

Integrated Analysis of Genetic Diversity in *Costus pictus* through Phenotypic Characterization, Molecular Markers and UHPLC-MS-Based Diosgenin Profiling

P. Ragul

Department of Plantation, Spices, Medicinal and Aromatic crops, Horticultural college and Research Institute, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu

Chitra Rajagopal

Chitra.varadharaj@gmail.com

Department of Floriculture and Landscape, Horticultural college and Research Institute, Tamil Nadu Agricultural University, Coimbatore, Tamilnadu

D.Kesiar Loudursamy

Horticultural college and Research Institute, Tamil Nadu Agricultural University, Coimbatore, Tamilnadu

A. Thanga Hemavathy

Department of Pulses & Centre for Plant Breeding and Genetics, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu

S. Vellaikumar

Department of Biotechnology, Centre Plant Molecular Biology & Biotechnology, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu

Rajagopal B

Department of Centre Plant Molecular Biology & Biotechnology, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu

P Karthikeyan

Field Application Specialist, Spinco Pvt Ltd, Perungudi, Chennai

Research Article

Keywords: Genetic Diversity, PCoA, Diosgenin, AMOVA, UHPLC-MS, ISSR markers

Posted Date: May 13th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-6538032/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Genetic Resources and Crop Evolution on June 12th, 2025. See the published version at <https://doi.org/10.1007/s10722-025-02494-w>.

Abstract

Insulin plant (*Costus Pictus* D. Don.), one of the priceless medicinal plants, has the ability to lower blood sugar levels. Despite the pharmaceutical industry's constant demand, this species is not being used as much at the molecular level. Therefore, the present study aimed to assess the genetic diversity among 20 accessions of *C. pictus* collected from various geographical regions across South India using Inter-Simple Sequence Repeat (ISSR) molecular markers and Quantification of Diosgenin using Ultra High Performance Liquid Chromatography. A total of 12 ISSR primers were used in the present study. The utilization of principal component analysis and hierarchical clustering to examine morphological diversity facilitates the selection of progenitors for breeding schemes and simplifies genotype classification. Each of the first five components accounted for roughly 74–45% of the variance. Cluster analysis was used to separate the 20 genotypes into 5 clusters, which represented the genetic diversity of the group. Principal coordinate analysis further supported this grouping, showing that the first three coordinates accounted for 43.05 percent of the total variation. Acc. genotypes. KLCP3, TNCP20, TNCP13, and IISRNAGS9101 were chosen for the insulin plant breeding program, while APCP11 and KACP8 were chosen for their high yield. Also the diosgenin genotype quantification showed impressive variability (3.365–354.05 µg/g), with the greatest content being of TNCP9 (354.05 µg/g) and then TNCP10 (148.93 µg/g). The UHPLC-MS system was very accurate (calibration curve: $Y = 22576.5X + 22520.6$, $R^2 = 0.9988$) diosgenin content estimation for future pharmacological applications, particularly in relation to their antidiabetic activities for genotypes are likely to be valuable sources for the development of natural drugs and by analyzing ISSR markers, we identified the genetic relationships between accessions, which can guide targeted conservation efforts.

Introduction

Insulin plant (*Costus pictus*) wide across India throughout the world enormous antidiabetic activity of bioactive compounds for pharmaceutical industry. Cultivation of now days very rare wide range allelic diversity of genepool study of molecular work. *C. pictus* recently introduced in India from the American countries as an herbal care for diabetes. (Benny, 2004). The monocot family Costaceae has close to seven genera and containing about 143 known species (Christenhusz *et al.*, 2016) This plant is believed to be indigenous to the tropical parts of Asia, Central America, South America, and Africa. (Raju *et al.*, 2008) It has also become of great medicinal significance in the past few years because of its therapeutic application, exhibiting diverse pharmacological activities such as anti-diabetic, diuretic, antioxidant, and anticancer activity, and with potential bioactivity. (Hegde *et al.*, 2014). Diosgenin, a well-known steroidal sapogenin, is derived through the hydrolysis of the saponins, dioscin and can be extracted from various plant species, including *Dioscorea*, *Trigonella*, *Costus*, and *Smilax*. Traditionally used in medicine to treat a wide range of ailments, diosgenin holds considerable industrial significance due to its role as a precursor in the synthesis of steroidal drugs. (Selim *et al.*, 2015; Yan *et al.*, 2015) It is widely cultivated as an ornamental plant in South India. Various important phytochemical features have been reported from *C. pictus* plant like ascorbic acid, α -tocopherol, β -carotene, steroids, triterpenoids, alkaloids, tannins, saponins and flavonoids. (Urooj, 2010). Plant leaves have also been found to possess 21.2% fibers like

high levels of the elements K, Ca, Cr. Given its high therapeutic potential, the use of *C. pictus* in agroforestry practices would increase its accessibility to poor rural communities who rely on it, while, concurrently, also promoting efforts towards its genetic conservation. (Nag et al., 2015) Additionally, its genetic diversity will be essential towards improving its breeding and cultivation practices. Traditionally, morphological traits have been employed in the classification of different genera and species. These traits are often of restricted variability, environmentally influenced, and will usually entail cultivation of the plants to maturity in order to achieve meaningful identification. At present, a significant number of medicinally important plant species are under serious threat of extinction and genetic erosion, yet comprehensive data on them remain limited. (Kala, 2000) For many of these endangered species, conservation efforts are minimal, and only a small amount of genetic material is preserved in gene banks. Additionally, the growing focus on identifying novel drug compounds from plant sources has further accelerated the depletion of natural genetic resources. Determination of the genetic variability of these species is therefore essential to ensure that only quality accessions are utilized for propagation and conservation. Recent advancements in molecular marker technologies have been shown to be effective tools for analyzing and evaluating genetic diversity. In addition, tools enable the clarification of genetic relationships within and among species, hence aiding breeders in genetic improvement of valuable medicinal plants. (Ganie et al., 2015). The molecular approach offers a more effective means of identifying specific plant accessions or genotypes compared to traditional morphological markers, as it directly targets the plant's hereditary information and enables a clearer understanding of genetic relationships among individuals. The hypervariable nature of ISSRs combined with minimal equipment requirements and ease of use has made them extremely useful and cost-effective molecular markers for many ecological and systematic investigations. (Yang et al., 1996). The amplification and data-scoring protocols used for ISSR markers are similar to those used for random amplified polymorphic DNA (RAPD) markers with the exception that the annealing temperature for ISSR amplification is generally higher, resulting in a higher degree of stringency for amplified fragments (Wolfe *et al.*, 1998; Bhattacharyya *et al.*, 2015; Prajapat et al., 2015) The Inter Simple Sequences repeats DNA (ISSR) marker system has been widely employed to assess genetic variation at the molecular level in various medicinal and aromatic plant species. In plants it is unlikely that ISSR markers result from amplification of plastid DNA because the microsatellites found in this genome are predominantly mononucleotide repeats. Further the Ultra-High Performance Liquid Chromatography-Mass Spectrometry (UHPLC-MS) is a powerful analytical technique that combines the high-resolution separation of UHPLC with the sensitive detection of mass spectrometry. This method enables rapid, precise, and accurate analysis of complex mixtures in pharmaceuticals, metabolomics, environmental monitoring, and food safety. (Guillarme *et al.*, 2012) With superior speed, resolution, and sensitivity compared to traditional HPLC, UHPLC-MS is widely used for targeted and untargeted profiling, structural elucidation, and trace-level quantification. (Fekete et al., 2014) Its applications range from drug development and biomarker discovery to contaminant detection, making it indispensable in modern analytical laboratories. The present study focused on evaluating the effectiveness of ISSR markers in uncovering genetic diversity among *Costus pictus* (Insulin plant) accessions. Additionally, the study placed high emphasis on the detection of elite germplasm through genetic divergence analysis coupled with phenotypic and molecular information to

assist future breeding programs. Furthermore diosgenin content of the accessions was quantified using UHPLC-MS to evaluate their pharmacological potential.

