



The complete mitochondrial genome of *Arabidopsis thaliana* (Brassicaceae) isolated in Korea

Jongsun PARK*, Hong XI and Yongsung KIM¹

InfoBoss Inc. and InfoBoss Research Center, Seoul 06278, Korea

¹*Honam National Institute of Biological Resources, Mokpo 58762, Korea*

(Received 7 May 2021; Revised 21 May 2021; Accepted 5 June 2021)

ABSTRACT: *Arabidopsis thaliana* (L.) Heynh. is a small plant species that serves as a model organism of plant biology and genetics. Here, we present the first complete mitochondrial genome of Korean *A. thaliana* natural isolate (named as 180404IB4), which is 368,875 bp long and contains 58 genes (33 protein-coding genes, 22 tRNAs, and three rRNAs), with a GC ratio of 44.8%. Sixty-four single-nucleotide polymorphisms and 11 insertion and deletion regions (1,089 bp in length) are identified against the Col-0 ecotype, showing one large insertion of 1,069 bp without structural variation. Phylogenetic trees constructed from 30 conserved genes indicate that the 180404IB4 mitochondrial genome is clustered with Col-0 and three East Asian ecotypes.

Keywords: *Arabidopsis thaliana*, Brassicaceae, Korean individual, intraspecific variations, mitochondrial genome

Arabidopsis thaliana (L.) Heynh. is a small plant species distributed in Eurasia and Africa (Hoffmann, 2002). It is a model organism of plant biology and genetics (Rensink and Buell, 2004) because of its rapid life cycle (ca. 10 weeks) (Mandel and Yanofsky, 1995) and relatively small genome size (ca. 119 Mb) (The Arabidopsis Genome Initiative, 2000). As next-generation sequencing technologies have been established, above one thousand whole genomes have been sequenced (Gan et al., 2011; Schmitz et al., 2013): e.g., 1,135 natural isolates from Eurasian and North African (The 1001 Genomes Consortium, 2016) and 118 strains from Yangtze River in China (Zou et al., 2017). In contrast, only six complete chloroplast (Park et al., 2020a) and three mitochondrial genomes (mitogenomes) (Unseld et al., 1997; Davila et al., 2011; Sloan et al., 2018) are available even though they are valuable to investigate phylogenetic relationship (Park et al., 2020a, 2020c, 2021). Here, we presented the complete mitogenome of *A. thaliana* isolated in Korea together with three East Asian isolates to understand their phylogenetic relationship.

Materials and Methods

Plant material

We collected an individual accession of *A. thaliana* in the population located in Yeonggwang-gun, Jeollanam-do province in Korea (180404IB4). A voucher and isolated DNA were deposited in the InfoBoss Cyber Herbarium (IN, the voucher number IB-00925).

DNA extraction and mitochondrial genome determination

Its total DNA was extracted from fresh leaves by using a DNeasy Plant Mini Kit (QIAGEN, Hilden, Germany). Genome sequencing was performed using HiSeqX at Macrogen Inc., Korea, and *de novo* assembly was done by Velvet v1.2.10 (Zerbino and Birney, 2008) and GapCloser v1.12 (Zhao et al., 2011). Assembled sequences were modified and confirmed by BWA v0.7.17 (Li, 2013) and SAMtools v1.9 (Li et al., 2009). All bioinformatic analyses were conducted in the Genome Information System (<http://geis.infoboss.co.kr/>) like the previous studies (Kwon et al., 2019a, 2019b; Park et al., 2019a, 2019d, 2020b; Min et al., 2020; Choi et al., 2021a, 2021b;

*Author for correspondence: starflr@infoboss.co.kr

Kim et al., 2021). The same method was applied to assemble mitogenomes of three East Asian *A. thaliana* isolates based on public raw reads (Table 1).

Genome annotation was conducted based on Col-0 mitogenome (NC_037304) (Sloan et al., 2018) with Geneious Prime 2020.2.4 (Biomatters Ltd., Auckland, New Zealand). A circular map of *A. thaliana* mitogenome was drawn using OGDRAW v1.31 (Greiner et al., 2019).

