



Data in Brief

Whole genome sequence and genome annotation of *Colletotrichum acutatum*, causal agent of anthracnose in pepper plants in South Korea



Joon-Hee Han^a, Jae-Kyung Chon^b, Jong-Hwa Ahn^b, Ik-Young Choi^c, Yong-Hwan Lee^d, Kyoung Su Kim^{a,*}

^a Applied Biology Program, Division of Bioresource Sciences, Kangwon National University, Chuncheon 24341, South Korea

^b National Instrumentation Center for Environmental Management, Seoul National University, Seoul 08829, South Korea

^c Department of Agriculture and Life Industry, Kangwon National University, Chuncheon 24341, South Korea

^d Department of Agricultural Biotechnology, Center for Fungal Genetic Resources, Center for Fungal Pathogenesis, Plant Genomics and Breeding Institute, Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 08829, South Korea

ARTICLE INFO

Article history:

Received 21 March 2016

Received in revised form 22 March 2016

Accepted 22 March 2016

Available online 28 March 2016

Keywords:

Colletotrichum acutatum

Anthracnose

Ascomycete

Pepper

Genome sequence

ABSTRACT

Colletotrichum acutatum is a destructive fungal pathogen which causes anthracnose in a wide range of crops. Here we report the whole genome sequence and annotation of *C. acutatum* strain KC05, isolated from an infected pepper in Kangwon, South Korea. Genomic DNA from the KC05 strain was used for the whole genome sequencing using a PacBio sequencer and the MiSeq system. The KC05 genome was determined to be 52,190,760 bp in size with a G + C content of 51.73% in 27 scaffolds and to contain 13,559 genes with an average length of 1516 bp. Gene prediction and annotation were performed by incorporating RNA-Seq data. The genome sequence of the KC05 was deposited at DDBJ/ENA/GenBank under the accession number LUXP00000000.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Specifications

Organism/cell line/tissue	<i>Colletotrichum acutatum</i>
Strain	KC05
Sequencer or array type	PacBio RS II and Illumina MiSeq system
Data format	Raw and processed
Experimental factors	DNA extracted from a field strain, no treatment
Experimental features	Genome sequencing
Consent	Data are publicly available
Sample source location	Pepper field in Chuncheon, Kangwon, South Korea

1. Direct link to deposited data

Deposited data can be found here: <http://www.ncbi.nlm.nih.gov/nuccore/LUXP00000000.1/>

2. Experimental design, materials, and methods

Colletotrichum acutatum J.H. Simmonds is an ascomycete pathogenic fungus which causes the most destructive plant disease called

anthracnose in a wide range of plants including vegetables, fruits, legumes, and perennial trees worldwide [1–4]. In particular, *C. acutatum* causes severe economic losses in the production of peppers in tropical and subtropical countries [5–8]. Missing genomic sequences of *C. acutatum* are a major hindrance to understand the development and pathogenicity of this fungus. Here we present the whole genome sequence of the *C. acutatum* strain KC05 isolated from an infected pepper fruit in South Korea; its mitogenome was announced recently [9].

The whole genome shotgun sequence of *C. acutatum* KC05 was generated using the PacBio RS II system with a PacBio Single Molecule Real-Time (SMRT) sequencer (Pacific Biosciences, Menlo Park, CA) and a MiSeq system (Illumina, San Diego, CA). A total of 3034 Mbp from a 20-kb whole genome shotgun library was sequenced by SMRT sequencing in 233,599 long single pass reads which were assembled using the hierarchical genome assembly process (HGAP) [10]. The sequence reads were assembled with an average consensus concordance of 99.97% and an average coverage depth of 49.7×. Low-quality contig sequences were corrected by mapping to the contigs with high quality (10,338,346 reads; 2 × 250 bp) generated from a 600-bp paired-end library using the Illumina MiSeq system. A total of 39 contigs produced by HGAP were assembled for scaffolding with the Illumina 5-kb mate-pair library sequence data using SSPACE V3.0 software [11]. The final assembly of the entire *C. acutatum* KC05 genome sequence was determined to be 52,190,760 bp with a G + C content of 51.73% in 27 scaffolds

* Corresponding author at: Applied Biology Program, Division of Bioresource Sciences, Kangwon National University, Chuncheon, South Korea.

E-mail address: kims@kangwon.ac.kr (K.S. Kim).

(minimum, 14,215 bp; maximum, 6,776,476 bp; average, 1,932,991; N_{50} , 4,416,304 bp) and was submitted to GenBank (LUXP00000000).

