

Complete chloroplast genome sequence of turnip (*Brassica rapa* ssp. *rapa*): genome structure and phylogenetic analysis

Rui Han^a, Minyu Tian^a, Guangnan Zhang^{a,b}, Dengkui Shao^{a,b} and Yanjing Ren^{a,b}

^aAcademy of Agriculture and Forestry Sciences of Qinghai University (Qinghai Academy of Agriculture and Forestry Sciences), Xining, P. R. China; ^bQinghai Key Laboratory of Vegetable Genetics and Physiology, Xining, P. R. China

ABSTRACT

Turnip (*Brassica rapa* ssp. *rapa*) is considered worldwide to be one of the most important leaf and root vegetable crops in the Brassicaceae family. However, to date, few chloroplast (cp) genomic resources have been reported for this genus. Here, we determined the complete cp genome sequences of *Brassica rapa* ssp. *rapa*. A 153,621 bp quadripartite cycle without any gap was obtained with a large single-copy region (LSC) of 83,512 bp, a small single-copy region (SSC) of 17,683 bp, and two inverted repeat (IR), IRa and IRb of 26,213 bp. A total of 132 genes were identified, including 87 protein-coding genes (PCG), 37 transfer RNA (tRNA), and 8 ribosomal RNA (rRNA). The phylogenetic analysis of ten other crops selected showed that the turnip was most closely related to the *Brassica rapa*.

ARTICLE HISTORY

Received 25 July 2020
Accepted 18 September 2020

KEYWORDS

Turnip; chloroplast genome; genome structure; phylogenetic analysis



Turnip (*Brassica rapa* ssp. *rapa* L. $2n=2\times=20$), one of the most important leaf and root vegetable crops in the Brassicaceae family, is widely spread planted in China and throughout East Asia. As a vegetable crop, turnip is rich in glucosinolates (Rochfort et al. 2006; Zhang et al. 2008), dietary phenolic (Chung et al. 2016), dietary fiber, vitamin C and other bioactive compounds (Thiruvengadam and Chung 2015). The complete chloroplast (cp) genome of angiosperms usually comprises four parts: a large single-copy region (LSC), a small single-copy region (SSC) and two inverted repeats (IR), IRa and IRb that reside between LSC and SSC, finally forming a typical quadripartite cycle (Li et al. 2017). Compared with the nuclear genome, the cp genome is inherited from the maternal parent, which is highly conserved in gene content and genome structure. Usually, the length of cp genomes ranges from 120 to 160 kb and the number of unique genes range from 110 to 130 (Du et al. 2020). Genes in the cp genome play crucial roles in photosynthesis and the biosynthesis of starch, amino acids, fatty acids, and pigments (Rodríguez-Ezpeleta et al. 2005).

Traditionally, turnip is considered to be closest to the *Brassica rapa* even if its phenotype is similar to the *Raphanus sativus*. In recent years, the cp genomes of many species in the Brassicaceae family were reported, and several phylogenetic analysis were performed to show the phylogenetic relationships among Brassicaceae species (Prabhudas et al. 2015; Seol et al. 2015; Du et al. 2020). However, to our knowledge, this is the first complete turnip cp genome sequence presented. The phylogenetic relationships were performed between seven Brassicaceae species, three other species

(*Oryza sativa*, *Solanum lycopersicum* and *Nicotiana tabacum*) and turnip. These results of this research will provide better understanding on evolutionary relationships among Brassicaceae species and turnip.

The fresh and healthy leaf of *B. rapa* ssp. *rapa* was collected from an individual turnip plant, W21, which planted in the field of Qinghai university (N36°42'; E 101°45'), Xining, China, and its genomic DNA was stored at Qinghai university. The DNA extraction followed by the modified CTAB (cetyl trimethyl ammonium bromide) method (Porebski et al. 1997). DNA purification, library construction and assessment were performed as detailed in the manufacturer's protocol. The qualified DNA were sequenced using an Illumina HiSeq2500 with an average read length of 1363 bp (Shaanxi Breeding Biotechnologies Co., Ltd). After raw reads were filtered to obtain high quality data by removing the adaptor reads and reads of low quality, a total of 8.38 Gb clean data with a Q30 value of 91.65% was obtained. The draft cp genome of turnip was assembled using the MITObim software (Hahn et al. 2013) with *Brassica nigra* (NC_030450) as a reference. The draft genome was further corrected by PE read mapping. Genes in the complete cp were annotated by the DOGMA (Wyman et al. 2004) and Mitofy (Alver Son et al. 2010) software.

