Euphytica (2011) 178:261–272
DOI 10.1007/s10681-010-0315-8

Marker-assisted backcross selection in an interspecific *Cucumis* population broadens the genetic base of cucumber (*Cucumis sativus* L.)

Tusar K. Behera · Jack E. Staub · Snigdha Behera · Isabelle Y. Delannay · Jin Feng Chen

Received: 28 March 2010 / Accepted: 22 November 2010 / Published online: 16 December 2010

Abstract Cucumber (*Cucumis sativus* L.) is a major cucurbit vegetable species whose genetic base has been drastically reduced during its domestication. The crop's narrow genetic base (3–12% DNA polymorphism) has resulted from the use of limited genetic material and intense selection during plant improvement. Recently, however, interspecific hybridization has been successful in *Cucumis* via mating of *C. hystrix* Chakr. and *C. sativus*, which resulted in the amphidiploid *C. hytivus*. We report herein a marker-assisted strategy for increasing genetic diversity in cucumber through introgression backcrossing employing *C. hytivus*. The comparatively late-flowering but high-yielding, indeterminate, monoecious line WI 7012A (P_1; donor parent) derived from a *C. hytivus* × *C. sativus*-derived line (long-fruited Chinese *C. sativus* cv. Beijingjietou) was initially crossed to the determinate, gynoecious *C. sativus* line WI 7023A (P_2; recurrent parent 1), and then advanced backcross generation progeny (BC_2) were crossed with the gynoecious indeterminate line WI 9-6A (P_3; recurrent parent 2). More specifically, a single F_1 individual ($P_1 \times P_2$) was backcrossed to P_2, and then BC progeny were crossed to P_2 and P_3, where marker-assisted selection (MAS) for genetic diversity (8 mapped and 16 unmapped markers; designated Sel) or no selection (designated NSel) was applied to produce BC_3P_2 (Sel) and BC_3P_3 (Sel), and BC_2P_2 (NSel) and $BC_2P_2S_1$ (NSel) progeny. Relative vegetative growth, number of lateral branches (LB), days to flowering (DF), yield (fruit number), and fruit quality [as measured by length:diameter (L:D) and endocarp:total diameter (E:T) ratios] were assessed in parents and cross-progeny. DF varied from ~20 (BC_3P_2Sel) to ~25 days (BC_2P_3Sel) among the populations examined, where progeny derived from P_2 possessed the shortest DF. Differences in cumulative yield among the populations over six harvests were detected, varying from ~8 fruits per plant in BC_3P_2 (Sel) to ~39 fruits per plant in BC_2P_3 (Sel). Although the vigorous vegetative growth of line P_1 was observed in its backcross progeny, highly heterozygous and polymorphic backcross progeny derived

T. K. Behera (✉)
Division of Vegetable Science, Indian Agricultural Research Institute, PUSA Campus, New Delhi 110 012, India
e-mail: tusar@rediffmail.com

J. E. Staub · S. Behera · I. Y. Delannay
USDA-ARS, Vegetable Crops Unit, Department of Horticulture, University of Wisconsin, 1575 Linden Drive, Madison, WI 53706, USA

Present Address:
J. E. Staub
USDA-ARS, Forage and Range Research Laboratory, Utah State University, Logan, UT 84322-6300, USA

J. F. Chen
State Key Laboratory of Crop Genetics and Germplasm Enhancement, Key Laboratory of Southern Vegetable Crop Genetic Improvement, Nanjing Agricultural University, 210095 Nanjing, China

from P_3 were comparatively more vigorous and bore many high-quality fruit. Response to selection was detected for LB, DF, L:D, and E:T, but the effectiveness of MAS depended upon the parental lines used. Data indicate that the genetic diversity of commercial cucumber can be increased by introgression of the *C. hystrix* genome through backcrossing.

Keywords Genetic diversity · *C. hytivus* · Molecular markers · Morphological traits

Introduction

Cucumber (*C. sativus* L.; $2n = 2x = 14$) improvement is a complex process, often involving refinement of populations derived from intercrossing elite and/or exotic (unadapted) germplasm, extraction of inbred lines from such populations, and subsequent identification of commercially acceptable F_1 hybrids (Staub et al. 2008). Genetic diversity in *C. sativus* is relatively low (3–8%) when compared with other *Cucumis* species (10–25%) (Kupper and Staub 1988; Horejsi and Staub 1999; Luan et al. 2008). In fact, the genetic diversity of elite cucumber lines is extremely low (1–3%) (Dijkhuizen et al. 1996; Staub et al. 2002b; Behera et al. 2010), highlighting the need for continued introgression of exotic germplasm to ensure broad-based improvement of cucumber.

The concept of gene pools (primary, secondary, and tertiary) and their utilization for broadening species diversity was proposed by Harlan and de Wet (1971). The primary gene pool of *C. sativus* comprises primarily two interfertile botanical varieties, var. *sativus* L. and var. *hardwickii* R. Alef. (Lebeda et al. 2007). The secondary *Cucumis* gene pool houses *C. hystrix* Chakr. (H), which is sparely cross-compatible with *C. sativus* (C) (Chen et al. 2003) and has been suggested as a source for broadening the genetic base of cucumber (Bates and Robinson 1995). A synthetic amphidiploid species, *C. hytivus* Chen and Kirkbride 2000 ($2n = 4x = 38$; HHCC), was developed by interspecific mating between *C. sativus* ($2n = 2x = 14$; CC) and *C. hystrix* ($2n = 2x = 24$; HH) (Chen and Kirkbride 2000; Chen et al. 2003). Subsequently, allotriploids ($2n = 3x = 36$) and backcross-derived, fully fertile diploid ($2n = 14$; BC_1S_3) derivatives were produced from amphidiploid × diploid (*C. sativus* var. *sativus*; recurrent parent) mating. These genetic stocks may be useful for broadening the genetic base of commercial cucumber.

Novel genes, such as those that condition disease resistance [i.e., gummy stem blight (causal agent *Didymella bryoniae*)], are not found in cultivated cucumber but are present in *C. hystrix* (Chen et al. 2003). Backcross introgression has been suggested for incorporation of such genes (Wehrhahn and Allard 1965) and has shown potential for improving the genetic diversity and yield of cucumber (Owens et al. 1985).

Moderately saturated linkage maps have been developed for cucumber, and genomic regions associated with economically important quantitative and qualitative trait loci have been identified (Bradeen et al. 2001; Fazio et al. 2003a; Ren et al. 2009). The efficiency and effectiveness of selection during line and population development in cucumber can be increased through MAS (Fan et al. 2006; Robbins and Staub 2009). In fact, backcrossing with concurrent initial molecular-based genotyping and selection for genetic diversity has proven useful in cucumber (Fan et al. 2006; Delannay 2009; Delannay and Staub 2010a; 2010b). Therefore, a study was designed that employed marker-assisted backcrossing to increase genetic diversity in *C. sativus* by introgression of the *C. hystrix* genome into elite breeding lines. This process consisted of using mapped and unmapped markers (Fazio et al. 2003a; Kong et al. 2006; Nam et al. 2005; Ritschel et al. 2004) to survey *C. hystrix* × *C. sativus*-derived BC_1 progeny for the construction of marker-genotyped BC_2 progeny. These progenies were then backcrossed and/or successively self-pollinated to produce BC_3 or BC_2S_1 lines, respectively, and compared with nonselected progeny. Such genetically diverse lines will have potential for direct use in cucumber improvement if they possess adequate yield and fruit quality characteristics.

