

## Article

# Integrating Multivariate Analysis and DNA Barcoding for Amaranth Germplasm Characterization and Promising Genotype Selection

Adnan Kanbar <sup>1,2,\*</sup> , Yaman Jabbour <sup>3</sup>  and Peter Nick <sup>4,\*</sup> 

<sup>1</sup> National Unit for Environmental Research and Services (NUERS), Research Sector, Kuwait University, P.O. Box 5969, Al-Shadadiya 13060, Kuwait

<sup>2</sup> Department of Field Crops, Faculty of Agriculture, University of Damascus, Damascus P.O. Box 30621, Syria

<sup>3</sup> Field Crop Department, Aleppo Research Center, General Commission for Scientific Agriculture Research (GCSAR), Damascus P.O. Box 12573, Syria; yaman.jab@gmail.com

<sup>4</sup> Molecular Cell Biology, Joseph Kölreuter Institute for Plant Sciences, Karlsruhe Institute of Technology, Fritz-Haber-Weg 4, 76131 Karlsruhe, Germany

\* Correspondence: kanbaradnan@gmail.com or adnan.kanbar@ku.edu.kw (A.K.); peter.nick@kit.edu (P.N.)

## Abstract

Amaranth (*Amaranthus* spp.) is a multifaceted genus of C4 plants with significant nutritional and agronomic potential, yet it remains underutilized in mainstream agriculture. Despite growing interest in Amaranth, most germplasm studies have used either phenotypic or molecular approaches alone, lacking integration. Multivariate methods have not been systematically applied to identify promising genotypes, and species-specific selection indices for grain Amaranth remain unexplored. To address these gaps, this study comprehensively characterized 84 Amaranth genotypes representing multiple species (*A. caudatus*, *A. cruentus*, *A. hypochondriacus*, *A. hybridus*, *A. spinosus*, *A. powellii*, *A. tricolor*, and 38 accessions of unknown taxonomic status) using field experiments in a randomized complete block design with three replications and DNA barcoding with chloroplast (*psbA-trnH*) and nuclear (ITS) markers. Analysis of variance revealed highly significant differences ( $p < 0.01$ ) among genotypes for all six agronomic traits evaluated, confirming substantial genetic variability with grain yield exhibiting the widest variation (CV = 28.55%), ranging from 0.25 to 125.56 g/plant. High broad-sense heritability estimates (0.79–0.99) coupled with high genetic advance, particularly for grain yield (117.54%), indicated that these traits would respond favorably to selection. Path analysis and stepwise regression identified early flowering, long inflorescences, and heavy seeds as the primary determinants of grain yield, collectively explaining 27% of yield variation. Mahalanobis  $D^2$  analysis identified nine multivariate outliers with distinct phenotypic profiles, among which G39 emerged as the most promising breeding candidate, combining exceptional yield (90.50 g/plant) with desirable architecture, long inflorescence, and large seeds. Principal component analysis further resolved trait complexes, identifying 11 PC1-selected promising genotypes as donors for plant architecture and three PC2-selected promising genotypes as donors for seed size characteristics. Molecular analysis revealed distinct genetic relationships. *A. caudatus* (*kiwicha*) exhibited limited haplotype diversity indicating a narrow genetic base, while *A. cruentus* and *A. hypochondriacus* showed broader diversity, with the nuclear ITS network providing clearer resolution than chloroplast markers due to biparental inheritance. Outlier genotypes, including G82, G83, G13, G10, and G39, occupied unique haplotype positions, confirming that their phenotypic distinctiveness corresponds to genuine genetic differentiation. The novelty of this study lies in integrating multivariate biostatistical techniques (heritability, path analysis, Mahalanobis  $D^2$ , PCA, and stepwise regression)



Academic Editors: Zhanying Zhang, Xingming Sun and Kun Zhang

Received: 4 April 2026

Revised: 6 May 2026

Accepted: 9 May 2026

Published: 13 May 2026

**Copyright:** © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

with two complementary DNA barcode systems (chloroplast and nuclear) within a single germplasm collection. This integrated approach provides breeders with well-characterized germplasm, validated selection criteria, and prioritized parental materials for Amaranth improvement. Further multi-location and multi-season evaluations are recommended to ensure the stability and adaptability of these promising germplasm accessions.

**Keywords:** *Amaranthus*; genetic diversity; heritability; path coefficient; Mahalanobis  $D^2$ ; multivariate statistics; *psbA-trnH*; ITS marker

## 1. Introduction

Amaranth (*Amaranthus* spp.) is a multifaceted genus of C4 plants that has served humanity for millennia as both a leafy vegetable and a pseudo-cereal grain. Originating in the Americas, where it was a staple food of pre-Columbian civilizations including the Aztecs and Incas, Amaranth has since spread to diverse agroecological regions across Africa, Asia, and Europe [1,2]. The genus encompasses approximately 60–70 species, including three principal grain Amaranths (*A. caudatus*, *A. cruentus*, and *A. hypochondriacus*) and numerous leafy vegetable types such as *A. tricolor*, *A. viridis*, and *A. dubius* [3,4]. These three grain species represent the primary Amaranth types utilized for food production globally and are now cultivated across diverse geographical regions spanning North America (USA and Canada), South America (Guatemala, Peru, and Argentina), Europe (Germany and France), Asia (India and China), and Africa (Ethiopia).

This remarkable geographic expansion reflects the crop's exceptional adaptability to adverse environmental conditions, including tolerance to saline and alkaline soils, resilience to high temperatures, capacity to thrive at high altitudes, and ability to withstand prolonged periods of water scarcity [5–7]. Moreover, Amaranth qualifies as an environmentally sustainable crop due to its inherent pest resistance [8]. Despite its historical significance, nutritional excellence, and remarkable adaptability, Amaranth remains underutilized in mainstream agriculture, often relegated to the category of orphan crops grown in marginal areas by resource-limited farmers [9–11].

Amaranth exhibits remarkable morphological diversity across the genus, with plant forms ranging from 0.3 to 3.0 m in height and displaying extensive variation in growth habits (erect, prostrate, or decumbent), stem color polymorphism (green, red, purple, or striped), branching intensity, pubescence, leaf arrangement (alternate or opposite), and inflorescence architecture (terminal and axillary spikes, panicles, or glomerules differing in density, length, orientation, and color) [2,3,12]. Seed characteristics also contribute to this diversity, with colors spanning yellow, white, brown, black, and pink, alongside considerable interspecific variation in seed size and surface luster [1,12]. Phenological traits such as days to flowering and maturity show wide variation across the genus, ranging from early- to late-maturing types, enabling adaptation to diverse photoperiod regimes and growing seasons [13–15]. This extensive morphological and phenological plasticity, while valuable for adaptation to diverse environments across tropical, subtropical, and temperate regions, complicates taxonomic classification as characteristics such as pigmentation show wide segregation and plant size varies with environmental factors including sunlight hours, soil conditions, and cultural practices [14]. Furthermore, frequent hybridization between cultivated and weedy species, particularly with *A. hybridus*, produces intermediate morphological forms, challenging species identification based solely on morphological descriptors [6,8,16,17].

Amaranth is predominantly self-pollinated, yet it exhibits a notable rate of natural cross-pollination (approximately 20–30%) mediated by wind [8]. This relatively high out-crossing rate justifies the application of molecular markers alongside phenotypic evaluation to accurately assess genetic diversity and population structure within germplasm collections [18,19]. Furthermore, Amaranth is a warm-season C4 crop typically cultivated from spring to early summer. Its exceptional adaptability to increasing temperatures and water scarcity positions it as a promising crop for climate change mitigation and adaptation strategies [5,6]. The present study, therefore, provides a valuable assessment of Amaranth germplasm that could be utilized across diverse European agro-climatic zones, as well as in other regions facing increasing environmental stresses.

Despite these attributes, Amaranth breeding programs have lagged behind those of major cereals, with most cultivation still relying on landraces and limited systematic improvement [6]. Genetic resources conserved in gene banks remain largely uncharacterized and underutilized, yet their effective utilization requires comprehensive evaluation of phenotypic diversity, genetic parameters, and molecular relationships [19]. Traditional breeding relying on visual selection for one or two traits often fails to capture complex yield architecture. While indirect selection based on simple phenotypic correlations is common, this approach reveals only the magnitude of association without distinguishing between direct and indirect effects [20]. Path coefficient analysis addresses this limitation by partitioning correlation coefficients into direct and indirect effects, enabling identification of traits with genuine influence on yield [21]. Building on this foundation, multivariate approaches have become indispensable: Mahalanobis  $D^2$  analysis quantifies genetic divergence to identify distinct accessions for hybridization [22]; principal component analysis reduces dimensionality to reveal trait association patterns [23,24]; and stepwise regression selects influential traits for predicting yield, providing focused selection indices [25]. The integration of these techniques transforms complex trait interrelationships into actionable criteria for efficient genetic improvement.

While phenotypic characterization and multivariate analysis provide valuable insights into trait variation and relationships, they cannot fully capture the evolutionary history and genetic relatedness among accessions. This limitation has been addressed through the application of DNA barcoding, which utilizes standardized genetic markers to assess biodiversity and evolutionary relationships [19,26]. In plants, the combination of chloroplast and nuclear DNA sequences has proven particularly informative for resolving taxonomic uncertainties and tracing evolutionary trajectories [12,26–28]. Among the various analytical approaches for visualizing genetic relationships, median-joining networks constructed from DNA barcode sequences have emerged as powerful tools for depicting haplotype diversity and evolutionary connections [29]. Unlike traditional phylogenetic trees, which assume bifurcating branching patterns, network approaches accommodate the reality of population-level processes including incomplete lineage sorting, hybridization, and recombination, making them particularly suitable for closely related species and intraspecific studies [18,30]. These networks display unique haplotypes as circles proportional to their frequency, with mutational steps represented as connecting lines, effectively illustrating both genetic distances and evolutionary relationships among accessions.

The integration of molecular data with phenotypic and multivariate analyses creates a comprehensive framework: multivariate approaches identify phenotypically distinct genotypes, while molecular networks validate whether this distinctiveness corresponds to genetic differentiation. This synergy is particularly valuable in *Amaranthus*, where morphological plasticity and hybridization complicate taxonomy [6].

Accessions identified through both approaches that occupy unique haplotype positions while exhibiting superior agronomic performance represent prioritized materials for conserva-

tion and breeding. Additionally, clustering patterns of unknown accessions alongside reference species provide provisional taxonomic assignments. DNA barcoding thus serves not as a standalone tool but as a complement that reinforces insights gained from phenotypic characterization [26]. Several knowledge gaps remain in Amaranth germplasm characterization. First, most studies have used either phenotypic or molecular approaches alone, lacking an integrated framework. In those cases, where molecular phylogeny has been applied, this was based on statistically inferred distances only. However, the specific SNPs required to change a given genotype to its neighbor are informative by themselves (homologisation based on specific quality). This can be addressed by inferring haplotype networks. Second, multivariate methods have not been systematically applied to identify promising genotypes with both superior agronomic performance and unique genetic profiles. Third, species-specific selection indices for grain Amaranth are largely unexplored. The study addresses these gaps by integrating multivariate biostatistical techniques with chloroplast and nuclear DNA barcoding.

The central hypothesis of the present study is that genotypes exhibiting extreme phenotypic profiles (multivariate outliers) will occupy unique haplotype positions in molecular networks, indicating that phenotypic distinctiveness corresponds to genuine genetic differentiation rather than environmental plasticity alone. Unlike previous studies that used only phenotypic or molecular approaches [14,24], the current work integrates both approaches within a single germplasm collection to test this hypothesis.

Accordingly, we undertook to comprehensively characterize 84 Amaranth genotypes representing multiple species and diverse origins. The specific objectives were to: (i) assess the extent of phenotypic variation for six agronomic traits; (ii) estimate genetic parameters including heritability and genetic advance; (iii) analyze trait associations and identify direct and indirect effects on grain yield; (iv) identify multivariate outliers and promising genotypes using  $D^2$  and PCA approaches; and (v) evaluate genetic relationships using chloroplast and nuclear DNA barcodes. The integration of these complementary approaches aims to provide breeders with well-characterized germplasm, validated selection criteria, and prioritized parental materials for accelerated Amaranth improvement.

## 2. Results

### 2.1. Phenotypic Variation Among 84 Amaranth Genotypes

The analysis of variance for days to flowering, plant height, inflorescence length, grain yield, cross-sectional seed area, and 1000-seed weight (TSW) revealed highly significant differences ( $p < 0.01$ ) among the 84 Amaranth genotypes (Table S1). Grain yield and 1000-seed weight data were square-root transformed prior to analysis to meet normality assumptions. These results indicate that sufficient genetic variability exists within the germplasm panel selected for this study. Block effects were significant only for days to flowering ( $p < 0.01$ ), while remaining non-significant for all other traits, indicating that environmental heterogeneity within the experimental field had minimal influence on most agronomic traits.

