A Highly Contiguous Genome Assembly of Arthrinium puccinoides

Sørensen, Trine; Petersen, Celine; Fechete, Lavinia Ioana; Nielsen, Kåre Lehmann; Sondergaard, Teis Esben

Published in:
Genome Biology and Evolution

DOI (link to publication from Publisher):
10.1093/gbe/evac010

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

A Highly Contiguous Genome Assembly of *Arthrinium puccinoides*

Trine Sørensen, Celine Petersen, Lavinia I. Fechete, Kåre L. Nielsen*, and Teis E. Sondergaard*

Department of Chemistry and Bioscience, Aalborg University, Denmark

*Corresponding authors: E-mails: tes@bio.aau.dk; kln@bio.aau.dk.

Accepted: January 15, 2022

Abstract

The phylogenetic relationship of the *Arthrinium* genus has changed throughout the years. For many years, the *Arthrinium* genus included the *Apiospora* genus as well. New evidence has now showed that these two genera in fact are phylogenetically different and belong to two different clades. Here, we present the first genome draft within the *Arthrinium* genus. This genome was sequenced using the MinION platform from Oxford Nanopore Technologies and the assembly was contiguous. The assembly comprises ten contigs totaling 39.8 Mb with an N50 length of 7.9. In the assembly, 11,602 genes were predicted whereof 10,784 were functionally annotated. A total of 37 rRNA genes were observed in the assembly and repeat elements spanning 7.39% of the genome were found. A total of 99 secondary metabolite gene clusters were predicted, showing a high potential of novel secondary metabolites. This genome sequence will not only be useful for further investigation of the *Arthrinium* clade, but also for discovery of novel secondary metabolite compounds that could be of high interest for the food, agricultural, or pharmaceutical industry.

Key words: MinION, de novo assembly, genome annotation, secondary metabolites, phylogenetics.

Significance

The fungal clade of *Arthrinium*, which mainly is observed in temperate, cold, and alpine areas, was seen as one group containing not only the *Arthrinium* genus, but also the genus of *Apiospora*. Recent evidence, however, separate these two clades in to two different new clades as they were shown to be phylogenetic different. No genomes of this new *Arthrinium* clade are currently available. In this work, we present the first available genome draft of an *Artirinium* sp. Our results support the recently reported separation of the current *Arthrinium* clade into the two new clades—*Arthrinium* and *Apiospora*.

Introduction

The genus *Arthrinium* has had a schizophrenic existence located in different genera as time has passed. *Arthrinium* is mainly found in temperate, cold, and alpine areas (Pintos and Alvarado 2021), which was also the case for *Arthrinium caricicola* that is the first proposed and validated *Arthrinium* (Kunze and Schmidt 1817; Fries 1832). Due to different conidial shape, some *Arthrinium* spp. were considered to belong to other genera, such as *Camptoum*, *Gonatosporium*, *Goniosporium*, and *Sporophleum* (Willdenow 1824), but later again associated with *Arthrinium* (Höhnel 1925; Cooke 1954). Other different genera synonymized with *Arthrinium* are described in literature (Johnston and Stevenson 1917; Subramanian 1956; Hughes 1953; Cole and Samson 1979; Minter 1985). Notably, one of these is the association with *Apiospora* through *Papularia* Fr. because it is the asexual morph of *Apiospora*. Hence, the biological relation between *Arthrinium* and *Apiospora* was established (Saccardo 1875; Höhnel 1919; Petrak 1925; Hudson 1960, 1963; Ellis 1965). Subsequently, other genera were shown to have a biological relation to *Apiospora* and therefore likewise considered as synonyms of *Arthrinium* (Liu et al. 1976; Dyko and Sutton 1979; Samuels et al. 1981; Sivanesan 1983; Kirk 1986).

