



사인과 초두구의 엽록체 유전체 조립 및 변이지역 발굴

정지우¹ · 김민정² · 전은경³ · 구현조⁴ · 나경주^{5†}

Chloroplast Genome Assembly and DNA Variation Detection of *Amomum villosum* Lour. and *Alpinia kwangsiensis* T.L.Wu & S.J.Chen

Ji Woo Jung¹, Minjung Kim², Eun Kyoung Jun³, Hyun Jo Koo⁴, and Gyoungju Nah^{5†}

ABSTRACT

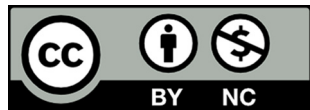
Received: 2025 October 24
1st Revised: 2025 November 10
2nd Revised: 2025 December 08
3rd Revised: 2025 December 11
Accepted: 2025 December 11
Published: 2025 December 30
This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background: *Amomum villosum*, known for its anti-inflammatory and antibacterial effects, is a traditional medicinal plant used for gastrointestinal disorders. However, *Alpinia kwangsiensis*, a related species with lower pharmacological efficacy, is often used as an adulterant, warranting accurate molecular identification.

Methods and Results: In this study, we analyzed the complete chloroplast genomes of *A. villosum* and *A. kwangsiensis* using an Illumina NextSeq 1000 sequencer and a NOVOPlasty assembler to identify polymorphic regions for DNA marker development. The genome of *A. villosum* (164,070 bp) comprised large and small single-copy regions and two inverted repeats, whereas that of *A. kwangsiensis* (162,449 bp) displayed a similar structure with size differences. Moreover, the gene content of the two species slightly varied, with *A. villosum* and *A. kwangsiensis* encoding 133 and 130 genes, respectively. Our comparative analysis involving related *Amomum* (*Wurfbainia*) and *Alpinia* accessions clarified the taxonomic relationships. We identified small insertions/deletions and single-nucleotide polymorphisms as potential species-specific DNA marker candidates

Conclusion: The results of this study provide essential chloroplast genomic resources for authenticating medicinal materials and support future marker-assisted methods for the identification of *A. villosum* and related species.

Key Words: *Alpinia kwangsiensis*, *Amomum villosum*, Complete Chloroplast Genome, Next-Generation Sequencing, Phylogenetic Tree, DNA Variation



서 언

생강과 (Zingiberaceae Martinov)는 주로 열대 및 아열대 지역에 분포하며, 향기 나는 근경과 독특한 꽃 구조로 인해 약용, 식용, 향신료 및 관상용으로 널리 이용되는 다년생 초본 식물이다 (Delin and Larsen, 2000; Kress *et al.*, 2002; Li *et al.*, 2020). 본 과의 식물은 일반적으로 발달된 엽초와 두 줄로 배열된 (distichous) 잎, 엽설 (ligule)을 가지며, 좌우대칭 (zygomorphic)의 화서 구조와 함께 단일 가임 수술, 하위 3실 (inferior, 3-locular)의 자방을 특징으로 한다. 이들 형태적 특성은 분류학적 및 계통학적 연구에 있어 중요한 형질로 간주

된다.

이 가운데 사인 (*Amomum villosum* Lour.)은 중국 남부 (광둥, 광시, 푸젠, 윈난)와 인도차이나반도 등 동아-동남아시아 전역에 분포하는 대표적인 생강과 식물로, 전통의학 및 향신료로 오래전부터 이용되어 온 유용 자원이다 (Feng *et al.*, 2024; Liao *et al.*, 2024). 또한 동일한 *Amomum* 속에 속하는 *Amomum muricarpum*도 그 추출물이 염증성 사이토카인 발현과 STAT3·NF-κB 신호를 억제하여 염증 반응을 감소시키는 효과를 보인다고 보고되었다 (Park *et al.*, 2023). 이는 *Amomum*속 식물이 염증 조절에 잠재적 치료 자원임을 시사한다.

사인은 *Amomum* Roxb. 속에 포함되어 있었으나, 형태 및

[†]Corresponding author: (Phone) +82-02-880-4971 (E-mail) gnah.nicem@snu.ac.kr

¹서울대학교 농생명과학공동기기원 유전체분석센터 선임분석원 / Senior analyst, Genome Analysis Center, NICEM, Seoul National University, Seoul 08826, Korea

²서울대학교 농생명과학공동기기원 유전체분석센터 연구원 / Researcher, Genome Analysis Center, NICEM, Seoul National University, Seoul 08826, Korea

³서울대학교 농생명과학공동기기원 유전체분석센터 선임연구원 / Senior researcher, Genome Analysis Center, NICEM, Seoul National University, Seoul 08826, Korea

⁴서울대학교 농생명과학공동기기원 유전체분석센터 책임연구원 / Principal researcher, Genome Analysis Center, NICEM, Seoul National University, Seoul 08826, Korea

⁵서울대학교 농생명과학공동기기원 유전체분석센터 센터장 / Head, Genome analysis Center, NICEM, Seoul National University, Seoul 08826, Korea

분자계통학적 분석 결과에 따라 *Wurfbainia* Giseke 속으로 재분류되기도 한다 (de Boer *et al.*, 2018). 최근의 분자계통학 연구에 따르면, 기존에 하나의 속 (屬)으로 묶여 있던 *Amomum*은 실제로는 같은 조상에서 비롯된 단일 집단이 아니라 여러 갈래로 나뉘는 비단일 계통군 (non-monophyletic) 그룹이라는 사실이 드러났다. 그 중 Clade VII에 속하는 종들은 예전부터 “*Amomum-Villosum group*”으로 불리던 집단인데, 형태적 특징뿐 아니라 유전적 분석에서도 다른 *Amomum* 종들과 확실히 구분되는 독립된 그룹을 이루는 것으로 나타났다. 이처럼 Clade VII가 뚜렷하게 분리된 계통이라는 점을 근거로, 과거에 이름만 제안되고 거의 쓰이지 않았던 *Wurfbainia*라는 속 이름을 다시 공식적으로 사용해야 한다는 주장이 대두되었다. 또한 *Amomum*을 기존처럼 광범위한 범위로 유지하려면, *Alpinia*나 *Elettaria* 같은 다른 속들까지 함께 합쳐야 하는 문제가 생기는데, 이는 식물 분류 체계의 안정성과 명명 규칙에 좋은 방식이 아니라는 점도 지적되었다. 따라서 Clade VII에 속하는 종들을 *Wurfbainia* 속으로 재분류하는 것이 가장 타당하며, 이를 통해 *Amomum* 속의 다계통성 문제를 해결하는 가장 합리적인 접근이라는 결론을 제시했다 (de Boer *et al.*, 2018). 따라서, 현재 사인은 *W. villosa* (Lour.) Škorničk. & A.D.Poulsen의 학명으로 인식되고 있으며, 기존의 *A. villosum*은 명명법적 이명으로 처리되어 인식된다.

