Association mapping, trait variation, interaction and population structure analysis in cucumber (Cucumis sativus L.)

Rahul Kumar · Anilabh Das Munshi · Tusar Kanti Behera · Gograj Singh Jat · Harshawardhan Choudhary · Akshay Talukdar · Prasanta Dash · Deepak Singh

Abstract In several regions of the world, low productivity in this crop is attributed to several factors including poor understanding of the genomic complexity of important traits associated with fruit quality and yield. Therefore, genome wide association analysis was performed for important traits using simple sequence repeats (SSR) markers. Significant variation was recorded for all the studied traits in 78 cucumber genotypes under two environments (open field and net house) which indicated that the constituted association panel was suitable for association mapping. Genotyping was done using 60 highly polymorphic SSRs. By performing genome scanning out of 60 SSR markers, using mixed linear model (MLM) approach 4 and 6 markers explained an average of 23.93% and 17.37% of the trait variation under net house and open field condition, respectively. Based on MLM approach two markers on 3rd chromosome (UW084942) and 4th chromosome (UW062953) found to be associated with the average fruit weight (g) under both net house and open field condition. Population structure analysis revealed four distinct sub-populations that corroborated with the geographical origin as well as fruit quality and quantitative traits. The four sub-populations (A–D) had fixation index percentage equal to 24.35 29.48, 37.17 and 8.97 respectively, supporting the existence of moderate population structures. Therefore, the extensive phenotypic and genotypic characterization, population structure, and markers associated with important traits provided in this study will facilitate marker assisted improvement programs in cucumber.

Keywords Cucumis sativus · Marker–trait association · Population structure · SSR markers · Fruit traits

Abbreviations
MLM Mixed linear model
SSR Simple sequence repeats
Fst Fixation index percentage
MAS Marker assisted selection
QTL Quantitative trait loci
LD Linkage disequilibrium
**Introduction**

Cucumber (*Cucumis sativus* L. 2n = 2x = 14) is one of the most important member of the family Cucurbitaceae. Wide variation for most of the traits including growth habit, sex expression, fruit size and colour were recorded in Indian germplasm (Sebastian et al. 2010). As a result of continuous selection on the basis of local preferences, a large number of landraces and feral forms have been accumulated in different growing regions throughout the world (Grumet et al. 2021). Most of the commercially important traits of cultivated crops are quantitatively inherited and hence, genes responsible for inheritance of these traits are difficult to detect. Cucumber has a narrow genetic base with 3–12% polymorphism (Staub et al. 2005; Yang et al. 2013). Cucumber is considered as a model plant for genetic research because of its smaller genome size (367 Mb), rich diversity of gene expression and short life cycle (Ren et al. 2009). A large number of SSR markers from whole genome sequences have been developed (Cavagnaro et al. 2010; Ren et al. 2009). Draft genomes of three cucumber lines (‘9930’ ‘Gy14’ and ‘B10’) are now available and they are instrumental in understanding the structure and organization of cucumber genome (Yang et al. 2013).

The genetic variance for important horticultural yield traits is directly related to the genetic distance between the parents. Genome mapping has also indicated that the frequency of polymorphism increases progressively with increased genetic distance between the parents (Serquen et al. 1997). In order to use marker-assisted selection (MAS) effectively, the saturated linkage maps (5–10 cM) must be constructed and judiciously applied to maximize cost–benefit ratio (Staub et al. 2005). SSRs were used in genetic mapping, MAS, genetic diversity (Dar et al. 2017; Jat et al. 2018; Kumar et al. 2020) and population structure analysis in cucumber (Dar et al. 2017). The sequencing of the whole cucumber genome (Huang et al. 2009) made it possible to use more breeder friendly molecular markers such as SSRs for genetic mapping and MAS in cucumber. The advantages of these cucumber SSR markers have already been reported in several recent linkage mapping and marker-trait association studies (Zhang et al. 2010; Weng et al. 2010). But only a few important genes or QTLs of cucumber have been mapped with these markers (Yuan et al. 2008a, b; Miao et al. 2011).

Presently, association mapping approach is used in identifying major QTLs and candidate genes for array of economically important traits in different crop species. This approach offers unique opportunities to exploit genetic variation in natural populations with high-resolution mapping of complex agronomic traits (Zhu et al. 2008). Association mapping relies on the linkage disequilibrium (LD) which is maintained over generations between loci which are genetically linked to one another (Neumann et al. 2011). It has been identified as a tool to resolve complex trait variations down to the sequence level (Nordborg and Tavare 2002), and to identify novel mutations causing specific phenotypes (Palaisa et al. 2004). SSRs are used quite often in identifying complex traits in different crop species.

Apart from the use of SSR markers, several analytical advances are available for the identification of loci governing particular traits. These advances include the stratification of population structure and statistical models such as generalized linear model (GLM) and mixed linear model (MLM) approaches. A unified, mixed-model approach for association mapping combined with a population structure analysis is a dependable and robust system for identifying reliable quantitative trait loci (Yu and Buckler 2006).

The present study focused on analyzing the population structure of 78 cucumber genotypes and identifying linked molecular markers for earliness and yield traits. These marker-QTL associations identified in our study would be immensely useful in cucumber improvement through MAS.

**Materials and methods**

**Plant materials**

The experimental material for the present study consisted of 78 diverse cucumber genotypes comprising of Indian as well as exotic germplasm maintained at Division of Vegetable Science, ICAR-IARI, New Delhi (Supplementary Table 1).
Phenotyping of the germplasm

These 78 genotypes were evaluated in net house during rainy season (August to November, 2017) and for two consecutive seasons under open field condition i.e. rainy season (August to November, 2017) and summer season (February to May, 2018). Out of total 15 plants grown in each genotype, 5 plants were randomly chosen and tagged to record data of 10 important quantitative traits under open field condition (Table 1). However, under net house we were able to record data on 8 traits (Table 2). Data of 2 seasons recorded under open field condition were pooled for analysis. Data of all the traits were recorded with three replications using a randomized block design.

Genomic DNA extraction

From fresh young leaves of the 78 diverse cucumber genotypes genomic DNA was extracted following a modified CTAB method (Doyle and Doyle 1990). DNA quality was checked through 0.8% agarose gel electrophoresis and quantity by using spectrophotometer at wavelength of 260 nm, and stored at −20 °C for future use.

