THAI NGUYEN UNIVERSITY
UNIVERSITY OF EDUCATION

PHO THI THUY HANG

THE STUDY OF CHLOROPLAST GENOME CHARACTERISTICS AND BIOACTIVE COMPOUNDS OF SOME *Adinandra* SPECIES

Speciality: Genetics
Code: 9420121

DISSERTATION SUMMARY

THAI NGUYEN, NĂM 2024

The dissertation was finished at:

UNIVERSITY OF EDUCATION - THAI NGUYEN UNIVERSITY

Supervisors:

1. Assoc. Prof. Dr. Nguyen Huu Quan
2. Dr. Nguyen Thi Thu Nga

Reviewer 1:……

Reviewer 2:……

Reviewer 3:…….

The dissertation will be defended in the university committee:

UNIVERSITY OF EDUCATION - THAI NGUYEN UNIVERSITY

At ……., 2024

The dissertation can be read at:

- National library of Vietnam;
- Digital Center - Thai Nguyen University;
- Library of Education

INTRODUCTION

1. Problem statement

The genus *Adinandra* has been identified with approximately 85 species worldwide, of which about 17 species are distributed in Vietnam [16], [97], [159]. According to the Vietnam Red Book, some species of the *Adinandra* genus are valuable and rare genetic resources that are at risk of extinction. They are currently classified as "Vulnerable" (VU), such as *A. megaphylla Hu* [2]. Therefore, accurate species identification is essential for the conservation and development of these rare species. Molecular biology methods using DNA barcodes provide higher accuracy in species identification and differentiation compared to traditional morphological comparison methods.

Although the chloroplast genome is highly conserved, there are still regions that are prone to variation. Differences in the nucleotide sequences of genes in these variable regions form the basis for distinguishing one species from another and determining genetic relationships between species at the molecular level. Currently, there is very limited information on the chloroplast genome of species within the *Adinandra* genus, with only four out of 85 species having their chloroplast genome fully sequenced. Publications on the use and proposal of DNA barcodes for species identification within the *Adinandra* genus are scarce, with only the matK gene being proposed as a DNA barcode to identify species like *A. megaphylla* and *A. lienii* [13], [102].

In a project screening the biological activities of plants in Vietnam, extracts from several species of the genus *Adinandra* (family Pentaphylacaceae) have been identified to possess anti-cancer activities [14], [108], [136]. Additionally, some studies have shown that species of the *Adinandra* genus exhibit antibacterial, anti-inflammatory, and antioxidant effects, as well as being used in the treatment of sprains and snake bites [3], [6], [38], [106]. However, global research on the chemical composition of the *Adinandra* genus has primarily focused on *A. nitida*, with many other species in the genus remaining understudied. In Vietnam, research on isolating and testing the biological activities of new compounds has been conducted on *A. hainanensis*, *A. poilanei*, and *A. lienii* out of the 17 species identified.

Based on these reasons, the dissertation is carried out with the topic: "Study on the characteristics of the chloroplast genome and bioactive compounds of some *Adinandra* species."

2. Research objectives

- Analyze the characteristics of the chloroplast genome of *A. bockiana*.

- Analyze the genetic relationships between species and propose DNA barcode candidates to support species identification within the *Adinandra* genus.

- Identify the chemical composition and biological activities of compounds isolated from three species of the *Adinandra* genus.

3. Research Content

Content 1: Study the characteristics of the chloroplast genome of A. bockiana

- Perform a detailed analysis of the chloroplast genome characteristics of *A. bockiana*.

- Compare the chloroplast genome of *A. bockiana* with other species of the *Adinandra* genus available on GenBank.

Content 2: Study the phylogeny of the Adinandra genus and search for potential genes to propose as chloroplast DNA barcodes

- Construct a phylogenetic tree based on chloroplast genome sequences and the *matK, trnL,* and *rbcL* gene sequences of species within the *Adinandra* genus.

- Analyze the phylogenetic tree diagrams and search for DNA barcode candidates for species identification within the *Adinandra* genus.

Content 3: Study the chemical composition and evaluate the biological activities of compounds isolated from three selected species

- Isolate compounds using chromatographic methods.

- Determine the chemical structure of isolated compounds based on physical parameter measurements, spectroscopic methods, and by referencing literature.

- Evaluate the biological activities (antibacterial, cytotoxicity against cancer cells, α-glucosidase inhibition) of selected compounds isolated from the three studied species.

4. New contributions of the dissertation

(1) This dissertation is a pioneering study both in Vietnam and globally, providing a detailed and comprehensive analysis of the chloroplast genome characteristics of *A. bockiana*. It proposes the *matK* and *rbcL* gene regions as potential DNA barcode candidates for species identification within the *Adinandra* genus.

(2) The dissertation is the first study to isolate 37 compounds from the leaves of *A. megaphylla*, *A. bockiana*, and *A. glischroloma*, including two new compounds (debutyldorycnic acid and adinanquercetiside, isolated from the leaves of *A. megaphylla*).

(3) For the first time, the compound 23-hydroxyursolic acid from *A. glischroloma* has been found to inhibit α-glucosidase and exhibit cytotoxicity against liver cancer (HepG2) and breast cancer (MCF-7) cell lines. The compound ursolic acid from *A. megaphylla*, *A. bockiana*, and *A. glischroloma* has shown strong inhibitory effects on the growth of the bacterium *Pseudomonas aeruginosa*. Additionally, isoquercetin (from *A. megaphylla* and *A. glischroloma*) strongly inhibits the growth of *Citrobacter freundii* and *Streptococcus milleri*.

5. Scientific and practical significance of the dissertation

Scientific significance

The research findings of the dissertation provide a foundation for applying the proposed DNA barcodes in species identification and analyzing genetic relationships among species within the *Adinandra* genus. The dissertation has identified the chemical composition of the leaves of three *Adinandra* species in Vietnam, highlighting differences from the *Adinandra* species in China. Specifically, *Adinandra* species in Vietnam are rich in triterpenoid compounds, whereas Chinese *Adinandra* species are rich in flavonoid compounds.

The results of the biological activity tests of the isolated compounds provide a scientific basis for explaining the antibacterial and cytotoxic activities of the extracts, as well as the use of certain *Adinandra* species in cancer treatment in Vietnam.

The research articles published in domestic and international scientific journals, along with the gene sequences submitted to GenBank, are valuable references for research and teaching.

Practical significance

The discovery of the α-glucosidase inhibitory activity and the cytotoxicity against HepG2 and MCF-7 cell lines by the compound

23-hydroxyursolic acid may provide a basis and open up opportunities for the development of new treatments for diabetes, liver cancer, and breast cancer.

The discovery of the strong inhibitory effect of ursolic acid on the growth of *P. aeruginosa*, and the inhibition of *C. freundii* and *S. milleri* by isoquercetin, may open opportunities for utilizing plant-derived compounds in treating diseases caused by these bacteria.

