Supplementary Table 1. Notes on multivariate morphometric studies of *Anacardium occidentale* and *A. humile* published between 1999-2017, tabled in chronological order. Some studies were not obtainable online and not seen, but are reported here as cited by other authors. Only postgraduate and undergraduate studies that are available online are reported here.

|  |  |
| --- | --- |
| Barros (1991) | One of the first multivariate analysis of morphometric and molecular (isozyme) characters in cashew. This is summarized in Paiva et al. (2003) [see below] |
| Archak et al. (2003) | Compared DNA fingerprinting results and 27 morphometric variables (replicated in 6 trees per accession) on 19 genotypes of Indian cashew using RAPDs, ISSR and AFLP markers. AFLP discriminated the genotypes best, but the other two techniques also discriminated all genotypes. AFLP similarity matrix was very different from those of RAPD and ISSR, the latter two showing significant correlation (Mantel tests). A morphological similarity matrix was uncorrelated with those of any of the molecular markers. Concluded that AFLP was best option for fingerprinting. Lack of correlation likely due to pronounced allogamy producing diverse genetic backgrounds. |
| Paiva et al. (2003) | These authors included a brief summary of the PhD thesis of Barros (1991) as follows: Barros used botanical and agronomic descriptors and isoenzyme markers of 67 accessions of the Embrapa germplasm collection in Pacajus, Ceará, Brazil, and is an important early attempt at a multivariate analysis of *Anacardium* genotypes. Barros concluded that isoenzyme markers could be successfully used for characterizing and identifying cashew genotypes and those of other species of the genus. Concluded that further studies using more advanced molecular marker techniques were needed in order to improve characterization and also to provide useful genetic information for genetic improvement of the cashew.  Euclidean distances were estimated within each group of accessions, discriminated using PCA with 30 botanical-agronomic variables for all accessions. PCA allowed the identification of six groups, which were then characterized by minimum and maximum mean intra-group Euclidean distances between pairs of genotypes: group 1: seed-propagated early dwarf cashews from Ceará; group 2: seed-propagated common cashews from Pacajus; 3) seed-propagated common cashews from India; 4) seed-propagated common cashews from São Paulo; 5) seed-propagated common cashews from Venezuela; vegetatively propagated common cashews from Ceará.  The greatest [minimum] mean value among pairs of accessions was in group G6, corresponding to common cashew collected in Ceará and propagated vegetatively (accessions 54 and 67). This group also showed the greatest within-group diversity, as shown by the greatest Euclidean distance recorded for intragroup pairs of accessions. |
| Samal et al. (2003) | Analysed genetic relationships in 20 cashew genotypes in an orchard at Orissa University, India (sample of ca. ten per genotype), using 8 morphological characters. Morphological data presented as means of four individuals per genotype, with three replications of each. Gower's coefficient was used to compute distance and UPGMA cluster analysis was then carried out. RAPD markers obtained were analysed using Dice coefficient on binary matrix and UPGMA cluster analysis performed. Four clusters obtained from morphological data. Differences found with molecular clustering; conclusion favoured molecular methods. |
| Lucena (2006) | Studied genetic diversity in cashew genotypes from a natural population [or maybe an informal orchard??] at the edge of the town of Boa Vista, Roraima, using 14 morphological and physico-chemical characteristics of the apple and nut. Two separate data sets were gathered, one of 30 plants and the other of 10 plants. Data analysis used ANOVA (for each variable among genotypes), canonical variables [looks as if she used PCA really as there is no sign of a priori groupings used] and Tocher cluster analysis based on Mahalanobis distance. Environmental and Genetic Coefficients of Variation and Heritability were calculated. Detailed discussion given of genetic variation for each variable in each data set. She found three groups using Tocher method, two of which consisted of a single individual. Further analysis of the main group of 28 individuals yielded 9 subgroups, of which three consisted of a single individual. She cites "(var. nanum)" for early dwarf cashew and says it is natural to the cerrados of Amazonia. |
| Aliyu & Awopetu (2007) | Aliyu & Awopetu (2007) studied 59 accessions of cashew at the Nigerian germplasm collection in Ibadan originating from Nigeria, India and Brazil. They scored 36 quantitative and 33 qualitative characters [> 50% of characters are from the apple and nut) and analysed them with Complete linkage cluster analysis and PCA. They found four groups. Brazilian genotypes clustered together in two groups (but also some were found in the other two clusters), and Indian and Nigerian genotypes predominated in the other two. Each cluster had characteristic fruit and growth habits. |
| Pessoni (2007) | The study is an evaluation of the genetic diversity in species of *Anacardium* (*A. occidentale* comum and anão precoce, *A. othonianum* and *A. microcarpum*).  Chapter 1 analysed quantitative phenotypic variablesusing 69 accessions in the Pacajus (Ceará) germplasm collection and a sample of 23 individuals of a subspontaneous population from a restinga in northern Rio state. 13 quantitative IBPGR phenotypic descriptors were used (10 measurements per plant when possible): inflorescence length, inflorescence width, number of branches in inflorescence, hypocarp length, diameters of base and apex of hypocarp, hypocarp weight, drupe length, width and thickness, drupe weight, maximum height of plant, mean diameter of canopy (N-S, E-W). Discriminant analysis and PCA, cluster analysis using PCs, UPGMA and Tocher optimization were used on quantitative phenotypic data. Genetic divergence by taxon was investigated by discriminant analysis based on principal components and Anderson classification functions. The accessions were classified into three species/ecotypes: common cashew (*A. occidentale*), dwarf cashew (*A. occidentale*) and *A. othonianum*. 14 accessions were left unclassified *a priori* and were classified *a posteriori* using Anderson classification functions. Divergence by geographical origin was investigated: six *a priori* groups were formed: Nordeste (Ceará and Paraíba), Dwarf (Maranguape, CE), Rio de Janeiro (from RJ subspontaneous population), Roraima, India, othonianum (from cerrado of Central Brasil). 