Molecular Genetic Markers and Their Legal Relevance for Cultivar Discrimination: A Case Study in Cucumber

J.E. Staub¹ and V. Meglic²

Additional index words. cucumber, proprietary rights, isozymes, restriction fragment length polymorphisms, random amplified polymorphic DNA, genetic distance, genetic diversity

Summary. Scientific disagreement about criteria for accurate classification of similar, if not seemingly identical, cultivars has led to spirited debate in legal and agricultural communities. The lack of universally acceptable working definitions of functional genetic distance and difference, as well as insufficient data on genetic diversity, has made it difficult to define a legal framework for cultivar discrimination. In order to satisfy the “distinctness” criterion during plant patenting, genetic diversity and difference must be described unequivocally in measurable terms. Moreover, the number of markers or other characteristics needed to identify the “nonobvious” nature of the cultigen will determine the breadth of protection under the patent. Increasingly, patent examiners must interpret novelty and distinctness in terms of molecular as well as gross phenotypic (flower color, plant habit, etc.) information. A description of difference using molecular markers may be more difficult compared to a description of function (i.e., how many markers are required to assign difference). Consequently, the effective use of molecular marker information in the legal community will require scientific agreement on the meaning of genetic distance as it relates to genetic difference.

In the past decade, there has been increasing interest and debate over ownership of intellectual property—patents, copyrights, trade secrets, and plant proprietary rights in both the private and public sectors (Williams and Weber, 1989). To the horticulturist involved in the technological development and protection of plant germplasm, patent issues have become increasingly important and complex. Interest in plant proprietary rights issues has arisen because the protection of research products is necessary to provide incentive for continued investments by private and, in some cases, public institutions. Moreover, recent advances in molecular biology (e.g., plant transformation) and the complex ramifications it holds for gene manipulation [i.e., potential specific

insertion of a foreign gene(s)] has caused increased scrutiny by investors. The purpose of this narrative is to examine the utility of molecular markers for cultivar discrimination in light of current patent law.

A patentable entity includes “any new and useful process, machine, manufacture or composition of matter, or any new and useful improvement thereof.” 35 U.S.C. 161. At present, there are three main avenues of plant protection: 1) certificates of Plant Variety Protection (PVP); 2) Utility Patents, and; 3) Plant Patents. PVP certificates are provided through the USDA Plant Variety Protection Office, and Utility and Plant Patents are administered by the Patent and Trademark Office. The United States Plant Variety Protection Act (PVPA) was created to protect “new” or “novel” varieties of sexually propagated germplasm, other than fungi, bacteria, or first-generation hybrids, “so as to afford adequate encouragement for research, and for marketing...to yield for the public the benefits of new varieties.” 7 U.S.C. sec.

U.S. Department of Agriculture, Agricultural Research Service, Department of Horticulture, University of Wisconsin, Madison, WI 53706

¹Research Horticulturist and Professor.

²Graduate student assistant.

Because phenotypes are subject to environmental modification, molecular markers such as isozymes, restriction fragment length polymorphisms (RFLPs), and random amplified polymorphic DNAs (RAPD) will likely play an increasingly important role in augmenting and enhancing the genotypic description of cultivars.

2402(a), 2581. A protectable "novel variety" must be: 1) "distinct," 2) "uniform," and; 3) "stable." Utility Patents are available for any "invention" (including plants) if it is "useful," "novel," and "nonobvious." 35 U.S.C. sec. 100 *et seq.* The requirement that an invention be "novel" also requires that it be "distinct." As with PVP, the law does not specify what characteristics are to be used for the determination of "distinctness" or the quantitative difference necessary to satisfy the "novel" characteristic. Cultivars that are asexually propagated (grafting, budding, cutting, layering, cloning) can be protected under the amended Plant Patent Act (PPA) of 1930. 35 U.S.C. Sec. 161-164. An applicant must "clearly and precisely" describe characteristics that define the new variety. 35 U.S.C. 162.

Cultivars are identified by measurable traits that are controlled by genes. They are biologically distinct because of the action of genes. Gene action that controls trait expression can be described as either monogenic or polygenic. Monogenic, or qualitative inheritance, is the result of the expression of alleles at one locus (gene). For example, in the case of complete dominance, if the measurements taken on a trait controlled by one locus are sufficiently accurate, both classes [dominant (AA, Aa) and recessive (aa)] governed by alleles at that locus are distinct [e.g., resistance (*bw*) or susceptibility (*Bw*) to bacterial wilt in cucumber (Iezzoni and Peterson, 1980)]. In contrast, polygenic, or quantitative inheritance, reflects the segregation of many genes such that a trait (e.g., fruit yield and quality in cucumber) forms a continuous phenotypic distribution where distinct classes are difficult, if not impossible, to identify.

Both types of genetic variation (monogenic and polygenic) have been used to establish the "distinctness" criterion for cultivar identification. Nevertheless, the use of phenotypic traits for measurement of genetic diversity and the estimation of genetic distance suffers from several shortcomings. Many economically important phenotypes are not simple expressions of genotype. Inter- and intra-locus interactions (epistasis), as well as genotype by environmental interactions, can make an accurate assessment of genetic variation between individuals and populations difficult. As mentioned above, many phenotypic traits are not discrete, but are under poly-

genic control. This imparts not only genetic variation to the trait's expression, but also variation due to measurement and statistical errors. Thus, traits under polygenic control may not be an adequate descriptor of difference.

