Genome Sequence Resource for *Erysiphe necator* NAFU1, a Grapevine Powdery Mildew Isolate Identified in Shaanxi Province of China

Xingyuan Zhang[1,2#], Bo Mu[1,2#], Kaicheng Cui[1,2], Min Liu[1,2], Guihua Ke[1,2], Yongtao Han[1,2], Ying Wu[3], Shunyuan Xiao[3], Ying-Qiang Wen[1,2*]

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

[2] Key Laboratory of Horticultural Plant Biology and Germplasm Innovation in Northwest China, Ministry of Agriculture and Rural Affairs, Yangling, 712100, Shaanxi, China

[3] Institute for Bioscience and Biotechnology Research & Department of Plant Sciences and Landscape Architecture, University of Maryland College Park, Rockville, MD, 20850 USA

[*]Corresponding author: wenyq@nwsuaf.edu.cn (Y.-Q. Wen)

[#]These authors contribute equally to this study.

Key words: *Erysiphe necator*, grapevine, genome, powdery mildew

Funding**:** This work was supported by the National Natural Science Foundation of China (Grant No. 31772264, 31972986) and the National Key Research and Development Program of China (2018YFD1000300) to YQW and NSF support (IOS-1901566) to S.X.

Abstract:

Erysiphe necator is an economically important biotrophic fungal pathogen responsible for powdery mildew disease on grapevine. Currently, genome sequences are available for only a few *Erysiphe necator* isolates from USA. Based on the combination of Nanopore and Illumina sequencing technologies, we present here the complete genome assembly for an isolate of *En.* NAFU1 identified in China. We acquired a total of 15.93 Gb raw reads. These reads were processed into a 61.12 Mb genome assembly containing 73 contigs with the N50 of 2.06 Mb and a maximum length of 6.05 Mb. Combining the results of three gene-prediction modules, i.e. an evidence-based gene modeler (EVidenceModeler or EVM), an ab initio gene modeler, and a homology-based gene modeler, we predicted 7235 protein-coding genes in the assembled genome of *En.* NAFU1. This information will facilitate studies of genome evolution and pathogenicity mechanisms of *E. necator* and other powdery mildew species through comparative genome sequence analysis and other

molecular genetic tools.

Key words: *Erysiphe necator*, grapevine, genome, powdery mildew

Genome Announcement

Powdery mildew caused by ascomycete fungi in the order of Erysiphales is an important and widespread disease of numerous plant species including wheat, barley, bean, rose, rubber, tomato, strawberry and grape (Braun and Cook 2012; Wu et al. 2018). The Eurasian grape, *Vitis vinifera* L., is the most widely cultivated and economically important fruit crop in the world. However, the Eurasian grape is susceptible to many oomycete and fungal diseases, including downy mildew caused by *Plasmopora viticola*, grey mold caused by *Botrytis cinerea*, and powdery mildew caused by *Erysiphe necator* (Gadoury et al. 2012; Gessler et al. 2011). Among these three types of pathogens, *E. necator* does not require specific humidity and temperature conditions for infection (Dry et al. 2010), therefore powdery mildew is the most frequent disease of grapevines in many areas. *E. necator* can infect all green tissues of a grapevine. Like in other plants, grapevine leaves infected by powdery mildew

64 often show reduced photosynthesis and suffer from premature senescence and
65 abscission (Han et al. 2016). Even a low-level infection can reduce the quality
66 of berries, which may affect the flavor of wine, table or raisin grapes. Severe
67 infection can cause berry cracking and dropping, or even result in significant
68 loss of harvest (Gadoury et al. 2003; Qiu et al. 2015). Despite its importance,
69 relatively little information is currently available regarding pathogenicity
70 mechanisms of *E. necator* as well as other powdery mildew pathogens. Apart
71 from the genetic intractability of all powdery mildew fungi, the lack of well-
72 assembled genomes for the identification of key effectors of host-adapted
73 powdery mildew (sub) species or isolates also hinders mechanistic studies of
74 powdery mildew.

