STATISTICAL ANALYSIS OF GENETIC DIVERSITY USING TOMATO DATABASE

NIKOLAYA VELCHEVA
Institute of Plant Genetic Resources "K. Malkov", Sadovo, Bulgaria
E-mail: nikolaya_velcheva@abv.bg

Abstract

System R is an environment for creating advanced applications in statistics, data analysis and graphics. This software is successfully used to analyze large databases from investigation of genetic diversity stored in *ex situ* collections.

During the period 2007 – 2010 at the IPGR – Sadovo 63 determinant tomato accessions with a different geographical origin from the Bulgarian National Collection were studied. Important 15 morphological and biochemical parameters were evaluated. The obtained results were processed by statistical package System R. The analysis of the studied traits showed a great genetic diversity and it grouped accessions in the collection. This makes the genotypes an interesting genetic resource for future research activities and inclusion in tomato breeding programs.

The evaluation of the experimental data has been supported by the SEEDNet project as a part of specialization ,,Statistical Analysis of Genetic Diversity and Scientific Publication" in Swedish Biodiversity Centre, Alnarp.

Key words: tomato database, genetic diversity, statistical analysis, System R

INTRODUCTION

R is a language for statistical programming, which gained popularity worldwide. Its advantage over other software systems with graphical interface and programming languages is that it is applicable to both traditional statistical analysis and application of non-traditional statistical solutions. This advantage is most clearly manifested in science research and in particular to analyze large data sets in the study of genetic diversity, stored in *ex situ* collections in the world centers of genetic resources. (Crawley, 2005; Onkov, Ganchev, 2010).

The preservation of the plant biodiversity is the main priority in the research work in the IPGR –

Sadovo, related to the implementation of the National Programme for Plant Genetic Resources. Germplazm collections from the main crops, which fully represent the genetic diversity in the context of the contemporary criteria for the future significance of the initial material in the selection programmes, have been created (Krasteva et al., 2007).

Tomato is the most common vegetable crop in Bulgaria. This is explained by the suitable climate – soil conditions and its ecological plasticity, high fertility, multi-purpose use and the biological value of the fruit (Krasteva, 1995). The structure of the tomato collection includes: introduced cultivars; new Bulgarian cultivars; old primitive cultivars, removed from the national cultivar list; local populations; wild relatives; and breeding lines (Kratseva, Todorova, 2003).

The purpose of this study is a comparative evaluation of 63 tomato accessions from the Bulgarian National collection and to determinate genetic diversity in order to review the current potential of the conserved gene fund in the direction of further selection.

There were two hypotheses in the beginning of this experiment: the null – all genotypes are the same in the studied characteristics, the second – there is at least one of them that is different.

MATERIAL AND METHODS

Plant material and study parameters. During the period 2007 – 2010 at the Institute of Plant Genetic Resources – Sadovo, 63 determinant tomato accessions with a different geographical origin from the Bulgarian National Collection were studied. Important 15 morphological and biochemical parameters by tomato descriptor (IPGRI, 1996) were evaluated. 45 introduced accessions, 15 local forms from expeditions and 3 Bulgarian cultivars as standards were the object of the study. The seedlings production was done in the glasshouse of the Institute. The plants were sown in May on the experimental vegetable field. The tomatoes were grown based on a common technology for middle - early field production (Krasteva et al., 2007a). During the different stages of the crop development, biometric measurements of 10 plants from all tested accessions were made. The biochemical evaluations of the fruit were performed in the Biochemistry Laboratory of the Institute (Stanchev, Boboshevska, 1974).

Statistical data analysis. The obtained results were processed using statistical package System R (Quick, 2010). Variability of the study traits and cluster analysis of the experimental data was made (Ganeva et al., 2006; Mokreva, 2007; Ivanova et al., 2008), using prepared specific algorithms and graphics.

RESULTS AND DISCUSSION

The analysis shows a relatively high genetic diversity in the collection and the variability of each stud-

ied morphological and biochemical characteristics are presented by histograms (Figure 1).

The fruit mass is an important characteristic with a high economical importance to the culture. In terms of this it is seen more diverse in the large-fruited tomatoes (weight over 100 g), which are suitable for direct consumption and also for canning. In the small-fruited accessions (100 g and less), which are suitable for processing, the diversity of weight is less.

