

Whole-genome sequencing and SNP analysis of Thai *Cannabis sativa* cultivar ‘Hang Kra Rog Phu Phan’ (Cannabaceae)

THAVATCHAI KAMOLTHAM^{1,*}, NOPHADON LUANGPIROM¹, NAPAPORN KUAMSAB²,
TANAWAN KUMMALUE³, TANAWAT CHAIPHONGPACHARA⁴

¹Department of Cannabis Health Sciences, College of Allied Health Sciences, Suan Sunandha Rajabhat University. 111/1-3 Rama 2 St., Samut Songkhram 75000, Thailand. Tel.: +66-34-773-905, *email: thavatchai.ka@ssru.ac.th

²Faculty of Health Science and Technology, Community Public Health Program, Southern College of Technology. 124/1 Thungsong-Huayyod St., Thung Song, Nakhon Si Thammarat 80110, Thailand

³Faculty of Medicine Siriraj Hospital, Mahidol University. 2 Wang Lang St., Bangkok Noi, Bangkok 10700, Thailand

⁴Department of Public Health and Health Promotion, College of Allied Health Sciences, Suan Sunandha Rajabhat University. 111/1-3 Rama 2 St., Samut Songkhram 75000, Thailand

Manuscript received: 18 June 2025. Revision accepted: 4 October 2025.

Abstract. Kamoltham T, Luangpirom N, Kuamsab N, Kummalue T, Chaiphongpachara T. 2025. Whole-genome sequencing and SNP analysis of Thai *Cannabis sativa* cultivar ‘Hang Kra Rog Phu Phan’ (Cannabaceae). *Biodiversitas* 26: 4946-4953. *Cannabis* (*Cannabis sativa*) is a herbaceous plant valued for its medicinal and therapeutic uses. In Thailand, the indigenous cultivar ‘Hang Kra Rog Phu Phan’ has long been applied in traditional medicine and is recognized for its high Tetrahydrocannabinol (THC) content. This study provides the first genomic characterization of this traditional Thai cultivar using Whole-Genome Sequencing (WGS). Fresh leaf samples were collected from a licensed cultivation site, and paired-end libraries were prepared with the Illumina TruSeq DNA PCR-Free Kit. Sequencing on the NovaSeq 6000 platform produced 63.8 million raw reads (9.57 Gb), yielding 63.3 million high-quality reads (9.50 Gb) after trimming. Clean reads showed a 94.43% alignment rate to the ‘Pink Pepper’ reference genome, with an average depth of 11.39×. Variant calling identified 23.0 million genomic variants, including 18.5 million SNPs and 4.5 million Indels, with 6.04 million high-confidence SNPs retained after stringent filtering. Phylogenetic and principal component analyses revealed unexpected genomic proximity between ‘Hang Kra Rog Phu Phan’ and the CBD-dominant ‘CBD Shark’ cultivar, highlighting its distinct lineage among high-THC cultivars. These findings provide valuable genomic resources for precise cultivar authentication, marker-assisted breeding, conservation of native Thai germplasm, and functional genomics to advance cannabis-based therapeutics.

Keywords: *Cannabis sativa* cultivar Hang Kra Rog Phu Phan, SNP analysis, Thailand, whole-genome sequencing

INTRODUCTION

Cannabis (*Cannabis sativa* L., Cannabaceae) is an herbaceous species with broad applications in medicine, agriculture, and biotechnology. Originating from Central and Southeast Asia, it has been cultivated for over 5,000 years for fiber, food, and therapeutic purposes (Adams et al. 2021). The cannabis plant is characterized by extensive phytochemical diversity, including >550 compounds, such as cannabinoids, flavonoids, and terpenes, which contribute to its insecticidal, fungicidal, and pharmacological potential (Hussain et al. 2021; Fordjour et al. 2023; El Oihabi et al. 2024). Among its bioactive molecules, Cannabidiol (CBD) and Δ^9 -Tetrahydrocannabinol (THC) are particularly relevant due to their diverse therapeutic and psychoactive properties.

Environmental factors including climate, soil, and geography strongly influence genotypic and chemotypic variation of *C. sativa* (Chen et al. 2024; Younas et al. 2024). As regional conditions shape phenotypic traits, genomic studies of local cultivars are essential for optimizing cultivation and selecting varieties with desired properties (Balant et al. 2022; Chen et al. 2022). Recent advances in genomics, especially the Whole-Genome Sequencing (WGS) and Single Nucleotide Polymorphism (SNP) analysis have

transformed cannabis research by enabling exploration of genetic diversity, trait-associated polymorphisms, and evolutionary relationships (Sawler et al. 2015; Weiblen et al. 2015). *C. sativa* is diploid with 9 autosomes and XY sex chromosomes, comprising an estimated genome size of ~808-900 Mb (Shujat et al. 2024). Nevertheless, indigenous cultivars remain underrepresented in genomic datasets, creating challenges in cultivar authentication, standardized breeding, and medical cannabis development (Hurgobin et al. 2021). SNP variation has been shown to occur in genes regulating cannabinoid biosynthesis, including THCAS, CBDAS, and Olivetolic Acid Cyclase (OAC). Such polymorphisms influence THC/CBD ratios and overall chemotype diversity (Hurgobin et al. 2021). Variants enriched in transcription factor families, such as WRKY, MYB, and bHLH further suggest roles in stress adaptation, pathogen resistance, and heat tolerance (Fordjour et al. 2023). These findings underscore the importance of expanding genomic studies to local landraces, which may harbor unique alleles contributing to both phytochemical and adaptive traits.