사인과 유사한 형태를 가지는 또 다른 종인 초두구 (*Alpinia kwangsiensis* T.L.Wu & S.J.Chen)는 광둥성, 광시성, 구이저우성, 중국 윈난성 및 베트남 북부 산악 지역에 분포하며 마찬가지로 다년생 초본으로 성장한다 (Tran *et al.*, 2018). 이 식물은 위줄기 (pseudostem)를 포함하여 높이 1.5-3.0m에 달하며, 굵게 발달한 근경과 끝부분에 발달하는 정생 화서를 특징으로 한다. 꽃차례는 정생으로 나타나며 이삭상 또는 총상 형태를 이루고, 꽃은 백색에서 연분홍색이며 순판 (labellum)에는 적색 또는 황색의 무늬가 나타난다. 또한 각 포엽은 하나의 꽃을 받치고 있으며, 과실은 삭과로 발달한다 (Delin and Larsen, 2000).

사인은 특히 중국 남부 영남 지역에서 식·약용 작물로 재배되며, 건조한 열매인 'Fructus Amomi'는 전통 중의학에서 위장 질환 치료에 널리 사용되어 왔다. 이 종은 당대 (唐代) 이래로 약효가 기록되어 왔으며, 현재에도 다양한 소화기계 증상 개선에 이용되고 있다 (Feng *et al.*, 2024; Yang *et al.*, 2022; Zhang *et al.*, 2025). 이러한 사인의 경제적 가치와 약리적 활용 가능성에도 불구하고, 건조한 과실 형태가 비슷한 생강과의 초두구가 유사재료종으로 활용되어 정확한 식별에는 여전히 한계가 존재한다. 더구나 한약재 및 건강식품의 가공 과정에서 형태적 특징이 소실되거나 약화됨에 따라 DNA 기반 마커 개발법이 매우 절실한 상황이다.

이러한 배경에서 최근 엽록체 유전체 (chloroplast genome)

분석이 종 동정 및 계통분류 연구에 있어 유용한 도구로 부각되고 있다. 엽록체 유전체는 비교적 보존된 구조를 가지며, Insertion/Deletion (InDel), Single Nucleotide Polymorphism (SNP), Simple Sequence Repeat (SSR) 등 다양한 형태의 DNA 마커 개발에 활용될 수 있는 유용한 유전적 정보원을 제공한다 (Li *et al.*, 2020; Song *et al.*, 2022; Zhu *et al.*, 2023; Wang *et al.*, 2024).

이에 본 연구에서는 생강과 식물인 사인과 초두구 두 종의 완전한 엽록체 유전체를 시퀀싱하여 조립하고, 유전자 기능과 위치를 주석화하였다. 이를 바탕으로 두 종 간의 계통유전학적 관계를 규명하고, 변이 분석을 통해 마커 개발 가능성을 평가하였다. 변이 분석 결과 확인된 1,635개의 변이 (SNP-InDel)는 large single copy (LSC) 영역에 집중되어 나타나 종간 차이를 구분할 수 있는 유용한 후보 마커로 활용될 가능성이 높다. 특히 LSC에 분포한 117개의 InDel과 1,026개의 SNP는 변이 밀도가 높아 종 분류 및 계통학적 구분을 위한 고해상도 분자마커 개발에 적합한 구간으로 판단된다. 또한 SSR 분석에서 확인된 반복 유형 차이와 함께 이러한 변이 정보를 통합하면, 두 종을 효과적으로 식별할 수 있는 다중 마커 패널 구축이 가능할 것으로 보인다. 본 연구 결과는 약용식물의 진위 판별을 위한 중요한 유전체 자원을 제공하며, *A. villosum* 및 그 근연종의 향후 마커 기반 동정 (marker-assisted identification)을 위한 기초 자료로 활용될 수 있을 것이다.

재료 및 방법

1. 식물 재료 수집

사인 (*Amomum villosum* Lour., 유전자원번호: WKU-2-19-0927)과 초두구 (*Alpinia kwangsiensis*, 유전자원번호: WKU-2-18-0459)의 건조된 열매 (Fig. 1)는 원광대학교 한의과대학 본초학교실 (<http://foramedica.net>)로부터 분양받았다.

2. DNA 추출 및 라이브러리 제작

두 종의 신선한 잎 조직으로부터 유전체 DNA는 Kuo *et*



Amomum villosum

Alpinia kwangsiensis

Fig. 1. Dried fruits of *A. villosum* and *A. kwangsiensis*. Photo by Dr. Hyun Jo Koo).

al. (2022)이 보고한 방법을 기반으로 한 수정된 CTAB (cetyltrimethylammonium bromide) 방법을 사용하여 추출하였다. 추출된 DNA의 농도와 순도는 NanoDrop 분광광도계 (Thermo Scientific Inc., Waltham, MA, USA) 및 1% 아가로스 겔 전기영동을 통해 확인하였다.

Illumina paired-end (PE) 유전체 라이브러리는 제조사의 매뉴얼에 따라 NEXTflex® Rapid DNA Sequencing Kit (Bioo Scientific, Austin, TX, USA)를 이용하여 제작하였다.

3. 엽록체 유전체 조립 및 계통수 분석

전체 엽록체 유전체 시퀀싱은 NextSeq 1000 플랫폼 (Illumina Inc., San Diego, CA, USA)을 이용하여 수행되었다. 시퀀싱된 원시 데이터는 Trimmomatic v0.39 (Bolger et al., 2014)를 사용하여 어댑터 서열 및 저품질 염기를 제거하여 고품질 PE 리드 (read)를 확보하였다. 확보된 리드는 NOVOPlasty v4.3.1 (Dierckxsens et al., 2017)을 이용하여 비지도 조립 (de novo assembly)을 수행하였고, 조립된 서열에 대해 동일 시퀀싱 리드를 다시 매핑하는 self-mapping 방식을 적용하여 Kim et al. (2015)의 절차에 따라 수작업으로 오차를 교정하였다.

유전체 주석 (annotation)은 GeSeq (Tillich et al., 2017)을 사용하여 수행되었으며, protein-coding genes (PCGs), transfer RNA (tRNA) genes, ribosomal RNA (rRNA) genes를 포함한 주요 유전자 구성을 확인하였고 필요에 따라 수동으로 추가 수정하였다. 최종 완성된 엽록체 유전체 서열은 GenBank에 등록되었으며, 사인은 PV669993, 초두구는 PV669995의 등록번호를 부여받았으며 관련된 BioProject, BioSample, SRA 정보는 다음과 같다: 사인의 경우 BioProject PRJNA1301522, BioSample SAMN50446599, SRA SRR34862458로 등록되어 있으며 초두구의 경우 BioProject PRJNA1301523, BioSample SAMN50446624, SRA SRR34863081로 등록되어 있다.

