1   **The genome of the butternut canker pathogen, *Ophiognomonia clavigignenti-***

2   ***juglandacearum* shows an elevated number of genes associated with secondary metabolism**

3   **and protection from host resistance responses.**

4   Guangxi Wu[a], Taruna A. Schuelke[b], Gloria Iriarte[a,c], Kirk Broders[a,c]

5   [a] Department of Agricultural Biology, Colorado State University, Fort Collins, Colorado, United

6   States

7   [b] Ecology, Evolution, and Marine Biology Department, University of California, Santa Barbara,

8   California

9   [c]Smithsonian Tropical Research Institute, Apartado 0843-03092, Balboa, Ancon,

10 Republic of Panamá.

11

14 **Abstract**

15 *Ophiognomonia clavigignentijuglandacearum* (*Oc-j*) is a plant pathogenic fungus that causes

16 canker and branch dieback diseases in the hardwood tree butternut, *Juglans cinerea*. *Oc-j* is a

17 member of the order ~~of~~ Diaporthales, which includes many other plant pathogenic species,

18 several of which also infect hardwood tree species. In this study, we sequenced the genome of

19 *Oc-j* and achieved a high-quality assembly and delineated ~~the~~ its phylogeny ~~of Oc-j~~ within the

20 Diaporthales order using a genome-wide multi-gene approach. We also further examined

21 multiple gene families that might be involved in plant pathogenicity and degradation of complex

22 biomass, which are relevant to a pathogenic life-style in a tree host. We found that the *Oc-j*

23 genome contains a greater number of genes in these gene families compared to other species in

24 the Diaporthales. These gene families include secreted CAZymes, kinases, cytochrome P450,

25 efflux pumps, and secondary metabolism gene clusters. The large numbers of these genes likely

26 provide *Oc-j* with an arsenal to cope with the specific ecological niche as a pathogen of the

27 butternut tree.

28 **Introduction**

29   *Ophiognomonia clavigignenti-juglandacearum* (*Oc-j*) is an Ascomycetous fungus in the

30 family Gnomoniaceae and order Diaporthales. Like many of the other species within the

31 Diaporthales, *Oc-j* is a canker pathogen, and is known to infect the hardwood butternut (*Juglans*

32 *cinerea*) (**Figure 1**). The Diaporthales order is composed of 13 families [1], which include

33 several plant pathogens, saprophytes, and enodph~~ytt~~ytes [2]. Numerous tree diseases are caused

34 by members of this order. These diseases include dogwood anthracnose (*Discula destructiva*),

35 butternut canker (*Ophiognomonia clavigignenti-juglandacearum*), apple canker (*Valsa mali*

36 and~~&~~ *Valsa pyri*), Eucalyptus canker (*Chrysoporthe autroafricana, C. cubensis* and *C.*

Commenté [LB1]: I suggest adding "likely" here, which would be in line with the caution you exercised when discussing your results.

37   *deuterocubensis*), and perhaps the most infamous and well-known chestnut blight

38   (*Cryphonectira parastica*). Furthermore, several species ~~of~~ in the Diaporthales also cause

39   important disease of crops including soybean canker (*Diaporthe aspalathi*), soybean seed decay

40   (*Diaporthe langicolla*) and sunflower stem canker (*Diaporthe helianthi*). In addition to

41   pathogens, there are also a multitude of species that have a saprotrophic or endophytic life

42   strategy [3]. However, the saprophytic and endophytic species have not been studied as

43   extensively as the pathogenic species in the Diaporthales.

44   Like many of the tree pathogens in Diaporthales, *Oc-j* is an invasive species, introduced

45   into the U.S. from an unknown origin. This introduction caused extensive damage among the

46   butternut population in North America during the latter half of the 20th century. The first report

47   of butternut canker was in Wisconsin in 1967 [4], and in 1979, the fungus was described for the

48   first time as *Sirococcus clavigignenti-juglandacearum* (*Sc-j*) [5]. ~~Recent~~ Pphylogenetic studies

49   have determined the pathogen ~~which~~ that causes butternut canker is a member of the genus

50   *Ophiognomonia* and was reclassified as *Ophiognomonia clavigignenti-juglandacearum* (*Oc-j*)

51   [6]. The sudden emergence of *Oc-j*, its rapid spread in native North American butternuts, the

52   scarcity of resistant trees, and low genetic variability in the fungus point to a recent introduction

53   of a new pathogenic fungus that is causing a pandemic throughout North America [7].

