **Photogrammetric engineering and remote sensing**, Maryland, v. 43, n. 12, p. 1541-1552, 1977.

SCOTT, A. J.; KNOTT, M. A cluster analysis method for grouping means in the analysis of variance. **Biometrics**, Washington, v. 30, n. 3, p. 507-512, 1974.

STEEL, R. G. D.; TORRIE, J. H.; DICKEY, D. A. **Principles and procedures of statistics: a biometrical approach**. New York: McGraw-Hill Book Company, 1997.

TOBACCO GROWER'S ASSOCIATION OF BRAZIL (AFUBRA). Rio Grande do Sul, 2021. Available at: < https://afubra.com.br/fumicultura-brasil.html>. Accessed on: July 6, 2021.

TUCKER, C. J. Red and photographic infrared linear combinations for monitoring vegetation. **Remote Sensing of Environment**, v. 8, n. 2, p. 127-150, 1979.

VENABLES, W. N.; RIPLEY, B. D. **Modern Applied Statistics with S**. 4 ed. New York: Ed. Springer, 2002.

XAVIER, A. et al. Genetic architecture of phenomic-enabled canopy coverage in *Glycine max*. **Genetics**, Austin, v. 206, p. 1081-1089, 2017.

YANG, W. et al. Crop phenomics and high-throughput phenotyping: past decades, current challenges, and future perspectives. **Molecular Plant**, v. 13, n. 2, p. 187-214, 2020.

ZHOU, X. et al. Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 130, p. 246-255, 2017.

APPENDIX A

Table 1 – Summary of the analyses of variance for cured leaf yield (kg plot^{-1}) for Experiment 1 and green leaf yield (kg plot^{-1}) for Experiments 2 and 3.

Experiment 1			Experiment 2			Experiment 3		
SV	**DF**	**MS**	**SV**	**DF**	**MS**	**SV**	**DF**	**MS**
Hybrids	14	0.75**	**Lines**	59	12.61^{ns}	**Lines**	103	12.19**
Replicates	3	0.08^{ns}	**Replicates**	2	55.20**	**Replicates**	2	133.10**
Error	42	0.18	**Error**	118	9.92	**Error**	206	7.38

** and ^{ns}: significant ($p \leq 0.01$) and non-significant ($p > 0.05$) according to the F test, respectively.

Source: from the author (2021).

Table 2 – Summary of the analyses of variance for the canopy coverage (CC), brightness index (BI), normalized green-red difference index (NGRDI), and visible atmospherically resistant index (VARI) at different phenology stages for experiment 1.

			MS				
	SV	**DF**	**Initial II**	**Crop development II**	**Pre-flowering III**	**Flowering I**	**Flowering II**
CC	**Hybrids**	14	1.61^{ns}	24.59^{ns}	44.61**	51.04**	49.54**
	Replicates	3	0.97^{ns}	17.81^{ns}	16.99^{ns}	107.60**	50.45**
	Error	42	0.92	15.05	10.16	9.31	9.58
BI	**Hybrids**	14	7.32^{ns}	35.94^{ns}	16.04**	20.29**	16.73**
	Replicates	3	43.68**	46.83^{ns}	7.71^{ns}	10.53^{ns}	6.42^{ns}
	Error	42	6.42	30.66	2.97	5.47	5.36
NGRDI	**Hybrids**	14	0.00001^{ns}	0.00004^{ns}	0.00001*	0.00006**	0.00006*
	Replicates	3	0.00007**	0.00021**	0.00001^{ns}	0.00002^{ns}	0.00001^{ns}
	Error	42	0.00001	0.00003	0.00000	0.00002	0.00002
VARI	**Hybrids**	14	0.00003^{ns}	0.00013^{ns}	0.00003*	0.00016**	0.00021**
	Replicates	3	0.00026**	0.00054**	0.00005*	0.00004^{ns}	0.00006^{ns}
	Error	42	0.00003	0.00011	0.00002	0.00003	0.00008

**, *, and ^{ns}: significant ($p \leq 0.01$), significant ($p \leq 0.05$), and non-significant ($p > 0.05$) according to the F test, respectively.

