Statistical classification of major landraces in Thai *Dioscorea alata* with aboveground morphological data

Napat Jantaraprasit, Ekaphan Kraichak*

Department of Botany, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand

*Corresponding author. E-mail address: fsciepk@ku.ac.th

Abstract

Dioscorea alata L., greater yam, is a climber crop, producing edible tubers and bulbils. *D. alata* has many landraces worldwide from artificial selection. Accurate identification of different *D. alata* landraces from the aboveground part is challenging due to their twinning habit, polyploidy, and high morphological variation. To distinguish major landraces of *Dioscorea alata* in Thailand, we measured various aboveground morphological characters of nine major landraces and performed Principal Component Analysis (PCA) and K-means clustering analysis on the measured characters. In PCA, the first principal component explained only 35.1% of the total variance and exhibited ambiguous groupings for the four landraces. In addition, the K-means analysis showed overlapping clusters among the identified groups, which did not correspond to any of the known landraces. The results showed that the selected morphological characters could not discriminate Thai *D. alata's* landraces. This study exemplified the need for additional methods to identify the diversity among these landraces, which will likely become important Thai food crops.

Keywords: K-means, PCA, Phenetics, Yam

Introduction

Dioscorea alata L., greater yam, is a climber crop, producing nontoxic edible tuber and bulbil. This species is one of the most common yams cultivated in Thailand and Southeast Asia (Wilkin et al., 2009; Sugihara et al., 2021). Through thousand years of artificial selection, *D. alata* has many local landraces worldwide, based on different tuber colors, sizes, and shapes (Chaïr et al., 2016; Sharif et al. 2020). In the East and Northeast of Thailand, many landraces are cultivated as a cash crops and consumed in traditional cuisines. In contrast, the other landraces, such as Man Sao with characteristically columnar tuber, are only found in the wild. It is still unclear whether these diverse landraces originated in Thailand or were imported long ago and then became naturalized as a feral crop (Wilkin et al., 2009; Sharif et al., 2020; Sugihara et al., 2021; Wu et al., 2021). In Thailand, at least three cultivars are formally recognized only by their tuber shapes (Wilkin et al., 2009), even though the locals have named many more landraces for *D. alata*.

Accurate identification of different *D. alata* landraces has been challenging due to their twinning habit, polyploidy, continuous morphological variation, and small dioecious flower. Although landraces are mainly identified by their tuber characters, their accurate and consistent identification is difficult. Many tuber characters overlap (Hasan et al., 2008; Jyothy et al, 2017; Abeywickrama et al., 2018; Sheikh et al., 2019), but these landraces differ in their growing conditions, tuber yields and tuber tastes. Our inability to correctly identify these landraces will affect consumers' confidence and result in agricultural investment failure. To overcome this problem, researchers used multivariate analysis to classify each species and landraces into proper groups. The correct classification will allow us to cultivate, conserve genetic, and perform selective breeding with high precision in the future (Hasan et al., 2008; Wilkin et al., 2009; Jyothy et al, 2017; Abeywickrama et al., 2018).

In this work, we focused on the aboveground morphological characters, namely stems and leaves, for the statistical classification of the local landraces of *D. alata* in Thailand. These characters are easy to measure in the field at the early stages of cultivation and to compare with the reference specimens in the herbarium. We measured 14 aboveground morphological characters of nine major landraces. The data were subjected to multivariate analyses (principal component analysis, PERMANOVA, and k-means clustering) to determine whether these aboveground characters could classify the known landraces. The results would provide additional tools for future researchers interested in working with multiple landraces within *D. alata*.