Materials and methods

Parents

Three parents were used herein to produce genetically diverse BC_2, BC_2S_1, and BC_3 lines for evaluation. Line

WI 7012A (P_1), which is a late-flowering, indeterminate, monoecious BC_1S_3 line derived from a *C. hytivus* × *C. sativus* (long-fruited Chinese *C. sativus* cv. Beijingjietou; recurrent backcross parent) mating, was received from Nanjing Agricultural University, Nanjing, China (Chen et al. 2003). Although this line possesses a sequential fruiting habit and a comparatively high number of primary lateral branches (>4) and yield, it bears warty, light-green fruit of commercially unacceptable shape and quality. Gynoecious determinate line WI 7023A (P_2) and indeterminate line WI 9-6A (P_3) were created through selection and backcrossing {Gy-7 (recurrent parent; University of Wisconsin) and H-19 [donor parent; University of Arkansas (Fayetteville)]} to identify a small-statured genotype that develops high-quality fruit for once-over mechanical harvest operations (Robbins and Staub 2009). These lines originated from the same populations that were used to develop recombinant inbred lines for the mapping of quantitative trait loci (QTL) in US processing cucumber at the Agricultural Research Service, Madison, WI (Staub et al. 2002a; Robbins and Staub 2009).

Development of backcross families

The inbred backcross line (IBL) breeding method (Wehrhahn and Allard 1965) was employed to introgress *C. hystrix* genomic components into *C. sativus* via the *C. hytivus*-derived line WI 7012A. Line WI 7012A (considered as an inbred but "exotic" line) was crossed with WI 7023A to develop F_1 plants, which were subsequently mated with WI 7023A (recurrent parent) to create BC_1P_2 lines without selection [designated BC_1P_2 (NSel)] (Fig. 1). Three hundred ninety-two BC_1P_2 plants were evaluated for fruit yield (number per plant), number of primary lateral branches, fruit length:diameter ratio (L:D), and seed cavity diameter (endocarp:total diameter) ratio (E:T) in a field nursery [Plainfield loamy sand (Typic Udipasamment) soil] at the University of Wisconsin Experimental Station in Hancock, WI (UWESH)

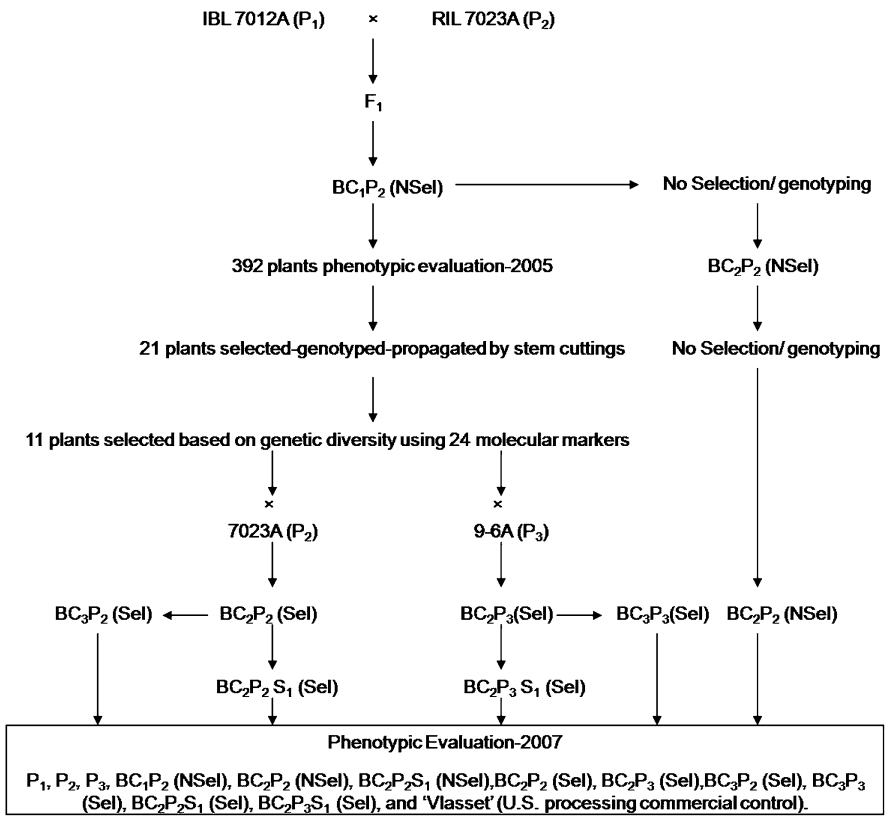

Fig. 1 Schematic representation of the development of F_1 and backcross populations in cucumber using *C. hytivus* × *C. sativus*. Evaluated at Hancock, WI in 2007. *IBL* inbred backcross line, *RIL* recombinant inbred line (Fazio et al. 2003a), *Sel* individuals selected for genetic diversity by 24 molecular makers, *NSel* individuals not receiving marker selection

during the summer of 2005 (Delannay 2009). Based on this field evaluation, 21 BC_1P_2 lines (plants) were selected [designated BC_1P_2 (Sel)] given their comparative fruit yield and quality over six harvests (i.e., base population for subsequent MAS or NSel). These selections were multiplied by stem cutting, grown in the greenhouse to produce healthy mature plants, and then genotyped with 24 DNA markers [8 mapped and 16 unmapped (Delannay 2009); see below]. Based on their comparative genetic diversity, 11 BC_1 genotypes were selected for crossing to elite lines WI 7023A (P_2) and WI 9-6A (P_3), the progeny of which were designated as BC_2P_2 (Sel) and BC_2P_3 (Sel), respectively. These MAS BC_2 lines were then backcrossed and/or self-pollinated without selection to produce $BC_2P_2S_1$ (Sel) and $BC_2P_3S_1$ (Sel) and also backcrossed to create BC_3P_2 (Sel) and BC_3P_3 (Sel) progeny (Fig. 1). In addition, non-MAS BC_1P_2 progeny were backcrossed once to P_2 and then self-pollinated to produce BC_2P_2 (NSel) and $BC_2P_2S_1$ (NSel) cross-progeny.

Molecular marker analysis

To assess genetic diversity, parent lines (WI 7012A, WI 7023A, and WI 9-6A) and derived cross-progeny populations were genotyped using nine single-nucleotide polymorphic (SNP; AI4SNP, AT1SNP, D11SNP, L1LSNP, M4LH2SNP, M7LG3SNP, M7LH3SNP, M8SNP, and W7SNP), nine sequence amplified characterized region (SCAR; AK5SCAR, BC231SCAR, BC526SCAR, P14SCAR, S_AV14SCAR, S_AV16-3SCAR, S_AU18-1SCAR, S_E9SCAR, S_M5SCAR), and six simple sequence repeat (SSR; CSWTAAA01SSR, CSWACC02SSR, CSWATT02SSR, CSWCT03SSR, CSWGAAA02SSR, and CSWTAA11BSSR) markers (Fazio et al. 2003a; Kong et al. 2006; Nam et al. 2005; Ritschel et al. 2004). The number of plants examined in each backcross population, along with their percentage polymorphism, average heterozygosity, and number of polymorphic loci, are presented in Table 1.

