Genetic distance and specific combining ability in cassava

Citation:

Ceballos, Hernán; Becerra López-Lavalle, Luis Augusto; Calle, Fernando; Morante, Nelson; Ovalle Rivera, Tatiana melissa; Hershey, Clair. 2016. *Genetic distance and specific combining ability in cassava*. Euphytica 14 p.

Publisher's DOI:

http://dx.doi.org/10.1007/s10681-016-1701-7

Access through CIAT Research Online:

http://hdl.handle.net/10568/73430

Crop Science
A Journal Serving the International Community of Crop Scientists
Crop Science SOCIETY OF AMERICA

GENETIC DISTANCE AND HETEROSIS IN CASSAVA

Journal:	*Crop Science*
Manuscript ID:	CROP-2015-05-0320-ORA
Manuscript Type:	1. Original Research Articles
Divisions:	C1 crop breeding & genetics
Date Submitted by the Author:	25-May-2015
Complete List of Authors:	Ceballos, Hernan; CIAT, Cassava Project; CIAT, Cassava Project Becerra Lopez-Lavalle, Luis; CIAT, Cassava Program Calle, Fernando Morante, Nelson Ovalle, Tatiana Hershey, Clair
Keywords:	Crop genetics, Other tubers, Vegetable crops

GENETIC DISTANCE AND HETEROSIS IN CASSAVA

H. Ceballos ✉, L.A. Becerra López-Lavalle, F. Calle, N. Morante, T.M. Ovalle, and C. Hershey

International Center for Tropical Agriculture (CIAT). Recta Cali-Palmira km 17. Valle del Cauca, Colombia

✉ Corresponding author:
e-mail: h.ceballos@cgiar.org
Telephone: (57-2)445-0125
Fax: (57-2) 445-0073

Abstract

Cassava (*Manihot esculenta* Crantz) is an important source of energy in the tropics. Its starchy roots are valuable for food security as well as for different industries. Cassava is an outcrossing crop and its breeding is based on the use of heterozygous progenitors. A strategy for a more efficient genetic improvement of cassava is required to meet the projected increases in demand from populations (particularly in Africa) that depend on this crop. An alternative would be based on the exploitation of non-additive genetic effects (heterosis) through reciprocal recurrent selection. Results from three diallel studies (with 9-10 progenitors), conducted at three different environments (sub-humid, acid soils and mid-altitude valleys) in Colombia, have already been published for fresh root yield. For each environment two locations were used with three replications per location. In this article, the diallels phenotypic data was linked to Nei's genetic distances which were estimated through a set of 95 SNPs diagnostic of the cassava diversity. In addition phenotypic analysis of dry matter yield was conducted. Results indicated inconsistent correlations between genetic distances and performance of the F1 families *per se* and specific combining ability effects for the two variables analyzed.

Key words: heterosis; non-additive effects; specific combining ability effects; genetic distances.

Abbreviations: specific combining ability (SCA); general combining ability (GCA); fresh root yield (FRY); dry matter yield (DMY)

5585 Guilford Rd., Madison WI 53711

1. Introduction

Cassava (*Manihot esculenta* Crantz) is a perennial shrub originated in the neotropics. Its most important product is the starchy roots used as a source of energy by millions of people, particularly in Sub-Saharan Africa. It is also a competitive source of starch; cassava is the second most important source of starch worldwide, after maize (Stapleton, 2012; Norton, 2014), and its starch is the most traded internationally. Dried cassava root chips are also used at industrial levels for animal feeding and ethanol production. Commercial cassava is multiplied through the use of stem cuttings. What farmers grow are clonally propagated hybrids. As such, cassava can be used as model for clonal crops with the advantage that is grown annually and, being diploid (Wang et al., 2011), avoids the complication of polyploidy that several other clonally propagated species have.

Cassava breeding is based on the production of segregating progenies. Full-sib families are produced through direct crosses, whereas open pollinations result in half-sib families. These segregating families are then evaluated through a phenotypic mass selection (Jennings and Hershey, 1985; Jennings and Iglesias, 2002; Ceballos et al. 2012). Little or no attention is paid to family structure in the selection process. Breeders focus their attention on evaluating and selecting individual genotypes regardless of the family they belong to. It is these individual selected genotypes that will be eventually released by breeders and grown by farmers. A key feature of this process is that additive, dominance and epistatic genetic effects not only influence the breeders decisions (although they are not ordinarily measured), but also can be exploited in the

5585 Guilford Rd., Madison WI 53711

cloned genotypes grown by farmers. The clonal reproduction of cassava allows multiplication of individual genotypes in such a way that environmental and genetic factors affecting their performance can be separated. This is important because within family genetic effects can be properly estimated.

An important and distinctive characteristic of cassava breeding is that it uses heterozygous progenitors to produce the varieties grown by farmers: clonally propagated hybrids. This fact places cassava in a unique position compared with autogamous or commercial hybrid crops (such as maize, sorghum and sunflower) whose breeding is based on the use of homozygous progenitors. Breeding of many other clonal crops is also based on heterozygous progenitors.

For cassava to remain competitive with other crops used for different agro-industrial end uses (particularly maize) and to meet the projected increase in demands for food security crop (particularly in Africa) more efficient breeding approaches would be required. In spite of the large efforts and financial investments in identifying molecular markers to make cassava genetic enhancement more efficient through marker-assisted selection MAS, the practical application has been negligible (de Oliveira et al., 2012). Genetic transformation has, so far, failed to deliver products that can help farmers. An alternative to make cassava breeding more efficient is to partition genetic variation for their adequate exploitation. Heterosis has been found to be a key phenomenon not only for cross pollinated crops but also, and surprisingly, for autogamous crops such as rice (Cheng et al., 2004; Spielman et al., 2013). The exploitation of heterosis requires

5585 Guilford Rd., Madison WI 53711

special breeding approaches such as reciprocal recurrent selection (RRS) or line improvement from two different heterotic groups. RRS has been used to exploit heterosis in many different crops (Bernardo, 2014): maize, cotton, eucalyptus, gourd, oil palm, pearl millet, rice, sorghum, and tomato.

The identification or creation of heterotic patterns in cassava germplasm is an important step that is urgently needed because they are the backbone of successful hybrid breeding and RRS (Melchinger and Gumber, 1998). Heterotic patterns can be defined as a pair of populations (or individual inbred genotypes) which express high heterosis and, consequently, high hybrid performance in their cross (Hallauer and Miranda, 1981; Melchinger and Gumber, 1998; Parentoni et al., 2001).

