

The complete chloroplast genome of weedy rye *Secale cereale* subsp. *segetale*

Tianyu Du^{a,b}, Yiyu Hu^b, Yanqing Sun^b, Chuyu Ye^b and Enhui Shen^{a,b,c}

^aShandong (Linyi) Institute of Modern Agriculture of Zhejiang University, Linyi, China; ^bInstitute of Crop Sciences, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou, China; ^cThe Rural Development Academy, Zhejiang University, Hangzhou, China

ABSTRACT

Weedy rye (*Secale cereale* subsp. *segetale* Zhukov 1928) is a problematic weed species in wheat field. However, it can potentially provide valuable genetics resources to increase the genetic variations and introduce desirable genes for rye and wheat breeding. Here, we assembled the complete chloroplast genome of *S. cereale* subsp. *segetale*. The chloroplast genome is 137,051 bp in length, containing a large single copy region (81,090 bp), a small single copy region (12,795 bp) and two separated inverted repeat regions (21,583 bp). A total of 131 unique genes were annotated, consisting of 82 protein-coding genes, 41 tRNA genes, and 8 rRNA genes. The phylogenetic analysis showed that *Secale cereale* subsp. *segetale* (weedy rye) and *S. cereale* subsp. *cereale* (rye) clustered together as sisters to other *Secale* species.

ARTICLE HISTORY

Received 10 November 2021
Accepted 12 May 2022

KEYWORDS

Chloroplast genome; weedy rye; *Secale cereale* subsp. *segetale*; phylogenetic analysis

Weedy rye occurs as a weed in cereal fields, mainly in the Near East and Central Asia, and is fully interfertile with cultivated rye (Zohary et al. 2012). Although weedy rye is generally considered to be a malignant weed, as tertiary gene pool of *Triticum aestivum* L. (Santos et al. 2016), it can potentially provide valuable genetic resources (Sun et al. 2022), such as resistance to insects and disease (rust, mildew, aphids, etc.), high yield and resistance to abiotic stress (Che et al. 2008; Hagenblad et al. 2016). In view of climate change and new biotic and abiotic stresses, there is also a need to investigate wild species of rye, which is critical to improve yields and quality of that cereal (Feuillet et al. 2008). Chloroplast genome sequences are useful for understanding plant origin and evolution. However, there is one chloroplast genome in *Secale* currently.

In this study, we assembled the complete chloroplast genome of *S. cereale* subsp. *segetale*. The seed of *S. cereale* subsp. *segetale*, collected in California (38.5475, -121.7393), United States, was acquired from Germplasm Resources Information Network (GRIN) (accession number: CISE 102) and cultured in field (Hangzhou, Zhejiang Province China). The plant was deposited at Herbarium of Zhejiang University (YuPing Ma, 3160105887@zju.edu.cn) under the voucher number: HZU60244001. Total genomic DNA was sequenced by DNBSEQ-T7 platform. Approximately 80.4 Gbp of clean data was obtained in this study. NGSQCToolkit v2.3 (Patel and Jain 2012) was used for quality control. The clean data was applied in *de novo* assembly by NOVOPlasty v3.6 (Dierckxsens et al. 2017) using the complete chloroplast genome of *Triticum aestivum* (GenBank accession number NC_002762) as a reference. GeSeq online (Tillich et al. 2017)

was used for genome annotation. The assembled genome sequences and annotation information have been deposited in GenBank under the accession number LC645358.1.

The total length of *S. cereale* subsp. *segetale* chloroplast genome is 137,051 bp. The genome exhibited a distinct quadripartite structure containing a pair of inverted repeats (IRa and IRb, 21,583 bp each), a large single-copy region (LSC, 81,090 bp) and a small single-copy region (SSC, 12,795 bp). The GC contents of the IR, LSC, and SSC regions are 43.86, 36.22, and 32.17%, respectively. A total of 131 unique genes were annotated, including 82 protein-coding genes, 41 tRNA genes, and 8 rRNA genes.

To understand the phylogenetic relationship between *Secale cereale* subsp. *segetale* and other *Triticeae* species, we built a phylogenetic tree of nine *Triticeae* species based on complete chloroplast genome sequences (NC_024764.1 *Triticum timopheevii*, NC_046698.1 *Triticum zhukovskiyi*, NC_024831.1 *Aegilops comosa*, NC_021761.1 *S. cereale*, NC_024831.1 *Aegilops bicornis*, LC_645210.1 *S. strictum* subsp. *kuprijanovii*, LC_649171 *S. sylvestre*, NC_056985 *Hordeum vulgare*) downloaded from NCBI GenBank database. We first performed multiple sequence alignments using MAFFT v7.310 (Katoh et al. 2002) with the parameter '-auto -reorder -phyloipout'. Then a maximum-likelihood tree was constructed using IQ-tree v1.6.12 (Nguyen et al. 2015) with recommended model TVM+I and 1000 bootstrap values. The tree was illustrated and modified by iTOL (Letunic and Bork 2019). The phylogenetic tree showed that *S. cereale* subsp. *segetale* was first clustered with *S. cereale* forming as a monophyletic group (Figure 1). *Secale sylvestre* and *S. strictum* subsp. *kuprijanovii* are the wild types of *S. cereale*, and the phylogenetic