Because phenotypes are subject to environmental modification, molecular markers such as isozymes, restriction fragment length polymorphisms (RFLPs), and random amplified polymorphic DNAs (RAPD) will likely play an increasingly important role in augmenting and enhancing the genotypic description of cultivars. Molecular markers are useful estimators of genetic distance because they have a predictable genetic basis, distinguish specific segments of DNA or their products, and are relatively insensitive to environmental fluctuations.

Molecular marker techniques

All three forms of protection require some measure of distinctness. The PVP and Plant Patent specify that distinctiveness is a necessary characteristic for protection. The characteristic of distinctiveness also is implied in the Utility Patent because novelty is a prerequisite. Although isozyme, RFLP, and RAPD analysis can provide information on varietal distinctiveness, the procedures required and type of genetic information obtained are different.

Isozymes. In isozyme analysis, histochemical staining techniques for detecting specific enzymes are linked with electrophoretic procedures (usually starch gel) to separate differentially charged protein molecules. Because enzymes catalyze specific biochemical reactions, it is possible to visualize the location of a particular enzyme on a gel by: 1) supplying the appropriate substrate and cofactors, and 2) involving a product of the enzymatic reaction in a color-producing reaction. The colored product becomes deposited on the gel, forming a visible band where a particular enzyme has been localized electrophoretically.

These patterns have a genetic basis because protein products are encoded by alleles (Tanksley, 1983). The gene products of a locus combine at random to produce the active form of the enzyme. For example, a diploid individual heterozygous for a dimeric enzyme encoded by a single locus with two alleles (*A* and *a* or 1 and 2) produces a three-banded phenotype (denoted as *Aa* or 12) on the gel (Fig. 1). This heterozygous genotype can be contrasted to the single-banded phenotypes produced by the two possible homozygous genotypes *aa* (11) and *AA* (22).

RFLPs. The ultimate source of genetic variation is at the level of DNA sequences.

Fig. 1. Isozyme variants produced by dimeric- and monomeric-encoded proteins in diploid organisms.

Polymorphisms in DNA sequences can be detected by the use of restriction enzymes, which cut DNA molecules at sites of particular nucleotide sequences (restriction sites), resulting in fragments. Like enzymes, DNA fragments can be separated by electrophoresis. Identification of genomic DNA fragments requires special procedures such as the Southern blot (Southern, 1975). In this procedure, genomic DNA restriction fragments that have been separated by electrophoresis are transferred to nitrocellulose or a nylon filter. Filter-immobilized DNA is allowed to hybridize to radioactively labeled probe DNA. Probes are usually small (500 to 3000 base pairs) cloned nuclear DNA segments. The filter is placed against photographic film, where radioactive disintegrations from the probe result in visible bands. Polymorphisms resulting in the presence or absence of restriction sites can be identified because they will change the lengths of the characteristic restriction fragments. These restriction fragment length polymorphisms are visualized as banding patterns that are similar to those of isozymes, but represent variation at the DNA rather than the protein level.

RAPDS. Many laboratories recently have begun using a technique derived from polymerase chain reaction (PCR) methodology to develop random DNA polymorphisms (Williams et al., 1990). In RAPD technology, short oligomers of usually 9 or 10 nucleotides in length are hybridized to genomic DNA. The primer/target complexes then are used as substrates for DNA polymerase to copy the genomic sequences 3' to the primers. Iteration of this process yields a discrete set of amplified DNA products that represent target sequences bounded by opposite-oriented, primer-annealing sites. This set of DNA products can be visualized by electrophoresis in agarose. Polymorphisms can be identified easily as band losses or band gains relative to a standard pattern. These band differences result from point mutations in the primer-annealing sites or in length differences due to insertions or deletions within the amplified segments. This technique is quite powerful in the detection of DNA polymorphisms because of its speed, sensitivity, and randomness.

Comparisons among marker systems. Each marker system has certain technical/operational advantages and disadvantages that characterize its value as an estimator of genetic variation. Morphological markers are of limited utility because of the paucity of single-locus traits controlling easily and accurately scored phenotypes. Isozyme analysis, while relatively inexpensive, reproducible, and predictive, detects limited variation. Although RFLP and RAPD detect more variation than

is present in isozymes, they have certain shortcomings. RFLP methodology employs radioactive isotopes, requires relatively large amounts of plant material, and is time-consuming and expensive. RAPD methodology requires precision and reproducibility. Small changes in reaction conditions or purity of target DNA can have profound effects on results; especially on the detection of relatively weak bands and bands >3.0 kilobases long.

Although the operational costs of isozyme and RAPD analysis are relatively low when compared to RFLP analysis, results may not accurately reflect actual differences if experimental procedures are not well-defined and controlled. For example, in contrast to enzymes that catalyze specific reactions, class enzymes such as esterase, catalase, and peroxidase are involved with biochemical reactions in a number of pathways. Not only is the genetic basis of these enzymes often difficult to ascribe, isozyme banding patterns are frequently tissue-dependent and susceptible to environmental modification. In the case of RAPDs, comparisons of primer products to discriminate two cultivars may be invalid if the chemical reaction is not uniform.