Thailand has a long history of cannabis cultivation, with diverse landraces adapted to ecological gradients (Woods et al. 2022; Dal Martello et al. 2024). In recent years, five cultivars have been officially registered with the Department

of Agriculture, Ministry of Agriculture and Cooperatives, namely Issara 01 (*C. sativa* cultivar ‘Issara 01’), Hang Kra Rog Phu Phan ST1 (*C. sativa* cultivar ‘Hang Kra Rog Phu Phan ST1’), Hang Suea Sakonnakhon TT1 (*C. sativa* cultivar ‘Hang Suea Sakonnakhon TT1’), Tanao Si Kan Khaw WA1 (*C. sativa* cultivar ‘Tanao Si Kan Khaw WA1’), and Tanao Si Kan Dang RD1 (*C. sativa* cultivar ‘Tanao Si Kan Dang RD1’). Among them, *C. sativa* ‘Hang Kra Rog Phu Phan’ is notable for its distinctive squirrel-tail-shaped inflorescence, high THC content, and long-standing use in traditional Thai medicine (Sunanta et al. 2024). Despite its cultural and agricultural significance, this *C. sativa* cultivar in Thailand remains genetically uncharacterized, limiting its integration into global cannabis genomics. Comparative genomics with established references, such as the ‘Pink Pepper’ cultivar (Grassa et al. 2021), provides an opportunity to assess genetic similarity, discover functional SNPs, and support conservation and breeding. Profiling native cultivars like ‘Hang Kra Rog Phu Phan’ is not only vital for preserving genetic biodiversity but also for identifying markers associated with cannabinoid biosynthesis, stress adaptation, and agronomic performance.

This study presents the first genomic characterization of *C. sativa* ‘Hang Kra Rog Phu Phan’ using WGS and SNP profiling. The specific objectives are to evaluate genomic similarity with the reference genome ‘Pink Pepper,’ identify and annotate genome-wide SNPs, and analyze genetic relatedness and diversity using phylogenetic and principal component analysis. The results are expected to contribute to cultivar authentication, marker-assisted breeding, and conservation of native Thai cannabis germplasm, while providing valuable genomic resources for medical and agricultural applications.

MATERIALS AND METHODS

Plant materials

Fresh leaf samples of *C. sativa* cultivar ‘Hang Kra Rog Phu Phan’ were collected from a licensed cultivation site authorized by the Department of Thai Traditional and Alternative Medicine, Thailand. In total, three independent plants were sampled and approximately 100 mg of leaf tissue from each individual was subjected to DNA extraction. Immediately after collection, the samples were flash-frozen in liquid nitrogen and stored at -80°C to preserve genetic integrity.

DNA extraction

Genomic DNA was extracted using a modified version of the Cetyltrimethylammonium Bromide (CTAB) method (Doyle and Doyle 1987), with slight adjustments to enhance DNA purity and yield. Approximately 100 mg of leaf tissue was ground in liquid nitrogen and incubated at 65°C for 30 minutes in a CTAB extraction buffer containing 2% CTAB, 100 mM Tris-HCl (pH 8.0), 20 mM EDTA, 1.4 M NaCl , and 2% β -mercaptoethanol. DNA was subsequently purified using chloroform:isoamyl alcohol (24:1) extraction, followed by isopropanol precipitation. The resulting DNA pellet was washed with 70% ethanol, air-dried, and resuspended in TE buffer. The quality and concentration of

the extracted DNA were assessed using a Nanodrop spectrophotometer, a Qubit fluorometer, and agarose gel electrophoresis to ensure that the DNA was of high quality for downstream analysis.

Whole-genome sequencing

To enhance reproducibility and transparency, a summary workflow diagram outlining the sequencing, data processing, and analysis steps is provided (Figure 1). Whole-genome sequencing of *C. sativa* ‘Hang Kra Rog Phu Phan’ was performed using the Illumina NovaSeq 6000 platform. Paired-end libraries were prepared using the Illumina TruSeq DNA Library Prep Kit according to the manufacturer’s protocol. Paired-end libraries (150 bp read length) were constructed using the Illumina TruSeq DNA PCR-Free Library Prep Kit, following the manufacturer’s instructions. The raw sequencing output yielded a total of 150 bp paired-end reads with high base quality suitable for downstream analyses. Initial quality assessment of raw reads was conducted using FastQC v0.11.9 to evaluate per-base quality scores, GC content, and adapter contamination (Bolger et al. 2014). Adapter trimming and quality filtering were carried out using Trimmomatic with the following parameters: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15MINLEN:36. These settings ensured the removal of Illumina adapter sequences, low-quality bases at read ends, and reads shorter than 36 bp. Following trimming, reads were re-evaluated with FastQC to confirm quality improvement. Only clean and high-quality reads passing all filters were retained. The final dataset consisted of approximately 9.50 Gb of clean sequence data, which were then used for genome alignment and variant discovery. This corresponds to an average genome coverage of $\sim 11.4\times$ based on the reference genome size of approximately 820 Mb.

Reference genome and sequence alignment

Raw reads were trimmed and filtered using the Trimmomatic to remove adapters and low-quality bases. Cleaned reads were aligned to the reference genome of *C. sativa* cultivar ‘Pink Pepper’ (GenBank assembly: GCA_029168945.1) using BWA-MEM. Post-alignment processing included evaluation with SAMtools (Danecek et al. 2021) and removal of duplicate reads with Picard tools to minimize bias and ensure data integrity.

