



International Journal of Research in Agronomy

E-ISSN: 2618-0618
P-ISSN: 2618-060X
© Agronomy
NAAS Rating (2026): 5.20
www.agronomyjournals.com
2026; 9(6): 155-167
Received: 27-05-2026
Accepted: 30-06-2026

J Marimuthu
Ph.D. Scholar, Department of Vegetable
Science, Horticultural College and Research
Institute, Tamil Nadu Agricultural University
(TNAU), Coimbatore, Tamil Nadu India

R Gangai Selvi
Professor (Statistics), Department of Physical
Sciences and Information Technology,
Agricultural Engineering College and Research
Institute, Tamil Nadu Agricultural University
(TNAU), Coimbatore, Tamil Nadu India

B Gokulakrishnan
Ph.D. Scholar, Department of Vegetable
Science, Horticultural College and Research
Institute, Tamil Nadu Agricultural University
(TNAU), Coimbatore, Tamil Nadu India.

S Thaher Ahmed
Ph.D. Scholar, Department of Vegetable
Science, Horticultural College and Research
Institute, Tamil Nadu Agricultural University
(TNAU), Coimbatore, Tamil Nadu India

Monishsriram R R
Ph.D. Scholar, Department of Vegetable
Science, Horticultural College and Research
Institute, Tamil Nadu Agricultural University
(TNAU), Coimbatore, Tamil Nadu India

S Pushparaj
Ph.D. Scholar, Department of Agronomy,
Agricultural College and Research Institute,
Tamil Nadu Agricultural University (TNAU),
Coimbatore, Tamil Nadu India

Gokulnath G
Ph.D. Scholar, Department of Agronomy,
Agricultural College and Research Institute,
Tamil Nadu Agricultural University (TNAU),
Coimbatore, Tamil Nadu India.

R J Parthiban
Ph.D. Scholar, Department of Agronomy,
Agricultural College and Research Institute,
Tamil Nadu Agricultural University (TNAU),
Coimbatore, Tamil Nadu India

V V V Vimal Rajan
Ph.D. Scholar, Department of Genetics and
Plant Breeding, Agricultural College and
Research Institute, Tamil Nadu Agricultural
University (TNAU), Coimbatore, Tamil Nadu,
India

Corresponding Author:

R. Gangai Selvi
Professor (Statistics), Department of
Physical Sciences and Information
Technology, Agricultural Engineering
College and Research Institute, Tamil
Nadu Agricultural University (TNAU),
Coimbatore, Tamil Nadu India

Deconstructing vegetable productivity: a review on correlation and path coefficient analysis for yield components

**J Marimuthu, R Gangai Selvi, B Gokulakrishnan, S Thaher Ahmed
Monishsriram RR, S Pushparaj Gokulnath G, RJ Parthiban and V V V Vimal
Rajan**

DOI: <https://www.DOI.org/10.33545/2618060X.2026.v9.i7c.5911>

Abstract

Vegetable crop productivity is a complex quantitative trait governed by the interaction of multiple morphological, physiological, and reproductive components. Correlation and path coefficient analyses are widely used biometrical tools in plant breeding to understand the nature and magnitude of associations among yield and its contributing traits. This review critically examines the role of correlation and path analysis in vegetable crops for identifying key yield-determining traits and improving selection efficiency in breeding programmes. Across major vegetable groups including solanaceous, cucurbitaceous, leguminous, cole, and root-bulb crops, traits such as fruit or pod number, fruit weight, and plant or canopy architecture consistently exhibit strong positive associations with yield. Path coefficient analysis further reveals that these traits exert the highest direct effects, while vegetative traits such as plant height, leaf area, and branching pattern generally influence yield indirectly through enhanced photosynthetic efficiency and assimilate supply. The review also highlights that trait relationships are highly influenced by genetic factors, environmental variability, genotype $\times$ environment interactions, management practices, and biotic and abiotic stress conditions. Despite their wide application, limitations exist due to environmental instability and lack of integration with modern breeding tools. Therefore, recent advances such as molecular markers, quantitative trait loci (QTL) mapping, genome-wide association studies (GWAS), genomic selection, high-throughput phenotyping, and artificial intelligence-based predictive models are increasingly being integrated with classical correlation and path analysis approaches. This integration enables more precise identification of yield-contributing traits and accelerates genetic improvement in vegetable crops. Overall, correlation and path analysis remain fundamental tools in vegetable breeding, and their combination with modern technologies offers a powerful strategy for sustainable yield enhancement under changing environmental conditions.

Keywords: Vegetable crops; correlation analysis; path coefficient analysis; yield components; plant breeding; genetic variability; genotype $\times$ environment interaction; genomic selection; qtl mapping; artificial intelligence

Introduction

Vegetable crops occupy a unique and essential position in global agriculture due to their significant contributions to food security, nutritional well-being, livelihood enhancement, and economic development. Unlike staple cereals that primarily supply calories, vegetables are rich sources of vitamins, minerals, dietary fiber, antioxidants, and a wide range of health-promoting phytochemicals. With increasing global population, rapid urbanization, changing dietary preferences, and rising awareness of healthy diets, the demand for vegetables has increased substantially worldwide. As a result, improving vegetable productivity has become a central goal of agricultural research and breeding programs. However, yield improvement in vegetable crops is challenging because yield is a complex quantitative trait governed by multiple interacting components and strongly influenced by environmental factors. Therefore, understanding the relationships among yield and its contributing traits is essential for developing efficient breeding strategies, where correlation and path coefficient analyses serve as important biometrical tools for

identifying key yield-determining characters. Vegetables play a vital role in global food and nutritional security because they provide essential micronutrients required for human health. Dias and Ryder (2011) ^[13] emphasized that vegetables are important components of balanced diets due to their high levels of vitamins, minerals, fiber, and phytochemicals. Similarly, Schreinemachers *et al.* (2018) ^[42] highlighted that vegetables are among the most affordable sources of micronutrients and are crucial for improving dietary quality, especially in developing countries. Beyond nutrition, vegetable cultivation supports rural livelihoods through employment generation, income creation, and value chain activities such as processing, packaging, and marketing. Ebert (2020) ^[14, 15, 16] further noted that vegetables are central to combating micronutrient deficiencies and hidden hunger globally, particularly in vulnerable populations.

Vegetables are also rich in bioactive compounds such as carotenoids, flavonoids, phenolics, and glucosinolates, which are associated with reduced risks of chronic diseases including cardiovascular disorders, diabetes, obesity, and certain cancers. Dias (2019) ^[11, 12] reported that regular vegetable consumption contributes significantly to disease prevention and overall health improvement. In this context, investments in vegetable research and development are considered essential for achieving nutritional security and sustainable food systems (Keatinge *et al.*, 2014) ^[25].

The global vegetable sector has experienced substantial growth over recent decades due to advances in breeding, protected cultivation, irrigation technologies, and crop management practices. Asia dominates global vegetable production, with China and India being the leading contributors. The adoption of high-yielding hybrids, drip irrigation, fertigation, and precision farming techniques has significantly increased productivity. However, Schreinemachers *et al.* (2018) ^[42] observed that considerable yield gaps still exist between actual farm yields and genetic potential, particularly in developing regions where research investment, infrastructure, and post-harvest systems remain limited. Increasing population pressure and climate variability further intensify the demand for higher and more stable vegetable production. Despite technological progress, vegetable yield improvement faces multiple constraints. Climate change, including rising temperatures, irregular rainfall, drought, flooding, and salinity stress, severely affects plant growth, flowering, and fruit development. Biotic stresses such as insect pests, diseases, and nematodes also contribute to significant yield losses. Additionally, the narrow genetic base of many cultivated vegetables limits breeding progress and reduces adaptability. Ebert and Engels (2020) ^[14, 15, 16] emphasized the importance of conserving plant genetic resources for sustaining long-term improvement. Poor soil fertility, limited irrigation, and post-harvest losses further aggravate productivity challenges. Most importantly, yield itself is a complex trait influenced by multiple genes and environmental interactions, making direct selection inefficient.

Yield in vegetable crops is considered one of the most complex quantitative traits because it results from the interaction of numerous morphological, physiological, and biochemical processes. Wright (1921) ^[49] described such traits as polygenic in nature, requiring statistical approaches for proper analysis. Yield may be expressed as fruit yield, pod yield, root yield, bulb yield, or head yield depending on the crop. It is determined by multiple components such as plant height, branching pattern, vine length, flowering behavior, fruit set, fruit number, and fruit size. Dewey and Lu (1959) ^[10] emphasized that direct selection for yield is often ineffective due to its complex nature and environmental

sensitivity. Yield component analysis therefore becomes essential in plant breeding as it helps identify traits that contribute most significantly to yield. Falconer and Mackay (1996) ^[18] noted that indirect selection through correlated traits is often more efficient than direct selection for complex traits. Correlation analysis measures the degree and direction of association between traits, while path coefficient analysis, introduced by Wright (1921) ^[49] and applied in plant breeding by Dewey and Lu (1959) ^[10], partitions correlations into direct and indirect effects. This allows breeders to identify traits that directly influence yield as well as those that act indirectly through other components.

