

SUDRÉ CP; LEONARDECZ E; RODRIGUES R; AMARAL JÚNIOR AT; MOURA MCL; GONÇALVES LSA. 2007. Genetic resources of vegetable crops: a survey in the Brazilian germplasm collections pictured through papers published in the journals of the Brazilian Society for Horticultural Science. *Horticultura Brasileira* 25: 496-503.

Genetic resources of vegetable crops: a survey in the Brazilian germplasm collections pictured through papers published in the journals of the Brazilian Society for Horticultural Science

Cláudia P Sudré¹; Eduardo Leonardecz²; Rosana Rodrigues¹; Antônio T do Amaral Júnior¹; Maria da CL Moura³; Leandro SA Gonçalves¹

¹UENF, LMGV, Av. Alberto Lamego, 2000. Pq. Califórnia, 28013-602 Campos dos Goytacazes-RJ; ²UCB, SGAN 916, Av. W5 Norte, 70929-970 Brasília-DF; ³AGERP/SEAGRO/MA, Av. Guaxenduba 199, Centro, 65015-560 São Luís-MA; cpombo@uenf.br; rosana@uenf.br

ABSTRACT

The research on plant genetic resources is essential for the conservation of genetic diversity and accessions' divergence studies, as a basis for plant breeding. Aiming to know the state of art in this subject, a historical survey was carried out in *Revista de Olericultura* and in *Horticultura Brasileira*, from 1961 to 2006, searching for papers dealing with vegetable crops genetic resources. In each of the papers, the species studied, first author institution, publication year, applied software, number of accessions and descriptors, and the multivariate techniques used were registered. Based on these characteristics, papers were grouped using multivariate analysis. Sixty-one papers dealt somehow with genetic resources in the time covered by the survey, from which 91.8% were published after 1990 (60.7% from 2001 to 2005). The use of multivariate analysis was reported in 57.3% of the papers, with an average of 2.3 and maximum of 6 multivariate procedures per paper. The Tocher Method, reported in 34% of the papers, was the most frequently used multivariate analysis. Twenty-five species were studied. *Capsicum* was the most frequently studied genus, either considering number of papers (seven) or accessions (664). Research institutions located in the Southeast region concentrated the highest number of papers. UFV (*Federal University of Viçosa*), UENF (North Fluminense State University *Darcy Ribeiro*), Embrapa Vegetables, and UNESP (São Paulo State University) – *Campus* of Jaboticabal responded for 45% of the papers. There was an improvement in the adequacy of the statistical techniques used along time, due mainly to the development of free-access software. The software GENES was the most frequently reported in the papers surveyed. Nevertheless, almost 50% of the authors did not mention the software used for data analysis. Quantitative morphoagronomic and evaluation descriptors were the most often used. The multivariate analysis allowed grouping the papers in nine clusters.

Keywords: review, multivariate analysis, diagnosis, genetic diversity, vegetables.

RESUMO

Recursos genéticos de hortaliças: as atividades nas coleções brasileiras de germoplasma retratadas nas publicações da Associação Brasileira de Horticultura

A pesquisa com recursos genéticos vegetais é essencial tanto para a conservação da diversidade genética, quanto para o estudo da divergência entre acessos, base para programas de melhoramento. Com objetivo de conhecer o estado-da-arte nesse tema, foram identificados os trabalhos relacionados a recursos genéticos de hortaliças publicados na “Revista de Olericultura” e “Horticultura Brasileira”, de 1961 a 2006. Foram anotadas as espécie(s) estudada(s), a instituição de origem do primeiro autor, o ano de publicação, os softwares utilizados, o número de acessos, o número de descritores e o número de técnicas multivariadas aplicadas em cada artigo. Com base nessas características, os artigos foram agrupados utilizando análise multivariada. No período pesquisado, 61 artigos trataram, sob algum aspecto, de recursos genéticos. Desses, 91,8% foram publicados a partir de 1990, com concentração de 2001 a 2005 (60,7%), com 57,3% deles utilizando pelo menos uma técnica multivariada, com média de 2,3 e máximo de seis técnicas por artigo. O método de agrupamento de Tocher foi o mais utilizado, relatado em 34% dos artigos. Os artigos cobriram 25 espécies. Dentre elas, *Capsicum* foi o gênero mais pesquisado, tanto em relação ao número de artigos (sete), quanto em número de acessos estudados (664). As instituições de pesquisa da região Sudeste concentraram o maior número de artigos. Se somadas, UFV, UENF, Embrapa Hortaliças e UNESP – *Campus* de Jaboticabal foram responsáveis por 45% dos trabalhos publicados no tema. Houve um aprimoramento das técnicas estatísticas utilizadas na análise dos dados, sobretudo devido ao desenvolvimento e uso de softwares de fácil compreensão. O programa GENES foi o mais referenciado. Entretanto, quase a metade dos autores não citou o programa estatístico utilizado para a análise dos dados. Os descritores de caracterização morfoagronômica quantitativos e de avaliação foram os mais estudados. A análise multivariada permitiu classificar os artigos em nove grupos.

