Received: 11 Nov 2023 | Revised: 18 Jun 2024 | Accepted: 24 Jun 2024

Electronic Journal of Plant Breeding

Research Note

Correlation and principal component analysis of bixin content and yield-related traits in Annatto (*Bixa orellana* L.)

Keisham Bindyalaxmi*, K. Kumaran, S. Vennila and R. Saranya Kumari

Department of Forest Biology and Tree Improvement, Forest College and Research Institute, Tamil Nadu Agricultural University, Mettupalayam, Tamil Nadu, India.
***E-Mail:** keishambindyalaxmi@gmail.com

Abstract
Annatto (*Bixa orellana* L.) is an important dye-yielding species harbouring huge variations in qualitative as well as quantitative traits. Identification of factors influencing bixin content and seed yield is an important breeding aspect, considering the economic importance of annatto for selection of elite genotypes based on morphological traits. Thirty-one trees plus were selected for the study, and sixteen quantitative traits were recorded for two years (2021 and 2022) to understand their association with bixin content and seed yield. Annatto has bivalve pods; however, trivalve pods were also observed during the study. A positive and significant association (r = 0.65) between bixin content and seed yield was observed from the correlation analysis. The study revealed that bixin content was highly correlated and influenced directly by the number of primary and secondary branches and hundred seed weight, while tree height, basal diameter, and the number of primary and secondary branches exhibit a significantly positive direct effect on seed yield, suggesting these traits as a reliable morphological marker for selecting highly productive genotypes. The thirty-one plus trees were grouped through hierarchical clustering into six clusters, with maximum bixin content and seed yield in Cluster I and Cluster III. The genotypes TNBi 0020, TNBi 0021, TNBi 0022, TNBi 0023, TNBi 0024 and TNBi 0037 can be utilized for mass multiplication and popularization for commercial cultivation of *Bixa orellana*.

Keywords: *Bixa orellana*, bixin, yield, correlation, PCA

Bixa orellana L. belongs to the family Bixaceae and is an important dye-yielding species. It originates from the Neotropics (Rivera-Madrid *et al*., 2006). It is now well-adapted and naturalized in most tropical countries. The popularity of it increased in Southeastern Asian countries after its introduction by the Spaniards during the 17th century (Ulbricht *et al*., 2012). In India, it is commonly known as 'sinduri' and is commercially cultivated in Kerala, Karnataka, Tamil Nadu, Andhra Pradesh, Orissa, West Bengal, Gujarat, Maharashtra, Madhya Pradesh, and Chhattisgarh. It harbours huge variations for various qualitative and quantitative traits. Seeds of *Bixa orellana* yield a reddish-orange dye known as annatto; the principal pigment is bixin, which accounts for 80% of the total pigment (Hirko and Getu, 2022). The pods may be indehiscent or dehiscent; indehiscence leads to a significant increase in yield, while dehiscent pods lead to bixin photo-oxidation, lowering bixin content (Moreira *et al*., 2015; Valdez-Ojeda *et al*., 2008). Due to its nontoxic attributes, annatto is the second most demanded natural dye. The demand for annatto increased with the growing movement of society towards sustainable, green and environmental friendly products as an alternative or co-partner to synthetic dyes.

The compound annual growth rate of the annatto market is predicted to be 5.2% in the period of 2024-2032, and its value in 2023 is reported to be 215.49 million USD (Annatto Market Report, 2023). Major producers of annatto globally are Latin America, holding 61% of the world's total production, followed by Africa, contributing 27%, and Asia with 12% of the global production; while Brazil is the largest contributor to it with an annual production of 12,000 tons (Baptistão *et al*., 2023; Lopes de Melo *et al*., 2023; Kapoor and Ramamoorthy, 2021; Dequigiovanni *et al*., 2018). The demand for annatto in the global market is accelerating, with its highest demand in the food and cosmetic industries, with Asia and Europe being the fastest-growing markets for it (Cárdenas-Toro *et al*., 2024). For any commercial export of the seeds, the range of bixin content must be more than 2.5% (ITC, 1993). The commercial traits of bixin content and seed yield per plant are directly proportional to the price of annatto. These traits are determined by the genotype, the environment and their interaction. Coupled with the growing demand for natural dye and colourants, there arises a need to select annatto genotypes with maximum bixin content and seed yield. However, the seed yield of plants is a complex trait influenced directly or indirectly by numerous quantitative traits such as plant height, girth, number of primary and secondary branches, seeds per pod, number of pods per plant, etc. Hence, for the selection of genotypes for higher yield, seed yield should not be the only selection criterion; the interaction of other yield-influencing traits must be considered (Khan *et al*., 2022). Therefore, multivariate analysis like correlation analysis and path coefficient analysis are crucial for identifying and selecting the traits that directly or indirectly influence seed yield and such traits can be further utilized for the indirect selection of elite genotypes in any breeding program (Saroj *et al*., 2021; Navaneetha *et al.,* 2019) Hence, the present study investigated the quantitative traits influencing the productive traits of bixin content and seed yield and evaluated the strength of these factors on diversity.

Table 1. Area from which the Plus tree was collected and its location in Gene Bank and Dye Park

