

Comparative Analysis of Whole Chloroplast Genomes of Three Common Species of Nekemias from Vine Tea

Feng Yang

Guizhou Normal University

Shaoxiong Wu

Xi'an International University

Zhengwen Yu (✉ yuzhengwen2001@126.com)

Guizhou Normal University

Research Article

Keywords: Nekemias grossedentata, complete chloroplast genome, phylogenetic tree, chromatographic fingerprint

Posted Date: February 1st, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-3909748/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Nekemias grossedentata is a plant species that has been utilized for centuries as both a traditional medicine and food source due to its high levels of bioactive compounds. Specifically, the plant's young stems and leaves are used to make vine tea, which has been used to treat various ailments. While *Nekemias cantoniensis* and *Nekemias megalophylla* are also utilized for similar purposes in various regions, limited research on their chloroplast genomes and chromatography has significantly hindered the study of their evolutionary and phylogenetic relationships, as well as the market supply of these materials. In order to address these gaps in knowledge, this study used the Illumina sequencing platform to sequence the complete chloroplast (cp) genomes of three *Nekemias* species, which were then compared to that of *Nekemias grossedentata*. Additionally, high-performance liquid chromatography (HPLC) was utilized to construct chromatographic fingerprints of the three plants. The annotation results demonstrated that the three chloroplast genomes were typical quadripartite structures, with lengths of 162,147 bp (*N. grossedentata*), 161,981 bp (*N. megalophylla*), and 162,500 bp (*N. cantoniensis*), respectively. A total of 89 (*N. grossedentata*)/86 (*N. megalophylla*, *N. cantoniensis*) protein-coding genes, 37 tRNA genes, and 8 rRNA genes were annotated. The IR/SC boundary regions were relatively conserved across the three species, although three regions (*rps19-rpl2*, *rpl32-trnL-UAG*, *ccsA-ndhD*) exhibited nucleotide diversity values (Pi) of variable sites higher than 1%. Phylogenetic analysis indicated that *N. grossedentata* had a closer genetic relationship with *N. megalophylla* than *N. cantoniensis*. Moreover, the chromatographic fingerprints revealed that the functional components were highly similar to the genetic relatedness of the species. Overall, this study provides valuable information for species identification, phylogeny, quality assessment, and rational use of three medicinal plants of the genus *Nekemias*. Furthermore, the findings of this study will contribute to the standardization of vine tea raw materials and the optimization of product efficacy.

1. Introduction

Nekemias grossedentata is a common species in the *Vitaceae* family, and is a traditional medicinal and edible plant rich in bioactive compounds, particularly dihydromyricetin, with a content of more than 35%¹. The *Ampelopsis* A. Rich ex Mich genus contains around 30 plant species worldwide, distributed in Asia, North America, and Central America, with 17 of these species currently identified in China². Some species have gained attention due to their medicinal value, particularly as ingredients in vine tea³, which has been used for centuries to treat cold fever, sore throat, jaundice hepatitis, among others^{4–8}. However, other species within the same genus have similar morphology and functions, and contain many of the same functional compounds such as dihydromyricetin, myricetin, and others⁹. The most commonly related species include *Nekemias megalophylla* and *Nekemias cantoniensis*¹⁰. Therefore, the close relationship between *Nekemias grossedentata* and its related species in terms of morphology, chemical composition, potency, and therapeutic properties is controversial, particularly regarding their quality standards and interchangeability. Previous studies have primarily focused on morphology, chemical composition, and pharmacological activity. However, relying solely on morphological characters for species identification has led to inaccurate results in *Ampelopsis* and *Nakemias*¹¹. Therefore, a new and accurate method to study the relationship between *Nekemias* species and their differences in ingredients will undoubtedly resolve taxonomic issues with vine tea and provide a reliable basis for raw material identification and quality evaluation.

Chloroplasts (Cp) possess an independent, circular genome that plays a vital role in the photosynthesis, development, and physiology of green plants¹². Generally, the Cp genome has a quadripartite structure (120–160kb) and contains 110–130 unique genes. In most angiosperms, the typical quadripartite cyclic structure comprises a large single-copy (LSC) region and a small single-copy (SSC) region separated by a pair of inverted repeats (IRA and IRB)¹³. Due to their highly conserved structure, general recombination-free, uniparental inheritance, and small effective population sizes¹⁴, chloroplast (cp) DNA sequences have been extensively utilized to resolve plant phylogenies¹⁵. The advent of next-generation sequencing has made the process cheaper and faster than before¹⁶. Consequently, phylogenetic analyses based on cp genome data are increasingly prevalent, even among small taxonomic populations, due to advances in next-generation sequencing technologies^{17,18}. Despite several species of *Nekemias* cp genomes being sequenced, the chloroplast genomes of the three plants used as raw materials for vine tea have not been fully developed.

In previous studies, there was a lack of comprehensive phylogenetic comparison and examination of chemical composition differences among the three species. In the present study, we employed a combination of long PacBio reads and short Illumina reads to de novo assemble the complete chloroplast (cp) genomes of the three *Nekemias* plants, and fully elucidated the features of these cp genomes. Subsequently, we analyzed the phylogenetic relationships of 29 *Nekemias* species from GenBank. Finally, the chemical composition differences among the three target plants were compared using HPLC fingerprinting. Our research provides valuable information for species identification, phylogeny, quality assessment, and rational use of the three medicinal plants within the genus *Nekemias*, which will ultimately facilitate the standardization of raw material supply and optimization of vine tea products.

2. Results

Cp genome organizations

The chloroplast genome information of the three *Nekemias* species compared in this section has been uploaded to the NCBI database, and the GenBank accession numbers are shown in Table 1. As in other flowering plants, the chloroplast genome of *Nekemias* is a typical quadripartite structure which contains two inverted repeats (IRs), a large copy region (LSC) and a small copy region (SSC). As shown in Fig. 1 and Table 1, the chloroplast genome length of the three species was between 161,981 bp and 162,500 bp, while the whole genome length of the chloroplast of the *Nekemias grossedentata* had an intermediate length 162,147 bp, and the maximum length difference among the three species was only 519 bp.

By comparing the GC content of the three species in Tables 1 and 2, we found that the GC ratio of the IR region is greater than that of the other two independent regions, which is the same as that of most land plants. This is because the eight ribosomal RNAs (rRNAs) of the chloroplast genome are located in the IR region of the two inverted repeats, and ribosomal RNA is rich in GC bases, which causes this result¹⁹. In addition, the AT content of the three targeted plant protein-coding regions (CDS regions) was greater than their GC content, which was also consistent with other land plants^{20,21}. We mapped and annotated the chloroplast genomes of these three species. Combined with Table 2, we can find that although the chloroplast genome of *N. grossedentata* has an intermediate length, it has 134 whole genes and 89 protein coding genes, 3 more protein coding genes and a unique gene—*ycf15*, compared to *N. megalophylla* and *N. cantoniensis*. This gene belongs to a protein-coding gene with unknown function, and no specific functional classification has been found in similar studies. It may be used as a molecular marker to identify *Nekemias* species in the supply of vine tea products. In addition, the three *Nekemias* species showed a high degree of similarity in 37 tRNAs and 8 rRNAs, but the length of the CDS region was different. The CDS region of *Nekemias grossedentata* was more than 1500 bp longer than the other two species, but the the total GC content of the CDS region was not significantly different, both were about 37.36%. The difference of nucleotide composition at each site was also small. The GC content of the first, second, and third codon positions in *N. grossedentata* was 45.57%, 38.19% and 30.19%, respectively. The GC content of the first, second, and third codon positions in *N. cantoniensis* was 45.73%, 38.28% and 30.02%, respectively. The GC content of the first, second, and third codon positions in *N. megalophylla* was 45.71%, 38.28% and 30.04%, respectively. Although the difference between species is small, we can still find that the GC content of the third codon is low. According to the literatures^{22,23}, it can be found that the rule of AT base preference of the third codon is also reflected in other angiosperms, which indicates that our data mining and analysis are reliable.

