Euphytica (2010) 175:65–78
DOI 10.1007/s10681-010-0183-2

Use of molecular markers aids in the development of diverse inbred backcross lines in Beit Alpha cucumber (*Cucumis sativus* L.)

Isabelle Y. Delannay · Jack E. Staub

Received: 4 January 2010 / Accepted: 28 April 2010 / Published online: 16 May 2010
© Springer Science+Business Media B.V. 2010

Abstract Beit Alpha cucumber (*Cucumis sativus* L.) is a Mediterranean fresh-market type with a relatively narrow genetic base. To broaden its base for plant improvement, 42 diverse accessions were compared employing a previously defined standard marker array to choose wide-based parental lines for use in backcross introgression. Inbred backcross lines (IBL) were developed by crossing Beit Alpha line '04HD5' (De Ruiter Seeds, The Netherlands; recurrent parent) and PI 285606 (Poland; donor parent), and then selecting the most genetically diverse BC_1 and BC_2 progeny based on molecular marker profiles, followed by three generations of single-seed descent to produce 117 IBL. Molecular genotyping of IBL was then performed, and IBL were evaluated for days to anthesis, sex expression, pistillate flowers per node, lateral branch number, fruits per plant, fruit length, and fruit weight in the US, The Netherlands, Israel, and Turkey. Multivariate analyses and genetic distance comparisons indicate that IBL possessed considerable inter-line morphological and genotypic diversity. These diverse IBL will be useful in genetic studies and to evaluate Beit Alpha cross-progeny derived from IBL × elite germplasm created to broadened genetic base of this market type.

Keywords Genetic diversity · Genetic distance · Morphological traits · Multivariate analysis · Inbred backcross lines · Cucumber

I. Y. Delannay (✉) · J. E. Staub
Vegetable Crops Unit, U.S. Department of Agriculture, Agricultural Research Service, Department of Horticulture, University of Wisconsin, 1575 Linden Dr, Madison, WI 53706, USA
e-mail: idelannay@yahoo.com

Present Address:
J. E. Staub
U. S. Department of Agriculture, Agricultural Research Service, Forage and Range Research Laboratory, Utah State University, Logan, UT 84322-6300, USA

Introduction

Beit Alpha is a cucumber (*Cucumis sativus* L.; $2n = 2x = 14$) type widely cultivated in Mediterranean regions for fresh market consumption (Shaw et al. 2000). Although its cultivation in other production areas is relatively recent, knowledge about its historical roots is sparse (Shaw et al. 2004). It is thought to have originated in Israel (Shaw et al. 2000; Villalta et al. 2003) where it is well suited to extreme environmental conditions, especially high temperatures (35–40°C) (Shaw et al. 2004). Despite its popularity, there have been no efforts by public breeding programs to improve this market type (Shaw et al. 2000; Shaw et al. 2004; Shaw et al. 2007).

Traditionally, the Beit Alpha market type has been grown under protected "hoop house" conditions, which allows for environmental control such that continuous cropping can occur in some growing environments (Shaw et al. 2000; Shaw et al. 2004). Depending on the cropping season, fruits from single

plants can be harvested up to 30 times and yield as many as 65 moderately small (15–18 cm in length), slightly curved, smooth-skinned, dark green fruit (Shaw et al. 2000; Shaw et al. 2004). Such prolific fruit production is due to the presence of multiple pistillate flowers per node and a relatively high number of primary lateral branches. These and culinary characteristics are critical considerations when developing best management practices for this crop (Shaw et al. 2000; Nandgaonkar and Baker 1981).

The narrow genetic diversity among cucumber market types and exotic accessions has been well documented {3–8% among elite and exotic germplasm and 12% between botanical varieties [*C. sativus* var. *sativus* L. and var. *hardwickii* (R.) Alef.]} (Dijkhuizen et al. 1996; Horejsi and Staub 1999; Meglic and Staub 1996; Meglic et al. 1996, Staub et al. 1992; Staub et al. 1997; Staub et al. 1999). However, the genetic base of Mediterranean fresh-market cucumbers, such as Beit Alpha, is not necessarily as narrow as other market classes. Mediterranean-type cucumbers (including the Beit Alpha type) have a relatively broad genetic base (genetic distance = GD = 0.09–0.55) when compared to other cucumber market classes such as the European Long type (GD = 0.00–0.24) (Dijkhuizen et al. 1996), but possess considerable within market class genetic affinities (Horejsi and Staub 1999). Given its unique genetic nature and putative recent single-source origin (Israel), a study was designed to diversify this market class by: (1) developing a genetically diverse array of inbred backcross lines (IBL; BC_2S_3) based on marker-assisted selection (MAS); (2) determining efficacy of MAS in IBL establishment, and; (3) estimating the genetic diversity within and among IBL and evaluating their potential breeding value. The creation of a genetically diverse set of Beit Alpha-type IBL with differing morphological characteristics will allow cucumber breeders to improve this market class, and to define and characterize quantitative trait loci (QTL) for MAS.

Materials and methods

Identification of parental lines and IBL development

In the fall of 2004, 42 accessions [20 elite cucumber lines, 17 diverse PIs from the US National Plant Germplasm System (NPGS) (Horejsi and Staub 1999), and five breeding lines from the US Department of Agriculture, Agricultural Research Service (USDA, ARS) cucumber breeding project, Madison, Wisc.] were evaluated in preliminary experiments to select parents (donor and recurrent) for IBL development using a standard marker array [22 simple sequence repeat (SSR; 17) and sequence characterized amplified region (SCAR; 5) markers] previously described by Staub et al. (2005). Markers were added to this array throughout the study based on continued parental screening for use in IBL creation and genotyping.

Twenty seeds from each accession were germinated in vermiculite in a greenhouse at the University of Wisconsin-Madison. Tissue from young expanding leaves of each accession was harvested and bulked for DNA analysis. This tissue was immediately lyophilized, and subsequent DNA extraction, polymerase chain reactions (PCR) for SCAR and SSR markers, electrophoresis, and product visualization were according to Fazio et al. (2003).

The standard marker array was employed to provide an initial estimate of genetic distance (GD; Jaccard 1908) for use in a multivariate analysis (multidimensional scaling; Kruskal 1964a, b) to identity parental lines using NTsys version 2.01 computer software (Rohlf 1998). Genetic relationships were examined and genetically distinct parents [recurrent (line ‘04HD5’) and donor (PI 285606)] were chosen and crossed to produce the F_1 progeny (Delannay 2009).

Gynoecious line ‘04HD5’ is an elite Beit Alpha type inbred obtained from De Ruiter Seeds (Bergschenhoek, The Netherlands) that typically possesses several pistillate flowers per node (2–3). The monoecious landrace PI 285606 originates from Warsaw, Poland (obtained by the NPGS in 1963). Line ‘04HD5’ flowers approximately 1 week later than PI 285606, and, in contrast to PI 285606, produces fruits ($\sim$length = 15 cm) that are slightly longer than those of PI 285606 ($\sim$length = 12 cm). While smooth, fine-spined (white) fruits of line ‘04HD5’ remain green beyond optimal commercial maturity, PI 285606 fruit possess thick, black-spines and turn orange upon maturity. Their F_1 progeny are predominantly female and develop orange fruit with thin, black spines (Fig. 1).