Number of Plus tree	Plus tree code	Collected from	Place of Location	Latitude	Longitude
1	TNBi 002	Coimbatore	Gene Bank	11^{0}326166'N	76.935640'E
2	TNBi 008	Nadukkuppam	Gene Bank	11^{0}326208'N	76.935563'E
3	TNBi 009	Sirumugai	Gene Bank	11^{0}326912'N	76.935944'E
4	TNBi 0020	Sirumugai	Gene Bank	11^{0}326877'N	76.936067'E
5	TNBi 0021	Sirumugai	Gene Bank	11^{0}326479'N	76.935058'E
6	TNBi 0022	Coimbatore	Gene Bank	11^{0}326363'N	76.935883'E
7	TNBi 0023	Coimbatore	Gene Bank	11^{0}326365'N	76.935909'E
8	TNBi 0024	Karaikudi	Gene Bank	11^{0}326328'N	76.935982'E
9	TNBi 0025	Coimbatore	Gene Bank	11^{0}326319'N	76.935982'E
10	TNBi 0026	Nedunkunam	Gene Bank	11^{0}326286'N	76.935938'E
11	TNBi 0027	Kallar	Gene Bank	11^{0}326242'N	76.935908'E
12	TNBi 0028	Kallar	Gene Bank	11^{0}326178'N	76.935898'E
13	TNBi 0029	Coimbatore	Gene Bank	11^{0}326178'N	76.935908'E
14	TNBi 0030	Coimbatore	Gene Bank	11^{0}326188'N	76.935898'E
15	TNBi 0031	Coimbatore	Gene Bank	11^{0}326161'N	76.935859'E
16	TNBi 0032	Nadukkuppam	Gene Bank	11^{0}326163'N	76.935854'E
17	TNBi 0033	Coimbatore	Gene Bank	11^{0}326264'N	76.935653'E
18	TNBi 0034	Melkondagalore	Gene Bank	11^{0}326258'N	76.935635'E
19	TNBi 0035	Coimbatore	Gene Bank	11^{0}326256'N	76.935669'E
20	TNBi 0036	Coimbatore	Gene Bank	11^{0}326207'N	76.935638'E
21	TNBi 0037	Coimbatore	Gene Bank	11^{0}326204'N	76.935637'E
22	TNBi 0038	Coimbatore	Gene Bank	11^{0}325462'N	76.935412'E
23	TNBi 0039	Coimbatore	Gene Bank	11^{0}326225'N	76.935554'E
24	TNBi 0040	Coimbatore	Gene Bank	11^{0}326221'N	76.935550'E
25	TNBi 0041	Coimbatore	Gene Bank	11^{0}326172'N	76.935872'E
26	TNBi 0042	Mettupalayam	Dye Park	11^{0}326989'N	76.939090'E
27	TNBi 0043	Mettupalayam	Dye Park	11^{0}326802'N	76.940004'E
28	TNBi 0044	Mettupalayam	Dye Park	11^{0}326814'N	76.939997'E
29	TNBi 0045	Mettupalayam	Dye Park	11^{0}327307'N	76.940704'E
30	TNBi 0046	Mettupalayam	Dye Park	11^{0}327288'N	76.940834'E
31	TNBi 0047	Mettupalayam	Dye Park	11^{0}326124'N	76.940506'E

TNBi: Tamil Nadu Bixa

The experiment was conducted in the Annatto fields, Department of Forest Biology and Tree Improvement, Forest College and Research Institute, Mettupalayam, Tamil Nadu, India. Thirty-one plus trees were selected from two populations, twenty-five trees from Gene Bank ($11^0$326'N latitude and $76^0$935'E longitude) while six plus trees were chosen from Dye Park ($11^0$327'N latitude and $76^0$940'E longitude) (**Table 1**). The trees were planted at 3m x 3m in 2017. Sixteen quantitative traits *viz*., tree height (H), basal diameter (BD), number of primary branches (NPB), number of secondary branches (NSB), pod length (PL), pod width (PW), pod thickness (PT), pod fresh weight (PFW), pod dried weight (PDW), number of seeds per pod (NSPP), seed length (SL), seed width (SW), seed thickness (ST), hundred seed weight (HSW), seed oil content (SOC) and norbixin content (NC), as well as bixin content (BC) and yield (Y), were taken for two years (2021 and 2022) to evaluate the factors influencing the bixin content and seed yield. Seed oil content was evaluated using the Soxhlet apparatus and n-hexane (Sharma *et al*., 2016). Norbixin and bixin content were estimated as given by Wallin (2006).

All the statistical analyses were performed using RStudio. The variance analysis, Pearson correlation analysis and path coefficient analysis were carried out using the 'Agricolae' package (Mendiburu, 2015). The principal component analysis was done using the packages 'Factoextra' (Kassambara and Mundt, 2016) and 'FactoMineR' (Le *et al*., 2008). The optimum number of clusters was identified using the 'NbClust package' (Charrad *et al*., 2014), while the 'Factoextra' (Kassambara and Mundt, 2016) and the 'ggplot2' (Wickham, 2009) packages were used to visualise the clusters. For hierarchical cluster analysis, Ward's minimal variance method (Williams, 1976) with squared Euclidean distance as a measure of dissimilarity was used (Ward, 1963).

Significant variations were observed in the selected sixteen quantitative traits with bixin content and seed yield, except for basal diameter and bixin content (**Table 2**). The maximum coefficient of variance was observed in pod length (66.10%), followed by yield (46.90%) and bixin content (40.13%). During the study, few trivalve pods of annatto were observed in the plus trees of TNBi 0035 and TNBi 0037 (**Fig. 1**); however, the maximum pods were bivalve, as reported in various studies of *Bixa orellana* (Pandey *et al*., 2019; Singh *et al*., 2016; Akshatha *et al*., 2011). The number of seeds per pod ranged from 19 to 50. Numerous

Table 2. Results of sixteen quantitative traits, bixin content and seed yield of thirty-one plus trees of *Bixa orellana* L.

Traits	p-value	CV %	SD	Mean	Max	Min
H	<0.001 ***	15.29	0.58	2.44	3.80	1.70
BD	0.33	18.20	4.51	24.38	29.20	18.03
NPB	0.004**	21.32	0.78	3.00	4.00	2.00
NSB	0.02 *	24.62	2.58	9.00	12.00	5.00
PL	<0.001 ***	66.10	0.58	3.57	5.41	2.66
PW	<0.001 ***	11.46	0.48	3.31	4.20	2.86
PT	<0.001 ***	6.49	0.21	2.32	2.89	2.01
PFW	<0.001 ***	12.06	1.38	6.80	9.77	4.99
PDW	<0.001 ***	22.78	0.81	2.46	4.23	1.33
NSPP	<0.001 ***	14.24	7.10	37.00	50.00	19.00
SL	<0.001 ***	7.46	0.65	4.14	5.55	3.27
SW	<0.001 ***	10.25	0.50	2.90	3.78	1.99
ST	<0.001 ***	8.78	0.51	3.29	4.00	2.37
HSW	<0.001 ***	12.87	1.45	7.82	10.23	5.56
SOC	<0.001 ***	31.01	0.97	1.32	3.47	0.27
NC	<0.001 ***	1.87	0.39	1.38	2.40	0.90
BC	0.54	40.13	0.86	1.64	2.37	0.94
Y	<0.001 ***	46.90	0.53	0.80	1.50	0.21

* Significant at 0.05 level, ** Significant at 0.01 level, *** Significant at 0.001 level

CV: Coefficient of variation; SD: Standard deviation; H: Tree height (m); BD: Basal diameter (cm); NPB: Number of primary branches; NSB: Number of secondary branches; PL: Pod length (cm); PW: Pod width (cm); PT: Pod thickness (cm); PFW: Pod fresh weight (g); PDW: Pod dried weight (g); NSPP: Number of seeds per pod; SL: Seed length (mm); SW: Seed width (mm); ST: Seed thickness (mm); HSW: Hundred seeds weight; SOC: Seed oil content (%); NC: Norbixin content (%); BC: Bixin content (%); Y: Seed yield (kg)

