*Original Research Article*

**Multivariate Analysis for Enhancing Seed Yield and Oil Quality in Sunflower (*Helianthus annuus* L.): Correlation Dynamics and Principal Component Dissection of Oil Content and Fatty Acid Profiles**

.

ABSTRACT

|  |
| --- |
| To investigate the complex relationships between seed yield and oil content, with their related agronomic traits in sunflower, this correlation analysis study was done at Punjab Agricultural University (30.9010º N and 75.85.73º E) using 113 genotypes, which were grown in spring 2022 in randomised block design with three replications. Eighteen quantitative traits were analysed, revealing key correlations, like seed yield per plant (economic yield) positively associated with structural traits like plant height, head diameter, stem girth, number of leaves per plant, days to maturity, biological yield per plant, number of seeds per head, hundred seed weight and seed volume weight, while oil content was positively correlated with seed volume weight and inversely correlated with hundred seed weight, indicating a trade-off between yield and oil content. Additionally, oleic acid content demonstrated strong positive correlations with palmitic and linoleic acid content. These insights provide direction for breeding programs aiming to optimise sunflower yield and quality.Furthermore, this research applied principal component analysis (PCA) to elucidate the genetic and phenotypic interrelations influencing oil content and fatty acid composition, the results of which identified five principal components, with the first three accounting for 93% of the total variation, out of which first two components capturing about 77% of the variation. The scatter plot and biplot analysis revealed that oleic acid and linoleic acids dominate the first component with a strong inverse correlation, while the second component is significantly influenced by the positive correlations of palmitic and stearic acids, thereby explaining 77% of variations in oil parameters, and offering actionable insights for enhancing sunflower breeding programs targeting oil quality optimisation. |

*Keywords:* Sunflower, correlation analysis, principal component analysis, yield, oil parameters, scatter-plot, biplot

1. INTRODUCTION

Sunflower (*Helianthus annuus* L.) is an oilseed crop belonging to the family Asteraceae, with a significant contribution to edible oil production and biofuel industries. Therefore, enhancing seed yield and oil content is a critical goal for breeders, particularly in the face of rising, here, correlation analysis serves as a pivotal tool for breeders, facilitating the comprehension of relationships between diverse traits inherent in sunflower and guiding breeders to gain invaluable insights into the associations among traits such as yield, disease resistance, flowering patterns, and seed quality through meticulous correlation scrutiny. This understanding guides the selection of optimal parental lines by emphasizing combinations that exhibit favourable correlations, thereby enhancing the probability of yielding progeny with desired traits. Moreover, correlation analysis expedites the breeding process by focusing on traits strongly correlated with improved sunflower performance, leading to the development of superior varieties.

The oil accounts for 80% value of this crop and is considered to be of superior quality because it contains about 90% fatty acid, 9% phytosterols and 1% vitamin E (alpha-tocopherol), and oleic and linoleic (typically 90% unsaturated fatty acids) are the primary fatty acids present in sunflower oil along with 10% saturated fatty acids like palmitic and stearic acid (Škorić and Sakač, 2013). This is primarily used in the form of salad dressings, cooking oil or in preparation of margarine and non-dairy creamers, bakery applications, and spray coating oils for cereal, crackers and dried fruit, thereby it has huge market demand. However, the complexity of sunflower traits, influenced by genetic and environmental factors, poses challenges in breeding programs aimed at improving the oil content and oil quality of this crop. Therefore, principal component analysis (PCA), a multivariate technique that simplifies the complexity by reducing the dimensionality of the data and highlights the most influential traits, has been used in this experiment to understand, which fatty acids significantly contribute to overall oil content and how different acids interrelate.

2. material and methods

In this experiment, seven inbred lines (CMS 67A, CMS 73A, CMS 82A, CMS 84A, CMS 86A, CMS 88A and CMS 103A), thirteen tester lines (HOHAL-17-1, HOAL-2-Pɜ, HOHAL-11-4, HOAL-30-2, HOAL-21-2, HOHAL-75-4, HOHAL-85-1, HOHAL-28-1, HOHAL-21-1, HOHAL-70-2, HOAL-41-2, HOHAL-85-3 and HOAL-17-2) and their ninety-one hybrids (crossed in Line × Tester fashion), and two commercial checks (PSH 1962 and PSH 2080), i.e., a total of 113 genotypes were sown in spring 2022 in three replications in randomized complete block (RCB) design at sunflower breeding area of the department of plant breeding and genetics at Punjab Agricultural University (30.9010º N and 75.85.73º E).

The five randomly selected plants from each treatment were tagged for recording the observations for the quantitative characters like: days to flowering initiation, days to 50% flowering, plant height, plant diameter, stem girth, number of leaves per plant, days to maturity, biological yield per plant, seed yield per plant (economic yield), number of seeds per head, hundred seed weight, seed volume weight and harvest index.). The oil content of the seed samples was calibrated using Nuclear Magnetic Resonance Spectrometry (NMR), and fatty acid composition (stearic acid (C18:0), palmitic acid (C16:0), linoleic acid (C18:2) and oleic acid (C18:1) was calibrated using Gas Liquid Chromatography (GLC)

* 1. **Correlation coefficient analysis**

The correlation coefficient of eighteen agronomic and biochemical characters was calibrated statistically among themselves, as per the procedure given by Pearson (1901) mathematically and Al-Jibouri *et al.,* (1958) and Singh and Chaudhary, (1977), who explained its usage in plant breeding. and its significance was calculated using a t-test (Panse and Sukhatme, 1985). The R software was used in this experiment to calibrate the correlation coefficient, which explained the degree of the linear relationship between various variables, thereby a positive value indicating the direct or similar relationship between two variables, while a negative value means inverse relations between variables. The equation for the same is:

