




# Draft genome sequence of *Neonectria* sp. DH2 isolated from *Meconopsis grandis* Prain in Tibet

Xiaojing Lin<sup>1,3</sup> · Hui Xu<sup>3</sup> · Lan Liu<sup>1,2</sup> · Huixian Li<sup>1</sup> · Zhizeng Gao<sup>1,2</sup> 

Received: 28 December 2018 / Accepted: 17 July 2020 / Published online: 21 July 2020  
© King Abdulaziz City for Science and Technology 2020

## Abstract

In the current study, we report the high-quality draft genome sequence of *Neonectria* sp. DH2, an endophytic fungus isolated from *Meconopsis grandis* Prain in Tibet. The whole genome is about 45.8 Mbp, with a GC content of 53%. A total of 14,163 genes are predicted to encode proteins, and 557 of them are considered as unique, as no matches are found in five gene databases. A neighbor-joining phylogenetic tree based on internal transcribed spacer (ITS) region sequences shows that *Neonectria* sp. DH2 was most closely related to *Neonectria ramulariae*. 47 biosynthetic gene clusters (BGC) were identified in *Neonectria* sp. DH2 genome, and only 5 BGCs shows significant similarities to previously reported BGCs. The presence of 42 unique BGCs in *Neonectria* sp. DH2 suggests that it has great potential to produce novel secondary metabolites.

**Keywords** *Neonectria* sp. DH2 · Endophytic fungi · Draft genome · Secondary metabolites

## Introduction

The ascomycete fungi genus *Neonectria* are well-known plant pathogens associated with beech and fruit trees bark canker (Calic et al. 2017; Wenker et al. 2017; Ramsfield et al. 2013). They are also a common genus of endophytic fungal (Zhong et al. 2017; Yang et al. 2015). Some members of this genus produce bioactivity secondary metabolites. For example, endophytic fungus *Neonectria ramulariae* Wollenw KS-246 produces cytotoxic pyrrolicidines and pyrrospirones (Shiono et al. 2008). Another *Neonectria* sp. isolated from a soil sample from the Qinghai-Tibetan plateau produces oxaphenalenones and Neonectrolide A–E, which showed cytotoxic effects against human tumor cell lines (Ren et al. 2012).

A fungus isolated from a *Nasutitermes corniger* nest and identified as *Neonectria discophora* SNB-CN63 displays a potent antibacterial activity (Nirma et al. 2014). In contrast to the isolation and bioactivity study of their secondary metabolites, the genome information of *Neonectria* sp. remained under-exploited. Currently, only three genome sequences have been reported: *Neonectria ditissima* (Genebank Accession LDPL00000000, Deng et al. 2015), *Neonectria punicea* (Genebank Accession QGQA00000000), and *Neonectria hederæ* (Genebank Accession QGQB00000000). Here, we reported a high-quality draft genome sequence of *Neonectria* sp. DH2 to better understand the genome of genus *Neonectria* and exploit its second metabolites production potential.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s13205-020-02345-8>) contains supplementary material, which is available to authorized users.

✉ Zhizeng Gao  
gaozhizeng@mail.sysu.edu.cn

<sup>1</sup> School of Marine Sciences, Sun Yat-sen University, Guangzhou 510006, China

<sup>2</sup> Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Zhuhai 519000, China

<sup>3</sup> Research Center of Chinese Herbal Resource Science and Engineering, Guangzhou University of Chinese Medicine, Guangzhou 510006, China

## Materials and methods

### Strains and cultivation conditions

*Neonectria* sp. DH2 was isolated from *Meconopsis grandis* Prain, a rare Tibetan traditional medicinal herb, in Tibet, China. The strain grew slowly on PDA medium (200 g/L of potato extract, 20 g/L of glucose, and 20 g/L agar) at 20 °C and formed white mycelium and spore gradually.