There are 89 protein-coding genes in the chloroplast genome of *N. grossedentata*. It had a total of 27,007 codons, while the 86 protein-coding genes of *N. cantoniensis* and *N. megalophylla* had a total of 26,487 and 26,482 codons, respectively. Among all the codons in the protein coding region, the codon responsible for isoleucine (Ile), AUU (1125/1103/1103), was the most frequently used in all three species, while the codon responsible for cysteine (Cys), UGC (85/80/81), was the least used in all three species. As shown in Fig. 2, the relative synonymous codon usage frequencies (RSCU) in the protein-coding regions of the chloroplast genomes of the three species are slightly different but generally similar. All amino acids except methionine (Met) and threonine (Thr) had multiple paired codons, up to six of them, including arginine, leucine and serine. Most amino acids have two paired codons, and the single preference is obvious, but in multi-codon amino acids, there are several preferential codons at the same time (for example, three codons in arginine have RSCU > 1). However, we can find that the higher rate of synonymous substitutions provides more informative characters for phylogenetic studies.

(Fig. 1. Gene map of the *N. grossedentata* cp genomes. The genes inside and outside the outer circle are transcribed in the direction from 0K to 15K. Genes are classified into 14 groups according to their biological function and are shown by different colored boxes. And gene maps of the other two species are submitted in the supplementary material.)

Table 1
Summary of the base composition of the cp genomes of three *Nekemias* species

Species	Accession number	Total length (bp)	LSC region(bp)	IR region(bp)	SSC region(bp)	GC cotent(%)	GC content in LSC region(%)	GC content in IR region(%)	GC content in SSC region(%)
N. grossedentata	MT267294.1	162,147	89,244	27,232	18,439	37.36	35.35	42.52	31.82
N. cantoniensis	ON616406.1	162,500	89,333	27,042	19,083	37.36	35.36	42.66	31.70
N. megalophylla	NC_068499.1	161,981	89,236	27,153	18,439	37.37	35.36	42.58	31.81

Table 2
Gene number and CDS nucleotide composition of the cp genomes of the three *Nekemias* species

Species	Number of		CDS (bp)				ATCG contents in CDS (%)						
	Total gene	Unique gene	P-cgs	tRNAs	rRNAs		T(U)	C		A		G	
<i>N. grossedentata</i>	134	114	89	37	8	81,021	31.30	17.72		30.72		20.26	
<i>N. cantoniensis</i>	131	113	86	37	8	79,461	31.29	17.72		30.70		20.29	
<i>N. megalophylla</i>	131	113	86	37	8	79,446	31.28	17.72		30.71		20.29	
Species	Base contents in 1st position (%)					Base contents in 2nd position (%)				Base contents in 3rd position (%)			
	T(U)	C	A	G	T	C	A	G	T	C	A	G	
<i>N. grossedentata</i>	23.66	18.83	30.77	26.74	32.34	20.34	29.47	17.85	37.90	13.98	31.91	16.21	
<i>N. cantoniensis</i>	23.57	18.85	30.70	26.88	32.31	20.38	29.41	17.90	37.98	13.94	32.00	16.09	
<i>N. megalophylla</i>	23.58	18.83	30.71	26.87	32.30	20.38	29.42	17.90	37.95	13.93	32.00	16.11	

Repeat and simple sequence repeat (SSR) analyses

In this paper, the analysis of the repeat sequences of three *Nekemias* species was based on the online software REPuter 1.0 and Tandem v4.09, and the detected repeat sequences included forward repeat sequences, palindromic repeat sequences and tandem repeat sequences (the rest of the repeat sequences were not included in the experiment due to too few or no repeat sequences). The results are shown in Fig. 3, a, b, c and d. The characteristics of simple sequences repeat (SSR) were compared and analyzed through the online website MISA, and the results are shown in

Figure 4.

((a) Repeat types in three cp genomes; (b) tandem repeats in three cp genomes; (c) forward repeats in three cp genomes; (d) palindromic repeats in three cp genomes. The ordinate indicates the number of repeats, while the abscissa represents the name of the species. In (a), different colors mean different repetition types, in (b–d), different colors show different lengths. As marked in the upper right corner.)

In the analysis of repeat sequences, it can be found from Figure.3a that the number of forward repeat and palindromic repeat sequences of the three *Nekemias* species is 49, but the number of tandem repeat sequences of *Nekemias grossedentata* (72) is more than that of the other two species (69/70). Many of forward and palindromic repeats were found in the IR region of the tetrad structure. There were 6 forward and palindromic repeats from the intronic region which starting from the IR region to the *rp/23* spacer. However, most of the tandem repeats were located in the LSC region, and the longest segment of all the tandem repeats was 79 bp, which was located in the intronic region between *rps19* and *rp/2* genes in the IR region of all three species. The analysis of the chloroplast genomes of three *Nekemias* species showed that the types and numbers of repeat sequences of the three species were similar, and the length of each repeat sequence was slightly different. The length of tandem repeat sequences of each species was generally concentrated in the range of 10–30 bp, while the length of forward and palindromic sequences was not significantly different and relatively scattered. However, *N. cantoniensis* had no repeat sequences larger than 65 bp.

Simple sequence repeats (SSRs) have a high frequency of variation within the same species, so they are often used as a genetic map construction, correction and mapping study²⁴. According to the parameter setting, 217 simple repeats were observed in the chloroplast genome of *N. grossedentata*, including 140 mono-nucleotide repeats, 58 di-nucleotide repeats, 6 tri-nucleotide repeats and 10 tetra-nucleotide repeats, 2 penta- and 1 hexa-, respectively. The number and type of simple repeat sequences of *N. megalophylla* were very similar to that of *N. grossedentata*, with only one additional mono-typic sequence, but *N. cantoniensis* showed a slight quantitative difference. SSR analysis of the chloroplast genomes of the three species showed that the genomes of the three species were very similar and did not reflect obvious variation level. Single nucleotide repeats (SNPS) were the main simple repeat type in all three species, and A/T type SNPS were the dominant type. The number of oligonucleotide repeat of three or more was small but the types were very rich.

IR expansion and contraction

Expansion and contraction of IR region boundaries is the main reason for the size changes of the cp genome and play an important role in species evolution²⁵. We compare the IR/SSC and IR/LSC junction areas of three species to determine whether they have expansion or contraction. It can be seen from

Figure 5 that the *rpl22* gene of the *N. grossedentata* is always located in the LSC region at the JLB boundary, while the *rpl22* gene of the other two species are extended to the IRb region. The *ycf1* gene was identical in all three species. This gene is divided into two parts and spans three regions (IRa-SSC-IRb). As the gene with the longest sequence composition, the function of *ycf1* is still unknown²⁶, but some studies have shown that this gene family is indispensable in plants. And we might be able to leave a signature on this gene for identification, for example, through molecular markers. Besides, there was no significant differences between the chloroplast genomes of *N. grossedentata* and *N. megalophylla*, while there was a certain shift in the gene loci of *N. cantoniensis*, indicating evolutionary expansion at their boundaries.

Sequence divergence analysis

In Fig. 6, the three complete cp genomes were compared with mVISTA using *N. grossedentata* as a reference. We can first find that the similarity in the non-coding sequence(CNS) is relatively low, such as the region between *trnR* and *atpA*, *rps19* and *rpl2*, *psbM* and *trnD*, which indicates that the substitution rate of non-coding regions is faster than that of coding regions²⁷. It also indicated that the coding region of closely related species is highly conserved and the rate of variation is extremely low. It is worth noting that the alignment results of the three sequences show that the similarity of them is generally greater than 90%, but the coding region of *ycf1* gene has a slight degree of divergence. Although the function of this gene is unknown, gene knockout studies have proved that this gene is an essential gene for plant cell survival²⁸. In order to clarify the variation in the higher regions, the nucleotide diversity values (PI) were calculated using the DnaSP v.6.10 software (Fig. 8.). The variation in intergenic regions ranged from 0 to 9.59%, with an average of 0.39%, which was sixfold higher than that in the CDS regions (0.06% on average). Three divergent loci in intergenic regions (*rps19-rpl2* *rpl32-trnL-UAG* *ccsA-ndhD*) had a Pi exceeding 1%, while no gene in the CDS region had a PI value greater than 1%. These three divergence hotspot regions should be applied to the development of molecular markers for phylogenetic and phylogeographic analyses, as well as identification of *Nekemias* species. Finally, according to relevant studies²⁹, the phylogenetic relationship of closely related species is often judged by the molecular evolution rate of their non-coding regions, so the results can really provide a theoretical basis for phylogenetic analysis and species identification.

