RESEARCH ARTICLE

A whole genome atlas of 81 *Psilocybe* genomes as a resource for psilocybin production. [version 1; peer review: 1 not approved]

Kevin McKernan , Liam Kane , Yvonne Helbert, Lei Zhang, Nathan Houde, Stephen McLaughlin

Research and Development, Medicinal Genomics, Beverly, Massachusetts, 01915, USA

V1 First published: 23 Sep 2021, 10:961
<https://doi.org/10.12688/f1000research.55301.1>
Latest published: 17 Dec 2021, 10:961
<https://doi.org/10.12688/f1000research.55301.2>

Abstract

The *Psilocybe* genus is well known for the synthesis of valuable psychoactive compounds such as Psilocybin, Psilocin, Baeocystin and Aeruginascin. The ubiquity of Psilocybin synthesis in *Psilocybe* has been attributed to a horizontal gene transfer mechanism of a ~20Kb gene cassette. A recently published highly contiguous reference genome derived from long read single molecule sequencing has underscored interesting variation in this Psilocybin synthesis gene cassette. This reference genome has also enabled the shotgun sequencing of spores from many *Psilocybe* strains to better catalog the genomic diversity in the Psilocybin synthesis pathway. Here we present the de novo assembly of genomes of 81 *Psilocybe* genomes compared to the *P. envy* reference genome. Surprisingly, the genomes of *Psilocybe galindoi*, *Psilocybe tampanensis* and *Psilocybe azurescens* lack sequence coverage over the previously described Psilocybin synthesis pathway but do demonstrate amino acid sequence homology to an alternative pathway and may illuminate previously proposed convergent evolution of Psilocybin synthesis.

Keywords

Psilocybe cubensis, Genome, Single molecule sequencing, Psilocybin

Open Peer Review

Approval Status

1

version 2

(revision)
17 Dec 2021

version 1

23 Sep 2021

[view](#)

1. **Bryn T. M. Dentinger**, University of Utah,
Salt Lake City, USA

Alexander Bradshaw , University of Utah,
Salt Lake City, USA

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, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Supervision, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Kane L:** Methodology, Project Administration, Writing – Review & Editing; **Helbert Y:** Investigation, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Zhang L:** Methodology, Project Administration, Resources; **Houde N:** Investigation, Methodology; **McLaughlin S:** Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Software, Validation, Visualization, Writing – Original Draft Preparation

Competing interests: No competing interests were disclosed.

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 L, Helbert Y *et al.* **A whole genome atlas of 81 Psilocybe genomes as a resource for psilocybin production. [version 1; peer review: 1 not approved]** F1000Research 2021, 10:961 <https://doi.org/10.12688/f1000research.55301.1>

First published: 23 Sep 2021, 10:961 <https://doi.org/10.12688/f1000research.55301.1>

Introduction

Psilocybin has recently been awarded breakthrough drug status by the FDA (Tullis 2021). Psilocybin has also been shown to be a safe alternative to traditional selective serotonin re-uptake inhibitors (SSRIs) for the treatment of depression (Griffiths *et al.* 2016; Ross *et al.* 2016; Bogenschutz *et al.* 2018; Carhart-Harris *et al.* 2021; Davis *et al.* 2021). SSRIs have also been shown to be effective at reducing SARS-CoV-2 viral load (Schloer *et al.* 2020; Creeden *et al.* 2021; Dechaumes *et al.* 2021; Schloer *et al.* 2021; Zimniak *et al.* 2021). The mechanism of the viral interference is hypothesized to be related to the co-expression of ACE2 and DOPA decarboxylase (DDC) which catalyzes the synthesis of serotonin (Attademo and Bernardini 2021). It is possible psilocybin's serotonin agonism will show promise in COVID treatment and prevention in the future, thus understanding the biological mechanism of synthesis is of increasing importance.

Two models have been proposed for psilocybin production in *Psilocybe* spp. Horizontal gene transfer has been thoroughly described by Reynolds *et al.* and Fricke *et al.*, while convergent evolution has been proposed by Awan (Fricke *et al.* 2017; Awan 2018; Reynolds *et al.* 2018). Given the broad geographic distribution of psilocybin-producing mushrooms and the history of human cultivation and selection for psilocybin-producing strains, these models may not be mutually exclusive.

To address this question, we whole-genome sequenced 81 *Psilocybe* spp. spore samples to assess the sequence coverage over the previously well-characterized psilocybin synthesis cassette.

Results

In a whole-genome sequencing survey of 81 *Psilocybe* spp. genomes we noticed a lack of sequence coverage over the psilocybin gene cluster described by Reynolds *et al.* for five of the 81 Psilocybin producing genomes (three *Psilocybe tampanensis*, one *Psilocybe galindoi*, one *Psilocybe azurescens*). These genomes also exhibited genome-wide poor read mapping efficiency to the *Psilocybe cubensis* “P.envy” reference genome, with the exception of the mitochondria and ITS regions. These non-conical psilocybin-synthesizing genomes share 99%, 100%, and 91% ITS sequence identity to a *Psilocybe tampanensis* sequence described by Rockefeller *et al.* (NCBI accession number: MH220315.1). The sequence from Rockefeller *et al.* was obtained in the wild and verified with photography of morphological features. The sequences described in this study were derived from spores that can be legally sold for taxonomy purposes but cannot be legally cultured to obtain morphological or chemical verification.

One other genus was labelled as *Panaeolus copelandia* but delivered two 100% identical ITS sequence to *Psilocybe cubensis* and *Aspergillus fumigatus*. This samples was omitted from further analysis. Sequencing and variant calling statistics are displayed in Figure 1 and Supplement Table 1.