Extensive phenotypic variation was observed among the 84 Amaranth genotypes for all six agronomic traits evaluated under irrigated conditions (Table 1; Figure S1). The descriptive statistics revealed wide ranges and substantial variability, indicating considerable genetic diversity within the germplasm panel. Days to flowering ranged from 68.0 to 101.0 days (mean = 80.8 days), with a low coefficient of variation (CV = 0.87%), indicating moderate consistency in flowering behavior across genotypes. Plant height varied from 80.0 cm to 295.0 cm (mean = 179.1 cm), with substantial diversity (CV = 4.91%), encompassing both dwarf and exceptionally tall accessions. Inflorescence length ranged from 10.0 cm to 170.0 cm (mean = 91.0 cm), displaying a CV of 8.74%. Grain yield exhibited the widest variation, ranging from 0.25 g plant<sup>-1</sup> to 125.56 g plant<sup>-1</sup> (mean = 30.1 g plant<sup>-1</sup>), with an exceptionally high CV of 28.55%, demonstrating extraordinary diversity in productivity potential.

**Table 1.** Descriptive statistics for six agronomic traits measured in 84 Amaranth (*Amaranthus* spp.) genotypes evaluated under irrigated conditions with three replications ( $n = 252$  observations per trait).

Statistic	Mean $\pm$ SE	Min.	Max.	Range	Skewness	Kurtosis	CV%
Days to flowering (day)	80.82 $\pm$ 0.41	68.00	101.00	33.00	0.72	2.54	0.87
Plant height (cm)	179.13 $\pm$ 5.08	80.00	295.00	215.00	0.44	2.66	4.91
Inflorescence length (cm)	90.96 $\pm$ 4.59	10.00	170.00	160.00	0.26	4.00	8.74
Grain yield (g/plant)	30.14 $\pm$ 4.97	0.25	125.56	125.32	0.83	4.37	28.55
Cross-seed area (mm <sup>2</sup> )	1.22 $\pm$ 0.05	0.50	1.79	1.29	−0.20	3.67	7.70
1000-seed weight (g)	0.69 $\pm$ 0.04	0.15	1.11	0.96	−0.61	5.02	9.00

SE = Standard Error of the mean; CV = Coefficient of Variation (%).

Cross-sectional seed area ranged from 0.50 mm<sup>2</sup> to 1.79 mm<sup>2</sup> (mean = 1.22 mm<sup>2</sup>), with moderate variability (CV = 7.70%), while 1000-seed weight varied from 0.15 g to 1.11 g (mean = 0.69 g), with considerable variation (CV = 9.00%). The frequency distributions (Figure S1) showed continuous variation for all traits, with days to flowering and plant height exhibiting near-normal distributions (skewness = 0.72 and 0.44, respectively), while grain yield displayed positive skewness (0.83) and elevated kurtosis (4.37), indicating that while most genotypes produced moderate yields, a subset of high-yielding accessions contributed to the extended tail of the distribution. The wide ranges and CV values, particularly for grain yield (28.55%), inflorescence length (8.74%), and plant height (4.91%), confirm extensive genetic diversity within the germplasm, thereby providing significant scope for genetic improvement through selective breeding.

## 2.2. Genetic Parameters: Variance Components, Heritability, and Genetic Advance

Genetic parameter estimates revealed important insights into the inheritance patterns and selection potential of the six agronomic traits (Table 2). Phenotypic ( $\sigma^2_p$ ) and genotypic ( $\sigma^2_g$ ) variance were highest for plant height ( $\sigma^2_p = 2066.72$ ,  $\sigma^2_g = 1989.25$ ) and grain yield ( $\sigma^2_p = 431.18$ ,  $\sigma^2_g = 357.12$ ), indicating substantial genetic variability available for selection in these traits. The genotypic coefficient of variation (GCV) and phenotypic coefficient of variation (PCV) followed similar patterns across traits, with grain yield exhibiting the highest GCV (62.70%) and PCV (68.89%), thereby confirming exceptional genetic variability for productivity enhancement. Inflorescence length (GCV = 26.86%, PCV = 28.25%) and plant height (GCV = 24.90%, PCV = 25.38%) also showed high coefficients, while days to flowering displayed the lowest values (GCV = 9.83%, PCV = 9.87%). Minimal differences between GCV and PCV across all traits indicated that environmental influence on trait expression was limited, consistent with the ANOVA results.

**Table 2.** Genetic parameters for six agronomic traits in 84 Amaranth (*Amaranthus* spp.) genotypes evaluated under irrigated conditions ( $n = 252$  observations per trait).

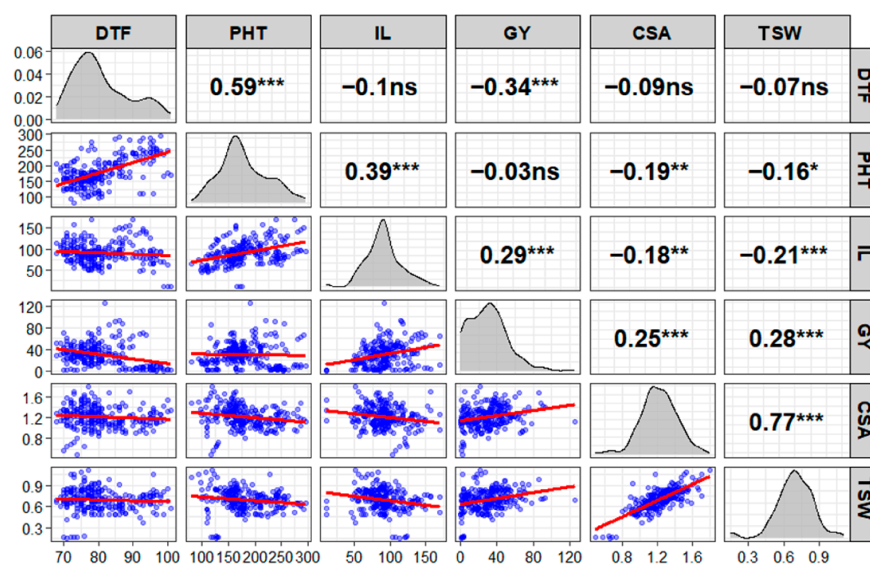
Statistic	$\sigma^2_g$	$\sigma^2_p$	GCV (%)	PCV (%)	H <sup>2</sup> b (%)	GA	GAM
Days to flowering	63.08	63.58	9.83	9.87	0.99	16.30	20.16
Plant height	1989.25	2066.72	24.90	25.38	0.96	90.14	50.32
Inflorescence length	597.08	660.26	26.86	28.25	0.90	47.87	52.63
Grain yield	357.12	431.18	62.70	68.89	0.83	35.43	117.54
Cross-seed area	0.03	0.04	15.00	16.86	0.79	0.33	27.49
1000-seed weight	0.02	0.02	20.56	22.46	0.84	0.27	38.77

$\sigma^2_p$  = phenotypic variance;  $\sigma^2_g$  = genotypic variance; GCV = genotypic coefficient of variation (%); PCV = phenotypic coefficient of variation (%); H<sup>2</sup>b = broad-sense heritability; GA = genetic advance; GAM = genetic advance as a percentage of the mean (%).

Broad-sense heritability ( $H^2b$ ) ranged from moderate to very high across all traits (Table 2). Days to flowering showed the highest heritability ( $H^2b = 0.99$ ), followed closely by plant height (0.96), inflorescence length (0.90), 1000-seed weight (0.84), grain yield (0.83), and cross-seed area (0.79). These high heritability values ( $>0.75$  for all traits) indicate that phenotypic selection would be effective for genetic improvement. Genetic advance as percentage of mean (GAM) revealed the expected gain from selection. Grain yield exhibited remarkably high GAM (117.54%), indicating that selecting the top-performing genotypes would result in more than doubling of mean yield. Inflorescence length (GAM = 52.63%) and plant height (GAM = 50.32%) also showed substantial expected gains, while days to flower (20.16%), and 1000-seed weight (38.77%) displayed moderate to high GAM values.

### 2.3. Trait Associations and Path Coefficient Analysis

Phenotypic and genotypic correlation analyses revealed complex interrelationships among the six agronomic traits (Figures 1 and S2), thereby providing insights into trait combinations that could facilitate or constrain simultaneous selection. Days to flowering showed significant negative phenotypic correlation with grain yield ( $r = -0.34$ ,  $p < 0.001$ ; Figure 1). This relationship was even stronger at the genotypic level ( $r = -0.36$ ,  $p < 0.001$ ; Figure S2). Days to flowering also exhibited significant positive correlations with plant height ( $r = 0.59$ ,  $p < 0.001$ ).

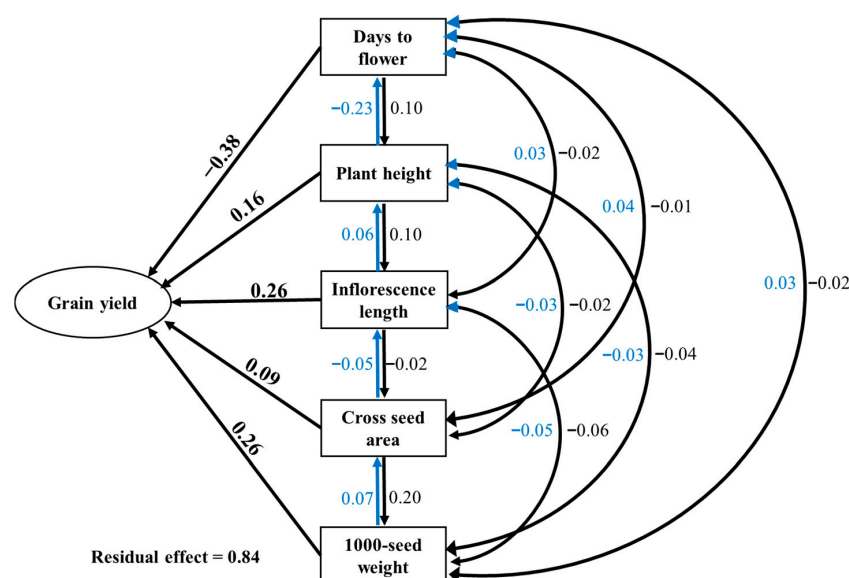


**Figure 1.** Phenotypic correlation matrix of six agronomic traits measured in 84 *Amaranthus* spp.) genotypes evaluated under irrigated conditions. The upper diagonal panels display Pearson correlation coefficients with significance levels: \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , ns = not significant. The lower diagonal panels show scatterplots with linear regression lines (red) indicating the direction and strength of relationships. Diagonal panels present density distributions for each trait. Trait abbreviations: DTF = days to flower, PHT = plant height (cm), IL = inflorescence length (cm), GY = grain yield ( $\text{g plant}^{-1}$ ), CSA = cross-seed area ( $\text{mm}^2$ ), TSW = 1000-seed weight (g).

Plant height demonstrated significant positive correlations with inflorescence length at both phenotypic ( $r = 0.39$ ,  $p < 0.001$ ) and genotypic ( $r = 0.39$ ,  $p < 0.01$ ; Figure S2) levels, indicating that taller plants develop longer inflorescences. However, plant height showed no significant direct correlation with grain yield. Inflorescence length exhibited the strongest positive correlation with grain yield among all traits (phenotypic  $r = 0.29$ ,  $p < 0.001$ ; genotypic  $r = 0.28$ ,  $p < 0.01$ ), establishing inflorescence development as a key determinant of yield potential. This trait also correlated positively with plant height. Cross-seed area and 1000-seed weight were strongly and positively correlated with each other (phenotypic

$r = 0.77, p < 0.001$ ; genotypic  $r = 0.82, p < 0.001$ ), as expected for related seed size parameters. Both traits showed significant positive correlations with grain yield, confirming that larger seeds contribute positively to overall yield.

Path coefficient analysis partitioned the correlation coefficients into direct and indirect effects to identify traits with genuine influence on grain yield versus those whose correlations arise indirectly through other traits (Figure 2; Table S2). Multicollinearity diagnostics confirmed acceptable levels for all traits (VIF < 3, tolerance > 0.40; Table S3), ensuring stability of path coefficients. Phenotypic path analysis (Figure 2) revealed that 1000-seed weight exerted the strongest positive direct effect on grain yield (0.26), despite its moderate phenotypic correlation ( $r = 0.28$ ). Inflorescence length showed a substantial positive direct effect (0.26) on grain yield, consistent with its significant positive correlation ( $r = 0.29$ ). Days to flower exhibited a strong negative direct effect ( $-0.38$ ) on grain yield, even stronger than its phenotypic correlation ( $r = -0.34$ ), confirming that early flowering directly enhances yield rather than acting through correlated traits. Plant height displayed direct effect (0.16) on grain yield despite its correlation with inflorescence length, indicating that its apparent association with yield operates entirely through inflorescence development.



**Figure 2.** Phenotypic path analysis showing direct and indirect effects (path coefficients) of five agronomic traits on grain yield based on phenotypic data from 84 Amaranth genotypes under irrigated conditions. Numbers adjacent to arrows are standardized path coefficients. The bold number represents the direct effect of each trait on the grain yield. The other numbers represent the indirect effect of each trait on the grain yield via other independent traits. Blue and black color numbers and arrows show the direction of indirect effect of the trait via other traits on the grain yield. The residual effect is 0.84.