Source: from the author (2021).

Table 3 – Summary of the analyses of variance for the canopy coverage (CC), brightness index (BI), normalized green-red difference index (NGRDI), and visible atmospherically resistant index (VARI) at different phenology stages for Experiment 2.

			MS			
	SV	**DF**	**Initial II**	**Crop development II**	**Pre-flowering III**	**Flowering II**
CC	**Lines**	59	0.15^{ns}	11.28^{ns}	25.67^{**}	75.64^{ns}
	Replicates	2	2.95^{**}	90.09^{**}	84.76^{**}	418.13^{**}
	Error	118	0.12	8.45	14.62	68.07
BI	**Lines**	59	24.25^{*}	52.60^{ns}	17.02^{**}	35.81^{**}
	Replicates	2	395.97^{**}	93.33^{ns}	38.81^{**}	142.43^{**}
	Error	118	16.25	38.83	3.72	19.16
NGRDI	**Lines**	59	0.00001^{ns}	0.00003^{ns}	0.00003^{**}	0.00009^{*}
	Replicates	2	0.00003^{ns}	0.00010^{**}	0.00075^{**}	0.00084^{**}
	Error	118	0.00001	0.00002	0.00001	0.00006
VARI	**Lines**	59	0.00003^{ns}	0.00010^{ns}	0.00008^{**}	0.00029^{*}
	Replicates	2	0.00007^{ns}	0.00014^{ns}	0.00274^{**}	0.00306^{**}
	Error	118	0.00004	0.00009	0.00004	0.00018

**, *, and ^{ns}: significant ($p \leq 0.01$), significant ($p \leq 0.05$), and non-significant ($p > 0.05$) according to the F test, respectively.

Source: from the author (2021).

Table 4 – Summary of the analyses of variance for the canopy coverage (CC), brightness index (BI), normalized green-red difference index (NGRDI), and visible atmospherically resistant index (VARI) at different phenology stages for Experiment 3.

			MS			
	SV	**DF**	**Initial II**	**Crop development II**	**Pre-flowering III**	**Flowering II**
CC	**Lines**	103	0.23^{ns}	14.64^{**}	31.90^{**}	70.08^{**}
	Replicates	2	1.55^{**}	132.60^{**}	357.66^{**}	69.13^{ns}
	Error	206	0.22	8.61	12.74	29.12
BI	**Lines**	103	20.46^{**}	41.59^{**}	46.36^{**}	79.78^{**}
	Replicates	2	118.59^{**}	292.03^{**}	201.22^{**}	169.54^{**}
	Error	206	13.31	18.33	16.45	28.49
NGRDI	**Lines**	103	0.00001^{**}	0.00005^{**}	0.00006^{**}	0.00012^{**}
	Replicates	2	0.00009^{**}	0.00031^{**}	0.00041^{**}	0.00258^{**}
	Error	206	0.00001	0.00001	0.00002	0.00005
VARI	**Lines**	103	0.00005^{**}	0.00014^{**}	0.00018^{**}	0.00030^{**}
	Replicates	2	0.00028^{**}	0.00077^{**}	0.00139^{**}	0.00572^{**}
	Error	206	0.00003	0.00004	0.00006	0.00011

**, *, and ^{ns}: significant ($p \leq 0.01$), significant ($p \leq 0.05$), and non-significant ($p > 0.05$) according to the F test, respectively.

Source: from the author (2021).

Table 5 – Summary of the analyses of variance for the Z index for Experiment 1 (CC Flowering II + Yield) and Experiment 3 (CC Flowering II + Yield).

Experiment 1			**Experiment 3**		
SV	**DF**	**MS**	**SV**	**DF**	**MS**
Hybrids	14	9.31**	**Lines**	103	4.79**
Replicates	3	0.00ns	**Replicates**	2	0.00ns
Error	42	1.85	**Error**	206	2.42
$r_{g\hat{g}}$	0.89		$r_{g\hat{g}}$	0.70	

** and ns: significant ($p \leq 0.01$) and non-significant ($p > 0.05$) according to the F test, respectively.
Source: from the author (2021).