54   While *Oc-j* is a devastating pathogen of ~~a hardwood tree species~~ the butternut, *Juglans*

55   *cinerea*, ~~many of the~~ there are also several species in the genus *Ophiognomonia* that are

56   endophytes or saprophytes of tree species in the order *Fagales* and more specifically the

57   *Juglandaceae* or walnut family [8,9]. This relationship may support the hypothesis of a host

58   jump, where the fungus may have previously been living as an endophyte or saprophyte before

59   coming into contact with butternut. In fact, a recent study from China reported *Sirococcus*

Mis en forme : Police :Italique

60 *(Ophiognomonia) clavigignenti-juglandacearum* as an endophyte of *Acer truncatum*, which is a

61 maple species native to northern China [10]. The identification of the endophytice strain was

62 made based on sequence similarity of the ITS region of the rDNA. A recent morphological and

63 phylogenetic analysis of this isolate determined that while it is not *Oc-j*, this isolate is indeed

64 more closely related to *Oc-j* than any other previously reported fungal species [11]. The

65 endophyicte isolate also did not produce conidia in culture in comparison to *Oc-j* which produces

66 abundant conidia in culture. It is more likely that these organisms share are common ancestor

67 and represent distinct species.

68   While the impact of members in of the Diaporthales on both agricultural and forested

69 ecosystems is significant [2], there has been limited information regarding the genomic evolution

70 of this order of fungi. Several species have recently been sequenced and the genome data made

71 public. This includes pathogens of trees and crops as well as an endophytic and saprotrophic

72 species [12]. However, these were generally brief genome reports and a more thorough

73 comparative analysis of the species within the Diaporthales has yet to be completed.

74   Here we report the genome sequence of *Oc-j* and use it in comparative analyses with

75 those of tree and crop pathogens within the Diaporthales. Comparative genomics of several

76 members of the Diaporthales order provides valuable insights into fundamental questions

77 regarding fungal lifestyles, evolution and phylogeny, and adaptation to diverse ecological niches,

78 especially as they relate to plant pathogenicity and degradation of complex biomass associated

79 with tree species.

80 **Methods**

81 **DNA extraction and library preparation**

82   For this study, the ex-type culture of *Oc-j* (ATCC 36624) ~~recovered~~ isolated from an infected

83   butternut tree in Wisconsin in 1978, was sequenced. For DNA extraction, isolates were grown on

84   cellophane-covered potato dextrose agar for 7-10 d, and mycelia was collected and lyophilized.

85   DNA was extracted from lyophilized mycelia using the CTAB method as outlined by the Joint

86   Genome Institute for whole genome sequencing (Kohler A, Francis M. Genomic DNA

87   Extraction, Accessed 12/12/2015 http://1000.fungalgenomes.org/home/wp-

88   content/uploads/2013/02/genomicDNAProtocol-AK0511.pdf). The total DNA quantity and

89   quality were measured using both Nanodrop and Qubit, and the sample was sent to the Hubbard

90   Center for Genome Studies at the University of New Hampshire, Durham, New Hampshire.

91   DNA libraries were prepared using the paired-end Illumina Truseq sample preparation kit, and

92   were sequenced on an Illumina HiSeq 2500.

93   **Genome assembly and annotation**

94   We corrected our raw reads using BLESS 0.16 [33] with the following options: -kmerlength 23 -

95   verify –notrim. Once our reads were corrected, we trimmed the reads at a phred score of 2 both

96   at the leading and trailing ends of the reads using Trimmomatic 0.32 [34]. We used a sliding

97   window of four bases that must average a phred score of 2 and the reads must maintain a

98   minimum length of 25 bases. Next, *de novo* assembly was built using SPAdes-3.1.1 [35] with

99   both paired and unpaired reads and the following settings: -t 8 -m 100 --only-assembler.

100 Genome sequences were deposited at ncbi.nlm.nih.gov under Bioproject number PRJNA434132.

101 Gene annotation was performed using the MAKER2 pipeline [36] in an iterative manner as is

102 described in [37], with protein evidence from related species of *Melanconium* sp., *Cryphonectria*

103 *parasitica*, *Diaporthe ampelina*, (from jgi.doe.gov) and *Diaporthe helianthi* [19], for a total of

104 three iterations. PFAM domains were identified in all the genomes using hmmscan with trusted

105 cutoff [38]. Only nine species were included in the PFAM analysis. For the three *Chrysoporthe*

106 species, *D. aspalathi*, and *D. longicolla*, protein sequences were not readily available for

107 downloading online and e-mail requests were unsuccessful. Secondary metabolism gene clusters

108 were identified using antiSMASH 4.0 [39].

109

110 **Species phylogeny**

111 Core eukaryotic proteins identified by CEGMA [40] were first aligned by MAFFT [41] and then

112 concatenated. Only proteins that were present in all genomes and all sequences were longer than

113 90% of the *Saccharomyces cerevisiae* ortholog were used. This resulted in a total of 340 single-

114 copy genes that were concatenated into a single alignment. Phylogeny was then inferred using

115 maximum likelihood by RAxML [42] with PROTGAMMA used to estimate protein evolution,

116 and with 100 bootstraps and then midpoint rooted.