12 accessions were left ungrouped and later classified using Anderson DA. ANOVA was carried out on the variables using the factors taxon and geography. DA and unrooted UPGMA tree based on Mahalanobis distance were used to analyse the data set. PCA with standardized data was used. Tocher optimization was also used and the groups obtained compared with those obtained by Anderson DA. Among taxa, common cashew showed greatest variation in most variables, but *othonianum* had largest means (but mostly not significant) for inflorescence variables and significantly smaller means for apple and nut variables. No significant differences between common and dwarf cashew was found in any variable. DA showed clear separation of *othonianum* from the other two taxa and weak overlapping separation of the other two. Among geographical groups, *othonianum* had highest means of inflorescence characters and lowest means of apple and nut characters and Rio de Janeiro the lowest means for variables of the other five groups. These were the two groups that were most distinct in the DA ordination and cross-validation. India was also 100% distinct in cross validation. UPGMA using Mahalanobis distance showed strong separation of *othonianum*, and Nordeste and Dwarf the closest. Seven groups were detected with Tocher optimization [which didn't corresppond very well to the a priori groupings by taxon or geography] and presented in a PCA. Most accessions were measured in same environment (Pacajus) which reduces variability due to environment [unlike taxonomic measurement of natural populations]. Variable measurements used are means of multiple measurements per individual. Using different multivariate methods helps to find consistent patterns. The analyses highlighted the difference of *othonianum* from common and dwarf cashew. Results support the idea that dwarf cashew originated in Roraima. The accession identified as *A. microcarpum* was assigned to common cashew. One accession from Piauí matched *othonianum*. In geographical groups the dwarf cashew was no longer distinguishable from common cashew. *Othonianum*  accessions are more diverse than those from Rio, due to their origin in various localities while RJ is a true single population sample. Results support recognition of *othonianum* as separate species.  The ISSR molecular analysis (Chapter 2) used 91 accessions from Pacajus germplasm collection and 45 individuals of the Rio restinga subspontaneous population. Sorensen and Nei & Li coefficients of similarity were used. Minimum number of markers needed to estimate genetic dissimilarity between accessions was computed (n=155) by correlation of dissimilarity matrices. Non-parametric discriminant analysis using k-nearest neighbours was applied using same two grouping models as before: taxa (species, ecotypes, included *humile* in this analysis) and geographical origin. He used a method (projeção gráfica das distancias in tres dimensões) to visualize relationships, which sounds like Non-metric multidimensional scaling. AMOVA was computed using the geographical grouping models. Pairwise ΦST values and Nei & Li distance were used to construct UPGMA and Ward trees and tested using cophenetic correlation. KNN analysis grouped common and dwarf cashews groups very successfully but very poorly for *othonianum* and *humile*. The two accessions originally identified as *microcarpum* (of unknown origin) were grouped with dwarf cashew. In geographical groups KNN also produced poor discrimination for *othonianum* and *humile* and Nordeste but good for Dwarf, Rio, Roraima and India. In the NDMS analysis the Rio population was the most strongly separate with Dwarf a not very good second; high stress values suggest a weak result. AMOVA showed 72.5% intragroup (geographical) variation and pairwise ΦST values were almost all significant. UPGMA and Ward cluster analysis showed distinctness of Rio and similarity of *humile* and *othonianum*. Pessoni notes that the accessions of *humile* and *othonianum* are small or single samples from a variety of locations, which would tend to obscure any natural population structure within each species in this study, e.g. compare RJ sample from a single population. The AMOVA result is similar to those obtained in other arboreal perennial allogamous species.  Chapter 3 is an analysis using 18 qualitative phenotypic characters from IBPGR (1986) list, and 69 accessions at the Pacajus germplasm collection. The characters were: leaf shape, leaf apex shape, angle of insertion of leaf on stem, cross sectional shape of leaf, hypocarp shape, hypocarp base shape, hypocarp apex shape, hypocarp apex cavity, hypocarp mature colour, drupe base shape, drupe suture shape, drupe position of suture relative to apex, drupe shape of sides, drupe apex shape, inflorescence general shape, inflorescence compactness, inflorescence branching type, plant growth habit. Data were coded as binary for each character state. Character state relative frequency was calculated for each character and normalized Shannon-Wiener diversity indexes were estimated. A dissimilarity index which ignored zero matches was used ((b+c)/ (a+b+c)). Discriminant analysis by k-nearest neighbour and mean nearest neighbour was used on the dissimilarity matrix, using model of two species, *occidentale* and *othonianum* and ignoring *microcarpum* and *spp.* accessions. NMDS was also used to project the points in 3D space. UPGMA and Ward's cluster analysis were also used. Most qualitative phenotypic characters varied considerably according to Shannon-Wiener diversity index. K-nearest neighbours DA separated the two taxa better than k-mean neighbours according to apparent error rate, but neither provided good discrimination for both species simultaneously. NMDS showed no separation of the two taxa. UPGMA failed to differentiate the two taxa. Thus unsupervised classification failed.  