Cuttings were taken from a ‘04HD5’ × PI 285606-derived F_1 progeny and the original

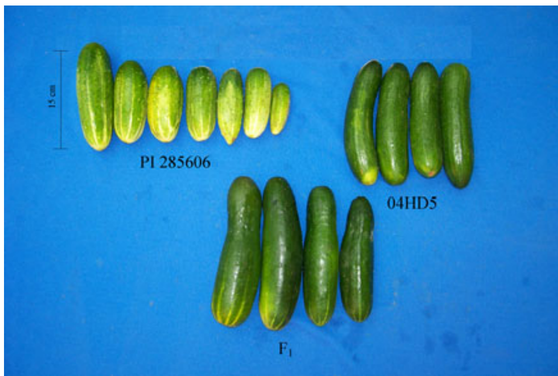


Fig. 1 Fruit of PI 285606 (donor parent), line '04HD5' (recurrent parent), and their F_1 progeny used in the development of inbred backcross lines of Beit Alpha cucumber (*Cucumis sativus* L.)

'04HD5' recurrent parent. These clones were rooted and greenhouse-grown to reproductive maturity, and then mated to generate the BC_1 generation where a cloned '04HD5' was used as the maternal parent and the F_1 progeny was used as the paternal parent. Tissue from young expanding leaves of F_1 and '04HD5' plants and 384 BC_1 seedlings at the first leaf stage was collected, and DNA was extracted as described above (Fazio et al. 2003).

Fifty BC_1 individuals were selected (selection intensity = 13%) for pollination based on their high heterozygosity at 24 mapped SSR (7), SCAR (8), and SNP (7) marker loci (Table 1; Fazio et al. 2002; Fazio et al. 2003). These BC_1 individuals were crossed with pollen from the cloned '04HD5' recurrent parent to produce BC_2 progeny. Eight seeds from each of the BC_2 families (50 × 8–384 seeds) were planted, sampled for DNA at the seedling stage and greenhouse-grown for pollination. One hundred-twenty BC_2 individuals having the greatest heterozygosity as defined at 35 marker loci (Table 1) were self-pollinated to produce BC_2S_1 lines. One plant from each BC_2S_1 line was then subsequently self-pollinated twice without further selection to generate 117 BC_2S_3 IBL by single seed decent (Wehrhahn and Allard 1965; Tanksley et al. 1996).

Phenotypic data collection

During the summer of 2008, parents and cross progeny were evaluated in Wisconsin, USA; Enkhuizen, The Netherlands; Beit Hanan, Israel; and Antalya, Turkey. In the USA, seeds of parental lines ('04HD5' and PI 285606), F_1 and F_2 progenies, and 117 BC_2S_3 IBL were sown in a field nursery [Plainfield loamy sand (Typic Udipasamment) soil] at the University of Wisconsin Experimental Station in Hancock, Wisconsin (UWESH). The experimental design was a randomized complete block (RCBD) with four blocks. Each treatment block (entry) had ten plants and consisted of plants spaced 15 cm apart in single rows positioned on 1.5 m centers (i.e., ~44,400 plants/ha) with edge borders. The average high temperature and relative humidity conditions were 26°C and 68%, respectively between May and August.

In Enkhuizen, The Netherlands; Beit Hanan, Israel; and Antalya, Turkey the recurrent parent ('04HD5') and IBL were evaluated at experimental stations maintained by Enza Zaden, Nickerson-Zwaan, and Nunhems seed companies, respectively. Germplasm was evaluated under protected-field conditions in "hoop houses" under a randomized (entries) but non-replicated design (i.e., single block). Each entry consisted of ten plants per plot. The soil types at Enkhuizen, Beit Hanan, and Antalya were Marine clay (leekeerd), sandy loam, and red Mediterranean lithosol, respectively. During August and September, the average high temperature and relative humidity conditions were 24°C and 78%, 32°C and 71%, and 34°C and 60% at Enkhuizen, Beit Hanan, and Antalya, respectively. Parents and cross progeny were germinated in greenhouses and transplanted at the 2- to 3-leaf stage, 15 cm apart within rows positioned on 1.5 m centers with end borders (i.e., ~44,400 plants/ha).

At all locations, plants were evaluated for the following traits: days to anthesis, sex expression, number of pistillate flowers per node, number of lateral branches per plant, fruit number, fruit length, and fruit weight. Days to anthesis, sex expression, pistillate flowers per node, and lateral branch number were scored once for each plant within a plot. Days to anthesis was recorded for individual plants as the number of days between either direct seeding or transplanting and the appearance of the corolla of the first fully expanded flower. For sex expression, individual plants within plots were given a numerical value based on their relative gynoecy as gynoecious (2), predominantly female (1), or monoecious (0). A plant was considered gynoecious if all flowers within

Table 1 Allelic frequency expectation fits at molecular marker loci assessed in BC_2S_3 Beit Alpha market class cucumber (*Cucumis sativus* L.) lines, where loci were used for selection at BC_1 and BC_2, and for genotyping of derived BC_2S_3 lines

