E

BMT/6/10
ORIGINAL: English
DATE: February 11, 2000

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

GENEVA

WORKING GROUP ON BIOCHEMICAL AND MOLECULAR TECHNIQUES AND DNA-PROFILING IN PARTICULAR

Sixth Session
Angers, France, March 1 to 3, 2000

THE GEVES SOFTWARE PACKAGE FOR ESTIMATING GENETIC DISTANCES BETWEEN VARIETIES, WITH OR WITHOUT LINKAGE MAP INFORMATION, AND ANALYZING THE GENETIC DIVERSITY OF A COLLECTION OF VARIETIES THROUGH MOLECULAR DATA

prepared by experts from France

THE GEVES SOFTWARE PACKAGE FOR ESTIMATING GENETIC DISTANCES BETWEEN VARIETIES, WITH OR WITHOUT LINKAGE MAP INFORMATION, AND ANALYZING THE GENETIC DIVERSITY OF A COLLECTION OF VARIETIES THROUGH MOLECULAR DATA

P. Dubreuil, V. Lombard & C.P. Baril

The GEVES software package LCDMV (Logiciel de Calcul de Distances Moléculaires entre Variétés)

LCDMV is a software package developed with SAS using SAS-STAT and SAS-IML. It allows the analysis of biochemical (isozymes) or molecular (RFLP, STS, SSR, RAPD, AFLP…) data obtained on homogeneous or heterogeneous varieties. Its first vocation is to estimate genetic distances between varieties and to analyze the structure of a collection of varieties through molecular data.

Three files are necessary to use LCDMV describing the varieties, the markers and the frequencies of markers (table 1).

Six estimators of distance between varieties can be worked out according to the type of varieties and the type of markers (table 2).

LCDMV successively reads the entry files and automatically identifies the type of varieties (homogeneous or heterogeneous) and the type of makers (allelic or banding data). If different types of markers are present in the same file, all the markers are analyzed as banding data. The following outputs are provided:

- histogram of markers frequencies in the analyzed collection,
- Principal Component Analysis (PCA) on varieties and individuals (for heterogeneous varieties),
- genetic distances adapted to the combination type of varieties – type of markers with the precision of the estimates,
- confidence interval calculated through analytical or bootstrap approach (the re-sampling procedure is computed on markers/locus and individuals for heterogeneous varieties),
- histogram of genetic distances estimated between all pairs of varieties and plot of very closed varieties,
- hierarchical clustering (three possible criteria) based on genetic distances with the tree and the associated robustness of nodes (permutation procedure),
- for chosen pairs of inbred lines, the genetic map is given with identification of locus for which the two lines are different,
- polymorphism of markers in the collection analyzed and within each heterogeneous variety.

The diagram of the calculation chain of LCDMV is showed on figure 1.

Improvement of the precision of genetic distance estimates using a consensus linkage map.

In the context of plant registration and protection, estimation of pedigree relatedness between cultivars is a crucial aspect of the protection of the intellectual property of plant breeders (plant breeders' rights). For that purpose, several genetic distance estimators based on molecular data have been proposed. The knowledge of the precision of the estimates is also particularly important to judge if a candidate variety is too closely related to a protected cultivar.

Seventeen primer combinations revealed a total of 324 polymorphic bands among 83 rapeseed cultivars. 218 AFLP markers were mapped and 115 markers were retained for the calculation of Rogers and BLUE pairwise distances. The average distance between AFLP markers on the consensus map was 17 cM, with a standard deviation of 14.4 cM.

The Rogers distance estimator and the BLUE estimator were highly correlated ($r=0.98$, $p<0.001$) and led to similar groupings (results not showed). The comparison between the sampling standard deviations of the two estimators showed that BLUE standard deviation is 25% lower than Rogers standard deviation (figure 2). This gain of precision was strictly due to the distribution of the markers on our linkage map.

The following figures show the influence of the marker density on the standard deviation improvement obtained with the BLUE estimator. Figure 3 shows the improvement of the standard deviation (Std) ((Rogers Std – BLUE Std) / Rogers Std) obtained with the BLUE estimator when the density of markers ranged from 1 marker per 90 cM to 1 marker per 5 cM. With a density of 1 marker per 20 cM which was closed to the density of our map, the theoretical gain of precision is 45%. The discrepancy between the simulations and the results based on our mapped AFLPs may come from the difference of the marker distributions between the simulated map (uniform) and the experimental map.