117 **Secreted CAZyme prediction**

118 SignalP [43] was used to predict the presence of secretory signal peptides. CAZymes were

119 predicted using CAZymes Analysis Toolkit [44] based on the most recent CAZY database

120 (www.cazy.org). Proteins that both contain a signal peptide and are predicted to be a CAZyme

121 are annotated as a secreted CAZyme.

122 **Results and Discussion**

123 **Genome assembly and annotation of *Oc-j***

124 The draft genome assembly of *Oc-j* contains a total of 52.6 Mbp and 5,401 contigs, with an N50

125 of 151 Kbp (Table 1). The completeness of the genome assembly was assessed by identifying

126 universal single-copy orthologs using BUSCO with lineage dataset for Sordariomycota [13]. Out

127 of 3,725 total BUSCO groups searched, we found that 3,378 (90.7%) were complete and 264

Commenté [LB2]: Seems awkaward. Maybe remove "were"?

128   (7.1%) were fragmented in the *Oc-j* genome, while only 83 (2.2%) were missing. This result

129   indicates that the *Oc-j* genome is relatively complete.

Genome-wide multi-gene phylogeny of Diaporthales

131   A total of 340 genes were used to generate a multigene phylogeny of the order Diaporthales. All

132   sequenced species in Diaporthales (one genome per species for 12 species) were used. These

133   species include: *Cryphonectria parasitica* (jgi.doe.gov), *Chrysoporthe cubensis* [14],

134   *Chrysoporthe deuterocubensis* [14], *Chrysoporthe austroafricana* [15], *Valsa mali* [12], *Valsa*

135   *pyri* [12], *Diaporthe ampelina* [16], *Diaporthe aspalathi* [17], *Diaporthe longicolla* [18],

136   *Diaporthe helianthi* [19], and *Melanconium* sp. The outgroups used were *Neurospora crassa*

137   [20] and *Magnaporthe grisea* [21]. We show that the order Diaporthales is divided into two main

138   branches. One branch includes *Oc-j*, *C. parasitica,* and the *Chrysoporthe* species (**Figure 2**), in

139   which *Oc-j* is located basally~~the outlier~~, indicating its early divergence from the rest of the

140   branch. The other branch includes the *Valsa* and *Diaporthe* species.

Gene content comparison across the Diaporthales

142   To examine the functional capacity of the *Oc-j* gene repertoire, the PFAM domains were

143   identified in the protein sequences (**Supplementary Table 1**). For comparison, we also

144   examined eight of the above mentioned 13 related species, where protein sequences were

145   successfully retrieved (see **Methods** for details).

146   CAZymes are a group of proteins that are involved in degrading, modifying, or creating

147   glycosidic bonds and contain predicted catalytic and carbohydrate-binding domains [22]. When

148   secreted by fungal pathogens, CAZymes can participate in degrading plant cell walls during

149   colonization ~~by fungal pathogens~~; therefore, a combination of CAZyme and protein secretion

150   prediction was used to identify and classify enzymes likely involved in cell wall degradation in

151 plant pathogens [22]. Overall, the *Oc-j* genome contains 576 putatively secreted CAZymes, more

152 than any of the other eight species included in this analysis, and 60 CAZymes more than *D.*

153 *helianthi*, the species with the second most (**Figure 3, Suppl. Table 1**). Given that all except the

154 bread mold species *N. crassa*, which has the lowest number of secreted CAZymes, are plant

155 pathogens, this result indicates that *Oc-j* contains an especially large gene repertoire for cell wall

156 degradation. However, it is not the greatest among members of the kingdom fungi. For instance,

157 the saprophytic species *Penicillium subrubescens* and *Podospora anserina* have been found to

158 have 719 and 590 CAZymes, respectively [23]. This pattern also holds true when compared to

159 other tree pathogens that have a latent period and an extended host range that . For instance,

160 *Botryosphaeria dothidea*, which is a pathogen of many woody species and has a prolonged latent

161 and endophytic phase during the infection process, was reported to have 623 CAZymes [24].

162 Other studies have also found an elevated number of CAZymes in multi-host pathogens of

163 woody plant species including *Eutypa lata* and *Neofusicoccum parvum* which had 484 and 413

164 CAZymes, respectively [22][2, 3]. In a recent comparative genomic analysis of tree pathogens,

165 it was reported that the black walnut pathogen, *Geosmithia morbida*, had 406 CAZymes, the

166 lodgepole pine pathogen *Grosmania clavigera* had 530 CAZymes, and the plane tree canker

167 pathogen, *Ceratocystis platani*, had 360 CAZymes [23]. Since secreted CAZymes are involved

168 in degrading plant cell walls [22], the large number of secreted CAZymes in *Oc-j* and other

169 pathogens of woody plants might help facilitate its a life-style infecting multiple species of

170 perennial woody plants with a wide array of both preformed and infection-induced resistance

171 responses. and colonizing butternut trees.

