Research Article

Genetic Diversity of Hemp Germplasm in Northern Thailand

Kanjanabhorn Aksonsit and Sompid Samipak*
Department of Genetics, Faculty of Science, Kasetsart University, Bangkok, Thailand

Tanapon Chaisan
Department of Agronomy, Faculty of Agriculture, Kasetsart University, Bangkok, Thailand

Wanchai Pluempanupat
Department of Chemistry, Faculty of Science, Kasetsart University, Bangkok, Thailand

Piyada Juntawong
Texas Department of State Health Services, Austin, Texas, United States

* Corresponding author. E-mail: sompid.s@ku.th DOI: 10.14416/j.asep.2025.07.003
Received: 17 March 2025; Revised: 13 May 2025; Accepted: 29 May 2025; Published online: 4 July 2025

Abstract
Cannabis sativa L., commonly known as hemp, is a plant native to Central Asia. It is well known for its cannabinoid compounds, which have significant potential for medical applications. Recognizing the economic and medical value of hemp, the Thai government has permitted its cultivation for commercial, medical, and research purposes. However, a comprehensive understanding of hemp genetics is crucial to support industry expansion and enhance future breeding programs. This study investigated the genetic diversity of 37 hemp accessions collected in northern Thailand, along with two reference varieties (RPF1 and RPF2). Using DArTSeq-based genotyping-by-sequencing and whole-genome sequencing technologies, we identified 3,609 single nucleotide polymorphisms (SNPs). STRUCTURE analysis, principal component analysis (PCA), and neighbor-joining analysis consistently identified three genetic clusters; however, these clusters did not correlate with geographical locations. Genetic differentiation among clusters was observed (fixation index, F_{ST} = 0.064-0.079; Nei's coefficient of genetic differentiation, Nei's G_{ST} = 0.058-0.078). Total genetic diversity estimated (expected heterozygosity, H_E = 0.348; observed heterozygosity, H_O = 0.092). Global inbreeding (F_{IT} = 0.033) and molecular variance (4.83%) suggested low to moderate genetic differentiation, while the high inbreeding coefficient (F_{IS} = 0.737) indicated substantial inbreeding within clusters. The genetic data from this study provide a resource for developing molecular markers to distinguish hemp varieties, supporting selective breeding efforts. These findings will contribute to improving agronomic traits, conserving genetic diversity, and ensuring the sustainable use of hemp genetic resources.

Keywords: *Cannabis sativa*, DArTseq genotyping, Genetic diversity, Single nucleotide polymorphisms, STRUCTURE

1 Introduction

The genus *Cannabis* (2n=20) has long been domesticated and has coexisted with mankind. It is a dioecious, annual herb belonging to the family Cannabaceae [1]. The center of origin is believed to have been in Central Asia, from where it spread to India and the Middle East before reaching other countries [2]. The prominent species in the Cannabaceae family are *Cannabis sativa* (hemp) and *Cannabis indica* (marijuana). Besides these two main species, crossbreeding between them has produced hybrid varieties with varying compositions and properties [3]. Traditionally, cannabis plants have

been classified based on their morphology and chemical composition [4]. Hemp is a fiber-type variety of cannabis that is rich in cannabidiol (CBD) and related compounds while containing little to no psychoactive delta-9-tetrahydrocannabinol (THC). The drug-type variety, commonly referred to as marijuana, has high levels of THC and has long been used for medicinal or recreational purposes. Hemp has been historically utilized for textiles and food production [5]. CBD has gained popularity in clinical research due to its non-intoxicating effects. It is used to treat epilepsy, substance use disorders, and to manage pain associated with rheumatoid arthritis, and various skin conditions [6], [7]. THC, which has psychoactive effects, has recently been found to be effective in the treatment of Alzheimer's disease [8] and glaucoma [9].

With the increasing recognition of hemp's importance and benefits, the Thai government has permitted its cultivation for commercial, medical, and research purposes [10], [11]. However, no official hemp breeding program currently exists in Thailand. Molecular techniques and morphological markers have been used to study hemp's genetic diversity. DNA sequence-based approaches are now more widely adopted because morphological characteristics are influenced by environmental factors, and molecular markers can be technically challenging to analyze [12], [13].

Plant diversity can be examined using various molecular techniques, such as Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), and Simple Sequence Repeats (SSR). Each technique has distinct advantages and limitations, including increased complexity, lengthy processing time, or high costs [14]-[17]. For these reasons, this study utilized the Diversity Arrays Technology (DArT) method, which originated from the solid-state method and is used to analyze DNA polymorphisms across the genome [18].

Genetic diversity is a crucial parameter for assessing the evolutionary history, conservation status, and breeding potential of plant varieties. Separating cannabis strains into genetically distinct groups is essential for understanding their evolutionary history and practical potential. Identifying clear population structure allows researchers to more effectively associate genetic backgrounds with important traits such as fiber quality, cannabinoid content, and stress tolerance. This separation also provides a foundation for more accurate genome-wide association studies (GWAS) and targeted breeding strategies [19]. For example, Sawler *et al*., [20] demonstrated strong genetic differentiation between hemp and drug-type cannabis, with implications for both breeding and regulation. Gao *et al*., [21] focused on fiber-type hemp varieties with shared traits, offering insights into industrial improvement. Similarly, Vergara *et al*., [22] analyzed groups with similar chemotypes to highlight the risk of genetic erosion due to uncontrolled hybridization. Studying well-defined groups not only enhances breeding efficiency but also supports the conservation of genetic resources, ultimately advancing both scientific research and commercial development.

