BRUNA LINE CARVALHO

BREEDING FOR MULTIPLE TRAITS USING PHENOTYPIC AND GENOTYPIC INFORMATION IN TOBACCO FOR TROPICAL CONDITIONS

LAVRAS - MG

2016

BRUNA LINE CARVALHO

BREEDING FOR MULTIPLE TRAITS USING PHENOTYPIC AND GENOTYPIC INFORMATION IN TOBACCO FOR TROPICAL CONDITIONS

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 a obtenção do título de Doutor.

Orientador

Dr. Magno Antonio Patto Ramalho

LAVRAS - MG

2016

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

Carvalho, Bruna Line.
Breeding for multiple traias using phenotypic and genotypic information in tobacco for tropical conditions / Bruna Line Carvalho. – Lavras : UFLA, 2016.
157 p. : il.

Tese(doutorado)–Universidade Federal de Lavras, 2015.
Orientador(a): Magno Antonio Patto Ramalho.
Bibliografia.

1. Plant Breeding. 2. Quantitative Genetics. 3. Combining Ability. 4. Selection Index. 5. Genomic Prediction. I. Universidade Federal de Lavras. II. Título.

BRUNA LINE CARVALHO

BREEDING FOR MULTIPLE TRAITS USING PHENOTYPIC AND GENOTYPIC INFORMATION IN TOBACCO FOR TROPICAL CONDITIONS

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 a obtenção do título de Doutor.

APROVADA em 18 de dezembro de 2015

Dr. Ramsey Steven Lewis	NCSU
Dr. Carlos Eduardo Pulccineli	BAT
Dr. César Augusto Brasil Pereira Pinto	UFLA
Dr. Adriano Teodoro Bruzi	UFLA

Orientador

Dr. Magno Antonio Patto Ramalho

LAVRAS - MG

2016

To my parents for being the base of everything in my life.

OFEREÇO

In memoriam of my grandpa, who construct a progeny to be proud of.

DEDICO

AKNOWLEGMENTS

A Coordenação de Aperfeiçoamento de Pessoal de nível Superior (CAPES), pela concessão da bolsa de estudos.

A Souza Cruz S/A - BAT, por fornecer recursos para realização dessa pesquisa. Agradeço aos colaboradores que participaram da elaboração desse trabalho, em especial ao Dr. Carlos Eduardo Pulcinelli, pelo envolvimento e por abrir as portas para a interação entre a universidade e a empresa. Agradeço a toda equipe do Centro de Melhoramento de Tabaco (CMT) pela participação e condução do projeto.

Agradeço a Universidade Federal de Lavras e ao programa de Pós-Graduação em Genética e Melhoramento de Plantas, pela oportunidade da realização de um sonho. Ao meu orientador, professor Magno A. P. Ramalho, agradeço por cada minuto dispendido comigo, por toda atenção, ajuda e ensinamentos. Você é exemplo de dedicação e entusiasmo. Espero sinceramente que essa jornada não termine por aqui.

To the North Carolina State University (NCSU), where I spent some time during my phD. I recognize to everyone who helped me in some way. My sincerely acknowledge to professor Ramsey Lewis who gave me the opportunity to be at NCSU. Thank you for trusting on me and providing me experience for expanding my knowledge in Plant Breeding and Genetics. You are an example of professor, advisor and researcher. And I appreciate for all of his team, Katherine, Sherry, Jessica, Justin, Travis, Carrie, Katlyn and others, for the support and nice time.

Agradeço a todos os professores que participaram da minha formação. Á professora Édila V. R. Von Pinho que foi meu primeiro exemplo profissional, por todo incentivo para iniciar a jornada de pesquisa. À pesquisadora Dr. Ângela F. B. Abreu, por toda ajuda, apoio, carinho e amizade. Ao prof. Adriano T. Bruzi, pela co-orientação no doutorado, pelo apoio, conselhos e amizade. Aos professores Daniel Furtado, Marcio Balestri, Antonio Augusto Garcia por estar sempre de portas abertas quando precisei de apoio e por toda a contribuição. A todos os professores da UFLA que participaram da minha formação, especialmente ao professor César Brasil, pelas discussões e conversas proveitosas e por aceitar o convite para ser membro da banca de defesa dessa tese.

A toda minha família por estarem sempre presentes, mesmo em minha ausência. Em especial aos meu pais, Edna e Custódio, pelo grande amor e suporte.

Aos amigos e colegas da UFLA, pelo companheirismo. Em especial a Scheila, Lala, Narjara, Marcela, Cynthia, Jessica, Samira e Suellen. Ao colegas do GEN pelos momentos de discussão e debate, e por tornar nosso ambiente de estudo mais agradável. Às amigas de Lavras pela distração, apoio e amizade.

E a todos, que de algum modo, direto ou indireto, contribuíram para que fosse possível a realização desse trabalho, meus sinceros agradecimentos.

“Nature’s stern discipline enjoins mutual help at least as often as warfare. The fittest may also be the gentlest.”

Theodosius Dobzhansky

LIST OF ABREVIATIONS

Ac: Accuracy
ALK: Total alkaloid content
BAT: British American Tobacco
BB: Bayes-B
BL: Bayes Lasso
BLY: Burley
BRR: rr-BLUP
CI: Coincidence index
CMV: Canonic multivariable
CV: Coefficient of variation
CVS: Ratio of conversion from nicotine to nornicotine
DF: Degrees of freedom
FCV: Flue-Cured Virginia
GBS: Genotyping-by-Sequencing
GQI: General Quality Index
GCA: General Combining Ability
GWS: Genome-wide selection
SLR: Steam by leaf lamina ratio
L1: Location 1
L2: Location 2
MS: Mean Square
r^2: repeatability
R^2: Coefficient of determination
SCA: Specific Combining Ability
SS: Sum of Squares
SSV: Sum of the standardized variable
SUG: Total sugar content
TRN: training population
TST: testing population
YLD: Yield

LIST OF FIGURES

FIRST PART

SECOND PART – ARTICLES

ARTICLE 1

ARTICLE 2

ARTICLE 3

LIST OF TABLES

SECOND PART – ARTICLES

ARTICLE 1

SUMMARY

FIRST PART

1 INTRODUCTION

The tobacco crop has an enormous socio-economic importance in Brazil. Economic because it is one of the ten agribusiness products that most earn export revenue for Brazil (AFUBRA, 2015; FAO, 2012). And social because it is a crop that involves some hundreds of thousands of family farmers whose main source of income is tobacco, since it is a highly profitable crop (BTY, 2014; SCHUCH, 2003; PRIEB, 2005).

To guarantee that tobacco presents better quality and that the producers have higher economic returns, companies produce every year new cultivars by plant breeding programs that present superior phenotypes for the characteristics of interest compared to the cultivars already on the market. However, it takes approximately 10 years to release a new cultivar, so long-term planning is necessary for success in the breeding programs (COOPER et al., 2014).

In the current system of tobacco seed production, hand crossings are made to combine phenotypes from traits of interest present in different lines and especially to preserve the intellectual property (IP) of the cultivar produced by the company using cytoplasmic male sterility (AYCOCK; MANN; MATZINGER, 1963; MANN; JONES; MATZINGER, 1962; SCHNABLE; WISE, 1998). In this way, in the farmer's field, the plants do not produce seeds. Consequently the use of hybrids does not incur any additional labor cost. Therefore, the exploration of heterosis, even of small magnitude, would be advantageous.

To adopt hybrids, it is necessary to identify superior lines that perform well when used in crosses. This is the most time spending and expensive part of plant breeding programs. This is because even with a few lines, the number of

hybrid combinations is large and these hybrids should be assessed in field experiments to identify the best ones. The diallelic crosses is the methodology most used when there are some lines and it is required to identify those that combine best for hybrid production (CRUZ; CARNEIRO; REGAZZI, 2014). This methodology has been widely used in many species, including tobacco (BUTORAC; BELJO; GUNJACA, 2004; PARKES et al., 2013; QI et al. 2013; MATZINGER; MANN; COCKERHAM, 1972; XIAO et al., 2007).

Several traits should be considered in breeding programs. In the case of tobacco, in addition to yield, traits related to leaf chemical and physical aspects are extremely important because they directly affect the end product quality. There are some alternatives when selecting for several traits (BERNARDO, 2014; FALCONER; MACKAY, 1996; LYNCH; WALSH, 1998). The most efficacious is obtaining a linear index that considers all the traits of interest, with their due weights (BAKER, 1986; RESENDE; SILVA; AZEVEDO, 2014).

It is important to mention that the lines and hybrids developed in tobacco breeding programs are high performing, therefore it becomes even more difficult to detect differences among the genotypes. A more recent approach is combining the phenotypic information assessed in experiments with genetic information obtained by the new high-throughput sequencing techniques, that is, carry out genomic-wide selection (GWS) (HESLOT; JANNINK; SORRELLS, 2015; MEUWISSEN; HAYES; GODDARD, 2001). One of the proposals would be to obtain a predictive model in the first cycle of the breeding program based on genomic and phenotypic information of the population, and using this model for selection in posterior cycles based only genomic information of individual/progenies (CROSSA et al., 2010). Another application of GWS would be to predict the hybrid performance using only the genetic and phenotypic information of lines (RIEDELSHEIMER; TECHNOW; MELCHINGER, 2012; SU et al., 2012). Thus only the genetically superior

hybrids would be tested in the field and consequently resources could be concentrated only on these crosses (COOPER et al., 2014).

The chapters in this thesis have the objective of approaching three aspects of tobacco breeding. i) Assess the general combining ability and specific combining ability by diallel crosses; ii) Identify an index for simultaneous multitrait selection, iii) Verify the feasibility of using GWS in the tobacco breeding program.

2 LITERATURE REVIEW

2.1 Tobacco crop and breeding

The *Nicotiana tabacum* species (tobacco) has been cultivated since pre historical times and it is considered the crop with the largest economic value in the world among the non-foodstuffs species (BELOGRADOVA et al., 2009; KNAPP et al., 2004). *N. tabacum* is an allotetraploid, 2n=4x=48 chromosomes and was the result of a crossing event between the species *N. sylvestris* (2n=24) and *N. tomentosiformis* (2n=24), followed by chromosome duplication (CLARKSON et al., 2005; LEWIS; NICHOLSON, 2007; LIM et al., 2004; MURAD et al., 2002)

The tobacco plant has enormous advantages in terms of research in breeding and/or molecular biology because of its characteristics. They include sexual/asexual propagation; easy manual hybridization; it is an autogamous species with a relatively short cycle, wide variability and it produce a large number of seeds per plant. For these and other reasons, it has been widely used as a model plant in various research programs worldwide (MUELLER et al., 2005; LU et al., 2012).

The crop is divided in several varietal groups, based on morphology, quality of leaf, curing method and other criteria. The main groups are Virginia, Burley, Oriental, Dark and Cigar. The groups that detain the larger slice of the market are Flue-Cured Virginia (FCV) and Burley (BLY). In FCV varietal group the harvest should be scaled according to ripeness and curing is carried out in chambers where the temperature and humidity are very precisely controlled. After curing, the leaves present a yellow and orange coloring, with a high concentration of reducing sugars (WERNSMAN; RUFTY, 1987). The BLY varietal group presents a color ranging from pale green to yellow, due to

the presence of recessive alleles that effect chlorophyll production in the plants (CLAUSEN; CAMERON, 1944). It is harvested by cutting the plant close to soil level. Later the whole plant is hung in open sheds for the curing process. When it ends, the leaves have a tan to reddish brown coloring, with low reducing sugar concentration (WERNSMAN; RUFTY, 1987). As noted in the two examples described, each group presents particular characteristics for industry and so all of them have been used in the breeding programs. Plants derived from crosses among lines of different groups usually present characteristics related to undesirable quality (GARNER; ALLARD; CLAYTON, 1936; CLAYTON, 1958) and therefore this strategy has been avoided.

Brazil is the second largest world producer and biggest exporter of tobacco in the world (FAO, 1012), additionally tobacco is one of the ten most important products in Brazilian exportation. The total production in the country in the 2013/14 growing season was 751 mil tons, but the FCV represented more than 80% of this production (AFUBRA). Tobacco is cropped by some hundreds of thousands of producers, predominantly family farmers, who represent 92% of the total (SCHUCH, 2003), and the crop is their main source of income. In the southern region alone, there are more than 150,000 family producers in this activity (AFUBRA, 2015). These data reflect the magnitude of the economic and social importance of tobacco for the development of the country.

The mean annual increase in yield in the last 50 years has been approximately 26 kg/ha per year (Figure 1; AFUBRA, 2014; FAO, 2012), from this total 38% can be attributed to genetic progress (BOWMAN et al., 1984). In addition, there has been a significant improvement in the physical quality of the leaves of the cultivars developed, which has resulted in an additional annual gain of $26/kg (SARCEVIC et al., 2013).

Tobacco genetic breeding, with scientific focus, started in the 20th century (EAST; JONES, 1921; SHAMEL; COBEY, 1907). Gains in yield and

increase in resistance to several pathogens have been reached but the minimum quality requirements have limited these gains due to correlation between yield and quality traits (MOON et al., 2009). The quality components of the tobacco leaf have complex genetic control (WERNSMAN; RUFTY, 1987), and so it is not simple to transfer favorable alleles from exotic germplasm to commercial lines. Thus the ideal is to identify genotypes and pre-breeding to reach acceptable yield levels before introducing them in breeding programs.

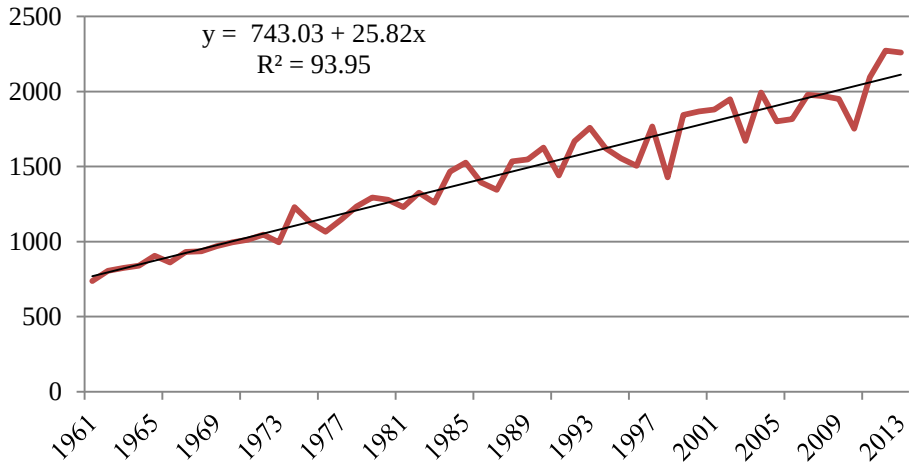

Figure 1 Linear regression equation and coefficient of determination (R^2) for cured leaf yield (kg/ha) in Brazil. Data obtained from 1961 to 2013.
Source: AFUBRA (2014) and FAO (2012)

The performance stability of the cultivars is also important for the traits of interest. Yield and stability are genetically independent (SPRAGUE; FEDERER, 1951; TOLLENAAR; LEE, 2002), that is, lines with genetic potential to reach high yields in specific environments may not be adapted to a wide range of environments. Several studies have been published showing the significant effects of the genotype x year interaction and genotype x environment interactions for several traits such as yield and leaf number and size

(POVILAITIS, 1970; BOWMAN et al., 1986; CASTELLI et al., 1994). It is important to mention that the hybrids present larger genetic homeostasis because they have higher heterozygosis and therefore they would theoretically be more stable when cultivated in different environments (ALLARD; BRADSHAW, 1964; BRUZI; RAMALHO; FERREIRA, 2007; RAMALHO et al., 2012a)

2.2 Heterosis in autogamous plants

Heterosis, or hybrid vigor, is the phenomenon when the F_1 generation of the cross between two parents presents superior performance to the mean of the parents (CHEN, 2013; RAMALHO et al., 2012b). This phenomenon was discovered by Shull (1908) and according to many, is one of the main discoveries of humanity (GRUN, SIMPSON, 2005) This discovery was the starting point for the creation of all the seed industry that occurred in the 20^{th} century. In spite of its enormous importance and the advances in molecular biology, the molecular mechanisms of its occurrence are still unknown (BIRCHLER et al, 2010).

The use of commercial hybridization is predominant in allogamous plants (SOUZA JR, 2011), such as corn, onion, carrots and many others. In autogamous plants, its use has been restricted to some species, including rice (GOFF; ZHANG, 2013), sorghum (BEN-ISRAEL et al., 2012) and tomato (KRIEGER; LIPPMAN; ZAMIR, 2010). The main reason is that, due to the reproduction mode, the contribution of heterosis in relation to the mean of the parents is normally low (RAMALHO et al., 2012a). Thus hybrid production is only viable when they can be used at a cost consistent with their performance (FEHR, 1987). It should be mentioned that in addition to exploiting heterosis, hybrids can present larger homeostasis, because loci in heterozygosis contribute to the buffer effect in contrasting environments and therefore they are more

stable (ALLARD; BRADSHAW, 1964; BRUZI; RAMALHO; FERREIRA, 2007; RAMALHO et al., 2012a).

To adopt hybrids, the line complementarity should be assessed based on the combining ability estimates. One of the methodologies most used are the diallelic crosses (HALLAUER; CARENA; MIRANDA FILHO, 2010). This methodology has been adopted in several species to identify superior combinations (QI et al. 2013; PARKES et al., 2013). There are several published studies that report this kind of information in temperate conditions for the different tobacco varietal groups (DEAN, 1974; IBRAHIM; SLAVÍK; AVRATOVSCUKOVÁ, 1984; MATZINGER; MANN; COCKERHAM, 1972; MATZINGER; WERNSMAN; ROSS, 1971). However, it is believed that in tropical conditions the results may be different, because tobacco is a species significantly influenced by the environmental conditions where it is grown, mainly in the qualitative aspects.

Although heterosis has not been deeply exploited in the tobacco crop (ALEKSOSKI, 2010), the hybrid vigor obtained by line crossing within and among different varietal groups has been reported for several traits in some published articles (BUTORAC et al., 2000a, 2000b; KARA; ESENDAL, 1995; KORUBIN-ALEKSOSKA; ALEKSOSKI, 2012; KRISHNAMURTHY et al., 1994; WILKINSON et al., 1994). In most of the studies predominance of additive effects for traits related to yield was found. Among the varietal groups, BLY has the highest crossing rates, and can reach up to 19% (LITTON; STOKES, 1964). There is no information whether this fact can affect the heterosis magnitude in this varietal group.

Several studies have reported positive and significant correlation between the divergences assessed by molecular markers and heterosis for several species, including corn (LEGESSE et al., 2008), wheat (KRYSTKOWIAK et al., 2009) and sorghum (JORDAN, 2003). Some authors have suggested that

molecular divergences of the lines could be a good predictor of the hybrid performance (BETRAN et al., 2003), however none of these authors confirmed that genetic markers can accurately predicted heterosis, because the magnitudes of these estimates, although significant, are not high. It should be mentioned that in almost all of these studies it were used molecular markers that have a low cover of the genome such as SSR, RAPD, AFLP; and in most cases, the markers used were not associated to the trait of interest. An alternative would be the use of SNP type markers for a higher coverage of the genome, so that a more accurate prediction of the hybrid performance could be obtained. This fact has received enormous attention in molecular breeding, in the area called genome wide selection (GWS) (HEFFNER et al., 2010; MASSMAN et al., 2013; MEUWISSEN; HAYES; GODDARD, 2001).

2.3 Indices for multitrait selection

The genetic gains in a breeding program do not only depend on a specific trait but rather on several (CERÓN-ROJAS; CROSSA; SAHAGUN-DASTELLANOS, 2015). Some crops, such as tobacco for example, this is very clear, because in addition to agronomic characteristics, such as yield and disease resistance, the traits linked to physical chemical quality are also extremely important. This is because the end product is directly used by the consumer. Under this condition, the breeders have three alternatives for selection (BERNARDO, 2014; FALCONER; MACKAY, 1996; YAN; FRÉGEAU-REID, 2008). Selection *in tandem*, when selection is made initially for one trait until the desired level is reached, then for a second trait and so on. Another alternative, perhaps the most used in practice, is selection by independent culling levels. In this case, a limit is considered for each trait and the genotypes above or below these limits are eliminated. Finally, the third option is the use of

selection indices, and this is considered the most efficient method (BAKER, 1986 LYNCH; WALSH, 1998; RESENDE; SILVA; AZEVEDO, 2014; YAN; FRÉGEAU-REID, 2008).

There are several reports in the literature on the theory of applying the selection indexes to plants (BAKER, 1986; BERNARDO, 2010, 2014; WRICKE; WEBER, 1986). Firstly, it is necessary to know the traits of interest individually to establish the most appropriate economic weights to obtain the desired gains (HIDALGO et al., 2014). It is also important to be aware of the association among the traits, which can benefit or hinder the work of the breeder in selecting superior genotypes for multi traits in breeding programs. Genetic correlation can occur due to two events, genetic linkage or pleiotropy (FALCONER; MACKAY, 1996). If the genes responsible for the control of the two traits are linked, recombinant genotypes can be obtained by assessing a large number of individual/progenies. On the other hand, if the genes are pleiotropic, it means that the same gene controls both traits and so it is not possible to obtain recombinant genotypes. It should be emphasized that if the trait is quantitative, i.e. controlled by a large number of genes, some genes may be linked, pleiotropic or distributed independently.

When there is correlation, to obtain the maximum gain, the breeding strategies should focus on breaking unfavorable linkage or substituting alleles with pleiotropic effects with others that have less influence on yield (LERNER, 1958). An example in tobacco is the negative correlation observed between the yield components and the percentage of total alkaloids in some countries, e.g. the United States (LEGG; MATZINGER; MANN, 1965; LEWIS, 2006; MATZINGER; MANN, 1964; MATZINGER; MANN; COCKERHAM, 1972; MATZINGER; MANN; ROBINSON, 1960; MATZINGER; WERNSMAN, 1968; MATZINGER; WERNSMAN; WEEKS, 1989). As the cultivars developed should present high yield and level of alkaloids in that country, this

correlation hinders obtaining cultivars with the desired phenotype. Lewis (2006) identified not only genotypes with acceptable total alkaloid proportions and good yield but also genotypes with low association between yield and total alkaloid percentage. This shows that, in the breeding programs, a screening could be carried out in the search for lines with characteristics similar to those found by this author. However, it should be emphasized that in the environmental conditions of Brazil, the same correlation is not observed.

The pioneer index for selecting traits that are correlated was called the Smith-Hazel method (SMITH, 1936; HAZEL, 1943). Theoretically, it is the most accurate in theory, but it requires that reliable estimates of phenotypic and genetic variances to be obtained, which is not always possible. An alternative would be the use of canonic correlations to construct an index (RESENDE; SILVA; AZEVEDO, 2014). Canonic analysis is a multivariate technique based on the correlation between the linear combinations of a set of variables. (AKBAS; TAKMA, 2005; CERÓN-ROJAS; CROSSA; SAHAGUN-DASTELLANOS, 2015; HIDALGO et al., 2014; VENTURA et al., 2011). According to Resende; Silva; Azevedo (2014), the canonical transformation of the original variables allows the construction of a selection index of maximum efficiency.

A very important point for an index to be adopted is that it should be easy to interpret and analyze. In this context we can cite the sum of the standardized variables index (RAMALHO et al., 2012a). For its application, the variables are first standardized and later they are summed for each observation. The gain for each trait can be observed on a graph known as the full/empty ball, where the cultivars with homogeneous performance for all the traits form a graph of full ball shape or when there is deficiency in one or more variable, the empty ball shape graph (RAMALHO et al., 2012a). It has frequently been used in the literature because it is easy to visualize (REIS et al., 2011; MENDES;

RAMALHO; ABREU, 2009). There are other options of indices that are shown in a numerous texts (CRUZ; CARNEIRO; REGAZZI, 2014; FALCONER; MACKAY, 1996; LYNCH; WALSH, 1998; RESENDE; SILVA; AZEVEDO, 2014).

2.4 Genome wide selection (GWS)

Molecular markers have been used for decades as a tool for selection (LANDE; THOMPSON, 1990; STUBER; GOODMAN; MOLL, 1982; TANKSLEY et al., 1989). They generated a high expectation and many thought that they could substitute phenotype assessment. However, this success was restricted to simple heredity traits or traits controlled by a few large effects QTLs - Quantitative Trait Loci (XU; CROUCH, 2008). It is important to mention that most of the traits of agronomic interest have a complex genetic control, so most of genes are of small effect, and therefore application of the marker assisted selection (MAS) for these cases has been limited (BERNARDO, 2008; CROSBIE et al., 2003; DEKKERS; HOSPITAL, 2002; MOREAU; CHARCOSSET; GALLAIS, 2004). Furthermore, establishing linkage between the markers and QTLs is made with populations derived from bi-parental crosses so that when applied to other populations in the breeding program the markers are frequently not useful. Theoretically, the ideal would be to use all the QTLs associated to the trait of interest.

An alternative is the use of high-density markers that span the whole genome. This gave rise to GWS that is based on the simultaneous estimation of the effects for all the markers of the training population (TRN), in which the genotypes and phenotypes of the individuals/progenies are available (MEUWISSEN; HAYES; GODDARD, 2001). Thus with a dense molecular map, some markers are much closer to the QTL's of interest, resulting in linkage

disequilibrium. In this way, there are correlation between the effect of some markers and the trait assessed without the need to establish a linkage phase (MEUWISSEN; HAYES; GODDARD, 2001). However, the number of effects to be estimated would be larger than the number of observations and therefore there would be not sufficient degrees of freedom to jointly estimate the effects of the markers by the least-squares method (LANDE; THOMPSON, 1990; HASTIE; TIBSHIRANI; FRIEDMAN, 2009). A series of statistical models is described in the literature to estimate the effect and variance of the markers, which differ basically in the *priori* adopted (DE LOS CAMPOS et al., 2013). Simulation-based studies have shown differences in the predictive power among the methods adopted (CLARK; HICKEY; VAN DER WERF, 2011; COSTER et al., 2010; DAETWYLER et al., 2010;). Several factors can affect prediction, for example, the genetic architecture of the trait under selection, linkage disequilibrium between markers, sample size, heritability of the trait, and marker density (DAETWYLER et al., 2010; HABIER et al., 2010; ZHONG et al., 2009). The accuracy would be higher whenever the chosen model fits better to the trait genetic architecture (LUND et al., 2009; BASTIAANSEN et al., 2010; PSZCZOLA et al., 2011). However many times the population structure and the architecture of trait under selection are not known, and this hinders the choice of the most suitable model.

A frequently used method is rr-BLUP - Ridge Regression Best Linear Unbiased Prediction (MEUWISSEN; HAYES; GODDARD, 2001), where the marker effects can be estimated even if there are a higher number of markers than observations. For this it is assumed *a priori* that all the markers explain the same quantity of genetic variance (V_g), that is, the variance per locus will be V_g/n, where n is the number of markers. However, some markers are closer to the QTL and therefore, theoretically explain a high portion of the variance. Thus

this assumption is not realistic and leads to multi-collinearity between the markers from the grouping effect (ISHWARAN; RAO, 2011; MUIR, 2007).

On the other hand, in the Bayesian analysis, it is assumed that each marker explains a variance and therefore can vary among markers, making the model more realistic (EUWISSEN; HAYES; GODDARD, 2001). In this way, the markers that are not in linkage disequilibrium with the QTLs have variance equal to zero, so this method may theoretically improve the prediction power. In consequence, these models can lead to estimates with significant bias (HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

In conclusion, when the traits are influenced by a small number of QTLs or when the QTLs are not distributed uniformly along the genome, the methods that adopt individual variances for each marker are superior to the rr-BLUP method (DAETWYLER et al., 2010; HAYES et al., 2010). Otherwise, there are no significant differences in the prediction power of these methods. Some authors have suggested that when the predictions are made in various selective cycles of a breeding program, the methods that supplies the genomic position of a functional polymorphism may be more beneficial than the methods that distribute the marker effects throughout the genome equally (CLARK; HICKEY; VAN DER WERF, 2011; DAETWYLER et al., 2010; MEUWISSEN; GODDARD, 2010).

The first studies published using GWS were based on simulated data and showed very promising results. Later GWS was tested using real experimental data and only small differences in the prediction power were found between the methods used (HESLOT et al., 2012; PÉREZ-RODRÍGUEZ et al., 2012; RESENDE et al., 2012; RIEDELSHEIMER; TECHNOW; MELCHINGER, 2012; WIMMER et al., 2013). The general conclusion was that this fact reflects an infinitesimal genetic model behind the trait under study, i.e. the trait is controlled for a vast number of genes. This occurs because the

majority of the characteristics of agronomic interest are controlled by a very large number of QTLs, each one of them with little effect in the trait expression, and that are very influenced by the environment (SCHÖN et al., 2004). Therefore, to predict reproductive genetic values, rr-BLUP is a good option because it is more robust and efficient computationally (WIMMER et al., 2013).

Another aspect to consider refers to the marker density. When the marker density is not so high, the probability of finding markers strongly linked to all the QTLs is lower. Thus, the markers close to these QTLs can express a relative sign, inducing a less acute distribution of the effects of the QTL compared to that obtained with a high marker density. Thus models that present as *priori* that all the markers are associated with the expression of the trait are superior when the marker density is not high (MEUWISSEN et al., 2009; MEUWISSEN; GODDARD, 2010).

Information on markers of the individuals have shown to be the superior to the use of kinship information based on pedigree observations (DE LOS CAMPOS et al., 2009; CROSSA et al., 2010; HESLOT et al., 2012). This is easily explained, because pedigree is based on the mean of the progeny, and using markers the real kinship can be calculated. Another aspect considered in the GWS is the sample size of the training population (TRN), which has a direct effect on the predictive accuracy. The accuracy of marker effects estimates increase with the population size, or with the number of genotypes in the TRN population (LIU et al., 2011). This occurs because the bias and variance of the estimates of the marker effects decrease with the sample size (DE LOS CAMPOS et al., 2013). In addition, increase in the training population leads to a higher probability of finding genetic associations between the TRN population and the test population TST, and consequently better accuracy. It is important to mention that the smaller the size of the TRN population the larger the influence of the *priori* on the model, so that with small sample sizes, the marker effects

should be interpreted with caution (LIU et al., 2011). However, in practice, it is often difficult to assess a very large population because of limited resources, as occurs in the tobacco crop, which phenotypic data collection is not an easy task.

To construct the prediction model, the set of genotypes used should be as close as possible to the population where the GWS will be applied (HABIER; FERNANDO; DEKKERS, 2007). It is important that the individuals are assessed phenotypically in similar environments to those where the crop will be grown in the future. In addition, it is essential to obtain good-quality phenotypic data to obtain an accurate prediction otherwise random errors will impact on the GWS (DE LOS CAMPOS et al., 2013)

The GWS, when first proposed, generated numerous perspectives and hopes in the scientific animal and plant breeding community. Therefore, the results reached were not as encouraging as expected (HESLOT; JANNINK; SORRELLS, 2015). There is still the need for many studies to include this tool in the everyday of breeding programs. Each case should be seen independently, so the best strategy and the GWS model can be chosen. In tobacco it was not found any reports of GWS use in the literature.

REFERENCES

AFUBRA. **Associação dos Fumicultores do Brasil.** 2014. Disponível em: <http://www.afubra.com.br>. Acesso em: 6 dez. 2015.

ALEKSOSKI, J. Estimation of the heterotic effect in f1 generation of various tobacco genotypes and their diallel crosses. **Biotechnology & Biotechnological Equipment,** v. 24, n. 2, p. 407-4011, 2010.

ALLARD, R. W.; BRADSHAW, A. D. Implications of Genotype-Environmental Interactions in Applied Plant Breeding. **Crop Science,** v. 4, p. 503-508, 1964.

AKBAŞ, Y.; TAKMA, Ç. Canonical correlation analysis for studying the relationship between egg production traits and body weight, egg weight and age at sexual maturity in layers. **Czechoslov. Journal of Animal Science,** v. 50, p. 163–168, 2005.

AYCOCK, M. K.; MANN, T. J.; MATZINGER D. F. Investigations with a form of cytoplasmic male-sterility in flue-cured tobacco. Tobacco Science, v. 7, p. 130-135, 1963.

BAKER, J. R. **Selection indices in plant breeding**. Boca Raton: CRC, 1986. 218 p.

BASTIAANSEN, J. W. M. et al. Comparison of analyses of the QTLMAS XIII common dataset. I: Genomic selection. **BMC Proceedings,** v. 4(Suppl. 1): S1, p.1-7, 2010.

BELOGRADOVA, K. et al. An Overview on Tobacco Haploids. In: Touraev, A.; Forster, B. P.; Jain, S. M. **Advances in Haploid Production in Higher Plants.** Spring: Netherlands. 2009. p. 75-85.

BEN-ISRAEL, I. et al. Heterotic Trait Locus (HTL) Mapping Identifies Intra-Locus Interactions That Underlie Reproductive Hybrid Vigor in *Sorghum bicolor*. **Plos One** v.7, n. 6, p. e38993, 2012.