Marker analysis was preformed by bulk sampling (at least 10 samples) of young tissue in the three-leaf stage from parental lines and cross-progeny populations, where samples were held at −4°C (~2 h) until transfer to −80°C storage for DNA extraction according to Fazio et al. (2003a). Subsequently, polymerase chain reactions (PCR) and electrophoresis using SNP, SCAR, and SSR makers (as defined above) were performed according to Fazio et al. (2002) and Robbins et al. (2008). Banding morphotypes identified by markers that detected differences between 5 and 30 bp were visualized using 3% agarose gels run

Table 1 Statistical measures of genetic variation in various cucumber populations as measured by 24 molecular markers

Population[a]	N[b]	H[c]	I[d]	Number of polymorphic loci	Polymorphism (%)[e]	Average heterozygosity
BC_1P_2 (NSel)	98	0.29	0.44	29	78.4	0.290
BC_2P_2 (NSel)	96	0.25	0.38	28	75.7	0.254
$BC_2P_2S_1$ (NSel)	72	0.29	0.44	30	81.1	0.277
BC_1P_2 (Sel)	24	0.30	0.43	29	78.4	0.293
BC_2P_2 (Sel)	44	0.26	0.38	26	70.3	0.269
BC_3P_2 (Sel)	48	0.25	0.38	28	75.7	0.287
$BC_2P_2S_1$ (Sel)	48	0.28	0.43	31	83.8	0.286
BC_2P_3 (Sel)	44	0.31	0.45	32	85.7	0.306
BC_3P_3 (Sel)	48	0.29	0.40	33	85.7	0.296
$BC_2P_3S_1$ (Sel)	48	0.31	0.46	32	86.5	0.305

[a] *P_1* WI 7012A, *P_2* WI 7023A, *P_3* WI 9-6A, *Sel* selected for genetic diversity using SSR, SCAR, and SNP markers (Fazio et al. 2003a; Kong et al. 2006; Nam et al. 2005; Ritschel et al. 2004), *NSel* not selected for genetic diversity by markers

[b] *N* number of plants per population

[c] *H* Nei's (1973) gene diversity

[d] *I* Shannon's information index (Shannon and Weaver 1949)

[e] Percentage of polymorphic loci

at 250 V for 4 to 6 h, and amplicon differences greater than 30 bp were detected using 1.6% agarose gels run at 250 V for 2 h. For codominant markers that detected <5 bp differences, Alexa-labeled 2′-deoxyuridine 5′-triphosphates (dUTPs) was added to the PCR master mix, and then polymorphisms were identified using fragment analysis performed at the University of Wisconsin Biotechnology Center, Madison, WI. Banding and size differences were then analyzed using GeneMarker (version 1.2; Softgenetics, State College, PA).

Open field evaluation

Parental lines (WI 7012A, WI 7023A, and WI 9-6A) and cross-progeny populations as well as 'Vlasset' (Seminis Seed Company, Woodland, CA; control) were sown on May 18 in a greenhouse in Madison, WI, and then transplanted to a UWESH field nursery on June 20, 2007. The design was randomized complete block with three replications. Experimental plots consisted of 10 plants of each parental line (three) or cross-progeny family [BC_1P_2 (NSel), BC_1P_2 (Sel), BC_2P_2 (Sel), BC_2P_3 (Sel), BC_3P_2 (Sel), BC_3P_3 (Sel), BC_2P_2 (NSel), and $BC_2P_2S_1$ (NSel)] (Table 1) spaced 13 cm apart within rows (5.2 m long) on 1.5-m row centers (~51,000 plants/ha) with end and side borders of 'Vlasset.' Data were collected on relative plant vegetative growth, number of primary lateral branches (LB), days to first flower (DF), yield (number of fruit), and fruit L:D and E:T on per-plant basis. Plant vigor in melon is often associated with plant architecture, including sex expression (gynoecious/monoecious), flowering date, and plant stature (determinate/indeterminate) (Lebeda et al. 2007). Given the differing architecture of the parents, in this work vegetative growth (relative size of vines) was scored by visual observation, where disease-free plants were given comparative values from 1 to 4, such that 1 = limited growth (38–43 cm; usually manifest as early flowering in determinate, gynoecious lines), 2 = moderate growth (44–58 cm; usually manifest as early flowering in determinate, predominantly gynoecious lines), 3 = vigorous growth (59–73 cm; usually manifest as late flowering in indeterminate, monoecious lines), and 4 = extremely vigorous grow (>73 cm; usually manifest as extremely late flowering in indeterminate, monoecious lines) (Fig. 2). DF was recorded as the number of days between sowing and the appearance of the first fully expanded corolla. LB was recorded when individual plants reached anthesis, where only LB longer than 5 cm on the first 10 nodes was observed. Fruits per plant, fruit length and width, fruit endocarp width, and total fruit diameter were recorded at each of six harvests at 1-week intervals. Harvest began when a majority of the fruit on a plant were greater than 2 cm in diameter (equivalent to USDA 2A

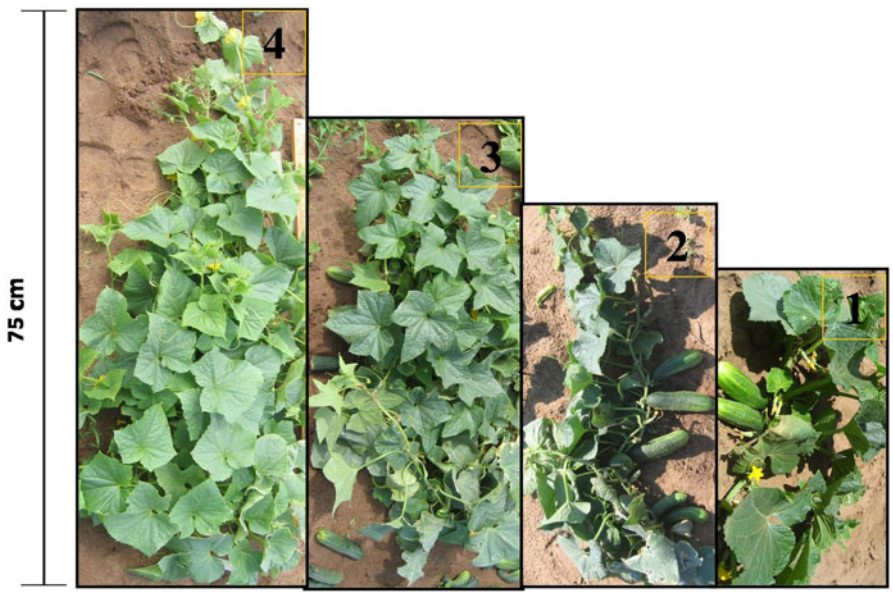

Fig. 2 Comparative cucumber plant growth recorded as relative score (values *1–4*) based on visual observation: *1* limited growth (38–43 cm), *2* moderate growth (44–58 cm), *3* vigorous growth (59–73 cm), and *4* extremely vigorous growth (>73 cm)

grade). Mean fruit L:D and E:T were obtained by measuring the length and diameter of fruit ranging between 2.5 and 3.0 cm in diameter (equivalent to USDA 2–3A grade). Cumulative numbers of fruit per plant were calculated by dividing the total number of fruit harvested by the number of plants.

