Characterization and phylogenetic analysis of the complete chloroplast genome sequence of *Pertya multiflora* (Asteraceae), a rare and endangered wild plant species in China

Ye Liu, Huijuan Zhang, Qian Wu, Wenxin Huang, Tao Wu, Ming Guan and Ming Jiang

College of Life Sciences, Taizhou University, Jiaojiang, China

**ABSTRACT**

*Pertya multiflora* (Asteraceae) is a rare wild plant species narrowly distributed in Zhejiang province, China. In our present study, we assembled its complete chloroplast genome using high-throughput sequencing data. The results indicated that the whole chloroplast genome of *P. multiflora* was 153,396 bp in length. Its large single copy, small single copy, and inverted region sequences were 84,575 bp, 18,451 bp, and 25,185 bp. The *P. multiflora* chloroplast genome was composed of 134 genes, including 87 protein-coding genes, 37 tRNA genes, eight rRNA genes, and two pseudogenes. Phylogenetic analysis results showed that *P. multiflora* was grouped with *Gerbera jamesonii*, with 100% bootstrap support.

*Pertya* is a small genus in the family Asteraceae, which consists of approximately 25 plant species. *Pertya* plants are divided into four series, including *Paniculatae*, *Phylicoides*, *Scandentes*, and *Sinenses*. *Pertya* plants are distributed in Asia countries like China, Japan, Thailand, and Afghanistan, but most of them are found in China (Wu et al. 2011). Plants in the genus *Pertya* are shrubs, subshrubs, or perennial herbs, rarely scandent shrubs (Murata and Ohi-Toma 2016). In recent years, the studies on *Pertya* have been focused on phytochemistry, mating system, and taxonomy (Nagai et al. 1975; Ohtsuka et al. 2005; Zhang et al. 2013). Nagai et al. (1975) isolated a new triterpene methyl ether designated as O-methyl pertyl from the roots of *P. robusta*, and it was identified as the first C33 triterpene methyl ether. Nagumo et al. (1982) isolated a new sesquiterpene dilactone, namely pertilide, from *P. multiflora*.

*Pertya multiflora* was collected from Shuimokeng valley (28°52.043N, 121°07.522E), Linhai county, Zhejiang province, China. The leaves were rinsed with distilled water and kept in a −80°C refrigerator for DNA extraction. A specimen was deposited at the Molecular Biology Laboratory at Taizhou University (Ming Jiang, jiangming1973@139.com) under the voucher number of CHS2016003. Genomic DNA was isolated following the cetyl trimethyl ammonium bromide (CTAB) protocol of Doyle and Doyle (1987). A DNA library was constructed following the standard protocol supplied by Illumina, and it was then sequenced on the Hiseq X Ten sequencing platform at 150 bp paired-end (PE). Approximately 2.98 Gb raw reads were generated, and they were then filtered by NGS QC Toolkit v2.3.3 to yield 2.96 Gb high-quality clean reads (Patel and Jain 2012). NOVOPlasty, a seed-extend based *de novo* assembler, was used to assemble the chloroplast genome (Dierckxsens et al. 2017). An online program DOGMA (Dual Organellar GenoMe Annotator, http://dogma.ccbb.utexas.edu/) was applied to annotate the chloroplast genome (Wyman et al. 2004).

Fresh leaves of *P. multiflora* were collected from Shuimokeng valley (28°52.043N, 121°07.522E), Linhai county, Zhejiang province, China. The leaves were rinsed with distilled water and kept in a −80°C refrigerator for DNA extraction. A specimen was deposited at the Molecular Biology Laboratory at Taizhou University (Ming Jiang, jiangming1973@139.com) under the voucher number of CHS2016003. Genomic DNA was isolated following the cetyl trimethyl ammonium bromide (CTAB) protocol of Doyle and Doyle (1987). A DNA library was constructed following the standard protocol supplied by Illumina, and it was then sequenced on the Hiseq X Ten sequencing platform at 150 bp paired-end (PE). Approximately 2.98 Gb raw reads were generated, and they were then filtered by NGS QC Toolkit v2.3.3 to yield 2.96 Gb high-quality clean reads (Patel and Jain 2012). NOVOPlasty, a seed-extend based *de novo* assembler, was used to assemble the chloroplast genome (Dierckxsens et al. 2017). An online program DOGMA (Dual Organellar GenoMe Annotator, http://dogma.ccbb.utexas.edu/) was applied to annotate the chloroplast genome (Wyman et al. 2004).