Materials and method

The research included 20 Insulin Plant Accession from various regions, namely. Tamil Nadu, Kerala, Karnataka, Andhra Pradesh, and Puducherry, as indicated in Table 2. Utilizing ultra-high performance liquid chromatography, the primary goal was to assess the molecular and morphological diversity of these accessions in order to confirm high alkaloid diosgenin content levels and identify them with high yield. A period of evaluation and characterization was conducted from January 2024 to March 2025. 13 characters are used in three replications of a randomized block design.

Observation recorded

Observations encompassed thirteen traits, specifically Plant height (cm), Number of leaves per plant(NLPP), Number of tillers per plant(NTPP), leaf length (LL), leaf breadth (LB), Length of inflorescence (LI), Number of flower per inflorescence (NFPI), Days after flower initiation (DAFI), Stem girth (SG), Length of Rhizome (LR), Girth of rhizome (GR), Inter nodal length of Rhizome (INLR), Rhizome yield per plant (gm)(RYPP).

DNA Isolation and PCR Amplification of Markers

Using the CTAB (Cetyl Trimethyl Ammonium Bromide) method, DNA was extracted from immature leaves. The extracted DNA purity was assessed using a 0–8% Agarose gel, and the concentration of the DNA was standardized to 10 ng/μ using a NanoQuantND-1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). The accession's molecular diversity was determined using 20 ISSR primers. PCR was performed using reaction mixtures consisting of 1μL of DNA, 1μL of primer, 3μL of sterile water, and 5μL of master mix, for a total of 10μL. Applied Bio system, Waltham, M.A, USA), PCR amplification was performed with the following parameters: a four-minute initial denaturation step at 94°C, followed by a one-cycle touch-down phase in which the annealing temperature dropped by stage two is twenty cycles (94°C for 30 s, 50°C for 30 s, and 72°C for 1 min), followed by twenty cycles in stage three (94°C for 15 s, 45°C for 30 s, and 72°C for 1 min) with the following parameters: 72°C for 10 min and 4°C hold for one cycle.

Agarose gel- electrophoresis

Agarose was dissolved in 100 milliliters of freshly made 1X TAE buffer to create a 3 percent agarose gel for ISSR analyzing. A microwave was used to heat the agarose solution until it completely dissolved. The solution was allowed to cool to the proper temperature before being thoroughly mixed with 2 μL of ethidium bromide (EtBr). The gel solution was then transferred into a casting tray that had a comb

attached to it, which was positioned between 0 and 5 mm above the gel's surface. After the gel had set for an hour at room temperature, the comb was carefully taken out to create wells. 1X TAE buffer was added to the electrophoresis tank, and the gel that had solidified was then put inside. Three microliters of 6X gel loading dye were added to each sample before the amplified ISSR PCR products were loaded into the wells. For size reference, an additional 8 μ L of a DNA marker was loaded into a different well. Three hours and thirty minutes were spent conducting the electrophoresis at 50 V. Using a gel documentation system, the gel was observed under ultraviolet (UV) light after separation.

Sample Preparation of Costus Rhizome:

Freshly harvested tubers (50 g) were peeled off gently to remove the outer skin and then cut into small, equal pieces to achieve even extraction. The cut tubers were hydrolyzed with acid by refluxing with 3.5 M hydrochloric acid (115 mL) for three hours.(Pazhanichamy et al., 2012) This ensured the cleavage of glycosidic linkages, releasing sapogenins from their naturally occurring glycosides. The solution was filtered after hydrolysis to isolate the solid residue from the acidic filtrate. The solid residue was then washed extensively with distilled water until neutrality was attained to eliminate any remaining acid. After washing, the residue was dried in an oven at a controlled temperature of 65–70°C overnight to remove moisture. After being fully dried, the residue was then subjected to Soxhlet extraction using petroleum ether as the solvent. The extraction done for six hours to ensure efficient extraction of non-polar sapogenins into the organic phase.

The petroleum ether extract was then concentrated under reduced pressure using a rotary evaporator to drive off the solvent, resulting in the precipitation of a solid fraction. This crude sapogenins solid harvested by filtration and dried to give the crude sapogenin extract. The resulting crude sapogenin extract was now ready for purification and characterization by high performance liquid chromatography.

Standard Diosgenin Solution

A solution of 1 mg of Diosgenin in 10 ml of chloroform was used to create the Diosgenin standard (100 μ g mL⁻¹).

Determination of Diosgenin using LC-ESI-MS/MS-UHPLC Conditions

The Shimadzu LCMS-8045 tandem quadrupole mass analyzer with electro Spray ionization source was connected to the degassing unit (DGU-405), solvent delivery pump (LC-40D xs), autosampler (SIL-40C xs), column oven (CTO-40S), and flow control valve (FCV-20AH2) that were used for the UHPLC analysis. The chromatographic separation was performed using the Velox C18 column with dimension of 2.1 mm $\times$ 150mm with 1.8 μ m particle size) (Shimadzu, Japan) and the column temperature was maintained at 40°C. The separation of diosgenin was performed in isocratic mode of the mobile phase (A & B). Mobile

phase B is composed of 90% methanol at a steady flow rate of 0.2 mL/min, while mobile phase A is composed of 0.1–1% formic acid in water (10 percent). A 1µL injection volume was employed for every sample analysis..

ESI-Triple quadrupole mass analyser

The LCMS-8045 tandem quadrupole mass analyzer with ESI ionization source that Shimadzu manufactured was used for the experiment. After the precursor and product ions were carefully optimized using standard reference material, the analysis was carried out in positive mode multiple reactions monitoring (MRM) mode. Temperatures were kept at 300°C for the interface, 280°C for the desolvation, and 500°C for the heatblock. Heating gas, drying gas, and nebulizing gas were all kept at 3, 10, and 10 L/min, respectively. The streamlined MRM values for Diosgenin is presented in Table 1:

Table 1
MRM condition optimized for Diosgenin analysis.

Name	Precursor ion (m/z)	Product ion (m/z)	Q1 Pre bias (V)	CE (V)	Q1 Pre bias (V)
Diosgenin	415.40	271.30	-21.0	-19.0	-19.0
Diosgenin	415.20	253.20	-17.0	-25.0	-17.0

Statistical Analysis

In order to qualitatively score the DNA bands, gel photos from ISSR and RAPD analyses were used. Only distinct and easily reproducible bands from each accession were taken into account for scoring. Because each band was considered a distinct character, its presence or absence was noted in a binary data matrix, where "1" denoted presence and "0" denoted absence. Following the methodology of Sharma et al., the resolving power (Rp), average polymorphic information content (PIC), and polymorphic percentage (P percent) were calculated using the collected data in based on the approach of Sharma et al. (2017)(Powell et al., 1996) say that in order to determine the marker index (MI). the ratio of the multiplex (MR). Darwin versus 6.0.(Perrier & Jacquemoud Collet, 2006) Neighbor joining tree and genetic distances were estimated using this method. (Saitou *et al.*, 1987).