Phylogenetic analysis

Maximum-Likelihood (ML) and Bayesian inference (BI) phylogenetic trees were constructed based on the concatenated alignment of the 30 aligned conserved genes from the seven *A. thaliana* including Korean individual and three ecotypes (Table 1) and *Arabis alpina* (Xu and Bi, 2018) mitogenomes (Table 1) by MAFFT v7.450 (Katoh and Standley, 2013). The ML tree was reconstructed in MEGA X with 1,000 bootstrap repeats (Kumar et al., 2018). In the ML analysis, a heuristic search was used with nearest-neighbor interchange branch swapping, TVM + F + R4 model determined by jModelTest v2.1.5 (Darriba et al., 2012), and uniform rates among sites. All other options used the default settings. The posterior

probability of each node was estimated by BI using MrBayes v3.2.6 (Huelsenbeck and Ronquist, 2001) plug-in implemented in Geneious Prime 2020.2.4 (Biomatters Ltd.). The HKY85 model with gamma rates was used as a molecular model. A Markov chain Monte Carlo algorithm was employed for 1,100,000 generations, sampling trees every 200 generations, with four chains running simultaneously. Trees from the first 100,000 generations were discarded as burn-in.

Data availability

Mitochondrial genome sequence of Korean *A. thaliana* can be accessed via accession number MK358445 in NCBI GenBank. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA727818, SAMN19040823, and SRR14458670, respectively.

Results and Discussion

The complete mitogenome of *A. thaliana* 180404IB4 (MK358445) is 368,875 bp long (Fig. 1A), longer than three previously assembled and one newly assembled *A. thaliana* mitogenomes (Table 1). Overall GC content is 44.8% and it

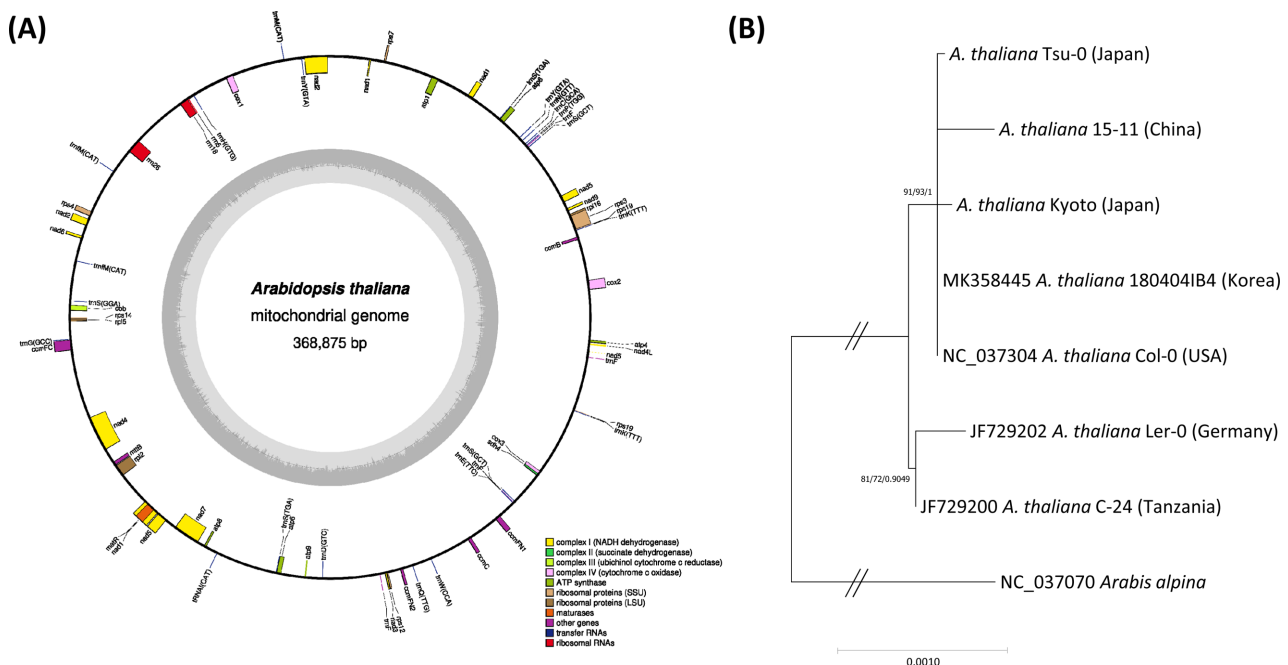


Fig. 1. **A.** Circular map of mitochondrial genome of *Arabidopsis thaliana* isolated in Korea. Gene shown outside are transcribed clockwise, and inside the circle are transcribed counter clockwise. Genes are colour-coded to distinguish different functional groups. The dark grey and the light grey plot in the inner circle correspond to the GC content and AT content, respectively. **B.** Maximum-Likelihood (ML) and Bayesian inference (BI) phylogenetic trees of 30 conserved genes from seven *Arabidopsis* and one *Arabis* complete mitochondrial genomes. Phylogenetic tree was drawn based on ML tree. The numbers above branches indicate bootstrap support value of ML tree and posterior probability of BI tree, respectively.