Genotypic path analysis (Table S2) confirmed these patterns, with 1000-seed weight (0.29), inflorescence length (0.25), and days to flower ( $-0.43$ ) showing the strongest direct effects at the genetic level. The residual effect (0.84) indicates that approximately 16% of the variation in grain yield was explained by the five traits included in the model, suggesting that additional unmeasured factors contribute to yield determination.

#### 2.4. Stepwise Multiple Regression for Yield Prediction

Stepwise multiple regression analysis validated the path analysis findings and established a predictive model for grain yield based on the most influential traits (Table 3). The stepwise procedure selected four traits as significant predictors, entering them in the

following order: days to flowering entered first, explaining 11% of the yield variation (partial  $R^2 = 0.11$ ,  $p < 0.0001$ ) with a negative parameter estimate ( $-0.94$ ), confirming that early flowering enhances yield. Inflorescence length entered second, adding 6% to the explained variation (partial  $R^2 = 0.06$ ,  $p < 0.0001$ ) with a positive parameter estimate ( $0.21$ ), consistent with its role as a key yield component. One thousand-seed weight entered third, contributing 10% to the model (partial  $R^2 = 0.10$ ,  $p < 0.0001$ ) with the largest positive parameter estimate ( $44.14$ ), underscoring its strong direct influence on yield. Plant height entered fourth, adding only 1% to the explained variation (partial  $R^2 = 0.01$ ,  $p = 0.0854$ ) with marginal significance, confirming its limited direct contribution to yield.

**Table 3.** Relative contribution (partial and model  $R^2$ ),  $F$  value, and probability values for predicting Amaranth grain yield by the stepwise procedure analysis ( $n = 84$ ).

Step	Variable Entered	Partial $R^2$	Model $R^2$	Parameter Estimate	Standard Error	$p > F$
1	Days to flowering	0.11	0.11	$-0.94$	0.18	$<0.0001$
2	Inflorescence length	0.06	0.17	0.21	0.05	$<0.0001$
3	1000-seed weight	0.10	0.27	44.14	7.38	$<0.0001$
4	Plant height	0.01	0.28	0.06	0.03	0.0854

The final predictive model was:

$$Y = 45.24 - 0.94 (\text{DTF}) + 0.21 (\text{IL}) + 44.14 (\text{TSW}) + 0.06 (\text{PHT})$$

where DTF = days to flowering, IL = inflorescence length, TSW = 1000-seed weight, and PHT = plant height.

The cumulative model  $R^2$  of 0.28 indicates that these four traits collectively explain 28% of the variation in grain yield, consistent with the path analysis residual effect. Days to flower, inflorescence length, and 1000-seed weight emerged as the three most important predictors, together accounting for 27% of the yield variation, while plant height contributed negligibly.

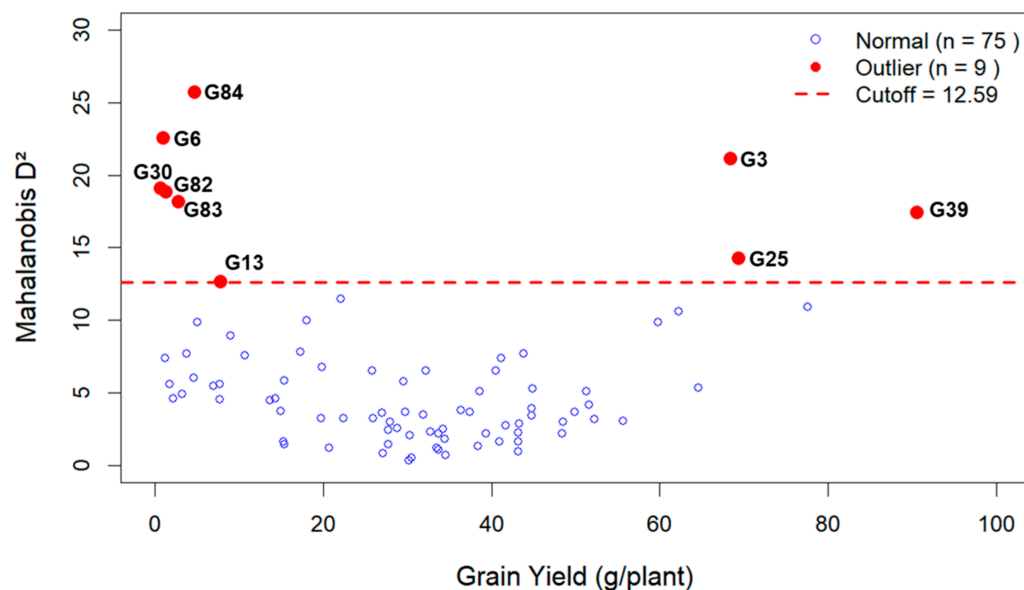
### 2.5. Multivariate Outlier Detection via Mahalanobis $D^2$ Analysis

Mahalanobis distance ( $D^2$ ) analysis based on six quantitative traits identified nine genotypes as multivariate outliers ( $D^2 > \chi^2 = 12.59$ ,  $df = 6$ ,  $\alpha = 0.05$ ), representing genetically distinct accessions with unique trait combinations (Table 4; Figures 3 and S3).

**Table 4.** Multivariate outlier genotypes identified by Mahalanobis  $D^2$  analysis in 84 Amaranth genotypes based on six agronomic traits evaluated under irrigated conditions.

Rank	Genotype ID	Mahalanobis $D^2$	$p$ Value	Significance	DTF	GY	PHT	IL	CSA	TSW	GY_CV
1	G84	25.72	0.0003	***	88.00	4.70	163.00	152.00	1.42	0.76	0
2	G6	22.57	0.0010	***	100.00	0.90	170.50	10.00	1.34	0.57	0
3	G3	21.12	0.0017	**	93.00	68.30	229.00	139.00	0.94	0.71	0
4	G30	19.09	0.0040	**	93.70	0.60	110.50	48.00	1.13	0.44	0
5	G82	18.89	0.0044	**	71.70	1.20	121.00	108.00	0.58	0.16	0
6	G83	18.18	0.0058	**	90.00	2.80	128.50	121.50	0.69	0.17	0
7	G39	17.47	0.0077	**	86.00	90.50	237.50	116.00	1.43	0.71	0
8	G25	14.29	0.0266	*	78.30	69.30	156.50	95.00	1.10	0.87	0
9	G13	12.65	0.0489	*	77.00	7.80	282.50	93.50	1.37	0.63	0

DTF = days to flower (days), GY = grain yield (g/plant), PHT = plant height (cm), IL = inflorescence length (cm), CSA = cross-seed area (mm), TSW = 1000-seed weight (g), GY\_CV = coefficient of variation for grain yield (%). Significance: \*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ . Critical value = 12.59. Genotypes were ranked according to  $D^2$  values. GY\_CV = 0 indicates perfect stability across replicates. The critical threshold and outlier classification were consistent between original and transformed grain yield data.



**Figure 3.** Scatter plot showing the relationship between grain yield and Mahalanobis  $D^2$  distance for 84 Amaranth genotypes. Blue circles = normal genotypes ( $n = 75$ ); red circles = outlier genotypes ( $n = 9$ ); red dashed line = critical value ( $\chi^2 = 12.59$ ,  $df = 6$ ,  $\alpha = 0.05$ ). Outlier genotypes are labeled. G39 (GY = 90.5 g/plant) is the top breeding candidate. The critical threshold and outlier classification were consistent between original and transformed grain yield data.

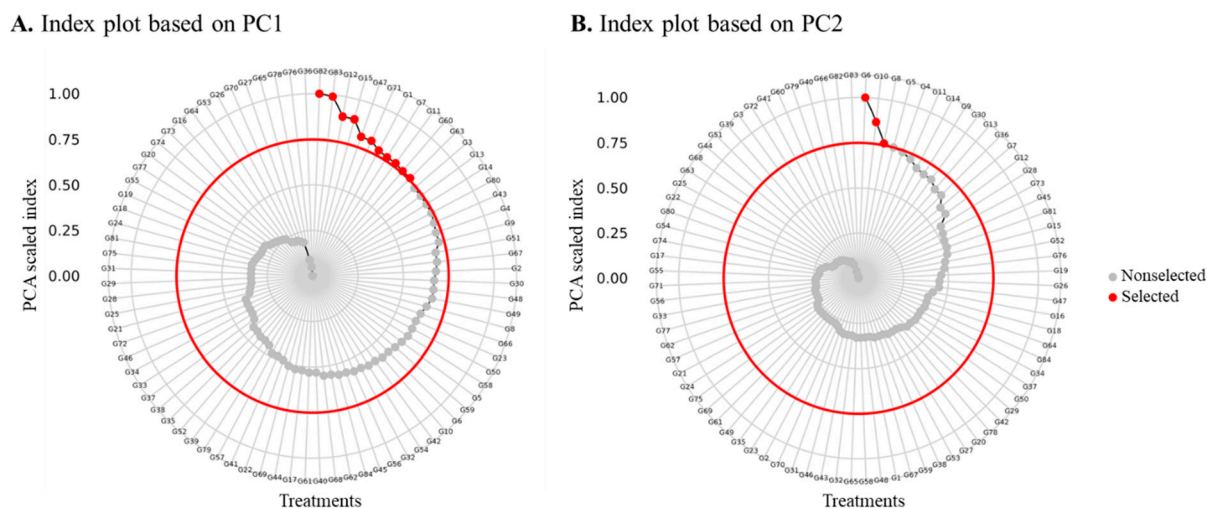
The outlier genotypes, ranked by  $D^2$  value, were G84 ( $D^2 = 25.72$ ,  $p < 0.001$ ) characterized by exceptionally long inflorescences (152.0 cm) and large seeds; G6 ( $D^2 = 22.57$ ,  $p < 0.001$ ) notable for late flowering (100 days) and extremely short inflorescence (10.0 cm); G3 ( $D^2 = 21.12$ ,  $p < 0.01$ ) combining tall stature (229.0 cm), long inflorescences (139.0 cm), and high yield (68.30 g/plant); G30 ( $D^2 = 19.09$ ,  $p < 0.01$ ) with late flowering and very low yield (0.60 g/plant); G82 ( $D^2 = 18.89$ ,  $p < 0.01$ ) exhibiting early flowering and small seeds; G83 ( $D^2 = 18.18$ ,  $p < 0.01$ ) also showing small seeds and low yield; G39 ( $D^2 = 17.47$ ,  $p < 0.01$ ) as the top breeding candidate with the highest yield among outliers (90.50 g/plant), tall stature (237.5 cm), long inflorescences (116.0 cm), and large seeds; G25 ( $D^2 = 14.29$ ,  $p < 0.05$ ) as high-yielding (69.30 g/plant) with large seeds; and G13 ( $D^2 = 12.65$ ,  $p < 0.05$ ) as extremely tall (282.5 cm) with moderate yield (Table 4).

The scatter plot of grain yield versus Mahalanobis distance (Figure 3) revealed that outlier status was associated with both exceptionally high-yielding (G39, G25, G3) and extremely low-yielding genotypes. Notably, G39 emerged as the most promising breeding candidate, combining outstanding yield with desirable traits including tall stature, long inflorescence, and large seeds, while maintaining perfect stability across replicates (GY\_CV = 0). These results corroborate the path analysis findings and provide a practical selection index: selection for early flowering, long inflorescences, and heavy seeds to maximize grain yield in Amaranth breeding programs.

## 2.6. Principal Component Analysis and Promising Genotype Selection

Principal component analysis (PCA) reduced the six agronomic traits to three principal components with eigenvalues  $>1.0$ , collectively explaining 85.2% of the total morphological variation (Figure S4). PC1 accounted for 37.7% of the variance and was primarily associated with plant height and inflorescence length, representing plant architecture and reproductive potential. PC2 explained 24.5% of the variance and was strongly loaded with cross-seed area and 1000-seed weight, representing seed size characteristics. PC3 contributed 23.0% of the variance and was associated with days to flower and grain yield, representing phenology and productivity.

Based on index scores standardized to a 0–1 scale, genotypes exceeding the 75th percentile (top 25%) were identified as promising for each principal component (Figure 4; Table S4). PC1-selected genotypes ( $n = 11$ ) exhibited superior plant architecture, characterized by exceptional height and inflorescence development, including G12 (scaled index = 0.89) at 286.5 cm tall with long inflorescence, G15 (scaled index = 0.89) at 236.0 cm, and G82 (scaled index = 1.00) showing extreme inflorescence length (108.0 cm) relative to stature. PC2-selected genotypes ( $n = 3$ ) exhibited superior seed size characteristics, with G6 (scaled index = 1.00) having large seeds (1.34 mm<sup>2</sup>), G8 (scaled index = 0.76) with large seeds (1.28 mm<sup>2</sup>), and G10 (scaled index = 0.87) possessing the largest seeds among the selected genotypes (1.38 mm<sup>2</sup>, 0.83 g/1000).