### Genome sequencing, assembly and annotation

*Neonectria* sp. DH2 was inoculated in 300 mL PDB medium (300 g/L potato, 20 g/L glucose) and cultivated at 20 °C. Mycelium was harvested after 7 days of cultivation. Genomic DNA was extracted by QIAGEN® Genomic DNA extraction kit (Cat#13323, QIAGEN) according to the standard operating procedure provided by the manufacturer. The extracted DNA was quantified by NanoDrop™ One UV–Vis spectrophotometer (Thermo Fisher Scientific, USA) for DNA purity (OD260/280 ranging from 1.8 to 2.0 and OD 260/230 is between 2.0 and 2.2). The genomic DNA was sheared into 20-kb fragments by g-TUBEs (Covaris, USA) and it was sequenced on Sequel Sequencing Kit 2.1 (Pacific Biosciences, USA). The coverage sequencing depth was 80×. The *Neonectria* sp. DH2 sequence generated 454,780 raw reads consisting of 3,696,611,013 raw nucleotides. The data were assembly using CANU assembler v2.0 (Koren et al. 2017). Repeats were masked by RepeatMasker (Tarailo-Graovac and Chen 2009). The draft genome was annotated using Augustus v3.2 (Stanke et al. 2004) and Genscan v3.0 (Stifanic and Batel 2007). BUSCO v3.0 (Simao et al. 2015) (ascomycota single-copy homologous gene databases) was used to assess the completeness of *Neonectria* sp. DH2 genome. Of the assembled nucleotides, 96.2% of complete genetic components can be found, and 0.7% of the complete BUSCO single-copy orthologues are duplication, indicating that it is a high-quality assembly. BLAST analysis was based on the non-redundant protein database (Nr), Swiss-Prot, gene ontology (GO), cluster of orthologous groups of eukaryotic proteins (KOG) and kyoto encyclopedia of genes and genomes (KEGG) databases. RNAmmer

**Table 1** General features of the *Neonectria* sp. DH2

Attribute	Value
Genome size (Mbp)	45
DNA G+C content (%)	53
DNA contigs	43
Contig N50 (bp)	1,899,891
Largest contig (Mbp)	5.08
CRISPR repeats	6
rRNA	72
tRNA	210
Genes with function prediction	13,606
Annotated in Nr	13,597 (96.00%)
Annotated in SwissProt	9369 (66.15%)
Annotated in GO	7329 (51.75%)
Annotated in KOG	4038 (28.51%)
Annotated in KEGG	2756 (19.46%)

(Lagesen et al. 2007) and tRNAscan-SE (Lowe and Eddy 1997) were used to identify rRNAs and tRNAs. The 18S rRNA gene sequence of *Neonectria* sp. DH2 and several other species were aligned using ClustalW (Oliver et al. 2005) and phylogenetic analysis based on the nearest neighbor-joining method was drawn using MEGA7 (Kumar et al. 2016). CRISPRs webserver was used to identify CRISPR repeats. AntiSMASH 3.0 (Medema et al. 2011) was used to predict secondary metabolite gene clusters.

## Results and discussion

The assembly genome is approximately 45.8 Mbp, including 43 contigs. Table 1 shows the detail of the genome annotation. The GC content of *Neonectria* sp. DH2 genome is 53.00%. The longest contig is 5.08 Mbp and the N50 length is 1,899,89 bp. Approximately 1.18 Mbp repeated regions were found in the genome, accounted for 2.57% of the genome size. Gene prediction analysis yielded a total of 14,163 protein-encoding genes. Of the 14,163 genes predicted, 13,606 (96.07%) were annotated by matching against the non-redundant protein database (Nr), Swiss-Prot, gene ontology (GO),

