J. Amer. Soc. Hort. Sci. 127(4):558–567. 2002.

The Formation of Test Arrays and a Core Collection in Cucumber Using Phenotypic and Molecular Marker Data

Jack E. Staub[1]
U.S. Department of Agriculture, Agricultural Research Service, Vegetable Crops Unit, Department of Horticulture, 1575 Linden Drive, University of Wisconsin, Madison, WI 53706

Fenny Dane
Department of Horticulture, 101 Funchess Hall, Auburn University, AL 36849

Kathleen Reitsma
U.S. Department of Agriculture, North Central Plant Introduction Station, Iowa State University, Ames, IA 50011-1170

Gennaro Fazio and Anabel López-Sesé
U.S. Department of Agriculture, Agricultural Research Service, Vegetable Crops Unit, Department of Horticulture, 1575 Linden Drive, University of Wisconsin, Madison, WI 53706

Additional index words. genetic resources, genetic diversity, isozymes, germplasm, principal component analysis, core subsets, situation specific arrays, disease resistance, abiotic stress

Abstract. **Genetic relationships among 970 cucumber (*Cucumis sativus* L.) plant introductions (PIs) in the U.S. National Plant Germplasm System (NPGS) were assessed by observing variation at 15 isozyme loci. Allozyme frequency data for these PIs were compared to allozyme variation in heirloom and modern (H&M) cultivars released from 1846–1985 (H&M cultivars; 178 accessions), and experimental commercial (EC) germplasm (EC germplasm; 82 accessions) in use after 1985. Multivariate analysis defined four distinct groups of accessions (Groups A–D), where Group A consisted of PIs received by the NPGS before 1992, Group B contained PIs from India and China obtained by NPGS after 1992, Group C consisted of EC germplasm, and Group D contained H&M cultivars. Morphological, abiotic stress (water and heat stress tolerance) and disease resistance evaluation data from the Germplasm Resources Information Network (GRIN) for the PIs examined were used in conjunction with estimates of population variation and genetic distance estimates to construct test arrays and a core collection for cucumber. Disease resistance data included the evaluation of angular leafspot [*Pseudomonas lachrymans* (E.F. Smith) Holland], anthracnose [*Colletotrichum lagenarium* (Ross.) Ellis & Halst], downy mildew [*Pseudoperonospora cubensis* (Berk. & Curt) Rostow], rhizoctonia fruit rot (*Rhizoctonia solani* Kuhn), and target leafspot [*Corynespora cassiicola* (Berk. & Curt) Wei] pathogenicity. The test arrays for resistance–tolerance to angular leafspot, anthracnose, downy mildew, rhizoctonia fruit rot, target leafspot, and water and heat stress consisted of 17, 16, 17, 16, 17, 16, and 16 accessions, respectively. The core collection consisted of accessions in these test arrays (115) and additional 32 accessions that helped circumscribe the genetic diversity of the NPGS collection. The core collection of 147 accessions (115 + 32) represents ≈11% of the total collection's size (1352). Given estimates of genetic diversity and theoretical retention of diversity after sampling, this core collection could increase curatorial effectiveness and the efficiency of end-users as they attempt to identify potentially useful germplasm.**

Genetic variability is essential for crop improvement, and therefore substantial resources have been directed towards collecting, preserving, and evaluating collections to access their genetic diversity (Peeters and Galwey, 1988; Plucknett et al., 1983). National genebanks, such as the U.S. National Plant Germplasm System (NPGS), maintain substantial collections of diverse genotypes of worldwide origin. These crop germplasm collections differ in size, scope, and reproductive biology, and as a consequence require different management strategies for their curation and genetic maintenance.

Resources at genebanks are often not sufficient to adequately conserve, manage, and evaluate germplasm collections of ever-increasing size. The concept of core collections was introduced to improve managerial effectiveness of germplasm curation, and increase the availability of unique genetic resources for plant improvement (Frankel, 1984; Brown, 1989 a, 1989b). A core collection is a representative sample from an entire collection that allows for more efficient evaluation of genetic diversity, and for the identification and resampling of accessions of potential economic value. In theory, a core collection should include the maximum genetic variance that exists in the entire collection with minimum genetic repetitiveness (Johnson and Hodgkin, 1999; van Hintum et al., 2000). A standard range collection (Beuselinck and Steiner, 1992) or test array as used herein because of its specific and defined characteristics is a subset of the entire collection that contains accessions varying in one specific attribute (e.g., a specific disease resistance). A series of test arrays can be extracted from a collection to provide researchers with germplasm sets for specific, often economically important comparisons (e.g., the comparative evaluation of disease resistance in newly acquired germplasm). Test arrays, in turn, can be used to develop core collections (Beuselinck and Steiner, 1992; Staub et al., 1989).

Received for publication 24 Jan. 2002. Accepted for publication 17 Apr. 2002. Mention of a trade name, proprietary product, or specific equipment does not constitute a guarantee or warranty by the USDA and does not imply its approval to the exclusion of other products that may be suitable.

[1]Author to whom all correspondence should be addressed; e-mail jestaub@facstaff.wisc.edu.

It is important for curatorial purposes to obtain estimates of genetic similarity among collection accessions. Such information enhances the management strategies for the entire collection [i.e., what needs to be added or removed (duplicates)] and what should be included in a core collection (i.e., accessions that circumscribe its genetic constitution). Similarity estimates can be obtained through pedigree information and/or by direct genetic analysis of a collection (van Hintum, 1993, 1994). Biochemical (protein and DNA) genetic markers have been used extensively to characterize accessions and define the structure of germplasm collections (Bretting and Widrlechner, 1995). The choice of genetic markers in germplasm management depends on the size of a collection (i.e., resource constraints), the germplasm's reproductive biology, and the amount and kind of genetic variation present.

The cucumber (*Cucumis sativus*) collection of NPGS is relatively small [1352 plant introductions (PI)], consisting of accessions of *Cucumis sativus* L. var. *sativus* (adapted germplasm; 2n = 2x = 14; hereafer referred to as *C. s.* var. *sativus*) and a cross-compatible, feral wild form, *C. sativus* var. *hardwickii* (Royle) Gabaev (hereafer referred to as *C. s.* var. *hardwickii*).). This crop is grown in most temperate regions, and is the fourth most widely grown vegetable crop in the world, after tomato, onion and cabbage (Tatlioglu, 1993). Despite its relatively small size, the NPGS cucumber collection is of considerable importance for future genetic improvements in cultivars given the extensive use of plant introductions (PIs) in crop improvement (Peterson 1975).

The greatest genetic diversity in cucumber resides at its center of origin, India (Leppik, 1966; Staub et al., 1999). After its initial domestication (≈3,000 BC), cucumber was likely disseminated via the Silk Road and Oceanic routes (Meglic et al., 1996; Staub et al., 1999). Genetic diversity in India is unique (Meglic et al. 1996), and recent acquisitions from the Indian states of Rajasthan, Madhya Pradesh, and Uttar Pradesh have enriched the NPGS cucumber collection (Staub et al., 1997a). Genetic variation in the secondary center for diversity, China, has been described (Staub et al., 1999), and Chinese accessions in the NPGS are genetically distant from India and all other cucumber germplasm in the collection (Staub et al., 1997a, 1999). The genetic diversity of the NPGS cucumber collection has also been significantly broadened by collections made during a 1994 U.S.–China government-sponsored expedition to China (Wehner et al., 1996), and by donations of Chinese germplasm made since 1994.

Isozymes, restriction fragment length polymorphisms (RFLPs) and randomly amplified polymorphic DNA (RAPD) markers have been used to assess genetic diversity in cucumber (Dijkhuizen et al., 1996; Knerr et al., 1989; Knerr and Staub, 1991; Meglic et al., 1996; Meglic and Staub, 1996a, 1996b; Staub et al., 1997b). Isozyme analysis of diverse cucumber germplasm by using relatively few polymorphic loci (14 to 21 depending on the study) has allowed for the discrimination of NPGS accessions grouped by geographic proximity (Knerr et al., 1989; Meglic et al., 1996; Staub et al., 1997a) (Table 1). Studies of Dijkhuizen et al. (1996) (RLPFs; U.S. and European commercial germplasm), Staub et al. (1997b) (RAPD markers–isozymes; inbred lines and hybrids), and Horesji and Staub (1999) (RAPD markers; PIs and commercial germplasm) indicated that accession groupings defined by isozyme markers were similar to those produced by the analysis of DNA marker loci, and that these groupings were biologically meaningful based on taxonomy and geographic origin. A comparative analysis of the genetic constitution of accessions in these studies and experimental commercial germplasm in use after 1985 would be useful for defining historical benchmarks. For example, the estimation of genetic distances among germplasm pools (i.e., by cultivar release date or acquisition as PIs) would define genetic diversity as a function of acquisition and use during recent cultivation (>1850).