BERNARDO, R. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. **Crop Science**, v. 48, p. 1649–1664, 2008.

BERNARDO, R. **Breeding for Quantitative Traits in Plants**. 2. ed.

Woodbury: Stemma Press, 2010. 390 p.

BERNARDO, R. **Essentials of Plant Breeding.** 1. ed. Woodbury: Stemma Press, 2014. 260 p.

BETRAN, F. J. et al. Genetic diversity, specific combining ability, and heterosis in tropical maize under stress and non stress environments. **Crop Science**, v. 43, p. 797–806, 2003.

BIRCHLER, J. A. et al. Heterosis. **The Plant Cell**, v. 22, p. 2105–2112, 2010

BOWMAN, D. T. et al. Contribution of genetics and production technology to long-term yield and quality gains in flue-cured tobacco. **Tobacco Science**, v. 28, p. 30–35, 1984.

BOWMAN, D. T. et al. Stability of flue-cured tobacco cultivars for yield and quality. **Tobacco Science**, v. 30, p. 16–19, 1986.

BTY. **Brazilian Tobacco Yearbook**. Gazeta, 2014.

BRUZI, A. T.; RAMALHO, M. A. P.; FERREIRA, D. F. Homeostasis of common bean populations with different genetic structures. **Crop Breeding and Applied BioTechnology,** v. 7, p. 111 – 116, 2007.

BUTORAC, J. BELJO, J. GUNJACA, J. Study of inheritance of some agronomic and morphological traits in burley tobacco by graphic analysis of diallel cross. **Plant Soil Environment**, v. 50, n. 4, p. 162-167, 2004.

BUTORAC, J. et al. Combining abilities of Agronomic and morphological traits in Burley tobacco. **Agriculturae Conspectus Scientificus**, v. 65, n. 3, p. 153-159, 2000a.

BUTORAC, J. et al. Inheritance of certain economic and agronomic traits in Burley tobacco. **Die Bodenkultur**, v. 51, n. 3, p. 151-156, 2000b.
CASTELLI, F. et al. Adaptation of aromatic burley tobacco varieties to northern Italian environments. **Journal of Agronomy and Crop Science,** v. 172, p. 342–351, 1994.

CERÓN-ROJAS, J. J.; CROSSA, J.; SAHAGUN-DASTELLANOS, J. Statistical Sampling Properties of the Coefficients of three Phenotypic Selection Indices. **Crop Science,** 2015

CHEN, Z. J. Genomic and epigenetic insights into the molecular bases of heterosis. **Nature Reviews: Genetics,** v. 14, p. 471-482, 2013.

CLARK, S.; HICKEY, J.; VAN DER WERF, J. Different models of genetic variation and their effect on genomic evaluation. **Genetics Selection Evolution**, v. 43, p. 18, 2011.

CLARKSON, J. J. et al. Long-term genome diploidization in allopolyploid *Nicotiana* section Repandae (Solanaceae). **New Phytologist,** v. 168, p. 241–252, 2005.

CLAUSEN, R. E.; CAMERON, D. R. Inheritance in Nicotiana tabacum. XVIII. monosomic analysis. **Genetics,** v. 29, p. 447-477, 1944.

CLAYTON, E. E. The genetics and breeding progress in tobacco during the last 50 years. **Agronomy Journal**, v. 50, p. 352–356, 1958.

COOPER, M. et al. Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. **Crop and Pasture Science**, 2014

COSTER, A. et al. Sensitivity of methods for estimating breeding values using genetic markers to the number of QTL and distribution of QTL variance. **Genetics Selection Evolution**, v. 42, n. 9, p. 1-11, 2010.

CROSBIE, T. M. et al. Plant breeding: Past, present, and future. In: LAMKEY, K. R.; LEE M. **Plant Breeding: The Arnel R. Hallauer Int. Symp**., Mexico City. 17–23 Aug. 2003. Oxford: Blackwell, 2003. p. 1–50.

CROSSA, J. et al. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. **Genetics**, v. 186, p. 713–724, 2010.

CRUZ, C. D.; CARNEIRO, P. C. S.; REGAZZI, A. J. **Modelos Biométricos Aplicados ao Melhoramento Genético**. v. 2, 3. ed. Viçosa: UFV, 2014. 668 p.

DAETWYLER, H. D. et al. The impact of genetic architecture on genome-wide evaluation methods. **Genetics,** v. 185, p. 1021 – 1031, 2010.

DEAN, C. E. Heterosis, Inbreeding Depression, and Combining Ability in Diallel Crosses of Cigar-Wrapper Tobacco. **Crop Science,** v. 14, n. 3, p. 482-484, 1974.

DEKKERS, J. C. M.; HOSPITAL, F. The use of molecular genetics in the improvement of agricultural populations. **Nature Reviews Genetics**, v. 3, p. 22–32, 2002.

DE LOS CAMPOS, G. et al. Predicting quantitative traits with regression models for dense molecular markers and pedigree. **Genetics**, v. 182, p. 375–385, 2009.

DE LOS CAMPOS, G. et al. Whole genome regression and pre- diction methods applied to plant and animal breeding. **Genetics**, v. 193, p. 327–345, 2013.

EAST, E. M.; JONES, D. F. Round tip tobacco: A plant "made to order." **Journal of Heredity**, v. 12, p. 50–56, 1921.

FALCONER, D. S.; MACKAY, T. F. C. **Introduction to quantitative genetics**. 4th ed. Malaysia: Pearson, 1996. 464 p.

FAOSTAT. **Food and Agriculture Organization**. 2012. Disponível em: <http://faostat.fao.org/>. Acesso em: 6 dez. 2015.

FEHR, W. R. **Principles of cultivar development**. New York: Macmillan, 1987. 761 p.

GARNER, W. W.; ALLARD H. A.; CLAYTON, E. E. Superior germplasm in tobacco. In: **Yearbook of agriculture**. 1st ed. Washington: USDA, 1936. p. 785–830.

GOFF, S. A.; ZHANG, Q. Heterosis in elite hybrid rice: speculation on the genetic and biochemical mechanisms. **Current Opinion in Plant Biology**, v. 16, n. 2, p. 221-227, 2013.

GRUN, B.; SIMPSON, E. **The Timetables of History: A Horizontal Linkage of People and Events.** 4. ed. Touchstone, 2005. 848 p.

HABIER, D.; FERNANDO, R. L.; DEKKERS, J. M. The impact of genetic relationship information on genome-assisted breeding values. **Genetics**, v. 177, p. 2389–2397, 2007.

HABIER, D. et al. The impact of genetic relationship information on genomic breeding values in German Holstein cattle. **Genetics Selection Evolution**, v. 42, p. 5, 2010.

HALLAUER, A. R.; CANENA, M. J.; MIRANDA FILHO, J. B. **Quantitative genetics in maize breeding**. 3rd ed. New York: Springer, 2010. 680 p.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction.** 2 ed., Stanford: Springer-Verlag, 2009.

HAYES, B. J. et al. Genetic architecture of complex traits and ac- curacy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. **PLoS Genetics,** v. 6, p. e1001139, 2010.

HAZEL, L.N. The genetic basis for constructing selection indexes. **Genetics,** v. 28, p. 476–490, 1943.

HEFFNER, E. L. Plant Breeding with Genomic Selection Gain per Unit Time and Cost. **Crop Science**, v. 50, p. 1681-1690, 2010.

HESLOT, N.; JANNINK, J-L.; SORRELLS, M. E. Perspectives for Genomic Selection Applications and Research in Plants. **Crop Science**, v. 55, p. 1-12, 2015.

HESLOT, N. et al. Genomic selection in plant breeding: a comparison of models. **Crop Science**, v. 52, p. 146–160, 2012.

HIDALGO, A. M. et al. Canonical-correlation analysis applied to selection-index methodology in quails. **Livestock Science,** v.169, p. 35-41, 2014.

IBRAHIM, H.; SLAVIK, B.; AVRATOVSCUKOVA, N. Yield and Yield Components in Flue-Cured Tobacco and Their Genetic Analysis. **Biologia Plantarum**, v. 26, p. 285-292, 1984.

ISHWARAN, H.; RAO, J. S. Generalized ridge regression: geometry and computational solutions when p is larger than n. **Technical report,** Cleveland Clinic and University of Miami, Miami. 2011.

JORDAN, D. R. et al. Prediction of hybrid performance in grain sorghum using RFLP markers. **Theoretical Applied Genetics,** v. 106, p. 559-567, 2003.

KARA S. M.; ESENDAL, E. Heterosis and combining ability analysis of some quantitative characters in Turkish tobacco. **Tobacco Research,** v. 21, p. 16-22, 1995.

KNAPP, S.; CHASE, M. W.; CLARKSON, J. J. Nomenclatural changes and a new sectional classification in Nicotiana (Solanaceae). **Taxon,** v. 53, p. 73-82, 2004.

KORUBIN-ALEKSOSKA, A.; ALEKSOSKI, J. Investigations of Combining Abilities in Diallel Set of Varieties Belonging to Several Tobacco Types. **Agriculturae Conspectus Scientificus,** v. 77, n. 4, p. 203-206, 2012.

KRIEGER, U.; LIPPMAN, Z. B.; ZAMIR, D. The flowering gene *SINGLE FLOWER TRUSS* drives heterosis for yield in tomato. **Nature Genetics,** v. 42, n. 5, p. 459-465, 2010.

KRISHNAMURTHY A. S. et al. Combining ability studies for yield, yield components and total alkaloids in flue-cured tobacco. **Tobacco Research,** v. 20, n. 1, p. 43-46, 1994.

KRYSTKOWIAK, K. et al. Relationship between phenotypic and genetic diversity of parental genotypes and the specific combining ability and heterosis effects in wheat (Triticum aestivum L.). **Euphytica,** v. 165, p. 419-434, 2009.

LANDE, R.; THOMPSON, R. Efficiency of marker-assisted selection in the improvement of quantitative traits. **Genetics**, v. 124, p. 743–756, 1990.

LEGESSE, B. W. et al. Relationship between hybrid performance and AFLP based genetic distance in highland maize inbred lines. **Euphytica**, v. 162, p. 313–323, 2008.

LEGG, P. D.; MATZINGER, D. F.; MANN T. J. Genetic variation and covariation in a Nicotiana tabacum L. synthetic two generations after synthesis. **Crop Science**, v. 5, p. 30–33, 1965.

LERNER, I. M. **The genetic basis of selection**. New York: John Wiley & Sons. 1958.

LEWIS, R. S. Identification of Germplasm of Possible Value for Confronting an Unfavorable Inverse Genetic Correlation in Tobacco. Crop Science, v. 46, p. 1764-1771, 2006.

LEWIS, R. S.; NICHOLSON, J. S. Aspects of the evolution of *Nicotiana tabacum* L. and the status of the United States *Nicotiana* germplasm collection. **Genetic Resources and Crop Evolution**, v. 54, p. 727–740, 2007.

LIM, K. Y. et al. Genome evolution in allotetraploid Nicotiana. **Biological Journal of the Linnean Society,** v. 82, p. 599-606, 2004.

LITTON, C. C.; STOKES, G. W. Outcrossing in burley tobacco. **Tobacco Science**, v. 8, p. 113– 115, 1964.

LIU, Z. et al. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. **Genetics Selection Evolution**, v. 43, n. 19, p. 1-9, 2011.

LU, X. P. et. al. Development of DArT markers for a linkage map of flue-cured tobacco. **Chinese Science Bulletin.** 2012.

LUND, M. S. et al. Comparison of analyses of the QTLMAS XII common dataset. I: Genomic selection. **BMC Proceedings,** v. 3 (Suppl. 1): S1, p. 1-8, 2009.

LYNCH, M.; WALSH, B. **Genetics and Analysis of Quantitative Traits***.* 1. ed. Sunderland, MA: Sinauer Associates, 1998. 980p.

RAMALHO, M. A. P. et al. **Aplicações da Genética Quantitativa no Melhoramento de Plantas Autógamas**. 1. ed. Lavras: UFLA, 2012a. 522p.

MANN, T. J.; JONES, G. L.; MATZlNGER D. F. The use of cytoplasmic male sterility in flue-cured tobacco hybrids. **Crop Science**, v. 2, p. 407-410, 1962.

MASSMAN, J. M. et al. Genome-wide prediction from maize single-cross data. **Theoretical Applied Genetics**, v. 126, p. 13-22, 2013.

MATZINGER, D. F.; MANN, T. J. Genetic studies on associations between nicotine content and yield of flue-cured tobacco. In: **Proc. 3rd World Tob. Sci. Congr.** Salisbury, 18–26 Fev. 1963, p. 357–365. 1964.

MATZINGER, D. F.; MANN T. J.; COCKERHAM, C. C. Recurrent family selection and correlated response in Nicotiana tabacum L. I.‘Dixie Bright 244’ 3 ‘Coker 139.’. **Crop Science**, v. 12, p. 40–43, 1972.

MATZINGER, D. F., MANN T. J., ROBINSON H. F. Genetic variability in flue-cured varieties of Nicotiana tabacum. I. Hicks Broadleaf 3 Coker 139. **Agronomy Jounal**, p. 52, p. 8–11, 1960.

MATZINGER, D. F.; WERNSMAN, E. A. Four cycles of mass selection in a synthetic variety of an autogamous species Nicotiana tabacum L. **Crop Science,** v. 8, p. 239–243, 1968.

MATZINGER, D. F.; WERNSMAN, W. A.; WEEKS W. W. Restricted index selection for total alkaloids and yield in tobacco. **Crop Science**, v. 29, p. 74–77. 1989.

MATZINGER, D. E.; WERNSMAN E. A.; ROSS H.F. Diallel crosses among burley varieties of nicotiana-tabacum L. in F1 and F2 generations. Crop Science, v. 1, p. 275-279, 1971.

MENDES, F. F.; RAMALHO, M. A. P.; ABREU, A. F. B. Índice de seleção para escolha de populações segregantes de feijoeiro-comum. **Pesquisa Agropecuária Brasileira**, Brasília, v. 44, n. 10, p. 1312-1318, 2009.

MEUWISSEN, T.; GODDARD, M. Accurate prediction of genetic values for complex traits by whole-genome resequencing. **Genetics**, v. 185, p. 623–631, 2010.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, p. 1819-1829, 2001.

MEUWISSEN, T. et al. A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value. **Genetics Selection Evolution,** v. 41, n. 2, p. 1-10, 2009.

MOON, H. S. et al. Changes in Genetic Diversity of U.S. Flue-Cured Tobacco Germplasm over Seven Decades of Cultivar Development. **Crop Science**, v. 49, 2009.

MOREAU, L.; CHARCOSSET, A.; GALLAIS. A. Experimental evaluation of several cycles of marker-assisted selection in maize. **Euphytica**, v. 137, p. 111–118, 2004.

MUELLER, L. A. et al. The SOL Genomics Network: a comparative resource for Solanaceae biology and beyond. **Plant Physiology**, v. 138, p. 1310–1317, 2005.

MUIR, W. M. Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic

parameters. **Journal of Animal Breeding and Genetics,** v. 124, p. 342-355, 2007.

MURAD, L. et al. The origin of the paternal genome of tobacco is traced to a particular lineage within Nicotiana tomentosiformis (Solanaceae). **American Journal of Botany,** v. 89, p. 921- 928, 2002.

PARKES, E. Y. et al. Combining ability of cassava genotypes for cassava mosaic disease and cassava bacterial blight, yield and its related components in two ecological zones in Ghana. **Euphytica,** v. 194, n. 1, p. 13-24, 2013.

PÉREZ-RODRÍGUEZ, P. et al. Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. **G3**, v. 2, p. 1595–1605, 2012.

POVILAITIS, B. Variance components in tobacco cultivar trials. **Canadian Journal of Genetics and Cytology**, v. 12, p. 331–339, 1970.

PRIEB, R. I. P. **Pluriatividade na produção familiar fumageira**. Universidade do Texas: Editora EDUNISC, 2005. 195 p.

PSZCZOLA, M. et al. Comparison of analyses of the QTLMAS XIV common dataset. I: Genomic selection. **BMC Proceedings** 5(Suppl. 3): S1. 2011

QI, H. et al. Identification of combining ability loci for five yield- related traits in maize using a set of testcrosses with introgression lines**. Theoretical and Applied Genetics**, v. 126, n. 2, p. 369-377, 2013.

RAMALHO, M. A. P. et al. **Aplicações da Genética Quantitativa no Melhoramento de Plantas Autógamas**. 1. ed. Lavras: UFLA, 2012a. 522p.

RAMALHO, M. A. P. et al. **Genética na agropecuária**. 5. ed. Lavras: UFLA, 2012b. 565 p.

REIS, C. A. F. et al. Selection of eucalyptus progenies by Z index through LSM and Blup. **Pesquisa Agropecuária Brasileira**,v. 46, p. 517-523, 2011.

RESENDE, M. D. V.; SILVA, F. F. ; AZEVEDO, C. F. **Estatística Matemática, Biométrica e Computacional**. 1. ed. Viçosa: UFV, 2014. 881 p.

RESENDE, M. F. R. et al. Accuracy of genomic selection methods in a standard dataset of Loblolly pine (Pinus taeda L.). **Genetics**, v. 190, p. 1503–1510, 2012.

RIEDELSHEIMER, C.; TECHNOW, F.; MELCHINGER, A. Comparison of whole-genome prediction models for traits with contrasting genetic architecture in a diversity panel of maize inbred lines. **BMC Genomics,** v. 13, p. 452, 2012.

ŠARČEVIC, H. et al. Long-term Genetic Improvement and Genetic Diversity of Croatian Flue-cured Tobacco (*Nicotiana tabacum* L.) Cultivars. **Crop Science**, v. 53, 2013.

SCHNABLE, P. S.; WISE, R. P. The molecular basis of cytoplasmic male sterility and ferlility restoration. **Trends in Plant Science**, v. 3, n. 5, p. 175-180, 1998.

SCHUCH, H. **Culturas Gaúchas: Fumo.** Porto Alegre: Assembléia Legislativa do Rio Grande do Sul, 2003.

SCHÖN, C.-C. et al. Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. **Genetics,** v. 167, p. 485– 498, 2004

SHAMEL, A. D.; COBEY, W. W. **Tobacco breeding.** Washington: U.S. Gov. Print. Office, n. 96, 1907.

SHULL, G. H. The composition of a field of maize. **Rep Am Breeders' Assoc,** v. 4, p. 296– 301, 1908.

SMITH, H. F. A discriminant function for plant selection. In: **Papers on Quantitative Genetics and Related Topics**. Department of Genetics, North Carolina State College, Raleigh, North Carolina, 1936. p. 466-476.

SOUZA JR, C. L. Cultivar development of allogamous crops. **Crop Breeding and Applied Biotechnology**. v. 11, n. spe, p. 8-15, 2011.

SPRAGUE, G. F.; FEDERER W. T. A comparison of variance components in corn yields trials: II. Error, year × variety, location × variety and variety components. **Agronomy Journal**, v. 43, p. 535–541, 1951.

STUBER, C. W.; GOODMAN, M. M.; MOLL. R. H. Improvement of yield and ear number resulting from selection at allozyme loci in a maize population. **Crop Science**, v. 22, p. 737–740, 1982.

SU, G. et al. Estimating Additive and Non-Additive Genetic Variances and

Predicting Genetic Merits Using Genome- Wide Dense Single Nucleotide Polymorphism Markers. **Plos One,** v. 7, n. 9, p. e45293, 2012.

TANKSLEY, S.D. et al. RFLP mapping in plant breeding: New tools for an old science. **Nature Biotechnology**, v. 7, p. 257–264, 1989.

TOLLENAAR, M.; LEE, E. A. Yield potential, yield stability and stress tolerance in maize. **Field Crop Research**, v. 75, p. 161– 169, 2002.

VENTURA, H. T. et al. A canonical correlation analysis of the association between carcass and ham traits in pigs used to produce dry-cured ham. **Genetics and Molecular Biology**, v. 34, p. 451–455, 2011.

WERNSMAN, E.; RUFTY, R. Tobacco. In: FEHR, W. **Principles of Cultivar Development** v. 2, New York: Macmillan Publishing Company Inc., 1987.

WILKINSON C. A.; JONES, J. L.; TILSON, W. M. Diallel analysis of crosses among Virginia flue-cured tobacco. **Tobacco Science**, v. 38, p. 21-24, 1994.

WIMMER, V. et al. Genome-Wide Prediction of Traits with Different Genetic Architecture Through Efficient Variable Selection. **Genetics**, v. 195, p. 573–587, 2013.

WRICKE, G.; WEBER, W. E. **Quantitative genetics and selection in plant breeding.** New York: Walter de Gruyter, 1986. 406p.

XIAO, B. G. et al. Analysis on genetic contribution of agronomic traits to total sugar in flue-cured tobacco (Nicotiana tabacum L.). **Field Crops Research,** v. 102, p. 98–103, 2007.

XU, Y.; CROUCH. J. H. Marker-assisted selection in plant breeding: From publications to practice. Crop Science, v. 48, p. 391–407, 2008.

YAN, W.; FRÉGEAU-REID, J. Breeding Line Selection Based on Multiple Traits. **Crop Science,** v. 48, p. 417-423, 2008.

ZHONG, S. et al. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. **Genetics**, v. 182, p. 355–364, 2009
.

SECOND PART - ARTICLES

ARTICLE 1 Combining ability and heterosis of flue-cured Virginia and Burley tobacco in tropical conditions

Combining ability and heterosis of flue-cured Virginia and Burley tobacco in tropical conditions

Bruna Line Carvalho, Carlos Eduardo Pulcinelli, Adriano Teodoro Bruzi, Ramsey Steven Lewis, Magno Antonio Patto Ramalho

ABSTRACT

Tobacco hybrids are already being produced by many tobacco seed companies. However their use have not been well explored. The only strategy breeders have applied to the development of hybrids is to combine phenotypic expression of traits observed in distinct lines into a hybrid, for instance morphological characters and resistances to diseases. Thus, our goal was to study the combining ability of superior tobacco lines; and to trace a strategy for exploring heterosis in a breeding program. Two varietal groups were used in this study, Flue-Cured Virginia (FCV) and Burley (BLY). For the FCV 13 lines were crossed in a diallel design and 72 hybrids were obtained. For BLY 10 lines were crossed in the same scheme and 41 hybrids were obtained. The hybrids, parent lines and checks were evaluated in the field in 10x10 and 8x8 triple lattice design for FCV and BLY, respectively. The traits assessed were yield (YLD), general quality index (GQI), steam by leaf lamina ratio (SLR), total sugar content (SUG), total alkaloids (ALK) and conversion nicotine to nornicotine (CVS), this last just for the BLY. Using the diallel analysis was estimated the general and specific combining abilities (GCA and SCA) and heterosis. It was found that heterosis is higher for FCV than BLY. The mean heterosis is small, but there are some hybrid combinations with high heterosis. The contribution of SCA is larger than GCA for total variation. Therefore the use of hybrids must be encouraged not just for combining phenotypes but also for exploring heterosis.

The program must focus in increasing the performance of lines *per se* and then in testing all possible combinations for finding the best hybrid.

Keywords: Plant breeding; Quantitative Genetics; Diallel crosses; Hybrids; Seed production

INTRODUCTION

Tobacco is considered the crop with the highest economic value in the world among the non-feed species (BELOGRADOVA et al., 2009) and more than 5.9 million tons are produced annually (ITGA, 2015; AFUBRA, 2015). Brazil is the second largest world producer, accounting for approximately 11.5% of the production and tobacco is one of the ten agribusiness products that have the highest export earnings in Brazil (FAO, 2013). It is grown predominantly on family farms, that represent 92% of the total, and this activity is their main source of income (BTY, 2014; PRIEB, 2005; SCHUCH, 2003). Tobacco production, for industrialization purposes, is concentrated predominantly in the southern region of the country, although production for artisanal cigarettes is distributed throughout all Brazilian regions. Several varietal groups are cultivated, but the two main ones are Flue Cured Virginia (FCV), representing 83% of this production, and Burley (BLY) 12% (ITGA, 2015; AFUBRA, 2015).

There are a few tobacco breeding programs in progress in Brazil and in the world. Because it is an autogamous plant, the main objective of the breeding programs in Brazil has been to obtain cultivars consisting of a pure line, although there are also several commercial hybrids on the market. Tobacco flowers are complete and dioecious consisting of a stigma and five anthers. The floral size and morphology characteristics of the species make it fairly easy to collect and store pollen, so that it is feasible to produce line hybrids commercially. In current tobacco seed production, cultivar protection is obtained by using isogenic lines, one of them with cytoplasmic male sterility (AYCOCK; MANN; MATZINGER, 1963; MANN; JONES; MATZINGER, 1962; SCHNABLE; WISE, 1998). Thus the hybrid seed production system is practicable, and the cultivars are naturally protected, because they do not produce seeds in the farmer´s field unless they are manually pollinated. In

consequence of the artificial crossing to produce commercial seeds that has already been adopted, hybrid cultivar production should be encouraged even though heterosis is small. The additional work of the breeder would be to identify the lines that complement each other when crossed at a determined stage of the breeding program and the gain obtained would be incorporated as an additional advantage for the developed cultivar.

To adopt hybrids, the line complementarity should be assessed by estimates of the combining ability. One of the methodologies most used when there are endogamic lines and those with best combination need to be identified, are diallel crosses (HALLAUER; CARENA; MIRANDA FILHO, 2010). This methodology has been adopted in several species to identify superior combinations (PARKES et al., 2013; QI et al. 2013). There are several published papers that report this information under temperate conditions for the different tobacco varietal groups (DEAN, 1974; IBRAHIM; SLAVÍK; AVRATOVSCUKOVÁ, 1984; MATZINGER; MANN; COCKERHAM, 1962; MATZINGER; WERNSMAN; ROSS, 1971). However, it is believed that the results may be different under tropical conditions because tobacco is a species very influenced by the environmental conditions where it is produced, and it is especially sensitive to the photoperiod (HUBER; RUFTY; KERR, 1984; THOMAS et al., 1975; TSO; KASPERBAUER; SOROKIN, 1970), mainly in the qualitative aspects.

As there is little information on diallel crosses in tobacco under tropical conditions, the present study was carried out with the following objectives: estimate the general and specific combining ability of lines and hybrids of the FCV and BLY varietal groups, involving agronomic, physical and chemical traits; estimate the mean and specific heterosis for the crosses; and discuss the possible advantages of using hybrids in tobacco breeding programs under tropical conditions.

MATERIALS AND METHODS

Treatment description and experimental procedure

The data used in the present study were provided by the company Souza Cruz S.A., which is a subsidiary of the British American Tobacco (BAT) group. The lines of the FCV and BLY tobacco varietal groups were developed by the company and selected for this study because they were outstanding for yield and/or good agronomic and qualitative attributes. These were crossed in a diallel design to obtain the hybrid combinations. The descriptions of the experiments and the traits assessed for each varietal group are described below:

FCV varietal group: A cross was made among 13 lines and 72 hybrid combinations were obtained. The lines, hybrids and 15 checks were assessed in a 10 x 10 triple lattice design, in two locations in southern Brazil: 1. The company experimental station, located in Mafra, SC (26˚10'S, 49˚48'W and 848m altitude); 2. The farm of a FCV tobacco producer, situated in the locality of Ribeirãozinho, Mafra, SC (26˚06'S, 49˚56'W and 801m altitude).

The experiments were set up in the ideal cultivation period of the crop in this region, that is, at the beginning of October/2012. The seedlings were produced in a greenhouse in the system of expanded extruded polystyrene trays on a water bed, known as the floating system. The plots consisted of one row with 10 plants, spacing 1.2 m between-row and 0.5 m between plants, without borders. Crop treatments were as recommended in the company technological package. For base fertilization 600kg/ha of 10-16-10 was used and the side dressings applied in three portions with saltpeter; each portion had 133 kg/ha for area 1 and 80 kg/ha for area 2, the difference was due to the soil fertility. Topping of plants was carried out by plot according to the yield potential. The leaves were collected at the physiological maturity point in successive harvests,

depending on ripeness, beginning in January/2013. The harvest plots were placed in bags with adequate ventilation in the center of the standard barn for curing FCV tobacco. The cured leaves were sent to the Experiment Classification Centre of the company to assess the qualitative and quantitative traits. Later sampling was made proportionally to the position classes and quality produced, per plot, for assessment content of chemical compounds.

The following traits of economic interest were assessed for this varietal group: yield (*YLD*) in kg/ha, by weighing the total of leaves of the plot after curing; general quality index (*GQI*) is a classifying index that considers the position of the leaf on the plant, leaf shine intensity, color, maturity, oiliness and body. It is described in the Normative Instruction n˚10 of the Ministry of Agriculture, Fishing and Supply (MAPA/BRASIL, 2007), and is important in the final value of commercialized tobacco. It is given in percentage relative to the standardized ideal for the leaves of each position on the plant and at the end, a weighting is calculated according to the quantity of leaves in each position/class produced; steam by leaf lamina ratio (*SLR*) is calculated by the the percentage in weight of steams divided by the total leaf lamina,; total sugars (*SUG*) and total alkaloid (*ALK*) of cured leaves, measured in the company chemical laboratory by the CORESTA n˚62 method (2015), and given in percentage of mass in relation to the total sample. For this samples were taken per plot, proportional to the classes produced.

BLY varietal group: a cross was made among 10 lines and 41 hybrid combinations were obtained. Only seven of these lines were assessed in the experiments, together with hybrids and 16 checks. The treatments were assessed in a 8 x 8 triple lattice design, in two locations: 1. The company experimental station, located in Mafra, SC (26˚01’S, 49˚43’W and 839m altitude); 2. The farm of a tobacco producer, situated in the locality of Cascavel, in Campo do Tenente, PR (26˚06’S, 49˚56’W and 802m altitude).

The experiments were set up in the ideal cultivation period of the crop in this region, that is, beginning of October 2013. The seedlings were produced in a greenhouse in the floating system. The plot consisted of one line with 10 plants, 1.15 m between-row spacing and 0.45 m between-plants spacing, without borders. The crop treatments were used according to recommendations in the company technological package. For base fertilization 600kg/ha of 10-16-10 was used and side dressing applied in three portions of 135 kg/ha of urea. Topping of plants was carried out by plot according to the yield potential. The plants were harvested when they reached the physiological maturity point, approximately 40 days after topping, in March/2014 and placed in the center of the curing barn (low ceiling shed with a storage yard), with borders on the sides so that there was no interference in the curing process from the external environment. The cured leaves were sent to the Experiment Classification Center of the company to assess the qualitative and quantitative traits, and later a sample was taken proportionally to the position and quality classes produced, per plot, to assess the content of chemical compounds. The traits assessed were similar to those assessed for the FCV group, i.e. *YLD, GQI, SLR, SUG* and *ALK,* and a further trait, nicotine to nornicotine conversion (*CVS*), that is given by: *% nornicotine / (% nicotine + % nornicotine).*

Statistical analyses

First, the phenotypic data were submitted to analysis of variance by location and across locations. The coefficient of variation (*CV*), the accuracy of the model (*Ac*) and the coefficient of determination (R^2) were estimated in order to verify the experimental precision. The statistical model adopted in the joint analysis was:

$$Y_{ijkl} = m + b_{l/kj} + r_{k/j} + l_j + t_i + tl_{ij} + \bar{e}_{ijkl}$$

where Y_{ijkl} is the phenotypic observation of the plot of treatment *I*, location *j*, replication *k*, block *l*; *m* is the constant inherent to all the observations, that in this case refers to the general mean; $b_{l/kj}$ is the effect of block *l*, within replication *k* and location *j*; $r_{k/j}$ is the effect of replication *k*, within location *j*; l_j is the effect of location *j*; t_i is the effect of treatment *I*; tl_{ij} is the effect of the interaction between the treatment *I* and location *j*; and $\bar{e}_{ijkl}$ is the mean error associated to the plot estimates, with $e \cap N(0, \sigma_e^2)$. The source of treatment variation (t_i) was partitioned into the effects of hybrids (H), parents (P) and checks I, and the respective interactions with locations. The contrast treatments of the diallel *vs* checks and hybrid *vs* parents were also estimated.

The heterosis was calculated using the adjusted means of the treatments involved in the diallel. The mean heterosis is obtained by $\bar{h}(\%) = (\bar{H} - \bar{P})/\bar{P}\ x\ 100$. And specific heterosis, that is, of each hybrid combination, by: $h_{ij} = (H_{ij} - (P_i + P_j)/2)/((P_i + P_j)/2)\ x\ 100$. The parental *vs* hybrids contrast showed whether the mean heterosis was different from zero. The same procedure was adopted to test the nil hypothesis of the specific heterosis.

Diallel analysis was made of the hybrids and parents adopting fix model method II of Griffing (1956), by location and across location. The model used in the analysis per location was:

$$y_{ij} = m + g_i + g_j + s_{ij} + \bar{e}_{ij.}$$

where y_{ij} is the mean value of the diallel cross between the parents *i* and *j*; m is a constant, that in this case represents the general means of the treatments of the diallel; g_i and g_j are the effects of the general combining ability of the *i*-eth and *j*-eth parent, respectively; s_{ij} is the effect of the specific combining

ability for the cross between parents i and j; and $\bar{e}_{ij.}$ is the mean experimental error, where $e \cap N(0.\sigma_e^2)$.