PCR amplification

Sixty SSR markers (Supplementary Table 2) distributed across the cucumber genome were used in this study (Yang et al. 2013). The PCR reactions were carried out in a volume of 20 μl reaction mixture each containing 2 μl of 10× reaction buffer, 0.30 μl of 10 mM dNTPs, 1 μl each of forward and reverse primers, 2.0 μl of template genomic DNA (50 ng), 0.2 μl of Taq DNA polymerase (0.75 U) and 13.5 μl molecular grade water. The PCR thermal profiles were: DNA denaturation at 94 °C for 4 min followed by 35 cycles of 94 °C for 30 s, annealing temperature at 52–59 °C for 1 min and extension at 72 °C for 1 min, finally 72 °C for a final extension of 10 min and cooling at 4°C. The PCR products were resolved by electrophoresis in 3.0% agarose gels containing 0.1 μg/ml ethidium bromide in 1× TBE buffer at 120

Table 1  Summary statistics of horticultural traits evaluated in 78 cucumber genotypes under open field condition

| Variable                                      | Mean  | SD   | Minimum | Maximum | Variance |
|------------------------------------------------|-------|------|---------|---------|----------|
| Node number for first female flower            | 8.00  | 1.46 | 5.22    | 10.78   | 2.14     |
| Days taken for first female flower to anthesis | 51.00 | 3.24 | 39.50   | 58.81   | 10.48    |
| Number of fruits per plant                     | 6.04  | 1.22 | 4.28    | 11.93   | 1.49     |
| Days taken for first fruit harvest             | 61.94 | 4.56 | 47.65   | 73.26   | 20.80    |
| Fruit length (cm)                              | 12.49 | 2.83 | 7.20    | 22.87   | 8.01     |
| Fruit diameter (cm)                            | 5.13  | 0.50 | 4.03    | 6.39    | 0.25     |
| Fruit length/diameter ratio                    | 2.47  | 0.65 | 1.39    | 5.21    | 0.42     |
| Average fruit weight (g)                       | 135.08| 56.65| 63.35   | 347.22  | 3209.18  |
| Yield per plant (g)                            | 807.52| 354.27| 400.76 | 2136.00 | 125,503.99|
| Vine length (cm)                               | 148.79| 14.33| 80.48   | 176.37  | 205.35   |

Table 2  Summary statistics of horticultural traits evaluated in 78 cucumber genotypes under net house condition

| Variable                                      | Mean  | SD   | Minimum | Maximum | Variance |
|------------------------------------------------|-------|------|---------|---------|----------|
| Node number for first female flower            | 5.51  | 1.05 | 3.19    | 7.74    | 1.09     |
| Days taken for first female flower to anthesis | 46.70 | 3.76 | 35.55   | 59.06   | 14.12    |
| Days taken for first fruit harvest             | 58.00 | 3.94 | 46.05   | 64.49   | 15.49    |
| Fruit length (cm)                              | 15.29 | 4.77 | 9.73    | 35.29   | 22.74    |
| Fruit diameter (cm)                            | 5.26  | 0.53 | 3.85    | 6.89    | 0.28     |
| Fruit length/diameter ratio                    | 2.94  | 1.02 | 1.75    | 7.66    | 1.04     |
| Average fruit weight (g)                       | 179.96| 93.58| 59.37   | 487.89  | 8756.49  |
| Vine length (cm)                               | 145.95| 14.57| 78.24   | 164.47  | 212.34   |
Volts for 3 h. After electrophoresis DNA fragments were visualized and documented using ALPHA IMAGER gel documentation system (Alpha Innotech, USA).

Statistical analysis

The data were analyzed using the ANOVA procedure of SAS 9.2 (SAS Institute, Cary, NC, USA) to derive their summary statistics including mean, range, standard deviation, variance, and coefficient of variation. The positive and negative correlations among the yield and related traits in 78 cucumber genotypes were measured using Pearson correlation coefficient at 1 and 5% levels of significance in SPSS for Windows. The fixed effect model was used for combined analysis of variance for 8 phenotypic traits including genotype (G) × environment (E) interaction for both the environments (net house and open field). Genetic and environmental effects as well as the genetic variation among the genotypes were measured according to the various genetic parameters developed by Burton (1951), Burton and DeVane (1953) and Johnson et al. (1955).

For population structure analysis, we assumed an admixed model with a uniform prior probability and independent allele frequency of the number of populations, K. All the runs with 50,000 MCMC replicates after a burn-in of 50,000 replicates were conducted for K = 2–10. Five independent runs were performed for each value of K to generate our estimate of the true number of sub-populations as reported by Pritchard et al. (2000). The ‘Structure harvester’ program (http://taylor0.biology.ucla.edu) was used to obtain the final K value (K = 4 was optimum for this analysis), based on both the LnP (D) and Evanno’s DK (Evanno et al. 2005; Earl et al. 2012). Markers which were exhibiting a probability value (p value) less than 0.02 thresholds were considered significantly associated with the particular phenotypic trait. The association of each alleles of the locus with the trait of interest studied under experiment was tested using two approaches viz. GLM and MLM (Yu et al. 2006), using the TASSEL v2.0.1 software (Bradbury et al., 2007). The MLM approach exhibited the least variation of the Q (population structure) or K (kinship) model in GLM approach.

Results

Phenotypic evaluation and correlation among the important traits

The combined analysis of variance indicated significant effects of environment, genotypes and genotypes × environment (GE) interactions on 8 quantitative traits studied under net house and open field condition (Supplementary Table 3). Summary statistics including minimum, maximum and mean values of earliness and yield traits studied under open field condition (10 traits; 2 seasons mean) and net house (8 traits; 1 season mean) are given in the Tables 1 and 2 respectively. The superior performance for earliness and yield traits were recorded under net house grown crop compared to open field condition.

The correlations among the important earliness and yield traits are presented in Tables 3 and 4. Eight traits studied under net house showed node number for first female flower was positive and highly significantly correlated with days taken for first female flower to anthesis, days taken for fruit harvest and with vine length. Days taken for first female flower to anthesis was also had positive significant correlation with days taken for first fruit harvest and with vine length. Fruit length was positively correlated with the fruit length/diameter ratio, average fruit weight and fruit yield per plant. Average fruit weight had positive and highly significant correlation with vine length. Fruit diameter showed significant and negative correlation with fruit length/diameter ratio (Table 3).

Traits studied under open field condition also showed positive and highly significant correlation with the traits like days taken for first female flower anthesis with days taken for first fruit harvest and with vine length while with number of fruits per plant it was negatively correlated. Number of fruits per plant was positively correlated with the yield per plant. Fruit length showed significantly high and positive correlation with fruit length diameter ratio, average fruit weight, yield per plant and vine length. Average fruit weight showed high positive and significant
| Traits                                      | Node number for first female flower | Days taken for first female flower to anthesis | No of fruits per plant | Days taken for first fruit harvest | Fruit length (cm) | Fruit diameter (cm) | Fruit length/diameter ratio | Average fruit weight (g) | Yield per plant (g) | Vine length (cm) |
|---------------------------------------------|-------------------------------------|-----------------------------------------------|------------------------|-----------------------------------|-------------------|---------------------|--------------------------|------------------------|---------------------|------------------|
| Node number for first female flower         | 1.00                                |                                               |                        |                                   |                   |                     |                          |                        |                     |                  |
| Days taken for first female flower to anthesis | 0.24**                             | 1.00                                          |                        |                                   |                   |                     |                          |                        |                     |                  |
| Number of fruits per plant                  | 0.04NS                              | −0.37**                                       | 1.00                   |                                   |                   |                     |                          |                        |                     |                  |
| Days taken for first fruit harvest          | 0.17**                             | 0.518**                                       | −0.36**                | 1.00                              |                   |                     |                          |                        |                     |                  |
| Fruit length (cm)                           | 0.13*                              | 0.28**                                        | −0.11NS                | 0.16*                             | 1.00              |                     |                          |                        |                     |                  |
| Fruit diameter (cm)                         | 0.05NS                              | −0.06NS                                       | 0.28**                 | 0.03NS                            | −0.17**           | 1.00                |                          |                        |                     |                  |
| Fruit length/diameter ratio                 | 0.12NS                              | 0.25**                                        | −0.20**                | 0.12NS                            | 0.93**            | −0.49**             | 1.00                     |                        |                     |                  |
| Average fruit weight (g)                    | 0.07NS                              | 0.29**                                        | −0.11NS                | 0.25**                            | 0.69**            | −0.03NS             | 0.62**                   | 1.00                   |                     |                  |
| Yield per plant (g)                         | 0.07NS                              | 0.15*                                         | 0.28**                 | 0.11NS                            | 0.64**            | 0.11NS              | 0.52**                   | 0.91**                 | 1.00                |                  |
| Vine length (cm)                            | 0.30**                             | 0.51**                                        | −0.22**                | 0.47**                            | 0.29**            | 0.21**              | 0.16**                   | 0.36**                 | 0.26**              | 1.00             |