If appropriately documented, a PCR-based system such as RAPD analysis has considerable potential for use in cultivar discrimination for several reasons. Like isozyme analysis, the molecular phenotype of any plant can be determined quickly and accurately at the seedling stage using small amounts of tissue. Identification can be performed without growing plants to maturity, thus saving time, expense, and space. In contrast to RFLP analysis, PCR allows specific genomic segments to be assayed without radioactive probes. RAPD methodology (Williams et al., 1990) has proven useful in the construction of a genetic map in lettuce (Michelmore et al., 1991). RAPDs have been reported to be linked to *Pseudomonas* resistance (Martin et al., 1991, 1993) and, in lettuce, they have been used as linked-markers to downy mildew resistance (Michelmore et al., 1991; Paran et al., 1991).

A case study in cucumber

Traditionally, the plant protection prerequisites of distinctness or novelty have been met by providing pedigree explanations using qualitative and quantitative descriptors. In cucumber, cultivars have been characterized using such traits as disease resistance reaction, sex expression, plant habit, and commercial acceptability (Peterson et al., 1985, 1986). Isozyme analysis has also proven useful as a tool for line discrimination, for evaluating the level of population heterozy-

If appropriately documented, a PCR-based system such as RAPD analysis has considerable potential for use in cultivar discrimination.

Some morphological descriptors provide for unique characterization, but require additional information to be definitive.

gosity/homozygosity, and for monitoring allelic frequency (Knerr et al., 1989; Knerr and Staub, 1991; Staub et al., 1985, 1987, 1992). The isozyme data presented in this section are drawn from an evaluation of the U.S. germplasm collection (Knerr et al., 1989, Knerr and Staub, 1991) and a survey of cultivars (varieties and experimental lines) of U.S. and European seed companies (J.E.S., unpublished data). Variation was recorded at 14 loci in 753 plant introductions and 590 cultivars from 11 seed companies [six U.S. (273 sampled) and five European (317 sampled)]. In each case, 15 individuals were sampled from each coded cultigen or accession.

Accessions in the U.S. collection represent contributions from more than 40 countries. Accessions from Peoples Republic of China (14%) and Turkey (20%) constitute a sizable portion of the collection. Czechoslovakia (4%), India (7%), Iran (8%), Japan (6%) Spain (7%), Russia (7%), Yugoslavia (8%), and other countries have contributed to to a lesser extent. The complete isozyme data set of the U.S. germplasm collection can be found in the USDA Genetic Resources Information Network (GRIN).

Although sampling of commercial cultivars was not completely random, it did represent a broad array of elite U.S. processing, U.S. market, Mediterranean (Beit Alpha type), and Dutch glasshouse plant types. The important components of many, if not all, of this germplasm can be traced to plant introductions present in the U.S. germplasm collection (Peterson, 1975). Therefore, in many ways the commercial cultivars can be considered subsets of this collection and, as such, the observed allelic frequencies have comparative value.

Morphological and biochemical descriptors must be reliable and predictive, and their relationships to other loci should be documented. Regardless of the descriptor, the patent applicant must clearly and precisely define the new cultivar. Likewise, the patent examiner must have a knowledge of the diversity and genetic differences among cultivars that are most closely related to make decisions on patentability. These estimates of genetic diversity are obtained either from previously documented sources (e.g., journal publications, cultivar release announcements, etc.) or from the applicant at the time of patent application.

Morphological descriptors. Cucurbit is illustrative of a crop in which morphological descriptors have proven useful for cultivar identification. Morphological variation is present in cultivars as well as in the U.S. germplasm collection. At present, 105 qualitatively inherited morphological traits can be used as descriptors (Pierce and Wehner, 1990).

However, some descriptors are influenced by the environment and other major genes. For instance, the degree of gynoecey (femaleness) is controlled by the interaction of two loci-*M* and *F* and other major genes. Plants with at least one active allele at the *F* locus and one active *M* allele will be gynoeceous (*F_M_*). These loci are affected by other sex genes that modify their expression [e.g., *gy* (gynoeceous) and *In-F* (female intensifier)] and by daylength, total light intensity, temperature, and soil moisture status (Cantliffe et. al., 1981; Kubicki, 1969). Thus, a description of sex expression has significance only when accompanied by properly documented pedigree information and an account of environmental conditions. Further, other genes such as the determinate gene (*de*) also interact with the *F* locus to enhance pistillate expression. Hybrids made with determinate inbreds have more stable gynoeceous sex expression at high plant densities than indeterminate genotypes (Lower et al., 1983). Because factors such as these influence sex expression so dramatically, the *F* locus, although commercially important, has not been a particularly useful quantifier of identity.

Some morphological descriptors provide for unique characterization, but require additional information to be definitive. For example, determinate plants are homozygous recessive *dede*, which results in the premature termination of plant growth as a consequence of the conversion of the sympodial bud into floral tissue (Hutchins, 1940). However, the main stem of determinate plants may vary in vine length (ranging between 30 to 120 cm). Thus, precise descriptions of plant habit for determinate plant types requires data on main stem length taken in several environments. Moreover, it is sometimes difficult to distinguish between the bush (*bu*), dwarf (*dw*), and compact (*cp*) plant types (all indeterminate types with shortened or reduced internodes) without genetic analysis. Thus, in the case of cucumber, genetic conformation of plant habit should also be available at the time of PVP or patent application.