Statistical analyses

Trait data were subjected to analyses of variance (ANOVA) using a mixed models procedure (PROC Mixed) to define blocks, and germplasm (parents and cross-progeny) effects, and genotype-by-environment interactions using SAS software (SAS 2003; Littell et al. 1996). To determine trait relationships, pairwise phenotypic Pearson correlations were calculated using SAS (2003). Data from each planting were initially combined for ANOVA to define block, and line (parents and cross-progenies) effects using the "proc glm" procedure in SAS. All variables were treated as random effects, and least-square means were calculated for each line using the "lsmeans" option in the "proc glm" procedure in SAS.

Based on allelic differences at marker loci, gene frequency, percent polymorphic loci, and mean heterozygosity were estimated using Tools for Population Genetic Analyses (TFPGA) version 1.3 software (Miller 1997). Estimates of genetic distance (GD; Nei 1973) and population differentiation (G_{ST}; McDermott and McDonald 1993) were calculated using POPGENE version 1.32 software (Yeh and Boyle 1997).

Results and discussion

The marker-based genetic and phenotypic analyses conducted herein formed the basis for comparative examination of backcross progeny. Such comparisons were critical in determining whether changes in genetic diversity occurred as a result of MAS and if the alterations in allelic frequency detected were the result of selection primarily for genomic segments of *C. hytivus* during MAS. The assessment of genetic (marker-based) and phenotypic (economically important traits) diversity conducted herein also provided for an estimation of the potential value of backcross progeny for plant improvement.

Genetic diversity

The genetic variation among the backcross populations examined was not significant ($P < 0.05$; data not shown). Nevertheless, backcross progeny derived from line WI 9-6A (P_3) were more polymorphic than those derived from WI 7023A (P_2). Likewise, although the genetic diversity as measured by the gene diversity coefficient (H^c) and Shannon's information index (I) were relatively high in all backcross progeny derived from WI 9-6A, values for such estimates were comparatively low in progeny derived from WI 7023A (Table 1). In fact, the estimated percentage of polymorphism was relatively high in marker-selected backcross progeny, ranging from 70.3% in BC_2P_2 (Sel) to 86.5% in $BC_2P_3S_1$ (Sel). Such polymorphism levels might be expected given the use of the *C. hystrix*-derived amphidiploid *C. hytivus* in the creation of donor parent WI 7012A {[*C. hytivus* × *C. sativus*) × *C. sativus*] S_3}. Similar differences between distantly related *C. sativus* var. *sativus* and var. *hardwickii* (Kupper and Staub 1988) have been exploited to create high-yielding, multiple disease-resistant lines (Staub et al. 1992). The differences observed in the marker-selected backcross progeny described herein portend their potential value in broadening the narrow genetic base of cucumber (3–12% polymorphism; Dijkhuizen et al. 1996; Horejsi and Staub 1999).

The observed levels of heterozygosity between marker-selected and nonselected backcross progeny ranged from 0.25 in BC_2P_2 (NSel) to 0.31 in BC_2P_3 (Sel) (Table 1). In contrast, heterozygosity differences between marker-selected backcross genotypes ranged from 0.27 in BC_2P_2 (Sel) to 0.29 in BC_1P_2 (Sel). Moreover, the heterozygosity detected in BC_1P_2-derived progeny was increased when mated to WI 9-6A to create BC_2P_3 and $BC_2P_3S_1$ progeny (heterozygosity ~0.31), suggesting that this parent (originating from LB sequential fruiting *C. sativus* line H-19) be used in future experiments directed towards increasing the genetic diversity of cucumber in combination with advanced lines originating from the *C. hytivus*-derived line WI 7012A. Emphasis should be placed on alignment of complementary quantitative trait loci (QTL) associated with fruit yield and quality through continued joint phenotypic and MAS recurrent selection (Robbins et al. 2008; Robbins and Staub 2009).

Gain from selection is to a great extent dependent upon the complementary characteristics of often genetically diverse parental lines. The mean genetic identity (*I*; synom. genetic similarity) among pairs of parents was lower than the *I* among any two pairs of backcross populations (Table 1). When backcross populations are considered as a group, *I* predictably increased dramatically during advanced backcrossing [i.e., among any backcross pair (BC_2P_2 versus BC_3P_2)] to as high as 0.91, suggesting that relatively few alleles contributed significantly to the among-population genetic variation; for example, the lowest *I* (0.320) was detected between backcross populations derived from WI 7012A (*C. hytivus* derived) and WI 7023A (*C. sativus* derived), where a gradual increase in genetic similarity was detected over successive backcrossing and selfing generations [e.g., 0.49 (BC_1P_2 NSel versus WI 7012A) to 0.59 (BC_3P_2 Sel versus WI 7012A)]. In contrast, the highest identity (0.91; most similar) was detected between populations BC_2P_2 (Sel) and BC_3P_2 (Sel) and populations BC_3P_3 (Sel) and $BC_2P_3S_1$ (Sel). Shannon's information index (I^d) showed very close values (range 0.38–0.46) among all backcross populations. These data suggest close genetic affinities among all backcross populations examined, and confirm earlier reports of MAS that dramatic gains from selection for yield and quality components in cucumber often occur during early backcross [(*C. sativus* × *C. sativus*) × (*C. sativus* × *C. sativus*)] generations (Fazio et al. 2003b; Fan et al. 2006).

The largest genetic distance among the parents used was detected between the parents WI 7012A and WI 7023A (Nei's GD = 0.68; Table 2). Predictably, all marker-selected backcross progeny derived from WI 7012A were genetically more diverse than those derived from WI 7023A. Curiously, WI 7023A-derived backcross progenies were more similar to WI 9-6A-derived progeny than to WI 7012A-derived progeny. Nevertheless, the GD among backcross progeny, regardless of parental donor, was similar. Such genetic affinities among advanced backcross progeny are likely attributable to the progressive level of inbreeding (backcrossing and selfing) that occurs during line development. Tarter et al. (2004), in fact, argued that, in maize (*Zea mays* L. ssp. *mays*), intensive inbreeding and phenotypically based backcrossing within elite lines derived from exotic sources (i.e., comparable to the *C. sativus* parents used herein) could result in loss of exotic alleles. Appropriately designed MAS in cucumber during backcrossing has increased selection efficiency even after

Table 2 Genetic distance (Nei 1973) among cucumber parental lines and their derived backcross populations as measured by 24 molecular markers (SSR, SCAR, and SNP)