With the use of genetic data, Hyde et al. (1998), placed *Arthrinium* and *Apiospora* in their own family named *Apiosporaceae*. However, Crous and Groenewald (2013), simply suggest *Arthrinium* as genus name for them both based on the one fungus = one name policy. Both Crous and Groenewald (2013) and Wang et al. (2018) genetically identified multiple species of *Arthrinium* (now known as *Apiospora*). However, when Pintos et al. (2019), published marker genes from *Arthrinium cariciola*, *Arthrinium curvatum*, and *Arthrinium sporophleum*, they noted that these *Arthrinium* together with *Arthrinium japonicum* and *Arthrinium puccinioides* clustered in a clade separately from all other *Arthrinium* (now known as *Apiospora*). Thus, they proposed that *Apiospora* and *Arthrinium* could be phylogenetically different, but that further data were needed (Pintos et al. 2019). Shortly thereafter, the phylogenetic delimitation of *Arthrinium* and *Apoisopra* was confirmed by Pintos and Alvarado (2021).

The four *Arthrinium* genomes available in NCBI (*Apiospora malaysianum* [ASM650811v1], *Apiospora* KUC21332 [ASM1716395v1], *Apiospora saccharicola* [ASM1900006v1], and *Apiospora phaeospermum* [ASM650353v1]) are in fact *Apiospora* according to Pintos and Alvarado (2021), meaning that no genome draft is published for *Arthrinium*. We therefore present the first genome draft of an *Arthrinium* spp. and hereby give a genetic resource useful for genomic and evolutionary studies. This genome draft also gives insight into the biosynthesis potential of secondary metabolites of this genus which could be of great interest for discovery of novel compounds. We furthermore support the finding of Pintos and Alvarado (2021), through phylogenetic relation between *Arthrinium* and *Apoispora*.

Results and Discussion

Genome Sequencing and Assembly

A total of 3,325,171,178 bases, across 211,994 reads, were generated during Oxford Nanopore Technologies sequencing. Of these, 2,794,209,936 bases and 106,349 reads were used in the de novo assembly of *Arthrinium puccinoides* after quality filtering. The N50 of the read length and the mean read quality were 31,689 bases and 12 Phred score, respectively.

The de novo assembly resulted in a draft genome size of 39.8 Mb spanning 10 contigs (table 1). N50 of the assembly length is 7.9 Mb whereas N99 is 2.3 Mb, suggesting a highly contiguous assembly. Furthermore, Benchmarking Universal Single-Copy Orthologs (BUSCO) was used to assess the assembly completeness in gene space and 97.6% complete BUSCOs were observed, indicating an almost complete gene content. A low observed number of duplicated and fragmented BUSCOs further support a high-quality assembly.

Table 1
Assembly and Genome Annotation Statistics

Genome Characteristics		Value
Assembly	Draft genome size (Mb)	39.8
	Coverage (×)	70.2
	Number of contigs	10
	Longest contig (Mb)	9.5
	N50 (Mb)	7.9
	N99 (Mb)	2.3
	GC (%)	53
BUSCO	Complete BUSCOs (%)	97.6
	Complete and single-copy BUSCOs (%)	97.3
	Complete and duplicated BUSCOs (%)	0.3
	Fragmented BUSCOs (%)	0.8
	Missing BUSCOs (%)	1.6
Annotation	Number of predicted genes	11,602
	Number of predicted genes with a function	10,784

Genome Annotation and Repeat Elements

Annotation of the genome resulted in the identification of 11,602 protein-coding genes of which 10,784 genes (92.95%) could be assigned a function bioinformatically (table 1). Noncoding rRNA genes and tRNA genes were predicted across the genome comprising 37 rRNA genes and 174 tRNA genes. Repetitive elements were furthermore observed to span 7.39% of the genome where 1.73% and 1.86% of these were identified as tandem repeats and long-terminal repeat retrotransposons, respectively. Long-interspersed elements account for 0.31% and short-interspersed elements account for 0.13%.