The effects of the general and specific combining abilities (*GCA* and *SCA* respectively) were estimated by the least-squares method. Thus the solutions were obtained based on the equation $X'X\hat{\beta} = X'Y$.

RESULTS

To facilitate the presentation, the results of the FCV and BLY varietal groups are presented separately.

FCV varietal group

Information regarding the precision and homoscedasticity of the experiments are presented in table 1. The *CV* values were below 15% for most of the traits, except for some characteristics of the chemical composition (Table 1) that was expected because of the nature of those traits and so any sampling variation can be detected. The accuracy assessed the phenotype as indicator of the genotypes. For some traits it was not very high, for example, *YLD* (Table 1), however, it is emphasized that accuracy is associated to experimental precision when there is large variation among treatments, because it is obtained by the F test. Thus, very often, low accuracy does not mean low precision, and so, low accuracy does not signify low precision, but rather low variation among the treatments assessed. The R^2 estimate computes the proportion of the sum of squares (*SS*) of the model in relation to the total *SS*. Evidently, the closer to the unit, the smaller the contribution of the SS of the error to total SS, that is, the experiment is more precise. The values were similar to those of accuracy indicating intermediate precision for most of traits assessed.

Table 1 Estimate of the precision and homoscedasticity parameters for each environment of the traits assessed in the FCV tobacco/varietal group of tobacco.

Location	Parameter	YLD	GQI	SLR	SUG	ALK
1	CV (%)	12.23	12.36	4.03	20.32	8.76
	Ac	0.49	0.58	0.84	0.58	0.94
	R^2	0.68	0.62	0.72	0.71	0.84
2	CV (%)	14.85	16.75	5.21	22.51	9.62
	Ac	0.46	0.66	0.86	0.60	0.96
	R^2	0.53	0.62	0.74	0.60	0.89
MSE_{MAX}/MSE_{MIN}[1]		1.35	1.27	1.65	1.32	1.36

[1] MSE: Mean square of error - maximum (MAX) and minimum (MIN)

Joint analysis requires the assumption that the error variance of the experiments is similar in magnitude. The ratio between the two mean squares of error (MSE) for all the cases was very low, lower than two, (Table 1), indicating homogeneity of variances. The summary of the joint analysis of variance of the different traits assessed is shown in Table 2. The source of variation (*SV*) location was significant for all traits ($P<0.01$), except *SLR*. This difference can be understood by the means of the traits in each environment. For *YLD*, the mean of location 2 was 12.4% superior. For *GQI* the inverse occurred. The treatment SV was significant ($P<0.01$) for all traits, indicating that there is variation among them. In the partitioning of the *SV* treatments in hybrids, parents and checks, it was observed that the difference was also significant ($P\leq0.01$), except between checks for *YLD*. Most of the *SV* involving locations x treatment interaction, and their partitioning, was not significant. In the case of parents, the line yields ranged from 3060.2 kg/ha, line 5, to 4162.50 kg/ha, line 10 (Table 3). The variation was less for *GQI*, line 8 presented the lowest estimate 39.70%, and line 6 the highest, 55.08%. The best performing line for the *SLR* trait, the lowest value, was line 5. The same line presented the lowest yield and highest *SUG* content.

Table 2 Summary of the joint analysis of variance of two locations, for the traits *YLD* (kg/ha), *GQI* (%), *SLR* (%), *SUG* (%) and *ALK (%)* of the FCV varietal group.

		P value				
S.V.	**DF**	**YLD**	**GQI**	**SLR**	**SUG**	**ALK**
Locations – L	1	<0.01	<0.01	0.17	<0.01	<0.01
Replication /L	4	<0.01	<0.01	<0.01	<0.01	<0.01
Block /Rep	54	<0.01	0.03	0.24	<0.01	0.33
Treatment (T)	99	<0.01	<0.01	<0.01	<0.01	<0.01
Hybrids (H)	71	0.01	<0.01	<0.01	<0.01	<0.01
Parents (P)	12	<0.01	<0.01	<0.01	<0.01	<0.01
Checks (C)	14	0.11	<0.01	<0.01	0.01	<0.01
(H+P) vs C	1	<0.01	0.01	0.31	<0.01	<0.01
H vs P	1	0.01	0.53	0.08	0.02	<0.01
Diallel (H+P)	84	<0.01	<0.01	<0.01	<0.01	<0.01
GCA	12	<0.01	<0.01	<0.01	<0.01	<0.01
SCA	78	0.27	0.01	<0.01	0.02	<0.01
T x L	99	0.28	0.11	<0.01	0.94	<0.01
H x L	71	0.19	0.16	<0.01	0.97	<0.01
P x L	12	0.22	0.62	0.05	0.41	0.25
C x L	14	0.62	0.09	0.13	0.83	<0.01
Diallel x L	84	0.51	0.19	<0.01	0.87	<0.01
GCA x L	12	0.28	0.03	0.24	0.63	<0.01
SCA x L	78	0.78	0.53	<0.01	0.93	<0.01
L1[1] mean		3490	51.68	28.03	9.11	2.84
L2[2] mean		3922	45.95	27.89	7.97	3.18
Overall mean		3706	48.82	27.96	8.54	3.01
CV[3] (%)		13.79	14.49	4.66	21.35	9.26

[1,2] Location 1 and 2; [3] CV: Coefficient of variation

The contrast of treatments included in the diallel (hybrids and parents) *vs* checks was significant (P≤0.01), except for the *SLR* trait, indicating that the general mean of the diallel differed from the mean of the checks. A very important contrast is that which assesses the means of the hybrids in relation to

the mean of the parents, that is, estimates whether the mean heterosis was different from zero. This contrast was only not significant for *GQI*, which was due probably to the fact that the temperature curve in the chamber was the same for all treatments, as is not possible to cure the treatments individually to exploit the potential of each one. The means presented in Table 3 confirm this observation and show that the mean heterosis was different from zero, although small, for all traits assessed. It should be emphasized that this heterosis value refers to the mean of all hybrid combinations and because of this, there may be combinations with higher or lower heterosis.

Table 3 Mean of the parents, checks, hybrids and mean heterosis of the FCV varietal group, for the traits *YLD* (kg/ha), *GQI* (%), *SLR* (%), *SUG* (%) and *ALK* (*%),* average of two locations.

Parents	**YLD**	**GQI**	**SLR**	**SUG**	**ALK**
1	3297.50	43.40	26.82	9.13	3.27
2	4114.50	45.93	28.20	8.12	2.70
3	3581.17	47.78	27.62	8.55	2.89
4	3517.83	42.40	28.85	6.43	2.65
5	3060.17	52.70	24.51	11.30	3.02
6	3461.83	55.08	28.76	9.42	2.74
7	3752.83	45.25	29.59	6.73	2.75
8	3812.00	39.70	30.40	7.85	3.11
9	3836.33	51.12	28.84	7.55	4.01
10	4162.50	45.98	28.95	8.27	3.53
11	4025.67	40.98	30.26	6.10	3.75
12	3539.67	49.17	28.17	8.95	2.78
13	3067.67	53.37	26.35	7.73	3.47
Check mean	3398.64	50.59	27.83	7.98	3.39
Diallel mean	3761.06	48.45	27.99	8.62	2.94
Parents mean	3597.57	47.95	28.26	8.14	3.11
Hybrids mean	3790.58	48.54	27.95	8.71	2.91
Mean heterosis (%)	**5.37**	**1.24**	**-1.10**	**7.01**	**-6.72**

The heterosis of each hybrid combination was calculated. Figure 1 shows the frequency distribution of these values, which varied significantly according to the combination for all the traits and presents both negative and positive values. However, only 22% of the hybrids presented heterosis significantly above zero. For YLD, for example, some hybrids means were 30% lower than the mean of the parents. For more details, Table 4 shows the 10 most productive hybrids. The heterosis of the most productive hybrids ranged from 3%, hybrid *6x3*, to 15%, hybrids *8x1* and *5x2*, for this trait. The case of *SLR* and *ALK* should be highlighted, where the mean heterosis was negative. In this case, dominance was expressed in the sense of reducing expression of the trait, which is desirable.

Table 4 Mean of the 10 highest yielding hybrids (YLD in kg/ha), and respective specific heterosis estimates (%), $\hat{g}_i$, $\hat{g}_j$ and $\hat{s}_{ij}$. Data obtained in the diallel of the FCV varietal groups, mean of two locations.

Hybrid	**YLD**	**Heterosis**	$\hat{g}_i$	$\hat{g}_j$	$\hat{s}_{ij}$	**%GCA[1]**	**%SCA[2]**
3x2	4,346	9	123.5	176.9	271.6	52.5	47.5
11x7	4,224	12	-15.2	4.9	460.1	-2.3	102.3
8x1	4,194	15	91.3	-17.8	346.0	17.5	82.5
9x2	4,163	9	-36.6	176.9	248.3	36.1	63.9
8x3	4,134	8	91.3	123.5	145.5	59.6	40.4
5x2	4,130	15	-136.2	176.9	315.6	11.4	88.6
6x3	4,116	3	64.4	123.5	153.8	55.0	45.0
10x7	4,105	13	36.7	4.9	289.0	12.6	87.4
1x2	4,097	8	-17.8	176.9	164.1	49.2	50.8
3x12	4,057	14	123.5	17.5	142.4	49.8	50.2
Mean contribution of 10 most yielding hybrids						**34.1**	**65.9**

[1] *%GCA* = $(\hat{g}_i + \hat{g}_j) / (\hat{g}_i + \hat{g}_j + \hat{s}_{ij}) \times 100$; [2] *%SCA* = $\hat{s}_{ij} / (\hat{g}_i + \hat{g}_j + \hat{s}_{ij}) \times 100$

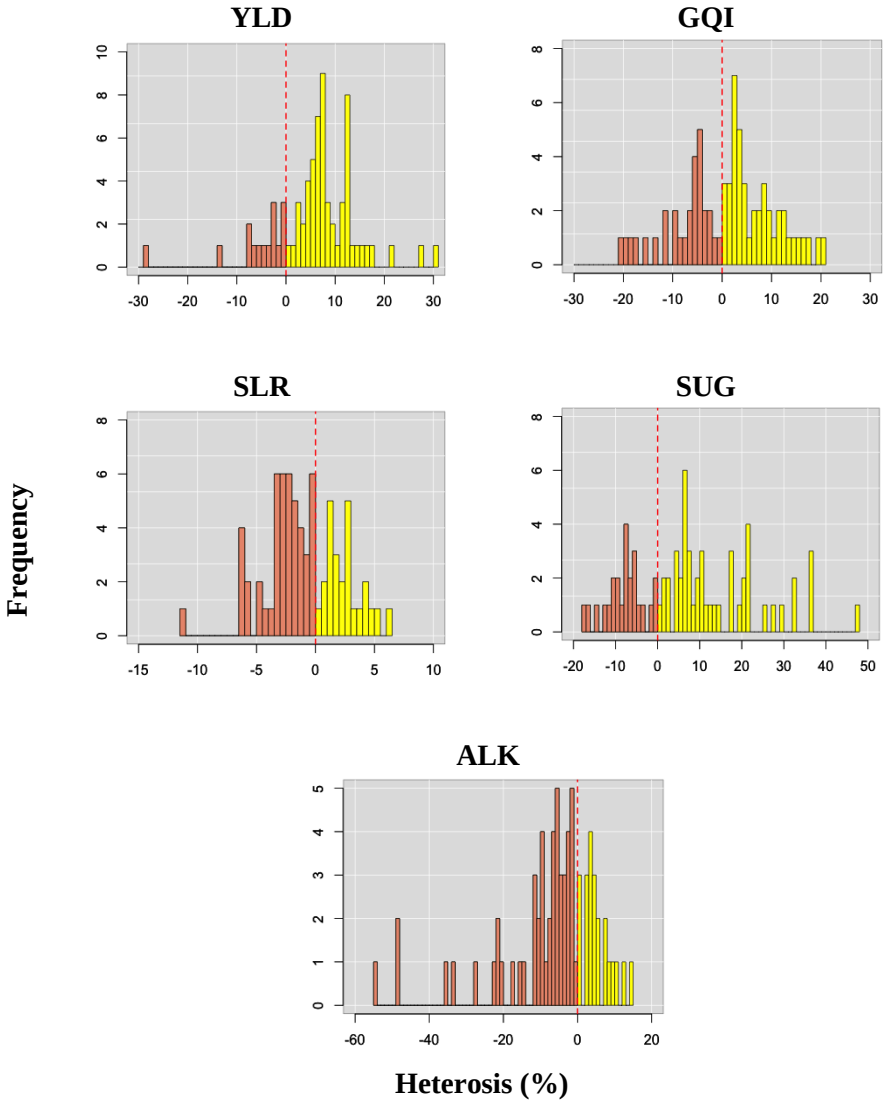

Figure 1 Frequency distribution of the heterosis estimates (%) of the hybrids involved in the diallel in relation to the mean of the parents for the traits YLD, GQI, SLR, SUG and ALK assessed in the FCV group. Data based on the mean of two locations.

In the analysis of the diallel the significance of the *GCA* effect (P<0.01) showed that the lines differed for the general combining ability ($\hat{g}_i$) for all traits (Table 2). Regarding the *SCA* effect, only the *YLD* trait did not present significance, highlighting the importance of the effect of dominance (*d*) in the genetic control of the physical and chemical quality traits. Further evidence of the importance of dominance in FCV tobacco is the proportion explained by the *SS* of the *SCA* compared to the *SS* of the diallel model. Table 5 shows that in the mean of all the traits, the *SCA* and *GCA* contribution was relatively similar.

Table 5 Relative contribution of *SCA* to the *SS* of the model of the traits *YLD* (kg/ha), *GQI* (%), *SLR* (%), *SUG* (%), *ALK* (*%)* and *CVS* (%) of the FCV varietal group

	Parameter	**YLD**	**GQI**	**SLR**	**SUG**	**ALK**
FCV	SS GCA	11442553	4892	456.2	329.8	78.4
	SS SCA	26149400	6116	209.5	417.3	46.4
	SCA contribution (%)	**69.56**	**55.56**	**31.47**	**55.86**	**37.17**

As most of the *GCA x L* and *SCA x L* interactions were not significant (Table 2) emphasis will be directed to the estimates of the genetic parameters of the model in the mean of the two locations. The effects of $\hat{g}_i$ and the amplitude of the effects of $\hat{s}_{ij}$ are shown in Table 6 and Figure 2, respectively. There was variation for all the traits in the $\hat{g}_i$ estimates of the lines and $\hat{s}_{ij}$ estimates of the hybrid combinations. The $\hat{g}_i$ estimate varied between the lines and among the traits as already mentioned. When considering yield, the two lines with largest $\hat{g}_i$ were *2* and *3;* for *GQI*, lines *1* and *5*; and for *SLR*, lines 4 and 5 were outstanding, because decrease in the trait is desirable.

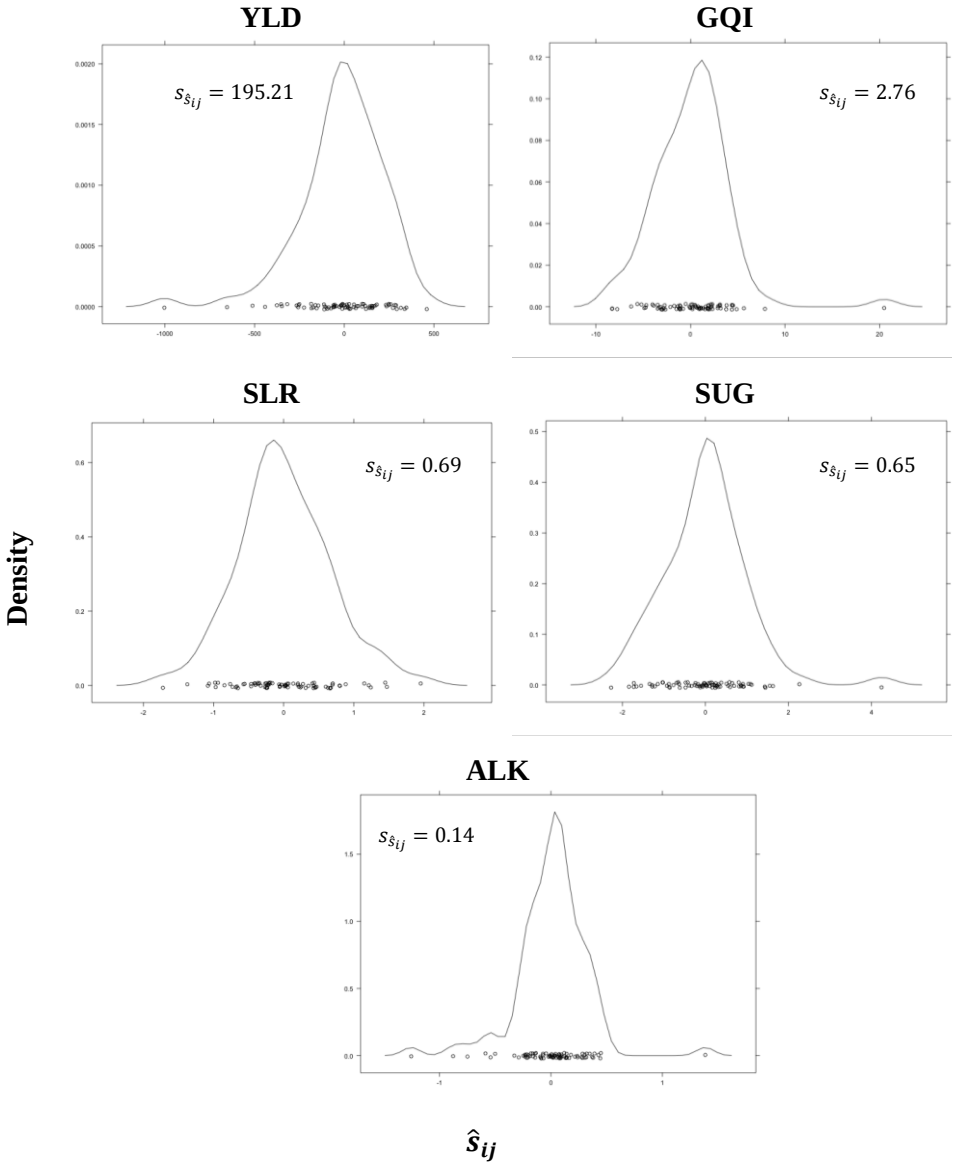

Figure 2 Frequency and probability distribution of the $\hat{s}_{ij}$ estimates of hybrids and lines involved in the diallel of traits *YLD* (kg/ha), *GQI* (%), *SLR* (%), *SUG* (%) and *ALK* (*%)* of the FCV varietal group. Data based on the mean of two locations.

Table 6 Estimates of $\hat{g}_i$ of the parents involved in the diallel of the FCV varietal group for the traits *YLD* (kg/ha), *GQI* (%), *SLR* (%), *SUG* (%) and *ALK (%)*, and the respective areas associated to the estimate of each parent and the between parent comparison ($s_{\hat{g}_i}$ and $s_{\hat{g}_i-\hat{g}_j}$). Data presented on the mean of two locations.

Parent	YLD	GQI	SLR	SUG	ALK
1	-17.97	6.94	-0.57	1.74	-0.21
2	163.01	-3.18	0.22	-0.18	-0.24
3	127.41	0.29	0.76	-0.23	0.40
4	-237.04	1.32	-1.22	0.00	0.21
5	-136.30	2.67	-1.25	0.98	0.07
6	68.49	1.19	0.47	-0.32	0.17
7	8.84	-1.46	0.57	-0.34	-0.19
8	95.25	-1.25	0.61	-0.24	0.22
9	-32.71	-0.32	0.11	0.15	0.04
10	39.62	-3.86	0.48	-0.93	-0.77
11	-30.55	-1.73	0.56	-0.85	0.20
12	18.76	-0.99	-0.45	0.34	0.15
13	-66.80	0.39	-0.28	-0.13	-0.05
$s_{\hat{g}_i}$	61.13	1.02	0.15	0.18	0.06
$s_{\hat{g}_i-\hat{g}_j}$	89.98	1.51	0.21	0.26	0.09

The frequency distributions of the $\hat{s}_{ij}$ estimates for the various traits show wide variation in the estimate between the hybrids assessed (Figure 2). For *YLD* the $\hat{s}_{ij}$ estimates ranged from -1.003 kg/ha (s_{2x11}) to 460 kg/ha (s_{7x11}). For *GQI*, the lower limit was -8.30% (s_{1x6}), and the upper limit was 20.48% (s_{1x1}). It is important to know the contribution of $\hat{g}_i$ and $\hat{s}_{ij}$ to the best performing hybrids. For example, Table 4 shows the ten best performing hybrids for *YLD*. For the highest yielding combination (*3x2*), $\hat{g}_i$ and $\hat{s}_{ij}$ contributed similarly to the hybrid performance. In the case of the second most productive hybrid (*11x7*), the good performance was a function principally of the $\hat{s}_{ij}$. In the mean of the

ten combinations, *SCA* contributed 65.9% to the hybrid performance and *GCA*, 34.1%. Therefore the superiority of the hybrid depended on the $\hat{g}_i$ of each parent and mainly, on the complementarity between them that is, the specific combining ability ($\hat{s}_{ij}$).

BLY varietal group

The results regarding the BLY varietal group were very similar to the FCV varietal group. The accuracy of the experiments was fairly high, with some exceptions, for the same reason mentioned for the FCV varietal group, that is, the lines assessed are part of best of the program, where the variation is low, reflected in the estimates of this parameter (Table 7). The *CV* was below 15% for all the traits in the two locations and, consequently, across locations except for *CVS*. However, the accuracy was high (*Ac* >0.80) for this trait, indicating that significant differences could be detected among the treatments, although the MSE deviation from the mean was high. Similarly to the FCV varietal group, the MSE of the two experiments met the assumption of homoscedasticity for all the traits.

Table 7 Estimate of the parameters of precision and homoscedasticity of each environment of the traits assessed in the BLY varietal group in tobacco.

Location	Parameter	YLD	GQI	SLR	SUG	ALK	CVS
1	CV (%)	9.29	7.88	3.28	7.57	7.36	49.41
	Ac	0.54	0.45	0.86	0.45	0.48	0.96
	R^2	0.72	0.63	0.80	0.67	0.71	0.90
2	CV (%)	11.38	9.51	2.65	6.09	8.94	42.16
	Ac	0.67	0.82	0.93	0.60	0.74	0.80
	R^2	0.81	0.75	0.82	0.57	0.75	0.92
MSE_{MAX}/MSE_{MIN}[1]		1.05	1.22	2.03	1.64	1.75	1.31

[1] MSE: Mean square of error - maximum (MAX) and minimum (MIN)

The summary of the joint analysis of variance of BLY is showed in table 8. The locations presented significant differences (P<0.01) for all the traits, except for *CVS*. This difference can also be observed in the table of trait means for each environment (Table 8). For *YLD* and *GQI*, for example, the mean of location 1 was approximately 15% above the mean of location 2. The treatments also differed for all the traits, indicating that there was variation among them. The partitioning of the treatments effect in hybrids, parents and checks showed significant difference (P<0.1) for all the three sources of variation. Considering only the lines, the amplitude of variation among the parents was 25% in relation to the mean for *YLD* (Table 9). The smallest variation occurred for *SUG* (11.4%) and the biggest for *CVS* (376%). The highest yielding parent, line *1*, also had the highest *GQI* and lowest *CVS*, and was a strong candidate for selection. The effect of the location x treatment interaction and the partitioning of this interaction for treatments did not coincide among the traits but it was not significant in most cases.

The mean of the diallel treatments, as in the FCV varietal group, was different from the mean of the checks for most of the traits (P<0.05), except for *SUG* (Table 8). This fact indicates the potential of these lines in hybrids for the traits assessed. The contrast between the parents and hybrids shows that on average, the hybrids presented the same performance as the parents, shown by the means presented in Table 9. Consequently the mean heterosis for these traits was nil, except for *CVS*, which presented heterosis, but of low magnitude. Nevertheless, it should be emphasized that this value refers to the mean of the combinations, i.e., there may be combinations with larger or smaller heterosis.

Table 8 Summary of the joint analysis of variance based on two locations, for the traits *YLD* (kg/ha), *GQI* (%), *SLR* (%), *SUG* (%), *ALK* (%) and *CVS* (%) of the BLY varietal groups.

		P value					
S.V.	**DF**	**YLD**	**GQI**	**SLR**	**SUG**	**ALK**	**CVS**
Locations – L	1	<0.01	<0.01	<0.01	<0.01	<0.01	0.80
Replication /L	4	<0.01	<0.01	<0.01	0.15	0.01	0.06
Block /Rep	42	0.01	0.04	<0.01	<0.01	<0.01	0.25
Treatment (T)	63	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
Hybrids (H)	40	<0.01	<0.01	<0.01	0.09	<0.01	<0.01
Parents (P)	6	<0.01	0.01	<0.01	0.01	0.01	<0.01
Checks (C)	15	<0.01	<0.01	<0.01	<0.01	<0.01	0.90
Diallel (H+P) vs C	1	<0.01	<0.01	0.04	0.33	<0.01	<0.01
H vs P	1	0.82	0.55	0.14	0.58	0.95	0.26
Diallel	47	<0.01	<0.01	<0.01	0.08	<0.01	<0.01
GCA	9	<0.01	<0.01	<0.01	0.49	<0.01	<0.01
SCA	38	0.13	0.09	<0.01	0.14	0.45	<0.01
T x L	63	0.10	<0.01	0.02	0.18	<0.01	0.35
H x L	40	0.11	0.04	0.01	0.09	0.15	0.44
P x L	6	0.67	<0.01	0.17	0.20	<0.01	0.01
C x L	15	0.07	0.01	0.33	0.62	0.08	1.00
Diallel x L	47	0.31	<0.01	0.04	0.11	0.06	0.26
GCA x L	9	0.04	<0.01	0.02	0.43	<0.01	0.99
SCA x L	38	0.75	0.08	0.40	0.16	0.60	0.31
L1[1] mean		2318	80.76	29.61	1.34	1.34	7.98
L2[2] mean		1995	70.26	30.14	1.28	1.28	7.88
Overall mean		2156	75.51	29.87	1.31	1.31	7.93
CV[3] (%)		0.81	8.64	2.98	6.91	8.31	45.97

[1,2] Location 1 and 2; [3] CV: Coefficient of variation

Table 9 Mean of the parents, checks, hybrids and mean heterosis in the BLY varietal group, for the traits *YLD* (kg/ha), *GQI* (%), *SLR* (%), *SUG* (%), *ALK* (%) and *CVS* (%), on the mean of two locations.

Parents	YLD	GQI	SLR	SUG	ALK	CVS
1	2418.67	80.28	29.05	1.27	4.99	2.88
2	2327.67	79.50	29.63	1.37	4.93	38.53
3	2247.83	75.97	29.27	1.30	5.31	4.53
4	2183.00	76.63	29.42	1.27	5.21	3.63
5	1853.00	78.73	30.30	1.27	4.71	12.00
6	2268.83	72.75	30.73	1.28	4.49	3.23
7	2261.17	67.25	32.47	1.42	4.90	3.21
Check mean	1920.35	71.53	29.70	1.30	4.53	5.23
Diallel mean	2235.49	76.69	29.93	1.31	4.97	8.85
Parents mean	2243.09	76.08	30.13	1.30	4.97	9.48
Hybrids mean	2234.19	76.79	29.90	1.31	4.97	8.74
Mean heterosis (%)	-0.40	0.93	-0.79	0.70	-0.10	-7.80

As in the FCV varietal group, the frequency distribution was plotted of the heterosis values for the hybrid combinations (Figure 3). Since parents *8*, *9* and *10* were not assessed in the experiment, the specific heterosis that involved them could not be calculated. There was variation for heterosis among the hybrid combinations and as in the FCV varietal group, positive and negative values were observed for all the traits. However, the magnitude of the heterosis values was low compared to the FCV, except for *CVS*, and 37% of the values were significantly different from zero for *CVS* and 17% for the other traits. Even so, a combination was observed with mean yield 15% above the mean of the parents. For SLR, the maximum heterosis value was 5%. Table 10 shows the ten highest yielding hybrids. The heterosis of these hybrids ranged from 5% to 16% and the mean of the most productive (*2x4*) was 13% superior to the mean of the parents.

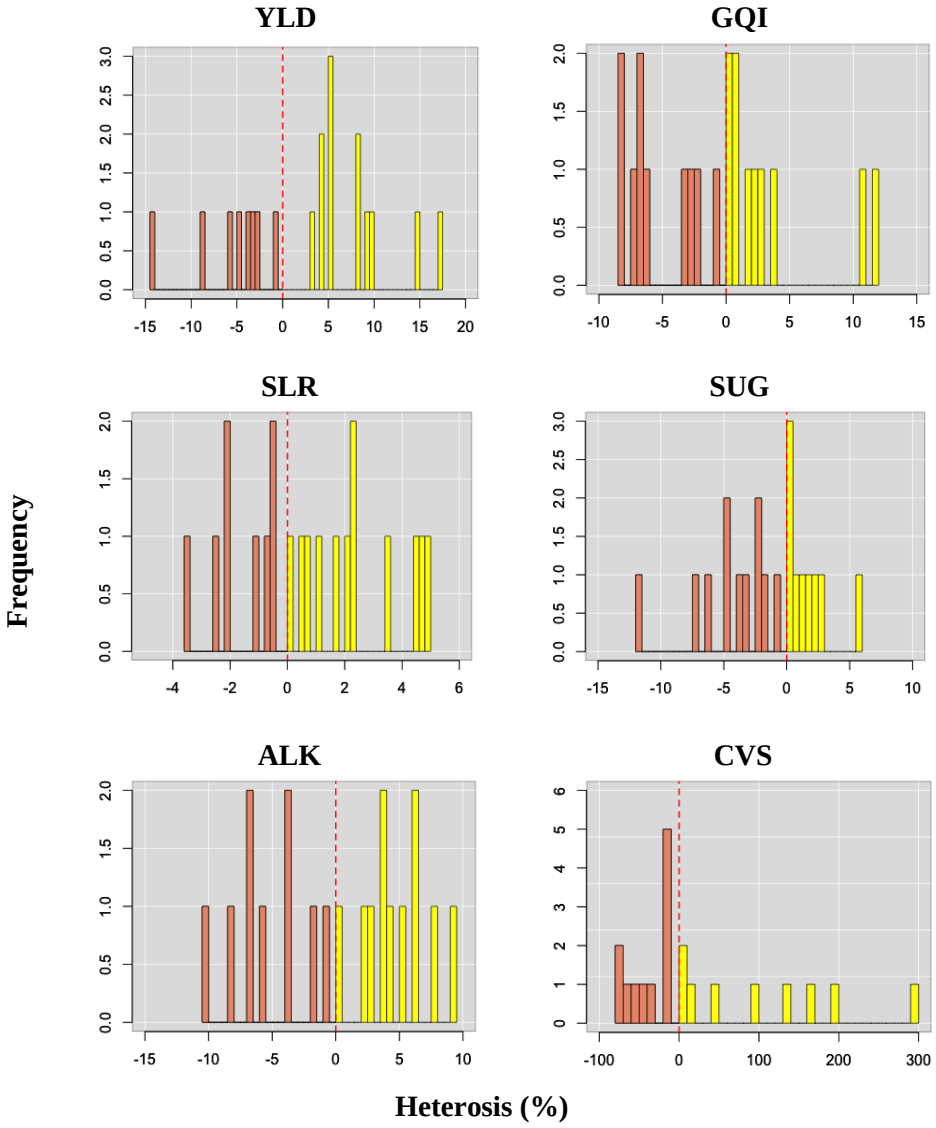

Figure 3 Frequency distribution of the heterosis estimates (%) of hybrids involved in the diallel in relation to the mean of the parents for the traits for the traits YLD, GQI, SLR, SUG and ALK assessed in the BLY varietal group. Data is based on the mean of two locations.

Table 10 Mean of the 10 highest yielding hybrids (kg/ha), and respective $\hat{g}_i$, $\hat{g}_j$ and $\hat{s}_{ij}$. Estimates. Data obtained on the diallel of the BLY varietal groups, mean of two locations.