Chapter 1. LITERATURE REVIEW

1.1. The genus *Adinandra* and the chloroplast genome

1.1.1. Characteristics of the genus Adinandra

1.1.1.1. Taxonomy of the genus Adinandra

1.1.1.2. Morphological characteristics of the genus Adinandra

1.1.1.3. Distribution of Adinandra species in Vietnam

1.1.2. Research on the chloroplast genome

1.1.2.1. Chloroplast genome of higher plants

1.1.2.2. Chloroplast genome of some species of the genus Adinandra

Currently, only four species of the *Adinandra* genus have had their entire chloroplast genomes sequenced and registered on GenBank: *A. megaphylla* (accession number MW697901.1), *A. millettii* (accession number NC_035678.1), *A. bockiana* (accession number MW699853.1), and *A. angustifolia* (accession number NC_035653.1) [104], [105], [145], [146].

The chloroplast genome has a typical structure with four regions: a Large Single Copy (LSC) region of about 86 kb, a Small Single Copy (SSC) region of about 18 kb, and a pair of Inverted Repeat (IRa and IRb) regions, each over 26 kb in size. The chloroplast genome size ranges from 156 to 156.5 kb, containing 129-132 genes. The average GC content is approximately 37.4% [104], [145], [146]. Apart from the studies by Nguyen et al. (2019, 2021) that proposed and utilized the *matK* gene for identifying *A. megaphylla* and *A. lienii*, no other studies have proposed new barcode candidates, despite the presence of numerous genes in the chloroplast genome [13], [102].

1.2. Molecular evolutionary genetics analysis

1.2.1. Genetic basis of molecular evolution

1.2.2. Molecular evolution analysis based on the chloroplast genome

1.2.2.1. Research on genetic relationships among plant species based on the chloroplast genome

The 2008 International Botanical Congress pointed out that the chloroplast genome contains a wealth of information similar to that of short mitochondrial barcode sequences used in animals. As a result, the chloroplast genome was proposed as a super barcode [37].

1.2.2.2. Research on genetic relationships among plant species based on chloroplast DNA barcodes

DNA Barcoding

Research on genetic relationships and species identification using chloroplast DNA barcodes

In the chloroplast genome, seven DNA regions have been selected as candidate DNA barcodes for land plants: the *matK, rbcL, rpoB,* and *rpoC1* genes, as well as the *psbK-psbI, atpF-atpH,* and *trnH-psbA* intergenic spacers. Among these, four regions are coding gene segments (*matK, rbcL, rpoB,* and *rpoC1*), and three are non-coding intergenic spacers (*atpF-atpH, trnH-psbA,* and *psbK-psbI*) [32], [60].

However, each species or genus may have specific DNA barcodes that are more suitable. Therefore, identifying potential genes to serve as DNA barcodes for the specific study subject is essential.

1.3. Chemical composition and biological activity of the genus *Adinandra*

1.3.1. Chemical composition of the genus Adinandra

There have been numerous studies worldwide and in Vietnam on the chemical components of species within the *Adinandra* genus. Most studies agree that *Adinandra* species contain major groups of compounds such as flavonoids, phenolics, triterpenoids, triterpenoid saponins, aldehydes, and coumarins, with flavonoids and triterpenoids being the primary components [84], [85], [86], [138], [153].

Currently, research on isolating new compounds has been conducted for only a few species, including *A. nitida*, *A. lienii*, *A. poilanei*, and *A. hainanensis*, out of the 85 species in the *Adinandra* genus. These studies have isolated 47 compounds, specifically: flavonoids (8 compounds), triterpenoid saponins (8 compounds), triterpenoids (17 compounds), sterols (4 compounds), and phenolics (3 compounds). Additionally, compounds from other groups (7 compounds) such as diterpenoids, coumarins, aldehydes, quinones,

lignans, tocopherols, and phytols have also been isolated from some *Adinandra* species.

1.3.2. Biological activity of the genus Adinandra

1.3.2.1. Biological activity of extracts and compounds from Adinandra species

Extracts and compounds from *Adinandra* species exhibit various biological activities, including antibacterial, antioxidant, anticancer, anti-allergic, lipid-lowering, antihypertensive, liver-protective, and gastric-protective effects. Notably, extracts from *Adinandra* species in Vietnam, such as *A. bockiana* and *A. megaphylla*, have been extensively studied for their antibacterial, antioxidant, and anticancer properties.

1.3.2.2. Biological activity of compounds isolated from Adinandra species

Globally, research on the biological activity of compounds isolated from *Adinandra* species primarily focuses on antioxidant and anticancer activities [49], [86], [149], [150]. Additionally, *Adinandra* species have demonstrated other biological activities such as anti-allergic, lipid-lowering, antihypertensive, liver-protective, and α-glucosidase-inhibitory effects [3], [11], [86], [134], [147], [148].

Chapter 2. MATERIALS AND METHODS

2.1. Research materials

2.1.1. Plant materials

Three species of the genus *Adinandra* were used: *A. megaphylla* Hu and *A. bockiana* E. Pritz. ex Diels collected in Liem Phu commune, Van Ban district, Lao Cai province. Specifically, *A. megaphylla* was collected at an altitude of 1200-1800 m, with coordinates 21°59'15"N; 104°19'28"E, and *A. bockiana* was collected at an altitude of 800 m, with coordinates 21°59′15"N; 104°19′28"E. The species *A. glischroloma* was collected in Y Ty commune, Bat Xat district, Lao Cai province at an altitude of 1844 m, with coordinates 103°37′42″E, 22°37′35″N. Leaves of these three species were used to prepare extracts, isolate compounds, and test the biological activities of the obtained compounds.

2.1.2. Bacterial strains for testing

The bacterial strains used in the study to determine the antibacterial activity of compounds obtained from species of the genus *Adinandra* include: *Citrobacter freundii*, *Escherichia coli*

(ATCC25922), *Pseudomonas aeruginosa* (ATCC15442), *Staphylococcus aureus* (ATCC13709), and *Streptococcus milleri*. These are pathogenic bacteria, with *S. aureus* being a Gram-positive bacterium, and the remaining strains being Gram-negative.

2.1.3. Cell lines for testing

The cell lines used for testing include: lung carcinoma cells (SK-LU-1), gastric carcinoma cells (MKN-7), liver carcinoma cells (HepG2), and breast carcinoma cells (MCF7), with human embryonic kidney cells (HEK-293A) used as controls.

2.1.4. Research data

Data from the chloroplast genome of several species published on GenBank were used, including *A. megaphylla* (Accession number MW697901.1) [104], *A. millettii* (Accession number NC_035678.1) [146], and *A. angustifolia* (Accession number NC_035653.1) [145], to compare with the chloroplast genome of *A. bockiana* regarding the genetic diversity of the chloroplast genome in the genus *Adinandra*. Additionally, data from other gene sequences were accessed on GenBank at the address https://www.ncbi.nlm.nih.gov/nucleotide/ [160].