DArTseq is a high-throughput genotyping technology designed to efficiently detect single nucleotide polymorphisms (SNPs) across multiple genomic locations. It employs restriction enzyme digestion to reduce genome complexity so that the predominantly active-low copy sequence areas are selected. These selected fragments are then ligated to adaptors, amplified, and sequenced using next-generation sequencing (NGS) platforms. By focusing on a representative fraction of the genome rather than sequencing the entire genome, DArTseq provides a cost-effective and scalable solution for genetic diversity studies, molecular breeding, and conservation research [23].

In Thailand, numerous researchers have studied hemp, focusing on its fiber properties and bioactive chemical compounds for medical applications [24]–[27]. However, genetic research on hemp remains limited. Understanding genetic diversity and variation is crucial for the development of future breeding programs. Comprehensive genetic studies are necessary to determine whether hemp cultivars collected from different regions in Thailand share genetic similarities or exhibit distinct genetic diversity. Such insights will contribute to effective germplasm management and the improvement of hemp varieties.

This study examined the genetic data of 37 hemp accessions collected from northern Thailand, along with two reference varieties (RPF1 and RPF2). The accessions were genotyped for SNP variation using DArTSeq-based genotyping-by-sequencing (GBS) technology and whole-genome sequencing (WGS) technologies. A phylogenetic dendrogram was constructed to visualize genetic relationships among accessions, and population structure analysis was performed to infer genetic subgroups.

2 Materials and Methods

2.1 *Hemp seed collection*

Hemp samples were collected from three provinces in northern Thailand (Figure 1). Seeds were germinated in seed trays and cultivated at a plantation in Doi Lo District, Chiang Mai, Thailand (GPS: 18°33'57.2"N 98°48'54.0"E). Leaf samples (55 days old) from the 37 accessions (KU) and two reference varieties (RPF) (Table S1) were harvested and immediately kept on ice until DNA extraction.

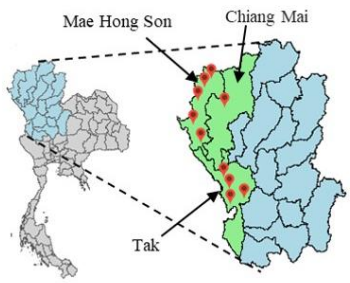

Figure 1: Map of Thailand showing northern Thailand (left) and the three provinces (right) where hemp accessions were collected. Red symbols represent the collection sites.

2.2 *DNA extraction*

DNA was extracted from 100 mg of young leaf tissue using the GenUP™ Plant DNA Kit (Biotechrabbit, Germany). DNA quality was assessed using a NanoDrop spectrophotometer (Thermo Scientific, USA) and by performing 1% agarose gel electrophoresis in TBE buffer at 120 V. For KU accessions, the DNA was digested with *EcoRI* restriction enzyme (New England Biolabs, USA), and the digested products were re-analyzed using 1% agarose gel electrophoresis.

2.3 *DNA sequencing*

For next-generation sequencing, the 37 KU accessions were sent to Diversity Arrays Technology Pty Ltd. (Canberra, ACT, Australia) for whole-genome profiling using DArTseq genotyping-by-sequencing (GBS) [28]. Meanwhile, the RPF accessions underwent whole genome sequencing at BGI Genomics (Shenzhen, China), where sequencing was performed using an Illumina Hiseq 2000 system (Illumina, CA, USA).

2.4 *Sequence analysis*

Sequencing data from DArTseq and whole-genome approaches were quality-checked using FastQC version 0.11.9 [29], and low-quality reads were trimmed from FASTQ files using Trimmomatic version 0.39 [30]. High-quality reads were aligned to the reference genome (hemp genome version cs10) and reformatted into BAM format using the Burrows-Wheeler Aligner (BWA) version 0.7.17 [31] and SAMtools version 1.10 [32]. The resulting BAM files were used for variant calling with GATK version 4.1.8.1 [33], generating a variant call format (VCF) file. The VCF file, containing single nucleotide polymorphisms (SNPs) and insertion/deletion (indel) data, was filtered using BCFtools version 1.10 [34] with a QUAL threshold greater than 20% and a minor allele frequency (MAF) greater than 5 %. Annotation of variants was subsequently performed using SnpEff version 5.1 [35]. The dendrogram was constructed using the weighted neighbor-joining method [36] with MEGA11 [37].

2.5 *Population structure and data analysis*

The population structure was analyzed using STRUCTURE version 2.3.4 software [38]. The number of hypothetical clusters (K) was set from 1 to 10, with five replicates for each K value. Subpopulations were inferred using an admixture model with a burn-in period of 50,000 cycles and 100,000 Markov Chain Monte Carlo (MCMC) cycles. The optimal K value was determined using the Evanno method [39] and visualized using the pophelper version 2.3.1 [40] package in R. The expected heterozygosity (H_E), observed heterozygosity (H_O), inbreeding coefficient (F_{IS}), global inbreeding coefficient (F_{IT}), and gene flow (Nm) were calculated using adegenet version 2.1.10 [41] and hierfstat version 0.5-11 [42] in R. Nei's coefficient of gene differentiation (G_{ST}) was estimated using vcfR version 1.15.0 [43]. Analysis of molecular variance (AMOVA) and the Shannon-Wiener diversity index (H) were computed using poppr version 2.9.5 [44]. Principal component analysis (PCA) and pairwise F_{ST} were performed using the SNPRelate package version 1.36.0 [45] in R.