Hybrid	YLD	Heterosis	$\hat{g}_i$	$\hat{g}_j$	$\hat{s}_{ij}$	%GCA[1]	%SCA[2]
2x4	2,538	13	117.3	2.1	183.7	39.4	60.6
2x9	2,460	-	117.3	43.9	64.7	71.3	28.7
2x6	2,432	6	117.3	-1.4	81.7	58.7	41.3
2x3	2,410	5	117.3	-42.8	101.1	42.5	57.5
6x9	2,392	-	-1.4	43.9	114.8	27.0	73.0
1x9	2,387	-	105.8	43.9	2.7	98.2	1.8
5x7	2,380	16	-90.1	51.0	184.6	-26.9	126.9
1x8	2,379	-	105.8	-76.2	114.8	20.5	79.5
7x9	2,372	-	51.0	43.9	42.9	68.9	31.1
1x10	2,351	-	105.8	-109.7	120.8	-3.3	103.3
Mean contribution of 10 most yielding hybrids						**39.6**	**60.4**

[1] *%GCA = $(\hat{g}_i + \hat{g}_j) / (\hat{g}_i + \hat{g}_j + \hat{s}_{ij})$ x 100*; [2] *%SCA = $\hat{s}_{ij} / (\hat{g}_i + \hat{g}_j + \hat{s}_{ij})$ x 100*

Regarding the diallel (Table 8) the significance of the *GCA* effect (P<0.01) shows that the lines had differences in the $\hat{g}_i$, except for the *SUG* trait. For the *SCA*, only traits *GQI*, *SLR* and *CVS* were significant (p<0.1), indicating the presence of dominance in the control of this trait. Table 11 shows the proportion of the *SS* of the *SCA* compared to the SS of the diallel. This proportion is relatively high for most of the traits, and some cases even larger than the SS of the *GCA*.

Table 11 Relative *SCA* contribution to SS in the model of the traits *YLD* (kg/ha), *GQI* (%), *SLR* (%), *SUG* (%), *ALK* (*%)* and *CVS* (%) of the BLY varietal group.

	Parameter	YLD	GQI	SLR	SUG	ALK	CVS
BLY	SS GCA	2562808	2612	277.6	0.09	13.71	21403
	SS SCA	3206164	2332	127.1	0.63	9.26	3148
	SCA contribuition (%)	**55.58**	**47.17**	**31.40**	**87.14**	**40.32**	**12.82**

The *GCA x L* interaction was significant for *YLD*, *GQI*, *SLR* and *ALK* (P<0.01; Table 8). The *SCA x L* was not significant for all the traits. The effects

of $\hat{g}_i$ and the amplitude of the $\hat{s}_{ij}$ effects on the mean of the locations are shown in Table 12 and Figure 4, respectively. There was variation among the lines for $\hat{g}_i$ and among hybrids for $\hat{s}_{ij}$ for all the traits, even though some hybrids presented non-significant *SCA* effect. The parents with the most significant $\hat{g}_i$ value for yield were lines *1* and *2*, coinciding with the highest yielding lines. For *GQI*, the lines with largest $\hat{g}_i$ were *1* and *7*, where only line *1* presented the highest mean for this parameter. Figure 4 shows the wide variation in the estimate between the hybrids assessed. For yield the $\hat{s}_{ij}$ estimate ranged from -237.08 kg/ha (s_{1x5}) to 395.88 kg/ha (s_{1x1}). For *GQI*, the lower limit was from -6.40% (s_{1x2}) to 13.17% (s_{1x1}).

As already mentioned, knowing the contribution of $\hat{g}_i$ and $\hat{s}_{ij}$ of the best performing hybrids is important to verify whether the performance of these hybrids is due to $\hat{g}_i$ or $\hat{s}_{ij}$. Taking into account *YLD* as an example, for hybrid *2x4* with highest mean, *GCA* contributed approximately 40% and *SCA* 60% (Table 10). However, for the second most yielding hybrid, the *GCA* contribution was larger, 70%, and that of *SCA* 30%. In the mean of the 10 best performing hybrids, the *GCA* contributed with 40% and *SCA* with 60%. Therefore, as in the BLY varietal group, both the *GCA* and the *SCA* contributed to the hybrid performance.

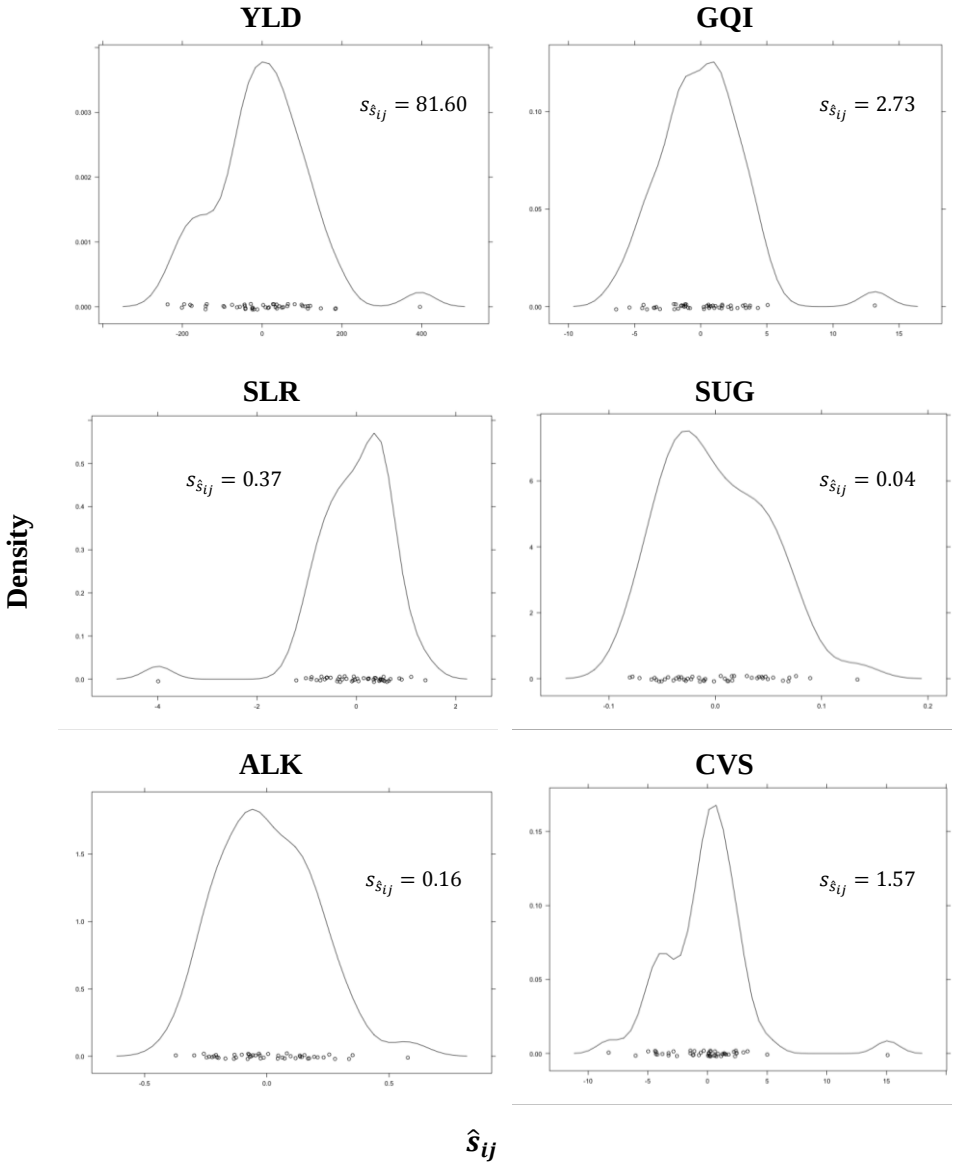

Figure 4 Frequency and probability distribution of the $\hat{s}_{ij}$ estimates of the hybrids and lines involved in the diallel for the traits *YLD* (kg/ha), *GQI* (%), *SLR* (%), *SUG* (%), *ALK* (%) and *CVS* (%) in the BLY varietal group. Data based on the mean of two locations.

Table 12 Estimates of $\hat{g}_i$ of the parents involved in the diallel of the BLY varietal group for the traits *YLD* (kg/ha), *GQI* (%), *SLR* (%), *SUG* (%), *ALK* (%) and *CVS* (%), and the respective errors associated to the estimates of each parent and in the between- parent comparison ($s_{\hat{g}_i}$ and $s_{\hat{g}_i-\hat{g}_j}$). Data presented on the mean of two locations.

Parent	YLD	GQI	SLR	SUG	ALK	CVS
1	105.85	4.87	-1.56	-0.02	0.28	-1.37
2	117.33	1.00	-0.04	0.00	-0.01	17.53
3	-42.76	-0.18	0.07	-0.01	0.10	-1.34
4	2.14	-0.73	-0.71	-0.01	0.21	-0.24
5	-90.14	-1.79	-0.09	0.00	-0.10	1.47
6	-1.41	-4.21	0.58	0.00	-0.30	-2.98
7	51.01	2.68	1.57	-0.01	0.10	-2.69
8	-76.24	-1.48	0.31	0.00	-0.03	-2.83
9	43.88	0.92	0.25	0.04	-0.14	-4.38
10	-109.66	-1.07	-0.38	0.00	-0.10	-3.17
$s_{\hat{g}_i}$	37.81	1.30	0.16	0.01	0.08	0.20
$s_{\hat{g}_i-\hat{g}_j}$	56.36	1.94	0.24	0.02	0.12	0.30

DISCUSSION

The two locations where the experiments with each varietal group were carried out, although close in distance, were different in several aspects, that contributed to significant differences being detected among them in the analysis of variance for most of the traits (Table 3 and 8). One of the locations common to the two groups is the Souza Cruz Company Experimental Station and the others belonged to farmers with a tradition of cultivating tobacco. It is important to mention that more than 92% of tobacco is produced in Brazil by small farmers (BTY, 2014; PRIEB, 2005; SCHUCH, 2003). For this, locations 2, both for the FCV and the BLY varietal groups were conducted on-farm.

In function of the variation between locations and treatments assessed, the treatment x location (TxL) interaction was significant for some traits (Tables 3 and 8). The occurrence of the genotype x environment (GxE) interaction is common in the tobacco crop in Brazil (PULCINELLI et al., 2014) and in the world (BOWMAN et al., 1986; CASTELLI et al., 1994; MATZINGER; WERNSMAN; ROSS, 1971; SADEGHI et al., 2011). When there is interaction, there are two options, either to consider the results for each environment or the mean of the environments. Considering that the biggest challenge of the breeder is identifying superior cultivars for planting in the farmer´s field (GAUCH; ZOBEL, 1988), and *a priori* the climatic conditions of the future are unknown, the best option is to consider the results on the mean of the environments so there is a better chance of representing the future cropping locations. Furthermore, when dealing with breeding, it would be impossible to develop and recommend cultivars for specific locations. Therefore, this strategy has been frequently used (DUDLEY, 1997; FERREIRA et al., 2015; LIMA et al., 2014;)

Tobacco is a very particular crop, where there is major concern with the physical quality and chemical composition of the leaves, in addition to other agronomic aspects, such as yield. There are some varietal groups that are defined predominantly by qualitative aspects and some are planted in specific regions in the world. In the case of Brazil, the main varietal groups are, in first place the FCV, that occupies the largest area (83%), and then the BLY, planted in 12% of the area (AFUBRA, 2015) so that these two groups were chosen to be assessed in the present study.

Considering these two groups, most of the important traits are common among them. Both for the FCV and the BLY varietal groups, it was clear that *YLD* (kg/ha) has enormous importance and for this reason it was assessed for the two groups. *GQI*, as mentioned, is a standard index presented in the average grade of the plot taking into account the physical characteristics of the leaf such

as color, shine and body (MAPA/BRASIL, 2007). This index is extremely important because it is used to assess the commercial value of the tobacco. Thus, the higher the *GQI*, the higher the value of the tobacco, that is, the more the producer earns. Regarding the steam weight against the total leaf weight, the smallest percentage as possible is required because it is not desirable that the plant uses its energy for steam production, but rather it should be used for lamina that represents a high quality product for industrial use. Another trait considered was *SUG*, which is fairly important, especially for the FCV group. In this group, the ideal is to obtain content close to 12%. *ALK* was assessed in the two groups and it ideal range is between 2.5% a 3.5%. Nicotine to nornicotine conversion occurs mainly in tobacco in the BLY varietal group and therefore was not assessed in FCV. Due to the undesirable properties of nornicotine, the smallest conversion percentage possible should be obtained (HECHT, 2003; BURNS et al., 2008). The lines or hybrids considered "low converters" by the industry have conversion below 2.5%. For this, only lines and hybrids below this range are selected as candidates for commercial planting.

The main challenge of breeding programs is to obtain new cultivars better than those already existing for all the characteristics of commercial interest regardless of the varietal group. In this context, the comparison of the performance of the lines and hybrids involved in the diallel against the checks assesses the potential of the germplasm used. For example, the mean of the diallel treatments for yield was 10.6% and 16.4% larger than the checks in the FCV and BLY varietal groups, respectively (Table 3). However, for *GQI* in the BLY varietal group there was slight reduction in the mean (4.4%) (Table 9). However, this is a mean value, and some lines and hybrids presented *GQI* larger than the checks. In the case of *ALK*, the mean of the diallel was 15.3% lower than the checks for the FCV varietal group. *ALK* of the diallel ranged from 1.33% to 4.00% for the hybrids and lines, with 1.94% mean. Thus lines/hybrids

can be identified within the ideal range (2.5% – 3.5%). It is common that the *ALK* does not reach this range in tobacco lines and hybrids cultivated in the United States, due to the negative correlation between *YLD* and *ALK* reported by many authors (LEGG; MATZINGER; MANN, 1965; LEWIS, 2006; MATZINGER; MANN, 1964; MATZINGER; MANN; COCKERHAM, 1972; MATZINGER; MANN; ROBINSON, 1960; MATZINGER; WERNSMAN, 1968; MATZINGER; WERNSMAN; WEEKS, 1989). Therefore, one of the highest difficulties of tobacco breeders in the USA is to increase yield while maintaining an acceptable alkaloid content level (BOWMAN, 1996; MOON et al., 2009).

For the BLY varietal group, the treatments involved in the diallel presented higher *GQI* compared to the checks and also a higher *CVS* percentage, which is not desirable as already mentioned, because of nornicotine is a precursor of risk associate compounds. It is pointed out that convertor parents were involved in the diallel, lines 2 and 5, that resulted in increase in the mean. However, many lines and hybrids presented a much lower *CVS* compared to the checks (Tabela 9), showing the productive and qualitative potential of the lines and hybrids in the breeding program.

One of the methods to assess the line combining ability is the methodology by Griffing (1956), where diallel crosses are used. According to the author, the expected value of a cross is the function of the sum of the general combining ability of the parental lines ($\hat{g}_i$). However, the true value of the hybrid may deviate from the expected value. This deviation is named specific combining ability (s_{ij}) of the parental lines. In statistical terms $\hat{g}_i$ represents the main effects and s_{ij}, the interaction (FALCONER; MACKAY, 1996). In this way the value of a hybrid combination (X) disregarding the error associated to the estimate of the means is:

$$X - \bar{X} = \hat{g}_i + \hat{g}_j + \hat{s}_{ij} \quad (1)$$

where $\bar{X}$ is the general mean; $\hat{g}_i$, $\hat{g}_j$ and $\hat{s}_{ij}$ are the general and specific combining abilities of parents *I* and *j*, and of the combination *I* x *j*, respectively. The $\hat{g}_i$ effect of each parent is estimated from the performance of parent *I* in crossings with all the others. Considering only one locus, Venkovsky; Barriga (1992) showed that $\hat{g}_i$ is supplied by:

$$\hat{g}_i = (p_i - \bar{p})[a_B + (1 - 2\bar{p})d_B] \qquad (2)$$

where p_i and $\bar{p}$ are the allelic frequencies of parent *i* and in the mean of the parents, respectively; a_B and d_B are the deviations of the homozygote and heterozygote, respectively, in relation to the mean of the homozygotes for locus *B*.

Expression (2) shows that $\hat{g}_i$ is equal to the contribution of the loci in homozygosis (*a*) only if $\bar{p}$=0.5 or if *d* is nil. That is, if $\bar{p}$=0.5 or *d=0*, the mean of the lines would explain all the combining ability and there would be no need to test the lines in hybrid combinations. In the present study, it was clear in the FCV that *a* alone was not sufficient to explain $\hat{g}_i$, because the three highest-yielding parents were lines 2, 10 and 11; but of these three, only parent 2 was among those that presented the highest $\hat{g}_i$ (Tables 3 and 6). This is a strong indication that the second part of the expression also contributed to $\hat{g}_i$ and therefore, *d* is different from zero and/or $\bar{p} \neq 0.5$.

Expression (1), shows that $\hat{s}_{ij}$ also contributed to the performance of the hybrid combination. It is significant when there are combinations that are relatively superior or inferior in relation to the mean performance of the parents (HALLAUER; CARENA; MIRANDA FILHO, 2010). It is represented by the following expression, also considering only one locus B (VENKOVSKY; BARRIGA, 1992):

$$s_{ij} = 2[(\bar{p} - p_i)(r_i - \bar{r})d_B]$$

where r_i and $\bar{r}$ have the same significance as p_i and $\bar{p}$, but for the other parent. In this expression it is evident that the *SCA* will only be significant if the parents are divergent and there is dominance in the control of the trait. The results obtained in the present study highlight the importance of *SCA* in explaining the total variation among the hybrids (Tables 5 and 11). This fact is not common in the autogamous species for reasons already given. It is emphasized that the lines that started the breeding program are very divergent, that may be one of the reasons for the high *SCA* contribution. This reflects once more the importance of heterosis in the two tobacco varietal groups involving the best lines available in the company breeding program and thus it should be exploited.

In the present study the proportion of the variation of *SCA* compared to *GCA* ranged from 13% to 87% (Tables 5 and 11). There are many reports in the literature where the *GCA* contribution is much larger than the *SCA* contribution in tobacco (DEAN, 1974; IBRAHIM; SLAVÍK; AVRATOVSCUKOVÁ, 1984; MATZINGER; MANN; COCKERHAM, 1962; MATZINGER; WERNSMAN; ROSS, 1971;). This is also frequent for other species, mainly the autogamous (ABREU; RAMALHO; FERREIRA, 1999; RIOS, 2015)

The proportion of *GCA* and *SCA* explaining the performance of the best hybrids was variable (Tables 5 and 11). As already commented, taking as example YLD of FCV, the main contribution to some combinations was from the *SCA* and for others from the *GCA*. This information is fundamental when choosing segregating populations to extract superior lines. If the *GCA* contribution is large, in principle it is an indication that the mean of the lines of that population in F_∞ will be high. However, the ideal is to associate high *GCA* with high *SCA*. Under this condition, the mean of the lines should be highly

associated to significant variation that is very desirable to identify elite lines (ABREU; RAMALHO; SANTOS, 2002; BURTON, *to be published*). In the example, the segregant population of pair *3x2* would be the best option, because it fits the criteria mentioned. In the population derived from hybrid *11x7*, substantial segregation is expected, because most of the loci should be in heterozygosis, but associated to the low means of the lines in F_∞. Thus the probability of obtaining lines with superior performance to the *3x2* hybrid population is small. Furthermore, the *GCA* estimate is also useful for choosing parents in a recurrent selection program. Preference should be given to parents with high $\hat{g}_i$ estimate for the traits of most interest.

One of the objectives of the present study was to verify the viability of producing commercial hybrids from the crosses of the best lines available in the company breeding program. Heterosis is estimated by the superiority of the F_1 generation compared to the mean of the parents (SHULL, 1908; HALLAUER; CARENA; MIRANDA FILHO, 2010). When there is a diallel, mean or specific heterosis can be estimated. Mean heterosis refers to the combination of all hybrid combinations compared to the mean of the lines. Specific heterosis is attributed to the heterosis of each pair individually.

Comparison of the line and hybrid means showed that the heterosis measured from the parental means varied from 1%, for *SLR*, to 7%, for *SUG*. However, the mean heterosis was practically nil in the BLY, except for *CVS* (7.8%). The mean of the parents was also larger than the mean of the checks (Table 3), showing the success in breeding the most recently obtained lines. In the case of FCV, the line performance contributed with more than 92% to the mean of the hybrids, considering all the traits. For the BLY varietal group the contribution was even bigger. Even in allogamous plants, such as corn, in which the mean heterosis for grain yield is high, the increase in trait expression has been associated with improvement of the inbred line *per se*, while the heterosis

percent has remained constant over the years (LI et al., 2013; TROYER; WELLIN, 2009). In rice was also shown by genomic analysis only a few loci with dominance or over-dominance, the superiority of hybrids against parental lines was due to the combination of superior alleles instead (HUANG et al., 2015)

The specific heterosis estimates reveal that there is enormous variation among the hybrids within each group for all the traits (Figures 1 and 3). However, a perfect association was not found between performance *per se* of the line and the hybrid performance. For example, for the trait *YLD* in FCV, based only on line performance *per se*, lines 2 and 10 would be selected for crossing (Table 3). However, hybrid *2x10* is not among the highest yielding. This fact shows that tobacco breeding programs should concentrate on selecting lines with good performance *per se* and in later stages, identify those which complement each other best, that is, obtain the best hybrid combinations.

Heterosis in autogamous plants is normally not of the same magnitude as that observed in alogamous plants (BERNARDO, 2010; CHEN, 2010). In the autogamous species, since self-pollination occurs naturally, probably with evolution, the frequency of deleterious alleles was reduced. (ALLARD, 1999; BERNARDO, 2014). As a consequence the inbreeding is low, close to zero. The mean heterosis obtained in the present study is of similar magnitude to that reported in other autogamous species, such as soybean (BURTON; BROWNIE, 2006), rice (LI et al., 2012) and wheat (KRYSTKOWIAK et al., 2009). Heterosis in tobacco has also been assessed in other countries, focused on *YLD* and *ALK*. The values of the estimates were comparable to those reported here (ALEKSOSKI, 2010; LEGG; COLLINS; LITTON, 1970; VANDENBERG; MATZINGER, 1970).

The hypotheses to explain heterosis were proposed at the start of the 20th century but to date there are still doubts about what is behind heterosis

(CHEN, 2013; SCHNABLE; SPRINGER, 2013). Falconer; Mackay (1996) showed for a single locus B: $h=d_B\ Y^2$, where h refers to heterosis ; d_B to the heterozygote deviation in relation to the mean point; and Y is the divergence in the allele frequencies of the parents. This expression shows that, for heterosis to occur, there must be dominance in the expression of the trait and the parent should be divergent. The initial questioning of researchers was whether only dominance was necessary for heterosis to occur, or whether there was "over dominance", that is, the advantage of the heterozygote in relation to the homozygote. The doubt still persists (BARANWAL et al., 2012; CROW, 1999, LI et al., 2008). Other explanations for the occurrence of heterosis have arisen, such as: presence of epistasis (LI et al., 2001; LUO et al., 2001; SCHENELL; COCKERHAM, 1992); genome complementarity (FU; DOONER, 2002; SCHNABLE; SPRINGER, 2013); and even epigenetic factors (CHEN, 2010, 2013; TSAFTARIS et al., 2005). However, all these hypotheses point to the need to identify parent lines that are complimentary. Therefore for maximum advantage with the use of hybrids in the breeding programs, the combining ability must be assessed of the available superior lines.

An interesting fact observed was the occurrence of some negative heterosis estimates. Given that heterosis is calculated by $h = F_1 - \frac{P_1+P_2}{2}$, when h is a function only of dominance, positive or negative values can be obtained, depending only on the d direction. Nevertheless, when positive and negative values are found for the same traits, d alone is not sufficient to explain the superiority of some hybrid combinations (GOODNIGHT, 1999). To explain this type of heterosis, the interlocus interaction can be considered, i.e. epistasis. The epistasis interactions result in changes of the allele mean effect according to the hybrid combinations or crosses (PRAY; GOODNIGHT, 1995). Thus, some combinations can express positive, negative or nil heterosis and so the hybrid

performance cannot be easily predicted by statistical models when this phenomenon occurs.

It should be remembered that hybrid commercialization is only feasible when they can be produced at a cost consistent with the hybrid performance, that is, the feasibility depends on the heterosis obtained in the cross (CIMMYT, 2000; FEHR, 1987;). However, in tobacco, the use of hybrid seeds should be considered from another angle, because artificial crossing is already made to protect the cultivars by male sterility and so no extra cost is incurred (MANN; JONES; MATZINGER, 1962; SCHNABLE; WISE, 1998). Therefore, even with small levels of heterosis, the hybrid between two different lines would outperform the inbreds.

Furthermore, with the use of hybrids, favorable phenotypic characteristics of interest can be combined, that are present in different lines, as for example, disease resistance, agronomic aspects and qualitative characteristics. The use of hybrids in tomato has been widely adopted in spite of the low heterosis of this species reported in the literature (LIPPMAN; ZAMIR, 2007). In this case, the major interest is to combine in the hybrid phenotypes of interest present in different lines. In the United States, tobacco hybrids use has also been common (MILLER; KENNEDY; RITCHEY, *year not available*; NCCIA, 2015). However, heterosis is not exploited in the best way, as described in the present study, and what it is being done is to combined lines with different observed phenotypes in a hybrid.

Another interesting finding in the present study was the heterosis direction. For traits in which increase in heterosis is desirable, i.e. *YLD*, *GQI* and *SUG*, the heterosis estimates were positive. For those that require reduced expression, i.e. *SLR*, *ALK* and *CVS*, the estimates were negative. Thus heterosis can be used in favor of the breeder in hybrids development in Brazil considering the traits of economic importance. Therefore the best strategy to be adopted in

the tobacco breeding programs is the use of diallel crossings to identify the best hybrid combinations among superior lines and also trying to complement the simply inherited traits present in distinct lines in the hybrids, such as morphological traits and disease resistance, especially to viruses, which are the major pathological problem in tobacco production (BURK; CHAPLIN, 1980).

CONCLUSIONS

SCA contributed substantially to the total variation of most of the traits, indicating the occurrence of dominance in the genetic control of the traits.

In the case of yield, heterosis was 5.37% in the mean of the 72 hybrid combinations of the FCV varietal group. For the BLY varietal group, the mean heterosis was nil. However, for both the groups, combinations were identified with heterosis above 15%.

The use of hybrids in tobacco should be encouraged not only to use the heterosis but also to associate other phenotypes of traits of interest, which are in different lines.

Hybrid performance usually depends on both *GCA* and *SCA*, so the breeding program should focus on obtaining lines with good performance *per se*, but that are complimentary, to obtain commercial hybrids.

REFERENCES

ABREU, A. F. B.; RAMALHO, M. A. P.; FERREIRA, D. F. Selection potential for seed yield from intra and inter racial populations in common bean. **Euphytica,** v. 108, p. 121-127. 1999.

ABREU, A. F. B.; RAMALHO, M. A. P.; SANTOS, J. B. Prediction of seed-yield potential of common bean populations. **Genetics and Molecular Biology,** v. 25, n. 3, p. 323-327, 2002.

AFUBRA. **Associação dos Fumicultores do Brasil**. Disponível em: <http://www.afubra.com.br>. Acesso em: 6 dez. 2015.

ALEKSOSKI, J. Estimation of the heterotic effect in f1 generation of various tobacco genotypes and their diallel crosses. **Biotechnology & Biotechnological Equipment,** v. 24, n. 2, p. 407-4011, 2010.

ALLARD, R. W. History of plant population genetics. **Annual Review of Genetics,** v. 33, p. 1–27, 1999.

AYCOCK, M. K.; MANN, T. J.; MATZINGER D. F. Investigations with a form of cytoplasmic male-sterility in flue-cured tobacco. Tobacco Science, v. 7, p. 130-135, 1963.

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

BELOGRADOVA, K. et al. An Overview on Tobacco Haploids. In: Touraev, A.; Forster, B. P.; Jain, S. M. **Advances in Haploid Production in Higher Plants.** Spring: Netherlands. 2009. p. 75-85.

BERNARDO, R. **Breeding for Quantitative Traits in Plants**. 2. ed. Woodbury: Stemma Press, 2010. 390 p.

BERNARDO, R. **Essentials of Plant Breeding.** 1. ed. Woodbury: Stemma Press, 2014. 260 p.

BOWMAN, D. T. History of the Regional Minimum Standards Program for the release of flue-cured tobacco varieties in the United States. **Tobacco Science**, v. 40, p. 99–110, 1996.

BTY. **Brazilian Tobacco Yearbook**. Gazeta, 2014.

BURK, L. G.; CHAPLIN, J. F. Variation among anther-de- rived haploids from a multiple disease-resistant tobacco hybrid. **Crop Science**, v. 20, p. 334-338, 1980.

BURNS, D. M. et al. Mandated lowering of toxicants in cigarette smoke: a description of the World Health Organization TobReg proposal. **Tobacco Control,** v. 17, p. 132–141, 2008.

BURTON, J. W.; BROWNIE, C. Heterosis and Inbreeding Depression in Two Soybean Single Crosses. **Crop Science**, v. 46, n. 6, p. 2643-2648, 2006.

BURTON, J. W.; BROWNIE, C. Heterosis and Genetic Variance in Four Soybean RIL Population. *To be published*

CASTELLI, F. et al. Adaptation of aromatic burley tobacco varieties to northern Italian environments. **Journal of Agronomy and Crop Science,** v. 172, p. 342–351, 1994.

CHEN, Z. J. Molecular mechanisms of polyploidy and hybrid vigor. **Trends Plant Science,** v. 15, p. 57–71, 2010.

CHEN, Z. J. Genomic and epigenetic insights into the molecular bases of heterosis. **Nature Reviews: Genetics,** v. 14, p. 471-482, 2013.

CORESTA. Determination of Nicotine in Tobacco Products by Gas Chromatographic analysis, **Coresta Recommended Method n° 62**. Fev. 2005. 9 p.

CROW, J. F. Dominance and overdominance. In: Coors, J.G.; Pandey, S. (Eds.) **The Genetics and Exploitation of heterosis in crops**. Madison: American Society of Agronomy, 1999, p.49–58.

CIMMYT**. The Genetics and Exploitation of Heterosis: An International Symposium** – Book of Abstracts, 2000.

DEAN, C. E. Heterosis, Inbreeding Depression, and Combining Ability in Diallel Crosses of Cigar-Wrapper Tobacco. **Crop Science,** v. 14, n. 3, p. 482-484, 1974.

DUDLEY, J. W. Quantitative Genetics and Plant Breeding. **Advances in Agronomy,** v. 59, p. 1-13, 1997.

FALCONER, D. S.; MACKAY, T. F. C. **Introduction to quantitative genetics**. 4th ed. Malaysia: Pearson, 1996. 464 p.

FAOSTAT. **Food and Agriculture Organization**. 2012. Disponível em: <http://faostat.fao.org/>. Acesso em: 6 dez. 2015.

FEHR, W. R. **Principles of cultivar development**. New York: Macmillan, 1987. 761 p.

FERREIRA, R. A. D. C. et al. Implications of the number of years assessment on recommendation of common bean cultivars. **Plant Breeding,** v. 134, p. 599-604, 2015.

FU, H.; DOONER, H. K. Intraspecific violation of genetic colinearity and its implications in maize. **PNAS,** v. 99, p. 9573–78, 2002.

GAUCH, H. G. Model selection and validation for yield trials with interaction. **Biometrics**, v. 44, p. 705–715, 1988.

GOODNIGHT, C. J. Epistasis and Heterosis. In: Coors, J.G.; Pandey, S. (Eds.) **The Genetics and Exploitation of heterosis in crops**. Madison: American Society of Agronomy, 1999, p.59–68.

GRIFFING, B. Concept of general and specific combining ability in relation to diallel crossing systems. **Australian Journal of Biological Sciences,** v. 9, p. 463–493, 1956.

HALLAUER, A. R.; CARENA, M. J.; MIRANDA FILHO, J. B. **Quantitative genetics in maize breeding**. 3rd ed. New York: Springer, 2010. 680 p.

HECHT, S. S. Tobacco carcinogens, their biomarkers and tobacco-induced cancer. **Nature Reviews Cancer**, v. 3, p. 733–744, 2003.

HUANG, X. et al. Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. **Nature Communications.** v. 6, p. 1-9, 2015.

HUBER, S. C.; RUFTY, T. W.; KERR, P. S. Effect of Photoperiod on Photosynthate Partitioning and Diurnal Rhythms in Sucrose Phosphate Synthase Activity in Leaves of Soybean (*Glycine max* L.[Merr.]) and Tobacco (*Nicotiana tabacum* L.). **Plant Physiology,** v. 75, p. 1080-1084, 1984

IBRAHIM, H.; SLAVIK, B.; AVRATOVSCUKOVA, N. Yield and Yield Components in Flue-Cured Tobacco and Their Genetic Analysis. **Biologia Plantarum**, v. 26, p. 285-292, 1984.

ITGA. **International Tobacco Grower's Association**. Disponível em: <www.tobaccoleaf.org>. Acesso em: 7 dez. 2015.

KRYSTKOWIAK, K. et al. Relationship between phenotypic and genetic diversity of parental genotypes and the specific combining ability and heterosis effects in wheat (Triticum aestivum L.). **Euphytica,** v. 165, p. 419-434, 2009.

LEGG, P. D.; MATZINGER, D. F.; MANN T. J. Genetic variation and covariation in a Nicotiana tabacum L. synthetic two generations after synthesis. **Crop Science**, v. 5, p. 30–33, 1965.