Chapter 4 compared the results of the three analytical approaches, quantitative and qualitative morphometrics and molecular. Data from 69 accessions at Pacajus were used. 13 quantitative morphometric variables, 18 qualitative morphological variables were used. Dissimilarity matrices were computed, Euclidean distance for quantitative, the dissimilarity measure used in Chapter 3 for qualitative, and complement of Sorensen or Nei & Li for ISSR. Mantel tests of correlation between the three matrices were used and the matrices were also summed. KNN analyses (using 3 taxon categories of common cashew, dwarf cashew and *othonianum*) were carried out on the three separate matrices, and on the three pooled combinations of two matrices and on pooled all three (see their Table 2). Molecular matrix showed overall least errors in KNN of the three taxa (qualitative matrix the most) but poor performance in discriminating *othonianum*. But *othonianum* was discriminated best by pooled quantitative and qualitative. Only quantitative and ISSR showed correlation, not very strong but significant. The KNN of the pooling of these two matrices gave best average discrimination of the three taxa. NMDS in 3D of distance matrix of the pooled matrices quantitative and ISSR showed clear distinction of *othonianum* from the rest. Tocher's method produced different results for the three matrices. Concluded that combining non-correlated data sets is not to be recommended, since they are expressing contrasting patterns of variation from different parts of the genome.  Natural populations from Rio and Roraima states were different from those of the Northeast.  Overall conclusions: DA adequate to discriminate taxa using morphometric data. *Othonianum* is distinct morphologically and because carried out on genotypes growing in same conditions (Pacajus) must mostly be genetically based. UPGMA of morphometric data reasonable. KNN of molecular data failed to distinguish species, but distinguished subgroups of *occidentale*. AMOVA results similar to those of other outcrossing tree species. Molecular data showed RJ population as most distinct molecularly but not morphometrically.  Further studies are needed to compare natural populations of *A. othonianum* and *A. humile* to evaluate their taxonomic status. Pessoni mentions the recognition by Mitchell & Mori of ecotypes for restinga, and cerrado. He cites Barros (1995, not listed here) recognizing 21-22 species in classical taxonomy. The drupe matures before the hypocarp. He cites Mitchell & Mori's assertion that in the cerrado species *occidentale, humile* and *nanum* there appear to be few barriers to interspecific crossing, which explains the natural occurrence of intermediate individuals in cerrado habitats. Brief summary of cytological studies. Goldblatt found range 2n=24, 2n=40, 2n=42 in *A. occidentale*; Leao et al. (2001) found 2n=16 in *A. othonianum*. Cites Johnson (1974 – portuguese translation) for description of restinga ecotype. Pessoni cites the differentiation of the restinga ecotype into cajueiro comum and cajueiro anão precoce [this seems a bit doubtful given that most accessions at Pacajus seem to be from cerrado habitats]. He talks of natural populations of anao precoce. He gives descriptions of *humile* and *othonianum* noting the latter as member of the cerrado ecotype of Mitchell & Mori.[there is a description of othonianum in the book chapter by Agostini et al. 2006 – see main paper References]. Summarizes Barros (1991) and says that he concluded that the early dwarf cashew is an ecotype of *A. occidentale*. |
| Desai (2008) | Studied 57 cashew genotypes collected from different parts of Goa state, India, and planted at the ICAR research complex at Old Goa. The 37 morphological descriptors used were selected from the IBPGR list. Phenotypic and genotypic coefficients of variation were computed (PCV, GCV) for 20 quantitative characters, as well as phenotypic and genotypic correlation, heritability, and genetic advance. 20 quantitative attributes were used in a PCA. RAPDs analysis was carried out and analysed with neighbour-joining trees. Extensive sampling and data presented with explicatory photos. Six clusters of [morpho-] genotypes were identified using non-hierarchical Euclidean cluster analysis. RAPDs analysis produced 2 clusters, one composed of 8 subclusters and the other of three. No strong correlation between RAPDs and [morph-] genotype clustering was found. Abundant genetic variability found using morphological features. |
| Rufino et al. (2008) | Used multivariate analysis to classify thirty genotypes of *Anacardium* ("cajuí") from native populations in Piauí, without specifying all their localities; four from the municipality of Monsenhor Gil, and one each from the municipalities of Teresina and Passagem Franca; these are all areas of transitional forest types in the interior of the state. From their introduction it seems likely that other genotypes came from the restinga ecotype populations of Ilha Grande and Parnaíba. The variables used were for apple: total soluble solids (SST), pH, total titratable acid (ATT) and SST/ATT, weight; apple and nut: total weight; nut: length, basal diameter. Data analysis included ANOVA univariate and multivariate with Scott-Knott test, comparison of means, CVA, relative importance of characters, genetic distances, cluster analysis (nearest neighbour with Tocher optimization) using Mahalanobis distance and selection index. Seven groups were detected of which five contained one individual each and two were larger. |
| Archak et al. (2009) | Studied the genetic (AFLP markers) and morphological (27 qualitative and quantitative markers) variation of cashews from populations at four regions along the coast of India, northern and southern zones on the west and east coasts, 91 genotypes in all. They aimed to reconstruct the history of its introduction since the 16th century, estimated to have been between 1563 and 1578 at Cochin. From there it spread along the eastern and western coasts of peninsular India. It is not known if many independent introductions were made at different ports during that early period. Material used was from the National Cashew Genebank at Puttur, Karnataka state [so not using a truly population level approach]. AFLP data analysis used STRUCTURE. Morphometric data gathered from trees older than 10 years, replicating data from three trees per genotype (accession) and was used to compute a pair-wise distance matrix to compare with the AFLP distance matrix using the Mantel test. The botanical characters they used (selected from IBPGR 1986) were: twig internode length, tree height, tree spread, extension growth of twigs, twig diameter, no. leaves per twig, leaf size (area), inflorescence length, inflorescence width, sex ratio, apple length, apple width, nut length, nut width; other characters were included. AMOVA found no genetic (AFLP) differentiation among regions or populations, 100% variance was within populations, so no genetic structure. Morphometric analysis found 88% variance within populations and 12% between. No significant correlation between morphometric and AFLP traits was found. Spread of cashew in India thought to have been primarily anthropogenic. Genetic diversity in Indian cashew estimated as moderate but substantial. Four distinct groups found by STRUCTURE, pointing to multiple original introductions. However, the four genetic groups were uncorrelated with geography, occurring almost equally in the four geographical groups. Therefore the four inferred original founder introductions could have been spread by human action from a single point, e.g. Cochin, as proposed by Johnson (1973). |