SNP calling and variant filtering

Variants were identified using GATK v4.2.0.0 following the Best Practices pipeline (McKenna et al. 2010). SNP calling was conducted with the Haplotype Caller in GVCF mode, followed by merging with CombineGVCFs and joint genotyping using Genotype GVCFs. Hard filtering was applied using the Variant Filtration tool with the following thresholds: Quality by Depth (QD) <2.0 , Fisher Strand bias (FS) >60.0 , Mapping Quality (MQ) <40.0 , mapping quality rank sum (MQRank Sum) <-12.5 , and read position rank sum (ReadPosRank Sum) <-8.0 . Further refinement with VCFtools v0.1.16 retained only biallelic SNPs, excluded variants with missing genotype data, and filtered out those with Minor Allele Frequency (MAF) <0.05 .

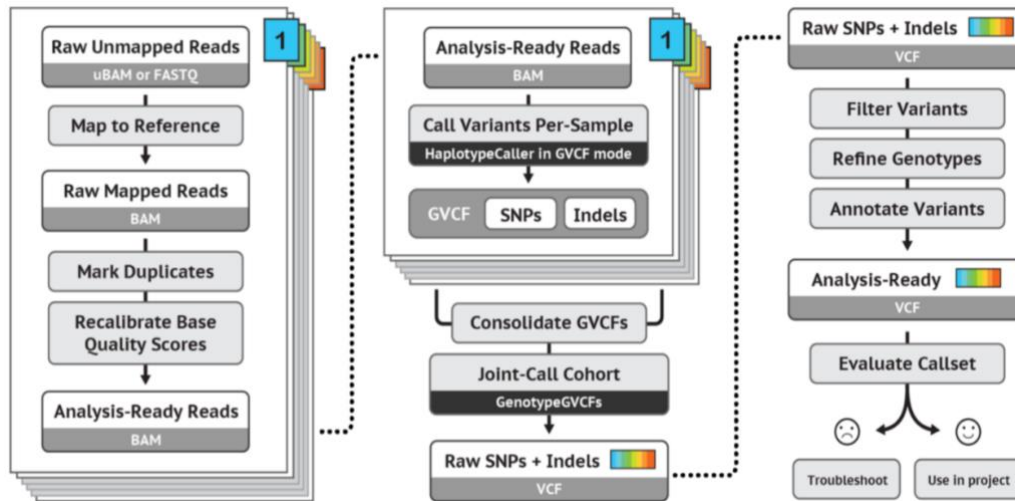


Figure 1. Workflow diagram illustrating the Genome Analysis Toolkit (GATK) Best Practices pipeline used for whole-genome SNP and indel discovery in *Cannabis sativa* 'Hang Kra Rog Phu Phan'

Data analysis

The high-confidence SNP dataset obtained after filtering was used for downstream analyses to investigate genetic diversity, population structure, and phylogenetic relationships among *C. sativa* cultivars. A phylogenetic tree was constructed using the neighbor-joining method in MEGA X software, incorporating SNP data from the 'Hang Kra Rog Phu Phan' cultivar and 21 other cannabis samples obtained from the NCBI database (Kumar et al. 2018). Principal Component Analysis (PCA) was conducted using the SNPRelate package in R to capture major patterns of genetic variation and visualize population structure. This provided insights into the genetic diversity and evolutionary dynamics of the studied cultivars.

RESULTS AND DISCUSSION

Genome sequencing and alignment quality

Genome-wide sequencing of *C. sativa* 'Hang Kra Rog Phu Phan' was conducted using WGS techniques. The initial

sequencing generated 63,775,668 raw reads. Following quality filtering to remove unreliable or ambiguous data, the number of clean reads was slightly reduced to 63,337,084 (Table 1). The initial total DNA sequence data amounted to 9.57 Gb, while 9.50 Gb of reliable, high-quality sequence data were retained after quality control. The overall alignment rate achieved was 94.43%, with an average read depth of 11.39x, reflecting robust coverage for further genomic analysis. The raw sequence data were aligned to the reference genome of the *C. sativa* 'Pink Pepper' cultivar (NCBI GenBank assembly: GCA_029168945.1). Detailed information on the reference genome is provided in Table 2. Post-alignment, the GATK suite was employed to identify SNPs across the genome. The whole-genome sequencing data of *C. sativa* 'Hang Kra Rog Phu Phan' generated in this study have been submitted to the NCBI Sequence Read Archive (SRA) under accession number SRR35341777.

Table 1. Summary of genetic data quality for the 'Hang Kra Rog Phu Phan' cultivar of *Cannabis sativa* using WGS

Metric	Value	Description/Unit
Total raw reads	63,775,668	Number of paired-end reads generated
Clean reads (after filtering)	63,337,084	Reads retained post quality trimming
Total base pairs (Gb)	9.50	Gigabases of cleaned sequence data
GC content (guanine/cytosine) (%)	34.52	Proportion of guanine and cytosine bases
Mean read length (bp)	150	Base pairs per read (billion pairs)
Genome size (estimated) (Mb)	~820	Estimated genome size of <i>Cannabis sativa</i>
Read alignment rate (%)	94.43	% of reads mapped to reference genome (Pink Pepper)
Mean genome coverage	11.39x	Average sequencing depth across the genome
N50 read length (bp)	149	Median read length such that 50% of reads are \geq n50
Q30 score (%)	92.30	% of bases with Phred quality score \geq 30
Total variants identified	23,025,604	Combined SNPs and INDELs before filtering
High-confidence SNPs retained	6,040,109	SNPs after quality and frequency filters (MAF \geq 0.05)

Table 2. Reference genome data for the cannabis cultivar ‘Pink Pepper’ obtained from the National Center for Biotechnology Information (NCBI) database

NCBI Ref Seq. assembly	GCF_029168945.1
Submitted GenBank assembly	GCA_029168945.1
Taxon	<i>Cannabis sativa</i>
Cultivar	Pink pepper
Isolate	KNU-18-1
WGS project	JAQSJK01
Assembly type	Haploid
Submitter	Kangwon National University
Date	14-Mar-23

SNP and indel discovery

Following alignment, genome-wide variant calling was performed using GATK software to identify SNPs and Indels. A comparative analysis with genomic data from 21 other cannabis cultivars (obtained from the NCBI database) revealed 23,025,604 variant positions, comprising 18,520,256 SNPs and 4,505,348 Indels (Table 3). The SNP dataset was further refined by selecting high-confidence SNPs with a minimum MAF of 5% and no missing data across samples. After stringent filtering, the number of retained SNP positions was reduced to 6,040,109, distributed across the 9 autosomal chromosomes and chromosome X (Figure 2). This refined dataset was then used for subsequent genetic analyses, including phylogenetic tree construction and PCA.