The poor population structure in cassava and the limited studies on its genetics (at the quantitative level), can partially explain why there is no information regarding heterotic groups in this crop. The relationship between genetic distance and heterosis, has been analyzed in different crops (Ali et al., 1995; Betrán et al., 2003; Cheres et al., 2000; Diers et al., 1996; Ghaderi et al., 1984; Lanza et al., 1997; Melchinger, 1999; Reif et al., 2003; Riday et al., 2003; Xiao et al., 1996; Zhang et al., 2006**)**, but not in cassava. With the development of new molecular tools, genetic distances can now be assessed efficiently. The objective of the present study was to analyze the relationship between Nei´s genetic distances (Nei and Li, 1979) and the specific combining ability effects among the progenitors from three different diallel studies conducted earlier (Cach et al., 2005; 2006; Calle et al., 2005; Jaramillo et al., 2005; Perez et al., 2005a; 2005b).

2. Materials and methods

2.1 Phenotypic data

Three diallel studies (9-10 progenitors) were conducted respectively for three different environments in Colombia: sub humid environment (Cach et al., 2005; 2006), acid soil savannas (Calle et al., 2005; Perez et al., 2005a) and mid-altitude valleys (Jaramillo et al., 2005; Perez et al., 2005b). The list of progenitors used in each of the three diallel studies is presented in Table 1. Unfortunately four progenitors (SM 2058–2, SM 1636-24, SM 1673-10 and SM 1657-12) were no longer available for measurement of genetic distances which, is the key additional information presented in this article along with the analysis of a variable (dry matter yield).

For each F1 cross, at least 30 genotypes (one plant per genotype derived from the germination of botanical seed) were produced and that was the number of genotypes representing each F1 family. The botanical seed was germinated and seedling transplanted to the field two months later. No analysis was made on these plants which were used only as source of cloned planting material. When the seedling plants were 11 months old the stems were collected and six vegetative cuttings for each of the 30 genotypes of each F1 family were obtained. The six cuttings from each genotype were planted in three replications at each of two representative locations of the target environments (three replications x two locations = six cuttings per genotype).

The 30 genotypes representing each F1 family were planted in the same plot (six rows and five plants per row). Because each individual genotype was cloned and planted in

5585 Guilford Rd., Madison WI 53711

the replicated trials, genetic variation within family could be estimated. This is a rather unique advantage offered by clonally propagated crops such as cassava.

The analysis of variance follows method 4 (direct and reciprocal crosses were combined for each F1 family and progenitors were not evaluated) proposed by Griffing (1956). Genotypes and environments were considered fixed and random effects, respectively. Analysis was done manually using Microsoft Excel (Microsoft, 2004; Nelson, 2000).

2.2 SNP genotyping

DNA was extracted as described by (Doyle and Doyle 1990) with the following minor adjustment: DNA was extracted from powdered leaf tissues using Qiagen Tissue Lyser (Venlo, Netherlands).

The samples were processed using a newly developed protocol for 96 single nucleotide polymorphism (SNP) genotyping in cassava with the EP1 system and SNP type assays of Fluidigm®, application version 3.1.2. (Peña-Venegas, et al. 2014). SNPs are an abundant type of DNA polymorphism. SNPs are biallelic in nature and therefore they are ideal for genetic studies of organisms and especially for assessing diversity in cassava (Kawuki, et al. 2009). The technique allowed to simultaneously collecting both end-point and real-time data from a unique chip cell with 97% confidence.

For the process 60 ng of DNA of each sample was used for DNA variant-site amplification. Two pre-amplification primers [locus specific primer (LSP) and specific target amplification (STA) primer] amplify the target region containing the SNP to be

genotyped. Subsequently, an additional PCR amplifies a portion of that target SNP region, using the LSP and two fluorescently labeled allele-specific primers ASP1 and ASP2; designed by aligning 10 cassava genomes against the cassava reference genome sequence information available at Phytozome v10. ASP1 and ASP2 are internal primers containing either the first or the second allele, respectively. All 96 SNPs are pre-amplified simultaneously in one multiplex PCR, for each DNA sample separately, on a MasterCycler® pro (Eppendorf, Germany).The specific target PCR cycling conditions in the thermocycler were 95°C for 15 min; followed by 14 cycles at 95 °C for 15 sec and 14 cycles at 60 °C for 4 min.

The last PCR is performed on a Fluidigm 96.96 Dynamic Array (SNP chip), where the reactions occur in separate nano-wells for each SNP and DNA sample combination, allowing simultaneous genotyping of 95 DNA samples and one water control at 96 SNP loci. Fluidigm 96.96 Dynamic Array is run on a BioMark HD System (Fluidigm), with the following PCR cycling conditions: 50 °C for 2 min, 70 °C for 30 min, 25 °C for 10 min and 95 °C for 5 min, followed by four touchdown cycles (95 °C for 15 s, from 64 °C to 61 °C for 45 s, 72 °C for 15 s) and 28 or 33 additional cycles (95 °C for 15 s, 60 °C for 45 s, 72 °C for 15 s) to discriminate properly homozygosis and heterozygosis in each sample tested. The PCR ends with 1 cycle at 20 °C for 10 s (see Fluidigm genotyping user guide). Fluorescence plots obtained for each SNP were analysed using the Fluidigm SNP genotyping analysis software.

2.3 SNP diversity analysis

DNA from 22 of 26 elite clonal cultivars used in three diallel studies (CM4574-7, CM523-7, CM5655-4, CM6740-7, CM6754-8, CM7033-3, CM8027-3, COL2737, ECU72, HMC-1, PER183, SM1219-9-3, SM1278-2, SM1411-5, SM1565-15, SM1565-17, SM1665-2, SM1741-1, SM2192-6, SM2219-11, SM805-15, and TAI8) was extracted as indicated above. Four genotypes had been eliminated since the original field assessment of the diallel studies: SM 1636-24 and SM 1673-10 (from mid-altitude valleys environment); SM 1657-12 (from sub-humid environment); and SM 2058-2 (from acid soil savannas). The analysis of the relationship between heterosis and genetic distances was, therefore, based on seven progenitor for the mid-altitude valleys, eight progenitors for the sub-humid conditions, and nine progenitors for the acid soil savannas.

The molecular analysis was based on 96 SNPs diagnostic of the cassava diversity in South America and the Caribbean Region (Peña-Venegas, et al. 2014). Alleles for each SNP were scored as present, absent, or missing (failed to amplify) and converted into a binary matrix to determine minor allele frequencies (MAF) for each SNP locus. The genetic distance among genotypes was calculated based on the matrices of allele frequencies using the Nei and Li's genetic distance (GD) matrix (Nei and Li 1979). The clustering criterion used was neighbor joining and the resulting dendogram was un-rooted. Robustness of the cladogram topology was assessed by bootstrap analysis using Winboot software. To ensure the accuracy of the bootstraping 10000 replicates of the data set were performed

3. Results and discussion

The most relevant trait where heterosis is likely to play an important role is fresh root yield (FRY). However, in the case of cassava, wide variations in dry matter content (DMC) can be observed. Using FRY alone to measure yield may be misleading if high "productivity" is associated to a low DMC. Therefore, this study will focus not only on FRY but also on dry matter yield (DMY), which is a combination of FRY and DMC. DMY is envisioned as more appropriate to quantify the overall effort made by the plant to store energy.

