Comparison of the chloroplast genome sequences of representative species in the traditional herbal medicinal genus *Aconitum* (Ranunculaceae) (#18521)

First submission

Please read the **Important notes** below, the **Review guidance** on page 2 and our **Standout reviewing tips** on page 3. When ready **submit online**. The manuscript starts on page 4.

Important notes

Editor and deadline

Michael Wink / 10 Jul 2017

Files	4 Figure file(s) 4 Table file(s) 1 Other file(s) Please visit the overview page to **download and review** the files not included in this review PDF.
Declarations	**One or more DNA sequences were reported.**

Review guidelines

Please read in full before you begin

How to review

When ready **submit your review online**. The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

To finish, enter your editorial recommendation (accept, revise or reject) and submit.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to **PeerJ standards**, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see **PeerJ policy**).

EXPERIMENTAL DESIGN

- Original primary research within **Scope of the journal**.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- Data is robust, statistically sound, & controlled.
- Conclusions are well stated, linked to original research question & limited to supporting results.
- Speculation is welcome, but should be identified as such.

The above is the editorial criteria summary. To view in full visit **https://peerj.com/about/editorial-criteria/**

7 Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that your international audience can clearly understand your text. I suggest that you have a native English speaking colleague review your manuscript. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Give specific suggestions on how to improve the manuscript	*Line 56: Note that experimental data on sprawling animals needs to be updated. Line 66: Please consider exchanging "modern" with "cursorial".*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Comparison of the chloroplast genome sequences of representative species in the traditional herbal medicinal genus *Aconitum* (Ranunculaceae)

Hanghui Kong [1], **Wanzhen Liu** [2], **Gang Yao** [3], **Wei Gong** Corresp. [2]

[1] Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China
[2] College of Life Sciences, South China Agricultural University, Guangzhou, China
[3] College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou, China

Corresponding Author: Wei Gong
Email address: wgong@scau.edu.cn

The herbal medicinal genus *Aconitum* L., belonging to the Ranunculaceae family, represents the earliest diverging lineages within the eudicots. It is currently composed of two subgenera, *A*. subgenus *Lycoctonum* and *A*. subg. *Aconitum*. To better understand the phylogenetic relationship and to provide molecular information for utilization of *Aconitum* species, the complete chroloplast (cp) genome sequences of three species *A. angustius*, *A. finetianum* and *A. sinomontanum* in subg. *Lycoctonum* were characterized and compared with other members in *Aconitum*. The cp genome sequences are 156,109 bp for *A. angustius*, 155,625 bp for *A. finetianum* and 157,215 bp for *A. sinomontanum*, respectively. All three species possess 126 genes with 84 protein coding genes (PCGs). Structure variations were detected in the LSC/IR/SSC boundaries among the *Aconitum* species. Five pseudogenes were identified, among which *Ψrps*19 and *Ψycf*1 were located in LSC/IR/SSC boundaries, *Ψrps*16 and *Ψinf*A in LSC region, and *Ψycf*15 in IRb region. Synteny analyses showed no gene rearrangement and inversion events in *Aconitum*. The nucleotide variability (*Pi*) of *Aconitum* was estimated to be 0.00549, with comparably higher variations in LSC and SSC regions than IRs regions. Eight intergenic regions are highly variable. Altogether 50 simple sequence repeats (SSRs) were detected in *A. finetianum* and *A. angustius*, while 57 SSRs in *A. sinomontanum*. More than 80% of SSRs were present in LSC region. Altogether, 62% of SSRs are mononucleotides in subg. *Lycoctonum*, and 46.81% in subg. *Aconitum*. Comparably, higher percentage of di-, tri-, tetra-, and penta- SSRs were present in subg. *Aconitum* than those in subg. *Lycoctonum*. The availability of the complete cp genome sequences of three species in subg. *Lycoctonum*, will benefit for further phylogenetic reconstruction and aid in the germplasm utilization of *Aconitum* species.

Comparison of the chloroplast genome sequences of representative species in the traditional herbal medicinal genus *Aconitum* (Ranunculaceae)

Hanghui Kong[1, 4], Wanzhen Liu [2], Gang Yao[3], Wei Gong[2, *]

[1] *Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China*

[2] *College of Life Sciences, South China Agricultural University, Guangzhou, 510614, China*

[3] *College of Forestry and Landscape Architecture, South China Agricultural University, Guangzhou, 510614, China*

[4] *Guangdong Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China*

Corresponding author

* Wei Gong

Address: College of Life Sciences, South China Agricultural University, Guangzhou 510614, China

E-mail: wgong@scau.edu.cn

Tel: +86 20-37088127; Fax: +86 20-37088127

Abstract

The herbal medicinal genus *Aconitum* L., belonging to the Ranunculaceae family, represents the earliest diverging lineages within the eudicots. It is currently composed of two subgenera, *A*. subgenus *Lycoctonum* and *A*. subg. *Aconitum*. To better understand the phylogenetic relationship and to provide molecular information for utilization of *Aconitum* species, the complete chroloplast (cp) genome sequences of three species *A. angustius*, *A. finetianum* and *A. sinomontanum* in subg. *Lycoctonum* were characterized and compared with other members in *Aconitum*. The cp genome sequences are 156,109 bp for *A. angustius*, 155,625 bp for *A. finetianum* and 157,215 bp for *A. sinomontanum*, respectively. All three species possess 126 genes with 84 protein coding genes (PCGs). Structure variations were detected in the LSC/IR/SSC boundaries among the *Aconitum* species. Five pseudogenes were identified, among which *Ψrps*19 and *Ψycf*1 were located in LSC/IR/SSC boundaries, *Ψrps*16 and *Ψinf*A in LSC region, and *Ψycf*15 in IRb region. Synteny analyses showed no gene rearrangement and inversion events in *Aconitum*. The nucleotide variability (*Pi*) of *Aconitum* was estimated to be 0.00549, with comparably higher variations in LSC and SSC regions than IRs regions. Eight intergenic regions are highly variable. Altogether 50 simple sequence repeats (SSRs) were detected in *A. finetianum* and *A. angustius*, while 57 SSRs in *A. sinomontanum*. More than 80% of SSRs were present in LSC region. Altogether, 62% of SSRs are mononucleotides in subg. *Lycoctonum*, and 46.81% in subg. *Aconitum*. Comparably, higher percentage of di-, tri-, tetra-, and penta- SSRs were present in subg. *Aconitum* than those in subg. *Lycoctonum*. The availability of the complete cp genome sequences of three species in subg. *Lycoctonum*, will benefit for further phylogenetic reconstruction and aid in the germplasm utilization of *Aconitum* species.

47

48 **Keywords:** *Aconitum*; chloroplast genome; herbal medicine; next generation sequencing;

49 Phylogenetic reconstruction; Ranunculaceae

INTRODUCTION

The chloroplast (cp) is an intracellular organelle, which plays an important role in the process of photosynthesis and is widely present in algae and plant (*Neuhaus & Emes, 2000; Inoue, 2011*). The cp genome in angiosperms is circular DNA molecule with typically quadripartite structure, consisting of two copies of a large inverted repeat (IR) region that separate a large-single-copy (LSC) region from a small-single-copy (SSC) region (*Raubeson & Jansen, 2005*; *Yang et al., 2010; Gree, 2011*; *Wicke et al., 2011*). Though highly conserved among plants, some differences of gene synteny, copy number and pseudogenes have been observed in the cp genome structures (*Shradha et al., 2010*; *Lei et al., 2016*; *Ivanova et al., 2017*). In the past years, the complete cp genome has extensively been used in plant taxonomical analyses, phylogenetic reconstruction, speciation process and biogeographical inference at different taxonomic levels. In particular, the cp genome is useful to investigate the maternal origin in plants, especially those with polyploid species, due to its haploid maternal inheritance and high conservation in gene content and genome structure (*Birky, 1995*; *Soltis & Soltis, 2000*; *Song et al, 2002*). High-throughput sequencing technologies speed up the achievement of cp genome sequences and advanced the shifts from phylogenetics to phylogenomics. Highly valuable informative universal markers based on indels, substitutions and inversions of cp genome have been further developed for various molecular studies in plants.

The genus *Aconitum* L. belongs to the tribe Delphinieae in the Ranunculaceae family and represents one of the earliest diverging lineages within the eudicots APG IV (*Wang et al., 2009*; *Sun et al., 2011*; *The Angiosperm Phylogeny Group, 2016*). The genus *Aconitum* is currently composed of two subgenera, *A*. subgenus *Lycoctonum* and *A*. subg. *Aconitum* (*Jabbour & Renner, 2012*; *Wang et al., 2013*). *Aconitum angustius*, which is tretraploid in subg. *Lycoctonum*,

possesses heterologous chromosomes and is suggested to be hybridized between *A. finetianum* and *A. sinomontanum*. Those species display intermediate morphological characteristics and overlapping geographical distributions (*Shang & Lee, 1984*; *Yuan & Yang, 2006*;*Gao, 2009*; *Gao, Ren & Yang, 2012*). Based on previous morphological analysis and phylogenetic inference, *A. finetianum* was inferred to be the potentially maternal progenitor of *A. angustius* (*Gao, 2009*; *Kong et al., 2017*).