SNP distribution and genetic variation

A total of 6,040,109 SNPs were identified in the *C. sativa* ‘Hang Kra Rog Phu Phan’. These SNPs were distributed across the genome, with significant representation on both the autosomal chromosomes and chromosome X (Figure 2). The phylogenetic tree constructed from the SNP data showed that ‘Hang Kra Rog Phu Phan’ was genetically closest to the ‘CBD Shark’ cultivar (Figure 3). Additionally, PCA further confirmed the clustering pattern observed in the phylogenetic tree, reinforcing the genetic distinctiveness and relationships among these cannabis cultivars (Figure 4). The Thai landrace ‘Hang Kra Rog Phu Phan’ (shown in red) is positioned in close proximity to the ‘CBD Shark’ cultivar, suggesting a high degree of genetic similarity between these two accessions. Other cannabis cultivars, including industrial hemp varieties (e.g., ‘FINOLA’, ‘Felina32’) and accessions from the United States, form separate clusters, reflecting differences in genetic background. Principal components 1 and 2 (PC1 and PC2) explain 9.25% and 8.63% of the total genetic variation, respectively.

Discussion

Whole-genome sequencing analysis of *C. sativa* cultivar ‘Hang Kra Rog Phu Phan’ revealed a high genomic similarity of 94.43% to the well-characterized reference genome of the ‘Pink Pepper’ cultivar. This high degree of similarity indicates a close genetic relationship between the two cultivars, suggesting the presence of shared functional

genes and similar biochemical pathways. These findings are consistent with previous studies reporting genetic similarities among cannabis cultivars, which often share common metabolic pathways and gene families related to important traits, such as CBD production (Sawler et al. 2015; Weiblen et al. 2015). The high sequencing depth (11.39×) and robust alignment rate (94.43%) further validate the quality and completeness of the genomic data, establishing it as a reliable reference for future genomic studies on Thai cannabis.

The identification of over 6 million high-confidence SNPs underscores the substantial genetic diversity present in cannabis populations. This extensive SNP dataset is invaluable, providing a wealth of genomic information that will serve as a critical resource for future cannabis breeding programs. It lays the groundwork for cultivar optimization and genetic studies aiming to enhance desirable agronomic traits, such as disease resistance and yield, as well as therapeutic characteristics, including CBD and terpene profiles (Vergara et al. 2016). The large number of identified SNPs also reflects the complex and diverse genetic makeup of cannabis, which can be leveraged to accelerate genetic improvement for both medicinal and industrial applications.

In this study, a total of 6,040,109 SNPs were identified in the *C. sativa* cultivar ‘Hang Kra Rog Phu Phan’. This data is lower than the 9.3 million high-quality SNPs reported by Busta et al. (2022), who analyzed *C. sativa* samples collected from 17 counties across Nebraska, spanning a wide range of climatic conditions. The greater number of SNPs in the Nebraska study likely reflects the broader geographic and environmental diversity of the sampled populations, which contributes to increased genetic variability. By contrast, the ‘Hang Kra Rog Phu Phan’ cultivar, derived from a single geographic location in Thailand, may have undergone limited selective breeding, leading to reduced genetic diversity. Therefore, the comparatively lower SNP count observed in this study underscores differences in sampling breadth, environmental heterogeneity, and domestication history between the two datasets (Ren et al. 2021; Busta et al. 2022).

By comparison, wild-type *C. sativa* accessions from Tibet, China, represent a distinct genomic lineage shaped by long-term adaptation to natural environments. As reported by Gao et al. (2020), these accessions were sequenced using PacBio and Hi-C technologies, yielding an assembled genome of approximately 808 Mb, with a high proportion of repetitive elements (74.75%) and notable heterozygosity. A total of 38,828 protein-coding genes were annotated, over 98% of which were functionally classified. This comprehensive annotation sheds light on ecological adaptation and cannabinoid biosynthetic pathways. While the Tibetan wild-type genomes offer insights into natural selection and wild-type diversity, the ‘Hang Kra Rog Phu Phan’ cultivar provides unique perspective on domesticated cannabis germplasm shaped by traditional human use under Southeast Asian conditions. Its long-standing application in Thai traditional medicine and high THC content underscore its value for applied breeding, pharmacological research, and cultural preservation.