| Chipojola et al. (2009) | Studied genetic diversity in 40 genotypes from 4 populations in Malawi using 14 quantitative and 4 qualitative morphological variables; three vegetative characters used: canopy spread, leaf size (area), number of leaves per twig. Inflorescence variables used were number of flowers per panicle, number of male flowers per panicle, number of hermaphrodite flowers, flower sex ratio. Other characters were of nut and apple. UPGMA found 4 clusters and in the PCA the most influential characters were apple length, apple nut ratio, nut weight, kernel weight, out turn percent and flower sex ratio. Cashew well adapted to poor soils and dry sandy locations and is drought resistant, grows best on well drained sandy soils. Samples were taken at two smallholder farmers' orchards and two commercial farmers' orchards, ten trees from each orchard. Flower sex ratio very variable between populations. Wide range of variation found in variables in general, especially in nut and kernel. |
| Carvalho (2011), Carvalho et al. (2012) | Studied genetic variability in plants of *A. humile* ("cajuzinho-do-cerrado") sampled from natural populations in the cerrados of Goiás and Mato Grosso using RAPD markers. High variability among the 122 genotypes was detected within the 11 municipalities sampled. UPGMA and PCoA analyses of Nei & Li distance matrices were carried out. Four groups were detected. Plants of *A. humile* are difficult to separate in the field due to their underground stems and branches. Canopies of 200 m2 are reported which suggests involvement of more than one individual [cf. restinga ecotype]. Carvalho (2011) studied morphological descriptors of *A. humile* in 10 municipalities of Goiás and one municipality in Mato Grosso. Sampled 122 plants [no. of populations not specified]. Eight morphological characters used: plant height, canopy diameter, trunk height (height of first branch), trunk diameter (at its middle), number of lateral branches, mature leaf length, mature leaf width, presence of disease lesions. PCA and then Tocher's method were used for estimating group number using morphological variables. UPGMA with Nei & Li distance and AMOVA, Nei diversity used for RAPD marker data. Five to seven groups found with morphology. AMOVA found 93% variation within 11 populations. The morphological variables with most influence in the PCA were canopy diameter, plant height and trunk height. |
| Gomes et al. (2011) | Used PCA to study 42 plants of *A. othonianum* (as "cajuí") from natural populations in cerrados of the watershed of the Rio dos Cochos (Municipios de Januária and Cônego Marinho, northern Minas Gerais. The plants are native to the cerrado and intensively used by local people for the fruits and bark. 18 characters were measured: plant height, mean diameter of canopy, diameter of trunk, number of fruits, total fruit mass, drupe mass, drupe length, drupe width, hypocarp mass, hypocarp length, basal diameter of hypocarp, apical diameter of hypocarp, total soluble solids (SST), titratable total acidity (ATT), ATT/SST, pH and dry mass of hypocarp. Three PCs explained 91.7% of variation and allowed the selection of the variables that contribute most to overall variability. The eight most influential variables were stem diameter, mean diameter of canopy, number of fruits, drupe mass, drupe length, drupe width, ATT/SST and pH. Two genotypes were highly divergent from the others. |
| Dasmohapatra (2012), Dasmohapatra et al. (2014) | Studied 25 genotypes of domesticated *A. occidentale* from the Indian germplasm collection at Bhubaneswar using RAPDs and ISSR markers with UPGMA cluster analysis and principal coordinate analysis. They identified 14 clusters of genotypes. They carried out variability studies of the genotypes with respect to agronomically important phenotypic characters and compared heritability, GCV, PCV, genetic advance and correlation with nut yield per tree for each phenotypic character. Genetic divergence was estimated using Mahalanobis D2 statistics. Using 20 characters the genotypes formed eight clusters under CVA using Tocher's method, and cluster analysis using Gower's similarity coefficient. There was no geographical correlation. There was only partial agreement between the clustering results of morphometric and molecular datasets. Molecular similarity did not correlate with their pedigrees, i.e. their known parental relationships. The morphological characters used were: plant height, plant girth, canopy spread (E-W), canopy spread (N-S), number of shoots per sq.m of canopy, number of leaves per twig, leaf area, leaf shape, leaf apex shape, colour of mature leaves, time of flowering, inflorescence shape, inflorescence compactness, inflorescence branching type, inflorescence length, inflorescence breadth, flower colour, number of flowering laterals per sq.m of canopy, flowering durations, sex ratio, number of nuts per panicle, nut shape, nut weight, nut breadth, nut attachment to apple, kernel weight, shelling %, yield per plant, apple shape, apple colour, apple weight, apple length, apple breadth, A/N ratio, apple juice %, apple TSS. |
| Maia et al. (2012) | Reported on genetic divergence in eleven genotypes of cajuí (as *A. spp.*) held in the germplasm collection of Embrapa Meio-Norte in Parnaíba, Piauí, using nine agronomically significant phenotypic variables: total fruit weight, hypocarp weight, drupe weight, basal diameter and apical diameter of the hypocarp, hypocarp firmness, vitamin C and total soluble solids. They used UPGMA cluster analysis and found three genetically divergent groups. |
| Frota Júnior et al. (2014) | Investigated 50 early dwarf cashew genotypes originating in Piauí and Ceará states, using four morphological characters in order to find divergent types which could be used in breeding programmes. They used UPGMA cluster analysis based on mean Euclidean distance and group determination by Tocher's method to determine eleven genotype groups and the three most divergent genotypes. |