Phylogenetic analysis

A phylogenetic tree was built based on 29 species of *Vitaceae* cp genomes (18 *Vitis* species, 2 species from *Tetrastigma*, 7 species from *Nekemias*), using the maximum likelihood (ML) method, with two outgroups. Most branches have a bootstrap value of 100%, and in the ML tree, all three mentioned species (marked in red) form a well-supported monophyletic group. The phylogenetic tree generated a total of 27 branches with branch support greater than 94%. Among all evolutionary nodes, a total of 24 branches have node values greater than 99%. *Nekemias. grossedentata* and *N. megalophylla* are sister groups are clustered together in one terminal branch, thereby representing a similar genetic relationship. These results indicated that the three target plants were not only closely related to each other, but also had a certain genetic distance from the other four species of the same genus. It is worth to know that another plant of the same genus, *N. chaffanjonii*, was also found in the experimental preparation of this study. Unfortunately, there is little genetic information and related research content of this species. This plant deserves more attention in future studies.

HPLC analysis of the biosynthesis related compounds

In terms of chemical constituents, HPLC fingerprints of the three species were obtained at 292nm wavelength by comparing and analyzing common functional constituents and superimposed on the control. Through the analysis and comparison of Fig. 9, It can be seen that the active components of the three species are almost the same, but the content is different. All the three species had obvious peaks at the baseline of dihydromyricetin standard, and the content of *N. grossedentata* and *N. megalophylla* was close to each other and much higher than another species. Both taxifolin and myricetin were present in the three species, but the content of taxifolin was generally low, while the content of myricetin was much higher in *N. cantoniensis* than the other two species. The overall results showed that the compound composition of *N. grossedentata* and *N. megalophylla* were very similar, and dihydromyricetin was their main functional compound. The compound composition of *N. cantoniensis* were not similar to others, myricitrin and myricetin were its main functional compound. This result was also consistent with the results of other contents in this experiment, and complementary to verify the close correlation between plant functional components and genetic evolution, which not only improved the information of biochemical components of species, but also further improved the identification content of *Nekemias* species.

3. Materials and Methods

Plant material and DNA extraction

The materials and genetic information of the three species have been uploaded to the public database of NCBI. Therefore, we take *N. grossedentata* as an example (Data uploaded by our laboratory)³⁰. The samples, young and healthy leaf, were collected from Dashuijing Village, Devo Town, Anlong Ciuentry, Guizhou Province, China (25°22'53.06"N, 105°38'17.70"E, 1141m above sea level). The total genomic DNA was extracted using DNAsecure Plant Kit (TIANGEN, Beijing, China), while the total amount of 700 ng DNA per sample was used as input material for the DNA sample preparations. Sequencing libraries were generated using NEB NextVRUltra DNA Library Prep Kit for Illumina (NEB, Ipswich, MA). Total DNA was used to generate libraries with an average insert size of 350 bp. The library preparations were sequenced on an Illumina platform and 150 bp pairedend reads were generated.

Assembly and annotation

We take *N. grossedentata* as a reference example. The filtered reads were assembled using the program GetOrganelle³¹ with *N. grossedentata* (GenBank accession number: MT267294.1) as the initial reference genome. The assembled cp genome was annotated using the online software GeSeq³². The circular maps of the five species were obtained using Chloroplot³³. Codon usage and cp genomic characteristics were analyzed using MEGA7.0³⁴.

Characterization of repeat sequences and SSRs

Repeat sequences containing the tandem, forward and palindromic. They were identified using REPuter³⁵ with a Hamming distance set at 3 bp and a minimum repeat size set at 30 bp. Tandem repeats were analyzed using the Tandem Repeats Finder with default parameters³⁶. SSRs were detected using MISA³⁷ with the following thresholds: eight repeat units for mono SSRs, four repeat units for di- and trinucleotide repeat SSRs, and three repeat units for tetra-, penta-, and hexanucleotide repeat SSRs.

Genome comparison

The mVISTA in the Shuffle-LAGAN mode³⁸ was used to compare the cp genomes of *N. grossedentata* with the cp genomes of the other two reported species cp genomes (The accession numbers of the 3 species are shown in Table 1) by using the annotation of *N. grossedentata* as the reference. The nucleotide variability was calculated with DnaSP 6.10³⁹. The online software IRSCOPE⁴⁰ (<https://irscope.shinyapps.io/irapp/>) was used to carry out boundary analysis on the IR regions of the three species.

Phylogenetic analysis

A total of 29 complete cp genome sequences were downloaded from the NCBI GenBank for the phylogenetic analysis. The cp genomes were aligned by using MAFFT v7.520 software^{41,42}, while MEGA 7.0 was used to construct the phylogenetic tree with the maximum-likelihood (ML) method. The ML tree based on TN + gamma + I model was performed using RAxML (Version 8.0.19) with 1,000 bootstrap replicates⁴³, while two different species were set as the outgroups.

HPLC analysis of the biosynthesis related compounds

An ultrasonic method to extract total compounds from leaves was used. Approximately 0.01 g of oven-dried leaf tissue was weighed per sample and placed into a centrifuge tube, to which 10 mL of methanol extract was added. After shaking evenly, filter to obtain the test solution. The contents of compounds were measured by high performance liquid chromatography (HPLC), which used methanol as solvent A (0.1% phosphoric acid as solvent B) 0 min, 26:74, 10 min, 50:50, 15 min, 55:45, 25–30 min, 26:74. The column temperature was 40°C, the sample volume injected was 5.0 µL, and the flow rate was 1 mL/min, with the detector set to 292 nm.

4. Discussion

In this study, we utilized high-throughput sequencing technology to obtain the complete chloroplast genomes of three *Nekemias* species used as raw materials for vine tea. We conducted comprehensive comparison and analysis of the obtained data through data mining and phylogenetic analysis. Firstly, we demonstrated the close relationship between the three *Nekemias* species by analyzing their chloroplast genome characteristics, including sequence length, gene number, gene map, base composition, coding region characteristics, and codon usage bias. This type of comparison is a rare finding in previous research. Secondly, using chloroplast genome repeat sequences and SSRs characteristics, we further verified the genetic relationship between the three *Nekemias* species used as raw materials for vine tea, and drew

a preliminary conclusion that the genetic relationship between *N. grossedentata* and *N. megalophylla* was closer than that between *N. cantoniensis* and *N. grossedentata*. SSRs are commonly used as genetic markers in community genetics and evolutionary studies ⁴⁴, and mononucleotide repeats were the most enriched, which is consistent with our findings. We used the complete chloroplast genome sequence characteristics, including the IR boundary characteristics and the complete genome sequence homology comparison of mVISTA, to verify the conclusion that *N. megalophylla* and *N. cantoniensis* are both related species of *N. grossedentata*, and *N. megalophylla* is more closely related to *N. grossedentata* with more similar composition. We calculated the nucleotide diversity values (Pi) to clarify the variation in the higher regions and identified three hotspot regions of divergence, which have important implications for the development of molecular markers. Additionally, we constructed a plausible phylogenetic tree of the Vitaceae using the chloroplast genomes of 29 species, which confirmed the phylogenetic relationships of the target plants. Finally, we compared the chemical composition of the three species and concluded, after qualitative analysis, chromatographic fingerprint analysis is an effective supplementary method to identify three *Nekemias* species.

Some studies have suggested that the taxonomic identification of *N. grossedentata* was hasty, and some varieties have been discovered ⁴⁵. Chromatographic fingerprint analysis is a rational method for identifying and assessing the quality of traditional Chinese herbal medicines ⁴⁶. It can effectively authenticate, evaluate, and distinguish the merits and demerits of these medicines, and increase the accuracy of species identification. Therefore, we performed qualitative research by conducting chromatographic analysis on samples with the same concentration. We found that the content of dihydromyricetin, the main functional component, was not significantly different between the two species, *N. grossedentata* and *N. megalophylla*. Dihydromyricetin has various medicinal properties ^{47–49}, such as anti-inflammatory, antibacterial, and antioxidant effects. As *N. megalophylla* has higher dihydromyricetin content, it may be a better genetic line for superior breeding and a better source species for vine tea. These findings can not only contribute to the current identification of *Nekemias* species, but also to future studies on the mechanisms of dihydromyricetin synthesis in these species.