In vegetable crops, traits such as number of fruits or pods per plant, average fruit weight, fruit diameter, branches per plant, and earliness consistently show strong associations with yield. These traits are widely used as selection criteria in breeding programs because they are often more heritable and stable than total yield itself. Singh and Chaudhary (1985) ^[46] highlighted that understanding trait relationships improves selection efficiency and accelerates genetic gain in crop improvement. The present review focuses on synthesizing available literature on correlation and path coefficient analyses across major vegetable crop groups, including solanaceous, cucurbitaceous, leguminous, cole, and root and bulb vegetables. It aims to identify key yield-contributing traits, summarize major research findings, and highlight their applications in breeding programs. The review also emphasizes integrating classical biometrical approaches with modern genomic, phenomic, and computational tools to enhance precision breeding and improve vegetable productivity under changing environmental conditions.

Concept of yield and yield components in vegetable crops

Yield in vegetable crops represents the final expression of plant productivity, resulting from the interaction of genetic, environmental, and physiological processes throughout the crop growth cycle. It is a complex quantitative trait controlled by multiple genes with small additive effects and is strongly influenced by genotype $\times$ environment interactions. Because of this complexity, direct selection for yield is often less efficient than selection based on its component traits, which are generally more heritable and stable across environments (Falconer & Mackay, 1996; Acquaah, 2012) ^[1, 28]. Yield in vegetables varies with crop type and may be expressed as fruit, pod, root, bulb, or head yield. Fehr (1987) ^[19] described yield as the cumulative outcome of vegetative growth, reproductive development, and assimilate partitioning. In fruit vegetables such as tomato, brinjal, and cucumber, yield is mainly determined by fruit number and fruit weight, while in root and bulb crops it depends on storage organ size and weight (Singh & Chaudhary, 1985) ^[46].

Morphological traits such as plant height, branching, vine length, and leaf area influence yield by regulating photosynthetic efficiency and assimilate production. Reproductive traits including flowering time, sex ratio, and fruit set percentage directly affect fruit formation and final productivity. Among yield attributes, number of fruits per plant, fruit size, and average fruit weight are the most important contributors to economic yield across vegetable crops (Dewey & Lu, 1959; Tanksley, 2004) ^[10, 48]. The interrelationship among growth traits, reproductive traits, yield components, total yield, and marketable yield is illustrated in Figure 1, which highlights the hierarchical contribution of various plant characteristics to overall vegetable productivity.

Environmental conditions such as temperature, soil fertility, and moisture availability significantly influence the expression of yield components. In addition, genotype $\times$ environment interactions often modify trait performance across locations and

seasons, making multi-environment evaluation essential for identifying stable and high-yielding genotypes (Blum, 2011; Ceccarelli, 2015) [3, 7]. As shown in Figure 1, marketable yield is the cumulative outcome of growth and reproductive processes acting through yield-contributing traits, while correlation and path coefficient analyses help identify the most influential components affecting final productivity.

Statistical approaches in yield component analysis

Statistical approaches are essential for understanding the complex relationships among yield and its component traits in vegetable crops, as summarized in Table 1. Since yield is a highly polygenic trait and is strongly influenced by environmental factors, direct selection for yield alone is often less effective. Therefore, biometrical and statistical tools such as variability analysis, correlation analysis, and path coefficient analysis are widely used in plant breeding. These approaches help in partitioning total variability into genetic and environmental components and in identifying the magnitude and direction of associations among traits, thereby improving selection efficiency in breeding programmes. According to Falconer and Mackay (1996) [18] and Acquah (2012) [1], statistical genetic tools are fundamental for interpreting quantitative variation and guiding effective selection strategies in crop improvement.

Role of biometrical genetics in crop improvement

Biometrical genetics provides the statistical framework for analyzing quantitative traits controlled by multiple genes. The development of quantitative genetics has significantly strengthened plant breeding by enabling the interpretation of complex inheritance patterns. Fisher (1918) [20] laid the foundation of modern statistical genetics, while Wright (1921) [49] further developed the concepts of correlation and path analysis. Singh and Chaudhary (1985) [46] emphasized that biometrical methods are essential for understanding complex traits like yield where multiple component traits interact simultaneously. In vegetable crops, biometrical genetics is widely used to estimate variability, heritability, and genetic advance, which are crucial parameters for selection. Johnson *et al.* (1955) [23] demonstrated that selection based on genetic parameters is more efficient than selection based solely on phenotypic performance. Thus, biometrical genetics serves as the backbone of modern vegetable breeding programmes by helping breeders distinguish genetic effects from environmental influences.

Variability studies

Variability is the basic requirement for crop improvement, as selection depends on the presence of sufficient genetic variation. Variability studies help in assessing the extent of variation in a population and determining the potential for selection. Phenotypic coefficient of variation (PCV) and genotypic coefficient of variation (GCV) are commonly used measures. Burton (1952) [4, 5] reported that high PCV and GCV values indicate greater scope for selection, while the difference between them reflects environmental influence. Burton and Devane (1953) [6] further emphasized the importance of genetic variability in plant breeding programmes. In vegetable crops, studies have consistently reported high variability in traits such as fruit number, fruit weight, branching pattern, and fruit size, indicating good scope for genetic improvement (Singh *et al.*, 2018) [47]. Heritability and genetic advance are also important parameters used alongside variability estimates. Allard (1999) [2] stated that traits with high heritability and high genetic advance are more reliable for effective selection in breeding programmes.

Association analysis in plant breeding

Association analysis is used to study the relationships among different traits, particularly between yield and its components. Correlation analysis, introduced by Pearson (1896) [36], is widely used in plant breeding to estimate the degree and direction of association between traits. Correlation coefficients are calculated at phenotypic and genotypic levels, where genotypic correlations are usually higher because environmental effects reduce phenotypic expression (Singh & Chaudhary, 1985) [46]. In vegetable crops, correlation analysis has been extensively used to identify traits associated with yield. Meena and Bahadur (2015) [32] reported strong positive correlations between fruit yield and traits such as fruit number and fruit weight in tomato. Similarly, Rani *et al.* (2015) [39] observed positive associations between yield and vine length in cucurbits. However, correlation analysis does not distinguish between direct and indirect effects, which limits its interpretability. To overcome this limitation, path coefficient analysis is used. Wright (1921) introduced path analysis to partition correlation into direct and indirect effects, while Dewey and Lu (1959) [10] applied it in plant breeding to identify key yield-contributing traits. In vegetable crops, traits such as fruit number per plant and fruit weight consistently show strong direct effects on yield.

Importance of trait relationship studies

Trait relationship studies are crucial for understanding the interaction among yield components and for improving selection efficiency. Since yield is a complex trait, indirect selection through correlated traits is often more effective than direct selection. Falconer and Mackay (1996) [18] highlighted that indirect selection is particularly useful when heritability of yield is low or when environmental variation is high. Path analysis further enhances selection efficiency by separating direct and indirect effects of traits on yield. Dewey and Lu (1959) [10] emphasized that traits with high direct effects should be prioritized in breeding programmes. In vegetable crops, fruit number, fruit weight, branching pattern, and fruit size are often identified as major yield contributors. Trait relationship studies also help in identifying undesirable associations, such as negative correlations between yield and late maturity or excessive vegetative growth. Johnson *et al.* (1955) [23] suggested that combining variability, correlation, and path analysis provides a powerful framework for selecting superior genotypes. Recent studies in vegetable crops further confirm the importance of these approaches. Kumar *et al.* (2019) [28] reported that integration of correlation and path analysis improves selection accuracy in cucurbits, while Sharma *et al.* (2020) [43] emphasized their role in developing high-yielding and stable vegetable varieties. Thus, statistical approaches remain fundamental tools in modern vegetable breeding for identifying key traits and improving productivity.

Correlation analysis in vegetable crops

Correlation analysis is one of the most widely used biometrical tools in plant breeding for understanding the nature and degree of association among quantitative traits. In vegetable crops, where yield is a complex outcome of many interrelated components, correlation analysis helps breeders identify traits that move together in a desirable or undesirable direction. This enables indirect selection of superior genotypes through associated characters rather than relying solely on yield, which is often highly influenced by environmental variation. The method forms the foundation for further multivariate approaches such as path coefficient analysis and regression-based selection indices.

Historical development of correlation analysis

The concept of correlation was first introduced by Galton (1888) [21] and later mathematically formalized by Pearson (1896) [36], who developed the product-moment correlation coefficient widely used in biological sciences. In plant breeding, correlation analysis gained importance with the development of quantitative genetics by Fisher (1918) [20], who integrated statistical methods into genetic studies. Later, Wright (1921) [49] extended correlation analysis into causal modeling through path coefficient analysis, enabling the partitioning of correlations into direct and indirect effects. The application of correlation in plant breeding was further strengthened by Robinson *et al.* (1951) [41] and Johnson *et al.* (1955) [23], who demonstrated its importance in selection studies. In vegetable crops, correlation analysis has since been extensively used to study yield components and their interrelationships in tomato, brinjal, chilli, cucumber, and other species.

Concept and definition of correlation

Correlation refers to the statistical measure that expresses the degree and direction of association between two variables. In plant breeding, it indicates how changes in one trait are associated with changes in another trait. According to Singh and Chaudhary (1985) [46], correlation analysis is essential for identifying traits that can be improved simultaneously or indirectly during selection. Correlation coefficients range from -1 to +1. A positive correlation indicates that both traits increase or decrease together, while a negative correlation indicates an inverse relationship. A zero correlation implies no association between traits. In vegetable crops, traits such as fruit number and yield often show strong positive correlations, whereas traits like days to maturity may show negative correlations with earliness-related yield advantages.

