A whole genome atlas of 81 Psilocybe genomes as a resource for psilocybin production.

Kevin J. McKernan, Liam T. Kane, Yvonne Helbert, Lei Zhang, Nathan Houde, Stephen McLaughlin.
Medicinal Genomics, 100 Cummings Center Beverly MA, 01915

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.

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). SSRI’s 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. Horizontal gene transfer has been thoroughly described by Reynolds et al. and Fricke et al., while convergent evolution has been proposed by Awan et al. (Fricke et al. 2017; Awan 2018; Reynolds et al. 2018). Given the broad geographic distribution of psilocybe 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 spore samples to assess the sequence coverage over the previously well characterized Psilocybe synthesis cassette.
Results
In a whole genome sequencing survey of 81 Psilocybe genomes we noticed a lack of sequence coverage over the Psilocybe gene cluster described by Reynolds et al. for 5 of the 81 Psilocybin producing genomes (3 *Psilocybe tampanensis*, 1 *Psilocybe galindoii*, 1 *Psilocybe azurescens*). These genomes also exhibited genome-wide poor read mapping efficiency to the P.Envy reference genome with the exception of the mitochondria and ITS regions. These non-conical psilocybe 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.

**Figure 1.** (Left) Photographs of strains were obtained from the vendors public websites. (Middle Left). Sequence coverage of all contigs larger than 1Kb where coverage is on the Log Y axis and contigs sorted by length 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.Envy reference. (Right Bottom) Number of Reads for each run according to Samtools flagstat.
### Assembly Coverage

| Assembly | Coverage (contigs) | Total Length (bp) | N50 (bp) |
|----------|-------------------|-------------------|----------|
| Mauve-1-PS | 1,500,000 | 29,652,014 | 22,003,067 |
| Mauve-2-PS | 1,500,000 | 29,652,014 | 22,003,067 |
| Maxasp-1-SW | 1,500,000 | 29,652,014 | 22,003,067 |

### Assembly Statistics

| Assembly | GC (%) | N50 (bp) |
|----------|--------|----------|
| Mauve-1-PS | 41.4% | 22,003,067 |
| Mauve-2-PS | 41.4% | 22,003,067 |
| Maxasp-1-SW | 41.4% | 22,003,067 |

### SNPs vs Hf170

| Reference | Homo % |
|-----------|--------|
| Mauve-1-PS | 79% |
| Mauve-2-PS | 79% |
| Maxasp-1-SW | 79% |

### Reads Mapped Back to Reference

| Reference | Homo % |
|-----------|--------|
| Mauve-1-PS | 79% |
| Mauve-2-PS | 79% |
| Maxasp-1-SW | 79% |

### Colony Image

| Colony Image | Assembly Coverage (contigs) | Total Length (bp) | N50 (bp) |
|--------------|-------------------------------|-------------------|----------|
| Penny-Coli-PS | 1,500,000 | 29,652,014 | 22,003,067 |
| Penny-Coli-SW | 1,500,000 | 29,652,014 | 22,003,067 |
| PES-Works-SW | 1,500,000 | 29,652,014 | 22,003,067 |

### Assembly Statistics

| Assembly | GC (%) | N50 (bp) |
|----------|--------|----------|
| Penny-Coli-PS | 41.4% | 22,003,067 |
| Penny-Coli-SW | 41.4% | 22,003,067 |
| PES-Works-SW | 41.4% | 22,003,067 |

### SNPs vs Hf170

| Reference | Homo % |
|-----------|--------|
| Penny-Coli-PS | 79% |
| Penny-Coli-SW | 79% |
| PES-Works-SW | 79% |

### Reads Mapped Back to Reference

| Reference | Homo % |
|-----------|--------|
| Penny-Coli-PS | 79% |
| Penny-Coli-SW | 79% |
| PES-Works-SW | 79% |

### Summary

- The assemblies show high coverage and quality, indicating successful sequencing.
- The SNPs vs Hf170 analysis indicates a high percentage of Homo SNPs.
- Mapped reads show high percentages, suggesting effective mapping.

