Theor Appl Genet (1993) 86:1028–1032

THEORETICAL AND APPLIED GENETICS

Genetic similarity among ecotypes of *Arabidopsis thaliana* estimated by analysis of restriction fragment length polymorphisms

G. King[1], J. Nienhuis[2], C. Hussey[1]

[1] AgriDyne Technologies Inc., 417 Wakara Way, Salt Lake City, UT 84108, USA
[2] Department of Horticulture, University of Wisconsin, Madison, WI 53706, USA

Received: 10 February 1992 / Accepted: 28 January 1993

Abstract. Genetic similarity was estimated among a sample of 28 ecotypes of *Arabidopsis thaliana*. Twenty-five previously mapped genomic clones were used as probes in Southern hybridizations to detect restriction fragment length polymorphisms (RFLPs). A total of 62 polymorphic restriction fragments were classified as to their presence or absence for each genotype. The genetic similarity between each pair of ecotypes was calculated as the ratio of concordant to total bands scored. The mean genetic similarity among the 28 ecotypes was 0.69 and ranged from 0.32 to near 1.0. No relationship was observed between genetic similarity and geographical origin of the 28 ecotypes. The ecotype most distantly related to the other 27 was Niederzenz, with a mean genetic similarity of 0.55 ± 13. A bootstrap procedure was used to generate 200 random samples of bands of size n ($n = 8, 12, 16, \ldots, 55$), and the coefficient of variance (CV) was estimated for each sample. The plot of the first two principal components provided a description of the relative genetic similarity among ecotypes. The results provide information useful to investigators interested in sampling the genetic variation among *Arabidopsis* ecotypes.

Key words: *Arabidopsis thaliana* – Bootstrap – RFLPs – Principal component analysis – Genetic similarity

Introduction

Arabidopsis has become the model system for plant molecular biologists because of its small size, short life cycle, small genome, and low level of interspersed repetitive DNA. Although thousands of accessions of *Arabidopsis* have been collected, much current research has focused on the utilization of only a very few ecotypes, e.g., Niederzenz, Columbia, and Landsberg (Chang et al. 1988). To our knowledge, only one previous study has capitalized upon the wide genetic variation available in *Arabidopsis* (Langridge and Griffing 1959). Genetic variation may exist among ecotypes that could prove useful in extending the possibilities of *Arabidopsis* as a model species in plant molecular biology research. Knowledge of the relative genetic similarity among ecotypes is useful because it permits characterization of germ plasm resources and provides more efficient sampling of genetic diversity.

Seed of many *Arabidopsis* ecotypes has been increased under varying environmental conditions in different laboratories at different times; hence it is possible that selection, migration (seed mixtures), or genetic drift (sampling errors) may have resulted in populations that are sufficiently divergent genetically to affect the reproducibility of experimental results. In addition, damage to storage facilities during World War II (M. Koornneef, personal communication) has raised questions regarding the purity and origin of some accessions stored in European repositories.

The genetic similarity between two populations is directly proportional to the concordance between their base pair sequences. In lieu of direct sequence data, indirect measures which reflect underlying base pair differences are commonly used for determining relatedness among ecotypes. In cultivated plants, pedigrees can be used to estimate the coefficient of parentage; however, this information is not commonly available in *Arabidopsis*. Phenotypic information on quantitative, morphological, or biochemical characteristics

Communicated by A. L. Kahler
Correspondence to: G. King

also can be used to estimate genetic similarity (Doebley et al. 1990; Miller and Tanksley 1990; Smith 1984; Smith et al. 1990). However, environmental effects as well as genotype by environmental interactions can result in the failure of the phenotype to accurately represent the genotype. RFLP (restriction fragment length polymorphism) maps have been developed for *Arabidopsis* (Chang and Meyerowitz 1986; Nam et al. 1989) and provide an opportunity to characterize *Arabidopsis* germ plasm. RFLP information has been used to determine the probable progenitor of maize (Doebley et al. 1990), to demonstrate phylogenetic relationships in the genus *Lycopersicon* (Miller and Tanksley 1990), and to estimate genetic similarity relative to known pedigrees among elite maize hybrids (Smith et al. 1990).