Karl Pearson correlation coefficient ( rxy or r) = $\frac{ Cov xy }{\sqrt{Vx × Vy} }$

= $\frac{N \sum\_{i=1 }^{N}X\_{1}Y\_{1}- \sum\_{}^{}\overbar{X}\sum\_{}^{}\overbar{Y}}{\sqrt{N\left(\sum\_{i=1}^{N}X\_{i}^{2}- \left( \sum\_{}^{}\overbar{X}\right)^{2}\right)\left(\sum\_{i=1}^{N}Y\_{i}^{2}- \left(\sum\_{}^{}\overbar{Y}\right)^{2}\right)}}$

= $\frac{\sum\_{i=1}^{N}\left(X\_{i}- \overbar{X}\right)\left(Y\_{i}- \overbar{Y}\right)}{\sqrt{\sum\_{i=1}^{N}\left(X\_{i}- \overbar{X}\right)^{2}\sum\_{i=1}^{N}\left(Y\_{i}- \overbar{Y}\right)^{2}}}$

where,

Covxy = correlation coefficient between characters X and Y

Vx = variance for character X

Vy = variance for character Y

N = total number of Samples

Xi = X variable samples

Yi = Y variable samples

$\overline{X}$ = mean of values in X variables

$\overline{Y}$ = mean of values in Y variables

The correlation coefficient is tested against the t-test value, as follows:

H0: $ρ = 0$ (no correlation)

H1: $ρ \ne 0$ (There is a correlation)

Standard Error (SE) = $\sqrt{\frac{1 - r^{2}}{n - 2}}$

t = $\frac{r}{SE}$ , d.f = n - 2

where,

r = sample correlation coefficient

n = number of observations

* 1. **Principal component analysis**

PCA was used for oil quality parameters for germplasm characterisation and categorising them according to their traits, by calculating the eigenvectors and eigenvalues of the covariance matrix. The eigenvectors represent the principal component and the eigenvalues represent the variance explained by each component. For this R software package was used, and a biplot was generated using the 'FactoMineR' (Factor analysis and data mining with R) package with the command 'biplot (Husson et al., 2009 and R Core Team, 2013). The equation for the same is;

$$Eigenvalues \left(PCi\right)=\left(X- \overbar{X}\right)×Wi$$

Where,

X = original data matrix for each plant genotype and for each recorded value of a trait

$\overbar{X}$ = mean vector for each trait in the data matrix

Wi = $i^{th}$ eigenvector of the covariance matrix (represents the direction of the principal component)

1. results and discussion
	1. **Correlation analysis of agronomic and biochemical traits**

The correlation analysis of various agronomic traits and oil content among themselves is depicted in Table 1, while the correlation of fatty acid composition with oil content is presented in Table 2. Correlation analysis becomes important as the positively correlated traits directly affect each other while negatively correlated traits have an inversely proportional relationship. Moreover, correlation gives a clear picture of whether the characters are associated, which may be through linkage, pleiotropy, or physiological and biochemical pathways followed by the plant during growth and development. Therefore, a change in one character will cause variation in the other character