3 Results and Discussion

3.1 *DNA extraction and Sequence analysis*

Hemp leaves collected from the Doi Lo plantation and transported under cold conditions to the laboratory exhibited signs of cold damage, where leaf tissue turned brown in areas directly exposed to ice. Such physical symptoms suggest underlying cellular stress, which can negatively impact molecular integrity. Rapid temperature fluctuations and frostbite can compromise DNA quality in plant tissues by disrupting cellular structures and inducing oxidative stress [46], [47]. These stress responses often result in DNA fragmentation or chemical modification, ultimately affecting the success of downstream genetic analyses. Sequence comparison among all accessions from DArTseq and whole-genome sequencing identified 3,609 SNPs. Most SNPs were mapped to chromosome 2 (12.86% or 464 SNPs), followed by an approximately equal distribution across chromosomes 1, 10, and 4 (10.20–11.86%), with lower distributions across the remaining chromosomes (7.8–9.8%).

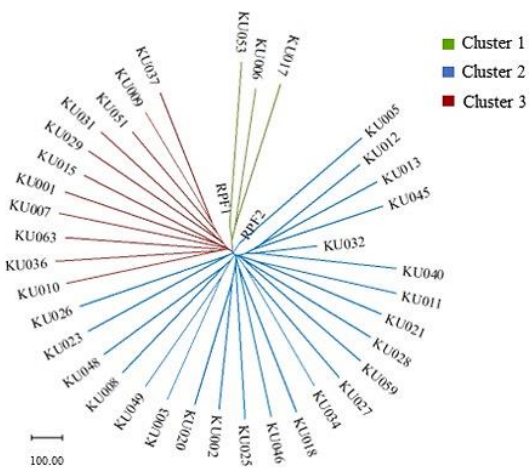

Figure 2: Dendrogram of 37 hemp accessions and two reference varieties constructed using the neighbor-joining method. The colors of the branches illustrate the clustering of hemp accessions.

A dendrogram constructed using the neighbor-joining method (Figure 2) revealed three distinct clades. However, the grouping did not correspond to the accessions' geographical origins. This suggests the presence of gene flow among populations, as accessions from Chiang Mai, Tak, and Mae Hong Son were distributed across all clades. The control variety RPF1, located in clade 1, exhibited the highest genetic similarity to KU053 from Mae Hong Son province, yet also formed a sister group with KU006 and KU017 from Chiang Mai and Tak, respectively. Conversely, RPF2 shared the highest similarity with KU005 from Chiang Mai in Cluster 2, which contained accessions from all three provinces. The clustering pattern, supported by bootstrap values exceeding 90% at multiple nodes, indicates the robustness of the three identified clades (bootstrap values not shown in Figure 2). Bootstrap support values are widely used to evaluate the reliability of inferred clades in phylogenetic trees. In general, values above 95% indicate strong support, values between 70–95% indicate moderate support, and values below 70% suggest weak support [48]. In this study, relationships between accessions supported by bootstrap values ≥70% were considered acceptable for interpreting sample clustering.

3.2 *Genetic structure*

The hemp accessions were grouped into populations based on STRUCTURE analysis using 3,609 SNPs. The highest Δ*K* value at *K*=3 indicated the presence of three genetically distinct subpopulations, which aligned well with the dendrogram (Figure 3). Each subpopulation was designated as Cluster 1, Cluster 2, and Cluster 3. Cluster 1 comprised 11 accessions (15.39%), Cluster 2 included 22 accessions (56.41%), and Cluster 3 contained 11 accessions (28.20%), as shown in Figure 4. Compared to the dendrogram, the only discrepancies were observed in the RPF2 and KU005 accessions, which were assigned to Cluster 1 according to STRUCTURE analysis but were placed in Cluster 2 in the dendrogram.

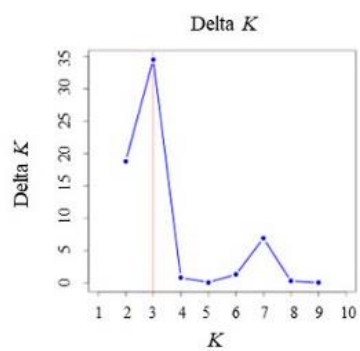

Figure 3: Population structure of 37 hemp accessions and two reference varieties. The number of subpopulations is indicated by the highest Δ*K* value.

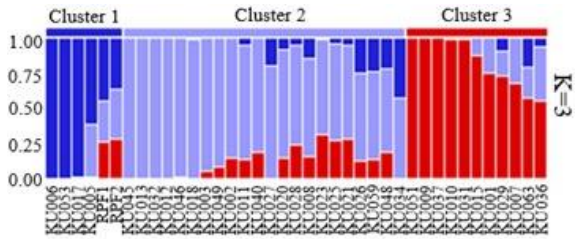

Figure 4: Population structure of 37 hemp accessions and two reference varieties. The proportion of individuals corresponding to each cluster is shown.

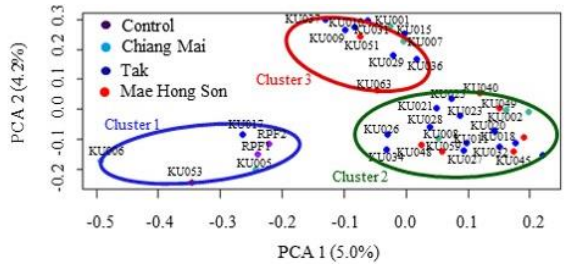

Figure 5: Principal component analysis (PCA) of 37 hemp accessions and two reference varieties using DArTseq-based SNPs. Colored dots represent hemp samples, and colored oval shapes indicate clustering.