2.2. Chemicals, equipment, and research locations

2.2.1. Chemicals and research equipment

2.2.2. Research locations

The experiments were conducted at the laboratories of the Department of Biology, Thai Nguyen University of Education; the Key laboratory of Genetic technology, Institute of Biotechnology; and the Marine Biochemistry Institute at the Vietnam Academy of Science and Technology.

2.3. Research methods

2.3.1. Method for studying the chloroplast genome characteristics

The chloroplast genomes of *A. angustifolia* (GenBank accession number MF179491) and *A. millettii* (GenBank accession number MF179492) [145], [146] were used for comparison with the chloroplast genomes of the studied species.

2.3.2. Method for molecular evolutionary genetic analysis

The phylogenetic tree was constructed based on the nucleotide sequences of the *matK*, *trnL*, and *rbcL* genes using the Maximum Likelihood method with bootstrap values repeated 1000 times using the Mega X software [79].

2.3.3. Method for studying chemical composition and biological activity

2.3.3.1. Methods for studying chemical composition

Extraction and residue preparation: Dry leaf powder of each species (*A. megaphylla* - 3.5 kg; *A. glischroloma* - 3.2 kg; and *A. bockiana* - 3.3 kg) was used to prepare total extracts and residues as outlined in Figure 2.2.

Compound isolation: Methods such as Thin Layer Chromatography (TLC), Column Chromatography (CC), and Gas Chromatography (GC) were employed [12].

Chemical structure determination: The chemical structures of the compounds were determined using physical parameter measurements and spectroscopy methods (NMR) with modern equipment, combined with analysis and literature reference searches.

2.3.3.2. Methods for determining the biological activity of compounds

Antibacterial activity testing: Conducted using the agar diffusion method according to the study by Mahesh and Satish (2008) [91].

Cytotoxic activity testing: Performed using the method described by Skehan et al. (1990) [121].

α-Glucosidase inhibition activity testing: According to Tran et al. (2014) [128].

2.3.4. Data processing and results analysis

Statistical analysis was performed using SPSS software and bioinformatics tools: BioEdit, BLAST in NCBI for gene analysis [58], [75], [85], [160].

Chapter 3. RESULTS AND DISCUSSION

3.1. Characteristics of the chloroplast genome of *A. bockiana*

3.1.1. Structure and composition of the chloroplast genome of *A. bockiana*

The complete chloroplast genome of *A. bockiana* is 156284 bp in size and has a typical structure consisting of four regions: a large single-copy region (LSC) of 85693 bp, a small single-copy region (SSC) of 18411 bp, and a pair of inverted repeat regions (IR) of 26090 bp, with a GC content of 37.4% (Figure 3.1).

Analysis of the chloroplast genome of *A. bockiana* reveals 129 genes, including 84 protein-coding genes (PCGs), 37 tRNA genes,

and 8 rRNA genes. Based on function, these 129 genes are categorized into 18 groups (Table 3.1).

3.1.2. Repetitive sequence data for A. bockiana

The total number of simple sequence repeats (SSRs) in the chloroplast genome is 51, with repeat types including A (18 SSRs), T (32 SSRs), and G or C (1 SSR) ranging from 10 to 19 bp in length. Most SSRs are located in the LSC region (35), with only a few SSRs found in the SSC and IR regions, with 6 and 4 SSRs, respectively. The chloroplast genome was found to have 70 repeat sequences, including 48 identical repeats, 20 direct repeats, and 2 inverted repeats, with no additional repeats (Figure 3.2).

3.1.3. Codon usage frequency of protein-coding genes in the chloroplast genome of A. bockiana

In the chloroplast genome of *A. bockiana*, 52057 codons were found in the coding regions of protein-coding genes (Table 3.2). Codons ending in A and U were found more frequently than those ending in G and C. Among the 64 codon types, 30 types were used more frequently than expected in a balanced state (RCSU > 1), while 29 types were used less frequently (RCSU < 1). The initiation codons AUG (encoding methionine) and UGG (encoding tryptophan) showed no deviation (RCSU = 1) from the expected codon usage in a balanced state. Termination codons include UAA, UGA, and UAG (Table 3.2).

3.1.4. Comparison of the chloroplast genome of A. bockiana with A. megaphylla, A. millettii, and A. angustifolia

3.1.4.1. Variation in size and gene number in the chloroplast genome

Comparing the complete chloroplast genome of *A. bockiana* [105] with those of *A. megaphylla* [104], *A. angustifolia* [145], and *A. millettii* [146] reveals diversity in genome size, the size of each region, and gene number (Table 3.3).

Table 3.3. Diversity in size and gene number in the chloroplast genome of some species of the genus *Adinandra*

TT	Genomic characteristics	Species name			
		A. bockiana	*A. megaphylla*	*A. millettii*	*A. angustifolia*
1	Genomic size (bp)	156284	156298	156311	156344

2	LSC size (bp)	85693	85688	85698	85743
3	SSC size (bp)	18411	18424	18421	18419
4	IR size (bp)	26090	26093	26096	26091
5	GC content (%)	37,4	37,4	37,4	37,4
6	Number of genes	129	131	132	132
7	Number of PCG	84	86	87	87
8	Number of tRNA	37	37	37	37
9	Number of rRNA	8	8	8	8

3.1.4.2. Variations in chloroplast genome sequences

Compared to the IR region, the LSC and SSC regions exhibit higher variation. Coding regions tend to be more conserved than non-coding regions, with most variations detected primarily in non-coding regions. The genes *matK*, *psaA*, *ndhK*, *ndhG*, and *rbcL* show different nucleotide sequences among the four species: *A. bockiana*, *A. megaphylla*, *A. millettii*, and *A. angustifolia* (Figure 3.3).

The nucleotide diversity (Pi) values among the chloroplast genome sequences of the four *Adinandra* species differ between species and also vary across regions of the chloroplast genome within the same species. The average Pi value for the four species *A. bockiana*, *A. megaphylla*, *A. millettii*, and *A. angustifolia* is 0.00105.

3.1.4.3. Contraction and expansion of the chloroplast genome

The size of each IR region in the four chloroplast genomes ranges from 26090 to 26096 bp. The analysis results indicate that there is no expansion or contraction of the IR region in the chloroplast genomes of the studied species. In the chloroplast genomes of *A. megaphylla*, *A. millettii*, and *A. angustifolia*, the *ycf1* gene spans 4543 bp in the SSC region (the gene's 5' end) and 1067 bp in the IRa region (the gene's 3' end), with the *ycf1* gene forming the boundary between IRa and SSC. However, the *ycf1* gene is not present in this region in the chloroplast genome of *A. bockiana* (Figure 3.5).

