MATERIALS AND METHODS
3. MATERIALS AND METHODS

The present investigation entitled “Studies on genetic divergence in morphological and quantitative characters of sweet basil (Ocimum basilicum L.)” was conducted during 2005-06 and 2006-07 at the Research Farm, Janta Vedic College, Baraut, Baghpat, U.P.

The detail of materials and methods are given below:

3.1 MATERIALS

Thirty genotypes of sweet basil (Ocimum basilicum) used for studying the genetic advance, variability, heritability character association, path coefficient and genetic divergence.

The list of thirty genotypes is presented as follows:

<table>
<thead>
<tr>
<th>S. No.</th>
<th>Genotypes</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>IC 110267</td>
<td>NBPG, New Delhi</td>
</tr>
<tr>
<td>2.</td>
<td>EC 338785</td>
<td>NBPG, New Delhi</td>
</tr>
<tr>
<td>3.</td>
<td>EC 388890</td>
<td>NBPG, New Delhi</td>
</tr>
<tr>
<td>4.</td>
<td>EC 388895</td>
<td>NBPG, New Delhi</td>
</tr>
<tr>
<td>5.</td>
<td>EC 388893</td>
<td>NBPG, New Delhi</td>
</tr>
<tr>
<td>6.</td>
<td>EC 387838</td>
<td>NBPG, New Delhi</td>
</tr>
<tr>
<td>7.</td>
<td>EC 388788</td>
<td>NBPG, New Delhi</td>
</tr>
<tr>
<td>8.</td>
<td>EC 338794</td>
<td>NBPG, New Delhi</td>
</tr>
<tr>
<td>9.</td>
<td>IC 326735</td>
<td>NBPG, New Delhi</td>
</tr>
<tr>
<td>10.</td>
<td>IC 333332</td>
<td>NBPG, New Delhi</td>
</tr>
<tr>
<td>11.</td>
<td>IC 336833</td>
<td>NBPG, New Delhi</td>
</tr>
<tr>
<td>12.</td>
<td>EC 312264</td>
<td>NBPG, New Delhi</td>
</tr>
<tr>
<td>13.</td>
<td>EC 388891</td>
<td>NBPG, New Delhi</td>
</tr>
</tbody>
</table>
3.2 Experimental details

Thirty genotypes of Ocimum basilicum, received from NBPGR, New Delhi and grown in randomized block design (RBD) with three replications at the Research Farm of Janta Vedic College, Baraut, Baghpat, U.P. during kharif season of 2005-06 and 2006-07, as per flowering lay out.

Number of genotypes : 30
Number of replications : 3
Plot size : 3 x 3 m
Line to line spacing : 45 cm
Plant to plant spacing : 30 cm
Number of lines per plot : 3
Number of plants per line : 10
Number of plants per plot : 30
3.4 Observation recorded

In each replication the data were recorded on five randomly selected competitive plants from each varieties on the following ten morphological characters and seventeen quantitative characters.

(a) Morphological characters

1. Stem colour
   Stem colour was recorded at initiation of flowering (visual scoring) using RHS colour chart.

2. Lamina colour
   Lamina colour was recorded at flowering stage (visual scoring) using RHS colour chart.

3. Lamina shape
   Lamina shape was recorded at flowering stage (visual scoring).

4. Lamina margin
   Lamina margin was recorded at flowering stage (visual scoring).

5. Lamina pubescence
   It was recorded at flowering stage (visual scoring).

6. Lamina surface
   It was recorded at flowering stage (visual scoring).

7. Inflorescence type
   It was recorded at full bloom stage (visual scoring).

8. Flower colour
   It was recorded at full bloom stage (visual scoring) using RHS colour chart.
9. **Seed shape**

   It was recorded on mature and dried seeds.

10. **Seed colour**

   It was recorded on mature and dried seeds (visual scoring) using RHS colour chart.

(b) **Quantitative characters**

   One character number of flowers per whorl is a quantitative character but at full bloom stage it was counted six flowers per whorl.

   In each replication the data were recorded on the randomly selected competitive plants from each variety on the following sixteen characters:

1. **Number of primary branches per plant**

   Number of primary branches per plant was counted at flowering stage (average of five random plants).