PCA illustrated clear separation into three distinct groups (Figure 5), which corresponded well with the STRUCTURE analysis, further supporting significant genetic divergence among clusters. The observed clustering patterns in both PCA and STRUCTURE likely reflect underlying biological and anthropogenic factors. The high level of admixture in Cluster 2 suggests extensive gene flow, which may result from local seed exchange or unintentional hybridization during cultivation. In contrast, Clusters 1 and 3 appeared more genetically distinct, which could reflect lineage conservation, localized selection, or reduced gene flow due to geographic or agronomic isolation. The inclusion of reference varieties within Cluster 1 further supports the possibility that this group represents more conserved or historically important genotypes. The lack of clear geographic separation among clusters indicates that human activity, particularly seed distribution practices, may play a more influential role in shaping genetic structure than geography alone [49], [50].

Table 1: Pairwise genetic differentiation between populations based on STRUCTURE analysis.

Cluster Pair	F_{ST}	Nei's G_{ST}	Nm
Cluster 1 and Cluster 2	0.079	0.078	2.951
Cluster 1 and Cluster 3	0.079	0.058	4.086
Cluster 2 and Cluster 3	0.067	0.067	3.504

F_{ST} = fixation index; Nei's G_{ST} = Nei's coefficient of gene differentiation; Nm = gene flow

The F-statistic is a useful metric for inferring genetic diversity within and between populations. The fixation index (F_{ST}), along with Nei's G_{ST}, is commonly used to estimate genetic differentiation, where a value close to 0 indicates minimal genetic differentiation, while a value close to 1 suggests very high genetic differentiation. [51], [52]. Pairwise comparisons are presented in Table 1. Both F_{ST} and Nei's G_{ST} values were relatively similar for each pair of clusters. The F_{ST} values ranged from 0.067–0.079, while Nei's G_{ST} values ranged from 0.058–0.078, indicating low to moderate differentiation. Gene flow (Nm) is another important parameter for measuring the level of genetic exchange between populations. A value below 1 suggests low gene flow, whereas a value greater than 1 indicates high gene flow [53]. In this study, high Nm values (ranging between 2.951-4.086) were observed between clusters (Table 1). Thus, the hemp accessions analyzed in this study were not genetically isolated, indicating that genetic exchange has occurred. Hemp dispersal likely occurs through the transport of pollen and seeds between populations, or over longer distances via humans, other animals, including birds [54], [55].

The AMOVA results indicated significant structure clustering (*p*-value = 0.01-0.04) and low genetic variation between populations (4.83%) as shown in Table 2. The high within-sample genetic variation (94.85%) suggested that most genetic diversity was present within individuals rather than between populations. This finding aligned with previous studies on Moroccan cannabis germplasm, which used SSR markers and reported that genetic differentiation accounted for only 16% of the total molecular variance. The greatest genetic variability was observed within individuals (65%), which was attributed to high heterozygosity among individuals [56].

Table 2: AMOVA results based on genetic clusters inferred from STRUCTURE analysis.

ΔK=3	Df	Mean Sq	Variance	%	Phi Statistic	*p*-value
Between Pops	2	918.96	29.48	4.83	0.052	0.01
Between samples Within Pops	7	585.16	1.95	0.32	0.003	0.04
Within-samples	29	578.36	578.36	94.85	0.048	0.01
Total	38	597.54	609.79	100.00		

Df = degrees of freedom; Mean Sq = mean square

3.3 *Genetic diversity*

The genetic diversity of hemp across the three clusters was evaluated using various indices, including H, H_E, H_O, F_{IS}, and F_{IT}, as shown in Table 3. Cluster 3 exhibited the highest genetic diversity, with an H value of 3.090, indicating a more genetically variable population compared to Clusters 1 and 2. H_E reflects the potential genetic variability under ideal conditions, ranging from 0.324 to 0.383 across clusters. H_O, measuring the actual observed variation, was significantly lower (0.120 to 0.245), indicating a deficiency in heterozygosity within all clusters. This deficit may result from inbreeding or historical genetic drift.

The large differences between H_E and H_O in Clusters 2 and 3, combined with high F_{IS} values (0.939 and 0.957, respectively), suggest reduced genetic variation, possibly due to limited gene flow or the repeated use of genetically similar parental lines. From a breeding perspective, this may constrain the development of improved cultivars and reduce population adaptability. For cannabis users, particularly in medical or industrial applications, such genetic limitations may lead to variability in cannabinoid composition, flowering time, or fiber characteristics. Preserving adequate heterozygosity is therefore essential to ensure the stability, uniformity, and overall quality of hemp-derived products.

Although F_{IS} values indicate non-random mating and a trend toward homozygosity, likely driven by selection for desirable traits (e.g., cannabinoid content or disease resistance), the low F_{IT} value (0.033) suggests that overall genetic diversity remains considerable at the population level. Selective breeding may explain the formation of distinct genetic clusters, yet substantial variation is still preserved across the broader hemp gene pool.

Table 3: Genetic diversity indices of hemp populations inferred from STRUCTURE analysis.

Cluster	N	H	H_E	H_O	F_{IS}	F_{IT}
Cluster 1	6	1.790	0.383	0.245	0.285	-
Cluster 2	11	2.400	0.333	0.017	0.939	-
Cluster 3	22	3.090	0.324	0.012	0.957	-
Total	39	3.660	0.348	0.092	0.737	0.033

4 Conclusions

This study collected and analyzed 37 hemp accessions from three provinces in northern Thailand. Sequence analysis, conducted alongside two reference varieties, identified 3,609 SNPs, with the majority mapped to chromosome 2. STRUCTURE analysis and PCA confirmed the presence of three genetically distinct clusters, consistent with the dendrogram constructed using MEGA11.

Genetic differentiation among clusters, based on F_{ST} and Nei's G_{ST} values, indicated low to moderate divergence, while the high gene flow (Nm) suggests continued genetic exchange. Low H_O and high F_{IS} values pointed to heterozygote deficiency and inbreeding, although the overall population retained considerable diversity (low F_{IT}). These patterns reflect the influence of selective breeding and potentially limited genetic input.