5. Conclusion

The comparison of the whole cp genomes of three *Nekemias* species used as raw materials for vine tea is reported for the first time. Through the data obtained from the NCBI public database, we confirmed that the gene sequence and genome structure of the *Nekemias grossedentata* are similar to the other two species, but there are still traces of independent evolution and distinct regions of divergence. We identified 217 SSR loci that could potentially be used as molecular markers to study the diversity of the genus *Nekemias*. Three different mutation hot spots (*rps19-rpl2 rpl32-trnL-UAG ccsA-ndhD*) with high nucleotide variability (Pi) could be considered as potential DNA barcodes of *Nekemias grossedentata*. In our phylogenetic tree, eight plants of genus *Nekemias* converge into two strong clades, and three species are closely related to *N. grossedentata*. The HPLC fingerprints of four *Nekemias* species increased the accuracy of species identification. It is noteworthy that the content of dihydromyricetin is not lower in *N. megalophylla* than in *N. grossedentata*, which can be considered as a raw material for this compound. Therefore, it is important to explore the molecular mechanism of high dihydromyricetin in some *Nekemias* species. These results not only revealed the structural characteristics and sequence diversity of *Nekemias* plants, but also strengthened the relationship between species and confirmed the kinship of *Nekemias* family. We believe that by analysis of cp genome and phylogenetic relationships, it will be helpful to understand the evolution of *Nekemias* and solve the problem that the existing closely related species are difficult to distinguish in taxonomy. At the same time, in terms of chemical composition content analysis, chromatographic fingerprint analysis is an effective supplementary method to identify a variety of TCMs. In future studies, more detailed references can be provided for the quality evaluation and rational use of plant raw materials by further improving the relevant data such as the content of Q markers.

Declarations

Author contributions

Zhengwen Yu and Shaoxiong Wu conceived and designed the research framework; Zhengwen Yu, Feng Yang and Chunyan Jiang collected and identified all the samples; Shaoxiong Wu, Chenju Yang and Xiayu Feng performed the experiments; Shaoxiong Wu and Feng Yang analyzed the data and wrote the paper. All the authors have read and approved the first manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Acknowledgments

This study was supported by the National Natural Science Foundation of China under Grant [Nos. U1812401 and 32060068].

Data Archiving statement

The mentioned chloroplast genome data that support the findings of this study are openly available in GenBank of NCBI at <https://www.ncbi.nlm.nih.gov>.

References

1. Hu, H., Luo, F., Wang, M., Fu, Z., and Shu, X. (2020). New Method for Extracting and Purifying Dihydromyricetin from *Nekemias grossedentata*. *ACS Omega* 5, 13955-13962. 10.1021/acsomega.0c01222.
2. Atlas of Chinese higher plants. (2002). *Journal of Botany*, 27.
3. Zhang, Q., Zhao, Y., Zhang, M., Zhang, Y., Ji, H., and Shen, L. (2021). Recent advances in research on vine tea, a potential and functional herbal tea with dihydromyricetin and myricetin as major bioactive compounds. *J Pharm Anal* 11, 555-563. 10.1016/j.jpha.2020.10.002.
4. Hou, X.L., Tong, Q., Wang, W.Q., Shi, C.Y., Xiong, W., Chen, J., Liu, X., and Fang, J.G. (2015). Suppression of Inflammatory Responses by Dihydromyricetin, a Flavonoid from *Nekemias grossedentata*, via Inhibiting the Activation of NF-kappaB and MAPK Signaling Pathways. *J Nat Prod* 78, 1689-1696. 10.1021/acs.jnatprod.5b00275.
5. Liang, H., He, K., Li, T., Cui, S., Tang, M., Kang, S., Ma, W., and Song, L. (2020). Mechanism and antibacterial activity of vine tea extract and dihydromyricetin against *Staphylococcus aureus*. *Sci Rep* 10, 21416. 10.1038/s41598-020-78379-y.
6. Wu, J., Miyasaka, K., Yamada, W., Takeda, S., Shimizu, N., and Shimoda, H. (2022). The Anti-Adiposity Mechanisms of Ampelopsin and Vine Tea Extract in High Fat Diet and Alcohol-Induced Fatty Liver Mouse Models. *Molecules* 27. 10.3390/molecules27030607.
7. Zhang, X., Xu, Y., Xue, H., Jiang, G.C., and Liu, X.J. (2019). Antioxidant activity of vine tea (*Nekemias grossedentata*) extract on lipid and protein oxidation in cooked mixed pork patties during refrigerated storage. *Food Sci Nutr* 7, 1735-1745. 10.1002/fsn3.1013.
8. Chen, S., Zhao, X., Wan, J., Ran, L., Qin, Y., Wang, X., Gao, Y., Shu, F., Zhang, Y., Liu, P., et al. (2015). Dihydromyricetin improves glucose and lipid metabolism and exerts anti-inflammatory effects in nonalcoholic fatty liver disease: A randomized controlled trial. *Pharmacol Res* 99, 74-81. 10.1016/j.phrs.2015.05.009.
9. Rongxin, C., Xican, L., Chunhou, L., Jiayi, Z., Jingyuan, Z., Jianwu, L., Boxu, T., Zheng, L., Shuqin, L., and Yan, Y. (2022). Standards-Based UPLC-Q-Exactive Orbitrap MS Systematically Identifies 36 Bioactive Compounds in *Nekemias grossedentata* (Vine Tea). *Separations* 9.
10. Xiong, L., Ling, L., Xu, W., and Huang, Z. (2022). Literature research and microscopic identification of hakka vine tea in Fujian. *Fujian Traditional Chinese Medicine* 53, 26-29. 10.13260/j.cnki.jfjtc.012514.
11. Shi, Y., Xiao, L., Nie, J., Hu, M., Wang, W., and Fang, J.G. (2020). Investigation on the resources of Chinese characteristic national medicinal herb vine tea. *Traditional Chinese medicine* 43, 1076-1080. 10.13863/j.issn1001-4454.2020.05.006.
12. Howe, C.J., Barbrook, A.C., Koumandou, V.L., Nisbet, R.E., Symington, H.A., and Wightman, T.F. (2003). Evolution of the chloroplast genome. *Philos Trans R Soc Lond B Biol Sci* 358, 99-106; discussion 106-107. 10.1098/rstb.2002.1176.
13. Shetty, S.M., Md Shah, M.U., Makale, K., Mohd-Yusuf, Y., Khalid, N., and Othman, R.Y. (2016). Complete Chloroplast Genome Sequence of Corroborates Structural Heterogeneity of Inverted Repeats in Wild Progenitors of Cultivated Bananas and Plantains. *Plant Genome* 9. 10.3835/plantgenome2015.09.0089.
14. Birky, C.W., Jr., Maruyama, T., and Fuerst, P. (1983). An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics* 103, 513-527. 10.1093/genetics/103.3.513.
15. Shaw, J., Shafer, H.L., Leonard, O.R., Kovach, M.J., Schorr, M., and Morris, A.B. (2014). Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. *Am J Bot* 101, 1987-2004. 10.3732/ajb.1400398.
16. Alkan, C., Sajjadian, S., and Eichler, E.E. (2011). Limitations of next-generation genome sequence assembly. *Nat Methods* 8, 61-65. 10.1038/nmeth.1527.
17. Ma, P.F., Zhang, Y.X., Zeng, C.X., Guo, Z.H., and Li, D.Z. (2014). Chloroplast phylogenomic analyses resolve deep-level relationships of an intractable bamboo tribe Arundinarieae (poaceae). *Syst Biol* 63, 933-950. 10.1093/sysbio/syu054.
18. Carbonell-Caballero, J., Alonso, R., Ibanez, V., Terol, J., Talon, M., and Dopazo, J. (2015). A Phylogenetic Analysis of 34 Chloroplast Genomes Elucidates the Relationships between Wild and Domestic Species within the Genus *Citrus*. *Mol Biol Evol* 32, 2015-2035. 10.1093/molbev/msv082.
19. He, Y., Xiao, H., Deng, C., Xiong, L., Yang, J., and Peng, C. (2016). The Complete Chloroplast Genome Sequences of the Medicinal Plant *Pogostemon cablin*. *Int J Mol Sci* 17, 820-820. 10.3390/ijms17060820.