Despite the low read mapping rates for *Psilocybe tampanensis*, *Psilocybe galindoi* and *Psilocybe azurescens*, the *de novo* assemblies of these genomes have high BUSCO completeness scores (93%) and matching ITS sequences suggesting they are indeed *Psilocybe* but differ enough at the species level to produce low cross species read mapping rates.

An alternative hypothesis is that these libraries are metagenomic or contaminated with non-*Psilocybe* spp. fungi but still provide enough sequence coverage of the ITS region due to its high copy number. Contaminated libraries are usually detectable with bimodal sequencing coverage as its rare for the organisms to be equimolar. All five non-*P. cubensis* species have very uniform sequence coverage across the contigs and high BUSCO completion scores with strong ITS sequence implying clean assemblies.

In combination, these data demonstrate close relatedness of *Psilocybe tampanensis* and *Psilocybe galindoi* but distant relatedness to the *Psilocybe cubensis* “P.envy” reference genome. Given the known psilocybin production in these alternative *Psilocybe* species, these data also imply an alternative synthesis pathway must exist in these species as suggested by Awan.

Sequence coverage analysis

Uniform sequence coverage and high read mapping rates were observed over most of the of *P. cubensis* “P.envy” genome for most of the strains (50 of the 81 are displayed in Figure 2 and Figure 3). A few notable higher coverage exceptions were the ITS region on Scaffold_9 and the mitochondrial genome represented by Scaffold_26. This is not a surprising result given the increased copy number of ITS and mitochondrial sequences in fungal genomes.

A few smaller, repeat-rich contigs (scaffolds 27-31) demonstrated more variable coverage across the *Psilocybe* genus. The ITS region on Scaffold_9 appears to be a tandem repeat which is collapsed in the reference sequence. Scaffolds 27, 28, 31, and 32 have telomeric ends and are small (171 kb, 117 kb, 69 kb, 33 kb). While these smaller contigs have

Figure 1. (Left) Photographs of strains were obtained from the vendors public websites. (Middle Left). Sequence coverage of all contigs larger than 1 kb where coverage is on the Log Y axis and contigs sorted by length (largest to smallest: left to right) are on the X axis. (Middle Right) Assembly statistics from Quast 5.0. (Right Top) Number of Heterozygous and Homozygous SNPs detected with reads mapped to the *P. cubensis* "P.envy" reference. (Right Bottom) Number of Reads for each run according to Samtools flagstat.

Figure 1. (continued)

Figure 1. (continued)

Figure 1. (continued)

Figure 1. (continued)

Figure 1. (continued)

Figure 1. (continued)

Figure 1. (continued)

Figure 1. (continued)

Figure 2. Read mapping efficiency of species specific reads mapped to the *P. cubensis* “P.envy” reference genome. Due to size limitations, only 50 genomes are presented. The remaining read mapping results can be seen in the underlying data.

variable coverage in *Psilocybe* spp., it is important to understand that they represent less than 2% of the genome. Scaffolds 29 and 30 lack telomeric sequence and are 74 kb and 69 kb respectively. Three scaffolds have telomeric sequences on both ends. (Scaffold_2: 4.3 Mb, Scaffold_3: 4.1 Mb; Scaffold_9: 2.3 Mb) suggesting these are tip-to-tip chromosome assemblies.

The genomes most diverged from the *P. cubensis* “P.envy” reference genome (*P. tampanensis*, *P. galindoi*, *P. azurescens*) exhibited very low read mapping rates. This is demonstrated by *P. azurescens* (OR-Coast) in Figure 3. This genome-wide low mapping rate was also observed in the psilocybin production cassette on Scaffold_7 (Figure 4 and Figure 5).

Despite the low read mapping rates to *P. cubensis* “P.envy”, *P. tampanensis* assembled into 23 kb N50 genome with a 93% BUSCO completeness score and a 99% identical match to Rockefeller *et al.* ITS sequence for *P. tampanensis*. These low mapping rates are reflective of a highly diverged species at the nucleotide level but to exhaust the exploration of psilocybin production, amino acid level homology was further explored.

To search for pathway redundancy, we utilized tBLASTn to identify alternative psilocybin-producing enzymes in the reference *P. cubensis* “P.envy” genome. PsiK exists in the conical cluster on scaffold_7 but also has a close homolog on scaffold_1. The copy in the conical cluster on scaffold_7 has a unique SnpEFF high impact variant at p.Arg173Gly that only exists in *P. cubensis* “P.envy” (Figure 6B). The copy on Scaffold_1 is missing the 5' exon in *P. tampanensis*, *P. galindoi* and *P. azurescens*, and is significantly diverged. The most closely related Scaffold_7 PsiK amino acid sequences are seen in a multiple sequence alignment seen in Figure 6A. Further cloning and expression is required to confirm if this additional copy provides any pathway redundancy. Of interest, these alternative alignments are only partially clustered with PsiM and often PsiM is located in the middle of 80 kb contigs in the absence of other pathway related genes (Table 1). This demonstrates a non-clustered psilocybin synthesis pathway in the non-*P. cubensis* strains.

We constructed a phylogenetic tree (Figure 7) of the various genomes. Many of the genomes share strain names but were acquired from different spore banks (PS *versus* SW *versus* ITW *versus* Mush). These phylogenetically cluster together as expected. Many samples were sequenced multiple times as biological replicas and are represented in the sample nomenclature as Name_replica_vendor.

Figure 3. Coverage plot of 50 genomes mapped to *P. cubensis* “P.envy” reference. Coverage (log) on the Y axis and scaffolds 1-32 on the X axis. Scaffolds labeled with a T have a single ~200 bp terminal telomeric sequence. Scaffolds with TT have telomeric sequence on both ends and are assumed to be tip to tip telomeric chromosomes. Scaffolds are ordered from largest to smallest from left to right. *P. azurescens* (OR-Coast) has low mapping frequency due to divergence.