Table 3. Genome-wide sequence data for 21 cannabis cultivars obtained from the NCBI database, used for comparative analysis to identify DNA sequence variability at single nucleotide polymorphism positions

SRA ID	Sample	Submitted by	Bases (Gb)	Published
SRR35341777	Hang Kra Rog Phu Phan	Suan Sunandha Rajabhat University	9.57	This study
SRR14857079	CBD Shark	University of Moncton	62.6	6/18/2021
SRR16053696	Kearny	University of Nebraska - Lincoln	24.3	9/24/2021
SRR16053699	Hooker	University of Nebraska - Lincoln	13.4	9/24/2021
SRR16053700	Greely	University of Nebraska - Lincoln	19.4	9/24/2021
SRR16053701	Dougl's	University of Nebraska - Lincoln	22.3	9/24/2021
SRR16053703	York	University of Nebraska - Lincoln	23.8	9/24/2021
SRR16053704	RedWil	University of Nebraska - Lincoln	20.5	9/24/2021
SRR16053706	Polk	University of Nebraska - Lincoln	27.9	9/24/2021
SRR16053708	Nemaha	University of Nebraska - Lincoln	22.4	9/24/2021
SRR16053711	Merick	University of Nebraska - Lincoln	22.1	9/24/2021
SRR16053712	Madisn	University of Nebraska - Lincoln	21.6	9/24/2021
SRR16053714	Loup	University of Nebraska - Lincoln	20.2	9/24/2021
SRR16053715	Lincln	University of Nebraska - Lincoln	23.2	9/24/2021
SRR16053717	Knox	University of Nebraska - Lincoln	20.0	9/24/2021
SRR16053719	Keith	University of Nebraska - Lincoln	20.8	9/24/2021
SRR16053721	Dakota	University of Nebraska - Lincoln	19.0	9/24/2021
SRR16053722	Custer	University of Nebraska - Lincoln	18.6	9/24/2021
SRR16242757	Nance	University of Nebraska - Lincoln	30.5	10/7/2021
SRR20856322	US_feral	Colorado State University	19.4	8/10/2022
SRR24187786	FINOLA	University College Dublin	26.0	5/1/2023
SRR24187787	Felina32	University College Dublin	26.9	5/1/2023

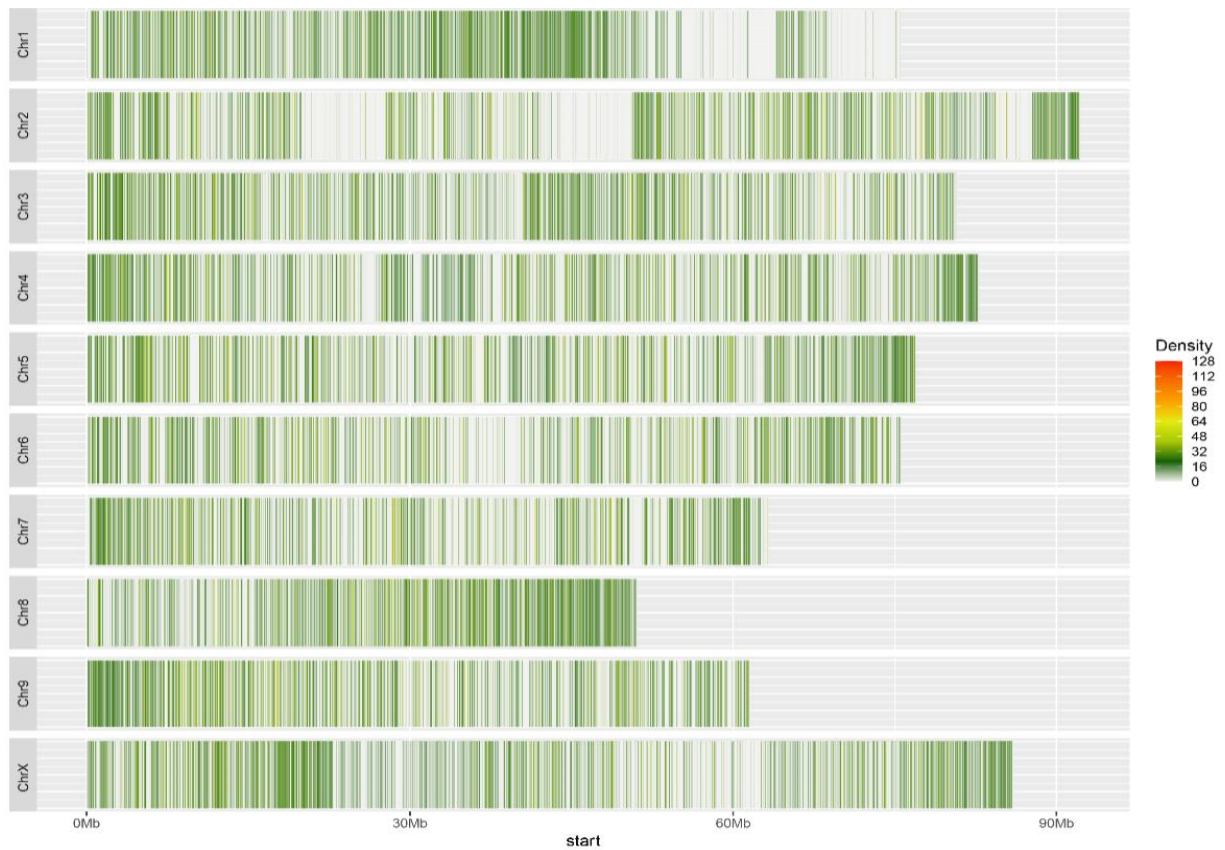


Figure 2. Genome-wide distribution of 6,040,109 Single Nucleotide Polymorphisms (SNPs) identified in *Cannabis sativa* 'Hang Kra Rog Phu Phan'. The plot shows SNP density across all 10 Chromosomes (Chr01-Chr10), calculated using a 1-kb sliding window. Color intensity represents SNP density, with green indicating low-density regions and red indicating high-density regions