Table 1. Cucumber (*Cucumis sativus* L.) populations evaluated for isozymic variation.

Population examined	No. of accessions examined	No. of isozyme loci evaluated	Citation	Major conclusions
NPGS[z]	757	18	Knerr et al. (1989)	Detection of genetic affinities between accessions, grouping by geographic proximity and identification of six most discriminatory loci
NPGS	757	14	Knerr and Staub (1991)	Reassessment of Knerr et al. (1989) using mapped loci to redefine geographic groupings
NPGS	757	21	Meglic et al. (1996)	Analysis separated NPGS accessions into groups by country and by continent or subcontinent
Heirloom/modern cultivars[y]	178	21	Meglic and Staub (1996)	Analyses partitioned cultivars into two distinct groups: those released before 1968 and those released after 1968
India collections before and after 1992[x]	146 (757)[w]	21	Staub et al. (1997)	India accessions received after 1992 differed significantly from Indian accessions acquired before 1992 and all other germplasm in the NPGS
China collections before and after 1994[v]	67 (903)[w]	21	Staub et al. (1999)	Chinese accessions received after 1994 were significantly different from Chinese accessions acquired before 1994, all Indian accessions, and other germplasm in the NPGS

[z]NPGS = U.S. National Plant Germplasm System.
[y]Prominent commercial varieties grown in the U.S. between 1846 and 1978.
[x]Collected during a U.S./India sponsored expedition during 1992 (86, 50, and 10 accessions from Rajasthan, Madhya Pradesh, and Uttar Pradesh, respectively).
[w]Number in parentheses represents the number of accessions from which data were taken from that previously used for comparative analyses to determine genetic relationships in the NPGS.
[v]Collected during a U.S./China sponsored expedition during 1994 (39 accessions), and those received as a gift to NPGS from J.F. Chen in 1996 (28).

The Genetic Resources Information Network (GRIN) contains passport information and data from evaluation studies (disease and abiotic stress resistance) of cucumber. These data, along with isozymic variation among accessions in the NPGS cucumber collection can be used to develop test arrays and a core collection for this crop. Therefore, we present herein an assessment of the genetic diversity of selected public and private proprietary commercial lines in breeding programs after 1985 of worldwide origin, and use this information in combination with evaluation data found in GRIN and previous isozymic assessments (Table 1) to form 1) test arrays for various economically important traits and 2) a core collection useful for tactical assessments of cucumber germplasm. The diversity analysis performed herein on experimental commercial germplasm in use after 1985 adds to previous genetic diversity analyses for a more comprehensive understanding of the breath of genetic variation in the NPGS cucumber collection. These data and those of Meglic and Staub (1996b) form a historically important baseline for future comparisons.

Materials and Methods

Plant material. The genetic diversity of cucumber in the NPGS was assessed using plant introduction (PI) allozyme frequency data obtained 1) during previous studies of the NPGS collection (Meglic et al., 1996; Staub et al., 1997a, 1999) (Table 1), 2) during the examination of heirloom and modern cultivars to 1985 (Meglic and Staub, 1996b) (Table 1), and 3) from polymorphisms observed among an array of commercial breeding lines used after 1985 (reported originally herein). Diversity assessment was performed using 18 populations grouped according to previous reports (Tables 1 and 2) with the addition of the population of 82 experimental commercial accessions in use after 1985 (presented herein). In total, 1230 entries were considered for analysis.

NPGS accessions (970) were grouped into 14 population groups (i.e., a population group is defined here as a distinct set of noninterbreeding individuals having the same geographic region), designated as U.S. (no. 1, 24), Eastern Europe (no. 2, 160), China Old (no. 3, 110), India Old (no. 4, 46), Iran (no. 5, 63), Japan (no. 6, 46), Unassociated (no. 7, 81), Turkey (no. 8, 178), *C. s.* var. *hardwickii* (no. 9, 4), Western Europe (no. 10, 45), China New-1 (no. 11, 39), China New-2 (no. 12, 28), India New-1 (no. 13, 37), and India New-2 (no. 14, 109) according to previous analyses (Staub et al., 1997a, 1999) (Table 2) (numbers in parenthesis indicate the analysis group number and the number of accessions examined, respectively). The grouping assigned unassociated consisted of accessions that did not fit any geographical grouping according to Meglic et al. (1996). India and China Old refer to those accessions received by the NPGS before 1992 and 1994, respectively (Staub et al., 1997a, 1999). India and China New refer to those accessions obtained by the NPGS after 1992 and 1994, respectively. The hyphenated number after the designation New indicates subgroupings according to Staub et al. (1997a, 1999).

Commercial cultivars developed from 1846–1937 (no. 15,

Table 2. Genetic variability measures for 12 *Cucumis sativus* L. populations as calculated by variation at 11 isozyme loci.

Population[z]	N[y]	P[x]	A[w]	Ae[v]	Ho (SD)[u]	He (SD)[t]
1) U.S.	24	52.4	1.71	1.29	0.087(0.052)	0.166(0.045)
2) Eastern Europe	160	71.4	1.76	1.27	0.121(0.021)	0.157(0.044)
3) China (Old)	110	66.7	1.71	1.16	0.055(0.021)	0.110(0.033)
4) India (Old)	46	66.7	1.86	1.40	0.108(0.039)	0.200(0.049)
5) Iran	63	57.1	1.62	1.29	0.132(0.034)	0.163(0.046)
6) Japan	46	47.6	1.48	1.15	0.049(0.030)	0.097(0.032)
7) Unassociated	81	76.2	2.00	1.32	0.126(0.031)	0.185(0.046)
8) Turkey	178	76.2	1.81	1.30	0.131(0.021)	0.175(0.045)
9) *C. s.* var. *hardwickii*	4	42.9	1.52	1.31	0.095(0.122)	0.165(0.046)
10) Western Europe	45	47.6	1.52	1.27	0.123(0.040)	0.156(0.043)
11) China (New-1)	39	57.1	1.57	1.25	0.155(0.032)	0.149(0.041)
12) China (New-2)	28	47.6	1.48	1.18	0.099(0.038)	0.106(0.037)
13) India (New-1)	37	81.0	1.81	1.42	0.284(0.211)	0.246(0.207)
14) India (New-2)[s]	109	90.5	2.01	1.59	0.213(0.175)	0.338(0.174)
15) Cultivars (1846 to 1937)[r]	40	46.7	1.47	1.17	0.038(0.028)	0.104(0.045)
16) Cultivars (1937 to 1968)[r]	27	60.0	1.60	1.19	0.040(0.033)	0.118(0.045)
17) Cultivars (1968 to 1985)[r]	111	46.7	1.53	1.22	0.052(0.019)	0.123(0.051)
18) Cultivars (after 1985)[r]	82	33.3	1.40	1.30	0.069(0.024)	0.156(0.059)
Mean		59.3	1.66	1.28	0.110	0.164
SD		15.3	0.18	0.11	0.063	0.066

[z]Populations are as described in Meglic et al., 1996, Staub et al., 1997a, and Staub et al., 1999, where India (Old) and India (New) designates accessions acquired before and after 1993, respectively, and China (Old) and China (New) designates accessions received by the U.S. National Plant Germplasm System before and after 1994, respectively. Unassociated are those accessions that did not group with respect to geographic proximity according to Meglic et al., 1996.
[y]Number of accessions examined.
[x]The percentage of polymorphic loci.
[w]The mean number of alleles per locus.
[v]The effective number of alleles.
[u]The observed heterozygosity, where SD = standard deviation.
[t]The expected proportion of heterozygous loci per individual, where SD = standard deviation.
[s]Accessions in Group 2 (102) and those not associated with any group (7) of Staub et al. (1997).
[r]Date of release.