The results provide a foundation for breeding programs aimed at improving economically valuable traits. SNPs identified in this research can be used for plant selections in future programs (Table S2). While hemp is not currently under threat in Thailand, proactive conservation is critical to maintain genetic diversity and prevent future erosion due to the hybridization of unregulated expansion [20], [22].

Furthermore, the sustainable use of cannabis resources should be guided by appropriate policy and regulation to prevent unintended consequences for public health and the environment [57], [58].

Acknowledgments

We thank Kasetsart University Research and Development Institute (KURDI) for supporting grant number FF(KU)10.64, 2021. We are grateful for the guidance from Dr. Pumipat Tongyoo and Dr. Passorn Wonnapinij.

Author Contributions

K.A.: investigation, data analysis, writing—original draft; S.S.: conceptualization, supervision, funding acquisition, project administration, research design,

writing—reviewing and editing; T.C.: conceptualization, funding acquisition, sample collection; W.P.: writing—reviewing and editing; P.J.: writing—reviewing and editing. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] A. Pollio, "The name of *Cannabis*: A short guide for nonbotanists," *Cannabis and Cannabinoid Research*, vol. 1, no. 1, pp. 234–238, 2016.

[2] A. C. Zeven and J. M. De Wet, *Dictionary of Cultivated Plants and Their Regions of Diversity: Excluding Most Ornamentals, Forest Trees and Lower Plants*. Wageningen, Netherlands: Pudoc, 1982.

[3] R. G. Pertwee, *Handbook of Cannabis*. Oxford: Oxford University Press, 2014.

[4] G. Grassi and J. M. McPartland, "Chemical and morphological phenotypes in breeding of *Cannabis sativa* L.," in *Cannabis sativa L. Botany and Biotechnology*, pp. 137–160, 2017.

[5] J. A. Toth, G. M. Stack, A. R. Cala, C. H. Carlson, R. L. Wilk, J. L. Crawford, D. R. Viands, G. Philippe, C. D. Smart, and J. K. C. Rose, "Development and validation of genetic markers for sex and cannabinoid chemotype in *Cannabis sativa* L.," *GCB Bioenergy*, vol. 12, no. 3, pp. 213–222, 2020.

[6] J. Teitelbaum, "A hemp oil, CBD, and marijuana primer: powerful pain, insomnia, and anxiety-relieving tools," *Alternative Therapies in Health and Medicine*, vol. 25, pp. 21–23, 2019.

[7] D. R. Blake, P. Robson, M. Ho, R. W. Jubb, and C. S. McCabe, "Preliminary assessment of the efficacy, tolerability and safety of a cannabis-based medicine (Sativex) in the treatment of pain caused by rheumatoid arthritis," *Rheumatology (Oxford, England)*, vol. 45, no. 1, pp. 50–52, 2006, doi: 10.1093/rheumatology/kei183.

[8] C. Cao, Y. Li, H. Liu, G. Bai, J. Mayl, X. Lin, K. Sutherland, N. Nabar, and J. Cai, "The potential therapeutic effects of THC on Alzheimer's disease," *Journal of Alzheimer's Disease*, vol. 42, no. 3, pp. 973–984, 2014.

[9] A. Passani, C. Posarelli, A. T. Sframeli, L. Perciballi, M. Pellegrini, G. Guidi, and M. Figus, "Cannabinoids in glaucoma patients: The never-ending story," *Journal of Clinical Medicine*, vol. 9, no. 12, p. 3978, 2020.

[10] S. Assanangkornchai, K. Thaikla, M. Talek, and D. Saingam, "Medical cannabis use in Thailand after its legalization: A respondent-driven sample survey," *PeerJ*, vol. 10, p. e12809, 2022, doi: 10.7717/peerj.12809.

[11] M. Laohavanich, "Political economy of cannabis in Thailand," *Chulalongkorn Medical Journal*, vol. 66, no. 1, Jul. 2023, doi: 10.58837/CHULA.CMJ.66.1.11.

[12] M. A. Nadeem, M. A. Nawaz, M. Q. Shahid, Y. Doğan, G. Comertpay, M. Yıldız, and F. S. Baloch, "DNA molecular markers in plant breeding: Current status and recent advancements in genomic selection and genome editing," *Biotechnology & Biotechnological Equipment*, vol. 32, no. 2, pp. 261–285, 2017, doi: 10.1080/13102818.2017.1400401.

[13] D. Vlk and J. Řepková, "Application of next-generation sequencing in plant breeding," *Czech Journal of Genetics & Plant Breeding*, vol. 53, no. 3, pp. 89–96, 2017.

[14] M. Agarwal, N. Shrivastava, and H. Padh, "Advances in molecular marker techniques and their applications in plant sciences," *Plant Cell Reports*, vol. 27, pp. 617–631, 2008.

[15] A. Karp, O. Seberg, and M. Buiatti, "Molecular techniques in the assessment of botanical diversity," *Annals of Botany*, vol. 78, no. 2, pp. 143–149, 1996.

[16] L. Mondini, A. Noorani, and M. A. Pagnotta, "Assessing plant genetic diversity by molecular tools," *Diversity*, vol. 1, no. 1, pp. 19–35, 2009.

[17] K. Semagn, Å. Bjørnstad, and M. N. Ndjiondjop, "An overview of molecular marker methods for plants," *African Journal of Biotechnology*, vol. 5, no. 25, pp. 2540–2568, 2006.