**Table 1.: Correlation coefficient for various agronomic traits and oil content**

| **Traits of Variation** | **DFI** | **DTF** | **HGT** | **HDIA** | **STG** | **NLPP** | **DTM** | **BYLD** | **SYP** | **NSH** | **HSW** | **VWT** | **HI** | **OC** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **DFI** | 1 | 0.709\*\*\* | -0.237\*\*\* | -0.184\*\*\* | -0.025 | -0.196\*\*\* | 0.041 | 0.014 | 0.001 | 0.052 | -0.15\*\* | -0.021 | -0.012 | -0.009 |
| **DTF** | 0.709\*\*\* | 1 | -0.144\*\* | -0.144\*\* | -0.154\*\* | -0.161\*\* | 0.227\*\*\* | -0.039 | -0.018 | 0.054 | -0.164\*\* | 0.134\* | -0.076 | -0.022 |
| **HGT** | -0.237\*\*\* | -0.144\*\* | 1 | 0.788\*\*\* | 0.476\*\*\* | 0.51\*\*\* | 0.39\*\*\* | 0.473\*\*\* | 0.505\*\*\* | 0.448\*\*\* | 0.248\*\*\* | 0.273\*\*\* | -0.209\*\*\* | 0.094 |
| **HDIA** | -0.184\*\*\* | -0.144\*\* | 0.788\*\*\* | 1 | 0.543\*\*\* | 0.467\*\*\* | 0.416\*\*\* | 0.57\*\*\* | 0.607\*\*\* | 0.507\*\*\* | 0.366\*\*\* | 0.162\*\* | -0.225\*\*\* | -0.078 |
| **STG** | -0.025 | -0.154\*\* | 0.476\*\*\* | 0.543\*\*\* | 1 | 0.185\*\*\* | 0.165\*\* | 0.526\*\*\* | 0.435\*\*\* | 0.385\*\*\* | 0.145\*\* | 0.031 | -0.215\*\*\* | -0.026 |
| **NLPP** | -0.196\*\*\* | -0.161\*\* | 0.51\*\*\* | 0.467\*\*\* | 0.185\*\*\* | 1 | 0.2\*\*\* | 0.326\*\*\* | 0.297\*\*\* | 0.233\*\*\* | 0.306\*\*\* | 0.145\*\* | -0.175\*\* | -0.042 |
| **DTM** | 0.041 | 0.227\*\*\* | 0.39\*\*\* | 0.416\*\*\* | 0.165\*\* | 0.2\*\*\* | 1 | 0.313\*\*\* | 0.239\*\*\* | 0.183\*\*\* | 0.257\*\*\* | 0.29\*\*\* | -0.29\*\*\* | -0.077 |
| **BYLD** | 0.014 | -0.039 | 0.473\*\*\* | 0.57\*\*\* | 0.526\*\*\* | 0.326\*\*\* | 0.313\*\*\* | 1 | 0.661\*\*\* | 0.584\*\*\* | 0.237\*\*\* | 0.158\*\* | -0.49\*\*\* | -0.057 |
| **SYP** | 0.001 | -0.018 | 0.505\*\*\* | 0.607\*\*\* | 0.435\*\*\* | 0.297\*\*\* | 0.239\*\*\* | 0.661\*\*\* | 1 | 0.922\*\*\* | 0.301\*\*\* | 0.156\*\* | 0.094 | 0.012 |
| **NSH** | 0.052 | 0.054 | 0.448\*\*\* | 0.507\*\*\* | 0.385\*\*\* | 0.233\*\*\* | 0.183\*\*\* | 0.584\*\*\* | 0.922\*\*\* | 1 | -0.026 | 0.163\*\* | 0.137\* | 0.042 |
| **HSW** | -0.15\*\* | -0.164\*\* | 0.248\*\*\* | 0.366\*\*\* | 0.145\*\* | 0.306\*\*\* | 0.257\*\*\* | 0.237\*\*\* | 0.301\*\*\* | -0.026 | 1 | 0.022 | -0.074 | -0.149\*\* |
| **VWT** | -0.021 | 0.134\* | 0.273\*\*\* | 0.162\*\* | 0.031 | 0.145\*\* | 0.29\*\*\* | 0.158\*\* | 0.156\*\* | 0.163\*\* | 0.022 | 1 | -0.082 | 0.112\* |
| **HI** | -0.012 | -0.076 | -0.209\*\*\* | -0.225\*\*\* | -0.215\*\*\* | -0.175\*\* | -0.29\*\*\* | -0.49\*\*\* | 0.094 | 0.137\* | -0.074 | -0.082 | 1 | 0.074 |
| **OC** | -0.009 | -0.022 | 0.094 | -0.078 | -0.026 | -0.042 | -0.077 | -0.057 | 0.012 | 0.042 | -0.149\*\* | 0.112\* | 0.074 | 1 |

*\*, \*\* and \*\*\* means significant at 1%, 5% and 10% level of significance respectively.*

*where,* ***DFI*** *= days to flowering initiation (in days),* ***DTF*** *= days to fifty percent flowering (in days),* ***HGT*** *= plant height (in cm),* ***HDIA*** *= head diameter (in cm),* ***STG*** *= stem girth (in cm),* ***NLPP*** *= number of leaves per plant,* ***DTM*** *= days to maturity (in days),* ***BYLD*** *= biological yield per plant (in grams),* ***SYP*** *= seed yield per plant or economic yield per plant (in grams),* ***NSH*** *= number of seeds per head,* ***HSW*** *= hundred seed weight (in grams),* ***VWT*** *= seed volume weight (in grams /100ml),* ***HI*** *= harvest index (in %),* ***OC*** *= oil content (in %).*

In this study, days to flowering initiation exhibited a strong positive correlation with days to fifty per cent flowering (0.709), aligning with the previous findings of Rao, (1987) and Tariq et al., (1992) while it was negatively associated with plant height (-0.237), head diameter (-0.184), number of leaves per plant (-0.196) and hundred seed weight (-0.15). These findings suggest that earlier flowering genotypes tend to have smaller plant structures, which is in opposition to the findings of Reddy and Kumar, (1996).

Similarly, days to fifty per cent flowering showed a positive correlation with days to maturity (0.227) and seed volume weight (0.134), which shows that delayed flowering might enhance the yield potential, in accordance with the findings of Chikkadevaiah et al., (2002) and Kaya et al., (2007). But still, this character shows a negative association with key structural traits like plant height (-0.144), head diameter (-0.144), stem girth (-0.154) number of leaves per plant (-0.161), hundred seed weight (-0.164), reinforcing that early flowering genotypes have smaller plant structure, similar to the findings of Singh and Kakar, (1997) and Habib et al., (2007).

Plant height was further a key determinant of yield-related traits, showing a positive correlation with the head diameter (0.788), stem girth (0.476), number of leaves per plant (0.51), days to maturity (0.39), biological yield per plant (0.473), seed yield per plant (0.505), number of seeds per head (0.448), hundred seed weight (0.248) and seed volume weight (0.243). These results align with the reports of Gill et al., (1997), Khan et al., (2007), Ilahi et al., (2009), Patil, (2011), Rao, (2013), Kumari et al., (2012), Iqbal et al., (2013) and Sincik and Goksoy, (2014) who also identified that positive relationships of plant height with these traits indicate that taller plants with robust structural features contribute directly to higher yield, however, it did show a negative correlation with harvest index (-0.012), due to increasing in biological yield with increase in plant height. However, this character had no significant correlation with oil content as opposed to the study of Arshad et al., (2019).