40), commercial cultivars developed between 1937–68 (no. 16, 27), and commercial cultivars developed between 1968–85 (no. 17, 111) were grouped according to Meglic and Staub (1996a). In 1994, 82 commercial cultivars released after 1985 (no. 18) representing all major market types were obtained from 11 vegetable seed companies having active breeding programs with diverse objectives specific to market type and growing location (Staub, 1999). Data sets and genetic distance matrices are given at the web site location http://www.hort.wisc.edu/usdavcru/staub. Seeds of F1 hybrids were obtained from the following companies: Asgrow Seed Co., U.S. (9), Campbell Soup Co., U.S. (7), Ferry Morse Seeds, U.S. (7), Harris Moran Seeds, U.S. (7), Petoseed Co., U.S. (7), Rogers NK Seeds, U.S. (7), De Ruiter Zonen, The Netherlands (7), Enza Zaden, The Netherlands (7), Nickerson Zwaan, The Netherlands (8), Nunhems Seeds, The Netherlands (9), and Royal Sluis, The Netherlands (7) (number in parenthesis refers to the number of accessions examined). These companies accounted for ≈60% to 80% of the commercial cucumber seed sold in North America and Europe at the time the study was initiated. Although the pedigrees of these cultivars are proprietary, these accessions represent diverse market types and target growing environments {i.e., U.S. and European processing and market, European glasshouse [long (length to diameter ratio >7) and midlength or middy (length to diameter ratio 4.5 to 6.5], and Beit Alpha (Middle Eastern)}.

Isozyme analysis. Isozymic examination of all cucumber accessions was performed according to Knerr and Staub (1992) using a random sample of 15 to 20 plants from each accession for bulk analysis (Meglic et al., 1996). Cotyledonary tissue samples (0.01 g) were bulked for analysis, and gels were stained using the protocols of Allendorf et al. (1977), Brewer (1970), and Shaw and Prasad (1970) to resolve isozyme banding patterns at 15 loci using 12 enzyme systems (Meglic et al., 1996). Those loci included *Gpi-1*, *Gr-1*, *G2dh*, *Mdh-1*, *Mdh-2*, *Mdh-3, Mpi-2*, *Pep-la*, *Pep-pap*, *Pgd-1*, *Pgd-2*, *Pgm*, *Per*, *Idh*, and *Skdh-1*.

Since the *C. s.* var. *sativus* line Gy 14a was used as a control in this and previous studies, alleles at loci were precisely located in isozyme gels by visual observation. Genetic nomenclature for describing allozymic variation followed a modified form (Knerr and Staub, 1992) previously described by Richmond (1972). For loci having two alleles, allelic frequency was assumed to be p = 0.5 and q = 0.5 in bulked samples of each accession. Estimates of allelic frequencies were calculated according to the protocol of Widrlechner et al. (1992), and allelic frequency data matrices from previous reports pertinent to this analysis (Table 1) are available through GRIN.

Diversity analysis. Allelic variation within 18 population groups (14 NPGS and 4 cultivar groups) was used to calculate the percentage of polymorphic loci (P), the mean number of alleles per locus (A), the effective number of alleles (A_e), the observed heterozygosity (H_o), and the expected proportion of heterozygous loci per individual (H_e) in each of the populations examined following Berg and Hamrick (1997). These estimates allowed for the quantification of population genetic diversity for comparative analysis (Hamrick et al., 1992).

In previous evaluations of cucumber PIs for allozymic variation within bulk samples (Staub et al., 1997a) and in random sampling of 15 individuals within an accession in our study (data not presented), allele frequency adequately fit Hardy-Weinberg expectations for loci possessing two alleles. Such observations have also been made in accessions of other *Cucumis* species where collections have been maintained under controlled pollination (López-Sesé et al., 2002). Thus, the 11 diallelic loci used herein for heterozygosity (H_o) estimation were deemed adequate for hypothesis testing. Loci possessing multiple alleles (>2; *Gpi*, *Gr*, *Mpi-2*, *Pep-la*) were not used.

Principal component analyses (PCA) were used to depict affinities among the accessions and populations examined (Harris, 1975). The correlation matrix of allele frequencies was subjected to PCA (SAS Institute, 1992), to generate linear combinations (principal components) of the original variables (loci) which maximally discriminate among the accessions. The first three principal components were plotted in three dimensions with STATISTICA software (1994) to depict genetic relationships. The relative spatial relationships among the 18 populations were examined, and distinctive accessions were identified by their unique isozymic profiles (15 isozyme loci) and their relationship to other accessions after PCA.

Groupings identified by visual assessment were tested for differences using chi-square analysis. Isozymes useful for discrimination were identified by two-sample, chi-square goodness-of-fit tests (Gibbons, 1976), and an overall comparison of the populations was made by using the combined probabilities from independence tests procedure of Sokal and Rohlf (1981). This procedure allowed for the test of difference between the populations examined and grouped them according to their inherent genetic similarities (Staub et al., 1997a, Staub et al., 1999). Relative distances between groups were further investigated using Nei's I identity coefficient (GI) (Nei, 1987) and genetic distance (GD) (Nei, 1973) as estimated by computations using a similarity matrix derived from allele frequencies using computer algorithms in POPGENE (Yeh et al., 1997).

Establishment of test arrays. Cucumber accessions in the NPGS were previously evaluated for their resistance to angular leafspot [*Pseudomonas lachrymans* (E.F. Smith) Holland], anthracnose [*Colletotrichum lagenarium* (Ross.) Ellis & Halst], downy mildew [*Pseudoperonospora cubensis* (Berk. & Curt) Rostow], rhizoctonia fruit rot (*Rhizoctonia solani* Kuhn), target leafspot [*Corynespora cassiicola* (Berk. & Curt) Wei], and water and heat stress tolerance (Staub and Krasowska, 1990; Staub and Palmer 1987; Staub et al., 1989; Wehner and Shetty, 1997). The U.S. Cucurbit Germplasm Committee (CGC) accepted these data for entry into GRIN based on the performance of accessions in replicated evaluation under controlled conditions. These tests adhere to published protocols and/or protocols specifically developed for germplasm evaluation and supported by the CGC.

GRIN evaluation data, as well as isozymic variation, were used to establish test arrays (Table 3). Initially, test arrays constructed were based on their relative resistance (disease as greenhouse seedling screening or field tests) or tolerance (environmental stress) reactions (GRIN) (Staub and Palmer, 1987; Staub et al., 1989; Staub and Krasowska, 1990; Wehner and Shetty, 1997). Phenotypic ratings ranged from 0 (resistance–tolerance) to 10 (complete susceptibility and/or death). Relatively low ratings (1 to 3) indicated a degree of resistance or tolerance, intermediate ratings (4 to 6) indicated some resistance or tolerance attributes, and high ratings (7 to 9) indicated varied susceptibility depending upon the disease or environmental stress. The mean, as well as the minimum and maximum ratings for each trait, were considered as critical decision criteria for the initial formation of test arrays.