---

**Note:** The data represents a partial analysis due to the complexity of the image. Further details and analysis would require a comprehensive examination of the dataset and the associated software outputs.
| Colony Image | Assembly Coverage (%) Large < Small Config (%) | Assembly Statistics (Quant 5.0) | SNVs vs Hifi-Fancy Reference | homo %  |
|--------------|-------------------------------------------------|---------------------------------|-----------------------------|--------|
| Pink-Buffalo-2-PS Glena | Pink_Ruffalo_2_P Glena | # contigs (<= 25000 bp) 148 | # contigs (<= 50000 bp) 266 | Total length (>= 20000 bp) 29,432,534 | Total length (>= 50000 bp) 22,024,000 | Contingent Length 1,683 | Largest contig 473,399 | Total length 45,103,984 | NC 45 | NS 46,816 | Contaminated |
| Purple-Mystic-3-PS Glena | Purple_Mystic_3_P Glena | # contigs (<= 25000 bp) 86 | # contigs (<= 50000 bp) 73 | Total length (>= 20000 bp) 28,973,593 | Total length (>= 50000 bp) 24,481,577 | Contingent Length 1,587 | Largest contig 1,798,965 | Total length 33,359,643 | NC 48 | NS 518,241 | Contaminated |
| Barren-2-PS Glena | Barren_2_P Glena | # contigs (<= 25000 bp) 379 | # contigs (<= 50000 bp) 187 | Total length (>= 20000 bp) 26,141,217 | Total length (>= 50000 bp) 19,544,790 | Contingent Length 1,679 | Largest contig 385,314 | Total length 42,682,113 | NC 46 | NS 41,618 | Contaminated |
| Tampa-1-SW Glena | Tampa_1_SW Glena | # contigs (<= 25000 bp) 360 | # contigs (<= 50000 bp) 182 | Total length (>= 20000 bp) 25,930,738 | Total length (>= 50000 bp) 14,944,249 | Contingent Length 8,126 | Largest contig 380,786 | Total length 40,472,653 | NC 47 | NS 23,583 | Contaminated |
| Tampa-2-SW Glena | Tampa_2_SW Glena | # contigs (<= 25000 bp) 265 | # contigs (<= 50000 bp) 162 | Total length (>= 20000 bp) 23,285,118 | Total length (>= 50000 bp) 15,047,879 | Contingent Length 8,090 | Largest contig 386,240 | Total length 45,534,020 | NC 47 | NS 29,499 | Contaminated |
| Tampa-3-SW Glena | Tampa_3_SW Glena | # contigs (<= 25000 bp) 389 | # contigs (<= 50000 bp) 172 | Total length (>= 20000 bp) 21,727,813 | Total length (>= 50000 bp) 13,402,298 | Contingent Length 7,874 | Largest contig 381,756 | Total length 45,100,156 | NC 47 | NS 24,338 | Contaminated |
| Treasure-coast-1-PS Glena | Treasure_coast_1_P Glena | # contigs (<= 25000 bp) 416 | # contigs (<= 50000 bp) 219 | Total length (>= 25000 bp) 29,185,305 | Total length (>= 50000 bp) 21,723,641 | Contingent Length 6,769 | Largest contig 691,902 | Total length 43,138,201 | NC 45 | NS 50,547 | Contaminated |
| Treasure-coast-2-PS Glena | Treasure_coast_2_P Glena | # contigs (<= 25000 bp) 497 | # contigs (<= 50000 bp) 237 | Total length (>= 25000 bp) 26,886,155 | Total length (>= 50000 bp) 21,533,597 | Contingent Length 4,002 | Largest contig 681,150 | Total length 43,392,649 | NC 46 | NS 54,295 | Contaminated |