[18] D. Jaccoud, K. Peng, D. Feinstein, and A. Kilian, "Diversity arrays: A solid state technology for sequence information independent genotyping," *Nucleic Acids Research*, vol. 29, no. 4, p. E25, 2001, doi: 10.1093/nar/29.4.e25.

[19] T. Gülck and B. L. Møller, "Phytocannabinoids: Origins and biosynthesis," *Trends in Plant Science*, vol. 25, no. 10, pp. 985–1004, 2020. doi: 10.1016/j.tplants.2020.06.010.

[20] J. Sawler, J. M. Stout, K. M. Gardner, D. Hudson, J. Vidmar, L. Butler, J. E. Page, and S. Myles, "The genetic structure of marijuana and hemp," *PLOS ONE*, vol. 10, no. 8, Aug. 2015, Art. no. e0133292, doi: 10.1371/journal.pone.0133292.

[21] C. Gao, P. Xin, C. Cheng, Q. Tang, P. Chen, C. Wang, G. Zang, and L. Zhao, "Diversity analysis in *Cannabis sativa* based on large-scale development of expressed sequence tag-derived simple sequence repeat markers," *PLOS ONE*, vol. 9, no. 10, 2014, Art. no. e110638, doi: 10.1371/journal.pone.0110638.

[22] D. Vergara, C. Feathers, E. L. Huscher, B. Holmes, J. A. Haas, and N. C. Kane, "Widely assumed phenotypic associations in *Cannabis sativa* lack a shared genetic basis," *PeerJ*, vol. 9, 2021, Art. no. e10672, doi: 10.7717/peerj.10672.

[23] D. A. Hoisington and R. K. Varshney, "Diversity Arrays Technology (DArT) low cost, generic molecular markers for breeding and research," *Biotech News*, vol. 2, no. 1, pp. 6–8, 2007.

[24] T. Chaisan, P. Thobunluepop, N. Thongthip, A. Rakpenthai, B. Puangsin, S. Samipak, and W. Pluempanupat, "Identification of morphological traits affecting high seed yield potential from new hemp germplasm collected in Thailand," *Chilean Journal of Agricultural Research*, vol. 85, no. 1, pp. 88–97, 2025. doi: 10.4067/S0718-58392025000100088.

[25] T. Saleepochn, S. Tharamak, S. Charanja, L. Pimsawang, T. Wisarutwanit, T. Chaisan, W. Songoen, and W. Pluempanupat, "Comparative fatty acid profiling of seed lines of hemp (*Cannabis sativa* L.) and inhibition of tyrosinase, α-glucosidase and acetylcholinesterase," *Agriculture and Natural Resources*, vol. 58, no. 5, pp. 547–554, 2024. doi: 10.34044/j.anres.2024.58.5.01.

[26] R. Rattanapakdee, P. Thobunluepop, S. Anusonpornperm, T. Chaisan, A. Pongtip, P. Maniin, W. Chitbanchong, and S. Gorinstein, "Effects of harvest time on medicinal qualities of hemp," *Journal of Current Science and Technology*, vol. 14, no. 3, Art. no. 63, 2024. doi: 10.59796/jcst.V14N3.2024.63.

[27] P. Chaowana, W. Hnoocham, S. Chaiprapat, P. Yimlamai, K. Chitbanyong, K. Wanitpinyo, T. Chaisan, Y. Paopun, S. Pisutpiched, S. Khantayanuwong, and B. Puangsin, "Utilization of hemp stalk as a potential resource for bioenergy," *Materials Science for Energy Technologies*, vol. 7, pp. 19–28, 2024. doi: 10.1016/j.mset.2023.07.001.

[28] F. S. Baloch, A. Alsaleh, M. Q. Shahid, V. Çiftçi, L. E. Sáenz de Miera, M. Aasim, M. A. Nadeem, H. Aktaş, H. Özkan, and R. Hatipoğlu, "A whole genome DArTseq and SNP analysis for genetic diversity assessment in durum wheat from Central Fertile Crescent," *PLOS ONE*, vol. 12, no. 1, 2017, Art. no. e0167821, doi: 10.1371/journal.pone.0167821.

[29] S. Andrews. "FastQC: A quality control tool for high throughput sequence data." bioinformatics. babraham.ac.uk. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed Mar. 17, 2025).

[30] A. M. Bolger, M. Lohse, and B. Usadel, "Trimmomatic: A flexible trimmer for Illumina sequence data," *Bioinformatics (Oxford, England)*, vol. 30, no. 15, pp. 2114–2120, 2014.

[31] H. Li and R. Durbin, "Fast and accurate short read alignment with Burrows-Wheeler transform," *Bioinformatics (Oxford, England)*, vol. 25, no. 14, pp. 1754–1760, 2009.

[32] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin, "The Sequence Alignment/Map format and SAMtools," *Bioinformatics (Oxford, England)*, vol. 25, no. 16, pp. 2078–2079, 2009.

[33] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo, "The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data," *Genome Research*, vol. 20, no. 9, pp. 1297–1303, 2010.

[34] P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, and H. Li, "Twelve years of SAMtools and BCFtools," *GigaScience*, vol. 10, no. 2, Feb. 2021, Art. no. giab008, doi: 10.1093/gigascience/giab008.

[35] P. Cingolani, A. Platts, L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and D. M. Ruden, "A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3," *Fly*, vol. 6, no. 2, pp. 80–92, 2012.

[36] W. J. Bruno, N. D. Socci, and A. L. Halpern, "Weighted neighbor joining: A likelihood-based approach to distance-based phylogeny reconstruction," *Molecular Biology and Evolution*, vol. 17, no. 1, pp. 189–197, 2000, doi: 10.1093/oxfordjournals.molbev.a026231.