| Vieira et al. (2014) | Compared leaf outline shapes of natural populations of the restinga ecotype in coastal Piauí with those of domesticated *A. occidentale* in the same region and the interior of the state. They used geometric morphometrics (landmark analysis) on samples from ten different localities and found the two taxa to be only weakly different under discriminant analysis. Sousa (unpublished) carried out a multivariate morphometrical analysis of 24 characters from leaves, inflorescence, flowers and fruits sampled from nine of the ten populations studied by Vieira et al. (2014). Using discriminant analysis with this data set, she found a stronger differentation of the restinga ecotype populations (as *A. microcarpum*) from those of cashew (*A. occidentale*). She provided botanical descriptions of the two taxa based on collections from coastal Piauí. |
| Borges (2015) | Studied genetic diversity of 18 cajuí (as *Anacardium spp.*) accessions from two experimental stations of the Active Germplasm Bank in Embrapa Meio-Norte, in the municipalities of Teresina and Parnaíba, Piauí state, using morphometrics and ISSR molecular markers. The plants were collected from six populations in the municipalities of Ilha Grande (Baixão, Cal, Labino) and Parnaíba (Embrapa experimental station, Fazenda Bom Jesus, Pedra do Sal). She states in chapter 2 that cajuí occurs in the Piauí littoral as part of the pioneer vegetation in restinga and carrasco, most concentrated in Labino, Pedra do Sal, Cal and Tatus. She notes that in Nigeria the cashew was introduced to control erosion and reforest degraded areas, prior to being recognized for its commercial potential. [This also true for India]. States that Barros (1991) was the first study of genetic diversity using multivariate techniques. Cites also Pessoni (2007), Rufino (2004), Maia et al. (2012) and Vieira et al. (2014). Chapter 3 is the morphometric study, morphoagronomic characterization using 14 morphological characters and four chemical characters. PCA explained 81.7% of total variance in 3 axes. Cluster analysis by UPGMA and Tocher optimization found substantial variation within and between populations. She used IBPGR descriptors, replicated ten times per plant: inflorescence length, maximum inflorescence width, branch number in inflorescence, hypocarp length, hypocarp diameter, hypocarp basal diameter, hypocarp apical diameter, hypocarp weight, drupe length, drupe width, drupe weight, maximum plant height, mean canopy diameter N-S and E-W (the latter two characters measured only once per individual). The chemical characters were: pH, total titratable acid (ATT), total soluble solids (SST), ratio SST/ATT.  PCA carried out on correlation matrix. UPGMA and Tocher used mean Euclidean distance between accessions. Found wide phenotypic diversity in hypocarp and drupe. Positive correlation between hypocarp and drupe characters; also between the inflorescence characters, but not between these character sets. Plant height and canopy width highly correlated. In general correlations between characters were high. Fruit size (drupe and hypocarp) had greatest loadings on PC1, chemical characters on PC2 and inflorescence characters (positive loadings) and plant stature (negative loadings) on PC3.  Mean Euclidean distance showed no correlation with locality. Within-locality variation in distance was just as high as between. Singh (1981)'s method used to find which characters contributed most to inter-accession distance. This method [see biotools R package] calculates a function for the importance of variables based on the squared generalized Mahalanobis distance. Tocher optimization found two groups, one of which consisted only of a single accession. The main group was subdivided into five subgroups by the same method, three consisting of a single individual. UPGMA showed four main groups (tree cut at the mean pair-wise distance value), of which two consisted of a single individual. Overall, high variability found in the measured characters indicating high inter-accession divergence.  The discriminant analyses used a group model of two groups, Parnaíba and Ilha Grande municipalities. Inflorescence length and width were strongly weighted and influenced the divergence between the accessions. Much overlap found between the two groups. Some of this variation could be due to some accessions having been introduced from other regions, given that the accessions were collected in rural properties. Human interference also likely to be explanation for variation in caju genotypes in a locality. Fruit bats are dispersers (Mitchell & Mori) and bees the pollinators (Freitas). Leads to wide dispersal between areas. Hybridization with caju may be another source of variation. High diversity shows high potential for commercial exploitation of cajuí and for genetic improvement of caju.  Chapter 4 is on ISSR analysis of genetic diversity. Evaluated genetic diversity and population structure in 25 accessions of cajuí grouped into two populations, Parnaíba and Ilha Grande municipalities. She used 9 primers with 104 markers. Genetic diversity was high. UPGMA mixed individuals from the two groups. AMOVA showed all variability was intra populational. Little differentiation between populations found. Interpreted as indicating high gene flow. STRUCTURE analysis corroborated these results, showing that the two a priori groups are not true biological groups. However, k=2 was the best grouping, suggesting other factors at work, e.g. human selection and possible hybridization [with caju]. The ISSR study included accessions of *humile, microcarpum, othonianum, giganteum and occidentale* from Pacajus.  The binary data was made into a similarity matrix using Dice-Sorensen coefficient using PAST. UPGMA used with cophenetic correlation coefficient and bootstrap. Genetic diversity parameters were computed with POPGENE. AFLP-SURV used to compute mean heterozygosity and Nei diversity using Taylor expansion with software TFPGA. AMOVA was computed with ARLEQUIN. Gene flow was computed with POPGENE. Population structure was analysed with STRUCTURE.  