Data article

Draft genome sequence of *Colletotrichum sansevieriae* Sa-1–2, the anthracnose pathogen of *Sansevieria trifasciata*

Masayuki Nakamura [a,*], Takashi Fujikawa [b], Daichi Nakamori [a], Hisashi Iwai [a]

[a] *Faculty of Agriculture, Kagoshima University, Kagoshima 890-0065, Japan*
[b] *Institute of Fruit Tree and Tea Science, National Agriculture and Food Research Organization (NARO), Fujimoto 2-1, Tsukuba, Ibaraki 305-8605, Japan*

A R T I C L E   I N F O

A B S T R A C T

Colletotrichum sansevieriae is an ascomycete fungus causing anthracnose disease on plants in the genus *Sansevieria*. Here, we report the draft genome sequence of isolate Sa-1–2 of this fungus. The genome size is > 51 Mb, and the assembly consists of 8647 contigs and contains 13,664 predicted protein-coding genes. Pathogenicity factors such as plant cell wall-degrading enzymes and effector proteins were also predicted. Additionally, the phylogenetic relationship of isolates from different *Colletotrichum* spp. was analyzed, revealing that the isolate belongs to a novel major clade consisting of species that infect succulent plants originating from Africa. The draft genome sequence has been deposited at GenBank under accession number NJHP00000000.

---

* Corresponding author.
  E-mail address: masa@agri.kagoshima-u.ac.jp (M. Nakamura).

Specifications Table

Subject area	*Biology*
More specific subject area	*Microbiology, Genomics*
Type of data	*Table, text file, figure*
How data was acquired	*Shotgun whole-genome DNA sequences using Ion PGM*
Data format	*Raw and assembled*
Experimental factors	*Genomic DNA extracted from Colletotrichum sansevieriae Sa-1–2*
Experimental features	*de novo assembly, gene prediction, phylogenetic analysis*
Data source location	*Kagoshima, Japan*
Data accessibility	*Deposited data is available at the National Center for Biotechnology Information (NCBI) under the accession number NJHP00000000 (https:// www.ncbi.nlm.nih.gov/nuccore/NJHP00000000).*

Value of the data

- The first draft genome sequence of *Colletotrichum sansevieriae*, a causal agent of anthracnose on sansevieria, is now available.
- Plant cell wall-degrading enzymes and effector proteins related to pathogenicity were predicted.
- The fungus belongs to a novel major clade in the genus *Colletotrichum*.
- These data will be useful for further research into the biology, evolution, and pathogenicity of anthracnose pathogens.

1. Data

Colletotrichum sansevieriae is an anthracnose pathogen identified as a new species in 2006 [1]. The fungus causes water-soaked lesions and leaf blight on sansevieria, which is one of the most important plants in subtropical regions in Japan as a potted ornamental and for cut leaves [1,2]. The fungus shows pathogenicity only on *Sansevieria* spp., thereby having high host specificity. We are interested in learning which factors determine their host specificity. In this work, we performed draft genome analysis of the fungus to obtain fundamental genetic information for elucidating its host specificity determinants.

The draft assembly consists of 8647 contigs (GenBank accession number NJHP00000000) with a total length of 51.2 Mb, a G+C content of 50.8%, an N50 of 15,122 bp and an average length of 5922 bp. The total number of predicted tRNAs and rRNAs was 339 and 75, respectively. There were 13,664 predicted protein-coding genes, of which 1334 were classified as secreted.

The secretory proteins were annotated and 144 were predicted to be plant cell wall-degrading enzymes for cellulose, hemicellulose and pectin. Putative effector proteins were also analyzed, with a total of 316 proteins predicted, including 14 that had no match to sequences of other filamentous fungi deposited in GenBank. We compared the predicted effectors to those derived from other fungi [3] (Table 1). The percentage of effectors from isolate Sa-1–2 in the secretome was almost the same as for other hemibiotrophic fungi such as *C. higginsianum*, *Magnaporthe oryzae* and *Fusarium graminearum*. The obligate biotrophic fungus *Puccinia graminis* has the largest number of effector candidates in the secretome. In contrast, the necrotrophic fungus *Botrytis cinerea* has a smaller set of effector candidates.

C. sansevieriae was reported as a new species for the first time in 2006 [1] and was classified based only on the internal transcribed spacer 2 region of an rRNA gene. Thus, we also analyzed the fungus phylogenetically using the genetic data obtained in this work. Thus far, it has revealed that the genus *Colletotrichum* with straight conidia comprises five major clades [4]. Based on this work, *C. sansevieriae* Sa-1–2 belongs to a novel major clade consisting of newly identified species *C. euphorbiae* [5], *C. ledebouriae* [6] and *C. neosansevieriae* [7] that also infect succulent plants originating from Africa.

