

# Chromosome-Level Reference Genome of *Venturia effusa*, Causative Agent of Pecan Scab

David J. Winter,<sup>1</sup> Nikki D. Charlton,<sup>2</sup> Nick Krom,<sup>2</sup> Jason Shiller,<sup>2</sup> Clive H. Bock,<sup>3</sup> Murray P. Cox,<sup>1,†</sup> and Carolyn A. Young<sup>2,†</sup>

<sup>1</sup> School of Fundamental Sciences and the Bio-Protection Research Centre, Massey University, Palmerston North 4442, New Zealand

<sup>2</sup> Noble Research Institute, LLC, Ardmore, OK 73401, U.S.A.

<sup>3</sup> United States Department of Agriculture–Agricultural Research Service Southeastern Fruit and Tree Nut Research Laboratory, Byron, GA 31008, U.S.A.

## Abstract

Pecan scab, caused by *Venturia effusa*, is a devastating disease of pecan (*Carya illinoensis*), which results in economic losses on susceptible cultivars throughout the southeastern United States. To enhance our understanding of pathogenicity in *V. effusa*, we have generated a complete telomere-to-telomere reference genome of *V. effusa* isolate FRT5LL7-Albino. By combining Illumina MiSeq and Oxford Nanopore MinION data, we assembled a 45.2-Mb genome represented by 20 chromosomes and containing 10,820 putative genes, of which 7,619 have at least one functional annotation. The likely causative mutation of the albino phenotype was identified as a single base insertion and a resulting frameshift in the gene encoding the polyketide synthase *ALM1*. This genome represents the first full chromosome-level assembly of any *Venturia* sp.

## Genome Announcement

*Venturia* spp. are widespread fungal pathogens that cause scab disease on fruit trees (González-Domínguez et al. 2017)—on apple (*Venturia inaequalis*), pear (*V. pyrina* and *V. nashicola*), peach and almond (*V. carpophila*), and olive (*V. oleaginea*)—as well as on pecan (*V. effusa*). Pecan scab is polycyclic during the growing season and most prominent on leaves, twigs, and shucks. Pecan scab-susceptible cultivars are managed by fungicide application throughout the growing season to reduce economic losses due to reduced nut size and number, and poor nutmeat quality as a result of the disease (Bock et al. 2017a). *V. effusa* was previously assumed to reproduce solely asexually but the recent identification of a heterothallic mating system and the fact that the mating types are in equilibrium throughout orchards in the southeastern United States indicates that a sexual stage may contribute to the disease cycle (Bock et al. 2018; Young et al. 2018). At present, there are

D. J. Winter and N. D. Charlton contributed equally to the work.

†Corresponding authors: C. A. Young; [cayoung@noble.org](mailto:cayoung@noble.org); and M. P. Cox; [m.p.cox@massey.ac.nz](mailto:m.p.cox@massey.ac.nz)

\*The e-Xtra logo stands for “electronic extra” and indicates that one supplementary figure and one supplementary table are published online.

The author(s) declare no conflict of interest.

Accepted for publication 17 October 2019.

## Funding

This study was supported by the Noble Research Institute, LLC. C. H. Bock was supported by the USDA-ARS National Programs through CRIS project 6042-21220-012-00.

## Keywords

genomics, metabolomics, proteomics

more than 90 genome assemblies available for *Venturia* spp., representing an active research area of fungal genomics. However, many of the genomes are highly fragmented, ranging from 66 scaffolds to more than 40,000 (Bock et al. 2016b; Chen et al. 2017; Deng et al. 2017; Johnson et al. 2019; Le Cam et al. 2019; Passey et al. 2018; Prokhorchik et al. 2019). Here, we present the first entire and finished genome sequence for the genus *Venturia*.