Table 1. List of the mitochondrial genomes used in this study.

Strain species	Origin	Complete genome	Accession number	Length (bp)	GC ratio (%)	Reference / Source
180404IB4	Korea	Y	MK358445	368,875	44.8	This study
Col-0	USA	Y	NC_037304	367,808	44.8	Sloan et al. 2018
Ler-0	Germany	Y	JF729202	345,418	44.8	Davila et al. 2011
C24	Tanzania	Y	JF729200	366,922	44.8	Davila et al. 2011
Kyoto	Japan	Y	N/A	350,265	44.8	SRR492307 (Schmitz et al. 2013)
Tsu-0	Japan	N	N/A	389,442	44.7	ERR031555 (Gan et al. 2011)
15-11	China	Y	N/A	368,875	44.8	SRR2204166 (Zou et al. 2017)
<i>Arabidopsis thaliana</i>		Y	NC_037070	323,159	41.3	Xu and Bi 2018

contains 58 genes (33 protein-coding genes, 22 tRNAs, and three rRNAs) (Fig. 1A), which are the same as the Col-0 mitogenome.

Sixty-four single nucleotide polymorphisms (SNPs; 0.017%) and 11 insertion and deletion (INDEL) regions (1,089 bp in length; 0.30%) were identified against Col-0 mitogenome. Interestingly, one 1,069-bp insertion was found, similar to the insertions identified in chloroplast genomes of *Coffea arabica* L. (Park et al., 2019c) and *Duchesnea chrysantha* (Zoll. & Moritz) Miq. (Park et al., 2019b) and mitogenomes of *Liriodendron tulipifera* L. (Park et al., 2019a) and *Populus alba* x *Populus glandulosa* (Park et al., 2019d). Proportion of intraspecific variations identified between 180404IB4 and Col-0 is similar to that of *Rosa rugosa* Thunb. (124 SNPs, 0.041% and 769-bp INDELs, 0.25%) (Park et al., 2020b), is less than those of *Liriodendron tulipifera* L. (365 SNPs, 0.07% and 2,117-bp INDELs, 0.38%) (Park et al., 2019a) and *Glycyrrhiza uralensis* Fisch. ex DC. (1,099 SNPs, 0.24% and 1,736-bp INDELs, 0.38%) (Baek et al., under review), but is greater in number than that of *Malus x domestica* Borkh. (140 SNPs, 0.035% and 6-bp INDELs, 0.0015%) (Goremykin et al., 2012; Ge et al., 2020).

No structural variation among 180404IB4, 15-11, and Col-0 mitogenomes was found, while the remaining *A. thaliana* genomes (Table 1) presented structural variations (Davila et al., 2011). It is also congruent to the recent analysis which presents a dynamic structure of *A. thaliana* mitogenome along with different ecotypes (Masutani et al., 2021).

Phylogenetic trees constructed from the conserved genes of the seven *A. thaliana* and *Arabidopsis alpina* mitogenomes (Table 1) show that four East Asian *Arabidopsis* including 180404IB4 were clustered with the Col-0 (USA) with high supportive

values and Ler-0 (Germany) and C-24 (Tanzania) were clustered in an independent clade (Fig. 1B). It is different from the phylogenetic tree constructed from the complete chloroplast genomes, displaying the intraspecific topology of the six *A. thaliana* ecotypes except for C-24 with high supportive values (Park et al., 2020a). It might be caused by the low coverage of the 30 conserved genes (6.98%) against the complete mitogenome in comparison to that of the chloroplast genomes. To improve the resolution of phylogenetic analysis using mitogenomes, more conserved areas except genic regions should be rescued and addition mitogenomes of East Asian *A. thaliana* are needed to be assembled to investigate their phylogenetic relationship. Throughout this study, this mitogenome presents that Korean *A. thaliana* is phylogenetically related to the other East Asian isolates and Col-0.

