DATA NOTE

REVISED A draft reference assembly of the *Psilocybe cubensis* genome [version 2; peer review: 2 approved]

Previously titled: A draft sequence reference of the *Psilocybe cubensis* genome

Kevin McKernan[1], Liam T. Kane[1], Seth Crawford[2], Chen-Shan Chin[3], Aaron Trippe[2], Stephen McLaughlin[1]

[1]R&D, Medicinal Genomics, Beverly, Mass, 01915, USA
[2]R&D, Oregon CBD, Corvallis, Oregon, 97330, USA
[3]Bioinformatics, DNANexus, Mountain View, CA, 94040, USA

v2 **First published:** 09 Apr 2021, **10**:281
https://doi.org/10.12688/f1000research.51613.1
Latest published: 15 Jun 2021, **10**:281
https://doi.org/10.12688/f1000research.51613.2

Abstract

We describe the use of high-fidelity single molecule sequencing to assemble the genome of the psychoactive *Psilocybe cubensis* mushroom. The genome is 46.6Mb, 46% GC, and in 32 contigs with an N50 of 3.3Mb. The BUSCO completeness scores are 97.6% with 1.2% duplicates. The Psilocybin synthesis cluster exists in a single 3.2Mb contig. The dataset is available from NCBI BioProject with accessions PRJNA687911 and PRJNA700437.

Keywords

Psilocybe cubensis, Genome, Single molecule sequencing, Psilocybin

This article is included in the Draft Genomes collection.

Open Peer Review

Reviewer Status ✓ ✓

	Invited Reviewers	
	1	**2**
version 2 (revision) 15 Jun 2021	✓ report	
version 1 09 Apr 2021	? report	✓ report

1. **Jason Slot**, The Ohio State University, Columbus, USA

2. **Anders Goncalves da Silva**, Lighthouse Genomics Inc., British Columbia, Canada
Philippe Henry, Egret Bioscience Ltd., West Kelowna, Canada
Lighthouse Genomics Inc., BC, Canada

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Kevin McKernan (Kevin.McKernan@medicinalgenomics.com)

Author roles: McKernan K: Conceptualization, Data Curation, Methodology, Project Administration, Software, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing; **Kane LT**: Data Curation, Methodology; **Crawford S**: Data Curation, Methodology; **Chin CS**: Software; **Trippe A**: Data Curation, Methodology; **McLaughlin S**: Data Curation, Formal Analysis, Software

Competing interests: No current conflicts but potential future conflicts should be disclosed. Medicinal Genomics offers DNA and RNA sequencing services in the Cannabis space and are likely to offer similar services in the P.cubensis space if market demand matures.

Grant information: The author(s) declared that no grants were involved in supporting this work.

How to cite this article: McKernan K, Kane LT, Crawford S *et al.* **A draft reference assembly of the *Psilocybe cubensis* genome [version 2; peer review: 2 approved]** F1000Research 2021, **10**:281 https://doi.org/10.12688/f1000research.51613.2

First published: 09 Apr 2021, **10**:281 https://doi.org/10.12688/f1000research.51613.1

REVISED **Amendments from Version 1**

To address the reviewers' very helpful comments

1. We have included more descriptions of the SNP calling including a Github version of the code one can run to reproduce this.
2. We have expanded the analysis to include variant calling on the HiFi reads mapped back to their own reference and in doing so recalled the SNPs from the MGC unknown strain mapped to the P.envy reference utilizing the same code for a controlled comparison. This leveraged different variant callers and produced more variants.
3. We have improved the readability of figure tracks.
4. Adjusted Title language, mushroom number and Fungi vs Fungus typo.
5. Addressed the Anexic growth and bacterial contamination concerns.
6. Clarified these citations are to NCBI/SRA submissions which currently do not have a DOI publication associated with them.
7. We have clarified the text to underscore the importance of other variants in the pathway. We have included his suggested reference here and added links to a SNPeff file that can be used to compare these variants to those he listed.
8. In performing SNP calling on the HiFi reads mapped back to the reference we note that 98% of the variants found are heterozygous variants with balanced alleles. Coverage maps can also be viewed in the CoGe genome browser provided to address additional questions regarding aneuploidy/multiple nuclei but coverage for all scaffold is very consistent with the exception of scaffold_26 which is mitochondria.

Any further responses from the reviewers can be found at the end of the article

Introduction

There are hundreds of mushrooms capable of synthesizing the psychoactive compound psilocybin. This compound has been classified as a "breakthrough therapy" for depression by the FDA (Johnson and Griffiths 2017). The psilocybin pathway was identified by Fricke et al., but to date no public references exist in NCBI with N50s longer than 50kb (Fricke et al. 2017; Blei et al. 2018; Fricke et al. 2019a; Fricke et al. 2019b; Blei et al. 2020; Demmler et al. 2020; Fricke et al. 2020). A more contiguous genome assembly can assist in further resolution of the genetic diversity in the fungi's secondary metabolite production.