Materials and Methods

Samples collection and species identification

During 2020–2021, we collected samples of the local landraces of *Dioscorea alata* from four locations in Eastern Thailand, where many species of *Dioscorea* served as forage foods and cultivated species. The locations included 1) Ban Tha Liap, Sanam Chai Khet District, Chachoengsao Province, 2) Local Eastern Agroforestry Community Enterprise (Herb Home and Forest), Sanam Chai Khet District, Chachoengsao

Province, 3) Local agroforestry area in Ban Klum Suai, Kaeng Hang Maeo District, Chanthaburi Province, and Aran Piyawong forest park, Kaeng Hang Maeo District, Chanthaburi Province. Fig. 1

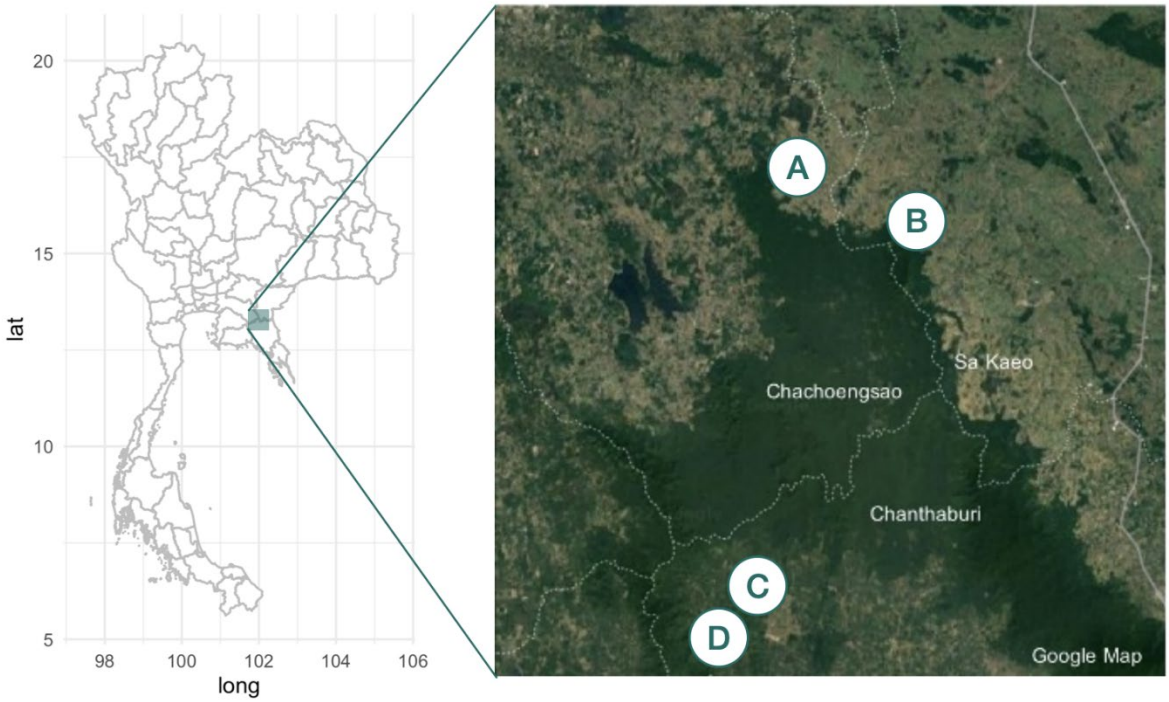


Fig. 1 The location of collecting sites of *D. alata* landraces in Eastern Thailand: (A) Ban Tha Liap; (B) Herb Home and Forest; (C) Ban Klum Suai; and (D) Aran Piyawong Forest Park

For each plant, we collected mature leaves and twining stems using pruning shears and pressed them with plant presses. The samples were dried in a herbarium drying cabinet. After completely dried, a total of 29 samples were preserved as voucher specimens at the Department of Botany, Faculty of Science, Kasetsart University. We identified landraces according to the vernacular name used by the locals in the collecting area. The identity at the species level (as *D. alata*) was verified with the description in the Flora of Thailand Volume 10 Part 1 (Wilkin et al, 2009) and the comparisons with voucher specimens in Bangkok Herbarium Forests (BKF) and digital images of the specimens through POWO (2022).

Multivariate morphological characters analysis

For each dried sample, we measured a total of 14 leaf and stem characters that could also be easily observable and reliably measured during field identification (Table 1). The majority of the characters were on petiole or lamina. We used the maximum value of each character to represent each sample because it was supposedly closest to the size at maturity. The landraces with at least three replicates were included in the subsequent multivariate classification.