The genus *Aconitum* is known as taxonomically and phylogenetically challenging taxa in the last decade years. Early divergence between subg. *Lycoctonum* and subg. *Aconitum* in Europe was suggested based on *trn*H-*psb*A and ITS (*Utelli, Roy & Baltisberger*; 2000). Though high morphological variability within and among populations was detected due to recent speciation, the morphological characters are not as valuable as systematic characters. *Jabbour & Renner* (2012) have conducted phylogenetic reconstruction focusing on Delphineae based on *trn*L-F and ITS, which suggested *Aconitum* to be monophyletic clade and sister group of *Delphinium*. Most recently, phylogenetic inferences of the relationship among the polyploid species in subg. *Lycoctonum* have been made using four cpDNA intergenic regions (*ndh*F-*trn*L, *psb*A-*trn*H, *psb*D-*trn*T and *trn*T-L) and two nrDNA regions (ITS and ETS) (*Kong et al., 2017*). *Aconitum finetianum* was inferred to be the potential maternal progenitor of *A. angustius*. With the same cpDNA intergenic regions, taxonomical revision has been conducted based on phylogenetic analyses of subg. *Lycoctonum* by *Hong et al.* (2017). The application of cpDNA markers with high informative loci seem to be limited in the previous research. So far, no genomic level of phylogenetic information has been provided. Unclear phylogenetic position of some species still exists in *Aconitum*.

Even though some species are highly toxic because of aconite alkaloid, many species in

Aconitum are proved to be essential in the formulations of traditional herbal medicine in Asia, possessing a variety of medicinal importance (*Zhao et al., 2010*; *Semenov et al., 2016*; *Wada et al., 2016*; *Liang et al., 2017*). The unclear phylogenetic relationship would prohibit the identification among the *Aconitum* species. In this study, we report the complete cp genome sequences of three species in subg. *Lycoctonum*. We established and characterized the organization of the complete cp genome sequences of tetraploid *A. angustius* as well as diploid *A. finetianum* and *A. sinomontanum*. We further compare the structure, gene arrangement and microsatellite repeats (SSRs) with the related species in both subgenera of *Aconitum*. Altogether 13 species and two varieties from *Aconitum* were used for phylogenetic reconstruction at genomic level. Evidence of maternal origin from *A. finetianum* was investigated for tetraploid *A. angustius*. Our result will provide abundant information for the research of taxonomical identification, phylogenetic inference or population history of *Aconitum* or Ranunculaceae, which can also aid in the utilization of the genetic resources of *Aconitum* as a traditional herbal medicine.

MATERIALS AND METHODS

Plant samples and DNA extraction

Fresh leaves were collected from *A. angustius*, *A. finetianum* and *A. sinomontanum* growing in the greenhouse of South China Botanical Garden, Chinese Academy of Sciences. Total genomic DNA was extracted from the fresh leaves of *A. angustius*, *A. finetianum* and *A. sinomontanum* using the modified CTAB method (*Dolye & Dolye, 1987*). The DNA concentration was quantified using a Nanodrop spectrophotometer (Thermo Scientific, Carlsbad, CA, USA). The final DNA concentration >30 ng/μL were chosen for further Illumina sequencing.

120

121 **Chloroplast genome sequencing, assembling and annotation**

122 We sequenced the complete cp genome of *A. angustius*, *A. finetianum* and *A. sinomontanum* with
123 the Illumina HiSeq 2000 sequencing platform at Beijing Genomics Institute (BGI) in Wuhan,
124 China. Genomic DNA was fragmented randomly and then required length DNA fragments were
125 retained by electrophoresis. Adapters were ligated to DNA fragments followed by cluster
126 preparation and sequencing. A paired-end library was constructed with 270 bp insert size, and
127 then 150 bp paired reads were sequenced using the Illumina HiSeq 2000 sequencing platform.

128 We assembled the cp genomes using Geneious 9.1.4 (Biomatters Ltd., Auckland, New
129 Zealand) with blast and map to reference tools, respectively. Using the program DOGMA
130 (http://dogma.ccbb.utexas.edu/) (*Wyman, Jansen & Boore, 2004*) and Geneious, annotation was
131 performed in comparison with the cp genomes of *A. barbatum* var. *hispidum* (KT820664) and *A.*
132 *barbatum* var. *puberulum* (KC844054) (*Chen et al., 2015*) in subg. *Lycoctonum*. The annotations
133 of tRNA genes were further confirmed using ARAGORN (*Laslett & Canback, 2004*) and then
134 manually adjusted using the program Geneious. Hitting contigs from blast and consensus
135 sequence from map to reference function were subsequently assembled manually to construct
136 complete chloroplast genomes. Finally, the circular genome maps of the four species were
137 illustrated using Organellar Genome DRAW tool OGDRAW (http://ogdraw.mpimp-
138 golm.mpg.de/) (*Lohse et al., 2013*). The annotated chloroplast genomic sequences of *A.*
139 *angustius*, *A. finetianum* and *A. sinomontanum* have been submitted to GenBank.

140

141 **Genome comparison and divergence hotspot**

142 The cp genome sequences from the finalized data set were aligned with MAFFT v7.0.0 (*Katoh &*

143 *Standley, 2012*) and adjusted manually when necessary. Altogether, 13 species and two varieties
144 in both subgenera of *Aconitum* were used for alignment (Table 1). Based on many other cp
145 genome studies, the IRs expansion/contraction could lead to changes in the structure of the cp
146 genome, leading to the length variation of angiosperm cp genomes and contributing to the
147 formation of pseudogenes (*Kim & Lee, 2004*; *Nazareno, Carlsen & Lohmann, 2015*; *Ivanova et*
148 *al., 2017*). Therefore, we conducted comparison analysis to detect the variation of the
149 LSC/IR/SSC boundaries among the species or varieties. Comparative analysis of the nucleotide
150 diversity (*Pi*) among the complete cp genomes of *Aconitum* was performed based on a sliding
151 window analysis using DnaSP 5.10 (*Librado & Rozas, 2009*). The window length was 600 bp
152 and step size was 200 bp. In order to test and visualize the presence of genome rearrangement
153 and inversions, gene synteny was performed by MAUVE as implemented in Geneious with
154 default settings based on thirteen species and two varieties in both subgenera.
155
156 **Simple sequence repeats analysis**
157 MISA (http://pgrc.ipk-gatersleben.de/misa/misa.html) (*Thiel et al., 2003*) is a tool for the
158 identification and location of perfect microsatellites and compound microsatellites (two
159 individual microsatellites, disrupted by a certain number of bases). We used per script MISA to
160 search for potential simple sequences repeats (SSRs) loci in the cp genomes of the three species.
161 The minimum numbers (thresholds) of the SSRs were set to be 10, 5, 4, 3, and 3 for mono-, di-,
162 tri-, tetra-, and penta-nucleotides SSRs. All of the repeats found were manually verified and
163 redundant results were removed.
164
165 **Phylogenetic analysis**

Three species and two varieties in subg. *Lycoctonum*, and ten species in subg. *Aconitum*, with *Megaleranthis saniculifolia and Clematis terniflora*v as outgroup, were used for phylogenetic reconstruction. The cp genome sequences from the finalized data set were aligned with MAFFT. The complete cp genome sequences and PCGs were used, respectively, for the phylogenetic reconstruction for both subgenera in *Aconitum*. Three different methods including Bayesian Inference (BI), Maximum Parsimony (MP) and Maximum Likelihood (ML) were employed. In all analyses, gaps were treated as missing.

Bayesian Inference (BI) of phylogenies was performed using MrBayes v.3.2 (*Huelsenbeck & Ronquist, 2001*; *Ronquist & Huelsenbeck, 2003*). The best model was determined for each sequence partition, after comparison among 24 models of nucleotide substitution using Modeltest v.3.7 (*Posada & Crandall, 1998*). We performed Maximum Parsimony (MP) by using PAUP* v.4.0b10 (*Swofford, 2002*). We calculated the bootstrap values with 1000 bootstrap replicates, each with 10 random sequence addition replicates holding a single tree for each run. We conducted Maximum Likelihood (ML) using RAxML (*Stamatakis, 2006*) and the RAxML graphical interface (raxmlGUI v.1.3 (*Silvestro & Michalak, 2012*) with 1000 rapid bootstrap replicates. The general time-reversible (GTR) model was chosen with a gamma model of the rate of heterogeneity.