Table 2
List of 20 Insulin plant germplasm and their source of collection

Sl.No.	Cultivar	Description	Latitude	Longitude
1	Karaikal local	Local cultivar, collected from PAJANCOA & RI college, Karaikal	10 ⁰ 95'05"N	79 ⁰ 77'78"E
2	TNAU local	Local cultivar, collected from Periyakulam Tamil Nadu	10 ⁰ 12'56"N	77 ⁰ 56'17"E
3	KAU local	Local cultivar, collected from AMPRS, Thrissur, Kerala	10 ⁰ 54'75"N	76 ⁰ 28'21"E
4	IISRNAGS9101	Accession ICAR-IISR, Kozhikode, Kerala	11 ⁰ 29'80"N	75 ⁰ 84'07"E
5	Kozhikode local	Local cultivar, collected from Pokkunnu, Kozhikode	11 ⁰ 40'02"N	75 ⁰ 80'56"E
6	Malapuram local	Local cultivar, collected from Malapuram, Kerala	10 ⁰ 84'12"N	75 ⁰ 99'38"E
7	TNAU local	Local cultivar, collected from Killikulam, Tamil Nadu	8 ⁰ 70'11"N	77 ⁰ 86'27"E
8	IIHR local	Local cultivar, collected from IIHR, Hassan, karanataka	13 ⁰ 13'05"N	77 ⁰ 48'45"E
9	Erode local	Local cultivar, collected from JKKMCAS College, Erode	11 ⁰ 50'76"N	77 ⁰ 40'60"E
10	Coimbatore local	Local cultivar, collected from Saravampatti, Coimbatore	11 ⁰ 04'36"N	77 ⁰ 00'02"E
11	Nellore local	Local cultivar, collected from KVK, Nellore	14 ⁰ 40'38"N	79 ⁰ 96'43"E
12	Salem local	Local cultivar, collected from SMPG mettur dam, Salem	11 ⁰ 78'64"N	77 ⁰ 79'69"E
13	Tiruvarur Local	Local cultivar, collected from Andipalayam, Tiruvarur	10 ⁰ 76'35"N	79 ⁰ 68'42"E
14	Palakkad local	Local cultivar, collected from Walaiyar,Kerala	10 ⁰ 84'28"N	76 ⁰ 83'88"E
15	Bangalore local	Local cultivar, collected from GKVK Campus, Bangalore	13 ⁰ 08'08"N	77 ⁰ 57'79"E
16	Andhra local	Local cultivar, collected from YSHR Campus, Andhra Pradesh	16 ⁰ 88'30"N	81 ⁰ 45'16"E
17	Dharmapuri local	Local cultivar, collected from Dharmapuri, TamilNadu	12 ⁰ 03'33"N	78 ⁰ 06'49"E
18	Chengalpattu local	Local cultivar, collected from Chengalpattu	11 ⁰ 03'33"N	76 ⁰ 41'67"E

Sl.No.	Cultivar	Description	Latitude	Longitude
19	Cuddalore local	Local cultivar, collected from Virudhachalam	11 ⁰ 30'52"N	79 ⁰ 19'31"E
20	Kanyakumari local	Local cultivar, collected from Thovalai, Kanyakumari	8 ⁰ 13'47"N	77 ⁰ 30'02"E

Table 2 (Continued)

Version 6.0 of the Darwin software was used to analyze the binary data matrix derived from ISSR profiles. Using bootstrap analysis, the robustness of the generated clusters was evaluated. To help interpret the molecular variation among the accessions, the dendrogram offered a graphical depiction of genetic diversity and clustering patterns. Using the GenAlEx V6.5, an analysis of molecular variance (AMOVA), principal coordinate analysis (PCoA), number of observed alleles (Na), number of effective alleles (Ne), Nei's gene diversity (H), Shannon's information index (I), and unbiased Nei's gene diversity (uH) were assessed. (Gogoi et al., 2023) Genetic dissimilarities among the 20 *Costus pictus* accessions were computed based on the simple matching coefficient. The resulting dissimilarity matrix was utilized to create a dendrogram using the Unweighted Neighbor-Joining (UNJ) technique in order to illustrate the genetic connections between the accessions. Dot PCA was used to evaluate morphological diversity in R Grapes, and a biplot was used to visualize the relationships between PC1, PC2, PC3, PC4, and PC5.

Results

Study of ISSR Marker Data and Genetic Grouping Patterns

Twelve of the fifteen primers used for the genotype screening were polymorphic, while the remaining primers were monomorphic. (Supplementary Figs. 1,2 and 3). All outcomes—P%, PIC, Rp, and PI were reported in the (Supplementary Table 1). Every polymorphic primer yielded distinct bands with an allele number ranging from 1 to 6. There were 84 alleles found in total. Primer UBC 841 exhibits the most alleles (12). This study's key finding is that every allele found had a polymorphic rate. The information of a marker for identifying polymorphism is valued by PIC. More usefulness in differentiating accession to support the optimal breeding program for desired traits was indicated by a higher PIC value. The primers used in this investigation had average PIC values ranging from 0.11 to 0.90. The primer UBC841 had the highest average PIC value, while UBC 879 had the lowest. The range of MI values was 0.018 to 0.88, with the highest UBC being 841. Although the Rp value varied from 0 to 25 to 5 to 05, UBC 841 showed the highest Rp and UBC 879 the lowest. The calculated overall average PIC was 0.72. Furthermore, the MR was 0.65 and the MI was 0.81. Using the Jaccard pairwise distance matrix, five distinct cluster neighbor joining techniques were developed based on ISSR marker data gathered from 20 accession regions. Three entities made up Cluster I, four entities made up Cluster II, seven entities made up Cluster III, three entities made up Cluster IV, and three entities made up Cluster V (Fig. 1) PYCP1 and TNCP2 showed the

lowest genetic dissimilarity (0.15), while KACP8 and TNCP20 showed the highest genetic dissimilarity (0.85).

Principal Coordinate Analysis

PCoA detects individual and group differences, identifies outliers, and confirms genetic correlations between accession. With respective contributions of 16.31 percent, 14.59 percent, and 12.14 percent, the results showed that PC1, PC2, and PC3 together account for 43.05 percent of the total variation (Table 3).

Table 3			
Proportion of the total variation explained by the first three axes			
Axis	1	2	3
%	16.31	14.59	12.14
Cum %	16.31	30.91	43.05

The cluster APCP16, KLCP14, TNCP20,KACP15, KLCP18, TNCP17, TNCP10,TNCP13, TNCP19, KLCP6, TNCP2, KLCP3,TNCP9 and IISRNAGS9101while the remaining ones are significantly more distant. (Fig. 2).

Analysis of Molecular variance (AMOVA)

To discover the molecular variation in Insulin plant accession, Amova was utilized. According to Table 4 and Supplementary Fig. 4), the analysis revealed that only 11% of the variation was attributed to variation across populations, with the vast majority of variation (89%) occurring within populations.

Table 4					
Synopsis of the Analysis of Molecular Variance					
Source	Df	SS	MS	Est. Var.	%
Among Pops	1	31.392	31.392	1.760	11%
Within Pops	18	260.958	14.498	14.498	89%
Total	19	292.350		16.257	100%

Genetic Differentiation Within and Among Populations

Genetic Variability within and across populations reveals a notable distinction, particularly in Sub-population II, which presented the most significant values in terms of Observed Alleles ($N_a = 1.83$) and Polymorphic Percentage (86.90%). In population I exhibited high value Nei's gene diversity ($h = 0.28$), no.of. effective alleles ($n_e = 1.49$), Shannon's information index (0.42), and Nei's unbiased diversity ($u_h =$

0.30). (Table 4). While the subpopulation II recorded lowest Na ((1.75), Ne (1.42), I (0.40), h (0.26) and uh (0.27)). The polymorphic percentage was lowest in population I (83.3%).

Table 5
The number of observed alleles (Na), number of effective alleles (Ne), Nei's gene diversity (H), Shannon's information index (I), and unbiased Nei's gene diversity (uH) were evaluated for the insulin plant population

Population	Mean					Polymorphic %
	Na	Ne	I	H	uh	
p1	1.750	1.490	0.427	0.285	0.304	83.33%
p2	1.833	1.421	0.406	0.262	0.274	86.90
Total	1.792	1.455	0.417	0.273	0.289	
SE	0.041	0.026	0.018	0.013	0.014	

Diversity based on biometrical Observation

To assess diversity, Principal Component Analysis (PCA) and cluster analysis were employed. PCA is a reliable multivariate statistical technique that can be used to pinpoint the primary causes of variation in datasets. Out of five PCs, 13 principal components had eigenvalues greater than 1, according to PCA (Table 6). This analysis was conducted to investigate the diversity among 20 insulin plant accessions collected from different regions of India. The PCA demonstrated that 13 principal components (PCs) accounted for 100% of the total variability observed in the insulin plant samples. Among these PCs, PC1 contributed 27.86%, PC2 accounted for 16.6%, PC3 for 12.01%, PC4 for 10.0%, and PC5 for 8.15%, each possessing eigenvalues exceeding one, together explaining 4.81% of the overall variance in the total variability. The results suggest that key traits associated with the primary principal components may be used to select for and develop improved insulin plant accessions. In the PCA, 13 characteristics—PH, NLPP, NTPP, LL, LB, LI, NFPI, DAFI, SG, LR, GR, ILR, and RYPP exhibited uniformity across all analyzed genotypes. Additionally, the scree plot illustrates the percentage of explained variance for various dimensions or components. Traits RYPP, NTPP, and DAFI play a significant role in the divergence observed in insulin plants, as depicted in (Supplementary Fig. 5). The biplot of PC1 and PC2 shows a graphic representation of the variability in traits, represented by the genotypes placed at the positive and negative favorable traits (Supplementary Table 2.) Accession like IISRNAGS9101, KLCP3, KACP8, TNCP13, TNCP20 are placed away from the center whereas the remaining are placed near to the mid center (Fig. 3). Hierarchical clustering analysis segregated 20 accession into four clusters, Cluster I having 7 accession, cluster II having 7, and cluster III had 3 and cluster IV with 3 accession. Mean performance evaluated for traits based on morphological data. The knowledge of these influential factors aids in discriminating among samples or populations in the study and sheds light on the most significant characteristics for classification or further analysis.