The multidirectional evaluation of the tomato accessions allowed a realistic clustering of the collection. In the first dendogram (Figure 2), the cluster analysis separates the genotypes by their differences between morphological and biochemical indices. The bright color corresponds to the dominant parameter in determining the groups. The major factor in the grouping has proven to be the fruit weight. It has formed two clusters with large-fruited tomatoes and two with small-fruited which are merged into two groups in a greater distance. The indicator of plant height is the second factor that affects clustering in the collection. The characteristics length and width of tomato leaves and vitamin C content in the fruit have also active participation with their values. These five indicators have in the most highly stressed genetic diversity in the collection.

More detailed visualization of the magnitude of each of the investigated indicators on each of the genotypes in the already described grouping is shown in Figure 3.

Color presentation in the small-group accessions as the best gene sources indicators detailed in dry matter content and sugar-acid coefficient in fruit are respectively A400033 (Israel) and A8E0258 (local, Bulgaria). The dry matter content is one of the most important parameters determining the quality of the fruit and its suitability as a material in the canning industry. The ratio sugars/acids is an important chemical characterization of the fruit and determines the taste of tomatoes.
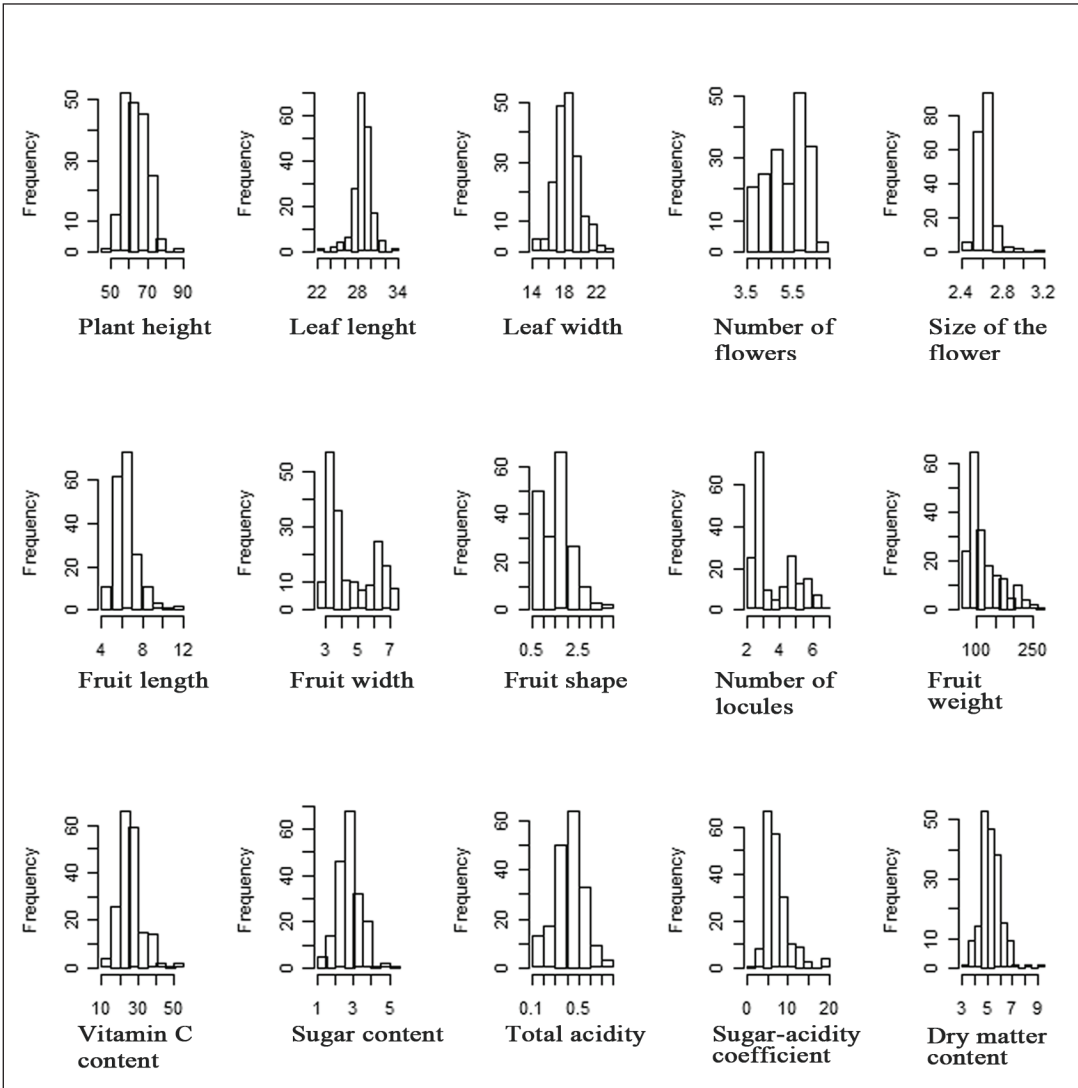