RESULTS AND DISCUSSION

Genome Organization and Features

Using Illumina HiSeq 2000 sequencing platform, a total number of 2x150 bp pair-end reads ranging from 9,879,068 to 27,530,148 bp were produced for three species in subg. *Lycoctonum*. Altogether, 1,270 Mb clean data were produced for *A. angustius*, 3,586 Mb for *A. finetianum,*

and 3,590 Mb for *A. sinomontanum*. The de novo assembly generated average 6713 contigs with N50 length of average 732 bp for *A. angustius*, average 6201 contigs with N50 length of average 801 bp bp for *A. finetianum* and average 6999 contigs with N50 length of average 769 bp for *A. sinomontanum*. Scaffolds from assembly with k-mer value 35 to 149 were matched to reference cp genome sequences, which were used to determine relative position and direction respectively. We generated a new draft chloroplast genome by manual identification of overlap regions. Double check and correction according to quality and coverage of each base position by reads remapping were conducted for further determination of the draft genome. The complete cp genome sequences of the three species with full annotations were deposited into GenBank (Accession Number: MF155664, MF155665 and MF155666).

The size of the cp genomes was 156,109 bp for *A. angustius*, 155,625 bp for *A. finetianum* and 157,215 bp for *A. sinomontanum* (Table 1). Chloroplast genomes displayed a typical quadripartite structure, including a pair of IRs (25,927-26,225 bp) separated by LSC (86,664-88,074 bp) and SSC (16,914-17,107 bp) regions (Fig. 1 and Table 1). The GC content of the three species is 38.00%, demonstrating congruence to that reported in *A. barbatum* var. *hispidum* and *A. barbatum* var. *puberulum* (38.00%) of subg. *Lycoctonum* as well as in the species of subg. *Aconitum* (38.00% or 38.10%) (Table 1).

When duplicated genes in IRs regions were counted only once, cp genomes of *A. angustius*, *A. finetianum* and *A. sinomontanum* encode 126 predicted functional genes, including 84 PCGs, 38 tRNA genes and four rRNA genes. The remaining non-coding regions include introns, intergenic pacers, and pseudogenes. Altogether 18 genes were duplicated in the IRs region, including seven PCGs and seven tRNA genes and four rRNA genes (Fig. 1; Table 2). Thirteen genes (8 PCGs and five tRNA genes) contain one interval, and three PCGs (*clp*P, *ycf*3 and *rps*12)

have two intervals (Table 2). The maturase K (*mat*K) gene in the cp genomes of the three species is located within *trn*K intron, which is similar in most other plants species. In the IRs regions, the four rRNA genes and two tRNA genes (*trn*I and *trn*A) are clustered as 16S-*trn*I-*trn*A-23S-4.5S-5S, as is found in cp genomes of *A. barbatum* var. *hispidum*and and *A. barbatum* var. *puberulum*as well as in many other plant species (*Mardanov et al., 2008*; *Wu et al., 2014*; *Chen et al., 2015*).

Comparative analysis of genomic structures

Synteny analysis has been performed in order to identify the potential genome rearrangement and inversions based on the cp genome sequences of *Aconitum* species. No gene rearrangement and inversion events were detected (Fig. S1). Genomic structure including gene number and gene order seems to be highly conserved among the *Aconitum* species. However, structure variations were still present in the LSC/IR/SSC boundaries (Fig. 2). The genes *rps*19-*rp*l2-*trn*H and *ycf*1-*ndh*F were located in the junction regions of LSC/IR and SSC/IR. The *rps*19 gene, crossing the LSC/IRa junction region in *A. sinomontanum*, *A. barbatum* var. *puberulum* and *A. barbatum* var. *hispidum* of subg. *Lycoctonum*, as well as in *A. jaluense*, *A. volubile*, *A. carmichaelii*, *A. kusnezoffii* and *A. monanthum* of subg. *Aconitum*, has apparently lost its protein-coding ability due to partial gene duplication in IRb region, thus producing pseudolized *Ψrps*19 gene. This is the same case with the *ycf*1 gene, as the IRb/SSC junction region is located within *ycf*1 CDS region and only partial gene is duplicated in IRa region, resulting in a pseudogene. This is a general structure among the dicots. The length of pseudogene *Ψycf*1 in IRs regions was 1279 bp for two varieties in subg. *Lycoctonum* and seven species in subg. *Aconitum*. However, it showed length variation among the remaining six species, which are *A. angustius* (1292 bp), *A.*

finetianum (1165 bp) and *A. sinomontanum* (1292 bp) in subg. *Lycoctonum*, as well as *A. chiisanense* (1274 bp), *A. volubile* (1356 bp) and *A. carmichaelii* (1263 bp) in subg. *Aconitum* (Fig. 2; Table 3).

Additional three pseudogenes *Ψycf*15, *Ψrps*16 and *Ψinf*A were identified (Table 3). The *Ψycf*15 gene appears to be pseudolized in *A. austrokoreense* and *A. chiisanense* with four bases insertions as well as in *A. monanthum* with one base insertion, contributing to several internal stop codons. The *Ψinf*A region is pseudolized with two nonsynonymous substitutions producing internal stop codons in all members of subg. *Lycoctonum*. The pseudolized *Ψinf*A gene has also been found in other angiosperm chloroplast genomes (*Raman & Park, 2015*; *Lu, Li & Qiu, 2017*). The gene *rps*16 is responsible for ribosomal protein S16 and coded in the cp genome in most higher plants. However, it has been detected to be functionally lost in various plant species (*Shradha et al, 2010*). A pseudogene *Ψrps*16 is revealed to be present in the cp genomes of *A. angustius* and *A. finetianum* in subg. *Lycoctonum* as well as nine species in subg. *Aconitum* due to the loss of one CDS region (Table 3). As is revealed in other studies, the functional loss of *rps*16 gene might be compensated by the dual targeting of the nuclear *rps*16 gene product (*Keller et al., 2017*).

Sequence divergence among the species in *Aconitum*

The nucleotide variability (*Pi*) values were estimated to be 0.00549, ranging from 0 to 0.03856, using comparative analysis of sequence divergence based on complete cp genome sequences of *Aconitum* species. The highest variation was found in LSC and SSC regions, with an average *Pi* = 0.007140 and 0.008368, respectively. The IRs regions showed much lower nucleotide diversity with *Pi* = 0.001079 and 0.001459. Eight intergenic regions (*trn*H-*psb*A, *trn*K-*rps*16, *trn*D-*trn*Y,

*trn*Y-*trn*E, *trn*E-*trn*T, *trn*T-*trn*L, *rpl*12-*clp*P and *trn*H-*trn*R) are highly variable, with *Pi* value around 0.023 (Fig. 3). The former eight loci are present in LSC region, while the pseudogene *Ψycf*1 is in SSC region. The single-copy regions have been demonstrated to be highly variable with loci clustered in 'hot spots' (*Kong & Yang, 2017*). Among the eight intergenic regions, *trn*H-*psb*A and *trn*T-*trn*L has been reported to be variable and useful for phylogenetic reconstruction within the subg. *Lycoctonum* (*Utelli, Roy & Baltisberger, 2000*; *Kong et al., 2017*). However, the other intergenic regions, even with higher nucleotide variability, have never been involved in the phylogentic analysis for the genus *Aconitum*. The highly variable loci detected in the current study may provide a basis for further deep phylogenetic reconstruction of this genus. The observed divergence hotspot regions provided abundant information for marker development in phylogenetic analyses or conservation genetics of *Aconitum*.

Characterization of simple sequence repeats

With MISA analysis, 50 SSRs with minimum 10 bp repeats in length were detected in the *A. finetianum* and *A. angustius*, but 57 SSRs were detected in *A. sinomontanum* (Table 4). This result is comparable with those reported in *A. barbatum* var. *hispidum* (53 SSRs) and *A. barbatum* var. *puberulum* (57 SSRs) in subg. *Lycoctonum*, but relatively higher than that of subg. *Aconitum* (an average of 47 SSRs). In both subgenera, most SSRs are located in LSC regions, with an average of 92.00% and 80.85% in subg. *Lycoctonum* and subg. *Aconitum*, respectively. Among all the SSRs, the mononucleotide A/T repeat units occupied the highest proportion with 62% and 46.82% of total SSRs in subg. *Lycoctonum* and subg. *Aconitum*, respectively. Though with lower number of SSRs, higher proportion of di-, tri-, tetra- and penta-nucleotide repeats were detected in subg. *Aconitum*. The SSRs have remarkably high A/T content with only seven

SSRs, including $(ATCT)_3$, $(TTCT)_3$, $(CTTT)_3$, $(TAAAG)_3$, $(TTTC)_3$, $(ATAC)_3$ and $(CATT)_3$, contain one C or G nucleotide.