[37] K. Tamura, G. Stecher, and S. Kumar, "MEGA11: Molecular evolutionary genetics analysis version 11," *Molecular Biology and Evolution*, vol. 38, no. 7, pp. 3022–3027, 2021, doi: 10.1093/molbev/msab120.

[38] L. Porras-Hurtado, Y. Ruiz, C. Santos, C. Phillips, A. Carracedo, and M. V. Lareu, "An overview of structure: Applications, parameter settings, and supporting software," *Frontiers in Genetics*, vol. 4, p. 98, 2013, doi: 10.3389/fgene.2013.00098.

[39] G. Evanno, S. Regnaut, and J. Goudet, "Detecting the number of clusters of individuals using the software Structure: A simulation study," *Molecular Ecology*, vol. 14, no. 8, pp. 2611–2620, 2005, doi: 10.1111/j.1365-294X.2005.02553.x.

[40] R. M. Francis, "pophelper: An R package and web app to analyse and visualize population structure," *Molecular Ecology Resources*, vol. 17, no. 1, pp. 27–32, 2017.

[41] T. Jombart and I. Ahmed, "adegenet 1.3-1: New tools for the analysis of genome-wide SNP data," *Bioinformatics (Oxford, England)*, vol. 27, no. 21, pp. 3070–3071, 2011, doi: 10.1093/bioinformatics/btr521.

[42] J. Goudet, "hierfstat, a package for R to compute and test hierarchical F-statistics," *Molecular Ecology Notes*, vol. 5, no. 1, pp. 184–186, Mar. 2005, doi: 10.1111/j.1471-8286.2004.00828.x.

[43] B. J. Knaus and N. J. Grünwald, "vcfr: A package to manipulate and visualize variant call format data in R," *Molecular Ecology Resources*, vol. 17, no. 1, pp. 44–53, 2017, doi: 10.1111/1755-0998.12549.

[44] Z. N. Kamvar, J. F. Tabima, and N. J. Grünwald, "Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction," *PeerJ*, vol. 2, p. e281, 2014, doi: 10.7717/peerj.281.

[45] X. Zheng, D. Levine, J. Shen, S. M. Gogarten, C. Laurie, and B. S. Weir, "A high-performance computing toolset for relatedness and principal component analysis of SNP data," *Bioinformatics (Oxford, England)*, vol. 28, no. 24, pp. 3326–3328, 2012, doi: 10.1093/bioinformatics/bts606.

[46] S.-H. Han, J. Y. Kim, J.-H. Lee, and C.-M. Park, "Safeguarding genome integrity under heat stress in plants," *Journal of Experimental Botany*, vol. 72, no. 21, pp. 7421–7435, Nov. 2021. doi: 10.1093/jxb/erab355.

[47] S. S. Gill and N. Tuteja, "Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants," *Plant Physiol. Biochem*, vol. 48, no. 12, pp. 909–930, 2010. doi: 10.1016/j.plaphy.2010.08.016.

[48] D. M. Hillis and J. J. Bull, "An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis," *Systematic biology*, vol. 42, no. 2, pp. 182–192, Jun. 1993. doi: 10.1093/sysbio/42.2.182.

[49] N. Patterson, A. L. Price, and D. Reich, "Population structure and eigenanalysis," *PLoS Genetics*, vol. 2, no. 12, p. e190, 2006. doi: 10.1371/journal.pgen.0020190.

[50] J. Novembre and M. Stephens, "Interpreting principal component analyses of spatial population genetic variation," *Nature genetics*, vol. 40, no. 5, pp. 646–649, 2008. doi: 10.1038/ng.139.

[51] Z. Luo, J. Brock, J. M. Dyer, T. Kutchan, D. Schachtman, M. Augustin, Y. Ge, N. Fahlgren, and H. Abdel-Haleem, "Genetic diversity and population structure of a *Camelina sativa* spring panel," *Frontiers in Plant Science*, vol. 10, p. 184, Feb. 2019, doi: 10.3389/fpls.2019.00184.

[52] S. J. G. Hall, "Genetic differentiation among livestock breeds-values for Fst," *Animals: an Open Access Journal from MDPI*, vol. 12, no. 9, p. 1115, 2022, doi: 10.3390/ani12091115.

[53] S. Wright, "The interpretation of population structure by F-statistics with special regard to systems of mating," *Evolution*, vol. 19, no. 3, pp. 395–420, Sep. 1965, doi: 10.1111/j.1558-5646.1965.tb01731.x.

[54] J. M. McPartland and S. G. Naraine, "Experimental endozoochory of *Cannabis sativa* achenes," *Medical Cannabis and Cannabinoids*, vol. 1, no. 2, pp. 96–103, 2018, doi: 10.1159/000492971.

[55] E. Small and T. Antle, "A preliminary study of pollen dispersal in *Cannabis sativa* in relation to wind direction," *Journal of Industrial Hemp*, vol. 8, no. 2, pp. 37–50, 2003, doi: 10.1300/J237v08n02_03.

[56] C. Benkirane, M. Charif, C. M. Müller, et al., "Population structure and genetic diversity of Moroccan cannabis (*Cannabis sativa* L.) germplasm through simple sequence repeat (SSR) analysis," *Genetic Resources and Crop Evolution*, vol. 71, pp. 2037–2051, 2024, doi: 10.1007/s10722-023-01754-x.

[57] World Health Organization. "Cannabis: Health and Regulation." who.int. https://www.who.int. (accessed Mar. 17, 2025).

[58] United Nations Office on Drugs and Crime. "World Drug Report 2023." unodc.org. https://www.unodc.org. (accessed Mar. 17, 2025).