20. XiWen, L., HuanHuan, G., YiTao, W., JingYuan, S., Robert, H., HeZhen, W., ZhiGang, H., Hui, Y., HongMei, L., Kun, L., et al. (2013). Complete chloroplast genome sequence of *Magnolia grandiflora* and comparative analysis with related species. *Science China(Life Sciences)* *56*, 189-200.
21. Vining, K.J., Johnson, S.R., Ahkami, A., Lange, I., Parrish, A.N., Trapp, S.C., Croteau, R.B., Straub, S.C.K., Pandelova, I., and Lange, B.M. (2017). Draft Genome Sequence of *Mentha longifolia* and Development of Resources for Mint Cultivar Improvement. *Mol Plant* *10*, 323-339. 10.1016/j.molp.2016.10.018.
22. Mariotti, R., Cultrera, N.G., Diez, C.M., Baldoni, L., and Rubini, A. (2010). Identification of new polymorphic regions and differentiation of cultivated olives (*Olea europaea* L.) through plastome sequence comparison. *BMC Plant Biol* *10*, 211. 10.1186/1471-2229-10-211.
23. Lu, C., Shen, Q., Yang, J., Wang, B., and Song, C. (2016). The complete chloroplast genome sequence of Safflower (*Carthamus tinctorius* L.). *Mitochondrial DNA A DNA Mapp Seq Anal* *27*, 3351-3353. 10.3109/19401736.2015.1018217.
24. Zhou, T., Wang, J., Jia, Y., Li, W., Xu, F., and Wang, X. (2018). Comparative Chloroplast Genome Analyses of Species in *Gentiana* section *Cruciata* (Gentianaceae) and the Development of Authentication Markers. *Int J Mol Sci* *19*, 1962-1962. 10.3390/ijms19071962.
25. Gui, L., Jiang, S., Xie, D., Yu, L., Huang, Y., Zhang, Z., and Liu, Y. (2020). Analysis of complete chloroplast genomes of *Curcuma* and the contribution to phylogeny and adaptive evolution. *Gene* *732*, 144355. 10.1016/j.gene.2020.144355.
26. De Las Rivas, J., Lozano, J.J., and Ortiz, A.R. (2002). Comparative analysis of chloroplast genomes: functional annotation, genome-based phylogeny, and deduced evolutionary patterns. *Genome Res* *12*, 567-583. 10.1101/gr.209402.
27. Widmer, A., and Baltisberger, M. (1999). Extensive intraspecific chloroplast DNA (cpDNA) variation in the alpine *Draba aizoides* L. (Brassicaceae): haplotype relationships and population structure. *Mol Ecol* *8*, 1405-1415. 10.1046/j.1365-294x.1999.00700.x.
28. Hu, Z.-Y., Hua, W., Huang, S.-M., and Wang, H.-Z. (2011). Complete chloroplast genome sequence of rapeseed (*Brassica napus* L.) and its evolutionary implications. *Genetic Resources and Crop Evolution* *58*, 875-887.
29. Qihui, Z., and Song, G. (2005). Phylogenetic relationships among A-genome species of the genus *Oryza* revealed by intron sequences of four nuclear genes. *The New phytologist* *167*, 249-265.
30. Yu, Z., Yi, Y., and Gu, L. (2021). The complete chloroplast genome of *Lonicera fulvotomentosa* Hsu et S. C. Cheng and its phylogenetic analysis. *Mitochondrial DNA B Resour* *6*, 842-843. 10.1080/23802359.2021.1884027.
31. Jin, J.J., Yu, W.B., Yang, J.B., Song, Y., dePamphilis, C.W., Yi, T.S., and Li, D.Z. (2020). GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol* *21*, 241-241. 10.1186/s13059-020-02154-5.
32. Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R., and Greiner, S. (2017). GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res* *45*, W6-W11. 10.1093/nar/gkx391.
33. Zheng, S., Poccai, P., Hyvonen, J., Tang, J., and Amiryousefi, A. (2020). Chloroplot: An Online Program for the Versatile Plotting of Organelle Genomes. *Front Genet* *11*, 576124. 10.3389/fgene.2020.576124.
34. Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* *33*, 1870-1874. 10.1093/molbev/msw054.
35. Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J., and Giegerich, R. (2001). REPuter: the manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res* *29*, 4633-4642. 10.1093/nar/29.22.4633.
36. Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* *27*, 573-580. 10.1093/nar/27.2.573.
37. Thiel, T., Michalek, W., Varshney, R.K., and Graner, A. (2003). Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* *106*, 411-422. 10.1007/s00122-002-1031-0.
38. Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res* *32*, W273-279. 10.1093/nar/gkh458.
39. Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E., and Sanchez-Gracia, A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Mol Biol Evol* *34*, 3299-3302. 10.1093/molbev/msx248.
40. Amiryousefi, A., Hyvonen, J., and Poccai, P. (2018). IRscope: an online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* *34*, 3030-3031. 10.1093/bioinformatics/bty220.
41. Katoh, K., Rozewicki, J., and Yamada, K.D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* *20*, 1160-1166. 10.1093/bib/bbx108.
42. Wong, K.M., Suchard, M.A., and Huelsenbeck, J.P. (2008). Alignment uncertainty and genomic analysis. *Science* *319*, 473-476. 10.1126/science.1151532.

43. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312-1313. 10.1093/bioinformatics/btu033.

44. Yang, Y., Zhou, T., Duan, D., Yang, J., Feng, L., and Zhao, G. (2016). Comparative Analysis of the Complete Chloroplast Genomes of Five *Quercus* Species. *Front Plant Sci* 7, 959. 10.3389/fpls.2016.00959.

45. Zhang, Y., S., Liu, J., and Song, S., Q. (2021). Ancient and understated vine tea. *Living world*, 4-13.

46. Xie, P., Chen, S., Liang, Y.Z., Wang, X., Tian, R., and Upton, R. (2006). Chromatographic fingerprint analysis—a rational approach for quality assessment of traditional Chinese herbal medicine. *J Chromatogr A* 1112, 171-180. 10.1016/j.chroma.2005.12.091.

47. Khan, S., Ullah, M.W., Siddique, R., Liu, Y., Ullah, I., Xue, M., Yang, G., and Hou, H. (2019). Catechins-Modified Selenium-Doped Hydroxyapatite Nanomaterials for Improved Osteosarcoma Therapy Through Generation of Reactive Oxygen Species. *Front Oncol* 9, 499. 10.3389/fonc.2019.00499.

48. Castaldo, L., Izzo, L., Gaspari, A., Lombardi, S., Rodriguez-Carrasco, Y., Narvaez, A., Grosso, M., and Ritieni, A. (2021). Chemical Composition of Green Pea (*Pisum sativum* L.) Pods Extracts and Their Potential Exploitation as Ingredients in Nutraceutical Formulations. *Antioxidants (Basel)* 11. 10.3390/antiox11010105.

49. Silvan, J.M., Guerrero-Hurtado, E., Gutierrez-Docio, A., Alarcon-Cavero, T., Prodanov, M., and Martinez-Rodriguez, A.J. (2021). Olive-Leaf Extracts Modulate Inflammation and Oxidative Stress Associated with Human *H. pylori* Infection. *Antioxidants (Basel)* 10. 10.3390/antiox10122030.