Table 2 presents the results for the sub-humid environment. FRY ranged from 26.5 to 45.7 t ha^{-1}. Interestingly, these two extremes involved progenitor (P) 1, crossed with P3 (26.5 t ha^{-1}) and P9 (45.7 t ha^{-1}). This type of result highlights the relevance of heterosis for traits such as FRY. As expected, genetic distance was smaller for the 1x3 cross than for 1x9 (0.124 and 0.195, respectively). Across the entire experiment cross 1x3 had the lowest FRY and showed the smallest genetic distance. However, there were 14 F1 families with genetic distances higher than that for the highest yielding cross (1x9). This result would suggest that genetic distance was more effective in identifying clones whose crosses are likely to show poor performance (perhaps as result of some degree of inbreeding depression) than for identifying clones expected to show positive heterosis. Similar results were observed for DMY (Table 2), which ranged from 7.5 to 12.4 t ha^{-1}, and involved the same crosses.

Measured values of specific combining ability (SCA) effects for the two traits are also presented in Table 2. These values are slightly different than those reported by Cach et

5585 Guilford Rd., Madison WI 53711

al. in 2006 (in the present study one of the original progenitors was missing and, therefore, SCA values needed to be recalculated for a more accurate analysis of the relationship between SCA and genetic distances). Similar situation will be observed for the diallels conducted in the remaining two environments.

The highest and lowest FRY coincide with the highest and lowest SCA values (-9.6 and 5.91 respectively for crosses 1x3 and 1x9). The clear contrast in SCA values for these two families suggests that indeed the performance of these hybrid families depended heavily in non-additive genetic effects. The same conclusions can be drawn for DMY, where SCA values ranged from -2.91 to 1.48, precisely for the same two crosses (1x3 and 1x9, respectively).

The main objective of this study was to analyze the relationship between Nei´s genetic distances (Nei and Li, 1979) and variation for FRY and DMY (measured both in t ha^{-1} and SCA units), which is presented in Figure 1. In every case there is a positive relationship indicating that genetic distances are indeed linked, to a certain degree, with heterosis (which in turn is closely associated with SCA effects) and yield performance. Based on the coefficients of determination (r^2), the relationship is stronger for DMY (Plots C and D, Figure 1), than for FRY (Plots A and B, Figure 1). The r^2 values for the relationship between genetic distances and the two variables were higher when FRY and DMY were expressed in SCA units (Plots B and D in Figure 1), rather than in t ha^{-1} (Plots A and C, Figure 1). This makes sense as SCA is more directly associated with

5585 Guilford Rd., Madison WI 53711

heterosis. Yield *per se*, measured in t ha^{-1} depends not only in SCA, but also in additive genetic effects (general combining ability or GCA effects in diallel terminology).

Table 3 and Figure 2 present the results for the acid soils environment of Colombia (Meta Department). Average FRY ranged from 12.5 to 26.5 t ha-1, considerably lower yields than those observed for the less stressful sub-humid environment. The lowest yielding F1 family was the cross 9x10 (12.5 t ha^{-1}). These two parents were also involved in two of the five lowest yielding crosses (5x10 and 2x9). Cross 7x8 was the highest yielding among the 36 families evaluated. Four of the best five yielding crosses involved progenitor P7. These results highlight that, in addition to SCA, GCA are also important in the performance of hybrids: crosses involving P7 tend to show an outstanding performance, whereas those from P9 would be expected to have a low yield. The relationship between Nei´s genetic distance and yield did not show a pattern in the acid soil savanna as was the case for the sub-humid environment. In fact, among the five crosses with lowest average FRY, two were among the five families with largest genetic distance among the respective progenitors (1x8 and 8x9). The genetic distances between the progenitors of the highest yielding crosses (7x8 and 1x7) were not particularly higher (0.238 and 0.197, respectively) than the average distance across the experiment (0.214). No clear pattern was evident when FRY was analyzed in term of SCA units either. Results were slightly better as the SCA value of the cross 8x9 exposing one of the largest genetic distances among the two progenitors (0.289) was not among the five lowest.

5585 Guilford Rd., Madison WI 53711

For DMY results were also disappointing as no clear association between genetic distances could be observed when the variable was analyzed in t ha^{-1} or in SCA units. The large genetic distance between progenitors P8 and P9 was met with a mediocre DMY (4.89 t ha^{-1}) of their cross, which was the second lowest. There were six families with average DMT > 8 t ha^{-1} and only cross 1x7 had a genetic distance below the average. The family with largest genetic distance was 5x9 (0.324) which yielded an mediocre average DMY of 5.81 t ha^{-1}. Similarly, the highest average DMY was observed in cross 7x10 (8.74 t ha^{-1}) which shows the 8^{th} largest genetic distance (-.238). The analysis of DMY through SCA effects did not improve its association with genetic distances. Cross 1x8 has one of the five highest genetic distance among progenitors (0.253) but its SCA value was among the worst five (-1.25). Similarly, but in the opposite direction, was the case of cross 5x8 with low genetic distance (0.170) but showing the highest SCA value among the hybrid families considered (1.36).

The associations between genetic distances and the two traits analyzed are presented in Figure 2. Plots A and C present the regressions for FRY and DMY in t ha^{-1}, respectively. In both cases there is a weak negative relationship, with negligible r^2 values. As expected, the association gets closer to the expected results when FRY and DMY were analyzed in terms of SCA units (Plots B and D, Figure 2), since the regression coefficients are not as negative as in the plots on the left of the figure. In every case, however, r^2 values were small. Perhaps the poor association between genetic distances and FRY or DMY may be the result of the strong selection pressure by two important diseases: bacterial blight (*Xanthomonas axonopodis* pv. *Manihotis*) and

super-elongation (*Sphaceloma manihoticola* (Teleomorph: *Elsinoe brasiliensis*). Reaction to these diseases has strong impact on yield but would not be reflected in the genetic distance measured.

The last diallel set was evaluated in the mid-altitude valleys environment and relevant results are presented in Table 4 and Figure 3. Only 21 families derived from 7 progenitors could be analyzed for this environment. The cross between P2 and P8 showed the second best FRY yield (60.5 tha^{-1}), which was among the three with largest genetic distance among progenitors (0.286). Equally promising was the fact that cross 1x9 had one of the lowest averages for FRY (38.0 t ha^{-1}) and also had a small genetic distance (0.207). The same comments can be made for the DMY performance of these two families (measured in t ha^{-1}). The best five performing families for FRY had an average genetic distance of 255, whereas the worst five had an average genetic distance among their respective progenitors of 214. Similar conclusions can be drawn from the analysis of DMY, with average genetic distances among progenitors of the best and worst five families of 255 and 220, respectively.