Similarly, head diameter had a positive influence on plant height (0.788), stem girth (0.543), number of leaves per plant (0.467), days to maturity (0.416), biological yield per plant (0.57), seed yield per plant (0.607), number of seeds per head (0.507), hundred seed weight (0.366) and seed volume weight (0.162), but negative influence on days to flowering initiation (-0.184), days to fifty per cent flowering (-0.144) and harvest index (-0.225), reinforcing its role as a critical yield parameter, similar to the reports of Doddamani et al., (1997), Lal et al., (1997) Amorim et al., (2008), Machikowa and Saetang, (2008), Darvishzadeh et al., (2011) and Sujatha and Nadaf, (2013).

Further, stem girth was another vital structural trait showing a significant positive association with yield parameters like biological yield per plant (0.526), seed yield per plant (0.435), number of seeds per head (0.385) and hundred seed weight (0.145), in accordance to the findings of Behradfar et al., (2009) and Jockovic et al., (2015) but negative correlation with harvest index (0.215).

Number of leaves on the main stem correlated positively with traits such as plant height (0.51), head diameter (0.467), stem girth (0.185), days to maturity (0.2), biological yield per plant (0.326), seed yield per plant (0.297), number of seeds per head (0.233), hundred seed weight (0.306) and seed volume weight (0.145) and negatively correlated with harvest index (-0.175), further underscoring its importance in sunflower growth and yield, and reinforcing the conclusions of Teklewold et al., (2000) and Vidhyavathi et al., (2005).

On the other hand, days to maturity exhibited strong positive correlations with biological yield per plant (0.313), seed yield per plant (0.661), number of seeds per head (0.584), hundred seed weight (0.237) and seed volume weight (0.29), suggesting that genotypes with extended growth periods can achieve better yields, as supported by Arshad et al., (2007), who linked extended maturity to enhanced seed yield.

Biological Yield per plant, a significant yield determinant, showed robust association with seed yield per plant (0.661), number of seeds per head (0.584), hundred seed weight (0.237) and seed volume weight (0.158). While, Seed yield per plant (economic yield) was positively correlated with plant height (0.505), head diameter (0.607), stem girth (0.435), biological yield per plant (0.661), number of seeds per head (0.922), hundred seed weight (0.301) and seed volume weight (0.156), indicating that these traits contribute significantly to economic yield and so improvement in these traits will lead direct improvement in the complex character like seed yield. A similar trend was also followed by the trait number of seeds per head. These associations align with studies by Yasin and Singh, (2010), Dan et al., (2012) and Akcay and Tekin, (2018), emphasizing the contribution of structural and yield components to productivity.

However, traits like hundred seed weight showed a trade-off, being positively correlated with the yield parameters but negatively associated with oil content (-0.149), as also reported by Gjorgjieva et al., (2015). While seed volume weight, an indicator of seed density, was positively associated with all the yield parameters and oil content (0.112).

Oil content showed a moderate positive correlation with seed volume weight (0.112) but was negatively associated with head seed weight (-0.149), highlighting the inherent trade-offs between seed size and oil yield, as presented in a similar study by Zia et al., (2013), Kaya and Atakisi, (2016), Singh et al., (2018) and Liu et al., (2020).

**Table 2.: Correlation coefficient among oil content various fatty acid parameters**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Traits of Variation** | **Oil Content (%)** |  **Palmitic Acid (%)** | **Stearic Acid (%)** | **Oleic Acid (%)** | **Linoleic Acid (%)** |
| **Oil Content (%)** | 1 | 0.03 | 0.01 | 0.06 | 0.05 |
| **Palmitic Acid (%)** | 0.03 | 1 | 0.12\* | 0.4\*\*\* | 0.34\*\*\* |
| **Stearic Acid (%)** | 0.01 | 0.12\* | 1 | 0.14\*\* | 0.03 |
| **Oleic Acid (%)** | 0.06 | 0.4\*\*\* | 0.14\*\* | 1 | 0.99\*\*\* |
| **Linoleic Acid (%)** | 0.05 | 0.34\*\*\* | 0.03 | 0.99\*\*\* | 1 |

*\*, \*\* and \*\*\* means significant at 1%, 5% and 10% level of significance respectively.*

 Fatty acid composition analysis revealed intricate interrelations among the components, while no correlation between oil content and fatty acid profile was observed. Palmitic acid was positively correlated with stearic acid (0.12), oleic acid (0.4), and linoleic acid (0.34), while stearic acid was positively correlated with palmitic acid (0.12), oleic acid (0.14) and linoleic acid (0.03). These relationships reflect similar trends reported by Rather et al., (1998) and Tilak et al., (2016).

The strongest correlation was observed between oleic acid and linoleic acid (0.99), indicating a tight relationship between these two key unsaturated fatty acids. This finding aligns with studies by Hladni et al., (2011) and Kholghi et al., (2011), who emphasized that the balance between oleic acid and linoleic acid is a critical factor in determining sunflower oil quality, as high oleic acid content typically corresponds to low linoleic acid levels and vice versa. These findings provide valuable insights into breeding strategies aimed at optimizing both oil content and fatty acid profiles in sunflower genotypes.

* 1. **Principal component analysis of oil content and other fatty acids**

The principal component analysis revealed that the eigenvalue of the first three components is more than 0.827, thereby contributing to about 93% of the variability in 113 genotypes, for five traits. Thereby, these three components capture the majority of the information present in the original variables, making them suitable for dimensional reduction while retaining most of the dataset’s variability. Similar studies on sunflower oil composition and PCA analyses haave confirmed the effectiveness of this method in identifying the key traits (Pérez-Vich et al., 1998 and Manin et al., 2021).