새로 시퀀싱한 사인과 초두구를 포함하여, NCBI GenBank에서 확보한 관련 종의 전체 엽록체 유전체를 합쳐 총 28종의 ingroup 서열을 사용하였고, 또한 계통수의 외군 (outgroup)으로 Zingiber 속 2종의 완전한 엽록체 유전체를 추가로 포함하였다. 모든 서열은 MAFFT v7.526 (Katoh et al., 2002)을 사용하여 다중서열정렬 (multiple sequence alignment)을 진행

하였다. 정렬된 서열을 기반으로 MEGA v11.0.13 (Tamura et al., 2021)을 이용하여 Maximum Likelihood (ML) 방법으로 계통수를 구축하였다. ML 분석에서는 Tamura-Nei 모델, Uniform rates, 그리고 모든 코돈 위치 (1st, 2nd, 3rd) 및 비번역 부위 (non-coding sites)를 포함한 전체 염기서열을 사용하였고, 신뢰도 검정을 위해 1,000회 bootstrap 분석을 수행하였다.

4. 엽록체 유전체 변이 분석

두 종 간의 엽록체 유전체 변이 분석은 MUMmer v3 (Kurtz et al., 2004)를 사용하여 수행하였다. Nucmer 명령어를 통해 기본 설정값으로 전체 유전체 정렬을 수행하였고, 생성된 delta 정렬 파일은 show-snps 명령어와 -ClrT 옵션을 적용하여 SNP 및 InDel을 추출하였다. 추출된 변이들은 기준 (reference) 또는 질의 (query) 서열 중 한쪽에 gap이 존재하는지를 기준으로 SNP와 InDel로 분류하였다.

또한, SSR 분석을 위해 misa.pl 스크립트 (https://github.com/cfljam/SSR_marker_design/blob/master/misa.pl)를 사용하였으며, mono-, di-, tri-, tetra-, penta-, hexa-nucleotide 반복 서열을 탐색하였다.

결과 및 고찰

1. 엽록체 유전체 데이터 생산 및 조립

사인의 엽록체 유전체 조립을 위한 Illumina 기반 리드 생산 및 리드 trimming 전후 데이터 생산량 및 quality 값은 아래 Table 1과 같다. 사인의 경우 Q30 이상 trimmed 리드 기준으로 총 17,478,334 리드가 생산되었고, 초두구의 경우 총 11,016,779 리드가 생산되었다.

Trimmed 리드를 기반으로 조립된 사인의 완전한 엽록체 유전체는 총 164,070 bp의 길이를 가지며, G+C 함량은 36%로 나타났다. 시퀀싱된 리드를 조립된 엽록체 유전체에 참조 (reference) 서열로 사용하여 self-mapping 방식으로 매핑하였으며, 그 결과 평균 커버리지는 약 305배로 확인되었다 (Fig. 2A). 한편, 초두구의 완전한 엽록체 유전체는 162,449 bp의 길이와 36.1%의 G+C 함량을 보였으며, 평균 커버리지는 약 365배로 나타났다 (Fig. 2B).

사인과 초두구의 엽록체 유전체는 모두 높은 시퀀싱 커버리

Table 1. Statistics of next generation sequencing data of *A. villosum* and *A. kwangsiensis*.

Species		Reads	Bases	Q30 Ratio	Q20 Ratio	GC Ratio
<i>Amomum villosum</i>	Raw data	18,906,756	2,854,920,156	94.15	98.3	40.6
	Trimmed data	17,478,334	2,639,228,434	97	99.3	40.3
<i>Alpinia kwangsiensis</i>	Raw data	12,090,251	1,825,627,901	92.62	97.81	45.52
	Trimmed data	11,016,779	1,663,533,629	95.8	98.98	47.12

사인과 초두구의 엽록체 전장유전체 분석

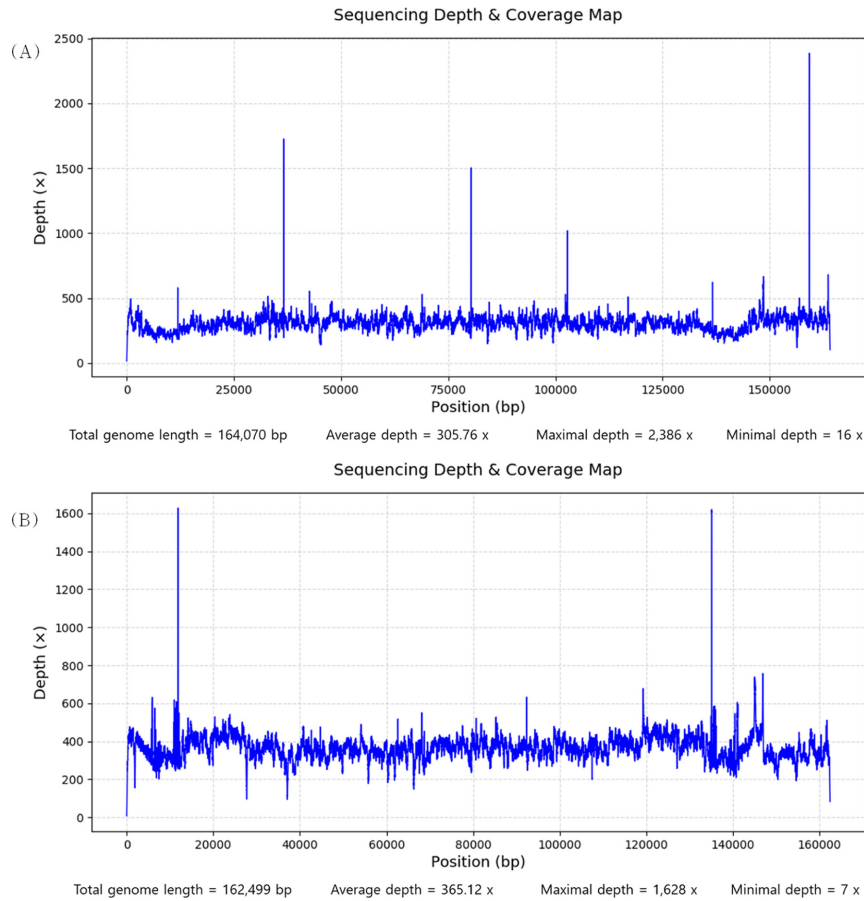


Fig. 2. Chloroplast genome coverage depth of (A) *A. villosum* and (B) *A. kwangsiensis*. Chloroplast genome coverage depth of (A) *A. villosum* and (B) *A. kwangsiensis*. For *A. villosum*, the maximum, minimum, and average coverage depths are 2,386 \times , 16 \times , and 305 \times , respectively. For *A. kwangsiensis*, the corresponding values are 1,628 \times , 7 \times , and 365 \times , respectively.

지와 안정적인 GC 함량을 보여 조립 신뢰도가 높으며, 두 종 간 유전체 크기 차이는 종 특이적 분자적 특성을 반영하는 것으로 판단된다. 이 결과는 서열 변이 분석의 기반 자료로서 두 유전체의 활용 가능성을 충분히 뒷받침한다.