**Table 2** Number of genes associated with general KOG functional categories

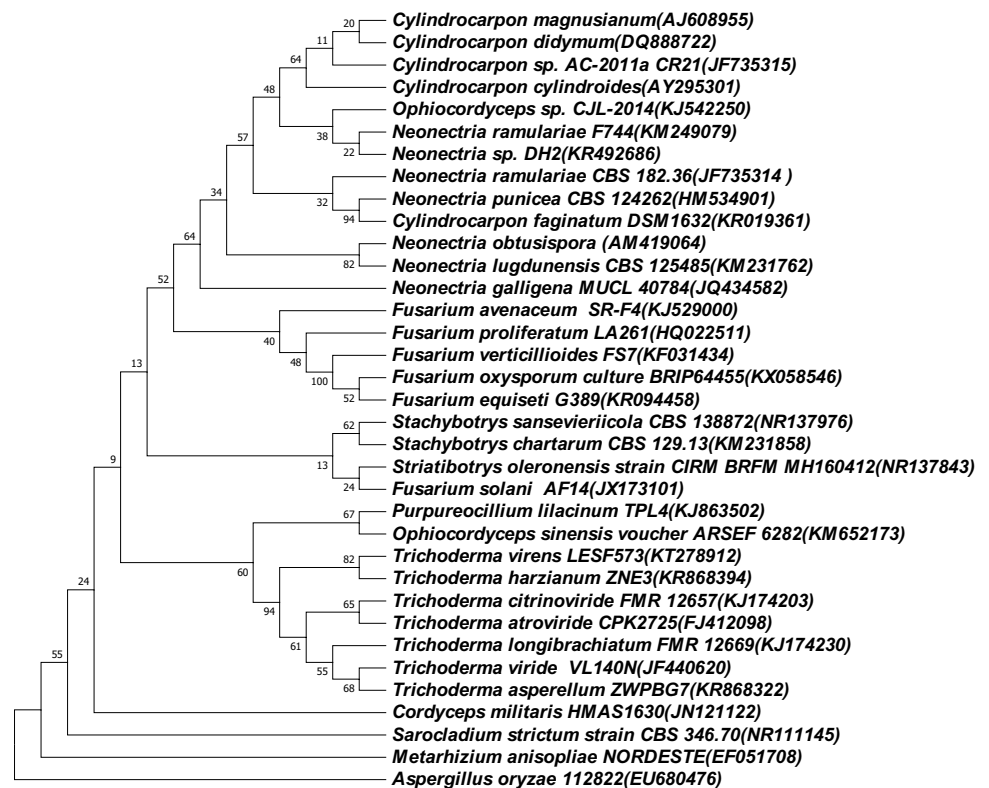
Code	Value	%age	feature
B	46	0.32	Chromatin structure and dynamics
C	363	2.56	Energy production and conversion
E	269	1.90	Amino acid transport and metabolism
A	99	0.70	RNA processing and modification
F	54	0.38	Nucleotide transport and metabolism
G	275	1.94	Carbohydrate transport and metabolism
H	77	0.54	Coenzyme transport and metabolism
J	195	1.38	Translation, ribosomal structure and biogenesis
K	134	0.95	Transcription
V	34	0.24	Defense mechanisms
L	84	0.59	Replication, recombination and repair
I	336	2.37	Lipid transport and metabolism
M	97	0.68	Cell wall/membrane/envelope biogenesis
N	2	0.01	Cell motility
O	328	2.32	Posttranslational modification, protein turnover, chaperones
Y	5	0.04	Nuclear structure
T	240	1.69	Signal transduction mechanisms
Q	423	2.99	Secondary metabolites biosynthesis, transport and catabolism
R	929	6.56	General function prediction only
Z	59	0.42	Cytoskeleton
S	206	1.45	Function unknown
P	115	0.81	Inorganic ion transport and metabolism
W	6	0.04	Extracellular structures
X	1	0.01	Unknown

The total is based on the total number of protein coding genes in the genome

Eukaryotic orthologous groups (KOG) and kyoto encyclopedia of genes and genomes (KEGG) databases, 557 (3.93%) were considered as unique, without any matches in five databases mentioned above. In addition, 72 rRNA and 210 tRNA genes were identified. Six CRISPR repeats were detected. Among the predicted genes, 28.51% were assigned to KOG. Table 2 presents the distribution of genes into KOGs functional categories. A neighbor-joining phylogenetic tree was build based on 18 s rRNA gene sequences of *Neonectria* sp. DH2 (from RNAmmer) and other 33 *Hypocreales* fungi with outgroup of *Aspergillus oryzae* (Fig. 1). *Neonectria* sp. DH2 was clustered together with *Cylindrocarpon/Neonectria* genus and was most closely related to *Neonectria ramulariae* (Deng et al. 2015).