A total of 11 cp SSRs loci were shared among the cp genomes of tetraploid *A. angustius* and diploid *A. finetianum*. No common cp SSRs were found between *A. angustius* and *A. sinomontanum*. This result provides evidence of the maternal origin of the tetraploid *A. angustius* from diploid *A. finetianum*, which is consistent with previous research (*Gao, 2009*; *Kong et al., 2017*). Among the three species, the highest number of unique SSRs loci were present in *A. sinomontanum* (11) followed by *A. angustius* (7) and *A. finetianum* (6).

Phylogenetic analyses

In the present study, two datasets including complete cp genome sequences and 84 to 86 PCGs of thirteen species and two varieties from subg. *Lycoctonum* and subg. *Aconitum* were used to perform phylogenetic analyses of *Aconitum*. The total aligned length with parsimony informative loci is 178,392 bp with 4,342 for the complete cp genome sequences, and 106,535 bp with 3,164 for PCGs, respectively. The topologies based on three different methods yielded mostly concordant tree topologies across all analyses, with high post Bayesian posterior probabilities and bootstrap values at each node (Fig. 4). All the phylogenetic trees support the monophyly of *Aconitum* comprising of two monophyletic subgenus of subg. *Aconitum* and subg. *Lycoctonum*, respectively.

Based on the phylogenetic tree, the tetraploid *A. angustius* was always closely related with *A. finetianum*, which is also supported by previous research (*Kong et al., 2017*). The two species co-occur on several mountains in southeast China and even grow very closely within a community (*Yuan & Yang, 2006*). The two species show very similar morphological

characteristics in having leaves 3-parted, the upper sepal cylindric and pedicels retrosely pubescent, often making them confused with each other (*Gao, Ren & Yang, 2012*). *A*conitum *finetianum* is supported to be the potential maternal progenitor of *A*. *angustius* based on both molecular and morphological evidence (*Kong et al., 2017*). Therefore, it is reasonable to understand that the two species demonstrated close phylogenetic relationship based on cp genome sequences.

Acknowledgement

Special thank goes to Dr. Tongjian Liu and Dr. El Mahdi Bendif for great assistance in lab work and data analyses.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the National Natural Science Foundation of China (31470319; 31470312) and Science and Technology Planning Project of Guangdong Province, China (2016A030303048). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests

The authors declare no competing interests.

Author Contributions

Hanghui Kong and Wei Gong conceived and designed the experiments, collected the samples, analyzed the data and wrote the paper. Wanzhen Liu performed the experiment and contributed to analysis tools. Gang Yao contributed to the discussion and reviewed the drafts of the paper.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences: GenBank accession number: MF155664, MF155665 and MF155666.

Data Availability

The following information was supplied regarding data availability: The raw data can be found in https://doi.org/10.6084/m9.figshare.5092414.v1, https://doi.org/10.6084/m9.figshare.5092420.v1 and with the GenBank accession numbers in Table 1.

Supplemental Information

Supplemental information for this article can be found online.

REFERENCES

Birky CW. 1995. Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proceedings of the National Academy of Sciences of the United States of America* **92(25):**11331–11338 DOI 10.1073/pnas.92.25.11331.

Chen XC, Li QS, Li Y, Qian J, Han JP. 2015. Chloroplast genome of *Aconitum barbatum* var. *puberulum* (Ranunculaceae) derived from CCS reads using the PacBio *RS* platform. *Frontiers in Plant Science* **6(42):**1–9 DOI 10.3389/fpls.2015.00042.

Doyle JJ, Doyle JL, Rausher J, Brown AHD. 2004. Diploid and polyploid reticulate evolution throughout the history of the perennial soybeans (*Glycine* subgenus *Glycine*). *New Phytologist* **161:**121–132 DOI 10.1046/j.1469-8137.2003.00949.x.

Gao Q. 2009. Taxonomic revision of *Aconitum* L. subgenus *Lycoctonum* (DC.) *Peterm.* (Ranunculaceae) from China. D. Phil. Thesis, Institute of Botany, the Chinese Academy

of Sciences, Beijing, China

Gao Q, Ren C, Yang QE. 2012. Taxonomic status and distributional range of *Aconitum angustius* (Ranunculaceae) based on cytological evidence. *Nordic Journal of Botany* **30:**1–13 DOI 10.1111/j.1756-1051.2012.01506.x.

Green BR. 2011. Chloroplast genomes of photosynthetic eukaryotes. *the Plant Journal* **66:**34–44 DOI 10.1111/j.1365-313X.2011.04541.x.

Hong Y, Luo Y, Gao Q, Ren C, Yuan Q, Yang QE. 2017. Phylogeny and reclassification of *Aconitum* subgenus *Lycoctonum* (Ranunculaceae). *PLoS ONE* **12(1):**e0171038 DOI 10.1371/journal.pone.0171038.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17:**754–755 DOI 10.1093/bioinformatics/17.8.754.

Inoue K. 2011. Emerging roles of the chloroplast outer envelope membrane. *Trends in Plant Science* **16(10):**550–557 DOI 10.1016/j.tplants.2011.06.005.

Ivanova Z, Sablok, G, Daskalova E, Zahmanova G, Apostolova E, Yahubyan G, Baev V. 2017. Chloroplast genome analysis of resurrection Tertiary relic *Haberlea rhodopensis* highlights genes important for desiccation stress response. *Frontiers of Plant Sciences* **8:**204 DOI 10.3389/fpls.2017.00204.

Jabbour F, Renner SS. 2012. A phylogeny of *Delphinieae* (Ranunculaceae) shows that *Aconitum* is nested within *Delphinium* and that Late Miocene transitions to long life cycles in the Himalayas and Southwest China coincide with bursts in diversification. *Molecular Phylogenetics and Evolution* **62:**928–942 DOI 10.1016/j.ympev.2011.12.005.

Katoh K, Frith MC. 2012. Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics* **28(23):**3144–3146 DOI

10.1093/bioinformatics/bts578.

Kearse M, Moir R, Wilson, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28:**1647–1649 DOI 10.1093/bioinformatics/bts199.

Keller J, Rousseau-Gueutin M, Martin GE, Morice J, Boutte J, Coissac E, Ourari M, Aïnouche M, Salmon A, Cabello-Hurtado F, Aïnouche A. 2017. The evolutionary fate of the chloroplast and nuclear *rps16* genes as revealed through the sequencing and comparative analyses of four novel legume chloroplast genomes from *Lupinus*. *DNA Research* in press DOI 10.1093/dnares/dsx006.

Kim KJ, Lee HL. 2004. Complete chloroplast genome sequences from Korean ginseng (Panax *schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Research* **11:**247–261 DOI 10.1093/dnares/11.4.247.

Kong WQ, Yang JH. 2017. The complete chloroplast genome sequence of *Morus cathayana* and *Morus multicaulis*, and comparative analysis within genus *Morus* L. *PeerJ* **5:**e3037 DOI 10.7717/peerj.3037.

Kong HH, Zhang Y, Hong Y, Barker MS. 2017. Multilocus phylogenetic reconstruction informing polyploid relationships of *Aconitum* subgenus *Lycoctonum* (Ranunculaceae) in China. *Plant Systematics and Evolution* **303:**727–744 DOI 10.1007/s00606-017-1406-y.

Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research* **32:**11–16 DOI 10.1093/nar/gkh152.

Lei WJ, Ni DP, Wang YJ, Shao JJ, Wang XC, Yang D, Wang JS, Chen HM, Liu C. 2016. Intraspecific and heteroplasmic variations, gene losses and inversions in the chloroplast genome of *Astragalus membranaceus*. *Scientific Reports* **6:**21669 DOI 10.1038/srep21669.

Liang X, Chen L, Song L, Fei W, He M, He C, Yin Z. 2017. Diterpenoid alkaloids from the root of *Aconitum sinchiangense* W. T. Wang with their antitumor and antibacterial activities. *Natural Product Research* **11:**1–8 DOI 10.1080/14786419.2016.1272113

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25:**1451–1452 DOI 10.1093/bioinformatics/btp187.

Lohse M, Drechsel O, Kahlau S, Bock R. 2013. OrganellarGenomeDRAW — a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Research* **41(W1):**W575–W581 DOI 10.1093/nar/gkt289.

Lu RS, Li P, Qiu YX. 2017. The Complete Chloroplast Genomes of Three *Cardiocrinum* (Liliaceae) Species: Comparative Genomic and Phylogenetic Analyses. *Frontiers of Plant Sciences* **7:**2054 DOI 10.3389/fpls.2016.02054

Mardanov AV, Ravin NV, Kuznetsov BB, Samigullin TH, Antonov AS, Kolganova TV, Skyabin KG. 2008. Complete sequence of the duckweed (*Lemna minor*) chloroplast genome: structural organization and phylogenetic relationships to other angiosperms. *Journal of Molecular Evolution* **66:**555–564 DOI 10.1007/s00239-008-9091-7

Nazareno AG, Carlsen M, Lohmann, LG. 2015. Complete chloroplast genome of *Tanaecium tetragonolobum*: the first Bignoniaceae plastome. *PLoS ONE* **10:**e129930 DOI 10.1371/journal.pone.0129930.