Fig. 1. Trivalve and bivalve pods of Annatto

studies have revealed that the maximum number of seeds per pod ranges from 30-60 (Pech-Hoil *et al*., 2024; Umadevi *et al*., 2020; Stringheta *et al*., 2018; Akshatha *et al*., 2011). The number of seeds in the trivalve pods ranges from 61 – 68. Norbixin content among the plus trees was recorded between 0.90 to 2.40%, which confirms the results by Raddatz-Mota *et al*. (2016), where the norbixin range was reported at 0.1 – 4.0%. Maximum bixin content of 2.37% and a minimum of 0.94% were observed. Various authors have reported bixin content in the range of 1 – 4% (Carvalho *et al*., 2024; Pech-Hoil *et al*., 2024; Cárdenas-Conejo *et al*., 2023; Stringheta *et al*., 2018). Seed yield among the plus trees was highly variable and the highest yield of 1.50kg and lowest of 0.21kg were measured. The average seed yield of a three-year plant is reported to be in the range of 0.5 to 1kg or a maximum of 4kg (Math *et al*., 2016; Kumaran, 2014).

Pearson correlation coefficient analysis was carried out to understand the direction and strength of the association between variables. In tree improvement, correlated traits are significant as improving one of the correlated traits leads to simultaneous change in the other (Mehta and Shirin, 2020). From the correlation matrix (**Table 3**), it was observed that bixin content has a strong and positive correlation with tree height (r = 0.46), number of primary branches (r = 0.65), number of secondary branches (r = 0.48) and hundred seed weight (r = 0.58). The hundred seed weight was found to have a positive influence on

Table 3. Pearson correlation coefficient of sixteen quantitative traits and yield, bixin content

Traits	H	BD	NPB	NSB	PL	PW	PT	PFW	PDW	NSPP	SL	SW	ST	HSW	SOC	NC	BC	Y
H	1																	
BD	0.54**	1																
NPB	0.06	0.04	1															
NSB	-0.12	0.17	0.72***	1														
PL	0.61***	0.26	-0.34	-0.28	1													
PW	-0.11	-0.14	-0.47	-0.27	0.33	1												
PT	0.21	0.18	-0.08	-0.09	0.08	0.44*	1											
PFW	0.30	0.38*	-0.10	-0.12	0.35	0.20	0.45*	1										
PDW	0.10	0.33	0.08	0.07	0.09	0.12	0.30	0.67***	1									
NSPP	0.33	0.17	0.12	0.26	0.37*	0.20	0.38*	0.22	0.21	1								
SL	0.01	0.28	-0.20	0.04	0.12	0.16	0.17	0.16	0.12	0.08	1							
SW	0.03	0.12	-0.23	0.10	0.28	0.45	0.23	0.29	0.20	0.51**	0.66***	1						
ST	-0.17	-0.10	0.12	0.15	-0.40	-0.23	0.11	-0.05	0.08	-0.09	0.33	0.22	1					
HSW	0.26	0.23	0.37*	0.37*	0.08	0.27	0.38*	0.19	0.17	0.29	0.00	-0.01	-0.26	1				
SOC	0.58***	0.20	0.05	-0.32	0.33	-0.14	-0.02	0.16	0.05	-0.16	-0.11	-0.34	-0.29	0.24	1			
NC	0.01	-0.23	-0.08	-0.16	0.26	-0.06	-0.13	0.06	-0.19	-0.197	-0.30	-0.35	-0.31	-0.19	0.22	1		
BC	0.46**	0.30	0.65***	0.48**	0.10	-0.28	0.20	0.02	0.11	0.33	-0.27	-0.31	-0.16	0.58**	0.33	-0.01	1	
Y	0.60***	0.71***	0.43*	0.40*	0.18	-0.18	0.23	0.24	0.30	0.35	0.07	0.02	0.05	0.34	0.23	-0.33	0.65***	1

* Significant at 0.05 level, ** Significant at 0.01 level, *** Significant at 0.001 level

H: Tree height (m); BD: Basal diameter (cm); NPB: Number of primary branches; NSB: Number of secondary branches; PL: Pod length (cm); PW: Pod width (cm); PT: Pod thickness (cm); PFW: Pod fresh weight (g); PDW: Pod dried weight (g); NSPP: Number of seeds per pod; SL: Seed length (mm); SW: Seed width (mm); ST: Seed thickness (mm); HSW: Hundred seeds weight; SOC: Seed oil content (%); NC: Norbixin content (%); BC: Bixin content (%); Y: Seed yield (kg) bixin content, which was also reported by Kala *et al*. (2017). For yield, it was found that basal diameter had the highest positive correlation of r = 0.71, followed by tree height (r = 0.60), number of primary branches (r = 0.43) and number of secondary branches (r = 0.40). A similar significant and positive correlation between tree height, girth, number of primary and number of secondary branches with seed yield of *Jatropha curcas* was reported by Kumar *et al*. (2015). Height, diameter and number of plant branches indicate good seed yield (Asati *et al*., 2021; Kumar *et al*., 2020; Rehman *et al*., 2020 ; Sumalini *et al.,* 2015). Joseph and Siril (2014) reported a highly significant correlation coefficient between the productive traits of bixin content and seed yield (r = 0.788). The present study also confirms the finding, as bixin content and seed yield were observed to have a significant and positive correlation (r = 0.65).

Path analysis was used to understand the association between various quantitative traits. Path analysis, developed in the 1920s, is an extension of the regression model, which aims at examining the casual patterns of a set of variables on dependent variables and is of great importance in crop science (Kozak and Kang, 2006). Hiremata *et al*. (2022) expressed that understanding the association of components is extremely significant for any selection program based on complex traits such as yield. The analysis found pod length to give the maximum direct effect on bixin content. The correlation coefficient was 0.33. The number of primary branches, number of secondary branches and hundred seed weight had positive direct effects of 0.28, 0.27 and 0.21, respectively (**Table 4**). Seed width and pod width had a direct negative effect on bixin content, with a correlation coefficient of -0.27 and -0.21, respectively. Tree height, basal diameter, number of primary branches, number of secondary branches, pod thickness and hundred seed weight had significant positive effects through other traits. A negative indirect effect on bixin content was found from pod length (**Table 4**). Similar results of the positive influence of pod thickness and negative correlation of seed diameter with bixin content were reported by Nolasco Chumpitaz *et al*. (2020).