Initial test arrays were comprised of the most resistant–tolerant plant introductions to pathogen–environmental stresses [ratings between 0 and 3.5 (replication mean) in greenhouse tests or 0 to 2.0 in field tests]. These were subsequently reduced in size

Table 3. Test arrays composed of plant introductions (PI) drawn from the U.S. National Cucumber (*Cucumis sativus* L.) Germplasm Collection based phenotypic response to pathogen[z] or environmental stress and genetic diversity.[y]

Angular leafspot		Anthracnose		Downy mildew		Rhizoctonia fruit rot	
PI no.	Country	PI no.	Country	PI no.	Country	PI no.	Country
163213	Pakistan	169381	Turkey	164173	India	163213	Pakistan
164465	India	200815	Myanmar	175121	India	164679	India
167223	Turkey	227209	Japan	178886	Turkey	197086	India
175688	Turkey	234517	USA	197086	India	197087	India
197088	India	249562	Thailand	227209	Japan	206953	Turkey
209064	USA	267745	Brazil	234517	USA	211979	Iran
209069	USA	279466	Japan	249562	Thailand	234517	USA
249562	Thailand	279468	Japan	288238	Egypt	2495561	Thailand
255938	Netherlands	288238	Egypt	321010	Taiwan	269480	Pakistan
279466	Japan	358813	Malaysia	357857	Yugoslavia	285607	Poland
288238	Egypt	385967	Kenya	358814	Malaysia	321008	Taiwan
321011	Taiwan	390260	Japan	390243	Japn	358813	Malaysia
390240	Japan	391573	China	422182	Netherlands	368558	Yugoslavia
432855	China	422182	Netherlands	426169	Philippines	390261	Japan
432865	Japan	432867	China	432877	China	391570	China
483339	South Korea	432867	China	466922	Russian Fed.	418989	China
489754	China	483343	South Korea	483339	South Korea		

[z]Causal agents are angular leafspot [*Pseudomonas lachrymans* (E.F. Smith) Holland], anthracnose [*Colletotrichum lagenarium* (Ross.) Ellis & Halst], downy mildew [*Pseudoperonospora cubensis* (Berk. & Curt) Rostow], rhizoctonia fruit rot (*Rhizoctonia solani* Kuhn), target leafspot causal agent [*Corynespora cassiicola* (Berk. & Curt) Wei].
[y]According to Staub and Krasowska, 1990; Staub and Palmer 1987; Staub et al., 1989.

by examination of their genetic similarity and geographical grouping following Meglic et al. (1996) (Table 2). Accessions that were genetically most dissimilar and had been assigned high resistance–tolerance ratings formed the final test array for each trait considered. The final number in each test array differed, and selection was based on relative differences in disease rating and genetic similarity (i.e., some accessions from the same geographic region possessed similar ratings and/or similarities and were eliminated).

Formation of a core collection. The core construction process considered the resistance–tolerance to the disease–environmental stresses examined (i.e., accessions with ratings of 0 to 2) and relative genetic diversity. Susceptible PIs are not included in each test array because 1) except for accessions that are repeated in more than one array (i.e., multiple resistance), accessions in contrasting test arrays are susceptible and thus can be used for experimentation as susceptible controls; 2) accessions and cultivars susceptible to one or more of the traits evaluated have been identified and documented (release announcements and GRIN) and are ubiquitous (Wehner, 1994); 3) susceptible accessions are not usually of interest to primary end-users (i.e., breeders), and 4) susceptible accessions can be easily added by end-users based on their situational preference (e.g., susceptibility to local pathogens). The size and scope of the core collection was determined by examining the total genetic diversity of the 970 NPGS accessions (Knerr et al., 1989; Meglic et al., 1996) and the existing geographic groupings (Meglic et al., 1996), and then comparing this information to test array grouping by origin. All accessions in the test arrays were included in the core collection, as well as specific accessions (i.e., accessions from India and China acquired after 1992) that circumscribed the breath of genetic variation found in an extensive sampling of the presently constituted NPGS collection (1352 total accessions) (Horejsi and Staub, 1999; Meglic et al., 1996).

Results

Diversity analysis. Based on allelic frequency differences, the accessions examined could be partitioned into four distinct groups based upon visual examination (Groups A–D; $P > 0.05$) after PCA (Fig. 1). The genetic structure within populations (18) was defined by estimates of population-specific genetic variability (Table 2), and the genetic relationships between populations was characterized by pairwise comparisons of genetic distances (Table 4) and chi-square comparisons between Groups A–D (data not presented).

Variability within populations. The mean percentage of polymorphic loci (*P*) was 59.3% (Table 2). Relatively high levels of polymorphism were detected in accessions from Eastern Europe (71.4%), Turkey (76.2%), India New-1 (81.0%), India New-2 (90.5%), and those accessions not associated in geography (designated unassociated). The India New-1 and New-2 populations were considerably more heterogeneous than the other populations examined. Although the level of polymorphism in *C. s.* var. *hardwickii* accessions was relatively low (42.9%), the accessions examined were genetically distant from the other Group A accessions examined (Fig. 1, Table 4). The level of polymorphism in India New-1 and New-2 accessions (81% to 91%) was higher than that detected among China New-1 and New-2 accessions (48% to 57%). The level of polymorphism in India New-1 and India New-2 was principally due to variation at *Ak-3*, *Fdp-2*, *Mdh-2*, *Mpi-2*, *Pep-la*, *Pgd-2*, *Per*, and *Pgm* (data not presented). Likewise, estimated heterozygosity in India New-1 and India New-2 accessions (0.25 to 0.34) was higher than China New-1 and New-2 accessions (0.11 to 0.15). Cultivars released between 1937 and 1968 possessed a relatively high level of polymorphism (60%) when compared to the other cultivars examined (33.3% to 46.7%) (Groups C and D). Nevertheless, these cultivars did not differ appreciably in their relative heterozygosity (He).

The mean number of alleles per locus was 1.7, ranging from

Target leafspot		Water stress		Heat stress	
PI no.	Country	PI no.	Country	PI no.	Country
109484	Turkey	163222	India	118279	Brazil
137856	Iran	169392	Turkey	197086	India
181755	Lebanon	169395	Turkey	200818	Myanmar
212599	Afganistan	200815	Myanmar	211962	Iran
227209	Japan	204568	Turkey	211985	Iran
271328	India	211962	Iran	227209	Japan
288996	Hungary	249550	Iran	249562	Thailand
330628	Pakistan	263079	Russian Fed.	257486	China
351139	Former USSR	279468	Japan	269480	Pakistan
369717	Poland	292012	Israel	296121	Egypt
379283	Yugoslavia	308915	Russian Fed.	369717	Poland
422184	Czech Repub.	314426	Pakistan	385967	Kenya
426169	Philippines	344438	Iran	390240	Japan
432855	China	344445	Iran	422200	Czech Repub.
432857	China	422181	Czech Repub.	426169	Phillipines
432865	Japan	525075	Mauritius	466922	Russian Fed.

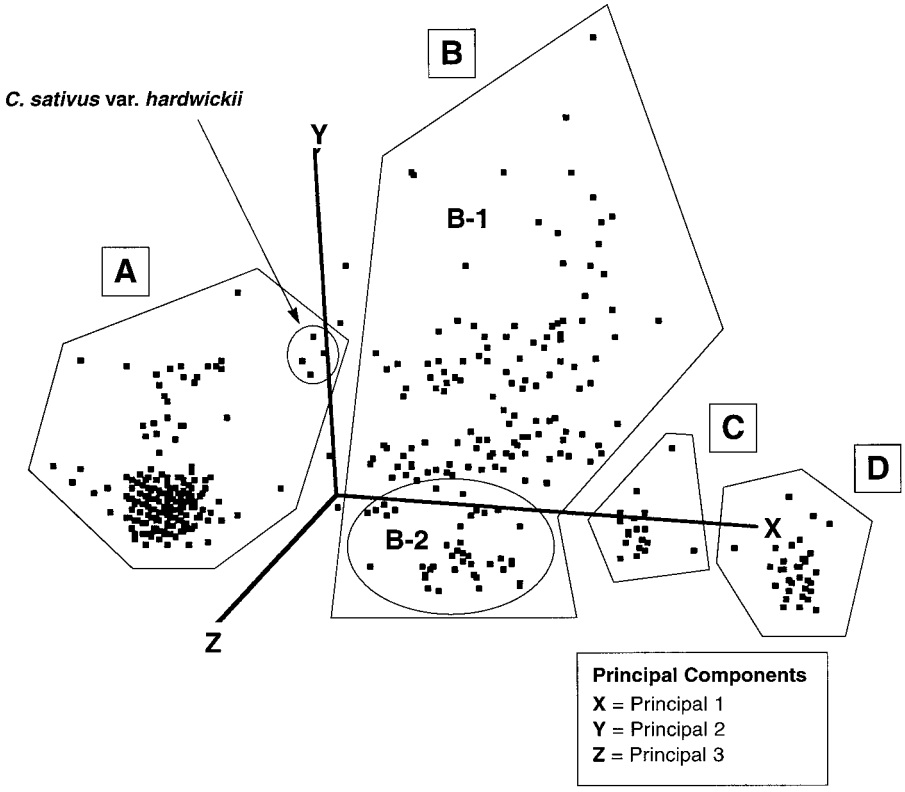
1.4 (population no. 18) to 2.1 (population no. 14) (Table 2). India New-2 accessions possessed the highest number of effective alleles per locus (1.6), and Japanese accessions (population no. 6) the lowest (1.2). The relatively high number of polymorphic loci and alleles estimated for India New-2 accessions were key factors that contributed to their comparatively high heterozygosity (He = 0.338). In contrast, Japanese accessions were relatively uniform (Ho = 0.049) when compared to the mean homozygosity (Ho = 0.110) of the populations examined.