We performed two types of multivariate analysis to classify the collected landrace statistically. First, we subjected the data to the Principal Component Analysis (PCA) using the "prcomp" function. The first two

principal components were plotted and annotated with the known names of landraces to see if they were separable. The permutational multivariate analysis of variance (PERMANOVA) was performed using the "adonis2" function in the vegan package (Oksanen et al, 2020) to test whether the landraces were significantly different from each other. The other approach involves the classification without a priori information about the landrace through k-means clustering. This method clusters the samples into groups to reduce the distance to the group's centroid. We set the number of groups (k) to four, the same number as the known landrace, to see if the unsupervised procedure could independently classify the landraces. The analysis and visualization were performed in R version 4.1.1 (R Core Team, 2022).

Results and Discussion

From a total of 29 samples, we were able to identify 9 landraces of *Dioscorea alata* (Table 1) Fig. 2, including 1) Lueat (มันเลือด), 2) Lueat Chom-poo (มันเลือดชมพู), 3) Lueat Nok (มันเลือดนก), 4) Tapab (มันตะพาบ), 5) Talab (มันตลับ), 6) Chao Maphrao (มันจาวมะพร้าว), 7) Chao Maphrao Daeng (มันจาวมะพร้าวแดง), 8) Sao (มันเสา), and 9) Sak (มันสาก). Based on our fieldwork in Eastern of Thailand, we found that Chao Maphrao and Chao Maphrao Daeng were two main landraces cultivated as cash crops. Additional eight landraces were cultivated as homegrown vegetable supplies and forage crops. However, more landraces of *D. alata* are expected to be found in other regions of Thailand (Wilkin et al., 2009). Additional detailed taxonomic studies will be required to gain a complete list of landraces in the country.

Lueat, Lueat Chom-poo, and Lueat Nok were broadly similar and therefore were recognized as one group, "Lueat," for the subsequent analyses. The four landraces with at least three samples, including Lueat, Talab, Chao Maphrao, and Chao Maphrao Daeng, were included in the multivariate analysis.

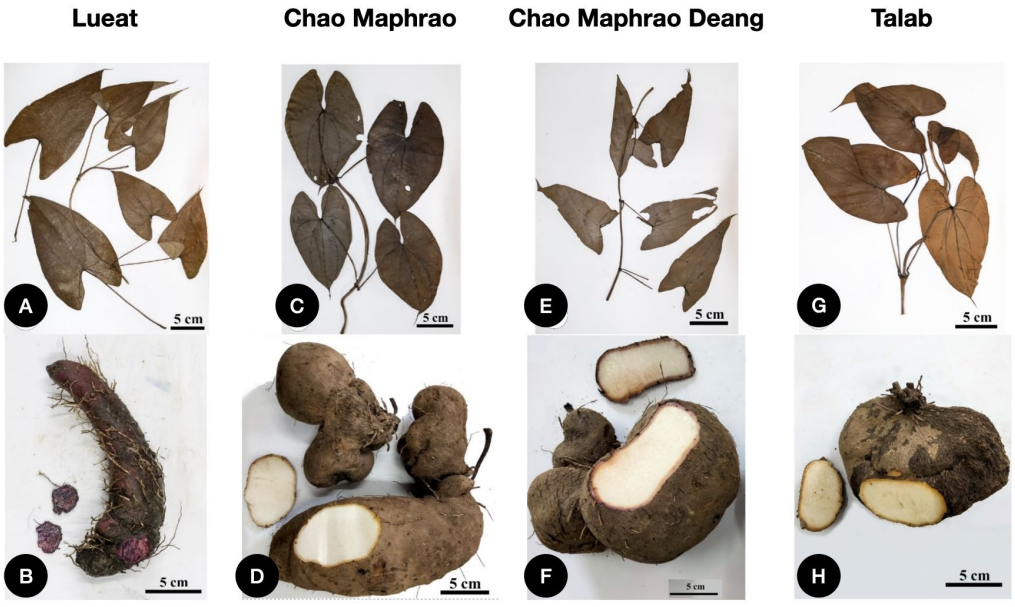


Fig. 2 Examples of *Dioscorea alata* landraces. The top panels are the dried herbarium specimens, and the bottom panels are the tubers. (A), (B) Lueat, (C), (D) Chao Maphrao, (E), (F) Chao Maphrao Deang, and (G), (H) Talab