In every case the relationship between Nei´s genetic distance and FRY or DMY showed a positive regression line (Figure 3). There is no apparent improvement when the association was analyzed for these variables based on SCA values (Plots B and D, Figure 3), compared with similar analyses based on t ha^{-1} (Plots A and C, Figure 3). There was no improvement of the association for DMY compared with FRY as had been

5585 Guilford Rd., Madison WI 53711

312 observed for the sub-humid environment. As was the case for the two previous
313 environments, r^2 values for the mid altitude valleys were small.
314
315 Results presented in Table 5 summarize those from the three diallel studies reported
316 earlier for FRY and presents new information for DMY which had not been analyzed
317 previously. A striking feature of the information presented in this table is the relatively
318 large magnitude of non-additive genetic effects estimated by σ^2_D. These diallel studies
319 made a significant contribution by implementing, for the first time (to the best knowledge
320 of the authors), the test for epistasis in diallel crosses. This test was significant in most
321 cases. Epistasis has also been found to be relevant for grain yield in maize (Lamkey et
322 al., 1995; Wolf and Hallauer, 1997; Crow, 2000, Kang, 2002 among many more reports
323 in the literature). As in the case of cassava, additive and dominance genetic effects
324 explain a great proportion of genetic variation. Performance of the best hybrids (in
325 maize as well as in cassava), therefore, depends mainly on additive and dominance
326 variance, but gets an extra boost from epistasis. In other words, what distinguishes the
327 success of best commercial maize hybrids or cassava clones from the rest is the extra
328 bit of genetic superiority derived from epistatic effects (Crow, 2000). More recent
329 research at the molecular level have exposed unexpected phenomena related to
330 heterosis such as a high degree of non-colinearity among progenitors and unequal
331 expression of alleles (Hochholdinger and Hoecker, 2007). All these effects acting
332 together explain the high complexity of heterosis.
333

5585 Guilford Rd., Madison WI 53711

334 The asexual propagation of cassava allows for the estimation of within-family genetic
335 variation. This is not possible in cereal and legume crops that cannot be propagated
336 asexually. In cassava, on the other hand, all genetic variation can be partitioned into
337 between and within family components. Results from the three diallel studies showed a
338 large proportion for the within-family genetic variation (≈ 90 %). The large within-family
339 component of variation reflects what breeders observe in the field during the selection
340 process. Selection based on GCA or genomic estimated breeding values (Meuwissen et
341 al., 2001; Heffner et al., 2009) would have limited value when such a large variation
342 occurs within families.
343
344 FRY and DMY are two key traits of economic relevance that can be improved through
345 conventional breeding approaches (Kawano et al., 1998). However, if non-additive
346 genetic effects are important for these traits (data presented in Table 5 highlight how
347 important they are), then a method to exploit these effects more efficiently would be
348 highly desirable. RRS has been used successfully and consistently for the exploitation
349 of heterosis for many years and in many different crops (Bernardo, 2014). For the
350 proper implementation of RRS, however, two or more heterotic populations are required
351 (Melchinger and Gumber, 1998; Hallauer and Miranda, 1981; Melchinger and Gumber,
352 1998; Parentoni et al., 2001). Unfortunately, so far, no heterotic patterns have been
353 reported in cassava, in spite of its relevance.
354
355 Results from this study would suggest that genetic distances cannot be used as reliable
356 predictors for those specific crosses where heterosis would occur. Moreover, in several

5585 Guilford Rd., Madison WI 53711

cases there was even a negative association between genetic distances and the yield of cassava genotypes in t ha^{-1} or estimated as SCA. The coefficients of determination (r^2) of the regression analyses presented in Figures 1-3 ranged from 0.000 to 0.280 (SCA for DMY in the acid soils savannas and the sub-humid environment, respectively). In some cases the association between genetic distances was slightly better when the response variable was estimated as SCA rather than in t ha^{-1}. This was expected as SCA estimates deviations from the expectations based on GCA of the two progenitors and are more closely associated with the non-additive component of heterosis. However, the differences were small and irrelevant for the potential identification of heterotic groups based on genetic distances. These observations agree with those for other crops (Cress, 1966; Crossa et al., 1987; Diers et al., 1996; Fu et al., 2014; Ghaderi et al., 1984; Pérez-Velázquez et al., 1995; Riday et al., 2003; Zhang et al., 2006). Nei's genetic distances used in this study assigns equal weight to each of the 95 SNPs used. However, only certain regions of the genome are responsible for the expression of heterosis and, therefore, genetic distances using markers linked to these specific regions would result in considerably better predicting capabilities (Riday et al., 2003).

In many cases genetic distances (based on different definitions and using different type of molecular markers) have been indeed positively associated with heterosis (Ali et al., 1995; Betrán et al., 2003; Kang, 2002; Lanza et al., 1997; Reif et al., 2003). The use of inbred progenitors and the availability of pre-existing heterotic groups have facilitated these positive associations between genetic distances and heterosis. The positive

5585 Guilford Rd., Madison WI 53711

association between genetic distance and heterosis (when found) has been more commonly applied for assignment of new germplasm to (pre-existing) heterotic groups, rather than for yield prediction. Population structure, therefore, influences the relationship between genetic distances and heterosis (Cheres et al., 2000; Crossa et al., 1987; Melchinger, 1999; Pérez-Velásquez et al., 1995; Xiao, 1996). In the case of cassava, there is no appropriate population structure and therefore, genetic distance based on the markers used failed to consistently explain heterosis. Two alternatives would change this situation: **a)** develop a population structure that will allow the emergence of heterotic patterns that genetic distances can detect; and/or **b)** identifying non-neutral markers that are closely associated with heterosis.

A strategic effort needs to be made in cassava to develop a population structure that would facilitate the creation or identification of heterotic groups. Identification of heterotic groups could better focus on diverse gene pools that have evolved isolated from each other over a long period of time (Saxena and Sawargaokar, 2014). Melchinger suggested in 1999 an approach for identifying and using these "diverse gene pools" taking advantage of molecular markers: "*When a large number of germplasm exists but no established heterotic groups are available, it is important to first identify groups of genetically similar germplasm….this can be accomplished most accurately and reliably by genetic distance estimates based on DNA markers. In a second step, one can then produce and evaluate diallel or factorial crosses among representative genotypes from each group….Finally promising groups can be selected as heterotic groups or patterns based on mean hybrid performance and other criteria.*"

5585 Guilford Rd., Madison WI 53711

CIAT has been working on the definition of diverse gene pools from its large germplasm collection using SNPs markers. Eight subpopulations have emerged from this diversity study (Becerra López-Lavalle, 2015). Representatives of each pool could be used for Melchinger's second step. Alternatively, progenitors of successful hybrids (such as the widely grown clone KU50 developed in Thailand but grown in many countries in SE Asia) can be used as a source of partially (or fully) inbred lines that can eventually lead to an approximation of the gametes that gave rise to that particularly outstanding hybrid. When promising heterotic groups are identified, the relative contribution of each SNP to the expression of heterosis could be analyzed which could lead to the identification of non-neutral markers.

5585 Guilford Rd., Madison WI 53711

4. References

Becerra Lopez-Lavalle, L.A. (2015). Revisiting cassava genetic diversity reveals eco-geographic signature of the crop's domestication. Plant and Animal Genome XXIII. San Diego, January 2015.