Relationships among populations. The first three principal components accounted for ≈48% of the observed variation (principal components 1, 2, 3 explained 32%, 9%, and 7%, respectively), and were used to examine genetic relationships in the four major accession groupings consisting of 18 populations (Groups A–D, Fig. 1; Table 2). Group A consisted exclusively of populations 1 to 10, Group B contained India populations New-1 and New-2 (no. 11 and 12, respectively) and China populations New-1 and New-2 (nos. 13, and 14, respectively), Group C consisted of experimental commercial germplasm in use after 1985 (population no. 18) and five cultivars released after 1968 (population no. 17), and Group D contained population 15 to 17 accessions (cultivars released from 1846–1985). Those accessions that were not associated with a group (Fig. 1; 7 total, where 4 are visually apparent and 3 are hidden in Fig. 1) were those India New accessions that were not associated with any group in Staub et al. (1997a).

These interpopulation relationships (Groups A–D) were supported by chi-square analysis. Independence tests at each locus (combined P values) taken collectively indicated that Groups A–D (Fig. 1) were different (overall probability of difference = 0.05 < P < 0.01). *Gpi*, *Gr*, *G2dh*, *Mdh-2*, *Mpi-2*, *Pep-la*, *Pgd*, *Idh*, and *Skdh* were important in elucidating these major groups (data not presented). Rare alleles (frequency <0.01) in GPI [*Gpi-(3)*], GR [*Gr-(3)*], MDH-1 [*Mdh-1-(2)*], MPI-2 [*Mpi-2-(3),(4)*], PEP-LA [*Pep-la-(3), (4), (5)*], PEP-PAP [*Pep-pap-(2)*], and PGM [*Pgm-(3)*] were detected in several populations.

Accessions in populations 1 to 10 (Group A) were closely related (GI from 0.91 to 0.99, Table 4; Fig. 1). In contrast, accessions of New India-1 and New India-2 (population no. 11 and 12) accessions differed from the newly acquired Chinese accessions (population no. 13 and 14) in Group B (New 1 and 2; GI from 0.65 to 0.77; new *Gpi* , *Gr*, and *Pep-la* alleles detected). Thus, Group B could be further partitioned (P < 0.05) into two subgroupings (Group B-1 and B-2). Subgroup B-1 contained

Fig. 1. Groups after principal component analysis of accessions (970) in the U.S. National Plant Germplasm System (NPGS) cucumber (*Cucumis sativus* var. *sativus* and var. *hardwickii*) collection, cultivars released between 1846–1985 (Group D), and experimental commercial germplasm in use after 1985 (Group C). Group A consists of NPGS accessions acquired before 1992, and Group B consists of NPGS accessions obtained from India (B-1) after 1992 and China (B-2) after 1994.



India New-1 and India New-2 population accessions and subgroup B-2 consisted of China New-1 and China New-2 (Fig. 1). These groupings are consistent with results of Staub et al. (1999). Cultivars released between 1846 and 1985 (population no. 15 to 17; Group D) were genetically similar (GI from 0.96 to 0.99), and possessed genetic affinities with commercial germplasm used after 1985 (population no. 18; GI from 0.89 to 0.97; Group C).

Establishment of test arrays. Upon examination of resistance–tolerance ratings in GRIN, accessions that had mean trait ratings between 0 to 3.5 and a reaction range between 0 and 6 were considered for inclusion in test arrays. Initial considerations for the creation of test arrays involved 43, 23, 78, 26, 102, 19, and 53 accessions evaluated for angular leafspot, anthracnose, downy mildew, rhizoctonia fruit rot, target leafspot, and water and heat stress resistance–tolerance, respectively. These accessions were used based on their reaction to abiotic and biotic stresses according to Staub and Palmer (1987), Staub et al. (1989), and Staub and Krasowska(1990), and are exclusively from Fig. 1 Group A. Based on their genetic similarity (Fig. 1; Tables 2 and 4) and relative mean rating (those ≤2.0), the number of accessions in these groups were reduced to 17, 16, 17, 16, 17, 16 and 16 for angular leafspot, anthracnose, downy mildew, rhizoctonia fruit rot, target leafspot, and water and heat stress resistance–tolerance, respectively (Table 3).

Several accessions possessed multiple resistances and were included simultaneously in several arrays (Table 3). These included PI 163213 (Pakistan; angular and target leafspot, water stress), PI 164173 (India; downy mildew and target leafspot), PI 164679 [India; angular leafspot and fruit rot (synom. belly rot)], PI 167223 (Turkey; angular leafspot, downy mildew), PI 197086 (India; downy mildew, fruit rot), PI 197088 (India; angular leafspot, downy mildew), PI 200815 (Myanmar; anthracnose, water stress), PI 209064 (U.S.; angular and target leafspot), PI 209069 (U.S.; angular and target leafspot, anthracnose), PI 211979 (Iran; downy mildew, fruit rot), PI 227209 (Japan; anthracnose, target leafspot, downy mildew), PI 234517 (US; anthracnose, angular leafspot, downy mildew), PI 249562 (Thailand, anthracnose, angular and target leafspot), PI 267942 (Japan; downy mildew, target leafspot), PI 271328 (India; target leafspot, fruit rot), PI 279466 (Japan; anthracnose, angular leafspot, downy mildew), PI 279468 (Japan; anthracnose, downy mildew, target leafspot, water stress), PI 288238 (Egypt; anthracnose, downy mildew, angular leafspot), PI 330628 (Pakistan; downy mildew, fruit rot, target leafspot), PI 358813 (Malaysia; anthracnose, downy mildew, fruit rot), PI 390240 (Japan; angular and target leafspot, fruit rot), PI 390259 (Japan; anthracnose, angular leafspot, downy mildew), PI 422182 (The Netherlands; anthracnose, downy mildew), PI 426169 (Philippines; downy mildew, target leafspot), PI 432855 (China; angular and target leafspot, downy mildew), PI 432865 (Japan; angular and target leafspot, downy mildew), and PI 483339 (South Korea; angular and target leafspot, downy mildew). These multiple resistance–tolerance accessions are unique because of their reaction level (rating ≤2.0) and geographic grouping (Fig. 1; Table 2). To increase the number of accessions in a potential core collection, however, some of these are not represented in each test array for which they possess resistance or tolerance.

Formation of a core collection. Where appropriate information exists, decisions on the inclusion of accessions in a core collection should consider their potential economic importance, their genetic constitution, and their genetic relationship relative to other accessions in the collection. The isozyme analysis of the populations examined defined their interpopulation relationships (Fig. 1; Table 2) and intrapopulation structure (data not presented). This information, along with isozymic analysis of Staub et al. (1999) (21 loci) and the evaluation data from GRIN (Table 3) allowed for the construction of a core collection for the NPGS cucumber collection.

Accessions possessing economically important traits present in the test arrays initially defined this core collection. Their

Table 4. Estimates of genetic identity (above diagonal) and distance (below diagonal) for 18 cucumber (*Cucumis sativus* L.) populations.