172   *Kinase*

173   Among aA total of 77 kinase related PFAM domains were found in the nine above mentioned

174   species. Of these,, 61 domains are present in more genesmore abundant in *Oc-j* than the average

175   of the eight related species, while only eight domains are present in fewer genesless abundant in

176   *Oc-j* (**Figure 3, Suppl. Table 1**). Given that kinases are involved in signaling networks, this

177   result might indicate a more complicated signaling network in *Oc-j*. One example of kinase gene

178   family expansion is the domain family of fructosamine kinase (PF03881). *Oc-j* contains 29 genes

179   with this domain, while *N. crassa* has none and the other related species have between 3-8 genes

180   (**Figure 3**). Almost nothing is known about the function of f-ructosamine kinases in fungi. In

181   other eukaryotes, it is involved in protein deglycation by mediating phosphorylation of

182   fructoselysine residues on glycated proteins. While plants and fungi both produce an array of

183   glycated proteins it is unclear if *Oc-j* is producing fructosamine kinase to function in deglycation

184   of proteins produced by the plant host. Fructosamine kinases are extracellular serine-proteases,

185   and are implicated in the pathogenic activity of several fungal pathogens of mammals [25],

186   including *Aspergillus fumigatus* [26]. -Whether or not fructosamine kinase functions are

187   important for plant pathogens remains unknown.

188   Cytochrome P450s (CYPs) are a superfamily of monooxygenases that play a wide range

189   of roles in metabolism and adaptation to ecological niches in fungi [27] [24]. Due to their

190   participation in a large number of detoxification reactions as well as in the metabolism of

191   specific xenobiotics which may be co-assimilated as carbon source, CYPs are thought to be

192   critical for the colonization of new ecological niches [ref28]. Among the nine species included in

193   the PFAM analysis, *N. crassa* has only 31 CYPs, while the other eight plant-pathogenic species

194   have between 104 – 223 CYPs. This result likely reflects that the ecological niche of *N. crassa*,

195   breada saprophyte, is more favorable for fungal growth than alive plants with active defense

Commenté [LB3]: A relevant reference would seem appropriate

Mis en forme : Police :Italique

Mis en forme : Police :Italique

196   mechanisms, thus fewer CYPs are needed for *N. crassa* to cope with relatively simple substrates.

197   The *Oc-j* genome contains 223 CYPs (**Figure 3**, **Suppl. Table 1**), more than any of the other

198   seven plant-pathogenic fungal species. This may be an evolutionary response to the number and

199   diversity of secondary metabolites, such as juglone, produced by butternut as well as other

200   *Juglans* species, that would need to be metabolized by any fungal pathogen trying to colonize the

201   tree. Other pathogens of woody plants species are also known to have an increased number of

202   CYPs. The grapevine canker pathogens *E. lata* and *N. parvum* have 205 and 212 CYPs

203   respectively [22]. According to the Fungal Cytochrome P450 Database (FCPD) [29], very few

204   fungal species have more CYPs than *Oc-j*. The brown rot and pathogenic fungi in the

205   basidiomycota have the greatest number of CYPs in the fungal kingdom. The brown rot fungus

206   *Postia placenta* has the greatest number of CYPS with 345 CYPs, while the cacao pathogen

207   *Moniliophthora perniciosa* and root rot pathogen *Armillaria mellea* have 300 and 245 CYPs,

208   repectively [29]. The one characteristic all of these fungi have in common is their association

209   with dead, dying or living trees and the need to breakdown or resist the multitude of chemical

210   compounds produced by tree species.

211

212   *Efflux pumps.* To overcome host defenses, infect and maintain colonization of the host,

213   fungi employ efflux pumps to counter intercellular toxin accumulation [30] [25]. Here, we

214   examine the presence of two major efflux pump families, the ATP-binding cassette (ABC)

215   transporters and transporters of the major facilitator superfamily (MFS) [30] in the Diaporthales

216   genomes. ABC transporters (PF00005, pfam.xfam.org) are present in all of the nine species

217   included in the PFAM analysis, ranging from 25 genes in *N. crassa* to 51 genes in *V. mali*, while

218   *Oc-j* has 47 genes (**Figure 3**, **Suppl. Table 1**). The major facilitator superfamily are membrane

219   proteins are which are expressed ubiquitously in all kingdoms of life for the import or export of

220   target substrates. The MFS is a clan that contains 24 PFAM domain families (pfam.xfam.org).

221   The *Oc-j* genome contains 653 MFS efflux pump genes, the most among all nine species

222   included in the PFAM analysis (**Figure 3, Suppl. Table 1**). *Cryphonectria. parasitica* has 552

223   genes, the second most. Interestingly, it was recently shown that a the secondary metabolite

224   juglone extracted from *Juglans* spp. could be used as potential efflux pump inhibitors in

225   *Staphylococcus aureus*, inhibiting the export of antibiotics out of the bacterial cells [31]. In line

226   with this previous finding, our results suggest that *Oc-j* genome contains a large arsenal of efflux

227   pumps likely due to the need to cope with hostile secondary metabolites such as the efflux pump

228   inhibitor juglone.