The objective of the research presented here was to use RFLP molecular markers to estimate genetic similarity among a sample of *Arabidopsis* ecotypes. The term "Ecotype" refers to a population that has maintained it's identity through isolation or selection in a specific environment.

Materials and methods

Plant material

Seeds of 28 *Arabidopsis thaliana* (L.) ecotypes representing a wide range of geographical diversity were obtained from several sources, including: Dr. A. Kranz, *Arabidopsis* Information Service, Frankfurt, Germany; Dr. R. Last, Boyce Thompson Institute, Ithaca, NY, and Dr. E. Meyerowitz, Cal Tech, Pasedena, Calif. Seeds of each ecotype were produced through two cycles of bulk increase in a Conviron growth chamber at 22 °C with a 20-h photoperiod. The names, geographical origins, and other available information regarding the ecotypes are provided in Table 1. The most common laboratory strains of *Arabidopsis*. Columbia and Landsberg, were included as standard references. Phenotypic differences (size of leaves, length of flower stalk, leaf surface, etc) were observed among the 28 ecotypes.

Isolation and treatment of plant DNA

DNA was extracted from 30 plants of each of the 28 ecotypes after 4 weeks of growth in a conviron chamber at 22 °C with a 12-h photoperiod. Lyophilized plant material was extracted with a CTAB (mixed alkyltri-methyl-ammonium bromide) buffer at 60 °C (Saghai-Maroof et al. 1984). This was followed by a chloroform/octanol extraction and then an extraction with 1/10th vol. of 10% CTAB in 0.7 *M* NaCl. The supernatant fluid was precipitated with buffer containing 5 m*M* Tris (pH 8.0), 10 m*M* EDTA, and 1% CTAB, followed by a precipitation in 95% ethanol. RNase was included in the restriction enzyme digestions. Five micrograms of DNA were digested with either *Bgl*II or *Eco*RI and hybridized after Southern blotting with lambda clones containing mapped RFLP markers.

RFLP markers

Twenty-five lambda clones containing *Arabidopsis* RFLP markers were received from the laboratory of Dr. E. Meyerowitz and subsequently grown from single plaques to high titer stocks in *E. coli* c600 by the plate lysate method (Sambrook et al. 1989). The phage stocks were purified from the lysate by DEAE chromatography (White and Rosenzweig 1989). DNA was purified following PEG precipitation of the phage and phenol extraction. The intact phage DNA was labeled with [^{32}P] with T7 polymerase (Stratagene Prime-It Kit) and hybridized to Southern blots of *Eco*RI and *Bgl*II digestions.

Genetic similarity analyses

A total of 62 polymorphic restriction fragments (bands) were scored. Each band was treated as a unit character, and the genotype was scored for the presence (1) or absence (0) of a band. From 1 to 4 polymorphic bands were scored for each of the 25 probes. Genetic similarity between each pair of ecotypes (378) was calculated using a method similar to that reported by Nei and Li (1979), i.e.,

$$GS(XY) = C(XY)/N(XY)$$

where GS(XY) is the measure of genetic similarity between lines X and Y, C(XY) is the number of concordant bands between lines X and Y, and N(XY) is the total number of bands scored.

Table 1. *Arabidopsis* ecotypes included in RFLP characterization of genetic similarity