2. 엽록체 유전체 구조 분석

*A. villosum*의 엽록체 유전체는 총 164,070 bp로, 88,799 bp의 LSC 영역, 15,353 bp의 small single copy (SSC) 영역, 그리고 각각 29,959 bp의 inverted repeats (IRa 및 IRb) 영역으로 구성되어 있었다 (Fig. 3A). Annotation 결과, 본 유전체는 총 133개 유전자를 포함하며, 이 중 87개의 PCGs, 38개의 tRNA 유전자, 8개의 rRNA 유전자로 구성되어 있다 (Table 2).

*A. kwangsiensis*의 엽록체 유전체 (PV669995) 또한 유사한 구조를 보였으며, 총 162,499 bp, LSC 91,537 bp, SSC 19,308 bp, IRa 및 IRb 각각 25,802 bp로 구성되어 있었다 (Fig. 3B). 해당 종은 총 130개 유전자 (86 PCGs, 36 tRNAs,

Table 2. Details of the complete chloroplast genome of *A. villosum* and *A. kwangsiensis*.

	<i>A. villosum</i>	<i>A. kwangsiensis</i>
Aligned reads	280,524	352,648
Genome size (bp)	164,070	162,449
Organelle coverage (x)	258	328
LSC length (bp)	88,799	91,537
SSC length (bp)	15,353	19,308
IR length (bp)	29,959	25,802
LSC GC content (%)	33.70	36.30
SSC GC content (%)	29.90	38.20
IR GC content (%)	41.10	32.90
GC content (%)	36.00	36.10
Total number of genes	133	130
PCGs	87	86
rRNA	8	8
tRNA	38	36

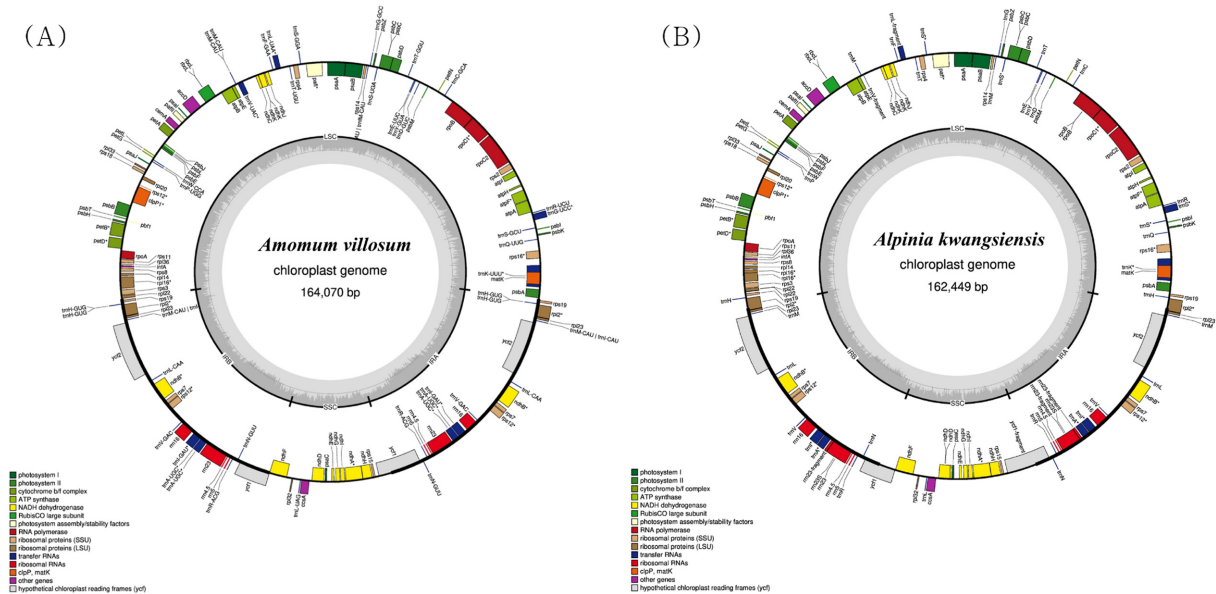


Fig. 3. Circular maps of the chloroplast genomes of (A) *A. villosum* and (B) *A. kwangsiensis*. Genes located inside the circle are transcribed in the counterclockwise direction, whereas those outside are transcribed in the clockwise direction. Gene colors represent different functional categories. The large single-copy (LSC), small single-copy (SSC), and inverted repeat (IRA and IRB) regions are also indicated.

8 rRNAs)를 포함하고 있다 (Table 2).

두 종의 엽록체 유전체는 전반적으로 유사한 구조적 틀을 공유하면서도 LSC·SSC·IR 구간의 길이와 유전자 수에서 차이를 보여, 종 수준의 특성을 반영하는 미세한 구조적 변이가 존재함을 시사한다. 이러한 구조적 차이는 두 종 간 진화적 분화 과정을 이해하는 데 중요한 단서를 제공하며, 나아가 종 식별을 위한 변이 탐색의 잠재적 지표로 활용될 수 있다.

3. 유전자 주석 결과

사인과 초두구의 엽록체 유전체는 각각 133개, 130개의 유전자를 포함하며, PCGs 각각 87개, 86개, tRNA 유전자 각각 38개, 36개, rRNA 유전자 모두 8개로 구성되어 있다 (Table

3 and 4). 예측된 유전자들은 19개의 기능 범주로 분류되었으며, *accD*, *ccsA*, *cemA*, *clpP1*, *infA*, *matK*, *pf1*, *rbcL*의 8개 유전자는 single-copy로, 나머지 유전자는 모두 multi-copy로 존재하는 것으로 나타났다. 아울러 사인과 초두구의 계통수 분석 (Fig. 4)에 사용된 *Wurfbainia* 속 16종 및 *Alpinia* 속 10종, 그리고 outgroup으로 사용된 *Zingiber* 속 2종의 엽록체 유전체 크기 및 유전자 수 정보를 비교하였다 (Table 5). 그 결과 엽록체 유전체 크기는 160,100 bp~164,069 bp 범위에, 유전자 수는 125개~139개 범위에 있었다. 본 분석에서 도출된 사인 및 초두구의 엽록체 크기 및 유전자 수도 위와 같은 범위에 있음을 확인하였다.

사인과 초두구의 유전자 구성은 기능 범주 분포와 single-

Table 3. Gene annotation of the chloroplast genome of *A. villosum*.

Category	Group of Genes	No. of Genes	Gene List
Photosynthesis	ATP synthase	6	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> , <i>atpH</i> , <i>atpI</i>
	Photosystem I	7	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i> , <i>psaK</i> , <i>psaL</i>
Photosynthesis	Photosystem II	15	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbT</i> , <i>psbZ</i> , <i>psb1</i>
	Cytochrome b6/f complex	6	<i>petA</i> , <i>petB</i> , <i>petD</i> , <i>petG</i> , <i>petL</i> , <i>petN</i>
Photosynthesis	c-type cytochrome synthesis gene	1	<i>ccsA</i>
	Rubisco (large subunit)	1	<i>rbcL</i>
Photosynthesis	NADH dehydrogenase	12	<i>ndhA</i> , <i>ndhB</i> (x2), <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>

Table 3. Continued.