Conclusions

These data support both Reynolds *et al.* and Awan in that psilocybin synthesis appears to have evolved both a conserved ~20 kb cassette seen in many *P. cubensis* fungi but also a less clustered pathway in regards to a non-contiguous PsiM gene that still needs further characterization and scrutiny. Alternative psilocybin production has also been suggested in the cicada-infesting *Massospora* spp. fungi (Boyce *et al.* 2019). Taken together, these data underscore the need for further exploration of psilocybin genomics for alternative or redundant synthesis pathways. Given the divergence of the other psilocybin-producing mushrooms, simply mapping reads from other *Psilocybe* species to the *P. cubensis* “P.envy” reference genome can be misleading. Until more complete references exist for the *P. tampanensis*, *P. azurescens* and *P. galindoi*, searching for conserved amino acid sequences more tolerant to synonymous DNA mutations will be required. The genetic variation in the species is substantial and these genomes will help to further correlate genotype to chemotype associations for this nascent field.

Methods

DNA isolation

Spores were obtained from four vendors (sporeworks.com, Premiumspores.com, Mushrooms.com and Inoculate TheWorld.com). Spore preparations utilized a modified DNA isolation procedure described in McKernan *et al.* (McKernan 2021). Briefly 1.4 mL of spores was centrifuged, decanted and resuspended in 200 µL of ddH₂O. 25 µL

Figure 4. Integrated Genome Viewer (IGV) coverage plot over the psilocybin synthesis pathway described by Reynolds *et al.* Strains *Huautla* and *Mazatapec* have complete sequence coverage while *P. azurescens* is lacking sequencing coverage over most of the genes.

of a thaumatin-like protein was added and incubated at 37°C (Medicinal Genomics part #420206) for 30 minutes. 12.5 µL of MGC lysis buffer was added and incubated at 65°C for 30 minutes with 9 steel beads. Vortexing was performed every 7 minutes. Lysed sample were micro-centrifuged and 200 µL of supernatant was aspirated and added to 250 µL of Medicinal Genomics (MGC) binding buffer (MGC part# 420001) for magnetic bead isolation. The samples were incubated with the MGC magnetic bead mixture for 10 minutes, magnetically separated and washed two times with 70% ethanol. The beads were dried at 37°C for 5 minutes to remove excess ethanol and eluted with 25 µL of ddH₂O

Library construction for whole genome sequencing

Fragmentation

Genomic DNA (gDNA) was quantified with a Qubit (Thermo Fisher Scientific) and normalized to reflect 4–8 ng/µL in 13 µL of TE buffer. Libraries were generated using enzymatic fragmentation with the NEB Ultra II kits (NEB part # E7103). Briefly, 3.5 µL of 5× NEB fragmentation buffer and 1 µL of Ultra II fragmentation enzyme mix are added to 13 µL of DNA. This reaction was tip-mixed 10 times, vortexed, and quickly centrifuged. Fragmentation was performed in a BioRad CFX96 thermocycler at 3.5 minutes at 37°C, 30 minutes at 65°C. The reaction was kept on ice until ready for adaptor ligation.

Adaptor ligation

The master mix for ligation was prepared on ice using 0.75 µL of Agilent SureSelect Adaptor Oligo Mix, 0.5 µL of ddH₂O, 15 µL of New England Biolabs (NEB) Ultra II Ligation Master Mix, 0.5 µL of NEB Ligation enhancer for a total reaction volume of 16.75 µL.

Ligation was performed by the addition of 16.75 µL of ligation master mix to the 17.5 µL Fragmentation/End Prep DNA reaction mixture, incubate for 15 minutes at 20°C. To purify excess adaptors and adaptor dimers, vortex at room temp AMPure XP beads (Beckman Coulter #A63881) to resuspend and add 16 µL (approximately 0.45×) of resuspended AMPure XP beads to the ligation reactions. Mix well by pipetting up and down at least 10 times. Incubate the mixture for 5 minutes at 25°C. Put the PCR plate on an appropriate magnetic stand (Medicinal Genomics #420202) to separate beads from supernatant. After the solution is clear (about 5 minutes), carefully remove and discard the supernatant. Be careful not to disturb the beads that contain DNA targets. Wash the magnetic beads by adding 200 µL of 70% ethanol to the PCR plate while in the magnetic stand. Incubate at room temperature for 30 seconds, and then carefully remove and discard the supernatant. Repeat the ethanol wash once for a total of 2 washes. Remove trace amounts of ethanol from the beads. Air dry beads for ~7 minutes while the PCR plate is on the magnetic stand with the lid open. Remove PCR plate from magnet and elute DNA target from beads into 10 µL of H₂O, transfer 9 µL cleaned DNA to a fresh well.

Sequence coverage over Psilocybin synthesis cluster

Figure 5. Top-Heatmap of Illumina un-normalized raw sequence coverage of the psilocybin synthesis pathway across 43 genomes. Low coverage genomes of the same strain (*Costa Rica* and *Samui*) were summed together. Bottom- IGV view of three strains with close homology to the reference genome.

PCR amplification

Add 12.5 μL $2\times$ NEBNext Q5 Hot Start Master Mix (New England Biolabs #M0492S) to 9 μL ligated DNA, then add 3.5 μL NEB 8-bp index primer/universal primer. The reaction was run the cycling program using 98°C for 30 seconds as an initial denaturization step. 6 cycles of denaturization, annealing and extension were performed cycling between 98°C for 10 seconds and 65°C for 75 seconds. A final 65°C for 5 minutes was performed with a final 4°C forever step.