Types of correlation

Correlation in plant breeding is generally classified into phenotypic, genotypic, and environmental correlations, each representing different levels of association.

Phenotypic correlation

Phenotypic correlation refers to the observable association between two traits as influenced by both genetic and environmental factors. It is calculated using phenotypic variances and covariances. According to Falconer and Mackay (1996) [18], phenotypic correlation reflects the combined effect of genotype and environment and is generally lower than genotypic correlation due to environmental masking effects. In vegetable crops, phenotypic correlations are commonly used in field-based selection studies where environmental variation is unavoidable. Studies by Meena and Bahadur (2015) [32] in tomato and Rani *et al.* (2015) [39] in cucurbits reported significant phenotypic correlations between yield and traits such as fruit number, fruit weight, and plant height.

Genotypic correlation

genotypic correlation measures the association between traits at the genetic level, excluding environmental influences. It is considered more reliable for selection because it reflects true genetic relationships. Singh and Chaudhary (1985) [46] reported that genotypic correlations are usually higher than phenotypic correlations due to the removal of environmental variance. In vegetable crops, high genotypic correlation between yield and its components indicates strong genetic linkage or pleiotropy. For example, Kumar *et al.* (2018) [26, 27] observed strong genotypic

correlations between fruit yield and fruit weight in brinjal, while Sharma *et al.* (2020) [43] reported similar findings in cucumber and bitter gourd.

Environmental correlation

Environmental correlation arises due to the influence of environmental factors affecting two traits simultaneously. It reflects non-genetic association caused by shared environmental conditions. According to Falconer and Mackay (1996) [18], environmental correlations can either enhance or reduce phenotypic correlations depending on the direction of environmental effects. In vegetable crops, irrigation, temperature, soil fertility, and pest incidence often influence environmental correlations among yield traits.

Estimation of correlation coefficients

Correlation coefficients are estimated using variance and covariance components. The genotypic, phenotypic and environmental correlation coefficient is calculated as:

Genotypic correlation coefficient

$$r_g(xy) = \frac{\text{cov. g}(xy)}{\sqrt{\sigma^2 g_x \times \sigma^2 g_y}}$$

Where,

$r_g(xy)$ = Genotypic correlation coefficient between traits x and y

$\text{Cov.g}(xy)$ = Genotypic covariance between traits x and y

$\sigma^2 g_x$ = Genotypic variance of trait x

$\sigma^2 g_y$ = Genotypic variance of trait y

Phenotypic correlation coefficient

$$r_p(xy) = \frac{\text{cov. p}(xy)}{\sqrt{\sigma^2 p_x \times \sigma^2 p_y}}$$

Where,

$r_p(xy)$ = Phenotypic correlation coefficient between traits x and y

$\text{Cov}_p(xy)$ = Phenotypic covariance between traits x and y

σ^2_{px} = Phenotypic variance of trait x

σ^2_{py} = Phenotypic variance of trait y

Interpretation of correlation values

The magnitude of correlation coefficients determines the strength of association between traits. Generally:

- 0.00-0.19 → very low correlation
- 0.20-0.39 → low correlation
- 0.40-0.59 → moderate correlation
- 0.60-0.79 → high correlation
- 0.80-1.00 → very high correlation

In vegetable crops, high positive correlations between yield and traits such as fruit number, fruit weight, and branches per plant indicate their importance in selection programs. Negative correlations often indicate trade-offs between vegetative growth and reproductive development. Allard (1999) [2] emphasized that interpretation of correlation should always be done in conjunction with biological relevance rather than statistical magnitude alone.

Advantages of correlation analysis

Correlation analysis offers several advantages in plant breeding: It

helps identify traits that are genetically associated with yield and can be used for indirect selection. It simplifies selection by reducing the number of traits to be considered simultaneously. It provides insights into the direction of selection (positive or negative association). It is relatively simple and cost-effective compared to advanced genomic tools. According to Singh and Chaudhary (1985) ^[46], correlation analysis is particularly useful in early-generation selection where direct yield measurement may be unreliable. Despite its usefulness, correlation analysis has certain limitations. It does not dist

Limitations of correlation analysis

A high correlation between two traits does not necessarily imply a causal relationship. Environmental influences may distort true genetic associations. Correlation values may vary across environments and populations. Wright (1921) ^[49] highlighted that correlation alone cannot explain causal relationships, which led to the development of path coefficient analysis. Therefore, correlation should always be interpreted along with other biometrical tools.

Applications of correlation analysis in vegetable breeding

Correlation analysis has wide applications in vegetable crop improvement. It is used to identify yield-contributing traits, select parental lines, and develop selection indices. Meena and Bahadur (2015) ^[32] reported strong positive correlations between fruit yield and number of fruits per plant. In brinjal, Kumar *et al.* (2018) ^[26, 27] observed significant associations between yield and fruit weight. In chilli, Yatung *et al.* (2014) ^[51] identified positive correlations between yield and branching traits. In cucurbits such as cucumber, bottle gourd, ridge gourd, and snake gourd, traits like vine length, fruit number, and fruit weight were found to be strongly associated with yield (Rajkumari *et al.*, 2021; Priyanka *et al.*, 2020) ^[37]. Similarly, in leguminous vegetables, pod number per plant showed strong correlation with yield (Patel *et al.*, 2018). Overall, correlation analysis remains a fundamental tool in vegetable breeding programs for identifying important selection criteria and improving yield potential through indirect selection strategies.

Path Coefficient Analysis: Theory and Applications

Path coefficient analysis is a powerful biometrical tool used in plant breeding to partition correlation coefficients into direct and indirect effects of yield-contributing traits. In vegetable crops, where yield is controlled by multiple interrelated characters, path analysis provides deeper insight into cause-effect relationships among traits, helping breeders identify the most important selection criteria for improving productivity. Unlike simple correlation analysis, which only measures association, path analysis explains the actual contribution of each trait to yield, making it highly valuable in crop improvement programmes. A schematic representation of the direct and indirect effects of yield-contributing traits on vegetable crop productivity through path coefficient analysis is presented in Figure 2.

Historical background

The concept of path analysis was first developed by Wright (1921) ^[49] in the field of genetics to study causal relationships among biological variables. He introduced the idea of path coefficients to quantify direct and indirect influences among traits. Later, this method was adapted in plant breeding by Dewey and Lu (1959) ^[10], who applied it to identify yield components in crop plants. Subsequently, Li (1975) ^[30] and Singh and Chaudhary (1985) ^[46] further refined the methodology and popularized its use

in agricultural research. In vegetable crops, path coefficient analysis has been widely used to understand the genetic architecture of yield in tomato, brinjal, chilli, cucumber, and various cucurbits. Studies by Meena and Bahadur (2015) and Kumar *et al.* (2018) ^[26] demonstrated its effectiveness in identifying key traits influencing yield.

Concept of path analysis

Path analysis is a standardized partial regression technique that decomposes correlation coefficients into direct and indirect effects. It explains how a dependent variable (yield) is influenced by several independent variables (yield components). According to Wright (1921) ^[49], each correlation coefficient between two traits can be expressed as the sum of direct effects and indirect effects through other traits. In vegetable crops, this allows breeders to determine whether a trait influences yield directly or indirectly through other characters such as fruit number, fruit weight, or branching pattern. For example, fruit yield in tomato may be influenced directly by fruit weight and indirectly by plant height through increased branching and fruit production.

Direct effects

Direct effects represent the immediate influence of an independent trait on yield without the mediation of other traits. Traits with high positive direct effects are considered important selection criteria in breeding programmes. Dewey and Lu (1959) ^[10] emphasized that traits with strong direct effects should be given priority during selection because they contribute most significantly to yield improvement. In vegetable crops, traits such as fruit number per plant, average fruit weight, and branches per plant often exhibit strong direct effects on yield. For instance, Meena and Bahadur (2015) ^[32] reported that fruit weight had the highest direct effect on yield in tomato. Similarly, Rani *et al.* (2015) ^[39] observed strong direct contributions of vine length and fruit number in cucurbits.

Indirect effects

Indirect effects occur when one trait influences yield through its association with another trait. These effects may either enhance or reduce the overall correlation between traits. According to Singh and Chaudhary (1985) ^[46], indirect effects are crucial in understanding the complex relationships among yield components. For example, plant height may not directly influence yield but may contribute indirectly through increased branching or leaf area, which enhances photosynthesis and fruit production. In cucurbit crops, Rajkumari *et al.* (2021) observed that vine length influenced yield indirectly through fruit number and fruit weight. Similarly, Kumar *et al.* (2018) ^[26, 27] reported that days to flowering affected yield indirectly via fruit set percentage and fruit number. Understanding indirect effects helps breeders avoid misleading interpretations of correlation coefficients. The direct and indirect effects of yield-contributing traits across vegetable crops are further summarized in Table 2.

