Genome Sequence Resource for *Elsinoë ampelina*, the Causal Organism of Grapevine Anthracnose

Zhi Li,^{1,2} Yanchun Fan,³ Pingping Chang,^{1,2} Linlin Gao,^{1,2} and Xiping Wang^{1,2,†}

¹ State Key Laboratory of Crop Stress Biology in Arid Areas, College of Horticulture, Northwest A&F University, Yangling, Shaanxi 712100, China

² Key Laboratory of Horticultural Plant Biology and Germplasm Innovation in Northwest China, Ministry of Agriculture, Northwest A&F University

³ College of Forestry, Northwest A&F University

Abstract

Elsinoë ampelina is an ascomycetous fungus that causes grape anthracnose, a potentially devastating disease worldwide. Here, we report a 28.29 Mb high-quality genome sequence of *E. ampelina* YL-1 that encodes 8,057 predicted protein-coding genes and represents the first sequenced genome assembly of *E. ampelina*. This study adds to the current genomic resources for the genus *Elsinoë* and paves the way for research on comparative genomic studies, *E. ampelina*–grape interactions, and improvement of management strategies.

Funding

This work was supported by the National Natural Science Foundation of China (31501740 and 31272136) and the Program for Innovative Research Team of Grape Germplasm Resources and Breeding grant number 2013KCT-25.

Keywords

Elsinoë, genome, fungal pathogenesis

Genome Announcement

Genus *Elsinoë* is now known to contain 75 species from 67 host genera (Fan et al. 2017). Among them, some species of *Elsinoë* cause anthracnose on economically important crops such as grapevines, citrus, apple, pear, and avocado (Li et al. 2019). Disease symptoms are characterized by a scab-like appearance with distortions and cracks of infected tissues. In many cases, the main impact is on the appearance of the harvested product. Little is known about the genome of most *Elsinoë* species. To date, only two species, *Elsinoë fawcettii* and *Elsinoë australis*, have been sequenced (Shanmugam et al. 2020).

Grape anthracnose, which is caused by *E. ampelina*, is an economically relevant fungal disease on cultivated *Vitis* species worldwide. *E. ampelina* belongs to the class Dothideomycetes, subclass Dothideomycetidae, order Myriangiales, family Elsinoaceae, and its teleomorph is anamorph *Sphaceloma ampelinum* de Bary. *E. ampelina* infects leaves, berries, tendrils, rachises, petioles, and stems of the grapevine and causes more damage on younger tissues (Santos et al. 2018). Typically, anthracnose symptoms are characterized by small necrotic spots (Thind 2015). *E. ampelina* causes serious losses on fruit quality and yield, especially in areas with high humidity (Magarey et al. 1993). Therefore, revealing the *E. ampelina* genome sequence is a fundamental step toward increasing our understanding of grape–*E. ampelina* interactions and provides a reference for facilitating both pathogenicity mechanism studies and potential development of disease control methods.

†Corresponding author: X. Wang; wangxiping@nwsuaf.edu.cn

*The *e-Xtra* logo stands for “electronic extra” and indicates that three supplementary figures and five supplementary tables are published online.

The author(s) declare no conflict of interest.

Accepted for publication 31 January 2020.

Table 1. Features of the *Elsinoë ampelina* YL-1 genome

Genome features	Statistics
Genome size (bp)	28,296,786
Gene number	8,057
Gene length (bp)	11,958,416
GC content of coding DNA (%)	55.3
Genes (% of total genome)	42.26
Gene average length (bp)	1,484
Gene internal length (bp)	16,338,370
Gene internal GC Content (%)	45.29
Gene internal (% of total genome)	57.74
Transfer RNA number	30
Ribosomal RNA number	17
Small nuclear RNA number	13
Repetitive sequences (%)	3.06
Contig	13
Maximum contig length (Mb)	3.8
N50 contig length (Mb)	2.7

We report the genome sequencing and assembly of the *E. ampelina* YL-1 strain, which was isolated from diseased leaves of three-year-old *Vitis vinifera* cv. Red Globe, was purified by single-spore isolation, was confirmed by the internal transcribed spacer region (accession number MG994971.1), and was maintained on potato dextrose agar (PDA) media at 25°C in constant darkness, as previously described (Li et al. 2018). The *E. ampelina* mycelia were collected from 7-day-old PDA plates and were stored at –80°C prior to DNA extraction. Genomic DNA was extracted from mycelia using the cetyl trimethyl ammonium bromide (CTAB) method (Weiland 1997). Genome of *E. ampelina* isolate YL-1 was sequenced using a combination of Illumina short-read and PacBio long-read sequencing technologies, and sequencing was performed at the Beijing Novogene Bioinformatics Technology Co., Ltd (Beijing, China). For short-read sequencing, one Illumina library with an insertion size of 350 bp was constructed and quantified using the Agilent 2100 Bioanalyzer (Agilent Technologies) and was sequenced on the Illumina HiSeq2500 platform. For long-read sequencing, a 20-kb insert library was constructed following the standard single-molecule real-time (SMRT) cell-construction protocol and was sequenced with PacBio RS II SMRT cells. Low-quality reads were filtered out using SMRT Link v5.0.1 software and the remaining reads were assembled to generate contigs without gaps (Li et al. 2010). Finally, the PacBio assembly was polished using the Illumina short reads to correct base errors.

