Genome sequences of *Knoxdaviesia capensis* and *K. proteae* (Fungi: Ascomycota) from *Protea* trees in South Africa

Janneke Aylward1*, Emma T. Steenkamp2, Léanne L. Dreyer1, Francois Roets3, Brenda D. Wingfield4 and Michael J. Wingfield2

Abstract

Two closely related ophiostomatoid fungi, *Knoxdaviesia capensis* and *K. proteae*, inhabit the fruiting structures of certain *Protea* species indigenous to southern Africa. Although *K. capensis* occurs in several *Protea* hosts, *K. proteae* is confined to *P. repens*. In this study, the genomes of *K. capensis* CBS139037 and *K. proteae* CBS140089 are determined. The genome of *K. capensis* consists of 35,537,816 bp assembled into 29 scaffolds and 7940 predicted protein-coding genes of which 6192 (77.98 %) could be functionally classified. *K. proteae* has a similar genome size of 35,489,142 bp that is comprised of 133 scaffolds. A total of 8173 protein-coding genes were predicted for *K. proteae* and 6093 (74.55 %) of these have functional annotations. The GC-content of both genomes is 52.8 %.

Keywords: *Knoxdaviesia*, Gondwanamycetaceae, Microascales, Ophiostomatoid fungi, *Protea*

Introduction

Two lineages of the polyphyletic assemblage known as ophiostomatoid fungi [1] are associated with the fruiting structures (infructescences) of serotinous *Protea* L. plants [2]. *Protea* species are a key component of the fynbos vegetation in the Core Cape Subregion (CCR) of South Africa [3] and the genus is predominantly encountered in South Africa [4, 5]. The *Protea*-associated ophiostomatoid fungi are, therefore, believed to be endemic to this region, similar to their hosts. This association of ophiostomatoid fungi with a keystone plant genus in a biodiversity hotspot is intriguing [6], as many ophiostomatoid fungi are notorious pathogens of trees [7–10], yet the *Protea* ophiostomatoid species are not associated with disease symptoms [11].

Ophiostomatoid fungi are characterized by the flask-shaped morphology of their sexual fruiting structures and their association with arthropods [1, 12]. The *Protea*-associated members of this assemblage are primarily dispersed by mites that come into contact with fungal spores in the *Protea* infructescences [13, 14]. These mites have limited dispersal ability, but use beetles and possibly larger vertebrates (such as birds) as vehicles for long-distance dispersal [15, 16].

The three *Knoxdaviesia* M.J. Wingf., P.S. van Wyk & Marasas species associated with *Protea* have intriguing host ranges. *K. capensis* M.J. Wingf. & P.S. van Wyk occurs in at least eight different *Protea* hosts, whereas *K. proteae* M.J. Wingf., P.S. van Wyk & Marasas and *K. wingfieldii* (Roets & Dreyer) Z.W. de Beer & M.J. Wingf. are confined to single host species, respectively *P. repens* L. and *P. caffra* Meisn.[17–20]. An investigation of the population biology of *K. proteae*, revealed that this fungus has a high level of intra-specific genetic diversity and that it is extensively dispersed within the CCR of South Africa [16, 21]. However, other than host range and dispersal mechanisms, little is known about the biology and ecology of *Knoxdaviesia* in general [11]. Here we present the description of the first drafts of the genome sequences of the two CCR species, *K. capensis* and *K. proteae*, as well as their respective annotations.

Organism information

Classification and features

The one lineage of *Protea*-associated ophiostomatoid fungi resides in the *Ophiostomataceae* (Ophiostomatales,
Ascomycota), while the second resides in the Gondwanamycetaceae (Microascales, Ascomycota) [11, 22]. The latter group includes three closely related Protea-associated species in the genus Knoxdaviesia (Fig. 1). This genus was initially described to accommodate the asexual state of the first species in the genus, *K. proteae* [23]. Under the dual nomenclature system of fungi, the sexual state of this fungus was described in the same paper as *Ceratocystis proteae* M.J. Wingf., P.S. van Wyk & Marasas [23]. A new genus, *Gondwanamyces* G.J. Marais & M.J. Wingf., was later described to accommodate the sexual state of this species and that of another species, *Ophiostoma capense* M.J. Wingf. & P.S. van Wyk [24]. The asexual states of both remained to be treated as species of *Knoxdaviesia*. Since the abolishment of the dual nomenclature system of fungi, the oldest genus name takes preference, irrespective of morph [25, 26]. The name *Knoxdaviesia*, therefore, has priority and all species previously treated in *Gondwanamyces* were transferred to *Knoxdaviesia* [27].