*Significant at 1% level of confidence and ** Significant at 5% level of confidence. NS=Non-significant
Table 4  Pearson correlation coefficient among horticultural traits evaluated in 78 cucumber genotypes under open field condition

| Traits | Node number for first female flower | Days taken for first female flower to anthesis | No of fruits per plant | Days taken for first fruit harvest | Fruit length (cm) | Fruit diameter (cm) | Fruit length/diameter ratio | Average fruit weight (g) | Yield per plant (g) | Vine length (cm) |
|--------|-------------------------------------|-----------------------------------------------|------------------------|-----------------------------------|------------------|-------------------|--------------------------|--------------------------|------------------|-----------------|
| Node number for first female flower | 1.00 | 0.19** | 0.01NS | 0.15* | 0.12NS | 0.02NS | 0.11NS | 0.07NS | 0.05NS | 0.27** |
| Days taken for first female flower to anthesis | 1.00 | 0.45** | − 0.31** | − 0.32** | 0.24** | − 0.02NS | 0.19** | − 0.10NS | 0.26** | 0.13* |
| Number of fruits per plant | 0.01NS | 1.00 | 0.45** | − 0.32** | 0.24** | − 0.02NS | 0.19** | − 0.10NS | 0.26** | 0.13* |
| Days taken for first fruit harvest | 0.15* | 0.45** | − 0.31** | 1.00 | 0.12NS | 0.02NS | 0.11NS | 0.07NS | 0.05NS | 0.27** |
| Fruit length (cm) | 0.19** | 0.45** | − 0.31** | 1.00 | 0.24** | − 0.02NS | 0.19** | − 0.10NS | 0.26** | 0.13* |
| Fruit diameter (cm) | 0.01NS | − 0.31** | 1.00 | 0.16* | 0.10NS | 0.18** | 0.10NS | 0.88** | 0.46** | 0.11NS |
| Fruit length/diameter ratio | 0.01NS | − 0.31** | 1.00 | 0.16* | 0.10NS | 0.18** | 0.10NS | 0.88** | 0.46** | 0.11NS |
| Average fruit weight (g) | 0.01NS | − 0.31** | 1.00 | 0.16* | 0.10NS | 0.18** | 0.10NS | 0.88** | 0.46** | 0.11NS |
| Yield per plant (g) | 0.01NS | − 0.31** | 1.00 | 0.16* | 0.10NS | 0.18** | 0.10NS | 0.88** | 0.46** | 0.11NS |
| Vine length (cm) | 0.01NS | − 0.31** | 1.00 | 0.16* | 0.10NS | 0.18** | 0.10NS | 0.88** | 0.46** | 0.11NS |

*Significant at 1% level of confidence and ** Significant at 5% level of confidence. NS=Non-significant
correlation with yield per plant as well as with vine length. Fruit length was negatively correlated with fruit diameter. Fruit diameter showed negative correlation with fruit length/diameter ratio (Table 4).

Under both environments correlation among traits like node number for first female flower, days taken to first female flower anthesis and vine length was positive and significantly high. Fruit length was positively correlated with fruit length/diameter, average fruit weight and with vine length while negative correlation was found with fruit diameter in both the environment. Average fruit weight was positively correlated with vine length. These wide phenotypic trait variations under both the environments among 78 cucumber genotypes indicated that the constituted association panel was suitable for association mapping for earliness and yield traits.

Variability and heritability analysis

Heritability estimate is an important parameter to the breeder for selecting the desired genotypes for further use. The genetic parameters of variability estimates such as phenotypic coefficient of variation (PCV), the genotypic coefficient of variation (GCV), broad sense heritability, and genetic advance as percent of means are presented in Table 5. The GCV values were smaller than PCV values for the traits like node number of first female flower, days taken for first female flower to anthesis, days taken for first fruit harvest and fruit length showing that these characters were more influenced by their surrounding environments. The high (> 20%) GCV values were observed for node number of first female flower (23.49), fruit length (23.33), fruit length/diameter ratio (24.95) and average fruit weight (32.38) where as low (0–10%) GCV values were observed for all other traits studied. The lowest GCV was observed for vine length (5.47). Among the studied traits, broad sense heritability ranged from 0.68 for fruit diameter to 0.95 for node number of first female flower. Heritability is grouped as low (<0.2), moderate (<0.2–0.4) and high (>0.4). In our study all traits were found highly heritable traits. High heritability estimates showed a high response to selection for those traits.

Population structure analysis

A total of 60 highly polymorphic SSR markers were used in determining the population structure and marker–trait association analysis. The Bayesian model-based program STRUCTURE 2.2 was used

| Traits | Node number of first female flower | Days taken for first female flower to anthesis | Days taken for first fruit harvest | Fruit length (cm) | Fruit diameter (cm) | Fruit length/diameter ratio | Average fruit weight (g) | Vine length (cm) |
|--------|-----------------------------------|---------------------------------------------|---------------------------------|------------------|-------------------|-------------------------|----------------------|----------------|
| EV ($\sigma^2_E$) | 0.14 | 1.06 | 1.76 | 1.07 | 0.05 | 0.06 | 652.35 | 20.95 |
| ECV | 13.47 | 5.17 | 5.42 | 18.32 | 10.79 | 23.07 | 39.75 | 7.60 |
| GV($\sigma^2_G$) | 2.51 | 11.79 | 13.14 | 10.46 | 0.11 | 0.46 | 2598.02 | 65.13 |
| GCV (%) | 23.49 | 7.03 | 6.04 | 23.33 | 6.40 | 24.95 | 32.38 | 5.47 |
| PV ($\sigma^2_P$) | 2.65 | 12.86 | 14.91 | 11.53 | 0.16 | 0.52 | 3250.37 | 86.08 |
| PCV (%) | 24.13 | 7.34 | 6.44 | 24.50 | 7.77 | 26.67 | 36.22 | 6.29 |
| $h^2_B$ (%) | 0.94 | 0.92 | 0.88 | 0.91 | 0.68 | 0.87 | 0.80 | 0.76 |
| GA at 5% | 3.18 | 6.77 | 7.01 | 6.34 | 0.56 | 1.31 | 93.87 | 14.46 |
| GAat 1% | 4.07 | 8.68 | 8.99 | 8.13 | 0.72 | 1.67 | 120.30 | 18.53 |
| GA as percent of mean at 5% | 47.19 | 13.87 | 11.69 | 45.77 | 10.87 | 48.08 | 59.64 | 9.80 |
| GA as percent of mean at 1% | 60.38 | 17.77 | 14.98 | 58.66 | 13.93 | 61.62 | 76.43 | 12.56 |