Isolate FRT5LL7 was originally collected on 23 June 2015 from a commercial orchard (Graham Pecan Farm) of the pecan cultivar Desirable located west of Albany, GA. The pecan orchard received fungicide applications typical for pecan management in Georgia (Wells 2019). FRT5LL7 was isolated from a lesion on a leaflet in the lower tree canopy (5 to 8 m). FRT5LL7 was a regular melanized scab isolate that spontaneously produced an albino sector once cultured on potato dextrose agar. The albino form of FRT5LL7 (FRT5LL7-Albino) was isolated on 4 September 2015 and maintained as a separate culture. FRT5LL7-Albino was chosen for genome sequencing because the quality of the extracted DNA was superior to that of melanized isolates. In addition, FRT5LL7-Albino has been used in sexual crosses, where it has provided a useful phenotype for determining recombination frequencies during sexual crosses (Charlton et al. 2019). No other phenotypic anomalies of FRT5LL7-Albino have been noted but pathogenicity of this isolate has not yet been tested.

The FRT5LL7-Albino genome was sequenced using both short- and long-read chemistry with Illumina MiSeq and Oxford Nanopore MinION, respectively. Genomic DNA was extracted from lyophilized mycelium using the Quick Fungal/Bacterial miniprep kit (mycelial pulverization with the beads was omitted) (ZymoResearch, Irvine, CA, U.S.A.) for Illumina MiSeq sequencing by the Bluegrass Technical and Community College (Lexington, KY, U.S.A.). The Taha DNA extraction method (Al-Samarrai and Schmid 2000) was used on the same mycelium for Oxford Nanopore MinION sequencing. DNA samples were purified using a SPRI bead mixture (Schalamun and Schwessinger 2017) to reduce the number of short DNA fragments. The library for MinION sequencing was prepared using the 1D library SQK-LSK09 kit and sequenced on one R9.4 MinION flowcell with the basecalling option turned on.

Initial assemblies were produced from the basecalled nanopore data using two programs with alternative assembly algorithms, CANU v1.8 (Koren et al. 2017) and NECAT v20190307 (Xiao et al. 2017). The different algorithms complimented each other, with discontinuities present in one assembly being well resolved in the other. The final assembly was polished with Pilon v1.23 (Walker et al. 2014), using 250 bp PE Illumina MiSeq reads as input. The assembly of FRT5LL7-Albino contains 20 complete chromosomes totaling 45.2 Mb of sequence and a circular mitochondrial genome (139 kb). This is a finished reference genome; it stretches from telomere to telomere and contains no ambiguity bases or gaps. The 20 chromosomes range in size from the largest at 4.12 Mb to the smallest at 0.56 Mb (Supplementary Fig. S1). The FRT5LL7-Albino assembled genome is a vast improvement over the earlier sequenced genome of *V. effusa* isolate 3Des10b, which consists of 40.6 Mb represented by 545 scaffolds (Bock et al. 2016b).

The genome was annotated using the funannotate pipeline (v1.4) (Palmer 2018), which makes use of AUGUSTUS v3.3.1 (Stanke et al. 2008) for gene calling. Having identified putative genes, funannotate uses InterProScan v5.22 (Jones et al. 2014), antiSMASH v4.20 (Blin et al. 2017), MEROPS (Rawlings et al. 2008), Pfam (Bateman et al. 2004), and dbCAN (Huang et al. 2018) databases to generate functional annotation. In addition, we identified putative effector genes using SignalP v4.1 (Petersen et al. 2011) and EffectorP v2.0 (Sperschneider et al. 2016), and located AT-rich regions of the genome using OcculterCut v1.1 (Testa et al. 2016). Our annotated genome contains 10,820 putative protein coding genes, of which 7,619 have at least one functional annotation. This includes 4,749 genes associated with Gene Ontology terms, 426 putative effectors, and 207 genes associated with secondary metabolite clusters (Table 1). The BUSCO score (Simão et al. 2015; Waterhouse et al. 2017) was 96.8%, with 1,272 complete BUSCOs of 1,315 identified from the Ascomycota database. The full functional annotation of the genome is included in Supplementary Table S1.