**Table 3.: EIGEN values of 5 Principal components**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Principal Components** | **Dim.1** | **Dim.2** | **Dim.3** | **Dim.4** | **Dim.5** |
| Oil Content  | 0.236 | -0.782 | 0.490 | 0.305 | -0.003 |
| Palmitic Acid | -0.199 | 0.615 | 0.753 | -0.119 | 0.011 |
| Stearic Acid | 0.534 | 0.734 | -0.111 | 0.405 | 0.007 |
| Oleic Acid  | 0.975 | -0.122 | 0.013 | -0.163 | 0.091 |
| Linoleic Acid | -0.979 | -0.034 | -0.083 | 0.156 | 0.091 |
| **EIGEN value** | 2.290 | 1.545 | 0.827 | 0.322 | 0.017 |
| **variance (%)** | 45.802 | 30.890 | 16.541 | 6.431 | 0.335 |
| **cumulative variance (%)** | 45.802 | 76.693 | 93.234 | 99.665 | 100.000 |

The significance of characteristics related to the principal components becomes apparent when examining the associated eigenvalues as outlined in Table 3. The eigenvalues, thereby variability decreases gradually in association with each principal component, and the first principal component (Dim.1) captures the highest variance (45.802%) and is the most significant component in explaining the variance in data, and it strongly shows high negative correlations with linoleic Acid and high positive correlations with oleic Acid, followed by stearic acid. This indicates that the primary axis of variation is driven by the balance between oleic acid and linoleic acid, with contribution from stearic acid in this component. These findings are consistent with the work of the Acuña and Natera, (2007) and Ullah, (2018), who also identified that oleic acid and linoleic acid are key determinants of first principal component, however, it contradicts the results of Ahmadian et al., (2019), who found that palmitic and stearic acid were primary contributors to the first principal component, while oleic and linoleic acid dominated the second component. Moreover, strong negative correlation of linoleic acid with saturated acids like palmitic, stearic and arachidic acid was also explained by Hosni et al., (2022) through PCA.

While, the second principal component (Dim.2) captures the second-highest variance (30.890%) and contributes to a cumulative variance of 76.693% when combined with Dim.1, and indicates a high negative correlation with oil content and positive correlations with palmitic acid and stearic acid. When compared with Dim.1, the Dim.2 represents an entirely different set of combinations of variables of being positively or negatively correlated with the trait, therefore, capturing another significant aspect of the data variation, and suggesting that the axis of variation is dominated by the contrast between oil content and a combination of palmitic and stearic acid, in accordance to the pattern suggested by Chernova et al., (2021).

The third principal component (Dim.3) explains 16.541% of the variance and, when combined with the first two components, accounts for 93.234% of the total variance, and is influenced particularly by positive correlations with oil content and palmitic acid, therefore axis of variation is focused on their relationship. Additionally, the fourth principal component (Dim.4) seems to have very little variability when compared with the first three components and shows relatively lower correlations with these traits, however, it also represents an aspect of data that these traits contribute. The fifth principal component (Dim.5) might not significantly contribute to explaining the variance compared to the others and seems to have very low correlations with all traits except for a small positive correlation with oleic acid and linoleic acid.

|  |  |
| --- | --- |
| **Figure 1.: Scatter - Plot between first principal component (Dim.1) and second principal component (Dim.2)** | **Figure 2.: Biplot between first principal component (Dim.1) and second principal component (Dim.2)** |
|  |  |

Further, as the first two principal components capture about 76.7% of the variance, the scatter-plot analysis (Figure 1) and bi-plot analysis (Figure 2) is being done to indicate the strength and the type of correlation of each variable with these two principal components. In the scatter plot, the variable oil content (OC) and linoleic acid (LA) point towards the left, showing a strong negative correlation with Dim.2, and oleic acid (OA) points towards the right, showing a positive correlation with Dim.1, while the arrow for the palmitic acid (PA) points towards top-left, showing a negative correlation with the Dim.1 and positive correlation with the Dim.2. Similarly, variable stearic acid (SA) points toward the top right, thereby, suggesting positive correlation with both the components. Moreover, data points for 113 genotypes show that the data points closer to the origin have a low score for both these components, thereby they have a closer average to the original variables, while data points away from the origin have a high score for both the components, thereby they have extreme values of the original variable. From this, the inference can be drawn that genotypes (or datatypes) at the right of the plot (positive Dim.1) are associated with high oleic acid and lower linoleic acid content, while at the left of the plot (negative Dim.1) are associated with high linoleic acid content and low oleic acid content. Similarly, genotypes (or datatypes) at the top of the plot (positive Dim.2) are associated with high palmitic and stearic acid content, while at the bottom of the plot (negative Dim.2) are associated with low oil content and varied levels of other fatty acids. Similarly, Zimmerman and Fick, (1973) also explained that decrease in oleic acid content with increasing linoleic and palmitic acid content, through this method.

Similar, to the above biplot analysis describes the relation of various components with the first two principal components, where the direction and colour of the arrow show the type and strength of correlation, thereby similar to the above results oleic acid (OA) has strong positive correlation, linoleic acid (LA) has strong negative correlation and oil content (OC) has moderate negative correlation with the first principal component (Dim.1), indicating that high oleic acid and high linoleic acid have reverse relationship i.e., high oleic acid will be accompanied by low linoleic acid and vice versa. While both, palmitic acid (PA) and stearic acid (SA) have a positive correlation with the second principal component (Dim.2), thereby, indicating that genotypes with high palmitic acid will have high stearic acid also.