PCR reaction cleanup

Resuspend room temp AMPure XP beads with a brief vortex. Add 15 μL of resuspended AMPure XP beads to the PCR reactions (~25 μL). Mix well by pipetting up and down at least 10 times. Incubate mixture for 5 minutes at room temperature. Put the PCR plate on an appropriate magnetic stand to separate beads from supernatant. After the solution is

Figure 6A. Multiple sequence alignment using CLUSTALW of the most closely related PsiK genes in *P. tampanensis*, *P. galindoi* and *P. azurensens*. The R173G mutation is also unique to *P. cubensis* "P.envy" when considering non-cubensis species.

clear (about 5 minutes), carefully remove and discard the supernatant. Be careful not to disturb the beads that contain DNA targets. Add 200 μ L of 70% ethanol to the PCR plate while in the magnetic stand. Incubate at room temperature for 30 seconds, and then carefully remove and discard the supernatant. Repeat the ethanol wash once. Air dry beads for 7 minutes while the PCR plate is on the magnetic stand with the lid open. Elute DNA target from beads into 15 μ L of nuclease free H₂O, transfer 15 μ L into a fresh well.

Sample quality control and sequencing

Libraries were evaluated on an Agilent Tape Station prior to pooling for Illumina sequencing. Sequencing was performed by GeneWiz, Cambridge MA. A total of 690 million paired reads (2 $\times$ 150 bp) were generated, averaging over 12 million read pairs per sample and a total sequence of 207 Gb.

JR316_008011-PSID-scaffold_7

JR316_008010-PSIM-scaffold_7

Figure 6B. (continued)

The VCF files were merged using bcftools and left-aligned to split up multi-allelic entries

```
bcftools merge *gvcf.vcf.gz --output-type v > merged.gvcf.gz
bcftools norm -f/NGS/genewiz-june1st-
10spores/ref/fungus/Psilocybe_cubensis_Envy_scaffolds.fa -m-any merged-norm.vcf.gz
```

The annotation from the *funannotate* pipeline was converted from gff3 format into SnpEff (v4.3t. 2017-11-24) database as described here (https://pcingola.github.io/SnpEff/se_buildingdb/)

Phylogenetic analysis utilized APE (v5.3.) and SNPRelate (v. 1.20.1) (Zheng *et al.* 2012; Paradis and Schliep 2019).

Table 1. tBLASTn search of *P. tampanensis*, *P. galindoi*, *P. cubensis* and *P. azurensis* for PsiK, PsiH, PsiD and PsiM reveals a fragmented psilocybin synthesis cluster with PsiM most often being on large but independent contigs disrupting the contiguity of the psilocybin synthesis cluster.

NCBI Name	JAFHE01.1	JAILYQ01.1	JAILYP01.1	JAIFHD01.1	SRX10250988
	<i>P. azurensis</i>	<i>P. tampanensis_1</i>	<i>P. tampanensis_2</i>	<i>P. mexicana</i>	<i>P. cubensis</i> (GT)
PsiK_Contig	JAIFHE010009952.1	JAILYQ010013513.1	JAILYP010003074.1	JAIFHD010015065.1	K141_42154
Contig Size	4,254	9,322	8,836	4,016	76,612
PsiH_Contig	JAIFHE010002467.1	JAILYQ010013513.1	JAILYP010003074.1	JAIFHD010015065.1	K141_42154
Contig Size	40,681	9,322	8,836	4,016	76,612
PsiD_Contig	JAIFHE010002467.1	JAILYQ010012987.1	JAILYP010013169.1	JAIFHD010017873.1	K141_42154
Contig Size	40,681	44,435	44,435	3,528	76,612
PsiM_Contig	JAIFHE010002467.1	JAILYQ010005515.1	JAILYP010014511.1	JAIFHD010002461.1	K141_42154
Contig Size	40,681	81,911	82,507	880	76,612

Figure 7. Phylogenetic tree of 81 psilocybe strains displayed using iTOL (<https://itol.embl.de>). *Panaeolus Copelandia* samples has an *P. cubensis* ITS sequence is likely a mislabeled sample.

Coverage analysis

Coverage analysis was performed using BAM files mapped to the *P. cubensis* “P.envy” genome reference using bwa-mem as described above (Li 2018). BAM files were analyzed using pileup.sh from the BBmap version 4/30/2020 (http://bib.irb.hr/datoteka/773708.Josip_Maric_diplomski.pdf). The Heatmap was generated using Heatmap.2 from the gplots version 3.1.1 in R Studio (<https://www.rdocumentation.org/packages/gplots/versions/3.1.1/topics/heatmap.2>).

Data availability

Underlying data

Zenodo: Underlying data for ‘A whole genome atlas of 81 *Psilocybe* spp. genomes as a resource for psilocybin production’. <https://doi.org/10.5281/zenodo.5062843> (McKernan *et al.* 2021).

The project contains the following underlying data:

Accession number: NCBI, Project ID PRJNA700437 and PRJNA687911

Root URL: <https://www.ncbi.nlm.nih.gov/search/all/?term=PRJNA700437>

Accession number URL: <https://www.ncbi.nlm.nih.gov/search/all/?term=PRJNA687911>

Data are available under the terms of the [Creative Commons Attribution 4.0 International license](#) (CC-BY 4.0).