LEWIS, R. S. Identification of Germplasm of Possible Value for Confronting an Unfavorable Inverse Genetic Correlation in Tobacco. Crop Science, v. 46, p. 1764-1771, 2006.

LI, L. et al. Dominance, Overdominance and Epistasis Condition the Heterosis in Two Heterotic Rice Hybrids. **Genetics,** v. 180, p. 1725-1742, 2008.

LI, Y. et al. Contributions of Parental Inbreds and Heterosis to Morphology and Yield of Single-Cross Maize Hybrids in China. **Crop Science**, v. 54, p. 76-88, 2013.

LI, Z.K. et al. Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. I. Biomass and grain yield. **Genetics,** v. 158, p. 1737–1753, 2001.

LIMA, L.K. et al. Implications of predictable and unpredictable environmental factors in common bean VCU trials in Minas Gerais. **Crop Breeding and Applied Biotechnology,** v. 14, p. 160-165, 2014.

LIPPMAN, Z. B.; ZAMIR, D. Heterosis: revisiting the magic. **Trends in Genetics,** v. 23, n. 2, p. 60-66, 2007.

LUO, L. J. et al. Overdominant epistatic loci are the primary genetic basis of inbreeding depression and heterosis in rice. II. Grain yield components. **Genetics,** v. 158, p. 1755–1771, 2001.

MANN, T. J.; JONES, G. L.; MATZlNGER D. F. The use of cytoplasmic male sterility in flue-cured tobacco hybrids. **Crop Science**, v. 2, p. 407-410, 1962.

MAPA/BRASIL. Instrução Normativa n˚10 do Ministério da Agricultura, Pecuária e Abastecimento, de 16 de abril de 2007. Institui a classificação de Tabaco**. Diário Oficial [da] República Federativa do Brasil**, Brasília, 23 mai. 2006. Disponível em: <http://www.agricultura.gov.br/legislacao>. Acesso em: 7 dez. 2015.

MATZINGER, D. F.; MANN, T. J. Genetic studies on associations between nicotine content and yield of flue-cured tobacco. In: **Proc. 3rd World Tob. Sci. Congr.** Salisbury, 18–26 Fev. 1963, p. 357–365. 1964.

MATZINGER, D. F.; MANN T. J.; COCKERHAM, C. C. Recurrent family selection and correlated response in Nicotiana tabacum L. I.'Dixie Bright 244' 3 'Coker 139.'. **Crop Science**, v. 12, p. 40–43, 1972.

MATZINGER, D. F., MANN T. J., ROBINSON H. F. Genetic variability in flue-cured varieties of Nicotiana tabacum. I. Hicks Broadleaf 3 Coker 139. **Agronomy Jounal**, p. 52, p. 8–11, 1960.

MATZINGER, D. F.; WERNSMAN, E. A. Four cycles of mass selection in a synthetic variety of an autogamous species Nicotiana tabacum L. **Crop Science,** v. 8, p. 239–243, 1968.

MATZINGER, D. F.; WERNSMAN, W. A.; WEEKS W. W. Restricted index selection for total alkaloids and yield in tobacco. **Crop Science**, v. 29, p. 74–77. 1989.

MATZINGER, D. E.; WERNSMAN E. A.; ROSS H.F. Diallel crosses among burley varieties of nicotiana-tabacum L. in F1 and F2 generations. Crop Science, v. 1, p. 275-279, 1971.

MILLER, R. D.; KENNEDY, B. S.; RITCHEY, E. L. Tobacco Breeding and Genetics. **Tobacco Research,** p. 31-32. *Year not available.*

MOON, H. S. et al. Changes in Genetic Diversity of U.S. Flue-Cured Tobacco Germplasm over Seven Decades of Cultivar Development. **Crop Science**, v. 49, 2009.

NCCIA. 2015 Blue Seed List: List of varieties, acreage, and producers of certified Forage Grass, Turf Grass, Peanut, Soybean, Sweetpotato and Tobacco

in 2043 for 2015**. North Carolina Crop Improvement Association**. Disponível em: <http://www.nccrop.com/forms.php>. Acesso em: 7 dez. 2015.

PARKES, E. Y. et al. Combining ability of cassava genotypes for cassava mosaic disease and cassava bacterial blight, yield and its related components in two ecological zones in Ghana. **Euphytica,** v. 194, n. 1, p. 13-24, 2013.

PRAY, L. A.; GOODNIGHT, C. J. Genetic variation in inbreeding depression in the red flour beetle *Tribolium castaneum*. **Evolution**, v. 49, p. 176–188, 1995.

PRIEB, R. I. P. **Pluriatividade na produção familiar fumageira**. Universidade do Texas: Editora EDUNISC, 2005. 195 p.

PULCINELLI, C. E. et al. Experimental strategies in carrying out VCU for the tobacco crop II: dimension of the experimental network. **Genetics and Molecular Research**, v. 13, n. 3, p.. 5541-5554, 2014.

QI, H. et al. Identification of combining ability loci for five yield- related traits in maize using a set of testcrosses with introgression lines**. Theoretical and Applied Genetics**, v. 126, n. 2, p. 369-377, 2013.

RIOS, R. O. Breeding Self-Fertilizing Plants: From Inbred to Hybrid Cultivars. In: Rios, R. O. (Ed.) **Plant Breeding in the Omics Era**. Springer, 2015, p. 141-171.

SADEGHI, S. M. et al. Additive main effects and multiplicative interactions (AMMI) analysis of dry leaf yield in tobacco hybrids across environments. **African Journal of Biotechnology,** v. 10, n. 21, p. 4358-4364, 2011.

SCHNABLE, P. S.; SPRINGER, N. M. Progress Toward Understanding Heterosis in Crop Plants. **Annual Review of Plant Biology**, v. 64, p. 71-88, 2013.

SCHNABLE, P. S.; WISE, R. P. The molecular basis of cytoplasmic male sterility and ferlility restoration. **Trends in Plant Science**, v. 3, n. 5, p. 175-180, 1998.

SCHENELL, F. W.; COCKERHAM, C. C. Multiplicative vs. arbitrary gene action in heterosis. **Genetics,** v. 131, n. 2, p. 461-469. 1992.

SCHUCH, H. **Culturas Gaúchas: Fumo.** Porto Alegre: Assembléia Legislativa do Rio Grande do Sul, 2003.

SHULL, G. H. The composition of a field of maize. **Rep Am Breeders' Assoc,** v. 4, p. 296– 301, 1908.

THOMAS, J. F. et al. Time of floral initiation in tobacco as a function of temperature and photoperiod. **Canadian Journal of Botany**, v. 53, n. 14, p. 1400-1410, 1975.

TSAFTARIS A. et al. Epigenetic mechanisms in plants and their implications in plant breeding. p. 157–171. In: Tuberosa, R.; Philips R.; Gale M. (eds.) **In the wake of the double helix: From the green revolution to the gene revolution**. Bologna, Italy: Avenue Media, 2005.

TSO, T. C.; KASPERBAUER, M. J.; SOROKIN, T. P. Effect of Photoperiod and End-of-Day Light Quality on Alkaloids and Phenolic Compounds of Tobacco. **Plant Physiology,** v. 45, p. 330-333, 1970.

TROYER, A. F.; WELLIN, E. J. Heterosis decreasing in hybrids: Yield test inbreds. **Crop Science**, v. 49, p. 1969–1976, 2009.

VANDENBERG, P.; MATZINGER, D. F. Genetic Diversity and Heterosis in Nicotiana. III. Crosses among Tobacco Introductions and Flue-Cured Varieties. **Crop Science,** v. 10, p. 437-440, 1970.

VENKOVSKY, R.; BARRIGA, P. **Genética Biométrica no Fitomelhoramento.** 1. ed. Ribeirão Preto: Revista Brasileira de Genética, 1992. 496 p.

ARTICLE 2 Application of multivariate and standardized selection indices for multiple trait selection in Flue-cured Virginia and Burley tobacco

Application of multivariate and standardized indices for multi-trait selection in Flue-cured Virginia and Burley tobacco

Bruna Line Carvalho, Carlos Eduardo Pulcinelli, Ramsey Lewis, Adriano Teodoro Bruzi, Magno Antonio Patto Ramalho

ABSTRACT

Several traits must be considered in selection of tobacco cultivars for attending the requirements of farmers, industry and consumers. Breeders should know the association of these traits to trace the best strategy to have gains for all them. The selection indices have been shown as the most efficient. In tobacco no reports were found for a multiple trait selection. In this study we aimed to show the efficiency of two indices: i) sum of the standardized variables index (SSV); ii) multivariate canonic variables index (CMV). The first do not consider the correlation among the traits, and it is of simple and easy for application. The second it's a multivariate index, which the correlation among traits is accounted for its estimation. It were evaluated 13 lines of the Flue-Cured Virginia (FCV) and 10 lines of Burley (BLY) varietal groups of tobacco. For the FCV 13 lines were crossed in a diallel design and 72 hybrids were obtained. For BLY 10 lines were crossed in the same scheme and 41 hybrids were obtained. The hybrids, parent lines and checks were evaluated in the field in 10x10 and 8x8 triple lattice design for FCV and BLY, respectively. The traits assessed were yield (YLD), general quality index (GQI), steam by leaf lamina ratio (SLR), total alkaloids (ALK), total sugar content (SUG) for the FCV, and conversion nicotine to nornicotine (CVS) for the BLY. We considered the weights according to the economic importance of each trait that were: 0.4 for YLD, 0.3 for GQI and 0.1 for the others. The general and specific combining abilities (GCA and SCA) and heterosis were estimated for each index. We identified that both varietal groups

and indices have a high contribution of SCA variation for total variation; and there are hybrid combinations with significant heterosis. We also found that besides the MV index shows higher accuracy for discriminate the lines and hybrids, the SV index is the most appropriate for usage in tobacco breeding programs since it provides higher gains considering all the traits.

Keywords: Plant breeding; Quantitative Genetics; Diallel crosses; Heterosis; Selection Index

INTRODUCTION

A new cultivar will only be adopted if it meets the requirements of the producers, industry and consumers. These players have very different needs and their interests do not coincide most of the cases, so the most difficulty faced by plant breeders is that selection is hardly ever directed for a single trait. In the case of the tobacco crop, this fact is very evident.

There are some strategies for selection of traits simultaneously (SIMMONDS; SMARTT, 1999): (i) *in tandem* selection, when different traits are selected in different generations; (ii) *independent culling levels*, when the traits are selected simultaneously but independently; and (iii) *selection index*, when the traits are selected simultaneously using an index that, generally, represents a linear function of the traits considered. Each trait is weighted based on its importance. Several published studies and books have demonstrated that the maximum efficiency is obtained by constructing a selection index (BERNARDO, 2014; LONG et al., 2006; MILLIGAN; BALZARINI; WHITE, 2003; SHARMA; DUVEILLER, 2003). In the tobacco crop no publication was found where an index was constructed for simultaneous multiple trait selection. We pointed out that, in tobacco, there are some estimates of negative association among important traits that complicate the work of the plant breeder. However, these estimates were predominantly obtained under temperate climate conditions (LEWIS, 2006; WHITE; PANDEYA; DIRKS, 1979). In addition, the handling of these associations varies among varietal groups.

Although tobacco is considered a typically autogamous plant, obtaining hybrids is desirable even though heterosis is of low magnitude. This is because companies already carry out manual crosses to combine phenotypes of interest present on different lines and especially to introduce male sterility in the commercialized cultivars and thus obtain natural protection for the

breeder/company. Therefore carrying out diallel crosses is a priority to identify the lines with good combining ability (FALCONER; MACKAY, 1996).

Although the use of diallel crosses was frequently adopted in the past for individual traits in tobacco (BUTORAC; BELJO; GUNJACA, 2004; MATZINGER; MANN; COCKERHAM, 1962; PANDEYA; DIRKS; POUSHINSKY, 1983; VANDENBERG; MATZINGER, 1970), multitrait analysis is not common in the literature. Furthermore, no report was found of diallel analysis involving a selection index for any species.

There are innumerable methodologies for applying selection indices (RESENDE; SILVA; AZEVEDO, 2014). The most efficient is called the optimum index or the Smith-Hazel index (BERNARDO, 2010; HAZEL, 1943; RESENDE; SILVA; AZEVEDO, 2014; SMITH, 1936). However it has as restriction the estimation of genetic and phenotypic variances and covariances with high accuracy, which it is not always possible in practice. Therefore this index can only be applied when the effect of parents is considered random. This is rare in the case of diallels, because the parents used in the crosses are only those previously selected as superior. Several factors should be considered when deciding which index to use, such as the accuracy of the estimates, simplicity of application and interpretation, and obtaining results that are compatible with the requirements of the breeders.

Thus the present study was carried out with the objective of assessing the efficiency of selection indices of the Sum of the Standardized Variables (SSV) and the Canonic Multivariate Variables (CMV), involving some important traits in two varietal tobacco groups, Flue-Cured Virginia (FCV) and Burley (BLY), using data from diallel crosses.

MATERIALS AND METHODS

Description of the treatment and experimental procedures

As described previously, the data used in the present study were provided by Souza Cruz S/A, subsidiary of BAT (British American Tobacco). The same data as described in chapter 1 were used. Therefore the experimental procedure and the traits assessed are the equivalent.

Statistical analyses

First, the phenotypic data were submitted to the analysis of variance (ANOVA), using the interblock information, by location and across locations, considering the traits individually (STEEL; TORRIE; DICKEY, 1997). The phenotypic correlations were estimated pairwise among the traits and the *t* test was applied to verify whether the estimates were different from zero.

Later the estimates of the selection indices were obtained. For SSV index, the variables were first standardized per plot taking as reference the replication. The following estimate was used:

$$Z_{ijl}^{k} = \frac{X_{ijl}^{k} - \bar{X}_{jl}^{k}}{\sigma_{jl}^{k}}$$

Where, Z_{ijl} is the standardized variable of treatment *i*, in replication *j*, in location *l*, for trait *k*; X_{ijl} is the observation of treatment *i*, in replication *j*, in location *l*, for trait *k*; $\bar{X}_{jl}$ is the general mean of the treatments of replication *j*, in location *l*, for trait *k*; σ_{jk} is the standard deviation of replication *j*, in location *l*, for trait *k*;.

For the variables that the reduction is desired, that is, SLR, SUG, ALK and CVS, the values obtained were multiplied by (-1). In this way the criterion is

the same for all, i.e. the higher the value the better. As the variables obtained were a deviation from 0, negative and positive values were found, therefore a constant (5) was added to all so that there was no value below zero. The variables were then summed for each observation to obtain the SVV index. At this point a weight was assigned to each trait based on their individual economic and commercial importance, which it was 0.4 for YLD, 0.3 for GQI and 0.1 for the remainders.

Another index obtained in the present study was the CMV index. It was calculated by obtaining a matrix of the quadratic genetic components (G), as treatments were considered as fixed, and a matrix of residual variances and co-variances (R) of traits pairwise by covariance analysis (ANCOVA).

The ratio between G and R given below was maximized to obtain a canonic index of maximum efficiency (FERREIRA, 2011):

$$\underset{\{x: x^T R\, x = 1\}}{max}\, \lambda(x) = \frac{x^T G\, x}{x^T R\, x}$$

Considering that G and R are symmetric matrixes and R is positively defined, the maximum of $\lambda\,(x)$, under the restriction $x^T R\, x = 1$, is given by the largest value of the λ_i eigenvalue from $S_R^{-1} G (S_R^{-1})^T$, *i=1,2,...,5*, and by the corresponding eigenvector $x_i = (S_R^{-1})^T z_i$, in which S_R is the Cholesky factor of R and z_i is the i-th eigenvector of $S_R^{-1} G (S_R^{-1})^T$. Then, $\lambda_1 / \sum \lambda_i$ is the contribution of the eigenvalue λ_1 to the total genetic variation. Thus only eigenvector x_i of order i=1 was used for the CMV selection index as the corresponding eigenvalue explained 66% or more of the genetic variance. A weighting was attributed to this vector for the corresponding values of each trait, with the same weights as the SSV index.

The ANOVA of the indices was carried out and as the parent lines were crossed pairwise, diallel analysis of the indices was also performed for each index, according to Griffing fix model, method II (1956), based on the following model:

$$y_{ik} = m + g_i \ + g_k \ + s_{ik} + \ \bar{e}_{ik.}$$

where y_{ik} is the observation of the hybrid between parents *i* and *k*; m is a constant, which in this case represents the general mean of the treatments of the diallel; g_i and g_k are the effects of the general combining ability of the *i*-th and *k*-th parent, respectively; s_{ik} is the effect of the specific combining ability for the cross between parents *i* and *k*; and $\bar{e}_{ik.}$ is the mean experimental error, where $e \cap N(0.\sigma_e^2)$. The effects of the general combining ability and specific combining ability (*GCA* and *SCA*) were calculated by using the least-squares method.

The mean heterosis ($\bar{h}$), heterosis of the parent of order *i* (h_i) and specific heterosis of order *ik* (h_{ik}) were estimated given that:

$$\bar{h} = \frac{\bar{F}_1 - \bar{P}}{\bar{P}} x100$$

$$h_{ik} = \frac{F_{1ik} - \frac{(P_i + P_k)}{2}}{\frac{(P_i + P_k)}{2}} x100$$

$$h_i = \bar{h}_{i.}$$

where $\bar{F}_1$ is the general mean of the hybrids; $\bar{P}$ is the general mean of the parents; F_{1ik} is the mean of each hybrid combination *ik*, and P_i and P_k are the mean of the parents involved in each combination.

RESULTS

The treatment effects for both varietal groups were significantly different (P≤0.01) for all traits assessed (Table 1), an essential condition to reach the objectives of the present study, that is, to select for multiple traits based on an index.

Table 1 Summary of the joint analysis of variance of the locations, for traits *YLD* (kg/ha), *GQI* (%), *SLR* (%), *SUG* (%) and *ALK (%)* of the FCV group, and for the traits *YLD* (kg/ha), *GQI* (%), *SLR* (%), *ALK* (%) and *CVS (%)* of the BLY group.

			P value				
Group	**S.V.**	**DF**	**YLD**	**GQI**	**SLR**	**SUG**	**ALK**
FCV	Locations (L)	1	<0.01	<0.01	0.17	<0.01	<0.01
	Treatment (T)	99	<0.01	<0.01	<0.01	<0.01	<0.01
	T x L	99	0.28	0.11	<0.01	0.94	<0.01
	Ac*		0.48	0.62	0.85	0.59	0.95
			YLD	**GQI**	**SLR**	**ALK**	**CVS**
BLY	Locations (L)	1	<0.01	<0.01	<0.01	<0.01	0.80
	Treatment (T)	63	<0.01	<0.01	<0.01	<0.01	<0.01
	T x L	63	0.10	<0.01	0.02	<0.01	0.35
	Ac*		0.61	0.64	0.90	0.53	0.88

*Accuracy based on the mean of locations

When establishing an index it is necessary to know the association among the traits considered. Table 2 shows the phenotypic correlations

coefficients (*r*) among the traits pairwise, in the mean of the environments. Even for the pairs where *r* was significant, the magnitude was not very high. The highest estimate for the BLY group was 0.67 between ALK and GQI. The highest estimates in the FCV group were between GQI and SUG (*r* = 0.53) and GQI and SLR (*r* = -0.53). In general, the magnitude and the direction of association among the traits pairwise were similar in the two groups.

Table 2 Estimates of correlation between traits pairwise in the FCV varietal group, upper diagonal, and in the BLY varietal group, lower diagonal. Estimates obtained from the mean of two locations.

Traits	**YLD**	**GQI**	**SLR**	**ALK**	**SUG**
YLD		-0.15	0.33**	-0.07	0.03
GQI	0.53**		-0.53**	0.43**	0.53**
SLR	0.09	-0.21		-0.18*	-0.47**
ALK	0.27*	0.67**	-0.41		0.07
CVS	0.26*	0.04	-0.13	0.03	

**,* Significant at 1% and 5% probability by the t-test, respectively

Significant differences were detected (P≤0.01) among treatments in ANOVA by location for both indices (Table 3) in the two varietal groups. The accuracy estimates, which refer to an average of the locations, were all of high magnitude. The lowest estimate was 0.69 for the SSV index in the FCV group. These accuracy estimates allowed inferring that we had a good discrimination among treatments for both indices. The treatment x location (TxL) interaction was significant only for the SSV index in both groups (P≤0.01).

Table 3 Summary of the joint analysis of variance across locations and diallel analysis by the Griffing model (1956,) method II. Values based on the Snedecor's F test for the SSV and CMV indices of the FCV and BLY varietal groups.

	FCV			BLY		
S.V.	**DF**	**SSV**	**CMV**	**DF**	**SSV**	**CMV**
Locations - L	1	0.11	1.64	1	0.00	6.31***
Replication /L	4	0.06	13.34***	4	0.15	4.99***
Block /Rep	54	1.82***	1.57***	42	1.42***	1.32
Treatment (T)	99	1.92***	10.77***	63	4.84***	23.47***
Hybrids (H)	71	1.60***	7.69***	40	2.36***	23.90***
Parents (P)	12	1.93**	5.69***	6	3.93***	45.22***
Checks (C)	14	3.55***	12.36***	15	7.06***	2.80***
(H+P) vs C	1	3.40*	207.86***	1	76.09***	137.87***
H vs P	1	3.45*	40.39***	1	0.29	1.24
Diallel (H+P)	84	1.48***	8.44***	47	2.72***	27.05***
GCA	12	2.92***	41.02***	9	5.92***	122.01***
SCA	78	1.13	3.35***	38	2.00***	3.36***
T x L	99	1.55***	1.06	63	1.60***	0.94
H x L	71	1.61***	1.14	40	1.58**	1.14
P x L	12	1.71*	0.54	6	1.32	0.67
C x L	14	1.34	1.19	15	1.74**	0.52
Diallel x L	84	1.40**	1.13	47	1.43**	1.10
GCA x L	12	2.76***	1.65*	9	3.53***	0.44
SCA x L	78	1.07	1.00	38	0.87	1.11
Check mean		4.92	2.07		4.68	1.28
Diallel mean		5.02	2.34		5.11	1.10
Parents mean		4.93	2.23		5.09	0.99
Hybrids mean		5.03	2.36		5.11	1.11
Ac[1]		0.69	0.95		0.89	0.98

***,**,* Significant at 1%, 5% and 10% of probability by the *t*-test, respectively.
[1] Accuracy based on the mean of locations.

The partitioning of the source of variation (S.V.) treatments into parental (*P*), hybrid (*H*) and check (*C*) effects showed a similar result for both groups

(Table 3). The *C vs* diallel (*P*+*H*) contrast was also significant (P≤0.07). Considering the mean of the two locations, for the SSV index, the hybrids and parents presented an estimate ~2% larger than the checks for FCV and ~9% for BLY. Considering the CMV index, the superiority of the diallel treatments compared to the checks was similar for the two groups, ~13% for FCV and ~16% for BLY. The *H vs P* contrast estimates whether the mean heterosis was different from zero for the selection indices. Heterosis was only significant for the FCV group, considering the two indices. Although significant, the mean heterosis was of low magnitude, 2.2% for the SSV and 5.8% for the CMV index.

Tables 4 and 5 show the estimates of the mean of the parents and the hybrids. In the FCV group, the two parents with the highest mean for the SSV index were numbers *3* and *6*. The best performance hybrids were combinations *1x5* and *1x7*. Using the CMV index, the outstanding parents were numbers *2* and *10*, therefore different from the previous group. There was no agreement either when considering hybrid performance; the combinations of best performance for this index were *5x10* and *7x10*.

Similarly, for the BLY group, the indices assessed did not present similar results. The parents with the highest mean for the SSV index were *1* and *5*, and the hybrids that were outstanding came from the cross of parent number *1*, i.e. combinations *1x9* and *1x3*. For the CMV index parents *2* and *10* were superior, while the best hybrids were combinations *2x10*, *2x4* and *6x10*, i.e. the best hybrids also had at least one parent with a high mean in their composition.

Table 4 Means of hybrids and lines from FCV group for the two selection indices, SSV above the diagonal and CMV below the diagonal. Data based on the mean of two locations.

Parent	1	2	3	4	5	6	7	8	9	10	11	12	13	Mean SSV
1		5.36	5.11	4.43	5.45	5.39	5.45	5.30	5.13	5.23	5.13	5.32	4.85	5.05
2	2.73		5.09	4.91	5.16	5.23	5.06	4.79	5.15	5.24	3.93	-	-	5.18
3	2.31	2.49		5.02	5.09	5.28	5.26	5.36	5.22	5.01	5.09	5.10	5.17	5.34
4	2.12	2.47	2.19		4.98	5.00	4.81	5.00	4.77	4.99	4.95	5.07	5.14	4.89
5	2.41	2.43	2.12	1.97		5.12	4.95	4.78	4.84	4.51	5.25	5.04	5.04	4.43
6	2.43	2.34	2.25	2.20	2.15		5.07	5.03	5.08	4.56	5.11	4.98	-	5.22
7	2.54	2.50	2.33	2.31	2.42	2.39		4.96	4.85	5.14	5.15	4.96	4.67	4.70
8	2.41	2.24	2.26	2.24	2.15	2.26	2.46		4.87	5.13	5.27	5.24	5.23	4.80
9	2.44	2.40	2.19	2.23	2.29	2.43	2.35	2.28		4.63	5.23	5.09	5.04	4.93
10	2.57	2.70	2.54	2.41	2.88	2.77	2.87	2.50	2.50		-	4.74	4.91	5.06
11	2.29	2.55	2.09	2.22	2.30	2.24	2.41	2.30	2.19	-		-	-	4.95
12	2.44	-	2.28	2.20	2.27	2.17	2.36	2.24	2.25	2.69	-		4.84	4.49
13	2.37	-	2.25	2.14	2.23	-	2.29	2.26	2.28	2.64	-	2.35		4.98
Mean CMV	2.31	2.49	1.99	2.01	2.09	2.20	2.37	2.27	2.33	2.40	2.11	2.09	2.30	

Table 5 Means of hybrids and lines from BLY group for the two selection indices, SSV above the diagonal and CMV below the diagonal. Data based on the mean of two locations.

Parent	1	2	3	4	5	6	7	8	9	10	Mean SSV
1		4.89	5.63	5.49	4.69	5.51	5.33	5.36	5.64	5.28	5.57
2	0.24		5.15	5.44	4.82	5.22	-	4.77	4.95	5.56	5.11
3	1.28	0.56		5.37	4.81	5.14	4.74	4.73	4.70	5.09	4.99
4	1.23	0.65	1.24		4.85	5.07	5.05	5.14	4.91	4.96	4.96
5	1.13	0.49	1.11	1.22		4.53	5.26	5.28	4.89	5.12	5.14
6	1.26	0.81	1.08	1.03	1.14		4.91	5.05	4.83	5.44	4.91
7	1.30	-	1.23	1.00	1.29	1.23		5.16	5.03	5.30	4.78
8	1.25	1.35	1.23	1.18	1.23	1.19	1.25		-	-	-
9	1.25	0.84	1.28	1.13	1.13	1.34	1.25	-		-	-
10	1.22	0.85	1.24	1.31	1.19	1.24	1.30	-	-		-
Mean CMV	1.21	0.29	0.64	1.34	1.23	1.31	1.24	-	-	-	

The heterosis associated to each parent was also assessed, that is, the mean heterosis of hybrids to which the referred parent was common (Table 6). In the FCV group, considering the SSV index, the largest parental heterosis was in parents *5* and *12*. With the CMV index, the largest heterosis was in both parents *3* and *10*. For the BLY group, again there was no agreement in the heterosis estimates between the indices, i.e. for SSV the largest heterosis estimate was for parents *4* and *7* and for CMV parents *3* and *7*. Several *h* estimates were negative, that is, some hybrids performed worse than the mean of respective parents.

The frequency distribution of the specific heterosis measured as a percentage of the parental mean showed that the variation was high, ranging in the FCV group from -22.4% to 14.8% for the SSV index and from -5.8% to 28% for the CMV index (Figure 1). It is emphasized that the two hybrids with largest heterosis also presented the highest means in this group, i.e. hybrids *1x5* and *1x7*. In the BLY group the heterosis ranged from -12.4% to 7.98% for the SSV index and from -67.8% to 38.2% for the CMV index. In this group, the combinations with largest heterosis did not correspond to the hybrids with the highest mean.

This fact indicates that even when several traits are simultaneously assessed the parents show complementarity in hybrid combinations. Therefore data analysis was performed following the diallel scheme. There was significant difference in GCA for both indices of the two groups. Similar result was obtained for the SCA, except for the SSV index in the FCV group (Table 3). Again, there was no good agreement among the parents in the GCA estimates between the two indices applied. For the FCV group, parents *1* and *3* had the most significant $\hat{g}_i$ estimates for the SSV index. On the other hand, for the CMV index, lines *2* and *10* were superior for this same parameter. In the BLY group,

the lines with largest $\hat{g}_i$ estimates were *1* and 9 for SSV, and *9* and *10* for CMV (Table 6).

Table 6 Mean heterosis and $\hat{g}_i$ effect of each parent of order *i* in the FCV and BLY varietal groups.

		SSV		CMV	
Group	**Parents**	**h_i**	**$\hat{g}_i$[1]**	**h_i**	**$\hat{g}_i$[2]**
FCV	**1**	4.01	0.20	6.90	0.08
	2	-1.25	0.01	5.77	0.13
	3	0.73	0.14	7.38	-0.11
	4	0.35	-0.12	4.50	-0.15
	5	6.79	-0.03	6.07	-0.07
	6	0.42	0.11	5.21	-0.04
	7	4.26	-0.04	6.09	0.07
	8	4.39	0.04	2.33	-0.05
	9	1.29	-0.02	2.03	-0.02
	10	-1.31	-0.14	14.22	0.29
	11	1.45	-0.06	6.80	-0.08
	12	6.95	-0.06	6.91	-0.05
	13	1.52	-0.05	1.24	0.00
BLY	**1**	-0.32	0.37	-5.85	0.02
	2	-0.09	-0.02	-20.65	-0.51
	3	2.07	-0.07	23.77	0.05
	4	3.79	-0.01	-8.06	0.00
	5	-5.22	-0.16	-5.61	-0.02
	6	1.24	-0.08	-4.03	0.06
	7	2.21	0.07	2.88	0.08
	8	-	-0.16	-	0.09
	9	-	0.15	-	0.12
	10	-	-0.10	-	0.11

[1/] $\text{Var}(\hat{g}_i)_{FCV} = 0.00598$; $\text{Var}(\hat{g}_i)_{BLY} = 0.00693$. [2/] $\text{Var}(\hat{g}_i)_{FCV} = 0.00042$; $\text{Var}(\hat{g}_i)_{BLY} = 0.00009$

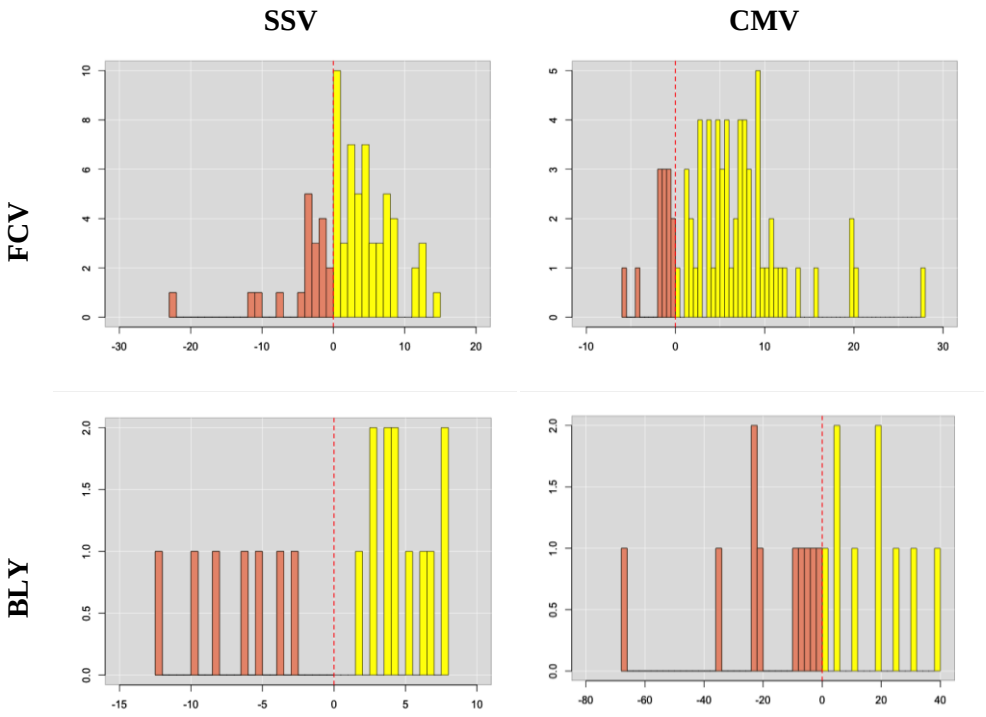

Figure 1 Frequency distribution of the specific heterosis estimates of each hybrid in the FCV and BLY groups. Data presented on the mean of two locations.