Methods

DNA isolation

Dried stems from *Psilocybe cubensis* strain P.envy. The strain name is anecdotal reported to have been grown axenically (unknown conditions) and obtained in Somerville, MA, US. These samples were used for isolation of high molecular weight (HMW) DNA using a modified CTAB/Chloroform and SPRI protocol. Briefly, 300mg of stem sample were ground to a fine powder using a -80C frozen mortar and pestle. 150 mg of powder was then aliquoted into 2 mL conical tubes (USA Scientific) with 1.5 mL cetrimonium bromide. These tubes were then incubated at room temperature on a tube rotator for 10 minutes. 6 uL of RNase A (Promega 4 mg/mL) was then added and both tubes were incubated at 37°C for one hour, vortexing every 15 minutes. Following this incubation, 7.5 uL Proteinase K (New England Biolabs 20 mg/mL) was added and the tubes were incubated at 60°C for 30 minutes, vortexing every 10 minutes. At the conclusion of the Proteinase K incubation, both tubes were incubated on ice for 10 minutes. The samples were then centrifuged for 5 minutes at 14000 rpm. 600 uL of supernatant was removed from each tube and added to 600 uL chloroform. The tubes were then vortexed until opaque and spun for 5 minutes at 14000 rpm. 400 uL of the aqueous layer was removed using a wide bore tip and added to a 1.5 mL Eppendorf tube. 400 uL MIP (marijuana infused products) Solution B and 400 uL DNA Binding Beads (Medicinal Genomics PN 420004) were added to the Eppendorf tube and inverted to homogenize. The tubes were then incubated at room temperature on the tube rotator for 15 minutes. The tubes were then removed from the rotator and placed on a magnetic tube rack for 3 minutes. The supernatant was removed, the beads were washed twice with 1 mL of 70% ethanol and allowed to dry for 5 minutes. The beads were then eluted in 100 uL of 56°C Elution Buffer (Medicinal Genomics PN 420004) using a wide bore tip and incubated at 56°C for 5 minutes. Following this incubation, the tubes were returned to the magnetic rack, the supernatant of both tubes were removed using a wide bore tip and pooled in a fresh Eppendorf tube. HMW DNA length was quantified on an Agilent TapeStation and produced a DIN of 8.1. Qubit Fluorometer (Thermo Fisher Scientific) quantified 55ng/ul. Nanodrop Spectrophotometer (Thermo Fisher Scientific) quantified 104ng/ul with 260/280nm ratio of 1.85 and 260/230nm of 0.95.

Sequencing

Sequencing libraries were constructed according to the manufacturer's instructions for Pacific Biosciences Sequel II HiFi sequencing. 773,735 CCS reads were generated. Quast (Gurevich et al. 2013) was used to assess the quality of the input fasta sequence file (N50 = 13.9Kb) and the output assembly fasta file (3.33Mb N50).

Assembly and annotation

The unfiltered CCS data was assembled using the Peregrine assembler (pg_asm 0.3.5,arm_config5e69f3d+) (Chin 2019). Reads were assembled into 32 contigs with lengths ranging from 32 kilobases to 4.6 megabases (Figure 1). The Peregrine assembler requires at least 2 HiFi reads to substantially overlap to contribute to a contig and as a result we did not observe any bacterial contamination in the assembly BUSCO v3.0.2 completeness scores (97.6%) were measured using agaricales_odb10.2020-08-05 BUSCO lineage database (Table 1) (Simao et al. 2015; Waterhouse et al. 2018). FunAnnotate v1.8.4 was used to annotate the genome (Li and Wang 2021) resulting in 13,478 genes.