Most filamentous fungi produce secondary metabolites with diverse biological activities such as cholesterol-lowering, anti-tumor, and antibiotic activities. Based on domain searching of the putative biosynthetic core genes and associated genes, 47 BGCs were predicted by antiSMASH (Table 3 and Additional file 1: Table S1): 16 polyketide synthase (PKS), 15 nonribosomal peptide synthetase (NRPS), 6 polyketide–nonribosomal peptide hybrid synthase (PKS-NRPS) and 4 terpene synthase genes. Of the 47 BGCs, 5 BGCs show similarities with previously reported gene clusters (Table 4 and Additional file 2). The antiSMASH prediction suggests that *Neonectria* sp. DH2 may produce small molecules similar to desmethylbassianin (Fisch et al. 2011), bikaverin (Arndt et al. 2015), fujikurins (Von Bargen et al. 2015), fusaric

**Fig. 1** Neighbor-joining phylogenetic tree of 18 s rRNA gene sequences of *Neonectria* sp. DH2 and its taxonomic neighbors. *Aspergillus oryzae* was used as the outgroup. The evolutionary history was inferred using the Neighbor-Joining method. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The analysis involved 35 nucleotide sequences. Codon positions included were 1st + 2nd + 3rd + Noncoding. All positions containing gaps and missing data were eliminated. There are 235 positions in the final dataset. Evolutionary analyses were by MEGA7



**Table 3** The BGCs of *Neonectria* sp. DH2 predicted by antiSMASH

Type	Count
PKS	15
NRPS	13
T1PKS-NRPS hybrid	6
Terpene	5
Others	8
total	47

acid (Brown et al. 2012) and LL-Z1272 $\beta$  (Li et al. 2016). The other 42 BGCs do not have significant homologous to previously reported BGCs, suggesting that *Neonectria* sp. DH2 has a remarkable potential to produce many novel secondary metabolites that might have interesting biological activities.

**Table 4** Comparison of 5 BGCs from *Neonectria* sp. DH2 with previously reported BGCs

BGCs from <i>Neonectria</i> sp. DH2	Reported BGCs	Percentage of similarity (%)
cluster3	Desmethylbassianin biosynthetic gene cluster	60
cluster14	Bikaverin biosynthetic gene cluster	42
cluster17	Fujikurins biosynthetic gene cluster	33
cluster 22	Fusaric acid biosynthetic gene cluster	45
cluster 37	LL-Z1272 $\beta$ biosynthetic gene cluster	100

## Genome accession number and CCTCC Patent number

The whole genome project has been deposited at DDBJ/EMBL/Genbank under the accession RQWH00000000 (bioproject PRJNA507358). The strain has been submitted to China Center for Type Culture Collection (CCTCC) for patent and reservation under the patent number CCTCC M 2015499.

**Acknowledgements** This study was supported by the Science and Technology Program of Guangzhou (No. 201804010476), National Natural Science Foundation of Guangdong (No. 2018A030313334) and Fundamental Research Funds for the Central Universities (No. 17lgpy64).