2. **Lamina length (cm)**

   Lamina length was measured on fully developed leaves at flowering stage (average of five random plants).

3. **Lamina width (cm)**

   Lamina width was measured on fully developed leaves at flowering stage (average of five random plants).

4. **Leaf stem ratio**

   Leaf stem ratio was recorded on fully developed plants at flowering stage (average of five random plants).

5. **Days to flower initiation**

   It was recorded as number of days from sowing/planting to flower initiation.
6. **Number of spikes per plant**

   Number of spikes per plant was counted at full bloom stage (average of five random plants).

7. **Spike length (cm)**

   Spike length was measured at full bloom stage (average of five random spikes).

8. **Number of flower whorls per spike**

   Number of flower whorls per spike was counted at full bloom stage (average of five random spikes).

9. **Number of flowers per whorl**

   Number of flowers per whorl were observed at full bloom stage.

10. **Plant height (cm)**

    Plant height was measured from ground level to the tip of the plant at complete flowering stage (average of five random plants).

11. **Fresh herbage yield per plant (g)**

    Fresh herbage yield per plant was recorded on fresh weight basis at full flowering stage (average of five random plants).

12. **Dry herbage yield per plant (g)**

    Dry herbage yield per plant was recorded on dry weight basis at shade drying (average of five random plants).

13. **Days to seed maturity**

    Days to seed maturity were counted as number of days from sowing/planting to complete seed maturity.
14. **Seed yield per plant (g)**

   Seed yield per plant was recorded at complete maturity stage (average of five random plants).

15. **1000-seed weight (g)**

   1000-seed weight was recorded on mature and dry seeds.

16. **Essential oil content (%)**

   Essential oil content was estimated by distillation of aerial parts on fresh weight basis and expressed on dry weight basis.

17. **Essential oil yield per plant (ml)**

   Essential oil yield per plant was extracted on fresh weight basis by steam distillation at full bloom stage (average of five random plants).

**STATISTICAL ANALYSIS**

Plot means were used for the following statistical analysis:

**Analysis of variance (ANOVA)**

Data recorded from experiment was utilized for analysis of variance (ANOVA) following Panse and Sukhatme (1967). The details of the analysis of variance are given below:

**Analysis of variance (ANOVA) Table for Randomized block design**

<table>
<thead>
<tr>
<th>source of variation</th>
<th>d.f.</th>
<th>Mean squares</th>
<th>Expected mean square</th>
<th>F calculated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replication</td>
<td>(r-1)</td>
<td>Sr</td>
<td>MSr</td>
<td>MSr/MSe</td>
</tr>
<tr>
<td>Genotype</td>
<td>(g-1)</td>
<td>Sg</td>
<td>MSg</td>
<td>MSe/MSe</td>
</tr>
<tr>
<td>Error</td>
<td>(r-1)(g-1)</td>
<td>Se</td>
<td>MSE</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>(rg-1)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Plate 1. Clevenger apparatus involved in the hydro-distillation
Where,

\[ r = \text{number of replication} \]
\[ g = \text{number of genotypes} \]
\[ MSg = \text{mean square due to genotype} \]
\[ MSr = \text{mean square due to replication} \]
\[ MSE = \text{mean square due to error} \]

Parameters of variability

Mean (\( \bar{X} \))

The mean value of each character was worked out by dividing the total by corresponding number of observations.

\[ \bar{X} = \frac{\sum x_{ij}}{N} \]

Where,

\( \sum x_{ij} = \text{any observation in } i^{th} \text{ varieties and } j^{th} \text{ replications} \)
\( N = \text{total number of observations} \)

Range

Lowest and highest values for character were recorded.

Standard error

Standard error of difference of two means was calculated with the help of error mean square from the analysis of variance table.

\[ \text{Standard error (SEM±)} = \sqrt{\frac{MSE}{r}} \]

Where,

\( MSE = \text{error mean sum of square} \)
\( r = \text{number of replications} \)
Critical difference (CD)

Critical differences for all the characters were calculated to compare treatment means. Critical differences were calculated with the help of standard error for the differences of two means and tabulated value of 't' at 5 per cent level of significance and at error degree of freedom. Critical difference (CD) = $\text{SE}_d \times t$ at 5 per cent probability at error degree of freedom.