4. Conclusion

This study elucidates the interrelationship among yield, oil content and fatty acid composition in sunflower, emphasizing the importance of selecting structural traits like plant height, head diameter and biological yield to be reliable indicators of advanced seed yield while addressing the trade-off with the oil quality. The fatty acid composition analysis underscored the critical balance between oleic acid and linoleic acid. The principal component analysis further streamlined trait evaluation by identifying the most influential variables, with the first two components accounting for 76.7% of the variation in oil-quality traits, thereby reinforcing the utility of multivariate techniques in simplifying complex datasets.

In conclusion, the combined use of correlation analysis and PCA offers a robust framework for understanding sunflower trait relationships, enabling the development of genotypes that balance yield, oil content and oil quality.

Consent

Not applicable

Ethical approval

Not applicable

References

1. Acuña, Auristela. Malavé., & Natera, Jesús. Rafael. Méndez. (2007). Lipid Composition of Sunflower Seeds (*Helianthus Annuus* L.) Using Multivariate Analysis. *UDO Agricultural Review*, 6 (1), 27-32. <http://hdl.handle.net/1807/45363>.
2. Ahmadian, S., Enferadi, T., & Alemzadeh, A., (2019). Assessment of Genetic Diversity of Cultivated Sunflower in Terms of Oil Content, Fatty Acid Compositions and Seed Traits. *Helia,* 42, 229 - 246. <https://doi.org/10.1515/helia-2019-0009>.
3. Akcay, S., & Tekin, M., (2018). Evaluation of relationships between yield and yield components in sunflower (*Helianthus annuus L*.). *Turkish Journal of Field Crops,* 23(1), 54-62.
4. Al-Jibouri, H. A., Miller, P. A., & Robinson, H. F., (1958). Genotypic and Environmental Variances and Covariances in an Upland Cotton Cross of Interspecific Origin. *Journal of Agronomy,* 50(10), 633-36. <https://doi.org/10.2134/agronj1958.00021962005000100020x>
5. Amorim, E. P., Ramos, N. P., Ungaro, M. R. G., & Kiihl. A., (2008). Correlation and path analysis in sunflower. Bragontia, *Campions,* 67(2), 307-16.
6. Arshad, M., Jan, S., Awan, S., Azam, S., Khalid, S., & Khan, M. A. (2019). Investigation of genetics divergence in newly developed local sunflower (*Helianthus annuus* L.) hybrids. *Pakistan Journal of Agriculture Research,* 32(1), 33-40. <http://dx.doi.org/10.17582/journal.pjar/2019/32.1.33.40>.
7. Arshad, M., Kashif, M., & Ayub, Khan. M. (2007). Genetic Divergence and Path Coefficient Analysis For Seed Yield Traits In Sunflower (*Helianthus annuus* L.) Hybrids. *Pakistan Journal of Botany*, 39(6), 2009-2015. [http://www.pakbs.org/pjbot/PDFs/39(6)/PJB39(6)2009.pdf](http://www.pakbs.org/pjbot/PDFs/39%286%29/PJB39%286%292009.pdf)
8. Behradfar, A., Gorttapeh, A. H., Zardashty, M. R., & Talat, F. (2009). Evaluation Correlated Traits for Seed and Oil Yield in Sunflower (*Helianthus annuus* L.) Through Path Analysis in under Condition Relay Cropping*. MEDWELL Research Journal of Biological Sciences,* 4(1), 82-85. <https://www.researchgate.net/publication/259383353>
9. Chernova, A. I., Gubaev, R. F., Singh, A., Sherbina, K., Goryunova, S. V., Martynova, E. U., Goryunov, D. V., Boldyrev, S. V., Vanyushkina, A. A., Anikanov, N. A., Stekolshchikova, E. A., Yushina, E. A., Demurin, Y. N., Mukhina, Z. M., Gavrilova, V. A., Anisimova, I. N., Karabitsina, Y. I., Alpatieva, N. V., Chang, P. L., Khaitovich, P., Mazin, P. V., & Nuzhdin, S. V. (2021). Genotyping and lipid profiling of 601 cultivated sunflower lines reveals novel genetic determinants of oil fatty acid content. *BMC Genomics,* 22, 505 <https://doi.org/10.1186/s12864-021-07768-y>.
10. Chikkadevaiah, H. C., Sujatha, H. L., &Nandini, R. C. (2002). Correlation and Path Analysis in Sunflower. *Helia*, 25(37), 109-118. <https://doi.org/10.2298/hel0237109c>.
11. Dan, G., Manivannan, N., & Vindhiyavarman, P. (2012). Genetic divergence of inbred lines in sunflower (*Helianthus annuus* L.). *Electronic Journal of Plant Breeding*, 3(1), 652-56. [*https://www.ejplantbreeding.org/index.php/EJPB/article/view/523*](https://www.ejplantbreeding.org/index.php/EJPB/article/view/523)*.*
12. Darvishzadeh, R., Maleki, H. H., & Sarrafi, A. (2011). Path Analysis of the Relationships between Yield and Some Related Traits in Diallel Population of Sunflower (*Helianthus annuus* L.) under Well-Watered and Water-Stressed Conditions. *Australian Journal of Crop Science* 5(6), 674-80. [https://search.informit.org/doi/abs/10.3316/informit. 281912150560452](https://search.informit.org/doi/abs/10.3316/informit.%20281912150560452)