## Compliance and ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Arndt B, Studt L, Wiemann P, Osmanov H, Kleigrew K, Kohler J, Krug I, Tudzynski B, Humpf HU (2015) Genetic engineering, high resolution mass spectrometry and nuclear magnetic resonance spectroscopy elucidate the bikaverin biosynthetic pathway in *Fusarium fujikuroi*. *Fungal Genet Biol* 84:26–36. <https://doi.org/10.1016/j.fgb.2015.09.006>
- Brown DW, Butchko RAE, Busman M, Proctor RH (2012) Identification of gene clusters associated with fusaric acid, fusarin, and perithecial pigment production in *Fusarium verticillioides*. *Fungal Genetics Biol* 49:521–532. <https://doi.org/10.1016/j.fgb.2012.05.010>
- Calic I, Koch J, Carey D, Addo-Quaye C, Carlson JE, Neale DB (2017) Genome-wide association study identifies a major gene for beech bark disease resistance in American beech (*Fagus grandifolia* Ehrh.). *BMC Genomics* 18:1. <https://doi.org/10.1186/s12864-017-3931-z>
- Deng CH, Scheper Reiny WA, Thrimawithana AH, Bowen JK (2015) Draft genome sequences of two isolates of the plant-pathogenic fungus *Neonectria ditissima* that differ in virulence. *Genome Announc* 3:1–2. <https://doi.org/10.1128/genomeA.01348-15>
- Fisch KM, Bakeer W, Yakasai AA, Song ZS, Pedrick J, Wasil Z, Bailey AM, Lazarus CM, Simpson TJ, Cox RJ (2011) Rational domain swaps decipher programming in fungal highly reducing polyketide synthases and resurrect an extinct metabolite. *J Am Chem Soc* 133:16635–16641. <https://doi.org/10.1021/ja206914q>
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>
- Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874. <https://doi.org/10.1093/molbev/msw054>
- Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35:3100–3108. <https://doi.org/10.1093/nar/gkm160>
- Li C, Matsuda Y, Gao H, Hu D, Yao XS, Abe I (2016) Biosynthesis of LL-Z1272beta: discovery of a new member of NRPS-like enzymes for aryl-aldehyde formation. *ChemBioChem* 17:904–907. <https://doi.org/10.1002/cbic.201600087>
- Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964. <https://doi.org/10.1093/nar/25.5.955>
- Maja Tarailo-Graovac, Nansheng Chen (2009) Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinf* 1:1. <https://doi.org/10.1002/0471250953.bi0410s25>
- Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R (2011) antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res* 39:W339–346. <https://doi.org/10.1093/nar/gkr466>
- Nirma C, Eparvier V, Stien D (2014) Antibacterial ilicicolinic acids C and D and ilicicolinal from *Neonectria discophora* SNB-CN63 isolated from a termite nest. *J Nat Prod* 78:159–162. <https://doi.org/10.1021/np500080m>
- Oliver T, Schmidt B, Nathan D, Clemens R, Maskell D (2005) Using reconfigurable hardware to accelerate multiple sequence alignment with ClustalW. *Bioinformatics* 21:3431–3432. <https://doi.org/10.1093/bioinformatics/bti508>
- Ramsfield TD, Power MWP, Kimberley MO (2013) The relationship between pruning and the incidence of *Neonectria fockeliana* in *Pinus radiata*. *N Z J For Sci*. <https://doi.org/10.1186/1179-5395-43-13>
- Ren JW, Zhang F, Liu XY, Li L, Liu G, Liu XZ, Che YS (2012) *Neonectrolide A*, a new oxaphenalenone spiroketal from the fungus *Neonectria* sp. *Org Lett* 14:6226–6229. <https://doi.org/10.1021/ol302979f>
- Shiono Y, Shimanuki K, Hiramatsu F, Koseki T, Tetsuya M, Fujisawa N, Kimura K (2008) Pyrrospirones A and B, apoptosis inducers in HL-60 cells, from an endophytic fungus, *Neonectria ramulariae* Wollenw KS-246. *Bioorg Med Chem Lett* 18:6050–6053. <https://doi.org/10.1016/j.bmcl.2008.10.032>
- Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>
- Stanke M, Steinkamp R, Waack S, Morgenstern B (2004) AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res* 32:W309–W312. <https://doi.org/10.1093/nar/gkh379>
- Stifanic M, Batel R (2007) Genscan for Arabidopsis is a valuable tool for predicting sponge coding sequences. *Biologia* 62:124–127. <https://doi.org/10.2478/s11756-007-0037-0>
- Von Bargaen KW, Niehaus EM, Krug I, Bergander K, Wurthwein EU, Tudzynski B, Humpf HU (2015) Isolation and Structure Elucidation of Fujikurins A-D: products of the PKS19 Gene Cluster in *Fusarium fujikuroi*. *J Nat Prod* 78:1809–1815. <https://doi.org/10.1021/np5008137>
- Wenneker M, de Jong PF, Joosten NN, Goedhart PW, Thomma BPHJ (2017) Development of a method for detection of latent European fruit tree canker (*Neonectria ditissima*) infections in apple and pear nurseries. *Eur J Plant Pathol* 148:631–635. <https://doi.org/10.1007/s10658-016-1115-3>
- Yang HR, Hu XP, Jiang CJ, Qi J, Wu YC, Li W, Zeng YJ, Li CF, Liu SX (2015) Diversity and antimicrobial activity of endophytic fungi isolated from *Cephalotaxus hainanensis* Li, a well-known medicinal plant in China. *Lett Appl Microbiol* 61:484–490. <https://doi.org/10.1111/lam.12483>
- Zhong LY, Zou L, Tang XH, Li WF, Li X, Zhao G, Zhao JL (2017) Community of endophytic fungi from the medicinal and edible plant *Fagopyrum tataricum* and their antimicrobial activity. *Trop J Pharm Res* 16:387–396. <https://doi.org/10.4314/tjpr.v16i2.18>