The dotted and plain curves in figure 4 show coefficient of variation of Rogers and BLUE estimators respectively, in terms of the number of markers (or markers density) for 4 genetic distances. For closely related cultivars ($d=0.05$), 150 markers are required to achieve a satisfactory level of precision (CV=10%) with BLUE distance, whereas more than 5 times this number is required with Rogers distance.

This study showed that the knowledge of the distribution of markers on a linkage map improved the precision of estimates of genetic distances between cultivars. The simulation of a genetic linkage map provides useful information to determine the density of molecular markers required for a given level of precision. The BLUE estimator is a convenient statistical tool to deal with aspects of plant registration such as the protection of plant breeders' rights.

References

Dillmann C., Charcosset A., Goffinet B., Smith J.S.C., Dattée Y. 1997. Best linear unbiased estimator of the molecular genetic distance between inbred lines. Advances in Biometrical Genetics. Proceedings of the Tenth Meeting of EUCARPIA Section Biometrics in Plant Breeding, Posnan, 14-16 may 1997, P. Krajewski and Z. Kaczmarek (eds.), pp. 105-110

Foulley J.L. and Hill W.G. 1999. On the precision of estimation of genetic distance. Genet Sel Evol 31, 457-464

Ghérardi M., Mangin B., Goffinet B., Bonnet D. and Huguet T. 1998. A method to measure genetic distance between allogamous populations of alfalfa (*Medicago sativa*) using RAPD markers. Theor Appl Genet 96: 406-412

Jaccard P. 1900. Contribution au problème de l'immigration post-glaciaire de la flore alpine. Bulletin de la Société Vaudoise des Sciences Naturelles 37:547-579

Jaccard P. 1908. Nouvelles recherches sur la distribution florale. Bulletin de la Société Vaudoise des Sciences Naturelles 44: 223-270

Nei M. and Li W.H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Nat Acad Sci USA 76: 3269-3273

Rogers J.S. 1972. Measures of similarities and genetic distances. Studies in genetics VII. Univ Texas Publ 7213:145-153