The final genome assembly was aligned to three other public *Psilocybe cubensis* datasets (Fricke et al. 2017; Torrens-Spence et al. 2018; Reynolds et al. 2018) and one different *Psilocybe* species (*Psilocybe cyanescens)* to verify taxonomic identification (Table 2). In total, 96-98.75% of these *Psilocybe cubensis* sequences align to the new HiFi generated

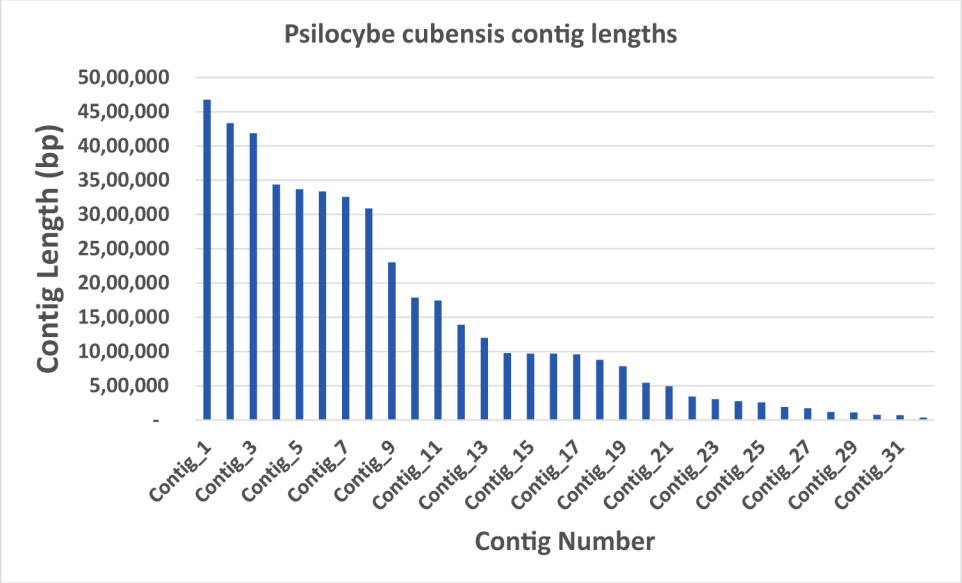

Figure 1. Psilocybe cubensis P.envy contig length distribution (n = 32).

Table 1. BUSCO completeness scores using agaricales_odb10.2020-08-05.

Total BUSCOs	Single-copy	Duplicated	Fragmented	Missing
3870	3729	45	9	87
97.60%	96.40%	1.20%	0.20%	2.20%

Table 2. Three *Psilocybe cubensis* data sets in NCBI and JGI were aligned to the P.envy HiFi reference. A different Psilocybe species (*Psilocybe cyanescens*) genome was also mapped with much lower mapping efficiency.

Author	Accession	Data type	Mapping rate	Tool	Species
Fricke et al. 2017	https://mycocosm.jgi.doe.gov/Psicub1_1/Psicub1_1.home.html	Illumina Assembly	98.8%	Minimap2	*P. cubensis*
McKernan et al. 2020	NCBI Project: PRJNA687911	Illumina FastQ	96.0%	bwa-mem	*P. cubensis*
Torrens-Spence et al. 2018	NCBI Project: PRJNA450675	RAN-Seq Assembly	98.5%	Minimap2	*P. cubensis*
Reynolds et al. 2018	NCBI Project: PRJNA387735	Illumina Assembly	56.8%	Minimap2	*P. cyanescens*

Psilocybe cubensis P.envy reference using minimap2 and bwa-mem (Li and Durbin 2010; Li 2018). Mapping rates were determined using samtools flagstat on bam files (Li et al. 2009). Alignments were visualized with MUMmer V4.0.0beta2 and Integrative Genomics Viewer v2.4.16 (Delcher et al. 2003; Robinson et al. 2011; Thorvaldsdottir et al. 2013).

Three Illumina genome assemblies (Reynolds et al., McKernan et al., Fricke et al.) were additionally aligned using MUMmer for whole genome alignment plots (Figure 2).

Polymorphisms

Illumina whole-genome shotgun data (McKernan et al. NCBI Project: PRJNA687911) was mapped to the P. envy HiFi reference assembly using bwa-mem (version0.7.17-r1188), samtools (version 1.8), sorted with sambamba (version 0.6.7)

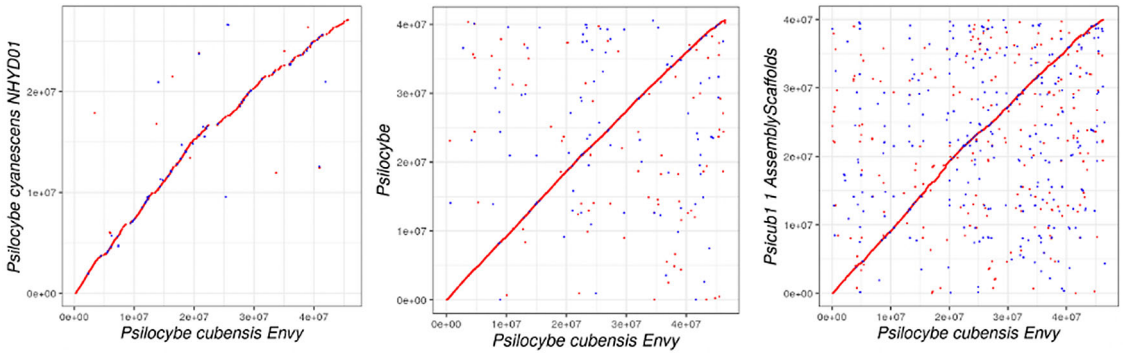

Figure 2. Whole genome alignments of short read Illumina assemblies to *Psilocybe cubensis* strain P. envy. Left is *Psilocybe cyanescens* from Reynolds et al. Middle is McKernan et al. (MGC) Illumina assembly. Right is Fricke et al. or JGI assembly.