Table 1 Aboveground morphological characters of *Dioscorea alata's* landraces collected in this study. When the replicates are available, the numbers are displayed as mean ± 1 standard deviation (SD)

Landrace	Lueat* (มันเลือด)	Lueat Chom-poo* (มันเลือดชมพู)	Lueat Nok* (มันเลือดนก)	Tapab (มันตะพาบ)	Talab (มันตลับ)	Chao Maphrao (มันจาวมะพร้าว)	Chao Maphrao Daeng (มันจาวมะพร้าวแดง)	Sao (มันเสา)	Sak (มันสาก)
Number of Samples	1	1	2	1	3	3	3	2	2
Stem width (cm)	3.2	3	3 ±0.7	2.8	4.5±0.3	3.4±1.7	2.8±0.4	2.4±0.5	2.8±0.01
Internode length (cm)	16.5	14	13.5±3.5	12.5	13.2±1.8	18.8±2.5	12.5±0.7	14.2±8.1	13.2±1.1
Node width (cm)	6.2	8.5	4.6±0.6	4.8	8.7±0.3	5.3±0.8	4.4±0.4	4.4±0.9	4.8±1
Petiole length (cm)	16	11.5	10.5±2.8	12.5	10.2±1.8	10.5±1.4	7±0.71	9.8±1.8	10.4±0.9
Petiole width (cm)	1.5	1.9	1.2±0.10	1.3	1.4±0.2	1.6±0.2	1.2±0.1	0.9±0.1	1.2±0.1
Leaf length (cm)	16.5	15.5	11.8±1.1	12.5	13.8±2.6	13.7±1.3	15.8±0.4	13.3±5.2	9.5±5.4
Leaf width (cm)	11.5	9.5	7.4±1.3	12	11±1.2	10±1.4	7.6±0.1	9.4±4.7	10.5±0.7
Lobe length (cm)	6.4	5.5	3.9±0.6	4.4	4.8±0.4	5±0.8	4.2±0.4	4.9±2.7	5
Lobe gap width (cm)	7.3	8.5	4.6±1.2	7	5.5±3.5	3.2±0.5	6.6±0.2	8.4±3.3	6.6±1.3
Leaf caudate depth (cm)	1.5	1.2	0.8±0.3	0.7	1.4±0.4	1.2±0.3	1.2±0.1	0.8±0.2	0.9±0.4
Number of veins	7	7	7	7	7, 9	7	7	7	7
Number of veins reach the apex	5	5	3	5	5	5	3	3, 5	3, 5
Three middle vein distance (cm)	4.8	4.5	3.8±0.6	5	4±0.2	4±0.6	3.6±0.4	3.6±2	4.6±0.1
Scalar veinlet distance (cm)	0.8	0.7	0.7±0.2	0.8	0.6±0.1	0.6±0.1	0.6±0.1	0.6±0.4	0.6±0.1

*These three landraces were grouped as "Lueat" for the multivariate analysis.

Multivariate morphological characters analysis

The PCA result showed that most of the studied landraces were not readily separated Fig. 3. The first principal component (PC1) explained about 35.1% of which was higher than the PCA using only tuber morphological characters in the other study at 19.5% (Hasan et al., 2008). Chao Maphrao Daeng landrace appeared to separate clearly from the other landraces. At the same time, the samples from Chao Maphrao and Lueat distributed widely in the morphospace, suggesting that these two landraces were highly variable. The PERMANOVA showed that these landraces were significantly different from each other ($F_{3,10}$ = 1.12, P = 0.011). The result was consistent with the visualization of the first two PCs. The significant difference among the landrace was likely due to the separation of Chao Maphrao Daeng from the other landraces.