13. Doddamani, I. K., Patil, S. A., & Ravikumar, R. L. (1997). Relationship of autogamy and self fertility with seed yield and yield components in sunflower (*Helianthus annuus* L.). *Helia* 20(26), 95-102. <https://agris.fao.org/agris-search/search.do?recordID=YU1998000021>.
14. Gill, H. S., Sheoran, R. K., Chandra, N. & Kumar, L. (1997). Inter-character association and path coefficient analysis in sunflower (*H. annuus* L.). *Annals of Biology,* 13(2), 275-277.
15. Gjorgjieva, B., Karov, I., Mitrev, S., Ruzdik, N. M., Kostadinovska, E., & Kovacevik, B. (2015). Correlation and Path Analysis in Sunflower (*Helianthus annuus* L.). *Helia* 38(63), 201-10. <https://doi.org/10.1515/helia-2015-0008>.
16. Habib, H., Sadaqat, Mehdi. S., Ashfaq, Anjum. M., & Ahmad, R. (2007). Genetic Association and Path Analysis for Oil Yield in Sunflower (*Helianthus annuus* L.). *International Journal of Agriculture and Biology* 9(2), 359-61. <http://www.fspublishers.org>.
17. Hladni, N., Jocic, S., Miklic, V., Saftic-Pankovic, D., & Kraljevic-Balalic, M. (2011). Interdependence of yield and yield components of confectionary sunflower hybrid. *Genetika*, 43(3), 583-94.
18. Hosni, T., Abbes, Z., Abaza, L., Medimagh, S., Salah, B., & Kharrat, M. (2022). Biochemical Characterization of Seed Oil of Tunisian Sunflower (*Helianthus annuus* L.) Accessions with Special Reference to Its Fatty Acid Composition and Oil Content*. Journal of Food Quality*. <https://doi.org/10.1155/2022/2875072>.
19. Husson, F., Josse, J., Lˆe, S., & Mazet, J. (2009). FactoMineR: Multivariateexploratory data analysis and data mining with R. R package version 1.12, URL http://factominer.free.fr.
20. Ilahi, F., Nadeem, M. H. & Sadaqat, H. A. (2009). Correlation and path co-efficient analysis for achene yield and yield components in sunflower. *Pakistan Journal of Agricultural Science* 46(1), 20-24.
21. Iqbal, M., Ijaz, U., Smiullah, I. M., Mahmood, K., Najeebullah, M., Abdullah, N. S., & Sadaqat, H. A. (2013). Genetic divergence and path coefficient analysis for yield related attributes in sunflower (*Helianthus annuus* l.) Under less water conditions at productive phase. *Plant Knowledge Journal* 2(1), 20-23.
22. Jockovic, M., Jocic, S., Jeromela, A. M., Ciric, M., Canak, P., Miklic, V., & Cvejic, S. (2015). Biomorphological association and path analysis in sunflower (*Helianthus annuus* L.). *Helia* 38(63), 189-99. <https://doi.org/10.1515/helia-2015-0004>.
23. Kaya, Y., & Atakisi, I. K. (2016). The effects of seed weight on the yield and yield components of sunflower (*Helianthus annuus L.).* *Helia,* 39(65): 87-98.
24. Kaya, Y., Evc, G., Durak, S., Pekcan, V., & G cer,T. (2007). Determining the Relationships between Yield and Yield Attributes in Sunflower. *Turkish Journal of Agriculture*, 31(4), 237-44. <https://journals.tubitak.gov.tr/cgi/viewcontent.cgi?article=2009andcontext=agriculture>.
25. Khan, H., Muhammad, S., Shah, R., & Iqbal, N. (2007). Genetic analysis of yield and some yield components in sunflower. *Sarhad Journal of Agriculture*, 23(4), 985-90.
26. Kholghi, M., Bernoussi, I., Darvishzadeh, R., & Pirzad, A. (2011). Correlation and path coefficient analysis of seed yield and yield related trait in Iranian confectionery sunflower populations. *African journal of Biotechnology,* 10(61), 13058-63. <https://doi.org/10.5897/AJB11.1452>.
27. Kumari, Soni., Sheoran, R. K., Singh, Amit., & Avtar, Ram. (2012). Genetic variability, heritability and character association in sunflower. *Research on Crops*, 13(2), 566-72. <http://www.cropresearch.org>
28. Lal, G. S., Bhadoriya, V. S., & Singh, A. K. (1997). Genetic association and path analysis in elite lines of sunflower. *Crop Research,* 13(3), 6. <https://www.cabidigitallibrary.org/doi/full/10.5555/19981604778>.
29. Liu, Z., Zhang, S., & Su, Y. (2020). Genetic analysis of oil content and fatty acid composition in sunflower (*Helianthus annuus* L.). *Industrial Crops and Products* 145, 112101.
30. Machikowa, T., & Saetang, Chirporn. (2008). Correlation and path coefficient analysis on seed yield in sunflower. *Suranaree Jouranl of Science and Technology*, 13(2), 243-48. <https://www.researchgate.net/publication/242390885>.
31. Manin, L., Pizzo, J., Rydlewski, A., Santos, P., Galuch, M., Zappielo, C., Santos, O., & Visentainer, J. (2021). Evaluation of the Adulteration of Edible and Cosmetic Sunflower Oils by GC-FID and ESI-MS. *Research on Crops*, 32 (11), 2086-2094. <https://doi.org/10.21577/0103-5053.20210101>.
32. Panse, V. G., and Sukhatme, P. V. (1985). Statistical methods for agricultural research. ICAR, New Delhi, 8, 308-318.
33. Patil, L. C. (2011). Research Note Correlation and path analysis in sunflower populations. *Electronic Journal of Plant Breeding*, 2(3), 442-47. <http://sites.google.com/site/ejplantbreeding>.