$EV$ environmental variance, $ECV$ environmental coefficient of variation, $GV$ genotypic variance, $GCV$ genotypic coefficient of variation, $PV$ phenotypic variance, $PCV$ phenotypic coefficient of variation, $h^2_B$, broad sense heritability, $GA$ genetic advance
to infer the population structure. Admixture model-based simulations were carried out by varying K from 2 to 10 with ten runs for each K, using all 78 cucumber genotypes which revealed evident knees at K = 4. The average LnP (D) (log-likelihood) value increased continuously with the increase in K from 2 to 10. However, its most apparent inflection was obtained at one of the best replicates of K = 4. The results of population numbers (K) were further confirmed using DK estimation. A sharp peak with the maximum value of DK was obtained at K = 4, thereby confirming the classification of 78 cucumber genotypes into four distinct population groups (Fig. 1). Using this approach, 78 cucumber genotypes were assigned to the corresponding A–D subpopulations, representing 24.35% (19), 29.48% (23), 37.17% (29) and 8.97% (7) of the genotypes (Table 6). The four sub-populations (A–D) had fixation index percentage (Fst) equal to 24.35, 29.48, 37.17 and 8.97 respectively, supporting the existence of moderate population structures.

Table 6 Significant divergence between groups (subpopulations) and average distances (expected heterozygosity) between cucumber populations

| Population | Percent mean value of Fst within population | Heterozygosity (He) | Number of genotypes |
|------------|--------------------------------------------|--------------------|---------------------|
| Fst_A      | 24.35                                      | 0.27               | 19                  |
| Fst_B      | 29.48                                      | 0.28               | 23                  |
| Fst_C      | 37.17                                      | 0.25               | 29                  |
| Fst_D      | 8.97                                       | 0.20               | 7                   |

Furthermore, the genetic distances among these four sub-populations by Fst value were also measured which showed the variable level of genetic differentiation between inferred populations. The Fst value ranged from 0.129 to 0.137 with an average 0.133, revealing smallest genetic distances between sub-pop A and C the largest between sub-population C and D (Table 7). Genotypes with more than 0.80 Fst considered as pure and less than 0.80 as admixture (Singh et al. 2016). There were mixed proportions of accesses with admixture with parental varieties in sub-populations defined by the structure, which might have grouped in different defined clusters. In the sub-population A, the percentage of mixed individuals was found to be 21.05%, likewise 26.08% for B, and 17.24% for C while sub population D consisted of 100% pure genotypes. Based on DK estimation cucumber genotypes were divided into four distinct populations. The first group (A) consisted of most of the genotypes collected from different districts of West Bengal along with four genotypes from IARI, New Delhi. This group also contains two genotypes.

![Model based population structure plot for each variety with K = 4, using Structure software with 60 polymorphic SSR loci. Color codes are as follows: Population A red, Population B green, Population C blue and Population D yellow. The single vertical line represents an individual variety and different segments of each vertical line show extent of admixture in an individual. (Color figure online)](image-url)
from China and one from USA. The second group (B) comprised of mixed 23 genotypes collected mainly from IARI, New Delhi, along with one genotype from USA and two from China. The third group (C) consisted of 29 mixed genotypes from different parts of India as well as abroad therefore being the largest group. This group mainly consisted of National Bureau of Plant Genetic Resources (NBPGR), New Delhi (8) and IARI, New Delhi (10) genotypes. The four groups (D) being smallest among all containing 7 genotypes. Genetic distances between different populations from structure analysis revealed that group D and C had highest cluster distance while group B and C had least distance (Table 8). Study revealed that genotypes from IARI, New Delhi are distantly related to that of genotypes present in group D. From the structure analysis it was clear that IARI, New Delhi genotypes are least variable among themselves as group (B) consisted of 8 genotypes while group (C) consisted of 10 genotypes of IARI, New Delhi.

Association analysis

Marker–trait association was performed for important earliness and yield traits in two different environments using MLM approach in the TASSEL software. By performing genome scanning through MLM approach, 4 and 6 SSR markers were associated with 8 and 10 earliness and yield traits studied under net house and open condition respectively (Tables 9 and 10). In combination, all four and six markers explained an average of 23.93% and 17.37% of the trait variation, respectively. Among all 8 traits studied under net house, UW084942 and UW062953 makers were associated with average fruit weight (g) with 29 and 15% of the phenotypic variance, respectively. While days taken to first fruit harvest and fruit diameter (cm) were associated with marker UW085408 and UW083899, respectively. Ten traits studied under open field condition revealed that a single marker UW084942 was associated with two traits namely average fruit weight (g) and fruit length (cm) while marker UW062953 separately associated with the average fruit weight (g). Other traits like number of fruits per plant, days taken to first female flower anthesis, vine length (cm) and days taken for first female flower to anthesis were associated singly with following markers UW085408, SSR06011, SSR03552 and UW084541, respectively.

The Manhattan plots of marker UW084942, common for traits studied under net house and open field condition, in which the genomic coordinates were displayed along the X-axis and the negative logarithm of the association p value for each SSR are displayed on the Y-axis, clarify that each dot on the Manhattan plot depicts a SSRs. The Manhattan plot displaying the significant association of UW084942 with average fruit weight (g) in both the approaches under two environment viz. under net house and open field condition (Fig. 2a, b).

On the other hand, the results of MLM-QQ plots clearly showed nearly perfect distribution of scores with slight deviation (Fig. 3a, b). This reveals the significance of MLM scans for the identification of candidate genes for such complex traits like average fruit weight (g). Therefore, in this analysis the evaluation of QQ plots suggests the importance of test of both general and mixed linear models in genome wide association mapping.

Discussion

In this study, 78 cucumber genotypes were collected from 17 locations across India and abroad for marker trait association. The morphological traits revealed a considerable variability in flowering behavior, growth habit and fruit characters. Node number to first female flower which determines the earliness of a genotype showed significant variation under both environments. Identification of genotypes with lower node to female flower would be useful in extending the availability of fruit. Judicious planting of early and late genotypes will help in sustainable marketing of cucumber cultivars for a longer period. Very wide range of diversity was recorded for fruit weight and fruit diameter under both the growing condition. Fruit length, fruit diameter and average fruit weight are considered as major yield contributing traits. For slicing cucumber, thin and long fruits are desirable and small fruited genotypes may be desirable for processing industry. In the present study wide range of variation was recorded for most of the other traits. Wide variations for important traits in cucumber were earlier reported by several workers (Mliki et al. 2003; Pandey et al. 2013; Kumar et al. 2020).
Table 8  List of cucumber genotypes used as association mapping panel with their source in which they are grouped