Table 6
Eigenvalues, variance and cumulative variability of Insulin plant accessions

PCA	Eigenvalue	Variance Contribution (%)	Cumulative Proportion of Variance
PC1	3.622	27.863	27.863
PC2	2.16	16.618	44.481
PC3	1.574	12.108	56.588
PC4	1.31	10.076	66.664
PC5	1.06	8.155	74.819
PC6	0.862	6.633	81.452
PC7	0.688	5.293	86.745
PC8	0.599	4.605	91.351
PC9	0.505	3.884	95.235
PC10	0.275	2.116	97.351
PC11	0.184	1.415	98.765
PC12	0.104	0.803	99.568
PC13	0.056	0.432	100

Calibration and Sample analysis

Reference standard materials were dissolved in a 100% methanol solution further it was diluted to 1–100 ng/mL for the sample analysis and the obtained peak area was plotted against the concentration to obtain a calibration curve. The observed calibration was yielded good R^2 value of 0.9988 with the linear equation of $Y = (22576.5)X + (22520.6)$ as shown in a Fig. 4.

Discussion

Molecular genetics has developed high-throughput markers which identify polymorphism in microsatellite loci. These markers have provided further techniques and are effective in studying the genetic variation in plants. Genetic markers are vital in determining the genetic variation in plant populations. Owing to the extremes of environmental changes, ISSR markers are efficient as the traditional methods selecting germplasm ISSR markers works best with morphological marker using single primers anchored in di- or trinucleotide repeats, generating multilocus profiles without requiring prior sequence information highly polymorphic however the primer expression of phenotype features. Each plant genotype possesses a unique DNA sequence thus, the number and size of amplified bands will vary among individuals.(Ganie et al., 2015) These variations depend both on the genetic makeup of the plant and the specific primers used during molecular analysis. Only clearly dominant bands were

included in the analysis to minimize scoring errors and enhance reproducibility. The reassurance provided by ISSR markers concerning the genetic variability of insulin plant accession in this study is particularly notable with respect to amplification efficiency. In contrast to morphological markers, DNA markers provide a higher level of accuracy in assessing genetic purity, thereby enhancing the overall efficiency of breeding programs. The availability of molecular biology methods has made DNA-based markers as a necessary tool for numerous applications such as germplasm characterization, genetic diversity estimation, variety and hybrid identification, testing of clonal fidelity, genetic relationship and analysis, systematics, evolution and phylogeny, gene tagging, and marker-assisted selection. (Huang *et al.*, 2014). Molecular data enable conservationists to investigate critical issues such as genetic erosion and gene pool fragmentation. (Barrett *et al.*, 1991). In our present investigate the polymorphic information content of ISSR markers ranged from 0.11 to 0.90, with an average PIC value of 0.73 The marker showed that ISSR (82.32 percent) produced the higher level of polymorphism, whereas (Naik *et al.*, 2017) PIC values for primers UBC 818 and UBC 812 were 0.91 and 0.77. Based on Dissimilarity index showed high genetic diversity between genotypes under study, with very high dissimilarity particularly between KACP8 and TNCP20 (Jaccard index ≈ 0.85). Such high divergence suggests the genotypes have very different genetic architectures. Such high dissimilarity may suggest different evolutionary lineages or adaptive specializations and thus these genotypes are worth to be included in breeding programs to introduce new genetic traits. In contrast, low dissimilarity of PYCP1 and TNCP2 (Jaccard index ≈ 0.15) indicates extremely similar marker profiles. Similar study occurs This genetic closeness may either reflect common descent or recent common descent and reflects their substitutability in situations where genetic homogeneity is essential. The increased gene fixation brought on by allopolyploidy could be the cause of the higher genetic similarity scores (0.99) *C. Speciosus*. (Naik *et al.*, 2017; Kuttappety *et al.*, 2023) The ISSR similarity/dissimilarity index revealed that CP-8 and CP-13 were the two most closely related accessions (0.95). The analysis revealed distinct genetic patterns among genotypes, with PCoA clearly separating divergent genotypes from the homogeneous core population. While the first two coordinates captured major genetic structure, unexplained variance suggests additional dimensions may exist. Peripheral genotypes showed significant divergence, potentially harboring unique alleles of breeding value, while central genotypes represented the population's genetic backbone. This differentiation highlights their complementary roles - divergent genotypes as sources of novel traits and central genotypes for maintaining genetic stability. These findings demonstrate the utility of combining multivariate and clustering approaches for comprehensive germplasm characterization, providing a foundation for targeted breeding strategies and conservation efforts.

PCoA study that first three 43.05% Similar says that (Das *et al.*, 2017) the analysis of principal coordinates (PCoA) and cluster dendrogram revealed distinct genetic groupings of ginger accessions that correlated strongly with their geographical origins. The Analysis of Molecular Variance (AMOVA) showed that there was considerable genetic variation present among the populations of the insulin plant. The variation observed seems to be largely determined by gene transfer and genetic drift within the groups. Yet, a low degree of variation was observed among groups, indicating a low degree of

genetic differentiation. The scenario may be attributed to gene flow, historical migration, or common ancestral population among the populations.

Significant genetic variation existed between the insulin plant populations, according to the Analysis of Molecular Variance (AMOVA). The variation observed seems to be largely determined by gene transfer and genetic drift (89%) within the groups. Yet, a low degree of variation was observed among groups (11%) indicating a low degree of genetic differentiation. Similar study (Peakall *et al.*, 2006) The scenario may be attributed to gene flow, historical migration, or common ancestral population among the populations. (van Caspel *et al.*, 2021) It can be inferred that *C. pictus* has evidently migrated from one geographic locality to another, progressively expanding over great distances in range. As it did, the species was able to acclimate and show a correlation to the prevailing environmental conditions of a new habitat. (Li *et al.*, 2004) A species' ability to change and evolve more, however, depends to a great extent on the degree of genetic diversity its populations possess, reflecting the wealth of its germplasm for a given place or environment. For *C. pictus* the percent of polymorphism can serve as an informative indicator of genetic diversity. (Li *et al.*, 2004) By considering PCoA 43.05% and PCA 74.8% value highly genetic diversity and ISSR Molecular markers 82.32% polymorphism Genetically speaking, high genetic diversity increases the potential of a species to change and evolve against varying environmental conditions. (Marotti *et al.*, 2007) ISSR banding patterns enhance the reliability of the results by enabling the screening of a larger number of genomic regions. (Tar'an *et al.*, 2005) This multivariate approach was used to complement the information obtained from the cluster analysis methods because it is more informative regarding distances among major groups. Due to its valuable medicinal properties, *C. pictus* warrants greater attention in conservation efforts. Developing sustainable management strategies requires a clear understanding of the species' population structure. Although limited molecular data are currently available to support this analysis, morphological identification alone may not be sufficient to accurately distinguish between different accessions of the same species. (Hasan *et al.*, 2024)

Notably, genotypes with high diosgenin (TNCP9, TNCP10) clustered distantly from others in PCA/PCoA, implying unique different genotypes genetic markers may underlie their metabolic superiority. These accessions are prime candidates genotypes. The quantification of diosgenin in 20 *Costus pictus* accessions via UHPLC-MS revealed striking variability (3.365–354.05 µg/g), with TNCP9 exhibiting the highest content (354.05 µg/g), followed by TNCP10 (148.93 µg/g). (Supplementary Table .3) This 100-fold variation suggests a strong genetic influence on diosgenin biosynthesis, corroborated by ISSR-based clustering (e.g., TNCP9 and TNCP10 grouped separately from low-yielding genotypes). Such variability aligns with studies on other *Costus* species, where environmental and allelic differences drive secondary metabolite production. (Rawat *et al.*, 2021). The UHPLC-MS analysis demonstrated excellent sensitivity for diosgenin quantification, with a limit of detection (1 ng/g) and limit of quantification (5 ng/g), while maintaining a linear calibration range from 1-100 ng/mL ($R^2 = 0.9988$).