**Figure 4.** Index plots of the (A) first principal component (PC1) and (B) second principal component (PC2) for 84 Amaranth genotypes evaluated under irrigated conditions, based on six agronomic traits. Index scores were standardized to a 0–1 scale for direct comparability. The red line marks the 75% cutoff threshold. Genotypes above the 75th percentile (top 25%) are shown as red circles (promising genotypes), and genotypes below as gray circles (non-selected). The selected genotypes for PC1 and PC2 are presented in Supplementary Table S4.

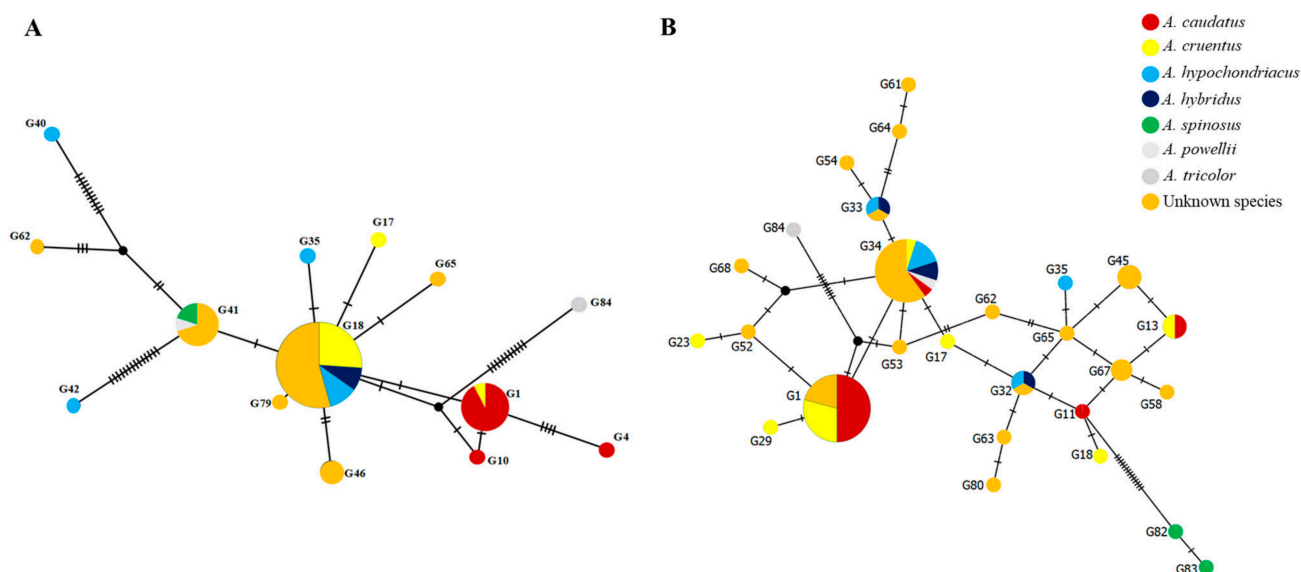
Integration with the outlier analysis revealed that three outlier genotypes were also identified among the PC-selected promising genotypes: G82 and G83 as top performers in PC1, and G6 as the top performer in PC2. Notably, G39, the top-yielding outlier, showed balanced performance across components but did not exceed the 75th percentile threshold in either PC1 or PC2 individually.

### 2.7. Molecular Diversity: Haplotype Network Analysis

The median-joining networks based on two DNA barcode markers, the intergenic spacer region separating the conserved *psbA* and *trnH* plastidic genes (chloroplast DNA) and the nuclear internal transcribed spacer (ITS, including ITS1, 5.8S, and ITS2), revealed the genetic relationships and haplotype distribution among the 84 Amaranth genotypes (Figure 5A,B).

The *psbA-trnH* chloroplast DNA network resolved multiple distinct haplotypes, with circle size proportional to haplotype frequency (Figure 5A). Several haplotypes were shared among multiple accessions, while others were unique to individual genotypes. Mutational steps between haplotypes revealed the evolutionary relationships and genetic distances among accessions. *A. caudatus* accessions formed a distinct cluster with limited haplotype diversity, suggesting a relatively narrow genetic basis within the cultivated grain Amaranth group. *A. cruentus* and *A. hypochondriacus* accessions showed greater haplotype dispersion,

indicating broader genetic diversity. Unknown species were distributed across multiple haplotypes, some clustering with known species (suggesting putative taxonomic affinities) and others occupying distinct positions (representing potentially novel genetic groups).



**Figure 5.** Median-joining networks based on DNA barcode sequences of 84 Amaranth genotypes. (A) *psbA-trnH* chloroplast DNA spacer region. (B) Nuclear internal transcribed spacer (ITS), including ITS1, 5.8S, and ITS2). Different colors represent different *Amaranthus* species: dark red = *A. caudatus*, yellow = *A. cruentus*, light blue = *A. hypochondriacus*, dark blue = *A. hybridus*, green = *A. spinosus*, gray = *A. powellii*, light gray = *A. tricolor*, and orange = unknown species. Circle size corresponds to the number of individuals sharing the same haplotype. Dashes along the connections represent single mutational steps (whereas black circles indicate haplotypes not observed in this study). The networks were constructed using PopART and edited using Microsoft PowerPoint.

The nuclear ITS DNA barcode network provided additional insights into the genetic diversity and relationships among the 84 Amaranth genotypes (Figure 5B). The ITS network revealed distinct haplotype groupings, with circle sizes reflecting the frequency of each haplotype within the germplasm collection. Multiple accessions shared common haplotypes, while others displayed unique haplotypes, indicating varying levels of genetic relatedness. *A. caudatus* accessions clustered together with limited haplotype variation, confirming the narrow genetic base observed in the chloroplast analysis. *A. cruentus* and *A. hypochondriacus* accessions were distributed across multiple haplotypes, reflecting broader genetic diversity within these species. Unknown species accessions were dispersed throughout the network, with some grouping alongside known species (providing clues for taxonomic assignment) and others occupying distinct positions (representing potentially unique genetic lineages). *A. hybridus*, *A. powellii*, *A. spinosus*, and *A. tricolor* occupied peripheral positions in the network, consistent with their genetic distinctiveness.

### 3. Discussion

#### 3.1. Amaranth Genotypes Display Extensive Phenotypic Variation

Analysis of variance revealed highly significant differences ( $p < 0.01$ ) among the 84 Amaranth genotypes for all six agronomic traits, confirming substantial genetic variability within the germplasm panel. These findings align with previous studies documenting significant genotypic variation in Amaranth collections from Ethiopia [14], Malawi [24], and exotic vegetable Amaranth genotypes [31]. The highly significant genotypic effects, coupled with non-significant block effects for most traits, indicate that observed variation

reflects true genetic differences rather than environmental artifacts, thereby providing a reliable foundation for genetic improvement.

Productivity potential showed the widest variation, ranging from very low to exceptionally high values with a high coefficient of variation, comparable to findings from Singh [32], Yeshitila et al. [14] and Sefasi et al. [24]. The presence of genotypes at both extremes of the distribution provides breeders with valuable materials for understanding yield determination and direct selection. Architectural traits also displayed remarkable variation, with plant stature ranging from dwarf to tall types and inflorescence characteristics varying from extremely short to exceptionally long, consistent with reports from Yeshitila et al. [14] and Sefasi et al. [24]. Yeshitila et al. [15] similarly reported extensive variation in plant height among Ethiopian Amaranth genotypes.

Phenological traits showed relatively lower variation, suggesting tighter genetic control within cultivated Amaranth. Nevertheless, the observed range provides sufficient variation for adaptation to different photoperiod regimes and growing seasons [15,33]. The negative association between flowering time and productivity indicates that early-maturing genotypes tend to be more productive, likely due to their ability to complete grain filling before terminal stresses.

The wide ranges and CV values confirm that this germplasm collection captures significant genetic diversity, which is essential for sustained genetic gain through selection. The inclusion of multiple species and unknown accessions further enriches the genetic base, offering opportunities for interspecific hybridization and introgression of desirable traits [16,17].

### *3.2. High Heritability Enables Effective Selection for Key Traits*

Genetic parameter estimates revealed high broad-sense heritability for all six agronomic traits, indicating that a substantial proportion of the phenotypic variation is attributable to genetic factors and would therefore respond favorably to selection. Similar high heritability estimates have been reported in Amaranth by Shrivastav et al. [20], Yadav et al. [13], Venkatesh et al. [34], and Patial et al. [35]. The high heritability coupled with low environmental influence suggests that phenotypic selection would be effective for genetic improvement. The highest heritability was observed for phenological traits, indicating strong genetic control with minimal environmental modification. This finding is consistent with observations reported by other researchers [13,35–37]. The combination of high heritability and a strong negative association with grain yield suggests that selection for early-flowering genotypes would be an effective yield improvement strategy.

Architectural traits also showed high heritability values, in agreement with Yeshitila et al. [14]. The combination of high heritability and high genetic advance for these traits indicates additive gene action and a favorable response to selection. Similarly, high genetic advance for morphological traits in Amaranth variability studies has been reported by Oduwaye et al. [37], Yadav et al. [13], and Patial et al. [35]. Grain yield exhibited high heritability coupled with remarkably high genetic advance, a combination particularly favorable for breeding programs as it indicates additive gene control and substantial potential for genetic gain. The combination of high heritability and high genetic advance for productivity and architectural traits suggests these traits are governed by additive gene action and would respond favorably to selection, thereby emerging as the most promising candidates for direct genetic improvement in Amaranth breeding programs.

### *3.3. Grain Yield Is Determined by Flowering Time, Inflorescence Length and Seed Weight*

Understanding the interrelationships among traits and their effects on productivity is essential for developing efficient selection strategies in Amaranth breeding programs.

Correlation and path coefficient analyses revealed that productivity is primarily determined by three key traits: phenology, inflorescence development, and thousand-seed weight.

A significant negative association with productivity was observed for phenological traits, indicating that earlier-maturing genotypes tend to produce higher yields [38,39]. Path analysis confirmed that this relationship is due to strong direct effects rather than indirect influences. Under the temperate conditions of Germany, where this study was conducted, temperatures drop considerably after August, and late-flowering genotypes experience cold stress during flowering and grain filling, which negatively impacts fertilization, seed set, and ultimately yield. Early-maturing genotypes complete their reproductive cycle before the onset of cold conditions, directly enhancing their yield potential. Similar findings have been reported by Henderson et al. [40], Artemyeva et al. [41], Dmitrieva and Ivanov [42], and Kanbar et al. [7].

From a physiological perspective, early flowering in Amaranth confers a significant adaptive advantage under temperate conditions. Flowering and grain filling are highly sensitive to low temperatures, which can disrupt pollen viability, reduce fertilization success, and limit seed set and development [43,44]. Early-maturing genotypes complete their reproductive cycle before the onset of cold stress, thereby avoiding these physiological constraints. Ecologically, this strategy represents a classic stress-escape mechanism, allowing plants to synchronize their life cycle with favorable seasonal windows [45]. Amaranth originated in warm tropical and subtropical regions of the Americas, and early-flowering genotypes are therefore better pre-adapted to temperate growing seasons such as those in Central Europe. This ecological adaptation has direct implications for breeding: early flowering should be prioritized when developing Amaranth varieties for temperate or high-latitude environments, whereas longer-season genotypes may be more suitable for tropical conditions where temperature is not limiting.

The strongest positive association with productivity was exhibited by inflorescence characteristics, with path analysis revealing substantial direct effects. This indicates that inflorescence development is a primary determinant of yield, functioning as the major sink structure where seeds develop. Longer inflorescences provide greater surface area for seed production. This findings are consistent with observations reported by other researchers [13,20,24,37,38].

The strongest positive direct effect on productivity was exerted by thousand-seed weight, despite its moderate phenotypic correlation. Path analysis revealed that the true contribution of thousand-seed weight is partially masked by opposing indirect effects through other traits, underscoring the importance of path analysis in revealing genuine trait relationships. Oduwaye et al. [37] and Venkatesh et al. [46] similarly found that thousand-seed weight had a high positive direct effect on grain yield, while Akther et al. [47] reported its significance in genetic divergence studies.

No significant direct association with productivity was observed for plant height, with path analysis revealing that its apparent relationship with yield operates entirely through inflorescence development, as documented by Yeshitila et al. [39], Shukla and Singh [48], and Venkatesh et al. [46].