Parental line or population[a]	WI 7012A	WI 7023A	WI 9-6A	BC_1P_2 NSel	BC_2P_2 NSel	$BC_2P_2S_1$ NSel	BC_1P_2 Sel	BC_2P_2 Sel	BC_2P_3 Sel	BC_3P_2 Sel	BC_3P_3 Sel	$BC_2P_2S_1$ Sel	$BC_2P_3S_1$ Sel
WI 7012A (P_1)	0.00												
WI 7023A (P_2)	0.68	0.00											
WI 9-6A (P_3)	0.53	0.32	0.00										
BC_1P_2 NSel[a]	0.51	0.19	0.29	0.00									
BC_2P_2 NSel	0.41	0.21	0.36	0.15	0.00								
$BC_2P_2S_1$ NSel	0.43	0.23	0.27	0.18	0.17	0.00							
BC_1P_2 Sel	0.43	0.26	0.30	0.10	0.16	0.21	0.00						
BC_2P_2 Sel	0.49	0.15	0.27	0.14	0.15	0.13	0.17	0.00					
BC_2P_3 Sel	0.43	0.17	0.22	0.14	0.14	0.13	0.16	0.14	0.00				
BC_3P_2 Sel	0.41	0.20	0.36	0.16	0.19	0.16	0.24	0.09	0.15	0.00			
BC_3P_3 Sel	0.44	0.29	0.19	0.21	0.20	0.19	0.25	0.19	0.12	0.24	0.00		
$BC_2P_2S_1$ Sel	0.48	0.25	0.40	0.22	0.15	0.21	0.25	0.17	0.17	0.14	0.23	0.00	
$BC_2P_3S_1$ Sel	0.44	0.28	0.26	0.20	0.16	0.23	0.23	0.17	0.13	0.22	0.09	0.11	0.00

[a] P_1 WI 7012A, P_2 WI 7023A, P_3 WI 9-6A

Sel selected for genetic diversity using SSR, SCAR, and SNP markers (Fazio et al. 2003a, Kong et al. 2006, Nam et al. 2005, Ritschel et al. 2004), *NSel* not selected for genetic diversity by markers

intense initial recurrent phenotypic selection (Fan et al. 2006). Thus, MAS in cucumber could be employed as a tool to mitigate loss of exotic alleles during inbreeding after relatively wide crossing (e.g., *C. hytivus* × *C. sativus*). The relatively high estimated level of gene diversity, Shannon's information index, and polymorphic loci defined herein lend support to this contention.

Phenotypic diversity

For exotically derived germplasm to be employed in agriculture it must possess acceptable commercial attributes such as high yield and quality. The inbred parents employed herein were chosen based on their unique and complementary architectural (e.g., determinate/indeterminate, sequential fruiting, and multiple lateral branching) and fruit quality (e.g., fruit dimensions and interior quality) characteristics (Table 3). The horticultural evaluation of the cross-progeny developed herein provides a preliminary assessment of their commercial potential. Parents and backcross progeny were assessed for their relative vegetative vigor (Fig. 1), LB, DF, fruits per plant, and fruit L:D and E:T. Nevertheless, although parental line differences were detected for these traits, several selected and nonselected cross-progeny populations were similar ($P < 0.05$). Such results indicate the need for the development of breeding strategies that consider trait correlations, epistatic interactions, and the strength of marker–trait associations (Fazio et al. 2003a; Robbins et al. 2008; Robbins and Staub 2009).

The strength and direction of fruit yield and quality correlations utilized herein has been well documented in a wide range of genetic backgrounds

Table 3 Least-square means (*ls* means) and least significant differences (LSD) for several traits as evaluated in cucumber parental lines and their derived backcross populations in the open field at Hancock, WI, 2007

Population	PG[a]	(LB)[b]	DF[c]	NF[d]	(L:D)[e]	(E:T)[f]
P_1 (WI 7012A)	3.72 ± 0.23	4.54 ± 0.28	27.5 ± 0.87	45.45 ± 2.40	3.58 ± 0.07	0.67 ± 0.01
P_2 (WI 7023A)	1.00 ± 0.23	1.12 ± 0.28	19.7 ± 0.87	12.55 ± 2.40	3.27 ± 0.07	0.64 ± 0.01
P_3 (WI 9-6A)	2.99 ± 0.23	3.24 ± 0.28	24.8 ± 0.87	36.65 ± 2.40	3.81 ± 0.07	0.57 ± 0.01
BC_1P_2 (NSel)	2.01 ± 0.16	2.17 ± 0.20	21.1 ± 0.61	25.47 ± 1.70	3.34 ± 0.05	0.65 ± 0.008
BC_2P_2 (NSel)	1.70 ± 0.16	2.01 ± 0.20	21.0 ± 0.61	23.84 ± 1.70	3.23 ± 0.05	0.65 ± 0.008
$BC_2P_2S_1$ (NSel)	1.64 ± 0.16	2.01 ± 0.20	21.0 ± 0.61	21.44 ± 1.70	3.31 ± 0.05	0.67 ± 0.008
BC_2P_2 (*Sel)	1.54 ± 0.16	2.19 ± 0.20	21.2 ± 0.61	24.51 ± 1.70	3.30 ± 0.05	0.65 ± 0.008
BC_3P_2 (Sel)	1.25 ± 0.16	1.51 ± 0.20	20.4 ± 0.61	17.94 ± 1.70	3.24 ± 0.05	0.64 ± 0.008
$BC_2P_2S_1$ (Sel)	1.55 ± 0.16	1.90 ± 0.20	21.3 ± 0.61	21.21 ± 1.70	3.15 ± 0.05	0.65 ± 0.008
BC_2P_3 (Sel)	2.87 ± 0.16	3.06 ± 0.20	24.8 ± 0.61	39.10 ± 1.70	3.65 ± 0.05	0.62 ± 0.008
BC_3P_3 (Sel)	2.59 ± 0.16	3.37 ± 0.20	24.1 ± 0.61	32.90 ± 1.70	3.76 ± 0.05	0.61 ± 0.008
$BC_2P_3S_1$ (Sel)	2.19 ± 0.16	2.89 ± 0.20	22.1 ± 0.61	24.92 ± 1.70	3.56 ± 0.05	0.62 ± 0.008
VLASSET	1.99 ± 0.23	3.12 ± 0.28	22.7 ± 0.87	29.80 ± 2.40	2.86 ± 0.07	0.59 ± 0.01
CV (%)	24.38	44.2	8.9	33.3	10.7	6.7
LSD (0.05)	0.17	0.39	0.70	3.15	0.13	0.02

* *Sel* selected for genetic diversity using SSR, SCAR, and SNP markers (Fazio et al. 2003a; Kong et al. 2006; Nam et al. 2005; Ritschel et al. 2004), *NSel* not selected for genetic diversity by markers

[a] *PG* Vegetative growth based on visual score, where plants were given comparative values from 1 to 4, based on vine length: *1* limited growth (38–43 cm), *2* moderate growth (44–58 cm), *3* vigorous growth (59–73 cm), and *4* extremely vigorous growth (>73 cm)

[b] *LB* Lateral branch number recorded as the number of lateral branches on the first ten nodes

[c] *DF* Days to flowering recorded as the number of days between transplanting and the appearance of the first fully expanded corolla

[d] *FN* Cumulative number of fruit per plant over six harvests

[e] *L:D* Average length (L) to diameter (D) ratio as determined from 5–10 fruits per plot over six harvests

[f] *E:T* Endocarp (E) to total diameter (T) ratio (E:T) as determined from 5–10 fruits per plot over six harvests

(Kupper and Staub 1988; Serquen et al. 1997; Cramer and Wehner 1999; Cramer and Wehner 2000b; Fazio et al. 2003a). For instance, lateral branch number was positively correlated ($r = 0.58$ to 0.42) with the number of fruit per plant in one processing cucumber population derived from the multiple branching line H-19 (Fazio 2001). Likewise, significant, positive correlations between yield and lateral branch number have been detected in a diverse array of cucumber populations (Fredrick and Staub 1989; Cramer and Wehner 1999, 2000a). This correlation, in fact, often increases during continued selection (from $r = 0.67$ to 0.82), which led Cramer and Wehner (2000b) to suggest that efforts to improve yield in cucumber should focus on increasing lateral branch number. Selection could be effective for this trait, since it is controlled by relatively few additive genes (~4–5; Wehner 1989; Serquen et al. 1997; Fazio et al. 2003a) that possess varying degrees of heritability [narrow-sense heritability (h^2) from 0.00 to 0.61] depending on the population exploited (Robbins and Staub 2009, 2008).