75 To date, the whole-genome sequences of five isolates of *Erysiphe necator*
76 (Branching, C-strain, e1-101, Lodi, and Ranch9) have been reported. There are
77 apparent sequence differences among the genomes of the five powdery mildew
78 isolates with different genetic and geographical backgrounds (Jones et al.
79 2014). Another report indicated that there is sequence polymorphism between
80 isolates from different regions and hosts in the eastern USA, let alone when the
81 US isolates are compared with those from southern France and Italy (Brewer
82 et al 2010; Frenkel et al. 2012). Viticulture has a long history in China and there
83 is rich wild grapevine germplasm (Gao et al. 2016) in many different regions of
84 China. It is conceivable that the long-time grapevine-powdery mildew co-

evolution must have shaped the genomes of powdery mildew pathogens in these regions. Hence, obtaining the whole-genome sequence of a grapevine powdery mildew isolate identified in China would provide valuable sequence information for future investigation of host-adaptation of powdery mildew in different regions.

Here, we reported the genome sequence of an *Erysiphe necator* isolate NAFU1 (*En.* NAFU1), which was isolated from *Vitis vinifera* cv. Rizamat, Shaanxi Province in China and maintained on the susceptible grapevine 'Thompson Seedless' (Gao et al. 2016). Spores *En.* NAFU1 were collected from infected leaves of grapevine at 10-15 days post-inoculation using a small vacuum and used for genomic DNA extraction using the CTAB method (Feehan et al. 2017). About 12 µg of pure DNA with an average size of 20 Kb was used for genome sequencing by the Oxford Nanopore and Illumina technologies. The genome sequences of *En.* NAFU1 were assembled by using a combination of short-reads (~6,657,136,974 bp) generated by Illumina sequencing (NOVASeq6000 platform, PE150, read length 150 bp, paired-end reads) and long-reads (~13,268,751,137 bp) by Oxford Nanopore sequencing (PromethION) performed at the Biomarker Technologies (Beijing, China). A total of 13.27 Gb of Nanopore long reads, representing ~217x coverage of the *En*. NAFU1 genome, and 6.66 Gb of Illumina NOVASeq6000 short reads (~109x) were generated. The Nanopore reads of low quality and less than 2,000

bp were filtered out. NECAT was used to assemble the Nanopore subreads after filtering, and Pilon was used to correct the assembled sequences using the second-generation sequence data (Koren et al. 2017; Walker et al. 2014).

The final genome assembly of *En*. NAFU1 is 61.12 Mb in length with 48.5% GC content. The assembly contains 73 contigs with the N50 length of 2.06 Mb (the longest contig length is 6.05 Mb) and 98.3% BUSCO completeness (based on 1315 conserved Ascomycota orthologs). This indicates that the whole-genome assembly of *En.* NAFU1 is of high quality, which ensures accurate prediction of the protein-coding genes in its genome. Genscan (Chris and Samuel 1997), Augustus v2.4 (Stanke and Waack 2003), GlimmerHMM v3.0.4 (Majoros et al. 2004), GeneID v1.4 (Blanco et al. 2007), and SNAP (version 2006-07-28) (Ian 2004) were used for ab initio gene prediction and GeMoMa v1.3.1 (Keilwagen et al. 2016) was used for homology-based gene prediction. Finally, EVM v1.1.1 (Haas et al. 2008) was used to integrate the above two methods to obtain 7235 protein-coding genes. In total, 7,235 protein-coding genes were predicted with an estimated BUSCO completeness being 96.7%, and average length of protein-coding gene being 2169 bp. The sequence features of the genome assembly of *En.* NAFU1 and other isolates were shown in Table 1. A comparison with the assembled genomes of other isolates (Jones et al. 2014) suggests a much deeper sequence depth and higher genome coverage, with fewer contigs for that of *En.* NAFU1.

To ensure a successful infection on their host plants, fungal pathogens produce a suite of carbohydrate-active enzymes (CAZymes) to digest polysaccharides of the plant cell wall and send hundreds of secreted effector proteins into the host cell to suppress plant immunity (Adachi et al. 2020; Yin et al. 2012). To assess the size of the CAZymes of *En.* NAFU1, genes encoding such enzymes were predicted by the webtools at the CAZymes database (http://www.cazy.org/). A total of 327 CAZyme genes were predicted in the genome of *E. necator* NAFU1. These CAZymes belong to five superfamilies, including 138 (42.2%) glycosyl hydrolases, 96 (29.35%) glycoside transferases, 53 (16.2%) carbohydrate esterases, 26 (7.95%) enzymes with auxiliary activities, and 14 (4.28%) other carbohydrate-binding proteins. To identify candidate genes encoding secreted proteins, SignalP v3.0 (Bendtsen et al. 2004) was used for the prediction of an N-terminal signal peptide and TMHMM (Krogh et al. 2001) was used for the prediction of transmembrane (TM) domains. A total of 453 genes in the genome were predicted to encode secreted proteins that contain a signal peptide but no transmembrane domains. Using EffectorP (Sperschneider et al. 2018), 41 genes were predicted to encode candidate effector proteins in the genome of *En.* NAFU1.