Name	Origin[a]	Source[b]
An-1	Belgium, E4-5/N51-52	A.I.S. 1777
Rsch-4	USSR, E34/N56-57	A.I.S. 1795
Mt-0	Libya, E38/N46	A.I.S. 2263
En-2	Germany, E8-9/N50	A.I.S. 2590
Np-0	Germany, E11-12/N52-53	A.I.S. 2254
Wil-2	USSR, E25/N55	A.I.S. 1783
Ct-1	Italy, E15/N37-38	A.I.S. 2225
Eil-0	Germany, E12-13/N51-52	A.I.S. 2635
Edi-0	Scotland, E3/N56	A.I.S. 1791
Lip-0	Poland, E19-20/N50	A.I.S. 1830
Tsu-0	Japan, E36-37/N34-35	A.I.S. 958
Bla-12	Spain, E3/N42	A.I.S. 1818
Sv-0	Denmark	A.I.S. 1518
Nw-4	Germany, E8-9/N50-51	A.I.S. 1741
No-0	Germany, E13-14/N51	A.I.S. 1687
As-0	Germany, E10/N51	A.I.S. 2128
Sf-2	Spain, E3/N41-42	A.I.S. 2166
Oy-0	Norway	A.I.S. A2226
Ita-0	Ithaca, NY	Last
Co-4	Portugal	Last
Ms-0	USSR, E38/N56	Last
Landsberg	Landsberg, Poland	Meyerowitz
Columbia	Selection from Landsberg	A.I.S. 1986
La-0	Selection from Landsberg	A.I.S. 2061
Ni	Niederzenz	Last
Be-0	also known as Bensheim	Last
C-24	see Valvekens et al. (1988)	Last
Rld	Selection from Rsch-4	Last

[a] Geographical origin and coodinates, if known.
[b] A.I.S., Arabidopsis Information Service, Frankfurt, Germany; Last, Rob Last, Boyce Thompson Institute, Ithaca, NY; Meyerowitz, Elliot Meyerowitz, Cal Tech, Pasadena, Calif.

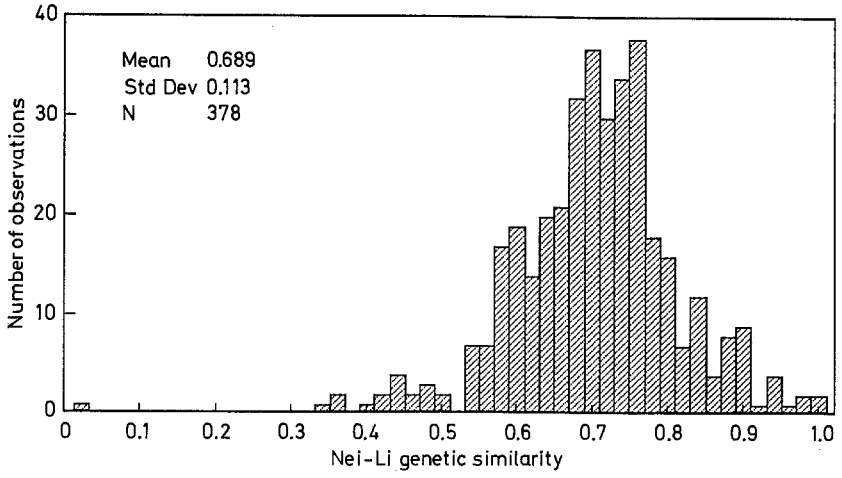

Fig. 1. Distribution of Nei-Li genetic similarity values among 28 *Arabidopsis* ecotypes

To estimate the variance of genetic similarity values, 200 random bootstrap samples each of size n ($n = 8, 12, 16, 20, 26, 33, 44$ and 55) were drawn from the data set (Efron and Tibshirani 1986). The genetic similarity between all pairs of genotypes (378) was calculated for each bootstrap sample. The variance was standardized to the coefficient of variance (CV) by dividing the variance by the bootstrap sample mean.

Results

The genetic similarity values for each pair of ecotypes (378) are approximately normally distributed about a mean of 0.69 and ranged from 0.32 to near 1.0 (Fig. 1). The original data and the matrix of genetic similarity values are too lengthy to be presented here, but are available from the authors. The first and second principal components of the genetic similarity matrix accounted for 70.3% and 6.5% of the variance, respectively (Fig. 2). No distinct clusters of ecotypes were observed in the plot of the first two principal components. Moreover, no correspondence was observed among the relative positions of the 28 ecotypes on the principal component plots and their geographical origin. Of the 378 pairs of ecotypes, the 2 most distantly related are Ct-1 and No-0, which have a genetic similarity of 0.32. Ni (Niederzenz) is an outlier in the principal component plot and has a mean genetic similarity value of 0.55 ± 13 with the other 27 ecotypes. Genetic similarity contrasts with Niederzenz rank among the most distantly related comparisons, including Columbia (0.41), No-0 (0.38), Landsberg (0.36), and La-0 (0.36). Two selections from Landsberg, La-0 and Columbia, are grouped in close proximity relative to the other ecotypes. The genetic similarities between La-0 and both Landsberg and Columbia are 0.91 and 0.62, respectively. Rld is a selection from Rsch-4; the genetic similarity between these two ecotypes is 0.96.