Marker	Marker type	Polymorphism type	Linkage Group[a]	BC_1^b	BC_2^c	Used in final evaluation[d]	Expected allele frequency[e]	Observed allele frequency[f]	G-test[g]
A08SCAR	SCAR	Codominant		Yes	Yes	Yes	0.86	0.85	0.045 n.s.[h]
AA9b	SCAR	Dominant	F4	No	No	No			
AB14SNPG1H1	SNP	Codominant		Yes	Yes	Yes	0.86	0.80	22.17**
AC17SNPG1H2	SNP	Codominant	F4	Yes	Yes	Yes	0.86	0.91	2.44 n.s.
AD12	SCAR	Dominant		Yes	Yes	No			
AF15 SCAR	SCAR	Dominant	F4	Yes	Yes	No			
AI4SNPG1H1	SNP	Codominant		No	Yes	Yes	0.86	0.78	10.63**
AS5SCAR	SCAR	Dominant		No	No	No			
AT1SNPG3H3	SNP	Codominant		Yes	Yes	Yes	0.86	0.60	70.07**
AW14	SCAR	Codominant	F3	Yes	Yes	Yes	0.86	0.64	89.33**
BC519	SCAR	Codominant	F3	No	Yes	No	0.86	0.56	86.26**
BC231	SCAR	Dominant	F7,Y7	Yes	Yes	No			
BC523	SCAR	Dominant	F1,Y2	No	Yes	No			
C8RH	SNP	Dominant		Yes	Yes	No			
CM15	EST-SSR	Codominant		No	No	Yes	0.86	0.78	10.7**
CM39	EST-SSR	Codominant		No	No	No			
CM46	EST-SSR	Codominant		No	No	No			
CM49	EST-SSR	Codominant		No	No	Yes	0.86	0.85	3.65 n.s.
CS31	EST-SSR	Codominant		No	No	No			
CS39	EST-SSR	Codominant		No	No	No			
CS53	EST-SSR	Codominant		No	No	Yes	0.86	0.65	45.37**
CS-AT1	SCAR	Codominant		Yes	Yes	Yes	0.86	0.95	4.77 n.s.
CSWACC02	SSR	Codominant		Yes	Yes	Yes	0.86	0.78	10.99**
CSWATT02	SSR	Codominant		No	Yes	Yes	0.86	0.64	54.08**
CSWCT11	SSR	Codominant	Y7	Yes	Yes	Yes	0.86	0.71	24.23**
CSWCT13-1	SSR	Codominant	F3	Yes	No	Yes	0.86	0.74	18.28**
CSWCT13-2	SSR	Codominant	F3	No	Yes	Yes	0.86	0.78	6.81*
CSWCT13B	SSR	Codominant	F3	No	No	Yes	0.86	0.69	64.94**
CSWCT16	SSR	Codominant	F1	No	No	No	0.86	0.59	65.16**
CSWCT25	SSR	Codominant	F1,Y2	No	Yes	Yes	0.86	0.65	46.21**
CSWCTT02	SSR	Codominant		Yes	No	No			
CSWCTT11	SSR	Codominant		No	No	Yes	0.86	0.68	28.93**
CSWGAAA02	SSR	Codominant		No	Yes	Yes	0.86	0.57	110.93**
CSWGATT01C	SSR	Codominant		Yes	Yes	No	0.86	0.84	1.76 n.s.
CSWTA05	SSR	Codominant		Yes	Yes	Yes	0.86	0.87	2.28 n.s.
CSWTAAA01	SSR	Codominant	F4,S4	No	Yes	Yes	0.86	0.81	10.2**
CT269	SSR	Codominant		Yes	Yes	Yes	0.86	0.62	68.14**
D11SNPG3H1	SNP	Codominant		No	Yes	Yes	0.86	0.88	0.92 n.s.
EST-201	EST	Dominant		No	No	No			
EST-2057	EST	Codominant		No	No	Yes	0.86	0.65	84.02**
EST-210	EST	Dominant		No	No	No			
EST-211	EST	Dominant		No	No	No			
EST-2181	EST	Codominant		No	No	No	0.86	0.66	316.57**

Table 1 continued

Marker	Marker type	Polymorphism type	Linkage Group[a]	BC_1^b	BC_2^c	Used in final evaluation[d]	Expected allele frequency[e]	Observed allele frequency[f]	G-test[g]
EST-223	EST	Dominant		No	No	No			
EST-252	EST	Dominant		No	No	No			
EST-302	EST	Dominant		No	No	No			
EST-315	EST	Codominant		No	Yes	Yes	0.86	0.69	29.82**
EST-327	EST	Codominant		No	Yes	Yes	0.86	0.64	43.54**
EST-336	EST	Codominant		No	Yes	Yes	0.86	0.71	34.02**
F4	SCAR	Dominant	F2	Yes	Yes	No			
F05-75	SSR	Codominant		No	No	No			
F08-90A	SSR	Codominant		No	No	Yes	0.86	0.77	15.29**
L18-3	SCAR	Codominant	F1	Yes	Yes	Yes	0.86	0.73	32.85**
M4LH2	SNP	Dominant		Yes	Yes	No			
M-5-BE-L	BAC end	Codominant		No	Yes	No	0.86	0.61	153.66**
M-6-BE-R	BAC end	Codominant		No	No	No			
M7LG3	SNP	Dominant		Yes	Yes	No			
P14	SCAR	Codominant	F4,Y6	No	No	Yes	0.86	0.76	16.11**
S_AB14_2	SCAR	Codominant		No	Yes	Yes	0.86	0.81	17.03**
S_BC79	SCAR	Dominant		No	No	No			
S_DC1EM5	SCAR	Codominant		No	Yes	No	0.86	0.68	82.08**
S_ME1EM9	SCAR	Dominant		No	Yes	No			
W7SNPG1H3	SNP	Codominant		Yes	Yes	Yes	0.86	0.88	1.04 n.s.

[a] Linkage groups F1, F2, F3, F4, F5, F6, and F6 are from Fazio et al. (2003); linkage groups S1, S2, S3, S4, S5, S6, and S7 are from Sun et al. 2006a; linkages groups Y1, Y2, Y3, Y4, Y5, Y6, and Y6 are from Yuan et al. (2008)

[b] BC_1 indicates markers used during the marker assisted selection (MAS) of the BC_1 generation

[c] BC_2 indicates markers used during MAS of the BC_2 generation

[d] *Yes* indicates the use of that marker during selection or evaluation, *No* indicates the exclusion of the marker

[e] Expected allele frequency of the homozygous genotype of the recurrent parent in the BC_2S_3 generation

[f] Observed allele frequency towards the homozygous genotype of the recurrent parent at BC_2S_3 generation

[g] G-test [$G = 2 \sum O * \ln (O/E)$] on observed allelic frequencies of the BC_2S_3 lines compared to their expected BC_2S_3 allelic frequencies

[h] *n.s.*, *, and ** are not significant, $P \leq 0.05$, and $P \leq 0.01$, respectively

the first ten nodes of the plant were pistillate. Plants were classified as predominately female if greater than 51% of flowers on the first ten nodes were pistillate. If plants possessed 50% or fewer pistillate flowers within the first ten nodes, they were designated monoecious. The trait, pistillate flowers per node, was quantified by recording the maximum number of pistillate flowers on a single node in the first ten nodes of each plant. Lateral branch number was recorded when individual plants reached anthesis, and only lateral branches measuring longer than 5 cm on the first ten nodes were recorded.

Fruit number, length and weight were recorded on a per-plot basis at each of three harvests. Harvests occurred at 1 week intervals when a majority of the fruits were greater than 3 cm in diameter. All fruit greater than 2 cm in diameter were harvested within a plot. Five to ten commercially mature fruits (4–5 cm in diameter) from each plot were selected randomly and their lengths averaged. Fruit number was calculated by dividing the total number of fruits per plot at each harvest by the number of plants within a plot. Average weight of fruit per harvest was calculated per entry by dividing fruit weight by fruit number per

plot. Cumulative three-harvest fruit number and average three-harvest length and weight per plant were also calculated.

Phenotypic evaluation

Variance analyses and mean separation

Data from each location were initially combined for analysis of variance (ANOVA) to define location, block, and line effects and genotype-by-environment interactions using SAS software (version 9.1 for Windows; SAS Institute, Cary NC). Location and lines were treated as random effects and block was treated as a fixed effect. The only evaluation presented herein that defined the effect of block was conducted in the USA (i.e., replicated within location). Spearman rank correlations were performed for each USA block separately using the Spearman option within the *proc corr* procedure in SAS (2003). Significant correlations ($P \leq 0.05$) between blocks allowed for USA blocks to be averaged in further analyses (e.g., rank correlation, principal component analysis, and repeatability measures as given below).