ORCID: Jongsun PARK <http://orcid.org/0000-0003-0786-4701>;
Yongsung KIM <https://orcid.org/0000-0002-5349-9226>

Acknowledgments

This work was supported by InfoBoss Research Grant (IBG-0023).

Conflict of Interest

The authors declare that there are no conflicts of interest.

Literature Cited

Choi, S. S., V. A. Bakalin, W. Kwon and J. Park. 2021a. The com-

- plete mitochondrial genome of *Douinia plicata* (Lindb.) Konstant. et. Vilnet (Scapaniaceae, Jungermanniales). Mitochondrial DNA Part B 6: 789–791.
- Choi, S. S., J. Min, W. Kwon and J. Park. 2021b. The complete mitochondrial genome of *Scapania ampliata* Steph., 1897 (Scapaniaceae, Jungermanniales). Mitochondrial DNA Part B Resources 6: 686–688.
- Darriba, D., G. L. Taboada, R. Doallo and D. Posada. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9:772.
- Davila, J. I., M. P. Arrieta-Montiel, Y. Wamboldt, J. Cao, J. Haggmann, V. Shedge, Y.-Z. Xu, D. Weigel and S. A. Mackenzie. 2011. Double-strand break repair processes drive evolution of the mitochondrial genome in *Arabidopsis*. BMC Biology 9: 64.
- Gan, X., O. Stegle, J. Behr, J. G. Steffen, P. Drewe, K. L. Hildebrand, R. Lyngsoe, S. J. Schultheiss, E. J. Osborne, V. T. Sreedharan, A. Kahles, R. Bohnert, G. Jean, P. Derwent, P. Kersey, E. J. Belfield, N. P. Harberd, E. Kemen, C. Toomajian, P. X. Kover, R. M. Clark, G. Rättsch and R. Mott. 2011. Multiple reference genomes and transcriptomes for *Arabidopsis thaliana*. Nature 477: 419–423.
- Ge, D., J. Dong, L. Guo, M. Yan, X. Zhao and Z. Yuan. 2020. The complete mitochondrial genome sequence of cultivated apple (*Malus domestica* cv. 'Yantai Fuji 8'). Mitochondrial DNA Part B Resources 5: 1317–1318.
- Goremykin, V. V., P. J. Lockhart, R. Viola and R. Velasco. 2012. The mitochondrial genome of *Malus domestica* and the import-driven hypothesis of mitochondrial genome expansion in seed plants. The Plant Journal 71: 615–626.
- Greiner, S., P. Lehwark and R. Bock. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Research 47: W59–W64.
- Hoffmann, M. H. 2002. Biogeography of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae). Journal of Biogeography 29: 125–134.
- Huelsenbeck, J. P. and F. Ronquist 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754–755.
- Katoh, K. and D. M. Standley 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30: 772–780.
- Kim, M., H. Xi and J. Park. 2021. Genome-wide comparative analyses of GATA transcription factors among 19 *Arabidopsis* ecotype genomes: Intraspecific characteristics of GATA transcription factors. PLoS One 16: e0252181.
- Kumar, S., G. Stecher, M. Li, C. Knyaz and K. Tamura. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution 35: 1547–1549.
- Kwon, W., Y. Kim and J. Park. 2019a. The complete mitochondrial genome of *Dumortiera hirsuta* (Sw.) Nees (Dumortieraceae, Marchantiophyta). Mitochondrial DNA Part B Resources 4: 1586–1587.
- Kwon, W., Y. Kim and J. Park. 2019b. The complete mitochondrial genome of Korean *Marchantia polymorpha* subsp. *ruderalis* Bischl. & Boisselier: inverted repeats on mitochondrial genome between Korean and Japanese isolates. Mitochondrial DNA Part B Resources 4: 769–770.
- Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at: <https://arxiv.org/abs/1303.3997> (2013).
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin and 1000 Genome Project Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25: 2078–2079.
- Mandel, M. A. and M. F. Yanofsky. 1995. A gene triggering flower formation in *Arabidopsis*. Nature 377: 522–524.
- Masutani, B., S.-I. Arimura and S. Morishita. 2021. Investigating the mitochondrial genomic landscape of *Arabidopsis thaliana* by long-read sequencing. PLoS Computational Biology 17: e1008597.
- Min, J., W. Kwon, H. Xi and J. Park. 2020. The complete mitochondrial genome of *Riccia fluitans* L. (Ricciaceae, Marchantiophyta): Investigation of intraspecific variations on mitochondrial genomes of *R. fluitans*. Mitochondrial DNA Part B Resources 5: 1220–1222.
- Park, J., Y. Kim and M. Kwon. 2019a. The complete mitochondrial genome of tulip tree, *Liriodendron tulipifera* L. (Magnoliaceae): Intra-species variations on mitochondrial genome. Mitochondrial DNA Part B Resources 4: 1308–1309.
- Park, J., Y. Kim and K. Lee. 2019b. The complete chloroplast genome of Korean mock strawberry, *Duchesnea chrysantha* (Zoll. & Moritz) Miq. (Rosoidae). Mitochondrial DNA Part B Resources 4: 864–865.
- Park, J., Y. Kim, H. Xi and K.-I. Heo. 2019c. The complete chloroplast genome of ornamental coffee tree, *Coffea arabica* L. (Rubiaceae). Mitochondrial DNA Part B Resources 4: 1059–1060.
- Park, J., Y. Kim, H. Xi, W. Kwon and M. Kwon. 2019d. The complete chloroplast and mitochondrial genomes of Hyunsasi tree, *Populus alba* x *Populus glandulosa* (Salicaceae). Mitochondrial DNA Part B Resources 4: 2521–2522.
- Park, J., J. Min, Y. Kim and Y. Chung. 2021. The comparative analyses of six complete chloroplast genomes of morphologically diverse *Chenopodium album* L. (Amaranthaceae) col-