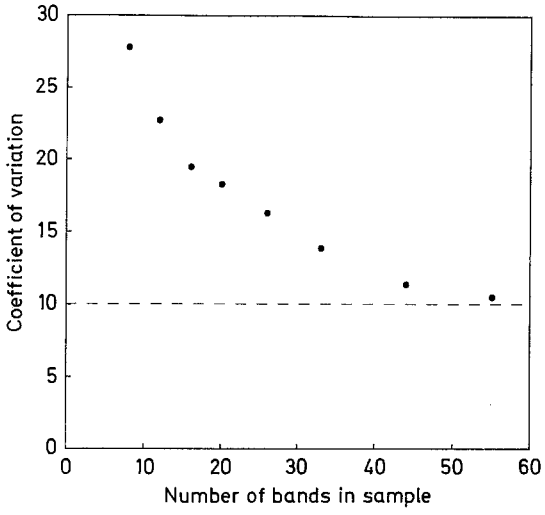

Fig. 2. Relationship between the mean coefficient of variation for genetic similarity among 28 *Arabidopsis* ecotypes and the number of RFLP bands scored. Each data point represents the mean of 200 bootstrap samples

The relationship between the coefficient of variance (CV) for genetic similarity and the number of bands sampled is illustrated in Fig. 3. Inspection of the plot indicates that CVs as low as 10% for estimating genetic similarity among the genotypes in this study could be achieved by sampling as few as 55 bands. Nevertheless, because of missing data, genetic similarity between some ecotypes was estimated with less precision than for others. Ecotypes Ct-1, Sv-0, Sf-2, Ita-0, and Rld had only 25, 16, 44, 37, and 28 scored bands, respectively. Thus, comparisons involving these ecotypes should be interpreted with caution, as the CVs for genetic similarity involving these ecotypes will range from approximately 12% to over 20% (Fig. 3).

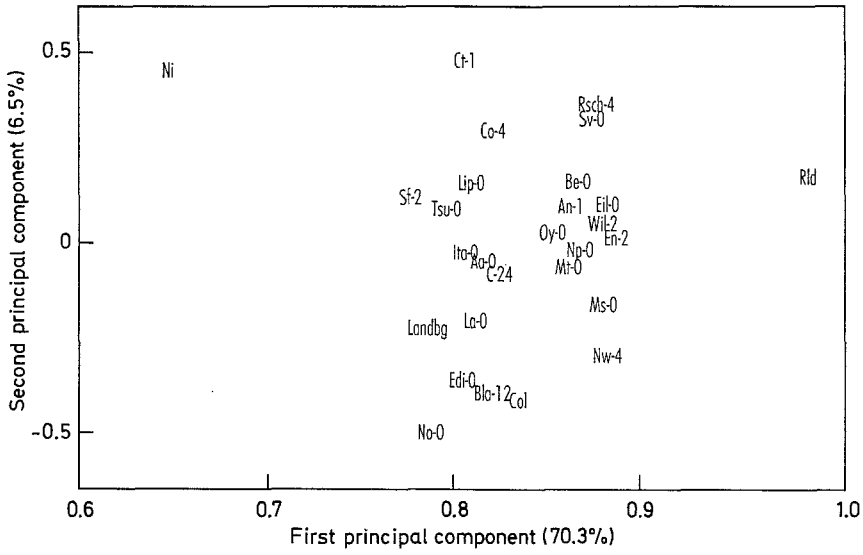

Fig. 3. Plot of first and second principal components of genetic similarity among 28 *Arabidopsis* ecotypes. The numbers in parentheses represent the percentage of variation accounted for by each principal component

Discussion

RFLPs are due to insertions/deletions or base pair substitutions at the nucleotide level. Therefore, comparison of RFLP patterns can provide an estimation of genetic relationships between ecotypes. Studies of maize (Melchinger et al. 1990; Smith et al. 1990), tomato (Miller and Tanksley 1990), and *Brassica oleracea* (Nienhuis et al. 1993) have demonstrated that relationships measured by RFLPs are highly correlated with known genetic relationships or pedigrees.