References

- Attademo L, Bernardini F: **Are dopamine and serotonin involved in COVID-19 pathophysiology?** *Eur J Psychiatry*. 2021; **35**: 62–63.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Awan: **Convergent evolution of psilocybin biosynthesis by psychedelic mushrooms.** *BioRxiv*. 2018.
[Publisher Full Text](#)
- Bogenschutz MP, Podrebarac SK, Duane JH, *et al.*: **Clinical Interpretations of Patient Experience in a Trial of Psilocybin-Assisted Psychotherapy for Alcohol Use Disorder.** *Front Pharmacol*. 2018; **9**: 100.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Boyce GR, Gluck-Thaler E, Slot JC, *et al.*: **Psychoactive plant- and mushroom-associated alkaloids from two behavior modifying cicada pathogens.** *Fungal Ecol*. 2019; **41**: 147–164.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Carhart-Harris R, Giribaldi B, Watts R, *et al.*: **Trial of Psilocybin versus Escitalopram for Depression.** *N Eng J Med*. 2021; **384**: 1402–1411.
[Publisher Full Text](#)
- Creeden JF, Imami AS, Eby HM, *et al.*: **Fluoxetine as an anti-inflammatory therapy in SARS-CoV-2 infection.** *Biomed Pharmacother*. 2021; **138**: 111437.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Danecek P, Bonfield JK, Liddle J, *et al.*: **Twelve years of SAMtools and BCFtools.** *GigaScience*. 2021; **10**.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Davis AK, Barrett FS, May DG, *et al.*: **Effects of Psilocybin-Assisted Therapy on Major Depressive Disorder: A Randomized Clinical Trial.** *JAMA Psychiatry*. 2021; **78**: 481–489.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dechaumes A, Nekoua MP, Belouzard S, *et al.*: **Fluoxetine Can Inhibit SARS-CoV-2 In Vitro.** *Microorganisms*. 2021; **9**.
[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](#)
- Griffiths RR, Johnson MW, Carducci MA, *et al.*: **Psilocybin produces substantial and sustained decreases in depression and anxiety in patients with life-threatening cancer: A randomized double-blind trial.** *J Psychopharmacol*. 2016; **30**: 1181–1197.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free 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](#)
- Hashemi H, Haghdoust AA, Haji-Aghajani M, *et al.*: **A Successful implementation of an idea to a nationally approved plan: Analyzing Iran's National Health Roadmap using the Kingdon model of policymaking.** *Med J Islam Repub Iran*. 2018; **32**: 46.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Li D, Liu CM, Luo R, *et al.*: **MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph.** *Bioinformatics*. 2015; **31**: 1674–1676.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Li D, Luo R, Liu CM, *et al.*: **MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices.** *Methods*. 2016; **102**: 3–11.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Li H: **Minimap2: pairwise alignment for nucleotide sequences.** *Bioinformatics*. 2018; **34**: 3094–3100.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- McKernan K, Liam TK, Yvonne H, *et al.*: **A whole genome atlas of 81 Psilocybe genomes as a resource for psilocybin production.** (1.0). *Zenodo*. 2021.
[Publisher Full Text](#)
- McKernan K: **Whole genome sequencing of colonies derived from cannabis flowers and the impact of media selection on benchmarking total yeast and mold detection tools.** *Zenodo*. 2021.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Paradis E, Schliep K: **ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R.** *Bioinformatics*. 2019; **35**: 526–528.
[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](#)
- Ross S, Bossis A, Guss J, *et al.*: **Rapid and sustained symptom reduction following psilocybin treatment for anxiety and depression in patients with life-threatening cancer: a randomized controlled trial.** *J Psychopharmacol*. 2016; **30**: 1165–1180.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Schloer S, Brunotte L, Goretzko J, *et al.*: **Targeting the endolysosomal host-SARS-CoV-2 interface by clinically licensed functional inhibitors of acid sphingomyelinase (FIASMA) including the antidepressant fluoxetine.** *Emerg Microbes Infect*. 2020; **9**: 2245–2255.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Schloer S, Brunotte L, Mecate-Zambrano A, *et al.*: **Drug synergy of combinatory treatment with remdesivir and the repurposed drugs fluoxetine and itraconazole effectively impairs SARS-CoV-2 infection in vitro.** *Br J Pharmacol*. 2021; **178**: 2339–2350.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Tullis (2021): **How ecstasy and psilocybin are shaking up psychiatry.** *Nature*. 2021; **589**, 506–509.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Zheng X, Levine D, Shen J, *et al.*: **A high-performance computing toolset for relatedness and principal component analysis of SNP data.** *Bioinformatics*. 2012; **28**: 3326–3328.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Zimniak M, Kirschner L, Hilpert H, *et al.*: **The serotonin reuptake inhibitor Fluoxetine inhibits SARS-CoV-2 in human lung tissue.** *Sci Reports*. 2021; **11**: 5890.
[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 19 October 2021

<https://doi.org/10.5256/f1000research.58864.r95528>

© 2021 Dentinger B 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.

Bryn T. M. Dentinger

Natural History Museum of Utah & School of Biological Sciences,, University of Utah, Salt Lake City, UT, USA

Alexander Bradshaw

Natural History Museum of Utah & School of Biological Sciences, University of Utah, Salt Lake City, UT, USA

The authors present genomic data from 81 samples of putative *Psilocybe* spp. The source of the samples used for sequencing is problematic in several ways, not the least being clearly documented misidentification and contamination. This reality limits the accuracy of the interpretations made. Moreover, these results are not reproducible without some form of voucher material, but the authors make no mention that such material exists. Therefore, the interpretation of the data presented is entirely dependent on identifications provided by the retail vendors that supplied the spore samples. Yet there is little to suggest that these have the appropriate authority or independently verifiable reference library to justify these identifications.

Ultimately, then, what the authors present are genomic data for 81 spore samples from what are putatively identified as *Psilocybe* spp., but which most appear to be strains of the widely cultivated *P. cubensis*. Still, these are only verified by corroborating data from spore supplies sourced from a few vendors. In my opinion, this is not very rigorous science (although I must sympathize with the difficulty presented by legal authorities to do much better). Nonetheless, the genomic data do provide some preliminary insight into the variation of the psilocybin biosynthesis pathway that may be expected from variants of *P. cubensis* and confirms earlier reports of variation in other *Psilocybe* spp., including the potential existence of an alternative, non-canonical pathway.