Stepwise multiple regression corroborated these findings, with phenology, inflorescence length, and thousand-seed weight emerging as the three most important predictors. The identification of early maturity, long inflorescences, and heavy seeds provides breeders with potential selection criteria. However, the residual effect of 0.84 indicates that these three traits collectively explain only 16% of the variation in grain yield (path analysis) or 27% (stepwise regression). The majority of yield variation (72–84%) therefore remains unexplained by the model, suggesting that other unmeasured factors, such as seed number per inflorescence, branching architecture, harvest index, or physiological traits, likely con-

tribute substantially to grain yield. Therefore, while early flowering, long inflorescences, and heavy seeds appear to be positive contributors, they should not be considered the definitive or sole determinants of yield. The absence of antagonistic relationships among these traits suggests that simultaneous improvement is feasible, thereby allowing the development of ideotypes combining all three characteristics. The importance of early maturity, long inflorescences, and thousand-seed weight as key yield determinants has been demonstrated across multiple *Amaranthus* species using stepwise multiple regression and path analysis [35,37,46,49]. However, these studies also emphasized that selection criteria should be species-specific: thousand-seed weight and stem diameter were effective for *A. hypochondriacus*, while thousand-seed weight and number of leaves were more suitable for *A. cruentus*.

### 3.4. Promising Genotypes Outperform the Germplasm Collection

Multivariate approaches including Mahalanobis  $D^2$  analysis and PCA enabled the detection of nine outlier genotypes with distinct phenotypic profiles and selection of promising accessions for specific trait complexes, consistent with approaches used by Sefasi et al. [24] and Yeshitila et al. [50]. Among the outliers, G39 emerged as the most promising breeding candidate, combining outstanding grain yield with tall stature, long inflorescence, and large seeds. This yield considerably exceeds those previously reported genotypes [32,35]. Two additional high-yielding outliers, G3 and G25, exhibited favorable trait combinations representing alternative ideotypes for different production systems. The exceptional performance of these three genotypes makes them ideal candidates for direct use as parents in hybridization programs or for potential release as varieties following further multi-location testing.

Principal component analysis provided a complementary framework for the selection of promising genotypes. PC1, representing plant architecture, identified 11 genotypes as valuable donors for improving inflorescence development. PC2, representing seed size characteristics, identified three genotypes possessing the largest seeds in the collection.

Similar PCA applications for trait identification and elite genotype selection in grain *Amaranthus* have been reported [24,39]. The limited overlap between  $D^2$  outliers and PC-selected elites reflects their different objectives:  $D^2$  identifies genotypes with extreme combinations across all traits, while PCA targets those excelling in specific trait complexes. This explains why G39, despite top yield, did not exceed the 75th percentile in individual PCs; its superiority derives from optimal combination of traits rather than from the extreme expression of any single trait.

The field experiment was conducted at a single location (Karlsruhe, Germany) during one growing season (2020). Consequently, the heritability estimates reported reflect variance components under these specific conditions and may not be transferable to other environments without accounting for genotype  $\times$  environment ( $G \times E$ ) interaction. Similarly, the identification of G39 as a high-yielding genotype does not guarantee its performance under different soil types, climate conditions, or management practices. Therefore, the term “promising” is used here to denote superior performance under the conditions tested, not as a validated recommendation for release. Trials across multiple seasons and locations are essential to assess  $G \times E$  interactions and confirm the stability of G39 and other candidate genotypes (G3, G25) before any breeding or variety release decisions can be made.

### 3.5. Phenotypic and Molecular Data Together Guide Amaranth Breeding

The integration of phenotypic and molecular data in the present study provides a comprehensive framework for developing efficient breeding strategies in Amaranth. The complementary nature of these approaches offers breeders multiple entry points

for genetic improvement: morphological characterization for trait evaluation, genetic parameter estimation for understanding inheritance patterns, multivariate analyses for genotype selection, and DNA barcoding for assessing genetic relationships.

The close phylogenetic relationships and the tendency for hybridization pose limitations to the resolution of barcoding markers. However, the individual SNPs required to interchange between two given genotypes are informative and can improve resolution, because statistical measures of difference are complemented by specific patterns, comparable to a fingerprint. This strategy has been widely used to infer evolutionary relationships between closely related taxa under the designation criterion of specific quality [51]. We made use of this strategy by inferring haplotype networks based on chloroplast *psbA-trnH* and nuclear ITS sequences that revealed distinct genetic relationships among the 84 Amaranth genotypes and provided molecular evidence for the complex domestication history of the genus. Clear separation of the major grain Amaranth species was observed, with *A. caudatus* accessions forming a compact cluster with limited haplotype diversity, suggesting a relatively narrow genetic base within this cultivated species. This finding aligns with Jamalluddin et al. [19], who reported reduced genetic diversity in cultivated Amaranth species due to domestication bottlenecks.

Greater haplotype dispersion was observed for *A. cruentus* and *A. hypochondriacus* accessions, indicating broader genetic diversity within these species. This pattern reflects the complex domestication history of Amaranth, a crop considered incompletely domesticated. Amaranth was domesticated independently twice in the Americas: once in Peru, giving rise to *A. caudatus* (*kiwicha*) from the wild ancestor *A. quitensis*, and once in Mexico, producing *A. cruentus* and *A. hypochondriacus* from the common progenitor *A. hybridus* [1,6]. Following these independent domestication events, there was subsequent gene flow and exchange between these centers, further complicating the genetic architecture of cultivated forms. Thus, when researchers or breeders refer to “Amaranth,” they may be working with genetically distinct crop types originating from different ancestral species with unique evolutionary trajectories.

Clear differences were observed between the two marker systems. The nuclear ITS network exhibited clearer separation and greater haplotype diversity than the chloroplast *psbA-trnH* network, due to fundamental differences in inheritance patterns. Nuclear markers are biparentally inherited, capturing genetic contributions from both parents and providing a comprehensive view of evolutionary relationships. In contrast, chloroplast DNA is maternally inherited, tracing only seed-mediated gene flow and maternal lineages. The higher mutation rate and faster evolution of nuclear ribosomal DNA make ITS particularly informative for distinguishing closely related accessions and revealing recent evolutionary divergence.

Building on our previous study using the same germplasm [12], which focused on species authentication, the present study revealed a notable concordance between phenotypic and molecular diversity. The nine genotypes identified as multivariate outliers by Mahalanobis  $D^2$  analysis, including the top-yielding candidate G39, consistently occupied unique or peripheral positions in both chloroplast and nuclear ITS haplotype networks. This convergence provides independent validation that phenotypic distinctiveness corresponds to genuine genetic differentiation, a pattern similarly observed in integrated studies of other crops such as rice [52,53], sorghum [54], sugarcane [55], wheat [56], and Brassica [57]. Compared to these studies, our work uniquely combines multiple multivariate techniques (path analysis, stepwise regression,  $D^2$ , PCA) with two complementary DNA barcode systems (chloroplast and nuclear) within a single germplasm collection, offering a more comprehensive framework for linking phenotypic performance to genetic identity. This integrated approach is particularly valuable in *Amaranthus*, where morphological

plasticity and frequent hybridization complicate species identification based solely on visual traits.

It is acknowledged that the *psbA-trnH* and ITS markers used in this study have known limitations in resolving closely related species in certain plant groups. However, these markers remain among the most informative and widely used barcodes for plants [58–64]. In our previous work using the same markers [12], we successfully identified a species-specific Single Nucleotide Polymorphism (SNP) in the *psbA-trnH* region that separated *A. caudatus* from all other *Amaranthus* species, enabling the development of a sequencing-free Amplified Refractory Mutation System (ARMS) authentication assay. This demonstrates that, despite general limitations, these markers were informative for our germplasm collection. Nevertheless, higher-resolution genotyping (e.g., SSR markers, SNP arrays) would be required for definitive taxonomic assignment of the 38 unknown accessions. Therefore, our haplotype assignments for unknown accessions are presented as provisional and should be interpreted with caution.

The 38 accessions of unknown taxonomic status were distributed across multiple haplotypes, with some clustering alongside known species and others occupying distinct positions. This pattern provides initial clues for taxonomic assignment and highlights the value of molecular markers for resolving taxonomic uncertainties in *Amaranthus*, a genus where morphological identification is often challenging due to phenotypic plasticity and ongoing gene flow between domesticated and weedy forms [65]. The integration of phenotypic and molecular diversity in this study is primarily qualitative; a quantitative approach, such as a Mantel test correlating phenotypic and genetic distance matrices, would provide a more rigorous assessment and is recommended for future studies.

## 4. Materials and Methods

### 4.1. Plant Materials

A total of 84 Amaranth genotypes were evaluated (Table 5), consisting of 15 *A. caudatus*, 14 *A. cruentus*, 9 *A. hypochondriacus*, 4 *A. hybridus*, 2 *A. spinosus*, 1 *A. tricolor*, 1 *A. powellii*, and 38 genotypes of uncertain taxonomic classification. The germplasm encompasses commercially utilized species, landraces, and agriculturally bred varieties, thereby capturing a substantial portion of the genetic diversity currently in human use. All accessions were raised to flowering and their taxonomic identities were determined following digital documentation of floral traits and plant habitus according to taxonomic keys of the Flora of China ([http://www.efloras.org/florataxon.aspx?flora\\_id=2&taxon\\_id=101257](http://www.efloras.org/florataxon.aspx?flora_id=2&taxon_id=101257)) (accessed on 13 March 2000), the Flora of North America ([http://www.efloras.org/florataxon.aspx?flora\\_id=1&taxon\\_id=101257](http://www.efloras.org/florataxon.aspx?flora_id=1&taxon_id=101257)) (accessed on 13 March 2000) [66].

**Table 5.** List of Amaranth genotypes used in the study, including species designation, number of accessions per species, and genotype identifiers.

Species Name	No. of Genotypes	Name of Genotypes
<i>A. caudatus</i>	15	G1 ( <i>A. caudatus</i> _7469), G2 ( <i>A. caudatus</i> _3807), G3 ( <i>A. caudatus</i> _8102), G4 ( <i>A. caudatus</i> _8600), G5 ( <i>A. caudatus</i> cv. Beige_3707), G6 ( <i>A. caudatus</i> cv. Fox_Peter 4_8299), G7 ( <i>A. caudatus</i> cv. Marron_2496), G8 ( <i>A. caudatus</i> cv. Negro_9442), G9 ( <i>A. caudatus</i> cv. Oscar Blanco 1_9444), G10 ( <i>A. caudatus</i> cv. Oscar Blanco 2_9445), G11 ( <i>A. caudatus</i> cv. Oscar Blanco 3_9425), G12 ( <i>A. caudatus</i> cv. Peter 1_8300), G13 ( <i>A. caudatus</i> cv. Peter 2_8301), G14 ( <i>A. caudatus</i> cv. Peter 3_8302), G15 ( <i>A. caudatus</i> cv. Rosado_9408)

Table 5. Cont.

Species Name	No. of Genotypes	Name of Genotypes
<i>A. cruentus</i>	14	G16 ( <i>A. cruentus</i> cv. Andy_8055), G17 ( <i>A. cruentus</i> _7470), G18 ( <i>A. cruentus</i> cv. A 10_8043), G19 ( <i>A. cruentus</i> cv. Amar_8091), G20 ( <i>A. cruentus</i> cv. Anna_8081), G21 ( <i>A. cruentus</i> cv. Bärnkrafft_8080), G22 ( <i>A. cruentus</i> cv. K 266_8037), G23 ( <i>A. cruentus</i> cv. K 436_8040), G24 ( <i>A. cruentus</i> cv. K72_8049), G25 ( <i>A. cruentus</i> cv. MT3_8041), G26 ( <i>A. cruentus</i> cv. Nu World_8042), G27 ( <i>A. cruentus</i> cv. Puerto Moult_8056), G28 ( <i>A. cruentus</i> cv. Suvarna_8046), G29 ( <i>A. cruentus</i> cv. Villarica_8057)
<i>A. hybridus</i>	4	G30 ( <i>A. hybridus</i> _8939), G31 ( <i>A. hybridus</i> cv. Pastewny_8052), G32 ( <i>A. hybridus</i> cv. Turkiestan_8053), G33 ( <i>A. hybridus</i> cv. Ural_8054)
<i>A. hypochondriacus</i>	9	G34 ( <i>A. hypochondriacus</i> _8104), G35 ( <i>A. hypochondriacus</i> cv. Anderer Typ_8085), G36 ( <i>A. hypochondriacus</i> cv. Anderer Typ_8096), G37 ( <i>A. hypochondriacus</i> cv. K343_8038), G38 ( <i>A. hypochondriacus</i> cv. K343 bunt_8083), G39 ( <i>A. hypochondriacus</i> cv. Mittlerer Typ (8084), G40 = <i>A. hypochondriacus</i> cv. Mittlerer Typ OR dunkel_8095), G41 ( <i>A. hypochondriacus</i> cv. Mittlerer Typ OR hell_8090), G42 ( <i>A. hypochondriacus</i> x <i>hybridus</i> cv. K432_8039)
<i>A. powellii</i>	1	G43 ( <i>A. powellii</i> _8938)
Unknown species	38	G44 ( <i>A. species</i> _8079), G45 ( <i>A. species</i> _8601), G46 ( <i>A. species</i> cv. Arma_8082), G47 ( <i>A. species</i> cv. Arora 2_8298), G48 ( <i>A. species</i> cv. C2_8060), G49 ( <i>A. species</i> cv. C3_8061), G50 ( <i>A. species</i> cv. C4_8044), G51 ( <i>A. species</i> cv. C5_8062), G52 ( <i>A. species</i> cv. C6_8045), G53 ( <i>A. species</i> cv. Energiety_8094), G54 ( <i>A. species</i> cv. K10_8063), G55 ( <i>A. species</i> cv. K25_8064), G56 ( <i>A. species</i> cv. K39_8065), G57 ( <i>A. species</i> cv. K41_8066), G58 ( <i>A. species</i> cv. K42_8067), G59 ( <i>A. species</i> cv. K47_8068), G60 ( <i>A. species</i> cv. K48_8069), G61 ( <i>A. species</i> cv. K50_8070), G62 ( <i>A. species</i> cv. K51_8071), G63 ( <i>A. species</i> cv. K53_8072), G64 ( <i>A. species</i> cv. K61_8047), G65 ( <i>A. species</i> cv. K62_8073), G66 ( <i>A. species</i> cv. K63_8048), G67 ( <i>A. species</i> cv. K64_8074), G68 ( <i>A. species</i> cv. K67_8058), G69 ( <i>A. species</i> cv. K71_8075), G70 ( <i>A. species</i> cv. K78_8050), G71 ( <i>A. species</i> cv. K80_8076), G72 ( <i>A. species</i> cv. K88_8077), G73 ( <i>A. species</i> cv. K91_8051), G74 ( <i>A. species</i> cv. Mittlerer Typ Rot_8089), G75 ( <i>A. species</i> cv. Mittlerer Typ Rot_8093), G76 ( <i>A. species</i> cv. Neuer Typ_8087), G77 ( <i>A. species</i> cv. Neuer Typ_8092), G78 ( <i>A. species</i> cv. NTCX_8088), G79 ( <i>A. species</i> cv. Poleski Ukrai_8078), G80 ( <i>A. species</i> cv. Sierra Leone_8059), G81 ( <i>A. species</i> cv. Typ X_8086)
<i>A. spinosus</i>	2	G82 ( <i>A. spinosus</i> _3809), G83 ( <i>A. spinosus</i> _8290)
<i>A. tricolor</i>	1	G84 ( <i>A. tricolor</i> _8103)