Category	Group of Genes	No. of Genes	Gene List
Self-replication	RNA polymerase	4	<i>rpoA, rpoB, rpoC1, rpoC2</i>
	Large ribosomal proteins (LSU)	11	<i>rpl14, rpl16, rpl2 (x2), rpl20, rpl22, rpl23 (x2), rpl32, rpl33, rpl36</i>
	Small ribosomal proteins (SSU)	15	<i>rps11, rps12 (x2), rps14, rps15, rps16, rps18, rps19 (x2), rps2, rps3, rps4, rps7 (x2), rps8</i>
	Ribosomal RNAs	8	<i>rrn16 (x2), rrn23 (x2), rrn4.5 (x2), rrn5 (x2)</i>
	Transfer RNAs	38	<i>trnA-UGC (x2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnG-UCC, trnH-GUG (x2), trnI-CAU (x2), trnI-GAU (x2), trnK-UUU, trnL-CAA (x2), trnL-UAA, trnL-UAG, trnM-CAU, trnN-GUU (x2), trnP-UGG, trnQ-UUG, trnR-ACC (x2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC (x2), trnV-UAC, trnW-CCA, trnY-GUA, trnM-CAU</i>
Other genes	Acetyl-CoA carboxylase	1	<i>accD</i>
	Protease	1	<i>clpP1</i>
	Translation initiation factor	1	<i>infA</i>
	Maturase	1	<i>matK</i>
	Envelope membrane protein	1	<i>cemA</i>
	Conserved hypothetical ORFs (ycf)	4	<i>ycf1 (x2), ycf2 (x2)</i>

Table 4. Gene annotation of the chloroplast genome of *A. kwangsiensis*.

Category	Group of Genes	No. of Genes	Gene List
Photosynthesis	ATP synthase	6	<i>atpA, atpB, atpE, atpF, atpH, atpI</i>
	Photosystem I	7	<i>psaA, psbA, psbC, psal, psaj, psal, psalI</i>
	Photosystem II	15	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbT, psbZ, pbf1</i>
	Cytochrome b6/f complex	6	<i>petA, petB, petD, petG, petL, petN</i>
	c-type cytochrome synthesis gene	1	<i>ccsA</i>
	Rubisco (large subunit)	1	<i>rbcl</i>
	NADH dehydrogenase	12	<i>ndhA, ndhB (x2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
	Self-replication	RNA polymerase	4
Large ribosomal proteins (LSU)		11	<i>rpl14, rpl16, rpl2 (x2), rpl20, rpl22, rpl23 (x2), rpl32, rpl33, rpl36</i>
Small ribosomal proteins (SSU)		15	<i>rps11, rps12 (x2), rps14, rps15, rps16, rps18, rps19 (x2), rps2, rps3, rps4, rps7 (x2), rps8</i>
Ribosomal RNAs		8	<i>rrn16 (x2), rrn23 (x2), rrn4.5 (x2), rrn5 (x2)</i>
Transfer RNAs		36	<i>trnA-UGC (x2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnH-GUG (x2), trnI-GAU (x2), trnK-UUU, trnL-TAG, trnL-CAA (x2), trnM-CAU (x4), trnN-GUU (x2), trnP-UGG, trnQ-UUG, trnR-UCU, trnR-ACC (x2), trnS-GCA, trnS-TGA, trnS-GCT (x2), trnT-UGU, trnT-GGU, trnV-GAC, trnV-UAC, trnW-CCA, trnY-GUA</i>
Other genes	Acetyl-CoA carboxylase	1	<i>accD</i>
	Protease	1	<i>clpP1</i>
	Translation initiation factor	1	<i>infA</i>
	Maturase	1	<i>matK</i>
	Envelope membrane protein	1	<i>cemA</i>
	Conserved hypothetical ORFs	3	<i>ycf1, ycf2 (x2)</i>

Table 5. Summary of the complete chloroplast genomes used in the phylogenetic tree.

Species (Accession)	Genome size (bp)	Gene (all)	PCCs	rRNA	tRNA
<i>Wurfbainia compacta</i> (MG000589.1)	162,176	127	89	8	30
<i>Wurfbainia longiligularis</i> (MK889505.1)	163,608	125	87	8	30
<i>Wurfbainia longiligularis</i> (NC_044774.1)	164,068	133	87	8	38
<i>Wurfbainia longiligularis</i> isolate Y19019 (MN067434.1)	163,608	133	87	8	38
<i>Wurfbainia longiligularis</i> isolate Y19020 (MN067435.1)	163,608	133	87	8	38
<i>Wurfbainia neoaurantiaca</i> (MW145134.1)	163,534	133	87	8	38
<i>Wurfbainia neoaurantiaca</i> voucher Y20180726015 (NC_057208.1)	163,534	133	87	8	38
<i>Wurfbainia villosa</i> (MH161416.1)	164,068	133	87	8	38
<i>Wurfbainia villosa</i> (MH161418.1)	164,069	133	87	8	38
<i>Wurfbainia villosa</i> (MK262730.1)	163,608	133	87	8	38
<i>Wurfbainia villosa</i> (MK389642.1)	163,733	126	88	8	30
<i>Wurfbainia villosa</i> isolate Y19017 (MN067431.1)	164,068	133	87	8	38
<i>Wurfbainia villosa</i> isolate Y19021 (MN067432.1)	164,069	133	87	8	38
<i>Wurfbainia villosa</i> (MN931250.1)	164,069	133	87	8	38
<i>Wurfbainia villosa</i> var. <i>xanthioides</i> (MH161417.1)	163,985	133	87	8	38
<i>Wurfbainia villosa</i> var. <i>xanthioides</i> isolate Y19018 (NC_044747.1)	163,981	133	87	8	38
<i>Alpinia chinensis</i> (NC_050165.1)	163,590	136	91	8	37
<i>Alpinia galanga</i> (MZ066611.1)	160,100	133	87	8	38
<i>Alpinia hainanensis</i> (MK262728.1)	162,387	133	87	8	38
<i>Alpinia japonica</i> (NC_083220.1)	161,906	133	87	8	38
<i>Alpinia katsumadae</i> (NC_048461.1)	162,387	133	87	8	38
<i>Alpinia katsumadai</i> (NC_048461.1)	162,387	133	87	8	38
<i>Alpinia kwangsiensis</i> (MZ066612.1)	162,323	133	87	8	38
<i>Alpinia officinarum</i> (MT254526.1)	162,137	132	86	8	38
<i>Alpinia oxyphylla</i> (KY985237.1)	161,990	139	92	8	39
<i>Alpinia pumila</i> (NC_048462.1)	161,920	133	87	8	38
<i>Zingiber officinale</i> (NC_044775.1)	162,621	133	87	8	38
<i>Zingiber yingjiangense</i> (OP869983.1)	163,623	133	87	8	38

copy 유전자 배열에서 높은 보존성을 보이면서도 전체 유전자 수에서 미세한 차이를 나타내어, 두 종이 엽록체 수준에서 안정적인 기본 구조를 유지하되 종 특이적 변이를 지닌다는 점을 시사한다.