In the case of the USA evaluation, least square means were calculated using *lsmeans* within the *proc glm* procedure in SAS. However, due to the lack of replication among the other test locations (3), lsmeans were only calculated for days to anthesis, sex expression, pistillate flowers per node, and lateral branch number for those locations. Least square means were not calculated for fruit number, fruit length, or fruit weight in the three locations since data for these variables were collected on a per-plot basis providing only one data point per entry. Likewise, the sex type data [i.e., gynoecious (2), predominantly female (1), and monoecious (0)] allowed for the computation of a sex score by giving a quantitative value (i.e., 0, 1, and 2) for each individual plant and taking the average value for each plot. Cumulative three-harvest fruit length and weight averages per plant were also used for ANOVA.

Spearman rank location correlation

The relative correspondence among locations for each variable was determined using Spearman rank correlation analyses (r_s) as computed by the *proc corr* procedure in SAS (i.e., Spearman statement option). For this analysis, days to anthesis, sex expression, pistillate flowers per node, and lateral branch number values were averaged for each line within a location. In addition to the rank correlations, each trait for a subset of the lines was plotted against location to visually identify dominant and consistent trends between locations.

Principal component analyses

Principal component analysis (PCA) was employed to describe among and within location entry relationships using the *proc princomp* procedure of SAS. Average sex expression, pistillate flowers per node, lateral branch number, fruit number, fruit length, and fruit weight were calculated by line for use in PCA. For the USA evaluation, blocks were pooled and all variables were averaged for PCA. Days to anthesis was not included in PCA due to the large discrepancy between the USA location and the other test locations (3) resulting from differences in plant-placement techniques (i.e., direct seeding or transplanting). Principal components (1–3) were visualized using the 3-D plot option within NTsys (Rohlf 1998).

Trait repeatability measures

Repeatability measures were performed for all variables to predict future performances within the IBL population (Falconer and Mackay 1989). Replication data (blocks) from the USA evaluation were averaged and the mean values for days to anthesis, sex expression, pistillate flowers per node, and lateral branch number for each line within a location were used for analysis. Estimates of variance for lines and locations (4) were obtained using the *covtest* option in the *proc mixed* procedure in SAS. Both location and lines were treated as random variables for repeatability estimations. Repeatability (r) was calculated according to Falconer and Mackay (1989) and standard error (SE) was calculated as adapted from Hallauer and Miranda (1988).

Molecular data collection

Each IBL was sampled for DNA extraction and analysis at the UWESH (USA) when plants had reached approximately five nodes in length by

harvesting the smallest leaf from each of 20 plants in the first two blocks. The leaves were bulked within each IBL and DNA was extracted as described above to provide 117 IBL samples for molecular genotyping. PCR was performed using these and samples of the original two parental lines (‘04HD5’ and PI 285606) as template DNA primed with 59 markers (39 codominant markers and 20 dominant markers; Fazio et al. 2003; Kong et al. 2006; Ritschel et al. 2004) using the conditions described above (Fazio et al. 2003; Table 1).

Molecular analysis

The markers employed were evaluated for predicted segregation for codominant (1:2:1) and dominant marker (3:1) ratios by goodness of fit testing. The G-test was employed for assessment of marker distributions (Sokal and Rohlf 1994), and the degrees of freedom (df) equaled two in all tests due to the possibility of identifying either homozygosity or heterozygosity at a locus. Homogeneity tests were also performed using the Pearson Chi-square test (Pearson, 1900) with df = (number of lines-1)*(number of classes-1) to test whether allelic frequencies were distributed similarly across markers.

Based on the goodness of fit analyses and the marker consistency during genotyping, 31 codominant markers were deployed for multidimensional scaling analysis of parents and IBL using NTsys (Table 1; Rohlf 1998). A genetic distance matrix for the IBL (117) and parents (2) was calculated using the Simgend procedure in NTsys with Rogers genetic distance formula (Rogers 1972) modified by Wright (Wright 1978). The GD matrix was then used in multidimensional scaling, where results were projected in three dimensions using the 3-D plot option in NTsys (Rohlf 1998). The within and between group genetic distances of selected IBL (based on trait analysis) were calculated by employing Rogers genetic distance (Rogers 1972) modified by Wright (1978) formula using the TFPGA computer software package (Miller, 1997).

Results

Phenotypic analyses of IBL

Variance analyses and mean separation

Statistical analyses (ANOVA; location, block within location, lines, and location-by-lines) of combined data from the USA, The Netherlands, Israel, and Turkey for all traits evaluated are represented in Table 2. The effect of location was significant ($P \leq 0.05$ or 0.01) for all traits, except sex expression. The traits days to anthesis, lateral branch number, and fruit length were significant ($P \leq 0.05$ or 0.01) for block effects. Lines were significantly different ($P \leq 0.05$ or 0.01) for all traits, and location-by-line interactions were detected for all

Table 2 Analysis of variance of traits of Beit Alpha parental cucumber (*Cucumis sativus* L.) line ‘04HD5’, PI 285606, and their derived F_1 and F_2, and inbred backcross progeny (BC_2S_3) when grown in Hancock, Wisconsin; USA, Enkhuizen, the Netherlands; Beit Hanan, Israel; and Antalya, Turkey in the summer of 2008

Trait	Location			Block			Line			Location-by-line		
	df_n^a	df_d^b	F value	df_n	df_d	F value	df_n	df_d	F value	df_n	df_d	F value
Days to anthesis	3	3.5	1032.00**[c]	3	5172	14.73**	123	353.6	2.61**	332	5172	2.99**
Sex expression	3	36	2.59 n.s.	3	5326	0.37 n.s.	123	372	40.55**	350	5326	2.92**
Pistillate flowers per node	3	143	85.47**	3	5325	0.85 n.s.	123	327	3.42**	321	4974	10.39**
Lateral branch number	3	18	215.57**	3	5671	4.35**	123	352	1.82**	346	5671	10.74**
Fruits per plant	3	2.26	91.50**	3	341	2.14 n.s.	123	429	2.42**	353	341	0.74 n.s.
Fruit length	3	3.08	19.61*	3	341	8.07**	123	381	6.17***	336	341	1.08 n.s.
Fruit weight	3	3.97	109.37**	3	341	1.69 n.s.	123	376	1.88**	337	341	1.24*

[a] df_n degrees of freedom for the numerator used in *F*-tests for the combined analysis

[b] df_d degrees of freedom for the denominator used in *F*-tests for the combined analysis

[c] *,**, *n.s.* indicates that the effect is significant at $P \leq 0.05$, $P \leq 0.01$, and not significant, respectively

traits ($P \leq 0.05$ or 0.01), except fruit number and fruit length.