Residual effects

Residual effects represent the unexplained variation in the dependent variable (yield) after considering all independent variables in the path model. It indicates the influence of factors not included in the analysis. According to Li (1975) ^[30], a low residual effect suggests that the selected variables adequately explain yield variation, whereas a high residual effect indicates the presence of other important unmeasured factors. In vegetable crops, residual effects may arise due to environmental factors, genetic interactions, or unmeasured physiological traits such as

photosynthetic efficiency or nutrient uptake capacity.

Methodology of path coefficient analysis

Path coefficient analysis is carried out through the following steps:

1. Estimation of correlation coefficients among all traits.
2. Selection of yield as the dependent variable and other traits as independent variables.
3. Construction of a path diagram representing causal relationships.
4. Partitioning of correlation coefficients into direct and indirect effects using simultaneous equations.
5. Calculation of residual effects.

The general formula used is:

$$r_{ij} = P_{iy} + \sum r_{ij} P_{iy}$$

The components of this formula are defined as:

- r_{iy} = The total correlation coefficient between an independent trait i and crop yield y
- P_{iy} = The standardized direct effect of trait i on crop yield.
- r_{ij} = The correlation coefficient between trait i and another independent trait j .
- P_{jy} = The standardized direct effect of trait j on crop yield.
- $\sum r_{ij} P_{jy}$ = the sum of all indirect effects of trait i on yield through other traits j .

Singh and Chaudhary (1985) ^[46] highlighted that path analysis requires accurate estimation of correlation coefficients and careful interpretation of biological relationships among traits.

Advantages of path analysis over correlation analysis

Path coefficient analysis offers several advantages over simple correlation analysis. It provides detailed information about the nature of relationships among traits by separating direct and indirect effects. This helps in identifying true cause-effect relationships rather than simple associations. According to Wright (1921) ^[49], correlation alone cannot distinguish between direct and indirect influences, whereas path analysis overcomes this limitation. In vegetable breeding, this allows identification of traits that directly influence yield. Dewey and Lu (1959) ^[10] emphasized that path analysis helps breeders prioritize traits with maximum direct effects for selection. It also assists in eliminating traits that show high correlation with yield but exert only indirect effects. Furthermore, path analysis improves the efficiency of selection by identifying key yield determinants such as fruit number, fruit weight, and branching pattern in vegetable crops.

Limitations of path analysis

Despite its advantages, path coefficient analysis has certain limitations. It assumes a predefined causal relationship among variables, which may not always be biologically accurate. Incorrect selection of variables can lead to misleading conclusions. According to Li (1975) ^[30], multicollinearity among independent variables can affect the reliability of path coefficients. Additionally, path analysis does not account for gene interactions or epistatic effects. In vegetable crops, environmental variability can also influence path coefficients, making results less stable across locations and seasons. Therefore, results should always be interpreted cautiously and validated across multiple environments.

Importance in crop improvement programmes

Path coefficient analysis plays a crucial role in crop improvement programmes by identifying key yield-contributing traits. It helps

Breeders select traits with maximum direct influence on yield, thereby increasing selection efficiency. In vegetable crops, numerous studies have demonstrated the importance of path analysis. Meena and Bahadur (2015) ^[32] identified fruit weight and fruit number as major direct contributors to yield in tomato. Similarly, Rani *et al.* (2015) ^[39] reported strong direct effects of vine length and fruit number in cucurbits. In brinjal, chilli, cucumber, and gourds, path analysis has consistently identified fruit number per plant and average fruit weight as the most important yield determinants (Kumar *et al.*, 2018; Rajkumari *et al.*, 2021) ^[26, 27]. These findings indicate that selection for these traits can effectively improve yield potential. Path analysis is also useful in identifying secondary traits that influence yield indirectly, such as plant height, leaf area, and days to flowering. By integrating both direct and indirect effects, breeders can design more efficient selection strategies. Overall, path coefficient analysis is an indispensable tool in vegetable breeding programmes for improving yield, understanding trait relationships, and accelerating genetic gain.

Correlation and path analysis in solanaceous vegetable crops

Correlation and path coefficient analyses are widely used in solanaceous vegetable crops such as tomato (*Solanum lycopersicum* L.), brinjal (*Solanum melongena* L.), chilli, and capsicum (*Capsicum* spp.) to understand the interrelationships among yield and its component traits. Since yield is a complex quantitative trait governed by multiple genes and strongly influenced by environmental conditions, direct selection is often less effective. Therefore, indirect selection through associated traits identified using correlation and path analysis is considered more efficient for improving yield in these crops (Singh & Chaudhary, 1985; Acquah, 2012) ^[1, 46]. In tomato, correlation studies have consistently shown that fruit yield per plant is positively associated with number of fruits per plant, average fruit weight, plant height, and number of primary branches. Meena and Bahadur (2015) ^[32] reported strong positive correlations between fruit yield and fruit number per plant, indicating that fruit number is a major yield-determining trait. Kumar *et al.* (2017) also observed significant genotypic correlations between fruit yield and fruit weight, suggesting strong genetic control. Path analysis further revealed that fruit number per plant and fruit weight exert the highest direct effects on yield, while plant height and branching contribute indirectly through their influence on reproductive traits (Dewey & Lu, 1959) ^[10]. Similar trends have been reported in brinjal, where fruit yield is positively associated with fruit number, fruit weight, plant height, and branching pattern. Kumar *et al.* (2018) ^[26, 27] found that fruit weight and fruit number are the most important direct contributors to yield, while vegetative traits act mainly through indirect effects. In chilli and capsicum, correlation studies also indicate strong positive associations of yield with fruit number per plant, fruit weight, fruit length, plant height, and branching habit. Yattung *et al.* (2014) ^[51] reported that fruit number is the most important yield-contributing trait in chilli. Sharma *et al.* (2020) ^[43] observed similar results in capsicum, where vegetative vigor traits supported yield indirectly through reproductive development. Path coefficient analysis confirmed that fruit number and fruit weight have the highest direct effects on yield, while plant height and branching influence yield indirectly (Fehr, 1987) ^[19]. Overall, studies across solanaceous crops consistently identify fruit number per plant and fruit weight as the most important selection criteria, making them key targets for yield improvement programmes in vegetable breeding.

Correlation and path analysis in cucurbitaceous vegetable crops

Cucurbitaceous vegetables such as cucumber, bitter gourd, bottle gourd, ridge gourd, sponge gourd, snake gourd, pumpkin, and ash gourd are important vine crops showing wide variability in growth habit, sex expression, and yield components. Because yield in these crops is highly complex and influenced by multiple vegetative and reproductive traits, correlation and path coefficient analyses are widely used to identify key yield-contributing characters and to separate their direct and indirect effects (Singh & Chaudhary, 1985; Dewey & Lu, 1959) ^[10].

In cucumber (*Cucumis sativus* L.), fruit yield is positively associated with fruit number per vine, fruit weight, vine length, and branching habit. Rani *et al.* (2015) ^[39] reported that fruit number is a major yield-determining trait. Path analysis confirmed that fruit number and fruit weight exert the highest direct effects, while vegetative traits like vine length contribute indirectly through improved photosynthetic efficiency. In bitter gourd (*Momordica charantia* L.), yield is strongly influenced by sex expression and fruit-bearing traits. Fruit number, female flower production, and fruit set percentage show strong positive correlations with yield. Kumar *et al.* (2019) ^[28] observed that fruit number has a major direct effect on yield, while vine length and branching act indirectly through reproductive efficiency.

Bottle gourd (*Lagenaria siceraria*) shows positive associations of yield with vine length, fruit length, fruit diameter, and fruit weight. Rajkumari *et al.* (2021) reported that fruit weight and fruit number are the most important contributors to yield. Path analysis indicated that fruit weight has the highest direct effect, while vegetative traits influence yield indirectly. In ridge gourd and sponge gourd, yield is mainly governed by fruit number, fruit weight, vine length, and branching habit. Priyanka *et al.* (2020) ^[37] and Meena and Bahadur (2015) ^[32] reported that fruit number per plant has a strong positive correlation and direct effect on yield in these crops. Snake gourd and pumpkin also show similar trends where fruit number and fruit weight are the major yield determinants, while vine length and leaf area contribute indirectly by enhancing assimilate supply (Kalloo, 1993) ^[24]. In ash gourd, fruit weight and fruit size are the most important traits influencing yield, with vegetative traits playing a supporting role. Overall, studies across cucurbitaceous crops consistently indicate that fruit number per plant and fruit weight are the most important direct contributors to yield, while vegetative traits such as vine length and branching mainly exert indirect effects.

Correlation and path analysis in leguminous vegetable crops

Leguminous vegetable crops such as cowpea, French bean, dolichos bean, and garden pea are important sources of dietary protein and also play a significant role in improving soil fertility through biological nitrogen fixation. Yield improvement in these crops is complex because pod yield is influenced by multiple interacting vegetative and reproductive traits. Therefore, correlation and path coefficient analyses are widely used in legume breeding programmes to identify key yield-contributing characters and to understand the nature of relationships among traits by separating direct and indirect effects (Singh & Chaudhary, 1985; Dewey & Lu, 1959) ^[10, 46].