Neuhaus H, Emes M. 2000. Nonphotosynthetic metabolism in plastids. *Annual Review of Plant Biology* **51:**111–140 DOI 10.1146/annurev.arplant.51.1.111.

Posada D, Crandall KA. **1998.** Modeltest: testing the model of DNA substitution. *Bioinformatics* **14:**817–818 DOI 10.1093/bioinformatics/14.9.817.

Raubeson LA, Jansen RK. 2005. Chloroplast genomes of plants. In: Henry RJ, ed. *Plant Diversity and Evolution: Genotypic and Phenotypic Variation in Higher Plants*. Cambridge: CABI Press, 45–68.

Roman G, Park S. 2015. Analysis of the Complete Chloroplast Genome of a Medicinal Plant, *Dianthus superbus* var. *longicalyncinus*, from a Comparative Genomics Perspective. **10(10):** e0141329 DOI 10.1371/journal.pone.0141329.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19:**1572–1574 DOI 10.1093/bioinformatics/btg180.

Shradha Roy, Ueda M, Kadowaki K, Tsutsumi N. 2010. Different status of the gene for ribosomal protein S16 in the chloroplast genome during evolution of the genus *Arabidopsis* and closely related species. *Genes and Genetic Systems* **85:**319–326 DOI 10.1266/ggs.85.319.

Semenov AA, Enikeev AG, Snetkova LV, Permyakov AV, Sokolova NA, Dudareva LV. 2016. Ortho-phthalic acid esters in lipophilic extract from the cell culture of *Aconitum baicalense* Turcz ex Rapaics 1907. *Doklady Biochemistry and Biophysics* **471:**421–422 DOI 10.1134/S1607672916060120.

Shang XM, Lee CL. 1984. Chromosome studies of ten species of *Aconitum* in China. *Acta Phytotaxonomica Sinica* **22:**378–385

Silvestro D, Michalak I. 2012. raxmlGUI: a graphical front-end for RAxML. *Organisms*

Diversity & Evolution **12:**335–337 DOI 10.1007/s13127-011-0056-0.

Soltis PS, Soltis DE. 2000. The role of genetic and genomic attributes in the success of polyploids. *Proceedings of the National Academy of Sciences of the United States of America* **97(13):**7051–7057 DOI 10.1073/pnas.97.13.7051.

Song BH, Wang XQ, Wang XR, Sun LJ, Hong DY, Peng PH. 2002. Maternal lineages of *Pinus densata*, a diploid hybrid. *Molecular Ecology* **11:**1057–1063 DOI 10.1046/j.1365-294X.2002.01502.x.

Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:**2688–2690

Sun G, Dilcher DL, Wang HS, Chen ZD. 2011. A eudicot from the Early Cretaceous of China. *Nature* **471:**625–628 DOI 10.1038/nature09811.

Swofford DL. 2002. PAUP*: Phylogenetic analysis using parsimony (* and other methods). Version 4.0b10, Sinauer Associates, Sunderland, Massachusets, USA

The Angiosperm Phylogeny Group. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* **181:**1–20 DOI 10.1111/boj.12385.

Thiel T, Michalek W, Varshney R, Graner A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeumvulgare* L.). *Theoretical Application of Genetics* **106:**411–422 DOI 10.1007/s00122-002-1031-0.

Utelli AB, Roy BA, Baltisberger M. 2000. Molecular and morphological analyses of European *Aconitum* species (Ranunculaceae) *Plant Systematics and Evolution* **224:**195–212 DOI 10.1007/BF00986343.

Wada K, Takeda K, Haraguchi M, Abe Y, Kuwahara N, Suzuki S, Terui A, Masaka T, Munakata N, Uchida M, Nunokawa M, Yamashita H, Goto M, Lee KH. 2016. Four new diterpenoid alkaloids from *Aconitum japonicum*. *Planta Medica* **81(S01):**S1–S381 DOI 10.1055/s-0036-1596757

Wang W, Liu Y, Yu SX, Gao TG, Chen ZD. 2013. *Gymnaconitum*, a new genus of Ranunculaceae endemic to the Qinghai-Tibetan Plateau. *Taxon* **62:**713–722 DOI 10.12705/624.10.

Wang W, Lu AM, Ren Y, Endress ME, Chen ZD. 2009. Phylogeny and classification of Ranunculales: Evidence from four molecular loci and morphological data. *Perspectives in Plant Ecology Evolution and Systematics* **11:**81–110 DOI 10.1016/j.ppees.2009.01.001.

Wicke S, Schneeweiss GM, dePamphilis, CW, Müller, KF, Quandt D. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology* **76:**273–297 DOI 10.1007/s11103-011-9762-4.

Wu Z, Gui S, Quan Z, Pan L, Wang S, Ke W, Liang D, Ding Y. 2014. A precise chloroplast genome of *Nelumbo nucifera* (Nelumbonaceae) evaluated with Sanger, Illumina MipSeq, and PacBio RS II sequencing platforms: insight into the plastid evolution of basal eudicots. *BMC Plant Biology* **14:**289 DOI 10.1186/s12870-014-0289-0.

Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **20(17):**3252–3255 DOI 10.1093/bioinformatics/bth352.

Yang M, Zhang XW, Liu GM, Yin YX, Chen KF, Yun QZ, Zhao DJ, Al-Mssallem IS, Yu J. 2010. The Complete Chloroplast Genome Sequence of Date Palm (*Phoenix dactylifera* L.). *PLoS ONE* **5(9):**e12762 DOI 10.1371/journal.pone.0012762.

488 **Yuan Q, Yang QE. 2006.** Polyploidy in *Aconitum* subgenus *Lycoctonum* (Ranunculaceae).
489 *Botanical Journal of the Linnean Society* **150:**343–353 DOI 10.1111/j.1095-
490 8339.2006.00468.x.

491 **Zhao Y, Bu G, Zhou Y, Lv L, Yan G, Chen B, Wang L, Cen X. 2010.** Mechanism study
492 of *Aconitum*-induced neurotoxicity in PC12 cells: involvement of dopamine release
493 and oxidative damage. *Neurotoxicology* **31:**752–757 DOI
494 10.1016/j.neuro.2010.06.005.

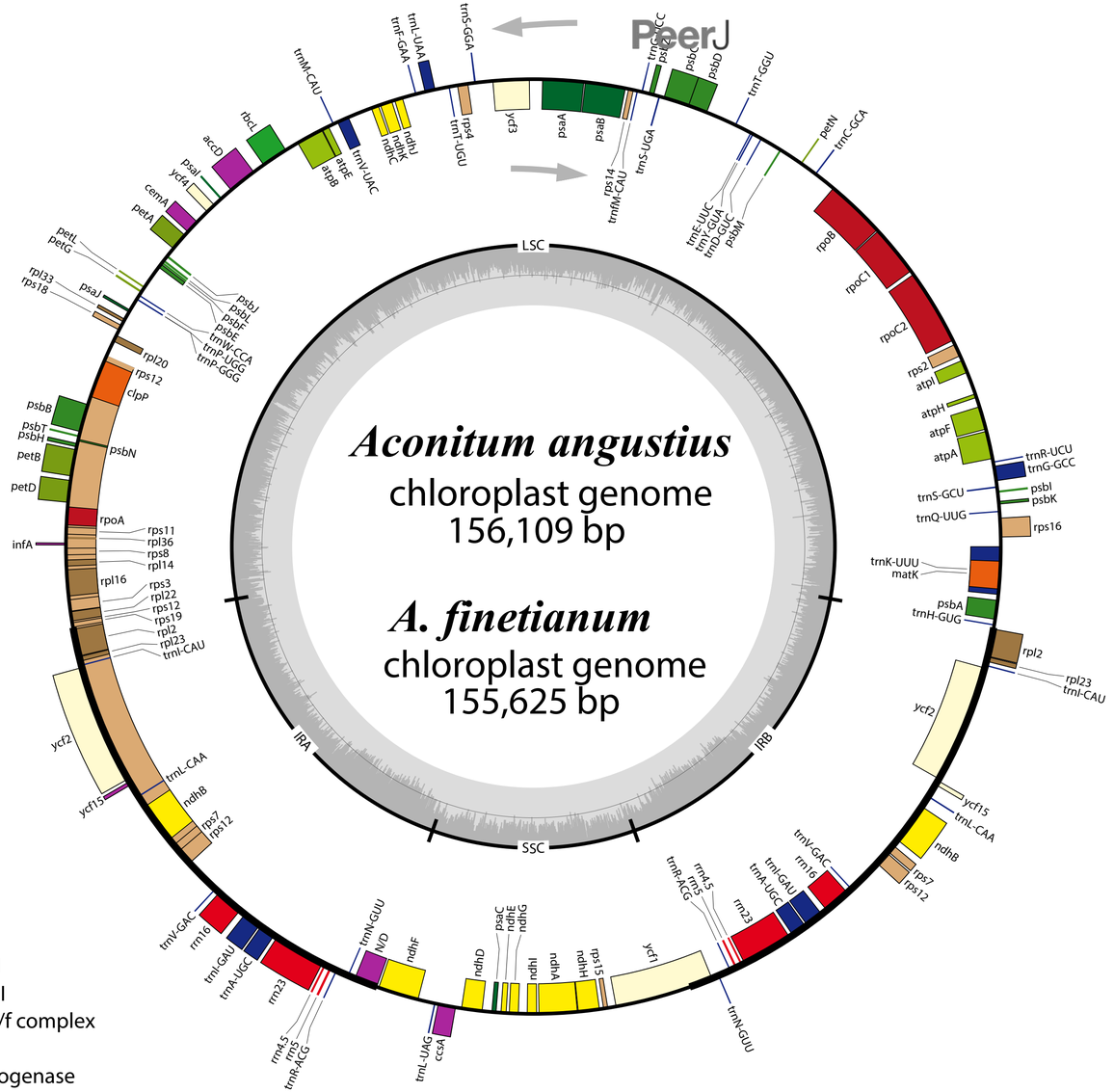
Figure 1(on next page)