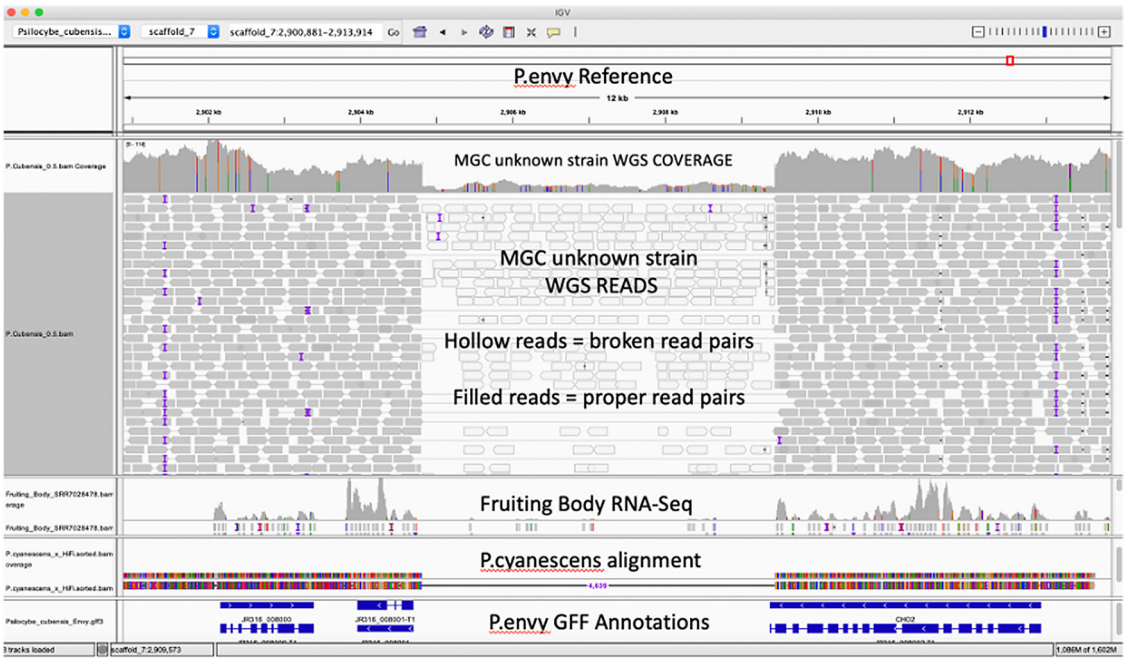

Figure 3. IGV display of Illumina reads mapped to HiFi *Psilocybe cubensis* P.envy assembly. Top track is Medicinal Genomics Illumina whole genome shotgun data of a different *P. cubensis* (strain name unknown: NCBI Project: PRJNA687911) mapped to the HiFi *P. cubensis* strain P.envy. Second track contains RNA-Seq data from a third *P. cubensis* genome (strain name also unknown: NCBI Project: PRJNA450675) hosted at JGI. Third track is *Psilocybe cyanescens* genome mapped to HiFi *P. cubensis* P.envy reference genome. Fourth track is FunAnnotate GFF3 annotation of the HiFi *P. cubensis* P.envy genome.

and variants were identified using GATK HaplotypeCaller (version 4.1.6.0) with default arguments. The annotation from the *funannotate* pipeline was converted from gff3 format into SnpEff (version 4.3t 2017-11-24) database as described here (https://pcingola.github.io/SnpEff/se_buildingdb/) and the variants that came out of HaplotypeCaller were annotated. 553,716 variants (471,443 SNPs and 82,273 small insertions and deletions) were called and annotated equating to aSNP every 99bp and a variant every 83bp including indels. Of these, 375,896 (67.9%) are heterozygous and 177,820 (32.1%) are homozygous with a ratio of just over 2 to 1 heterozygous:homozygous variants. Lastly, as a quality check, the original Pacific Biosciences CCS corrected shotgun reads were mapped back to the reference with minimap2 (version 2.17-r941) and variants were called again using GATK HaplotypeCaller. A total of 15,963 variants are identified and 15,674 (98.2%) are heterozygous with only 289 homozygous variants called. Whole genome shotgun reads mapped back to their consensus reference should produce predominantly heterozygous calls in a diploid organism. Scripts utilized to for variant calling are in github and described in the Data availability section.