Palavras-chave: revisão, análise multivariada, diagnóstico, diversidade genética, olerícolas.

(Recebido para publicação em 27 de junho de 2007; aceito em 31 de outubro de 2007)

Genetic resources are related to the variability of plants, animals, and microorganisms of present and potential socioeconomic interest for breeding programs, biotechnology, and allied fields of research, with special emphasis in the use and conservation of the biodiversity (Nass, 2001). In 1996, FAO (*Food and Agriculture Organization*)

passed the Action Global Plan for Genetic Resources related to Food and Agriculture. In this plan, priority is given to the survey and inventory of genetic resources related to food and agriculture and to the expansion of characterization, evaluation, and number of core collections, aiming at broadening the use of genetic resources (Cooper, 1998).

Brazil hosts a vast biodiversity. To manage and preserve this asset, the country relies on more than 200 *ex situ* germplasm banks that preserve over 250,000 accessions. However, most of the germplasm on these banks is exotic (Valois, 2005). Among the main custodian institutions, we may cite the Agronomic Institute of Campinas (IAC), which manages collections and active germplasm banks since the decade of 1930. IAC was one of the first Brazilian institutions to carry out germplasm conservation. Worth mentioning are also the Federal University of Viçosa (UFV) and the Brazilian Corporation for Agricultural Research (Embrapa). UFV created in 1966 the Germplasm Bank of Vegetable Crops – BGH (Silva *et al.*, 2001), where currently 6,560 accessions, of 106 species, are maintained. Embrapa coordinates a Curator System with 137 germplasm banks and holds the responsibility for the national base collections. These germplasm banks, specially the active banks, are starting points for research on accession characterization, evaluation, and conservation. As technology and statistical methods advance, there is a steady improvement of the amount and quality of the information drawn from these banks.

The increasingly frequent use of multivariate analysis is one of the tools that are boosting the studies carried out with accessions from germplasm banks. Multivariate analyses are based on algorithms that consider simultaneously all or nearly all characteristics assessed in germplasm characterization and evaluation experiments. This way, it is possible to integrate the multiple information drawn from experimental evaluations (Amaral Júnior, 1999), generating new information, such as clusters, canonic correlations, and distance projections in

the plan (Cruz & Carneiro, 2003). Pearson (1901) started this statistical approach when he described the principal component procedure.

Taking it all into consideration, this work aimed at surveying the papers published on *Revista de Olericultura* and *Horticultura Brasileira*, respectively previous and present scientific official journals of the Brazilian Association of Horticultural Science that dealt with genetic resources. In addition, by characterizing these papers, we intend to draw a portrait of the research on genetic resource on vegetable crops. Finally, the papers were grouped in similarity clusters.

The working material and the methods used

We reviewed 684 papers from *Revista de Olericultura*, which were published between 1961 and 1980, and 24 volumes of *Horticultura Brasileira*, which appeared from 1983 to 2006. The criteria used to select papers were to identify those that dealt with any aspect related to genetic resources, as such germplasm characterization, evaluation, and conservation. Selected papers were described for:

ESP: common name of the species studied in the paper, except fungi, summing up 26 categories: 1= pumpkins and squashes, 2= basil, 3= garlic, 4= potato, 5= sweet-potato, 6= eggplant, 7= onion, 8= *Colletotrichum lagenarium*, 9= kale, 10= cocona (*Solanum sessiliflorum* Dunal), 11= dotted smartweed (*Polygonum punctatum* Ell.), 12= peas, 13= bush bean/butter bean, 14= scarlet eggplant (*Solanum gilo*), 15= cassava, 16= watermelon, 17= melon, 18= sweet corn, 19= strawberry, 20= hot and sweet pepper, 21= bamboo piper (*Piper aduncum* L.), 22= black pepper, 23= okra, 24= radish, 25= taro, 26= tomato;

INST: first author institution, with 30 categories: 1= AGÊNCIARURAL, EEAnápolis (Rural Agency – Anápolis Experimental Station), 2= Amazonic National Institute of Research of, 3= Embrapa Tropical Agroindustry, 4= Embrapa Oriental Amazon, 5= Embrapa Temperate Agriculture, 6= Embrapa Vegetables, 7= Embrapa Rondônia, 8= Embrapa Tropical Semi-Arid, 9=

EPAMIG (Minas Gerais State Corporation for Agricultural Research), 10= ESALQ (Agriculture College Luiz de Queiróz), 11= FCAP, 12= IAC (Agronomic Institute of Campinas), 13= IB (Biological Institute), 14= PESAGRO-RIO (Rio de Janeiro State Company for Agricultural Research), 15= UENF (North Fluminense State University *Darcy Ribeiro*), 16= UESB (South Bahia State University), 17= UFES (Federal University of Espírito Santo), 18= UFG (Federal University of Goiás), 19= UFGD (Federal University of the Metropolitan Region of Dourados), 20= UFLA (Federal University of Lavras), 21= UFPB (Federal University of Paraíba), 22= UFPE (Federal University of Pernambuco), 23= UFPEL (Federal University of Pelotas), 24= UFRPE (Rural Federal University of Pernambuco); 25= UFRRJ (Rural Federal University of Rio de Janeiro); 26= UFS (Federal University of Sergipe); 27= UFV (Federal University of Viçosa); 28= UNB (University of Brasília); 29= UNESP-Jaboticabal (São Paulo State University at Jaboticabal), 30= UNIMONTES (Minas Gerais State University of Montes Claros);