229   *Secondary metabolism gene clusters.* Secondary metabolite toxins play an important role in

230   fungal nutrition and virulence [32]. To identify gene clusters involved in the biosynthesis of

231   secondary metabolites in *Oc-j* and related species, we scanned their genomes for such gene

232   clusters. We found that the *Oc-j* genome contains a remarkably large repertoire of secondary

233   metabolism gene clusters, when compared to the closely related *C. parasitica* and the

234   *Chrysoporthe* species (**Figure 3**, **Suppl. Table**). While *C. parasitica* has 44 gene clusters and the

235   *Chrysoporthe* species have 18 – 48 gene clusters, the *Oc-j* genome has a total of 72 gene clusters

236   (**Supplementary Table 1**), reflecting a greater capacity to produce various secondary

237   metabolites. Among the 72 gene clusters in *Oc-j*, more than half are type 1 polyketide synthases

238   (t1pks, 39 total, including hybrids), followed by non-ribosomal peptide synthetases (nrps, 14

239   total, including hybrids) and terpene synthases (9) (**Supplementary Table 1**).

240   When compared to all species included in this study, we found that *Oc-j* has the third highest

241   number of clusters, only after *D. longicolla* and *Melanconium* sp.; however, when normalized by

242   total gene number in each species, *Oc-j* is shown to have 6.4 secondary metabolism gene clusters

Mis en forme : Police :Italique

243   per 1000 genes, the highest among all species included in this study. The *D. helianthi* genome

244   contains only six clusters, likely due to the poor quality of the genome assembly with an N50 of

245   ~6 kb~~ps~~ (**Figure 3**, **Suppl. Table 1**). Cluster numbers can vary greatly within the same genus.

246   **Conclusion**

247   We constructed a high-quality genome assembly for *Oc-j*, and delineated the phylogeny of the

248   Diaporthales with a genome-wide multi-gene approach, revealing two major branches. We then

249   examined several gene families relevant to plant pathogenicity and complex biomass

250   degradation. We found that the *Oc-j* genome contains large numbers of genes in these gene

251   families. These genes might be essential for *Oc-j* to cope with its niche in the hardwood butternut

252   (*Juglans cinerea*). Future research will need to focus on understanding the prevalence of these

253   genes associated with complex biomass degradation among other members of the

254   *Ophiognomonia* genus which include endophytes, saprophytes and pathogens. It will be

255   interesting to know which of these different classes of genes are important in the evolution of

256   distinct fungal lifestyles and niche adaptation. Future research will also focus on comparing the

257   butternut canker pathogen to canker pathogens of other tree species which produce large

258   numbers of secondary metabolites known to be important in host defense. Our genome also

259   serves as an essential resource for the *Oc-j* research community.

260   **Conflicts of interest**

261   The authors have no conflict of interest.

262

268

References

[1] Voglmayr H, Castlebury L.A, Jaklitsch W.M, Juglanconis gen. nov. on Juglandaceae, and the new family Juglanconidaceae (Diaporthales). 2017. Persoonia Mol. Phylogeny Evol. Fungi. 38:136–155. doi:10.3767/003158517X694768.

[2] Rossman A.Y, Farr D.F, Castlebury L.A. 2007. A review of the phylogeny and biology of the Diaporthales, Mycoscience. 48:135–144. doi:10.1007/S10267-007-0347-7.

[3] Castlebury L.A, Rossman A.Y, Jaklitsch W.J, Vasilyeva L.N. 2002. A preliminary overview of the Diaporthales based on large subunit nuclear ribosomal DNA sequences, Mycologia. 94:1017–1031.

[4] Renlund D.W, Forest pest conditions in Wisconsin, Annual Report, 53 p., Wisconsin Department of Natural Resources, Madison, WI, 1971.

[5] Nair V.M.G, Kostichka C.J, Kuntz J.E. 1979. *Sirococcus clavigignenti-juglandacearum:* An undescribed species causing canker on butternut, Mycologia. 71:641–646. doi:10.2307/3759076.

[6] Broders K.D, Boland G.J, Reclassification of the butternut canker fungus, *Sirococcus clavigignenti-juglandacearum*, into the genus *Ophiognomonia*. 2011. Fungal Biol. 115:70–79. doi:10.1016/j.funbio.2010.10.007.

[7] Broders K.D, Boraks A, Sanchez A.M, Boland G.J, Population structure of the butternut canker fungus. 2012. *Ophiognomonia clavigignenti-juglandacearum*, in North American forests, Ecol. Evol. 2:2114–2127. doi:10.1002/ece3.332.

[8] Sogonov M.V, Castlebury L.A, Y. R.A, Mejia L.C, White J.F. 2008. Leaf-inhabiting genera of the Gnomoniaceae, Diaporthales, Stud. Mycol. 62:1-79.