Population[z]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	****	0.97[y]	0.97	0.98	0.98	0.96	0.99	0.98	0.97	0.99	0.66	0.70	0.70	0.73	0.37	0.36	0.37	0.42
2	0.03[x]	****	0.91	0.96	0.99	0.91	0.98	0.99	0.94	0.99	0.65	0.72	0.68	0.71	0.37	0.35	0.35	0.42
3	0.03	0.10	****	0.98	0.93	0.99	0.96	0.92	0.96	0.92	0.65	0.62	0.70	0.75	0.35	0.35	0.35	0.38
4	0.02	0.05	0.02	****	0.97	0.98	0.99	0.97	0.97	0.96	0.67	0.66	0.72	0.75	0.37	0.37	0.37	0.41
5	0.02	0.01	0.07	0.03	****	0.92	0.99	0.99	0.96	0.99	0.66	0.71	0.70	0.72	0.37	0.35	0.35	0.41
6	0.04	0.10	0.01	0.02	0.08	****	0.95	0.92	0.93	0.91	0.64	0.61	0.71	0.74	0.35	0.36	0.36	0.38
7	0.01	0.02	0.04	0.01	0.01	0.05	****	0.99	0.97	0.99	0.67	0.70	0.71	0.74	0.38	0.37	0.37	0.42
8	0.02	0.01	0.08	0.03	0.01	0.08	0.01	****	0.95	0.99	0.68	0.72	0.69	0.72	0.37	0.35	0.36	0.42
9	0.03	0.06	0.04	0.03	0.05	0.07	0.03	0.05	****	0.95	0.66	0.66	0.70	0.74	0.36	0.34	0.34	0.40
10	0.01	0.01	0.08	0.05	0.01	0.09	0.02	0.01	0.05	****	0.67	0.73	0.68	0.71	0.37	0.36	0.36	0.43
11	0.42	0.42	0.43	0.40	0.42	0.45	0.40	0.39	0.41	0.41	****	0.85	0.66	0.77	0.60	0.58	0.59	0.65
12	0.36	0.33	0.47	0.41	0.34	0.50	0.35	0.32	0.42	0.33	0.16	****	0.65	0.74	0.58	0.55	0.56	0.66
13	0.36	0.38	0.36	0.33	0.36	0.34	0.34	0.38	0.36	0.39	0.41	0.43	****	0.90	0.58	0.58	0.58	0.61
14	0.32	0.35	0.29	0.29	0.33	0.30	0.30	0.33	0.31	0.34	0.26	0.31	0.10	****	0.66	0.64	0.64	0.70
15	0.99	1.00	1.00	0.98	0.99	1.00	0.97	0.99	1.00	0.99	0.51	0.54	0.54	0.42	****	0.98	0.99	0.89
16	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.55	0.59	0.55	0.45	0.04	****	0.99	0.95
17	1.00	1.00	1.00	0.99	1.00	1.00	0.99	1.00	1.00	1.00	0.54	0.58	0.55	0.44	0.04	0.01	****	0.97
18	0.88	0.87	0.97	0.89	0.88	0.97	0.86	0.86	0.92	0.84	0.42	0.42	0.49	0.36	0.04	0.04	0.04	****

[z]Groupings by geographic proximity where 1 = USA, 2 = Eastern Europe, 3 = China (pre-1994), 4 = India (pre-1992), 5 = Iran, 6= Japan, 7 = Unassociated (no geographic proximity), 8 = Turkey, 9 = *C. sativus* var. *hardwickii*, 10 = Western Europe, according to Meglic et al. (1996), and 11 = China New-1, 12 = China New-2, according to Staub et al. (1999), and 13 = India New-1, 14 = India New-2, according to Staub et al. (1999). Populations 15, 16, 17, and 18 are commercial germplasm that was released between 1846 to 1937, 1937 to 1968, 1968 to 1985, and used experimentally after 1985, respectively.

[y]Estimate according to Nei (1973).

[x]Estimate according to Nei (1987).

resistance–tolerance reactions to biotic and abiotic stresses (Table 3) and genetic differences (i.e., grouping by geographic proximity; Meglic et al., 1996) provided a first approximation for core construction. In addition, PIs of unique genetic character as determined by isozymic analysis of recently acquired Indian [Staub et al., 1997a; India New-1 and India New-2 (populations 13 and 14 herein)] and Chinese [Staub et al., 1999; China New-1 and China New-2 (populations 11 to 12 herein)] accessions were considered for inclusion. Based on a strategic assessment of their relative relationship to PIs in NPGS populations 1 to 10 and to each other by isozymic analysis [Staub et al., 1997a, 1999, and analysis herein (Group B, Fig. 1)], the following Indian and Chinese accessions (32) were included in the core collection: 1) Indian origin, PI 605005, PI 605926, PI 605927, PI 605930, PI 605944, PI 606009, PI 606017, PI 606021, PI 606026, PI 606027, PI 606031, PI 606038, PI 606038, PI 606042, PI 606053, PI 605996, PI 606029, and PI 606030, and 2) Chinese origin, PI 618893, PI 618900, PI 618909, PI 618913, PI 618922, PI 618924, PI 618926, PI 618929, PI 618932, PI 618933, PI 618946, PI 618954, PI 618956, and PI 618957. Thus, the total number of accessions in this cucumber core is 147 [test arrays (115; Table 3) + an a priori selection of unique PIs (32) that included rare alleles].

Discussion

Previous assessments of genetic diversity of cucumber used 21 isozyme loci (Meglic et al., 1996; Staub et al., 1997a, 1999). Berg and Hamrick (1997) suggest that 10 to 20 loci are usually adequate to describe differences among populations in most species. Indeed, data from Staub et al. (1997b) indicated that 15 loci are adequate for defining genetic relationships among diverse cucumber germplasm. In our study, isozymic variation at these same 15 loci in experimental commercial germplasm in use after 1985 and cultivars released between 1846 and 1985 was added to previous isozymic data matrices of NPGS accessions (Meglic et al., 1996) for population analysis.

Like Meglic et al. (1996) and Staub et al. (1997a, 1999), we used PCA to describe relationships among cucumber germplasm and observed genetic concordance with those studies (Fig. 1). The relationships between Groups A and B (Fig. 1) were consistent with those reported by Staub et al. (1999) employing isozyme analysis (21 loci). Likewise, Horejsi and Staub (1999) used RAPD loci (71) to analyze selected PIs identified by Meglic et al. (1996) and additional present-day commercial cultivars, and found that the PIs examined did indeed circumscribe genetic variation in the NPGS. The relative proximity of Groups C and D to Groups A and B in our study provides a refined appraisal of the genetic relationships between germplasm in use between 1846 to the present and that held presently by the NPGS.

Relationships among accessions can change when additional data are added to or deleted from a database for reanalysis. In addition, interpretations may change when each reanalysis is performed using different numbers of loci. In our analysis, we added isozymic data from 260 accessions (population no. 15 to 18; Table 2) not present in previous analyses (Meglic et al., 1996, Staub et al., 1997a and 1999). In addition, we surveyed variation at 15 of the 21 isozyme loci used in previous studies. Graphic depiction of relationships after PCA in our study also allowed for the identification of four distinct gene pools (Groups A–D; Fig. 1).

These gene pools were not the same as those identified by Staub et al. (1997a, 1999). Earlier isozyme evaluation of introductions in the NPGS received after 1992 from India (Staub et al., 1997a) and China (Staub et al., 1999) identified two accession groupings in each country (India New-1 and India New-2, China New-1 and China New-2). These groupings were distinct from each other and from the other accessions in the NPGS forming four distinct groups (Staub et al., 1999). In our study, recently acquired Indian and Chinese accessions (India New-1, China New-1, India New-2, and China New-2) formed a single group (Group B, Fig. 1) positioned between previously acquired NPGS accessions (before 1992, Group A) and commercial experimental lines and cultivars (Groups C and D). Although they were distinct from other PIs in the NPGS (population no. 1 to 10) and formed two subgroups (B-1 and B-2), the genetic difference between these two germplasm pools was less obvious than in a previous analysis (Staub et al., 1999).

These spatial arrangements after PCA suggest that recently collected Indian and Chinese accessions are genetically closer to experimental commercial germplasm and cultivars (adapted germplasm) than are the NPGS accessions received before 1992 (Fig. 1). Genetic analyses by Meglic and Staub (1996a) and Dijkhuizen et al. (1996) indicated that genetic diversity in commercial cucumber is relatively low (i.e., 3% to 8% among elite adapted germplasm and 12% between commercial *C. s.* var. *sativus* germplasm and *C. s.* var. *hardwickii*). Our data suggest that breeders could select genotypes within accessions from either of two distinct gene pools (Groups A and B) in order to increase the genetic diversity of commercial germplasm. Indeed, the spatial position of experimental commercial germplasm in use after 1985 (Group C) and that of cultivars released between 1846–1985 (Group D) suggests that there may have been recent efforts to incorporate unadapted germplasm into elite germplasm. This hypothesis is supported by a comparison of the relative increase in heterozygosity in the cultivars examined and the time of their release (Table 2, population 15 < 16 < 17 < 18). Accessions in the NPGS collected before 1992 and cultivars released between 1846 and 1985 are genetically most distant (Fig. 1). It is intriguing to consider whether accessions from these groups or those in Group B should be considered in efforts to increase the diversity of commercial cucumber. Although Groups A and D are genetically distinct from modern commercial germplasm (Group C), recently collected accessions from India and China (Group B) are genetically closer to Group C (i.e., may be introgressed more easily) and have unique alleles (data not presented) when compared to accessions in Groups A and D.