The complete cp genome of turnip (GenBank accession no. MT409177) displayed a quadripartite cycle of 153,621 bp without any gaps, and consisting of four regions: a large single-copy region (LSC) of 83,512 bp, a small single-copy region (SSC) of 17,683 bp, and two inverted repeat (IR), IRa and IRb, of 26,213 bp. In the turnip cp genome, 132 distinctive genes

CONTACT Yanjing Ren  renyan0202@163.com  Academy of Agriculture and Forestry Sciences of Qinghai University (Qinghai Academy of Agriculture and Forestry Sciences), Xining, P. R. China

© 2020 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 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

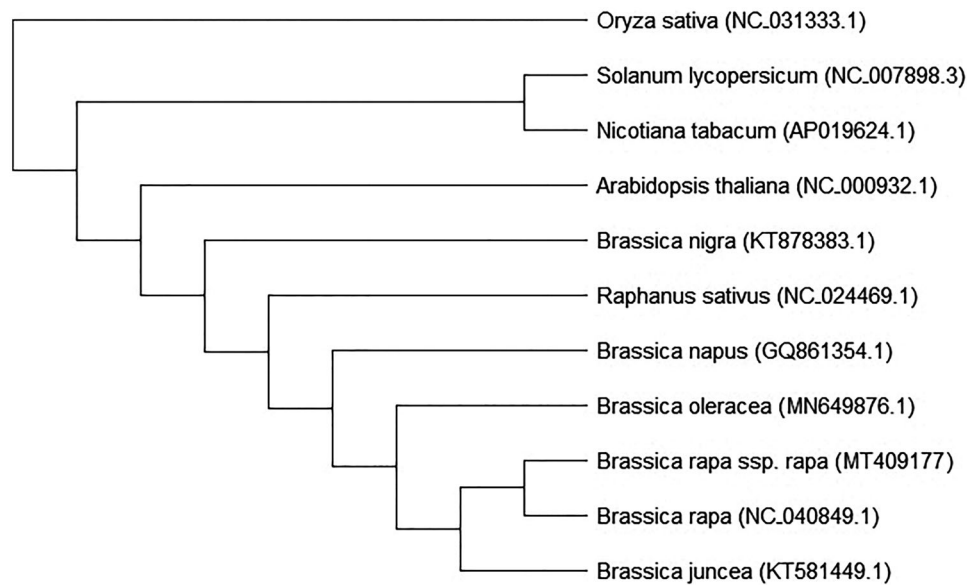


Figure 1. Phylogenetic analysis of 11 species based on the chloroplast protein-coding sequences. The chloroplast sequence of *Oryza sativa* (NC 031333.1) was used as the outgroup.

were identified, including 87 protein-coding genes, 37 tRNAs, and 8 rRNA. Of these 132 genes, there were 15 genes with two copies in the IR regions. They were *ndhB*, *rrn4.5*, *rrn5*, *rrn16*, *rrn23*, *rps7*, *rps12*, *rpl2*, *rpl23*, *trnI-CAU*, *trnV-GAC*, *trnN-GUU*, *ycf1*, *ycf2*, and *ycf15*, respectively.

The base composition of the complete cp genome sequence was analyzed and the overall GC content was 36.3%. The content of G and C was 18.6 and 17.8%, respectively. The overall GC content in IR regions (42.3%) were higher than that in the LSC (34.1%) and SSC regions (29.3%). The distribution of GC content in turnip cp genome was similar with that in *Lagerstroemia* (Zheng et al. 2020). Conversely, the overall AT content in SSC region (71.9%) was higher than that in the IR region (57.7%), and the content of A and T were 31.4 and 32.3%, respectively.