Isozyme variation. When isozyme fingerprinting is used for PVP or patent protection, the examiner will be faced with determining, based on public and private (presented by the applicant) information, whether the new variety can be distinguished from others in use. Cumulative isozymic frequencies across all three cucumber populations are presented in Table 1. Thirty alleles were examined using 14 polymorphic loci where the most common allele was ≥ 0.45 . No cultigen examined possessed a banding pattern that uniquely separated it from all other cultivars.

Allele frequency differed between and among the three populations examined. There were instances where alleles were frequent in the U.S. collection, but infrequent in commercial cultivars. For example, the frequency of the Gpi-1 (1) and Pep-pap 2 allele was high (96% to 99%) in the collection, but lower in the commercial cultivars (2% to 39%). Heterozygotes were infrequent (0% to 0.5%) for these enzyme systems, resulting in a high frequency of homozygotes (11 or 22) in commercial cultivars (data not presented). Some frequency changes were not as dramatic, but do indicate effects of selection on allele (allozyme) frequency (Ak-1, Ak-3, Mpi-2, and Pgm).

In some instances, the frequency of heterozygotes in commercial cultivars increased (4% to 11%) when compared to the U.S. collection (Ak-1, Mdh-2, and Per), but in the main they have decreased (3% to 52%) (Ak-3, Gr-1, Mpi-2, Mdh-3, Pep-gl, Pep-pap, Pgm, and Pgd-2) or remained relatively unchanged (Fdp-3, Gpi-1, and Gr-2) (data not presented). An increase in the frequency of polymorphisms in commercial cucumber may indicate that these isozymes are associated with some economically desirable attributes (Allard, 1992). Likewise, isozymes for which the level of polymorphism was decreased might indicate instances where an isozyme is associated with undesirable attributes. There were a large number of monomorphic loci (20-25) in all cultivars examined. These loci are said to be fixed, and may be important for the general adaptiveness (fitness) of cucumber (Allard, 1992). Because of their ubiquitous nature, monomorphic loci do not add to the description of a cultivar.

There were instances where striking differences in allele frequencies existed among varieties from different seed companies. For example, all cultigens of two U.S. companies (Asgrow and Rogers NK) were Gpi-1 (22), while cultigens of other companies were entirely Gpi-1 (11) (data not presented). A similar situation existed for Pep-pap, where, in contrast to other companies, cultigens of De Ruiter and Royal Sluis were largely Pep-pap (22). The magnitude of the observed differences indicates that isozymic variation may be potentially useful in plant patenting.

Knowledge of genetic diversity, such as that described above, is essential for describing the distinctiveness of a cucumber variety during PVP/patent application and examination. Without information on allele frequency, it is difficult to substantiate the unique character of an isozyme genotype. Given the allelic frequency of these isozyme loci, the probability of occurrence of various isozyme combinations can be calculated (Table 2). If information on allelic frequency is known

Table 1. *Allelic frequency at 14 enzyme coding loci in cucumber.*

Enzyme	Locus ¹	Allele	U.S. cultigens ²	European cultigens	U.S. germplasm collection
Adenylate kinase	Ak-1	1	47	60	87
		2	53	40	13
Adenylate kinase	Ak-3	1	35	22	59
		2	65	78	41
Fructose diphosphatase	Fdp-3	1	47	33	57
		2	53	67	43
Glucosephosphate isomerase	Gpi-1	1	67	92	4
		2	33	8	96
Glutathione reductase	Gr-1	1	0	3	8
		2	100	97	92
Glutathione reductase	Gr-2	1	1	1	11
		2	99	99	89
Malate dehydrogenase	Mdh-2	1	36	5	1
		2	64	95	99
Malate dehydrogenase	Mdh-3	1	87	82	99
		2	13	18	1
Mannosephosphate isomerase	Mpi-2	1	54	49	52
		2	46	51	45
		3	0	0	2.9
		4	0	0	0.1
Peptidase/glycyl-leucine	Pcp-gl	1	93	68	64
		2	7	32	36
Peptidase/phenyl-alanyl-proline	Pep-pap	1	98	61	1
		2	2	39	99
Peroxidase	Per	1	79	97	99.9
		2	21	3	0.1
Phosphoglucumutase	Pgm	1	26	58	29
		2	74	42	71
Phosphogluconate dehydrogenase	Pgd-2	1	38	34	40
		2	62	66	60

²U.S. = 273 cultigens from six seed companies, Europe = 317 cultigens from five Dutch seed companies, and U.S. germplasm collection = 753 plant introductions.