India New-2 (Fig. 1, Group B-1) contains accessions from the states of Rahjasthan and Madhya Pradesh, and possesses the largest intrapopulation diversity among all groups sampled (Staub et al., 1997) (Table 2). Since India is the center of diversity for cucumber, and Group B-1 accessions are measurably more diverse than the other Indian populations (population no. 4 and 13) examined, we hypothesize that a center of origin for cucumber might well be in central-western India. However, this hypothesis is difficult to test since extensive collections of cucumber have not been made in ethnobotantically well-defined agricultural and rural areas of eastern and southern India. The fact that landraces are slowly being replaced by American and European F1 hybrids in many Indian production areas (Staub et al., 1997a), speaks to the urgency of further collection in these areas.

Harlan (1972) advocated the use of active working collections for genetic resources management. The use of situation-specific subsets of germplasm collections (synom. test arrays) can in-

crease the effective use of genetic resources (Beuselinck and Steiner, 1992; Rana and Kochhar, 1996). In many cases, national germplasm programs do not have core collections associated with species under curation. In those national germplasm programs where emphasis is given to cooperative networking (i.e., distribution and collaborative experimentation) of active subsets of core collections, multilocation evaluation information can enhance the decision-making capabilities of end-users (e.g., breeders, curators, geneticists). By using phenotype (reaction to abiotic and biotic stresses) and genetic data (isozyme variation), we identified test arrays that could afford such possibilities.

Priorities of order for the construction of test arrays (i.e., which accessions to include) were made by considering the mean reaction ratings of accessions, the range of reaction within the accession, and the genetic differences among candidate accessions. For some groups of accessions, degree of resistance–tolerance was relatively high but the level of genetic diversity was comparatively low. For instance, although the diversity among Japanese accessions was relatively low (Tables 2 and 4), several accessions possessed multiple resistance–tolerance to abiotic and biotic stresses (Table 3). Thus, the test arrays constructed herein represent a compromise between the resource allocation considerations, the relative reaction of phenotypes (inclusion resistance–tolerance accessions with minimal reaction variances), and relative genetic difference. Such compromises are inevitable during test array construction and highlight the need for their continued refinement as new accessions and data become available.

Brown (1989a) suggested that core collection formation should strive to increase the effectiveness of the curator in collection management, endeavor to mirror the structure of the broader collection, and seek to maximize the potential benefit to intended users. The core concept and such circumscript suggestions for its creation have been generally accepted by both curatorial staff and potential users (Beuselinck and Steiner, 1992; Peeters and Galwey, 1988). Core collections in the NPGS have been constructed for various crop species of diverse origin, sexual behavior, and intended utility (Allen Stoner, NPGS, Beltsville, Md., personal communication). These include alfalfa (*Medicago sativa* L.), barley (*Hordeum vulgare* L.), bean (*Phaseolus ssp*. L.), bluegrass (*Poa pratensis* L.), clover (*Trifolium pratense* L. and *T. repens* L.), corn (*Zea mays* L.), okra [*Abelmoschus esculentus* (L.) Moench], medics (*Medicago lupulina* L.), pea (*Pisum sativum* L.), peanut (*Arachis hypogaea* L.), potato (*Solanum tuberosum* L.), ryegrass (*Lolium perenne* L.), tomato (*Lycopersicon esculentum* Mill.), sweet clover (*Melilotus alba* L.), sunflower (*Helianthus annuus* L.), trefoil (*Lotus corniculatus* L.), and wheat (*Triticum aestivum* L.). The cucumber core collection described herein is intended to increase curatorial effectiveness by facilitating the distribution of accessions having either a genetic diversity representative of that found in the collection as a whole or situation-specific arrays possessing traits of economic importance. In either case, these constructs form a starting point for end-users desiring a prescribed subset of genetic variation found in the whole collection.

The structure of a core collection (i.e., type and frequency of alleles) is dependent on the genetic constitution of the accessions that are included in it (Schoen and Brown, 1993). Theoretical estimates of retention of genetic variation at most loci indicate that inclusion of about 10% of the entire collection in the core collection will, in most cases, allow for the recovery of ≈70% of the alleles with 95% certainty (Brown, 1989a). However, the number of accessions in each of the existing core collections varies based on curatorial constraints, resources, and the rationale involved in developing an optimized breeder's core structure (e.g., sexual behavior, genetic information). For instance, while the corn core collection consists of 249 (1.7 %) of the available accessions (15,019), 2,304 accessions (51.8%; Ames, Iowa) are included in the potato core collection (4,444 accessions, Sturgeon Bay, Wisc.) (Allen Stoner, personal communication). Beyond these extremes the number of accessions in other NPGS core collections range between 7% to 15% of the entire collection, except for tomato (5.4%), bluegrass (5.3%), ryegrass (5.1%), alfalfa (5.6%), and clover (6.5%). We have constructed a cucumber core collection consisting of genetically distinct germplasm that represents 10.8% of the current collection.

There have been several strategies proposed and evaluated for improving core construction (Basigalup et al., 1995; Johnson and Hodgkin, 1999; Schoen and Brown, 1993; van Hintum, 1994; van Hintum et al., 2000; van Hintum and Haalman, 1994; Zeuli and Qualset, 1993). Genotype by environment interactions can be a major factor when considering core formation from agronomic trait data (Charmet et al., 1993). Like the core collections of trefoil and bean (Allen Stoner, personal communication), the cucumber core collection constructed herein was based initially on the evaluation of geographical groupings and morphological–disease resistance traits. This germplasm array was then refined using genetic information obtained by molecular markers. Given the genetic information about this collection [GRIN, Meglic et al. (1996), Staub et al. (1999)] and theoretical estimates of allelic retention after sampling (Staub, 1999; Widrlechner et al., 1992), we believe that the size of this core collection is adequate for initial or preliminary evaluation by end-users and amenable for future refinement as unique accessions are identified that have economically important characteristics.

Literature Cited

Allendorf, F.W., N. Mitchell, N. Remain, and G. Stahl. 1977. Isozyme loci in brown trout (*Salmon trutta* L.): Detection and interpretation from population data. Hereditas 86:179–190.

Basigalup, D.H., D.K. Barnes, and R.E. Stucker. 1995. Development of a core collection for perennial *Medicago* plant introductions. Crop Sci. 35:1163–1168.

Berg, E.E., and J.L. Hamrick. 1997. Quantification of genetic diversity at allozyme loci. Can. J. For. Res. 27:415–424.

Beuselinck, P.R. and J.J. Steiner. 1992. A proposed framework for identifying, quantifying and utilizing plant germplasm resources. Field Crops Res. 29:26–272.

Bretting, P.K. and M.P. Widrlechner. 1995. Genetic markers and plant genetic resource management. Plant Breeding Rev. 13:11–86.

Brewer, G.B. 1970. An introduction to iozyme techniques. Academic Press, New York. p. 186.

Brown, A.H.D. 1989a. The case for core collections, p. 136–156. In: A.H.D. Brown, O.H. Frankel, D.R. Marshall, and J.T. Williams (eds.). The use of plant genetic resources. Cambridge University Press, Cambridge, U.K.

Brown, A.H.D. 1989b. Core collections: A practical approach to genetic resources management. Genome 31:818–824.

Charmet, G., F. Balfourier, C. Ravel, and J.B. Denis. 1993. Genotype × environment interactions in a core collection of french perennial ryegrass populations. Theor. Appl. Genet. 86:731–736.

Dijkhuizen, A., W.C. Kennard, M.J. Havey, and J.E. Staub. 1996. RFLP variability and genetic relationships in cultivated cucumber. Euphytica 90:79–89.