Comparative Genome Analysis

Phylogenetic trees based on whole-genome alignment and alignment of orthogroups were constructed (fig. 1*A*) to support the phylogenetic delimitation of *Arthrinium* and *Apiospora* as proposed by Pintos and Alvarado (2021). The four *Apiospora* spp. previously stated as *Arthrinium* spp. (*Apiospora malaysianum* [ASM650811v1], *Apiospora* KUC21332 [ASM1716395v1], *Apiospora saccharicola* [ASM1900006v1], and *Apiospora phaeospermum* [ASM650353v1]) forms in both cases a separated cluster from *Arthrinium puccinoides* supporting the phylogenetic delimitation made with marker genes by Pintos and Alvarado (2021). Notably, the two phylogenetic trees are similar even though the phylogenetic tree based on whole-genome alignment contains both genetic and intronic information whereas the tree based orthogroups contains genetic information only.

Secondary Metabolite Potential

Arthrinium puccinoides shows high capacity for synthesizing numerous novel secondary metabolites (fig. 1*B*). A total of 99 secondary metabolite genes were predicted in the genome draft assembly, showing the highest number of predicted

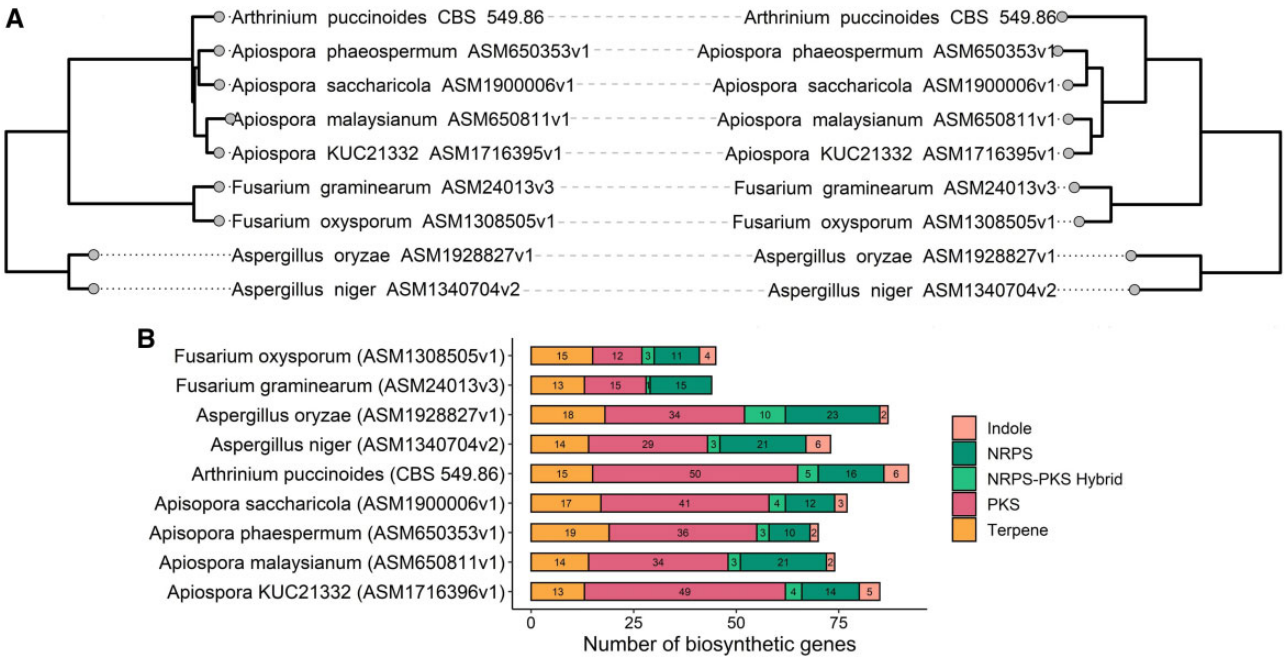

Fig. 1.—(A) The phylogenetic relation between *Apiospora* spp. and *Arthrinium puccinoides*. The *Apiospora* spp. were formerly known as *Arthrinium* spp. *Fusarium graminearum* (ASM24013v3) and *Fusarium oxysporum* (ASM1308505v1) were included as well. The phylograms were rooted with *Aspergillus niger* (ASM1928827v1) and *Aspergillus oryzae* (ASM1340704v2). A phylogenetic relationship based on whole-genome alignment is showed on the left, whereas a phylogenetic relationship based on alignment of 5,057 orthogroups is showed on the right. (*B*) Number of each type of predicted secondary metabolite genes in the nine different genomes using AntiSMASH.