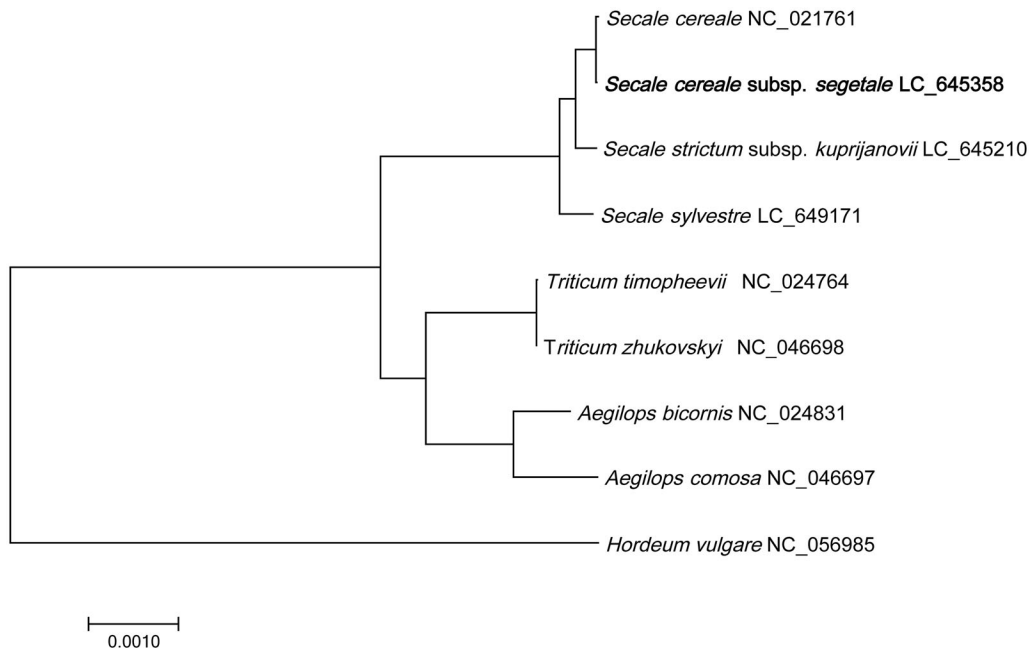


Figure 1. Maximum-likelihood (ML) tree based on 9 *Triticeae* species, using *Hordeum vulgare* as an outgroup. The numbers on the node are the fast bootstrap value based on 1,000 replications.

relationship showed that these wild types may be the progenitor of *S. cereale* subsp. *segetale*. The tree also supported the schematic phylogeny of genus *Secale* raised by Schreiber et al. (2019). The complete chloroplast genome sequence of *S. cereale* subsp. *segetale* will provide valuable information for genetic studies of *Secale* species.

Ethical approval

The collection of plant material (weedy rye) for this study was carried out in strict accordance with the guidelines provided by the herbarium of Zhejiang University, and the research process for this plant material was carried out in accordance with the guidelines provided by the Institute of Crop Science of Zhejiang University and Chinese regulations.

Author contributions

E.S. designed and supervised the study. T.D. and Y.H. analyzed and interpreted the data; Y.H. and Y.S. prepared the plant material; T.D., C.Y. and E.S. wrote and revised the manuscript; All authors agree to be accountable for all aspects of the work.

Disclosure statement

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

Funding

This work was supported by the Fundamental Research Funds for the Central Universities [2021QN81013] to E Shen; the Jiangsu Collaborative Innovation Center for Modern Crop Production; and the Innovation Method Project of China [2018IM0301002].

Data availability statement

The genome sequence data of this study is available in GenBank of NCBI at (<https://www.ncbi.nlm.nih.gov>) under the accession number LC645358.1.

The associated BioProject, SRA, and BioSample numbers are PRJNA749911, SRS9744726 and SAMN20425456 respectively.

References

- Che YH, Yang XM, Yang YP, Zhang SS, Li XQ, Li LH. 2008. Genetic diversity of *Secale cereale* subsp. *segetale* populations in Xinjiang. *J Triticeae Crops*. 28:959–758.
- Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res*. 45(4):e18.
- Feuillet C, Langridge P, Waugh R. 2008. Cereal breeding takes a walk on the wild side. *Trends Genet*. 24(1):24–32.
- Hagenblad J, Oliveira HR, Forsberg NEG, Leino MW. 2016. Geographical distribution of genetic diversity in *Secale* landrace and wild accessions. *BMC Plant Biol*. 16(1):1–20.
- Katoh K, Misawa K, Kuma KI, Miyata T. 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 30(14):3059–3066.
- Letunic I, Bork P. 2019. Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*. 47(W1):W256–59.
- Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 32(1):268–274.
- Patel RK, Jain M. 2012. NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLOS One*. 7(2):e30619.
- Santos E, Matos M, Silva P, Figueiras AM, Benito C, Pinto-Carnide O. 2016. Molecular diversity and genetic relationships in *Secale*. *J Genet*. 95(2):273–281.
- Schreiber M, Himmelbach A, Börner A, Mascher M. 2019. Genetic diversity and relationship between domesticated rye and its wild relatives as revealed through genotyping-by-sequencing. *Evol Appl*. 12(1):66–77.
- Sun YQ, Shen EH, Hu YY, Wu DY, Feng Y, Lao ST, Dong CF, Du TY, Hua W, Ye CY, et al. 2022. Population genomic analysis reveals domestication of cultivated rye from weedy rye. *Molec Plant*. 15(3):552–561.
- Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. 2017. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 45(W1):W6–W11.
- Zohary D, Hopf M, Weiss E. 2012. Domestication of plants in the old world: the origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin. Oxford University Press.