| Sl. no. | Genotypes     | Origin          | Fst1   | Fst2   | Fst3   | Fst4   | Population | Admixture/Pure |
|---------|---------------|-----------------|--------|--------|--------|--------|------------|-----------------|
| 1       | 702-6-B76     | China           | 0.016  | 0.032  | 0.949  | 0.003  | C          | Pure            |
| 2       | AS 1          | Karnataka       | 0.011  | 0.006  | 0.978  | 0.005  | C          | Pure            |
| 3       | Barapata Sasha| West Bengal     | 0.025  | 0.022  | 0.944  | 0.01   | C          | Pure            |
| 4       | BarshaMagal   | West Bengal     | 0.051  | 0.006  | 0.939  | 0.003  | C          | Pure            |
| 5       | BC 1          | Karnataka       | 0.007  | 0.932  | 0.056  | 0.005  | B          | Pure            |
| 6       | CCUH-3        | Uttarakhand     | 0.008  | 0.978  | 0.012  | 0.002  | B          | Pure            |
| 7       | CCUH-5        | Uttarakhand     | 0.951  | 0.02   | 0.016  | 0.013  | A          | Pure            |
| 8       | CL 746        | China           | 0.31   | 0.024  | 0.655  | 0.011  | A          | Admixture       |
| 9       | CL 758        | China           | 0.012  | 0.82   | 0.165  | 0.002  | B          | Pure            |
| 10      | CL 773        | China           | 0.94   | 0.04   | 0.013  | 0.007  | A          | Pure            |
| 11      | DARL-106      | Uttarakhand     | 0.008  | 0.98   | 0.01   | 0.002  | C          | Pure            |
| 12      | DC 21         | Western Uttar Pradesh | 0.944  | 0.019  | 0.032  | 0.005  | A          | Pure            |
| 13      | DC 22         | Western Uttar Pradesh | 0.007  | 0.008  | 0.979  | 0.006  | C          | Pure            |
| 14      | DC-102        | Western Uttar Pradesh | 0.003  | 0.968  | 0.017  | 0.012  | B          | Pure            |
| 15      | DC-103        | Western Uttar Pradesh | 0.072  | 0.016  | 0.91   | 0.002  | C          | Pure            |
| 16      | DC106         | Western Uttar Pradesh | 0.025  | 0.616  | 0.354  | 0.004  | B          | Admixture       |
| 17      | DC-27         | Western Uttar Pradesh | 0.021  | 0.048  | 0.914  | 0.017  | C          | Pure            |
| 18      | DC-505        | Western Uttar Pradesh | 0.313  | 0.245  | 0.413  | 0.029  | C          | Admixture       |
| 19      | DC92-1        | Western Uttar Pradesh | 0.007  | 0.023  | 0.964  | 0.006  | C          | Pure            |
| 20      | DG-4          | Karnataka       | 0.449  | 0.159  | 0.371  | 0.022  | A          | Admixture       |
| 21      | Dharwad Green | Karnataka       | 0.003  | 0.046  | 0.008  | 0.944  | D          | Pure            |
| 22      | Dofasli       | Western Uttar Pradesh | 0.9    | 0.007  | 0.05   | 0.043  | A          | Pure            |
| 23      | EC 636504     | USA             | 0.017  | 0.967  | 0.005  | 0.011  | B          | Pure            |
| 24      | EC 636505     | USA             | 0.158  | 0.01   | 0.827  | 0.005  | C          | Pure            |
| 25      | EC 636506     | USA             | 0.023  | 0.738  | 0.031  | 0.207  | B          | Admixture       |
| 26      | EC 636509     | USA             | 0.278  | 0.095  | 0.621  | 0.007  | C          | Admixture       |
| 27      | EC 636510     | USA             | 0.032  | 0.945  | 0.02   | 0.003  | B          | Pure            |
| 28      | EC 636511     | USA             | 0.009  | 0.944  | 0.023  | 0.024  | B          | Pure            |
| 29      | EC 636512     | USA             | 0.002  | 0.001  | 0.002  | 0.994  | D          | Pure            |
| 30      | EC 636513     | USA             | 0.005  | 0.012  | 0.003  | 0.98   | D          | Pure            |
| 31      | EC 636514     | USA             | 0.003  | 0.003  | 0.003  | 0.991  | D          | Pure            |
| 32      | EC 636517     | USA             | 0.005  | 0.007  | 0.985  | 0.003  | C          | Pure            |
| 33      | EC 757855     | USA             | 0.028  | 0.231  | 0.454  | 0.287  | C          | Admixture       |
| 34      | EC 757857     | USA             | 0.954  | 0.039  | 0.006  | 0.002  | A          | Pure            |
| 35      | EC 757858     | USA             | 0.229  | 0.142  | 0.618  | 0.011  | C          | Admixture       |
| 36      | EC757856      | USA             | 0.011  | 0.479  | 0.507  | 0.003  | C          | Admixture       |
| 37      | Faizabad Nutan| Eastern Uttar Pradesh | 0.018  | 0.095  | 0.883  | 0.004  | C          | Pure            |
| 38      | GLK 1         | IARI, New Delhi | 0.671  | 0.286  | 0.038  | 0.006  | A          | Pure            |
| 39      | GS 4          | Western Uttar Pradesh | 0.129  | 0.009  | 0.828  | 0.034  | C          | Pure            |
| 40      | GS-2          | Western Uttar Pradesh | 0.954  | 0.014  | 0.023  | 0.009  | A          | Pure            |
| 41      | GS-3          | Western Uttar Pradesh | 0.01  | 0.028  | 0.956  | 0.006  | B          | Pure            |
| 42      | Gy 14         | USA             | 0.011  | 0.977  | 0.01   | 0.002  | B          | Pure            |
| 43      | Gy-421        | USA             | 0.942  | 0.046  | 0.007  | 0.005  | A          | Pure            |
| 44      | HS-1          | Haryana         | 0.008  | 0.006  | 0.974  | 0.012  | C          | Pure            |
| 45      | HS3           | Haryana         | 0.008  | 0.866  | 0.06   | 0.066  | C          | Pure            |
| 46      | HS4           | Haryana         | 0.027  | 0.766  | 0.103  | 0.104  | B          | Admixture       |
The high heritability was recorded for all 8 traits studied under net house and open field condition. High heritability estimates for number of traits were reported in cucumber (Munshi et al. 2007; Kumar et al. 2008; Yogesh et al. 2009 and Bisht et al. 2010). Selection can be exercised on the basis

Table 8 (continued)