secondary metabolite genes compared with the other genome draft assemblies include in this report. Variants of different classes were predicted: 50 to be polyketide synthases, 16 to be nonribosomal peptide synthetases, 15 to be terpenes, and six to be indoles. Only a few of the secondary metabolite gene clusters could be predicted to code for a known secondary metabolite. With this in mind and the newly delimitation of the two clades (*Arthrinium* and *Apiospora*), it opens up for a whole new cluster of fungus to investigate and potential make discovery of novel compounds of high potential from this clade.

Materials and Methods

Growth Conditions, DNA Extraction, and Genome Sequencing

Arthrinium puccinoides (CBS 549.86) was obtained from the Westerdijk Fungal Biodiversity Institute (Utrecht, The Netherlands). The fungus was cultivated on solid Yeast Peptone Glucose (YPG) (10 g/l yeast extract, 20 g/l peptone, and 25 g/l glucose) at 25°C for 5 days, and transferred to liquid YPG media and grown in a rotary shaker for 4 days at 25°C and 100 rpm. The biomass was filtered using a microfilter, followed by lyophilization and subsequently grounded in a mortar.

Genomic DNA was extracted from the lyophilized and grounded mycelium using the Genomic Buffer Set and the Genomic-tips 20/G according to the manufacturer's protocol. The quality and quantity of the extracted DNA were evaluated using NanoDrop One (ThermoFisher), Qubit 3.0 (Invitrogen) with Qubit dsDNA HS Assay Kit, and 2200 TapeStation (Agilent) with Genomic DNA ScreenTape Analysis according to the manufacturer's instructions.

A library was constructed using the Genomic DNA by ligation (SQK-LSK109) protocol from Oxford Nanopore Technologies (Oxford, UK) and sequenced on a R9.4.1 flowcell.

De Novo Assembly and Genome Annotation

The raw data was basecalled using Guppy version 3.4.5 (Oxford Nanopore Technologies 2021) in GPU mode using the dna_r9.4.1_450bps_hac.cfg model. The reads were filtered using Filtlong version 0.2.0 (Wick 2018) to a minimum length of 10 kb and a minimum basecall quality of 80 (Q7). Minimap2 version 2.17 (Li 2018) and Miniasm version 0.3 (Li 2016) were used to create the assembly, which subsequently were polished using Racon version 1.3.3 (Vaser et al. 2017) with default settings and two rounds of Medaka version 1.0.1 (Oxford Nanopore Technologies 2018) with default settings. The completeness was assessed with BUSCO version 3.0.2 (Seppey et al. 2019) using the Ascomycota BUSCO data set.

The genome was annotated using AUGUSTUS version 3.4.0 (Stanke et al. 2004) and InterProScan version 5.38-76.0 (Jones et al. 2014) was used for functional annotation

with default settings and the following databases MobiDBLite, Pfam, CDD, ProSiteProfiles, PANTHER, Gene3D, ProSitePatterns, PRINTS, SUPERFAMILY, TIGRFAM, SMART, Coils, PIRSF, Hamap, and SFLD. Noncoding RNA genes were predicted using Barrnap version 0.9 (Seemann 2018) and tRNA-scan version 2.0.5 (Chan and Lowe 2019). Repeat sequences were identified using RepeatMasker version 4.1.2 (Tarailo-Graovac and Chen 2009) with eukaryote as species.