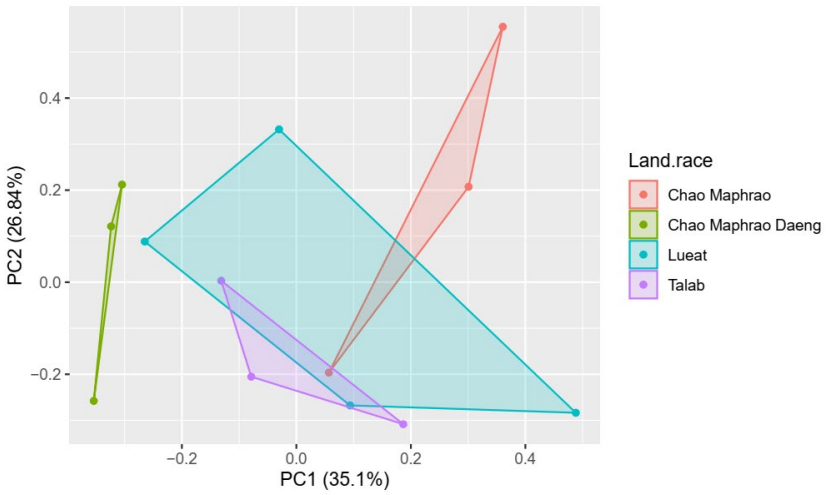
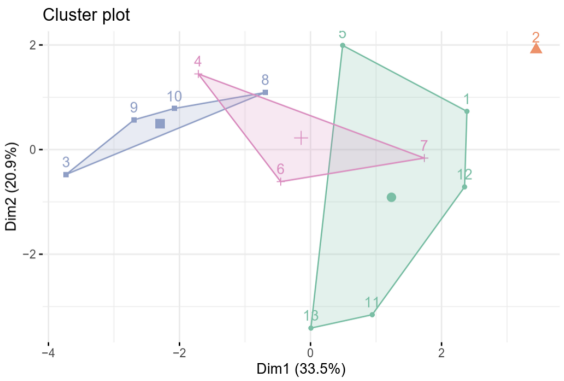


Fig. 3 The first two principal components from the PCA of the aboveground morphological characters with the known landraces annotated onto the results. The four landraces differed significantly from each other (PERMANOVA, $p = 0.011$). Chao Maphrao Daeng landrace is separate from the other, while the other landraces overlap in the morphospace

K-means results showed that none of the landraces was clustered in its group Fig. 4A. The first cluster comprised the samples of Lueat, Chao Maphrao, and Talab. The second cluster only comprised the samples of Talab. The third cluster comprised the samples of Lueat and Chao Maphrao Daeng. Lastly, the fourth cluster comprised the samples of Lueat and Chao Maphrao Fig. 4B. Interestingly, all of the samples from Talab and Chao Maphrao Deang belonged to their respective cluster, while Lueat and Chao Maphrao were more variable than the others. The current classifications of Lueat and Chao Maphrao landraces may still be quite broad, encompassing a high degree of morphological variation. Alternatively, these landraces may be experiencing hybridization among themselves and therefore are yet to form genetically and morphologically unique groups within the species.

While the statistical classification of aboveground characters could distinguish some landraces from others, it was clear that using aboveground morphological characters could not fully discriminate a large number of landraces of *Dioscorea alata.* Along with the increased sample size, the combination of underground characters, stems, and other qualitative characters may better classify the local landraces of *D. alata.* Additional molecular data could be generated to improve the classification. Our preliminary work on single-locus markers showed relatively poor results due to low DNA quality. Polyploidy was also reported in *D. alata*, with the number of ploidies being diploid, triploid, and tetraploid. However, no clear correlation between ploidy level and reproductive status or morphological character has been found (Egesi et al., 2002; Obidiegwu et al., 2010; Arnau et al., 2009). Future work on molecular markers using genome-wide SNPs information should also be conducted to increase the accuracy of the classification and identification of these important native crop plants. This study exemplified the need for additional methods to identify the diversity among these landraces, which will likely become important Thai food crops.



k-means cluster	Lueat	Chao Maphrao	Chao Maphrao Deang	Talab
1	1	5	-	11,12,13
2	2	-	-	-
3	3	-	8,9,10	-
4	4	6,7	-	-

Fig. 4 (A) K-means clustering results superimposed onto the ordinated space. The numbers represent the sample number. (B) The four clusters from the K-means clustering did not match with the known landraces. The numbers in the tables correspond to the same sample number in (A)

Conflict of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

We thank Mrs. Sarinya Kumpila and her family for providing *Dioscorea* samples. This study was funded by The National Science and Technology Development Agency (NSTDA), Thailand (P-19-52237), and the Development and Promotion of Science and Technology scholarship.

