



The complete chloroplast genome of *Secale sylvestre* (Poaceae: Triticeae)

Lidia Skuza^{1,2} · Romain Gastineau³ · Anna Sielska^{1,2}

Received: 21 June 2021 / Revised: 19 July 2021 / Accepted: 27 July 2021 / Published online: 15 August 2021
© The Author(s) 2021

Abstract

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. Here we present the complete, annotated chloroplast genome sequence of *Secale sylvestre* Host introd. no. 6047. The genome is 137116 base pair (bp) long. It is the first complete chloroplast genome that can be used as a reference genome for further analysis. The genome can be accessed on GenBank with the accession number (MW557517).

Keywords Chloroplast genome · Rye · *Secale sylvestre* · Sequencing

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 (Bennett and Smith 1976; Singh and Röbbelen 1977; Shang et al. 2006; Zhou et al. 2010; Skuza et al. 2019a).

Among the 8 chloroplast genome of *Secale* spp. available on GenBank, none is complete strictly speaking, with the second copy of the inverted repeat (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 was suggested by our previous studies (Doyle and Doyle 1990; Skuza et al. 2019b).

Material 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 (Bankevich et al. 2012). Sequencing took place in BGI Shenzhen's facilities on a. An amount of ca. 40 million clean 100 bp paired-end reads was obtained and assembled using SPAdes 3.14.0 (Gordon and Green 2013) with a k-mer of 85. The contigs corresponding to the chloroplast genome were joined together using Consed (Tillich et al. 2017). Annotations were performed with the help of GeSeq (Katoh and Standley 2013) and manually curated.

Communicated by: Izabela Pawłowicz

✉ Lidia Skuza
lidia.skuza@usz.edu.pl

¹ Institute of Biology, University of Szczecin, PL-71-415, Szczecin, Poland

² Centre for Molecular Biology and Biotechnology, University of Szczecin, PL-71-415, Szczecin, Poland

³ Institute of Marine and Environmental Sciences, University of Szczecin, PL-70-383 Szczecin, Poland

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

Results and discussion

The genome is 137116 bp long (Table 1). The large single copy (LSC) is 81132 bp long, the short single copy (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 was rendered difficult by the presence among 7 out of 8 of the other available genomes of numerous non-attributed bases. Instead, analyses focused on the presence of indels, which were found to be common and often phylogenetically informative (e.g., Chen et al. 2016). To do so, chloroplast genomes were partitioned by sub-units, aligned using MAFFT 7 (Kumar et al. 2018), and then visualized using MEGAX (Mo et al. 2020).

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 others (KC912691, KY636135, KY636136, KY636132, KY636134, KY636133, KY636138). This number is not very high compared to the results presented by other authors (e.g., Shaw et al. 2007), but considering that only two species were analyzed, the result was very satisfactory. 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 (Liu et al. 2019; Chen et al. 2020), especially for the purpose of genetic distinction between closely related species.

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.

Author contribution LS initiated the project, and 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.

Data availability 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>.

Code of availability Not applicable.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bankevich A, Nurk S, Antipov D et al (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Bennett MD, Smith JB (1976) Nuclear DNA amounts in angiosperms. *Philos Trans R Soc Lond B Biol Sci* 274:227–274
- Chen Z, Feng K, Grover CE et al (2016) Chloroplast DNA structural variation, phylogeny, and age of divergence among diploid cotton species. *PLoS ONE* 11(6):e0157183. <https://doi.org/10.1371/journal.pone.0157183>

- Chen S, Ishizuka W, Hara T et al (2020) Complete Chloroplast Genome of Japanese Larch (*Larix kaempferi*): insights into intraspecific variation with an isolated northern limit population. *Forests* 11:884. <https://doi.org/10.3390/f11080884>
- Doyle JJ, Doyle JL (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:39–40
- Gordon D, Green P (2013) Consed: a graphical editor for next-generation sequencing. *Bioinformatics* 29:2936–2937. <https://doi.org/10.1093/bioinformatics/btt515>
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>
- Kumar S, Stecher G, Li M et al (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35:1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Liu E, Yang C, Liu J et al (2019) Comparative analysis of complete chloroplast genome sequences of four major *Amorphophallus* species. *Sci Rep* 9:1–4. <https://doi.org/10.1038/s41598-018-37456-z>
- Mo Z, Lou W, Chen Y et al (2020) The chloroplast genome of *Carya illinoensis*: genome structure, adaptive evolution, and phylogenetic analysis. *Forests* 11:207. <https://doi.org/10.3390/f11020207>
- Shang HY, Wei YM, Wang XR et al (2006) Genetic diversity and phylogenetic relationships in the rye genus *Secale* L. (rye) based on *Secale cereale* microsatellite markers. *Genet Mol Biol* 29:685–691. <https://doi.org/10.1590/S1415-47572006000400018>
- Shaw J, Lickey EB, Schilling EE et al (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the Tortoise and the Hare III. *Am J Bot* 94(3):275–288. <https://doi.org/10.3732/ajb.94.3.275>
- Singh RJ, Röbbelen G (1977) Identification by Giemsa technique of the translocations separating cultivated rye from three wild species of *Secale*. *Chromosoma* 59:217–225
- Skuza L, Szućko I, Filip E, Adamczyk A (2019a) DNA barcoding in selected species and subspecies of Rye (*Secale*) using three chloroplast loci (*matK*, *rbcL*, *trnH-psbA*). *Not Bot Horti Agrobot* 47:54–62. <https://doi.org/10.15835/nbha47111248>
- Skuza L, Szućko I, Filip E, Strzała T (2019b) Genetic diversity and relationship between cultivated, weedy and wild rye species as revealed by chloroplast and mitochondrial DNA non-coding regions analysis. *PLoS ONE* 14(2):e0213023. <https://doi.org/10.1371/journal.pone.0213023>
- Tillich M, Lehwark P, Pellizzer T et al (2017) GeSeq—versatile and accurate annotation of organelle genomes. *Nucleic Acids Res* 45(W1):W6–W11. <https://doi.org/10.1093/nar/gkx391>
- Zhou J, Yang Z, Li G et al (2010) Diversified chromosomal distribution of tandemly repeated sequences revealed evolutionary trends in *Secale* (*Poaceae*). *Plant Syst Evol* 287:49–56. <https://doi.org/10.1007/s00606-010-0288-z>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.