Comparative Genome Analysis and Prediction of Secondary Metabolite Genes

Eight genomes were downloaded from NCBI (*Apiospora malaysianum* [ASM650811v1], *Apiospora* KUC21332 [ASM1716395v1], *Apiospora saccharicola* [ASM1900006v1], *Apiospora phaeospermum* [ASM650353v1]), *Fusarium graminearum* (ASM24013v3), *Fusarium oxysporum* (ASM1308505 v1), *Aspergillus oryzae* (ASM1928827v1), and *Aspergillus niger* (ASM1340704v2). Whole-genome alignment based on entire genomes in nucleotide space of the eight genomes and *A. puccinoides* genome was generated in CLC Genomics Workbench version 20.0 (Qiagen, Århus) using default setting. The whole-genome alignment was used as input to create an average nucleotide identity comparison with default setting in CLC Genomics Workbench. Afterwards a neighbor joining tree was constructed from the average nucleotide identity comparison in CLC Genomics Workbench. OrthoFinder (Emms and Kelly 2019) was used to identify orthogroups with default settings. Afterwards, OrthoFinder was used to infer a rooted species tree from 5,057 orthogroups using default settings (Emms and Kelly 2017, 2018). The trees was visualized using R version 4.1.1 in RStudio version 2021.09.0 (R Studio Team 2020) using ggtree 3.0.4 (Yu et al. 2016). Secondary metabolite genes were predicted using AntiSMASH version 5.1 (Blin et al. 2019) and visualized using R version 4.1.1 in RStudio version 2021.09.0 (R Studio Team 2020) using ggplot2 3.3.5 (Wickham 2016).

Acknowledgments

This work was supported by the Novo Nordisk Foundation (Grant No. NNF18OC0034952). The authors thank Elsa Sverrisdóttir for helpful comments on the manuscript.

Author Contributions

T.S.: Conceptualization, methodology, investigation, writing—original draft, writing—review and editing, visualization; C.P.: Conceptualization, methodology, investigation, writing—original draft, writing—review and editing, visualization; L.F.: Investigation; K.L.N.: Conceptualization, supervision, funding acquisition, methodology, review and editing; T.E.S.: Conceptualization, supervision, funding acquisition, methodology, review and editing. All authors have read and agreed to the published version of the manuscript.

Data Availability

The raw reads, the assembly, and the annotation can be found at NCBI BioProject PRJNA764836 with the GenBank accession number JAIUNF000000000 for the assembly and the SRA accession number PRJNA764836 for the raw reads.

Literature Cited

Blin K, et al. 2019. AntiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 47:81–87.
Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol. 1962:1–14.
Cole GT, Samson RA. 1979. Patterns of development in conidial fungi. London: Pitman.
Cooke WB. 1954. The genus arthrinium. Mycologia 46(6):815–822.
Crous PW, Groenewald JZ. 2013. A phylogenetic re-evaluation of *Arthrinium*. IMA Fungus 4(1):133–154.
Dyko B, Sutton B. 1979. New and interesting dematiaceous hyphomycetes from Florida. Mycotaxon 8:119–124.
Ellis MB. 1965. Dematiaceous hyphomycetes. VI. Mycological Papers 103:1–46.
Emms DM, Kelly S. 2018. STAG: Species tree inference from all genes. bioRxvi.
Emms DM, Kelly S. 2017. STRIDE: Species tree root inference from gene duplication events. Mol Biol Evol. 34(12):3267–3278.
Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20(1):238.
Fries EM. 1832. Systema mycologicum. vol. 3(2). Griefswald: Sumptibus Ernesti Mauritii.
Hudson HJ. 1963. Pyrenomycetes of sugar cane and other grasses in Jamaica. II. Conidia of *Apiospora montagnei*. Trans Br Mycol Soc. 46(1):19–23.
Hudson HJ. 1960. Pyrenomycetes of sugar cane and other grasses in Jamaica. I. Conidia of *Apiospora camptospora* and *Leptosphaeria sacchari*. Trans Br Mycol Soc. 43(4):607–616.
Hughes SJ. 1953. Conidiophores, conidia, and classification. Can J Bot. 31(5):577–659.
Hyde K, Fröhlich J, Taylor J. 1998. Fungi from palms. XXXVI. Reflections on unitunicate ascomycetes with apiospores. Sydowia 50:21–80.
Höhnel F von. 1919. Fragmente zur Mykologie. XXII Mitteilungen, nr. 1092 bis 1153. Sitzungsberichte Der Akademie Der Wissenschaften in Wien Mathematisch-Naturwissenschaftlichen Klasse. 128:549–634.
Höhnel F von. 1925. Über die Gattung Arthrinium Kunze. Mitteilungen Aus Dem Botanischen Institut Der Technischen Hochschule in Wien. 2:9–16.
Johnston JR, Stevenson JA. 1917. Sugar-cane fungi and diseases of Porto Rico. J Agric Univ P R. 1:177–264.
Jones P, et al. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30(9):1236–1240.
Kirk PM. 1986. New or interesting microfungi: XV. Miscellaneous hyphomycetes from the British Isles. Trans Br Mycol Soc. 86(3):409–428.
Kunze G, Schmidt JC. 1817. Mykologische hefte. vol. 1. Leipzig: G. Voss.
Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics 32(14):2103–2110.
Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34(18):3094–3100.
Liu B, Li Z, Du F. 1976. A new species of *Arthrinium*. Acta Microbiol Sin. 16:302–303.