3.2. Analysis of genetic relationships and phylogeny of the genus *Adinandra*

3.2.1. Analysis of phylogenetic relationships based on complete chloroplast genome sequences

The phylogenetic tree established based on complete chloroplast genome sequences shows very high reliability and stability, with bootstrap values of 100% at all branches. The four species *A. bockiana*, *A. megaphylla*, *A. millettii*, and *A. angustifolia* all form a single clade with a bootstrap value of 100% (Figure 3.6).

Figure 3.6. Phylogenetic tree of *A. bockiana* and other species based on complete chloroplast genome sequences

3.2.2. Analysis of genetic relationships based on matK, trnL, and rbcL gene sequences

3.2.2.1. Characteristics of matK, trnL, and rbcL genes in A. bockiana

3.2.2.2. Analysis of genetic relationships based on matK gene sequences

A. bockiana shows 100% sequence similarity with *A. megaphylla* and *A. nitida*. The *matK* gene sequence of *A. bockiana* is highly similar to the *matK* gene sequences of other species in the genus *Adinandra*, with similarities ranging from 99.27% to 100% (Table 3.5).

The sequence divergence coefficient of the *matK* gene between *A. bockiana* and other species in GenBank ranges from 0.001 to 1.100 (Table 3.6). The smallest divergence coefficient is 0.001 (with *A. formosana*), followed by 0.003 (with *A. integerrima* and *A. angustifolia*). *A. megaphylla* and *A. nitida* show no sequence divergence in the *matK* gene compared to *A. bockiana* (divergence coefficient is 0.000).

The phylogenetic tree based on *matK* gene sequences provides very high reliability and stability, with bootstrap values mostly greater than 90% for the majority of branches (Figure 3.7).

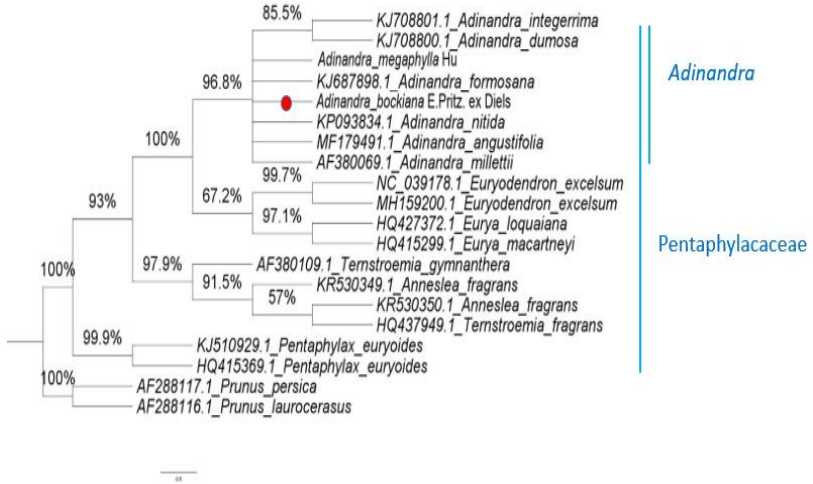

Figure 3.7. Phylogenetic tree of *A. bockiana* and related species based on *matK* gene sequences

3.2.2.3. Analysis of genetic relationships based on trnL gene sequences

The *trnL* gene sequence of *A. bockiana* shows a very high similarity (ranging from 99.33% to 100%) with other species in the genus *Adinandra*. Among them, *A. bockiana* has the highest similarity (100%) with *A. glischroloma* and *A. hainanensis* (Table 3.7).

Analysis of sequence divergence in the *trnL* gene between *A. bockiana* and other species in GenBank shows that the divergence coefficient ranges from 0.002 to 1.130. The *trnL* gene sequence of *A. bockiana* has the least divergence (divergence coefficient of 0.002) with other species in the genus *Adinandra*, such as *A. millettii*, *A. bockiana* (HM061582.1), *A. glischroloma*, *A. hirta* (AF534657.1), *A. hirta* (AF499817.1), *A. hainanensis*, *A. lasiostyla*, and *A. formosana* (Table 3.8).

The phylogenetic tree based on the *trnL* gene sequence (Figure 3.8) shows that the *trnL* gene provides low reliability and stability at the branch of the genus *Adinandra* (bootstrap value of 51%). In Figure 3.8, the genus *Adinandra* is divided into two subgroups. Within each subgroup, the stability is also very low, with bootstrap values of 26, 32, and 69%, respectively.

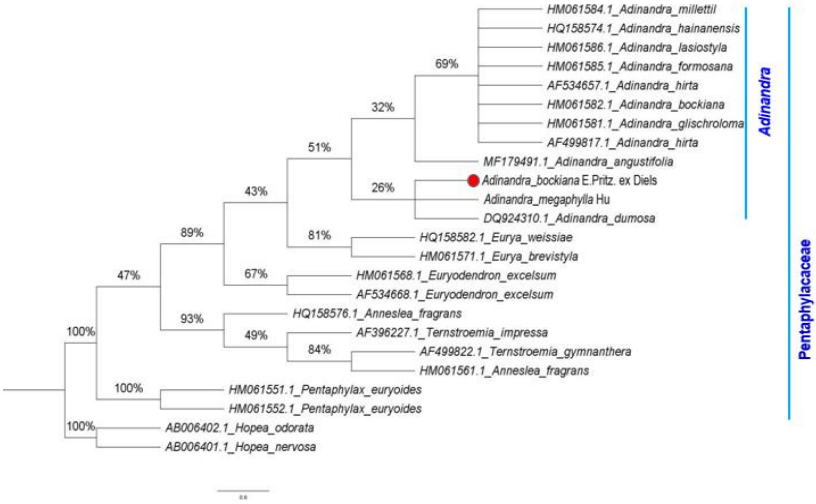

Figure 3.8. Phylogenetic tree of *A. bockiana* and related species based on *trnL* gene sequences

3.2.2.4. Analysis of genetic relationships based on rbcL gene sequences

A. bockiana shows the highest similarity (99.72%) with *A. megaphylla*, *A. millettii*, and *A. angustifolia*, and the lowest similarity (98.18%) with species in the genus *Camellia* (family Theaceae). No *rbcL* sequence is 100% identical to *A. bockiana* (Table 3.9).

The sequence divergence coefficient of the *rbcL* gene between *A. bockiana* and 19 other species ranges from 0.003 to 0.018 (Table 3.10). *A. millettii*, *A. angustifolia*, and *A. megaphylla* have the closest genetic relationships with *A. bockiana* (divergence coefficient of 0.003), followed by *A. glischroloma* and *A. formosana* with divergence coefficients of 0.004 and 0.006, respectively.