Fig. 1. Histograms of variability for each trait in the tomato collection
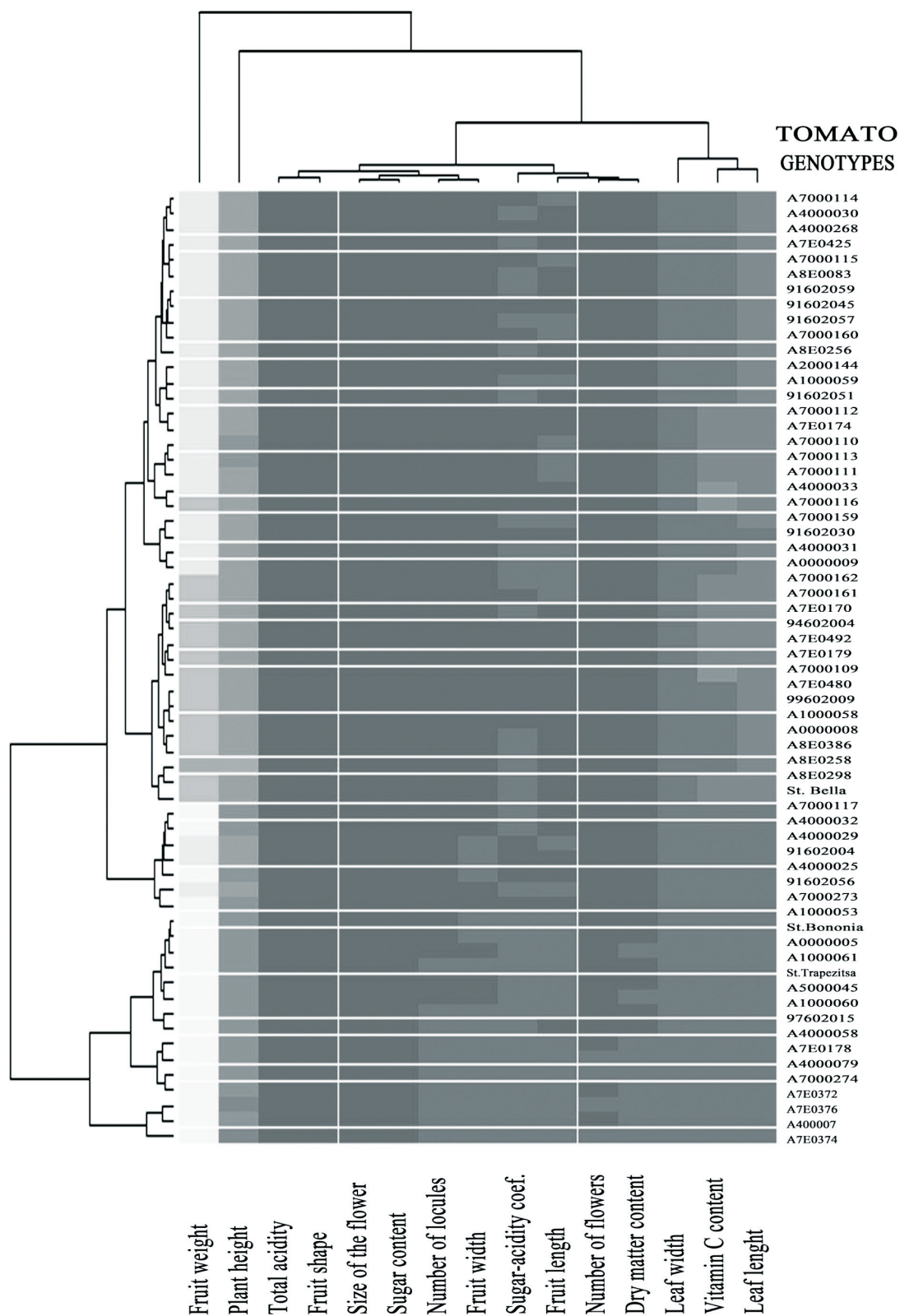
Fig. 2. Dendogram of tomato genotypes