[9] Walker D.M, Castlebury L.A, Rossman A.Y, Mejia L.C, White J.F. 2012. Phylogeny and taxonomy of *Ophiognomonia* (Gnomoniaceae, Diaporthales), including twenty-five new species in this highly diverse genus, Fungal Diversity. 57: 85-147.

[10] Sun X, Guo L.-D, Hyde K.D. 2011. Community composition of endophytic fungi in *Acer truncatum* and their role in decomposition, Fungal Divers. 47:85–95. doi:10.1007/s13225-010-0086-5.

[11] Broders K, Boraks A, Barbison L, Brown J, Boland G.J. 2015. Recent insights into the pandemic disease butternut canker caused by the invasive pathogen *Ophiognomonia clavigignenti-juglandacearum*, For. Pathol. 45:1–8. doi:10.1111/efp.12161.

300   [12] Yin Z, Liu H, Li Z, Ke X, Dou D, Gao X, Song N, Dai Q, Wu Y, Xu J.-R, Kang Z, Huang
301   L. 2015. Genome sequence of *Valsa canker* pathogens uncovers a potential adaptation of
302   colonization of woody bark, New Phytol. 208:1202–1216. doi:10.1111/nph.13544.
303   [13] Simão F.A, Waterhouse R.M, Ioannidis P, Kriventseva E.V, Zdobnov E.M. 2015. BUSCO:
304   assessing genome assembly and annotation completeness with single-copy orthologs,
305   Bioinformatics. 31:3210–3212. doi:10.1093/bioinformatics/btv351.
306   [14] Wingfield B.D, Barnes I, Wilhelm de Beer Z, De Vos L, Duong T.A, Kanzi A.M, Naidoo
307   K, Nguyen H.D.T, Santana Q.C, Sayari M, Seifert K.A, Steenkamp E.T, Trollip C, van der
308   Merwe N.A, van der Nest M.A, Markus Wilken P, Wingfield M.J. 2015. IMA Genome-F 5:
309   Draft genome sequences of *Ceratocystis eucalypticola, Chrysoporthe cubensis, C.*
310   *deuterocubensis*, *Davidsoniella virescens, Fusarium temperatum,Graphilbum fragrans,*
311   *Penicillium nordicum*, and *Thielaviopsis musarum*, IMA Fungus. 6:493–506.
312   doi:10.5598/imafungus.2015.06.02.13.
313   [15] Wingfield B.D, Ades P.K, Al-Naemi F.A, Beirn L.A, Bihon W, Crouch J.A, de Beer Z.W,
314   De Vos L, Duong T.A, Fields C.J, Fourie G, Kanzi A.M, Malapi-Wight M, Pethybridge S.J,
315   Radwan O, Rendon G, Slippers B, Santana Q.C, Steenkamp E.T, Taylor P.W.J, Vaghefi N,
316   van der Merwe N.A, Veltri D, Wingfield M.J. 2015. IMA Genome-F 4: Draft genome
317   sequences of *Chrysoporthe austroafricana, Diplodia scrobiculata, Fusarium nygamai,*
318   *Leptographium lundbergii, Limonomyces culmigenus, Stagonosporopsis tanaceti,* and
319   *Thielaviopsis punctulata,* IMA Fungus. 6:233–248. doi:10.5598/imafungus.2015.06.01.15.
320   [16] Savitha J, Bhargavi S.D, Praveen V.K. 2016. Complete genome sequence of the endophytic
321   fungus *Diaporthe* (Phomopsis) *ampelina*, Genome Announc. 4 e00477-16.
322   doi:10.1128/genomeA.00477-16.
323   [17] Li S, Song Q, Martins A.M, Cregan P. 2016. Draft genome sequence of *Diaporthe aspalathi*
324   isolate MS-SSC91, a fungus causing stem canker in soybean, Genomics Data. 7:262–263.
325   doi:10.1016/j.gdata.2016.02.002.
326   [18] Li S., Darwish O, Alkharouf N, Matthews B, Ji P, Domier L.L, Zhang N, Bluhm B.H. 2015.
327   Draft genome sequence of *Phomopsis longicolla* isolate MSPL 10-6, Genomics Data. 3:55–
328   56. doi:10.1016/j.gdata.2014.11.007.

329 [19] Baroncelli R, Scala F, Vergara M, Thon M.R, Ruocco M. 2016. Draft whole-genome

330   sequence of the *Diaporthe helianthi* 7/96 strain, causal agent of sunflower stem canker,

331   Genomics Data. 10:151–152. doi:10.1016/j.gdata.2016.11.005.

332 [20] Galagan J.E, Calvo S.E, Borkovich K.A, Selker E.U, Read N.D, Jaffe D, FitzHugh W, Ma