Parnaíba population had greater genetic diversity than Ilha Grande. Heterozygosity was similar in the two populations. UPGMA found three distinct groups one with most accessions and two with a single individual. *A. humile* came out as one of the latter two groups, but *othonianum* and *giganteum* were embedded with the cajuí. AMOVA found no difference whatever between the two populations. But STRUCTURE suggested two groups as optimal. These two molecular groups contained individuals from both populations. *A. microcarpum* (from Pacajus) did not group closely with the Piauí cajuí. "Entretanto a insuficiência de estudos de caracterização botânica para as plantas de cajuí do litoral piauiense, a proximidade genética entre cajuí proveniente da mesorregião sudoeste piauiense com as espécies *A. othonianum* e *A. occidentale* (PESSONI, 2007) e os resultados apresentados por este estudo indicam a possibilidade de outras espécies do gênero serem ocorrentes na região. Necessitando, por isso, de estudos taxonômicos, a fim de classificar a verdadeira espécie ou espécies de *Anacardium* ocorrente no litoral do Piauí."  The results show that the Parnaíba and Ilha Grande groups are not distinguishable and support the inference of intense gene flow between these areas. Dispersal by river and bats (13 species recorded from Piauí littoral, four of which are frugivorous, foraging for distances of up to 13 km). Cajuí has high germination (Correa et al. 2002). Pollen also well dispersed by bees. The two genetic groups detected by STRUCTURE suggest the existence of hybrids or migrants. The possible import of genotypes into the localities sampled by people cannot be ruled out. This would contribute to the high diversity and lack of structure.  [Morphometric data set presented in Appendix A] |
| Santos & Santos Junior (2015) | Studied 38 genotypes from a natural population of *A. humile*  in the cerrado of the municipality of Patrocinio, Minas Gerais, using multivariate analysis to estimate phenotypic diversity. Their estimated genetic diversity was interpreted as satisfactory for future genetic improvement studies. The species is known as cajuzinho-do-campo, cajuzinho-do-cerrado or cajuí. They used quantitative and qualitative characters of the phenotype to characterize the genotypes sampled. The study area was a cerrado of a transition type between cerrado campo sujo and cerrado campo rupestre. 132 plants were selected randomly, marked by GPS and each considered a genotype and the 38 which reached fruiting stage in the study period were finally used. The characters were evaluated every month in the vegetative period, every week during the reproductive period, and every day during peak flowering and fruiting. 33 characters were selected from the IPBGR descriptor list, 13 quantitative and 20 qualitative. Ten measurements were made for each character per individual, whenever possible. Standardized Mean Squared Euclidean Distance was used as distance measure for the quantitative variables, with PCA. Jaccard's distance was used for qualitative variables. Tocher's clustering by optimization was used with UPGMA to cluster the genotypes using both data sets. They studied proportion of hermaphrodite to masculine flowers per inflorescence. They found high within-population variability among the genotypes. |
| Sethi (2015) | Studied morphological and molecular diversity in hybrids of *A. occidentale* from India using multivariate analysis (PCA, cluster analysis) and DNA fingerprinting with RAPD markers (PCoA, UPGMA). The study found a significant positive association between nut yield and canopy spread, number of flowering laterals per sq.m. and nuts per panicle. Genetic divergence/variability was greatest in number of perfect flowers and total number of flowers. This author provides an extensive literature review of previous studies comparing genetic value of agronomic characters and molecular and multivariate studies, focusing especially on Indian research. The 24 characters used are explained in detail. |
| Jena et al. (2016) | Studied genetic diversity of twelve cashew cultivars in India using morphometrics and RAPD DNA markers. Material was obtained from the cashew orchard in the Orissa University of Agriculture and Technology. Fifteen individuals of each genotype were measured for morphometric attributes. Dice's dissimilarity coefficient was used to compute a UPGMA cluster analysis. DNA was collected from emerging leaves of the 12 cultivars. Fifteen primers were selected for the RAPD analysis. Jaccard's coefficient was used to compute a UPGMA analysis which was bootstrapped. Morphometric variation was computed for each cultivar for each variable: height, girth, canopy spread (E-W), canopy spread (N-S), vegetative flushing initiation (week), first flowering initiation (week), number of laterals per sq.m (E-W), number of flowering laterals per sq.m (E-W), number of laterals per sq.m (N-S), number of flowering laterals per sq.m (N-S), panicle length, apple colour, apple weight, nuts per panicle, nut weight, kernel weight, shelling percentage, yield (kg per plant). UPGMA analysis showed low level of correlation between molecular and morphometric data. |
| Paikra (2016) | Made a study of 14 Indian cultivated genotypes of *A. occidentale* which included their characterization using 68 morphological characters, finding significant among-genotype variation for most characters. Among characters used were plant height, canopy spread, leaf size, flowering intensity, inflorescence length, percentage fruit set, number of nuts per panicle, nut weight, shell thickness, kernel weight, apple weight and TSS of juice showed significant positive correlation with nut yield per tree. There is a thorough literature review of cashew floral biology. Detailed univariate and pairwise correlation analyses of characters. Uses Mahalanobis D2 and Tocher's clustering method as a multivariate approach. |
| Matos Filho (2017) | Studied common cashew (cajueiro comum) genotypes in commercial plantations of four localities in the interior of Piauí. The objective was to characterize the genotypes using morphological characters of agronomic importance and estimate genetic diversity with a view to genetic improvement. The study was based on genotypes from plantations made from seed, maintained using basic techniques and without consistent spacing between plants. The morphological characters used were of the drupe, kernel and hypocarp and physico-chemical features [?of the hypocarp]. Methods of analysis were univariate summary statistics, UPGMA, biplots, coefficient of repeatability and PCA, linear mixed models, restricted maximum likelihood. Genetic variation, accuracy of selection, heritability, residual variation and means were estimated for the variables. |