Table S1: Origin of the 37 hemp accessions collected from the North of Thailand

No.	Name	Place	District	Province	No.	Name	Place	District	Province
1	KU001	Samoeng Tai	Samoeng	Chiang Mai	20	KU026	Mae Sot	Mae Sot	Tak
2	KU002	Samoeng Tai	Samoeng	Chiang Mai	21	KU027	Tha Sai Luat	Mae Sot	Tak
3	KU003	Samoeng Tai	Samoeng	Chiang Mai	22	KU028	Tha Sai Luat	Mae Sot	Tak
4	KU005	Samoeng North	Samoeng	Chiang Mai	23	KU029	Mae Tao	Mae Sot	Tak
5	KU006	Samoeng North	Samoeng	Chiang Mai	24	KU031	Mae Tao	Mae Sot	Tak
6	KU007	Samoeng North	Samoeng	Chiang Mai	25	KU032	Mae Tuen	Mae Ramat	Tak
7	KU008	Samoeng North	Samoeng	Chiang Mai	26	KU034	Mae Tuen	Mae Ramat	Tak
8	KU009	Mae Ramat	Mae Ramat	Tak	27	KU036	Mae Tuen	Mae Ramat	Tak
9	KU010	Mae Charao	Mae Ramat	Tak	28	KU037	Mae Tuen	Mae Ramat	Tak
10	KU011	Mae Tuen	Mae Ramat	Tak	29	KU040	Khun Yuam	Khun Yuam	Mae Hong Son
11	KU012	Sam Muen	Mae Ramat	Tak	30	KU045	Mae Yuam Noi	Khun Yuam	Mae Hong Son
12	KU013	Phra That	Mae Ramat	Tak	31	KU046	Mueang Pon	Khun Yuam	Mae Hong Son
13	KU015	Khiri Rat	Phop Phra	Tak	32	KU048	Mueang Pon	Khun Yuam	Mae Hong Son
14	KU017	Phop Phra	Phop Phra	Tak	33	KU049	Mae Ngao	Khun Yuam	Mae Hong Son
15	KU018	Phop Phra	Phop Phra	Tak	34	KU051	Sop Moei	Sop Moei	Mae Hong Son
16	KU020	Chong Khaep	Phop Phra	Tak	35	KU053	Mae Yuam	Mae Sariang	Mae Hong Son
17	KU021	Chong Khaep	Phop Phra	Tak	36	KU059	Huai Pha	Mueang	Mae Hong Son
18	KU023	Na Bot	Wang Chao	Tak	37	KU063	Pang Mapha	Pang Mapha	Mae Hong Son
19	KU025	Mae Sot	Mae Sot	Tak					

Table S2: Significant SNPs were identified between clusters; (A) Cluster 1 and 2 had 23 positions, (B) Cluster 2 and 3 had 33 positions, (C) Cluster 1 and 3 had 7 positions. REF: in reference genome, ALT: in accessions.

(A) Significant SNPs between Cluster 1 and 2

No.	Chromosome	Position	REF	ALT	No.	Chromosome	Position	REF	ALT
1	Chr1	68562436	G	C	13	Chr6	16236114	C	G
2	Chr1	98654960	G	A	14	Chr6	17889052	A	G
3	Chr4	6117913	A	G	15	Chr7	63538013	T	A
4	Chr4	16644678	A	C	16	Chr7	65773279	T	C
5	Chr4	16646808	T	C	17	Chr8	56035484	A	G
6	Chr4	16646838	G	A	18	Chr9	9211379	T	C
7	Chr4	16646871	A	T	19	Chr9	16173299	C	A
8	Chr4	16646889	G	A	20	Chr9	16173302	C	T
9	Chr4	39578448	T	C	21	Chr9	16173356	A	G
10	Chr4	39578461	A	G	22	Chr9	16173365	G	A
11	Chr4	39578475	T	C	23	ChrX	9368698	G	A
12	Chr4	64913004	T	C					

(B) Significant SNPs between Cluster 2 and 3

No.	Chromosome	Position	REF	ALT	No.	Chromosome	Position	REF	ALT
1	Chr2	8926327	C	T	18	Chr8	47416709	G	A
2	Chr2	8926333	C	T	19	Chr8	47912299	T	C
3	Chr2	82381590	G	C	20	Chr8	48686648	C	G
4	Chr3	1143943	G	T	21	Chr8	57814538	C	T
5	Chr3	94175626	T	C	22	Chr8	58998798	G	C
6	Chr4	3202367	C	T	23	Chr9	23280070	A	T
7	Chr5	27020508	T	A	24	Chr9	32019340	G	A
8	Chr5	27020526	T	A	25	Chr9	32019358	T	G
9	Chr6	1689840	C	T	26	Chr9	56095392	A	G
10	Chr6	3615222	G	A	27	Chr9	56095431	A	G
11	Chr6	3615224	G	A	28	Chr9	56095441	C	T
12	Chr6	54834092	C	T	29	ChrX	20856648	C	T
13	Chr6	74013034	C	T	30	ChrX	58858194	G	C
14	Chr6	74013094	C	T	31	ChrX	99186616	C	T
15	Chr6	74271333	A	G	32	ChrX	103949415	G	A
16	Chr7	1714320	T	C	33	ChrX	104310035	T	C
17	Chr7	30896330	A	T					

(C) Significant SNPs between Clusters 1 and 3

No.	Chromosome	Position	REF	ALT
1	Chr3	1049343	T	A
2	Chr3	1049352	G	C
3	Chr3	1049356	T	A
4	Chr8	13842404	A	G
5	Chr8	13842440	C	T
6	ChrX	20856648	C	T
7	ChrX	41760755	C	G