Coefficient of variation (CV)

Genotypic and phenotypic coefficients of variation were estimated by the formula suggested by Burton and De Vane (1953) for each character as:

Genotypic coefficient of variation (G.C.V.) = \( \sqrt{\frac{V_g}{X}} \) x 100

Phenotypic coefficient of variation (P.C.V.) = \( \sqrt{\frac{V_p}{X}} \) x 100

Where,

\( \overline{X} \) is the mean of that particular character and \( V_g \) and \( V_p \) are the genotypic and phenotypic variances, respectively.

3.5.2.6 Heritability in broad sense

Heritability in broad sense was calculated according to the formula suggested by Allard (1960) for each character:

\[ H = \frac{V_g}{V_p} \times 100 \]

Where,

\( H \) = Heritability (broad sense),

\( V_g \) = Genotypic variance and

\( V_p \) = Phenotypic variance
Genetic advance expressed as percentage of mean

Estimates of appropriate variance components were substituted for the parameters to predict expected genetic gain as suggested by Allard (1960).

The expected genetic advance was calculated at 5 per cent selection intensity for each character as:

\[ GA = (k) \sqrt{\sigma^2_p(H)} \]

Genetic advance (% of mean) = \[ \frac{GA}{X} \times 100 \]

Where, \( K \) = Selection differential, the value of which is 2.06 at 5% selection intensity

\( H \) = Heritability in broad sense

\[ \sqrt{\sigma^2_p} \] = Phenotypic standard deviation

\( \bar{X} \) = Mean value for that character over all the varieties

Correlation coefficient analysis

Correlated characters are of interest to find out the genetic causes of correlation through the pleiotropic action of genes, to know how selection for one character will cause simultaneous change in other characters and to find out correlation between characters and fitness.

Phenotypic \( r(p) \) correlation coefficients for all possible pairs of characters were calculated from the already obtained variance and covariances according to Johnson et al. (1955).

The phenotypic correlation was measured by:

\[ r(p) = \frac{\sigma_{xy(p)}}{\sqrt{\sigma^2x(p) \times \sigma^2y(p)}} \]

Where,

\( \sigma_{xy(p)} \) = Phenotypic covariances between character \( x \) and \( y \)
\( \sigma^2 x (p) = \) Phenotypic variance of character \( x \)

\( \sigma^2 y (p) = \) Phenotypic variance of character \( y \)

The phenotypic correlation coefficients were tested against standardized tabulated significant values of 'r' with \((g-2)\) degrees of freedom as per the procedure described by Fisher and Yates (1938).

**PATH-COEFFICIENT ANALYSIS**

The correlation coefficients were used to work out path-coefficient analysis. The estimates of direct and indirect effects of various characters were calculated through path coefficient analysis. Path-coefficients were obtained according to Dewey and Lu (1959). A set of simultaneous equations in the following form were solved:

\[
 r_{ny} = p_{ny} + r_{n2} P_{2y} + r_{n3} P_{3y} + \ldots + r_{nx} P_{xy}
\]

Where,

- \( r_{ny} \) = Correlation coefficient of one character and yield
- \( p_{xy} \) = Path-coefficient between the character and yield
- \( r_{n2} \ldots r_{nx} \) = Represent correlation coefficient of the character and each of other yield components in turn.

The following correlation matrices were formed:

<table>
<thead>
<tr>
<th>Matrix A</th>
<th>Matrix B</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r_{1y} )</td>
<td>1 ( r_{12} ) ( r_{13} ) \ldots ( r_{1n} )</td>
</tr>
<tr>
<td>( r_{2y} )</td>
<td>1 ( r_{23} ) \ldots ( r_{2n} )</td>
</tr>
<tr>
<td>( r_{3y} )</td>
<td>1 \ldots ( r_{3n} )</td>
</tr>
<tr>
<td>( r_{ny} )</td>
<td>1</td>
</tr>
</tbody>
</table>
The technique given by Goulten (1954) was followed for inversion of \((B^{-1})\) of B matrix. Path coefficients \(p_{iy}\) were obtained as follows:

\[ p_{iy} = (B^{-1}) \times (A) \]

The indirect effects for a particular character through other characters were obtained by multiplication of direct path and particular correlation coefficient between these two characters respectively.