Overall, we yielded 9,864,304 bp PE clean reads after trimming poor quality bases and filtering low-quality reads, and they were used to produce a circular plastid by NOVOPlasty. The complete chloroplast genome sequence of *P. multiflora* is 153,396 bp in length with an LSC region (84,575 bp), an SSC region (18,451 bp), and two copies of IRs (25,185 bp). Annotation results demonstrated that the *P. multiflora* chloroplast genome harbored 134 genes, including 87 protein-coding genes, 37 tRNA genes, eight rRNA genes, and two pseudogenes. Among these genes, six protein-coding
genes (ycf2, ndhB, ycf15, rps7, rpl23, and rpl2), eight tRNA genes (trnG-UCC, trnN-GUU, trnR-ACG, trnA-UCC, trnL-GAU, trnV-GAC, trnL-CAA, trnL-CAU), and four rRNA genes (23S rRNA, 5S rRNA, 4.5S rRNA, and 16S rRNA) contain two copies. The 3' ends of ycf1 and rps19, locating at 3' and 5' ends of IRa respectively, were found to be truncated and thus formed two pseudogenes. We have submitted the annotated chloroplast genomic sequence to GenBank under an accession number of MW148616.

The complete chloroplast genome sequences of five Ligularia species (MF539932, MN783367, NC_039383, NC_039382, and KT988070), five Cynara (KP299292, KP842705, KP842706, KP842707, KP842715), five Artemisia (MG951482, MG951484, MG951490, MG951492, MG951496), four Saussurea (NC_044727, NC_044732, NC_044734, NC_044739), four Echinacea (KX548219, KX548222, KX548224, KX548223), six Atractylodes (NC_044671, NC_037484, NC_037483 MW301113, MT834523, MT834524), and three other Asteraceae plants together with P. multiflora and the outgroup Adenophora stricta (NC_036223) were aligned to construct a phylogenetic tree. The sequences were aligned using MAFFT v7.388 (Katoh and Standley 2013), and the best-fitting nucleotide substitution model for maximum likelihood (ML) analysis was selected by running jModelTest (Darriba et al. 2012). GTR + G + I was turned out to be the best DNA substitution model. A maximum likelihood (ML) tree was then generated using PhyML 3.1 (Guindon et al. 2010). Phylogenetic analysis results were well matched to taxonomic classification, and plant species within the same genus clustered on the same clade. Our results indicated that P.

**Figure 1.** A maximum likelihood tree based on the complete chloroplast genome sequences of *Pertya multiflora* (Asteraceae) and related species, with *Adenophora stricta* as the outgroup. The numbers next to nodes indicate bootstrap support values.
multiflora grouped with seven plants from genera G jamesonii, with 100% bootstrap support (Figure 1).

Characterization and phylogenetic analysis of the plastid sequence of P. multiflora will provide insights into genetic diversity studies and conservation of this rare plant species in the future.

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

Funding
This work was supported by National Undergraduate Training Program for Innovation and Entrepreneurship [202010390030].

Data availability statement
The data that support the findings of this study are openly available in GenBank of NCBI at https://www.ncbi.nlm.nih.gov/nuccore/MW148616. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA688619, SRR13329723, and SAMN17185260, respectively.

References
Darriba D, Taboada GL, Doallo R, Posada D. 2012. JModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 9(8): 772–772.

Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45:e18.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 19:11–15.

Funk VA, Bayer RJ, Keeley S, Chan R, Watson L, Gemeinholzer B, Schilling E, Panero JL, Baldwin BG, Garcia-Jacas N, et al. 2005. Everywhere but Antarctica: using a supertree to understand the diversity and distribution of the Compositae. Biol Skr. 55:343–374.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59(3):307–321.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.

Murata J, Ohi-Toma T. 2016. Correct name of the Japanese endemic species Pertya triflora (Asteraceae). J Jap Bot. 91:79–83.

Nagai M, Nagumo S, Izawa K. 1975. O-methyl pertyol, a C33 triterpene methyl ether. Tetrahedron Lett. 16(42):3655–3656.

Nagumo S, Nagai M, Inoue T. 1982. New sesquiterpene dilactone from Pertya glabrescens. Chem Pharm Bull. 30(2):586–592.

Ohtsuka A, Watanabe M, Yahara T. 2005. Inbreeding coefficients in six species of Ainsliaea and two species of Pertya (Asteraceae). Plant Syst Evol. 251(2–4):143–151.

Patel RK, Jain M. 2012. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS One. 7(2):e30619.

Wu ZY, Raven PH, Hong DY. 2011. Flora of China (Vol. 20-21). Beijing and St. Louis: Science Press and Missouri Botanical Garden Press.

Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organelar genomes with DOGMA. Bioinformatics. 20(17):3252–3255.

Zhang CF, Li LQ, Gao TG. 2013. Pertya multiflora sp. nov. (Asteraceae, Pertyeae) from Zhejiang, China. Nord J Bot. 31(5):626–630.