DATA NOTE

REVISeD A draft reference assembly of the *Psilocybe cubensis* genome [version 2; peer review: 1 approved, 1 approved with reservations]

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

Kevin McKernan1, Liam T. Kane1, Seth Crawford2, Chen-Shan Chin3, Aaron Trippe2, Stephen McLaughlin1

1R&D, Medicinal Genomics, Beverly, Mass, 01915, USA
2R&D, Oregon CBD, Corvallis, Oregon, 97330, USA
3Bioinformatics, 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.
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; Tripe 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.

Copyright: © 2021 McKernan K et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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: 1 approved, 1 approved with reservations] 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
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 -80°C 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.

<table>
<thead>
<tr>
<th>Total BUSCOs</th>
<th>Single-copy</th>
<th>Duplicated</th>
<th>Fragmented</th>
<th>Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>3870</td>
<td>3729</td>
<td>45</td>
<td>9</td>
<td>87</td>
</tr>
<tr>
<td>97.60%</td>
<td>96.40%</td>
<td>1.20%</td>
<td>0.20%</td>
<td>2.20%</td>
</tr>
</tbody>
</table>

**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.

<table>
<thead>
<tr>
<th>Author</th>
<th>Accession</th>
<th>Data type</th>
<th>Mapping rate</th>
<th>Tool</th>
<th>Species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fricke et al. 2017</td>
<td><a href="https://mycocosm.jgi.doe.gov/Psicub1_1/Psicub1_1.home.html">Link</a></td>
<td>Illumina Assembly</td>
<td>98.8%</td>
<td>Minimap2</td>
<td>P. cubensis</td>
</tr>
<tr>
<td>McKernan et al. 2020</td>
<td>NCBI Project: PRJNA687911</td>
<td>Illumina FastQ</td>
<td>96.0%</td>
<td>bwa-mem</td>
<td>P. cubensis</td>
</tr>
<tr>
<td>Torrens-Spence et al. 2018</td>
<td>NCBI Project: PRJNA450675</td>
<td>RAN-Seq Assembly</td>
<td>98.5%</td>
<td>Minimap2</td>
<td>P. cubensis</td>
</tr>
<tr>
<td>Reynolds et al. 2018</td>
<td>NCBI Project: PRJNA387735</td>
<td>Illumina Assembly</td>
<td>56.8%</td>
<td>Minimap2</td>
<td>P. cyanescens</td>
</tr>
</tbody>
</table>
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).
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


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


References
Open Peer Review

Current Peer Review Status: ? ✓

Version 1

Reviewer Report 21 May 2021

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

© 2021 Henry P et al. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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.