Table 1: three entry files of LOCIMAP

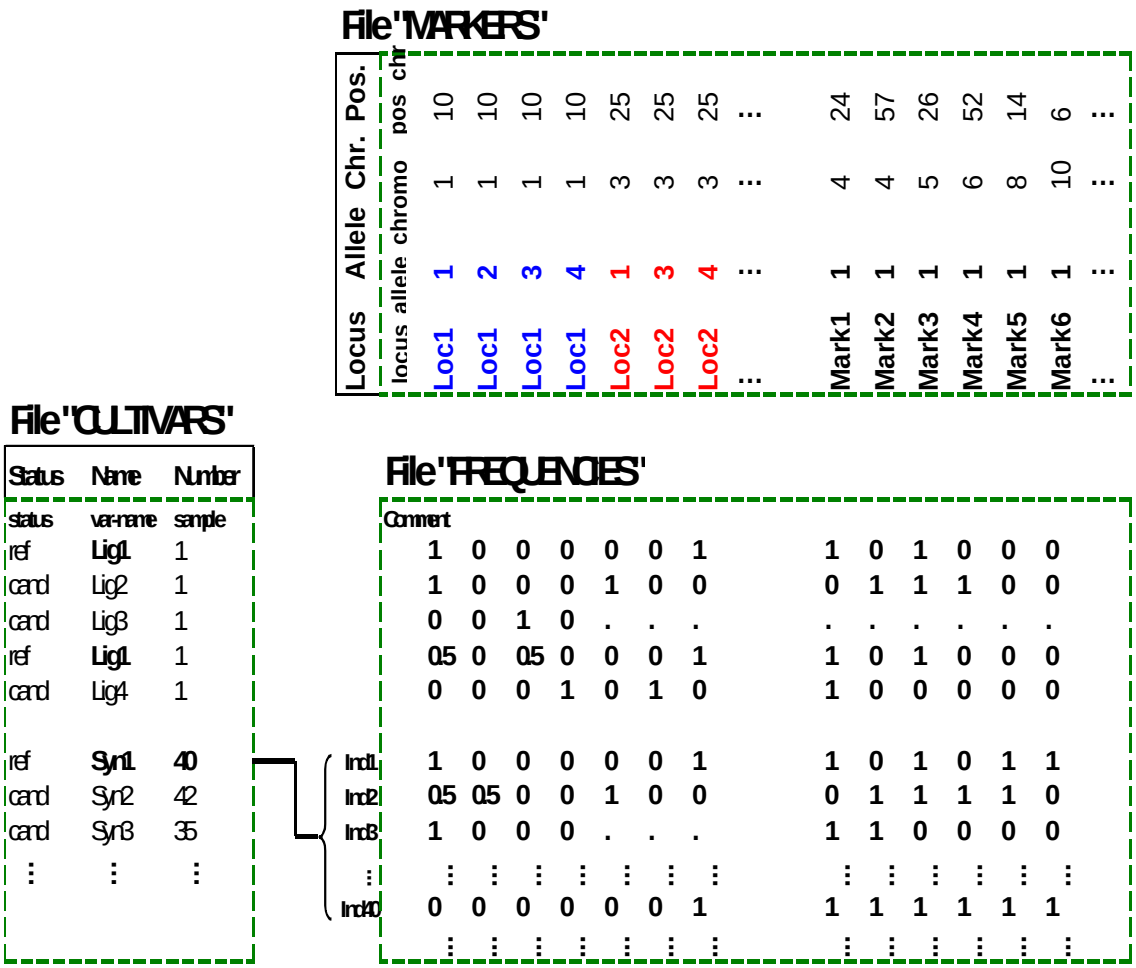

Table 2: 6 estimators of genetic distance	Molecular information « locus-alleles »	Molecular information « bands »
Homogeneous varieties	Rogers distance **(1972)** If linkage map: BLUE **distance (Dillmann et al., 1997)**	Nei&Li distance **(1979)** Jaccard distance **(1900 ;1908)**
Heterogeneous varieties	**Rogers distance** (unbiased estimator ; Ghérardi et al., 1998) Sanghvi distance **(Foulley et Hill, 1999)**	**Rogers distance** (unbiased estimator ; Ghérardi et al., 1998) under the biallelism hypothesis