Genome annotation and functional annotation included the following. i) For genome annotation, the Augustus 2.7 program (Stanke et al. 2008) was used to predict protein-coding genes. The repetitive sequences were identified in the assembly using RepeatMasker (Saha et al. 2008). Tandem repeats were detected by Tandem Repeat Finder (Benson 1999). Transfer RNA (tRNA) and ribosomal RNA (rRNA) genes were predicted with the tRNAscan-SE program (Lowe and Eddy 1997) and rRNAmmer (Lagesen et al. 2007), respectively. ii) For functional annotation of predicted protein-coding genes, we compared their protein against the gene ontology (GO) (Ashburner et al. 2000), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al. 2006), EuKaryotic Orthologous Groups (KOG) (Koonin et al. 2004), GenBank nonredundant (nr) (Pruitt et al. 2005), Transporter Classification Database (TCDB) (Saier et al. 2006), and Swiss-Prot (Bairoch and Apweiler 2000) databases. iii) To identify virulence-associated genes, carbohydrate-active enzymes (CAZymes) were predicted by the CAZymes database. Secondary metabolism gene clusters were annotated using antiSMASH (Blin et al. 2017). Fungal virulence genes were determined with the database of fungal virulence factors (DFVF) (Lu et al. 2012).

A total of 7.17 Gb of PacBio long reads, representing 253.7× coverage of the *E. ampelina* genome, and 1.19 Gb of Illumina short reads (38.5×) were generated. The final genome assembly comprised 28.29 Mb containing 13 contigs, with an N50 length of 2.7 Mb and a maximum length of 3.8 Mb (Supplementary Fig. S1; Supplementary Table S1). The proportion of repeat sequences was 3.06%, of which interspersed repeats and transposable elements accounted for 1.89%, while 1.17% was predicted to be tandem repeats (Supplementary Table S2).

A total of 8,057 protein-coding genes were identified, with an average length of 1,484 bp, with the predicted genes comprising 42% of the assembled genome (Table 1). General feature format of *E. ampelina* protein-coding gene models are listed in Supplementary Table S3. Among these genes, 367 (5%) were mapped to the TCDB, 1,902 (24%) were classified using the KOG database, 2,696 (34%) were mapped to the Swiss-Prot database, 5,308 (66%) were assigned GO terms, 6,118 (76%) were mapped to the KEGG database, and 6,399 (79%) were mapped to the National Center for Biotechnology Information nr database (Supplementary Figs. S2 and S3; Supplementary Table S4).

CAZymes are important components in the arsenal of enzymes used by pathogenic fungi to digest polysaccharides (Yin et al. 2012). A total of 407 proteins were predicted as coding for CAZymes from all six major CAzyme superfamilies, 203 (49.88%) glycosyl hydrolases, 57 (14.01%) glycoside transferases, 52 (12.78%) auxiliary activities, 48 (11.79%) carbohydrate-binding modules, 33 (8.11%) carbohydrate esterases, 14 (3.44%) polysaccharide lyases in the *E. ampelina* genome. Additionally, plant-pathogenic fungi can produce toxic secondary metabolites (SMs) and genes involved in SM biosynthesis are often located in clusters (Khaldi et al. 2010). In total, 20 SM clusters were identified in the *E. ampelina* genome, including six nonribosomal peptide synthetase (NRPS) clusters, three polyketide synthase (PKS) clusters, four terpene clusters, one PKS-NRPS cluster, and six other clusters. The annotation analysis predicted 385 genes defined by DFVF with involvement in pathogenesis (Supplementary Table S5).

The genome presented here is the first available genome sequence of *E. ampelina*. The sequence of the *E. ampelina* YL-1 genome can be used in intraspecies and interspecies comparative genomic studies. Moreover, this new data should help elucidate the molecular mechanisms underlying genome evolution, host adaptation, and pathogenesis of *E. ampelina*. The YL-1 genome sequence data has been deposited at DDBJ/ENA/GenBank database under the accession number SMYM00000000 (BioProject: PRJNA526997, BioSample: SAMN11125508).