References in alphabetical order:

ALIYU, O.M. & AWOPETU, J.A. 2007: Multivariate analysis of cashew (*Anacardium occidentale* L.) germplasm in Nigeria. − Silvae Genetica56(3-4): 170−179.

ARCHAK, S.; GAIKWAD, A.B.; GAUTAM, D.; RAO, E.V.V.B.; SWAMY, K.R.M. & KARIHALOO, J.L. 2003: Comparative assessment of DNA fingerprinting techniques (RAPD, ISSR and AFLP) for genetic analysis of cashew (*Anacardium occidentale* L.) accessions of India. − Genome 46: 362−369.

ARCHAK, S.; GAIKWAD, A.B.; SWAMY, K.R.M. & KARIHALOO, J.L. 2009: Genetic analysis and historical perspective of cashew (*Anacardium occidentale* L.) introduction into India. − Genome 52: 222−230.

BARROS, L.M. 1991: Caracterização morfológica e isoenzimática do cajueiro (Anacardium occidentale L.) tipos comum e anao-precoce, por meio de tecnicas multivariadas. − PhD thesis, Universidade de São Paulo, Escola Superior de Agricultura Luiz de Queiroz (ESALQ), Piracicaba, Brazil.

BORGES, A.N.C. 2015: Caracterização genética em germoplasma de cajuí (Anacardium spp.) por meio de marcadores morfoagronômicos e moleculares ISSR. − MSc thesis, Programa de Pós-graduação em Genética e Melhoramento, Universidade Federal do Piauí, Teresina, Brazil.

CARVALHO, R.S. 2011: Estudos de acessos de Anacardium humile A.St.Hil. por meio de caracterização morfológica e de marcadores RAPD. − MSc thesis, Universidade Federal de Goiás, Jataí, Brazil.

CARVALHO, R.S.; PINTO, J.F.N.; REIS, E.F.; SANTOS, S.C. & DIAS, L.A.S. 2012: Variabilidade genética de cajuzinho-do-cerrado (*Anacardium humile* St.Hil.) por meio de marcadores RAPD. − Rev. Bras. Frutic. Jaboticabal, São Paulo 34(1): 227−233.