In cowpea (*Vigna unguiculata*), correlation studies have shown that pod yield is positively associated with traits such as number of pods per plant, pod length, number of seeds per pod, plant height, and number of branches. Among these, number of pods per plant is consistently identified as a major yield-determining trait. Path analysis has further revealed that number of pods per plant and pod weight exert the highest direct effects on yield,

while plant height and branching contribute indirectly by enhancing photosynthetic efficiency and pod formation (Kumar *et al.*, 2018) ^[26, 27]. In French bean (*Phaseolus vulgaris*), pod yield shows strong positive associations with number of pods per plant, pod length, pod weight, and number of clusters per plant. Studies by Meena and Bahadur (2015) ^[32] and Sharma *et al.* (2020) ^[43] reported that pod number per plant is the most important yield component. Path coefficient analysis confirmed that pod number and pod weight have the strongest direct effects on yield, while vegetative traits influence yield indirectly. In some cases, early flowering traits show negative association with yield, indicating a trade-off between earliness and productivity. In dolichos bean (*Lablab purpureus*), yield is mainly governed by pod number, pod length, pod weight, and branching pattern. Correlation studies indicate that vigorous vegetative growth supports higher reproductive output. Path analysis shows that pod number and pod weight are the main direct contributors to yield, while vine growth contributes indirectly through assimilate production (Kalloo, 1993) ^[24]. In garden pea (*Pisum sativum*), pod yield is positively associated with number of pods per plant, pod length, number of seeds per pod, and plant height. Path analysis has shown that pod number per plant exerts the highest direct effect on yield, followed by seed number per pod, while vegetative traits influence yield indirectly through reproductive development (Falconer & Mackay, 1996) ^[18]. Overall, studies in leguminous vegetables consistently indicate that pod number per plant and pod weight are the most important direct determinants of yield, while vegetative traits mainly contribute indirectly by supporting reproductive efficiency and assimilate supply.

Correlation and path analysis in cole crops

Cole crops, belonging to the family Brassicaceae, include important vegetables such as cabbage (*Brassica oleracea* var. capitata), cauliflower (*Brassica oleracea* var. botrytis), and broccoli (*Brassica oleracea* var. italica). These crops are valued for their compact edible structures such as heads and curds, and yield in these crops is a complex quantitative trait influenced by multiple vegetative and reproductive components. Therefore, correlation and path coefficient analyses are widely used to identify key yield-contributing traits and to understand their direct and indirect effects, which helps improve selection efficiency in breeding programmes (Singh & Chaudhary, 1985; Dewey & Lu, 1959) ^[46]. In cabbage, correlation studies have shown that head yield is positively associated with plant height, number of outer leaves, head diameter, and leaf area. Singh *et al.* (2018) ^[47] reported strong positive correlations between head weight and head diameter, indicating that head size is a major determinant of yield. Sharma *et al.* (2020) ^[43] also observed that vegetative traits such as leaf number and plant vigor contribute indirectly to yield by enhancing photosynthetic efficiency. Path analysis further revealed that head diameter and head weight exert the highest direct effects on yield, while vegetative traits influence yield indirectly through assimilate production. In cauliflower, curd yield shows positive associations with curd diameter, curd weight, plant height, and number of leaves. Kumar *et al.* (2018) reported that curd weight is strongly correlated with total yield, making it a key yield component. Path coefficient analysis confirmed that curd weight has the highest direct effect on yield, while vegetative traits such as leaf number contribute indirectly by supporting assimilate supply. In some studies, early curd initiation has shown negative association with yield, suggesting a trade-off between earliness and productivity. In broccoli, head yield is positively associated with head weight, head diameter, plant height, and leaf number. Priyanka *et al.* (2020) ^[37] reported

that head weight is the most important trait influencing yield. Path analysis further confirmed that head weight and head diameter exert strong direct effects on yield, while vegetative traits contribute indirectly by improving photosynthetic efficiency. Overall, studies in cole crops consistently show that head or curd weight and diameter are the most important direct determinants of yield, while vegetative traits mainly act indirectly.

Correlation and path analysis in root and bulb vegetables

Root and bulb vegetables such as carrot, radish, beetroot, onion, and garlic form an important group of crops in which economic yield is stored in modified underground structures. Yield formation in these crops is a complex process influenced by vegetative growth, assimilate partitioning, and environmental conditions. Hence, correlation and path coefficient analyses are widely used to identify key yield-contributing traits and to understand their direct and indirect effects for effective selection in breeding programmes (Singh & Chaudhary, 1985; Dewey & Lu, 1959) ^[10]. In carrot (*Daucus carota*), root yield is positively associated with traits such as root length, root diameter, plant height, and number of leaves per plant. Singh *et al.* (2018) ^[47] reported that root diameter and root length are major determinants of yield. Path analysis further revealed that root diameter exerts the highest direct effect on yield, while plant height and leaf number influence yield indirectly through improved photosynthetic efficiency and assimilate translocation (Meena & Bahadur, 2015) ^[32]. In radish (*Raphanus sativus*), correlation studies have shown strong positive associations between root yield and traits such as root weight, root diameter, leaf number, and plant vigor. Kumar *et al.* (2019) ^[28] observed that root diameter is strongly correlated with yield. Path analysis confirmed that root weight and root diameter have strong direct effects, while vegetative traits influence yield indirectly through assimilate supply. In some studies, early maturity has shown negative association with yield (Sharma *et al.*, 2020) ^[43]. In beetroot (*Beta vulgaris*), yield is influenced by root size, root weight, leaf area, and plant height. Correlation studies have shown that root yield is positively associated with root diameter and leaf number. Rani *et al.* (2015) ^[39] reported that leaf area contributes significantly to yield. Path analysis indicated that root weight has the highest direct effect, followed by root diameter, while vegetative traits contribute indirectly.

In onion (*Allium cepa*), bulb yield is determined by bulb diameter, number of leaves, plant height, and neck thickness. Correlation analysis has shown strong positive associations between bulb yield and bulb size. Path analysis confirmed that bulb weight and bulb diameter exert strong direct effects on yield, while vegetative traits contribute indirectly through assimilate production. In garlic (*Allium sativum*), yield is mainly influenced by bulb weight, clove number, and bulb diameter. Correlation studies show positive associations between bulb yield and these traits. Path analysis revealed that bulb weight has the highest direct effect on yield, while vegetative traits contribute indirectly. Overall, root and bulb crops consistently show that storage organ weight and diameter are the most important direct yield determinants, while vegetative traits act mainly through indirect effects.

Comparative analysis of yield-contributing traits across vegetable crops

Across vegetable crop groups, correlation and path coefficient analyses consistently demonstrate that yield is a complex quantitative trait governed by multiple interacting morphological, reproductive, and physiological components. Although the

magnitude of trait associations varies among species, several yield-contributing characters show relatively stable relationships with yield across solanaceous, cucurbitaceous, leguminous, cole, and root-bulb vegetables. Comparative evaluation of these traits helps in identifying both universal selection criteria and crop-specific breeding strategies for improving productivity (Singh & Chaudhary, 1985; Acquaah, 2012) ^[1, 46].

Among vegetative traits, plant height, number of branches per plant, leaf area, and general plant vigor frequently exhibit positive associations with yield across most vegetable crops. These traits enhance photosynthetic efficiency and assimilate production, thereby supporting better reproductive development and yield formation. Falconer and Mackay (1996) ^[18] emphasized that vegetative vigor plays an important supporting role in quantitative trait expression, particularly in complex traits such as yield. Among reproductive traits, the number of fruits or pods per plant consistently shows a strong positive correlation with yield across crops such as tomato, brinjal, cucumber, cowpea, and pea. Meena and Bahadur (2015) ^[32] and Kumar *et al.* (2018) ^[26, 27] reported that fruit or pod number is one of the most reliable yield indicators in vegetable crops. Similarly, average fruit or pod weight is another highly stable trait positively associated with yield across cucurbits, solanaceous vegetables, and legumes. In root and bulb crops, root or bulb diameter and weight are the most important yield-determining traits (Sharma *et al.*, 2020) ^[43].

Path coefficient analysis across vegetable crops has identified fruit/pod number per plant and fruit/pod weight as the most important direct contributors to yield. According to Dewey and Lu (1959) ^[10], these traits represent primary yield components and should be prioritized in selection programmes. In solanaceous crops, fruit number and fruit weight exert the highest direct effects on yield, while in cucurbits, fruit number per vine and fruit weight are most important. In legumes, pod number per plant and pod weight are key determinants, whereas in cole crops, head or curd weight is the primary contributor. In root and bulb crops, root or bulb weight and diameter show the strongest direct effects on yield (Fehr, 1987; Kalloo, 1993) ^[19, 24]. Several traits contribute to yield indirectly through their influence on primary yield components. Plant height, branching pattern, leaf area, vine length, and vegetative growth vigor often show indirect effects by improving photosynthetic capacity and assimilate supply. For example, vine length in cucurbits and plant height in tomato and legumes enhance yield indirectly by supporting reproductive development. Similarly, leaf area in root and bulb crops contributes indirectly to root or bulb enlargement (Gardner *et al.*, 1985) ^[22]. Phenological traits such as days to flowering and maturity also influence yield indirectly, either positively or negatively depending on crop type and environmental conditions. Early flowering may enhance yield in short-duration crops, while delayed flowering may favor biomass accumulation in others. Overall, despite crop-specific variations, a common physiological principle emerges across vegetable crops: yield is primarily determined by the efficiency of assimilate partitioning into economically important sink organs such as fruits, pods, heads, roots, or bulbs.

