


The complete plastid genome sequence of garlic *Allium sativum* L

Mikhail A. Filyushin^a , Alexey V. Beletsky^a, Alexander M. Mazur^a and Elena Z. Kochieva^{a,b}

^aInstitute of Bioengineering, Research Center of Biotechnology of the Russian Academy of Sciences, Moscow, Russia; ^bFaculty of Biology, Lomonosov Moscow State University, Moscow, Russia

ABSTRACT

The complete plastid genome sequence of garlic *Allium sativum* was determined using Illumina sequencing. The plastid DNA is 153,172 bp in length and includes a large single copy region (LSC) of 82,035 bp and a small single copy region (SSC) of 18,015 bp, which are separated by a pair of 26,561 bp inverted repeat regions (IRs). In total, 134 genes are identified, containing 82 protein-coding genes, 38 tRNA genes, eight rRNA genes and six pseudogenes. Most of genes occur as a single copy, while 19 genes are duplicated in IRs. Among 15 intron-containing genes, *clpP* and *ycf3* contain two introns and the rest have one intron.

ARTICLE HISTORY

Received 28 September 2016
Accepted 10 October 2016

KEYWORDS

Garlic; chloroplast genome; next-generation sequencing; *Allium sativum*

Garlic (*Allium sativum* L.) is the second most important crop of the genus *Allium* after the bulb onion. It is cultivated and consumed worldwide and is popular for its nutritional and medicinal properties. Garlic production worldwide is estimated at more than 24 million tons and is steadily growing. Garlic cultivars are sterile and thus propagate only asexually. It was proposed that garlic originated in Central Asia and due to high ecological plasticity as well as to active trading, has spread throughout the world (Vavilov 1951; Hong & Etoh 1996).

Allium sativum is a monocotyledonous plant and belongs to section *Allium* genus *Allium* (family Amariyllidaceae order Asparagales), which contains more than 750 species (Friesen et al. 2006).

For sequencing *A. sativum* accession from Uzbekistan was chosen (specimen voucher VSRI: 31, Vavilov All-Russian Scientific Research Institute of Plant Industry). The complete garlic plastid genome was estimated by the high-throughput sequencing on the Illumina HiSeq 1500 Sequencing System (Illumina, CA). The plastid genome was assembled with SPAdes v3.8 (Bankevich et al. 2012) and manually finished with additional sequencing and *Allium cepa* (KF728079) as the reference. The resultant plastid genome was annotated by using the DOGMA program (<http://dogma.cccb.utexas.edu>) (Wyman et al. 2004) and by comparing with those of *A. cepa* (KF728079, KF728080, KM088013, KM088014) (von Kohn et al. 2013; Kim et al. 2015). A physical map of the *A. sativum* plastid genome was generated using the web tool OGDRAW