YEAR: paper publication year, within classes for five-year periods (1= 1961-65, 2= 1966-70, 3= 1971-75, 4= 1976-80, 5= 1981-85, 6= 1986-90, 7= 1991-95, 8= 1996-2000, 9= 2001-05, and 10= 2006);

PROG: software used in the identified papers to analyze the data (FITOPAC, GENES, NTSYS, SAEG, SANEST, SAS);

ACES: number of accessions characterized and/or evaluated in each paper;

DESC: total number of descriptors used in each paper;

DMQL: number of qualitative morphoagronomic descriptors used in each paper;

DMQT: number of quantitative morphoagronomic descriptors used in each paper;

DCBI: number of biochemical descriptors used in each paper;

DCMO: number of molecular descriptors used in each paper;

AVAL: number of evaluation descriptors used in each paper;

MULT: number of multivariate techniques descriptors used in each paper. Each uni- and multivariate technique was considered a binary qualitative descriptor, summing up 21 binary descriptors: 1= conglomeration analysis, 2= multidimensional scale analysis, 3= Anderson discriminant analysis, 4= centroid, 5= principal components, 6= canonic correlation, 7= variable discard, 8= LSD-Student, 9= Duncan, 10= relative importance of characters – Singh, 11= average linkage, 12= non-specified method, 13= Tocher optimization, 14= distance projection in the plan, 15= SAHN clustering, 16= Scott-Knott, 17= Tukey, 18= UPGMA, 19= canonic variables, 20= nearest-neighbor, 21= Ward;

To classify the papers, each one was considered as a treatment, summing up 61 papers. Descriptive statistics and multivariate analysis were used to interpret the data. Considering that the data reported on these papers were from several natures (binary, multicategorical, quantitative, and discrete), we chosen to perform a multivariate analysis that consider all data simultaneously. Although the multivariate analysis is more often employed on binary and quantitative data, the development of new routines on the software made it possible to associate other sorts of data and to use techniques in a single analysis that allow a more efficient identification of the differences and similarities among treatments.

The estimation of similarities among papers were obtained for each pair of paper (i, j). Similarities were then transformed in distances and the similarity coefficient was calculated by SAS (SAS, 1998), through the routine established by Victória *et al.* (2001), using the similarity index of Jaccard (Jaccard, 1901). Cluster knots were grouped using the UPGMA method. The consensus dendrogram was obtained after 1,000,000 bootstrap recalculations.

Published papers and reported methods of analysis

There were no papers dealing with genetic resources in the surveyed

journals before 1976. From the 684 papers from *Revista de Olericultura* reviewed, the first one reporting results in genetic resources appeared in 1976 and dealt with the morphoagronomic characterization of 100 okra accessions of the germplasm bank of UFV (Federal University of Viçosa), without using multivariate analysis (Pedrosa & Mizubuti, 1976). From 1983 ahead, year in which the first volume of *Horticultura Brasileira* came out, 60 papers were published concerning genetic resources, 35 using multivariate analysis and, the remaining, multiple comparisons and/or descriptive statistics. Taking into consideration the survey time lag (37 years), the number of papers is low, averaging less than two per year. On the other hand, in view of the number of surveyed issues of the scientific journals (78), the number of selected papers seems to be reasonable, with nearly one paper concerning genetic resources published per issue. The largest part of the papers, precisely 91.8%, was published from 1990 ahead, with a concentration in the period ranging from 2001 to 2005. In this period, 60.7% of the papers under evaluation were published and 54% employed multivariate analysis. The high frequency of papers in this period is due to the high appreciation genetic resources started receiving at that time and also to the development of more friendly software with open-access, focused on the researcher needs.

The diagnosis of the analyzed papers based on the statistical procedures used, revealed that 57.3% of the papers employed at least one multivariate technique, with an average of 2.3 and a maximum of six techniques per paper. The Tocher Method, for clustering, was the most frequently used, reported in 34% of the papers. This method makes mutually exclusive groups (Cruz & Carneiro, 2003) through accurate clustering, and uses less physical space than the dendrograms. The relative importance of characters (Singh, 1981) was present in 16.4% of papers and allowed the identification of the descriptors that contributed the most for the expression of genetic variability. Canonic variables appeared in 14.7% of

the papers and did not differ from the use of the nearest-neighbor method, used on 13.1% of the papers. Most of the authors did not mention performing the co-linearity test, information that would increase the impact of the results. Regarding means test, Tukey was the most regularly used (23%), mainly in papers that employed evaluation descriptors and a small quantum of genotypes. When a high number of genotypes were studied, Scott-Knott test was preferred (13.1%) because it forms mutually exclusive groups, which simplifies result interpretation.