¹Loci coding for enzymes designated by the first letter being uppercase and the rest lowercase. If an enzyme is coded by multiple loci, these are designated by hyphenated numerals and numbered from the most cathodal to most anodal.

and kept current, then it might be relevant information for use by the legal and horticultural community. For example, the probability of two cultigens being homozygous 11 for Per and Pgm can be calculated, and a range of probabilities can be given based on current data. The probability ranges between 0.19 to 0.51 for these unlinked loci. Some probabilities may be low enough to assist in legal discussions [e.g., range for Per and Pgd-2 (12) = 0.0022 to 0.0002]. Naturally, as the number of combinations increases, the probability decreases [e.g., Per, Pgm, and Pgd-2 (12) = 0.0001]. Such information adds a quantitative measure to a descriptor (i.e., rarity of isozymic patterns) and can be used to augment other types of descriptive data that are being proffered for patent examination. It must be emphasized that probabilities may be useful only in the absence of other information. If a party can document, say through

Table 2. *Probability of occurrence of isozyme combinations in cucumber using three enzymes each conditioned by one locus with two alleles.*

Enzyme locus combinations	Genotype	U.S. cultigens ^a	European cultigens	U.S. germplasm collection
Per/Pgm	11/11	0.1851	0.5103	0.2993
	12/12	0.0018	0.0025	0.0003
	22/22	0.1269	0.0070	0.0004
Per/Pgd-2	11/11	0.2776	0.2599	0.2893
	12/12	0.0022	0.0007	0.0002
	22/22	0.1034	0.0113	0.0004
Pgm/Pgd-2	11/11	0.0828	0.1458	0.0870
	12/12	0.0033	0.0023	0.0352
	22/22	0.4165	0.2259	0.2640
Per/Pgm/Pgd-2	11/11/11	0.0652	0.1391	0.0868
	12/12/12	0.0001	0.0001	0.0001
	22/22/22	0.0739	0.0042	0.0002

^aU.S. = 273 cultigens from six seed companies, Europe = 317 cultigens from five Dutch seed companies, and U.S. germplasm collection = 753 plant introductions.

pedigree information, that a certain isozyme phenotype is possible, then discussion of probability becomes moot.

If the samples examined represent a random assortment of inbred lines and hybrids within companies and the identified differences are a consistent reflection of genetic diversity, then genetic differences can be established among companies using isozyme analysis. Because the observed relationships among companies may change as new data become available, only polymorphic loci were used in the analysis (Nei and Li, 1979; Table 3). These relative genetic differences between companies (not genetic distance) can be depicted visually using monotonic multidimensional scaling techniques (Fig. 2).

It is clear that genetic differences exist among companies based on isozyme variation (Fig. 2). This variation could be used to create lines possessing unique isozyme banding patterns, which would allow for more-complete characterization during patent examination. Because the life cycle of cucumber is relatively short (3 to 4 months), lines with

simple, but unique, isozymic combinations could be constructed through backcrossing in a short period of time ($BC_6 = 2.5$ to 5 years) if the proper donor parent (i.e., possessing contrasting unique alleles) was available. Donor lines that differ in allelic constitution at all possible loci are not currently available in an array of commercially acceptable genetic backgrounds. We are constructing donor BC lines differing at 21 isozyme loci using prominent U.S. processing, U.S. market, and Dutch glasshouse lines. This requires incorporation of alleles, present in exotic germplasm, that do not exist in the commercial cultivars we have examined.

Mapping. In order to effectively interpret the legal relevance of markers, knowledge of their linkage relationships is critical. When two or more genes reside in the same chromosome, they are said to be linked. The strength of this linkage relationship is an important determinant of their continued association during the reproductive process. In meiosis, each chromosome replicates, forming two identical sister chromatids. Homologous chromosomes pair (synapse) and crossing over occurs at some physical point (chiasma) between non-sister chromatids. Crossing over involves breakage and reunion of only two of the four strands on the chromosome. When crossing over occurs, gametes assort such that allelic arrangements in the progeny consist of a combination of original parental alleles. This recombination process may result in the disassociation of linked genes. Such dissociations change the physical relationship between genes on the chromosome. If the description of a cultivar is partially defined by relationships of genes on a chromosome (e.g., isozymic banding patterns and probabilities of association), then linkage information is important in the plant patenting process.

The physical distance between genes determines the strength of their linkage. The closer they are to each other on the chromo-

Table 3. *Estimates of genetic differences among seed companies using isozyme variation at 14 cucumber loci.*

Company	Asgrow	Campbells	Ferry Morse	Harris Moran	Petoseed	Rogers NK	De Ruiter Z.	Enza	Nickerson	Nunhems	Royal Sluis
Asgrow	0.0 ^a	0.259	0.241	0.254	0.298	0.152	0.281	0.285	0.275	0.250	0.283
Campbells		0.0	0.165	0.160	0.198	0.255	0.235	0.190	0.167	0.179	0.213
Ferry Morse			0.0	0.159	0.195	0.250	0.213	0.178	0.167	0.175	0.204
Harris Moran				0.0	0.200	0.249	0.223	0.187	0.166	0.173	0.205
Petoseed					0.0	0.272	0.246	0.204	0.192	0.202	0.229
Rogers NK						0.0	0.274	0.268	0.252	0.242	0.274
De Ruiter Zonen							0.0	0.180	0.208	0.205	0.181
Enza								0.0	0.171	0.183	0.191
Nickerson									0.0	0.171	0.203
Nunhems										0.0	0.213
Royal Sluis											0.0

^aEstimates obtained using analysis of Nei on isozyme variation at 14 loci in 590 cultigens.

some, the tighter they are linked. Unit distance along the chromosome is expressed as the probability that crossing over will occur between two genes. One map distance or unit (centiMorgan or cM) is therefore equivalent to 1% crossing over. If two genes are so far apart in the chromosome that the probability of a chiasma forming between them is 100%, then 50% of the gametes will be parental type (non-crossover) and 50% recombinant (crossover) type. Recombination between two linked genes cannot exceed 50%, even when multiple crossovers occur between them. While this unit does not provide a physical distance along a chromosome, it does serve as a relative distance, and thereby makes it possible to determine the order of loci on a chromosome (i.e., genetic mapping).