#### 4.2. Greenhouse Planting and Field Transplanting

A field experiment was conducted under temperate conditions at the Botanical Garden of Karlsruhe Institute of Technology (KIT), Germany (49°0′24.8004″ N, 8°24′13.1508″ E; 119 m above sea level), characterized by an oceanic climate with mean temperatures ranging from −1 °C in winter to 26 °C in summer. On 15 April 2020, surface-sterilized seeds of all 84 accessions were sown in 100-well trays filled with a 1:1:1 mixture of peat moss, perlite, and soil. Six seeds were planted per well and later thinned to one seedling per well after germination, with 50 seedlings raised per accession in the greenhouse at 25 ± 3 °C under a 12 h photoperiod. Three weeks post-germination, seedlings were transplanted to the field in a randomized complete block design with three replications. Each plot measured 2.5 m in length with row spacing of 0.65 m and plant spacing of 0.40 m within rows. Each block contained three rows per accession; the two outer rows served as borders, while data were collected from plants in the middle row. Based on soil analysis, organic fertilizer (Hauert

Hornoska<sup>®</sup> Special) at 100 g/m<sup>2</sup> and NPK at 90:60:40 kg/ha were applied. Soil moisture was maintained at 70–80% field capacity through irrigation until seed maturity.

#### 4.3. Phenotyping and Morphological Evaluation

For morphological observations, three plants were randomly selected from the center of the second row in each plot. This sample size is consistent with standard protocols for Amaranth germplasm characterization, where three to five representative plants per replication are commonly used to balance statistical reliability with practical feasibility [33,67]. Data recorded included days to flower, plant height (cm), inflorescence length (cm), grain yield (g/plant), 1000-seed weight (g), and seed cross area (mm<sup>2</sup>). Inflorescence length was measured from the lowest to the highest flower on each plant. Seed cross area was determined from three randomly chosen seeds per plant sample in each replicate. Individual seeds were imaged at a fixed orientation using a Keyence VHX-950F digital microscope (Keyence, Neu-Isenburg, Germany), and area measurements were obtained from digital images using ImageJ 1.53 Version software (<https://imagej.net/ij/>; accessed on 11 January 2023). At full maturity, panicles were hand-harvested with pruning shears, dried in a hoop house for two weeks, and manually threshed. The extracted seeds were winnowed and evaluated for grain yield and 1000-seed weight.

#### 4.4. DNA Extraction, PCR Amplification and Sequencing

Genomic DNA of the 84 Amaranth accessions was extracted following the CTAB protocol described in our previous work [12]. Briefly, 100 mg of frozen leaf tissue was ground, incubated with CTAB extraction buffer containing β-mercaptoethanol at 65 °C for one hour, followed by chloroform-isoamyl alcohol purification and isopropanol precipitation. The extracted DNA was dissolved in nuclease-free water with RNase A, and its concentration and purity were assessed spectrophotometrically.

Two DNA barcode markers were amplified by PCR: the chloroplast intergenic spacer region separating the *psbA* and *trnH* genes, and the nuclear internal transcribed spacer (ITS) comprising ITS1, 5.8S, and ITS2 (See Table S5). Each 30 μL PCR reaction contained 20.4 μL nuclease-free water, 3 μL Thermopol Buffer, 3 μL bovine serum albumin, 0.6 μL dNTPs, 0.6 μL each of forward and reverse primers (200 nM), 0.3 μL Taq polymerase, and 1.5 μL template DNA (50 ng/μL). Thermal cycling conditions for both markers are detailed in Table S5. Amplification products were visualized by gel electrophoresis using 1.5% agarose stained with Midori Green, with fragment sizes determined against a 100 bp DNA ladder. PCR products were purified using the MSB Spin PCRapace Kit prior to sequencing, which was outsourced to MacroGen Europe B.V. (Amsterdam, The Netherlands) and Eurofins (Ebersberg, Germany). Sequence quality was assessed using FinchTV Version 1.4.01 software, and all sequences were deposited in the NCBI database.

#### 4.5. Data Processing

Mean value, standard error, maximum value, minimum value, range, skewness, kurtosis, and coefficient of variation of 6 measurable morphological traits were calculated using R studio software (V4.5.2). These descriptive statistics provide a basic understanding of data distribution and variability, which is fundamental for subsequent in-depth analyses. Before the analysis, the data were subjected to normality testing to ensure conformity to basic assumptions of analysis of variance (ANOVA) [68]. Grain yield and 1000-seed weight data were transformed using the square root transformation method in R studio software (V4.5.2). Consequently, two-way ANOVA (genotype and block) was performed under a randomized complete block design (RCBD) to evaluate the significance of differences among genotypes [69]. The significance of differences between means was determined at  $p < 0.05$  using the Tukey's test [70].

The phenotypic variance ( $\sigma^2_p$ ), genotypic variance ( $\sigma^2_g$ ), phenotypic coefficient of variation (PCV), and genotypic coefficient of variation (GCV) were computed as per Burton [71]. Broad-sense heritability ( $h^2_b$ ) was estimated using the formula described by [72]. Genetic advance (GA) was calculated by the formula suggested by Lush [73]. Genetic advance as a percent of the mean (GA%) was assessed as defined by Johnson et al. [74].

Phenotypic and genotypic correlations were estimated using the standard procedure suggested by Miller et al. [75] from the corresponding variance and covariance components. Significant correlation coefficients were tested at  $(n - 2)$  degrees of freedom on a 't' table from Fisher and Yates at 5% and 1% significance levels. Simple phenotypic and genotypic correlation coefficients were separated into direct and indirect effects using path analysis [21].

Before proceeding with path analysis, the data were examined for multicollinearity using the variance inflation factor (VIF) and tolerance statistics. This step is critical in path analysis to ensure the stability and accuracy of the estimated path coefficients. Small tolerance values (much lower than 0.1) or high VIF values ( $>10$ ) show high multicollinearity [76]. Following the confirmation of non-severe multicollinearity, phenotypic and genotypic path analyses were carried out using RAISINS (R and AI Solutions for INferential Statistics [77]) to study the direct and indirect contributions of traits to the associations. Stepwise multiple regression analysis was carried out as a confirmatory tool to validate the path analysis findings and to assess the combined contribution of the selected traits to grain yield [78].

To identify multivariate outliers among the 84 Amaranth accessions, Mahalanobis distance squared ( $D^2$ ) analysis was performed based on six agronomic traits evaluated under irrigated conditions [22]. Mahalanobis distance measures the multivariate distance of each genotype from the centroid of the trait space, accounting for correlations among traits. Under multivariate normality, the  $D^2$  values follow a chi-square ( $\chi^2$ ) distribution with degrees of freedom equal to the number of traits ( $df = 6$ ). Genotypes with  $D^2$  values exceeding the critical chi-square value at  $\alpha = 0.05$  ( $\chi^2 = 12.59$ ) were classified as multivariate outliers. The analysis was conducted using R studio software (V4.5.2).

To examine the agronomic significance of multivariate outlier status, a scatter plot was constructed illustrating the relationship between grain yield and Mahalanobis  $D^2$  distance. For each of the 84 Amaranth genotypes evaluated under irrigated conditions, the Mahalanobis  $D^2$  value (derived from six agronomic traits) was plotted on the abscissa ( $x$ -axis) against grain yield (g/plant) on the ordinate ( $y$ -axis). The scatter plot was generated using R software (version 4.2.5) with the ggplot2 package.

Principal component analysis (PCA) was conducted to reduce the dimensionality of the dataset and elucidate patterns of morphological variation among 84 Amaranth genotypes. The analysis was based on six quantitative traits. PCA was performed using the correlation matrix to standardize the variables and account for differences in measurement scales. The eigenvalues were calculated for each principal component (PC), and the proportion of total variance explained by each PC was determined. Components with eigenvalues greater than 1.0 were retained for interpretation based on the Kaiser criterion [79], and a scree plot was constructed to visualize the decreasing contribution of successive components to the total variance [80]. Following principal component analysis, index plots were constructed for the first two principal components (PC1 and PC2) to visualize the distribution of 84 Amaranth genotypes based on their multivariate phenotypic profiles. For each genotype, the PC1 and PC2 scores were extracted from the PCA conducted on six quantitative traits evaluated under irrigated conditions. To enable direct comparability between components and to

facilitate genotype selection, the PC scores were standardized to a 0–1 scale using the following normalization formula:

$$\text{Normalized score} = \frac{\text{Actual score} - \text{Minimum score}}{\text{Maximum score} - \text{Minimum score}}$$

This transformation rescales the original PC scores to the unit interval [0, 1], where values close to 1 indicate genotypes with high positive contributions to the respective principal component, and values near 0 represent genotypes with low or negative contributions [81]. All PCA computations and visualizations were performed using RAISINS (R and AI Solutions for Inferential Statistics).

To examine genetic relationships among the 84 Amaranth genotypes, haplotype networks were constructed using *psbA-trnH* chloroplast and ITS sequences. Sequences were aligned and exported in Nexus format. The median-joining network algorithm was employed to infer evolutionary relationships [29]. The network was generated using PopART version 1.7 with default parameters, including epsilon = 0 (no star contraction) [82]. The networks display unique haplotypes as circles proportional to their frequency, mutational steps as connecting lines with hash marks, and inferred intermediate haplotypes (median vectors) as small black nodes.

## 5. Conclusions

This study demonstrates that substantial genetic diversity exists among the 84 Amaranth genotypes, with grain yield showing the widest variation and high heritability estimates indicating a favorable response to selection. Path analysis identified early flowering, long inflorescences, and heavy seeds as the primary determinants of grain yield, providing breeders with a focused selection index for indirect improvement. Multivariate approaches enabled the identification of nine outlier genotypes with distinct phenotypic profiles, among which G39 emerged as the most promising breeding candidate combining exceptional yield with desirable architecture and large seeds, while principal component analysis further identified complementary promising donors for plant architecture and seed size characteristics. Molecular characterization using chloroplast and nuclear markers validated these findings, with outlier genotypes occupying unique haplotype positions and the nuclear ITS network providing clearer resolution due to biparental inheritance, confirming that phenotypically distinct genotypes like G82, G83, G13, G10, and G39 possess unique genetic combinations. The integration of phenotypic and molecular data provides breeders with well-characterized germplasm, validated selection criteria, and prioritized parental materials for Amaranth improvement programs. Further multi-location and multi-season evaluations are recommended to ensure the stability and adaptability of these promising germplasm accessions.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/plants15101493/s1>, Table S1. Genotypic means ( $\pm$  standard error) and ANOVA summary for six agronomic traits in 84 Amaranth (*Amaranthus* spp.) genotypes evaluated under irrigated conditions in a randomized complete block design with three replications ( $n = 252$  observations per trait); Table S2. Genotypic path coefficient analysis showing direct (bold diagonal) and indirect effects of five agronomic traits on grain yield in 84 Amaranth (*Amaranthus* spp.) genotypes under irrigated conditions; Table S3. Variance Inflation Factor (VIF) and Tolerance values for five agronomic traits in Amaranth (*Amaranthus* spp.) under irrigated conditions; Table S4. Selected genotypes based on PC1 and PC2 index scores from principal component analysis of 84 Amaranth genotypes evaluated under irrigated conditions; Table S5. Primers used to amplify DNA markers and the amplification protocol; Figure S1. Frequency distributions of six key agronomic traits in 84 Amaranth (*Amaranthus* spp.) genotypes evaluated under irrigated conditions; Figure S2. Genotypic

correlation matrix of six quantitative traits measured in 84 Amaranth (*Amaranthus* spp.) genotypes evaluated under irrigated conditions; Figure S3. Evaluation of multivariate distances; Figure S4. Scree plot showing the percentage of variance explained by each principal component (PC1 to PC6) from principal component analysis of six agronomic traits in 84 Amaranth genotypes evaluated under irrigated conditions. Refs. [83–85] are cited in Supplementary Materials.