Phylogenetic relationships between the ten species and turnip were conducted by MEGA7 using the maximum likelihood (ML) method and 1000 bootstrap replicates with the *Oryza sativa* (NC 031333.1) as an outgroup. The results showed that *B. rapa* ssp. *rapa* (MT409177) was closely related to *B. rapa* (NC040849.1), while *B. nigra* (KT878383.1) is more diverse than the neighboring species *Raphanus sativus* (NC022469.1) (Figure 1). These results were consistent with other studies (Jeong et al. 2014; Seol et al. 2015), and supports that turnip is a sub-species of *B. rapa*.

Disclosure statement

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

Funding

This research was financially supported by The Qinghai Natural Science Foundation [2019-ZJ-979Q], The Project of Qinghai Key Laboratory of Vegetable Genetics and Physiology [2020-ZJ-Y02] and The Special Project of Qinghai Academy of Agriculture and Forestry Sciences [2018-NKY-008].

Data availability statement

The data that support the findings of this study are openly available in NCBI at <https://www.ncbi.nlm.nih.gov/nuccore/MT409177> with the accession no. MT409177.

References

- Alver Son AJ, Wei X, Rice DW, Stern DB, Barry K, Palmer JD. 2010. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Mol Biol Evol.* 27(6):1436–1448.
- Chung IM, Rekha K, Rajakumar G, Thiruvengadam M. 2016. Production of glucosinolates, phenolic compounds and associated gene expression profiles of hairy root cultures in turnip (*Brassica rapa* ssp. *rapa*). *3 Biotech.* 6(2):175.
- Du X, Zeng T, Feng Q, Hu L, Luo X, Weng Q, He J, Zhu B. 2020. The complete chloroplast genome sequence of yellow mustard (*Sinapis alba* L.) and its phylogenetic relationship to other Brassicaceae species. *Gene.* 731:144340.
- Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucl Acids Res.* 41(13): e129–e129.
- Jeong YM, Chung WH, Mun JH, Kim N, Yu HJ. 2014. De novo assembly and characterization of the complete chloroplast genome of radish (*Raphanus sativus* L.). *Gen.* 551:01.
- Li Y, Zhou J, Chen X, Cui Y, Xu Z, Li Y, Song J, Duan B, Yao H. 2017. Gene losses and partial deletion of small single-copy regions of the chloroplast genomes of two hemiparasitic *Taxillus* species. *Sci Rep.* 7: 12834.
- Porebski S, Bailey LG, Baum BR. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol Biol Rep.* 15(1):8–15.
- Prabhudas SK, Raju B, Thodi SK, Parani M, Natarajan P. 2015. The complete chloroplast genome sequence of Indian mustard (*Brassica juncea* L.). *Mitochondr DNA.* DOI: 10.3109/19401736.2015.1101586
- Rochfort SJ, Imsic M, Jones R, Trenerry VC, Tomkins B. 2006. Characterization of flavonol conjugates in immature leaves of pak choi (*Brassica rapa* L. ssp. *chinensis* L. (Hanelt.)) by HPLC-DAD and LC-MS/MS. *J Agric Food Chem.* 54(13):4855–4860.
- Rodríguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Löffelhardt W, Bohnert HJ, Philippe H, Lang BF. 2005. Monophyly of

- primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. *Curr Biol.* 15(14):1325–1330.
- Seol YJ, Kim K, Kang SH, Perumal S, Lee J, Kim CK. 2015. The complete chloroplast genome of two Brassica species, *Brassica nigra* and *B. Oleracea*. *Mitochondr DNA.* <http://dx.doi.org/10.3019/19401736.2015.1115493>
- Thiruvengadam M, Chung IM. 2015. Selenium, putrescine, and cadmium influence health-promoting phytochemicals and molecular-level effects on turnip (*Brassica rapa* ssp. *rapa*). *Food Chem.* 173: 185–193.
- Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics.* 20(17):3252–3255
- Zhang H, Schonhof I, Krumbein A, Gutezeit B, Li L, Stützel H, Schreiner M. 2008. Water supply and growing season influence glucosinolate concentration and composition in turnip root (*Brassica rapa* ssp. *rapifera* L.). *J Plant Nutr Soil Sci.* 171(2):255–265.
- Zheng G, Wei L, Ma L, Wu Z, Gu C, Chen K, 2020. Comparative analyses of chloroplast genomes from 13 Lagerstroemia (Lythraceae) species: identification of highly divergent regions and inference of phylogenetic relationships.