In a study determining the genome sequence of any fungus, it is advisable to use a living isolate connected to the type specimen. However, the ex-type isolate of *K. proteae* (CMW738 = CBS486.88) is more than 20 years old and does not display the characteristic morphological features of the fungus in culture anymore. No living ex-type isolate exists for *K. capensis*. We thus collected fresh isolates of both species for this study in order to eliminate possible mutations or degradation that may have occurred though continual artificial propagation in culture media. The new isolates (Figs. 1 & 2) were collected from the same localities and hosts as the holotype specimens: *K. capensis* (CMW40890 = CBS139037) from the infructescences of *P. longifolia* Andrews in Hermanus, and *K. proteae* (CMW40880 = CBS140089) from *P. repens* infructescences in Stellenbosch, both locations in the Western Cape Province of South Africa. General features of these isolates are outlined in Table 1.

### Genome sequencing information

#### Genome project history

Considering the lack of ecological information on the genus *Knoxdaviesia* and the close relationship these Microascalean fungi have to important plant pathogens, two Protea-associated *Knoxdaviesia* species, believed to be native to the CCR in South Africa, were selected for genome sequencing. Both species were sequenced at Fasteris in Switzerland. The genome projects are listed in the Genomes OnLine Database [28] and the whole genome shotgun (WGS) project has been deposited at DDBJ/EMBL/GenBank (Table 2). Table 2 presents the project information and its association with the minimum information about a genome sequence version 2.0 compliance [29]. The full MIGS records for *K. capensis* and *K. proteae* are available in Additional file 1: Table S1 and Additional file 2: Table S2, respectively.

#### Growth conditions and genomic DNA preparation

Both *K. capensis* and *K. proteae* were cultured on Malt Extract Agar (MEA; Merck, Wadeville, South Africa) overlaid with sterile cellophane sheets (Product no.
Table 1 Classification and general features of *K. capensis* and *K. proteae* [29]

| MIGS ID | Property               | *K. capensis* Term                          | *K. proteae* Term                          | Evidence code a |
|---------|------------------------|---------------------------------------------|-------------------------------------------|-----------------|
|         | Classification         | Domain Fungi                                | Domain Fungi                              | TAS [19, 23]    |
|         |                        | Phylum Ascomycota                           | Phylum Ascomycota                         | TAS [19, 23]    |
|         |                        | Class Sordariomycetes                       | Class Sordariomycetes                     | TAS [19, 23]    |
|         |                        | Order Microascales                          | Order Microascales                        | TAS [22]        |
|         |                        | Family Gondwanamycetaceae                   | Family Gondwanamycetaceae                 | TAS [22]        |
|         |                        | Genus *Knoxdaviesia*                        | Genus *Knoxdaviesia*                      | TAS [27]        |
|         |                        | Species *K. capensis*                       | Species *K. proteae*                      | TAS [27]        |
|         |                        | Strain: CMW40890 = CBS139037                | Strain: CMW40880 = CBS140089              |                 |
|         | Cell shape             | septate, smooth-walled hyphae              | septate, smooth-walled hyphae             | TAS [19, 23]    |
|         | Motility               | Non-motile                                  | Non-motile                                | NAS             |
|         | Sporulation            | Unsheathed allantoid ascospores             | Falcate ascospores                        | TAS [19, 23]    |
|         | Temperature range       | 15–30 °C                                    | 15–30 °C                                  | TAS [19, 23]    |
|         | Optimum temperature    | 25 °C                                       | 25 °C                                     | TAS [19, 23]    |
|         | pH range; Optimum      | Unknown                                     | Unknown                                   |                 |
|         | Carbon source          | Unknown                                     | Unknown                                   |                 |
|         | MIGS-6 Habitat         | Seed cones (infructescences) of Protea spp. | Seed cones (infructescences) of *Protea repens* L. | TAS [19, 23]    |
|         | MIGS-6.3 Salinity      | Unknown                                     | Unknown                                   |                 |
|         | MIGS-22 Oxygen requirement | Aerobic; requirement/tolerance unknown    | Aerobic; requirement/tolerance unknown    |                 |
|         | MIGS-15 Biotic relationship | Plant-associated                        | Plant-associated                         |                 |
|         | MIGS-14 Pathogenicity  | None known                                  | None known                                |                 |
|         | MIGS-4 Geographic location | Hermanus, South Africa                  | Stellenbosch, South Africa                |                 |
|         | MIGS-5 Sample collection | February 2014                              | January 2014                              |                 |
|         | MIGS-4.1 Latitude      | -34.4093                                    | -33.9430                                  |                 |
|         | MIGS-4.2 Longitude     | 19.2150                                     | 18.8802                                   |                 |
|         | MIGS-4.4 Altitude      | 20 m                                        | 140 m                                     |                 |