The phylogenetic tree based on *rbcL* gene sequences shows that species in the genus *Adinandra* are clustered on the same branch and are divided into two subgroups with very close genetic relationships (bootstrap value of 95%). *A. bockiana* and *A. megaphylla* form subgroup 1, while subgroup 2 includes *A. glischroloma*, *A. angustifolia*, *A. millettii*, and *A. formosana* (Figure 3.9).

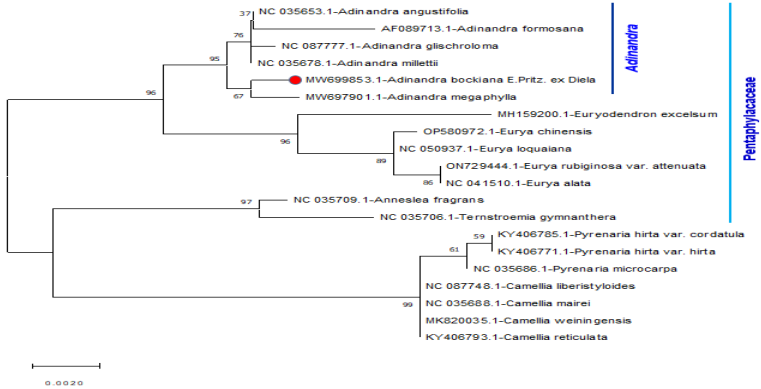

Figure 3.9. Phylogenetic tree of *A. bockiana* and related species based on *rbcL* gene sequences

The results of the similarity analysis, divergence coefficients, and phylogenetic trees for the genes indicate that the *matK* and *rbcL* gene sequences are suitable and recommended as DNA barcode candidates for species identification, classification, and determining the genetic relationships among species in the genus *Adinandra* and some species in the family Pentaphylacaceae.

3.3. Results of chemical composition and biological activity analysis of the three studied species

3.3.1. Chemical structure of some compounds isolated from the three studied species

3.3.1.1. Isolation and structural identification of compounds from A. megaphylla

Fifteen compounds were isolated from the leaves of *A. megaphylla* (Table 3.11, Figure 3.12). Among these, two new compounds are described for the first time: debutyldorycnic acid **(1)** and adinanquercetiside **(2)**; the remaining thirteen compounds (labeled (KH) from **3-15**) are known and include coniferyl aldehyde (3), ursolic acid **(4)**, 4,5-dihydroblumenol **(5)**, methyl gallate **(6)**, 24-hydroxytormentic acid **(7)**, gallic acid **(8)**, convoldorin **(9)**, scopolin **(10)**, isoquercitrin **(11)**, horridin **(12)**, pinoresinol-4'-O-β-D-glucopyranoside **(13)**, syringaresinol β-D-glucoside **(14)**, and camellikaempferoside B **(15)**.

Table 3.11. Results of isolation and chemical structure determination of compounds from the leaves of *A. megaphylla*

TT	Segment	Isolation method	KH compound	TLTK	Compound name	KL compound (mg)
1	F3.2	RP-18 (M:W 1:1)	**AHL2 (3)**	[98]	Coniferyl aldehyde	3,7
2	F5.2	Rửa acetone	**AHL3 (4)**	[19]	Ursolic acid	45,0
3	F5.3	RP-18 (M:W 1:1)	**AHL4 (5)**	[43]	4,5-dihydroblumenol	8,2
4	F5.3.1	Silicagel c, c (D:A 9:1)	**AHL6 (6)**	[118]	Methyl gallate	4,1
5	F6.7.2	Silicagel c, c (D:A 9:1)	**AHL7 (7)**	[61], [80]	24-hydroxytormentic acid	3,1
6			**AHL8 (8)**	[8]	Gallic acid	4,2
7	F9.1	RP-18 (M:W 1:3)	**AHL13 (9)**	[54]	Convoldorin	16,5
8	F9.2.1	RP-18 (M:W 1:4)	**AHL19 (10)**	[1]	Scopoline	5,1
9	F9.3	RP-18 (M:W 1:3)	**AHL15 (1)**		Debutyldorycnic acid	8,6
10	W3.2.3	RP-18 (M:W 1:1)	**WAM1 (11)**	[75]	Isoquercetine	3,6
11			**WAM2 (12)**	[48]	Horidin	4,2
12	W3.4.1	RP-18 (M:W 1:1)	**WAM11A (13)**	[112]	Pinoresinol-4'-O-*β*-D-glucopyranoside	7,1
13	W3.4.2	Silicagel c, c (D:M 1:9)	**WAM16 (14)**	[119]	Syringaresinol-4'-O-*β*-D-glucopyranoside	5,3
14	W3.8.4	RP-18 (M:W 1:3)	**WAM15 (15)**	[143]	Camellikaempferoside B	3,8
15			**WAM11.5 (2)**		Adinanquercetiside	5,1

Note: KH: Label; KL: Mass

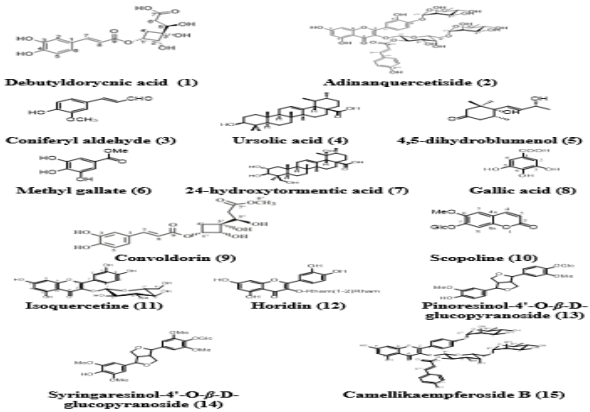

Figure 3.12. Chemical structures of compounds isolated from the leaves of *A. megaphylla*

3.3.1.2. Isolation and structural identification of compounds from A. bockiana

Eight compounds were isolated from the leaves of *A. bockiana* (Table 3.14, Figure 3.13). These compounds have all been found in other plant species, including ursolic acid, which was also isolated from *A. megaphylla*.