Bernardo, R. (2014). Essentials of plant breeding. Stemma Press Woodbury, Minnesota, USA. 252 p.

Cach, N.T., J.C. Pérez, J.I. Lenis, F. Calle, N. Morante and H. Ceballos (2005). Epistasis in the expression of relevant traits in cassava (*Manihot esculenta* Crantz) for subhumid conditions. Journal of Heredity 96(5):586-592.

Cach, T.N., J.I. Lenis, J.C. Pérez, N. Morante, F. Calle and H. Ceballos (2006). Inheritance of relevant traits in cassava (*Manihot esculenta* Crantz) for sub-humid conditions. Plant Breeding 125(2):177-182.

Calle, F., J.C. Pérez, W. Gaitán, N. Morante, H. Ceballos, G. Llano and E. Álvarez (2005). Diallel inheritance of relevant traits in cassava (*Manihot esculenta* Crantz) adapted to acid-soil savannas. Euphytica 144(1-2):177-186.

Ceballos, H., C. Hershey and L.A. Becerra-López-Lavalle (2012). New approaches to cassava breeding. Plant Breeding Reviews 36:427-504.

Cheng S.-H., L-Y. Cao, S.-H. Yang and H.-Q. Zhai (2004). Forty Years' Development of Hybrid Rice: China's Experience. Rice Science 11(5-6): 225–230.

Contreras Rojas, M., J.C. Pérez, H. Ceballos, D. Baena, N. Morante, and F. Calle (2009). Introduction of inbreeding and analysis of inbreeding depression in eight S1 cassava families. Crop Sci. 49:543-548.

437 Cress, C.E. (1966). Heterosis of the hybrid related to gene frequency differences
438 between populations. Genetics 53:269-274.
439 Crossa, J., C.O. Gardner and R.F. Mumm (1987). Heterosis among populations of
440 maize (*Zea mays* L.) with different levels of exotic germplasm. Theor. Appl. Genet.
441 73:445-450.
442 Crow, J.F. The rise and fall of overdominance (2000). Plant Breeding Reviews 17:225-
443 257.
444 de Oliveira, E.J., M.D. Vilela de Resende , V. da Silva Santos, C. Fortes Ferreira, G.
445 Alvarenga Fachardo Oliveira, M. Suzarte da Silva. L. Alves de Oliveira, and C.I.
446 Aguilar-Vildoso (2012). Genome-wide selection in cassava. Euphytica (2012)
447 187:263–276.
448 Doyle, J. and J. Doyle (1990). A rapid total DNA preparation procedure for fresh plant
449 tissue. Focus 12: 13-15.
450 Fu, D., M. Xiao, A. Hayward, Y. Fu, G. Liu, G. Jiang, and H. Zhang (2014). Utilization of
451 crop heterosis: a review. Euphytica 197:161–173.
452 Griffing, B. (1956). Concept of general and specific combining ability in relation to diallel
453 crossing systems. Australian J. Biol. Sci. 9, 463-93.
454 Hallauer, A.R. and J.B. Miranda Fo. (1981). Quantitative Genetics in Maize Breeding.
455 Second Edition. Iowa State Univ. Press. Iowa, USA. 468 pp.
456 Heffner, E.L., M.E. Sorrells, and J.-L. Jannink (2009). Genomic Selection for Crop
457 Improvement. Crop Sci. 49:1–12

5585 Guilford Rd., Madison WI 53711

458 Jaramillo, G., N. Morante, J.C. Pérez, F. Calle, H. Ceballos, B. Arias and A.C. Bellotti
459 (2005). Diallel analysis in cassava adapted to the midaltitude valleys environment.
460 Crop Science 45:1058–1063.
461 Jennings D.L. and C. Hershey (1985). Cassava breeding: a decade of progress from
462 international programmes. p. 89-116. **In**: G.E. Russel (ed.) Progress in plant
463 breeding. Butterworths Press, London, United Kingdom.
464 Jennings D.L. and C.A. Iglesias (2002). Breeding for crop improvement. p. 149-166. **In**:
465 R.J. Hillocks, J.M. Thresh, and A.C. Bellotti. (eds.), Cassava: biology, production
466 and utilization. CABI Publ., Wallingford, United Kingdom.
467 Kang, M.S. (2002). Quantitative genetics, genomics, and plant breeding. CABI
468 Publishing, Wallingford, UK. 400 pp.
469 Kawano, K., K. Narintaraporn, P. Narintaraporn, S. Sarakarn, A. Limsila, J. Limsila, D.
470 Suparhan, V. Sarawat, and W. Watananonta (1998). Yield improvement in a
471 multistage breeding program for cassava. Crop Sci 38: 325-332.
472 Kawuki, R.S., M. Ferguson, M. Labuschagne, L. Herselman and D.-J. Kim (2009).
473 Identification, characterisation and application of single nucleotide polymorphisms
474 for diversity assessment in cassava (*Manihot esculenta* Crantz). Molecular Breeding
475 23: 669-684. doi:10.1007/s11032-009-9264-0.
476 Hochholdinger F., and N. Hoecker (2007). Towards the molecular basis of heterosis.
477 Trends in Plant Science 12(9): 427-432.
478 Lamkey, K.R., B.J. Schnicker and A.E. Melchinger (1995). Epistasis in an elite maize
479 hybrid and choice of generation for inbred line development. Crop Sci. 35:1272-
480 1281.

5585 Guilford Rd., Madison WI 53711

Melchinger, A.E. (1993). Use of RFLP markers for analysis of genetic relationships among breeding materials and prediction of hybrid performance. **In**: D.R. Buxton, R. Shibles, R.A. Forsberg, B.L. Blad, K.H. Asay, G.M. Paulsen, and R.F. Wilson (eds) International Crop Science I. pp. 621-628.

Melchinger, A.E. and R.K. Gumber (1998). Overview of heterosis and heterotic groups in agronomic crops. **In**: K.R. Lamkey and J.E. Staub (Eds) Concepts and breeding of heterosis in crop plants. pp. 29-56. Crop Science Society of America, Madison, WI (USA).

Melchinger, A.E. (1999) Genetic diversity and heterosis. **In**: J.G. Coors and S. Pandey (eds.). Genetic and exploitation of heterosis in crops. pp 99-118. American Society of Agronomy, Inc. Madison, WI, USA

Meuwissen, T.H.E., B.J. Hayes and M.E. Goddard (2001). Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Microsoft Corporation, 2004. www.microsoft.com

Nei, M. and Li, W.H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences 76(10): 5269-5273.

Nelson, S.L., 2000. Office 2000. Manual de referencia. McGraw-Hill / Interamericana de España. Madrid, Spain.

Norton, R. (2014). Global starch market outlook and feedstock economics. Cassava World. Africa 2014. Centre for Management Technology (CMT). Lusaka, Zambia 20-20 March.