To mitigate negative correlations among yield and quality components in backcross progeny, complementary parental lines were used. In some cases, introgression of complementary parental characteristics resulted in backcross populations with unique characteristics (i.e., *C. sativus*- and *C. hytivus*-derived parents shared inferior and superior traits) (Table 3). For instance, even though a *C. sativus* parent (e.g., WI 7023A) was inferior for number of lateral branches (~1.00) and fruit yield (~13), it was superior for days to flowering (~20) and fruit quality [e.g., L:D (~3.2) and E:T (0.64) ratios]. As a result, derived backcross progeny (either Sel or NSel) often possessed commercially acceptable characteristics (i.e., yield and fruit quality) (Table 3). These findings are further supported by Delannay and Staub (2010b) who also found that backcross progeny derived from the complementary parents (i.e., WI 7023A and WI 7012A) exhibited more commercially acceptable traits than either parent alone. Likewise, the *C. sativus* inbred line WI 9-6A (P_3) used herein contributed superior yield and quality component traits [i.e., LB (~3), L:D (3.8), E:T (0.57), and cumulative FN (37)] but inferior DF (25) to its progeny. When used as a recurrent parent in advanced backcrossing (BC_2; $\{[(P_1 \times P_2) \times P_2] \times P_3\}$) to introduce complementary characteristics (BC_3P_3 and $BC_2P_3S_1$), progeny generally possessed an array of commercially acceptable attributes. Positive outcomes of MAS in cucumber, however, are not always predictable (Robbins and Staub 2009), and, thus, judicious use of marker–trait associations that employ appropriate breeding schemes (e.g., a combination of phenotypic and genotypic selection) is key to obtaining positive gain from selection.

Gain from selection for yield and quality traits in cucumber can be realized by altering plant architecture during selection. For instance, gynoecious, determinate, multiple lateral branching phenotypes with sequential fruiting habit have been proposed for increasing cucumber yield in once-over mechanical harvest operations (Staub et al. 2008; Cramer and Wehner 2000b). The multiple branching and sequential fruit habit is present in *C. sativus* var. *hardwickii*, *C. hystrix*, *C. hytivus*, and the *C. hytivus*-derived line WI 7012A used herein (Staub and Kupper 1985; Chen et al. 1997). These characteristics and the vigorous vegetative growth associated with WI 7012A make it attractive to breeding programs whose focus is yield improvement. Even though WI 7012A possesses some poor fruit quality attributes (i.e., predominant warts, blossom-end taper), the fruit of the WI 7023A and WI 9-6A-derived cross-progeny examined herein was commercially acceptable (Table 3). Moreover, some of these progeny possessed the plant architecture (i.e., gynoecious, determinate, multiple branching phenotypes) and yield attributes (e.g., early concentrated fruit set) required for mechanical harvest operations (data not shown).

Potential commercial value of backcross progeny

The maturity of backcross progeny varied due to differences in earliness and plant architecture (Table 3). For instance, although DF varied from ~20 (BC_3P_2S) to ~25 (BC_2P_3S) among primarily indeterminate backcross populations (i.e., BC_1 and BC_2), determinate and indeterminate plants were observed in early generations (data not presented). Regardless of generation, however, determinate plant types possessing early flowering habit were the consequence of WI 7023A alleles [P_2; determinate (*de*), gynoecious (*F*)]. Determinate gynoecious plant types typically flower earlier than indeterminate monoecious genotypes, and often provide a yield advantage over indeterminate genotypes in early (1–2) by not in late (4–6) harvests

(Staub et al. 2008). Consequently, differences in cumulative six-harvest yield were detected among backcross populations, varying from ~18 fruits per plant in BC_3P_2 (Sel) to ~39 fruits per plant in BC_2P_3 (Sel). In fact, the fruit yields of progeny with WI 7012A (monoecious, vigorous indeterminate) in their pedigree were consistently moderate to high (Table 3). Plant vigor (vine length) of both parental line WI 7012A (P_1; donor parent) (vigor score = 4; Fig. 2) and derived backcross progeny (Table 3; Fig. 3) were comparatively high (in relation to 'Vlasset'), especially in some advanced generations [e.g., BC_2P_3 (Sel), BC_3P_3 (Sel)]. These observations recapitulate similar findings reported by Delannay (2009) during advanced backcrossing in a *C. hytivus* genetic background with different recurrent *C. sativus* parental market types (i.e., Beit Alpha and Long European greenhouse). It is likely that the vigorous vegetative habit consistent with *C. hytivus*-derived progeny provided a yield advantage during later harvest dates, where its vigor-associated alleles (i.e., multiple lateral branching, indeterminate habit) were complementary but different from those of *C. sativus* recurrent parents used (Robbins et al. 2008). This was dramatically evident when WI 9-6A was used as the recurrent parent, since WI 9-6A-derived backcross progeny [e.g., BC_2P_3 (Sel), BC_3P_3 (Sel), BC_3P_3 (Sel), and $BC_2P_3S_1$ (Sel)] were more vigorous and at times