4. 계통수 분석 결과

본 연구에서는 Zingiberaceae와 그 대표 종들을 포함하여 엽록체 유전체 기반 계통분석을 수행하였다. Maximum Likelihood 기법으로 구축된 계통수 결과, 사인은 *W. villosa* 및 *W. longiligularis* 군과 함께 *Wurfbainia* 속에서 높은 지지값 (bootstrap 값 100)으로 강하게 클러스터링 되었다 (Fig. 4). 이는 기존 *Amomum*과 *Wurfbainia* 속 간 분류 경계가 명확하지 않다는 선행연구를 지지하며 (de Boer *et al.*, 2018), *A. villosum* 이 분자 계통학적으로는 *Wurfbainia* 속에 더 근접한 분류군임

을 강하게 시사한다.

한편, 초두구는 *A. kwangsiensis*와 함께 *Alpinia* 속에서 안정된 클레이드를 형성하였다 (bootstrap 값 100). 초두구와 *A. kwangsiensis*는 엽록체 유전체 기반 계통수 분석에서 높은 지지값 (bootstrap 100)으로 묶이는 안정적인 단일계통군을 형성하여, 두 종이 *Alpinia* 속 내에서 매우 근연한 진화적 관계임을 보여준다. 이는 두 종이 *Alpinia* 속의 독립된 분지군 (monophyletic clade)을 구성한다는 기존 분류체계를 강하게 지지하는 결과이다 (Zhang *et al.*, 2021).

본 연구에서 사용된 사인과 초두구의 엽록체 유전체 기반 계통 분석 결과, 두 시료는 GenBank에 등록된 근연종들과의 클러스터링에서 각각 *Amomum* (= *Wurfbainia*) 속과 *Alpinia* 속에 속하는 것으로 명확히 확인되었다. 이는 기존의 계통 분류와 일치하는 결과로, 시료 동정의 분자적 타당성을 지지한

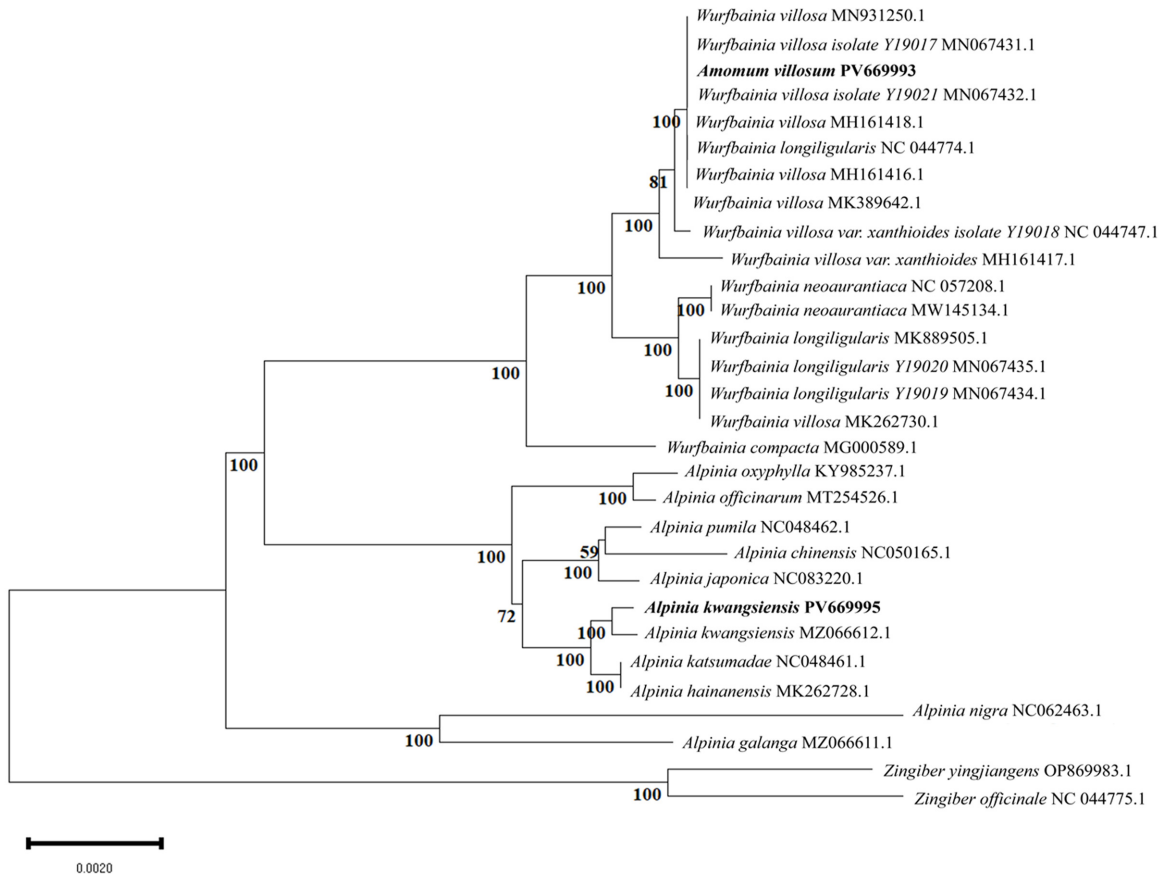


Fig. 4. Maximum-likelihood phylogenetic tree showing the relationships of *A. villosum* and *A. kwangsiensis* (highlighted in bold) with their close relatives. The analysis was based on complete chloroplast genome sequences of *A. villosum*, *A. kwangsiensis*, and 27 additional accessions from the genera *Amomum* and *Alpinia*, using *Zingiber yingjiangense* and *Zingiber officinale* as outgroups. Bootstrap support values based on 1,000 replicates are shown at each node.

다 (de Boer *et al.*, 2018; Liu *et al.*, 2022; Yang *et al.*, 2021). 따라서 두 종의 분자적 동정이 기존 분류체계와 일관됨을 확인함으로써 엽록체 유전체 기반 계통 정보가 해당 분류군들의 진화적 관계를 신뢰성 있게 반영하며, 약재 원식물의 정확한 종 판별에도 유효한 근거가 될 수 있음을 시사한다.

5. 후보 변이지역 분석 결과

*A. villosum*과 *A. kwangsiensis*의 엽록체 DNA 서열 간 변이 특성을 분석하기 위해 SSR, InDel, SNP 분석을 수행하였다. SSR 분석 결과, 두 종 간 차이를 보이는 SSR 반복 유형 차이가 존재하였으며, *A. villosum*에서 총 106개의 SSR이, *A. kwangsiensis*에서 69개의 SSR이 검출되었다. 두 종 모두에서 mononucleotide SSR이 가장 높은 비율을 차지하였으며, dinucleotide 및 trinucleotide 반복은 상대적으로 낮은 빈도로 나타났다 (Table 6).