Supplementary Material

Supplementary Material is not available with this version

Figures

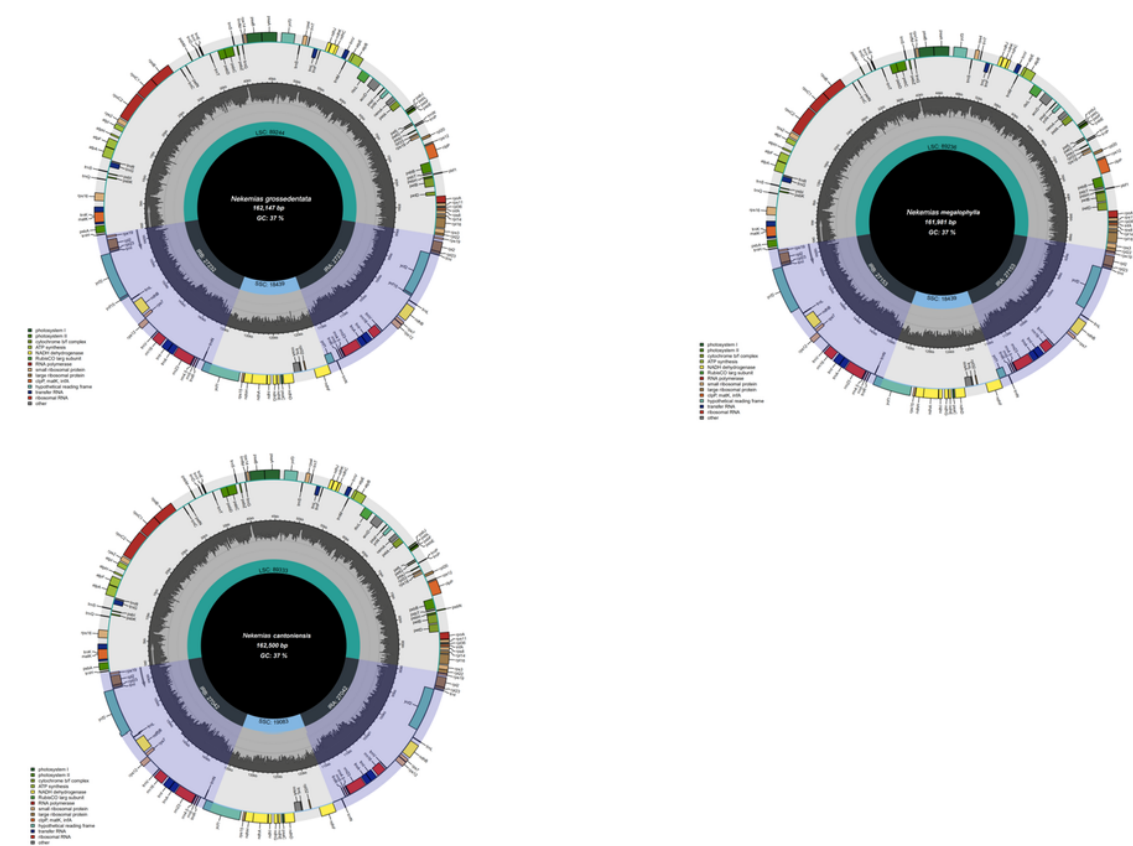


Figure 1

Gene map of the *N. grossedentata* cp genomes. The genes inside and outside the outer circle are transcribed in the direction from 0K to 15K. Genes are classified into 14 groups according to their biological function and are shown by different colored boxes. And gene maps of

the other two species are submitted in the supplementary material.)

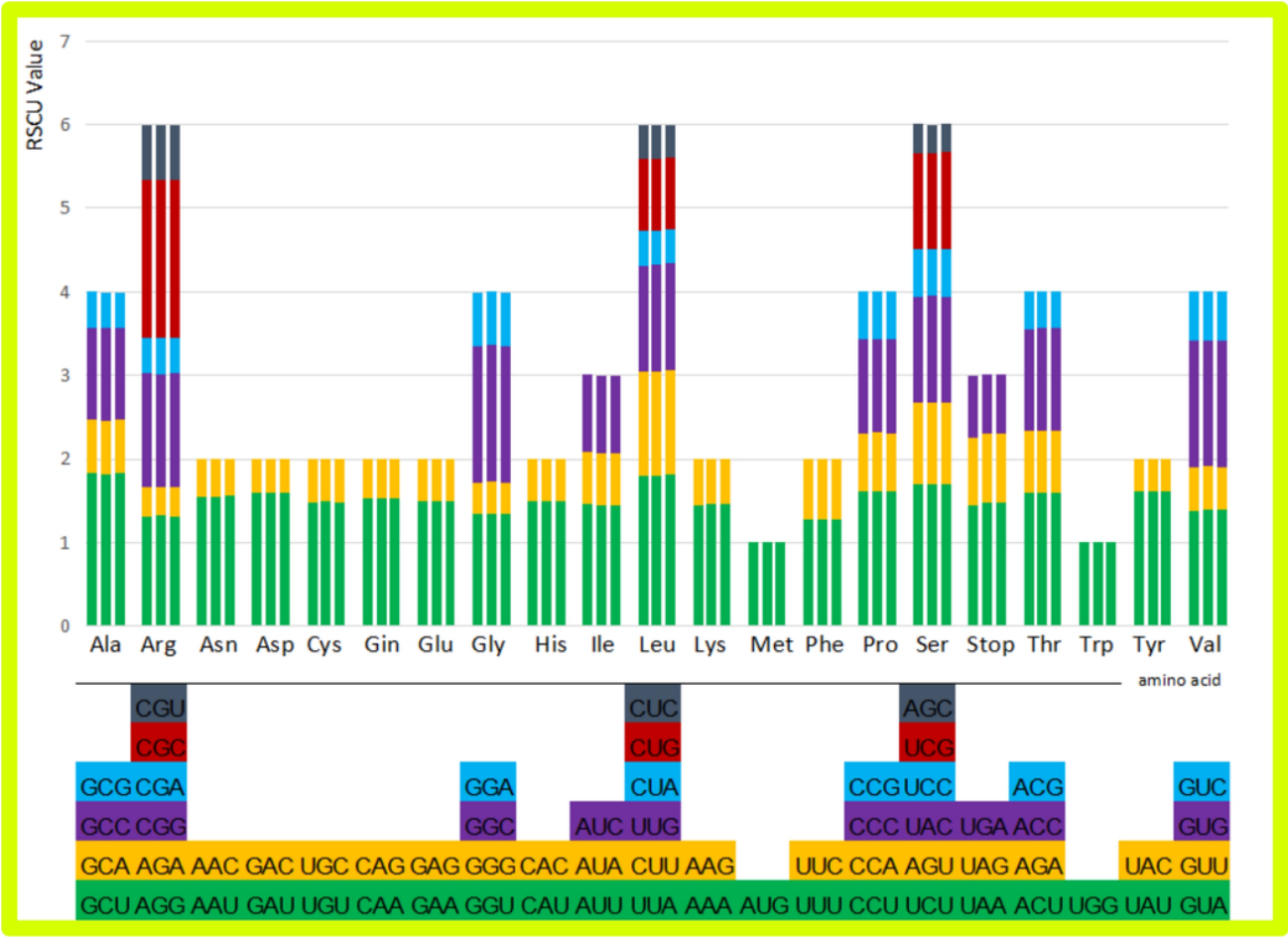


Figure 2

Codon content for the three *Nekemias* cp genomes.

(For each amino acid, the corresponding species from left to right are *N. grossedentata*, *N. cantoniensis*, *N. megalophylla*. The abscissa represents 20 amino acids and terminators, while the ordinate represents the RSCU value. The different colors of each amino acid correspond to the codon of the same color below. The names of all amino acids are replaced by abbreviations, such as Ala=Alanine)