Factors influencing correlation and path relationships

Correlation and path coefficient relationships among yield and its component traits in vegetable crops are not fixed; rather, they are dynamic and are strongly influenced by genetic constitution, environmental conditions, genotype × environment interactions, crop management practices, and biotic and abiotic stresses. These factors often modify both the magnitude and direction of trait

associations, thereby affecting selection efficiency in breeding programmes. Hence, careful interpretation of correlation and path analysis results is essential in vegetable crop improvement (Falconer & Mackay, 1996; Singh & Chaudhary, 1985) ^[18, 46].

Genetic factors form the primary basis of trait associations. The genetic architecture of a plant determines the expression of morphological, physiological, and reproductive traits contributing to yield. Genetic correlations arise mainly due to pleiotropy or linkage between genes controlling different traits. Singh and Chaudhary (1985) ^[46] reported that genotypic correlations are generally more stable than phenotypic correlations because they are less affected by environmental variation. Traits such as fruit number, fruit weight, and plant height often show strong genetic associations across vegetable crops, making them reliable selection criteria. Environmental factors also play a major role in modifying trait relationships. Temperature, rainfall, soil fertility, and photoperiod influence both vegetative and reproductive development. Gardner *et al.* (1985) ^[22] noted that environmental variation often reduces the magnitude of phenotypic correlations. For example, high temperature may reduce fruit set in cucurbits, weakening yield associations, while soil fertility can enhance vegetative growth and alter indirect trait relationships. Thus, environmental variation introduces instability in correlation estimates.

Genotype $\times$ environment (G $\times$ E) interaction further complicates trait relationships by altering the performance of genotypes across different environments. Allard (1999) ^[2] emphasized that G $\times$ E interactions can change both the magnitude and direction of correlations among traits. Yield components such as fruit number and fruit weight often show unstable associations across locations and seasons, making multi-environment testing essential for reliable selection. Agronomic management practices also influence correlation and path relationships. Irrigation, fertilization, spacing, and pruning can modify plant growth and yield components. Fehr (1987) ^[19] reported that nitrogen application enhances vegetative growth, indirectly affecting yield associations. Similarly, proper canopy management improves light interception and alters relationships between vegetative and reproductive traits.

Biotic and abiotic stresses significantly affect trait relationships by altering physiological processes. Stress conditions such as drought, salinity, heat, pests, and diseases often reduce yield and modify correlations among traits. Blum (2011) ^[3] reported that stress conditions can weaken existing trait associations and increase the importance of adaptive traits such as root depth and water-use efficiency. Overall, these factors highlight that correlation and path relationships are environment-dependent and must be interpreted carefully for effective vegetable crop improvement.

Applications in vegetable breeding programmes and integration with modern breeding tools

Correlation and path coefficient analyses are widely used in vegetable breeding programmes to improve selection efficiency by identifying key yield-contributing traits. These statistical tools help breeders understand the nature of associations among traits and determine whether their effects on yield are direct or indirect. This information is essential for selecting superior parents,

developing hybrids, constructing selection indices, and improving overall productivity in vegetable crops. Across most vegetable species, traits such as fruit or pod number, fruit weight, plant height, and branching pattern consistently emerge as major selection criteria due to their stable association with yield (Singh & Chaudhary, 1985; Acquah, 2012) ^[1, 46].

In parent selection, genotypes exhibiting strong and positive correlations between yield and its component traits are preferred as breeding material. Allard (1999) ^[2] emphasized that selecting parents based on genetically stable associations increases the probability of obtaining superior segregants in subsequent generations. Path coefficient analysis further strengthens this process by identifying traits with strong direct effects on yield, ensuring that breeders focus on biologically efficient and genetically reliable characters.

In hybrid development, correlation and path analysis assist in identifying complementary parental combinations that can produce superior hybrids. Fehr (1987) ^[19] noted that hybrids derived from parents with desirable trait combinations often exhibit enhanced heterosis for yield. Traits such as fruit number and fruit weight are particularly important in hybrid breeding of crops like tomato, brinjal, and cucurbits. Similarly, selection indices are developed using correlation and path estimates to assign appropriate weightage to multiple traits for simultaneous selection, thereby improving overall selection efficiency.

Early generation selection is often challenging due to strong environmental influence on yield expression. Therefore, breeders rely on correlated traits with higher heritability and stable genetic associations for indirect selection. Falconer and Mackay (1996) ^[18] highlighted that indirect selection is more effective when direct selection for yield is unreliable. In vegetable crops, traits such as vegetative vigor and fruit number are commonly used in early generation selection to ensure genetic gain. Modern breeding approaches have further enhanced the application of these statistical tools. Molecular marker-assisted selection integrates phenotypic correlations with genetic markers to improve selection accuracy (Collard & Mackill, 2008) ^[8]. Quantitative trait loci (QTL) mapping helps identify genomic regions controlling yield and its components, validating traits identified through path analysis (Lander & Botstein, 1989) ^[29]. Genome-wide association studies (GWAS) refine these associations by analyzing diverse germplasm populations (Yu *et al.*, 2006).

Genomic selection (GS) uses genome-wide marker data to predict breeding values for complex traits such as yield, and when combined with correlation and path analysis, it enhances prediction accuracy (Meuwissen *et al.*, 2001) ^[34]. High-throughput phenotyping technologies enable rapid and precise measurement of traits like canopy structure and leaf area, improving the reliability of correlation estimates (Fahlgren *et al.*, 2015) ^[17]. Artificial intelligence and machine learning approaches further enhance predictive breeding by analyzing complex, multi-trait datasets and modeling non-linear relationships among yield components (Liakos *et al.*, 2018) ^[31]. Overall, the integration of correlation and path analysis with modern genomic and computational tools has significantly improved the efficiency of vegetable breeding programmes by enabling precise identification of key yield-determining traits and accelerating genetic gain.

Research gaps and future perspectives in vegetable crop correlation and path analysis

Although correlation and path coefficient analyses have been widely applied in vegetable breeding, several limitations still restrict their full effectiveness in modern crop improvement. One major gap is the lack of multi-environment validation of trait associations. Most studies are conducted in a single location or season, which reduces the reliability of results because trait relationships are strongly influenced by genotype $\times$ environment interactions. As noted by Allard (1999) [2] and Falconer and Mackay (1996) [18], yield component relationships often vary across environments, making multi-location and multi-season testing essential for identifying stable selection criteria. Another limitation is the absence of large-scale meta-analyses across vegetable crops. Individual studies often report inconsistent results regarding key yield-contributing traits due to differences in genotypes, environments, and methodologies. Integrating

results across studies would help identify universally important traits such as fruit or pod number and fruit weight (Singh & Chaudhary, 1985) [46].

There is also limited integration of classical correlation and path analysis with modern omics technologies. While these statistical tools explain phenotypic relationships, they do not reveal underlying genetic mechanisms. Integration with genomics and molecular tools can improve understanding of complex traits like yield (Acquaah, 2012) [1]. Climate change further complicates trait stability, as abiotic stresses can alter correlations among yield components (Blum, 2011) [3]. Therefore, identifying stress-resilient traits is crucial for developing climate-smart vegetable varieties. Future research should integrate correlation and path analysis with high-throughput phenotyping, precision agriculture, and artificial intelligence tools to improve prediction accuracy and selection efficiency in vegetable breeding (Fahlgren *et al.*, 2015; Liakos *et al.*, 2018) [17, 31].

Table 1: Major correlation studies in vegetable crops and associated yield traits

S. No.	Crop	Traits studied	Major findings on correlation with yield	References
1	Tomato	Plant height, number of fruits/plant, fruit weight	Fruit yield showed strong positive correlation with number of fruits per plant and average fruit weight	Singh <i>et al.</i> (2019) [45]
2	Brinjal	Fruit length, fruit girth, branches/plant	Yield positively correlated with number of branches and fruit girth	Kumar & Prasad (2018) [26, 27]
3	Chilli	Fruit number, fruit length, capsaicin content	Fruit number had highest positive association with yield	Rao <i>et al.</i> (2020) [40]
4	Cucumber	Vine length, fruit number, fruit diameter	Yield significantly correlated with fruit number and vine length	Sharma <i>et al.</i> (2017) [44]
5	Bottle gourd	Fruit length, fruit weight, nodes/vine	Fruit weight showed strong positive correlation with yield	Patel <i>et al.</i> (2016) [35]
6	Bitter gourd	Fruit number, fruit length, leaf area	Yield highly dependent on fruit number per plant	Yadav <i>et al.</i> (2021)
7	Ridge gourd	Vine length, fruit weight, internode length	Fruit weight had major positive effect on yield	Meena <i>et al.</i> (2018) [33]
8	Okra	Plant height, pods/plant, pod weight	Pod number per plant showed strongest correlation with yield	Devi <i>et al.</i> (2020) [9]

Table 2: Important direct and indirect effects of yield-contributing traits identified through path coefficient analysis

