Complete chloroplast genome and phylogenetic analysis of a wild grass, *Hordeum roshevitzii* Bowden

Ruochen Zhang**, Xiuzhang Li**, Zhenjiang Chen¹, Xiang Yao², Tao Li¹ and Chunjie Li¹

¹State Key Laboratory of Grassland Agro-Ecosystems, College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, PR China; ²Qinghai Academy of Animal and Veterinary Science, Qinghai University, Xining, PR China; ³Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, Nanjing Botanical Garden Mem. Sun Yat-Sen, Nanjing, PR China

**These authors contributed equally to this work.

**CONTACT Chunjie Li chunjie@lzu.edu.cn College of Pastoral Agriculture Science and Technology, Lanzhou University; Lanzhou 730000, PR China

ABSTRACT

The complete circular chloroplast (cp) genome of wild *Hordeum roshevitzii* Bowden was sequenced and reported in this study. The complete chloroplast genome of wild *H. roshevitzii* was 12,753 bp in length, including a pair of inverted repeat regions (IRA/IRB) of 21,587 bp separated by one small single-copy (SSC) region of 12,753 bp and one large single-copy (LSC) region of 81,138 bp. A total of 133 genes, including 85 protein-coding genes (PCGs), 40 transfer RNA genes, and 8 ribosomal RNA genes were predicted from the cp genomes. The overall GC content of the cp genome was 38.25% and the corresponding values of the IR, SSC, and LSC were 43.89%, 32.19%, and 36.21%. The phylogenetic analysis of *H. roshevitzii* determined that *H. roshevitzii* was clustered closely with *Hordeum bogdanii*.

Wild *Hordeum roshevitzii* Bowden is an ancient and mysterious plant of the genus *Hordeum* in family *Gramineae*, with the common name tall wheatsgrass. *H. roshevitzii* originated with a history of 20,000 years in the Nile River Basin. It is a perennial plant, stems clump erect and smooth, generally 30–70 cm high, with three or four nodes. The base of leaf sheath is longer than the upper part, the length of leaf blade is 3–14 cm, often flat. The spikes are green or purplish in color, generally 4–7 cm in length, and the palea is equal to the lemma. Wild *H. roshevitzii* is mainly distributed in Inner Mongolia, northern Shanxi, Gansu, Qinghai, and Xinjiang in China. It is usual found in grassland with sandy soil along rivers. It is an important forage grass and also a valuable gene pool for cereal crops (Wang et al. 1991). However, *H. roshevitzii* is particularly rare and precious because it is extremely difficult to cultivate. *H. roshevitzii* has attracted widespread attention from medical experts and is even seen as a way to prevent cancer due to this grain’s purple chaff, which contains natural anthocyanin glycosides and used as antioxidants. Knowledge of genome constitution of *H. roshevitzii* was extremely valuable for forage breeding and pharmaceutical development.

Chloroplast genome as independent units of heredity, directly reflects the long-term accumulation of genetic variation of this plant in the process of evolution, and traces the origin of species and migration (Su et al. 2020). Chloroplast genome is widely used in plant phylogeny because chloroplast genome reflects relative interspecific genetic relationship and genetic diversity of plant. However, there is no research on genome of the wild *H. roshevitzii*, so obtaining the complete genome of wild *H. roshevitzii* cp would lay the foundation for the genetic research of this mysterious plant. In our study, we reported and annotated the complete cp genome sequence of wild *H. roshevitzii*, and submitted to the GenBank with the accession number MW266081.

Fresh sample was obtained from an alpine grazing meadow in Sunan Yugur Autonomous County (Gansu Province, N37°54′24″, E101°48′30″). The plant sample has been deposited in the College of Pastoral Agricultural Science and Technology, Lanzhou University, Lanzhou, China. Genomic DNA extraction and sequencing library construction were conducted by Benagen Technology Services Limited (Wuhan, China), and a 350 bp library of DNA samples was constructed by whole-genome shotgun method, the whole *H. roshevitzii* chloroplast genome was sequenced with 150 bp double-terminal reads on Illumina NovaSeq (University of Edinburgh 2017) sequencer. A total of 35,359,246 raw reads were obtained, they were filtered and trimmed using the SOAPnuke version 2.1.0 (Chen et al. 2018) to get the clean reads. Then 34,879,360 clean reads were spliced by SPAdes version 3.13.1 (Nicolas et al. 2017), the splicing results and reference genome were compared using the BLAST version 2.2.30+ (Kent 2002) tool with the cp genome of *H. bogdanii* (GenBank Accession No. NC_043839.1) as a reference. The cp genome sequence of wild *H. roshevitzii* was annotated by software GeSeq (Michael et al. 2017) with manual correction.