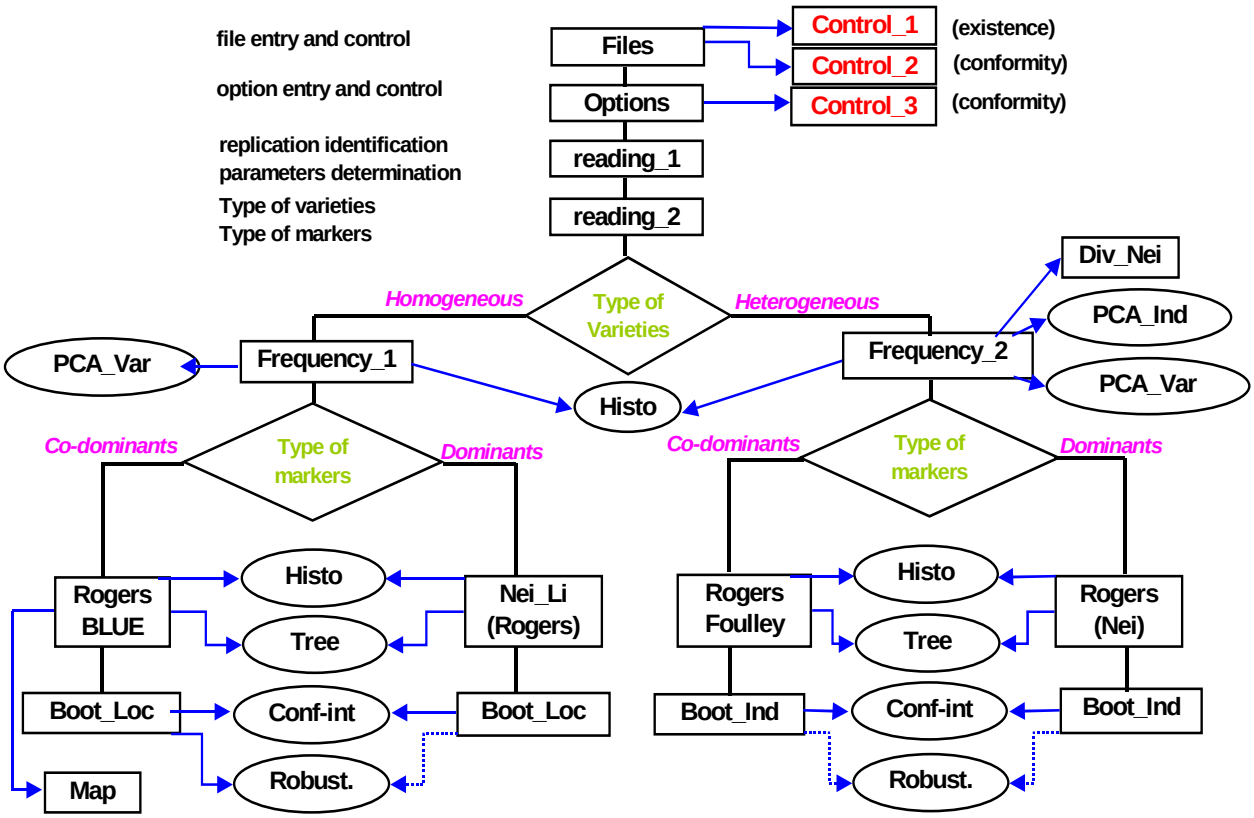

Figure 1: Diagram of the calculation chain of LCDMV package

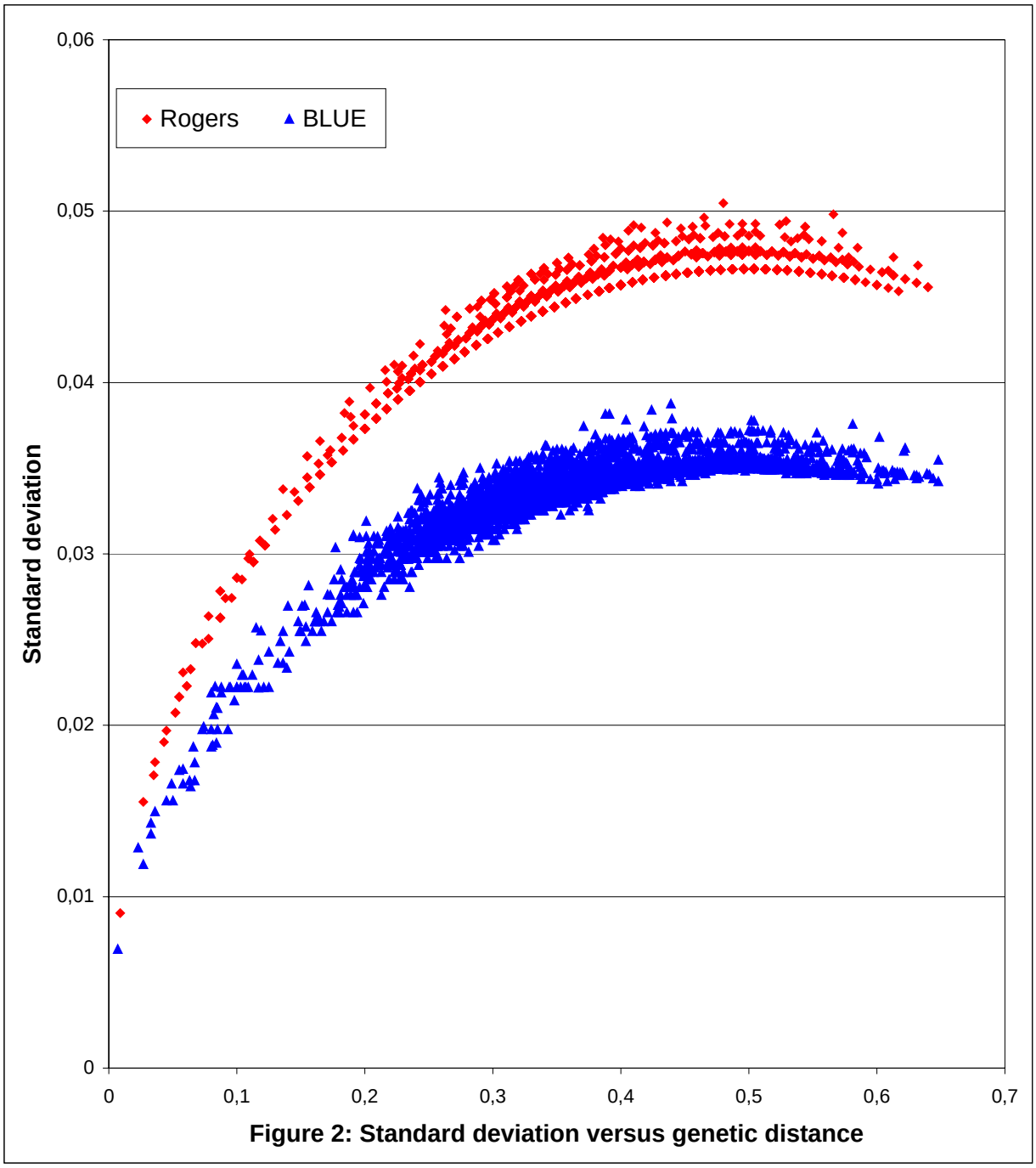