<table>
<thead>
<tr>
<th>NODE Position</th>
<th>Target gene</th>
<th>Ontogeny</th>
</tr>
</thead>
<tbody>
<tr>
<td>NODE_599 19295</td>
<td>sahH</td>
<td><em>S</em>-adenosyl-<em>l</em>-homocysteine hydrolase</td>
</tr>
<tr>
<td>NODE_599 19304</td>
<td>sahH</td>
<td><em>S</em>-adenosyl-<em>l</em>-homocysteine hydrolase</td>
</tr>
<tr>
<td>NODE_599 19415</td>
<td>sahH</td>
<td><em>S</em>-adenosyl-<em>l</em>-homocysteine hydrolase</td>
</tr>
<tr>
<td>NODE_599 19421</td>
<td>sahH</td>
<td><em>S</em>-adenosyl-<em>l</em>-homocysteine hydrolase</td>
</tr>
<tr>
<td>NODE_599 19744</td>
<td>sahH</td>
<td><em>S</em>-adenosyl-<em>l</em>-homocysteine hydrolase</td>
</tr>
<tr>
<td>NODE_599 19746</td>
<td>sahH</td>
<td><em>S</em>-adenosyl-<em>l</em>-homocysteine hydrolase</td>
</tr>
<tr>
<td>NODE_599 20318</td>
<td>sahH</td>
<td><em>S</em>-adenosyl-<em>l</em>-homocysteine hydrolase</td>
</tr>
<tr>
<td>NODE_599 20779</td>
<td>sahH</td>
<td><em>S</em>-adenosyl-<em>l</em>-homocysteine hydrolase</td>
</tr>
<tr>
<td>NODE_599 20801</td>
<td>sahH</td>
<td><em>S</em>-adenosyl-<em>l</em>-homocysteine hydrolase</td>
</tr>
<tr>
<td>NODE_599 20840</td>
<td>sahH</td>
<td><em>S</em>-adenosyl-<em>l</em>-homocysteine hydrolase</td>
</tr>
<tr>
<td>NODE_6392 313322</td>
<td>samS</td>
<td><em>S</em>-adenosyl-<em>l</em>-methionine synthetase</td>
</tr>
<tr>
<td>NODE_6392 314122</td>
<td>samS</td>
<td><em>S</em>-adenosyl-<em>l</em>-methionine synthetase</td>
</tr>
<tr>
<td>NODE_712 1234504</td>
<td>metS</td>
<td>l-methionine synthetase</td>
</tr>
<tr>
<td>NODE_755 63920</td>
<td>psiD</td>
<td>tryptophan decarboxylase</td>
</tr>
<tr>
<td>NODE_755 64575</td>
<td>psiD</td>
<td>tryptophan decarboxylase</td>
</tr>
<tr>
<td>NODE_755 64576</td>
<td>psiD</td>
<td>tryptophan decarboxylase</td>
</tr>
<tr>
<td>NODE_755 65088</td>
<td>psiD</td>
<td>tryptophan decarboxylase</td>
</tr>
<tr>
<td>NODE_755 65181</td>
<td>psiD</td>
<td>tryptophan decarboxylase</td>
</tr>
<tr>
<td>NODE_755 61250</td>
<td>psiM</td>
<td>methyltransferase</td>
</tr>
<tr>
<td>NODE_755 61711</td>
<td>psiM</td>
<td>methyltransferase</td>
</tr>
<tr>
<td>NODE_755 62335</td>
<td>psiM</td>
<td>methyltransferase</td>
</tr>
<tr>
<td>NODE_755 58208</td>
<td>psiT2</td>
<td>major-facilitator-type transporters</td>
</tr>
<tr>
<td>NODE_755 58407</td>
<td>psiT2</td>
<td>major-facilitator-type transporters</td>
</tr>
<tr>
<td>NODE_755 58772</td>
<td>psiT2</td>
<td>major-facilitator-type transporters</td>
</tr>
<tr>
<td>NODE_755 58883</td>
<td>psiT2</td>
<td>major-facilitator-type transporters</td>
</tr>
<tr>
<td>NODE_755 59054</td>
<td>psiT2</td>
<td>major-facilitator-type transporters</td>
</tr>
<tr>
<td>NODE_755 59321</td>
<td>psiT2</td>
<td>major-facilitator-type transporters</td>
</tr>
<tr>
<td>NODE_755 59372</td>
<td>psiT2</td>
<td>major-facilitator-type transporters</td>
</tr>
<tr>
<td>NODE_755 59426</td>
<td>psiT2</td>
<td>major-facilitator-type transporters</td>
</tr>
<tr>
<td>NODE_755 59484</td>
<td>psiT2</td>
<td>major-facilitator-type transporters</td>
</tr>
<tr>
<td>NODE_755 59492</td>
<td>psiT2</td>
<td>major-facilitator-type transporters</td>
</tr>
<tr>
<td>NODE_755 59504</td>
<td>psiT2</td>
<td>major-facilitator-type transporters</td>
</tr>
<tr>
<td>NODE_755 59522</td>
<td>psiT2</td>
<td>major-facilitator-type transporters</td>
</tr>
<tr>
<td>NODE_755 59537</td>
<td>psiT2</td>
<td>major-facilitator-type transporters</td>
</tr>
<tr>
<td>NODE_755 59540</td>
<td>psiT2</td>
<td>major-facilitator-type transporters</td>
</tr>
<tr>
<td>NODE_755 59566</td>
<td>psiT2</td>
<td>major-facilitator-type transporters</td>
</tr>
</tbody>
</table>
References

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.
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.
The benefits of publishing with F1000Research:

• Your article is published within days, with no editorial bias
• You can publish traditional articles, null/negative results, case reports, data notes and more
• The peer review process is transparent and collaborative
• Your article is indexed in PubMed after passing peer review
• Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com