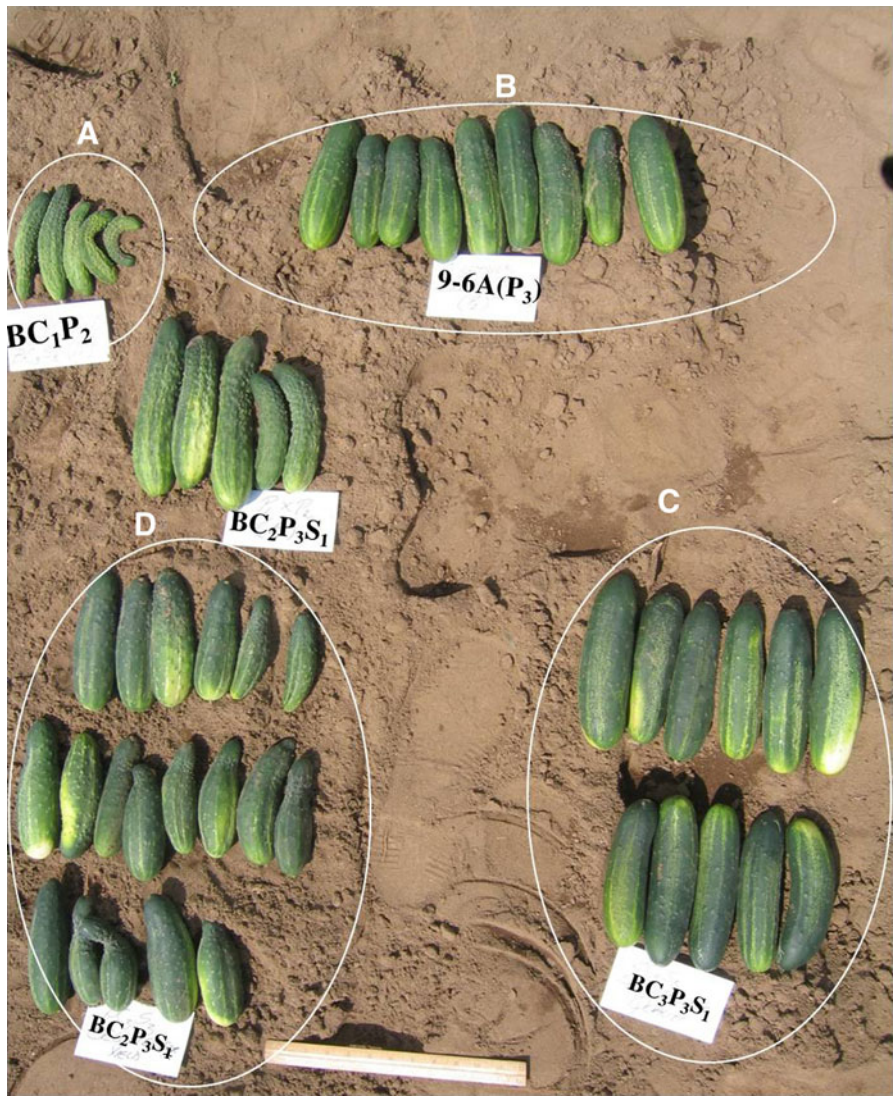

Fig. 3 Fruit of *Cucumis hytivus* WI 7012A-derived inbred backcross lines and *C. sativus* L. line 9-6A. **a** Fruit irregular in shape and spiny, **b** nonuniform dark-green fruit at edible maturity with low E:T ratio (0.57), **c** uniform dark-green fruit harvested at edible maturity from a single plant, **d** nonuniform dark-green fruit harvested at edible maturity from a single plant

higher yielding than BC_2P_2 progeny (Table 3; Fig. 3). These vigor and yield responses were likely potentiated by the use of WI 9-6A as an alternate parent in backcrossing [i.e., $BCP_2 \times P_3$] given the increase in heterozygosity and polymorphism level detected (Table 2). Line WI 9-6A originated from *C. sativus* crosses where line H-19 contributed sequential fruiting in a multiple lateral branching background (Staub et al. 2008). The yield and quality marker–trait associations were employed during MAS herein (Fan et al. 2006; Robbins et al. 2008) and likely led to the strategic pyramiding of yield-related QTL alleles (Robbins et al. 2008) that contributed to the documented yield increases. Recent increased saturation of the historic RIL-based map (Gy-7 × H-19) with SSR markers defined relatively tight marker–trait linkages with *de* (1.4 and 4.2 cM) and vigor-associated traits [e.g., little leaf (*ll*) at 4.2 and 3.6 cM] (Weng et al. 2010). These linkages will assist in MAS where the goal is recovery of determinate germplasm with increased vigor (increased lateral branching associated with *ll*).

The genotypic and phenotypic diversity of the backcross populations were significant for all the traits examined. Where direct comparisons could be made [i.e., BC_2P_2(NSel) versus BC_2P_2 (Sel), $BC_2P_2S_1$ (NSel) versus $BC_2P_2S_1$ (Sel)], marker-selected progenies were vigorous and early flowering (Table 3). Likewise, backcross progeny varied dramatically in morphology depending on the recurrent parent used (Fig. 2). Whereas multiple lateral branching (~3.5 laterals/plant) BC_3P_3 (Sel) progeny produced, as a cumulative average, 33 fruits/plant with mean L:D of 3.7, unilateral branching (~1.5 laterals/plant) BC_3P_2 (Sel) progeny produced comparatively fewer (~18 fruits/plant) and shorter (L:D of 3.2) fruit. Given the morphological diversity detected in these backcross populations and the potential benefit of using alternate recurrent parents (e.g., WI 9-6A) to increase diversity and improve population characteristics (e.g., plant vigor, LB, and fruit number and L:D), it is likely that the positive attributes of *C. hytivus*-derived WI 7012A can be capitalized upon during breeding. This assertion is supported by the fact that several backcross families [e.g., BC_2P_3 (Sel), BC_3P_3 (Sel), and $BC_2P_3S_1$ (Sel)] lacked the negative attributes associated with WI 7012A (i.e., late flowering, and warty, oblong fruit). If progeny with positive attributes could be recovered in such advanced backcross families, then MAS might have utility for further line advancement (fruit yield and quality; Fazio et al. 2003b; Fan et al. 2006) and subsequent assembly of diverse, heterozygous, commercially acceptable hybrids.

Acknowledgments The fund provided by the Department of Biotechnology, Ministry of Science and Technology, government of India for sponsoring T.K. Behera to carry out research work at the Department of Horticulture, UW Madison, USA is acknowledged and appreciated.

References

Bates DM, Robinson RW (1995) Cucumbers, melons and water-melons. In: Smartt J, Simmonds NW (eds) Evolution of crop plants. Longman, London

Behera TK, Staub JE, Behera S, Mason S (2010) Response to phenotypic and marker-assisted selection for yield and quality component traits in cucumber (*Cucumis sativus* L.). Euphytica 171:417–425

Bradeen JM, Staub JE, Wyse C, Antonise R, Peleman J (2001) Towards an expanded and integrated linkage map of cucumber (*Cucumis sativus* L.). Genome 44:111–119

Chen JF, Kirkbride JH (2000) A new synthetic species *Cucumis* (Cucurbitaceae) from inter-specific hybridization and chromosome doubling. Brittonia 52:315–319

Chen JF, Staub JE, Tashiro Y, Isshiki S, Miyazaki S (1997) Successful interspecific hybridization between *Cucumis sativus* L. and *C. hystrix* Chakr. Euphytica 96:413–419

Chen JF, Staub JE, Qian C, Jiang J, Luo X, Zhuang F (2003) Reproduction and cytogenetic characterization of interspecific hybrids derived from *Cucumis hystrix* Chakr. × *Cucumis sativus* L. Theor Appl Genet 106:688–695

Cramer CS, Wehner TC (1999) Little heterosis for yield and yield components in hybrids of six cucumber inbreds. Euphytica 110:99–108

Cramer CS, Wehner TC (2000a) Fruit yield and yield component correlations of four pickling cucumber populations. Cucurbit Genet Coop Rep 23:12–15

Cramer CS, Wehner TC (2000b) Path analysis of the correlation between fruit number and plant traits of cucumber populations. Hortic Sci 35:708–711

Delannay IY (2009) Use of molecular markers to increase genetic diversity of Beit Alpha, European long, and U.S. Processing market classes of cucumber (*Cucumis sativus* L.) through marker-assisted selection. PhD Dissertation, University of Wisconsin at Madison