Table 1

Effector prediction of *Colletotrichum sansevieriae* Sa-1–2 and comparison with various fungal species [3].

Fungal species	No. of predicted secretome	No. of predicted effectors
C. sanevieriae [This study]	1334	316 (23.6%)[a]
C. higginsianum	1528	471 (30.8%)
C. graminicola	1339	268 (20.0%)
Puccinia graminis f.sp. *tritici*	1946	846 (43.5%)
Magnaporthe oryzae	1576	485 (30.7)
Botrytis cinerea	933	183 (18.4%)

[a] Percentage of predicted effectors in the secretome.
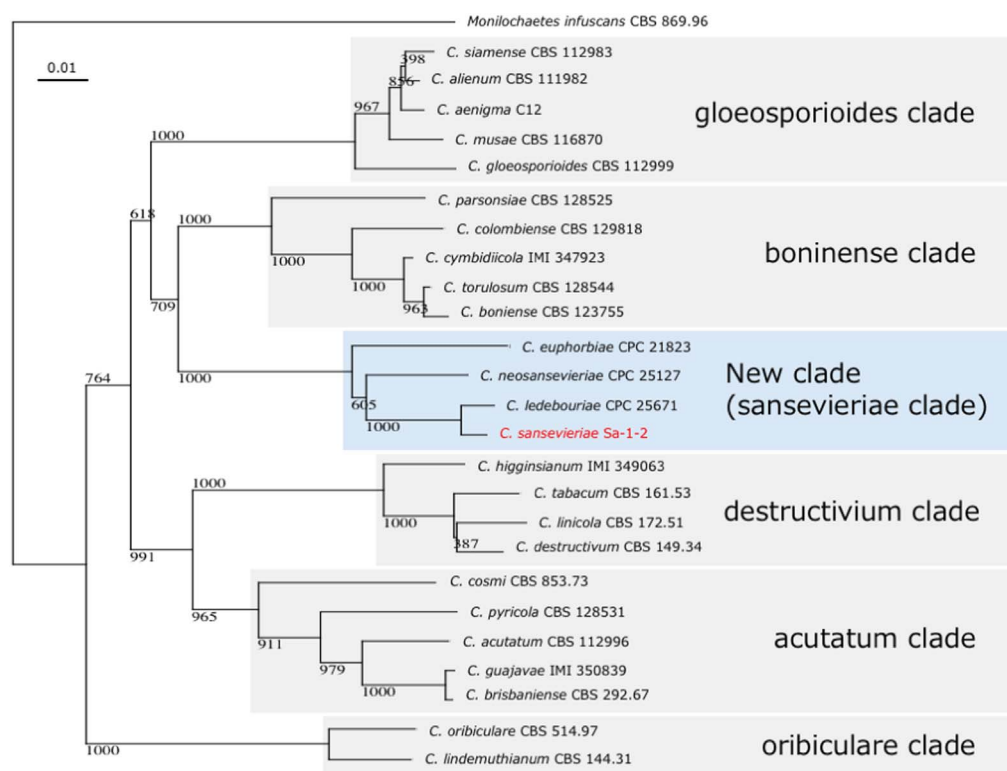
Fig. 1. Phylogenetic tree derived from the neighbor-joining method of concatenated alignment of rRNA genes, histone H3 genes and chitin synthase genes. The bootstrap value of 1000 replications is given for each node. Source accession numbers are described after the scientific names.

2. Experimental Design, Materials and Methods

2.1. Sequencing and assembly

Genomic DNA of the isolate Sa-1–2 was extracted from hyphae grown in potato dextrose broth (Difco, BD Diagnostic Systems, Sparks, MD, USA) and ultrasonicated by a Bioruptor (Cosmo Bio Co., Ltd., Tokyo, Japan). The genomic fragments were processed to template samples using the Ion Plus Fragment Library Kit (Thermo Fisher Scientific, Waltham, MA, USA) and the Ion PGM Hi-Q OT2 Kit with the Ion OneTouch 2 System (Thermo Fisher Scientific). Then, the template samples were

sequenced using the Ion PGM Sequencing Hi-Q Kit (Thermo Fisher Scientific) and a 318 Chip with the next generation sequencer Ion PGM (Thermo Fisher Scientific). In total, 8,914,863 single reads with an average length of 247.35 bp were obtained. After these single reads were filtered with a Phred score cut-off of < 20, the genome was *de novo* assembled using the CLC Genomics Workbench (Qiagen, Valencia, CA, USA).

2.2. Gene prediction

Protein coding genes were predicted with AUGUSTUS [8,9] using *Fusarium graminearum* and *Magnaporthe oryzae* as species parameters and annotated with FunctionAnnotator [10] and BlastKOALA [11]. Secretory proteins were classified using SignalP 4.1 [12] and then used to predict effector proteins through EffectorP [13]. Prediction of RNA-encoding genes was performed for tRNAs using tRNAScan-SE [14] and for rRNAs using HMMER 3.0 [15].