S. No.	Crop	Traits analyzed	Major direct effects on yield	Important indirect effects via other traits	References
1	Tomato	Plant height, fruit number, fruit weight	Fruit weight showed highest direct effect	Fruit number indirectly influenced yield via fruit weight	Singh <i>et al.</i> (2019) [45]
2	Brinjal	Branches, fruit size, fruit number	Number of fruits/plant had strong direct effect	Branching influenced yield indirectly via fruit number	Kumar & Prasad (2018) [26, 27]
3	Chilli	Fruit length, fruit number	Fruit number showed maximum direct effect	Fruit length acted indirectly through fruit weight	Rao <i>et al.</i> (2020) [40]
4	Cucumber	Vine length, fruit diameter	Fruit diameter showed strong direct effect	Vine length influenced yield indirectly via fruit number	Sharma <i>et al.</i> (2017) [44]
5	Bottle gourd	Fruit weight, nodes/vine	Fruit weight had highest direct contribution	Nodes/vine affected yield indirectly via fruit number	Patel <i>et al.</i> (2016) [35]
6	Bitter gourd	Fruit number, leaf area	Fruit number showed highest direct effect	Leaf area influenced yield indirectly via photosynthesis	Yadav <i>et al.</i> (2021)
7	Ridge gourd	Vine length, fruit weight	Fruit weight had major direct effect	Vine length showed indirect effects via fruit number	Meena <i>et al.</i> (2018) [33]
8	Okra	Pods/plant, pod weight	Pods per plant had strongest direct effect	Plant height influenced yield indirectly via pods/plant	Devi <i>et al.</i> (2020) [9]

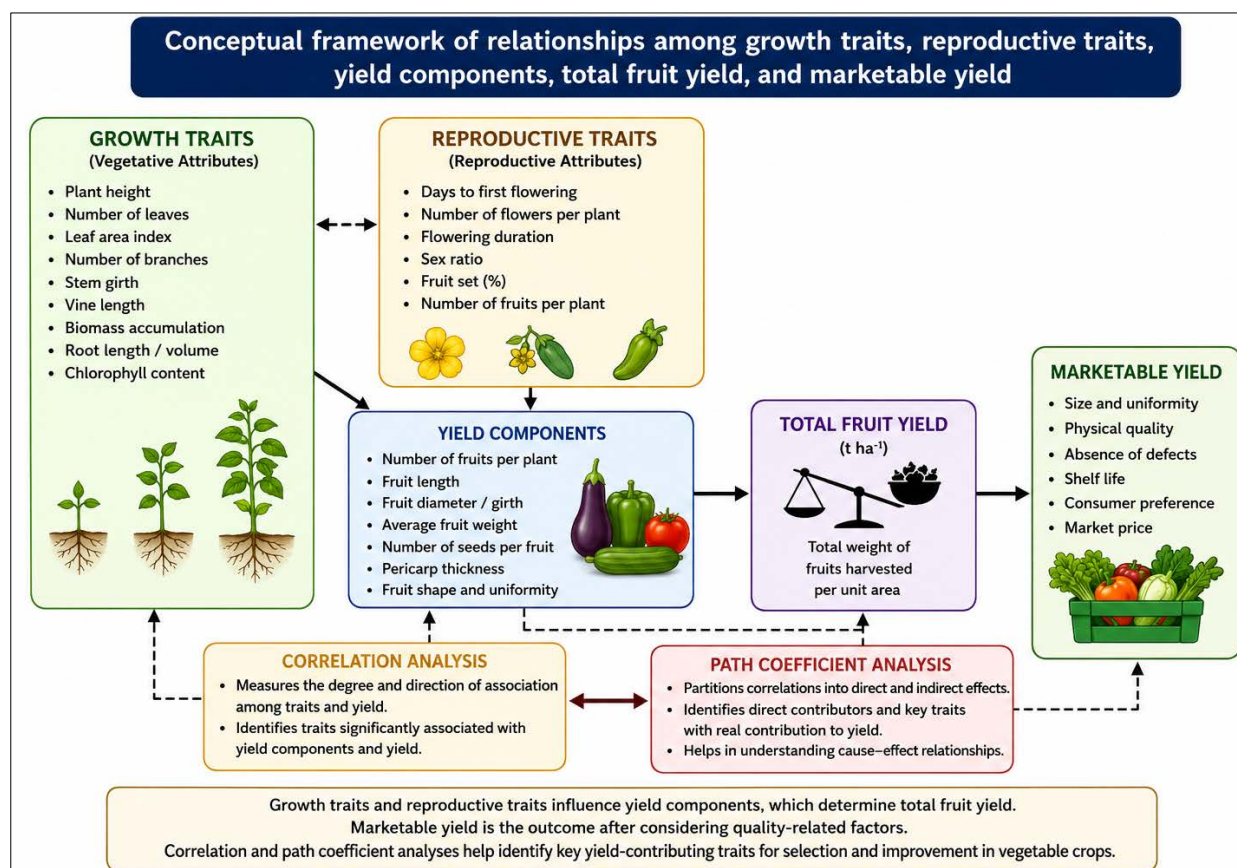


Fig 1: Conceptual framework showing the relationship between growth traits, reproductive traits, yield components and marketable yield

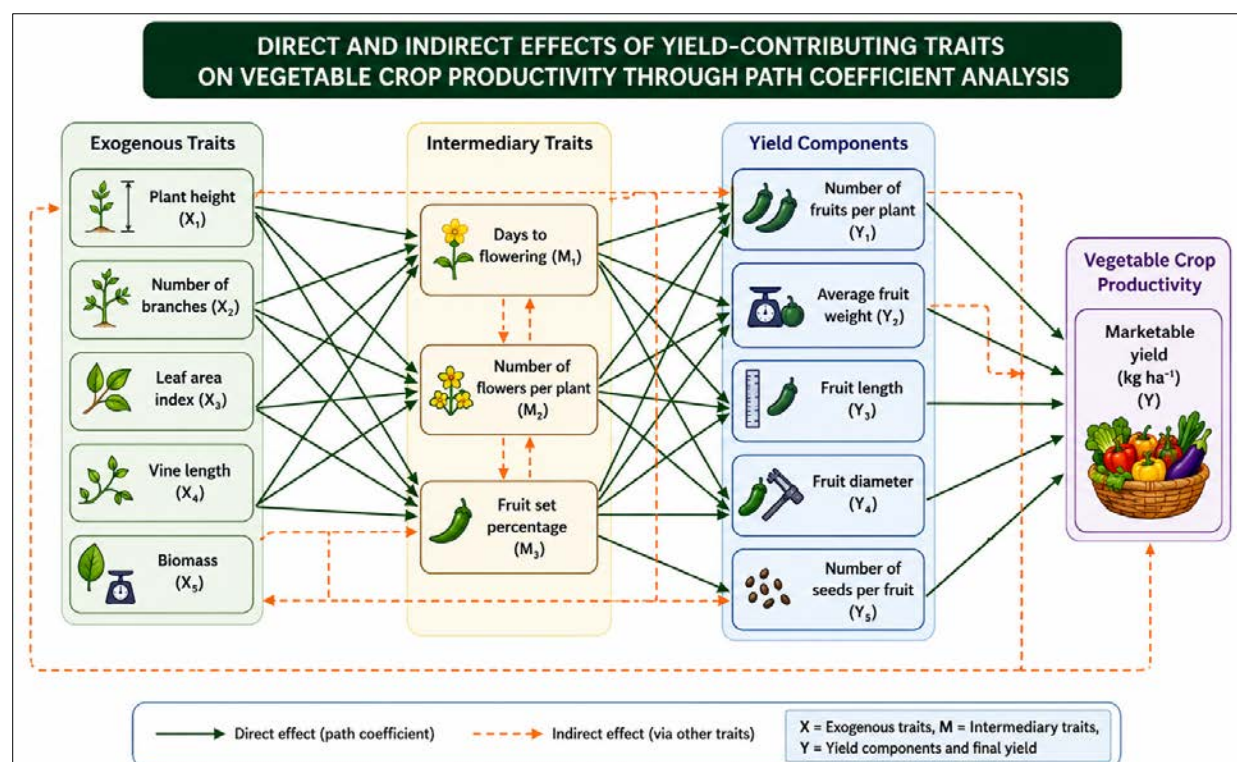


Fig 2: Schematic representation of direct and indirect effects of yield-contributing traits on vegetable crop productivity through path coefficient analysis

Conclusion

Vegetable crop productivity is a complex quantitative trait governed by multiple interacting morphological, physiological, and reproductive components. Correlation and path coefficient analyses have proven to be powerful statistical tools for dissecting

these relationships and identifying key yield-contributing traits. Across diverse vegetable crops, traits such as fruit or pod number, fruit or pod weight, and plant or canopy architecture consistently emerge as major determinants of yield, while vegetative traits generally exert indirect effects through improved assimilate

production and partitioning. The reviewed studies clearly indicate that understanding both the magnitude and direction of trait associations is essential for effective selection in breeding programmes. Path analysis further refines this understanding by separating direct and indirect effects, thereby enabling breeders to prioritize traits with true causal influence on yield. However, the stability of these relationships is often influenced by genetic background, environmental conditions, genotype $\times$ environment interactions, and crop management practices. The integration of classical biometrical approaches with modern breeding tools such as molecular markers, QTL mapping, genomic selection, high-throughput phenotyping, and artificial intelligence is transforming vegetable breeding into a more precise and predictive discipline. This combined approach enhances selection efficiency, accelerates genetic gain, and improves the development of high-yielding and resilient vegetable varieties. Overall, correlation and path coefficient analysis remain indispensable in vegetable crop improvement, and their continued integration with advanced technologies will play a crucial role in ensuring sustainable yield enhancement under changing agricultural environment

Statements and declarations

Funding

The authors received no financial support for the research, authorship and/ or publication of this article.