333   L.-J, Smirnov S, Purcell S, Rehman B, Elkins T, Engels R, Wang S, Nielsen C.B, Butler J,

334   Endrizzi M, Qui D, Ianakiev P, Bell-Pedersen D, Nelson M.A, Werner-Washburne M,

335   Selitrennikoff C.P, Kinsey J.A, Braun E.L, Zelter A, Schulte U, Kothe G.O, Jedd G, Mewes

336   W, Staben C, Marcotte E, Greenberg D, Roy A, Foley K, Naylor J, Stange-Thomann N,

337   Barrett R, Gnerre S, Kamal M, Kamvysselis M, Mauceli E, Bielke C, Rudd S, Frishman D,

338   Krystofova S, Rasmussen C, Metzenberg R.L, Perkins D.D, Kroken S, Cogoni C, Macino G,

339   Catcheside D, Li W, Pratt R.J, Osmani S.A, DeSouza C.P.C, Glass L, Orbach M.J, Berglund

340   J.A, Voelker R, Yarden O, Plamann M, Seiler S, Dunlap J, Radford A, Aramayo R, Natvig

341   D.O, Alex L.A, Mannhaupt G, Ebbole D.J, Freitag M, Paulsen I, Sachs M.S, Lander E.S,

342   Nusbaum C, Birren B. 2003. The genome sequence of the filamentous fungus *Neurospora*

343   *crassa*, Nature. 422:859–868. doi:10.1038/nature01554.

344 [21] Dean R.A, Talbot N.J, Ebbole D.J, Farman M.L, Mitchell T.K, Orbach M.J, Thon M,

345   Kulkarni R, Xu J.-R, Pan H, Read N.D, Lee Y.-H, Carbone I, Brown D, Oh Y.Y, Donofrio

346   N, Jeong J.S, Soanes D.M, Djonovic S, Kolomiets E, Rehmeyer C, Li W, Harding M, Kim S,

347   Lebrun M.-H, Bohnert H, Coughlan S, Butler J, Calvo S, Ma L.-J, Nicol R, Purcell S,

348   Nusbaum C, Galagan J.E, Birren B.W. 2005. The genome sequence of the rice blast fungus

349   *Magnaporthe grisea,* Nature. 434:980–986. doi:10.1038/nature03449.

350 [22] Morales-Cruz A, Amrine K.C.H, Blanco-Ulate B, Lawrence D.P, Travadon R, Rolshausen

351   P.E, Baumgartner K, Cantu D. 2015. Distinctive expansion of gene families associated with

352   plant cell wall degradation, secondary metabolism, and nutrient uptake in the genomes of

353   grapevine trunk pathogens, BMC Genomics. 16:469. doi:10.1186/s12864-015-1624-z.

354 [23] Peng M, Dilokpirnol A, Makela M.R., Hilden K, Bervoets S, Riley R, Grigoriev I.V., Haiaut

355   M., Henrissat B., de Vries R.P., Granchi Z. et al.2017.The draft genome sequence of the

356   ascomycete fungus *Penicillium subrubescens* reveals a highly enriched content of plant

357   biomass related CAZymes compared to related fungi. Journal of Biotechnology. 246:1-3.

358   doi: 10.1016/j.jbiotec.2017.02.012.

359 [24] Marsberg A, Kemler M, Jami F, Nagel J.H., Postma-Smidt A., Naidoo S., Wingfield M. J.,
360   Crous P. W., Spatafora J. W., Hesse C. N., Robbertse B., Slippers. 2017. *Botryosphaeria*
361   *dothidea*: a latent pathogen of global importance to woody plant health. Molecular Plant
362   Pathology. 18:477-88. doi: 10.1111/mpp.12495

363 [25] da Silva B.A, dos Santos ALS, Barreto-Bergter E, Pinto MR. 2006. Extracellular peptidase
364   in the fungal pathogen *Pseudallescheria boydii*. Curr Microbiol 53: 18–22.

365 [26] Monod M, Capoccia S, Lechenne B, Zaugg C, Holdom M, Jousson O. 2002. Secreted
366   proteases from pathogenic fungi. Int J Med Microbiol 292: 405– 419.

367 [27] Chen W, Lee M.-K, Jefcoate C, Kim S.-C, Chen F, Yu J.-H. 2014. Fungal Cytochrome
368   P450 Monooxygenases: Their Distribution, Structure, Functions, Family Expansion, and
369   Evolutionary Origin, Genome Biol. Evol. 6:1620–1634.

370 [28] Moktali V, Park J, Fedorova-Abrams ND, Park B., Choi J., Lee Y, Kang S. 2012.
371   Systematic and searchable classification of cytochrome P450 proteins encoded by fungal
372   and oomycete genomes. BMC Genomics 13: 525.

373 [29] Nelson D.R. The cytochrome p450 homepage. 2009. Hum Genomics. 4:59.

374 [30] Coleman J.J, Mylonakis E, Efflux in Fungi: La Pièce de Résistance. 2009. PLOS Pathog. 5
375   e1000486. doi:10.1371/journal.ppat.1000486.