2.3. Phylogenetic analysis

The rRNA, histone H3 and chitin synthase genes of *C. sansevieriae* Sa-1–2 were aligned with other *Colletotrichum* spp. sequences deposited in the DDBJ/EMBL/GenBank database using ClustalX 2.0.5 [16]. A phylogenetic tree was constructed using ClustalX 2.0.5 and viewed by NJ plot [17] (Fig. 1).

Transparency document. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2018.03.083.

References

[1] M. Nakamura, M. Ohzono, H. Iwai, K. Arai, Anthracnose of *Sansevieria trifasciata* caused by *Colletotrichum sansevieriae* sp. nov, J. Gen. Plant Pathol. 72 (2006) 253–256.
[2] K. Kawase, Y. Tsukamoto, Sansevieria Thunb. nom. cons, in: Y. Tsukamoto (Ed.), Encyclopedia of Horticultural Plants, 1, Shogakukan, Tokyo, 1994, pp. 1098–1100.
[3] J. Sperschneider, D.M. Gardiner, P.N. Dodds, F. Tini, L. Covarelli, K.B. Singh, J.M. Manners, J.M. Taylor, EffectorP: predicting fungal effector proteins from secretomes using machine learning, New Phytol. 210 (2015) 743–761.
[4] P.F. Cannon, U. Damm, P.R. Johnston, B.S. Weir, *Colletotrichum* – current status and future directions, Stud. Mycol. 73 (2012) 181–213.
[5] P.W. Crous, M.J. Wingfield, J. Guarro, R. Cheewangkoon, M. van der Bank, W.J. Swart, A.M. Stchigel, J.F. Cano-Lira, J. Roux, N. Madrid, et al., Fungal Planet description sheets 154-213: higher order classification of taxonomic novelties, Persoonia 31 (2013) 186–296.
[6] P.W. Crous, M.J. Wingfield, D.M. Richardson, J.J. Le Roux, D. Strasberg, J. Edwards, F. Roets, V. Hubka, P.W.J. Taylor, M. Heykoop, et al., Fungal Planet description sheets: 400–468, Persoonia 36 (2016) 316–458.
[7] P.W. Crous, M.J. Wingfield, J. Guarro, M. Hernandez-Restrepo, D.A. Sutton, K. Acharya, P.A. Barber, T. Boekhout, R.A. Dimitrov, M. Duenas, Fungal Planet description sheets: 320-370, Persoonia 34 (2015) 167–266.
[8] K.J. Hoff, M. Stanke, WebAUGUSTUS - a web service for training AUGUSTUS and predicting genes in eukaryotes, Nucleic Acids Res. (2013) http://dx.doi.org/10.1093/nar/gkt418.
[9] K.J. Hoff, M. Stanke, TrainAUGUSTUS - a Webserver Application for Parameter Training and Gene Prediction in Eukaryoties International Plant & Animal XX Conference, 2012, U.S.A. (Poster).
[10] T.W. Chen, R.C. Gan, Y.K. Fang, K.Y. Chien, W.C. Liao, C.C. Chen, T.H. Wu, I.Y.F. Chang, C. Yang, P.J. Huang, Y.M. Yeh, C.H. Chiu, T.W. Huang, P. Tang, Function Annotator, a Versatile and Efficient Web Tool for Non-model Organism Annotation, 7, 2017, p. 10430.
[11] M. Kanehisa, Y. Sato, K. Morishima, BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences, J. Mol. Biol. 428 (2016) 726–731.
[12] T.N. Petersen, S. Brunak, G.V. Heijne, H. Nielsen, SignalP 4.0: discriminating signal peptides from transmembrane regions, Nat. Methods 8 (2011) 785–786.
[13] J. Sperschneider, D.M. Gardiner, P.N. Dodds, F. Tini, L. Covarelli, K.B. Singh, J.M. Manners, J.M. Taylor, EffectorP: predicting fungal effector proteins from secretomes using machine learning, New Phytol. 210 (2015) 743–761.
[14] T.M. Lowe, P.P. Chan, tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes, Nucleic Acids Res. (2016) http://dx.doi.org/10.1093/nar/gkw413.
[15] S.R. Eddy, A new generation of homology search tools based on probabilistic inference, Genome Inform. 23 (2009) 205–211.

[16] M.A. Larkin, G. Blackshelds, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, D.G. Higgins, W. Clustal, X. Clustal, version 2.0, Bioinformatics 23 (2007) 2947–2948.
[17] G. Perriére, M. Gouy, WWW-query: an on-line retrieval system for biological sequence banks, Biochimie 78 (1996) 364–369.