Frankel, O.H. 1984. Genetic perspectives of germplasm conservation, p. 161–170. In: W.K. Arber, K. Illimense, W.J. Peacock, and P. Starlinger (eds.). Genetic manipulation: Impact of man and society. Cambridge

University Press, Cambridge, U.K.
Gibbons, J.D. 1976. Nonparametric methods for quantitative analysis. Holt, Rinehart and Winston, New York. p. 463.
Hamrick, J.L., M.W. Godt, and S.L. Sherman-Broyles. 1992. Factors influencing levels of genetic diversity in woody plant species. New Forests 6:95–124.
Harlan, J.R. 1972. Genetic resources in sorghum, p. 1–3. In: N.G.P. Rao and L.R. House (eds.). Sorghum in the seventies. Oxford and IBH, New Delhi.
Harris, R.J. 1975. A primer of multivariate statistics. Academic Press, New York. p. 332.
Horejsi, T. and J.E. Staub. 1999. Genetic variation in cucumber (*Cucumis sativus* L.) as assessed by random amplified polymorphic DNA. Gen. Res. Crop Evol. 46:337–350.
Johnson, R.C. and T. Hodgkin (eds.). 1999. Core collections for today and tomorrow. IPGRI, Rome.
Knerr, L.D. and J.E. Staub. 1992. Inheritance and linkage relationships of isozyme loci in cucumber (*Cucumis sativus* L.). Theor. Appl. Genet. 84:217–224.
Knerr, L.D. and J.E. Staub. 1991. A multivariate re-evaluation of biochemical genetic diversity in *Cucumis sativus* L. Cucurbit Genet. Coop. 14:25–28.
Knerr, L.D., J.E. Staub, D.J. Holder, and B.P. May. 1989. Genetic diversity in *Cucumis sativ*us L. as assessed by variation at 18 allozyme coding loci. Theor. Appl. Genet. 78:119–128.
Leppik, E.E. 1966. Searching gene centers of the genus *Cucumis* through host–parasite relationship. Euphytica 15:323–328.
López-Sesé, A., J.E. Staub, N. Katzir, and M.L. Gomez-Guillamon. 2002. Estimation of between and within accession variation in selected Spanish melon germplasm using RAPD and SSR markers to assess strategies for large collection evaluation. Euphytica (in press).
Meglic, V. and J.E. Staub. 1996a. Genetic diversity in cucumber (*Cucumis sativus* L.): II. An evaluation of selected cultivars released between 1846–1978. Genet. Res. Crop Evol. 46:547–558.
Meglic, V. and J.E. Staub. 1996b. Inheritance and linkage relationships of allozyme and morphological loci in cucumber (*Cucumis sativus* L.). Theor. Appl. Genet. 92:865–872.
Meglic, V., F. Serquen, and J.E. Staub. 1996b. Genetic diversity in cucumber (*Cucumis sativus* L.): I. A reevaluation of the U.S. germplasm collection. Genet. Res. Crop Evol. 43:533–546.
Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. USA 70:3321–3323.
Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.
Peeters, J.P. and N.W. Galwey. 1988. Germplasm collections and breeding needs in Europe. Econ. Bot. 42:503–521.
Peterson, C.E. 1975. Plant introductions in the improvement of vegetable cultivars. HortScience 10:575–579.
Plucknett, D.L., N.J.H. Smith, J.T. Williams, and N.M. Anishetty. 1983. Crop germplasm conservation and developing countries. Science 220:163–169.
Rana, R.S. and S. Kochhar. 1996. Core subsets of base collections and priorities of national progammes: Indian perspective. Genet. Res. Crop Evol. 43:423–428.
Richmond, R.C. 1972. Enzyme variability in the *Drosophila willistoni* group. III. Amounts of variability in the superspecies *D. paulistorum*. Genetics 70:87–112.
SAS Institute. 1992. SAS/STAT user's guide. release 6.03 edition. SAS Inst., Cary, N.C. p. 1028.
Schoen, D.J. and A.H.D. Brown. 1993. Conservation of allelic richness in wild crop relatives as aided by assessment of genetic markers. Proc. Natl. Acad. Sci. 90:10623–10627.
Shaw, C.R. and R. Prasad. 1970. Starch gel electrophoresis of enzymes—A compilation of recipes. Biochem. Genet. 4:297–320.
Sokal, R.R. and F.J. Rohlf. 1981. Biometry. 2nd ed. W.H. Freeman, New York.
StatSoft, Inc. 1994. STATISTICA. StatSoft, Inc., Tulsa, Okla. p. 638.
Staub, J.E. 1999. Intellectual property rights, genetic markers, and hybrid seed production. J. New Seeds 1:39–65.
Staub, J.E., F.C. Serquen, T. Horejsi, and J.F. Chen. 1999. Genetic diversity in cucumber (*Cucumis sativus* L.): IV. An evaluation of Chinese germplasm. Genet. Res. Crop Evol. 46:297–310.
Staub, J.E., F.C. Serquen, and J.D. McCreight. 1997a. Genetic diversity in cucumber (*Cucumis sativus* L.): III. An evaluation of Indian germplasm. Genet. Res. Crop Evol. 44:315–326.
Staub, J.E., J. Box, V. Meglic, T.F. Horejsi, and J.D. McCreight. 1997b. Comparison of isozyme and random amplified polymorphic DNA data for determining intraspecific variation in *Cucumis*. Genet. Res. Crop Evol. 44:257–269.
Staub, J.E., H. Barcfzynska, D. Van Kleinwee, M. Palmer, E. Lakowska, and A. Dijkhuizen. 1989. Evaluation of cucumber germplasm for six pathogens. Proc. Cucurbitaceae 89: Evaluation and enhancement of cucurbit germplasm, p. 149–153.
Staub, J.E. and A. Krasowska. 1990. Screening of the U.S. germplasm collection for heat stress tolerance. Cucurbit Genet. Coop. Rpt. 13:4–7.
Staub, J.E. and M.J. Palmer. 1987. Resistance to downy mildew [*Pseudoperonospora cubenisis* (Berk. & Curt.) Rostow.] and scab (spot rot*) [Cladosporium cucumerinum* Ellis and Authur] in *Cucumis* ssp. Cucurbit Genet. Coop. Rpt. 10:21–23.
Tatlioglu, T. 1993. Cucumber *Cucumis sativus* L., p. 197–234. In: G. Kalloo and B.O. Begh (eds.). Genetic improvement of vegetable crops. Pergamon Press Ltd., Tarrytown, N.Y.
van Hintum, T.J.L., A.H.D. Brown, C. Spillane, and T. Hodgkin. 2000. Core collections of plant genetic resources. IPGRI Tech. Bul. 3.
van Hintum, T.J.L. 1993. Pedigree analysis for comparing a core collection of modern cultivars, with examples from barley (*Hordeum vulgare* S. Lat.). Theor. Appl. Genet. 88:70–74.
van Hintum, T.J.L. 1994. Comparison of marker systems and construction of a core collection in a pedigree of European spring barley. Theor. Appl. Genet. 89:991–997.
Wehner, T.C. 1994. A set of cucumbers to represent the American market. Cucurbit Genet. Coop. Rpt. 17:12–13.
Wehner, T.C., J.D. McCreight, B. Rhodes, and X. Zhang. 1996. Mutually beneficial cucurbit expedition to the People's Republic of China leads to continued collaboration with the United States. Diversity 12:13.
Wehner, T.C. and N.V. Shetty. 1997. Downy mildew resistance of the cucumber germplasm collection in North Carolina field tests. Crop Sci. 37:1331–1340.
Widrlechner, M.P., L.D. Knerr, J.E. Staub, and K.R. Reitsma. 1992. Biochemical evaluation of germplasm regeneration methods for cucumber, *Cucumis sativus* L. Plant Genet. Res. Newsletter 88/89:1–4.
Yeh, F.C., R.C. Yang, T. Boiley, Z.-H. Ye, and J.X. Mao. 1997. POPGENE, the user-friendly shareware for population genetic analysis. Mol. Biol. Biotechnol. Ctr., Univ. Alberta, Canada.
Zeuli, P.L.S., and C.O. Qualset. 1993. Evaluation of five strategies for obtaining a core subset from a large gene resourse collection of durum-wheat. Theor. Appl. Genet. 87:295–304.