Diosgenin in *Costus pictus* extracts showed a consistent retention time of 6.43 min (Fig. 5) confirming method specificity. comparable to protocols used in *Dioscorea* and *Trigonella spp.* (Chaudhary *et al.*, 2018; GM, 2024). The range of diosgenin (3.365–354.05 µg/g) observed here is as seen in other *Costus*

species (Chaudhary et al., 2023) suggesting genotype-specific biosynthesis variability. In this study our method's linearity ($R^2 = 0.9988$) meets ICH guidelines (2005) and matches protocols used for *Dioscorea* spp. High-diosgenin genotypes (e.g., TNCP9) are promising for steroid drug precursors particularly antidiabetics. (Chen et al., 2022). Depending on the sample type, material processing, and instrument handling, both approaches may be equally accurate for the study, as the developed method was not claimed to be novel for the identification of diosgenin etc., (Srivastava et al., 2019). Pharmacological studies, given diosgenin's role as a precursor for steroidal drugs (e.g., antidiabetics, contraceptives). Conversely, low-diosgenin accessions (e.g., TNCP15: 3.365 $\mu\text{g/g}$) may prioritize morphological traits (e.g., biomass) for cultivation. This dual approach (molecular + biochemical) bridges the gap between genotype selection and pharmaceutical utility, advancing *C. pictus* as a sustainable source of bioactive compounds. Additionally, the goal of our research is to document the diosgenin variation in a natural population of *C. pictus* for identifying the superior chemotype to support commercial species cultivation and supply high-quality material to the herbal drug industry and molecular-level studies are essential to gain deeper insights into the genetic diversity and relationships among *C. pictus* accessions.

Conclusion

In summary, the application of high-throughput ISSR markers proved to be a powerful tool for assessing the genetic diversity within *Costus pictus* accessions. The high level of polymorphism (82.32%) and the wide range of PIC values (0.11 to 0.90) underscore the significant genetic variation present in the studied population. The dissimilarity indices revealed both highly divergent genotypes, such as KACP8 and TNCP20, suggesting their potential as sources of novel traits for breeding programs, and highly similar genotypes, like PYCP1 and TNCP2, which may be valuable for maintaining genetic homogeneity in specific applications. Multivariate analyses, including PCoA, further elucidated the genetic structure of the *C. pictus* accessions, separating divergent and homogeneous groups. The correlation of these genetic groupings with geographical origins, as suggested by similar studies, highlights the influence of environmental factors and evolutionary history on the genetic makeup of the species. AMOVA results indicated that the majority of genetic variation resides within populations, likely due to gene flow and genetic drift. Crucially, the integration of molecular data with biochemical analysis of diosgenin content revealed a strong link between genotype and metabolite production. The significant variation in diosgenin levels, with genotypes like TNCP9 exhibiting remarkably high concentrations, underscores the potential for selecting elite chemotypes for commercial cultivation and the herbal drug industry. The distinct clustering of high-diosgenin genotypes in PCA/PCoA further supports the genetic basis of this valuable trait. Therefore, this study demonstrates the efficacy of combining molecular markers and biochemical profiling for a comprehensive characterization of *C. pictus* germplasm. These findings provide a robust foundation for targeted breeding strategies aimed at enhancing both genetic diversity and the production of valuable secondary metabolites like diosgenin, ultimately contributing to the conservation and sustainable utilization of this important medicinal plant.

Declarations

Funding

The author affirms that no grants, funds, or other forms of assistance were obtained in order to prepare this manuscript.

Data availability Information from the supplementary material and the manuscript. On reasonable request, the corresponding author will provide any necessary data.

Conflicting Interests The author has disclosed no conflicting interests.

Author Contribution

P.R. conducted the research wrote the main manuscript, R.C, D.K, A.H edited manuscript, S.V, R. B and P.K revised manuscript verified it. The manuscript was reviewed by each author

References

1. Barrett, S. C., and J. R. Kohn. 1991. "Genetic and evolutionary consequences of small population size in plants: implications for conservation."
2. Benny, M. 2004. "Insulin plant in gardens."
3. Bhattacharyya, P., and S. Kumaria. 2015. "Molecular characterization of *Dendrobium nobile* Lindl., an endangered medicinal orchid, based on randomly amplified polymorphic DNA." *Plant systematics and evolution*. 301:201-210.
4. Chaudhary, M. K., A. Misra, P. K. Srivastava, and S. Srivastava. 2023. "Influence of Seasonal Variation on Diosgenin Content in *Costus speciosus* (J. Koenig) Sm. Rhizome Quantified Through Validated RP-HPLC-PDA Method." *Pharmacognosy Magazine*. 19 (1):66-74.
5. Chaudhary, S. A., P. S. Chaudhary, B. A. Syed, R. Misra, P. G. Bagali, S. Vitalini, and M. Iriti. 2018. "Validation of a method for diosgenin extraction from fenugreek (*Trigonella foenum-graecum* L.)." *Acta scientiarum polonorum. Technologia Alimentaria*. 17 (4):377-385.
6. Chen, H., L. Wang, C. Wang, Y. Zhang, H. Yu, Z. Fu, X. Fu, and L. Han. 2022. "Strategy of combining offline 2D LC-MS with LC-DIA-MS/MS to accurately identify chemical compounds and for quality control of *Dioscorea septemloba* Thunb." *Phytochemical Analysis*. 33 (7):1135-1146.
7. Christenhusz, M. J., and J. W. Byng. 2016. "The number of known plants species in the world and its annual increase." *Phytotaxa*. 261 (3):201–217-201–217.
8. Das, A., M. Gaur, D. Barik, and E. Subudhi. 2017. "Genetic diversity analysis of 60 ginger germplasm core accessions using ISSR and SSR markers." *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology*. 151 (5):822-832.
9. Fekete, S., J. Schappler, J.-L. Veuthey, and D. Guillarme. 2014. "Current and future trends in UHPLC." *TrAC Trends in Analytical Chemistry*. 63:2-13.