Structural variation

The N-methyltransferase gene responsible for Psilocybin production in P.envy contains a structural variation not seen in previous *P. cubensis* surveys (Figure 3). Illumina read mapping of the McKernan et al. *P. cubensis* assembly in NCBI (NCBI Project: PRJNA687911) demonstrates multiple read pairs spanning a 4.6kb insertion in the HiFi *P. cubensis* strain P.envy (SRA submission SRP299420). This insertion extends the 3' end of the P.envy N-methyltransferase gene. The 4.6kb insertion is also observed as a deletion in *Psilocybe cyanescens* and as a deletion in RNA-Seq data from Torrens-Spence et al. (NCBI Project: PRJNA450675) (Reynolds et al. 2018). Other SNPs also exist in these genes and should be considered in context of this deletion. Further work is required to understand the biological significance of this variation.

Conclusions

A highly contiguous *Psilocybe cubensis* genome has been generated. The N50 contigs lengths are 75 fold more contiguous than the existing assembly available at JGI. This reference can aid in the identification of genetic variation that may impact psilocybin, psilocin, norpsilocin, baeocystin, norbaeocystin and aeruginascin production.

Data availability

GenBank: Psilocybe cubensis strain MGC-MH-2018, whole genome shotgun sequencing project, Accession number JAFIQS000000000.1: https://www.ncbi.nlm.nih.gov/nuccore/JAFIQS000000000.1/.

BioProject: Psilocybe cubensis, Accession number PRJNA687911: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA687911

BioProject: Psilocybe cubensis strain: MGC-MH-2018, Accession number PRJNA700437: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA700437

CoGe genome browser: Psilocybe cubensis (Psilocybe cubensis P.envy), https://genomevolution.org/coge/GenomeInfo.pl?gid=60487

Variant calling scripts: https://github.com/mclaugsf/mgc-public/tree/master/f1000_10-281. The final list of annotated variants and the accompanying SnpEff output files are available here (https://github.com/mclaugsf/mgc-public/tree/master/f1000_10-281/nextflow/annotated-variants). The gff3 file that was used to perform the SnpEff annotation is available for download (https://github.com/mclaugsf/mgc-public/blob/master/f1000_10-281/gff/P-Envy-05-25-2021.gff3.gz) as well as Dockerized workflows written in nextflow used to perform the mapping, variants calling and annotation analysis (https://github.com/mclaugsf/mgc-public/tree/master/f1000_10-281/nextflow).

References

Awan: **Convergent evolution of psilocybin biosynthesis by psychedelic mushrooms.** *BioRXIV.* 2018.
Publisher Full Text

Blei F, Baldeweg F, Fricke J, *et al.*: **Biocatalytic Production of Psilocybin and Derivatives in Tryptophan Synthase-Enhanced Reactions.** *Chemistry.* 2018.
PubMed Abstract | Publisher Full Text

Blei F, Dorner S, Fricke J, *et al.*: **Simultaneous Production of Psilocybin and a Cocktail of beta-Carboline Monoamine Oxidase Inhibitors in "Magic" Mushrooms.** *Chemistry.* 2020; **26**: 729–734.
PubMed Abstract | Publisher Full Text | Free Full Text

Chin CS: **Human Genome Assembly in 100 Minutes.** *bioRxiv.* 2019.
Publisher Full Text

Delcher AL, Salzberg SL, Phillippy AM: **Using MUMmer to identify similar regions in large sequence sets.** *Curr Protoc Bioinformatics.* 2003; **Chapter 10**: Unit 10 13.

Demmler R, Fricke J, Dorner S, *et al.*: **S-Adenosyl-l-Methionine Salvage Impacts Psilocybin Formation in "Magic" Mushrooms.** *Chembiochem.* 2020; **21**: 1364–1371.
PubMed Abstract | Publisher Full Text | Free Full Text

Fricke J, Blei F, Hoffmeister D: **Enzymatic Synthesis of Psilocybin.** *Angewandte Chemie.* 2017; **56**: 12352–12355.
PubMed Abstract | Publisher Full Text

Fricke J, Kargbo R, Regestein L, *et al.*: **Scalable Hybrid Synthetic/Biocatalytic Route to Psilocybin.** *Chemistry.* 2020; **26**: 8281–8285.
PubMed Abstract | Publisher Full Text | Free Full Text