As indicated in **Table 5**, the yield was directly influenced by tree height, basal diameter, number of primary branches, number of secondary branches and pod weight with correlation coefficients of 0.43, 0.44, 0.31, 0.46 and 0.59, respectively. Seed width, hundred seed weight, and norbixin content had a high negative direct effect on yield with correlation coefficients of -0.67, -0.48 and -0.32, respectively. Hundred seed weight, seed width, number of seeds per pod, pod fresh weight, pod dried weight, pod length and basal diameter had a high positive but indirect

Table 4. Path coefficient analysis of sixteen quantitative traits with bixin content

Traits	H	BD	NPB	NSB	PL	PW	PT	PFW	PDW	NSPP	SL	SW	ST	HSW	SOC	NC	Indirect effect	Direct effect	Total Correlation
H		0.03	0.02	-0.03	0.2	0.02	-0.07	0.01	0.04	0.07	0.00	-0.01	0.00	0.05	0.08	0.00	0.41	0.05	0.46
BD	0.03		0.01	0.05	0.09	0.03	-0.08	0.02	0.02	0.05	-0.03	-0.03	0.00	0.05	0.03	0.01	0.25	0.05	0.30
NPB	0.00	0.00		0.19	-0.11	0.1	0.02	0.01	0.02	-0.03	0.02	0.06	0.00	0.08	0.01	0.00	0.37	0.28	0.65
NSB	-0.01	0.01	0.2		-0.09	0.05	0.02	0.01	0.03	-0.03	0.00	-0.03	0.00	0.08	-0.04	0.01	0.21	0.27	0.48
PL	0.03	0.01	-0.1	-0.08		-0.07	-0.07	0.01	0.05	0.03	-0.01	-0.08	0.01	0.01	0.05	-0.02	-0.23	0.33	0.10
PW	-0.01	-0.01	-0.13	-0.07	0.11		-0.03	0.01	0.02	0.14	-0.02	-0.12	0.00	0.06	-0.02	0.00	-0.07	-0.21	-0.28
PT	0.01	0.01	-0.02	-0.02	0.03	-0.09		0.02	0.04	0.31	-0.02	-0.06	0.00	0.08	-0.01	0.01	0.29	-0.09	0.20
PFW	0.01	0.02	-0.03	-0.03	0.12	-0.04	-0.21		0.03	0.14	-0.02	-0.08	0.00	0.04	0.02	0.00	-0.03	0.05	0.02
PDW	0.00	0.02	0.02	0.02	0.03	-0.03	-0.14	0.08		0.09	-0.01	-0.05	0.00	0.03	0.01	0.01	0.08	0.03	0.11
NSPP	0.02	0.01	0.03	0.07	0.13	-0.04	-0.05	0.02	0.12		-0.01	-0.14	0.00	0.06	-0.02	0.01	0.21	0.12	0.33
SL	0.00	0.01	-0.06	0.01	0.04	-0.03	-0.03	0.01	0.01	0.05		-0.18	0.00	0.00	-0.01	0.02	-0.16	-0.11	-0.27
SW	0.00	0.01	-0.07	0.03	0.09	-0.09	-0.06	0.02	0.06	0.07	-0.07		0.00	0.00	-0.05	0.02	-0.04	-0.27	-0.31
ST	-0.01	-0.01	0.04	0.04	-0.13	0.05	0.01	0.01	-0.01	0.04	-0.04	-0.06		-0.06	-0.04	0.02	-0.15	-0.01	-0.16
HSW	0.01	0.01	0.11	0.1	0.02	-0.05	-0.04	0.01	0.03	0.12	0.00	0.00	0.01		0.03	0.01	0.37	0.21	0.58
SOC	0.03	0.01	0.01	-0.09	0.11	0.03	-0.03	0.00	-0.02	-0.01	0.01	0.09	0.01	0.05		-0.01	0.19	0.14	0.33
NC	0.00	-0.01	-0.02	-0.04	0.09	0.01	-0.01	-0.02	-0.02	-0.04	0.03	0.09	0.00	-0.04	0.03		0.05	-0.06	-0.01

Residual Effect = 0.17

H: Tree height (m); BD: Basal diameter (cm); NPB: Number of primary branches; NSB: Number of secondary branches; PL: Pod length (cm); PW: Pod width (cm); PT: Pod thickness (cm); PFW: Pod fresh weight (g); PDW: Pod dried weight (g); NSPP: Number of seeds per pod; SL: Seed length (mm); SW: Seed width (mm); ST: Seed thickness (mm); HSW: Hundred seeds weight; SOC: Seed oil content (%); NC: Norbixin content (%)

Table 5. Path coefficient analysis of sixteen quantitative traits with yield

Traits	H	BD	NPB	NSB	PL	PW	PT	PFW	PDW	NSPP	SL	SW	ST	HSW	SOC	NC	Indirect effect	Direct effect	Total Correlation
H		0.24	0.02	-0.06	-0.04	-0.07	0.06	-0.01	0.1	-0.01	0	-0.02	-0.02	-0.12	0.1	0	0.17	0.43	0.60
BD	0.23		0.01	0.08	-0.02	-0.08	0.08	-0.03	0.05	-0.01	0.06	-0.08	-0.01	-0.11	0.03	0.07	0.27	0.44	0.71
NPB	0.03	0.02		0.33	0.02	-0.28	-0.02	-0.01	0.04	0	-0.04	0.15	0.02	-0.18	0.01	0.03	0.12	0.31	0.43
NSB	-0.05	0.08	0.22		0.02	-0.16	-0.02	-0.01	0.08	0.01	0.01	-0.07	0.02	-0.18	-0.06	0.05	-0.06	0.46	0.40
PL	0.26	0.12	-0.1	-0.13		0.2	0.07	-0.01	0.12	-0.01	0.02	-0.19	-0.05	-0.03	0.06	-0.08	0.25	-0.07	0.18
PW	-0.05	-0.06	-0.14	-0.12	-0.02		0.04	-0.02	0.06	-0.03	0.03	-0.3	-0.03	-0.13	-0.02	0.02	-0.77	0.59	-0.18
PT	0.09	0.08	-0.02	-0.04	-0.01	0.26		-0.02	0.12	-0.07	0.03	-0.15	0.01	-0.18	0	0.04	0.14	0.09	0.23
PFW	0.13	0.17	-0.03	-0.06	-0.02	0.12	0.2		0.08	-0.03	0.03	-0.19	-0.01	-0.09	0.02	-0.02	0.30	-0.06	0.24
PDW	0.04	0.15	0.02	0.03	-0.01	0.07	0.14	-0.08		-0.02	0.02	-0.13	0.01	-0.08	0.01	0.06	0.23	0.07	0.30
NSPP	0.14	0.08	0.04	0.12	-0.03	0.12	0.04	-0.02	0.32		0.01	-0.33	-0.01	-0.14	-0.03	0.06	0.37	-0.02	0.35
SL	0	0.12	-0.06	0.02	-0.01	0.1	0.03	-0.01	0.03	-0.01		-0.44	0.04	0	-0.02	0.1	-0.11	0.18	0.07
SW	0.01	0.05	-0.07	0.05	-0.02	0.27	0.06	-0.02	0.16	-0.01	0.12		0.03	0.01	-0.06	0.11	0.69	-0.67	0.02
ST	-0.07	-0.04	0.04	0.07	0.03	-0.14	-0.01	-0.01	-0.03	-0.01	0.06	-0.15		0.13	-0.05	0.1	-0.08	0.13	0.05
HSW	0.11	0.1	0.11	0.17	-0.01	0.16	0.04	-0.01	0.09	-0.02	0	0.01	-0.03		0.04	0.06	0.82	-0.48	0.34
SOC	0.25	0.09	0.02	-0.15	-0.02	-0.08	0.03	0	-0.06	0	-0.02	0.23	-0.04	-0.12		-0.07	0.06	0.17	0.23
NC	0	-0.1	-0.02	-0.07	-0.02	-0.03	0.01	0.02	-0.06	0.01	-0.06	0.23	-0.05	0.09	0.04		-0.01	-0.32	-0.33