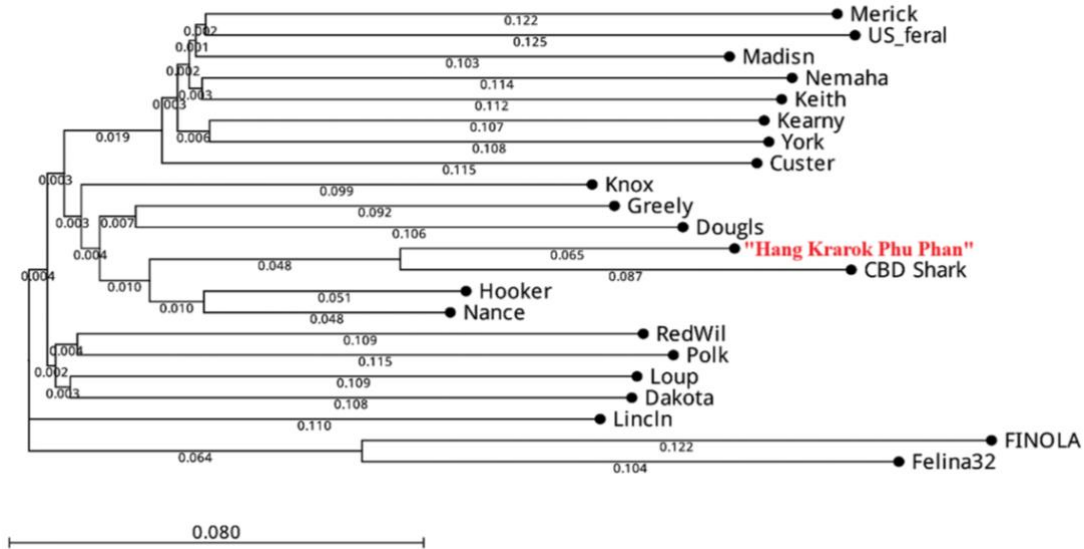


Figure 3. Phylogenetic relationships among *Cannabis sativa* accessions inferred using the neighbor-joining method based on genome-wide SNP data. The Thai landrace cultivar ‘Hang Kra Rog Phu Phan’ (highlighted in red; SRA ID = SRR35341777) is grouped closely with the ‘CBD Shark’ cultivar, indicating a high level of genomic similarity

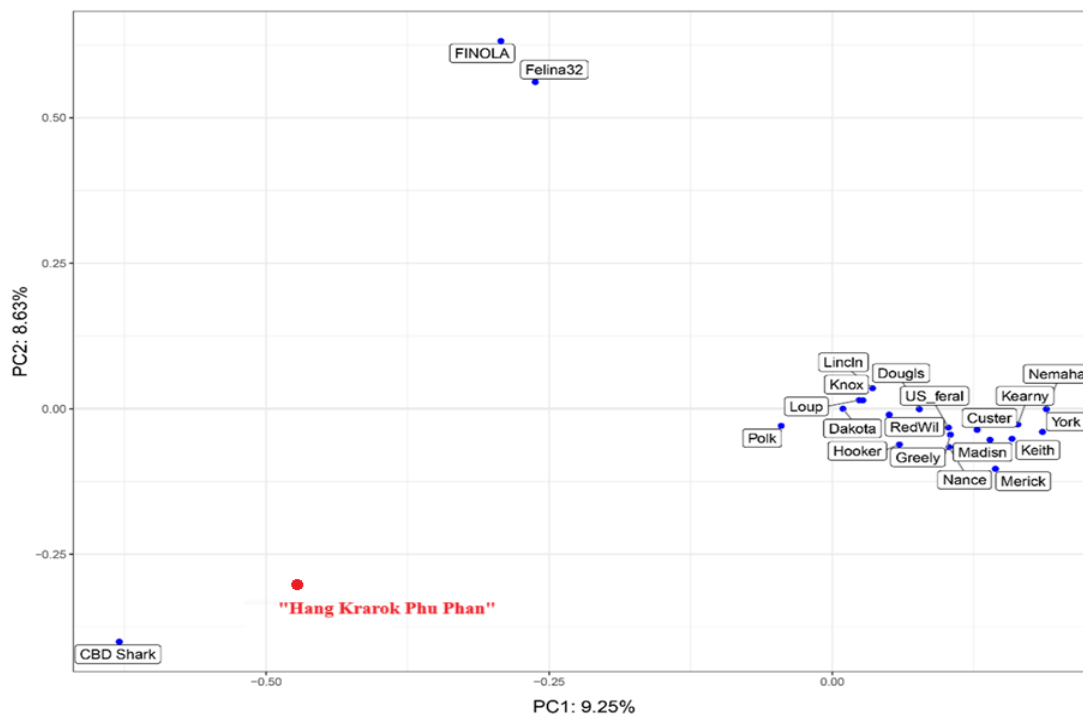


Figure 4. Principal Component Analysis (PCA) plot showing the genetic relationships among *Cannabis sativa* accessions based on genome-wide SNP data

Phylogenetic analysis and PCA clearly revealed that ‘Hang Kra Rog Phu Phan’ shares the closest genetic similarity with the CBD-rich ‘CBD Shark’ cultivar. This finding is particularly noteworthy, as ‘Hang Kra Rog Phu Phan’ is characterized by a high THC content, whereas ‘CBD Shark’ exhibits elevated CBD levels. However, genetic similarities between those two cultivars does not necessarily translate into chemical similarity. Although these cultivars may share a recent common ancestor or a

comparable genomic background, they differ at specific loci involved in CBD biosynthesis—particularly the THCAS (THC synthase) and CBDAS (CBD synthase) genes. The THC-to-CBD ratio is primarily regulated by variations in these genes, and even minor differences, such as SNPs or structural changes in coding or regulatory regions, can lead to markedly different chemotypes despite overall genomic similarity. These results are consistent with previous research highlighting considerable genomic

variability among cannabis cultivars, underscoring the importance of whole-genome analyses for understanding genetic relationships and diversity within species (Grassa et al. 2021).