The availability of the genome sequence of *En.* NAFU1 will facilitate intraspecific as well as interspecific comparative genome analyses of powdery mildew fungi for investigating how they have co-evolved with their respective

148 plant hosts in different habitats. One future study we hope to conduct is to
149 assess the impact of resistance from various wild Chinese grapevines on the
150 effector repertoire of *En.* NAFU1 in comparison with *E. necator* originated in the
151 USA and Europe. The *En*. NAFU1 genome has been deposited at
152 DDBJ/ENA/GenBank database under the accession number
153 JAFBAW000000000 (BioProject: PRJNA695796, BioSample:
154 SAMN17620199). The version described in this paper is version
155 JAFBAW010000000.

Literature Cited

Adachi, H., Biaas, A., and Kamoun, S. 2020. How to trick a plant pathogen? Biochemist 42:14-18.

Bendtsen, J. D., Nielsen, H., von Heijne, G., and Brunak, S. 2004. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340:783-95.

Blanco, E., Parra., G., and Guigó., R. 2007. Using geneid to identify genes. Current Protocols in Bioinformatics 18:4.3.1-4.3.28.

Braun, U., and Cook, R. 2012. "Taxonomic Manual of the Erysiphales (Powdery Mildews)."

Brewer, M.T., Milgroom, M.G. Phylogeography and population structure of the grape powdery mildew fungus, *Erysiphe necator*, from diverse *Vitis*species. *BMC Evol Biol* **10,** 268 (2010).

Chris, B., and Samuel, K. 1997. Prediction of complete gene structures in human genomic DNA. Journal of Molecular Biology 268:78-94.

Dry, I. B., Feechan, A., Anderson, C., Jermakow, A. M., Bouquet, A., Adam-Blondon, A. F., and Thomas, M. R. 2010. Molecular strategies to enhance the genetic resistance of grapevines to powdery mildew. Australian Journal of Grape and Wine Research 16:94-105.

Feehan, J. M., Scheibel, K. E., Bourras, S., Underwood, W., Keller, B., and Somerville, S. C. 2017. Purification of High Molecular Weight Genomic DNA from Powdery Mildew for Long-Read Sequencing. J Vis Exp 121:e55463.

Frenkel, O., Portillo, I., Brewer, M. T., Péros, J. P., Cadle-Davidson, L., and Milgroom, M. G. 2012. Development of microsatellite markers from the transcriptome of Erysiphe necator for analysing population structure in North America and Europe. Plant Pathology 61:106-119.

Gadoury, D. M., Cadle-Davidson, L., Wilcox, W. F., Dry, I. B., Seem, R. C., and Milgroom, M. G. 2012. Grapevine powdery mildew (Erysiphe necator): a fascinating system for the study of the biology, ecology and epidemiology of an obligate biotroph. Mol Plant Pathol 13:1-16.

Gadoury, D. M., Seem, R. C., Ficke, A., and Wilcox, W. F. 2003. Ontogenic resistance to powdery mildew in grape berries. Phytopathology 93:547-55.

Gao, Y. R., Han, Y. T., Zhao, F. L., Li, Y. J., Cheng, Y., Ding, Q., Wang, Y. J., and Wen, Y. Q. 2016. Identification and utilization of a new Erysiphe necator isolate NAFU1 to quickly evaluate powdery mildew resistance in wild Chinese grapevine species using detached leaves. Plant Physiol Biochem 98:12-24.

Gessler, C., Pertot, I., and Perazzolli, M. 2011. Plasmopara viticola: a review of knowledge on downy mildew of grapevine and effective disease management. . Phytopathologia Mediterranea 50:3-44.