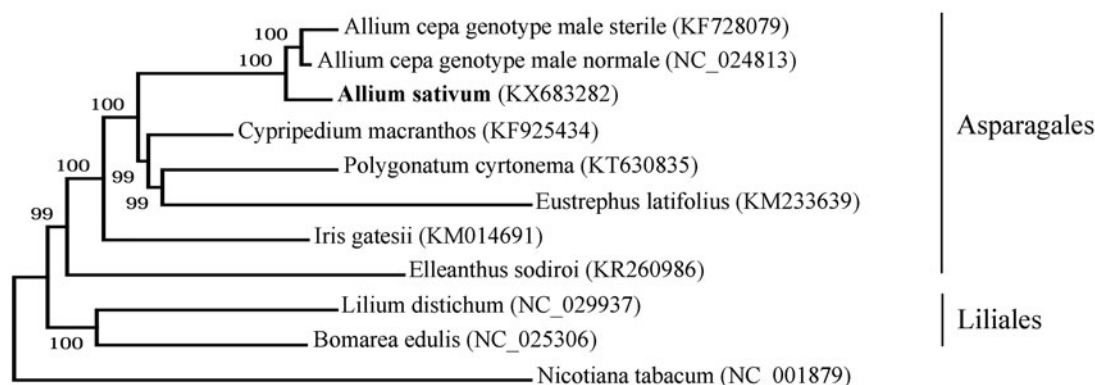


Figure 1. Phylogenetic tree inferred by maximum-likelihood using 82 protein-coding gene sequences of 10 species including seven species from the Asparagales order: *Allium cepa* (genotype male sterile KF728079 and genotype male fertile NC_024813), *Allium sativum* (KX683282), *Eustrephus latifolius* (KM233639), *Polygonatum cyrtonema* (KT630835), *Cyripedium macranthos* (KF925434), *Elleanthus sodiroi* (KR260986), *Iris gatesii* (KM014691); two species from Liliales order: *Bomarea edulis* (NC_025306), *Lilium distichum* (NC_029937); and *Nicotiana tabacum* (NC_001879) as an outgroup. PhyML 3.1 (Guindon et al. 2010) was used for the sequence alignment and construction of the tree. Bootstrap support values based on 1000 replicates are displayed on each node.

(<http://ogdraw.mpimp-golm.mpg.de>) (Lohse et al. 2013). The complete plastid genome sequence was submitted to GenBank with accession number KX683282.

The garlic plastid genome is 153,172 bp in length and comprises a large single copy region (LSC, 82,035 bp), small single copy region (SSC, 18,015 bp) and two inverted repeat regions (IRs, 26,561 bp).

The plastid genome harbors 134 genes that include 82 protein-coding genes, 38 tRNA genes, eight rRNA genes and six pseudogenes. Most of them are single copy genes, whereas 19 genes present in double copies, including six protein-coding genes (*rps19*, *rpl2*, *rpl23*, *ycf2*, *ndhB*, *rps7*), nine tRNA genes (*trnR-ACG*, *trnM-CAU*, *trnL-CAA*, *trnV-GAC*, *trnH-GUG*, *trnI-CAU*, *trnI-GAU*, *trnA-UGC*, *trnN-GUU*) and all four rRNA genes in IRs (*rrn4.5*, *rrn5*, *rrn16* and *rrn23*). Intron sequences are found in 15 genes, 13 (*atpF*, *rpoC1*, *trnL-UAA*, *trnV-UAC*, *ndhA*); four genes in IRs: *rpl2*, *ndhB*, *trnI-GAU*, *trnA-UGC*) of which contain a single intron while two (*clpP* and *ycf3*) have two introns. Six genes became pseudogenes due to internal stop codons identified in their coding sequences (*rps2*, *rps16*, *infA*, two *ycf15* in IRs) or because of incomplete duplication in the IRB/SSC junction region (*ycf1*).

Sequence comparison of *A. sativum* and *A. cepa* plastid genomes reveals similar gene order (von Kohn et al. 2013; Kim et al. 2015). Compared to *A. cepa*, in the plastid genome of *A. sativum* seven deletions (18–221 bp) in intergenic spacers and a number of short insertions (2–31 bp) are identified.

Phylogenetic analysis inferred from 82 protein-coding genes of plastid genome showed a close relationship of *A. sativum* and *A. cepa* (Figure 1).

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

Funding

This work was supported by the Russian Academy of Sciences, 10.13039/501100002674 Grant Funds [MCB 01201353319 and 0104-2014-0210].

ORCID

Mikhail A. Filyushin  <http://orcid.org/0000-0003-3668-7601>

References

- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 19:455–477.
- Friesen N, Fritsch RM, Blattner FR. 2006. Phylogeny and new intrageneric classification of *Allium* (Alliaceae) based on nuclear ribosomal DNA its sequences. *Aliso.* 22:372–395.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59:307–321.
- Kim S, Park JY, Yang TJ. 2015. Comparative analysis of complete chloroplast genome sequences of a normal male-fertile cytoplasm and two different cytoplasms conferring cytoplasmic male sterility in onion. *J Hortic Sci Biotechnol.* 90:459–468.
- Hong CJ, Etoh T. 1996. Fertile clones of garlic (*Allium sativum* L.) abundant around the Tien Shan Mountains. *Breed Sci.* 46:349–353.
- Lohse M, Drechsel O, Kahlau S, Bock R. 2013. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucl Acids Res.* 41:575–581.
- Vavilov NI. 1951. The origin, variation, immunity and breeding of cultivated plants. *Chronica Bot.* 13:1–364.
- von Kohn CM, Kielkowska A, Havey MJ. 2013. Sequencing and annotation of the chloroplast DNAs of normal (N) male-fertile and male-sterile (S) cytoplasms of onion and single nucleotide polymorphisms distinguishing these cytoplasms. *Genome.* 56:737–742.
- Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics.* 20:3252–3255.