Indirect effect = \(r_{ij} \times p_{iy}\)

Where,

\[ i = 1 \ldots \ldots n \]

\[ j = 1 \ldots \ldots n \]

\[ p_{iy} = p_{1y}, p_{2y}, \ldots, p_{ny} \]

\(r_{iy}\) = Correlation between two independent characters.

The residual effect i.e. the variation in yield unaccounted for those associated was calculated from the following formula.

Residual effect \((x) = \sqrt{1-R^2}\)

Where,

\[ R^2 = p_{1y} r_{1y} + p_{2y} r_{2y} + \ldots + p_{ny} r_{ny} \]

\(R^2\) is the squared multiple correlation coefficient and is the amount of variation in yield that can be accounted for by the yield component characters.

**3.5.8 Mahalanobis' \(D^2\)-statistics**

To measure the genetic divergence between two populations, Mahalanobis (1936) gave a statistics known as Mahalanobis's statistics. This statistics is useful in multivariate analysis of quantitative characters to assess the genetic divergence between populations.
i. Basic concept: If $x_1, x_2, \ldots, x_p$ be the multiple measurements available on each individual and $d_1, d_2, \ldots, d_p$ as $X_1^1 - X_1^2, X_2^1 - X_2^2, \ldots, X_p^1 - X_p^2$, respectively, be the difference in the means on $i^{th}$ character, for first and second population, then Mahalanobis $D^2$ statistics is defined as follows:

$$pd^2 = b_1d_1 + b_2d_2 + \ldots + bpd_p.$$  

Here, $b_i$ values are to be estimated such that the ratio of variance between the population to variance within the population is maximum. In terms of variance and covariance, the $D^2$ values is obtained as follows:

$$pd^2 = W_{ij} (X_i^1 - \bar{X}_i^1) (X_j^1 - \bar{X}_j^1)$$

where, $W_{ij}$ is the $(i, j)^{th}$ element in the inverse of estimated variance-covariance matrix.

ii. Computation of $D^2$ values: For each of the pairs of combinations, the mean deviation i.e. $\bar{Y}_i^1 - \bar{Y}_i^2$ with $i = 1, 2, 3, \ldots, p$ was computed and the $D^2$ was calculated as the sum of the squares of these deviations i.e.,

$$D^2 = \sum_{i=1}^{n} (\bar{Y}_i^1 - \bar{Y}_i^2)^2$$

E.g. in the present study, the $D^2$ value between two genotypes based on the measurement of $p$ character were measured as follows:

$$D^2 = (\bar{Y}_1^1 - \bar{Y}_1^2)^2 + (\bar{Y}_2^1 - \bar{Y}_2^2)^2 + \ldots + (\bar{Y}_n^1 - \bar{Y}_n^2)^2$$

iii. Testing the significance of $D^2$ values: The $D^2$ values were obtained for a pair of genotypes and was taken as the calculated value for $\chi^2$ and was tested against the tabulated value of $\chi^2$ for $n$ degree of freedom, where $n$ is the number of characters studied.
iv. Euclidean cluster analysis: Genetic divergence among 30 genotypes was worked out following the method of non-hierarchical Euclidean cluster analysis as suggested by Beale (1969). Spark (1973) has outlined the detailed algorithm for the method. In addition, the non-mathematical description of the method is given by Kendall (1980). Following assumptions are made in this method: (i) That the Euclidean distance 'D^2' separating 'n' points in p-dimensional space are proportional to the dissimilarities between the objects, and that no object can simultaneously belong to two clusters. (ii) That the solution for C1 (higher limit of) clusters does not provide better fit to the data than the solution for the C2 (lower limit of) clusters.