grouping based on differences between indices
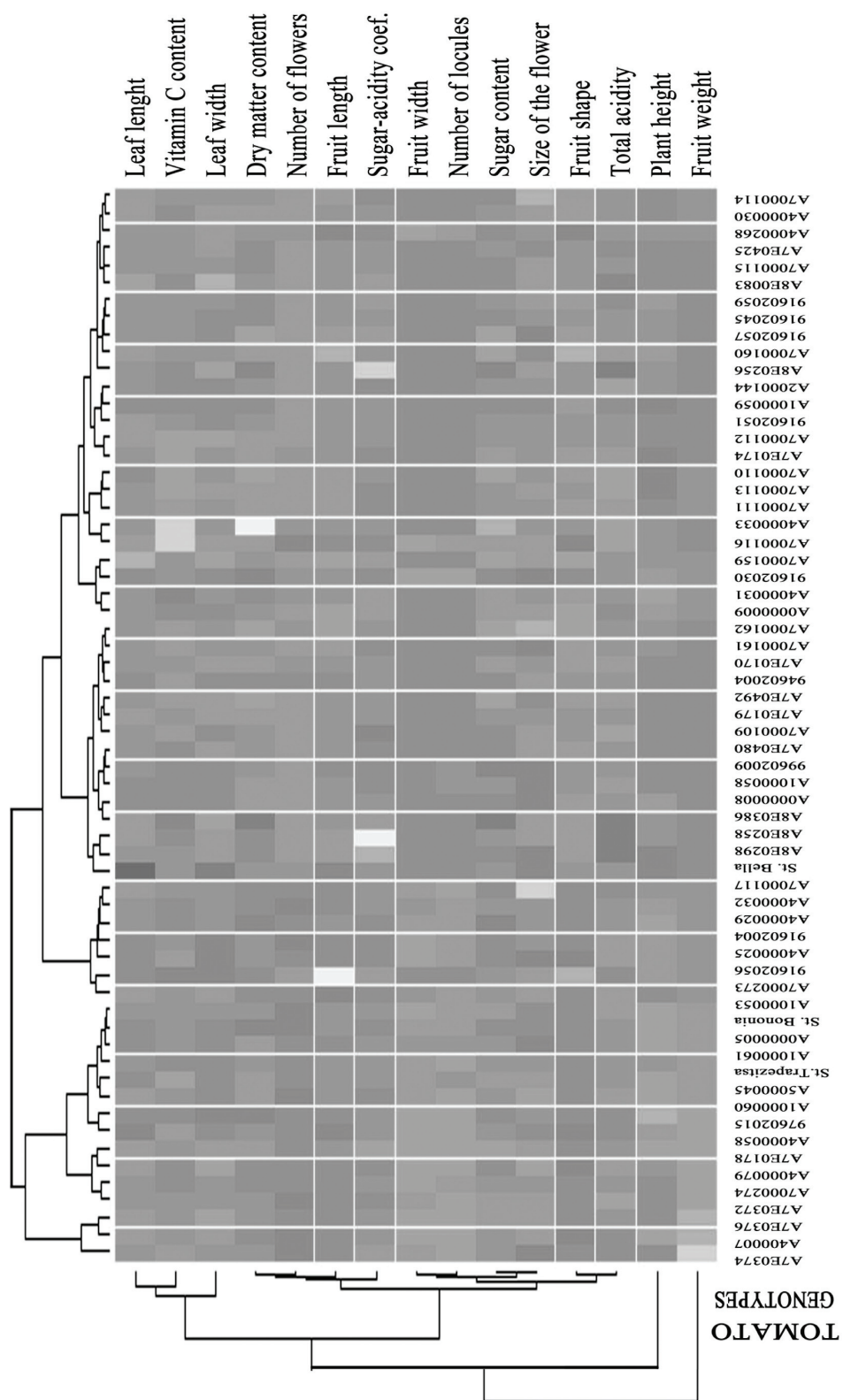
Fig. 3. Dendogram of magnitude of each parameter for each genotype