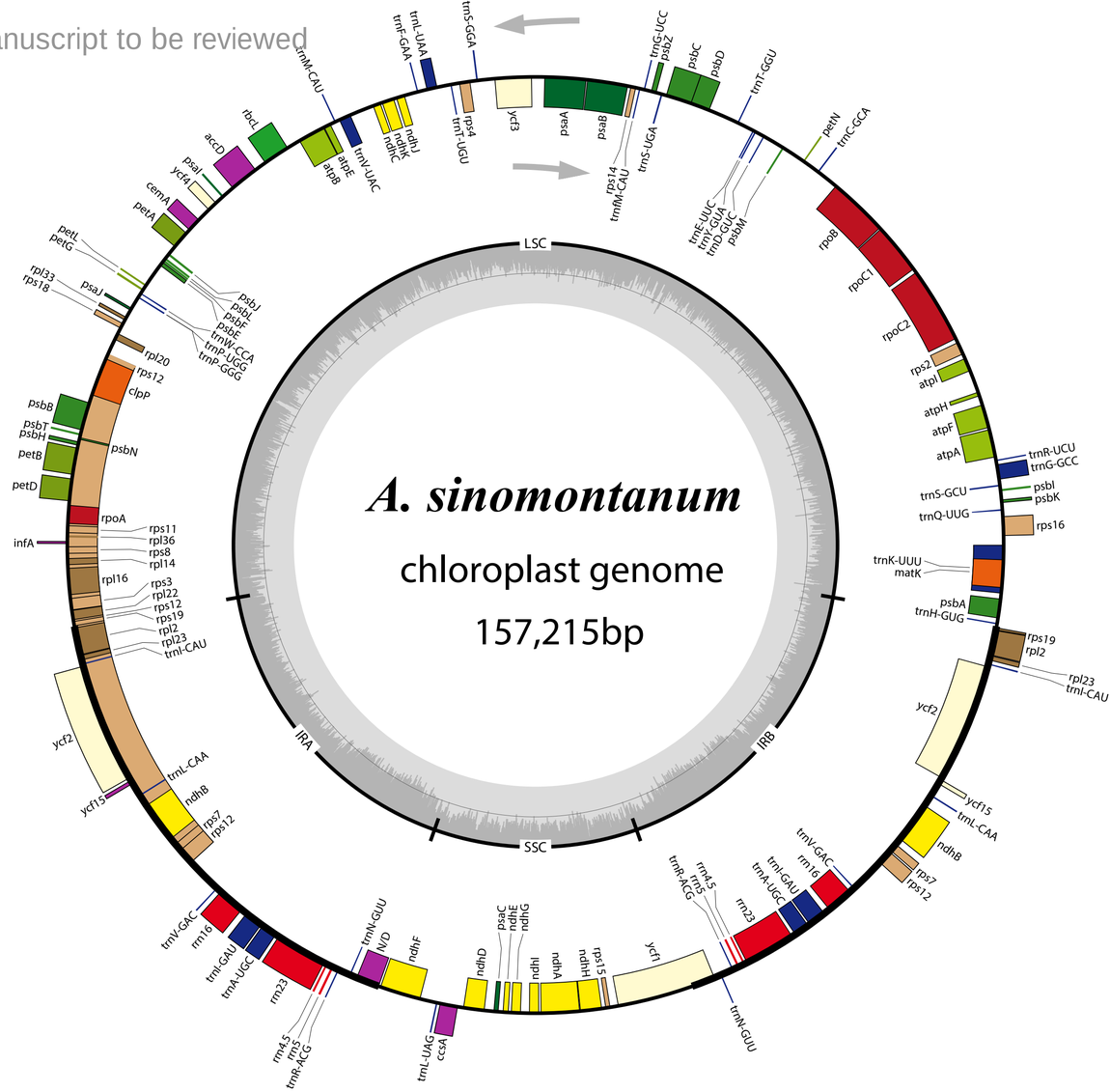
The gene maps of *Aconitum angustius* and *A. finetianum* (A) as well as *A. sinomontanum* (B).

Figure 1 The gene maps of *Aconitum angustius* and *A. finetianum* (A) as well as *A. sinomontanum* (B). The genes lying inside and outside the circles are transcribed in the clockwise and counterclockwise directions, respectively. Different colors denote the genes belonging to different functional groups. The thickness indicate the extent of the inverted repeats (IRa and IRb) that separate the small single-copy (SSC) region from the large single-copy (LSC) region. The dark gray in the inner circle corresponds to GC content, and the light gray to AT content.

PeerJ

Manuscript to be reviewed





- photosystem I
- photosystem II
- cytochrome b/f complex
- ATP synthase
- NADH dehydrogenase
- RubisCO large subunit
- RNA polymerase
- ribosomal proteins (SSU)
- ribosomal proteins (LSU)
- clpP, matK
- other genes
- hypothetical chloroplast reading frames (ycf)
- transfer RNAs
- ribosomal RNAs

Figure 2(on next page)

Comparison of the border positions of LSC, SSC and IRs repeat regions among thirteen species and two varieties in *Aconitum*

Figure 2 Comparison of the border positions of LSC, SSC and IRs repeat regions among thirteen species and two varieties in *Aconitum*. Genes are denoted by grey boxes and the gaps between the genes and the boundaries are indicated by the base lengths (bp). Extensions of the genes are also indicated above the boxes.