**Author Contributions:** Conceptualization, A.K. and P.N.; Methodology, A.K.; Investigation, A.K.; Bioinformatics and Formal Analysis, A.K. and Y.J.; Data Interpretation, A.K.; Data Curation, A.K.; Writing—Original Draft Preparation, A.K.; Writing—Review and Editing, P.N. and Y.J.; Supervision, P.N.; Project Administration, P.N.; Funding Acquisition, P.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** Open Access funding enabled and organized by Projekt DEAL. This work was supported by funds from the project “AMaranth as source of Omega fatty acids and gluten free grains: Marker-assisted breeding and cultivar selection to obtain functional healthy oils and raw materials for food-processing (AMOR)” through the German Federal Ministry for Economic Affairs and Climate Action (BMWK) in the Collective Research Networking (CORNET) Framework (Funding No. IGF 267 EN).

**Data Availability Statement:** Sequences were uploaded to the NCBI database. All data supporting the findings of this study are available within the paper and within its Supplementary Materials published online. Any other data are available from the corresponding author A.K. upon request.

**Acknowledgments:** Joachim Daumann and Anna-Luise Kuppinger are acknowledged for technical support in setting up and maintaining the Amaranth field plot.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Sauer, J.D. The Grain Amaranths and Their Relatives: A Revised Taxonomic and Geographic Survey. *Ann. Mo. Bot. Gard.* **1967**, *54*, 103. [\[CrossRef\]](#)
2. Joshi, B.D.; Rana, R.S. *Grain Amaranths: The Future Food Crop*; National Bureau of Plant Genetic Resources, Regional Station, Phagli: Shimla, India, 1991.
3. 2: Vegetables. In *Plant Resources of Tropical Africa: PROTA*; Grubben, G.J.H., Denton, O.A., Eds.; PROTA Foundation [u.a.]: Wageningen, The Netherlands, 2004.
4. Das, S. Systematics and Taxonomic Delimitation of Vegetable, Grain and Weed Amaranths: A Morphological and Biochemical Approach. *Genet. Resour. Crop Evol.* **2012**, *59*, 289–303. [\[CrossRef\]](#)
5. Alemayehu, F.R.; Bendevis, M.A.; Jacobsen, S.-E. The Potential for Utilizing the Seed Crop Amaranth (*Amaranthus* Spp.) in East Africa as an Alternative Crop to Support Food Security and Climate Change Mitigation. *J. Agron. Crop Sci.* **2015**, *201*, 321–329. [\[CrossRef\]](#)
6. Joshi, D.C.; Sood, S.; Hosahatti, R.; Kant, L.; Pattanayak, A.; Kumar, A.; Yadav, D.; Stetter, M.G. From Zero to Hero: The Past, Present and Future of Grain Amaranth Breeding. *Theor. Appl. Genet.* **2018**, *131*, 1807–1823. [\[CrossRef\]](#)
7. Kanbar, A.; Beisel, J.; Gutierrez, M.T.; Graeff-Hönninger, S.; Nick, P. Peruvian Amaranth (*kiwicha*) Accumulates Higher Levels of the Unsaturated Linoleic Acid. *Int. J. Mol. Sci.* **2023**, *24*, 6215. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Brenner, D.M.; Baltensperger, D.D.; Kulakow, P.A.; Lehmann, J.W.; Myers, R.L.; Slabbert, M.M.; Sleugh, B.B. Genetic Resources and Breeding of *Amaranthus*. In *Plant Breeding Reviews*; Janick, J., Ed.; Wiley: Hoboken, NJ, USA, 2000; pp. 227–285.
9. Mabhaudhi, T.; Chimonyo, V.G.P.; Chibarabada, T.P.; Modi, A.T. Developing a Roadmap for Improving Neglected and Underutilized Crops: A Case Study of South Africa. *Front. Plant Sci.* **2017**, *8*, 2143. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Mukuwapasi, B.; Mavengahama, S.; Gerrano, A.S. Grain Amaranth: A Versatile Untapped Climate-Smart Crop for Enhancing Food and Nutritional Security. *Discov. Agric.* **2024**, *2*, 44. [\[CrossRef\]](#)
11. Talabi, A.O.; Vikram, P.; Thushar, S.; Rahman, H.; Ahmadzai, H.; Nhamo, N.; Shahid, M.; Singh, R.K. Orphan Crops: A Best Fit for Dietary Enrichment and Diversification in Highly Deteriorated Marginal Environments. *Front. Plant Sci.* **2022**, *13*, 839704. [\[CrossRef\]](#)
12. Kanbar, A.; Beisel, J.; Wetters, S.; Gutierrez, M.T.; Graeff-Hönninger, S.; Nick, P. A Rapid, Simple, and Reliable Assay to Authenticate Peruvian *kiwicha* (*A. Caudatus*) for Food Applications. *Eur. Food Res. Technol.* **2022**, *248*, 2779–2797. [\[CrossRef\]](#)
13. Yadav, R.; Rana, J.C.; Ranjan, J.K. Analysis of Variability Parameters for Morphological and Agronomic Traits in Grain Amaranth (*Amaranthus* spp.) Genotypes. *Bioscan* **2014**, *9*, 1661–1665.

14. Yeshitila, M.; Gedebo, A.; Olango, T.M.; Tesfaye, B. Morphological Characterization, Variability, and Diversity among Amaranth Genotypes from Ethiopia. *Genet. Resour. Crop Evol.* **2023**, *70*, 2607–2636. [[CrossRef](#)]
15. Yeshitila, M.; Gedebo, A.; Tesfaye, B.; Degu, H.D. Agro-Morphological Genetic Diversity Assessment of Amaranthus Genotypes from Ethiopia Based on Qualitative Traits. *CABI Agric. Biosci.* **2024**, *95*. [[CrossRef](#)]
16. Gaines, T.A.; Ward, S.M.; Bukun, B.; Preston, C.; Leach, J.E.; Westra, P. Interspecific Hybridization Transfers a Previously Unknown Glyphosate Resistance Mechanism in *Amaranthus* Species. *Evol. Appl.* **2012**, *5*, 29–38. [[CrossRef](#)]
17. Thapa, R.; Edwards, M.; Blair, M.W. Relationship of Cultivated Grain Amaranth Species and Wild Relative Accessions. *Genes* **2021**, *12*, 1849. [[CrossRef](#)]
18. Suresh, S.; Chung, J.-W.; Cho, G.-T.; Sung, J.-S.; Park, J.-H.; Gwag, J.-G.; Baek, H.-J. Analysis of Molecular Genetic Diversity and Population Structure in *Amaranthus* Germplasm Using SSR Markers. *Plant Biosyst.—Int. J. Deal. All Asp. Plant Biol.* **2014**, *148*, 635–644. [[CrossRef](#)]
19. Jamalluddin, N.; Massawe, F.J.; Mayes, S.; Ho, W.K.; Symonds, R.C. Genetic Diversity Analysis and Marker-Trait Associations in *Amaranthus* Species. *PLoS ONE* **2022**, *17*, e0267752. [[CrossRef](#)] [[PubMed](#)]
20. Shrivastav, S.P.; Yadav, C.B.; Lal, K.; Singh, K.; Singh, V. Genetic Variability, Divergence, Correlation and Path Coefficient Analysis for Yield and Its Contributing Traits in *Amaranthus* (*Amaranthus paniculatus* L.). *Trends Biosci.* **2017**, *10*, 3257–3264.
21. Dewey, D.R.; Lu, K.H. A Correlation and Path-Coefficient Analysis of Components of Crested Wheatgrass Seed Production. *Agron. J.* **1959**, *51*, 515–518. [[CrossRef](#)]
22. Mahalanobis, P.C. On the generalized distance in statistics. *Sankhyā Indian J. Stat. Ser. A (2008-)* **2018**, *80*, S1–S7.
23. Pearson, K. LIII. On Lines and Planes of Closest Fit to Systems of Points in Space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1901**, *2*, 559–572. [[CrossRef](#)]
24. Sefasi, A.; Nyasulu, M.; Kamanga, R.M.; Yalaukani, L.; Pilanazo Katengeza, S.; Monjerezi, M.; Malidadi, C.; Masamba, K. Phenotypic Diversity and Multivariate Analyses of Yield and Yield-Related Traits in Amaranth Accessions from Malawi. *BMC Plant Biol.* **2025**, *25*, 1145. [[CrossRef](#)]
25. Efroymson, M. Multiple Regression Analysis. In *Mathematical Methods for Digital Computers*; Wiley: New York, NY, USA, 1960.
26. Cherian, P.; Nithaniyal, S.; Srinivasan, P. Molecular Identification and Phylogenetics Analysis of Southern Indian *Amaranthus* L. Using Chloroplast Barcodes. *Ecol. Genet. Genom.* **2025**, *34*, 100329. [[CrossRef](#)]
27. Waselkov, K.E.; Boleda, A.S.; Olsen, K.M. A Phylogeny of the Genus *Amaranthus* (Amaranthaceae) Based on Several Low-Copy Nuclear Loci and Chloroplast Regions. *Syst. Bot.* **2018**, *43*, 439–458. [[CrossRef](#)]
28. Xu, H.; Xiang, N.; Du, W.; Zhang, J.; Zhang, Y. Genetic Variation and Structure of Complete Chloroplast Genome in Alien Monoecious and Dioecious Amaranthus Weeds. *Sci. Rep.* **2022**, *12*, 8255. [[CrossRef](#)]
29. Bandelt, H.J.; Forster, P.; Rohlf, A. Median-Joining Networks for Inferring Intraspecific Phylogenies. *Mol. Biol. Evol.* **1999**, *16*, 37–48. [[CrossRef](#)]
30. Kong, S.; Sánchez-Pacheco, S.J.; Murphy, R.W. On the Use of Median-joining Networks in Evolutionary Biology. *Cladistics* **2016**, *32*, 691–699. [[CrossRef](#)]
31. Abrar, M.; Ahmad, T.; Iqbal, S.; Ur Rehman, R.N.; Bokhari, S.A.M.; Ahmad, Z.; Artyszak, A.; Hashem, A.; Alkahtani, J.; Abd-Allah, E.F. Multivariate Analysis for Agronomic, Physiological, Macro, and Micronutrient Traits of Exotic Vegetable Amaranth Genotypes. *BMC Plant Biol.* **2024**, *24*, 1137. [[CrossRef](#)]
32. Singh, R. Genetic Variability, Diversity and Stability Analysis in Grain Amaranth (*Amaranthus hypochondriacus* L.). Doctoral Thesis, Allahabad School of Agriculture, Sam Higginbottom Institute of Agriculture, Technology and Sciences, Prayagraj, India, 2010.
33. Baturaygil, A.; Schmid, K. Characterization of Flowering Time in Genebank Accessions of Grain Amaranths and Their Wild Relatives Reveals Signatures of Domestication and Local Adaptation. *Agronomy* **2022**, *12*, 505. [[CrossRef](#)]
34. Venkatesh, L.; Niranjana Murthy, N.M.; Nehru, S.D.; Manjappa, M. Genetic Variability, Heritability and Genetic Advance in Grain Amaranth (*Amaranthus* Spp.). *Asian J. Bio Sci.* **2014**, *9*, 67–70.
35. Patial, M.; Chauhan, A.; Singh, K.P.; Sharma, D. Character Association and Path Coefficient Analysis in Grain Amaranth (*Amaranthus* Spp.). *Int. J. Agric. Environ. Biotechnol.* **2014**, *7*, 101. [[CrossRef](#)]
36. Malaghan, S.N.; Revanappa, S.; Ajjappalavar, P.; Nagaraja, M.; Raghavendra, S. Genetic Variability, Heritability and Genetic Advance in Grain Amaranth (*Amaranthus* Spp.). *Int. J. Curr. Microbiol. Appl. Sci.* **2018**, *7*, 1485–1494. [[CrossRef](#)]
37. Oduwaye, O.A.; Porbeni, J.B.O.; Adetiloye, I.S. Genetic Variability and Associations Between Grain Yield and Related Traits in *Amaranthus Cruentus* and *Amaranthus Hypochondriacus* Grown at Two Locations. *J. Hortic. Res.* **2016**, *24*, 91–99. [[CrossRef](#)]
38. Nyasulu, M.; Sefasi, A.; Chimzinga, S.; Maliro, M. Agromorphological Characterisation of Amaranth Accessions from Malawi. *AJPS* **2021**, *12*, 1528–1542. [[CrossRef](#)]
39. Yeshitila, M.; Gedebo, A.; Degu, H.D.; Olango, T.M.; Tesfaye, B. Study on Characters Associations and Path Coefficient Analysis for Quantitative Traits of Amaranth Genotypes from Ethiopia. *Sci. Rep.* **2023**, *13*, 20981. [[CrossRef](#)]
40. Henderson, T.L.; Johnson, B.L.; Schneiter, A.A. Grain Amaranth Seeding Dates in the Northern Great Plains. *Agron. J.* **1998**, *90*, 339–344. [[CrossRef](#)]