Currently, the use of non-similarity matrixes based on multicategorical or on binary data is a frequent tool in papers concerning genetic resources (Cruz, 2006; Sudré *et al.*, 2006). Models that simultaneously use variables with distinct distributions, such as the Ward-MLM model (Bussab *et al.*, 1990; Franco *et al.*, 1998; Crossa & Franco, 2004), are also becoming more common. These analyses make a more thorough use of the variables in the assessment of genetic divergence and contribute to the identification of duplicated accessions, recommendation of highly heterotic crosses, and other purposes. Motta *et al.* (2006) used the Bussab *et al.* (1990) model to convert quantitative into binary data that could be analyzed simultaneously with molecular data. In their work, quantitative data came from physical-chemical evaluation, yield assessment, and morphological and genetic description of 12 garlic cultivars.

The characterization of the available germplasm, either native or not, aiming at studying its potential, as food, medicine or other, is becoming increasingly important. As the multidisciplinary approach is presently predominant, there is a need to carry out several experiments, in the field and at the laboratory, to identify and to quantify the socioeconomic potentialities of the germplasm. On these experiments, a number of characteristics are assessed, each one with its peculiar distribution, either continuous or discrete. However, it became obvious from the papers analyzed in the present survey that the techniques used not always allowed the

exploration of all available information. It should be considered that this way of analyzing data lies in the fact that univariate methods in general are enough to answer researchers' questions. Nevertheless, there are cases in which the multivariate analysis is not preferred due to (a) the absence of a specific routine in the software, (b) the absence of mathematical/statistical or statistical/genetic models that match the work developed, or (c) the excess of parameters to be estimated, which generates a large number of interactions and demands a huge computational effort. In this case, the analysis can be so time-consuming that is not possible to perform it.

The software used

Six statistical packages (FITOPAC, GENES, NTSYS, SAS, SAEG, SANEST) were reported in the identified papers. However, in 44% of the papers the software used was not mentioned. In general, these papers reported results of univariate analyses and it is possible that in some of them no software was used at all. Among the 35 papers that employed multivariate analysis, the software GENES was present in 60% of them, NTSYS in 17%, SAS in 5.7%, and FITOPAC in 2.9%. The remaining 14.3% of the papers did not inform the software used. It is likely that GENES was the most used software due to its adequacy to biometric genetics. GENES has a specific session to multivariate analysis with a range of procedures that fulfils most of researchers needs. In addition to that, the software popularity comes also from the open-access windows version (Cruz, 2006), the availability of its author to adjust its routines to the demand, and, since its author is a university professor, the continuous introduction of new users to the software. The first time GENES appeared in a *Horticultura Brasileira* paper was in 1994 (Amaral Jr. *et al.*, 1994).

NTSYS (*Numerical Taxonomy System*) appeared in the 60's, as a private software. Its multiplatform (Windows, Linux, MC-OS e Solaris) English version costs US\$ 300.00 (NTSYSpc, 2007). SAS, on its turn, is one of the most powerful packages when it comes to number and quality of available

procedures. Released in 1976, with a focus on Agronomy, currently SAS is used in a large number of areas. However, there are fees to use the services. In addition, to run properly the program, users need to be familiar to command lines (SAS, 2007). This is, at the same time, one of the strongest and weakest points of the software, and a challenge to the user. The program allows the user to precisely adjust the model of analysis to the demands. Nevertheless, this freedom represents a great barrier to those not familiar with programming and command lines. George Shepherd, from the Campinas University (UNICAMP), wrote FITOPAC. The first version appeared in 1988, emphasizing Phyto-sociologic and taxonomic analysis (Shepherd, 2001). Currently there is also the R software, with internet open-access. It is similar or even better than SAS, depending on the procedure selected. To date there were no papers concerning genetic resources reporting its use in *Horticultura Brasileira*, although in few years time R is likely to become common in science. SAEG and SANEST were used only in papers that reported results on multiple comparison, even though both packages perform multivariate analysis.

The species contemplated

In total, 25 species were studied in the identified papers, among them fruit, leaf, and tuber vegetables (Filgueira, 2000), as well as seasoning and medicinal herbs. *Capsicum* gathered the largest number of papers (seven), in which 664 accessions were evaluated. The location of a *Capsicum* diversity center in Brazil certainly contributed to the high frequency of papers dealing with the genus. In addition, there are wild and semi-domesticated species that have not been intensively studied (Bianchetti, 1996; Sudré *et al.*, 2005). Okra was investigated in six papers (329 accessions), while tomato appeared in another five (122 accessions). Bush beans (78 accessions), sweet-potato (452 accessions), and garlic (186 accessions) were present in four papers each; taro (108 accessions), melon (34 accessions), strawberry (41 accessions), watermelon (82 accessions), and potato

(42 accessions), were studied in three papers, and pumpkin (48 accessions), in two papers. The remaining vegetable crops and fungi had their genetic resources studied in one paper each. They are black pepper (18 accessions), dotted smartweed (eight accessions), bamboo piper (eight accessions), *Colletotrichum lagenarium* (19 accessions), onion (eight accessions), cassava (six accessions), radish (12 accessions), basil (55 accessions), cocona (*Solanum sessiliflorum* Dunal, 29 accessions), kale (seven accessions), peas (14 accessions), eggplant (92 accessions), scarlet eggplant (44 accessions), and sweet-corn (11 accessions). The final figure for number of accessions evaluated in each species might not be what is presented here, since there are cases in which the same experiment gave information to more than one paper. Nevertheless, the figures clearly indicate that a reasonable diversity of species were studied.