The higher-level structure of the genome is characterized by large blocks of AT-rich sequences interspersed between regions of approximately equal nucleotide content. The AT-rich regions are extremely gene poor, containing only 0.15% of the protein coding genes despite comprising 24.7% of the genome. This is similar to the genome structure seen in other *Venturia* spp., most notably the genomes of *V. inaequalis* which, on average, have 45.3% AT-rich sequence (Le Cam et al. 2019). This type of genome structure was first described in fungi

**Table 1.** Gene annotation statistics<sup>a</sup>

Annotation	Genes annotated
Gene	10,822
InterPro	7,619
PFAM	6,777
GO function	4,749
BUSCO	1,273
Signal peptide	1,130
Product	583
Name	564
EffectorP	426
CAZy	405
MEROPS	320
antiSMASH	207
SMCOG	49

<sup>a</sup> “Annotation” refers to the category of functional annotation provided by the funannotate pipeline. “SMCOG” refers to genes predicted to have secondary metabolite activities, while “antiSMASH” includes all genes that form part of a putative secondary metabolite cluster. “Name” and “Product” signify genes that have been assigned an NCBI gene name or protein product based on similarity to well-characterized genes in the UniProtKb database. The “Genes annotated” column provides the number of genes having at least one annotation of each type in our dataset.

for *Leptosphaeria maculans* (Rouxel et al. 2011). Subsequently, the evolution of effector genes underlying gene-for-gene interactions in the *L. maculans*–*Brassica napus* pathosystem has been shown to be associated with their proximity to or position within AT-rich regions (Rouxel et al. 2011; Van de Wouw et al. 2010). Gene-for-gene interactions have been described for other *Venturia* spp. (Bus et al. 2011; Deng et al. 2017) and effectors are likely to be important for the *V. effusa*–*Carya illinoensis* interaction. A gapless genome assembly of *V. effusa* spanning complete AT-rich regions is crucial to identifying and understanding the evolution of effector genes in the species. The sequence reported here will also underpin future work using RNAseq data and advanced computational techniques such as Coding-Quarry (Testa et al. 2015) and GeMoMa (Keilwagen et al. 2018) to identify these difficult-to-study genes. Key features of the genome, including AT-rich regions and the location of important genes, are illustrated in Supplementary Figure S1.

The albino phenotype is often due to mutation of a melanin biosynthesis gene. Evaluation of the putative melanin biosynthesis gene cluster found on chromosome 6 in FRT5LL7-Albino (Supplementary Fig. 1) revealed a single base insertion and a resulting frameshift in the gene encoding the polyketide synthetase *ALM1*. As found with other melanin-producing fungi, *ALM1* in *V. effusa* is clustered with two genes, *CRM1* and *BRM2*, that encode putative functions in melanin biosynthesis, with an additional gene, *SDH1*, located on chromosome 5 (Fetzner et al. 2014; Kimura and Tsuge 1993).

Although other *Venturia* genomes have been sequenced (Bock et al. 2016b; Chen et al. 2017; Deng et al. 2017; Johnson et al. 2019; Le Cam et al. 2019; Passey et al. 2018; Prokhorchik et al. 2019), the complete genome sequence of *V. effusa* FRT5LL7-Albino represents the first chromosomal-level assembly of a *Venturia* sp. This valuable resource is a basis for scientists in the *Venturia* research community to advance the comparative genomics of scab pathogens; it is a useful resource for the identification of effectors and will provide insights into the epidemiology and population biology of the pathogens and the evolution of the genus *Venturia*.

The genome sequence has been deposited into GenBank under genome accession CP042185 (BioProject PRJNA551043, BioSample SAMN12143997, and SRA accessions SRX6745934 and SRX6745935 for Illumina and Nanopore MinION data, respectively).

## Acknowledgments

We thank M. McDonald (Australian National University) and S. Florea (University of Kentucky) for their excellent advice and guidance before embarking on the MinION sequencing.