Table 3.14. Results of isolation and chemical structure determination of compounds from the leaves of *A. bockiana*

TT	Segment	Isolation method	KH compound	TLTK	Compound name	KL compound (mg)
1	E6.4.2	RP-18 (M:W 1,5:1)	**BA1 (16)**	[70]	*Ent*-kaur-16-en-19-oic-acid	5,0
2	E6.4.3	Silicagel c, c (H:E 9:1)	**BA2 (17)**	[123]	*β*-sitosterol	12,0
3	E10.8	Silicagel c, c (H:A 9:1)	**BA15 (21)**	[137]	Betulin	6,5
4	E10.10	Kết tinh	**BA16 (22)**	[65]	Betulinic acid	4,6
5	E16.1	Silicagel c, c (H:E 1:1)	**BA3 (18)**	[20]	Scopoletin	3,0
6	E16.4	RP-18 (M:W 3:1)	**BA5 (4)**	[19]	Ursolic acid	5,0
7	E21.1	Kết tinh	**BA7 (20)**	[47]	Daucosterol	15,0
8	E21.2	RP-18 (M:W 1,5:1)	**BA4 (19)**	[31]	Sumaresinolic acid	3,0

Note: KH: Label; KL: Mass

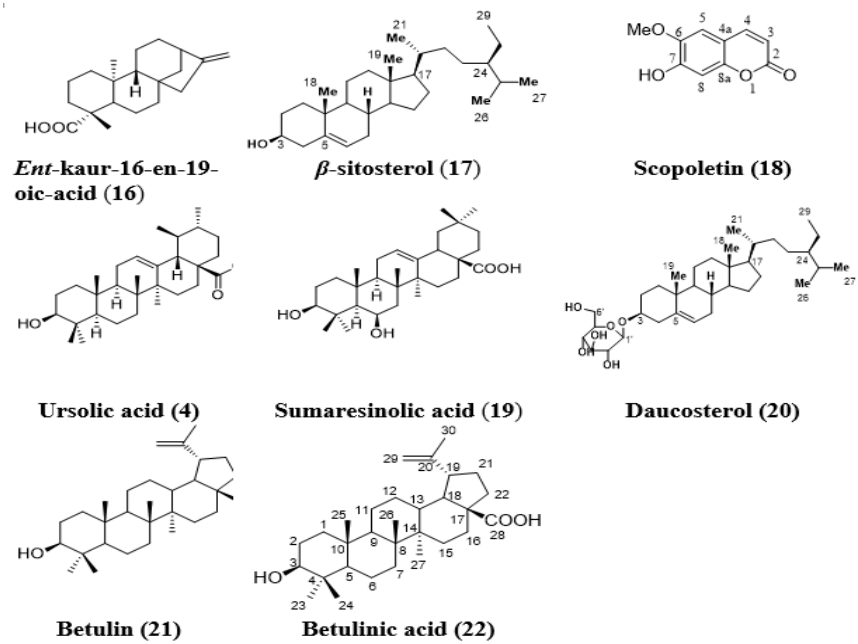

Figure 3.13. Chemical structures of compounds isolated from the leaves of *A. bockiana*

3.3.1.3. Isolation and structural identification of compounds from A. glischroloma

Fourteen compounds were isolated from the leaves of *A. glischroloma*, specifically: 28-nor-urs-12-ene-3β,17-β-diol **(23)**, micromeric acid **(24)**, 23-hydroxy ursolic acid **(25)**, euscaphic acid **(26)**, pomolic acid **(27)**, 3,13-dihydroxy ursolic acid -28,13-olide **(28)**, ursolic acid **(4)**, betulinic acid **(22)**, oleanolic acid **(29)**, ent-kaur-16-en-19-oic acid **(16)**, (3S, 5R, 6S, 9R)-megastigmane-3,9-diol **(30)**, quercetin-3-glucoside or isoquercetin **(11)**, syringaresinol **(31)**, and β-sitosterol **(17)** (Table 3.15, Figure 3.14).

Table 3.15. Results of isolation and chemical structure determination of compounds from the leaves of *A. glischroloma*

TT	Segment	Isolation method	KH compound	TLTK	Compound name	KL compound (mg)
1	E3.7	Silicagel c, c (H:E 9:1)	**AG16 (16)**	[70]	*Ent*-kaur-16-en-19-oic-acid	3,1
2	E4	Silicagel c, c (H:D 5:1)	**AG29 (17)**	[123]	*β*-sistosterol	7,2
3	E4.2	Silicagel c, c (H:A 30:1)	**AG2 (23)**	[28]	28-*nor*-urs-12-ene-3*β*,17 *β*-diol	3,4
4	E6.6.3	RP-18 (M:W 10:1)	**AG14 (22)**	[65]	Betulinic acid	6,2
5	E6.6.5	RP-18 (M:W 10:1)	**AG15 (29)**	[19]	Oleanolic acid	5,5
6	E9.4	Silicagel c, c (H:E 9:1)	**AG10 (28)**	[18]	3, 13-dihydroxy ursolic acid 28, 13-olide	2,2
7	E9.7.1	Sephadex (MeOH)	**AG18 (30)**	[122]	(3S, 5R, 6S, 9R)-megastigmane-3,9-diol	3,2
8	E9.7.4	Silicagel c, c (H:E 9:1)	**AG3 (24)**	[21]	Micromeric acid	2,3
9	E9.7.6	RP-18 (M:W 1:1)	**AG8 (27)**	[127]	Pomolic acid	10,0
10	E12.2	Silicagel c, c (H:A 4:1)	**AG28 (25)**	[22], [27]	23-hydroxyursolic acid	4,2
11			**AG11 (4)**	[19]	Ursolic acid	3,2
12	E15.2.1	Silicagel c, c (D:M 100:1)	**AG21 (31)**	[114]	Syringaresinol	4,0
13	E15.2.5	RP-18 (A:W 1:1)	**AG7 (26)**	[147]	Euscaphic acid	2,0
14	W3.7.1	Sephadex (MeOH)	**AG19 (11)**	[75]	Isoquercetine	6,0

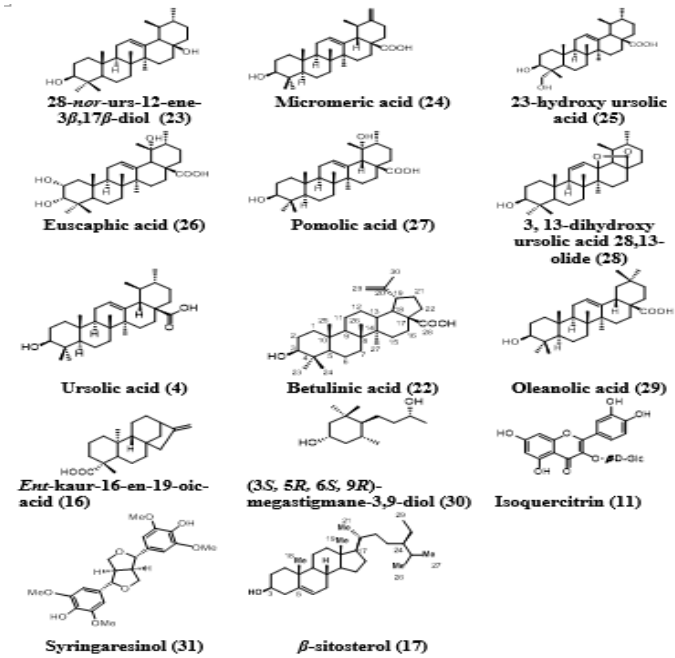

Figure 3.14. Chemical structures of compounds isolated from the leaves of *A. glischroloma*

3.3.2. Biological activity of compounds isolated from the three studied species

3.3.2.1. Antibacterial activity of some isolated compounds

The results of the antibacterial activity of the compounds are presented in Table 3.18.

