De Novo Assembly of a Bell Pepper Endornavirus Genome Sequence Using RNA Sequencing Data

Yeonhwa Jo, Hoseng Choi, Won Kyong Cho

Department of Agricultural Biotechnology, College of Agriculture and Life Sciences, Seoul National University, Seoul, Republic of Korea

The genus *Endornavirus* is a double-stranded RNA virus that infects a wide range of hosts. In this study, we report on the *de novo* assembly of a bell pepper endornavirus genome sequence by RNA sequencing (RNA-Seq). Our result demonstrates the successful application of RNA-Seq to obtain a complete viral genome sequence from the transcriptome data.

Address correspondence to Won Kyong Cho, wonkyong@gmail.com.

The members of the genus *Endornavirus*, of the family *Endornaviridae*, have been identified in a wide range of hosts, such as plants, fungi, and oomycetes (1, 2). In general, the genomes of endornaviruses are composed of a double-stranded RNA of 9.8 to 17.6 kb in length that encodes a single polyprotein (3). In particular, they do not form true virions, and they present in the host at low copy numbers (2). In addition, they are vertically transmitted by seeds and do not cause any observable disease symptoms (2). So far, the genomes for several different endornaviruses have been sequenced (3).

In our search for endornaviruses infecting various plant species, we found a long novel transcript (accession no. JW157405.1) from the transcriptome shotgun assembly (TSA) sequence database that showed a strong sequence identity to bell pepper endornavirus (4). The transcriptome was derived from a previous study in which transcriptome analysis was performed for three different pepper cultivars by next-generation sequencing (5). The mRNA libraries were sequenced by Illumina GAIIx and *de novo* assembled by combining three different assembly methods, Velvet, CLC Genomics Workbench, and CAP3 (6). To confirm *de novo* transcriptome assembly, we downloaded the raw data for the pepper cultivar Maor from the NCBI Short Read Archive (SRA) database and performed *de novo* transcriptome assembly using the Trinity method (7). A novel transcript similar to bell pepper endornavirus was identified. This result suggests the successful *de novo* assembly of a viral genome regardless of the transcriptome assembly method. After alignment of the obtained transcript on the bell pepper endornavirus reference sequence (GenBank accession no. NC_015781.2), we found that an adenine was inserted in the middle of the assembled transcript (at nucleotide [nt] 7633), which led to a frameshift. After removing the adenine nucleotide, the complete open reading frame was obtained. We named the newly identified endornavirus bell pepper endornavirus isolate Maor. Its complete genome is 14,659 bp long, which is 69 bp shorter than the reference genome. In particular, the 5′- and 3′-terminal noncoding regions of the obtained viral genome are 61 bp and 8 bp shorter, respectively, than those of the reference genome. In addition, the BLAST result showed

that the genome of bell pepper endornavirus isolate Maor is 88% identical to that of the reference genome. Specifically, the open reading frame region is highly variable between the two viral genomes. The G+C content for isolate Maor is 41.01%, while that of the reference genome is 40.41%. The bell pepper endornavirus isolate Maor encodes a polyprotein 4,815 amino acids in length and containing four viral domains, including methyltransferase, helicase, glycosyltransferase, and RNA-dependent RNA polymerase. Taken together, we demonstrated the successful application of *de novo* assembly by RNA sequencing (RNA-Seq) to reveal a complete viral genome sequence from the transcriptome data.

Nucleotide sequence accession number. The genome sequence of bell pepper endornavirus isolate Maor has been deposited in GenBank under the accession no. KP455654.

ACKNOWLEDGMENT

REFERENCES

1. **Roossinck MJ, Sabanadzovic S, Okada R, Valverde RA.** 2011. The remarkable evolutionary history of endornaviruses. J Gen Virol 92:2674–2678. http://dx.doi.org/10.1099/vir.0.034702-0.
2. **Fukuhara T, Koga R, Aoki N, Yuki C, Yamamoto N, Oyama N, Udagawa T, Horiuchi H, Miyazaki S, Higashi Y, Takeshita M, Ikeda K, Arakawa M, Matsumoto N, Moriyama H.** 2006. The wide distribution of endornaviruses, large double-stranded RNA replicons with plasmid-like properties. Arch Virol 151:995–1002. http://dx.doi.org/10.1007/s00705-005-0688-5.
3. **Song D, Cho WK, Park S-H, Jo Y, Kim K-H.** 2013. Evolution of and horizontal gene transfer in the *Endornavirus* genus. PLoS One 8:e64270. http://dx.doi.org/10.1371/journal.pone.0064270.
4. **Okada R, Kiyota E, Sabanadzovic S, Moriyama H, Fukuhara T, Saha P, Roossinck MJ, Severin A, Valverde RA.** 2011. Bell pepper endornavirus: molecular and biological properties, and occurrence in the genus *Capsicum*. J Gen Virol 92:2664–2673. http://dx.doi.org/10.1099/vir.0.034686-0.
5. **Ashrafi H, Hill T, Stoffel K, Kozik A, Yao J, Chin-Wo SR, Van Deynze A.**

2012. *De novo* assembly of the pepper transcriptome (*Capsicum annuum*): a benchmark for *in silico* discovery of SNPs, SSRs and candidate genes. BMC Genomics **13**:571. http://dx.doi.org/10.1186/1471-2164-13-571.

6. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res **18**:821–829. http://dx.doi.org/10.1101/gr.074492.107.

7. **Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N.** 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol **29**:644–652. http://dx.doi.org/10.1038/nbt.1883.