We can observe, in the probability distribution of the $\hat{s}_{ij}$ estimates, a variation in complementarity among lines in all cases (Figure 2). In the FCV group, the $\hat{s}_{ij}$ estimates ranged from -1.06 (hybrid *2x11*) to 0.41 (hybrid *5x11*) for the SSV index, and from -0.16 (hybrid *4x5*) to 0.33 (hybrid *10x13*) for the CMV index. In BLY group, the variations ranged from -0.58 (hybrid 1x5) to 0.37 (hybrid 5x7) for SSV index and from -0.39 (hybrid 1x2) to 0.15 (hybrid 2x5) for CMV index.

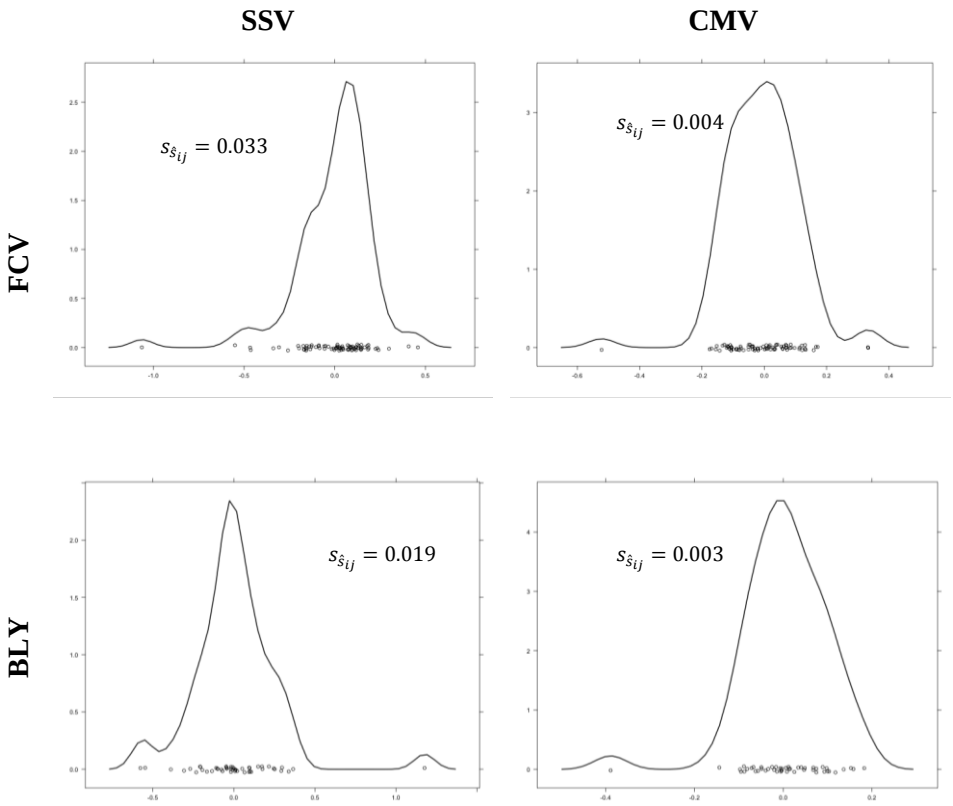

Figure 2 Probability distribution of the $\hat{s}_{ij}$ estimates for the SSV and CMV indices within each vegetable group.

Both indices presented different results, in this context, how to decide which would be the best option for plant breeders? One of the alternatives would be to observe if the correlated response for each trait in the index goes towards the direction intended by the breeder. For example, in tables 7 and 8, for FCV and BLY respectively, we show the five best performing hybrids and the two with the lowest performance for each index. For the SSV index, in all cases the best hybrids had YLD gain in both groups, which is highly desirable and the two worst, reduced expression of this trait. Regarding the GQI, an increase is required, that occurred in three of the five hybrids. For SLR, the objective is to

reduce the value and that was observed in only one case. However, for those that went in the opposite direction, the magnitude was low in percentage. In the case of SUG, in the FCV group, a deviation was estimated in relation to 12%, the intended value, that is, the smaller the deviation the better. There were only two cases in the desired direction, *1x7* and *1x5*. Reduction in ALK was required in both groups so that the selected hybrids were within the ideal range. In this case, for the FCV group, only one hybrid presented alkaloid above the mean. For BLY, the opposite occurred, only one hybrid presented alkaloid below the mean. And finally, the CVS trait, which occurs only in the BLY group, as already mentioned, the minimum expression as possible is required. In this case only one of the superior hybrids in the SSV index was above the mean.

The results of the CMV index, as already pointed out, did not coincide with the SSV index. The hybrids with highest index coefficients were all different from the SSV index. For the FCV group only three of the five best hybrids would have increase in YLD. Moreover, all would have reduced GQI and that in principle would make it impossible to apply this index. In the case of SLR, the result was similar to the SSV index. In the FCV group the result for ALK was in the intended direction for the breeders, but for SUG was in the opposite direction. In the BLY group, only one of the hybrids presented ALK in the desired direction, as occurred in the SSV index. For CVS, all the hybrids presented reduced expression of the trait.

Table 7 Deviation from the mean in the individual traits of the five best and two worst hybrids based on the SSV and CMV indices in the FCV varietal group. Values in blue are above the mean and in red are below the mean.

	Hybrid	Index	YDL	GQI	SLR	SUG	ALK
SSV	1x7	5.45	4.58	19.63	1.16	-3.67	-12.12
	1x5	5.45	1.38	24.52	-6.94	-3.27	-5.03
	1x6	5.39	5.40	1.15	7.00	7.98	-2.36
	3x8	5.36	15.66	-3.52	3.21	1.12	22.79
	1x2	5.36	9.91	-1.00	-0.04	4.21	-26.88
				...			
	1x4	4.43	-25.39	7.83	-12.69	4.10	-5.88
	2x11	3.93	-24.54	-27.43	6.33	3.95	-44.28
CMV	5x10	2.88	-3.83	-16.08	-2.68	6.13	-55.33
	7x10	2.87	8.43	-4.31	3.31	7.53	-40.50
	6x10	2.77	-3.15	-16.73	4.08	10.06	-43.92
	1x2	2.73	9.91	-1.00	-0.04	8.50	-26.88
	2x10	2.70	9.15	-5.30	1.73	11.03	-24.70
				...			
	3x11	2.09	-2.50	-1.18	3.07	19.49	20.96
	4x5	1.97	-10.79	12.44	-7.78	1.47	23.37

Table 8 Deviation from the mean in the individual traits of the five best and two worst hybrids based on the SSV and CMV indices of the BLY varietal group.

	Hybrid	Index	YDL	GQI	SLR	ALK	CVS
SSV	1x9	5.64	6.82	8.84	-2.85	6.87	-68.60
	1x3	5.63	1.71	10.76	-1.40	3.38	-64.55
	2x9	5.56	10.11	1.93	-1.40	-4.37	93.49
	1x6	5.51	2.71	4.29	-2.63	0.35	-66.46
	1x4	5.49	5.21	4.61	-7.80	7.05	-55.42
					...		
	1x4	4.70	-2.97	-0.31	3.21	-2.62	-65.41
	2x11	4.69	-9.91	-2.57	-3.46	2.93	20.41
CMV	3x10	1.34	-11.07	-0.48	-5.19	3.85	-62.46
	6x8	1.34	-7.82	-6.87	1.10	-3.19	-68.50
	4x9	1.31	-6.02	-7.00	-4.08	8.40	-65.30
	5x10	1.31	-10.92	0.50	-4.74	0.57	-61.01
	7x10	1.31	-1.29	2.54	2.77	1.13	-64.01
					...		
	2x5	0.49	0.14	-2.10	-0.07	-3.77	241.21
	1x2	0.24	3.63	-0.24	-5.41	0.39	344.73

DISCUSSION

The objective of any plant breeding program is to obtain cultivars that meet the requirements of farmers, industry and when it is the case, consumers, that is, breeders should observe various traits**.** Under this condition, applying an index would be the most efficient method for simultaneous multi-trait selection (RESENDE; SILVA; AZEVEDO, 2014; YAN; FRÉGEAU-REID, 2008). According to Resende; Silva; Azevedo (2014), a maximum efficiency selection index can be obtained by canonic transformation of the original variables. When

establishing the indices, an important decision is the choice of weights for the trait analysis (YAN; FRÉGEAU-REID, 2008, BERNARDO, 2010). In the case of tobacco, yield is extremely important, because tobacco producers have endeavored to increase their income and one of the forms of reaching this objective is by increasing the quantity of tobacco produced per area. Therefore, the company should present high yielding cultivars so that the grower chooses them, and not the competitors´, so that yield is considered the most important trait.

The GQI is also extremely important because it is directly linked to the quality of the end product. Furthermore, the larger the GQI the higher the price paid to the grower. Thus this index impacts the producer and consumer and therefore high weight should be given to this trait in the index. In the FCV group, the correlation between the most important variables, i.e. YLD and GQI, was not significant, suggesting no association between them (Table 2). This condition is favorable, because it is possible to obtain lines/hybrids with high YLD and GQI with breeding. In the BLY group, positive and significant correlation was observed, but of medium magnitude. Positive association, in this case, is obviously favorable to the breeder, because it means that more productive lines/hybrids tend to be of higher quality, and vice-versa. Selection for one of these two traits contributes in part to improvement in the other trait.

The SLR is undesirable for industry, because only a small percentage is used in manufacturing the industrialized product. However, this trait is considered secondary in selection, mainly because line/hybrids with larger leaves, that are more productive, tend to also have thicker leaf veins for better support. This fact was proven by the positive correlation estimate between SLR and YLD for the FCV group. However, in the BLY group, different from expected, SLR did not correlate with YLD, that is desirable. One hypothesis is

that the better yield of lines was due to the increase in the number of leaves, and not to the size of the leaves.

The requirement for ALK is that the lines/hybrids are within the desirable range (2.5% - 3.5%). It is important to note that this range is an average, since the alkaloid rates can vary depending on the leaf positions in the plant and maturity level in harvesting (DAVIS; NIELSON, 1999). Besides, there is a significant variation of alkaloid rates among the varietal groups, so this range was determined considering FCV and BLY. As most of the lines presented concentration close or above 3.5%, the index was applied to reduce this component. Positive and significant correlation was observed between ALK and GQI. In new cultivars released in Brazil, it is attempted to increase GQI and reduce ALK, so that this correlation is unfavorable to the breeder, but does not prevent lines/hybrids with reduced ALK and high GQI from being obtained, as the magnitude of the correlation estimate was medium. Positive correlation was also reported between ALK and YLD in the BLY group, and was unfavorable for the same reason, that is, increased YLD and decreased ALK are required. It is pointed out that, contrary to the situation in Brazil, in several other countries, e.g. the United States, Venezuela and others, one of the major challenges is increasing YLD, maintaining ALK at acceptable levels, because cultivars tend to present low ALK content when YLD is high, i.e. there is a negative and high correlation between YLD and ALK (Lewis, 2006). In the FCV group, negative correlation was observed between ALK and SLR, which is also unfavorable as it is required a decrease in both. Similarly, as magnitude was low, materials with low ALK and SLR can be obtained.

As already mentioned, SUG is important only for the FCV group because its concentration in the BLY group is very low. This component alters the taste of cigarettes and for this reason was considered in the index. Specialists indicate that 12% would be the ideal concentration for industry for FCV, and

thus the index was applied to reduce deviation based on this concentration. This trait presented positive correlation with IGQ and negative with SLR, both of which are unfavorable to the breeder. However, as the magnitude was also medium it was not limiting.

Finally, the CVS trait only occurred in the BLY group through the conversion of nicotine to nornicotine. Nornicotine is an undesirable component because it is a precursor of nitrosamines (NNN'S), which have harmful effects to human health (CAI et al., 2013; CARVALHO et al., 2014; HECHT, 2003). Thus the index was applied also to reduce expression of enzymes responsible for this conversion. This trait correlated positively and significantly with YLD but with low magnitude. To cater to these commented aspects and the importance of each trait, the weights used were selected as being 40% for YLD, 30% for GQI and 10% for the other traits. Unfortunately, no report was found in the literature where a selection index was used for tobacco to compare the weights adopted in the present study.

Another important aspect when adopting an index is the possibility of proceeding an univariate analysis, e.g. ANOVA, to verify whether the differences among the genotypes using the index are random or not. In this case, for both the indices the ANOVA was carried out and significant differences were observed among the treatments (Table 3). The TxL interactions were also significant. This means that the performance of the treatments did not coincide in the two locations. It is important to point out that there are more than 150,000 tobacco producers in the southern region of Brazil (AFUBRA, 2015), and they differ in several aspects including planting date, soil type and management technology. Consequently it is practically impossible to recommend a cultivar specifically adopted for each grower. The best option to mitigate the effect of the interaction is selection based on the mean of the environments (DUDLEY, 1997; FERREIRA et al., 2015; LIMA et al., 2014). For this reason, it was chosen to

concentrate the discussion of the results based on the mean of the two environments.

The lines chosen of both varietal groups are the product of several years of breeding by the company. It seems that the program has been successful, because in most cases the mean of the treatments of the diallel (H+P), involving the lines and the hybrid combinations among them, was superior to the mean of the checks (C), which are lines available on the market for commercial plantations. When the traits were assessed independently the result converges, that is, for all the traits, the diallel treatments presented means larger than the checks, demonstrating the potential of the lines in the breeding program of the company.

The use of hybrid seeds in tobacco has been adopted by companies to combine phenotypes present in different lines and to introduce male sterility to protect the cultivars developed. Thus, even with low heterosis, the use of hybrids should be encouraged to increase gains. For this diallel crossing was used to show the viability of employing hybrids when selecting for several traits at the same time. Although diallel crosses are widely used in plant breeding in innumerable species, including tobacco (BUTORAC; BELJO; GUNJACA, 2004; MATZINGER; MANN; COCKERHAM, 1962; PANDEYA; DIRKS; POUSHINSKY, 1983; VANDENBERG; MATZINGER, 1970), no report was found of its use, assessing several characteristics simultaneously using an index. The reports present in literature treating about this issue refer to a multivariate diallel analysis aiming at estimating the genetic and environmental correlation between the involved traits, but not to a practical application (LEDO; FERREIRA; RAMALHO, 2003; NDOUMBE; BIEYSSE; CILAS, 2001).

For the FCV group in both indices it was found that even when involving multiple traits simultaneously, the mean heterosis was different from zero. On the other hand, for the BLY group, this difference was not significant

(Table 3). This fact is also in agreement with the results obtained for each trait independently (CARVALHO et al., Chapter 1), that is, in the FCV group it was found that there was heterosis for all the traits assessed, demonstrating the importance of obtaining hybrids in this group. It must be pointed out that the mean heterosis estimate was lower than 6%, since the mean performance of the F_1 of a hybrid is given by: $\bar{F}_1 = \frac{P_1+P_2}{2} + h$, i.e. the hybrid performance depends on the mean of the parents $((P_1 + P_2)/2)$ and the heterosis (h) between them; thus it can be inferred that the hybrid performance depended predominantly on performance *per se* of the lines, which is expected in self-pollinated species (BERNARDO, 2010; RAMALHO et al., 2012). In the case of cross-pollinated species, especially corn, the mean heterosis is generally of high magnitude. However, there is evidence that over time hybrid performance has increased due to the improvement in the performance of the lines because heterosis, in percent, increased much less than the performance *per se* of the lines (TROYER; WELLIN, 2009; LI et al., 2014).

The general combining ability ($\hat{g}_i$) is the function of performance *per se* of the lines (p_i) and also of the mean heterosis of that parent in all the hybrid combinations (h_i), i.e. $\hat{g}_i = h_i + 1/2p_i$ (CRUZ; REGAZZI; CARNEIRO, 2014). For instance, the FCV group included the parents of one of the most productive hybrids for the SSV index, i.e. hybrid *1x5*. For parent *1*, the high $\hat{g}_i$ estimate was due to the parental heterosis (h_i) and also to the performance *per se* (p_i). For parent *5*, h_i was high but the $\hat{g}_i$ was negative, indicating that the performance *per se* of this line was inferior, that can be confirmed in table 4. In the case of the BLY group, as the mean heterosis was almost nil, the $\hat{g}_i$ estimate of the lines was more related to the performance *per se* of the lines than with h_i.

These results show that in the breeding programs for the FCV group, the application of performance *per se* of the parents alone was not sufficient to obtain good hybrids, the lines must be divergent. To assess this divergence, the

hybrid combinations must be assessed in the field. On the other hand, in the BLY group, the results were not the same as those obtained for the FCV group. The hybrid performance shows high association with the line *per se* performance due to the low estimates of heterosis obtained for BLY.

Finally, which criteria should be adopted to choose the best index to select genotypes? In principle accuracy could be used as a criteria. The accuracies of CMV index were higher than the SSV index. This is expected since for computing the canonical variable the genetic quadratic components are maximized over the error (FERREIRA, 2008; RESENDE; SILVA; AZEVEDO, 2014), which results in the accuracy increase and, consequently, the better discrimination among treatments. However, accuracy itself should not be the decisive factor. A good index is the one that enables choice of the best line/cultivar with the most phenotypic expression as possible, in the desired direction, of all the traits involved (Tables 7 and 8). The differences in the results can be better observed in Figures 3 and 4, that show the performance of the best hybrids based on the SSV and CMV indices using standardized data for each variable. In the superior hybrid combinations, in practically all cases, the SSV index was superior towards the direction intended by the breeder in both varietal groups. For example, in the case of the hybrid with the best estimate in the FCV group for the SSV index (*1x7*) the figure shows aspect closer to a "full-ball", and most of the traits had expression above the mean (Figure 3A). For the superior hybrid based on the CMV index (*5x10*), this did not occur, especially for one of the most important traits, GQI. In this case, the performance for this trait was below the general mean. The same finding can be observed for other comparisons among the indices for the superior hybrid combinations.

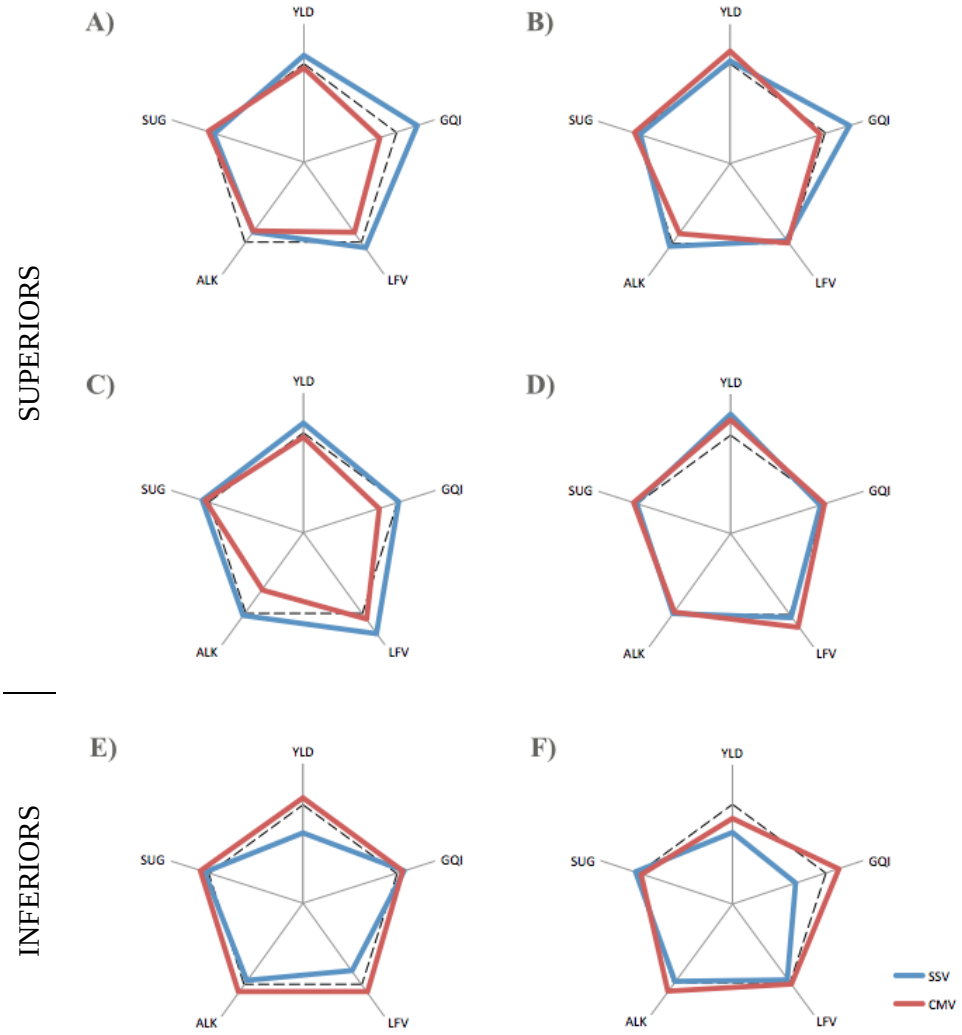

Figure 3 Graph of the standardized variables of the four superior hybrids (A-D) and the two inferior hybrids (E-F) for SSV and CMV indices in FCV group. The dotted line refers to the general mean. A) SSV: hybrid *1x7*; CMV: hybrid *5x10*. B) SSV: hybrid *1x5*; CMV: hybrid *7x10*. C) SSV: hybrid *1x6*; CMV: hybrid *6x10*. D) SSV: hybrid *3x8*; CMV: hybrid *1x2*. E) SSV: hybrid *1x4*; CMV: hybrid *3x11*. F) SSV: hybrid *2x11*; CMV: hybrid *4x5*.

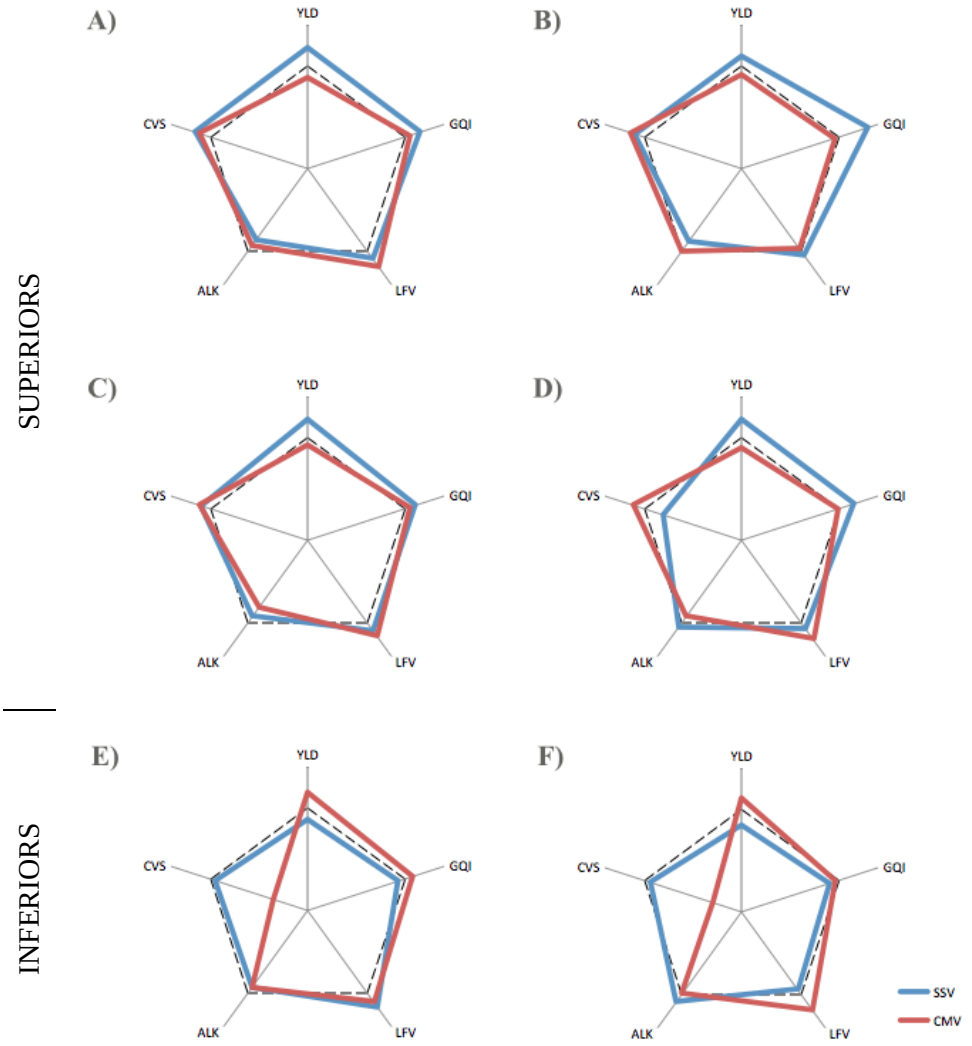

Figure 4 Graph of the standardized variables of the four superior hybrids (A-D) and the two inferior hybrids (E-F) for SSV and CMV indices in BLY group. The dotted line refers to the general mean. A) SSV: hybrid *1x9*; CMV: hybrid *3x10*. B) SSV: hybrid *1x3*; CMV: hybrid *6x8*. C) SSV: hybrid *2x10*; CMV: hybrid *4x10*. D) SSV: hybrid *1x6*; CMV: hybrid *1x7*. E) SSV: hybrid *1x5*; CMV: hybrid *2x5*. F) SSV: hybrid *3x8*; CMV: hybrid *1x2*.

When the two worst hybrid combinations for the SSV index were assessed, the hybrids presented phenotypic expression below the mean of most of the traits, characterizing an "empty ball". In this case, for the CMV index, the results were not in the expected direction, because the two worst hybrids for this index presented expression of several traits above the mean. It could be argued that the response of the traits can be altered if different weights for the traits were used in the construction of the indices. Attributing weights is not an easy task, because the breeder should understand not only the tobacco crop itself, but also all the productive chain, as the weights should reflect the importance of all traits on it. In the present study, the weights were attributed to both the indices, as already mentioned, aiming at the reality of the market as a whole.

Although the CMV index did not present the expected results for application as a selection index, there are other purposes where this index could be used efficiently, for example in selecting a trait using auxiliary variables to increase accuracy, or for genotype clustering (YAN; FRÉGEAU-REID, 2008). The reason why the CMV index did not go towards the direction intended by breeders in these results can be better understood when the correlations are observed between the canonic variable and the traits (Figure 5). As the first eigenvalue of both the groups explains more than 66% of the variation, only the vector corresponding to the first eigenvalue was used to compose the index, and because of this the focus will be on the horizontal positioning in the graphs. For the FCV group, the hybrids with the best estimate in the CMV index correlated positively with YLD and negatively with ALK and SLR, i.e. in the required direction. However, negative correlation was also observed between the CMV index and GQI, i.e. the superior hybrids presented a lower value for GQI and that becomes a limitation to applying this index since this is one of the most important traits.

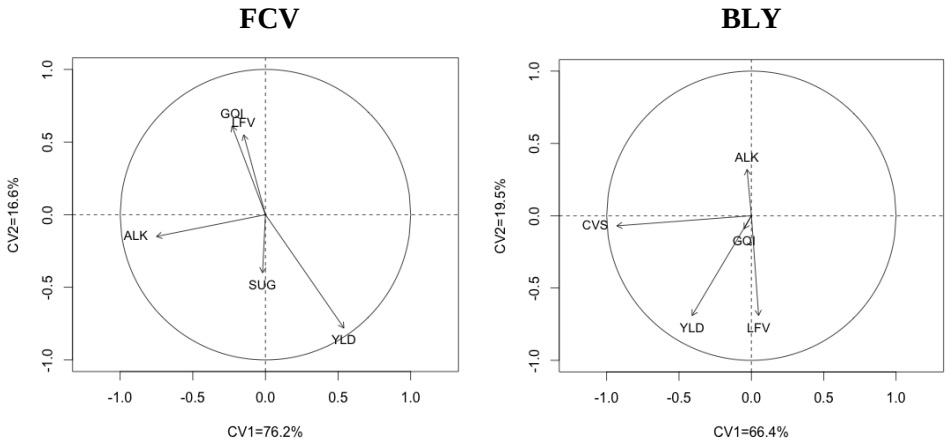

Figure 5 Correlation between the CMV (1 and 2) and the independent variables in the FCV and BLY varietal groups.

In the BLY group, the superior hybrids correlated negatively with CVS and YLD. In the case of CVS, the index is in agreement with the requirements but it would be unacceptable to apply an index that presents negative correlation with YLD, as this trait is the most important in tobacco selection. The other traits presented nil correlation with the CMV index that is also not desirable, because when an index is applied, the aim is to obtain gains for all the traits assessed. Thus, although the CMV index is more accurate, it is clear that the traits are not always selected in the direction required by the breeder and thus it is an index with practical restrictions.

CONCLUSIONS

The hybrids express heterosis for both indices in the FCV group. On the other hand, for the BLY group the mean heterosis is nil. Thus hybrid performance depends mainly on the lines *per se* performance.

The CMV index showed the highest accuracy, however it is practical limited for multi-trait selection. The SSV index, besides of its practically and simplicity in application, presented superiority in hybrid selection for all traits assessed.

Through the application of SSV index it is possible to obtain gains for all traits simultaneously, although they are of lower magnitude than selection for each trait individually.

REFERENCES

AFUBRA. **Associação dos Fumicultores do Brasil**. 2015 Disponível em: <http://www.afubra.com.br>. Acesso em: 6 dez. 2015.

BERNARDO, R. **Breeding for Quantitative Traits in Plants**. 2. ed. Woodbury: Stemma Press, 2010. 390 p.

BERNARDO, R. **Essentials of Plant Breeding.** 1. ed. Woodbury: Stemma Press, 2014. 260 p.

BUTORAC, J.; BELJO, J.; GUNJACA, J. Study of inheritance of some agronomic and morphological traits in burley tobacco by graphic analysis of diallel cross. **Plant Soil Environment,** v. 50, n. 4, p. 162-167, 2004.

CAI, B. (R)-nicotine biosynthesis, metabolism and translocation in tobacco as determined by nicotine demethylase mutants. **Phytochemistry**, v. 95, p. 188–196, 2013

CARVALHO, B. L. et al. Genetic Parameters Estimates Associated to Conversion of Nicotine to Nornicotine in Burley Tobacco. **American Journal of Plant Sciences,** v. 5, p. 3380-3388, 2014.

CRUZ, C. D.; CARNEIRO, P. C. S.; REGAZZI, A. J. **Modelos Biométricos Aplicados ao Melhoramento Genético**. v. 2, 3. ed. Viçosa: UFV, 2014. 668 p.

DAVIS, D. L.; NIELSEN, M. T. **Tobacco: Production, Chemistry and Technology**. 1. ed. Oxford: Blackwell Science Ltd., 1999. 467 p.

FALCONER, D. S.; MACKAY, T. F. C. **Introduction to quantitative genetics**. 4th ed. Malaysia: Pearson, 1996. 464 p.

FERREIRA, R. A. D. C. et al. Implications of the number of years assessment on

HAZEL, L.N. The genetic basis for constructing selection indexes. **Genetics,** v. 28, p. 476–490, 1943.

HECHT, S. S. Tobacco carcinogens, their biomarkers and tobacco-induced cancer. **Nature Review Cancer,** v. 3, p. 733–744, 2003.

LEDO, C. A.; FERREIRA, D. F.; RAMALHO, M. A. P. Análise de variância multivariada para os cruzamentos dialélicos. **Ciência e Agrotecnologia**, v. 27,

n. 6, p. 1214-1221, 2003.

LEWIS, R. S. Identification of Germplasm of Possible Value for Confronting an Unfavorable Inverse Genetic Correlation in Tobacco. Crop Science, v. 46, p. 1764-1771, 2006.

LI, Y. et al. Contributions of Parental Inbreds and Heterosis to Morphology and Yield of Single-Cross Maize Hybrids in China. **Crop Science**, v. 54, p. 76-88, 2014.

LIMA, L.K. et al. Implications of predictable and unpredictable environmental factors in common bean VCU trials in Minas Gerais. **Crop Breeding and Applied Biotechnology,** v. 14, p. 160-165, 2014.

LONG, J. et al. Responses to selection for partial resistance to crown rust in oat. **Crop Science**, v. 46, p. 1260–1265, 2006.

MATZINGER, D. F.; MANN T. J.; COCKERHAM, C. C. Recurrent family selection and correlated response in Nicotiana tabacum L. I.'Dixie Bright 244' 3 'Coker 139.'. **Crop Science**, v. 12, p. 40–43, 1972.

MILLIGAN, S. B.; BALZARINI, M.; WHITE, W. H. Broad-sense heritabilities, genetic correlations, and selection indices for sugarcane borer resistance and their relation to yield loss. **Crop Science**, v. 43, p. 1729–1735, 2003.

NDOUMBE, M.; BIEYSSE, D.; CILAS, C. Multi-trait selection in a diallel crossing scheme of cocoa. **Plant Breeding,** v. 120, p. 365-367, 2001.