CONTACT Chunjie Li chunjie@lzu.edu.cn College of Pastoral Agriculture Science and Technology, Lanzhou University; Lanzhou 730000, PR China

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
The complete size of wild *H. roshevitzii* cp genome was 137,065 bp, including a large single-copy region (LSC) with a length of 81,138 bp, and a small single-copy region (SSC) with a length of 12,753 bp, and a pair of inverted repeat regions (IRA/IRB), both of which were 21,587 bp. A total of 133 functional genes were contained in wild *H. roshevitzii* cp genome, including 85 protein-coding genes (PCGs) (79 PCG species), 40 transfer RNA genes (32 tRNA species), and 8 ribosomal RNA genes (4 rRNA species). Among them, trnK-UUU, rps16, atpF, trnG-UCC, trnV-UAC, trnL-UAA rpl2, ndhA, ndhB, trnI-GAU, and trnA-UGC have a single intron, however, two introns are contained in ycf3 and rps12. Most genes appeared in a single copy, whereas nineteen genes occurred as two copies, including seven PCG genes (*rps19, rpl2, rpl23, rps12, rps7, ndhB, and rps15*), eight tRNA genes (*trnH-GUG, trnL-CAU, trnL-GAU, trnL-CAA, trnV-GAC, trnA-UGC, trnR-ACG, and trnN-GUU*), and four rRNA genes (*rrn4.5, rrn5, rrn16, and rrn23*). The content of CG in whole chloroplast genome was 38.25%, including 36.21% in LSC region, 32.19% in SSC region and 43.89% in IR region.

Phylogenetic analysis was determined with *H. roshevitzii* cp PCGs and the cp PCGs of 24 species of *Gramineae* plants based on the maximum likelihood method, PCGs of 25 *Gramineae* plants were screened out from the complete cp genome, the sequences of PCGs were firstly aligned using MAFFT program (Katoh et al. 2002), and a maximum likelihood tree (Figure 1) was constructed by program RAxML version 8.2.10 (Stamatakis 2014).

The phylogenetic results showed that 25 species were clustered into two branches, and wild *H. roshevitzii* was closest to *Hordeum bogdanii* with strong bootstrap values (Figure 1). Complete cp genome of wild *H. roshevitzii* provides a

---

**Figure 1.** Maximum likelihood tree based on the chloroplast PCGs sequences of *H. roshevitzii* and other taxa within the *Gramineae*. The numbers shown next to the nodes are bootstrap support values.
theoretical basis for further elucidating the phylogenetic relationships of genus *Hordeum*.

**Disclosure statement**

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

**ORCID**

Chunjie Li [http://orcid.org/0000-0002-3287-2140](http://orcid.org/0000-0002-3287-2140)

**Data availability statement**

The genome sequence data that support the findings of this study are openly available in GenBank of NCBI at [https://www.ncbi.nlm.nih.gov/](https://www.ncbi.nlm.nih.gov/) under the accession no. MW 266081. The associated Bio-Project, SRA, and Bio-Sample numbers are PRJNA689374, SRR13355669, and SAMN17207691, respectively.

**References**

Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, Li Z, et al. 2018. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience. 7(1):1–6.

Katoh K, Misawa K, Kuma K-i, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30(14):3059–3066.

Kent WJ. 2002. Blat: the BLAST-like alignment tool. Genome Res. 12(4): 656–664.

Michael T, Pascal L, Tommaso P, et al. 2017. GeSeq – versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 45(W1): W6–W11.

Nicolas D, Patrick M, Guillaume S. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45(4):e18.

Stamatakis A. 2014. RAxML Version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30(9): 1312–1313.

Su X, Zhao J, Wang Z. 2020. The complete chloroplast genome of *Hordeum brevisubulatum*. Mitochondrial DNA B Resour. 5(3): 2988–3007.

University of Edinburgh. 2017. Edinburgh genomics to deploy Illumina’s NovaSeq platform. Edinburgh: University of Edinburgh.

Wang RC, Marburger JE, Hu CJ. 1991. Tissue-culture-facilitated production of aneupolyhaploid *Thinopyrum ponticum* and amphidiploid *Hordeum violaceum* x *H. bogdani* and their uses in phylogenetic studies. Theor Appl Genet. 81(2):151–156.