Delannay IY, Staub JE (2010a) Use of molecular markers aids in the development of diverse inbred backcross lines in Beit Alpha cucumber (*Cucumis sativus* L.). Euphytica. DOI 10.1007/s10681-010-0183-2

Delannay IY, Staub JE (2010b) Backcross introgression of the *Cucumis hystrix* genome increases genetic diversity in U.S. Processing cucumber. J Am Soc Hortic Sci 135: 351–361

Dijkhuizen A, Kennard WC, Havey MJ, Staub JE (1996) RFLP variability and genetic relationships in cultivated cucumber. Euphytica 90:79–89

Fan Z, Robbins MD, Staub JE (2006) Population development by phenotypic selection with subsequent marker-assisted selection for line extraction in cucumber (*Cucumis sativus* L.). Theor Appl Genet 112:843–855

Fazio G (2001) Comparative study of marker assisted and phenotypic selection and genetic analysis of yield components in cucumber. PhD dissertation, University of Wisconsin, Madison

Fazio G, Staub JE, Chung SM (2002) Development and characterization of PCR markers in cucumber (*Cucumis sativus* L.). J Am Soc Hortic Sci 127:545–557

Fazio G, Staub JE, Stevens MR (2003a) Genetic mapping and QTL analysis of horticultural traits in cucumber (*Cucumis sativus* L.) using recombinant inbred lines. Theor Appl Genet 107:864–874

Fazio G, Chung SM, Staub JE (2003b) Comparative analysis of response to phenotypic and marker-assisted selection for multiple lateral branching in cucumber (*Cucumis sativus* L.). Theor Appl Genet 107:875–883

Fredrick LR, Staub JE (1989) Combining ability analyses of fruit yield and quality in near-homozygous lines derived from cucumber. J Am Soc Hortic Sci 114:332–338

Harlan JR, de Wet JM (1971) Toward a rational classification of cultivated plants. Taxon 20:509–517

Horejsi T, Staub JE (1999) Genetic variation in cucumber (*Cucumis sativus* L.) as assessed by random amplified polymorphic DNA. Genet Res Crop Evol 46:337–350

Kong Q, Xiang C, Yu Z (2006) Development of EST-SSRs in *Cucumis sativus* from sequence database. Mol Ecol Notes 6:1234–1236

Kupper RS, Staub JE (1988) Combining ability between lines of *Cucumis sativus* L. and *Cucumis sativus* var. *hardwickii* (R.) Alef. Euphytica 38:197–210

Lebeda A, Widrlechner MP, Staub JE, Ezura H, Zalapa J, Krı′stkova′ E (2007) Cucurbits (Cucurbitaceae; *Cucumis* spp., *Cucurbita* spp., *Citrullus* spp.). In: Singh RJ (ed) Genetic resources, chromosome engineering, and crop improvement, vol 3. CRC Press, Boca Raton, pp 271–376

Littell RC, Milliken GA, Stroup WW, Wolfinger RD (1996) SAS system for mixed models. SAS Institute Inc., Cary

Luan F, Delannay I, Staub JE (2008) Melon (*Cucumis melo* L.) diversity analyses provide strategies for germplasm curation, genetic improvement, and evidentiary support of domestication patterns. Euphytica 164:445–461

McDermott JM, McDonald BA (1993) Gene flow in plant pathosystems. Annu Rev Phytopathol 31:353–369

Miller MP (1997) Tools for population genetic analysis (TEPGA), Version 3. Department of Biological Sciences. Northern Arizona University, Arizona

Nam YW, Lee JR, Song KH, Lee MK, Robbins MD, Chung SM, Staub JE, Zhang HB (2005) Construction of two BAC libraries from cucumber (*Cucumis sativus* L.) and identification of clones linked to yield component quantitative trait loci. Theor Appl Genet 111:150–161

Nei M (1973) Genetic distance between populations. Am Nat 106:283–292

Owens KW, Bliss FA, Peterson CE (1985) Genetic-Variation within and between cucumber populations derived via the inbred backcross line method. J Am Soc Hortic Sci 110:437–441

Ren Y, Zhang ZH, Liu JH, Staub JE, Han YH, Cheng ZC, Li XF, Lu JY, Miao H, Kang HX, Bie BY, Gu XF, Wang XW, Du YC, Jin WW, Huang SW (2009) An integrated genetic and cytogenetic map of the cucumber genome. PLoS One 4:e5795

Ritschel P, de Lima Lins T, Tristan R, Cortopassi-Buso G, Amauri-Buso J, Ferreira M (2004) Development of microsatellite markers from an enriched genomic library for genetic analysis of melon (*Cucumis melo* L.). BMC Plant Biol 4:9

Robbins MD, Staub JE (2009) Comparative analysis of marker-assisted and phenotypic selection for yield components in cucumber. Theor Appl Genet 119:621–634

Robbins MD, Casler M, Staub JE (2008) Pyramiding QTL for multiple lateral branching in cucumber using nearly isogenic lines. Mol Breed 22:131–139

SAS (2003) SAS Software, Version 9.1 for Windows. SAS Institute Inc., Cary

Serquen FC, Bacher J, Staub JE (1997) Genetic analysis of yield components in cucumber (*Cucumis sativus* L.) at low plant density. J Am Soc Hortic Sci 122:522–528

Shannon CE, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, Urnaba

Staub JE, Kupper RS (1985) Results of the use of *Cucumis sativus var. hardwickii* germplasm following backcrossing with *Cucumis sativus var. sativus*. Hortic Sci 20:436–438

Staub JE, Peterson CE, Crubaugh LK, Palmer MJ (1992) Cucumber population WI 6383 and derived inbreds WI 5098 and WI 5551. Hortic Sci 27:1340–1341

Staub JE, Crubaugh LK, Fazio G (2002a) Cucumber recombinant inbred lines. Cucurbit Genet Coop Rep 25:1–2

Staub JE, Dane F, Reitsma K, Fazio G, López-Sesé A I (2002b) The formation of test arrays and a core collection in cucumber (*Cucumis sativus* L.) using phenotypic and molecular marker data. J Am Soc Hortic Sci 127:558–567

Staub JE, Robbins MD, Wehner TC (2008) Cucumber. In: Prohens J, Nuez F (eds) Vegetables I: Asteraceae, Brassicaceae, Chenopodiaceae, and Cucurbitaceae. Springer, New York, pp 241–282

Tarter JA, Goodman MM, Holland JB (2004) Recovery of exotic alleles in semiexotic maize inbreds derived from crosses between Latin American accessions and a temperature line. Theor Appl Genet 109:609–617

Wehner TC (1989) Breeding for improved yield in cucumber. Plant Breed Rev 6:323–359

Wehrhahn C, Allard RW (1965) The detection and measurement of the effects of individual genes involved in the inheritance of a quantitative character in wheat. Genetics 51:109–119

Weng Y, Johnson S, Staub JES, Huang S (2010) An extended intervarietal microsatellite linkage map of cucumber, *Cucumis sativus* L. Hortic Sci 45:882–886

Yeh FC, Boyle TJB (1997) Population genetic analysis of co-dominant and dominant markers and quantitative traits. Belg J Bot 129:157