a Evidence codes - IDA inferred from direct assay, TAS traceable author statement (i.e., a direct report exists in the literature), NAS non-traceable author statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from http://www.geneontology.org/GO.evidence.shtml of the Gene Ontology project [58].
Z377597, Sigma-Aldrich, Steinham, Germany). After 10 days of growth at 25 °C, mycelia was scraped from the cellophane and DNA was extracted according to Aylward et al. [30]. Approximately 5 μg DNA from each species was used to prepare the three Illumina libraries (Table 2).

RNA was extracted from the K. proteae genome isolate to use as evidence for gene prediction. After growth on MEA at 25 °C for approximately 10 days, total RNA was isolated from the mycelia with the PureLink™ RNA Mini Kit (Ambion, Austin, TX, USA). Quality control was performed on the Agilent 2100 Bioanalyzer (Agilent Technologies, USA) using the RNA 6000 Nano Assay kit (Agilent Technologies, USA). The mRNA component of the total RNA was subsequently extracted with the Dynabeads® mRNA purification kit (Ambion, Austin, TX, USA).

### Genome sequencing and assembly

The genomes of K. capensis and K. proteae were sequenced with the Illumina HiSeq 2500 platform at Fasteris,

| Table 2 Project information |
|--------------------------------|
| **MIGS ID** | **Property** | **K. capensis Term** | **K. proteae Term** |
| MIGS 31 | Finishing quality | High quality draft | High quality draft |
| MIGS-28 | Libraries used | 2x paired-end (PE) (350 and 550 bp) and 1x mate-pair (MP) (3 kbp) | 2x paired-end (PE) (350 and 550 bp) and 1x mate-pair (MP) (3 kbp) |
| MIGS 29 | Sequencing platforms | Illumina Hiseq 2500 | Illumina Hiseq 2500 |
| MIGS 31.2 | Fold coverage | PE library 1: 91.6 x | PE library 1: 142 x |
| | | PE library 2: 80 x | PE library 2: 79.3 x |
| | | MP library: 17 x | MP library: 50.2 x |
| MIGS 30 | Assemblers | ABYSS 1.5.2; SSPACE 3.0 | ABYSS 1.5.2; SSPACE 3.0 |
| MIGS 32 | Gene calling method | MAKER 2.31.8 | MAKER 2.31.8 |
| GenBank ID | LNGK00000000 | LNGL00000000 |
| GenBank Date of Release | 11th January 2016 | 11th January 2016 |
| GOLD ID | Gp0093999 | Gp0110284 |
| BIOPROJECT | PRJNA246171 | PRJNA275563 |
| MIGS 13 | Source Material Identifier | CMW40890/CBS139037 | CMW40880/CBS140089 |
| Project relevance | Biodiversity, evolution | Biodiversity, evolution |