Several papers dealt with genetic resources of vegetable crops of large economic importance in the country. These papers came from different Brazilian institutes and were regularly published from 1976 to 2006. It was not observed any direct relation between economic importance of a vegetable crop and the frequency its genetic resources were studied. In 1996, Embrapa Vegetables sorted the national priority for research in vegetables crops (Embrapa, 1996). In this list, okra ranked 17th. Despite this, okra was the second most studied vegetable crop when it comes to genetic resources in the time lag of the present survey. On the other hand, important vegetable crops, such as carrot, lettuce, and cucumber, ranked 4th, 9th, and 11th in importance respectively (Embrapa, 1996), simply did not appeared in papers concerning genetic resources. This apparent inconsistency might be related to a major interest of the institutions in investigating vegetable crops with regional, instead of national, relevance. For instance, institutions in the Southeast region, in the states of Rio de Janeiro and Minas Gerais, were responsible for the papers dealing with okra because this a region where okra

is quite important. It is important to point out also that vegetables such as carrot and lettuce are intensively investigated and therefore research on genetic resources is no longer priority. For lettuce, for example, there is not even an official descriptors list.

The number of characterized and/or evaluated accessions was highly variable among crops, going from four to 366 accessions per paper, depending on the interest and the availability of human and financial resources, and facilities. In general, in characterization studies a larger number of accessions were used than in evaluation studies. The later demand experimental design and often replication in more than one environment, as it is the case for the evaluation of resistance to disease and pests, for instance. In addition, in most of the times, accessions under evaluation have already been characterized. Therefore, researchers usually make a pre-selection of the accessions to be studied according to their interests.

The descriptors used

The set of descriptors used in the papers varied extensively for the same reasons discussed for the *quantum* of accessions. In the period 1961-2006, the number of descriptors by paper ranged from four to 120. Qualitative morphoagronomic descriptors were present in 30% of the papers, ranging from one to 30 descriptors per paper, with an average of nine. Quantitative morphoagronomic descriptors were used in 64% of the papers, with an average of 10. Biochemical and molecular descriptors were reported in only 12% of the papers, with an average of 22 and 81 descriptors per paper, respectively. Evaluation descriptors were used in 62% of the papers, with an average of five. Therefore, the most frequently used descriptors were those regarding quantitative morphoagronomic characterization and accession evaluation. These descriptors are apparently cheaper to use in comparison to biochemical and molecular ones. However, they are more labor-intensive and demand larger and longer experiments. Qualitative descriptors are very often assessed in the experiments. However, as they are not

included as variables in the analysis, usually they are analyzed only by descriptive statistic procedures.

When describing or characterizing genetic resources, the ideal situation is to use the descriptor list of Biodiversity International, the institution that succeeded IPGRI (International Plant Genetic Resources Institute) and INIBAP (International Network for the Improvement of Banana and Plantain), in December 2006. The lists are of worldwide use and aim at setting standards to germplasm characterization and evaluation. In general, these lists gathered such a high number of descriptors that is difficult to use them in full. However, researchers should try to use at least the basic or core descriptors recommended by Bioversity International.

The enrolled institutions

The most prolific institutions on genetic resources papers were UFV (Federal University of Viçosa) and UENF (North Fluminense State University Darcy Ribeiro). Each one had nine papers published. The next institutions in number of papers, each one with five, were Embrapa Vegetables (Brazilian Agricultural Research Corporation – National Center for Vegetable Crops Research) and UNESP (São Paulo State University) – *Campus* of Jaboticabal. These four institutions produced together 28 papers (45.9% of the total). The remaining institutions published a maximum of two papers each. UFV owns one the largest germplasm bank of vegetable crops in the country, where more than 6,500 accessions are maintained (Silva *et al.*, 2001). UENF owns a smaller germplasm bank (1,600 accessions), although very representative. Embrapa Vegetables, on its turn, holds one of the largest germplasm collections of vegetable crops in the country and has a strong tradition on research on genetics and breeding, with several cultivars released to the market. UNESP at Jaboticabal has, since 1985, a graduation program on Genetics and Plant Breeding that makes constant use of genetic resources on its academic research. The simultaneous existence of graduation programs and research groups related to

genetic resources is a plausible explanation for finding three universities amongst the four institutions that scored higher in number of published papers.

There was no concentration of papers for a given species in none of the institutions. UFV published three papers on taro and one for each of the following crops: okra, collards, squash, *Capsicum*, dotted smartweed, and tomato. UENF published three papers dealing with okra, two with *Capsicum* and another two with tomato, one with sweet potato, and another one with bush beans. UNESP published two papers on potato, two on melon, and one on tomato. Embrapa Vegetables published one paper on each of the following vegetables: bush beans, *Capsicum*, peas, garlic, and sweet potato. Thus, considering the four institutions, there was a considerable diversity in relation to the vegetable crops that were used in genetic resources papers.