## Literature Cited

- Al-Samarrai, T. H., and Schmid, J. 2000. A simple method for extraction of fungal genomic DNA. *Lett. Appl. Microbiol.* 30:53-56.
- Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L. C., Studholme, D. J., Yeats, C., and Eddy, S. R. 2004. The Pfam protein families database. *Nucleic Acids Res.* 32: D138-D141.
- Blin, K., Wolf, T., Chevrette, M. G., Lu, X., Schwalen, C. J., Kautsar, S. A., Suarez Duran, H. G., de Los Santos, E. L. C., Kim, H. U., Nave, M., Dickschat, J. S., Mitchell, D. A., Shelest, E., Breitling, R., Takano, E., Lee, S. Y., Weber, T., and Medema, M. H. 2017. antiSMASH 4.0—Improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res.* 45:W36-W41.
- Bock, C. H., Brenneman, T. B., Wood, B. W., and Stevenson, K. L. 2017a. Challenges of managing disease in tall orchard trees—Pecan scab, a case study. *CAB Rev.* 12:1-18.
- Bock, C. H., Chen, C., Yu, F., Stevenson, K. L., and Wood, B. W. 2016b. Draft genome sequence of *Fusicladium effusum*, cause of pecan scab. *Stand. Genomic Sci.* 11:36.
- Bock, C. H., Young, C. A., Stevenson, K. L., and Charlton, N. D. 2018. Fine-scale population genetic structure and within-tree distribution of mating types of *Venturia effusa*, cause of pecan scab in the United States. *Phytopathology* 108: 1326-1336.
- Bus, V. G., Rikkerink, E. H., Caffier, V., Durel, C.-E., and Plummer, K. M. 2011. Revision of the nomenclature of the differential host-pathogen interactions of *Venturia inaequalis* and *Malus*. *Annu. Rev. Phytopathol.* 49:391-413.
- Charlton, N. D., Yi, M., Bock, C. H., Zhang, M., and Young, C. A. 2019. First description of the sexual stage of *Venturia effusa*, causal agent of pecan scab. *bioRxiv* 785790.
- Chen, C., Bock, C. H., and Wood, B. W. 2017. Draft genome sequence of *Venturia carpophila*, the causal agent of peach scab. *Stand. Genomic Sci.* 12:68.
- Deng, C. H., Plummer, K. M., Jones, D. A. B., Mesarich, C. H., Shiller, J., Taranto, A. P., Robinson, A. J., Kastner, P., Hall, N. E., Templeton, M. D., and Bowen, J. K. 2017. Comparative analysis of the predicted secretomes of Rosaceae scab pathogens *Venturia inaequalis* and *V. pirina* reveals expanded effector families and putative determinants of host range. *BMC Genomics* 18:339.
- Fetzner, R., Seither, K., Wenderoth, M., Herr, A., and Fischer, R. 2014. *Alternaria alternata* transcription factor CmrA controls melanization and spore development. *Microbiology* 160:1845-1854.
- González-Domínguez, E., Armengol, J., and Rossi, V. 2017. Biology and epidemiology of *Venturia* species affecting fruit crops: A review. *Front. Plant Sci.* 8:1496.
- Huang, L., Zhang, H., Wu, P., Entwistle, S., Li, X., Yohe, T., Yi, H., Yang, Z., and Yin, Y. 2018. dbCAN-seq: A database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic Acids Res.* 46:D516-D521.
- Johnson, S., Jones, D., Thrimawithana, A. H., Deng, C. H., Bowen, J. K., Mesarich, C. H., Ishii, H., Won, K., Bus, V. G. M., and Plummer, K. M. 2019. Whole genome sequence resource of the Asian pear scab pathogen, *Venturia nashicola*. *Mol. Plant-Microbe Interact.* 32:1463-1467.
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S. Y., Lopez, R., and Hunter, S. 2014. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* 30:1236-1240.
- Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O., and Grau, J. 2018. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinf.* 19:189.
- Kimura, N., and Tsuge, T. 1993. Gene cluster involved in melanin biosynthesis of the filamentous fungus *Alternaria alternata*. *J. Bacteriol.* 175:4427-4435.
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. 2017. Canu: Scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res.* 27:722-736.