The comprehensive mapping of SNPs across the cannabis genome offers valuable genetic markers that can be used for cultivar authentication, selective breeding, and functional genomic studies. By pinpointing genes associated with desirable phenotypic traits, such as CBD synthesis, disease resistance, and stress tolerance, this resource may significantly improve breeding efficiency (van Bakel et al. 2011; Braich et al. 2020). Furthermore, knowledge of genetic diversity will help optimize the genetic backbone of cannabis cultivars, facilitating the development of improved cultivars with tailored profiles for medicinal or industrial purposes. Future research should focus on characterizing functional SNPs associated with key biosynthesis pathways, particularly those related to CBDs and terpenes. These compounds are crucial for the therapeutic properties of cannabis, and understanding the genetic basis of their biosynthesis will enable more targeted breeding efforts. By identifying and utilizing SNPs linked to the regulation of these pathways, researchers optimize cannabis cultivars for improved chemical profiles, potentially enhancing the efficacy and safety of cannabis-based medicines (Hu et al. 2025). Additionally, such studies may lead to the discovery of novel genes involved in secondary metabolite production, further broadening the application of cannabis in the pharmaceutical and agricultural sectors (Hurgobin et al. 2021).

Although this study successfully identified and mapped over 6 million SNPs across the autosomal and X chromosomes, functional annotation of these variants has not yet been conducted. Upcoming analyses will incorporate tools such as SnpEff or VEP to predict the effects of these SNPs and classify them according to their genomic context—whether within coding regions, introns, or intergenic regions. Particular attention will be given to polymorphisms within cannabinoid synthase genes, such as THCAS and CBDAS, due to their direct impact on the pharmacological properties of the plant. These analyses will be essential for identifying functionally relevant variants to support breeding and improve the therapeutic value of Thai cannabis (Ingvarsdén and Brinch-Pedersen 2023).

In conclusion, this study provides the first comprehensive genomic characterization of the Thai cannabis landrace *C. sativa* cultivar ‘Hang Kra Rog Phu Phan’. Whole-genome sequencing produced 9.50 Gb of high-quality data, with a 94.43% alignment rate to the ‘Pink Pepper’ reference genome. Variant calling yielded over six million high-confidence SNPs across nine autosomes and the X chromosome. Phylogenetic and principal component analyses placed ‘Hang Kra Rog Phu Phan’ in close proximity to the CBD-rich ‘CBD Shark’ cultivar, highlighting its distinct genetic identity. These results confirm the research aims: to evaluate genomic similarity with a reference genome, identify genome-wide SNPs, and assess genetic relatedness. The SNP dataset provides a valuable resource for cultivar authentication, marker-assisted breeding, and

comparative genomics. It also supports conservation of native Thai germplasm, helping safeguard biodiversity and guide the sustainable development of Thailand’s cannabis industry.

ACKNOWLEDGEMENTS

The authors express their sincere gratitude to the Thai Traditional Medical Knowledge Fund, Department of Thai Traditional and Alternative Medicine, Ministry of Public Health, Thailand, for the financial support. Appreciation is also extended to the College of Allied Health Sciences, Suan Sunandha Rajabhat University, Thailand, for their generous support of this research.

REFERENCES

- Adams TK, Masondo NA, Malatsi P, Makunga NP. 2021. *Cannabis sativa*: From therapeutic uses to micropropagation and beyond. *Plants* 10 (10): 2078. DOI: 10.3390/plants10102078.
- Balant M, González RR, Garcia S, Garnatje T, Pellicer J, Vallès J, Vitales D, Hidalgo O. 2022. Novel insights into the nature of intraspecific genome size diversity in *Cannabis sativa* L. *Plants* 11 (20): 2736. DOI: 10.3390/plants11202736.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30 (15): 2114-2120. DOI: 10.1093/bioinformatics/btu170.
- Braich S, Baillie RC, Spangenberg GC, Cogan NOI. 2020. A new and improved genome sequence of *Cannabis sativa*. *GigaByte* 2020: gigabyte10. DOI: 10.46471/gigabyte.10.
- Busta L, Dweikat I, Sato SJ, Qu H, Xue Y, Zhou B, Gan L, Yu B, Clemente TE, Cahoon EB, Zhang C. 2022. Chemical and genetic variation in feral *Cannabis sativa* populations across the Nebraska climate gradient. *Phytochemistry* 200: 113206. DOI: 10.1016/j.phytochem.2022.113206.
- Chen L, Li H-L, Zhou H-J, Zhang G-Z, Zhang Y, Wang Y-M, Wang M-Y, Yang H, Gao W. 2024. Feature-based molecular network-assisted cannabinoid and flavonoid profiling of *Cannabis sativa* leaves and their antioxidant properties. *Antioxidants* 13 (6): 749. DOI: 10.3390/antiox13060749.
- Chen X, Guo H-Y, Zhang Q-Y, Wang L, Guo R, Zhan Y-X, Lv P, Xu Y-P, Guo M-B, Zhang Y, Zhang K, Liu Y-H, Yang M. 2022. Whole-genome resequencing of wild and cultivated cannabis reveals the genetic structure and adaptive selection of important traits. *BMC Plant Biol* 22 (1): 371. DOI: 10.1186/s12870-022-03744-0.
- Dal Martello R, Min R, Stevens CJ, Qin L, Fuller DQ. 2024. Morphometric approaches to *Cannabis* evolution and differentiation from archaeological sites: Interpreting the archaeological evidence from bronze age Haimenkou, Yunnan. *Veg Hist Archaeobot* 33 (4): 503-518. DOI: 10.1007/s00334-023-00966-6.
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10: giab008. DOI: 10.1093/gigascience/giab008.
- Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19: 11-15.
- El Oihabi M, Soutana M, Ammari M, Ben Allal L, Fakhil Lanjri A. 2024. Diversity and variability of bioactive compounds in *Cannabis sativa*: Effects on therapeutic and environmental uses and orientations for future research. *Case Stud Chem Environ Eng* 9: 100732. DOI: 10.1016/j.cscee.2024.100732.
- Fordjour E, Manful CF, Sey AA, Javed R, Pham TH, Thomas R, Cheema M. 2023. *Cannabis*: A multifaceted plant with endless potentials. *Front Pharmacol* 14: 1200269. DOI: 10.3389/fphar.2023.1200269.
- Gao S, Wang B, Xie S, Xu X, Zhang J, Pei L, Yu Y, Yang W, Zhang Y. 2020. A high-quality reference genome of wild *Cannabis sativa*. *Hortic Res* 7 (1): 73. DOI: 10.1038/s41438-020-0295-3.
- Grassa CJ, Weiblen GD, Wenger JP, Dabney C, Poplawski SG, Motley ST, Michael TP, Schwartz CJ. 2021. A new *Cannabis* genome assembly