Spearman rank location correlations

Rank correlations between locations are presented in Table 3. Generally, the rankings of lines across locations for all traits examined were similar. However, rank correlations were not significant ($P \leq 0.05$) between Israel and USA for days to anthesis, or between Israel and The Netherlands for pistillate flowers per node. Likewise, differences between Israel and USA, Israel and The Netherlands, and Israel and Turkey for lateral branch number, and between Israel and The Netherlands and between Israel with Turkey for fruit weight were detected. Although significant ($P \leq 0.05$) genotype-by-environment interactions were present for all traits, these interactions are most likely due to differences in trait magnitude by location since the Spearman rank

Table 3 Pearson rank correlations for horticultural traits evaluated in four locations using 117 BC_2S_3 Beit Alpha cucumber (*Cucumis sativus* L.) lines in 2008

	Trait	USA[a]	The Netherlands[b]	Israel[c]	Turkey[d]
USA	Days to anthesis	1	0.37**[e]	0.14 n.s.	0.43**
	Sex expression	1	0.81**	0.76**	0.82**
	Number of pistillate flowers per node	1	0.31**	0.66**	0.74**
	Lateral branch number per plant	1	0.23*	0.06 n.s.	0.22*
	Number of fruits per plant	1	0.23*	0.23*	0.19*
	Average fruit length per plant	1	0.59**	0.51**	0.60**
	Average fruit weight per plant	1	0.30**	0.21*	0.21*
The Netherlands	Days to anthesis	0.37**	1	0.27**	0.35**
	Sex expression	0.81**	1	0.90**	0.88**
	Number of pistillate flowers per node	0.31**	1	0.18 n.s.	0.19*
	Lateral branch number per plant	0.23*	1	0.11 n.s.	0.25**
	Number of fruits per plant	0.23*	1	0.41**	0.44**
	Average fruit length per plant	0.59**	1	0.59**	0.63**
	Average fruit weight per plant	0.30**	1	0.18 n.s.	0.27**
Israel	Days to anthesis	0.14 n.s.	0.27**	1	0.33**
	Sex expression	0.76**	0.90**	1	0.87**
	Number of pistillate flowers per node	0.66**	0.18 n.s.	1	0.60*
	Lateral branch number per plant	0.06 n.s.	0.11 n.s.	1	0.10 n.s.
	Number of fruits per plant	0.23*	0.41**	1	0.60**
	Average fruit length per plant	0.51**	0.59**	1	0.70**
	Average fruit weight per plant	0.21*	0.18 n.s.	1	0.17 n.s.
Turkey	Days to anthesis	0.43**	0.35**	0.33**	1
	Sex expression	0.82**	0.88**	0.87**	1
	Number of pistillate flowers per node	0.74**	0.19*	0.60*	1
	Lateral branch number per plant	0.22*	0.25**	0.10 n.s.	1
	Number of fruits per plant	0.19*	0.44**	0.60**	1
	Average fruit length per plant	0.60**	0.63**	0.70**	1
	Average fruit weight per plant	0.21*	0.27**	0.17 n.s.	1

[a] Hancock, Wisconsin, USA

[b] Enkhuizen, The Netherlands

[c] Beit Hanan, Israel

[d] Antalya, Turkey

[e] **, *, *n.s.*, significant at $P \leq 0.01$, $P \leq 0.05$, and not significant, respectively

correlations between most locations and traits were significant (Table 3).

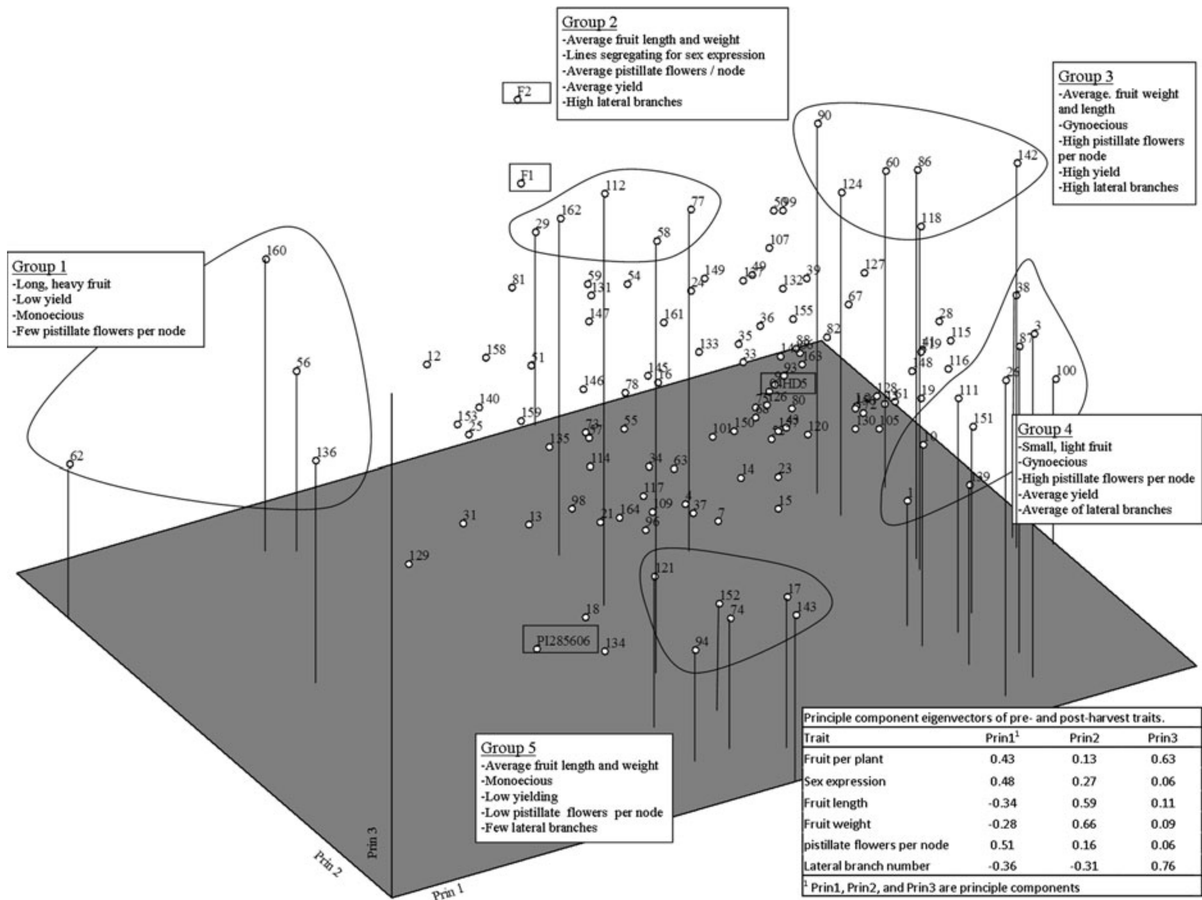
Principal component analysis

Results of PCA for all locations taken are given in Fig. 2. Principle components (PC) 1, 2, and 3 accounted for 36.5, 20.3, and 15.3% of the observed phenotypic variation, respectively (total = 72.1%). The traits pistillate flowers per node, sex expression, and fruit number were factors that accounted for most of the variation in PC 1 (Fig. 2). Likewise, while fruit length and fruit weight accounted for most of the variation in PC 2, lateral branch number and fruit number contributed to most of the variation in PC 3 (Fig. 2).