Two out of fourteen compounds (**(1)** and **(12)**) showed no inhibition of any bacterial strains (H = 0), with **(1)** being a new compound (Debutyldorycnic acid). The remaining compounds exhibited antibacterial activity at varying levels depending on the concentration and the bacterial strain tested. Most compounds at a concentration of 100 µg/mL did not inhibit the growth of any bacterial strains (except **(4)** and **(11)**). For each specific bacterial strain, the antibacterial activity of each compound is proportional to the test concentration, with the compounds showing the strongest activity at 400 µg/mL (Table 3.18).

Table 3.18. Results of antibacterial activity testing of some compounds isolated from the three studied species

KH compound	Concentration (μg/mL)	Bacterial strain and inhibition zone diameter (mm)				
		P. aeruginosa	***S. aureus***	***C. freundii***	***S. milleri***	***E. coli***
1	Did not show antibacterial activity at any concentration (H = 0)					
2	100	0	0	0	0	0
	200	0	0	0	0	0
	400	0	0	0	5,09±1,05	0
3	100	0	0	0	0	0
	200	0	0	0	0	0
	400	0	17,23±0,08	0	12,36±1,12	0
4	100	4,21±0,05	0	0	0	0
	200	20,22±0,98	0	0	15,12±0,89	0
	400	23,02±1,05	0	0	20,13±1,21	10,55±0,73
6	100	0	0	0	0	0
	200	0	0	0	0	0
	400	6,78±2,02	7,56±1,47	0	8,86±0,32	6,24±1,19
7	100	0	0	0	0	0
	200	0	0	0	0	0
	400	0	0	9,21±1,03	0	10,21±0,82
8	100	0	0	0	0	0
	200	16,11±1,03	0	0	17,47±1,14	12,55±0,84
	400	18,26±1,07	16,48±0,09	0	21,21±1,42	14,02±0,04
10	100	0	0	0	0	0
	200	0	0	0	0	0
	400	0	4,12±0,67	11,12±2,03	9,34±0,75	0
11	100	6,67±1,54	0	0	0	0
	200	17,09±1,04	0	17,31±0,25	18,26±1,36	5,08±1,57
	400	21,01±2,11	0	22,18±0,48	23,15±1,66	17,32±1,09
12	Did not show antibacterial activity at any concentration (H = 0)					
17	100	0	0	0	0	0
	200	0	4,56±0,93	0	5,22±0,99	6,36±1,13
	400	0	17,35±1,47	0	17,92±0,08	18,21±1,54
	100	0	0	0	0	0

KH compound	Concentration (μg/mL)	Bacterial strain and inhibition zone diameter (mm)				
		P. aeruginosa	*S. aureus*	*C. freundii*	*S. milleri*	*E. coli*
1	Did not show antibacterial activity at any concentration (H = 0)					
22	200	0	0	6,15±1,32	0	0
	400	8,05±1,04	0	9,65±0,86	0	7,23±0,76
27	100	0	0	0	0	0
	200	0	0	0	0	0
	400	0	6,34±1,05	10,22±1,75	8,58±0,05	5,08±1,52
31	100	0	0	0	0	0
	200	0	0	0	0	0
	400	0	0	0	6,12±0,98	0
DMSO	1%	0	0	0	0	0
Penicillin	200	22,85±0,67	23,01±1,65	18,14±0,21	26,57±1,02	23,00±0,86

*Note: **1**: Debutyldorycnic acid; **2**: Adinanquercetiside, **3**: Coniferyl aldehyde; **4**: Ursolic acid; **6**: Methyl gallate; **7**: 24-hydroxytormentic acid; **8**: Gallic acid; **10**: Scopoline; **11**: Isoquercetine; **12**: Horidin; **17**: β-sistosterol; **22**: Betulinic acid; **27**: Pomolic acid; **31**: Syringaresinol.*

Within the scope of this study, ursolic acid **(4)** and isoquercetin **(11)** were identified as the compounds with the strongest antibacterial activity. Specifically, compound **(4)** strongly inhibited *P. aeruginosa*, while compound **(11)** was effective against *C. freundii* and *S. milleri*.

3.3.2.2. Cytotoxic activity against cancer cells of some isolated compounds

The results of testing the cytotoxic activity of 18 compounds showed that 23-hydroxyursolic acid **(25)** exhibited strong cytotoxicity against both HepG2 (IC50 = 3.28 μg/mL) and MCF-7 (IC50 = 1.16 μg/mL) cell lines, while being non-toxic to normal kidney embryonic cells (HEK-293A).

Table 3.19. Results of cytotoxic activity evaluation of the isolated compounds against cancer cells

KH compound	Compound name	IC_{50} (μg/mL)				
		SK-LU-1	MKN-7	HEK-293A	HepG2	MCF-7
1	Debutyldorycnic acid	> 200	> 200	> 200	> 200	-
2	Adinanquercetiside	> 200	> 200	> 200	> 200	-
4	Ursolic acid	> 200	> 200	> 200	> 200	-
5	4,5-dihydroblumenol	> 200	> 200	> 200	> 200	-
7	24-hydroxytormentic	181,89±	189,03±	191,87±	179,37± 2,86	-

		3,11	3,02	2,94		
9	Convoldorin	> 200	> 200	> 200	> 200	-
10	Scopoline	> 200	> 200	> 200	32,00±0.75	43,63±1,34
12	Horidin	> 200	> 200	> 200	> 200	-
13	Pinoresinol-4'-O-*β*-D-glucopyranoside	> 200	> 200	> 200	> 200	-
14	Syringaresinol-4'-O-*β*-D-glucopyranoside	> 200	> 200	> 200	> 200	-
15	Camellikaempferoside B	> 200	> 200	> 200	> 200	-
16	Kaurenoic acid	-	-	>128	67,03±1,62	65,19±1,69
23	28-nor-urs-12-ene-3*β*,17 *β*-diol	-	-	>128	72,41±1,03	79,44±0,79
24	Micromeric acid	-	-	>128	>128	>128
25	23-hydroxyursolic acid	-	-	>128	3,28±0,17	1,16±0,03
26	Euscaphic acid	-	-	>128	>128	>128
29	Oleanolic acid	-	-	>128	>128	98,28±3,23
30	*3S, 5R, 6S, 9R*)-megastigmane-3,9-diol	-	-	>128	>128	>128
ĐC (+)	Ellipticine	0,34 ± 0,05	0,42 ± 0,05	0,32± 0,03	0,47± 0,06	0,43±0,02
ĐC (-)	DMSO	No cytotoxicity exhibited in any cell line				

Note: (-): Not tested; DC (+): Positive control (Ellipticine was tested at concentrations of 10, 2, 0.4, and 0.08 µg/mL); DC (-): Negative control (DMSO was tested at a concentration of 1%).

