

# The Complete Chloroplast Genome of *Secale Sylvestre* (Poaceae: Triticeae)

Lidia Skuza (✉ [lidia.skuza@usz.edu.pl](mailto:lidia.skuza@usz.edu.pl))

University of Szczecin <https://orcid.org/0000-0002-6362-844X>

Romain Gastineau

University of Szczecin: Uniwersytet Szczecinski

Anna Sielska

University of Szczecin: Uniwersytet Szczecinski

---

## Research note

**Keywords:** Chloroplast genome, rye, *Secale sylvestre*

**Posted Date:** March 19th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-322257/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Journal of Applied Genetics on August 15th, 2021. See the published version at <https://doi.org/10.1007/s13353-021-00656-x>.

# Abstract

## Objective

*Secale sylvestre* is a wild species of rye, morphologically distinct from domestic species. To draw comparisons between species based on molecular features, it is important to have high quality sequences, especially in the case of organellar genomes. For such reason, the complete chloroplast genome of *Secale sylvestre* Host introd. no. 6047 will provide useful data for ecological, agricultural and phylogenetic purposes.

## Results

Here we present the complete, annotated chloroplast genome sequence of *Secale sylvestre* Host introd. no. 6047. The genome is 137116 bp long. The genome can be accessed on GenBank with the accession number (MW557517).

# Introduction

*Secale* is a small but very diverse genus from the tribe *Triticeae* (family *Poaceae*). It includes annual, perennial, self-pollinating and open-pollinating, cultivated, weedy and wild species of various morphologies. The genus *Secale* includes for now four species whose phylogenetic relationships have not been fully determined (GRIN, <http://www.arsgrin.gov>). This causes a significantly reduction of progress in rye breeding that can be enriched with functional traits derived from wild rye species. In the genus, the wild species *Secale sylvestre* Host (1809) is singularized by several genetic peculiarities [1, 2, 3, 4, 5]

Among the 8 chloroplast genome of *Secale* spp. available on GenBank, none is complete strictly speaking, with the second copy of the IR missing all the time, and with the exception of *Secale cereale* KC912691, they all display several ambiguous and non-attributed bases, rendering it difficult to perform accurate SNPs comparisons. Thus, we presume that analysis of the complete chloroplast genome sequences of *Secale* spp., starting with *S. sylvestre*, will be useful and cost-effective for evolutionary and phylogenetic studies, as it was suggested by our previous studies [6, 7].

# Materials And Methods

Seeds of *Secale sylvestre* Host introd. no. 6047 were obtained from the Botanical Garden of the Polish Academy of Sciences in Warsaw. Total DNA was extracted from young sprouts following Doyle and Doyle [8]. Sequencing took place in BGI Shenzhen's facilities on a DNBSEQ platform. An amount of ca. 40 million clean 100 bp paired-end reads was obtained, and assembled using SPAdes 3.14.0 [9] with a k-mer of 85. The contigs corresponding to the chloroplast genome were joined together using Consed [10]. Annotations were performed with the help of GeSeq [11] and manually curated.

# Results And Discussion

The genome is 137116 bp long (Table 1). The LSC is 81132 bp long, the SSC is 12820 bp long and the IR is 21582 bp long. No ambiguous bases were found in the genome.

As stated above, SNP calling type of analysis were rendered difficult by the presence among 7 out of 8 of the other available genomes of numerous non-attributed bases. Instead, analyzes focused on the presence of indels. To do so, chloroplast genomes were partitioned by sub-units, aligned using MAFFT 7 [12] and then visualized using MEGAX [13].

Results provided evidences of the strong proximity between *S. sylvestre* Host introd. no. 6047 and *Secale strictum* voucher R 1108 (KY636137). A total of 16 indels were found to be common between these two strains, that discriminate them from all other (KC912691, KY636135, KY636136, KY636132, KY636134, KY636133, KY636138). The size of these indels ranges from 2 to 36 bp. Among these indels, 13 of were found in intergenic sequences (*rpl32 – trnA-L*; *psaC – ndhE*; *rrn16 – trnI-GAU*; *atpH – atpF*; *psaA – ycf3*; *trnT-UGU – trnL-UAA*; *trnF-GAA – ndhJ*; *atpB – rbcL*; *ycf4 – cemA*; *trnP-UGG – psaJ*; *psaJ – rpl33*; *clpP – psbB*; *rpl16 – rps3*). It is worth being underlined that the last three indels occurred in intronic sequences, one inside a tRNA (intron *trnK-UUU*), two inside protein-coding genes (intron *rps16*; intron *petD*), a feature that received recent attention [14, 15], especially for the purpose of genetic distinction between closely related species [16].