The most diverse IBL were identified by visual appraisal after PCA, and five phenotypically diverse groups were defined (i.e., at the graphic's periphery; Fig. 2). Group 1 contained IBL 56, 62, 136, and 160; Group 2 consisted of IBL 29, 58, 77. 112, and 162; Group 3 included IBL 60, 86, 90, 118, 124, and 142; Group 4 contained IBL 1, 3, 10, 26, 38, 87, 100, 111, 139, and 151, and; Group 5 consisted of IBL 17, 74, 94, 121, 143, and 152. Trait averages of these IBL by location are depicted in Fig. 3. An ANOVA performed on the IBL determined that these groups differed phenotypically from one another for all traits examined (Delannay 2009).

Trait repeatability measures

The location and line variance estimates and the repeatability measures for days to anthesis, sex expression, pistillate flowers per node, lateral branch number, fruit number, fruit length, and fruit weight



Principle component eigenvectors of pre- and post-harvest traits.

Trait	Prin1[1]	Prin2	Prin3
Fruit per plant	0.43	0.13	0.63
Sex expression	0.48	0.27	0.06
Fruit length	-0.34	0.59	0.11
Fruit weight	-0.28	0.66	0.09
pistillate flowers per node	0.51	0.16	0.06
Lateral branch number	-0.36	-0.31	0.76

[1] Prin1, Prin2, and Prin3 are principle components

Fig. 2 Principle component-based grouping (five unique clusters) of Beit Alpha market class cucumber (*Cucumis sativus* L.) parental lines ('04HD5' and PI 285606), and their derived F_1, F_2, and BC_2S_3 progeny framed by phenotypic descriptors as assessed in the USA, The Netherlands, Israel, and Turkey

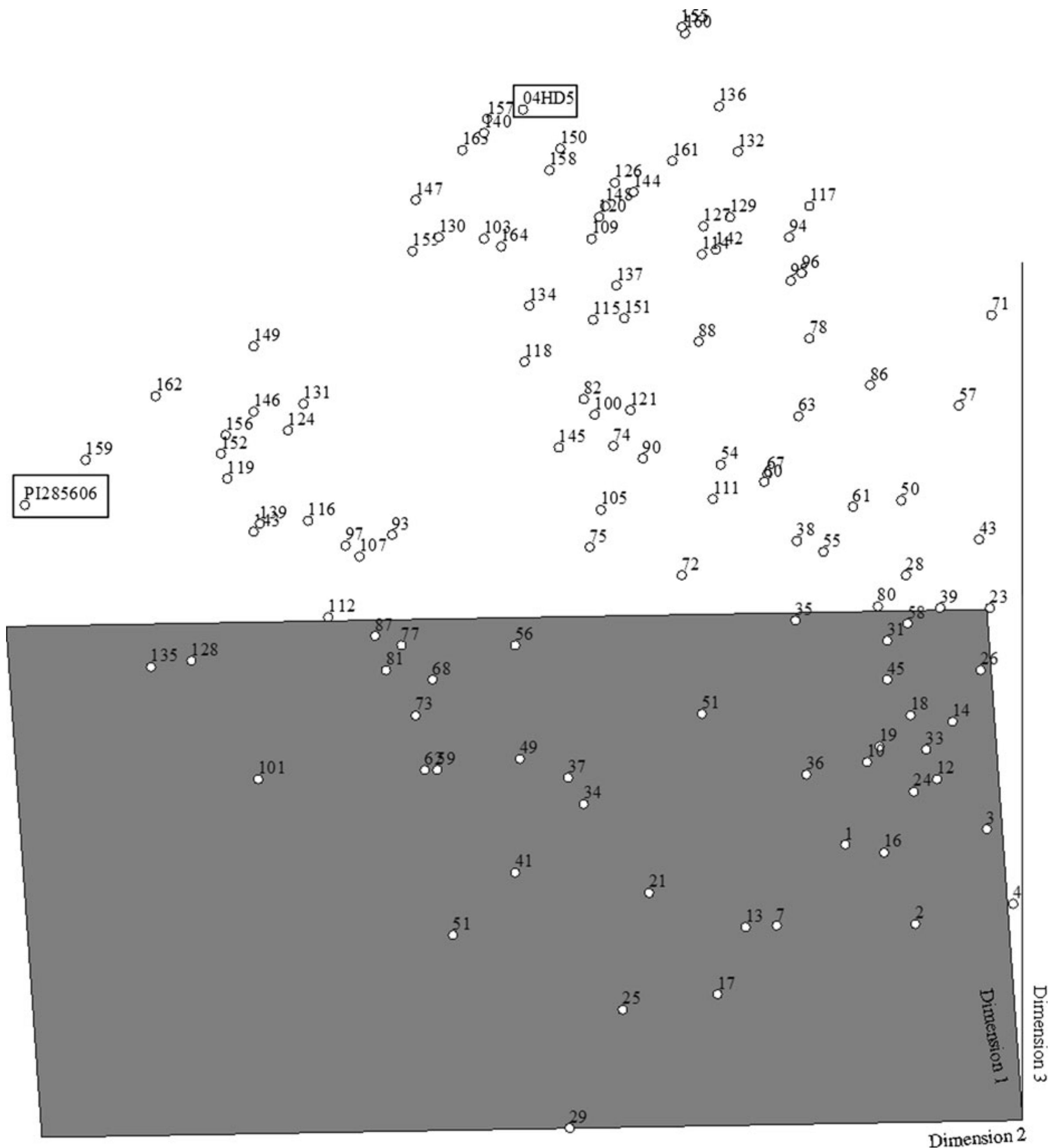


Fig. 3 Relationships among Beit Alpha market class cucumber (*Cucumis sativus* L.) parental lines '04HD5' and PI 285606, and derived BC_2S_3 lines after multidimensional scaling using 31 marker loci (SSR, SCAR, EST-SSR, SNP, and BAC-ends) for Rogers (1972) genetic distance estimation as modified by Wright (1978)

are given in Table 4. Repeatability measures were interpreted as significant for all traits since their values were at least twice that of their standard errors (Halluaer and Miranda 1988). When locations were taken collectively, repeatability was highest for sex expression (0.99 ± 0.13), followed by pistillate flowers per node (0.39 ± 0.0.78), and then fruit length (0.38 ± 0.07). Repeatability of measurements for fruit number (0.14 ± 0.03), fruit weight (0.08 ± 0.03), and days to anthesis (0.00 ± 0.00) were

Table 4 Components of variances (σ^2) and their standard errors and repeatability measures (r) for traits evaluated in inbred backcross (BC_2S_3) Beit Alpha market class cucumber (*Cucumis sativus* L.) lines averaged over four locations (USA, the Netherlands, Israel, and Turkey) in 2008