41. Artemyeva, E.P.; Valdyskikh, V.V.; Radchenko, T.A.; Belyaeva, P.A. Amaranthus Phenology during Its Introduction in the Middle Urals. *AIP Conf. Proc.* **2019**, *2063*, 030002. [[CrossRef](#)]
42. Dmitrieva, O.; Ivanov, S. Comparative Study of *Amaranth* Species (*Amaranthus* Spp.) in the Temperate Continental Climate of Russian Federation. *AAS* **2020**, *115*. [[CrossRef](#)]
43. Thakur, P.; Kumar, S.; Malik, J.A.; Berger, J.D.; Nayyar, H. Cold Stress Effects on Reproductive Development in Grain Crops: An Overview. *Environ. Exp. Bot.* **2010**, *67*, 429–443. [[CrossRef](#)]
44. Hedhly, A. Sensitivity of Flowering Plant Gametophytes to Temperature Fluctuations. *Environ. Exp. Bot.* **2011**, *74*, 9–16. [[CrossRef](#)]
45. Franks, S.J.; Sim, S.; Weis, A.E. Rapid Evolution of Flowering Time by an Annual Plant in Response to a Climate Fluctuation. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1278–1282. [[CrossRef](#)] [[PubMed](#)]
46. Venkatesh, L.; Niranjana Murthy, N.M.; Manjappa, M.; Nehru, S.D. Character Association and Path Co-Efficient Analysis for Various Traits in Grain Amaranth (*Amaranthus* Spp.). *Asian J. Bio Sci.* **2014**, *9*, 97–100.
47. Akther, C.; Hasan, M.; Raihan, M.; Hossain, M.; Mian, M. Genetic Divergence in Stem Amaranth (*Amaranthus tricolor* L.) Genotypes for Yield and Its Component Characters. *Agriculturists* **2013**, *11*, 82–88. [[CrossRef](#)]
48. Shukla, S.; Singh, S.P. Correlation and Path Analysis in Grain Amaranth (*Amaranthus* Spp.). *Indian J. Genet. Plant Breed.* **2003**, *63*, 163–164.
49. Alake, C.O.; Otusanya, G.O. Selection Criteria for Seed Yield Improvement in Multi-Species of *Amaranthus* Genotypes. *Int. J. Veg. Sci.* **2020**, *26*, 275–291. [[CrossRef](#)]
50. Yeshitila, M.; Gedebo, A.; Tesfaye, B.; Demissie, H.; Olango, T.M. Multivariate Analysis for Yield and Yield-Related Traits of Amaranth Genotypes from Ethiopia. *Heliyon* **2023**, *9*, e18207. [[CrossRef](#)] [[PubMed](#)]
51. Remane, A. 1: Theoretische Morphologie und Systematik. In *Die Grundlagen des Natürlichen Systems, der Vergleichenden Anatomie und der Phylogenetik*; Koeltz: Schmitten, Germany, 1971.
52. Mane, S.P.; Shashidhar, H.E.; Kanbar, A.; Hittalmani, S. Molecular Marker Analysis for Root Length in a Diverse Germplasm of Rice (*Oryza sativa* L.). *Indian. J. Genet. Plant Breed.* **2003**, *63*, 197–201.
53. Kanbar, A.; Shashidhar, H.E. Participatory Selection Assisted by DNA Markers for Enhanced Drought Resistance and Productivity in Rice (*Oryza sativa* L.). *Euphytica* **2011**, *178*, 137–150. [[CrossRef](#)]
54. Kanbar, A.; Shakeri, E.; Alhajturki, D.; Horn, T.; Emam, Y.; Tabatabaei, S.A.; Nick, P. Morphological and Molecular Characterization of Sweet, Grain and Forage Sorghum (*Sorghum bicolor* L.) Genotypes Grown under Temperate Climatic Conditions. *Plant Biosyst.—An. Int. J. Deal. All Asp. Plant Biol.* **2020**, *154*, 49–58. [[CrossRef](#)]
55. Perera, M.F.; Arias, M.E.; Costilla, D.; Luque, A.C.; García, M.B.; Romero, C.D.; Racedo, J.; Ostengo, S.; Filippone, M.P.; Cuenya, M.I.; et al. Genetic Diversity Assessment and Genotype Identification in Sugarcane Based on DNA Markers and Morphological Traits. *Euphytica* **2012**, *185*, 491–510. [[CrossRef](#)]
56. El-Esawi, M.A.; Elashtokhy, M.M.A.; Shamseldin, S.A.M.; El-Ballat, E.M.; Zayed, E.M.; Heikal, Y.M. Analysis of Genetic Diversity and Phylogenetic Relationships of Wheat (*Triticum aestivum* L.) Genotypes Using Phenological, Molecular and DNA Barcoding Markers. *Genes* **2022**, *14*, 34. [[CrossRef](#)]
57. Sharma, P.; Samkumar, A.; Rao, M.; Singh, V.V.; Prasad, L.; Mishra, D.C.; Bhattacharya, R.; Gupta, N.C. Genetic Diversity Studies Based on Morphological Variability, Pathogenicity and Molecular Phylogeny of the Sclerotinia Sclerotiorum Population From Indian Mustard (*Brassica juncea*). *Front. Microbiol.* **2018**, *9*, 1169. [[CrossRef](#)]
58. Yong, W.; Chunbo, X.; Laga, T.; Xiaoming, Z.; Mingjiu, W. Genetic evaluation and germplasm identification analysis on ITS2, *trnL-F*, and *psbA-trnH* of alfalfa varieties germplasm resources. *Open Life Sci.* **2023**, *18*, 20220582. [[CrossRef](#)]
59. Uncu, A.T.; Uncu, A.O. Plastid *trnH-psbA* Intergenic Spacer Serves as a PCR-Based Marker to Detect Common Grain Adulterants of Coffee (*Coffea arabica* L.). *Food Control.* **2018**, *91*, 32–39. [[CrossRef](#)]
60. Gao, T.; Ma, X.; Zhu, X. Use of the *psbA-trnH* Region to Authenticate Medicinal Species of Fabaceae. *Biol. Pharm. Bull.* **2013**, *36*, 1975–1979. [[CrossRef](#)]
61. Pang, X.; Liu, C.; Shi, L.; Liu, R.; Liang, D.; Li, H.; Cherny, S.S.; Chen, S. Utility of the *trnH-psbA* Intergenic Spacer Region and Its Combinations as Plant DNA Barcodes: A Meta-Analysis. *PLoS ONE* **2012**, *7*, e48833. [[CrossRef](#)] [[PubMed](#)]
62. Feng, S.; Jiao, K.; Zhu, Y.; Wang, H.; Jiang, M.; Wang, H. Molecular Identification of Species of *Physalis* (*Solanaceae*) Using a Candidate DNA Barcode: The Chloroplast *psbA-trnH* Intergenic Region. *Genome* **2018**, *61*, 15–20. [[CrossRef](#)] [[PubMed](#)]
63. Tripathi, A.M.; Tyagi, A.; Kumar, A.; Singh, A.; Singh, S.; Chaudhary, L.B.; Roy, S. The Internal Transcribed Spacer (ITS) Region and *trnH-psbA* Are Suitable Candidate Loci for DNA Barcoding of Tropical Tree Species of India. *PLoS ONE* **2013**, *8*, e57934. [[CrossRef](#)]
64. Lv, Y.-N.; Yang, C.-Y.; Shi, L.-C.; Zhang, Z.-L.; Xu, A.-S.; Zhang, L.-X.; Li, X.-L.; Li, H.-T. Identification of Medicinal Plants within the Apocynaceae Family Using ITS2 and *psbA-trnH* Barcodes. *Chin. J. Nat. Med.* **2020**, *18*, 594–605. [[CrossRef](#)]
65. Gerrano, A.S.; Van Rensburg, W.S.J.; Adebola, P.O. Genetic Diversity of *Amaranthus* Species in South Africa. *S. Afr. J. Plant Soil* **2015**, *32*, 39–46. [[CrossRef](#)]
66. Aichele, D.; Schwegleb, H.-W. Die Taxonomie Der Gattung Pulsatilla. *Feddes Reper.* **1957**, *60*, 1–230. [[CrossRef](#)]

67. Dinssa, F.F.; Hanson, P.; Ledesma, D.R.; Minja, R.; Mbwambo, O.; Tilya, M.S.; Stoilova, T. Yield of Vegetable Amaranth in Diverse Tanzanian Production Environments. *HortTechnology* **2019**, *29*, 516–527. [[CrossRef](#)]
68. Shapiro, S.S.; Wilk, M.B. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* **1965**, *52*, 591–611. [[CrossRef](#)]
69. Gomez, K.A.; Gomez, A.A. *Statistical Procedures for Agricultural Research*, 2nd ed.; Wiley: New York, NY, USA, 1984.
70. Tukey, J.W. Comparing Individual Means in the Analysis of Variance. *Biometrics* **1949**, *5*, 99. [[CrossRef](#)]
71. Burton, G.W.; DeVane, E.H. Estimating Heritability in Tall Fescue (*Festuca Arundinacea*) from Replicated Clonal Material. *Agron. J.* **1953**, *45*, 478–481. [[CrossRef](#)]
72. Hanson, C.H.; Robinson, H.F.; Comstock, R.E. Biometrical Studies of Yield in Segregating Populations of Korean Lespedeza. *Agron. J.* **1956**, *48*, 268–272. [[CrossRef](#)]
73. Lush, J.L. Heritability of Quantitative Characters in Farm Animals. *Hereditas* **1949**, *35*, 356–375. [[CrossRef](#)]
74. Johnson, H.W.; Robinson, H.F.; Comstock, R.E. Estimates of Genetic and Environmental Variability in Soybeans. *Agron. J.* **1955**, *47*, 314–318. [[CrossRef](#)]
75. Miller, P.A.; Williams, J.C.; Robinson, H.F.; Comstock, R.E. Estimates of Genotypic and Environmental Variances and Covariances in Upland Cotton and Their Implications in Selection. *Agron. J.* **1958**, *50*, 126–131. [[CrossRef](#)]
76. Hair, J.F.; Black, W.C.; Babin, B.J.; Anderson, R.E. *Multivariate Data Analysis*, 7th ed.; Pearson Prentice Hall: London, UK, 2010.
77. Hisham, M.M.; Chandran, J.; Gopinath, P.P. RAISINS: Integrating R and AI for Agricultural Data Analysis. *J. Sustain. Technol. Agric.* **2025**, *1*. [[CrossRef](#)]
78. Draper, N.R.; Smith, H. *Applied Regression Analysis*, 3rd ed.; Wiley series in probability and statistics. Texts and references section; Wiley: New York, NY, USA, 1998.
79. Kaiser, H.F. The Application of Electronic Computers to Factor Analysis. *Educ. Psychol. Meas.* **1960**, *20*, 141–151. [[CrossRef](#)]
80. Cattell, R.B. The Scree Test for the Number of Factors. *Multivar. Behav. Res.* **1966**, *1*, 245–276. [[CrossRef](#)] [[PubMed](#)]
81. Milligan, G.W.; Cooper, M.C. A Study of Standardization of Variables in Cluster Analysis. *J. Classif.* **1988**, *5*, 181–204. [[CrossRef](#)]
82. Leigh, J.W.; Bryant, D. POPART: Full-feature Software for Haplotype Network Construction. *Methods Ecol. Evol.* **2015**, *6*, 1110–1116. [[CrossRef](#)]
83. Sang, T.; Crawford, D.J.; Stuessy, T.F. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of Paeonia (*Paeoniaceae*). *Am. J. Bot.* **1997**, *84*, 1120–1136. [[CrossRef](#)] [[PubMed](#)]
84. Tate, J.A.; Simpson, B.B. Paraphyly of Tarasa (*Malvaceae*) and diverse origins of the polyploid species. *Syst. Bot.* **2003**, *28*, 723–737.
85. White, T.J.; Bruns, T.; Lee, S.; Taylor, J. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR Protocols: A Guide to Methods and Applications*; Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J., Eds.; Academic Press: San Diego, CA, USA, 1990; pp. 315–322.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.