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., White, O., Buell, C. R., and Wortman, J. R. 2008. Automated eukaryotic gene structure

196 annotation using EVidenceModeler and the Program to Assemble Spliced
197 Alignments. Genome Biol 9:R7.
198 Han, L., Weng, K., Ma, H., Xiang, G., Li, Z., Wang, Y., Liu, G., and Xu, Y. 2016.
199 Identification and Characterization of Erysiphe necator-Responsive
200 MicroRNAs in Chinese Wild Vitis pseudoreticulata by High-Throughput
201 Sequencing. Front Plant Sci 7:621.
202 Ian, K. 2004. Gene finding in novel genomes. BMC Bioinformatics 5:59.
203 Jones, L., Riaz, S., Morales-Cruz, A., Amrine, K. C., Mcguire, B., and Gubler, W. 2014.
204 Adaptive genomic structural variation in the grape powdery mildew pathogen,
205 erysiphe necator. BMC Genomics 15:1081.
206 Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., and Hartung, F.
207 2016. Using intron position conservation for homology-based gene prediction.
208 Nucleic Acids Res 44:e89.
209 Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A.
210 M. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer
211 weighting and repeat separation. Genome Res 27:722-736.
212 Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. 2001. Predicting
213 transmembrane protein topology with a hidden Markov model: application to
214 complete genomes. J Mol Biol 305:567-80.
215 Majoros, W. H., Pertea, M., and Salzberg, S. L. 2004. TigrScan and GlimmerHMM:
216 two open source ab initio eukaryotic gene-finders. Bioinformatics 20:2878-9.
217 Qiu, W., Feechan, A., and Dry, I. 2015. Current understanding of grapevine defense
218 mechanisms against the biotrophic fungus (Erysiphe necator), the causal agent
219 of powdery mildew disease. Hortic Res 2:15020.
220 Sperschneider, J., Dodds, P. N., Gardiner, D. M., Singh, K. B., and Taylor, J. M. 2018.
221 Improved prediction of fungal effector proteins from secretomes with EffectorP
222 2.0. Mol Plant Pathol 19:2094-2110.
223 Stanke, M., and Waack, S. 2003. Gene prediction with a hidden Markov model and a
224 new intron submodel. Bioinformatics 19:ii215–ii225.
225 Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo,
226 C. A., Zeng, Q., Wortman, J., Young, S. K., and Earl, A. M. 2014. Pilon: an
227 integrated tool for comprehensive microbial variant detection and genome
228 assembly improvement. PLoS One 9:e112963.
229 Wu, Y., Ma, X., Pan, Z., Kale, S. D., Song, Y., King, H., Zhang, Q., Presley, C., Deng,
230 X., Wei, C. I., and Xiao, S. 2018. Comparative genome analyses reveal
231 sequence features reflecting distinct modes of host-adaptation between dicot
232 and monocot powdery mildew. BMC Genomics 19:705.
233 Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. 2012. dbCAN: a web resource
234 for automated carbohydrate-active enzyme annotation. Nucleic Acids Res
235 40:W445–W451.

Xingyuan Zhang
Molecular Plant-Microbe Interactions

Genomic feature	NAFU1	C-strain	Lodi	Ranch9	Branching	e1-101
Sequencing platform	Nanopore	Illumina MiSeq	Illumina HiSeq	Illumina HiSeq	Illumina HiSeq	Illumina HiSeq
BioSample	SAMN17620199	SAMN02803834	SAMN02803901	SAMN02803894	SAMN02803892	SAMN02803896
BioProject	PRJNA695796	PRJNA247407	PRJNA248904	PRJNA248903	PRJNA248900	PRJNA248902
Total assembly size (bp)	61,122,667	52,505,057	49,793,988	49,465,130	50,658,153	49,942,550
Coverage	217×	76×	42×	24×	29×	42×
GC content (%)	48.5	39.0	38.9	38.8	38.5	38.8
Number of contigs	73	8,584	8,093	8,274	11,631	7,601
Maximum contig length (bp)	6,053,329	-	-	-	-	-
Contig N50 (bp)	2,063,233	16,949	13,724	13,213	12,413	15,756
Contig N90 (bp)	822,440	-	-	-	-	-

Contig L50	9	895	1,073	1,099	1,195	935
Total protein-coding genes	7,235	6,484	-	-	-	-
BUSCO completeness (%)	98.3	-	-	-	-	-
Transfer RNAs	244	-	-	-	-	-
Ribosomal RNAs	76	-	-	-	-	-
Secretome[a]	453	422				
Effectorome[b]	196	150	-	-	-	-

[a] Secretome: proteins have signal peptide, but without a transmembrane domain

[b] Effectorome: predicted by software EffectorP