A total of 13,559 genes in the *C. acutatum* KC05 genome sequence were annotated by BLAST using the non-redundant (nr) protein database. The evidence gene data were obtained through *de novo* assembly of the transcriptome using Trinity v.2.06 software [12], and were mapped to the genome using PASA v2.0.1 software (<http://pasapipeline.github.io/>). GeneMark-ES v.4.21 [13] and EvidenceModeler pipeline v.1.1.1 software [14] were used for self-training and for combining gene predictions, respectively. A total of 13,559 gene sets were composited, with a shortest gene length of 281 bp, a largest gene length of 27,642 bp, and an average gene length of 1516 bp in *C. acutatum* KC05.

3. Nucleotide sequence accession number

This Whole Genome project of *C. acutatum* KC05 has been deposited at DDBJ/ENA/GenBank under the accession LUXP00000000. The version described in this paper is version LUXP01000000.

Conflict of interest

The authors declare that there is no conflict of interests with respect to the work published in this paper.

Acknowledgments

This study was supported by a grant from the Strategic Initiative for Microbiomes in Agriculture and Food, Ministry of Agriculture, Food and Rural Affairs, Republic of Korea, by 2015 research grant from Kangwon National University (No. 520150109), and by BioHerb Research Institute, Kangwon National University, Republic of Korea.

References

- [1] B. Bernstein, E.I. Zehr, R.A. Dean, E. Shabi, Characteristics of *Colletotrichum* from peach, apple, pecan, and other hosts. *Plant Dis.* 79 (1995) 478–482.
- [2] N.A. Peres, L.W. Timmer, J.E. Adaskaveg, J.C. Correll, Lifestyles of *Colletotrichum acutatum*. *Plant Dis.* 89 (2005) 784–796.
- [3] J.H. Simmond, A study of the species of *Colletotrichum* causing ripe fruit rots in Queensland. *Queensl. J. Agric. Anim. Sci.* 22 (1965) 437–459.
- [4] H. Xia, X.L. Wang, H.J. Zhu, B.D. Gao, First report of anthracnose caused by *Glomerella acutata* on chili pepper in China. *Plant Dis.* 95 (2010) 219.
- [5] J.T. Kim, S.Y. Park, W.B. Choi, Y.H. Lee, H.T. Kim, Characterization of *Colletotrichum* isolates causing anthracnose of peppers in Korea. *Plant Pathol. J.* 24 (2005) 17–23.
- [6] S. Sreenivasaprasad, P. Talhinhas, Genotypic and phenotypic diversity in *Colletotrichum acutatum*, a cosmopolitan pathogen causing anthracnose on a wide range of hosts. *Mol. Plant Pathol.* 6 (2005) 361–378.
- [7] P.P. Than, R. Jeewon, K.D. Hyde, S. Pongsupasamit, O. Mongkolporn, P.W.J. Taylor, Characterization and pathogenicity of *Colletotrichum* species associated with anthracnose on chilli (*Capsicum* spp.) in Thailand. *Plant Pathol.* 57 (2008) 562–572.
- [8] R. Voorrips, R. Finkers, L. Sanjaya, R. Groenwold, QTL mapping of anthracnose (*Colletotrichum* spp.) resistance in a cross between *Capsicum annuum* and *C. chinense*. *Theor. Appl. Genet.* 109 (2004) 1275–1282.
- [9] J.O. Kim, K.Y. Choi, J.H. Han, I.Y. Choi, Y.H. Lee, K.S. Kim, The complete mitochondrial genome sequence of the ascomycete plant pathogen *Colletotrichum acutatum*. *Mitochondrial DNA early online* (2015) 1–2.
- [10] C.S. Chin, D.H. Alexander, P. Marks, A.A. Klammer, J. Drake, C. Heiner, et al., Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10 (2013) 563–569.
- [11] M. Boetzer, C.V. Henkel, H.J. Jansen, D. Butler, W. Pirovano, Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27 (2011) 578–579.
- [12] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, et al., Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* 29 (2011) 644–652.
- [13] A. Lomsadze, V. Ter-Hovhannisyan, Y.O. Chernoff, M. Borodovsky, Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res.* 33 (2005) 6494–6506.
- [14] B.J. Haas, S.L. Salzberg, W. Zhu, M. Pertea, J.E. Allen, J. Orvis, et al., Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* 9 (2008) R7.