Because genetic map information allows for genome characterization, it becomes useful in PVP and patent protection. If a unique trait is identified and its position on the chromosome is further characterized by its relationship to tightly linked marker loci, then the description of the trait and, consequently, the cultivar will be enhanced. Likewise, disputes involving alleged infringement of a gene can be resolved more effectively if the position of the gene and its relationship (linkage) to other genes is known.

Marker linkages are determined from segregation data by using a simple assumption: The closer together two genes are to one another, the less likely a recombination event will occur to change the inter-locus allele relationship. As the linkage relationships among genes become known, linkage maps can be constructed. Such maps are defined by the density or saturation of markers used to explain them. For the purpose of this discussion, two broad types of maps will be considered; low-density maps (markers between 5 to 20 cM) and high-density maps (markers between 0.5 to 2 cM). In order to construct highly saturated maps, genome mapping using molecular markers will be needed to augment maps previously constructed with morphological markers (linkage distances often 10 cM or more). Low-density maps presently are being constructed in a number of crop species using isozyme, RFLP, and RAPD markers (Table 4). It should be noted that linkage distances in Table 4 are given in terms of centiMorgans. It should not be concluded that the genome sizes of onion and pepper are similar. Actually, the physical size (in base pairs) of onion is about five times larger than pepper, which is, in turn, eight times larger than cucumber. Thus, physical and recombinant sizes are not correlated from one species to another.

The number of markers needed for a low-density genetic map of cucumber can be

Fig. 2. Genetic difference among European and U.S. seed companies as estimated by 14 cucumber loci and depicted using monotonic multidimensional scaling.

estimated. Cucumber is a diploid with seven chromosomes and 1.0 pg of DNA per haploid nucleus (Bennet and Smith, 1976), in which single-copy DNA accounts for 37% of the genome (Bhave et al., 1984). Using the well-tested hypothesis that one chiasma corresponds to 50 map units of genome (Swanson et al., 1981), and the observation of 15 chiasmata in a typical meiocyte (Fanourakis, 1984), it can be concluded that cucumber has a total genome length of ≈ 7.5 Morgans. Calculating from values in Beckman and Solter (1983), estimates of numbers of markers for adequate saturation of the cucumber genome are given in Table 4. With 100 markers averaging 10 cM apart, $\approx 90\%$ of the cucumber genome will be covered. In this way, the number of markers needed for genome coverage in crops can be compared.

The incidence of linkage in cucumber might be predicted to be high, given cucumber's low basic chromo-

Table 4. Genetic markers needed for genome coverage in several crops.

Crop/spacing	Proportion of genome coverage(%) ^a			
	70	80	90	95
Bean				
10 cM	72	95	137	178
20 cM	36	47	68	88
Carrot				
10 cM	64	87	123	160
20 cM	34	46	84	86
Cucumber				
10 cM	52	72	100	130
20 cM	29	38	54	71
Onion				
10 cM	67	76	86	90
20 cM	33	38	43	45
Potato				
10 cM	49	56	60	67
20 cM	25	28	30	34
Pepper				
10 cM	70	80	90	95
20 cM	35	40	45	48
Tomato				
10 cM	91	104	117	124
20 cM	45	52	59	62

^aInformation obtained from C.E. Vallejos (bean), Univ. of Florida; P. Simon (carrot) and M. Havey (onion), USDA, Univ. of Wisconsin; and M. Bonierbale (potato and tomato) and M. Kyle (pepper) Cornell Univ.

When considering quantitative traits, one way to define "difference" is to associate molecular markers with economic value.

some number, small chromosome size (Ramachandran et al., 1983), and low cross-over frequency (Fanourakis, 1984). However, very little information exists regarding linkages among genes in cucumber (Pierce and Wehner, 1990). Given the 105 described genes, linkage information is available for only 179 (3.2%) of the 5460 possible pairwise combinations. In a linkage study using morphological markers, Fanourakis and Simon (1987) placed 13 of 15 loci into three linkage groups. Joint segregation analyses among 14 genetically predictable polymorphic isozyme loci resulted in the assignment of 12 loci to four linkage groups (Knerr and Staub, 1992).

The placement of these and other descriptors on a genetic map has been time-consuming and expensive. To increase the likelihood of obtaining a low-density genetic map, we, along with other colleagues in our research unit, have been using RFLP and RAPD technology. Depending on the germplasm used, the frequency of polymorphic clones (using genomic and cDNA libraries) used in RFLP analysis is $\approx 15\%$, and polymorphisms detected by RAPD primers ranges between 10% to 15%. To date, we have recovered 33 RFLPs and 66 RAPDs that have been mapped using commercial cucumber. Low-density genetic maps are currently being constructed using isozyme, RFLP, and RAPD markers. These maps must be merged to provide the information need for effective use in plant variety protection. More markers will undoubtedly be needed to develop a highly saturated map in this species.