- associates elevated Cannabidiol (CBD) with hemp introgressed into marijuana. *New Phytol* 230 (4): 1665-1679. DOI: 10.1111/nph.17243.
- Hu J, Wang Z, Xu H, Wang Z, Li N, Feng R, Yin J, Liu F, Wang B. 2025. Transcriptomic analysis of wild *Cannabis sativa*: Insights into tissue- and stage-specific expression and secondary metabolic regulation. *BMC Genomics* 26 (1): 528. DOI: 10.1186/s12864-025-11697-5.
- Hurgobin B, Tamiru-Oli M, Welling MT, Doblin MS, Bacic A, Whelan J, Lewsey MG. 2021. Recent advances in *Cannabis sativa* genomics research. *New Phytol* 230 (1): 73-89. DOI: 10.1111/nph.17140.
- Hussain T, Jeena G, Pitakbut T, Vasilev N, Kayser O. 2021. *Cannabis sativa* research trends, challenges, and new-age perspectives. *iScience* 24 (12): 103391. DOI: 10.1016/j.isci.2021.103391.
- Ingvarsdson CR, Brinch-Pedersen H. 2023. Challenges and potentials of new breeding techniques in *Cannabis sativa*. *Front Plant Sci* 14: 1154332. DOI: 10.3389/fpls.2023.1154332.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol Biol Evol* 35: 1547-1549. DOI: 10.1093/molbev/msy096.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20 (9): 1297-1303. DOI: 10.1101/gr.107524.110.
- Ren G, Zhang X, Li Y, Ridout K, Serrano-Serrano ML, Yang Y, Liu A, Ravikanth G, Nawaz MA, Mumtaz AS, Salamin N, Fumagalli L. 2021. Large-scale whole-genome resequencing unravels the domestication history of *Cannabis sativa*. *Sci Adv* 7 (29): eabg2286. DOI: 10.1126/sciadv.abg2286.
- Sawler J, Stout JM, Gardner KM, Hudson D, Vidmar J, Butler L, Page JE, Myles S. 2015. The genetic structure of Marijuana and Hemp. *PLoS One* 10 (8): e0133292. DOI: 10.1371/journal.pone.0133292.
- Shujat S, Robinson GI, Norouzkhani F, Kovalchuk I. 2024. Using advanced biotechnological techniques to improve *Cannabis* cultivars. *Biocatal Agric Biotechnol* 60: 103250. DOI: 10.1016/j.bcab.2024.103250.
- Sunanta P, Sombat T, Moaphadungkul J, Chaemthet S, Nagle M, Bakshani C, Willats W, Sangta J, Sommano S. 2024. Evaluation of value adding components from postharvest biomass of Thai medicinal *Cannabis* var. Hang Kra Rog Phu Phan. *J Appl Res Med Aromat Plants* 43: 100593. DOI: 10.1016/j.jarmap.2024.100593.
- van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, Page JE. 2011. The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol* 12 (10): R102. DOI: 10.1186/gb-2011-12-10-r102.
- Vergara D, White KH, Keepers KG, Kane NC. 2016. The complete chloroplast genomes of *Cannabis sativa* and *Humulus lupulus*. *Mitochondrial DNA A DNA Mapp Seq Anal* 27 (5): 3793-3794. DOI: 10.3109/19401736.2015.1079905.
- Weiblen GD, Wenger JP, Craft KJ, ElSohly MA, Mehmedic Z, Treiber EL, Marks MD. 2015. Gene duplication and divergence affecting drug content in *Cannabis sativa*. *New Phytol* 208 (4): 1241-1250. DOI: 10.1111/nph.13562.
- Woods P, Price N, Matthews P, McKay JK. 2022. Genome-wide polymorphism and genic selection in feral and domesticated lineages of *Cannabis sativa*. *Genes Genom Genet* 13 (2): jkac209. DOI: 10.1093/g3journal/jkac209.
- Younas M, Qureshi R, van Velzen R, Mashwani Z-U-R, Saqib Z, Ali A, Rehman S, Farah MA, Al-Anazi KM. 2024. Geo-climatic factors co-drive the phenotypic diversity of wild hemp (*Cannabis sativa* L.) in the Potohar Plateau and Lesser Himalayas. *BMC Plant Biol* 24 (1): 1031. DOI: 10.1186/s12870-024-05730-0.