PeerJ
Manuscript to be reviewed

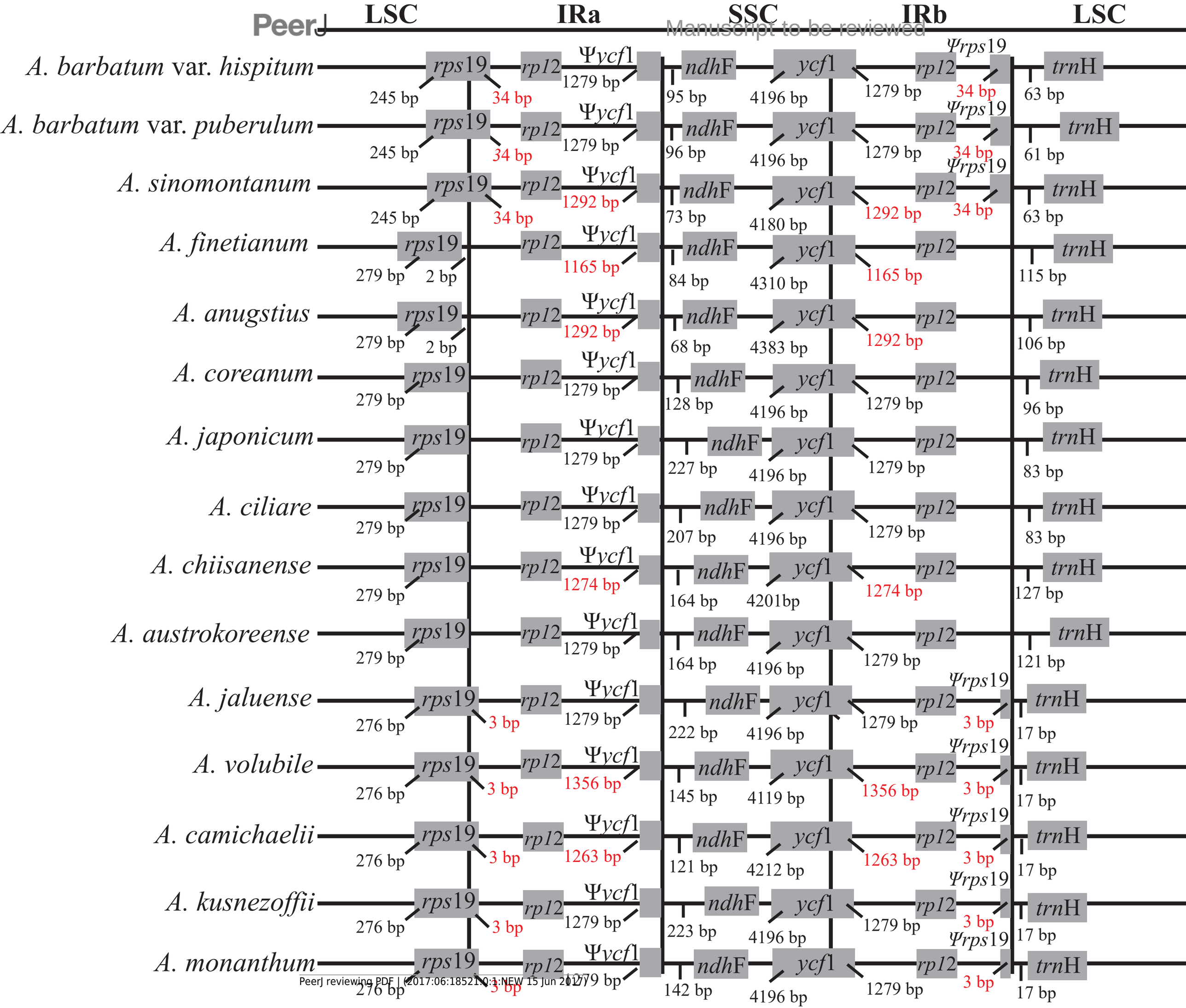


Figure 3(on next page)

Sliding window analysis of the whole cp genome for thirteen species and two varieties in *Aconitum*.

Figure 3 Sliding window analysis of the whole cp genome for thirteen species and two varieties in *Aconitum*. *X*-axis: position of the midpoint of a window; *Y*-axis: nucleotide diversity (*Pi*) of each window.

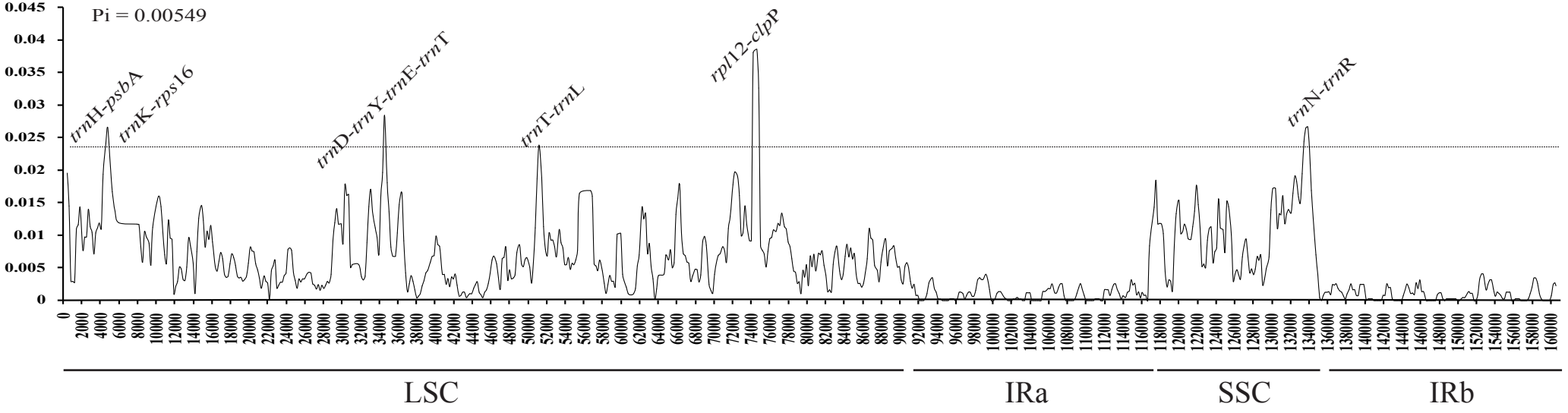

Pi = 0.00549
trnH-psbA
trnK-rps16
trnD-trnY-trnE-trnT
trnT-trnL
rpl12-clpP
trnN-trnR
LSC
IRa
SSC
IRb

Figure 4(on next page)

Phylogenetic relationship among *Aconitum* species.

Figure 4 Phylogenetic relationship among *Aconitum* species. Three methods of Bayesian Inference (BI), Maximum Parsimony (MP) and Maximum Likelihood (ML) have been employed based on complete cp genome sequences and PCGs, respectively. Numbers above the lines represent BI posterior probability, MP and ML bootstrap values.

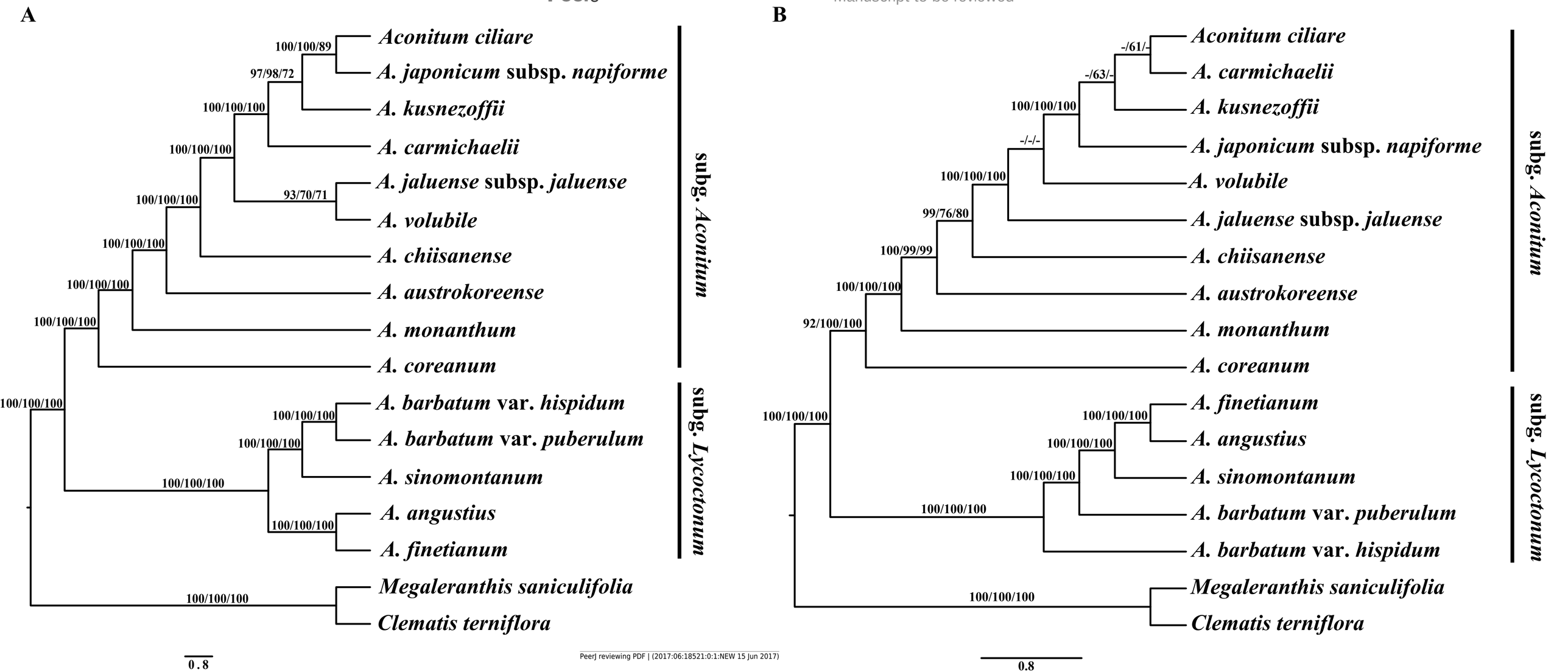

A
Aconitum ciliare
A. japonicum subsp. napiforme
A. kusnezoffii
A. carmichaelii
A. jaluense subsp. jaluense
A. volubile
A. chiisanense
A. austrokoreense
A. monanthum
A. coreanum
A. barbatum var. hispidum
A. barbatum var. puberulum
A. sinomontanum
A. angustius
A. finetianum
Megaleranthis saniculifolia
Clematis terniflora
100/100/89
97/98/72
100/100/100
100/100/100
93/70/71
100/100/100
100/100/100
100/100/100
100/100/100
100/100/100
100/100/100
100/100/100
100/100/100
100/100/100
subg. Aconitum
subg. Lycoctonum
0.8
B
Aconitum ciliare
A. carmichaelii
A. kusnezoffii
A. japonicum subsp. napiforme
A. volubile
A. jaluense subsp. jaluense
A. chiisanense
A. austrokoreense
A. monanthum
A. coreanum
A. finetianum
A. angustius
A. sinomontanum
A. barbatum var. puberulum
A. barbatum var. hispidum
Megaleranthis saniculifolia
Clematis terniflora
-/61/-
-/63/-
100/100/100
-/-/-
100/100/100
99/76/80
100/99/99
100/100/100
92/100/100
100/100/100
100/100/100
100/100/100
100/100/100
100/100/100
subg. Aconitum
subg. Lycoctonum
0.8

Table 1(on next page)

Summary of characteristics in chloroplast genome sequences of thirteen species and two varieties in Aconitum.