- Le Cam, B., Sargent, D., Gouzy, J., Amselem, J., Bellanger, M.-N., Bouchez, O., Brown, S., Caffier, V., De Gracia, M., Debuchy, R., Duvaux, L., Payen, T., Sannier, M., Shiller, J., Collemare, J., and Lemaire, C. 2019. Population genome sequencing of the scab fungal species *Venturia inaequalis*, *Venturia pirina*, *Venturia aucupariae* and *Venturia asperata*. *G3: Genes Genomes Genet.* 9: 2405-2414.
- Palmer, J. 2018. Funannotate: Eukaryotic Genome Annotation Pipeline. <https://funannotate.readthedocs.io>
- Passey, T. A. J., Armitage, A. D., and Xu, X. 2018. Annotated draft genome sequence of the apple scab pathogen *Venturia inaequalis*. *Microbiol. Resour. Announce.* 7: e01062-18.
- Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* 8: 785-786.
- Prokhorchik, M., Won, K., Lee, Y., Choi, E. D., Segonzac, C., and Sohn, K. H. 2019. High contiguity whole genome sequence and gene annotation resource for two *Venturia nashicola* isolates. *Mol. Plant-Microbe Interact.* 32:1091-1094.
- Rawlings, N. D., Morton, F. R., Kok, C. Y., Kong, J., and Barrett, A. J. 2008. MEROPS: The peptidase database. *Nucleic Acids Res.* 36:D320-D325.
- Rouxel, T., Grandaubert, J., Hane, J. K., Hoede, C., van de Wouw, A. P., Couloux, A., Dominguez, V., Anthouard, V., Bally, P., Bourras, S., Cozijnsen, A. J., Ciuffetti, L. M., Degrave, A., Dilmaghani, A., Duret, L., Fudal, I., Goodwin, S. B., Gout, L., Glaser, N., Linglin, J., Kema, G. H., Lapalu, N., Lawrence, C. B., May, K., Meyer, M., Olivier, B., Poulain, J., Schoch, C. L., Simon, A., Spatafora, J. W., Stachowiak, A., Turgeon, B. G., Tyler, B. M., Vincent, D., Weissenbach, J., Amselem, J., Quesneville, H., Oliver, R. P., Wincker, P., Balesdent, M. H., and Howlett, B. J. 2011. Effector diversification within compartments of the *Leptosphaeria maculans* genome affected by repeat-induced point mutations. *Nat. Commun.* 2:202.
- Schalamun, M., and Schwessinger, B. 2017. DNA size selection (> 1kb) and clean up using an optimized SPRI beads mixture. [protocols.io](https://protocols.io). doi:10.17504/protocols.io.idmca46
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210-3212.
- Sperschneider, J., Gardiner, D. M., Dodds, P. N., Tini, F., Covarelli, L., Singh, K. B., Manners, J. M., and Taylor, J. M. 2016. EffectorP: Predicting fungal effector proteins from secretomes using machine learning. *New Phytol.* 210:743-761.
- Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. 2008. Using native and syntetically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24:637-644.
- Testa, A. C., Hane, J. K., Ellwood, S. R., and Oliver, R. P. 2015. CodingQuarry: Highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. *BMC Genomics* 16:170.
- Testa, A. C., Oliver, R. P., and Hane, J. K. 2016. OcculterCut: A comprehensive survey of AT-rich regions in fungal genomes. *Genome Biol. Evol.* 8:2044-2064.
- Van de Wouw, A. P., Cozijnsen, A. J., Hane, J. K., Brunner, P. C., McDonald, B. A., Oliver, R. P., and Howlett, B. J. 2010. Evolution of linked avirulence effectors in *Leptosphaeria maculans* is affected by genomic environment and exposure to resistance genes in host plants. *PLoS Pathog.* 6:e1001180.
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., and Earl, A. M. 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963.
- Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., and Zdobnov, E. M. 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* 35: 543-548.
- Wells, L. 2019. Commercial Pecan Spray Guide. University of Georgia Extension, University of Georgia Athens, GA, U.S.A.
- Xiao, C.-L., Chen, Y., Xie, S.-Q., Chen, K.-N., Wang, Y., Han, Y., Luo, F., and Xie, Z. 2017. MECAT: Fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *Nat. Methods* 14:1072-1074.
- Young, C. A., Bock, C. H., Charlton, N. D., Mattupalli, C., Krom, N., Bowen, J. K., Templeton, M., Plummer, K. M., and Wood, B. W. 2018. Evidence for sexual reproduction: Identification, frequency, and spatial distribution of *Venturia effusa* (pecan scab) mating type idiomorphs. *Phytopathology* 108:837-846.