Figure 2: Standard deviation versus genetic distance

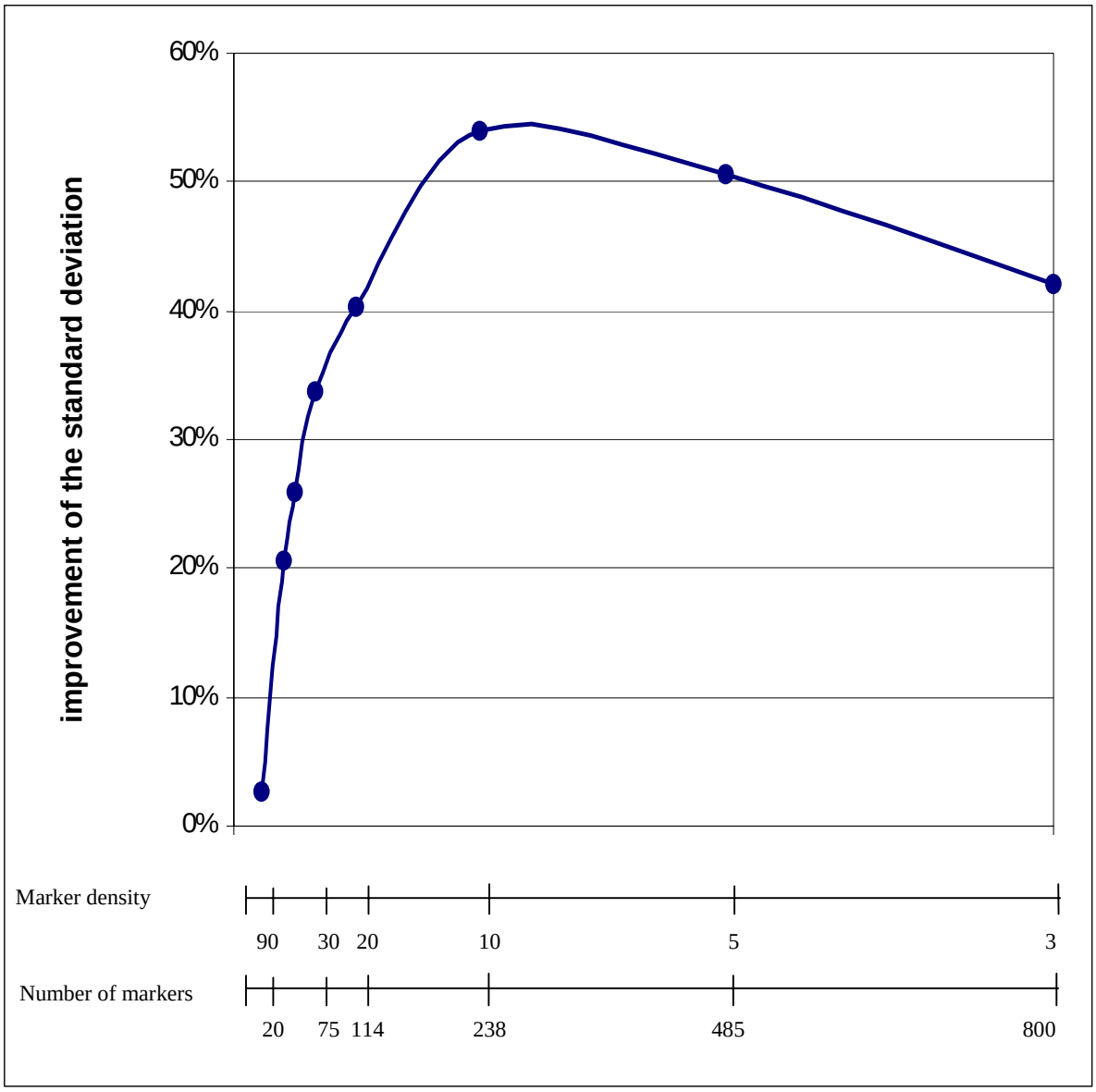

Figure 3: Improvement of the standard deviation according to the marker density