Author-Recommended Internet Resource

RepeatMasker program: <http://www.repeatmasker.org>

Literature Cited

- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. 2000. Gene ontology: Tool for the unification of biology. *Nat. Genet.* 25:25-29.
- Bairoch, A., and Apweiler, R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* 28:45-48.
- Benson, G. 1999. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* 27:573-580.
- Blin, K., Medema, M. H., Kottmann, R., Lee, S. Y., and Weber, T. 2017. The antiSMASH database, a comprehensive database of microbial secondary metabolite biosynthetic gene clusters. *Nucleic Acids Res.* 45 (D1): D555-D559.
- Fan, X. L., Barreto, R. W., Groenewald, J. Z., Bezerra, J. D. P., Pereira, O. L., Cheewangkoon, R., Mostert, L., Tian, C. M., and Crous, P. W. 2017. Phylogeny and taxonomy of the scab and spot anthracnose fungus *Elsinoë* (*Myriangiales*, *Dothideomycetes*). *Stud. Mycol.* 87:1-41.
- Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., and Hirakawa, M. 2006. From genomics to chemical genomics: New developments in KEGG. *Nucleic Acids Res.* 34:D354-D357.
- Khaldi, N., Seifuddin, F. T., Turner, G., Haft, D., Nierman, W. C., Wolfe, K. H., and Fedorova, N. D. 2010. SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genet. Biol.* 47:736-741.
- Koonin, E. V., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Krylov, D. M., Makarova, K. S., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Rogozin, I. B., Smirnov, S., Sorokin, A. V., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J., and Natale, D. A. 2004. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* 5:R7.
- Lagesen, K., Hallin, P., Rødland, E. A., Staerfeldt, H. H., Rognes, T., and Ussery, D. W. 2007. RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35:3100-3108.
- Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., Li, S., Yang, H., Wang, J., and Wang, J. 2010. *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20:265-272.
- Li, Z., Dang, H., Yuan, X., He, J., Hu, Z., and Wang, X. 2018. Morphological characterization and optimization of conditions for conidial production of *Elsinoë ampelina*, the causal organism of grapevine anthracnose. *J. Phytopathol.* 166:420-428.
- Li, Z., Zhang, S., Han, R., Zhang, H., Li, K., and Wang, X. 2019. Infection process and host responses to *Elsinoë ampelina*, the causal organism of grapevine anthracnose. *Eur. J. Plant Pathol.* 155:571-582.
- Lowe, T. M., and Eddy, S. R. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25:955-964.
- Lu, T., Yao, B., and Zhang, C. 2012. DFVF: Database of fungal virulence factors. *Database (Oxford)* 2012:bas032.
- Magarey, R. D., Emmett, R. W., Magarey, P. A., and Franz, P. R. 1993. Evaluation of control of grapevine anthracnose caused by *Elsinoë ampelina* by pre-infection fungicides. *Australas. Plant Pathol.* 22:48-52.
- Pruitt, K. D., Tatusova, T., and Maglott, D. R. 2005. NCBI reference sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* 33:D501-D504.
- Saha, S., Bridges, S., Magbanua, Z. V., and Peterson, D. G. 2008. Empirical comparison of *ab initio* repeat finding programs. *Nucleic Acids Res.* 36:2284-2294.
- Saier, M. H., Jr., Tran, C. V., and Barabote, R. D. 2006. TCDB: The transporter classification database for membrane transport protein analyses and information. *Nucleic Acids Res.* 34:D181-D186.

- Santos, R. F., Ciampi-Guillardi, M., Amorim, L., Massola Júnior, N. S., and Spósito, M. B. 2018. Aetiology of anthracnose on grapevine shoots in Brazil. *Plant Pathol.* 67:692-706.
- Shanmugam, G., Jeon, J., and Hyun, J. W. 2020. Draft Genome Sequences of *Elsinoë fawcettii* and *Elsinoë australis* causing scab diseases on citrus. *Mol. Plant-Microbe Interact.* 33:135-137.
- Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. 2008. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* 24:637-644.
- Thind, T. 2015. Anthracnose. Pages 17-19 in: *Compendium of Grape Diseases, Disorders, and Pests*, 2nd Ed., W. Wilcox, W. Gubler, and J. Uyemoto, eds. APS Press, St Paul, MN, U.S.A.
- Weiland, J. J. 1997. Rapid procedure for the extraction of DNA from fungal spores and mycelia. *Fungal Genet. Rep.* 44:60-63.
- Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. 2012. dbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 40 (W1):W445-W451.