Genetic distance. In the past, examiners at the PVP Office have relied on standards such as color charts for detection of qualitative difference and statistical significance established over time and/or locations to document quantitative difference. Molecular tech-

niques are attractive because they detect smaller genetic differences. In turn, molecular differences can be used to estimate genetic distance (Nei and Li, 1979); however, such estimations cannot be used to conclude that two cultivars are identical—they can only define differences that exist. In cases of alleged infringement of PVP or patents, the defendant has only to show that "sufficient" differences exist to prove "distinctness."

In the case of cucumber, isozymic data are useful in determining to what extent cultivars differ. For instance, consider the genetic distances (all monomorphic loci included) between a turn-of-the-century variety Windemoor Wonder (Seedsavers Exchange); a USDA inbred line, WI 2757, derived from a diverse number of exotic plant introductions; a relatively new U.S. market type, 'Dasher II' (Petoseed); a processing type, 'Transamerica' (Ferry Morse); and the West Indian Gerkin (*Cucumis anguria* L.) as measured by 14 enzyme coding loci (Fig. 3). Although all *C. sativus* cultigens are genetically distant (0.192) from the West Indian Gerkin (*C. anguria* L.), they possess isozymic similarities that define their intraspecific relationships. The genetic distance between 'Dasher II' and 'Windemoor Wonder' is equal to that between it and WI 2757. WI 2757 is genetically more distant from 'Windemoor Wonder' than it is from 'Dasher II'. Because of the relative genetic distances between 'Transamerica', 'Windemoor Wonder', and WI 2757, 'Transamerica' is placed in intermediate position between 'Dasher II' and WI 2757. Both 'Dasher II' and 'Transamerica' have 'Poinsett 76' in their pedigree, which may partially explain their close genetic relationship when compared to 'Windemoor Wonder' and WI 2757, which do not. Such relationships, if they are known, can more clearly define cultivar differences and, therefore, aid in plant patenting.

When considering quantitative traits, one way to define "difference" is to associate molecular markers with economic value. Molecular markers that affect the expression of quantitatively inherited traits have been shown to be located throughout the genome (Godshalk et al., 1990). In theory, there should be a point at which the cumulative effect of molecular markers in a mapped genome should account for a significant amount of the observed variation for a trait. If a highly saturated map exists for a crop, and extensive, well-documented information on markers and their associated economic performance are also available at a specific location, then differences in performance between cultivars might be attributed to a specific number of loci. This difference might be called "functional" genetic difference.

Fig. 3. Relationships between four cucumber cultivars using Nei's genetic distance measure based on 52 isozyme coding loci (14 polymorphic).

Functional genetic difference may be difficult to ascribe. This is especially true of quantitatively inherited traits, such as fruit number in cucumber, where genetic variance has been explained by additive and additive $\times$ additive epistatic effects. Moreover, since we know little about the linkage among quantitative traits in cucumber, it is difficult to estimate accurately the sample size necessary to detect quantitative trait loci. In order for the relationship between economic value and markers to become useful, field and laboratory experiments must be conducted to confirm values in diverse environments. Documented, statistically oriented computer algorithms coupled with crop-specific databases may be useful as simulations to explore the relationship between economic value and marker distribution. Nevertheless, it is doubtful that, without extensive documentation, markers associated with quantitative trait loci will be useful in plant variety protection.

The PVP Board and the legal community have not resolved the dilemma of defining clear genetic distance. It is believed that that this issue will need to be addressed on a crop-by-crop basis (Lerch, 1989). Although molecular marker data cannot be used to determine unequivocally that two varieties are identical, properly documented data can be used to measure genetic difference. Moreover, if minimum genetic distance were defined in a crop species, such data could provide quantifiable, comparative information for support of the distinctness criterion. As smaller differences are detected and used in certification or patenting, the degree of protection becomes narrower, and requirements for uniformity increase. It is clear that, in order for the horticultural industry to make informed decisions regarding the potential impact of plant protection laws, the scale, magnitude, and resolution of phenotypic and genotypic measures must be defined more clearly, and those measures interpreted in terms of current and pending patent law.