503 Parentoni, S.N., J.V. Magalhães, C.A.P. Pacheco, M.X. Santos, T. Abadie, E.E.G.
504 Gama, P.E.O. Guimarães, W.F. Merielles, M.A. Lopes, M.J.V. Vasconcelos and E.
505 Paiva (2001). Heterotic groups based on yield-specific combining ability data and
506 phylogenetic relationship determined by RAPD markers for 28 tropical maize open
507 pollinated varieties. Euphytica 121:197-208.
508 Peña-Venegas, C.T. Stomph, G. Verschoor,, L.A. Becerra Lopez-Lavalle, and P. Struik
509 (2014). Differences in Manioc diversity among five ethnic groups of the Colombian
510 Amazon. Diversity 6: 792-826.
511 Pérez-Velásquez, J.C., H. Ceballos, S. Pandey and C. Díaz-Amaris (1995). Analysis of
512 diallel crosses among Colombian landraces and improved populations of maize.
513 Crop Science 35:572-578.
514 Saxena, K.B. and S.L. Sawargaokar (2014). First information on heterotic groups in
515 pigeonpea [*Cajanus cajan* (L.) Millsp.]. Euphytica 200:187-196.
516 Spielman, D.J., D. E. Kolady and P.S. Ward (2013). The prospects for hybrid rice in
517 India. Food Security 5:651-665.
518 Stapleton, G. (2012). Global starch market outlook and competing starch raw materials
519 for starches by product segment and region. Cassava Starch World 2012. Centre for
520 Management Technology (CMT). Phnom Penh, Cambodia 22-24 February.
521 Wang, C., Z. Lentini, E. Tabares, M. Quintero, H. Ceballos, B. Dedicova, C. Sautter, C.
522 Olaya and Z. Peng (2011). Microsporogenesis and pollen formation in cassava
523 (*Manihot esculenta* Crantz). Biologia Plantarum 55(3):469-478.
524 Westwood, N.N. (1990). Maintenance and storage: clonal germplasm. Plant Breed.
525 Rev. 7:111-128.

5585 Guilford Rd., Madison WI 53711

526 Wolf, D.P. and A.R. Hallauer (1997). Triple testcross analysis to detect epistasis in
527 maize Crop Sci. 37:763-770.

528

5585 Guilford Rd., Madison WI 53711

Figure 1. Relationship between Nei's genetic distance (horizontal axis in each plot) with fresh root yield (top plots) and dry matter yield (bottom plots) in diallels conducted in the sub-humid environment. Plots on the left illustrate the relationship of Nei´s genetic distance with the actual units used to estimate fresh and dry matter yield. Plots on the right present the relationship between genetic distances and specific combining ability estimates for the two variables.

Figure 2. Relationship between Nei's genetic distance (horizontal axis in each plot) with fresh root yield (top plots) and dry matter yield (bottom plots) in diallels conducted in the acid soil savannas. Plots on the left illustrate the relationship of Nei´s genetic distance with the actual units used to estimate fresh and dry matter yield. Plots on the right present the relationship between genetic distances and specific combining ability estimates for the two variables.

Figure 3. Relationship between Nei's genetic distance (horizontal axis in each plot) with fresh root yield (top plots) and dry matter yield (bottom plots) in diallels conducted in the mid-altitude valleys. Plots on the left illustrate the relationship of Nei´s genetic distance with the actual units used to estimate fresh and dry matter yield. Plots on the right present the relationship between genetic distances and specific combining ability estimates for the two variables.

5585 Guilford Rd., Madison WI 53711

553 Table 1. List of progentiors used in the three diallels whose results were reported earlier
554 (Cach et al., 2005; 2006; Calle et al., 2005; Jaramillo et al., 2005; Perez et al., 2005a;
555 2005b).

Progenitor	Environment		
	Acid Soils	Mid-altitude valleys	Sub-humid
1	CM 4574 - 7	CM 6740-7	MTAI 8
2	CM 6740 - 7	SM 1219-9	CM 6754 - 8
3	CM 7033 - 3	SM 1278-2	CM 8027 - 3
4	SM 1219 - 9	SM 1636-24 [a]	SM 805 - 15
5	SM 1565 - 15	SM 1673-10 [a]	SM 1565 - 17
6	SM 2058 – 2 [a]	SM 1741-1	SM 1411 - 5
7	SM 2219 - 11	HMC 1	SM 1219 - 9
8	HMC 1	M ECU 72	SM 1657 - 12 [a]
9	MPER 183	MPER 183	SM 1665 - 2
10	MTAI 8		

556 [a] Progenitor no longer available for the measurement of genetic distances

557

558

5585 Guilford Rd., Madison WI 53711

559 Table 2. Nei`s genetic distances and results from a diallel conducted in the sub-humid
560 environment of Colombia.

Cross	Nei´s genetic distance	Fresh root yield		Dry matter yield	
		(t/ha)	SCA units	(t/ha)	SCA units
1x2	0.255	35.1	1.32	9.6	0.11
1x3	0.124	26.5	-9.60	7.5	-2.91
1x4	0.216	31.4	-0.92	9.1	0.10
1x5	0.216	42.3	1.80	11.3	0.64
1x6	0.176	36.5	-2.05	10.5	-0.56
1x7	0.191	42.4	3.55	12.1	1.14
1x9	0.195	45.7	5.91	12.4	1.48
2x3	0.261	32.8	0.02	9.6	0.05
2x4	0.239	27.6	-1.41	7.6	-0.53
2x5	0.194	35.5	-1.67	9.5	-0.26
2x6	0.195	38.0	2.71	10.8	0.63
2x7	0.186	34.8	-0.74	9.8	-0.29
2x9	0.264	36.2	-0.23	10.4	0.28
3x4	0.230	34.3	2.87	10.1	1.02
3x5	0.247	41.0	1.51	11.2	0.49
3x6	0.194	38.9	1.32	11.4	0.31
3x7	0.202	39.4	1.52	11.3	0.40
3x9	0.249	41.1	2.35	11.6	0.64
4x5	0.193	37.2	1.42	9.6	0.24
4x6	0.246	35.6	1.74	10.1	0.31
4x7	0.192	34.0	-0.09	9.5	-0.08
4x9	0.142	31.4	-3.60	8.6	-1.06
5x6	0.274	41.0	-1.00	11.3	-0.04
5x7	0.230	42.7	0.47	11.1	-0.08
5x9	0.157	40.7	-2.53	10.3	-0.98
6x7	0.193	37.5	-2.76	10.9	-0.69
6x9	0.246	41.3	0.04	11.7	0.04
7x9	0.194	39.6	-1.95	11.1	-0.40
Minimum	0.124	26.50	-9.60	7.50	-2.91
Maximum	0.274	45.70	5.91	12.40	1.48
Average	0.211	37.16	0.00	10.36	0.00