3.3.2.3. α-Glucosidase inhibition activity of the isolated compounds

The results in table 3.20 show that 7 out of 18 compounds have α-glucosidase inhibition activity at varying levels, while 11 compounds do not exhibit this activity (including two new compounds). Specifically, the compounds **(4), (5), (7), (16), (23), (25),** and **(29)** inhibit α-glucosidase with IC50 values ranging from 1.00 to 98.91 µg/mL. These compounds are more effective in inhibition compared to the control (IC50 = 147 µg/mL), with IC50 values for compounds **(4), (5), (7), (16), (23), (25)**, and **(29)** being 27.52, 13.67, 18.38, 13.08, 98.91, 1.00, and 4.16 µg/mL, respectively. Among them, 23-hydroxyursolic acid **(25)** exhibits the strongest activity with an IC50 of 1.00 µg/mL, approximately 147 times more potent than the control. The remaining compounds did not show α-glucosidase inhibition activity (IC50 > 128 µg/mL), including two new compounds.

Table 3.20. α-Glucosidase inhibition activity of the isolated compounds

KH compound	Compound name	IC_{50} (µg/mL)	KH compound	Compound name	IC_{50} (µg/mL)
1	Debutyldorycnic acid	> 128	**15**	Camellikaempferoside B	> 128
2	Adinanquercetiside	> 128	**16**	Kaurenoic acid	13,08 ± 0,94
4	Ursolic acid	27,52 ± 0,51	**23**	28-nor-urs-12-ene-3*β*,17 *β*-diol	98,91 ± 5,78
5	4,5-dihydroblumenol	13,67 ± 0,42	**24**	Micromeric acid	>128
7	24-hydroxytormentic	18,38 ± 0,41	**25**	23-hydroxyursolic acid	1,00 ± 0,07
9	Convoldorin	> 128	**26**	Euscaphic acid	>128
10	Scopoline	> 128	**29**	Oleanolic acid	4,16 ± 0,27
12	Horidin	> 128			
13	Pinoresinol-4'-O-*β*-D-glucopyranoside	> 128	**30**	*3S, 5R, 6S, 9R*)-megastigmane-3,9-diol	>128
14	Syringaresinol-4'-O-*β*-D-glucopyranoside	> 128	**ĐC**	Acarbose	147,86 ± 4,69

CONCLUSIONS AND RECOMMENDATIONS

1. Conclusions

1.1. The chloroplast genome of *A. bockiana* (accession number MW699853.1) has a size of 156284 bp and features the typical four-region structure: LSC (85693 bp), SSC (18411 bp), and a pair of IRa and IRb regions (26090 bp per region), with a GC content of 37.4%. It contains 129 genes, 70 repeat sequences, and 51 SSRs. Unlike other species in the genus *Adinandra*, the gene *ycf1* is not present at the boundary between IRa and SSC in the chloroplast genome of *A. bockiana*. There are variations in the number of chloroplast genes among species in the genus *Adinandra*: *A. bockiana* (129 genes), *A. megaphylla* (131 genes), and *A. millettii* and *A. angustifolia* (132 genes).

1.2. Sequence variations are primarily found in non-coding regions. Within coding regions, the genes *matK*, *psaA*, *ndhK*, *ndhG*, and *rbcL* show nucleotide differences among the four species: *A. bockiana*, *A. megaphylla*, *A. millettii*, and *A. angustifolia*. The LSC and SSC regions contain more variations than the IR region, with the LSC region showing the most variation. The genes *matK* and *rbcL*

are proposed as potential DNA barcode candidates to aid in the identification of species within the genus *Adinandra*.

1.3. Fifteen compounds were isolated from *A. megaphylla*, eight from *A. bockiana*, and fourteen from *A. glischroloma*. Two new compounds were identified: debutyldorycnic acid and adinanquercetiside. Most of the isolated compounds from the three species belong to the triterpenoid group. The compound 23-hydroxyursolic acid isolated from *A. glischroloma* shows the strongest cytotoxic activity against liver cancer cells (HepG2) and breast cancer cells (MCF-7) with IC50 values of 3.28 μg/mL and 1.16 μg/mL, respectively. 23-Hydroxyursolic acid also exhibits the strongest α-glucosidase inhibition activity (IC50 = 1.00 μg/mL). Ursolic acid shows the strongest antibacterial activity against *P. aeruginosa*, while isoquercetin shows activity against *C. freundii* and *S. milleri*.

2. Recommendations

2.1. Continue analyzing the chloroplast genome sequences of species within the genus *Adinandra* to provide additional data and identify potential gene regions for DNA barcoding to aid species identification.

2.2. Further evaluate the antioxidant, anti-inflammatory, and antiviral activities of the two new compounds (debutyldorycnic acid and adinanquercetiside) to determine their practical application potential.

PUBLICATIONS

1. **Pho Thi Thuy Hang**, Nguyen Thi Thu Nga, Sy Danh Thuong, Le Nguyen Thanh, Nguyen Van Phuong, Chu Hoang Mau, Nguyen Huu Quan (2024), "Chemical constituents of *Adinandra glischroloma* Hand.-Mazz. and their chemotaxonomic significance", *Biochemical Systematics and Ecology*,(SCIE, Q3, IF: 1,4), 113, 104803. https://doi.org/10.1016/j.bse.2024.104803
2. Quan N. H., **Hang P. T. T.**, Nga N. T. T., Duc H. V., Hue V. T., Thuong S. D., Thanh L. N., Mau C. H. (2023), "Chemical constituents and biological activities of the leaves of *Adinandra megaphylla*", *Phytochem. Lett*,(SCIE, Q2, IF: 1,873), 56, pp. 19-23. https://doi.org/10.1016/j.phytol.2023.06.002
3. Nga Thi Thu Nguyen, **Hang Thi Thuy Pho**, Quan Huu Nguyen, Nhung Thi Doan, Lan Thi Ngoc Nguyen, Huong Mai Pham, Lam Tung Le, Thuong Danh Sy, Ha Hoang Chu, Lien Thi Kim Vu, Mau Hoang Chu (2023), "Characteristics of the chloroplast genome of *Adinandra bockiana* and comparative analysis with species of *Pentaphylacaceae* family", *Plant Molecular Biology Reporter*, (SCIE, Q2, IF: 1,816), 41, pp. 611-621. https://doi.org/10.1007/s11105-023-01389-3.
4. **Pho Thi Thuy Hang**, Tran Dai Duong, Nguyen Thi Thu Nga, Tu Quang Tan, Nguyen Huu Quan, Chu Hoang Mau (2022), "Chemical composition and biological activity of the extract from the stem of *"*Sum lông*"*", *TNU Journal of Science and Technology*, 227(05), pp. 186-194. DOI: https://doi.org/10.34238/tnu-jst.5561