34. Pearson, Karl. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559-72. <https://pca.narod.ru/pearson1901.pdf>.
35. Pérez-Vich, B., Velasco, L., & Fernández-Martínez, J. (1998). Determination of seed oil content and fatty acid composition in sunflower through the analysis of intact seeds, husked seeds, meal and oil by near-infrared reflectance spectroscopy. *Journal of the American Oil Chemists' Society,* 75, 547-555. <https://doi.org/10.1007/S11746-998-0064-1>.
36. R Core Team. (2013). R: A language and environment for statistical computing [Software]. Available from <http://www.r-project.org/>.
37. Rao, N. G. L. (1987). Studies on correlation and path coefficient analysis in sunflower. *Mysore Journal of Agriculture Science*, 21, 94-95.
38. Rao, V. T. (2013). Association analysis in sunflower (*Helianthus annuus* L.). *International Journal of Applied Biology and Pharmaceutical Technology,* 4(1), 1-4. <https://www.researchgate.net/publication/349442091_Association_analysis_in_sunflower>.
39. Rather, A. G., Sandha, G. S., Shafiq, A., Wani., & Seikh, F. A. (1998). Correlation and path analysis in sunflower (*Helianthus annuus* L.). *Journal of Oilseed Research*, 15(1): 67-70.
40. Reddy, M. D., & Kumar, K. A. (1996). Performance of sunflower at different times of sowing during post rainy season in North Telangana Zone of Andhra Pradesh. *Journal of Oilseed Research*, 13(2): 313-16. <https://www.researchgate.net/publication/348482497>.
41. Škorić, D., & Sakač, Z. (2013). Significance of sunflower seeds (oil) in proper human nutrition. At the same time, nutrition, challenges, doubts, deceptions and fear. SASA, Branch of SASA, Novi Sad. Pp. 30-42. 41
42. Sincik, M., & Goksoy, A. T. (2014). Investigation of Correlation between Traits and Path Analysis of Confectionary Sunflower Genotypes. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca,* 42(1). <https://doi.org/10.15835/nbha4219429>.
43. Singh, R. K., & Chaudhary, B. D. (1977). Biometrical methods in quantitative genetic analysis. Kalyani Publishers, New Delhi.
44. Singh, R. K. & Kakar, S. N. (1997). Control on individual traits means during index selection. 3rd Congress SABRAO, Canberra, Pp. 22-25.
45. Singh, V. K., Sheoran, R. K. & Subhash, Chander. (2018). Correlation analysis for seed yield and its component traits in sunflower. *Journal of Pharmacognosy and Phytochemistry*, 7(3), 2299-2301. <https://www.phytojournal.com/archives/2018/vol7issue3/PartAE/7-3-228-855.pdf>.
46. Sujatha, K., & Nadaf, H. L. (2013), Correlation for yield and yield related trait in mutant and segregating genotypes in sunflower (*Helianthus annuus* L.). *Molecular Plant Breeding*, 4(32), 265-66. https://doi.org/[0.5376/mpb.2013.04.0032](http://dx.doi.org/10.5376/mpb.2013.04.0032).
47. Tariq, M., Drees, G., & Tahir, A. (1992). Genetic variability and correlation studies in sunflower. *Sarhad Journal of Agriculture*, 8(6), 659-63. <https://agris.fao.org/agris-search/search.do?recordID=PK9600698>.
48. Teklewold, A. H., Jayaramaiah., & Jagadeesh, B. N. (2000). Correlation and Path analysis of physio morphological characters of sunflower (*Helianthus annuus* L.) as related to breeding methods. *Helia,* 23(32), 105-114. https://www.cabdirect.org/ cabdirect/abstract/20013034487.
49. Tilak, I. S., Kisan, B., Patil, A., Shankergoud, I., Kadirvel, Ghante, V., & Kulkarni, V. (2016). Genotypic and phenotypic correlations and path coefficient analysis for oleic acid content in sunflower. *Karnataka Journal of Agricultural Science,* 29(4): 476-78. <https://www.researchgate.net/publication/325846189>.
50. Ullah, Z. (2018). Grouping and selection of 32 single cross sunflower hybrids using principal component analysis.  *Eurasian Journal of Agricultural Research*, 2, 4-12. <https://dergipark.org.tr/en/pub/ejar/issue/39001/449230>.
51. Vidhyavathi, R., Mahalakshmi, P., Manivannan, N., & Muralidharan, V. (2005). Correlation and path analysis in sunflower (*Helianthus annuus* L.). *Agricultural Science Digest,* 25(1): 6-10.
52. Yasin, A. B., & Singh, S. (2010). Correlation and path coefficient analyses in sunflower. *Journal of Plant Breeding and Crop Science*, 2(5): 129-33. <http://www.academicjournals.org/jpbcs>
53. Zia, U. Z., Sadaqat, H. A., Tahir, M. H. N., & Sadia, B. (2013). Path coefficient analysis of various traits in sunflower (*Helianthus annuus* L.). *Journal of Global Innovations in Agricultural and Social Sciences*, 1(1), 5-8.
54. Zimmerman, D., & Fick, G. (1973). Fatty acid composition of sunflower (*Helianthus annuus* L.) oil as influenced by seed position. *Journal of the American Oil Chemists' Society,* 50, 273-275. <https://doi.org/10.1007/BF02641352>.