Overall, the manuscript seems hastily written. The Introduction is brief and framed around what appears to be a misinterpretation of the published literature. There are many misspellings, grammatical errors, and awkwardly worded statements throughout. Capitalizations and italics are not used in a consistently appropriate way. The authors should also provide authorities of scientific names at first use to avoid nomenclatural ambiguity.

More specific comments are below:

- The potential therapeutic value of psilocybin for SARs-CoV-2 seems a bit hand-wavy.
- *“Two models have been proposed for psilocybin production in *Psilocybe* spp. Horizontal gene transfer has been thoroughly described by Reynolds et al. and Fricke et al., while convergent evolution has been proposed by Awan (Fricke et al. 2017; Awan 2018; Reynolds et al. 2018).”*

This is incorrect. None of these studies proposed either HGT nor convergent evolution for psilocybin biosynthesis IN *Psilocybe* spp. HGT has been proposed as the mechanism for the gain of psilocybin biosynthesis in other genera (e.g., *Panaeolus*, *Pluteus*) and convergent evolution has been proposed for the gain of psilocybin biosynthesis in the genus *Inocybe*. Thus, these statements are inaccurate representations of the published literature and need to be corrected.

- *“Given the broad geographic distribution of psilocybin-producing mushrooms and the history of human cultivation and selection for psilocybin-producing strains, these models may not be mutually exclusive.”*

I agree that these models are not necessarily mutually exclusive (indeed, none of the previous studies suggested so), but what does the “history of human cultivation and selection for psilocybin-producing strains” have to do with it? Moreover, can the authors cite the statement of “history of human cultivation” (what history)? And what evidence is there for “selection for psilocybin-producing strains”?

- *“These non-conical psilocybin-synthesizing genomes...”*

Do the authors mean “non-canonical”? N.B. the term “conical” appears throughout the manuscript where I suspect the term “canonical” is intended.

- *“...99%, 100%, and 91% ITS sequence identity to a *Psilocybe tampanensis* sequence described by Rockefeller et al. (NCBI accession number: MH220315.1)”*

First, “described” is not the correct term. Second, the sequence in GenBank was derived from a specimen identified as *P. tampanensis*. Third, the percent identities cited are not sufficient --- percent identity over what portion of the sequence? And was this one sequence the top hit for all three of the ITS sequences derived from the WGS?

- The degree to which the spore samples apparently consisted of contaminants and in at least one clear case a gross misidentification calls into question the identity, reliability and reproducibility of the results from the other “strains.”
- *“Contaminated libraries are usually detectable with bimodal sequencing coverage as its rare for the organisms to be equimolar. All five non-*P. cubensis* species have very uniform sequence coverage across the contigs and high BUSCO completion scores with strong ITS sequence implying clean assemblies.”*

This is one way to determine endogenous content, but other methods to detect contaminant DNAs in metagenomic datasets (e.g., taxon-annotated GC plots) should be

explored to corroborate the indirect inference from sequencing coverage (esp. b/c this assumes that target and contaminant sequencing coverage is bimodal, which is not necessarily the case if there are multiple contaminants and/or contaminants are at varying proportions relative to target).

- *"In combination, these data demonstrate close relatedness of *Psilocybe tampanensis* and *Psilocybe galindoi* but distant relatedness to the *Psilocybe cubensis* "P.envy" reference genome."*

The authors should be more cautious about this interpretation, as there is no evidence to support the identifications of the strains supplied nor their identity with authentic specimens (e.g., types or others). Who knows if these spores labeled as these species/strains are correct?

- *"Given the known psilocybin production in these alternative *Psilocybe* species, these data also imply an alternative synthesis pathway must exist in these species as suggested by Awan."*

It should be made clear that in Awan *et al.*, it was proposed that convergent evolution explains psilocybin biosynthesis in *Inocybe*, a distant relative of *Psilocybe*. On the other hand, Awan *et al.* did suggest there may be an alternative pathway present in the *Psilocybe* genomes. Which of these are the authors referring to here?

- *"PsiK exists in the conical cluster on scaffold_7 but also has a close homolog on scaffold_1."*

While this may be true, the high potential for contaminant DNAs in the samples sequenced suggests this may be the result of a metagenome assembly. This alternative needs to be ruled out.

- *"biological replicas"*

Replicates? And what evidence is there that these are truly biological replicates rather than the same spores being sourced from multiple places? (what's the likelihood that these companies are all deriving their spores independently?)

- *"These data support both Reynolds et al. and Awan in that psilocybin synthesis appears to have evolved both a conserved ~20 kb cassette seen in many *P. cubensis* fungi but also a less clustered pathway in regards to a non-contiguous PsiM gene that still needs further characterization."*

Please be more specific wrt what in these previous studies is being cited. At the beginning the authors incorrectly cite these papers as evidence of HGT and convergent evolution of psilocybin biosynthesis in *Psilocybe*. So what exactly is being cited here?

- *"Given the divergence of the other psilocybin-producing mushrooms, simply mapping reads from other *Psilocybe* species to the *P. cubensis* "P.envy" reference genome can be misleading."*

Why did the authors not bother to attempt to align these to the other existing *Psilocybe* genomes? Two other (potentially closely related) genomes are available (*P. cyanescens* and *P. serbica*), yet the authors made no attempt to align their reads from the putative *P. tampanensis*, *P. galindoi*, or *P. azurescens* to them (the latter is a potential synonym of *P.*

cyanescens). This could help further clarify the source of the low mapping rates to "P. envy" and may provide further insight into the non-canonical homologs of PsiK and PsiM that they report here.

- Methods: generally thorough enough to be reproducible, but in some parts it reads like a laboratory protocol rather than text written for a manuscript.
- Can the authors display the phylogenetic tree differently? As it currently is, it is difficult to interpret clustering because the branch lengths have been transformed in a way that is awkward to read. I would also suggest using an unrooted tree and provide branch support values.