CHIPOJOLA, F.M.; MWASE, W.F.; KWAPATA, M.B.; BOKOSI, J.M.; NJOLOMA, J.P. & MALIRO, M.F. 2009: Morphological characterization of cashew (*Anacardium occidentale* L.) in four populations in Malawi. − African Journal of Biotechnology 8(20): 5173−5181.

DASMOHAPATRA, R. (2012). Morphological and molecular characterization of different types of cashew (*Anacardium occidentale* L.). PhD thesis, Orissa University of Agriculture and Technology, Bhubaneswar, Orissa, India.

DASMOHAPATRA, R.; RATH, S.; PRADHAN, B.D. & ROUT, G.R. 2014: Molecular and agromorphological assessment of cashew (*Anacardium occidentale* L.) genotypes in India. − Journal of Applied Horticulture 16(3): 215-221.

DESAI, A.R. 2008: Molecular diversity and phenotyping of selected cashew genotypes of Goa, and physiological response of cv. GOA-1 to *in situ* moisture conservation. − PhD thesis, University of Agricultural Sciences, Dharwad, India.

FROTA JÚNIOR, J.I.; LIMA, E.N.; ARAÚJO, M.E.B.; LIMA, I.B. & LIMA, J.S. 2014: Divergência genética entre progênies de meio-irmãos de cajueiro anão precoce utilizando caracteres morfológicos. − Agrarian Academy, Centro Científico Conhecer, Goiânia 1(2): 47-59.

GOMES, J.G.; OLIVEIRA, N.C.C.; LOPES, P.S.N. & BALLESTEIRO, M. 2011: 12630 – Componentes principais no estudo da diversidade de cajuí na sub-bacia do Rio dos Cochos. − *Cadernos de Agroecologia* 6(2): 1-5. − Available at: <http://aba-agroecologia.org.br/revistas/index.php/cad/article/view/12630>. Accessed on 6 Dec. 2017.

JENA, R.A.; SAMAL, K.S.; PAL, A.; DAS, B.K. & CHAND, P.K. 2016: Genetic diversity among some promising Indian local selections and hybrids of cashew nut based on morphometric and molecular markers. − International Journal of Fruit Science 16(1): 69−93.

LUCENA, V.M.X. 2006: Diversidade genética entre genótipos de cajueiro (Anacardium occidentale L.) e qualidade do fruto e pseudofruto. MSc thesis, Universidade Federal de Roraima, Boa Vista. Brazil.

MAIA, M.C.C.; OLIVEIRA, L.C.; ALMEIDA, A.S.; ARAÚJO, L.B.; VASCONCELOS, L.F.L. & LACERDA, M.N. 2012: Divergência genética e correlações entre variáveis agroindustriais da população de trabalho de cajuí da Embrapa Meio-Norte. − Anais II Congresso Brasileiro de Recursos Genéticos, 24 a 28 de setembro de 2012, Belém, Pará, Trabalho 543.

MATOS FILHO, C.H.A. 2017: Diversidade genética e seleção em populações de cajueiro comum. PhD thesis, Programa de Pós-Graduação em Agronomia-Agricultura Tropical, Universidade Federal do Piauí, Teresina, Brazil.

PAIKRA, M.S. 2016: Studies on genetic divergence in cashew (*Anacardium occidentale* L.). − PhD thesis, Department of Horticulture, College of Agriculture, Indira Gandhi Krishi Vishwavidyalaya, Raipur, Chhattisgarh, India.

PAIVA, J.R.; CRISÓSTOMO, J.R. & BARROS, L.M. 2003: Recursos genéticos do cajueiro: coleta, conservação, caracterização e utilização. Documentos 65, 43p. − Embrapa Agroindústria Tropical,Fortaleza.

PESSONI, L.A. 2007: Estratégias de análise de diversidade em germoplasma de cajueiro *(Anacardium spp.)*. − PhD thesis, Universidade Federal de Viçosa, Minas Gerais, Brazil.

RUFINO, M.S.M.; ALVES, R.E. & ARAGÃO, F.A.S. 2008: Análise multivariada de genótipos de cajuizeiro em áreas nativas da região Meio-Norte do Brasil. − Proc. Interamer. Soc. Trop. Hort*.* 52: 140−143.

SAMAL, S.; ROUT, G.R. & LENKA, P.C. 2003: Analysis of genetic relationships between populations of cashew (*Anacardium occidentale* L.) by using morphological characterisation and RAPD markers. − Plant Soil Environ*.* 49(4): 176−182.

SANTOS, R.C. & SANTOS JUNIOR, J.E. 2015: Divergência genética por análise multivariada de caracteres fenotípicos de *Anacardium humile* (St. Hilaire). − Rev. Ceres, Viçosa 62(6): 553−560.

SETHI, K. 2015: Studies on morphological and molecular diversity in cashew (Anacardium occidentale L.) hybrids. − PhD thesis, Orissa University of Agriculture and Technology, Bhubaneswar, India.

VIEIRA, M.; MAYO, S.J. & ANDRADE, I.M. 2014: Geometric morphometrics of leaves of *Anacardium microcarpum* Ducke and *A. occidentale* L. (Anacardiaceae) from the coastal region of Piauí, Brazil. − Braz. J. Bot*.* 37(3): 315−327.