The Southeast geoeconomic region concentrate 57.4% of the papers. Not for coincidence, this region harbors the largest *quantum* of research groups in Brazil. The Mid-West region hold the second place, with 16.4% of the papers published. Embrapa Vegetables was the home institution of most of these papers. The Northeast published 14.7% of the papers, whereas the North and South regions contributed with 6.5 and 4.9% of the papers respectively.

The clustering of the papers

The 61 papers selected for dealing with genetic resources were grouped in nine clusters using the UPGMA method (Figure 1). This analysis, performed using simultaneously binary, multicategorical, and quantitative data, allowed the identification of similar and contrasting aspects among papers. The first cluster, with nine papers, gathered papers from the complete survey time lag. These papers were similar for not mentioning the software used for data analysis and for employing either LSD (Student) for mean comparison, or canonic correlation. Four papers, characterized by using only biochemical and evaluation descriptors and by comparing means using the Duncan test formed the second cluster. Cluster III had three papers that used SAS and

reported results on quantitative morphoagronomic and evaluation descriptors. In this cluster, multivariate analyses appeared in all papers. Nevertheless, the nearest-neighbor method was the only statistical procedure common to the three papers. The average linkage and the centroid method, as well as the conglomeration analysis appeared only in this group.

Ten papers were clustered in the fourth group. These papers were published from 1996 to 2005, with 80% reporting the use of the software GENES. These papers used basically quantitative morphoagronomic and evaluation descriptors and all performed multivariate analyses, with an average of three procedures per paper. All papers used the relative importance of characters and 90% the Tocher method. It is worth mentioning that the only papers to use the method of variable discard were grouped in this cluster. Cluster V assembled 13 papers, 92.3% published from 1996 to 2006, 69% developed at the Federal University of Viçosa (UFV) and the North Fluminense State University Darcy Ribeiro (UENF), and 92.3% employing GENES. Only one paper in this cluster did not mention the software used. All papers reported the use of multivariate analyses, with an average of 2.4 procedures per paper. The Tocher method was reported in 92.3% of the papers. Clusters IV and V confirmed the relevance of the software GENES for the increase in the frequency of use and amount of multivariate procedures employed in papers concerning genetic resources.

Clusters VI and VII were formed by only one paper each, published respectively in 2002 and 2004. The paper on cluster VI dealt with qualitative morphoagronomic descriptors in sweet potato and did not use any uni- or multivariate analysis, but only descriptive statistics. The paper on cluster VII reported the use of biochemical descriptors and presented a NTSYS dendrogram. Nevertheless, it did not mention the statistical procedure used. Cluster VIII was formed by five papers, published between 2001 and 2005. These papers used