Acknowledgement

We are grateful to Tamil Nadu Agricultural University for their ongoing support in the form of regular workshop that helps us with writing review articles.

Author contribution information

R.G and J.M are equally contributed. Idea: R.G, B.G, M.R.G, S.T and J.M; Analysis and writing: S.P, G.G, R.J.P and V.V.V; Final draft and revision: J.M, R.G, B.G, S.T and V.V.V

Consent for publication

All authors consent the publication of the current article.

Competing interests

The authors declare no competing interests.

References

- Acquaah G. Principles of plant genetics and breeding. 2nd ed. Oxford: Wiley-Blackwell; 2012.
- Allard RW. Principles of plant breeding. 2nd ed. New York: John Wiley & Sons; 1999.
- Blum A. Plant breeding for water-limited environments. Berlin: Springer; 2011. DOI:10.1007/978-1-4419-7491-4.
- Burton GW. Quantitative inheritance in grasses. In: Proceedings of the 6th International Grassland Congress. 1952; 1:277-283.
- Burton GW. Quantitative inheritance in grasses. In: Proceedings of the 6th International Grassland Congress. 1952; 1:277-283.
- Burton GW, Devane EH. Estimating heritability in tall fescue (*Festuca arundinacea*) from replicated clonal material. *Agronomy Journal*. 1953; 45:478-481. DOI:10.2134/agronj1953.00021962004500100005x.
- Ceccarelli S. Efficiency of plant breeding. USA: Crop Science Society of America; 2015.
- Collard BCY, Mackill DJ. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philosophical Transactions of the Royal Society B*. 2008; 363:557-572. DOI:10.1098/rstb.2007.2170.
- Devi R, Singh AK, Sharma P. Correlation and path coefficient analysis in okra (*Abelmoschus esculentus* L.). *Journal of Horticultural Sciences*. 2020;15(2):98-104.
- Dewey DR, Lu KH. A correlation and path coefficient analysis of components of crested wheatgrass seed production. *Agronomy Journal*. 1959; 51:515-518. DOI:10.2134/agronj1959.00021962005100090002x.
- Dias JS. Nutritional quality and health benefits of vegetables. *Food and Nutrition Sciences*. 2019; 10:1-12. DOI:10.4236/fns.2019.101001.
- Dias JS. Nutritional quality and health benefits of vegetables. *Food and Nutrition Sciences*. 2019; 10:1-12. DOI:10.4236/fns.2019.101001.
- Dias JS, Ryder EJ. World vegetable industry: production, breeding, trends and outlook. *Horticultural Reviews*. 2011; 38:299-356. DOI: 10.1002/9781118060795.ch5.
- Ebert AW. Vegetable crops for healthy diets in developing countries. *Horticulturae*. 2020;6(3):66. DOI:10.3390/horticulturae6030066.
- Ebert AW, Engels JMM. Plant genetic resources for food and nutrition security. *Plants*. 2020;9(5):536. DOI:10.3390/plants9050536.
- Ebert AW, Engels JMM. Plant genetic resources for food and nutrition security. *Plants*. 2020;9(5):536. DOI:10.3390/plants9050536.
- Fahlgren N, Gehan MA, Baxter I. Lights, camera, action: high-throughput plant phenotyping is ready for a close-up. *Current Opinion in Plant Biology*. 2015; 24:93-99. DOI: 10.1016/j.pbi.2015.02.006.
- Falconer DS, Mackay TFC. Introduction to quantitative genetics. 4th ed. Essex: Longman; 1996.
- Fehr WR. Principles of cultivar development: theory and technique. New York: Macmillan; 1987.
- Fisher RA. The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*. 1918; 52:399-433. DOI:10.1017/S008045680001216X.
- Galton F. Co-relations and their measurement. *Proceedings of the Royal Society of London*. 1888; 45:135-145.
- Gardner FP, Pearce RB, Mitchell RL. Physiology of crop plants. Iowa: Iowa State University Press; 1985.
- Johnson HW, Robinson HF, Comstock RE. Estimates of genetic and environmental variability in soybeans. *Agronomy Journal*. 1955; 47:314-318. DOI:10.2134/agronj1955.00021962004700070009x.
- Kaloo G. Vegetable breeding. Florida: CRC Press; 1993.
- Keatinge JDH, Waliyar F, Jamnadas RH, *et al.* Relearning old lessons for the future of food security in Asia. *Food Security*. 2014; 6:523-540. DOI:10.1007/s12571-014-0356-7.
- Kumar A, Prasad R. Genetic variability, correlation and path analysis in brinjal (*Solanum melongena* L.). *International Journal of Vegetable Science*. 2018;24(3):210-218. DOI:10.1080/19315260.2017.1396152.
- Kumar N, *et al.* Correlation and path coefficient analysis in brinjal. *Journal of Horticultural Sciences*. 2018;13(2):120-126.
- Kumar R, *et al.* Correlation and path analysis in cucurbits. *Vegetable Science*. 2019;46(1):1-8.
- Lander ES, Botstein D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*. 1989; 121:185-199.

- DOI:10.1093/genetics/121.1.185.
30. Li CC. Path analysis: a primer. California: Boxwood Press; 1975.
 31. Liakos KG, *et al.* Machine learning in agriculture: a review. *Sensors*. 2018;18(8):2674. DOI:10.3390/s18082674.
 32. Meena OP, Bahadur V. Correlation and path coefficient studies in tomato. *Vegetable Science*. 2015;42(2):145-150.
 33. Meena OP, Singh D, Bahadur V. Correlation and path analysis in ridge gourd (*Luffa acutangula* L.). *Vegetable Science*. 2018;45(1):45-50.
 34. Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 2001; 157:1819-1829. DOI:10.1093/genetics/157.4.1819.
 35. Patel JR, Patel MM, Chaudhary MG. Correlation and path analysis in bottle gourd (*Lagenaria siceraria*). *Agricultural Science Digest*. 2016;36(2):120-124. DOI:10.18805/asd.v36i2.10643.
 36. Pearson K. Mathematical contributions to the theory of evolution—III. *Philosophical Transactions of the Royal Society A*. 1896; 187:253-318. DOI:10.1098/rsta.1896.0007.
 37. Priyanka, *et al.* Correlation studies in cucurbits. *International Journal of Vegetable Science*. 2020;26(3):210-218.
 38. Rajkumari, *et al.* Genetic association studies in cucurbit crops. *Journal of Crop Improvement*. 2021;35(2):180-195. DOI:10.1080/15427528.2020.1837145.
 39. Rani P, *et al.* Correlation studies in cucurbits. *Indian Journal of Horticulture*. 2015;72(3):345-350.
 40. Rao MS, Kumar S, Reddy P. Association studies in chilli (*Capsicum annum* L.) for yield and quality traits. *Scientia Horticulturae*. 2020; 265:109115. DOI: 10.1016/j.scienta.2020.109115.
 41. Robinson HF, Comstock RE, Harvey PH. Genotypic and phenotypic correlations in corn and their implications in selection. *Agronomy Journal*. 1951; 43:282-287. DOI:10.2134/agronj1951.00021962004300060007x.
 42. Schreinemachers P, Pashley D, Newton P. Vegetable production and nutrition. *Global Food Security*. 2018; 17:33-45. DOI: 10.1016/j.gfs.2018.03.003.
 43. Sharma A, *et al.* Path analysis in vegetable crops. *Scientia Horticulturae*. 2020; 265:109115. DOI: 10.1016/j.scienta.2020.109115.
 44. Sharma N, Singh D, Verma S. Correlation and path coefficient analysis in cucumber (*Cucumis sativus* L.). *Journal of Crop Improvement*. 2017;31(4):520-528. DOI:10.1080/15427528.2017.1338901.
 45. Singh AK, Kumar R, Sharma P. Correlation and path coefficient analysis in tomato (*Solanum lycopersicum* L.). *Vegetable Science*. 2019;46(2):135-140.
 46. Singh BD, Chaudhary PK. Biometrical methods in quantitative genetic analysis. New Delhi: Kalyani Publishers; 1985.
 47. Singh S, *et al.* Variability studies in vegetable crops. *Horticultural Journal*. 2018;31(2):90-98.
 48. Tanksley SD. The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. *Plant Cell*. 2004;16(Suppl 1):S189. DOI:10.1105/tpc.018119.
 49. Wright S. Correlation and causation. *Journal of Agricultural Research*. 1921; 20:557-585.
 50. Yadav SK, Singh V, Kumar A. Correlation and path analysis in bitter gourd (*Momordica charantia* L.). *International Journal of Vegetable Science*. 2021;27(3):210-219. DOI:10.1080/19315260.2020.1739185.
 51. Yatung T, *et al.* Correlation studies in chilli. *African Journal of Agricultural Research*. 2014;9(20):1563-1570. DOI:10.5897/AJAR2014.8707.