- lected in Korea. *International Journal of Genomics* 2021: 6643444.
- Park, J., H. Xi and Y. Kim. 2020a. The complete chloroplast genome of *Arabidopsis thaliana* isolated in Korea (Brassicaceae): an investigation of intraspecific variations of the chloroplast genome of Korean *A. thaliana*. *International Journal of Genomics* 2020: 3236461.
- Park, J., H. Xi, Y. Kim, S. Nam and K.-I Heo. 2020b. The complete mitochondrial genome of new species candidate of *Rosa rugosa* (Rosaceae). *Mitochondrial DNA Part B Resources* 5: 3435–3437.
- Park, J., H. Xi and S.-H. Oh. 2020c. Comparative chloroplast genomics and phylogenetic analysis of the *Viburnum dilatatum* complex (Adoxaceae) in Korea. *Korean Journal of Plant Taxonomy* 50: 8–16.
- Rensink, W. A. and C. R. Buell. 2004. *Arabidopsis* to rice: applying knowledge from a weed to enhance our understanding of a crop species. *Plant Physiology* 135: 622–629.
- Schmitz, R. J., M. D. Schultz, M. A. Urich, J. R. Nery, M. Pelizola, O. Libiger, A. Alix, R. B. McCosh, H. Chen, N. J. Schork and J. R. Ecker. 2013. Patterns of population epigenomic diversity. *Nature* 495: 193–198.
- Sloan, D. B., Z. Wu and J. Sharbrough. 2018. Correction of persistent errors in *Arabidopsis* reference mitochondrial genomes. *The Plant Cell* 30: 525–527.
- The 1001 Genomes Consortium. 2016. 1,135 genomes reveal the global pattern of polymorphism in *Arabidopsis thaliana*. *Cell* 166: 481–491.
- The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.
- Unsold, M., P. Marienfeld, P. Brandt and A. Brennicke. 1997. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nature Genetics* 15: 57–61.
- Xu, Y. and C. Bi. 2018. The complete mitochondrial genome sequence of an alpine plant *Arabis alpina*. *Mitochondrial DNA Part B Resources* 3: 725–727.
- Zerbino, D. R. and E. Birney. 2008. Velvet: algorithms for *de novo* short read assembly using *de Bruijn* graphs. *Genome Research* 18: 821–829.
- Zhao, Q.-Y., Y. Wang, Y.-M. Kong, D. Luo, X. Li and P. Hao. 2011. Optimizing *de novo* transcriptome assembly from short-read RNA-Seq data: a comparative study. *BMC Bioinformatics* 12 Suppl 14: S2.
- Zou, Y.-P., X.-H. Hou, Q. Wu, J.-F. Chen, Z.-W. Li, T.-S. Han, X.-M. Niu, L. Yang, Y.-C. Xu, J. Zhang, F.-M. Zhang, D. Tan, Z. Tian, H. Gu and Y.-L. Guo. 2017. Adaptation of *Arabidopsis thaliana* to the Yangtze River basin. *Genome Biology* 18: 239.