References

Abeywickrama, K.G.T.A.K., Samarasinghe, W.L.G., Sumanasinghe, V.A., Edirisinghe, E.S.C. 2018. Analysis of diversity and determination of duplicates among twenty one *Dioscorea* accessions through morphological and molecular characterization. SLJFA. 4: 43–48. doi.org/10.4038/sljfa.v4i1.57.

Arnau, G., Nemorin, A., Maledon, E., Abraham, K. 2009. Revision of ploidy status of *Dioscorea alata* L. (Dioscoreaceae) by cytogenetic and microsatellite segregation analysis. Theor. Appl. Genet.18: 1239–1249. doi: 10.1007/s00122-009-0977-6

Chaïr, H., Sardos, J., Supply, A., Mournet, P., Malapa, R., Lebot, V. 2016. Plastid phylogenetics of Oceania yams (*Dioscorea* spp., Dioscoreaceae) reveals natural interspecific hybridization of the greater yam (*D. alata*). Bot. J. Linn. Soc. 180: 319–333. doi.org/10.1111/boj.12374.

Egesi, C.N., Pillay, M., Asiedu, R., Egunjobi, J.K. 2002. Ploidy analysis in water yam, *Dioscorea alata* L. germplasm. Euphytica 128: 225–230. doi.org/10.1023/A:1020868218902

Zain, H.S.M., Ngadin, A.A., Shah, R.M., Mohamad, N. 2008. Morphological variability of greater yam (*Dioscorea alata* L.) in Malaysia. Plant Genet. Res. 6: 52–61. doi.org/10.1017/s1479262108920050.

Jyothy, A., Sheela, M.N., Radhika, N.K., Anwar, I., Kumar, V., Abhilash, P.V. 2017. Morphological characterization of greater yam (*Dioscorea alata* L.) landraces in Kerala. J. root crops. 43: 3–10.

Obidiegwu, J., Rodriguez, E., Ene-Obong, E., Loureiro, J., Muoneke, C., Santos, C., Kolesnikova-Allen, M., Asiedu, R. 2010. Ploidy levels of *Dioscorea alata* L. germplasm determined by flow cytometry. Genet. Resour. Crop Evol. 57: 351–356. doi.org/10.1007/s10722-009-9473-8.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R. B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E., Wagner, H. 2020. vegan: Community Ecology. Package. R package version 2. 5–7.

POWO. 2022. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet; http://www.plantsoftheworldonline.org/.

R: A language and environment for statistical. R Foundation for Statistical Computing, Vienna, Austria.

Sharif, B.M., Burgarella, C., Cormier, F., et al. 2020. Genome-wide genotyping elucidates the geographical diversification and dispersal of the polyploid and clonally propagated yam (*Dioscorea alata*). Ann. Bot. 126: 1029–1038. https://doi.org/10.1093/aob/mcaa122.

Sheikh, N., Kumar, Y., Misra, A. K. 2019. Molecular characterization and phylogenetic assessment of a few *Dioscorea* (Dioscoreaceae) species of North-East India. Indian J Genet Plant Breed. 79: 82–88.

Sugihara, Y., Kudoh, A., Tamiru, M. O., et al. 2021. Population genomics of yams: evolution and domestication of *Dioscorea* species. Population Genomics. Springer, Cham. New York, USA.

Wilkin, P., Chirdsak, T., Santisuk, T., Larsen, K. 2009. Flora of Thailand: Dioscoreaceae, Part 1, Vol, 10. Forest Herbarium, Royal Forest Department. Bangkok. Thailand.

Wu, D., Lao, S., Fan, L. 2021. De-domestication: an extension of crop evolution. Trends Plant Sci. 26: 560–574. https://doi.org/https://doi.org/10.1016/j.tplants.2021.02.003.