Literature Cited

- Allard, R. W. 1992. Predictive methods for germplasm identification. In: H.T. Stalke and J.P. Murphy (eds.). Plant breeding in the 1990's. CAB International, Wallingford, Oxon, U.K.
- Beckman, J.S. and M. Soller. 1983. Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. Theor. Applied Genet. 67:35-43.
- Bennett, M.D. and J.B. Smith. 1976. Nuclear DNA amounts in angiosperms. Trans. R. Soc. London, Ser. B. 274:228-274.
- Bhave, M., M. Lagu, and P.K. Ranjekar. 1984. Molecular analysis of Cucurbitaceae genomes: I. Comparison of DNA reservation kinetics in six plant species. Plant Sci. Lett. 33:127-136.
- Cantliffe, D.J. 1981. Alteration of sex expression in cucumber due to changes in temperature, light intensity and photoperiod. J. Amer. Soc. Hort. Sci. 106:133-136.
- Fanourakis, N.E. 1984. Inheritance and linkage studies of the fruit epidermis structure, and investigation of linkage relationships of several traits and of meiosis in cucumber. PhD Diss., Univ. of Wisconsin, Madison.
- Fanourakis, N.E. and P.W. Simon. 1987. Inheritance and linkage studies of the fruit epidermal structure in cucumber. J. Hered. 78:369-371.
- Godshalk, E.B., M. Lee, and K.R. Lamkey. 1990. Relationship of restriction fragment length polymorphisms to single-cross hybrid performance of maize. Theor. Applied Genet. 80:273-280.
- Hutchins, A.E. 1940. Inheritance in the cucumber. J. Agr. Res. 60:117-128.
- Iezzoni, A.F. and C.E. Peterson. 1980. Linkage of bacterial wilt resistance and sex expression in cucumber. HortScience 15:257-258.
- Knerr, L.D., J.E. Staub, D.J. Holder, and B.P. May. 1989. Genetic diversity in *Cucumis sativus* L. assessed by variation at 18 allozyme coding loci. Theor. Applied Genet. 78:119-128.
- Knerr, L.D. and J.E. Staub. 1991. A multivariate re-evaluation of biochemical genetic diversity in *Cucumis sativus* L. Cucurbit Genet. Coop. Rpt. 14:25-28.
- Knerr, L.D. and J.E. Staub. 1992. Inheritance and linkage relationships of isozyme loci in cucumber (*Cucumis sativus* L.). Theor. Applied Genet. 84:217-224.
- Kubicki, B. 1969. Investigations of sex determination in cucumber (*Cucumis sativus* L.): IV. Multiple alleles of locus *Acr*. Genet. Polonica 10:23-68.
- Lerch, V.D. 1989. The plant variety protection act: a current look. Seed World 127:20-24.
- Lower, R.L., O.S. Smith, and A. Ghaderi. 1983. Effects of plant density, arrangement, and genotype on stability of sex expression in cucumber. HortScience 18:737-738.
- Martin, G. B., J.G.K. Williams, and S. D. Tanksley. 1991. Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near-isogenic lines. Proc. Natl. Acad. Sci. USA 88:2336-2340.
- Martin, G.B., M. Carmen de Vicente, and S. Tanksley. 1993. High-resolution linkage analysis and physical characterization of *Pto* bacterial resistance locus in tomato. Mol. Plant-Microbe Interact. 6:26-34.
- Michelmore, R.W., I. Paran, and R. V. Kesseli. 1991. Identification of markers linked to disease resistance genes by bulked segregant analysis: A rapid method to detect markers in specified genomic regions using segregating populations. Proc. Natl. Acad. Sci. USA 88:9828-9832.

It is clear that, in order for the horticultural industry to make informed decisions regarding the potential impact of plant protection laws, the scale, magnitude, and resolution of phenotypic and genotypic measures must be defined more clearly, and those measures interpreted in terms of current and pending patent law.

- Nei, M. and W. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA.* 76: 5256-5273.
- Paran, I., R.V. Kessli, and R.W. Michelmore. 1991. Identification of RFLP and RAPD markers linked to downy mildew resistance genes in lettuce using near-isogenic lines. *Genome* 34: 1021-1027.
- Peterson, C.E. 1975. Plant introductions in the improvement of vegetable cultivars. *HortScience* 10:575-579.
- Peterson, C.E., J.E. Staub, M. Palmer, and L. Crubagha. 1985. Wisconsin 2843, a multiple disease resistant cucumber population. *HortScience* 20:309-310.
- Peterson, C.E., J.E. Staub, P.H. Williams, and M.J. Palmer. 1986. Wisconsin 1983 cucumber. *HortScience* 21:1082-1083.
- Pierce, L.K. and T.C. Wehner. 1990. Review of genes and linkage groups in cucumber. *HortScience* 25:605-615.
- Ramachandran, C., R.A. Pai, and V.S. Seshadri. 1983. A Feulgen technique for identification of cucumber chromosomes. *Experientia* 39: 1155-1156.
- Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503-517.
- Staub, J.E., R.S. Kupper, D. Schuman, T.C. Wehner, and B. May. 1985. Electrophoretic variation and enzyme storage stability in cucumber. *J. Amer. Soc. Hort. Sci.* 110:426-431.
- Staub, J.E., L. Fredrick, and T. Marty. 1987. Electrophoretic variation in cross-compatible wild diploid species of *Cucumis*. *Can. J. Bot.* 65:792-798.
- Staub, J.E., L.D. Knerr, D.J. Holder, and B. May. 1992. Phylogenetic relationships among several African *Cucumis* species. *Can. J. Bot.* 70:509-517.
- Swanson C.P., T. Merz, and W.J. Young. 1981. Cytogenetics: the chromosome in division, inheritance and evolution. Prentice-Hall, Englewood Cliffs, N. J.
- Tanksley, S.D. 1983. Gene mapping, p. 109-138. In: S.D. Tanksley and T.J. Orton (eds.). *Isozymes in plant genetics and breeding, Part A.* Elsevier, Amsterdam.
- Williams, S.B. and K.A. Weber. 1989. Intellectual property protection and plants. *Intellectual Property Rights Associated with Plants.* Amer. Soc. of Agron. Spec. Publ. 52. p. 91-107.
- Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalski, and S. V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Res.* 18: 6531-6535.