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

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

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: genomics, phylogenetics

We confirm that we have read this submission and believe that we have an appropriate level of expertise to state that we do not consider it to be of an acceptable scientific standard, for reasons outlined above.

Author Response 09 Dec 2021

Kevin McKernan

Version 2 Per recommendation of the Moderator/Editor.

Response to Dentinger and Bradshaw

We appreciate the many helpful comments from Dentinger and Bradshaw. Many of them

are on point and the manuscript will be adjusted accordingly. A few comments appear to be a non-addressable given the legal status of live tissue of *Psilocybe cubensis*. As the reviewers themselves point out, this is not unique to this manuscript. We will address these points as well as we can, given the current legal climate.

1) The potential therapeutic value of psilocybin for SARs-CoV-2 seems a bit hand-wavy.

Fair point. The SARs connections have since strengthened and could in fact use more detail and citations to support. The 9K person clinical trial for Fluvoxamine just completed this month and is showing better efficacy than the vaccines. We have added more references and language to back this association and we have itemized the link as being associative to address the reviewers concerns over this nascent field.

2) This is incorrect. None of these studies proposed either HGT nor convergent evolution for psilocybin biosynthesis IN *Psilocybe* spp. HGT has been proposed as the mechanism for the gain of psilocybin biosynthesis in other genera (e.g., *Panaeolus*, *Pluteus*) and convergent evolution has been proposed for the gain of psilocybin biosynthesis in the genus *Inocybe*. Thus, these statements are inaccurate representations of the published literature and need to be corrected.

This has been corrected. This is an important and valid critique. A prior draft of this manuscript referred to this as "hallucinogenic mushrooms". In attempting to be more specific, we erroneously replaced this with *Psilocybe* spp. Which is too narrow and excludes the relevant families raised by the reviewers.

3) *"Given the broad geographic distribution of psilocybin-producing mushrooms and the history of human cultivation and selection for psilocybin-producing strains, these models may not be mutually exclusive."*

We have removed the discussion of human selection.

4) Do the authors mean "non-canonical"? N.B. the term "conical" appears throughout the manuscript where I suspect the term "canonical" is intended.

We have changed the conical/canonical language to "gene cluster" as this appears to be the more agreed upon language describing this recently described region of interest.

5) First, "described" is not the correct term. Second, the sequence in GenBank was derived from a specimen identified as *P. tampanensis*. Third, the percent identities cited are not sufficient --- percent identity over what portion of the sequence? And was this one sequence the top hit for all three of the ITS sequences derived from the WGS?

We have changed this to “reported to NCBI by”. We agree and clearly state the Rockefeller sequence is from *P.tampanensis* as you suggest. The listed percentages are across the full-length accession described by Rockefeller (946bp). We have annotated the length of the alignments and included an additional accession from Wesselink *et al.* from *P.mexicana*.

6) The degree to which the spore samples apparently consisted of contaminants and in at least one clear case a gross misidentification calls into question the identity, reliability and reproducibility of the results from the other “strains.”

We believe this is an unfair critique as our transparency is being weaponized against us. We could have easily omitted the contaminated sample from the manuscript but believe negative results are informative. It is important for readers to know that contamination does exist in the circulating spore preps and it is important to build tools to detect and screen for them. We should also be careful not to blame the spore providers for this contamination given endofungal bacteria are known to exist in many fungi and we can't be certain of the source of the contamination (*Aspergillus* is ubiquitous and our lab handles it). To attack the other sample veracity, because the authors were honest about one obvious contamination event out of 81 genomes is counter-productive and leads to more scientists hiding their negative results to avoid critiques like this.

7) This is one way to determine endogenous content, but other methods to detect contaminant DNAs in metagenomic datasets (e.g., taxon-annotated GC plots) should be explored to corroborate the indirect inference from sequencing coverage (esp. b/c this assumes that target and contaminant sequencing coverage is bimodal, which is not necessarily the case if there are multiple contaminants and/or contaminants are at varying proportions relative to target).

We provided read mapping rates for all libraries. Highly contaminated samples don't produce 90% read mapping rates unless the contamination is of the same species. This is evident in the assembly statistics we provided for every genome sequenced. Furthermore, assemblies are filtered with tBLASTn to evaluate genes with high amino acid sequence homology to *P.cubensis* specific genes (PsiK, PsiM, PsiD etc). Any additional minor contamination is thus filtered from this part of the analysis.

Nevertheless, we have added the suggested Taxon-annotated GC plots for the 5 of the *P.cubensis* libraries and 5 of the non-cubensis libraries per the reviewers recommendations (supplementary figure 1-4). There is some bacterial contamination evident in the *P.galindoi* sample but as mentioned before, we have filtered the contigs of interest out using tBLASTn for the relevant genes of interest. This isn't a relevant critique related to the assessment of the contiguity of the Psilocybin gene cluster in question.

8) The authors should be more cautious about this interpretation, as there is no evidence to support the identifications of the strains supplied nor their identity with authentic specimens (e.g., types or others). Who knows if these spores labeled as these

species/strains are correct?

We have added additional disclaimers to address the concerns raised over the spore providers veracity. In many ways these 81 genomes best address the very concerns these reviewers present. We were careful to sequence many strains with identical names from multiple vendors so this question could be addressed. There is no other biobank of *P.cubensis* that has this level of public sequence verification on the nomenclature system.