Trait	σ^{2a}_{loc}	σ^{2b}_{line}	σ^{2c}_{int}	r[d]
Days to anthesis	192.05 ± 157.35	1.60 ± 0.35	3.02 ± 0.31	0.0081 ± 0.002
Days to anthesis without USA[e]	4.97 ± 5.00	1.41 ± 0.39	2.91 ± 0.37	0.15 ± 0.04
Sex expression	0.00048 ± 0.0008	0.54 ± 0.07	0.00 ± 0.00	0.99 ± 0.13
Pistillate flowers per node	0.51 ± 0.42	0.32 ± 0.06	0.00 ± 0.00	0.39 ± 0.07
Lateral branch number per plant	7.91 ± 6.48	0.65 ± 0.22	2.78 ± 0.29	0.06 ± 0.02
Fruit number per plant	7.06 ± 5.79	1.59 ± 0.33	2.57 ± 0.27	0.14 ± 0.03
Average length per fruit	1.73 ± 1.42	1.46 ± 0.27	0.67 ± 0.16	0.38 ± 0.07
Average weight per fruit	788.52 ± 633.66	122.61 ± 40.39	645.42 ± 50.25	0.08 ± 0.03

[a] σ^2_{loc} = variance within location

[b] σ^2_{line} = variance within lines

[c] σ^2_{int} = variance of the genotype by environment interaction

[d] *r* repeatability of trait, where r = the σ^2_{line} divided by σ^2_p and $\sigma^2_p = \sigma^2_{lines} + \sigma^2_{loc} + \sigma^2_{int}$

[e] Days to anthesis using data from all locations except USA. USA excluded due to inconsistancies between locations in original planting

comparatively low. When the USA location was omitted from the analysis, the repeatability of days to anthesis rose to 0.15 ± 0.04 and values for other variables remain relatively constant (Table 4; Delannay 2009).

Molecular analyses of IBL

Allele frequency analysis

Results of the goodness of fit analyses are presented in Table 1. The expected frequencies for IBL being homozygous for recurrent parent (AA), heterozygous (Aa), and homozygous for donor parent (aa) were 0.86, 0.03, and 0.11, respectively. Pearson-based marker homogeneity tests indicated that allelic frequencies at marker loci were independent ($P \leq 0.01$). Only seven of the 38 codominant markers (AC17SNPG1H2, CM49, CS-AT1, CSWGATT01C, CSWTA05, D11SNPG3H1, and W7SNPG1H3) did not deviate from expected frequencies.

Comparisons among parents and IBL groupings

Graphical projection of IBL after multidimensional scaling is given in Fig. 3. The parental lines (PI 285606 and '04HD5') were genetically most distinct (GD = 0.90). Although the majority of IBL were more closely related to parental line '04HD5' than to PI 285606 (average GD = 0.41), IBL 29 (GD = 0.72) was most distant from line '04HD5'. The maximum GD detected (0.86) between entries occurred between IBL 3 and 29. In contrast, the GD between IBL 60 and 111 was estimated as zero.

With regards to PCA groupings, Groups 3 and 4 were the most similar (GD = 0.15), and Groups 1 and 4 were most dissimilar (GD = 0.32). Although the five IBL groups possessed little genetic affinity to the donor parent PI 285606 (GD = 0.66–0.78), genetic affinities with the recurrent parental line '04HD5' were substantial (GD = 0.23–0.35). While Group 3 IBL possessed little genetic affinity with the donor parent (GD = 0.78), Group 2 IBL were most similar to PI 285606 (GD = 0.66). Other IBL possessed varying genetic affinities to PI 285606 [Group 4 (GD = 0.77); Group 1 (GD = 0.70), and; Group 5 (GD = 0.68)].

Comparisons within IBL groupings

In Group 1, IBL 56 and 62 were most similar (GD = 0.32) and IBL 62 and 160 were most dissimilar (GD = 0.62). While Group 2 IBL 77 and 112 (GD = 0.31) were most similar, IBL 29 and 162 (GD = 0.85) were most dissimilar. In Group 3, IBL 60 and 142 were most similar (GD = 0.19), and IBL 86 and 124 (GD = 0.59) were most distinct. While Group 4 IBL 10 and 111 were similar (GD = 0.15),

IBL 3 and 151 (GD = 0.62) were most dissimilar. In Group 5, IBL 143 and 152 were most similar (GD = 0.41) and IBL 74 and 121 (GD = 0.70) were most distinct. When comparing IBL collectively (Groups 1–5), IBL 3 and 29 (GD = 0.86) were most distinct, and IBL 60 and 111 (0.00) were most similar.

Discussion

The data presented herein provide the first comprehensive genetic and phenotypic assessment of Beit-Alpha cucumber, which, in turn, led to the development and genotypic characterization of a wide-based IBL population through MAS. Because of their varying morphologies, genetic affinities, and contrasting allelic constitutions, these unique IBL will likely have utility in genetic studies (e.g., assessment of epistasis) and plant improvement (e.g., increasing genetic diversity) (Robbins et al. 2008).

Five distinct groups of IBL were identified by visual inspection after PCA (Fig. 2), and these differed significantly in plant morphology (Table 4; and Fig. 2). Since these groups diverge appreciably from the remaining IBL, they may be important germplasm pools for use in broadening the genetic base of this comparatively narrow market type.

In general, Group 3 and Group 4 IBL are gynoecious and possess a relatively high number of pistillate flowers per node (2.5–4.1), both of which are main factors used to define IBL orientation in PC 1. In contrast, the monoecious IBL present in Groups 1 and 5 possess comparatively few pistillate flowers per node (~1–2), which provides for their unique orientation as defined by PC 1. The IBL in Group 1 develop long (15.2–18.1 cm), large (143.1–168.3 g) fruit, where length and weight are moderately correlated ($r_s = 0.65$), and support findings of Yuan, et al. (2008), who detected high correlations between fruit weight and fruit length ($r_s = 0.76$–0.94) in Chinese cucumber. The IBL in Group 3 are relatively high yielding (~7–9 fruit/plant) and Group 2 possesses IBL that bear many lateral branches (~4–8). Group 4 IBL are typically gynoecious and possess an above-average number of pistillate flowers per node (2–4), but their yield is only average (5–8 fruits/plant and 90.5–113.1 gm/fruit) and they tend to possess only an average number lateral branches (3–7). Regardless of breeding strategy, the efficient development of Beit Alpha germplasm with improved yield will require the tactical use of genetically diverse, highly inbred lines such as Group 2 (high yield) and Group 3 (increased lateral branches) IBL.

The relative ranks between IBL for most of the traits examined tended to be significant over trial locations in the USA, The Netherlands, and Turkey (Table 3). The performance of some accessions for certain traits, however, was significant across some, but not all, locations (Delannay 2009). All traits, except sex expression, however demonstrated low to moderate repeatability herein, and thus appraisals of GxE interactions are critical during the improvement of Beit Alpha types for differing markets (i.e., USA vs. Europe vs. Middle-East).

Morphological and molecular appraisals of genetic diversity were found to be not equivalent. For instance, while IBL 7 and 143 were found to be phenotypically similar (Fig. 2), they do not share substantial genetic affinities (GD = 0.48) (Delannay 2009). This relationship becomes apparent in Fig. 3, where IBL 7 resides at the base of the graphic projection and IBL 143 is associated with PI 285606. An understanding of such relationships is critical since genetic and morphological appraisals of Beit Alpha cucumber lines will be necessary for the development of breeding strategies that seek to increase genetic diversity while concomitantly fix horticulturally important traits.