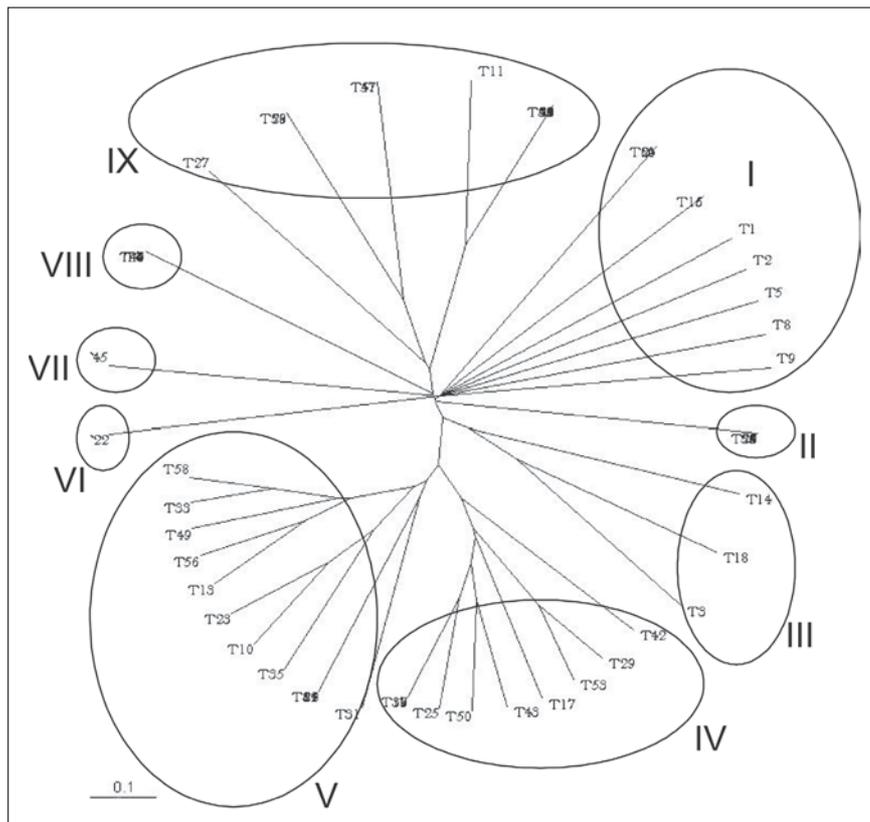


Figure 1. Circular dendrogram of distances based on data from 61 papers published in *Revista de Olericultura* and *Horticultura Brasileira*, using eight quantitative, four multicategorical qualitative, and 21 binary qualitative descriptors. Y axis corresponds to papers in genetic resources field (dendrograma circular de dissimilaridades com dados de 61 artigos da Revista de Olericultura e Horticultura Brasileira, utilizando oito descritores quantitativos, quatro qualitativos multicategóricos e 21 qualitativos binários. O eixo y corresponde aos artigos na área de recursos genéticos). UENF, Campos dos Goytacazes, 2007.

quantitative morphoagronomic and evaluation descriptors, except by one paper that employed qualitative morphoagronomic descriptors. In addition, none of the papers in this cluster used any multivariate procedures. Instead, all papers reported the use of Scott-Knott to perform the clustering.

Cluster IX gathered 15 papers, all published between 2001 and 2006, except by a single 1994 paper. Most of the papers in this cluster (53.3%) did not mention the software used for analysis. Those papers that give the information, reported the use of FITOPAC, NTSYS, and SANEST. The Tukey test was used in 73.3% of the papers, while the Duncan test was reported only once. All papers in the survey that used the SAHN Clustering were included in this cluster, as well as all papers that used molecular tools for

investigating genetic resources.

Based on the cluster analysis, a more robust statistical tool than the descriptive analysis, it was noticed in the papers studied a broad approach in the choice of both what descriptors and what statistical procedures to use. Even though, some trends were revealed. This paper diversity led to a lack of standard in collecting and analyzing the data, as well as in reporting the results. As consequence, readers do not have a comprehensive information on accession evaluation and on the procedures and software used. For instance, although several packages are available that perform simultaneous analysis of data with distinct nature, in 2006 there was still a paper published without mentioning the software and statistical procedure used. The information on the software used is an important stimulus to other authors to

search for the same or similar packages. There were also cases of authors limiting their analysis to mean comparison tests, even when working with descriptors that would have bear much more robust procedures, such as multivariate analysis.

Final Remarks

The present survey revealed an increase in the number of published papers regarding genetic resources during the period. In addition, it was evident that a steady rise in the use of multivariate techniques took place in more recent years, when most of the papers reported the use of more than one multivariate procedure. The availability of statistical software in Portuguese was certainly one of the reasons for that, with particular emphasis to the package GENES.

Several species appeared in the papers surveyed. The most frequent species were from *Capsicum*, as well as the largest number of studied accessions. Papers from institutions located in all five Brazilian geographic regions were identified in the survey. The Southeast contributed with the highest number of papers and, within this region, the Federal University of Viçosa (UFV) and the North-Fluminense State University Darcy Ribeiro (UENF) ranked first.

The analysis of the papers published in *Revista de Olericultura* (1961-1980) and *Horticultura Brasileira* (1983-2006) concerning genetic resources showed a clear sophistication of the analysis of data related to accession characterization in the course of time, due to the incorporation of more robust statistical procedures. On the other hand, descriptors almost did not change when earlier and more recent papers are compared. Nevertheless, Biodiversity International standardized the descriptors and sorted them by priority. As consequence, data of germplasm characterization increased in accuracy, results gained in discriminating power, and the experimental information turned out to be an efficient tool for duplicate identification in germplasm collections and for selection of the most relevant characteristics for genetic divergence studies, improving the effectiveness of predicting highly heterotic crosses,

amongst other applications.

Most of the information reported in the studied papers came both from quantitative and qualitative (binary and multicategorical data) characteristics, the first not suppressing the second. On the other way around, quantitative and qualitative characteristics were complimentary and concurred to a comprehensive description of the genetic variability among accessions.

In spite of the mounting number of papers regarding genetic resources published along time, there are still some obstacles to overcome. For instance, qualitative data are underexplored both in the calculation of distance matrixes and in joint analysis with quantitative data. It is also essential for future work that papers give more precise information on the methods, variables, and descriptors used. In addition to that, the germplasm of neglected vegetable crops, such as elephant ear (*Xanthosoma sagittifolium* (L.) Schott), West Indian gherkin (*Cucumis anguria* L.), yam (*Dioscorea* sp.), and common sowthistle (*Sonchus oleraceus* L.), which are of nutritional importance and cultural relevance for the Brazilian population, must be more thoroughly characterized and evaluated, in order to produce information that would contribute to disseminating their use.

ACKNOWLEDGMENTS

Authors thank Dr. Paulo Eduardo de Melo, from Embrapa Vegetables, for reviewing the paper.

REFERENCES

AMARAL JÚNIOR AT; SILVA DJH; SEDIYAMA MAN; CASALI VWD; CRUZ CD. 1994. Dissimilaridade genética de descritores botânico-agronômicos e isozimáticos em clones de couve-comum. *Horticultura Brasileira* 12: 113-117.