10. Ganie, S. H., P. Upadhyay, S. Das, and M. P. Sharma. 2015. "Authentication of medicinal plants by DNA markers." *Plant gene*. 4:83-99.
11. GM, R. D. 2024. "Extraction, Isolation, Identification and Estimation of Diosgenin by TLC Profiling and UHPLC-LC-SRM Analysis in Three Costus Species." *Journal of Drug Delivery & Therapeutics*. 14 (11).
12. Gogoi, A., S. Munda, M. Paw, T. Begum, M. H. Siddiqui, A.-R. Z. Gaafar, M. S. Kesawat, and M. Lal. 2023. "Molecular genetic divergence analysis amongst high curcumin lines of Golden Crop (*Curcuma longa* L.) using SSR marker and use in trait-specific breeding." *Scientific Reports*. 13 (1):19690.
13. Guillarme, D., and J.-L. Veuthey. 2012. *UHPLC in life sciences*. Royal Society of Chemistry.
14. Hasan, N., R. A. Laskar, S. A. Farooqui, N. Naaz, N. Sharma, M. Budakoti, D. C. Joshi, S. Choudhary, and M. S. Bhinda. 2024. "Genetic Improvement of Medicinal and Aromatic Plant Species: Breeding Techniques, Conservative Practices and Future Prospects." *Crop Design*.100080.
15. Hegde, P. K., H. A. Rao, and P. N. Rao. 2014. "A review on Insulin plant (*Costus igneus* Nak)." *Pharmacognosy reviews*. 8 (15):67.
16. Huang, X., and B. Han. 2014. "Natural variations and genome-wide association studies in crop plants." *Annual review of plant biology*. 65 (1):531-551.
17. Kala, C. P. 2000. "Status and conservation of rare and endangered medicinal plants in the Indian trans-Himalaya." *Biological conservation*. 93 (3):371-379.
18. Kuttappety, M., K. Surendran, and P. P. Pillai. 2023. "Cross section of genetic diversity in mainland and insular populations of *Costus speciosus* (Koen ex. Retz.) Sm. using SPAR markers reveal patterns linked to allopolyploidy and biogeography."
19. Li, H.-S., and G.-Z. Chen. 2004. "Genetic diversity of *Sonneratia alba* in China detected by inter-simple sequence repeats (ISSR) analysis." *Acta Botanica Sinica-English Edition*-. 46 (5):515-521.
20. Marotti, I., A. Bonetti, M. Minelli, P. Catizone, and G. Dinelli. 2007. "Characterization of some Italian common bean (*Phaseolus vulgaris* L.) landraces by RAPD, semi-random and ISSR molecular markers." *Genetic Resources and Crop Evolution*. 54:175-188.
21. Nag, A., P. S. Ahuja, and R. K. Sharma. 2015. "Genetic diversity of high-elevation populations of an endangered medicinal plant." *AoB Plants*. 7:plu076.
22. Naik, A., P. Prajapat, R. Krishnamurthy, and J. Pathak. 2017. "Assessment of genetic diversity in *Costus pictus* accessions based on RAPD and ISSR markers." *3 Biotech*. 7:1-12.
23. Pazhanichamy, K., K. Bhuvaneswari, B. Kunthavai, T. Eevera, and K. Rajendran. 2012. "Isolation, characterization and quantification of diosgenin from *Costus igneus*." *JPC–Journal of Planar Chromatography–Modern TLC*. 25:566-570.
24. Peakall, R., and P. E. Smouse. 2006. "GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research." *Molecular ecology notes*. 6 (1):288-295.
25. Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey, and A. Rafalski. 1996. "The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis."

Molecular breeding. 2:225-238.

26. Prajapat, P., N. Sasidharan, and A. Ballani. 2015. "Assessment of Genetic Diversity in Four Brassica species using randomly amplified polymorphic DNA markers." *International Journal of Agriculture, Environment and Biotechnology*. 8 (4):831.
27. Raju, J., and R. Mehta. 2008. "Cancer chemopreventive and therapeutic effects of diosgenin, a food saponin." *Nutrition and cancer*. 61 (1):27-35.
28. Rawat, P., M. Kumar, A. Srivastava, B. Kumar, A. Misra, S. Pratap Singh, and S. Srivastava. 2021. "Influence of Soil Variation on Diosgenin Content Profile in *Costus speciosus* from Indo-Gangetic Plains." *Chemistry & Biodiversity*. 18 (6):e2000977.
29. Saitou, N., and M. Nei. 1987. "The neighbor-joining method: a new method for reconstructing phylogenetic trees." *Molecular biology and evolution*. 4 (4):406-425.
30. Selim, S., and S. Al Jaouni. 2015. "Anticancer and apoptotic effects on cell proliferation of diosgenin isolated from *Costus speciosus* (Koen.) Sm." *BMC complementary and alternative medicine*. 15:1-7.
31. Srivastava, A., M. Kumar, A. Misra, P. K. Shukla, P. K. Agrawal, and S. Srivastava. 2019. "Evaluation of diosgenin content in *Costus speciosus* germplasm collected from Eastern Ghats of India and identification of elite chemotypes." *Pharmacognosy Magazine*. 15 (66).
32. Tar'an, B., C. Zhang, T. Warkentin, A. Tullu, and A. Vandenberg. 2005. "Genetic diversity among varieties and wild species accessions of pea (*Pisum sativum* L.) based on molecular markers, and morphological and physiological characters." *Genome*. 48 (2):257-272.
33. Urooj, A. 2010. "Nutrient profile and antioxidant components of *Costus speciosus* Sm. and *Costus igneus* Nak."
34. van Caspel, P. H., A. D. Poulsen, and M. Möller. 2021. "New chromosome counts of Asian costaceae and initial insights into the genome evolution of the family." *Edinburgh Journal of Botany*. 78:1-13.
35. Wolfe, A. D., and A. Liston. 1998. "Contributions of PCR-based methods to plant systematics and evolutionary biology." In *Molecular systematics of plants II: DNA sequencing*. 43-86. Springer.
36. Yan, C., T. You-Mei, Y. Su-Lan, H. Yu-Wei, K. Jun-Ping, L. Bao-Lin, and Y. Bo-Yang. 2015. "Advances in the pharmacological activities and mechanisms of diosgenin." *Chinese journal of natural medicines*. 13 (8):578-587.
37. Yang, W., A. C. de Oliveira, I. Godwin, K. Schertz, and J. L. Bennetzen. 1996. "Comparison of DNA marker technologies in characterizing plant genome diversity: variability in Chinese sorghums." *Crop science*. 36 (6):1669-1676.

Figures

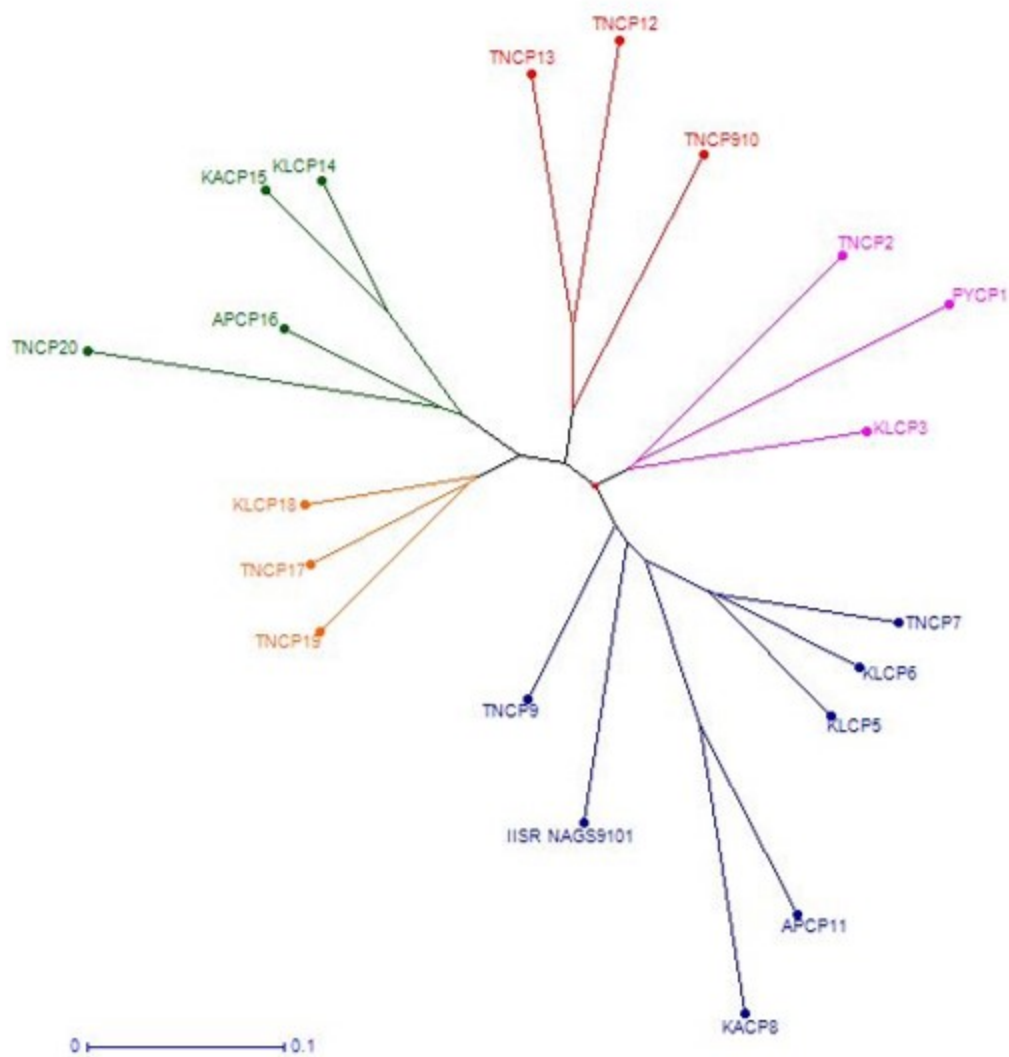


Figure 1

Using the Darwin software, a Neighbor-Joining tree was created, and the resulting dendrogram showed the existence of five different clusters. With three entities in cluster I (red), four in cluster II (green), seven in cluster III (blue), three in cluster IV (orange), and three in cluster V (violet), there were five different types of entities.