376 [31] Zmantar T, Miladi H, Kouidhi B, Chaabouni Y, Ben Slama R, Bakhrouf A, Mahdouani K,
377   Chaieb K. 2016. Use of juglone as antibacterial and potential efflux pump inhibitors in
378   *Staphylococcus aureus* isolated from the oral cavity, Microb. Pathog. 101: 44–49.
379   doi:10.1016/j.micpath.2016.10.022.

380 [32] Howlett B.J. 2006. Secondary metabolite toxins and nutrition of plant pathogenic fungi,
381   Curr. Opin. Plant Biol. 9:371–375. doi:10.1016/j.pbi.2006.05.004.

382 [33] Heo Y, Wu X.-L, Chen D, Ma J, Hwu W.-M. 2014. BLESS: Bloom filter-based error
383   correction solution for high-throughput sequencing reads, Bioinformatics. 30:1354–1362.
384   doi:10.1093/bioinformatics/btu030.

385 [34] Bolger A.M, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina
386   sequence data, Bioinformatics. 30:2114–2120. doi:10.1093/bioinformatics/btu170.

387 [35] Bankevich A, Nurk S, Antipov D, Gurevich A.A, Dvorkin M, Kulikov A.S, Lesin V.M,
388   Nikolenko S.I, Pham S, Prjibelski A.D, Pyshkin A.V, Sirotkin A.V, Vyahhi N, Tesler G,
389   Alekseyev M.A, Pevzner P.A. 2012. SPAdes: A New Genome Assembly Algorithm and Its

390   Applications to Single-Cell Sequencing, J. Comput. Biol. 19:455–477.

391   doi:10.1089/cmb.2012.0021.

392   [36] Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database

393   management tool for second-generation genome projects, BMC Bioinformatics. 12:491.

394   doi:10.1186/1471-2105-12-491.

395   [37] Yu J, Wu G, Jurick W.M, Gaskins V.L, Yin Y, Yin G, Bennett J.W, Shelton D.R. 2016.

396   Genome Sequence of *Penicillium solitum* RS1, Which Causes Postharvest Apple Decay,

397   Genome Announc. 4: e00363-16. doi:10.1128/genomeA.00363-16.

398   [38] Bateman A, Coin L, Durbin R, Finn R.D, Hollich V, Griffiths-Jones S, Khanna A, Marshall

399   M, Moxon S, Sonnhammer E.L.L, Studholme D.J, Yeats C, Eddy S.R. 2004. The Pfam

400   protein families database, Nucleic Acids Res. 32: D138–D141. doi:10.1093/nar/gkh121.

401   [39] Weber T, Blin K, Duddela S, Krug D, Kim H.U, Bruccoleri R, Lee S.Y, Fischbach M.A,

402   Müller R, Wohlleben W, Breitling R, Takano E, Medema M.H. 2015. antiSMASH 3.0—a

403   comprehensive resource for the genome mining of biosynthetic gene clusters, Nucleic Acids

404   Res. 43: W237–W243. doi:10.1093/nar/gkv437.

405   [40] Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in

406   eukaryotic genomes, Bioinformatics. 23:1061–1067. doi:10.1093/bioinformatics/btm071.

407   [41] Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple

408   sequence alignment based on fast Fourier transform, Nucleic Acids Res. 30:3059–3066.

409   doi:10.1093/nar/gkf436.

410   [42] Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of

411   large phylogenies, Bioinformatics. 30:1312–1313. doi:10.1093/bioinformatics/btu033.

412   [43] Petersen T.N, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal

413   peptides from transmembrane regions, Nat. Methods. 8:785. doi:10.1038/nmeth.1701.

414   [44] Park B.H, Karpinets T.V, Syed M.H, Leuze M.R, Uberbacher E.C. 2010. CAZymes

415   Analysis Toolkit (CAT): Web service for searching and analyzing carbohydrate-active

416   enzymes in a newly sequenced organism using CAZy database, Glycobiology. 20:1574–

417   1584. doi:10.1093/glycob/cwq106.

418

419

420

421   Table 1. Assembly statistics of the butternut canker pathogen and other Diaporthales analyzed.

422

423   **Supplementary Table 1. PFAM and antiSMASH annotation of *Oc-j* and related species**

424

425

426

427

428

429

430   **Figure Legend**

431   **Figure 1.** Symptoms of infection caused by *Oc-j* on the A) leaves, B) branches and C) trunk as

432   well as as D) signs of asexual fruting bodies on an infected branch.

433

434   **Figure 2.** Phylogeny of Diaporthales and related species inferred using maximum likelihood by

435   RAxML [37] with 1000 bootstraps and then midpoint rooted. The first and second numbers in

436   parentheses represent the genome sizes in Mb and the number of predicted protein models,

437   respectively

438

439   **Figure 3.** Abundance of genes in specific classes including ABC transporters, Cytochrome

440   P450, Kinases, CAZymes, and MFS Efflux pumps present in nine species within the order

441   Diaporthales

442

443
444

Wu et al., 19, Oc-j genome

445
446
447