Residual Effect = 0.12

H: Tree height (m); BD: Basal diameter (cm); NPB: Number of primary branches; NSB: Number of secondary branches; PL: Pod length (cm); PW: Pod width (cm); PT: Pod thickness (cm); PFW: Pod fresh weight (g); PDW: Pod dried weight (g); NSPP: Number of seeds per pod; SL: Seed length (mm); SW: Seed width (mm); ST: Seed thickness (mm); HSW: Hundred seeds weight; SOC: Seed oil content (%); NC: Norbixin content (%)

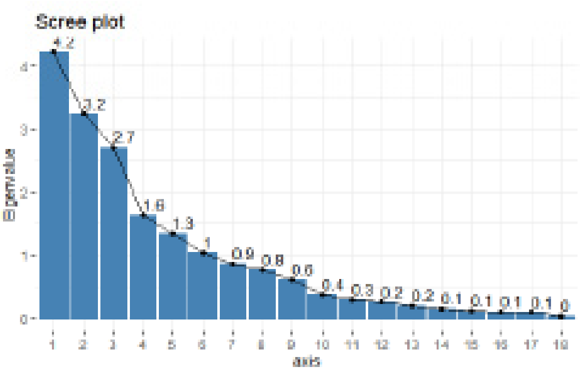

Fig. 2. Eigenvalues of the sixteen quantitative traits with bixin content and seed yield

effect on yield through other traits. While pod width had an indirect negative effect on yield through other traits.

Principal component analysis is an important technique that helps in determining the relationship among the variables and is widely used in agriculture. The PCA analysis revealed that the first seven components out of the eighteen contributed 83.67% of the total variability. The Eigenvalues, variability percentage and cumulative percentage contributions of the seven components are depicted in **Table 6 and Fig. 2**. The direction and contribution of each quantitative trait to the principal components are shown in **Fig. 3. From Fig. 3**, it can be seen that bixin content, yield, hundred seed weight, seed oil content, number of primary branches and number of secondary branches are closely related to each other and bixin content coupled with yield gave a high contribution to variability. The variations in the individuals and their relationship with the quantitative traits are represented in **Fig. 4. From Table 7**, height, basal diameter, pod length, seed oil content and yield were found to load strongly to the first principal component. The second principal component (FC2) was strongly loaded by the number of primary branches, number of secondary branches and bixin content.

Based on the cluster analysis of sixteen quantitative traits with bixin content and seed yield, the thirty-one plus trees were grouped into six clusters (**Fig. 5**). Clusters IV (TNBi 0034, TNBi 0040, TNBi 0041, TNBi 0036, TNBi 0039, TNBi 0033, TNBi 0038) and VI (TNBi 002, TNBi 0026, TNBi 0031, TNBi 009, TNBi 0027, TNBi 0025, TNBi 0028) had the maximum number of plus trees with seven members each. Variation in different quantitative traits among the six clusters and its deviation from the

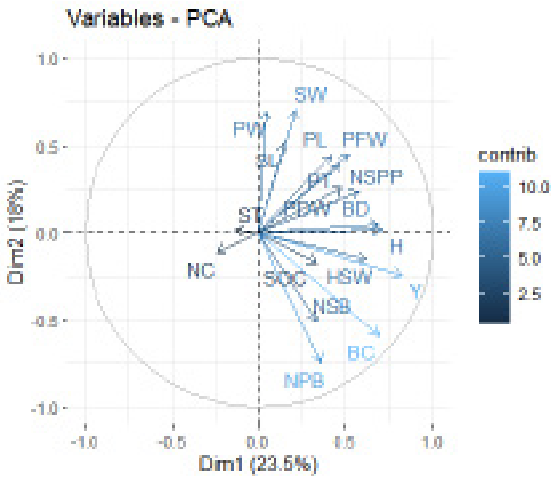

Fig. 3. Contribution of all the sixteen quantitative traits with bixin content and seed yield on the first two components

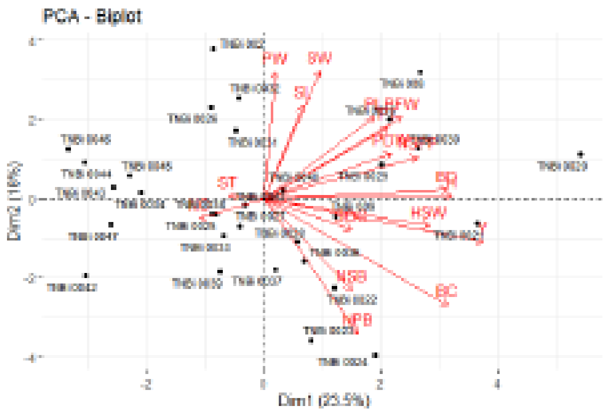

Fig. 4. Variations and directions of the individual plus trees with the variables (quantitative traits)

Table 6. Eigenvalues, total variance percentages and total cumulative percentages of all the sixteen quantitative traits with bixin content and seed yield