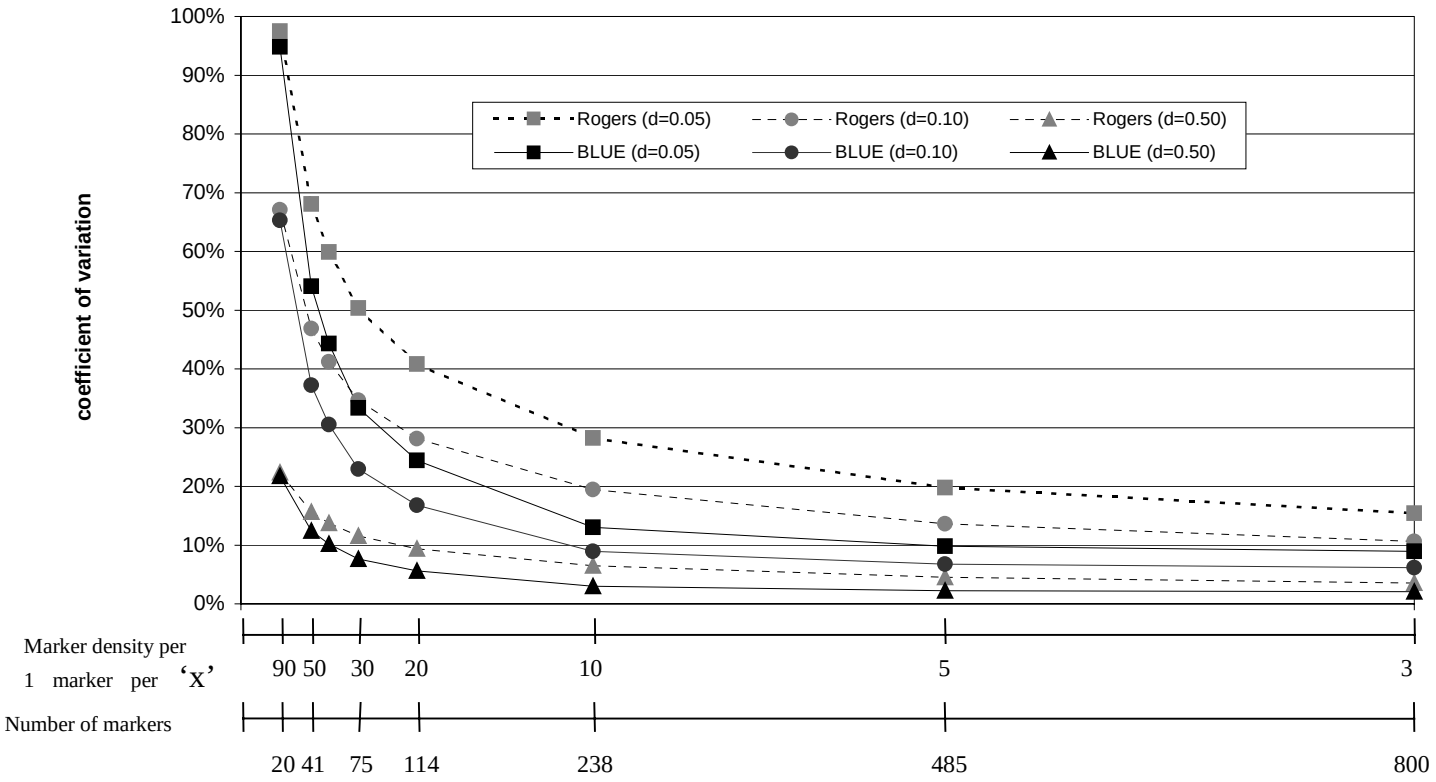

Figure 4: Coefficient of variation of Rogers and BLUE estimators in terms of markers density.

[End of document]