Minter DW. 1985. A re-appraisal of the relationships between *Arthrinium* and other hyphomycetes. Proc Indian Acad Sci. 94:281–308.
Oxford Nanopore Technologies. 2018. Medaka. Available from: https://github.com/nanoporetech/medaka [accessed 2021 Sept 21].
Oxford Nanopore Technologies. 2021. pyguppyclient. Available from: https://github.com/nanoporetech/pyguppyclient [accessed 2021 Sept 21].
Petrak F. 1925. Mykologische Notizen VIII. Annales Mycologici. 23:1–143.
Pintos Á, Alvarado P. 2021. Phylogenetic delimitation of *Apiospora* and *Arthrinium*. Fungal Syst Evol. 7:197–221.
Pintos Á, Alvarado P, Planas J, Jarling R. 2019. Six new species of *Arthrinium* from Europe and notes about *A. caricicola* and other species found in *Carex* spp. hosts. MycoKeys 49:15–48.
Samuels GJ, Mc kenzie EHC, Buchanan DE. 1981. Ascomycetes of New Zealand 3. Two new species of *Apiospora* and their *Arthrinium* anamorphs on bamboo. New Zeal J Bot. 19(2):137–149.
Seemann T. 2018. Barrnap. Available from: https://github.com/tseemann/barrnap [accessed 2021 Sept 21].
Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation. Methods Mol Biol. 1962:227–245.
Sivanesan A. 1983. Studies on ascomycetes. Trans Br Mycol Soc. 81(2):313–332.
Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. 32:309–312.
Subramanian CV. 1956. Microtypha. Proc Indian Acad Sci. B. 44(2):122–124.
Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinformatics 25:4.10.1–4.10.14.
R Team Studio. 2020. RStudio: Integrated Development for R. Available from: http://www.rstudio.com/ [accessed 2021 Sept 21].
Saccardo PA. 1875. Conspectus generum pyrenomycetum italicorum additis speciebus fungorum Venetorum novis vel criticis, systemate carpologico dispositorum. Atti Della Società Veneto-Trentina Di Scienze Naturali. 4:77–100.
Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 27(5):737–746.
Wang M, Tan X-M, Liu F, Cai L. 2018. Eight new *Arthrinium* species from China. MycoKeys 34:1–24.
Wick R. 2018. Filtlong. Available from: https://github.com/rrwick/Filtlong [accessed 2021 Sept 21].
Wickham H. 2016. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag.
Yu G, Smith D, Zhu H, Guan Y, Lam T. 2016. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 8:28–36.
Willdenow CL. 1824. Caroli a Linné Species plantarum, vol. 6, pars. 1. Berlin (Germany): GC Nauk.

Associate editor: Li-Jun Ma