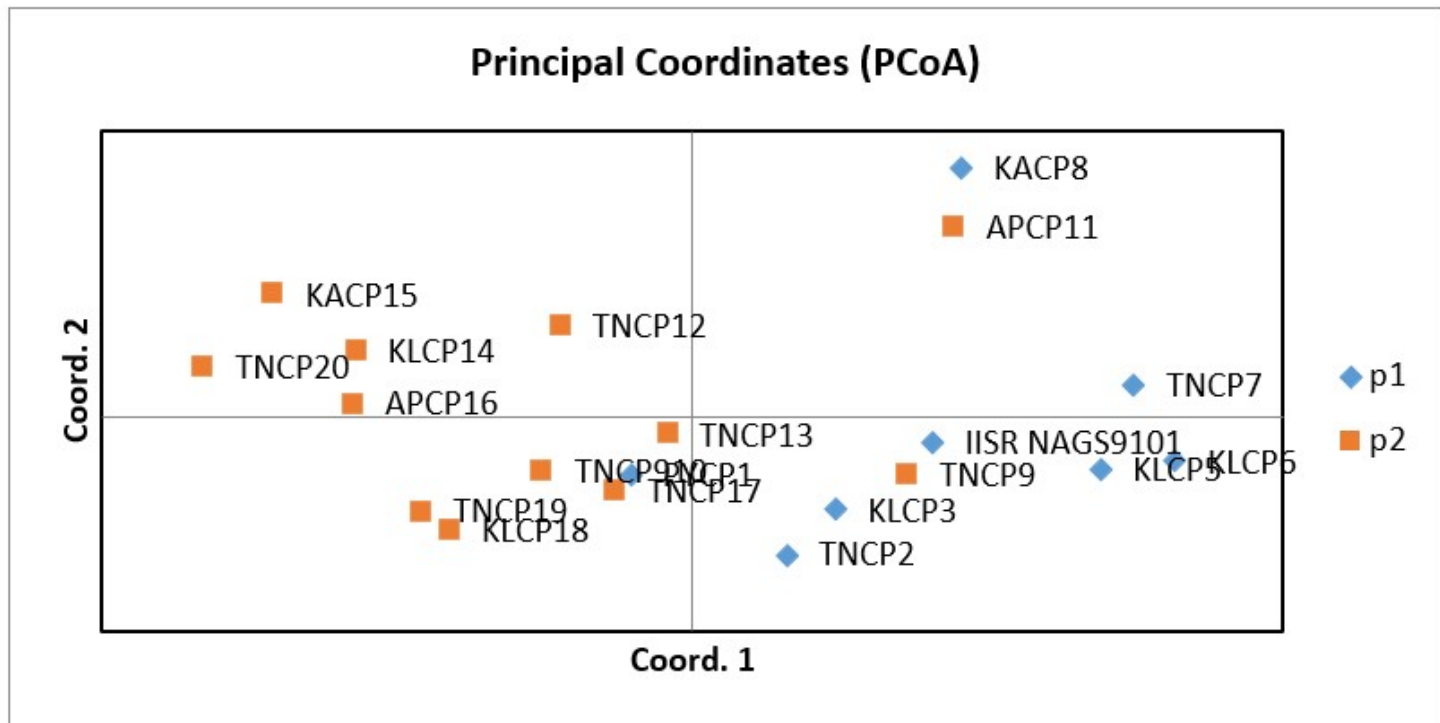


Figure 2

Principal Coordinate Analysis (PCoA) analysis on Insulin plant accession

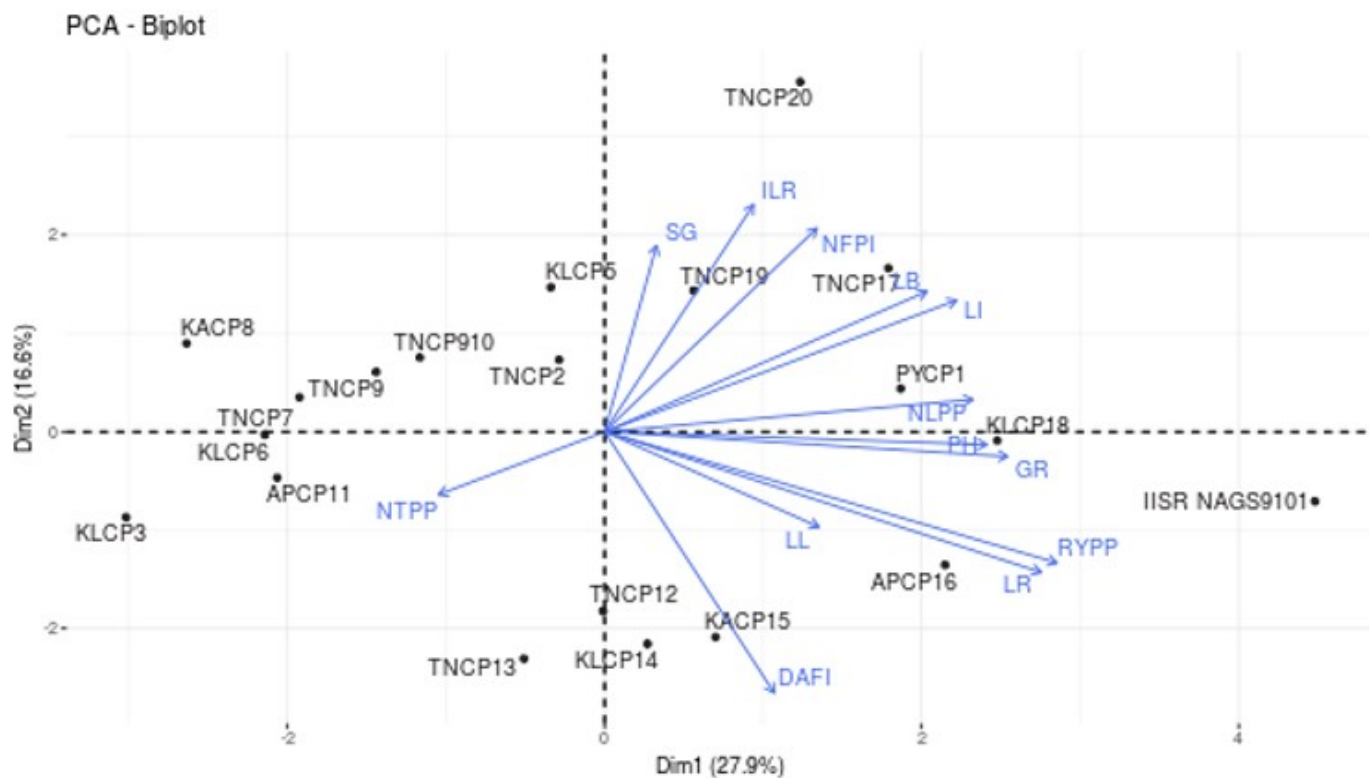


Figure 3

PC1 and PC2 are the biplots of the 20 insulin plants. Plant height (cm), number of leaves per plant (NLPP), number of tillers per plant (NTPP), leaf length (LL), leaf breadth (LB), length of inflorescence (LI), number of flowers per inflorescence (NFPI), days after flower initiation (DAFI), stem girth (SG), length of rhizome (LR), girth of rhizome (GR), inter nodal length of rhizome (INLR), and rhizome yield per plant (gm) are among the various characteristics.