PANDEYA, R. S.; DIRKS, V. A.; POUSHINSKY, G. Quantitative genetic studies in flue-cured tobacco (Nicotiana tabacum). I. Agronomic characters. **Canadian Journal of Genetics and Cytology**, v. 25, p. 336–345, 1983

RAMALHO, M. A. P. et al. **Aplicações da Genética Quantitativa no Melhoramento de Plantas Autógamas**. 1. ed. Lavras: UFLA, 2012a. 522p.

RESENDE, M. D. V.; SILVA, F. F. ; AZEVEDO, C. F. **Estatística Matemática, Biométrica e Computacional**. 1. ed. Viçosa: UFV, 2014. 881 p.

SHARMA, R. C.; DUVEILLER, E. Selection index for improv- ing helminthosporium leaf blight resistance, maturity, and kernel weight in spring wheat. **Crop Science**, v. 43, p. 2031–2036, 2003

SIMMONDS, N.; SMARTT J. **Principles of crop improvement**. 2. ed. Oxford, UK: Blackwell Science Ltd. Press, 1999. 424 p.

SMITH, H. F. A discriminant function for plant selection. In: **Papers on Quantitative Genetics and Related Topics**. Department of Genetics, North Carolina State College, Raleigh, North Carolina, 1936. p. 466-476.

STEEL, R. G. D.; TORRIE, J. H.; DICKEY, D. A. **Principles and Procedures of Statistics: A Biometrical Approach**. 3. ed. New York: McGraw-Hill, 1996. 672 p.

TROYER, A. F.; WELLIN, E. J. Heterosis decreasing in hybrids: Yield test inbreds. **Crop Science**, v. 49, p. 1969–1976, 2009.

VANDENBERG, P.; MATZINGER, D. F. Genetic Diversity and Heterosis in Nicotiana. III. Crosses among Tobacco Introductions and Flue-Cured Varieties. **Crop Science,** v. 10, p. 437-440, 1970.

WHITE, F. H.; PANDEYA, R. S.; DIRKS, V. A. Correlation studies among and between agronomic, chemical, physical and smoke characteristics in flue-cured tobacco (Nicotiana tabacum L.). **Canadian Jourbal of Plant Science**, v. 59, p. 111–120, 1979.

YAN, W.; FRÉGEAU-REID, J. Breeding Line Selection Based on Multiple Traits. **Crop Science,** v. 48, p. 417-423, 2008.

ARTICLE 3 Adding genome-wide information in a breeding program of flue-cured Virginia tobacco

Adding genome-wide information in a breeding program of flue-cured Virginia tobacco

Bruna Line Carvalho, Carlos Eduardo Pulcinelli, Ramsey Lewis, Adriano Teodoro Bruzi, Pedranne Barbosa, Christiano Simões, Magno Antonio Patto Ramalho

ABSTRACT

The advance and the cost reduction of the sequencing techniques have brought a perspective for applying genome-wide selection (GWS) in the breeding programs. Simulation studies have show that we can increase our gain per time using the GWS. But there still the need to test it in real data. No reports were published for tobacco crop in this matter. In this way, we aim to test the applicability of GWS in an established tobacco breeding program for: i) Verify the possibility to validate a model that could be use in future generation for selection without the phenotype; ii) Check whether the prediction of hybrids using GWS is feasible; iii) Correlate genetic divergence with hybrids performance and heterosis. For this we use superior 13 lines of Flue-Cured Virginia varietal group, they were crossed in a diallelic scheme obtaining 53 hybrids. The lines, hybrids and 15 checks were evaluated in the field, with 3 replications, in 2 locations. Traits assessed were yield (YLD), general quality index (GQI), steams by leaf lamina ratio (SLR), total alkaloid content (ALK) and total sugar content (SUG). GBS was used for sequence the genotypes. We tested three methods of prediction: rr-BLUP (BRR), Bayes B (BB) and Bayes Lasso (BL), considering just the additive effects in the model. Validation was performed with levels of missing information ranging from 10% to 90%. For predicting hybrids we used just the genotype and phenotype from the lines. We found that genetic divergence among lines were not associated to hybrid

performance or heterosis. The genotype x environment affects the GWS, however this information associated to the phenotype improves the accuracy of selection. The prediction of hybrids by GWS can be used in the sense of excluding not favorable combinations, so resources are focused in testing just promising hybrids. At last, genomic information can be used for complementing phenotypic information and increase the accuracy of selection.

Keywords: Plant breeding; Quantitative Genetics; Hybrid prediction; Genomic Selection; Genotype-by-Sequencing

INTRODUCTION

The plant breeding scenarios assume that selection of individuals or progenies should be based on genetic merit (HILL, 2014). For this, in recent decades the use of molecular markers to estimate it has opened up a new perspective of genetic breeding that has caught the attention of thousands of researchers worldwide (HESLOT; JANNINK; SORRELS, 2015). To have faster and efficient genetic progress, predictions must be the most accurate as possible. In this context, since differences to be detected among the genotypes are ever smaller, the molecular marker can be used to help the breeder in selection.

The first attempts to incorporate marker information in predictive models was based on the theory that markers could be identified associated to QTLs of large effect, that is known as marker assisted selection (MAS) (COOPER; PODLICH; SMITH, 2005; LANDE; THOMPSON, 1990; PODLICH; WINKLER; COOPER, 2004; SOLLER; PLOTKIN-HAZAN, 1977). This led to the discovery of some genes associated to genetic variations of traits. However, as most of the traits of economic interest are affected by hundreds of genes, the effect of each QTL on the trait is small (BUCKLER et al., 2009). Thus this technique has only been used effectively for traits controlled by a few genes and of difficult phenotypic assessment (XU; CROUCH, 2008). Furthermore, for applying MAS a mapping population has to be developed to identify the QTLs, that requires many additional resources applied to the breeding program. So the practical impact of MAS on breeding programs has been much less than first thought (DEKKERS, 2004; HESLOT; JANNINK; SORRELS, 2015; MOREAU; CHARCOSSET; GALLAIS, 2004; BERNARDO 2008).

With the advance and cost reduction of sequencing technologies, it has been possible to identify thousands and even millions of SNPs in the various

species (HEFFNER et al., 2010; MASSMAN et al., 2013; SEBASTIAN et al., 2010; SPINDEL et al., 2015; THAVAMANIKUMAR; DOLFERU; THUMMA, 2015). Therefore it is expected that some SNPs will be close or within coding regions and therefore in linkage disequilibrium with them. Consequently some markers are associated to the response of the QTL and could be used for selection without the need to establish the linkage phase. Thus the concept of genome-wide selection arose (GWS) (MEUWISSEN; HAYS; GODDARD, 2001), that basically consists of multiple regression of the phenotypes on all the markers available using a linear model (LORENZ, 2013). As all the markers are used, there is much less risk of not considering small effect QLTs in the predictive model (GUO et al., 2014) and consequently the accuracy of the genotype values is increased (BERNARDO; YU, 2007; MASSMAN et al., 2013; ENDELMAN et al., 2014; HESLOT; JANNINK; SORRELLS, 2015).

As number of parameters to be estimated in GWS, that is, the effect of the markers, is much higher than the number of phenotypic observations, there are no degrees of freedom to jointly adjust all effects by the least-squares method (LANDE; THOMPSON, 1990). Hence shrinkage-type procedures are adopted, such as mixed models (BLUP - Best Linear Unbiased Predictor) or Bayesian statistics to estimate the effects of the markers. Several models have been proposed, which differ in the restrictions imposed on the SNP distribution (GIANOLA et al., 2009; GIANOLA, 2013; HABIER et al., 2011; KOLLER; FRIEDMAN, 2009; LORENZ et al., 2011; PARK; CASELLA, 2008; MEUWISSEN; HAYS; GODDARD, 2001; VANRADEN, 2008;). The choice of which to use depends on multiple factors that include the genetic architecture of the trait, the marker density, sample size and so on (DE LOS CAMPOS et al., 2013).

In tobacco, this approach has not yet been explored. There are some studies in the literature associating markers to traits like diseases and yield

components (EIKHOLT; LEWIS, 2014; VIJAY; LEWIS, 2012; XIAO et al., 2013; YI et al., 1998), but reports were not found of GWS applied to this crop. It is important to emphasize that in Brazil tobacco is typically produced on small familiar farms in three states in the South of the country, that differ in many aspects, such as type of management, soil, climate and technology use. Since this makes phenotype assessment even more difficult in all the conditions imposed, GWS could be a useful tool to help the breeder choose the best lines/hybrids.

Another approach is hybrid prediction using GWS (SU et al., 2012; RIEDELSHEIMER; TECHNOW; MELCHINGER, 2012). As the tobacco seed companies already make manual crosses to combine phenotypes present in different lines and to introduce male sterility, the use of hybrids tends to increase in this species (CARVALHO et al., Article 1). As the number of lines obtained annually is increasing, the number of potential hybrid combinations is very large. In this situation GWS could be used to identify superior hybrid combinations from genetic and phenotypic information of the parent lines. The success of this strategy could open new perspectives on the exploitation of heterosis in tobacco even when it is of small magnitude.

Thus the objectives of the present study were: i) estimate the accuracy in predicting genotypic values by GWS, based on several prediction models; ii) identify whether the genotype x environment interaction (GxE) affects the breeding value using GWS; iii) propose a strategy to use GWS as an auxiliary tool in the process of line/hybrid selection, iv) estimate whether the genetic divergence between lines are associated with heterosis or the performance of the hybrids derived from them; v) predict hybrid combinations from the genetic and phenotypic value of the parent lines.

MATERIALS AND METHODS

The data used in the present study were supplied by the Souza Cruz S.A company, a subsidiary of British American Tobacco (BAT), and consisted of 13 tobacco lines of the Flue-Cured Virginia (FCV) varietal group, that were crossed in a diallel design and 53 hybrid combinations were obtained. The lines used in the crosses were selected due to good characteristics for physical and chemical composition of the leaves, plant morphology and high yield.

Phenotyping and Genotyping

The lines, hybrids and 15 checks were assessed as described in Chapter 1. The trails were installed in two locations. The following traits were assessed: YLD, GQI, SLR, SUG and ALK. Details about data collection are also described in Chapter 1. The sum of the standardized variables index (INDEX) was also estimated as proposed in Chapter 2.

Samples of young leaf tissue from one plant of each treatment were used to extract DNA. The procedure was carried out in the Molecular Biology Laboratory located in the Souza Cruz company, Rio Negro, PR, following the CTAB protocol, using an extraction kit. The samples were sequenced by the genotyping-by-sequencing technique (GBS) (ELSHIRE et al., 2011; POLAND et al., 2012) in the same lab. A total of 1193 SNPs were obtained. To control the data quality, markers with minor allele frequency (MAF) lower than 5% and markers with more than 70% of missing data were discarded. After processing, 861 SNP markers well distributed in the chromosomes were obtained to analyze the data. Imputation of missing data was performed using the mean frequency of the allele in the population. The SNP set of each individual represents its genetic value.

Statistical analyses

Analysis of variance was performed for all traits individually, by location and across locations, as described in Chapter 1. Adjusted means obtained were used for analysis described as following. Data fit was performed using the regression of phenotypes on the markers in each environment and later using the mean of the phenotypes of both environments in the marker regression. Three prediction methods were implemented: rr-BLUP (BRR), Bayes B (BB) and Bayes Lasso (BL), using the R-package BGLR (DE LOS CAMPOS; PÉREZ, 2015).

The equation used for the variable responses (*y*) is represented by: $y_i = \eta_i + \varepsilon_i$, where y_i is the observed mean of each treatment *i*; η_i is the linear predictor for each treatment *i*, i.e. the expectation of y_i given the predictors; and ε_i are the residues associated with the estimates, $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$. The estimated value is considered as the real genetic value of the individual, thus error variance (σ_ε^2) is equal to 1. The linear predictor is a function of the conditional expected value, given by:

$$\eta = 1\mu + \sum X\beta \quad (1)$$

where μ is the intercept, in this case the general mean; X is the incidence matrix of predictors *j*, i.e. the markers; β is the vector of the genetic effects treatments associated to each marker *j*.

In the BRR method, the effects of the hybrids are considered random and the variance for each locus is equal to $\sigma_{gi}^2 = \sigma_g^2/k$; where σ_g^2 is the total genetic variance and *k* is the number of markers.

The BB model is similar to that of the BRR model, except that the variance associated to the markers can vary according to the marker. The *a priori* distribution and data distribution are used to estimate the markers. Since

the majority of lo ci present null variance, i.e. they do not segregate, and just a few loci have genetic variance, the BB method considers the following *priori*'s for the marker effects:

$\sigma_{gi}^2 = 0$ with probability π,

$\sigma_{gi}^2 \sim \chi^{-2}(\nu, S)$ with probability (1- π),

where ν and S are degrees of freedom and scale parameter of the distribution, respectively, given that $\sigma_{gi}^2 > 0$. The adopted value were of ν = 4 and π = 0.95. The scale parameter was defined using the following estimator, according to Ober et al. (2011):

$$S = \frac{(\nu - 2)\sigma_g^2}{(1-\pi)\nu \sum_{j=1}^{S} 2p_j(1-p_j)}$$

The BL model is equivalent to the BB model but the π value is considered unknown, and the following *priori* was considered $\pi \sim Beta(p_0, \pi_0)$, in which $p_0 > 0$ and represents the sum of *priori*'s ("success" *priori* + "fail" *priori*), it was considered p_0=2; and $\pi_0 \in [0,1]$, that in the present study was considered the *priori* $\pi_0 = 0.5$.

For each analysis the number of Gibbs sampler interactions was 10000, 500 samples were discarded, and the amount used to estimate the *a posteriori* mean value was 5.

The means of the marker effects ($\hat{\beta}$) were extracted from each model, and the predicted genotypic values ($\hat{\eta}$) were estimated based on the sum of all the markers effects for each treatment (line/hybrid) as exposed in equation 1. Fitting jointly all markers ensures that the estimated effects are not biased, thus all the effects are captured and a multiple test is not necessary. The predictive accuracies (r_{GP}) were estimated by the correlation between the $\hat{\eta}$ and the mean values of observed phenotypes. The BRR method was selected for the latter

analyses because it has accuracy similar to the other methods, and it is simpler and quicker to compute.

The predicted genetic values ($\hat{\eta}$) and the observed phenotypic values ($\hat{y}$) were submitted to the analysis of variance considering the means of each location as replication. The percentage of the variation attributed to each source of variation (R^2) was determined from the ratio between the sum of squares (SS) of the source of variation in question and the total SS. The repeatability (r^2) was also determined of each model as follows: $r^2 = 1 - 1/F$, where F is the value of the Snedocors' F test for treatments (TREAT) obtained by analysis of variance.

Both values, $\hat{\eta}$ and $\hat{y}$, were also submitted to diallel analysis using the model by Griffing, method II (1956), by the least-squares method, to estimate the effects of the general combining ability (GCA) and the specific combining ability (SCA). The following model was adopted (CRUZ; CARNEIRO; REGAZZI, 2014):

$$y_{ik} = m \quad + g_i \quad + g_k \quad + s_{ik} + \quad \bar{e}_{ik.}$$

where y_{ik} is the observation of the hybrid between the parents *i* and *k*; m is a constant, that in this case represents the general mean of the diallel treatments; g_i and g_k are the effects of the general combining ability of the *i*-eth and *k*-eth parent, respectively; s_{ik} is the effect of the specific combining ability for the cross between the parents *i* and *k*; and $\bar{e}_{ik.}$ is the mean experimental error, where $e \cap N(0, \sigma_e^2)$.

To evaluate prediction accuracy the *k-fold* cross-validation was used which consists of partitioning of the population/treatments in two data sets: one used to train the model (TRN) and the other used to test the model (TST). In the present study, several proportions of TST were used in relation to the total number of treatments, ranging from 10% to 90%. A hundred random samples

from treatments were made for each proportion of missing values and for each trait.

Another strategy adopted was predicting hybrids considering only the genetic (SNP) and phenotypic information of the lines. In this context, the SNP value of each hybrid was determined by the combination of the values of the respective parental lines. Thus the prediction model was trained using the information from lines and then this model was applied to predict the hybrids. Accuracy was determined by correlating the predicted genotype values and the observed phenotype values. The regression equation between these values was also estimated.

The divergence was determined between lines by estimating the genetic distance pairwise using the Jaccard method (1908). The results were placed in a graph using the Neighbor Joining Tree method (SAITOU; NEI, 1987). Later the Pearson correlation was estimated between the genetic distances and the hybrid performance and the respective heterosis.

RESULTS

The accuracy estimates were obtained by the Pearson correlation between the predicted genotypic values and the observed phenotypic values (r_{GP}). The three methods used to estimate the effects of the markers, i.e. rr-BLUP (BRR), Bayes B (BB) and Bayes Lasso (BL) gave similar results. The estimates obtained from the average of the two location were larger than those obtained from each environment independently. Among the traits, the variation of correlation estimates was not high, except for the selection index (Table 1). As these three prediction methods presented similar accuracy, the BRR was chosen to estimate the marker effects, because it is simpler, faster and robust (WIMMER et al., 2013). The marker effects did not coincide between the

environments, reflecting the interaction GxE (Figure 1). Only for SUG and ALK the markers of the large effect coincided in both environments.

Table 1 Correlation between the genotypic predicted values and the observed phenotypic values (r_{GP}) for the traits YLD, GQI, SLR, SUG, ALK and INDEX, considering three prediction methods, rr-BLUP (BRR), Bayes B (BB) and Bayes Lasso (BL). Data obtained in each environment and across environments.

ENV	TRAIT	BRR	BB	BL
1	YLD	0.7639	0.7552	0.7174
	GQI	0.7608	0.7801	0.7511
	SLR	0.8849	0.8966	0.8592
	SUG	0.7230	0.7534	0.5857
	ALK	0.9018	0.8916	0.8943
	INDEX	0.6237	0.5936	0.5184
2	YLD	0.767	0.7302	0.7038
	GQI	0.8165	0.8205	0.8189
	SLR	0.8853	0.9063	0.8787
	SUG	0.7201	0.7081	0.5851
	ALK	0.9033	0.9196	0.8967
	INDEX	0.7848	0.8133	0.6304
MEAN	YLD	0.8548	0.7590	0.8197
	GQI	0.8525	0.7927	0.8242
	SLR	0.9307	0.9011	0.9237
	SUG	0.8166	0.7481	0.6600
	ALK	0.9261	0.9136	0.9252
	INDEX	0.7411	0.8381	0.6417

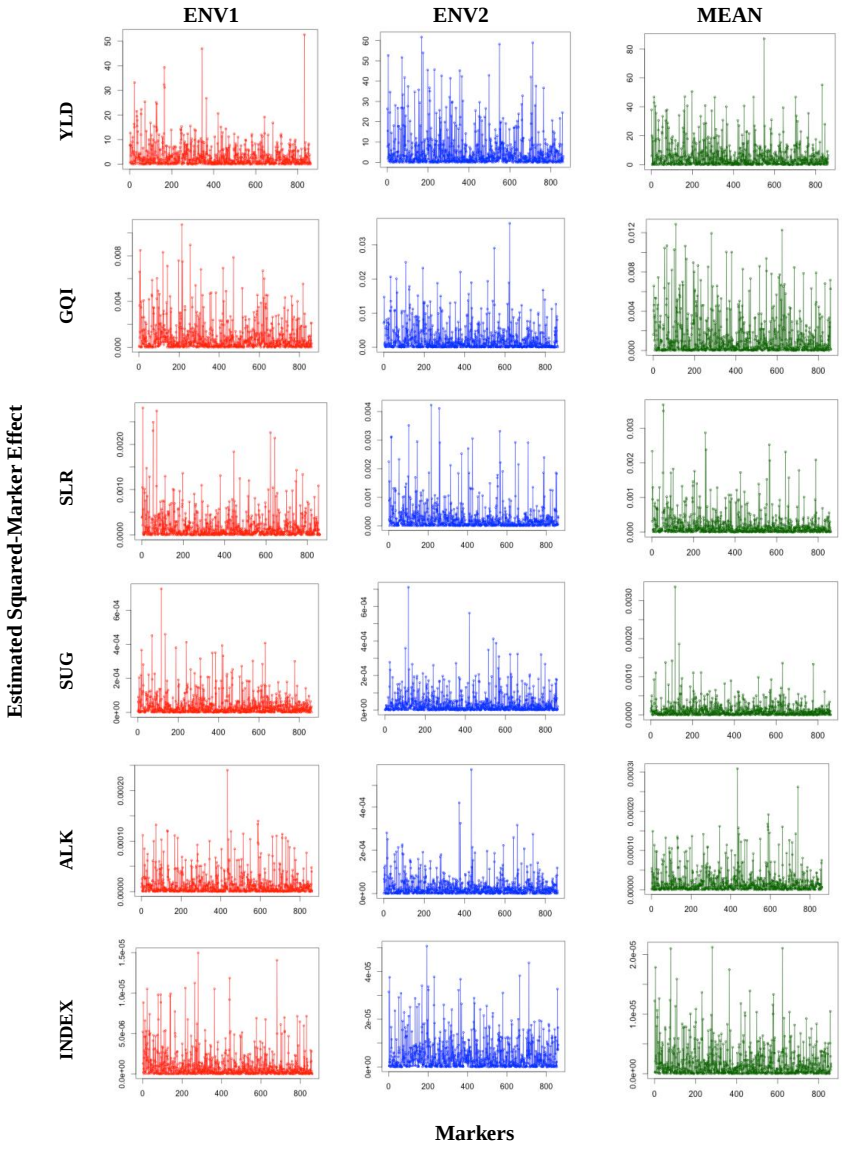

Figure 1 Estimated squared-marker effects of the 861 SNPs based on phenotypes of each environment (ENV1 and ENV2) and in the mean of the environments (MEAN) for the traits YLD (kg/ha), GQI (%), SLR (%), SUG (%), ALK (%) and INDEX.

The summary of the analysis of variance of the phenotypic (P) and genotypic (G) values (Table 2) showed that the contribution of TxE to the total variation was higher considering the phenotypic values compared to the genotypic values. For some traits the contribution of the source of variation TREAT in relation to the total variation was larger for the phenotypic data while for other traits it was larger for the genotypic data (G). However, since the interaction was lower for G, the accuracy $(\sqrt{r^2})$ of the model was larger in all cases.

Table 2 Summary of the analysis of variance based on the phenotypic (P) and genomic (G) data of the traits YLD (kg/ha), GQI (%), SLR (%), SUG (%), ALK (%) and INDEX. Data from two locations.

TRAIT	S.V.	TREAT (T)		ENV (E)		TxE		r^2**
		QM	R^2*	QM	R^2	QM	R^2	
YLD	P	214733	52.0	7081034	21.4	109653	26.6	0.6995
	G	47648	31.6	7714156	64.0	6601	4.4	0.9281
GQI	P	58.4	57.1	1585.1	19.4	24	23.5	0.7673
	G	20.2	47.7	1465.1	43.4	3.7	9.9	0.9022
SLR	P	3.47	72.7	1.32	0.3	1.28	26.9	0.7934
	G	1.66	81.7	1.38	0.9	0.36	17.5	0.8866
SUG	P	2.78	61.7	59.69	16.6	0.98	21.7	0.8053
	G	0.29	26.8	59.12	67.5	0.06	5.7	0.8867
ALK	P	0.54	82.6	3.88	7.4	0.07	10.0	0.9376
	G	0.31	81.1	4.22	13.9	0.02	5.0	0.9688
INDEX	P	0.1418	55.6	0.0568	0.3	0.1125	44.1	0.4543
	G	0.0222	62.3	0.0135	0.5	0.0133	37.2	0.6351

* Coefficient of determination; ** Repeatability

Another alternative to check the accuracy of the genotypic value in predicting the phenotype is the percentage of coincidence in the rank of the best treatments considering the two alternatives, G and P data (Table 3). The coincidence indices (CI) in rank varied among the traits, and were higher for

ALK (88.6%) and smallest for GQI (43.2%). Compared to the correlations presented in Table 1, there was relatively good agreement, although the accuracy classification considering the mean of the environments and the coincidence percentage was not perfect. It should be emphasized that the r_{GP} involves all the treatments, not only the 10 best, as the CI was estimated. It is understood that the genotypic and phenotypic values produce very similar results when all the treatments are considered in the predictive model.

Table 3 Ten treatments considered superior for the traits YLD, GQI, SLR, SUG, ALK and INDEX identified by the observed phenotype (P) and the genotypic predicted value (G). Values in the mean of the environments.

YLD		GQI		SLR		SUG		ALK		INDEX	
P	G	P	G	P	G	P	G	P	G	P	G
19	19	74	75	19	19	49	66	65	41	59	75
59	59	75	74	59	59	65	62	50	49	19	71
81	81	54	49	77	77	66	71	41	50	75	67
77	77	17	63	81	81	69	75	49	17	41	19
58	76	9	2	38	12	9	49	53	1	71	59
40	71	2	41	58	78	71	69	9	53	69	55
78	67	69	66	78	26	62	2	17	9	40	41
80	58	71	71	12	38	53	65	25	45	17	49
38	75	53	55	26	4	75	72	45	25	38	53
76	55	65	3	4	80	2	6	1	32	55	42
54.5%[1]		43.2%		77.3%		65.9%		88.6%		54.5%	

[1] Coincidence index - CI (%) obtained by CI=(C-R)/(S-R), where R is the number of progenies selected by random from the 10, i.e. 1; C is the number of progenies selected coincident for P and G; and S is the number of selected in each condition (10).

The effect of the GxE interaction in the prediction made by the GWS is a point to be questioned. Two circumstances are presented to answer this question: i) Correlation between the predicted genotypic value considering the mean of the environments and observed phenotypic value in each environment;

ii) Correlation of the genotypic value based on one location and the observed phenotypic value based on another location (Table 4). The r_{GP} estimates were low for the second case, that is, the predicted genotype value for one environment different from where the phenotype was assessed had low accuracy for most traits, except for ALK. These results are coherent with the estimates of the contributions of the interactions to the total variation (R^2), excluding the effect of error, for all the traits (Table 2), that is, the interaction corresponds to one quarter of the total phenotypic variation and therefore if data are not used from the environments where it is intended to make selection to "train" the model, the estimates will not be accurate.

Table 4 Accuracy of the prediction model for three approaches: i) Validation of the model by the mean of the environments (MEAN) for prediction in environment 1 (ENV1); ii) Validation of the model by MEAN for prediction in environment 2 (ENV2); iii) Validation of the model by the ENV1 for prediction in ENV2.

TRAITS	MEAN > ENV1	MEAN > ENV2	ENV1 > ENV2
YLD	0.6930	0.6715	0.4458
GQI	0.6699	0.7570	0.5857
SLR	0.6995	0.7920	0.5727
SUG	0.6222	0.6338	0.4351
ALK	0.8511	0.8634	0.8258
INDEX	0.3819	0.6759	0.2220

An important point to be assessed is the accuracy of the phenotype predictions of individuals that were not considered for training the model. For this the *k-fold* cross-validation was used with several percentage of missing observation, that is, with different treatment numbers removed from the population used in the validation process. Accuracy tended to decrease when the number of missing increased (Figure 2). This showed that when there are fewer numbers of treatments in the training population (TRN), i.e. phenotypically

assessed, the predictions of the testing population (TST) will be worse, i.e. individuals that were not assessed phenotypically.

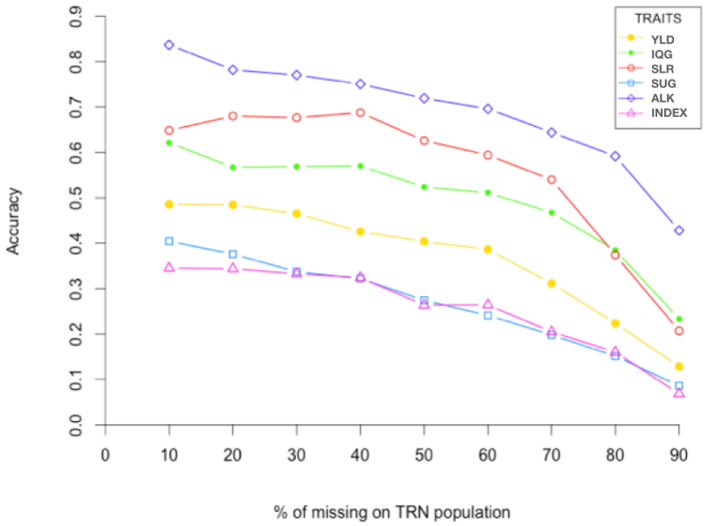

Figure 2 Accuracy of the cross validation process considering different percentages of missing values from the training population (TRN) and used for test the model (TST), for the traits YLD, GQI, SLR, SUG, ALK and INDEX. Mean of 100 replications.

Since the lines were pair-wise crossed in practically all combinations, the GCA and SCA were estimated of the predicted genetic values (G) and observed phenotype values (P). The lines differed regarding the GCA of the G and P values (Table 5). The ratios between the sum of squares (SS) of the phenotypic GCA and SS of the SCA were close to unit for most of the traits, indicating that GCA and SCA contributed equally to the total genetic variation. On the other hand, considering the genotypic values, the ratio between SS of GCA and SS of SCA was >4.7, indicating the predominance of additive effects. The correlations between the genotypic and phenotypic GCA (r_{GCA}) were positive and high for all traits (r > 0.9). The correlations between the genotypic and phenotypic SCA (r_{SCA}), although smaller than r_{GCA}, were also positive and significant (r >0.6).

Table 5 Estimates of general combining ability of the lines from phenotypic (P) and genotypic (G) data; sum of squares (SS) of GCA and SCA; proportion of GCA to SCA variation (SS GCA / SS SCA); correlation between phenotypic and genotypic GCA (r_{GCA}), and between phenotypic and genotypic SCA (r_{SCA}).

Traits	**YLD**		**GQI**		**SLR**		**SUG**		**ALK**		**INDEX**	
Lines	**P**	**G**	**P**	**G**	**P**	**G**	**P**	**G**	**P**	**G**	**P**	**G**
1	-17.81	-38.03	6.94	3.22	-0.57	-0.28	1.74	0.79	-0.21	-0.13	0.09	0.06
2	176.90	193.07	-3.18	-2.95	0.22	0.13	-0.18	-0.21	-0.24	-0.09	0.13	0.02
3	123.48	89.76	0.29	0.99	0.76	0.43	-0.23	0.01	0.40	0.33	-0.11	0.07
4	-240.97	-206.79	1.32	1.56	-1.22	-1.30	0.00	0.13	0.21	0.15	-0.15	-0.07
5	-136.19	-252.19	2.67	2.12	-1.25	-1.42	0.98	0.78	0.07	-0.02	-0.06	-0.12
6	64.42	80.03	1.19	-0.05	0.47	0.43	-0.32	-0.33	0.17	0.11	-0.04	0.05
7	4.92	19.03	-1.46	-0.29	0.57	0.49	-0.34	-0.24	-0.19	-0.13	0.07	0.03
8	91.33	81.49	-1.25	-0.70	0.61	0.67	-0.24	-0.10	0.22	0.25	-0.05	0.05
9	-36.64	-28.72	-0.32	-0.42	0.11	0.17	0.15	0.03	0.04	0.00	-0.02	-0.01
10	36.73	50.70	-3.86	-2.03	0.48	0.29	-0.93	-0.66	-0.77	-0.53	0.29	-0.05
11	-15.18	-13.58	-1.73	-0.86	0.56	0.38	-0.85	-0.65	0.20	0.31	-0.08	-0.01
12	17.49	-40.48	-0.99	-1.22	-0.45	-0.12	0.34	0.16	0.15	0.04	-0.05	-0.06
13	-68.49	-51.01	0.39	0.66	-0.28	-0.14	-0.13	0.01	-0.05	0.04	0.00	-0.01
SS GCA	3078737	1849139	634	389	69.8	51.7	42.1	23.8	9.85	7.51	1.41	0.50
SS SCA	2872613	387704	667	44	30.5	8.1	46.7	4.8	4.93	0.75	2.36	0.10
SS GCA/SS SCA	1.07	4.77	0.95	8.77	2.29	6.41	0.90	4.92	2.00	10.07	0.60	4.81
r_{GCA}	0.97		0.91		0.97		0.96		0.99		0.91	
r_{SCA}	0.81		0.74		0.84		0.87		0.82		0.64	

One of the advantages claimed in the use of GWS is the ability to predict the hybrid performance from the genetic and phenotypic value of the parent lines. In this case genotyping and phenotyping of hybrid would not be necessary, because the genotype would be estimated from the SNP combination of the respective parent lines, and the genotypic value would be predicted using the model validated based on the phenotypic information of the lines. Figure 3 shows the fit of predicted genotypic value to the observed phenotypic of hybrids using this approach. The blue line represents the linear regression equation obtained. The ideal would be obtaining an inclination (*b*) close to 1, which means that the genotypic values are similar to phenotypic observation. But the estimated *b* values were less than 0.30. Furthermore, the equation obtained was not sufficient to explain the variation in the data, especially for GQI, SUG and INDEX. This fact is proved by the estimates of the coefficient of determination (R^2). In the case of YLD, for example, only 1/3 of the phenotypic variation of the hybrids was explained by the predicted genotypic value. That is, hybrid performance cannot be predicted from training the model only with data from the lines.