561

562

5585 Guilford Rd., Madison WI 53711

563 Table 3. Nei`s genetic distances and results from a diallel conducted in the acid-soils
564 savannas of Colombia.

Cross	Nei´s genetic distance	Fresh root yield		Dry matter yield	
		(t/ha)	SCA units	(t/ha)	SCA units
1 x 2	0.239	25.90	3.44	8.57	1.13
1 x 3	0.226	19.99	-0.55	6.40	-0.27
1 x 4	0.181	23.53	0.63	7.75	0.18
1 x 5	0.233	17.82	-3.48	6.20	-0.97
1 x 7	0.197	26.50	1.13	8.44	0.13
1 x 8	0.253	16.24	-4.16	5.46	-1.25
1 x 9	0.206	21.49	2.91	6.78	0.92
1 x 10	0.231	21.27	0.07	7.27	0.12
2 x 3	0.172	19.12	-0.58	6.27	-0.12
2 x 4	0.208	18.59	-3.47	6.09	-1.21
2 x 5	0.171	21.38	0.92	7.31	0.42
2 x 7	0.231	24.96	0.43	8.05	0.02
2 x 8	0.211	17.39	-2.18	5.84	-0.60
2 x 9	0.207	16.83	-0.92	5.15	-0.43
2 x 10	0.172	22.73	2.37	7.65	0.78
3 x 4	0.208	21.52	1.38	6.97	0.44
3 x 5	0.190	19.80	1.26	6.67	0.54
3 x 7	0.181	18.94	-3.68	6.10	-1.17
3 x 8	0.134	17.11	-0.53	5.33	-0.34
3 x 9	0.225	18.72	2.89	5.85	1.04
3 x 10	0.199	18.25	-0.19	5.97	-0.13
4 x 5	0.188	21.89	0.98	7.38	0.35
4 x 7	0.240	25.13	0.15	8.25	0.08
4 x 8	0.211	18.68	-1.33	6.27	-0.31
4 x 9	0.221	17.13	-1.06	5.30	-0.42
4 x 10	0.191	23.52	2.72	7.89	0.89
5 x 7	0.179	20.61	-2.76	6.87	-0.90
5 x 8	0.170	22.76	4.36	7.54	1.36
5 x 9	0.324	18.68	2.09	5.81	0.50
5 x 10	0.229	15.83	-3.37	5.31	-1.30
7 x 8	0.225	26.53	4.05	8.50	1.19
7 x 9	0.272	19.06	-1.60	6.09	-0.36
7 x 10	0.238	25.54	2.27	8.74	1.00
8 x 9	0.289	15.35	-0.34	4.89	0.03
8 x 10	0.255	18.43	0.12	6.06	-0.08
9 x 10	0.187	12.51	-3.98	3.99	-1.29
Minimum	0.134	12.51	-4.16	3.99	-1.30
Maximum	0.324	26.53	4.36	8.74	1.36
Average	0.214	20.27	0.00	6.64	0.00

565

566 Table 4. Nei`s genetic distances and results from a diallel conducted in the mid-altitude
567 valleys of Colombia.

Cross	Nei´s genetic distance	Fresh root yield		Dry matter yield	
		(t/ha)	SCA units	(t/ha)	SCA units
1x2	0.208	50.63	-0.66	16.79	-0.79
1x3	0.181	42.91	1.40	15.49	0.55
1x6	0.195	50.02	5.37	17.57	1.69
1x7	0.211	48.38	3.22	16.36	0.76
1x8	0.236	53.49	4.24	18.18	1.71
1x9	0.207	37.98	-13.57	13.63	-3.91
2x3	0.211	48.56	1.39	16.94	0.56
2x6	0.238	44.83	-5.48	15.97	-1.36
2x7	0.211	44.75	-6.06	15.13	-1.91
2x8	0.286	60.48	5.58	19.50	1.58
2x9	0.221	62.44	5.23	20.92	1.92
3x6	0.230	39.51	-1.02	13.99	-0.70
3x7	0.218	36.36	-4.68	12.91	-1.50
3x8	0.277	46.06	0.93	15.97	0.69
3x9	0.279	49.41	1.98	16.76	0.41
6x7	0.267	45.37	1.19	15.88	0.53
6x8	0.236	41.48	-6.79	13.95	-2.27
6x9	0.244	57.30	6.74	19.42	2.12
7x8	0.365	50.15	1.38	16.41	0.47
7x9	0.289	56.03	4.95	18.66	1.65
8x9	0.278	49.83	-5.33	15.71	-2.18
Minimum	0.181	36.36	-13.57	12.91	-3.91
Maximum	0.365	62.44	6.74	20.92	2.12
Average	0.242	48.38	0.00	16.48	0.00

568

569

5585 Guilford Rd., Madison WI 53711

570

571 **Table 5**. Variance estimates (standard errors within parenthesis) for fresh root yield and
572 dry matter content in three different diallel sets evaluated in the three most relevant
573 environments for cassava in Colombia.

Genetic parameter	Fresh root yield (t ha^{-1})			Dry matter yield (t ha^{-1})		
	Acid soil	Sub-humid	Mid-altitude	Acid soil	Sub-humid	Mid-altitude
σ^2_G (Between)	1.65 (2.95)	13.09 (4.74)	42.78 (13.27)	0.24 (0.31)	0.69 (0.35)	3.56 (1.40)
σ^2_G (Within)	21.08 (2.30)	127.21 (7.65)	288.93 (1918)	2.06 (0.24)	9.97 (0.61)	33.88 (2.30)
σ^2_A	-1.49 (6.32)	17.82 (13.75)	11.88 (24.67)	-0.03 (0.66)	0.74 (0.93)	-1.64 (2.28)
σ^2_D	9.03 (7.93)	23.87 (11.15)	152.11 (49.08)	0.99 (0.85)	1.59 (0.92)	16.86 (5.81)
Epistasis test	15.05 (6.74)	100.40 (12.74)	168.91 (39.72)	1.33 (0.71)	8.40 (0.47)	22.06 (4.04)