## Limitations

The protocol itself showed no limitation, as it allowed to obtain complete and non-ambiguous genome sequence. However, far more clean genome sequences are needed in order to describe the most reliable molecular markers for species identification and phylogeny, especially for what concerns SNPs.

## Abbreviations

LSC: large single copy; SSC: short single copy; IR: inverted repeat; bp: base pair.

## Declarations

### Acknowledgements

Not applicable.

### Authors' contributions

LS, RG and AS conducted experiments and drafted the manuscript. Bioinformatic analyses were performed by RG.

### Funding

RG thanks the Horizon 2020 Research and Innovation Programme GHANA (The Genus *Haslea*, New marine resources for blue biotechnology and Aquaculture) under Grant Agreement No 734708/GHANA/H2020-MSCA-RISE-2016, and the 2017-2021 research funds granted for the implementation of a co-financed international research project from Polish Ministry of Science and Higher Education.

### **Availability of data materials**

The genome has been deposited on GenBank with the accession number MW557517. It is also available on Zenodo with the following link: <http://doi.org/10.5281/zenodo.4537281>

### **Ethics approval and consent to participate**

Not applicable.

### **Consent for publication**

Not applicable.

### **Competing interests**

The authors declare no competing interest.

## **References**

1. Bennett MD, Smith JB. Nuclear DNA amounts in angiosperms. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences.* 1976;274:227-74.
2. Singh RJ, Röbbelen G. Identification by Giemsa technique of the translocations separating cultivated rye from three wild species of *Secale*. *Chromosoma.* 1977;59:217-25.
3. Shang HY, Wei YM, Wang XR, Zheng YL. Genetic diversity and phylogenetic relationships in the rye genus *Secale* L.(rye) based on *Secale cereale* microsatellite markers. *Genetics and Molecular Biology.* 2006;29:685-91.
4. Zhou J, Yang Z, Li G, Liu C, Tang Z, Zhang Y, Ren Z. Diversified chromosomal distribution of tandemly repeated sequences revealed evolutionary trends in *Secale* (Poaceae). *Plant Systematics and Evolution.* 2010;287:49-56.
5. Skuza L, Szućko I, Filip E, Adamczyk A. DNA barcoding in selected species and subspecies of Rye (*Secale*) using three chloroplast loci (*matK*, *rbcl*, *trnH-psbA*). *Notulae Botanicae Horti Agrobotanici Cluj-Napoca.* 2019;47:54-62.
6. Skuza L, Szućko I, Filip E, Strzała T. Genetic diversity and relationship between cultivated, weedy and wild rye species as revealed by chloroplast and mitochondrial DNA non-coding regions analysis. *PloS one.* 2019;14:e0213023..
7. Doyle JJ, Doyle JL. Isolation of plant DNA from fresh tissue. *Focus.* 1990;12:39-40.

8. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology*. 2012;19:455-77.
9. Gordon D, Green P. Consed: a graphical editor for next-generation sequencing. *Bioinformatics*. 2013;29:2936-7.
10. Tillich M, Lehwarck P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic acids research*. 2017;45:W6-11.
11. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*. 2013;30:772-80.
12. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution*. 2018;35:1547-9.
13. Mo Z, Lou W, Chen Y, Jia X, Zhai M, Guo Z, Xuan J. The chloroplast genome of *Carya illinoensis*: genome structure, adaptive evolution, and phylogenetic analysis. *Forests*. 2020;11:207.
14. Chen S, Ishizuka W, Hara T, Goto S. Complete Chloroplast Genome of Japanese Larch (*Larix kaempferi*): Insights into Intraspecific Variation with an Isolated Northern Limit Population. *Forests*. 2020;11:884.
15. Liu E, Yang C, Liu J, Jin S, Harijati N, Hu Z, Diao Y, Zhao L. Comparative analysis of complete chloroplast genome sequences of four major *Amorphophallus* species. *Scientific reports*. 2019;9:1-4.

## Tables

**Table 1 Overview of the data files/data sets.**

Label	File types (file extension)	Data repository and identifier (DOI or accession number)
<i>Secale sylvestre</i> chloroplast, complete genome	FASTA, annotated GBK	MW557517