The mass, number of locules and width of the fruit are the dominant factors in the clusters, bringing large-fruited accessions together. In line with the histogram for mass, the cluster analysis also shows great diversity in fruit weight in the large-fruited group. For the other indicators in the two large-fruited clusters, a relatively high diversity in the magnitude of their values is also observed.

Accession with catalogue number 91602056, introduced by the USA as a canning tomato is distinguished. It is characterized by the longest fruit, but because of the relatively high weight of the fruit, it has fallen into the cluster of large-fruited accessions. In further studies it is advisable to refer this genotype to a group for processing.

While comparing the results it became clear that there is a wide genetic diversity in the studied tomato collection. Regarding the indicator of fruit weight, diversity in the large-fruited group is greater. For the rest morphological and biochemical parameters in the small-fruited group a greater genetic diversity is observed.

Based on the results of this study we found out that the future investigation of the collection would be separated in large-fruited and small-fruited groups. This is an important fact that would make the analysis more useful for the branches of the breeding of determinant tomato cultivars for direct consumptions and processing. Statistical analysis of the experimental data facilitates the choice of genotypes with a complex of valuable economical traits for each direction.

CONCLUSIONS

The investigation has shown relatively high genetic diversity in the determinant tomato collection. This makes the genotypes interesting genetic resource for future research activities.

The analysis has clustered the collection in two main groups: large-fruited and small-fruited accessions for processing.

The study of the genetic diversity in tomato collection indicates the applicability of System R to automatic processing of large databases in PGR and for appropriate visualization of the results. Its use for future evaluation of other collections stored at the National Genebank could be contributed for characterization of the conserved genetic diversity.

REFERENCES

Crawley, M. 2005. Statistics: An Introduction using R. Wiley. ISBN 0-470-02297-3.

Ganeva, D., I. Ivanova, G. Pevicharova. 2006. Identification of determinate tomato hybrids F1 using cluster analysis. Proc. of First Int. Sym. ,,Ecological Approaches towards the Production of Safety Food", Plovdiv, 205-210

Ivanova, I., D. Ganeva, G. Pevicharova. 2008. Determination of viable F1 hybrid tomatoes for industrial processing through multiobjective optimization. Proc. of VIIth Nat. Scientific Conf. with Int. participation ,,Environment and Health 2008", Plovdiv, 297-302

Krasteva, L. 1995. Plant resources of determinant tomatoes and their use. Sci. Works of the AU – Plovdiv, 40, 1, 13-16

Krasteva, L., T. Todorova. 2003. The Bulgarian *Solanaceae* collections. *Solanaceae* Genetic Resources in Europe. IPGRI, Rome, Italy, 15-27

Krasteva, L., S. Angelova, Y. Guteva, K. Varbanova. 2007. Management and sustainable use of plant genetic resources. 125 Years of Agricultural Science. IPGR, Sadovo, 1, 43-48

Krasteva, L., St. Masheva, D. Ganeva, M. Mihov. 2007a. Rules for good agricultural practices in the production of middle-early cultivars of tomatoes. IPGR, Sadovo.

Mokreva, T. 2007. Comparative characteristics of experimental criteria and algorithm for evaluation of experimental data in viticulture. PhD Thesis. Plovdiv, 145 p.

Onkov, K., D. Ganchev. 2010. Language for statistical programming R. 2010. Automation and Informatics, 2, 34-38

Quick, J. 2010. The Statistical Analysis with R Beginners Guide. Packt Publishing. ISBN 1849512086.

Stanchev, L., D. Boboshevska. 1974. Manual for Laboratory practice in agrochemistry. Plovdiv.

IPGRI. 1996. Descriptors for tomato *Licopersicon* spp. Rome, Italy.