Fricke J, Lenz C, Wick J, *et al.*: **Production Options for Psilocybin: Making of the Magic.** *Chemistry.* 2019a; **25**: 897–903.
PubMed Abstract | Publisher Full Text

Fricke J, Sherwood A, Kargbo R, *et al.*: **Enzymatic Route toward 6-Methylated Baeocystin and Psilocybin.** *Chembiochem.* 2019b; **20**: 2824–2829.
PubMed Abstract | Publisher Full Text

Gurevich A, Saveliev V, Vyahhi N, *et al.*: **QUAST: quality assessment tool for genome assemblies.** *Bioinformatics.* 2013; **29**: 1072–1075.
PubMed Abstract | Publisher Full Text | Free Full Text

Johnson MW, Griffiths RR: **Potential Therapeutic Effects of Psilocybin.** *Neurotherapeutics.* 2017; **14**: 734–740.
PubMed Abstract | Publisher Full Text | Free Full Text

Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics.* 2018; **34**: 3094–3100.
PubMed Abstract | Publisher Full Text | Free Full Text

Li H, Durbin R: **Fast and accurate long-read alignment with Burrows-Wheeler transform.** *Bioinformatics.* 2010; **26**: 589–595.
PubMed Abstract | Publisher Full Text | Free Full Text

Li H, Handsaker B, Wysoker A, *et al.*: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics.* 2009; **25**: 2078–2079.
PubMed Abstract | Publisher Full Text | Free Full Text

Li WC, Wang TF: **PacBio Long-Read Sequencing, Assembly, and Funannotate Reannotation of the Complete Genome of Trichoderma reesei QM6a.** *Methods Mol Biol.* 2021; **2234**: 311–329.
PubMed Abstract | Publisher Full Text

Reynolds HT, Vijayakumar V, Gluck-Thaler E, *et al.*: **Horizontal gene cluster transfer increased hallucinogenic mushroom diversity.** *Evol Lett.* 2018; **2**: 88–101.
PubMed Abstract | Publisher Full Text | Free Full Text

Robinson JT, Thorvaldsdottir H, Winckler W, *et al.*: **Integrative genomics viewer.** *Nat Biotechnol.* 2011; **29**: 24–26.
PubMed Abstract | Publisher Full Text | Free Full Text

Simao FA, Waterhouse RM, Ioannidis P, *et al.*: **BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs.** *Bioinformatics.* 2015; **31**: 3210–3212.
PubMed Abstract | Publisher Full Text

Thorvaldsdottir H, Robinson JT, Mesirov JP: **Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration.** *Brief Bioinform.* 2013; **14**: 178–192.
PubMed Abstract | Publisher Full Text | Free Full Text

Torrens-Spence MP, Liu CT, Pluskal T, *et al.*: **Monoamine Biosynthesis via a Noncanonical Calcium-Activatable Aromatic Amino Acid Decarboxylase in Psilocybin Mushroom.** *ACS Chem Biol.* 2018; **13**: 3343–3353.
PubMed Abstract | Publisher Full Text

Waterhouse RM, Seppey M, Simao FA, *et al.*: **BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics.** *Mol Biol Evol.* 2018; **35**: 543–548.
PubMed Abstract | Publisher Full Text | Free Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔

Version 2

Reviewer Report 22 June 2021

https://doi.org/10.5256/f1000research.57314.r87516

✔ **Jason Slot**
Department of Plant Pathology, The Ohio State University, Columbus, OH, USA

The authors have provided changes for all the comments and requested edits and I deem all of them to be acceptable, and to have much improved the manuscript. There remains one typo that the authors may wish to correct: In the last section of the Polymorphisms section, "utilized for" should replace "utilized to for".

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Fungal ecology, microbial genomics, evolution, metabolism

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 21 May 2021

https://doi.org/10.5256/f1000research.54802.r83076

✔ **Anders Goncalves da Silva**
Lighthouse Genomics Inc., British Columbia, Canada

Philippe Henry

[1] Egret Bioscience Ltd., West Kelowna, Canada
[2] Lighthouse Genomics Inc., BC, Canada

McKernan and colleagues present on the first highly contiguous draft genome for the magic mushroom *Psilocybe cubensis*. We commend their use of High accuracy long read sequencing and an advanced bioinformatics pipeline to build a much more complete picture of the *P. cubensis* genome and for making it openly available to the public with the promise the genetic architecture of tryptamine expression in magic mushrooms.