| Sl. no. | Genotypes  | Origin                  | Fst1 | Fst2 | Fst3 | Fst4 | Population | Admixture/Pure |
|--------|-------------|-------------------------|------|------|------|------|------------|----------------|
| 47     | HS-5        | Haryana                 | 0.104| 0.018| 0.876| 0.002| C          | Pure           |
| 48     | HS-9        | Haryana                 | 0.039| 0.021| 0.937| 0.003| C          | Pure           |
| 49     | Kalyanpur Green | Eastern Uttar Pradesh | 0.006| 0.821| 0.169| 0.003| B          | Pure           |
| 50     | KP/PKS 1325 | Meghalaya               | 0.015| 0.022| 0.961| 0.003| B          | Pure           |
| 51     | KP/PKS 1330 | Meghalaya               | 0.044| 0.031| 0.919| 0.005| C          | Pure           |
| 52     | KP/PKS 748  | Meghalaya               | 0.025| 0.019| 0.033| 0.922| D          | Pure           |
| 53     | Mirpur Selection | West Bengal        | 0.003| 0.002| 0.002| 0.993| D          | Pure           |
| 54     | Pa-5        | Eastern Uttar Pradesh   | 0.006| 0.987| 0.003| 0.004| B          | Pure           |
| 55     | Pahari 77   | Western Uttar Pradesh   | 0.003| 0.008| 0.986| 0.003| C          | Pure           |
| 56     | PahariBarsati | Western Uttar Pradesh  | 0.003| 0.025| 0.97  | 0.002| C          | Pure           |
| 57     | Panipat Local | Haryana               | 0.006| 0.004| 0.988| 0.002| C          | Pure           |
| 58     | PCUC-8      | Uttarakhhand            | 0.046| 0.819| 0.025| 0.11 | B          | Pure           |
| 59     | Poinsett    | USA                     | 0.004| 0.004| 0.003| 0.989| D          | Pure           |
| 60     | PusaSanyog  | IARI, New Delhi         | 0.005| 0.028| 0.964| 0.004| C          | Pure           |
| 61     | PusaUday    | IARI, New Delhi         | 0.006| 0.815| 0.297| 0.003| B          | Pure           |
| 62     | Ragini      | West Bengal             | 0.008| 0.009| 0.981| 0.002| C          | Pure           |
| 63     | RK-100      | Eastern Uttar Pradesh   | 0.035| 0.71 | 0.237| 0.018| B          | Pure           |
| 64     | RS 1        | China                   | 0.007| 0.984| 0.005| 0.004| B          | Pure           |
| 65     | SN-25       | Eastern Uttar Pradesh   | 0.947| 0.006| 0.035| 0.013| A          | Pure           |
| 66     | Swarna Ageti | Jharkhand              | 0.003| 0.007| 0.987| 0.003| C          | Pure           |
| 67     | Swarna Sheetal | Jharkhand            | 0.014| 0.095| 0.888| 0.004| B          | Pure           |
| 68     | WBC-1-1     | West Bengal             | 0.976| 0.01 | 0.009| 0.004| A          | Pure           |
| 69     | WBC-13      | West Bengal             | 0.991| 0.004| 0.003| 0.002| A          | Pure           |
| 70     | WBC-21      | West Bengal             | 0.012| 0.973| 0.011| 0.004| B          | Pure           |
| 71     | WBC-2-1     | West Bengal             | 0.635| 0.355| 0.004| 0.006| A          | Admixture      |
| 72     | WBC-27-1    | West Bengal             | 0.957| 0.019| 0.013| 0.011| A          | Pure           |
| 73     | WBC-37-2    | West Bengal             | 0.028| 0.963| 0.004| 0.005| B          | Pure           |
| 74     | WBC-39-2    | West Bengal             | 0.17 | 0.79 | 0.032| 0.008| B          | Pure           |
| 75     | WBC-5       | West Bengal             | 0.991| 0.004| 0.003| 0.002| A          | Pure           |
| 76     | WBC-55      | West Bengal             | 0.96 | 0.028| 0.006| 0.005| A          | Pure           |
| 77     | WBC-6       | West Bengal             | 0.976| 0.004| 0.018| 0.002| A          | Pure           |
| 78     | WBC-9       | West Bengal             | 0.99 | 0.003| 0.004| 0.003| A          | Pure           |

Table 9  Marker–trait association in 78 cucumber genotypes using mixed linear model (MLM) approach for 8 traits under net house

| Sl. no. | Trait                        | Marker   | Chromosome | Physical position | p-value | Marker R^2 |
|---------|------------------------------|----------|------------|-------------------|---------|------------|
| 1       | Days taken for first fruit harvest | UW085408 | 7          | 727,009           | 0.004885| 0.213985   |
| 2       | Average fruit weight (g)      | UW084942 | 3          | 11,313,147        | 0.011435| 0.293111   |
| 3       | Fruit diameter (cm)           | UW083899 | 4          | 10,801,560        | 0.016794| 0.198364   |
| 4       | Average fruit weight (g)      | UW062953 | 4          | 22,037,675        | 0.02192 | 0.158741   |
of phenotypic performance for highly heritable characters.

Usually, cucumber has a narrow genetic base with 3–12% polymorphism with low to moderate variability (Yang et al. 2013). However, because of the inclusion of different genotypes, including native and exotic collection of genotypes, a wide variability was observed in this study.

Population stratification contributes to false positive results in association analysis (Yu and Buckler 2006). The population structure with respect to geographical origin contributes toward pseudo-associations (Maskri et al. 2012). Therefore, adequate methods need to be implemented to control the effects of population structure to avoid the high rate of Type-I error (Agrama et al. 2007) using STRUCTURE 2.2.

In this study, a Bayesian model-based population structure analysis indicated the occurrence of three sub-populations largely according to the major geographic regions of their origin or genetic makeup.

Table 10 Marker–trait association in 78 cucumber genotypes using mixed linear model (GLM) approach for 10 traits under open field condition

| Sl. no. | Trait                        | Marker     | Chromosome | Physical position | p-value | Marker R²   |
|--------|------------------------------|------------|------------|------------------|---------|-------------|
| 1      | No of fruits per plant       | UW085408   | 7          | 727,009          | 0.004369| 0.23302     |
| 2      | Days taken for first female flower to anthesis | SSR06011 | 3          | 33,107,529       | 0.016321| 0.158685    |
| 3      | Vine length (cm)             | SSR03552   | 3          | 16,005,221       | 0.01634 | 0.176522    |
| 4      | Days taken for first female flower to anthesis | UW084542 | 1          | 13,796,672       | 0.017928| 0.147644    |
| 5      | Average fruit weight (g)     | UW084942   | 3          | 11,313,147       | 0.026404| 0.185587    |
| 6      | Fruit length (cm)            | UW084942   | 3          | 11,313,147       | 0.026805| 0.180577    |
| 7      | Average fruit weight (g)     | UW062953   | 4          | 22,037,675       | 0.028925| 0.134002    |

Fig. 2  a, b Manhattan plots for average fruit weight using MLM approach studied under net house and open field condition, respectively

Based on DK estimation cucumber genotypes were divided into four distinct populations. We found that most of the genotypes from IARI, New Delhi grouped together (B) and another group (C) consisted of genotypes from NBPR, New Delhi and remaining genotypes from IARI, New Delhi. Group B and C had least genetic distance among them suggesting their genetic relatedness with each other. Genotypes from Eastern part of India (Group A) were found to be genetically distant from the genotypes from IARI, New Delhi and NBPR, New Delhi. This study indicated the narrow genetic base of the Indian germplasm in spite of its origin in Indian part of Himalayan foothills. Therefore, there is urgent need to conserve highly variable accessions in order to minimize the genetic erosion. Similar reports were documented regarding genetic erosion of cucumber germplasm in India during the joint Indo-US expedition programme (Staub et al. 1997). Two genotypes from USA (Gy14 and Gy-421) and five genotypes from China (CL 773, CL 746, CL758, 702-6-B76C and Russian) were scattered
among four different groups suggesting their close association with Indian germplasm. Earlier, distribution of cucumber worldwide from Asia through silk route was proposed by several workers (Pandey et al. 2013). Contribution of Indian genotypes in the global cucumber germplasm has been validated by Dar et al. (2017) and separation based on their geographical location was evident among the selected cucumber genotypes. However, some of the genotypes could not be differentiated according to geographical
patterns and hence admixtures in the populations can be attributed primarily due to cross-pollination and gene flow from genotypes grown in vicinity. The information generated through population structure and ancestral background would facilitate germplasm conservation and management strategy and selection of suitable parents in cucumber breeding programs. The MLM approach was found to be more reliable as this approach simultaneously takes accounts for both population structure as well as kinship statistics (Neumann et al. 2011).