AMARAL JÚNIOR AT. 1999. Divergência genética entre acessos de moranga do banco de germoplasma de hortaliças da Universidade Federal de Viçosa. *Horticultura Brasileira* 17: 3-6.

BIANCHETTI LB. 1996. *Aspectos morfológicos, ecológicos e biogeográficos de dez táxons de Capsicum (Solanaceae) ocorrentes no Brasil*. Brasília: UNB. 174p. (Tese mestrado).

BUSSAB WO; MIAZAKI ES; ANDRADE DE. 1990. *Introdução à análise de agrupamentos*. São Paulo: ABE. 105 p.

COOPER HD; SPILLANE C; KERMAI I; ANYSHETY NM. 1998. Harnessing plant genetic resources for sustainable agriculture. *Plant Genetic Resources Newsletter* 114: 1-8.

CROSSA J; FRANCO J. 2004. Statistical methods for classifying genotypes. *Euphytica* 137: 19-37.

CRUZ CD. 2006. Programa GENES – *Análise Multivariada e Simulação*. Viçosa: UFV. 175p.

CRUZ CD; CARNEIRO PCS. 2003. *Modelos biométricos aplicados ao melhoramento genético*. Viçosa: UFV, 2003. 585p.

EMBRAPA. 1996. *Sistema de produção de frutas e hortaliças*. Brasília: Embrapa. 68p.

FILGUEIRA FAR. 2000. *Novo manual de olericultura: agrotecnologia moderna na produção e comercialização de hortaliças*. Viçosa: UFV. 402p.

FRANCO J; CROSSA J; VILLASEÑOR J; TABA S; EBERHART SA. 1998. Classifying genetic resources by categorical and continuous variables. *Crop Science* 38: 1688-1696.

JACCARD P. 1901. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull Soc Vandoice Sci Nat* 37: 547-579.

MOTTA JH; YURI JE; REZENDE GM; SOUZA RJ. 2006. Similaridade genética de cultivares de alho pela comparação de caracteres morfológicos, físico-químicos, produtivos e moleculares. *Horticultura Brasileira* 24: 156-160.

NASS LL (2001). Utilização de recursos genéticos vegetais no melhoramento. In: NASS LL; VALOIS ACC; MELO IS de; VALADARES-INGLIS MC (eds). *Recursos genéticos & melhoramento – plantas*. Rondonópolis: Fundação MT. p. 29-55.

NTSYSpc. 2007, October 11. *Numerical Taxonomy System*, version 2.2 for windows. Available in: <http://www.exetersoftware.com>.

PEARSON K. 1901. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* 2: 559-572.

PEDROSA JF; MIZUBUTI, A. 1976. Caracterização de 100 introduções de quiabeiro (*Abelmoschus esculentus*, Moench) do banco de germoplasma de hortaliças da Universidade Federal de Viçosa – MG. *Revista de Olericultura* XVI : 164-166.

SAS Institute Inc. 1998. Statistical analysis system. Release 8.02, Cary, NC.

SAS – Statistical Analysis Software. 2007, 13 de dezembro. Disponível em <http://www.sas.com/technologies/analytics/statistics/stat/index.html>.

SHEPHERD GJ. 2001. *Fitopac 1*. Manual do usuário. Campinas: Universidade de Campinas. 93 p.

SILVA DJH; Moura MCCL; CASALI VWD. 2001. Recursos genéticos do banco de germoplasma de hortaliças da UFV: histórico e expedições de coleta. *Horticultura Brasileira* 19: 108-114.

SINGH D. 1981. The relative importance of characters affecting genetic divergence. *The Indian Journal of Genetic and Plant Breeding* 14: 237-245.

- SUDRÉ CP; RODRIGUES R; RIVA EM; KARASAWAM; AMARAL JÚNIOR AT. 2005. Divergência genética entre acessos de pimenta e pimentão utilizando técnicas multivariadas. *Horticultura Brasileira* 23:22-27.
- SUDRÉ CP; CRUZ CD; RODRIGUES R; RIVA EM; AMARAL JÚNIOR AT; SILVA DJH; PEREIRA TNS. 2006. Variáveis multicategóricas na determinação da divergência genética entre acessos de pimenta e pimentão. *Horticultura Brasileira* 24: 88-93.
- VALOIS ACC. 2005. Acesso aos recursos genéticos e repartição de benefícios: uma visão atual e de futuro. In: LIMA MC (ed). *Recursos genéticos de hortaliças: riquezas naturais*. São Luís: Instituto Interamericano de Cooperação para a Agricultura. p. 15-54.
- VICTÓRIA DC, GARCIA AAF, SOUZA JÚNIOR AP. 2001. Desenvolvimento de um programa SAS para cálculo de coeficiente de similaridade de dados de marcadores moleculares utilizando *bootstrap*. In: CONGRESSO NACIONAL DE GENÉTICA, 47. *Resumos do 47º Congresso Nacional de Genética*. Águas de Lindóia: SBG (CD-ROM).
-