또한, 총 1,635개의 변이가 검출되었으며, 이 중 157개는 InDel, 1,478개는 SNP였다 (Table 7). 이러한 변이는 LSC,

SSC 및 IR 영역에 따라 뚜렷하게 다른 분포 양상을 보였으며, LSC에서 가장 많은 변이가 집중되어 나타났다 (Fig. 5). LSC 영역에서는 총 117개의 InDel과 1,026개의 SNP가 검출되었으며, SSC 및 IR 영역을 포함한 기타 구간에서는 40개의 InDel과 452개의 SNP가 분포하는 것으로 나타났다 (Table 8

Table 6. Simple sequence repeats (SSRs) identified in the chloroplast genomes of *A. villosum* and *A. kwangsiensis*.

Repeat motif type	<i>A. villosum</i>	<i>A. kwangsiensis</i>
Mononucleotide (≥10 bp)	91	58
Dinucleotide (≥6 repeats)	14	11
Trinucleotide (≥5 repeats)	1	0
Total	106	69

Table 7. Summary of sequence variants detected between *A. villosum* and *A. kwangsiensis*.

SNPs	Insertions (INS)	Deletions (DEL)	Total variants
1,478	81	76	1,635

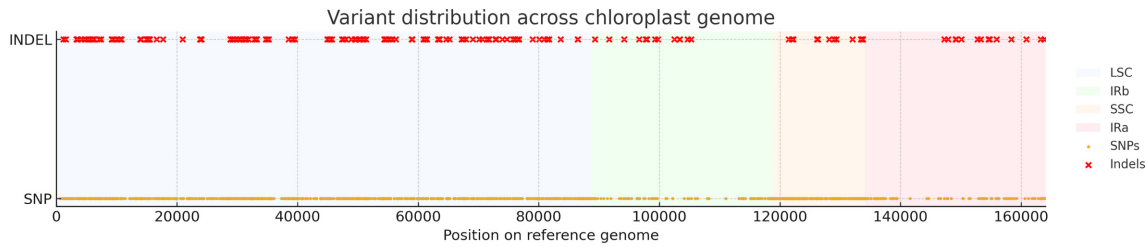


Fig. 5. Distribution of sequence variants across the chloroplast genomes of *A. villosum* and *A. kwangsiensis*. Single nucleotide polymorphisms (SNPs) and insertions/deletions (InDels) between *A. villosum* and *A. kwangsiensis* were identified using the complete chloroplast genome of *A. kwangsiensis* as the reference. Variants are plotted along the genome coordinate axis, with SNPs shown in orange and InDels in red. The structural regions of the chloroplast genome—the large single-copy (LSC), small single-copy (SSC), and inverted repeat (IR) regions—are indicated by semi-transparent background shading.

Table 8. Distribution of SNPs and InDels across chloroplast genomic regions of *A. villosum* and *A. kwangsiensis*.

Region	InDel	SNP	Total
LSC	117	1,026	1,143
SSC	12	261	273
IRb	13	102	115
IRa	15	89	104
Total	157	1,478	1,635

Table 9. Portion of representative sequence variants detected between *A. villosum* and *A. kwangsiensis*.

Position (<i>A. villosum</i>)	Region	Variant type	Base change
119	LSC	SNP	A → G
204	LSC	InDel	T → TA
1,532	LSC	SNP	C → A
32,110	LSC	InDel	AT → A
90,215	IRb	SNP	G → T
123,522	SSC	InDel	A → ATAT
150,993	IRa	SNP	T → C

and 9).

본 변이분석 결과 SSR, InDel, SNP에서 두 종 간 명확한 변이 패턴이 확인되었으며, 특히 LSC 영역에서 높은 변이 밀도가 관찰되어 종 간 구분을 위한 유용한 분자적 표지 영역임을 시사한다. 이러한 변이 특성은 후속 마커 개발과 약재 종 판별 연구에서 핵심적인 근거로 활용될 수 있다.

본 연구에서 발굴된 NGS 기반 SSR·SNP·InDel 마커들은 실제 변이와 동일한 반복 길이·유전자형인지 확인하기 위해 향후 PCR 및 Sanger sequencing으로 검증이 필요하다. 아울러 본 분석에서 발굴된 엽록체 유전체 염기서열 기반 종간 후보 변이들은 향후 약재별 추가적인 개체확보를 통해 종내 변이와 구별되는지에 대한 실험적 검증이 필요하다. 즉, 종별 추가 개체를 확보하여 후보 마커들의 증폭 안정성과 적용 가능성을

우선 평가한 후, 서열 분석을 통해 종내 변이와 종간 변이를 비교함으로써 판별력을 확인하는 방식으로 진행되어야 할 것이다. 후보 마커들의 증폭 특이성, 민감성, 종내 다형성 수준을 검토하여 종별로 신뢰도 높은 마커가 선정되면 실제 약재 시료에서 재현성 있는 종 판별이 가능함을 확인하여 최종적으로 약재 식물 종 동정에 적합한 마커 세트가 확립될 예정이다. 개발된 마커는 가공·건조 등으로 형태적 동정이 어려운 실제 유통 약재에서도 안정적으로 증폭되고 중간 분리도가 확보될 경우, 신뢰도 높은 판별법으로 활용될 수 있을 것이다. 또한 마커 조합을 이용한 다중 검증 체계를 구축하면 혼입·위변조 감별의 정확성을 높여 유통 관리와 품질 보증에 실질적인 기여를 할 것으로 기대된다.

본 연구에서는 사인과 초두구의 엽록체 유전체를 완전 조립하여 구조·유전자 구성을 비교한 결과, 두 종 모두 근연 분류군과 일관된 유전체 특성을 보였으며, 계통 분석에서도 각각 *Wurfbainia* 속과 *Alpinia* 속에 높은 지지값으로 안정적으로 클러스터링되어 시료 동정의 분자적 타당성이 확인되었다. 또한 SSR, InDel, SNP 분석을 통해 두 종 간 뚜렷한 변이 차이가 존재함을 규명하였고, 특히 LSC 영역에서 높은 변이 밀도가 관찰되어 종간 구분에 유용한 후보 변이 영역임을 제시하였다. 이러한 변이 정보를 바탕으로 도출된 후보 마커들은 향후 종내 변이 검증 및 증폭 특이성 평가를 통해 약재 감별용 실용 마커로 확립될 수 있으며, 궁극적으로 형태 판별이 어려운 실제 유통 약재에서도 신뢰도 높은 종 판별을 가능하게 하여 혼입·위변조 방지 및 품질 관리 향상에 기여할 것으로 기대된다.

감사의 글

본 연구는 2025년도 식품의약품안전처 부정·불량식품의 스마트 감시 시스템 및 판별기술 고도화 사업(22193생산안전471-3)의 지원에 의하여 이루어진 결과로 이에 감사드립니다. 그리고 본 연구에 사용한 시료를 보내주신 원광대학교 본초학과 관계자분께 감사드립니다.