Current study was carried out in two environment viz. under net house and open field condition revealed that traits exhibited significant marker–trait association in the MLM approach. Studies conducted in rapeseed (Cai et al. 2014), maize (Yu and Buckler 2006), bread wheat (Neumann et al. 2011), soybean (Kumar et al. 2014), mango (Lal et al. 2017) demonstrated the effectiveness of the MLM approach. As population structure can cause spurious associations (Kang et al. 2008) through MLM approach, incorporation of kinship in the association analysis allowing an improved control of type I and type II error rates over GLM due to relatedness and population structure (Yu and Buckler 2006).

The most intriguing part of the current study is that the many of the QTL previously published using the recombinant mapping approach were also validated and identification of novel genomic regions associated with the traits. Validation of the QTL identified by the other researchers makes the current study robust and very reliable. In the current study based on MLM analysis we found that 8 traits studied under net house confirm association of trait like days to first fruit harvest with marker UW085408 on chromosome 7 at 0.6 cM. While average fruit weight (g) was associated with two markers UW084942 and UW062953 on 3rd chromosome at 40 cM and on 4th chromosome at 105.1 cM respectively. Fruit diameter was associated with UW083899 on chromosome 3 at 55.6 cM. While 10 traits which were studied under open field condition revealed association of 6 traits through MLM analysis with the markers. A trait like number of fruits per plant was associated with UW085408 on chromosome 7 at 0.6 cM. Vine length, average fruit weight (g) and fruit length (cm) were found associated with SSR03552 at 53.7 cM, UW084942 at 40 cM and UW084942 at 40 cM on chromosome 3, respectively. Traits like days to first female flower anthesis was associated with two markers namely SSR06011 on chromosome 3 at 96.7 cM and UW084542 on chromosome 1 at 43 cM.

Among all traits studied under both net house and open field condition we found that there is common association for the trait like average fruit weight (g) on chromosome 3 with marker UW084942 at 40 cM. Earlier researcher have reported that fruit related traits like fruit weight and fruit shape index which includes fruit length, diameter and length/diameter ratio are controlled by quantitative trait loci (QTLs) in cucumber (Wei et al. 2016; Wenzel et al. 1995; Dijkhuizen and Staub 2002; Yuan et al. 2008a, b). Correlation study revealed that average fruit weight had significant and positive correlation with days taken for first fruit harvest, fruit length, fruit length/diameter ratio and yield per plant therefore associated marker UW084942 may be used to select fruit yield per plant over all.

MLM approach revealed that marker UW084942 on chromosome 3 was associated with the fruit length for both the environments. Wei et al. (2016) identified three QTLs for mature fruit length, with the largest effect displayed by mfl3.1, explaining 36.39% of the observed phenotypic variance on chromosome 3. Another QTL fl3.1 explained 39.10% of phenotypic variance and was also located on chromosome 3. The effects of other QTLs (fl1.1 and fl6.1) were comparatively smaller which were on linkage group 6. Similar finding were reported by Cheng et al. 2010 and Miao et al. 2011, they identified five QTLs (on LG1, LG4 and LG6) and three QTLs (on LG5 and LG6) that explained 7.1–14.1% of the phenotypic variation. Similarly, Yuan et al. 2008a, b detected 8 QTLs and dominant QTL on LG4 explained 23.32% of the fruit length variance. Three markers namely UW083899 on 6th chromosome, UW044712 on chromosome 1st and UW084186 on chromosome 2nd were associated with fruit diameter. UW062953 on 4th chromosome and SSR00607 on 3rd chromosome were associated with fruit length/diameter ratio. Similarly, Xu et al. 2015 identified a 0.19-Mb-long quantitative trait locus on chromosome 2 controlling fruit flesh thickness QTL ff2.1 (Fruit flesh thickness was calculated as the distance from the exocarp to the endocarp) and further they confirmed by SSR marker-based classical QTL mapping in 138 F2 individuals.

Based on MLM approach one marker each on chromosome 3rd (UW084942) and 4th chromosome
(UW062953) found to be associated with the average fruit weight (g) in the both environments. Wei et al. (2016) identified 9 QTLs for fruit weight of mature and immature fruits on three linkage groups. Three QTLs were detected on LG1; one was on LG6, and five on LG3 for fruit yield traits. Among all QTL fl3.2 explained 44.60% of the phenotypic variance and QTL fl3.1 explained 39.10% of phenotypic variance and both were located on linkage group 3. Yuan et al. 2008a, b also reported the same linkage group for immature fruit weight on LG3.

In this study, among the particular traits number of markers was found to be associated with more than one trait. This result indicates precise evaluation of traits. The different SSR loci, which exhibit overlapping association with diverse traits, signify the biological correlation among them. Therefore, it was presumed that the loci governing fruit length, fruit diameter, fruit weight, vine length, and yield per plant were probably located closest to each other on the same chromosome. But for other traits the differences in QTL number, position and phenotypic variance explained might be attributed to different mapping populations such as genotypes, population size etc. (Wei et al. 2016; Yuan et al. 2008a, b; Dijkhuizen and Staub 2002). Similarly, this is fact that crop lines are often assessed in a multi environment trial (MET), i.e. in different geographic locations, seasons, or years, in order to determine performance stability across environments (i.e. G · E). G · E is an important component of genetic variability (Crossa et al. 2011).

Conclusions

In conclusion, we reported the trait variation and genetic structure of 78 cucumber genotypes through which we could identify the potential germplasm for earliness and yield traits for open field as well as protected cultivation of cucumber. Moreover, two potential SSR markers, UW062953 and UW084942 can be used in breeding programs for yield improvement in cucumber because these markers were associated with average fruit weight (g) in MLM approach in two environments. The present study laid a solid foundation to develop more markers with close association with economically important traits in cucumber. Finally, this study shows that we should investigate more number of genotypes including parthenocarpic gynoecious lines and SSR markers to produce detailed picture on genetic structure of cucumber.

Acknowledgements The authors acknowledged the Department of Biotechnology (DBT), Ministry of Science and Technology, Government of India, and CRP on Molecular Breeding, Indian Council of Agricultural Research (ICAR) for providing financial support to this project.

Author contributions RK: methodology, data curation; ADM: methodology, data curation, project administration; TKB: supervision, writing, review and editing; GSJ: data curation, writing original draft, writing, review and editing; HC: writing, review and editing; AT: final editing; PD: provided lab facility; DS: statistical analysis.

Declarations

Conflict of interest No potential conflict of interest was reported by the author(s).

References

Agrama HA, Eizenga GC, Yan W (2007) Association mapping of yield and its components in rice. Mol Breed 19:341–356
Bisht B, Singh MP, Srivastava BK, Singh YV, Singh PK (2010) Evaluation of open-pollinated varieties and hybrids of cucumber for off-season production under naturally ventilated polyhouse. Indian J Horti 67(2):202–205
Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: Software for association mapping of complex traits in diverse samples. Bioinfo 23:2633–2635
Burton GW (1951) Quantitative inheritance in pearl millet (Pennisetum glaucum). Agron J 43:409
Burton GW, DeVane EH (1953) Estimating heritability in tall fescue (Festuca arundinacea) from replicated clonal material. Agron J 45:478–481
Cai D, Xiao Y, Yang W, Ye W, Wang B, Younas M, Wu J, Liu K (2014) Association mapping of six yield-related traits in rapeseed (Brassica napus L.). Theor Appl Genet 127:85–96
Cavagnaro PF, Senalik DA, Yang L, Simon PW, Harkins TT, Kodira CD, Huang S, Weng Y (2010) Genome-wide characterization of simple sequence repeats in cucumber (Cucumis sativus L.). BMC Gen 11(569):1–18
Cheng ZC, Gu XF, Zhang SP, Miao H, Zhang RW, Liu MM, Yang SJ (2010) QTL analysis for fruit length of cucumber. China Veg 12:20–25
Crossa JP, Perez G, de los Campos G, Mahuku G, Dreisigacker S (2011) Genomic selection and prediction in plant breeding. J Crop Imp 25:239–261
Dar AA, Mahajan R, Lay P, Sharma S (2017) Genetic diversity and population structure of Cucumis sativus L. by using SSR markers. 3Biot 307(7):2–12
Dijkhuizen A, Staub JE (2002) QTL conditioning yield and fruit quality traits in cucumber (*Cucumis sativus L*). effects of environment and genetic background. J New Seeds 4:1–30

Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Earl DA, Von H, Bridgett M (2012) STRUCTURE HARVEST: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Cons Genet Res 4:359–361

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14(8):2611–2620

Grumet R, McGreight JD, McGregor C, Weng Y, Mazourek M, Reitsma K, Labate J, Davis A, Fei Z (2021) Genetic resources and vulnerabilities of major cucurbit crops. Genes 12:1222. https://doi.org/10.3390 GENES12081222

Huang S, Li W, Zhang RQ, Li ZH, Gu L, Fan XF (2009) The genome of the cucumber (*Cucumis sativus L*). Nat Genet 41:1275–1281

Jat GS, Munshi AD, Behera TK, Choudhary H, Dash P, Amarnath R, Kumari S (2018) Genetics and molecular mapping of gynoecious (F) locus in cucumber (*Cucumis sativus L*). J Hort Sci Biotech 94(1):24–32

Johnson HW, Robinson HF, Comstock RE (1955) Estimates of genetic and environmental variability in soybeans. Agron J 47:314

Kang HM, Noah AM, Claire MW, Andrew K, David H, Mark JD, Eskin E (2008) Efficient control of population structure in model organism association mapping. Genetics 178(3):1709–1723

Kumar A, Kumar S, Pal AK (2008) Genetic variability and characters association for fruit yield and yield traits in cucumber. Indian J Horti 65(4):423–428

Kumar B, Talukdar A, Bala I, Verma K, Lal SK, Sapra RL, Namita B, Chander S, Tiwari R (2014) Population structure and association mapping studies for important agronomic traits in soybean. J Genet 93(3):775–784

Kumar R, Munshi AD, Behera TK, Jat GS, Choudhary H, Singh M, Talukdar A (2020) Genetic diversity analysis of cucumber (*Cucumis sativus L*.) accessions differing in quantitative traits and microsatellite markers. Ind J Agric Sci 90(11):2161–2167

Lal S, Singh AK, Singh SK, Srivastav M, Singh BP, Sharma N, Singh NK (2017) Association analysis for morphological traits in mango (*Mangifera indica L*.) by genic-SSR markers. Trees 31(5):1391–2140

Maskri YA, Sajjad M, Khan SH (2012) Association mapping: a step forward to discovering new alleles for crop improvement. Int J Agric Biol 14:153–160

Miao A, Staub J, Zhangyong S, Ghorbel A (2003) Genetic diversity in African cucumber (*Cucumis sativus L*.) provides potential for germplasm enhancement. Genet Res Crop Evol 50:461–468

Munshi AD, Pandita B, Behera TK, Kumar R, Bisht IS, Behera TK (2007) Genetic variability in *Cucumis sativus var. hardwickii* R. germplasm. Cucurbit Genet Cooperat Rep 30:5–10

Neumann K, Kobiljksi B, Dencic S, Varshney RK, Borner A (2011) Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum L*.). Mol Breed 27:37–58

Nordborg M, Tavare S (2002) Linkage disequilibrium: what history has to tell us? Trends Genet 18:83–90

Palaisa K, Morgante M, Williams M, Rafalski A (2004) Long range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep. Proc Nat Acad Sci, USA 101:9885–9890

Pandey S, Ansari WA, Mishra VK, Singh AK, Singh M (2013) Genetic diversity in Indian cucumber based on microsatellite and morphological markers. Biochem Syst and Ecol 51:19–27

Pritchard JK, Matthew S, Peter D (2000) Inference of population structure using multi-locus genotype data. Genetics 155:945–959

Ren Y, Zhang Z, Liu J, Staub JE, Han Y (2009) An integrated genetic and cytogenetic map of the cucumber genome. PLoS ONE 4(6):1–8

Sebastian P, Schaefer H, Telford IRH, Renner SS (2010) Cucumber (*Cucumis sativus*) and melon (*Cucumis melo*) have numerous wild relatives in Asia and Australia, and the sister species of melon is from Australia. Proc Nat Acad Sci USA 107(32):14269–14273

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

Staub JE, Serquen FC, McGreight JD (1997) Genetic diversity in cucumber (*Cucumis sativus L.*) III: An evaluation of Indian germplasm. Genet Res Crop Evol 44:315–326

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 Breed 124:44–53

Wei QZ, Fu WY, Yang L, Li D, Li Y, Gu X, Huang S, Mas JG, Weng YQ, Johnson S, Staub JE, Huang SW (2010) An extended microsatellite genetic map of cucumber (*Cucumis sativus L*.). Hort Sci 45:880–886

Wenzel G, Kennard WC, Havey MJ (1995) Quantitative trait analysis of fruit quality in cucumber: QTL detection, confirmation, and comparison with mating-design variation. Theor Appl Genet 91(1):53–61

Weng YQ, Johnson S, Staub JE, Huang SW (2010) An extended microsatellite genetic map of cucumber (*Cucumis sativus L*.). J Genet 90(3):371–377

Xu X, Lu L, Zhu B, Xu Q, Qi X, Chen X (2015) QTL mapping of cucumber fruit flesh thickness by SLAF-seq. Sci Rep 5:1–8

Yang L, Li D, Li Y, Gu X, Huang S, Mas JG, Weng Y (2013) A 1,681-locus consensus genetic map of cultivated cucumber including 67 NB-LRR resistance gene homolog and ten gene loci. BMC Plant Biol 13:1–13

Yogesh C, Yadav SK, Brijbal B, Dixit SK (2009) Genetic variability, heritability and genetic advance for some traits in cucumber. Indian J Horti 66(4):488–491
Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. Curr Opin Biotech 17:155–160
Yuan XJ, Li XZ, Pan JS, Wang G, Jiang S, Li XH, Deng SL, He HL, Si MX, Lai L, Wu AZ, Zhu LH, Cai R (2008a) Genetic linkage map construction and location of QTLs for fruit-related traits in cucumber. Plant Breed 127(2):180–188
Yuan XJ, Pan JS, Cai R, Guan Y, Liu LZ, Zhang WW, Li Z, He HL, Zhang C, Si LT, Zhu LH (2008b) Genetic mapping and QTL analysis of fruit and flower related traits in cucumber (Cucumis sativus L.) using recombinant inbred lines. Euphytica 164(2):473–491
Zhang WW, He HL, Guan Y, Du H, Yuan LH, Li Z, Yao DQ, Pan JS, Cai R (2010) Identification and mapping of molecular markers linked to the tuberculate fruit gene in the cucumber (Cucumis sativus L.). Theor Appl Genet 120:645–654
Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. Plant Genome 1:5–20

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.