The methods employed are state of the art and the authors provide sufficient access to the data to enable peers and the public to replicate the experiment. While they acknowledge that the HiFi sequencing approach comes with great advantages, in particular greatly improved contiguity and and BUSCO completeness scores compared to other *P. cubensis* genomes published to date, the authors did not acknowledge that fungi can have multiple nuclei in a cell, sometimes with completely different haplotypes. As such, we posit that their assembly could possibly be a metagenome assembly, rather than the assembly of a single genome, thus providing an alternative explanation to the large insertion detected in the norbaeocystin methyltranferase (psiM) gene.

Perhaps a means to provide a remedy to this is to provide some additional background on the *P. cubensis* Penis Envy (PE) strain, in particular the alleged origin of the PE mutant and its probable clonal propagation. While anecdotal at best, the fabled “mutation” of PE appeared and was selected from a phenotype of an amazonian *P. cubensis* accession, as a towering fruiting body with a pale cap and missing partial veil, which was then preserved via clonal propagation. The mutant is also described as being more potent that most other *P. cubensis* strain, leading to the hypothesis that it had a skewed drug to prodrug ratio (psilocyn/psilocybin) which would hint to a mutation in the psiK gene as opposed to the large insertion in psiM.

Other putative mechanisms could be polymorphisms at other loci involved in the psilocybin biosynthetic pathways as well as ancillary genes involved in the SAM salvage pathway (e.g. ref 1), a list of putative functional SNPs that may interact with the large insertion is shown here from an earlier version of the *P.cubensis* genome. Genotyping several strains at the 4.6kb insertion and ancillary SNPs may help shed light on the mechanism behind the higher perceived potency of PE compared to other *P. cubensis* strains and other species in the *Psilocybe* and *Panaeolus* genus, In that vein, the authors may gain additional insight by including the *P. serbica var bohemica* genome to their comparative analysis, provided that chemotypic information associated with each accessions is made available.

Node Position Target gene Ontogeny
NODE_599 19295 sahH *S-adenosyl-l-homocysteine hydrolase*
NODE_599 19304 sahH *S-adenosyl-l-homocysteine hydrolase*
NODE_599 19415 sahH *S-adenosyl-l-homocysteine hydrolase*
NODE_599 19421 sahH *S-adenosyl-l-homocysteine hydrolase*
NODE_599 19744 sahH *S-adenosyl-l-homocysteine hydrolase*
NODE_599 19746 sahH *S-adenosyl-l-homocysteine hydrolase*
NODE_599 20318 sahH *S-adenosyl-l-homocysteine hydrolase*
NODE_599 20779 sahH *S-adenosyl-l-homocysteine hydrolase*
NODE_599 20801 sahH *S-adenosyl-l-homocysteine hydrolase*

NODE_599 20840 sahH *S-adenosyl-l-homocysteine hydrolase*
NODE_6392 313322 samS *S-adenosyl-l-methionine synthetase*
NODE_6392 314122 samS *S-adenosyl-l-methionine synthetase*
NODE_6392 314300 samS *S-adenosyl-l-methionine synthetase*
NODE_712 1234504 metS l-methionine synthetase
NODE_755 63920 psiD tryptophan decarboxylase
NODE_755 64575 psiD tryptophan decarboxylase
NODE_755 64576 psiD tryptophan decarboxylase
NODE_755 65088 psiD tryptophan decarboxylase
NODE_755 65181 psiD tryptophan decarboxylase
NODE_755 61250 psiM methyltransferase
NODE_755 61711 psiM methyltransferase
NODE_755 62335 psiM methyltransferase
NODE_755 58208 psiT2 major-facilitator-type transporters
NODE_755 58407 psiT2 major-facilitator-type transporters
NODE_755 58772 psiT2 major-facilitator-type transporters
NODE_755 58883 psiT2 major-facilitator-type transporters
NODE_755 59054 psiT2 major-facilitator-type transporters
NODE_755 59321 psiT2 major-facilitator-type transporters
NODE_755 59372 psiT2 major-facilitator-type transporters
NODE_755 59426 psiT2 major-facilitator-type transporters
NODE_755 59484 psiT2 major-facilitator-type transporters
NODE_755 59492 psiT2 major-facilitator-type transporters
NODE_755 59504 psiT2 major-facilitator-type transporters
NODE_755 59522 psiT2 major-facilitator-type transporters
NODE_755 59537 psiT2 major-facilitator-type transporters
NODE_755 59540 psiT2 major-facilitator-type transporters
NODE_755 59566 psiT2 major-facilitator-type transporters
NODE_755 59594 psiT2 major-facilitator-type transporters
NODE_755 59648 psiT2 major-facilitator-type transporters
NODE_755 59694 psiT2 major-facilitator-type transporters
NODE_755 59771 psiT2 major-facilitator-type transporters
NODE_755 59869 psiT2 major-facilitator-type transporters
NODE_755 59875 psiT2 major-facilitator-type transporters
NODE_755 56543 psiH P450 monooxygenase
NODE_755 54126 psiK Kinase
NODE_755 54132 psiK Kinase
NODE_755 54136 psiK Kinase
NODE_755 54159 psiK Kinase
NODE_755 54213 psiK Kinase
NODE_755 54223 psiK Kinase
NODE_755 54311 psiK Kinase
NODE_755 51932 psiT1 major-facilitator-type transporters
NODE_755 52391 psiT1 major-facilitator-type transporters
NODE_755 52473 psiT1 major-facilitator-type transporters
NODE_755 52477 psiT1 major-facilitator-type transporters
NODE_755 52491 psiT1 major-facilitator-type transporters
NODE_755 52620 psiT1 major-facilitator-type transporters