REFERENCES

- Bolger A, Lohse M and Usadel B.** (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. 30:2114-2120.
- de Boer H, Newman M, Poulsen A, Droop A, Fér T, Hiên L, Hlavatá K, Lamxay V, Richardson J, Steffen K and Leong-Škorničková J.** (2018). Convergent morphology in *Alpinieae* (Zingiberaceae): Recircumscribing *Amomum* as a monophyletic genus. *Taxon*. 67:6-36.
- Dierckxsens N, Mardulyn P and Smits G.** (2017). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research*. 45:18.
- Delin W and Larsen K.** (2000). Zingiberaceae. In: *Flora of China*, Science Press. Beijing, China and Missouri Botanical Garden Press. St. Louis, USA. 24:322-377.
- Feng L, Wang Z, Lei Z, Zhang X, Zhai B, Sun J, Guo D, Wang D, Luan F, Zou J and Shi Y.** (2024). *Amomum villosum* Lour.: An insight into ethnopharmacological, phytochemical, and pharmacological overview. *Journal of Ethnopharmacology*. 335: 118615. <https://www.sciencedirect.com/science/article/pii/S0378874124009140> (cited by 2025 Dec 09).
- Kato K, Misawa K, Kuma K and Miyata T.** (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*. 30:3059-3066.
- Kim K, Lee SC, Lee J, Yu Y, Yang K, Choi BS, Koh HJ, Waminal NE, Choi HI, Kim NH, Jang W, Park HS, Lee J, Lee HO, Joh HJ, Lee HJ, Park JY, Perumal S, Jayakodi M, Lee YS, Kim B, Copetti D, Kim S, Kim S, Lim KB, Kim YD, Lee J, Cho KS, Park BS, Wing RA and Yang TJ.** (2015). Complete chloroplast and ribosomal sequences for 30 accessions elucidate evolution of *Oryza* AA genome species. *Scientific Reports*. 5:15655. <https://www.nature.com/articles/srep15655> (cited by 2025 Dec 09).
- Kress W, Prince L and Williams K.** (2002). The phylogeny and a new classification of the gingers(Zingiberaceae): Evidence from molecular data. *American Journal of Botany*. 89:1682-1696.
- Kuo P, Henderson IR and Lambing C.** (2022). CTAB DNA extraction and genotyping-by-sequencing to map meiotic crossovers in plants. *Methods in Molecular Biology*. 2484:43-53.
- Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C and Salzberg SL.** (2004). Versatile and open software for comparing large genomes. *Genome Biology*. 5:12.
- Li B, Pan J, Song F, Zeng Q, Guan M, Zhang W, Hu J, Wang A and Gao H.** (2020). Development of nuclear SSR and chloroplast genome markers in diverse *Liriodendron chinense* germplasm based on low-coverage whole genome sequencing. *Biological Research*. 53:21.
- Li D-M, Zhu G-F, Xu Y-C, Ye Y-J and Liu J-M.** (2020). Complete chloroplast genomes of three medicinal *Alpinia* species: Genome organization, comparative analyses and phylogenetic relationships in family Zingiberaceae. *Plants*. 9:286. <https://www.mdpi.com/2223-7747/9/2/286> (cited by 2025 Nov 28).
- Li X, Fan T, Zou P, Zhang W, Wu X, Zhang Y and Liao J.** (2020). Can the anatomy of abnormal flowers elucidate relationships of the androecial members in the ginger(Zingiberaceae)? *EvoDevo*. 11:12. <https://link.springer.com/article/10.1186/s13227-020-00157-8> (cited by 2025 Nov 28).
- Liao C, Hyde K, Chethana K, Dong W, Yang Y and Doilom M.** (2024). Three new *Periconia* species isolated from *Wurfbainia villosa* in Guangdong, China: A discussion on the doubtful taxa clustering in this genus. *Diversity*. 16:141. <https://www.mdpi.com/1424-2818/16/3/141> (cited by 2025 Nov 28).
- Park YG, Kim SS and Seo SR.** (2023). Immunosuppressive effects of *Amomum muricarpum* Elmer in Lipopolysaccharide triggered inflammatory signaling pathway. *Korean Journal of Medicinal Crop Science*. 31:107-115.
- Song W, Chuxuan J, Chen Z, Cai H, Wu X, Shi C and Wang S.** (2022). Comparative analysis of the complete chloroplast genomes of nine *Musa* species: Genomic features, comparative analysis, and phylogenetic implications. *Frontiers in Plant Science*. 13:832884. <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2022.832884/full> (cited by 2025 Dec 09).
- Tamura K, Stecher G and Kumar S.** (2021). MEGA11: Molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*. 38:3022-3027.
- Tillich M, Lehwerk P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R and Greiner S.** (2017). GeSeq: Versatile and accurate annotation of organelle genomes. *Nucleic Acids Research*. 45:6-11.
- Tran TM, Nguyen TMT and Tran TH.** (2018). Study on chemical constituents and biological activities of *Alpinia kwangsiensis* collected in Vietnam. *Vietnam Journal of Science and Technology-Hanoi University of Science and Technology*. 56:273-278.
- Wang C, Li Y, Yang G, Zhang W and Guo C.** (2024). Comparative analysis of chloroplast genomes and phylogenetic relationships in the endemic Chinese bamboo *Gelidocalamus* (Bambusoideae). *Frontiers in Plant Science*. 15:1470311. <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2024.1470311/full> (cited by 2025 Dec 09).
- Yang H, Wang L, Chen H, Jiang M, Wu W, Liu S, Wang J and Liu C.** (2021). Phylogenetic analysis and development of molecular markers for five medicinal *Alpinia* species based on complete plastome sequences. *BMC Plant Biology*. 21:431.
- Yang P, Zhao HY, Wei JS, Zhao YY, Lin XJ, Su J, Li FP, Li M, Ma DM, Tan XK, Liang HL, Sun YW, Zhan RT, He GZ, Zhou XF and Yang JF.** (2022). Chromosome-level genome assembly and functional characterization of terpene synthases provide insights into the volatile terpenoid biosynthesis of *Wurfbainia villosa*. *The Plant Journal*. 112:630-645.
- Zhang Y, Song M-F, Li Y, Sun H-F, Tang D-Y, Xu A-S, Yin C-Y, Zhang Z-L and Zhang L-X.** (2021). Complete chloroplast genome analysis of two important medicinal *Alpinia* species: *Alpinia galanga* and *Alpinia kwangsiensis*. *Frontiers in Plant Science*. 12:705892. <https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2021.705892/full> (cited by 2025 Dec 09).
- Zhang Y, Yang C, Li H and Li G.** (2025). A chromosome-level genome assembly of *Wurfbainia villosa*(Zingiberaceae). *Scientific Data*. 12:1490. <https://www.nature.com/articles/s41597-025-05874-0> (cited by 2025 Dec 09).
- Zhu Y, Zhang X, Yan S, Feng C, Wang D, Yang W, Daud MK, Xiang J and Mei L.** (2023). SSR identification and phylogenetic analysis in four plant species based on complete chloroplast genome sequences. *Plasmid*. 125:102670. <https://www.sciencedirect.com/science/article/pii/S0147619X2300001X> (cited by 2025 Dec 09).