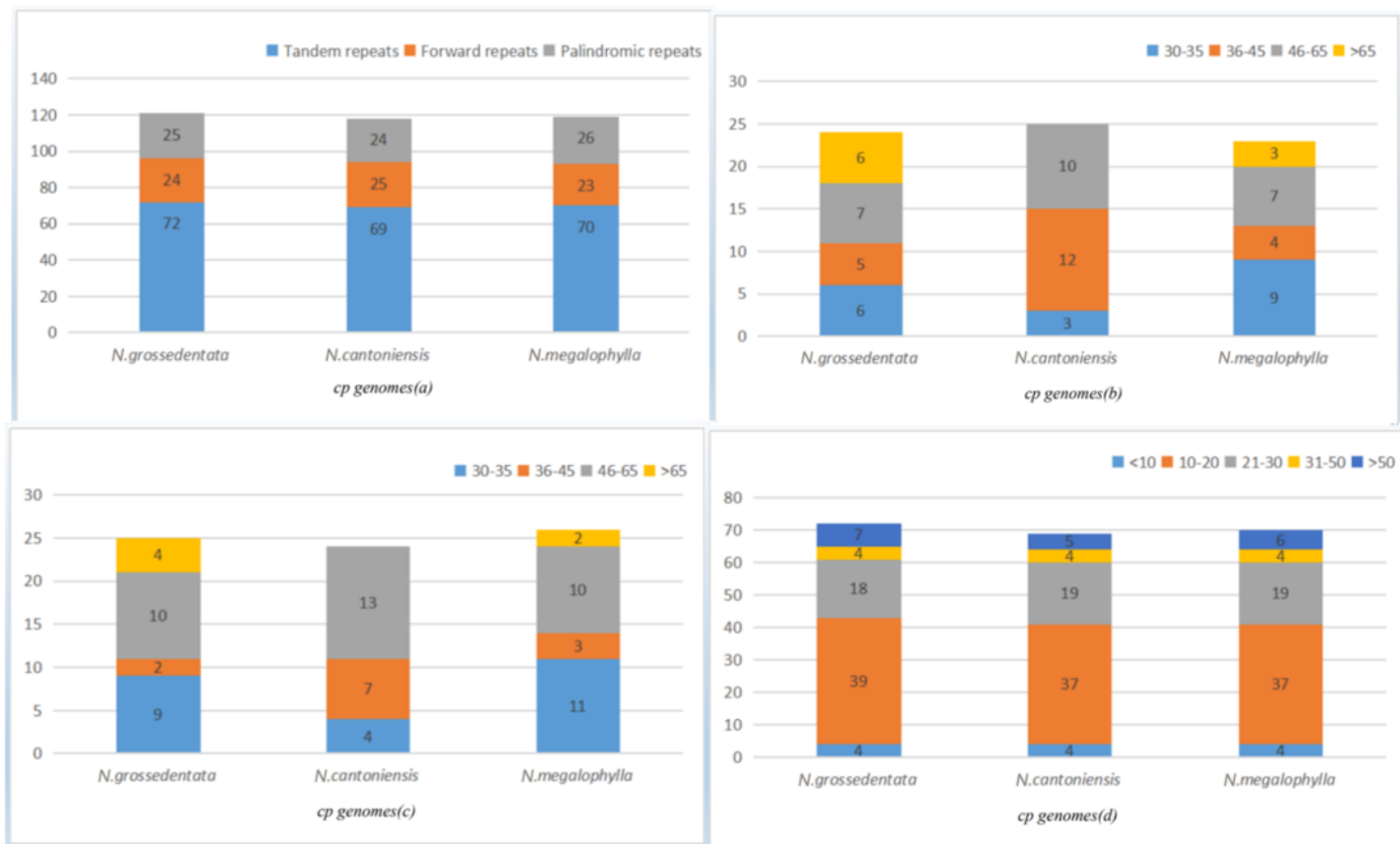


Figure 3

Repeat sequences analysis of three cp genomes.

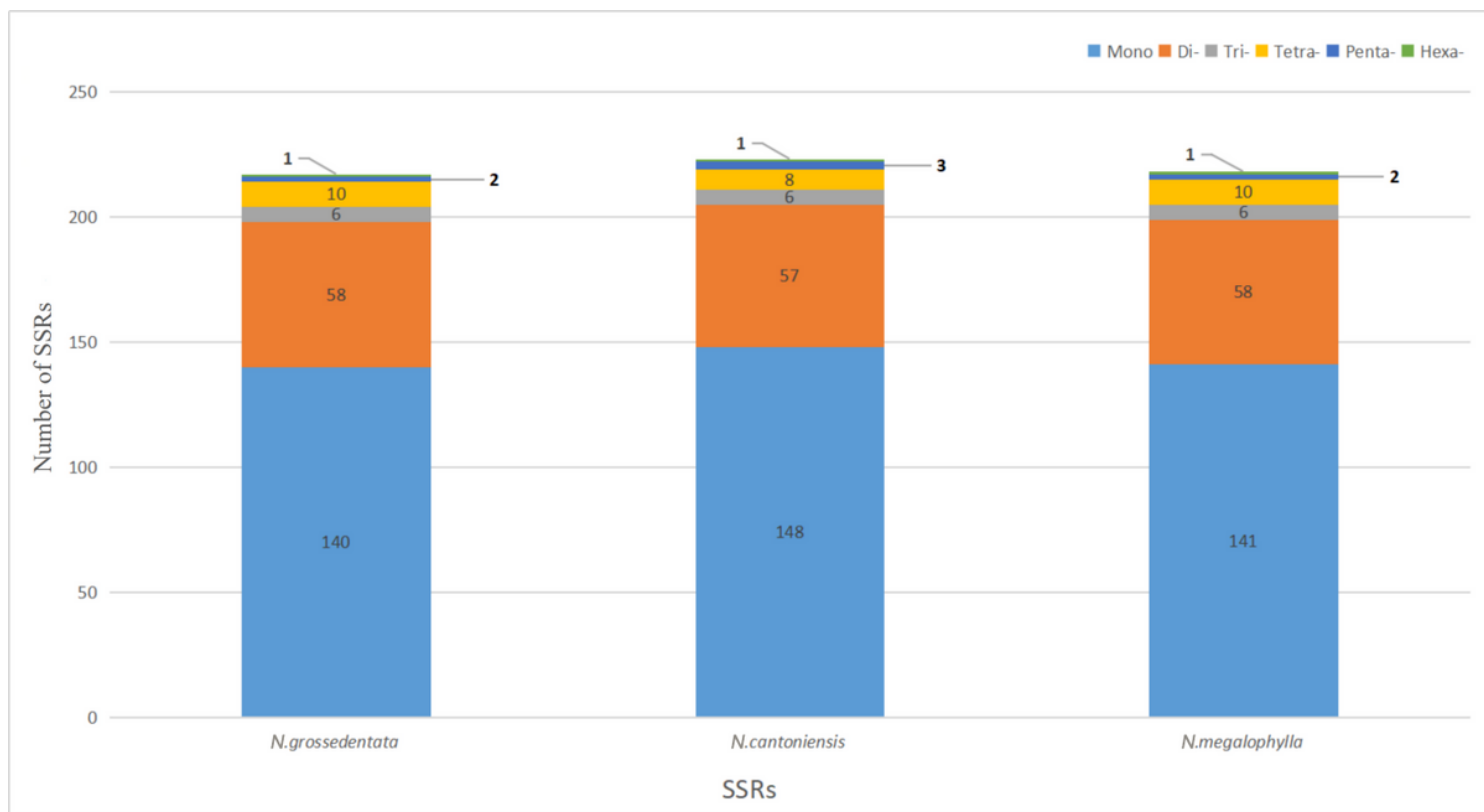


Figure 4

SSR analysis of three cp genomes. The ordinate represents the number of SSRs

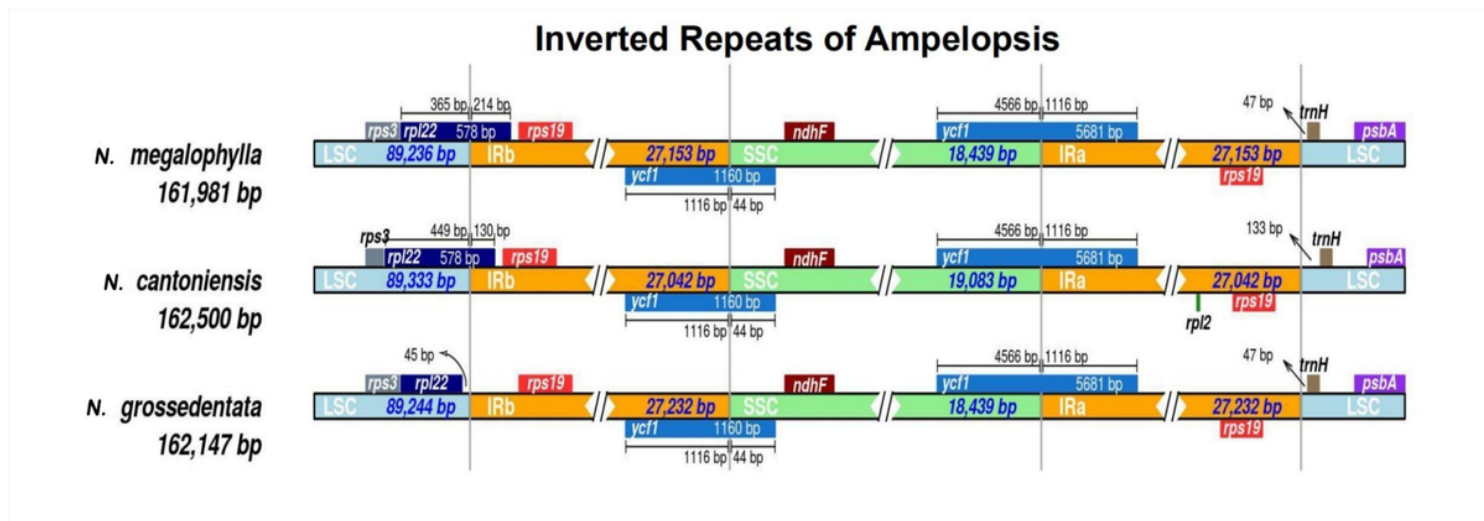


Figure 5

Comparison of LSC, SSC and IR regional boundaries of chloroplast genome between three species of *Nekemias*. JLB: junction line between LSC and IRb; JSB: junction line between IRb and SSC; JSA: junction line between SSC and IRa; JLA: junction line between IRa and SSC