| Table 3 Genome statistics |
|----------------------------|
| **Species** | **K. capensis** | **K. proteae** |
| Attribute | Value | % of Total | Value | % of Total |
| Genome size (bp) | 35,537,816 | 100.00 | 35,489,142 | 100.00 |
| DNA coding (bp) | 12,640,368 | 35.57 | 12,542,580 | 35.34 |
| DNA G + C (bp) | 18,774,628 | 52.83 | 18,745,365 | 52.82 |
| DNA scaffolds | 29 | | 133 | |
| Total genes | 8107 | 100.00 | 8316 | 100.00 |
| Protein coding genes | 7940 | 97.94 | 8173 | 98.28 |
| RNA genes | 167 | 2.06 | 143 | 1.72 |
| Pseudo genes | unknown | | unknown | |
| Genes in internal clusters | unknown | | unknown | |
| Genes with function prediction | 6192 | 77.98 | 6093 | 74.55 |
| Genes assigned to KOGs | 6059 | 76.31 | 6015 | 73.60 |
| Genes with Pfam domains | 5455 | 68.70 | 5335 | 65.28 |
| Genes with signal peptides | 354 | 4.46 | 335 | 4.10 |
| Genes with transmembrane helices | 1510 | 19.02 | 1527 | 18.68 |
| CRISPR repeats | N/A | | N/A | |

*The total is based on either the size of the genome in base pairs or the total number of protein-coding genes in the annotated genome.

**Based on tRNA and rRNA genes only.**
Switzerland, using two paired-end and one Nextera mate-pair library (Table 2). More than 60 million paired-end and 8 million mate-pair reads were obtained for each species. These reads were trimmed in CLC Genomics Workbench 6.5 (CLC bio, Aarhus, Denmark) so that the Phred Q (quality) score of each base was at least Q20. VelvetOptimiser (Gladman & Seeman, unpublished), a Perl script used as part of the Velvet assembler [31, 32], was initially used to optimize the assembly parameters. Assembly of contigs was performed in ABySS 1.5.2 [33] using the optimal parameters suggested by VelvetOptimiser as a starting point. Several assemblies were computed using kmer-values slightly higher and lower than the kmer-value suggested by VelvetOptimiser. The assembly with the lowest number of contigs was used to build scaffolds in SSPACE 3.0 [34], discarding scaffolds smaller than 1000 bp. Automatic gap closure was performed in GapFiller 1.10 [35]. The average genome coverage of each library was estimated using the Lander-Waterman equation (total sequenced nucleotides/genome size) (Table 2), which yielded a combined average coverage for the three libraries of 188.5x (K. capensis) and 271.5x (K. proteae).

The K. capensis genome consists of 29 scaffolds ranging between 1226 and 5,637,848 bp, whereas the 133 scaffolds of K. proteae are sized between 1022 and 2,610,973 bp. A search for the 1438 fungal universal single-copy ortholog genes with BUSCO 1.1b1 [36] identified 1355 complete and 67 partial genes in K. capensis and 1366 complete and 57 partial genes in K. proteae. The two genomes are therefore estimated to be >98 % complete.

The extracted mRNA of K. proteae was sequenced using an Ion PI™ Chip on the Ion Proton™ System (Life Technologies, Carlsbad, CA) at the Central Analytical Facility (CAF), Stellenbosch University, South
Africa. The >49 million raw RNA-Seq reads were mapped to the \textit{K. capensis} genome in CLC Genomics Workbench and assembled with Trinity 2.0.6 [37] using the genome-guided option.

**Genome annotation**

Genome annotation was performed with the MAKER 2.31.8 pipeline [38, 39], using custom repeat libraries for each species constructed with RepeatScout 1.0.5 [40] and two \textit{de novo} gene predictors, SNAP 2006-07-28 [41] and AUGUSTUS 3.0.3 [42]. The assembled \textit{K. proteae} RNA-Seq and predicted protein and/or transcript sequences from 22 sequenced Sordariomycete species (Additional file 3: Table S3), including two Microascalean fungi, were provided as additional evidence. AUGUSTUS was trained with the assembled \textit{K. proteae} RNA-Seq data and subsequently MAKER was used to annotate the largest scaffold of the \textit{K. capensis} and the largest scaffold of the \textit{K. proteae} assembly, independently. After manually curating all the gene predictions on these scaffolds with Apollo 1.11.8 [43], SNAP was trained with the curated gene predictions of each scaffold and the scaffolds were re-annotated. SNAP was retrained for each species individually and subsequently both genomes were annotated. EuKaryotic Orthologous Group (KOG) classifications were assigned to the predicted proteins through the WebMGA [44] portal that performs reverse-position-specific BLAST [45] searches on the KOG database [46]. Additional functional annotations were predicted with InterProScan 5.13-52.0 [47, 48], SignalP 4.1 [49] and TMHMM 2.0 [50].