574

575

576

577

578

579

580

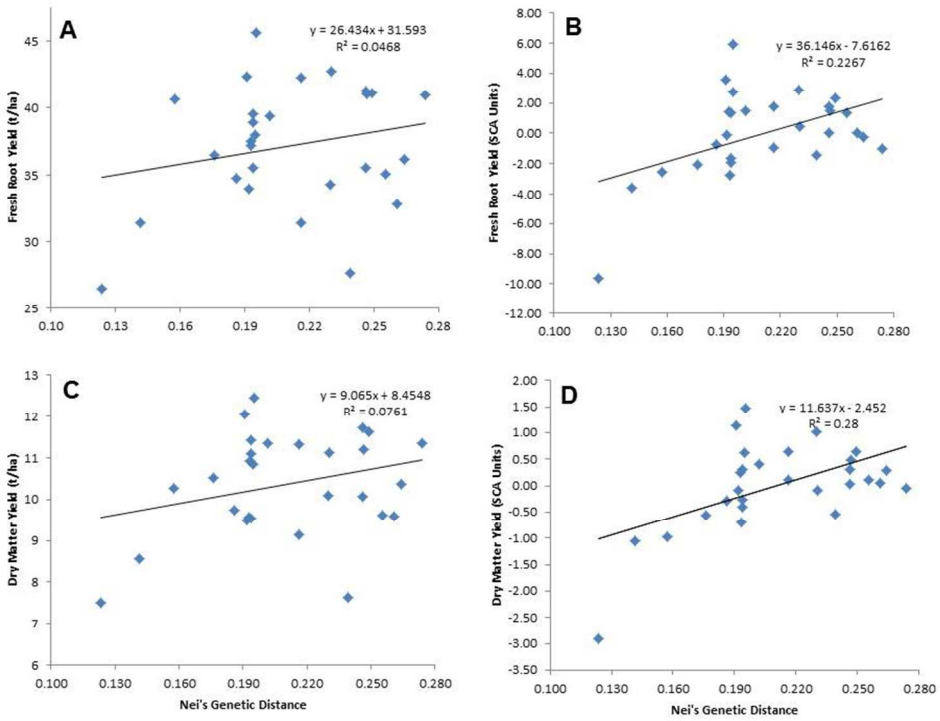

Figure 1. Relationship between Nei's genetic distance (horizontal axis in each plot) with fresh root yield (top plots) and dry matter yield (bottom plots) in diallels conducted in the sub-humid environment. Plots on the left illustrate the relationship of Nei´s genetic distance with the actual units used to estimate fresh and dry matter yield. Plots on the right present the relationship between genetic distances and specific combining ability estimates for the two variables.
254x190mm (96 x 96 DPI)

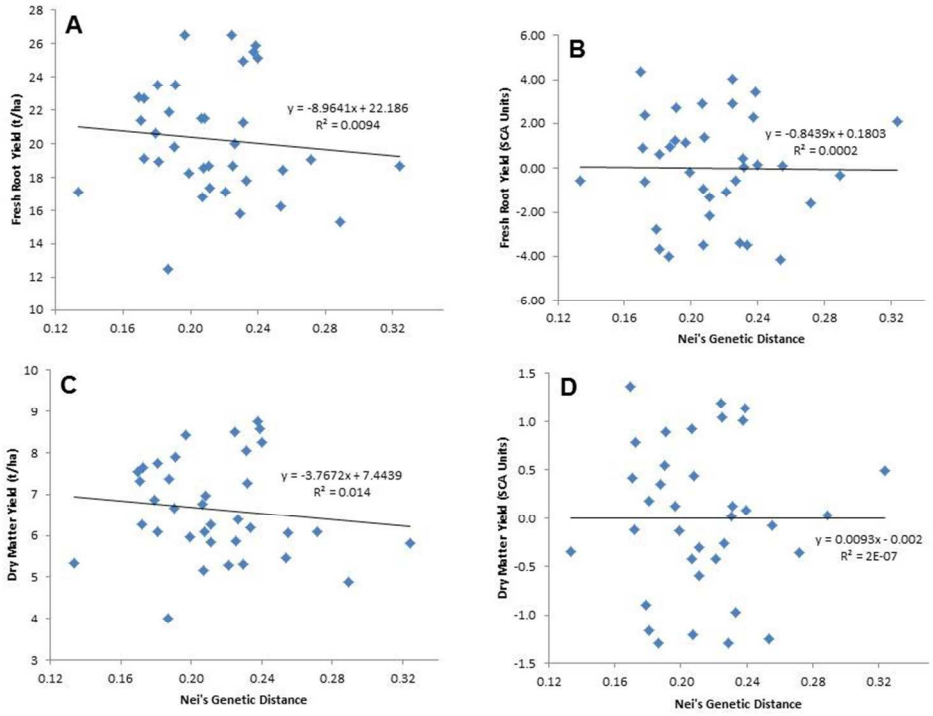

Figure 2. Relationship between Nei's genetic distance (horizontal axis in each plot) with fresh root yield (top plots) and dry matter yield (bottom plots) in diallels conducted in the acid soil savannas. Plots on the left illustrate the relationship of Nei´s genetic distance with the actual units used to estimate fresh and dry matter yield. Plots on the right present the relationship between genetic distances and specific combining ability estimates for the two variables.
254x190mm (96 x 96 DPI)

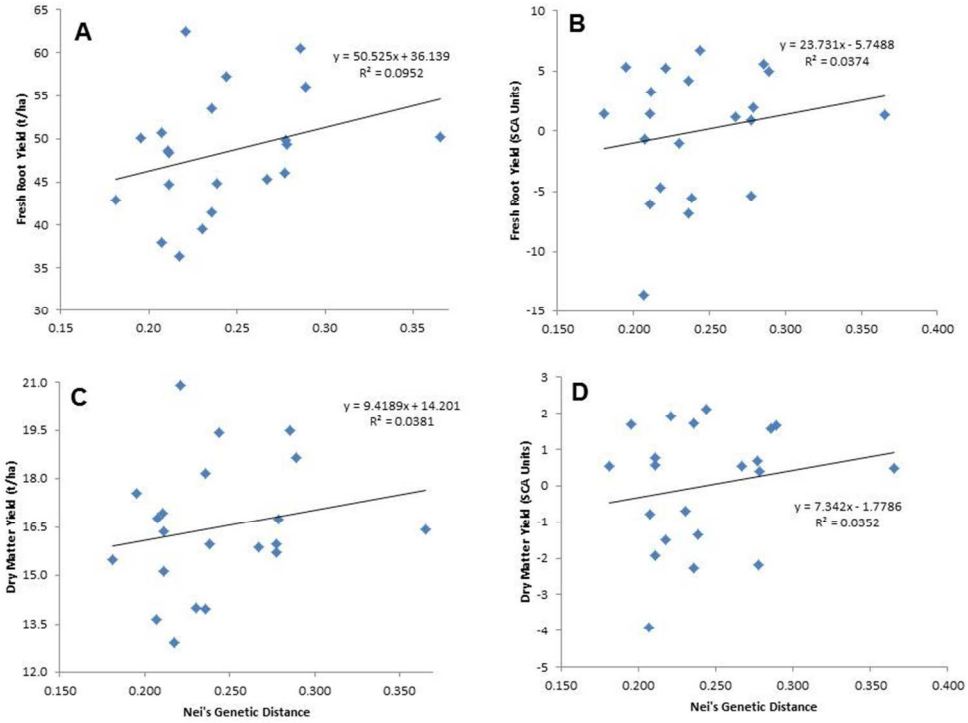

Figure 3. Relationship between Nei's genetic distance (horizontal axis in each plot) with fresh root yield (top plots) and dry matter yield (bottom plots) in diallels conducted in the mid-altitude valleys. Plots on the left illustrate the relationship of Nei´s genetic distance with the actual units used to estimate fresh and dry matter yield. Plots on the right present the relationship between genetic distances and specific combining ability estimates for the two variables.
254x190mm (96 x 96 DPI)