According to Beale (1969) initially each observation is allocated to its closest cluster center; the means of the cluster are then calculated and are taken to the new cluster centers at the same time; the sum of squared deviations of the observations from their respective clusters centers are computed. The observations are then checked in turn to see if a shift to a different cluster center results in a decrease in the total sum of squares. It is clearly an improvement to reassign an observation from cluster 'K' to cluster 'I' if it is nearer to the center of the latter. That is, if \( d_i^2 < d_k^2 \), where, \( d_i \) is the distance from the centre of cluster 'I'. However, a more effective criterion involves reassigning the observation if the squared deviation from the centre of the cluster 'I' is less than that form the centre of cluster 'K', even when the cluster centre are simultaneously repositioned. That is, when

\[
\frac{n_i}{n_i + 1} d_i^2 < \frac{n_k}{n_k - 1} d_k^2
\]

Where \( n_i \) is the number of observation in cluster 'I'.
In delimiting clusters usually average deviance rather than individual \( \frac{1}{2} \) \( m(m-1) \) deviances among a subset of \( 'm' \) points is considered, if the \( i^{th} \) variable on the \( j^{th} \) member is \( X_{ij} \), mean deviance of a set of \( 'n' \) points is as follows:

\[
\frac{1}{m(m-1)} \sum_{i=1}^{P} \sum_{j=1}^{m} \sum_{k=1}^{m} (X_{ij} - X_{ik})^2
\]

\[
\frac{1}{m(m-1)} \sum_{i=1}^{P} \sum_{j=1}^{n} \sum_{k=1}^{m} (X_{ij} - X_{i}) - (X_{ik} - X_{i})^2
\]

Where, \( X_{i} \) is the mean of \( X_{i} \) observations on \( 'm' \) members.

\[
\frac{1}{m(m-1)} \sum_{i=1}^{P} \left( \sum_{d} \sum_{k} (X_{ij} - X_{i}) + \sum_{k} \sum_{j} (X_{ik} - X_{i})^2 - 2 \sum_{j} \sum_{k} (X_{ij} - X_{i}) (X_{ik} - X_{i}) \right)
\]

The cross-product term vanishes and the other two are equal. Thus, average deviance:

\[
\frac{2}{m(m-1)} \sum_{i=1}^{P} \sum_{j=1}^{m} (X_{ij} - X_{ik})^2
\]

Now, instead of calculating \( 1/m (m-1) \) deviances, \( 'm' \) deviances from the centre of gravity are calculated initially, a given numbers of vectors of cluster centers are located in the \( 'p' \) space. The position of these centers can be chosen arbitrarily or randomly. However, a choice of initial cluster centers reduces the amount of computation to a considerable extent.

To start with \( 'n' \) cases are allotted to a predetermined maximum number of clusters (\( C \) maximum) according to the procedure suggested by Beale. The residual sum of squares RSS (\( C \)), for the solution involving \( 'C' \) clusters are calculated. Then the number of cluster \( 'C' \) is reduced by one cluster at a time until a predetermined lower (or upper) number of clusters is
reached. In each step RSS(C) min when RSS(C) values for C maximum (C1) and C minimum (C2) are available. These are used in carrying out pseudo F-ratio-test of the hypothesis, that the solution for C1 clusters does not provide better fit to the data that the solution for C2 clusters with C1>C2:

This F-ratio is calculated as:

$$F = \frac{RSS(C2) - RSS(C1)}{RSS(C1)} \left\{ \left( \frac{N-C2}{N-C1} \right) \left( \frac{C1}{C2} \right)^{2/p} - 1 \right\}$$

With P (C2-C1)° and p (n-C1) degree of freedom. If for a given C1, it is significant for any given C2, the null hypothesis is rejected and it is deduced that the representation in terms of C1 is adequate.

For reducing the number of cluster by one at a time until predetermined lowest (C min.) number of cluster is reached. Beale suggested certain producers. Instead of using Beale's procedure for merging two clusters, Doshi et al. (1991) have adopted a simple procedure: when a solution is found for 'C' clusters, 'C' vectors of new cluster centers are calculated. From this set of new cluster centre vector, last vector is dropped and (C1) vectors are used as initial vectors of cluster centers for arriving at C1 clusters. For determining the appropriate number of cluster, F-test gives rough guide in exploratory analysis.