Cucumber genetic maps have been constructed using narrow- and broad-based populations (Bradeen et al. 2001; Fazio et al. 2003; Sun et al. 2006a; Yuan et al. 2008). However, genetic mapping of economically important traits has not been performed using Beit Alpha cucumber germplasm. Due to their morphological diversity, the IBL described herein (e.g., Group 3 IBL) may have utility in genetic mapping that seeks to identify quantitative trait loci (QTL) that are difficult to characterize, such as parthenocarpy (Sun et al. 2006a, b) and other quality components (Fazio et al. 2003). MAS used alone or in combination with phenotypic selection has been successfully employed in cucumber breeding to increase lateral branch number and fruit number, and may provide assistance with introgression of yield- and quality-related QTL in Beit Alpha types (Fazio et al. 2003, Fan et al. 2006, Robbins and Staub

2009). The genotypic (heterozygosity) and phenotypic (yield components) characteristics of IBL 29 (Group 2; high fruit number and high lateral branch number), 90 (Group 3; high pistillate flowers per node, high lateral branch number, and high fruit number), and 142 (Group 3; high pistillate flowers per node, high lateral branch number, and high fruit number) are, in fact, indicative of their potential utility for trait introgression (e.g., backcrossing) or base population development (e.g., recurrent selection) during MAS-based improvement of this market class.

Acknowledgements This research project was supported by funding from Nunhems (Haelen, The Netherlands), DeRuiter (Bergschenhoek, The Netherlands), Enza Zaden (Enkhuizen, the Netherlands), and Nickerson-Zwaan (Made, the Netherlands) seed companies. Special thanks to Enza Zaden, Nickerson-Zwaan, and Nunhems for their help with field data collection in The Netherlands, Israel, and Turkey. Mention of trade name, proprietary product, or specific equipment does not constitute a guarantee or warranty by the USDA, and does not imply its approval to theexclusion of other products that may be suitable.

References

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

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. Ph.D. Dissertation, University of Wisconsin at Madison

Dijkhuizen A, Kennard WC, Havey MJ, Staub JE (1996) RFLP variation and genetic relationships in cultivated cucumber. Euphytica 90:79–87

Falconer DS, Mackay TFC (1989) Introduction to quantitative genetics. Benjamin Cummings, San Francisco

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 112:3–855

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 (2003) Genetic mapping and QTL analysis of horticultural traits in cucumber (*Cucumis sativus* L.) using recombinant inbred lines. Theory Appl Genet 107:864–874

Hallauer AR, Miranda JB (1988) Quantitative genetics in maize breeding, 2nd edn. Iowa State University, Ames, Iowa

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

Jaccard P (1908) Nouvelles rescherches sur la distribution florale. Bull Soc Vandoise des Sci Nat 44:223–270

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

Kruskal JB (1964a) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29:1–27

Kruskal JB (1964b) Nonmetric multidimensional scaling: a numerical method. Psychometrika 29:28–42

Meglic V, Staub JE (1996) Genetic diversity in cucumber (*Cucumis sativus* L): 2. An evaluation of selected cultivars released between 1846 and 1978. Genet Resour Crop Ev 43:547–558

Meglic V, Serquen F, Staub JE (1996) Genetic diversity in cucumber (*Cucumis sativus* L): 1. A reevaluation of the US germplasm collection. Genet Resour Crop Ev 43:533–546

Miller MP (1997) Tools for population genetic analyses (TFPGA) 1.3: A windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by author

Nandgaonkar AK, Baker LR (1981) Inheritance of multi-pistillate flowering habit in gynoecious pickling cucumber. J Am Soc Hortic Sci 106:755–757

Pearson K (1900) On a criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Cambridge University Press, Cambridge, UK

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

Rogers JS (1972) Measures of genetic similarity and genetic distance. In: Studies in genetics, vol VII. University Texas Publcation 7213, pp 145–153

Rohlf FJ (1998) NTSYS-Pc v2.0 numerical taxonomy and multivariable analysis system. Applied Biostatistics, New York

SAS (2003) SAS Software, version 9.1 for windows. SAS Institute Inc, Cary, NC

Shaw NL, Cantliffe DJ, Rodriguez ST, Spencer DM (2000) Beit Alpha cucumber: an exciting new greenhouse crop. Proc Fla State Hort Soc 113:247–253

Shaw NL, Cantliffe DJ, Funes J, Shine C (2004) Successful Beit Alpha cucumber production in greenhouse using pine bark as an alternative soilless media. HortTechnology 14:289–294

Shaw NL, Cantliffe DJ, Stoffella PJ (2007) A new crop for North American greenhouse growers: Beit Alpha cucumber: progress of production technology through university research trials. Acta Hortic 731:251–255

Sokal RR, Rohlf FJ (1994) Biometry: the principles and practice of statistics in biological research, 3rd edn. Freeman, New York

Staub JE, Knerr LD, Holder DJ, May B (1992) Phylogenetic relationships among several African *Cucumis* species. Can J Bot 70:509–517

Staub JE, Serquen FC, McCreight JD (1997) Genetic diversity in cucumber (*Cucumis sativus* L.): 3. An evaluation of Indian germplasm. Genet Resour Crop Ev 46:315–326

Staub JE, Serquen FC, Horejsi T, Chen JF (1999) Genetic diversity in cucumber (*Cucumis sativus* L.): IV. An evaluation of Chinese germplasm. Genet Resour Crop Ev 46:297–310

Staub JE, Chung SM, Fazio G (2005) Conformity and genetic relatedness estimation in crop species having a narrow genetic base: the case of cucumber (*Cucumis sativus* L.). Plant Breeding 124:44–53

Sun Z, Staub JE, Chung SM, Lower RL (2006a) Identification and comparative analysis of quantitative trait loci associated with parthenocarpy in processing cucumber. Plant Breeding 125:281–287

Sun ZY, Lower RL, Staub JE (2006b) Variance component analysis of parthenocarpy in elite US processing type cucumber (*Cucumis sativus* L.) lines. Euphytica 148: 331–339

Tanksley SD, Grandillo S, Fulton TM, Zamir D, Eshed Y, Petiard V et al (1996) Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L. pimpinellifolium*. Theor Appl Genet 92:213–224

Villalta AM, Sargent SA, Berry AD, Huber DJ (2003) Sensitivity of Beit Alpha cucumber (*Cucumis sativus* L.) to low temperature storage. Proc Fla State Hort Soc 116:364–366

Wehrhahn C, Allard RH (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

Wright S (1978) Evolution and the genetics of populations. Variability within and among natural populations. U. Chicago Press, Chicago, Illinois

Yuan X, Pan J, Cai R, Guan Y, Liu L, Zhang W et al (2008) Genetic mapping and QTL analysis of fruit and flower related traits in cucumber (*Cucumis sativus* L.) using recombinant inbred lines. Euphytica 164:473–491