We'd like to emphasize to the editors and readership that the reviewers correctly point out that the psilocybin legality issue plagues all studies in *P.cubensis* but this should not be used as a club to discredit manuscripts generating data that might address the very issue. It can even be seen in the reviewers own excellent work which we cite. Below is the methods section from Awan *et al.*

"Psilocybe cyanescens genome Illumina sequencing

DNA was extracted from fresh whole mushrooms of *P. cyanescens* by first freezing with liquid nitrogen and then grinding the tissue into a fine powder with a mortar and pestle in 2X

CTAB buffer following a modified protocol from Doyle (1991) DNA extracts were then cleaned through a CsCl density-gradient centrifugation and resuspended in TE buffer. DNAs were sent to Eurofins Scientific (Ebersberg, Germany) for 2 x 150 paired-end sequencing on an Illumina HiSeq 2500 platform."

No information is provided on how to acquire the samples Awan *et al.* sequenced. We have at least provided a vendor name and we were careful to collect samples with the same name from multiple vendors in replicates to test the veracity of the spore providers nomenclature. We believe our manuscript does a more thorough job to address such concerns than manuscripts prior as most previous studies simply sequence a few samples without replicates or guidance on spore providers.

While this is not the core focus of our manuscript, we would turn the reviewer's attention to the sequence and the SNP metrics provided for the samples with the same name but from multiple different spore providers as these metrics are informative to this question. This is better sample veracity than we observe in the cannabis market and we are unaware of any study that has this level of sequence data to authenticate the nomenclature system in *Psilocybe*.

We will reiterate that the nomenclature in the field is still very nascent and unvalidated. Sequencing studies like this are the first steps required to catalog it.

9) It should be made clear that in Awan *et al.*, it was proposed that convergent evolution explains psilocybin biosynthesis in *Inocybe*, a distant relative of *Psilocybe*. On the other hand, Awan *et al.* did suggest there may be an alternative pathway present in the *Psilocybe* genomes. Which of these are the authors referring to here?

We have clarified Awan *et al.* description of convergent evolution.

10) "*PsiK* exists in the conical cluster on scaffold_7 but also has a close homolog on scaffold_1."

While this may be true, the high potential for contaminant DNAs in the samples sequenced suggests this may be the result of a metagenome assembly. This alternative needs to be ruled out.

We believe this is a misunderstanding. Scaffold_1 is from a peer reviewed Pacific Biosciences P.envy HiFi assembly. This genome is an annotation grade, 32 contig assembly with a N50 over 3.3Mb. We have labelled each scaffold in that paragraph with P.envy scaffold # to help clarify which references we are referring to. P.envy has been annotated and is in NCBI. NCBI filters for such contaminant when importing reference grade genome assemblies and rarely to do bacterial contaminants consist of a 4.5Mb HiFi chromosome (with a telomere) that shares equal coverage to all of the other *Psilocybe* contigs. Scaffold_1 and Scaffold_7 in the P.envy reference genome have already been screened for contaminants ([source](#)).

11) Replicates? And what evidence is there that these are truly biological replicates rather than the same spores being sourced from multiple places? (what's the likelihood that these companies are all deriving their spores independently?)

Biological replicates from the same vendor are critical to understand how much the sequencing process varies on the same biological sample. The study was careful to select inter-vendor and intra-vendor replicates. This question is best addressed with the data now public with this manuscript. Of note is the very high genotyping concordance between samples with the same name from the same vendor (Malabar). This genotyping concordance is shared with Malabar samples from different vendors (SW and PS). Contrary to this, B-plus samples from different vendors have lower genotyping concordance. This suggests the spore providers are not clones of each other but do have some strains that appear to be clonal with other providers while also hosting some unique genetics under the same name. This data is in Figure 1. A complete genomic audit trail on the vendor veracity is outside the scope of this preliminary analysis but data does currently exist in Figure 1 to address these concerns. All of the FastQ files are in the SRA for people to explore these other questions.

12) Please be more specific wrt what in these previous studies is being cited. At the beginning the authors incorrectly cite these papers as evidence of HGT and convergent evolution of psilocybin biosynthesis in *Psilocybe*. So what exactly is being cited here?

We have clarified the Family/Genus regarding Reynolds and Awan.

13) Why did the authors not bother to attempt to align these to the other existing *Psilocybe* genomes? Two other (potentially closely related) genomes are available (*P. cyanescens* and *P. serbica*), yet the authors made no attempt to align their reads from the putative *P. tampanensis*, *P. galindoi*, or *P. azurescens* to them (the latter is a potential synonym of *P.*

cyanescens). This could help further clarify the source of the low mapping rates to "P. envy" and may provide further insight into the non-canonical homologs of PsiK and PsiM that they report here.

We have already compared the P.envy references to these submissions in McKernan *et al.* ([link](#)). We will reiterate that the gene cluster is intact in those species. Neither of these suggested references (*P.serbica* and *P.cyanescens*) have raw reads in the SRA. As a result, they cannot be run through blobtools for Taxon-annotated GC plots (coverage required from a mapped BAM file) to address the contamination problems on the same footing as the other data you are asking for comparisons to. Nevertheless we have mapped the reads from 5 non-cubensis strains to the above references and added a new table to demonstrate these results (Table 1). We also included 3 cubensis strains as a comparison. You can see from this work that *P.azurescens* is indeed more closely related to *P.cyanescens*, as the reviewers suggest, but *P.tampanensis* and *P.galindoi* are more closely related to each other but more distant to *P.cyanescens*.

14) Methods: generally thorough enough to be reproducible, but in some parts it reads like a laboratory protocol rather than text written for a manuscript.

We have reformatted this section to read more as a journal method than a lab protocol.

15) Can the authors display the phylogenetic tree differently? As it currently is, it is difficult to interpret clustering because the branch lengths have been transformed in a way that is awkward to read. I would also suggest using an unrooted tree and provide branch support values.

We have added a [link](#) to the newick file so readers can redisplay the phylogenetic tree multiple different ways.

We'd like to thank the reviewers for their careful review of the manuscript and for pointing out critical taxonomy considerations.

Competing Interests: None

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

F1000Research