Little is known regarding the genetic relationships among *Arabidopsis* ecotypes or how artificial selection within heterogenious populations might have resulted in different ecotypes. Thus, little correlative genetic or pedigree information is available. Nevertheless, it is known that Columbia was selected from a phenotypically variable Landsberg population (G. P. Redei, personal communication). The genetic similarity between Landsberg and Columbia is 0.62. Thus, the genetic similarity between a selection from a population and the population per se is comparable to the mean genetic similarity value among the 28 ecotypes in this sample. Moreover, a genetic similarity of 0.62 represents differences at 23 of the 62 bands scored; therefore, the development of Columbia through selection within Landsberg apparently resulted in changes at numerous loci. Two other ecotypes for which some pedigree information is available are Rsch-4 and Rld, (the latter is a selection from Rsch-4); the genetic similarity between these two ecotypes is 0.96. Niederzenz is the most distantly related relative to the other ecotypes included in this study; therefore, populations resulting from crosses with Niederzenz would be expected to have high levels of RFLP polymorphism. This agrees with the high levels of polymorphism observed in the Columbia × Niederzenz population used by Chang et al. (1988) to map RFLP probes.

References

Chang C, Meyerowitz E (1986) Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. Proc Natl Acad Sci USA 83:1408–1412

Chang C, Bowman J, DeJohn A, Lander E, Meyerowitz E (1988) Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. Proc Natl Acad Sci USA 85: 6856–6860

Doebley J, Stec A, Wendel J, Edwards M (1990) Genetic and morphological analysis of a maize-teosinte F_2 population: Implications for the origin of maize. Proc Natl Acad Sci USA 87:9888–9892

Efron B, Tibshirani R (1986) Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat Sci 1:54–77

Langridge J, Griffing B (1959) A study of high temperatures lesions in *Arabidopsis thaliana*. Aust J Biol Sci 12:117–135

Melchinger AE, Lee M, Lampkey KR, Hallauer AR, Woodman WL (1990) Genetic diversity for restriction fragment length polymorphisms and heterosis for two diallel sets of maize inbreds. Theor Appl Genet 80:488–496

Miller J, Tanksley S (1990) RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor Appl Genet 80:437–448

Nam HG, Giraudat J, den Boer B, Moonan F, Loos W, Hauge B, Goodman H (1989) Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana*. Plant Cell 1:699–705

Nei M, Li W (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76:5256–5273

Nienhuis J, Slocum MK, DeVos DA, Muren R (1993) Genetic similarity among *Brassica oleracea* genotypes as measured by restriction fragment length polymorphisms J Am Soc Hortic Sci (in press)

Saghai-Maroof M, Soliman K, Jorgensen R, Allard R (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci USA 81:8014–8018

Sambrook J, Fritsh EF, Maniatis T (1989) Molecular cloning 2nd edn. Cold Spring Harbor Press, Cold Spring Harbor, NY, pp 2.64–2.65

Smith JSC (1984) Genetic variability within US hybrid maize: multivariate analysis of isozyme data. Crop Sci 24:1041–1046

Smith OS, Smith JSC, Bowen SL, Tenborg RA, Wall SJ (1990) Similarities among a group of elite maize inbreds as measured by pedigree, F_1 grain yield, heterosis and RFLPs. Theor Appl Genet 80:833–840

Valvekens D, Van Montagu M, Van Lijsebettens M (1988) *Agrobacterium tumefaciens*-mediated transformation of *Arabidopsis thaliana* root explants by using kanamycin selection. Proc Natl Acad Sci USA 85:5536–5540

White BA, Rosenzweig S (1989) A reliable method for purification of bacteriophage L DNA. Biotechniques 7:694–695