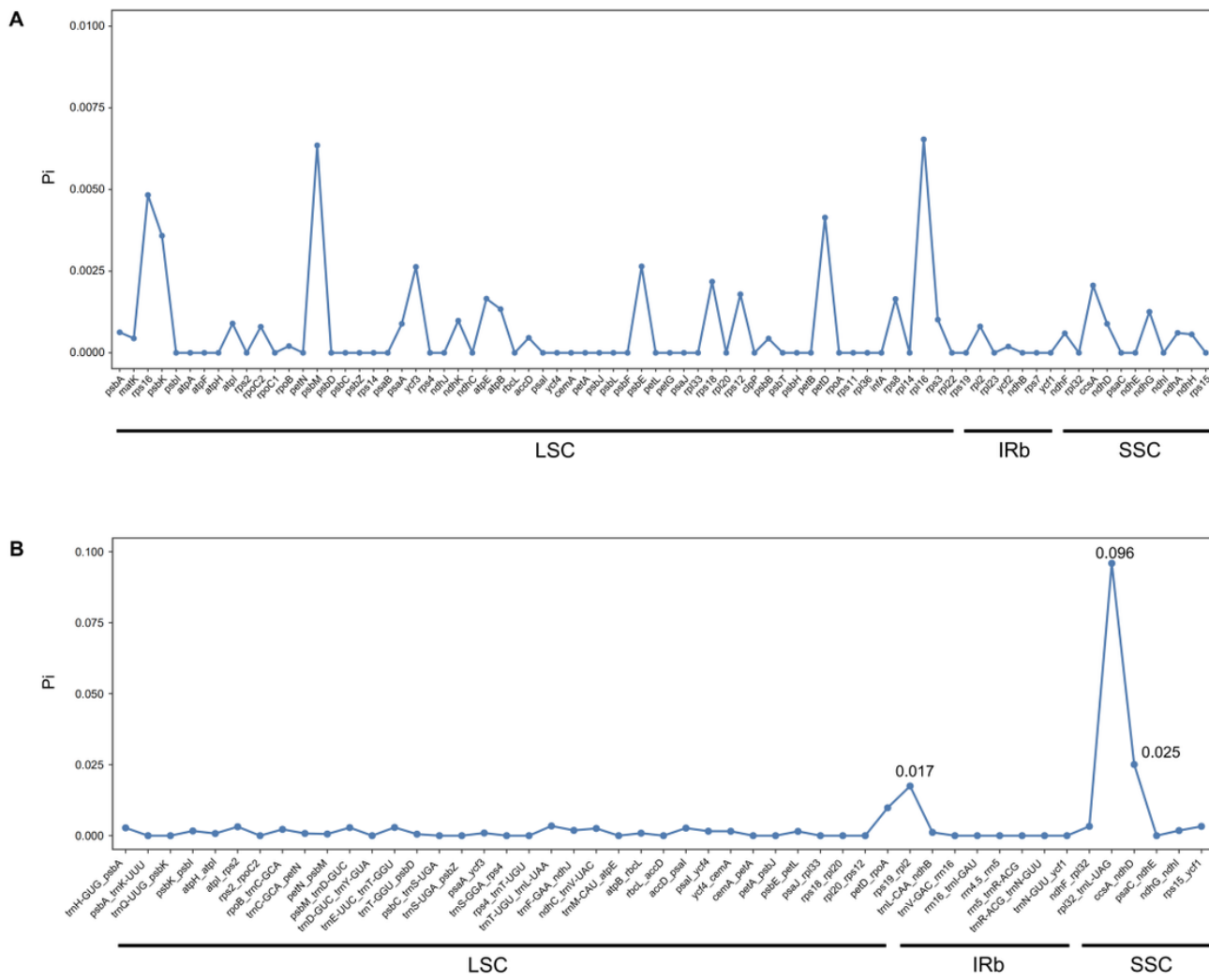


Figure 8

The nucleotide variability (P_i) values were compared among *N. grosedentata*, *N. megalophylla*, *N. cantoniensis*. (A) CDS region; (B) Intergenic spacer region.

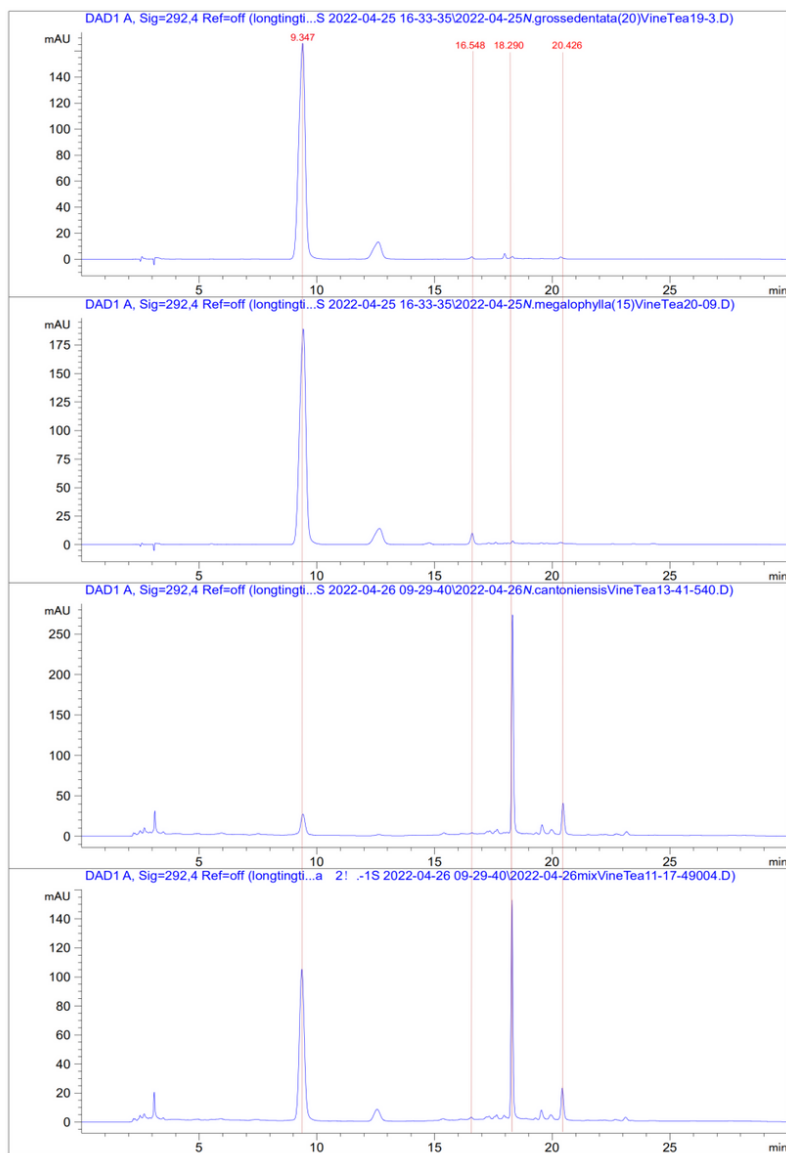


Figure 9

Fingerprinting of three *Nektomias* species based on HPLC. (The red baseline and time represent the peak time of the standard, from left to right are Dihydromyricetin (9.347), Taxifolin (16.548), Myricitrin (18.290), Myricetin (20.426); The three species from top to bottom were *N. grossedentata*, *N. megalophylla*, and *N. cantoniensis*. The bottom picture shows the mixed control. The wavelength of the characteristic spectra was 292nm, and the detailed spectra are shown in the Supplementary data.)