**Genome properties**

\textit{K. capensis} and \textit{K. proteae} have similar genome sizes at 35.54 and 35.49 Mbp, respectively. It was possible to assemble the \textit{K. capensis} genome into 29 scaffolds larger than 1000 bp, whereas the number of scaffolds above this threshold achieved for \textit{K. proteae} was 133. Both genomes had a GC content of 52.8%.

A total of 7940 protein-coding genes were predicted for \textit{K. capensis} and 8174 for \textit{K. proteae}. Additionally 137 and 116 tRNA and 30 and 27 rRNA genes were predicted for each species, respectively. More than 74% of the protein-coding genes of each species could be assigned to a putative function via the KOG and Pfam databases. The content of the two genomes are summarized in Tables 3 and 4.

**Conclusions**

At least six Microascalean fungi currently have publically accessible genomes [51–54]. \textit{K. capensis} and \textit{K. proteae}, however, represent the first sequenced genomes from the Microascalean family \textit{Gondwanamycetaceae}. The genomes of these two species will not only enable further understanding of the unique ecology of \textit{Protea}-inhabiting fungi, but will also be valuable in taxonomic and evolutionary studies.

**Additional files**

| Additional file 1 | Table S1. Associated MIGS record for \textit{K. capensis}. (DOC 75 kb) |
|------------------|---------------------------------------------------------------|
| Additional file 2 | Table S2. Associated MIGS record for \textit{K. proteae}. (DOC 73 kb) |
| Additional file 3 | Table S3. Sequenced Sordariomycete fungi used as evidence for genome annotations. (XLSX 12 kb) |

**Abbreviations**

CCR: core cape subregion; MEA: malt extract agar; KOG: EuKaryotic Orthologous Groups of proteins.

**Competing interests**

The authors declare that they have no competing interests.

**Authors’ contributions**

MJW, BDW and ETS conceived the study. LLD and FR supervised the study. JA performed the laboratory work. JA assembled and annotated the genomes with the help of BDW and ETS. JA drafted the manuscript with the help of LLD and FR. ETS revised the manuscript. All authors read and approved the final manuscript.

**Acknowledgements**

We are grateful to Quentin Santana and Dr. Lieschen Bahlmann for their guidance in the genome assembly and annotation procedures and to Dr. Wilhelm de Beer for the taxonomic information he contributed to this manuscript. This research was funded by the National Research Foundation (NRF) and the Department of Science and Technology/NRF Centre of Excellence in Tree Health Biotechnology. We also thank the Cape Nature Conservation Board for supplying the necessary collection permits.

**Author details**

1. Department of Botany and Zoology, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa. 2. Department of Microbiology and Plant Pathology, University of Pretoria, Pretoria 0002, South Africa. 3. Department of Conservation Ecology and Entomology, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa. 4. Department of Genetics, University of Pretoria, Pretoria 0002, South Africa.

**Received:** 13 November 2015  
**Accepted:** 18 February 2016  
**Published online:** 29 February 2016  

**References**

1. Spatafora JW, Blackwell M. The polyphyletic origins of ophiostomatoid fungi. Mycol Res. 1994;98:1–9.
2. Wingfield BD, Wijnen CD, Wingfield MJ. Phylogenetic relationships of ophiostomatoid fungi associated with Protea in Africam. Mycol Res. 1999;103:1616–20.
3. Manning J, Goldblatt P. Plants of the greater Cape floristic region, 1: The Cape Floristic Region. South African National Biodiversity Institute; 2012.
4. Gibbs RG. Analysis of the size and composition of the southern African flora. Bothalia. 1984;15:613–29.
5. Rebelo T. Proteae: a field guide to the Proteas of Southern Africa. Vlaeberg: Fernwood Press; 1995.
6. Mittemeier RA, Myers N, Thomsen JB, Da Fonseca GAB, Oliwier S. Biodiversity hotspots and major tropical wilderness areas: approaches to setting conservation priorities. Conserv Biol. 1999;12:516–20.
7. Paine TD, RaffaKF, Harrington TC. Interactions among scolytid bark beetles, their associated fungi, and live host conifers. Annu Rev Entomol. 1997;42:179–206.
8. Brasier CM. Ophiostoma novo-ulmi sp. nov., causative agent of current Dutch elm disease pandemics. Mycopathologia. 1991;115:151–61.