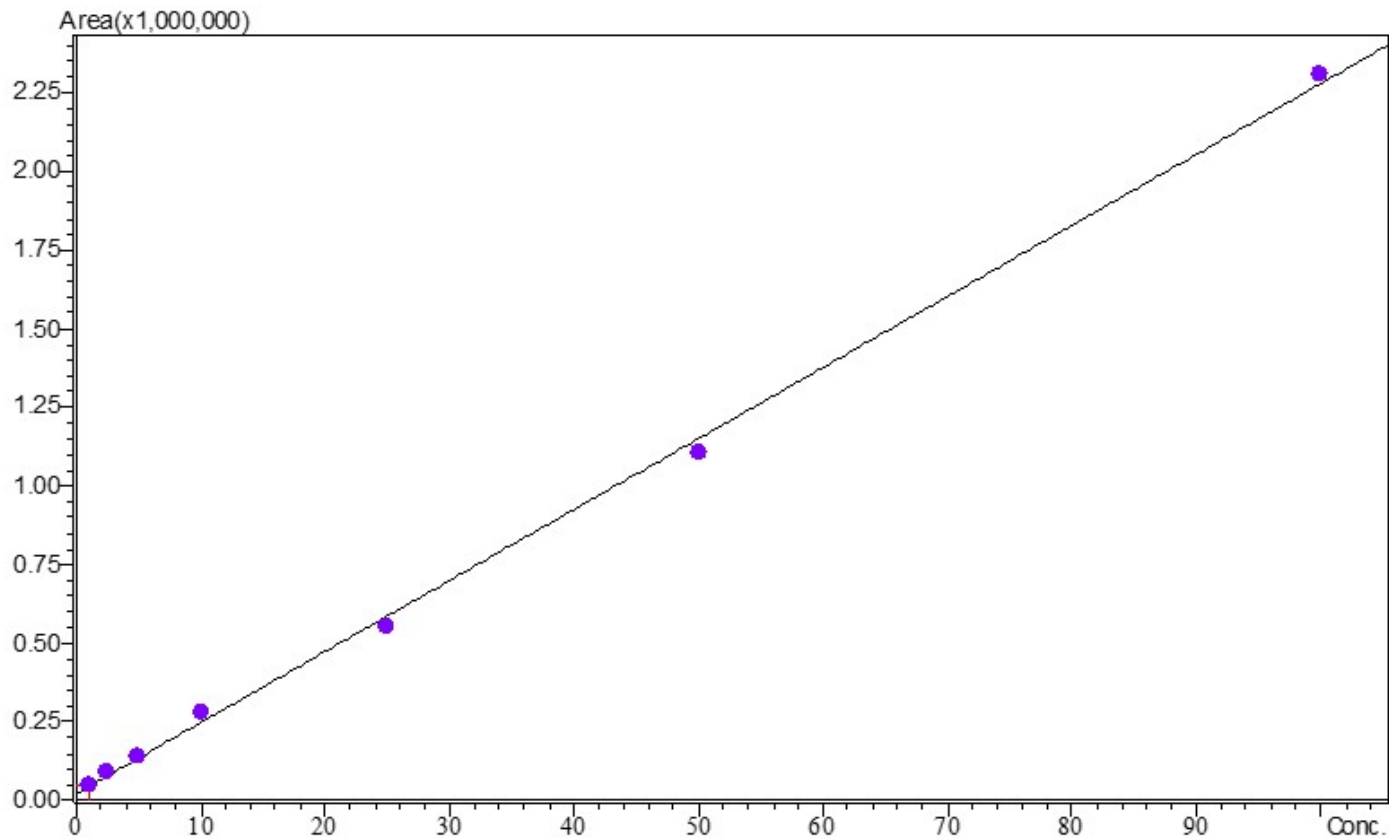


Figure 4
Linear calibration obtained for the diosgenin analysis from 1 – 100 ng/mL.

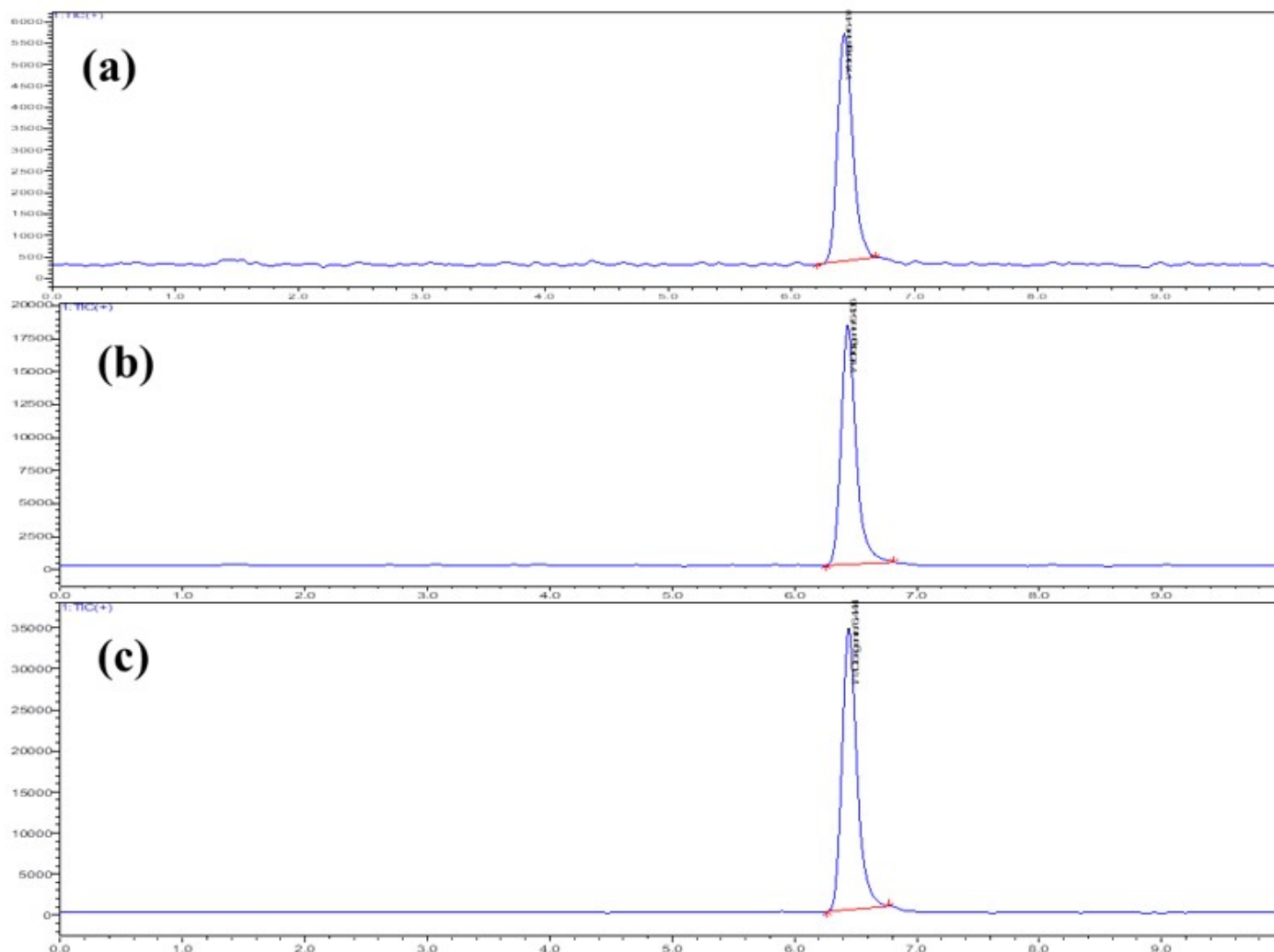


Figure 5

(a) Limit of detection Level Diosgenin (1 ng/g), (b) Limit of Quantification Level Diosgenin 5ng/g, and (c) Diosgenin in a *Costus pictus* plant sample extract with retention time of 6.43 min.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementaryfile1.docx](#)
- [supplementaryfile2.docx](#)
- [supplementaryfile3.docx](#)
- [supplementaryfile4.docx](#)
- [supplementaryfile5.docx](#)
- [supplementaryfile6.docx](#)
- [supplementaryfile7.docx](#)

- [supplementaryfile8.docx](#)