1 **Table 1 Summary of characteristics in chloroplast genome sequences of thirteen species and two varieties in *Aconitum*.**

	GenBank No.	Total genome size (bp)	LSC (bp)	SSC (bp)	IR (bp)	Total number of genes	Protein-coding genes	tRNA genes	rRNA genes	GC content
Aconitum* subg. *Lycoctonum										
A. angustius	MF155664	156,109	86,719	16,914	26,225	126	84	38	4	38%
A. finetianum	MF155665	155,625	86,664	17,107	25,927	126	84	38	4	38%
A. sinomontanum	MF155666	157,215	88,074	16,926	26,090	126	84	38	4	38%
A. barbatum var. puberulum	KC844054	156,749	87,630	16,985	26,067	127	85	38	4	38%
A. barbatum var. hispidum	KT820664	156,782	87,661	16,987	26,067	127	85	38	4	38%
Aconitum* subg. *Aconitum										
A. austrokoreense	NC_031410	155,682	86,388	17,054	26,120	126	83	39	4	38.1%
A. carmichaelii	NC_030761	155,737	86,330	17,021	26,193	124	83	37	4	38.1%
A. chiisanense	NC_029829	155,934	86,559	17,085	26,145	125	82	39	4	38.1%
A. ciliare	NC_031420	155,832	86,452	17,084	26,148	126	83	39	4	38.1%
A. coreanum	NC_031421	157,029	87,622	17035	26,186	128	86	38	4	38.0%
A. jaluense	KT820669	155,926	86,406	17,090	26,215	126	83	39	4	38.1%
A. japonicum	KT820670	155,878	86,480	17,104	26,147	127	84	39	4	38.1%
A. kusnezoffii	NC_031422	155,862	86,335	17,103	26,212	126	84	39	4	38.1%
A. monanthum	NC_031423	155,688	86,292	16,996	26,200	125	82	39	4	38.1%
A. volubile	KU556690	155,872	86,348	16,944	26,290	126	83	38	4	38.1%

2

Table 2(on next page)

Gene contained in the sequenced chloroplast genomes of three Aconitum species

1 **Table 2 Gene contained in the sequenced chloroplast genomes of three *Aconitum* species.**

Category	Gene group	Gene name				
Self-replication	Ribosomal RNA genes	*rrn16*[a]	*rrn23*[a]	*rrn4.5*[a]	*rrn5*[a]	
	Transfer RNA genes	*trnA-UGC*[a,,b]	*trnC-GCA*	*trnD-GUC*	*trnE-UUC*	*trnF-GAA*
		trnfM-CAU	*trnG-GCC*[b]	*trnH-GUG*	*trnI-CAU*[a]	*trnI-GAU*[a,b]
		trnK-UUU[b]	*trnL-CAA*[a]	*trnL-UAA*[b]	*trnL-UAG*	*trnM-CAU*
		trnN-GUU[a]	*trnP-UGG*	*trnQ-UUG*	*trnR-ACG*[a]	*trnR-UCU*
		trnS-GCU	*trnS-GGA*	*trnS-UGA*	*trnT-GGU*	*trnT-UGU*
		trnV-GAC[a]	*trnW-CCA*	*trnY-GUA*	*trnG-UCC*	*trnP-GGG*
		trnV-UAC				
	Small subunit of ribosome	*rps11*	*rps12*[c,d]	*rps14*	*rps15*	*rps18*
		rps19	*rps2*	*rps3*	*rps4*	*rps7*[a]
		rps8	*rps16*			
	Large subunit of ribosome	*rpl14*	*rpl16*[b]	*rpl2*[b]	*rpl20*	*rpl22*
		rpl23[a]	*rpl33*	*rpl36*		
	DNA-dependent RNA polymerase	*rpoA*	*rpoB*	*rpoC1*[b]	*rpoC2*	

Genes for photosynthesis	Subunits of photosystem I	*psaA*	*psaB*	*psaC*	*psaI*	*psaJ*
		ycf1	*ycf3*[c]	*ycf4*	*ycf15*	
	Subunits of photosystem II	*psbA*	*psbB*	*psbC*	*psbD*	*psbE*
		psbF	*psbH*	*psbI*	*psbJ*	*psbK*
		psbL	*psbM*	*psbN*	*psbT*	*psbZ*
	Subunits of cytochrome	*petA*	*petB*[b]	*petD*[b]	*petG*	*petL*
		petN				
	Subunits of ATP synthase	*atpA*	*atpB*	*atpE*	*atpF*[b]	*atpH*
		atpI				
	Large subunit of Rubisco	*rbcL*				
	Subunits of NADH dehydrogenase	*ndhA*[b]	*ndhB*[a,b]	*ndhC*	*ndhD*	*ndhE*
		ndhF	*ndhG*	*ndhH*	*ndhI*	*ndhJ*
		ndhK				
Other genes	Maturase	*matK*				
	Envelope membrane protein	*cemA*				
	Subunit of acetyl-CoA	*accD*				

	C-type cytochrome synthesis gene	*ccsA*
	Protease	*clpP*[c]
	Function unknown	*ycf2*[a]
	Translation initial factor	*infA*

a: Two gene copies in IRs; b: gene containing an interval; c: gene containing two intervals; d: gene divided into two independent transcription units.

Table 3(on next page)

The distribution of the five pseudogenes in Aconitum

Table 3 The distribution of the five pseudogenes in *Aconitum*.

Locations	LSC		LSC/IRa	IRa	IRa/SSC
Genes	*Ψrps*16	*Ψinf*A	*Ψrps*19	*Ψycf*15	*Ψycf*1
Aconitum* subg. *Lycoctonum					
A. angustius	+				+/1292bp
A. finetianum	+				+/1165bp
A. sinomontanum		+	+/34bp		+/1292bp
A. barbatum var. *puberulum*		+	+/34bp		+
A. barbatum var. *hispidum*		+	+/34bp		+
Aconitum* subg. *Aconitum					+
A. austrokoreense	+			+/4bp indel	+
A. carmichaelii	+		+/3bp		+/1263bp
A. chiisanense	+			+/4bp indel	+/1274bp
A. ciliare	+				+
A. coreanum					+
A. jaluense	+		+/3bp		+
A. japonicum	+				+
A. kusnezoffii	+		+/3bp		+
A. monanthum	+		+/3bp	+/1bp indel	+
A. volubile	+		+/3bp		+/1356bp

+: indicating the presence of pseudogenes

Table 4(on next page)

Number of chloroplast SSRs in different regions or different types present in Aconitum species

1 **Table 4 Number of chloroplast SSRs in different regions or different types present in *Aconitum* species.**

Species	Homo (>10)	Di (>5)	Tri (>5)	Te (>3)	Pen (>3)	Number of SSRs in different regions			
						LSC	SSC	IR	Total
Aconitum* subg. *Lycoctonum	31 (62.00%)	9 (18.00%)	3 (6.00%)	6 (12.00%)	0 (0.00%)	46 (92.00%)	6 (12.00%)	1 (2.00%)	50
A. angustius	32	7	2	8	1	43	6	1	50
A. finetianum	31	8	2	8	1	43	6	1	50
A. sinomontanum	30	10	2	8	0	50	6	1	57
A. barbatum var. *puberulum*	33	10	2	8	0	47	5	1	53
A. barbatum var. *hispidum*	31	10	7	7	0	49	5	1	55
Aconitum* subg. *Aconitum	22 (46.81%)	15 (31.91%)	1 (21.28%)	7 (14.89%)	1 (21.28%)	38 (80.85%)	7 (14.89%)	2 (4.36%)	47
A. austrokoreense	22	15	0	7	0	32	10	2	44
A. carmichaelii	21	16	1	7	0	37	6	2	45
A. chiisanense	21	16	1	7	2	39	6	2	47
A. ciliare	23	16	1	7	1	41	5	2	48
A. coreanum	39	14	1	7	1	50	10	2	62
A. jaluense	17	14	1	6	2	33	6	2	41
A. japonicum	20	16	1	7	1	37	6	2	46
A. volubile	17	15	1	6	1	35	3	2	40
A. kusnezoffii	19	16	1	7	1	37	5	2	44
A. monanthum	18	13	0	7	2	36	9	2	47

2