Components	**FC1**	**FC2**	**FC3**	**FC4**	**FC5**	**FC6**	**FC7**	**FC8**	**FC9**
Eigenvalue	4.22	3.24	2.71	1.63	1.34	1.04	0.86	0.77	0.60
Variance %	23.47	17.99	15.06	9.08	7.46	5.81	4.80	4.29	3.38
Cumulative variance %	23.47	41.46	56.52	65.60	73.06	78.87	83.67	87.96	91.34
Components	**FC10**	**FC11**	**FC12**	**FC13**	**FC14**	**FC15**	**FC16**	**FC17**	**FC18**
Eigenvalue	0.37	0.29	0.24	0.19	0.14	0.11	0.10	0.09	0.03
Variance %	2.06	1.59	1.37	1.08	0.79	0.61	0.53	0.47	0.16
Cumulative variance %	93.40	94.99	96.36	97.44	98.23	98.84	99.37	99.84	100.00

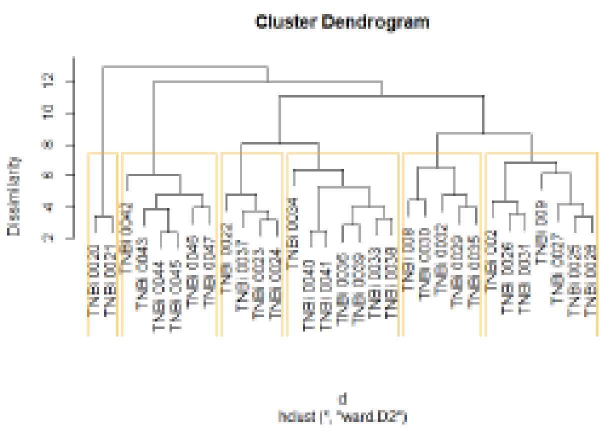

Fig. 5. Hierarchical cluster analysis of sixteen quantitative traits with bixin content and seed yield for thirty-one plus trees

total mean performance of thirty-one plus trees was observed (**Table 8**). A maximum average height of 2.93 m was observed in members of Cluster I (TNBi 0020, TNBi 0021). It can also be seen that members of it had the maximum average pod length (3.82cm), pod-dried weight (3.07g) and number of seeds per pod (40). The highest seed oil and bixin content were also recorded in this cluster, corresponding to 2.17% and 2.27%, respectively. The maximum average percentage of norbixin content of 1.82% was recorded in Cluster II containing six plus trees (TNBi 0042, TNBi 0043, TNBi 0044, TNBi 0045, TNBi 0046, TNBi 0047). Four members of Cluster III (TNBi 0022, TNBi 0037, TNBi 0023, TNBi 0024) accounted for the highest yield of 1.34 kg, with the maximum number of primary and secondary branches (4 and 11). Cluster V consisting of five members (TNBi 008, TNBi 0030, TNBi 0032, TNBi 0029, TNBi 0035) recorded maximum average basal diameter (25.99cm), pod width (3.74cm), pod thickness (2.58cm), pod fresh weight (7.91g), seed length (4.80mm), seed width (3.45mm) and hundred

Table 7. Correlation between the sixteen quantitative traits, bixin content and seed yield with the first seven principal components (Position: after 5th para of Results and Discussion)

Traits	FC1	FC2	FC3	FC4	FC5	FC6	FC7
H	0.91	0.04	-0.01	0.07	0.07	0.27	0.02
BD	0.59	0.22	0.45	-0.14	0.39	0.00	0.13
NPB	0.03	0.89	-0.14	-0.03	0.00	-0.09	-0.13
NSB	-0.22	0.88	0.18	-0.14	0.01	0.16	0.03
PL	0.52	-0.32	-0.07	-0.06	0.13	0.58	0.37
PW	-0.24	-0.41	0.18	0.57	0.09	0.28	0.43
PT	0.12	-0.04	0.06	0.81	0.31	0.20	-0.17
PFW	0.19	-0.12	0.00	0.16	0.87	0.18	0.05
PDW	0.02	0.12	0.11	0.13	0.88	0.00	-0.03
NSPP	0.11	0.28	0.06	0.27	0.08	0.82	0.02
SL	0.01	-0.20	0.78	-0.05	0.12	0.18	-0.10
SW	-0.19	-0.17	0.61	0.07	0.17	0.65	-0.02
ST	-0.17	0.05	0.29	0.03	0.02	-0.04	-0.86
HSW	0.18	0.49	0.12	0.59	0.07	-0.04	0.47
SOC	0.75	-0.11	-0.22	0.07	0.02	-0.30	0.19
NC	0.02	-0.21	-0.73	-0.26	0.06	0.09	0.19
BC	0.48	0.74	-0.22	0.24	-0.02	0.06	0.07
Y	0.62	0.55	0.27	0.07	0.22	0.11	-0.08
Variability %	23.47	17.99	15.06	9.08	7.46	5.81	4.80
Cumulative %	23.47	41.46	56.52	65.60	73.06	78.87	83.67

Table 8: Mean performance of the six clusters with the average performance of all the thirty-one plus trees

Traits	Cluster I	Cluster II	Cluster III	Cluster IV	Cluster V	Cluster VI	31 plus trees
H	2.93	2.10	2.48	2.13	2.47	2.62	2.44
BD	25.75	21.99	25.32	22.15	25.99	25.93	24.39
NPB	3.00	3.00	4.00	3.00	3.00	3.00	3.00
NSB	9.00	8.00	11.00	10.00	10.00	8.00	9.00
PL	3.82	3.56	2.90	3.40	3.72	3.56	3.57
PW	3.09	3.38	3.01	3.28	3.74	3.13	3.31
PT	2.39	2.20	2.35	2.27	2.58	2.25	2.32
PFW	6.94	6.24	5.77	6.62	7.91	6.82	6.80
PDW	3.07	2.08	2.33	2.38	2.93	2.55	2.46
NSPP	40.00	31.00	36.00	39.00	39.00	35.00	37.00
SL	3.58	3.74	3.61	3.95	4.80	4.66	4.14
SW	2.55	2.58	2.45	2.94	3.45	3.02	2.90
ST	2.97	3.08	3.38	3.25	3.52	3.56	3.29
HSW	8.35	6.53	8.34	8.37	8.68	7.18	7.81
SOC	2.17	1.32	1.05	0.88	1.01	1.57	1.32
NC	1.33	1.82	1.07	1.43	1.13	1.20	1.39
BC	2.27	1.34	2.15	1.68	1.65	1.38	1.64
Y	1.22	0.34	1.34	0.54	0.95	0.86	0.80

H: Tree height (m); BD: Basal diameter (cm); NPB: Number of primary branches; NSB: Number of secondary branches; PL: Pod length (cm); PW: Pod width (cm); PT: Pod thickness (cm); PFW: Pod fresh weight (g); PDW: Pod dried weight (g); NSPP: Number of seeds per pod; SL: Seed length (mm); SW: Seed width (mm); ST: Seed thickness (mm); HSW: Hundred seeds weight; SOC: Seed oil content (%); NC: Norbixin content (%); BC: Bixin content (%); Y: Seed yield (kg) seed weight (8.68g). Seven members of Cluster VI gave the highest average seed thickness of 3.56mm. For the improvement of bixin content and seed yield, the genotypes of Cluster I and Cluster III, TNBi 0020 TNBi 0021, TNBi 0022, TNBi 0037, TNBi 0023 and TNBi 0024 can be utilized as it is found to have the maximum bixin content, highest number of seeds per pod and greater seed yield.