In the present study both lines and hybrids were sequenced, so the association between the hybrid genetic value obtained from the sequencing and from the combination of the SNP's of the parent lines could be estimated. As was expected, the correlation estimate was high and positive (r = 0.86), although it was not equal to unit.

The SNP information was also used to calculate the genetic distance among lines pairwise (Figure 4). It was observed that some line pairs presented large divergence and the largest was between the pair *6* and *13*. The correlations between genetic distance among lines and the respective hybrid performances and heterosis were estimated. Contrary to the expected, it was observed in all the cases that the values were practically nil (Table 6).

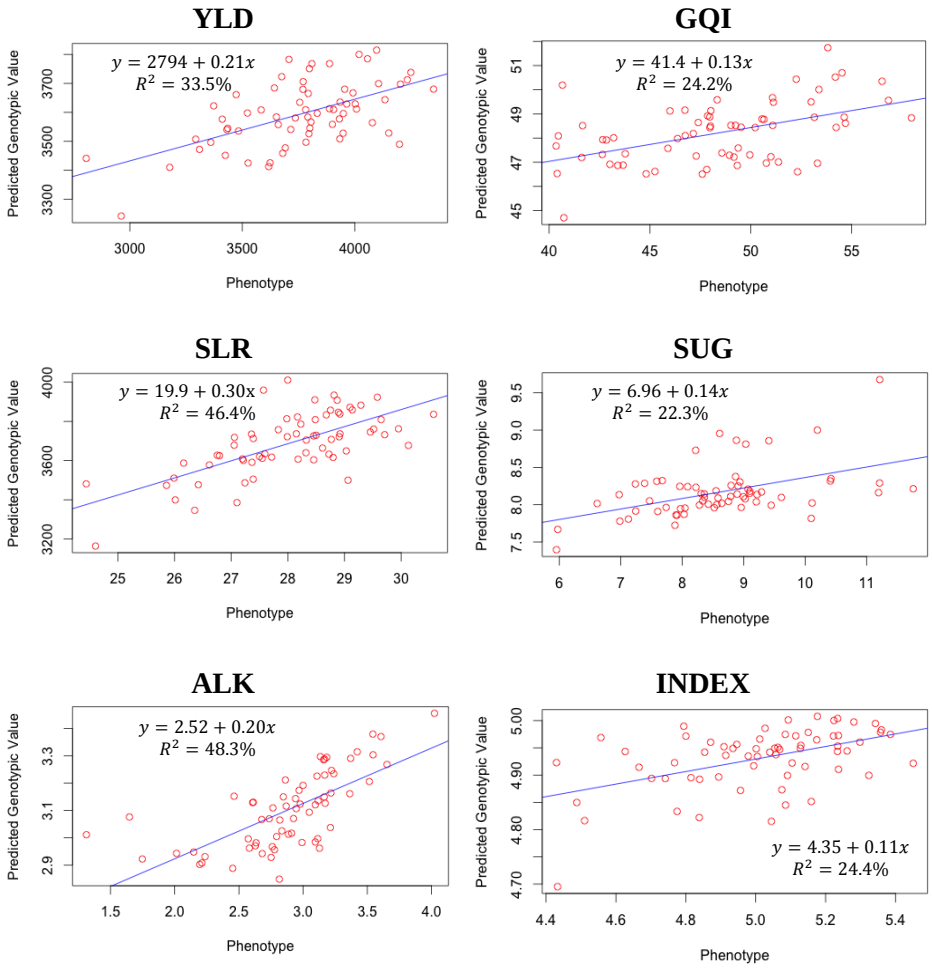

Figure 3 Regression equations of the predicted genotypic values by the observed hybrid phenotypes, using the genetic and phenotype data from the lines to validate the model for the traits YLD, GQI, SLR, SUG, ALK and INDEX.

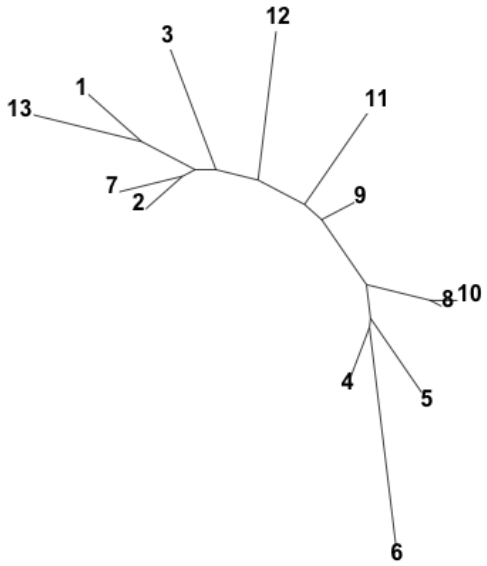

Figure 4 Neighbor joining tree of the genetic distance among lines, using the Jaccard method (1908).

Table 6 Correlation (*r*) between the line divergence with hybrid performance and the respective heterosis for the traits YLD, GQI, SLR, SUG and ALK. Values based on the mean of the environments.

	Hybrid		**Heterosis**	
Trait	***r***	**P-value**	***r***	**P-value**
YLD	0.13	0.2904	0.02	0.8864
GQI	0.09	0.4331	0.13	0.2762
SLR	-0.04	0.7618	-0.01	0.9435
SUG	-0.11	0.3783	-0.17	0.1481
ALK	0.12	0.3347	-0.05	0.6670

DISCUSSION

The total number of SNPs sequenced was 1193 and 861 were obtained after quality control. The GWS assumes the use of a large number of SNPs to cover all the genome (MEUWISSEN; HAYES; GODDARD, 2001). In the literature, this number has been very variable (WEIGEL et al. 2009; VAZQUEZ et al. 2010; MAKOWSKY et al. 2011). However, Lorenzana and Bernardo (2009) showed that the increase in the model accuracy is minimum starting from 100 markers, but it is important that the markers are well distributed in the genome. In the present study, a very uniform SNP distribution was obtained along the chromosomes.

To verify whether the GxE interaction has effect on the GWS efficiency, it is necessary that the environments where the treatments were assessed are contrasting. The two locations where the experiments were carried out, although close to each other, differ in several aspects. One of the locations was the experimental station of the Souza Cruz S/A Company and the other a producer´s farm with tradition of cultivating tobacco from the Flue-Cured Virginia varietal group. The difference between the locations is detected for all the traits based on the phenotypic (P) and genotypic (G) data (Table 2).

Several models of analysis are currently available for prediction using GWS (CROSSA et al., 2010; DE LOS CAMPOS et al., 2009; MEUWISSEN; HAYS; GODDARD, 2001), which jointly incorporate all the marker information available to predict the genotypic value (LORENZ, 2013). Each one has advantages and disadvantages. In the present study three alternatives were used: BRR, BB (MEUWISSEN; HAYS; GODDARD, 2001) and BL (PARK; CASELLA, 2008). The predicted accuracies of the three methods were very similar (Table 1). Although differences were found among these methods using simulated data (MEUWISSEN; HAYS; GODDARD, 2001; PARK; CASELLA

2008), the differences are minimal in real data for quantitative traits (DE LOS CAMPOS et al., 2013), as observed in the present and other studies (HESLOT et al., 2012; PÉREZ-RODRÍGUEZ et al., 2012; RESENDE et al., 2012; RIEDELSHEIMER et al., 2012; WIMMER et al., 2013). For this reason, we chose the BRR method considering only additive effects, because it is computationally simple and robust, thus it can be applied without restriction in the absence of detailed knowledge of the trait architecture (WIMMER et al., 2013). The main restriction is that this method assumes that the markers have the same variance (MEUWISSEN; HAYS; GODDARD, 2001). However, Ritland (1996) showed the goodness of fit of the infinitesimal additive model estimates and concluded that this model is not useful only in situations where there are some alleles with significant non-additive effects.

Several traits should be considered in tobacco breeding, contemplating the requirements of the producer, industry and the consumer. Cured leaf yield (kg/ha) is essential because it affects the economic success of the producer. The *GQI* is a standard quality index that considers the physical characteristics of the leaf including color, shine and body (MAPA/BRASIL, 2007). This index is extremely important, because its value is used to assess the commercial value of each lot that arrives at the factory. Therefore the higher the *GQI*, the better is the value of the tobacco. The SLR is the ratio between the main leaf nervure and total leaf weight. The lowest percent as possible is required because it is not desirable that the plant uses energy to produce nervures but rather it should use the energy to produce leaf blade, that is most used in the final product. The SUG and ALK traits are also important because they affect the chemical quality of the end product. GWS models usually focus on a single phenotype trait. However, most breeding programs assess a range of phenotypes of the individuals or progeny under selection. As all the traits mentioned should be considered simultaneously in a tobacco breeding program, GWS was also applied

considering a selection index (INDEX). Details on obtaining this index are presented by Carvalho et al. (2015, Article 1). We assumed that the association between the markers and phenotype could be made considering all the traits simultaneously. Although there are no conclusive results, it is evident that the number of genes involved in the expression of these traits must be variable and the same occurs for the effect of the environment in the phenotypic expression. The repeatability values obtained in the analysis of variance (Table 2) show that they were assessed with average to high precision (RESENDE; DUARTE, 2007).

The predicted accuracies (r_{GP}) were assessed by the correlation between the predicted genotypic value and the observed phenotypic value for each environment and on the mean of the environments. The estimates obtained varied among the traits, but in general they could be considered medium to high and very similar to that presented in the literature for other species, where estimates were obtained for other quantitative traits using different numbers of SNPs (HESLOT et al., 2012; PÉREZ-RODRÍGUEZ et al., 2012; RESENDE et al., 2012; RIEDELSHEIMER et al., 2012; WIMMER et al., 2013). The reliability of the correlation estimates depends on the heritability/repeatability with which the phenotypic were obtained, i.e. the experimental precision. As already mentioned, the accuracies ($\sqrt{r^2}$) with which the traits were assessed were relatively high (Table 2), that must have contributed to the high predictive accuracy estimates, especially when the mean of the environments was considered. Bernardo (2010) commented that the true genotypic values are not observable and that correlation is estimated from the observed performance and the predicted performance. The true value of the comparable accuracy in any situation would be r_{GP}/h, where h is the square root of the heritability (DEKKERS, 2007). The division by h corrects the effects in the estimated means. From this expression and the results shown in the literature, the GWS

efficiency is dependent on the quality of the phenotypic evaluation. It is very clear the need of assessing the experiments as precise as possible especially when the GWS will be applied in plant breeding.

From the breeders´ point of view the concern is to identify securely the superior genotypes - lines, progenies or hybrids. Since the r_{GP} considers the variation among the genotypes, those that are in a middle position, that are not of very much interest to the breeder, have a large effect on the results. For this reason we also estimate the coincidence of the ten best treatments assessed by the predicted genotypic value and observed phenotypic value. The coincidence was the larger for ALK, because it has a simpler genetic control, and the smallest for GQI, because it is an index that depends on the phenotypic expression of some traits and therefore it has a more complex genetic architecture. The information of the genotypic value, in training the model, in reality is an index that involves the phenotype and markers, that is, the genotype. In this context the index is always superior to the phenotype selection by itself, i.e. it is more accurate. It is emphasized that the estimates of the repeatability of the genome value among locations were much higher than the phenotypic value (Table 2). Repeatability in a random model is equivalent to the realized heritability (h_R^2), i.e. the variation that is effectively used in selection. Thus the markers function as additional information of the real kinship among the individuals assessed and therefore the use of the genome can improve the efficiency of the selective process.

Although the genetic information contributed to increase the h_R^2 values, the effect of the GxE interaction on GWS accuracy was clearly shown in some circumstances. The r_{GP} estimates were superior for all traits when the values of the mean of the environments were considered for estimating the marker effects compared with the values of each environment independently (Table 1). In addition, the r_{GP} estimates show that when the marker effect is estimated from

the mean value of the two environments, the prediction of the treatments in each environment is more accurate than when the marker effects are obtained by the phenotypic means of one environment and the prediction made on another (Table 4).

In genetic terms, the GxE interaction occurs when the contribution of the genes that control the trait or their expression level differ among the environments. The reason is that gene expression is influenced or regulated by the environments (KANG; GAUCH, 1995). When the genome information is used, the nucleotide sequence is the same for all environments, what changes is the relative expression of each marker on the phenotypic manifestation. The graphs that show the marker effects for the different traits are visually different between the environments and the mean of the environments (Figure 1). This corroborates to the comments above.

The GxE interaction effect on the GWS was well elucidated. However, the question of genetic sampling has not yet been discussed. The cross validation has been used to verify the results obtained in the GWS. For this part of the treatments are randomly chosen to compose the training population (TRN) to validate the model and estimate the marker effects and the remaining treatments are used to test the model, i.e. the test population (TST). In this context, we can check whether we can reduce the number of individuals phenotypically assessed in the field. Consequently only part of the individuals would be phenotypically assessed in the field and used to train the model. The remaining individuals would be predicted based on their marker incidence matrix. A question that arises is what would be the maximum number of missing individuals in TRN population to obtain an accurate prediction of the hybrids not assessed phenotypically. Different percentages of missing were considered, that is, numbers of treatments removed from the TRN population. When the level of missing decreased, the r_{GP} increased (Figure 2). Therefore the higher the

selection intensity applied, the larger should be the number of individuals assessed phenotypically for more accuracy in predicting the individuals not tested in the field. There are several reports in the literature that confirm this observation, using simulated data (VANRADEN; SULLIVAN, 2010) and real data (BASTIAANSEN et al., 2010; HEFFNER et al., 2011; HESLOT et al., 2012; LIU et al., 2011; LORENZANA; BERNARDO, 2009). This fact reflects in some consequences, because if the TRN is constituted by a very small sample, it certainly will not represent genotypically the reference population and therefore the TST hybrids will not be accurately predicted. To mitigate this effect it has been endeavored to use TRN with a large number of individual/lines (DE LOS CAMPOS et al., 2013; LIU et al., 2011), and that is frequently not viable in practice or increases the process cost (RINCENT et al., 2012).

What would be the implication of these results on the work of breeders for incorporating GWS as a routine activity in breeding programs? First, it should be pointed out that the phenotype should represent the future conditions in which the new cultivars will be grown. Taking tobacco as a reference, there are more than 160 thousands farmers in Brazil, normally family based, that use different management systems, and the climatic conditions, soil and biotic stresses also vary (AFUBRA/IBGE, 2015). Thus the experiments assessing the phenotype should be carried out under conditions that represent as much as possible those environments. Using GWS could contribute in the sense of incorporating more information to help the breeder in selection, obtaining estimates of the genotypic value as close as possible to the true genetic value for the conditions assessed and consequently obtaining higher h_R^2 values.

One of the concerns of any breeding program with the end objective of obtaining hybrids, as in the case of tobacco, is to be able to predict the hybrids obtained from a determined number of lines. This concern is obvious because the number of possible hybrid combinations is already large from a relatively

small number of lines. For example, with 20 lines, the number of possible combinations would be 190 without considering the reciprocal. In the present study, we tested to predict hybrids based on training the model with genetic and phenotypic value of the lines only. The genetic value of the hybrids was determined based on the combination of SNP value of its respective parent lines and the genotypic value was estimated. It was found that the correlation between the predicted genotypic value and the observed phenotypic value of hybrids was of average magnitude for all the traits (Figure 3). The fit of the linear regression equation was less than 50%, for all traits. These results show that using GWS would not be decisive in the hybrids to be selected, but it could be very useful as selective criteria to reduce the number of hybrids to be assessed in the field, eliminating those combinations with less probability of success.

In practice it is not common to sequence both the lines and hybrids, as it was done in this study, usually, only the lines are sequenced and the hybrid sequences are obtained based on the information of the respective parent lines. Consequently, the comparison between the hybrid genetic value obtained from sequencing and from the parent lines could be performed. That correlation was high and positive (r=0.86), but not equal to the unit, as expected in theory. There are some possible explanations for that. i) The analytical error, that it always possible to occur; ii) The lines are not completely homozygous for all loci, or have small genotype mixtures within each line. This last fact has been reported for some species, including tobacco itself (TOKATLIDIS, 2015).

Hybrid heterosis is a function of the divergence between the parents and the presence of dominance in the loci involved in the control of the trait (FALCONER; MACKAY, 1999). Since molecular markers have been implemented, more than 30 years ago (STUBER; GOODMAN; MOLL, 1982), there has been an intense search for the association between heterosis and genetic divergence of lines based on markers. In some studies a positive

correlation between line diversity and hybrid heterosis was observed (BETRAN et al., 2003; KIULA; LYIMO; BOTHA, 2008; WEGARY, VIVEK; LABUSCHAGNE, 2013; XU; LIU; LIU, 2004). In the present study the divergence was estimated among the 13 lines (Figure 4), and it was possible to discriminate the lines regarding their genetic constitution, although they were products of a breeding program submitted to intense selection in recent years. As heterosis is function of $h=dY^2$ (FALCONER; MACKAY, 1999), where d is the deviation of dominance and Y is the genetic divergence between the parent lines, it is expected in theory that the pairs of lines with larger divergence would present higher dominance. However, when the heterosis estimate or hybrid performance is associated to genetic divergence, the correlation estimates were nil for all traits. The fact that the genetic diversity was not considered a good heterosis predictor is in agreement with reports in the literature for several species including corn (LEGESSE et al., 2008; PARENTONI et al., 2001), wheat (KRYSTKOWIAK et al., 2009), sorghum (JORDAN et al., 2003). The most plausible explanation is that divergence was assessed by several markers that may not be associated to the trait under study. If were used only the markers associated to the trait, the estimated divergences would probably confirm what the theory shows for a locus.

In synthesis, the use of the GWS should be considered in some circumstances. For traits which the assessment is expensive or difficult to perform; and when there are too many genotypes to be phenotypically tested in the field, the GWS can be used as a screening to reduce the amount of genotypes, this is, the GWS would be used for excluding genotypes but not selecting (COOPER et al., 2014). In addition, the use of genetic information, as a kinship information, is an opportunity to increase the phenotypic selection accuracy (CROSSA et al., 2010, 2011; DE LOS CAMPOS et al., 2009; HESLOT et al., 2012; PEREZ et al., 2010). Thus phenotypic selection continues

to be essential in plant breeding and the genome-wide data can be used as additional information to increase the gains obtained (HESLOT; JANNINK; SORRELLS, 2015).

CONCLUSIONS

Correlations between the observed phenotypic value and the predicted genotypic value varie among the traits, but was high, indicating good accuracy of the models. However, the GxE interaction affected the use of GWS in predicting the genotypic values.

The lines showed divergence based on the markers. Nevertheless, the pair-wise line divergence was not associated to heterosis or to the hybrids derived from them.

The success of applying the prediction model varied significantly according to the proportion of missing treatments in the training (TRN) population. The accuracy estimate increased as higher the number of treatments in the TRN population.

The use of GWS was shown to be very promising in the sense of complementing the phenotypic information in the process of progeny/cultivar selection. In obtaining hybrids, it allows prioritizing some combinations to be evaluated in the field when not all of them can be assessed.

REFERENCES

AFUBRA. **Associação dos Fumicultores do Brasil.** 2014. Disponível em: <http://www.afubra.com.br>. Acesso em: 6 dez. 2015.

BASTIAANSEN, J. W. M. et al. Comparison of analyses of the QTLMAS XIII common dataset. I: Genomic selection. **BMC Proceedings,** v. 4(Suppl. 1): S1, p.1-7, 2010.

BERNARDO, R. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. **Crop Science**, v. 48, p. 1649–1664, 2008.

BERNARDO, R. **Breeding for Quantitative Traits in Plants**. 2. ed. Woodbury: Stemma Press, 2010. 390 p.

BERNARDO, R.; YU, J. Prospects for genome wide selection for quantitative traits in maize. **Crop Science**, v. 47, p. 1082–1090, 2007.

BETRAN, F. J. et al. Genetic diversity, specific combining ability, and heterosis in tropical maize under stress and non stress environments. **Crop Science**, v. 43, p. 797–806, 2003.

BUCKLER, E. S. et al. The genetic architecture of maize flowering time. **Science,** v. 325, p. 714–718, 2009.

COOPER, M.; PODLICH, D. W.; SMITH, O. S. Gene to phenotype models and complex trait genetics. **Australian Journal of Agricultural Research,** v. 56, p. 895–918, 2005.

COOPER, M. et al. Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction. **Crop and Pasture Science**, 2014.

CROSSA, J. et al. Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. **Genetics**, v. 186, p. 713–724, 2010.

CROSSA, J. et al. Genomic selection and prediction in plant breeding. **Journal of Crop Improvement**, v. 25, p. 239–261, 2011.

CRUZ, C. D.; CARNEIRO, P. C. S.; REGAZZI, A. J. **Modelos Biométricos Aplicados ao Melhoramento Genético**. v. 2, 3. ed. Viçosa: UFV, 2014. 668 p.

DEKKERS, J. C. Commercial application of marker-and gene- assisted selection in livestock: strategies and lessons. **Journal of Animal Science**, v. 82, p. E313–E328, 2004.

DEKKERS, J. C. Prediction of response to marker-assisted and genomic selection index theory. **Journal of Animal Breeding and Genetics**, v. 124, v. 331-341, 2007.

DE LOS CAMPOS, G.; RODRIGUEZ, P. P. **BGLR: Bayesian Generalized Linear Regression.** v. 1.0.4, 2015.

DE LOS CAMPOS, G. et al. Predicting quantitative traits with regression models for dense molecular markers and pedigree. **Genetics**, v. 182, p. 375–385, 2009.

DE LOS CAMPOS, G. et al. Whole genome regression and prediction methods applied to plant and animal breeding. **Genetics**, v. 193, p. 327–345, 2013.

EIKHOLT, D. P.; LEWIS, R. Effect of an Introgressed *Nicotiana tomentosa* Leaf Number QTL on Yield and Quality Characteristics in Flue-Cured Tobacco. **Crop Science,** v. 54, n. 2, p. 586-594, 2014.

ELSHIRE, R. J. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. **Plos One**, v. 6, n. 5, p. e19379, 2011.

ENDELMAN, J.B. et al. Optimal design of preliminary yield trials with genome-wide markers. **Crop Science**, v. 54, p. 48–59, 2014.

FALCONER, D. S.; MACKAY, T. F. C. **Introduction to quantitative genetics**. 4th ed. Malaysia: Pearson, 1996. 464 p.

GIANOLA, D. et al. Additive genetic variability and the Bayesian alphabet. **Genetics,** v. 183, p. 347–363, 2009.

GIANOLA, D. Priors in Whole-Genome Regression: The Bayesian Alphabet Returns. **Genetics,** v. 194, n. 13, p.573-596, 2013.

GRIFFING, B. Concept of general and specific combining ability in relation to diallel crossing systems. **Australian Journal of Biological Sciences,** v. 9, p. 463–493, 1956.

GUO M. et al. Maize ARGOS1 (ZAR1) transgenic alleles increase hybrid maize

yield. **Journal of Experimental Botany**, v. 65, n. 1, 2014.

HABIER, D. et al. Extension of the Bayesian alphabet for genomic selection. **BMC Bioinformatics**, v. 12, p. 186–198, 2011.

HEFFNER, E. L. et al. Plant breeding with genomic selection: Gain per unit time and cost. **Crop Science**, v. 50, p. 1681–1690, 2010.

HEFFNER, E.L. et al. Genomic selection accuracy for grain quality traits in biparental wheat populations. **Crop Science**, v.51, p. 2597–2606, 2011.

HESLOT, N.; JANNINK, J-L.; SORRELLS, M. E. Perspectives for Genomic Selection Applications and Research in Plants. **Crop Science**, v. 55, p. 1-12, 2015.

HESLOT, N. et al. Genomic selection in plant breeding: a comparison of models. **Crop Science**, v. 52, p. 146–160, 2012.

HILL, W. G. Applications of Population Genetics to Animal Breeding, from Wright, Fisher and Lush to Genomic Prediction. **Genetics,** v.196, p.1-6, 2014.

IBGE. **Instituto Brasleiro de Geografia e Estatística.** 2014. Disponível em: <http://www.ibge.gov.br/home/>. Acesso em: 6 dez. 2015.

JACCARD, P. Nouvelles recherches sur la distribution florale. **Bulletin de la Société vaudoise des sciences naturalles,** v. 44, p. 223-270, 1908.

JORDAN, D. R. et al. Prediction of hybrid performance in grain sorghum using RFLP markers. **Theoretical Applied Genetics,** v. 106, p. 559-567, 2003.

KANG, M. S.; GAUCH JR., H. G. **Genotype-by-environment interaction**. Boca Raton: CRC Press, 1995.

KIULA, B.A.; LYIMO, N. G.; BOTHA, A. M. Association between AFLP-based genetic distance and hybrid performance in tropical maize. **Plant Breeding**, v. 127, p. 140–144, 2008

KOLLER, D.; FRIEDMAN N. **Probabilistic Graphical Models: Principles and Techniques.** Cambridge: MIT Press, MA, 2009. 1280 p.

KRYSTKOWIAK, K. et al. Relationship between phenotypic and genetic diversity of parental genotypes and the specific combining ability and heterosis

effects in wheat (Triticum aestivum L.). **Euphytica,** v. 165, p. 419-434, 2009.

LANDE, R.; THOMPSON, R. Efficiency of marker-assisted selection in the improvement of quantitative traits. **Genetics**, v. 124, p. 743–756, 1990.

LEGESSE, B. W. et al. Genetic diversity of African maize inbred lines revealed by SSR markers. **Hereditas** , v. 144, p. 10–17, 2007.

LEGESSE, B. W. et al. Relationship between hybrid performance and AFLP based genetic distance in highland maize inbred lines. **Euphytica**, v. 162, p. 313–323, 2008.

LIU, Z. et al. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. **Genetics Selection Evolution**, v. 43, n. 19, p. 1-9, 2011.

LORENZ, A. J. Resource allocation for maximizing prediction accuracy and genetic gain of genomic selection in plant breeding: A simulation experiment. **G3,** v. 3, n. 3, p. 481–491, 2013.

LORENZ, A. J. Genomic selection in plant breeding: Knowledge and prospects. Advances in Agronomy, v. 110, p. 77–123, 2011.

LORENZANA, R. E.; BERNARDO, R. 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. **Theoretical Applied Genetics,** v. 120, p.151–161, 2009.

MAKOWSKY, R. et al. Beyond missing heritability: prediction of complex traits. **PLoS Genetics**, v. 7, p. e1002051, 2011

MAPA/BRASIL. Instrução Normativa n˚10 do Ministério da Agricultura, Pecuária e Abastecimento, de 16 de abril de 2007. Institui a classificação de Tabaco**. Diário Oficial [da] República Federativa do Brasil**, Brasília, 23 mai. 2006. Disponível em: <http://www.agricultura.gov.br/legislacao>. Acesso em: 7 dez. 2015.

MASSMAN, J. M. et al. Genome-wide prediction from maize single-cross data. **Theoretical Applied Genetics**, v. 126, p. 13-22, 2013.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, v. 157, p. 1819-1829, 2001.

MOREAU, L.; CHARCOSSET, A.; GALLAIS. A. Experimental evaluation of several cycles of marker-assisted selection in maize. **Euphytica**, v. 137, p. 111–118, 2004.

OBER, U. et al. Predicting genetic values: a kernel-based best linear unbiased prediction with genomic data. **Genetics**, v. 188, p. 695–708, 2011.

PARENTONI, S. N. et al. Heterotic groups based on yield-specific combining ability data and phylo- genetic relationship determined by RAPD markers for 28 tropical maize open pollinated varieties. **Euphytica**, v. 121, p. 197–208, 2001.

PARK, T.; CASELLA, G. The Bayesian Lasso. **Journal American Statistics Association**, v. 103, p. 681–686, 2008

PÉREZ, P. et al. Genomic-enabled prediction based on molecular markers and pedigree using the Bayesian Linear Regression package in R. **Plant Genetics**, v. 3, p. 106–116, 2010.

PÉREZ-RODRÍGUEZ, P. et al. Comparison between linear and non-parametric regression models for genome-enabled prediction in wheat. **G3**, v. 2, p. 1595–1605, 2012.

POLAND, J. et al. Genomic selection in wheat breeding using genotyping-by-sequencing. **The Plant Genome**, v. 5, p. 103–113, 2012.

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

RESENDE, M. F. R. et al. Accuracy of genomic selection methods in a standard dataset of Loblolly pine (Pinus taeda L.). **Genetics**, v. 190, p. 1503–1510, 2012.

RIEDELSHEIMER, C.; TECHNOW, F.; MELCHINGER, A. Comparison of whole-genome prediction models for traits with contrasting genetic architecture in a diversity panel of maize inbred lines. **BMC Genomics,** v. 13, p. 452, 2012.

RINCENT, R., D. et al. Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize. **Genetics**, v. 192, p. 715–728, 2012.

RITLAND, K., 1996 A marker-based method for inferences about quantitative

inheritance in natural populations. **Evolution**, v. 50, p. 1062–1073.

SAITOU, N.; NEI, M. The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees. **Molecular Biology and Evolution**, v. 4, n. 4, p. 406-425, 1987.

SOLLER, M.; PLOTKIN-HAZAN, J. The use of marker alleles for the introgression of linked quantitative alleles. Theoretical Applied Genetics, v. 51, p. 133–137, 1977.

SPINDEL J. et al. Genomic Selection and Association Mapping in Rice (Oryza sativa): Effect of Trait Genetic Architecture, Training Population Composition, Marker Number and Statistical Model on Accuracy of Rice Genomic Selection in Elite, Tropical Rice Breeding Lines. **PLoS Genetics,** v. 11, n. 6, p. e1005350, 2015

STUBER, C. W.; GOODMAN, M. M.; MOLL. R. H. Improvement of yield and ear number resulting from selection at allozyme loci in a maize population. **Crop Science**, v. 22, p. 737–740, 1982.

SU, G. et al. Estimating Additive and Non-Additive Genetic Variances and Predicting Genetic Merits Using Genome- Wide Dense Single Nucleotide Polymorphism Markers. **Plos One,** v. 7, n. 9, p. e45293, 2012.

THAVAMANIKUMAR, S.; DOLFERUS,R.; THUMMA, B. R. Comparison of Genomic Selection Models to Predict Flowering Time and Spike Grain Number in Two Hexaploid Wheat Doubled Haploid Populations. **G3,** v. 5, n. 10, p. 1991-1998, 2015.

TOKATLIDIS, I. S. Conservation Breeding of Elite Cultivars. **Crop Science,** v. 55, n. 6, p. 2417-2434, 2015.

VANRADEN, P. Efficient methods to compute genomic predictions. Journal of Dairy Science, v. 91, p. 4414–4423, 2008

VANRADEN, P. M.; SULLIVAN, P. G. International genomic evaluation methods for dairy cattle. **Genetics Selection Evolution**, v. 42, n. 7, p. 1-9, 2010

VAZQUEZ, A. I. et al. Predictive ability of subsets of single nu- cleotide polymorphisms with and without parent average in US Holsteins. **Jounal of Dairy Science**, v. 93, p. 5942–5949, 2010.

VIJAY, V.; LEWIS, R. S. Mapping of quantitative trait loci affecting resistance to *Phytophthora nicotianae* in tobacco (*Nicotiana tabacum* L.) line Beinhart-1000. **Molecular Breeding,** v.29, n. 1, p. 89-98, 2012

WEGARY, D.; VIVEK, B.; LABUSHAGNE, M. Association of parental genetic distance with heterosis and specific combining ability in quality protein maize. **Euphytica,** v.191, p. 205-216, 2013.

WEIGEL, K. A. et al. Predictive ability of direct genomic values for lifetime net merit of Holstein sires using selected subsets of single nucleotide polymorphism markers. **Journal of Dairy Science**, v. 92, p. 5248–5257, 2009.

WIMMER, V. et al. Genome-Wide Prediction of Traits with Different Genetic Architecture Through Efficient Variable Selection. **Genetics**, v. 195, p. 573–587, 2013.

XIAO, B. Location of Genomic Regions Contributing to *Phytophthora nicotianae* Resistance in Tobacco Cultivar Florida 301. **Crop Science,** v.53, n. 2, p. 473-481, 2013

XU, Y.; CROUCH. J. H. Marker-assisted selection in plant breeding: From publications to practice. Crop Science, v. 48, p. 391–407, 2008.

XU, S.; LIU, J.; LIU, G. The use of SSRs for predicting the hybrid yield and yield heterosis in 15 key inbred lines of Chinese maize. **Hereditas**, v. 141, p. 207–215, 2004

YI, B. et al. Mapping the Root-Knot Nematode Resistance Gene (*Rk*) in Tobacco with RAPD Markers. **Plant Disease,** v. 82, n. 12, p. 1319-1322, 1998.