NODE_755 52688 psiT1 major-facilitator-type transporters
NODE_755 52787 psiT1 major-facilitator-type transporters
NODE_755 52793 psiT1 major-facilitator-type transporters

References

1. Demmler R, Fricke J, Dörner S, Gressler M, et al.: S -Adenosyl-l -Methionine Salvage Impacts Psilocybin Formation in “Magic” Mushrooms. *ChemBioChem*. 2020; **21** (9): 1364-1371 Publisher Full Text

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?

Yes

Are the datasets clearly presented in a useable and accessible format?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Population genetics, genotype-chemotype

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 26 April 2021

https://doi.org/10.5256/f1000research.54802.r83396

?

Jason Slot
Department of Plant Pathology, The Ohio State University, Columbus, OH, USA

Summary:
The manuscript presents a high quality assembly of the historically and medicinally important fungus, *Psilocybe cubensis*. Best practices were observed in sequencing, assembly, and annotation. The manuscript notes a potentially important structural variation present in the *P. envy* strain psilocybin N-methyltransferase gene, which resembles the variation in the more potent, by psilocybin content, *Psilocybe cyanescens*.

- **Is the rationale for creating the dataset(s) clearly described?**

The study was undertaken in order to provide a high quality reference genome for the species. To date, the genomes in the species and genus are fragmented more than is desirable for basic and applied comparative investigations of genome content and architecture.

- **Are the protocols appropriate and is the work technically sound?**

The Pacific Biosciences Sequel II HiFi methods used for sequencing are among the best for generating near-chromosome level assembly. Assembly was performed with cutting-edge Peregrine Assembler, and the annotation appropriately used the fungus-specific FunAnnotate pipeline. Single nucleotide polymorphisms (SNP) were called with appropriate software, but parameters were not detailed in the text. This and structural variation were not intended to be exhaustive, but provide intriguing statistics and examples to warrant follow-up investigations.

- **Are sufficient details of methods and materials provided to allow replication by others?**

Parameters for SNP calling would have to be further detailed in order to allow replication of raw SNP numbers between two isolates.

- **Are the datasets clearly presented in a useable and accessible format?**

Figure 1 and both tables are clear and informative. Figure 2 readability would be improved by increasing the size of the axis titles. Figure 3 is not sufficiently informative or simple to acquire meaning as it is currently presented. This figure would benefit from marking the "tracks" clearly, but number and perhaps with additional labels for RNAseq, *P. cyanescens*, and annotation as the IGV display is too small to read as is. Is there significance to the locus that is presented in Figure 3? If not, then perhaps indicate it is a "representative" locus.

Other comments:

- The title might flow better as "A draft reference assembly of...".
- In the introduction, "several mushrooms" might better be stated "about 200 mushroom species".
- In the Introduction " fungi's " would be more syntactically correct as "fungus' "
- Given that dried stems were used for genomic DNA isolation, it is expected that some additional microbial DNA might be present. Authors should note if the mushroom was produced axenically, or if contaminant reads were filtered to either prevent or address presence of additional species' genomes in the assembly.
- Citations are incomplete in the last sentence of "Assembly and annotation" section, and second sentence of "Structural variation" section.

Is the rationale for creating the dataset(s) clearly described?

Yes

Are the protocols appropriate and is the work technically sound?

Yes

Are sufficient details of methods and materials provided to allow replication by others?

Partly

Are the datasets clearly presented in a useable and accessible format?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Fungal ecology, microbial genomics, evolution, metabolism

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 21 May 2021

Kevin McKernan, Medicinal Genomics, Beverly, USA

The reviewer makes excellent points. We will be making these suggested changes to the final manuscript.

Competing Interests: No competing interests were disclosed.