From the study, it is evident that *Bixa orellana* harbours considerable variation in quantitative traits. These variations are essential for the selection of elite genotypes. It confirms the significant positive correlation between bixin content and seed yield, indicating any improvement in yield will also lead to an increase in the bixin content of seeds. The bixin content was found to be highly correlated and influenced directly by the number of primary and secondary branches, and hundred seed weight. The study also revealed that tree height, basal diameter, and the number of primary and secondary branches exhibit a significant and positive direct effect on seed yield in *Bixa orellana*, suggesting that these traits can serve as reliable identification markers for high-quality genotypes with enhanced productivity. And for the commercialization and popularization of *Bixa orellana* L., with higher bixin content and seed yield, TNBi 0020, TNBi 0021, TNBi 0022, TNBi 0023, TNBi 0024 and TNBi 0037 can be promoted.

REFERENCES

Akshatha, V., Giridhar, P. and Ravishankar, G. A. 2011. Morphological diversity in *Bixa orellana* L. and variations in annatto pigment yield. *The Journal of Horticultural Science and Biotechnology*, **86**(4): 319-324.

Annatto Market Report 2023. Polaris Market Research. New York City, United States. Available from: https://www.polarismarketresearch.com/industry-analysis/annatto-market

Asati, B.S., Chandrakar, M.K. and Sengar, S.S. 2021. Genetic parameters studies for yield attributes in Moringa genotypes (*Moringa oleifera* Lam.). *Journal of Pharmacognosy and Phytochemistry*, **10**(1):1664-1668.

Baptistão, M., Lorençatto, R., Eberlin, M.N. and Simionato, A.V.C. 2023. Multielemental Characterization of *Bixa orellana* L.(urucum) seeds by inductively coupled plasma mass spectrometry. *Food and Bioprocess Technology*, **16**(11): 2521-2530.

Cárdenas-Conejo, Y., Narváez-Zapata, J.A., Carballo-Uicab, V.M., Aguilar-Espinosa, M., Us-Camas, R., Escobar-Turriza, P., Comai, L. and Rivera-Madrid, R. 2023. Gene expression profile during seed development of *Bixa orellana* accessions varying in bixin pigment. *Frontiers in plant science*, **14:** 1-12.

Cárdenas-Toro, F.P., Meza-Coaquira, J.H., Nakama-Hokamura, G.K. and Zabot, G.L. 2024. Obtaining bixin-and tocotrienol-rich extracts from Peruvian Annatto seeds using supercritical CO_2 extraction: Experimental and economic evaluation. *Foods*, **13**(10): 1-18.

Carvalho, P.R.N., Silva, M.A.A.D., Oliveira, J.V.S.D., Fabri, E.G. and Silva, M.G.D. 2024. Color measurements in annatto (*Bixa orellana* L.) seeds. *Brazilian Journal of Food Technology*, **27**: 1-8.

Charrad M., Ghazzali, N., Boiteau, V. and Niknafs, A. 2014. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software*, **61**: 1-36. [Cross Ref]

Dequigiovanni, G., Ramos, S.L.F., Lopes, M.T.G., Clement, C.R., Rodrigues, D.P., Fabri, E.G., Zucchi, M.I. and Veasey, E.A. 2018. New microsatellite loci for annatto (*Bixa orellana*), a source of natural dyes from Brazilian Amazonia. *Crop Breeding and Applied Biotechnology*, **18**: 116-122.

Hiremata, V., Narayanaswamy, M. and Shet, R. M. 2022. Assessment of growth and yield parameters in Arecanut (*Areca catechu* L.) through correlation and path analysis under hilly zone of Karnataka. *Journal of Horticultural Sciences*, **17**(2): 333-340. [Cross Ref]

Hirko, B. and Getu, A. 2022. *Bixa orellana* (Annatto Bixa) A review on use, structure, extraction methods and analysis. *Journal of Agronomy, Technology and Engineering Management*, **5**(1): 687-696.

ITC 1993. Market Brief on Annatto Seeds. An Over-view of the World Market. International Trade Centre, UNCTAD/WTO, Geneva, Switzerland.

Joseph, N. and Siril, E. A. 2014. Evaluation and selection of elite annatto (Bixa orellana L.) and identification of RAPD markers associated with yield traits. *Brazilian Journal of Botany*, **37**: 1-8. [Cross Ref]

Kala, S., Kumaran, K., Srimathi, P., Reeja, S. and Singh, R. K. 2017. Studies on variability, correlation and path analysis using important seed traits in *Bixa orellana* (L). *Journal of Tree Sciences*, **36**(1): 93-102. [Cross Ref]

Kapoor, L. and Ramamoorthy, S. 2021. Strategies to meet the global demand for natural food colorant bixin: A multidisciplinary approach. *Journal of Biotechnology*, **338:** 40-51.

Kassambara, A. and Mundt, F. 2016. Package 'factoextra'. Extract and visualize the results of multivariate data analyses.

Khan, M.M.H., Rafii, M.Y., Ramlee, S.I., Jusoh, M. and Al Mamun, M. 2022. Path-coefficient and correlation analysis in Bambara groundnut (*Vigna subterranea* [L.] Verdc.) accessions over environments. *Scientific reports*, **12**(1): 1-12.

Kozak, M. and Kang, M. S. 2006. Note on modern path analysis in application to crop science. *Communications in Biometry and Crop Science*, **1**(1): 32-34.

Kumar, A., Ahad, I. and Shahnaz, E. 2020. Heritability estimates and correlation studies in homogamous walnut (*Juglans regia* L.) accessions from South Kashmir. *International Journal of Minor Fruits, Medicinal and Aromatic Plants*, **6**(2): 36- 44.

Kumar, S., Pandey, S. K., Ranjan, R., Dhyani, S. K., Kumar, R. V. and Ahlawat, S. P. 2015. Variability, heritability and character association studies for growth and seed yield in Jatropha Curcas crosses. *Range Management and Agroforestry*, **36**(2): 229-232.

Kumaran, K. 2014. Production potential of annatto (*Bixa orellana* L.) as a source of natural edible dye. In: Proceedings of *International workshop on natural dyes*, March 5-7, 2014, Hyderabad.

Le S., Josse, J. and Husson, F. 2008. FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software*, **25**: 1-18. [Cross Ref]

Lopes de Melo, E., Matias Pereira, A.C., Lopes do Nascimento, A. and Tavares Carvalho, J.C. 2023. *Euterpe oleraceae* (Açaí), *Bixa orellana* (Annatto), and *Myrciaria dubia* (Camu-camu): A Review of Preclinical Evidence of Anti-senescence Potential. *Pharmacognosy Reviews*, **17**(34): 406-417.

Math, R.G., Ramesh, G., Nagender, A. and Satyanarayana, A. 2016 Design and development of annatto (*Bixa orellana* L.) seed separator machine. *Journal of Food Science and Technology*, **53**(1): 703–711.

Mehta, N. and Shirin, F. 2020. Correlation and path coefficient analysis of some economic traits in Mahua (Madhuca indica JF gmel.). *Applied Biological Research*, **22**(3): 253-260. [Cross Ref]

Mendiburu, F. D. 2015. Agricolae, Statistical Procedures for Agricultural Research. R Package version 1.2-3.

Moreira, P. A., Lins, J., Dequigiovanni, G., Veasey, E. A. and Clement, C. R. 2015. The domestication of annatto (*Bixa orellana*) from *Bixa urucurana* in Amazonia. *Economic Botany*, **69**: 127-135. [Cross Ref]

Navaneetha, J.S., Murugan, E. and Parameswari, C. 2019. Correlation and path analysis for seed yield and its components in sesame (*Sesamum indicum* L.). *Electronic Journal of Plant Breeding*, **10**(3): 1262-1268.

Nolasco Chumpitaz, J. L., Ccoyllo Llacsa, P. A., Koc Sanchez, G. E. and Medina Morales, P. E. 2020. Collection and morphological characterization of 149 accessions of achiote (*Bixa orellana* L.) from seven departments in Perú. *Peruvian Journal of Agronomy,* **4**: 93–103. [Cross Ref]

Pandey, S., Sharma, A., Panika, G. and Kumar, M. 2019. Morphological studies, traditional and industrial uses of *Bixa orellana*. A review. *Current Science International*, **8**(1): 70-4.

Pech-Hoil, R., Ferrer, M.M., Aguilar-Espinosa, M., Simpson, J., Valdez-Ojeda, R., Guzmán-Antonio, A., Gutiérrez-Pacheco, L.C. and Rivera-Madrid, R. 2024. Characterization and variability of morpho-genetic traits of commercial importance in achiote collection. *New Forests*, **55**(3: 523-541.

Raddatz Mota, D., Pérez Flores, L. J., Carrari, F. O., Insani, E. M., Asis, R., Mendoza Espinoza, J. A., Diaz, L. S. F. and Rivera Cabrera, F. 2016. Chemical characterization and quantification of the pigment extraction yield of seven Mexican accessions of *Bixa orellana*. *Rev Mex Ing Química,* **15**: 727–740. [Cross Ref]

Rehman, A., Mustafa, N., Du, X. and Azhar, M.T. 2020. Heritability and correlation analysis of morphological and yield traits in genetically modified cotton. *Journal of cotton research*, **3:** 1-9.

Rivera-Madrid, R., Escobedo-Gm, R. M., Balam-Galera, E., Vera-Ku, M. and Harries, H. 2006. Preliminary studies toward genetic improvement of annatto (*Bixa orellana* L.). *Scientia Horticulturae*, **109**: 165-172. [Cross Ref]

Saroj, R., Soumya, S.L., Singh, S., Sankar, S.M., Chaudhary, R., Yashpal, Saini, N., Vasudev, S. and Yadava, D.K. 2021. Unraveling the relationship between seed yield and yield-related traits in a diversity panel of *Brassica juncea* using multi-traits mixed model. *Frontiers in Plant Science*, **12**: 1-16.

Sharma, S. S., Islam, M. A., Malik, A. A., Kumar, K., Negi, M. S. and Tripathi, S. B. 2016. Seed traits, fatty acid profile and genetic diversity assessment in *Pongamia pinnata* (L.) Pierre germplasm. *Physiology and Molecular Biology of Plants*, **22**: 193-205. [Cross Ref]

Singh, S., Sharma, S. K., Chaturvedi, V., Upadhyay, R. K., Chauhan, P. S. and Pal, M. 2016. Growth, yield and quality of *Bixa orellana* L. morphotypes. *Research and Reviews: Journal of Botanical Sciences*, **5**: 40-42.

Stringheta, P. C., Silva, P. I. and Costa, A. G. 2018. Annatto/ Urucum—*Bixa orellana*. *Exot Fruits*, **2018**: 23-30. [Cross Ref]

Sumalini, K., Pradeep, T. and Manjulatha, G. 2015. Heritability, correlation and path coefficient analysis of grain yield and yield components in maize (*Zea mays* L.). *Madras Agricultural Journal*, **102**: 123 – 126.

Ulbricht C., Windsor, R. C. A., Brigham, Bryan, J. K., Conquer, J., Costa, D., Giese, N., Guilford, J., Higdon, E. R. B., Holmes, K., Isaac, R., Jingst, S., Kats, J., Peery, L., Rusie, E., Savinainen, A., Schoen, T., Stock, T., Tanguay-Colucci, S. and Weissner, W. 2012. An evidence-based systematic review of annatto (*Bixa orellana* L.) by the natural standard research collaboration. *Journal of Dietary Supplements*, **9:** 57-77. [Cross Ref]

Umadevi, M., Giridharan, S. and Kumaran, K. 2020. Floral, reproductive biology and morphological variation in annatto (*Bixa orellana* L.). *Electronic journal of plant breeding*, **11**(2): 439-446. [Cross Ref]

Valdez-Ojeda, R., Hernández-Stefanoni, J. L., Aguilar-Espinosa, M., Rivera-Madrid, R., Ortiz, R. and Quiros, C. F. 2008. Assessing morphological and genetic variation in annatto (*Bixa orellana* L.) by sequence-related amplified polymorphism and cluster analysis. *HortScience*, **43**(7):2013-2017. [Cross Ref]

Wallin, M. H. 2006. Annatto Extracts. *Chemical and Technical Assessment,* **21**: 1-21.

Ward Jr, J. H. 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association,* **58**: 236-244. [Cross Ref]

Wickham, H. 2009. Getting started with qplot. *Ggplot2: elegant graphics for data analysis*, 9-26. [Cross Ref]

Williams, W.T. 1976. Pattern Analysis in Agricultural Science, Elsevier Scientific Publishing Company, Melbourne.