| Colony Image | Assembly Coverage (%) Large < Small Config (%) | Assembly Statistics (Quant 5.0) | SNVs vs Hifi-Fancy Reference | homo %  |
|--------------|-------------------------------------------------|---------------------------------|-----------------------------|--------|
| Pink-Buffalo-3-PS Glena | Pink_Ruffalo_3_P Glena | # contigs (<= 25000 bp) 418 | # contigs (<= 50000 bp) 206 | Total length (>= 200000 bp) 29,432,534 | Total length (>= 500000 bp) 22,024,000 | Contingent Length 1,583 | Largest contig 473,399 | Total length 49,103,984 | NC 45 | NS 46,816 | Contaminated |
| Purple-Mystic-3-PS Glena | Purple_Mystic_3_P Glena | # contigs (<= 25000 bp) 86 | # contigs (<= 500000 bp) 73 | Total length (>= 200000 bp) 28,973,593 | Total length (>= 500000 bp) 24,481,577 | Contingent Length 1,587 | Largest contig 1,798,965 | Total length 33,359,643 | NC 48 | NS 518,241 | Contaminated |
| Barren-3-PS Glena | Barren_3_P Glena | # contigs (<= 250000 bp) 379 | # contigs (<= 500000 bp) 187 | Total length (>= 200000 bp) 26,141,217 | Total length (>= 5000000 bp) 19,544,790 | Contingent Length 8,126 | Largest contig 380,786 | Total length 40,472,653 | NC 47 | NS 23,583 | Contaminated |
| Tampa-1-SW Glena | Tampa_1_SW Glena | # contigs (<= 250000 bp) 360 | # contigs (<= 5000000 bp) 182 | Total length (>= 200000 bp) 25,930,738 | Total length (>= 50000000 bp) 14,944,249 | Contingent Length 8,126 | Largest contig 380,786 | Total length 40,472,653 | NC 47 | NS 23,583 | Contaminated |
| Tampa-2-SW Glena | Tampa_2_SW Glena | # contigs (<= 250000 bp) 265 | # contigs (<= 50000000 bp) 162 | Total length (>= 200000 bp) 23,285,118 | Total length (>= 500000000 bp) 15,047,879 | Contingent Length 8,090 | Largest contig 386,240 | Total length 45,534,020 | NC 47 | NS 29,499 | Contaminated |
| Tampa-3-SW Glena | Tampa_3_SW Glena | # contigs (<= 250000 bp) 389 | # contigs (<= 500000000 bp) 172 | Total length (>= 200000 bp) 21,727,813 | Total length (>= 5000000000 bp) 13,402,298 | Contingent Length 7,874 | Largest contig 381,756 | Total length 45,100,156 | NC 47 | NS 24,338 | Contaminated |
| Treasure-coast-1-PS Glena | Treasure_coast_1_P Glena | # contigs (<= 250000 bp) 416 | # contigs (<= 500000000 bp) 219 | Total length (>= 250000 bp) 29,185,305 | Total length (>= 5000000000 bp) 21,723,641 | Contingent Length 6,769 | Largest contig 691,902 | Total length 43,138,201 | NC 45 | NS 50,547 | Contaminated |
| Treasure-coast-2-PS Glena | Treasure_coast_2_P Glena | # contigs (<= 250000 bp) 497 | # contigs (<= 5000000000 bp) 237 | Total length (>= 250000 bp) 26,886,155 | Total length (>= 50000000000 bp) 21,533,597 | Contingent Length 4,002 | Largest contig 681,150 | Total length 43,392,649 | NC 46 | NS 54,295 | Contaminated |
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 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 5 non-cubensis Psilocybe 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 *et al.*

**Sequence Coverage Analysis**

Uniform sequence coverage and high read mapping rates were observed over most of the of Psilocybe 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 (171Kb, 117Kb, 69Kb, 33Kb). While these smaller contigs have variable coverage in Psilocybe, it is important to understand they represent less than 2% of the genome. Scaffolds 29 and 30 lack telomeric sequence and are 74Kb and 69Kb respectively. Three scaffolds have telomeric sequences on both ends. (Scaffold_2 = 4.3Mb, Scaffold_3 = 4.1Mb, Scaffold_9 = 2.3Mb) suggesting these are tip to tip chromosome assemblies.

The genomes most diverged from the 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.Envy, *P. tampanensis* assembled into 23Kb N50 genome with a 93% BUSCO completeness score and a 99% identical match to Rockefeller *et al.* ITS sequence for *Psilocybe 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.
**Figure 2.** Read mapping efficiency of species specific reads mapped to the P.Envy reference genome. Due to size limitations, only 50 genomes are presented. The remaining read mapping results can be seen in Supplement Table 1.
Figure 3. Coverage plot of 50 genomes mapped to P.Envy reference. Coverage (log) on the Y axis and scaffold 1-32 on the X axis. Scaffold labeled with a T have a single ~200bp 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.azurascens (OR-Coast) has low mapping frequency due to divergence.
Figure 4. Integrated Genome Viewer (IGV) coverage plot over the Psilocybin synthesis pathway described by Reynolds et al. Strains *Huautla* and *Mazatepec* have complete sequence coverage while *P. azurescens* is lacking sequencing coverage over most of the genes.
**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 3 strains with close homology to the reference genome.
To search for pathway redundancy, we utilized tBLASTn to identify alternative Psilocybin producing enzymes in the reference P.Env 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.Env (Figure 6B). The copy on Scaffold_1 is missing the 5’ Exon in P.tampanensis, P.galindoi and P.azurensceens, 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 80Kb contigs in the absence of other pathway related genes (Table 1). This demonstrates a non-clustered Psilocybin synthesis pathway in the non-Psilocybe cubensis strains.

**Figure 6A.** Multiple sequence alignment using CLUSTALW of the most closely related PsiK genes in *P.tampanensis, P.galindoi* and *P.azurensceens*. The R173G mutation is also unique to P.envy when considering non-cubensis species.
Table 1. tBLASTn search of *P. tampanensis*, *P. galindo* and *P. azurensens* 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.

|             | P. azurensens-OR-Coast | P. tampanensis_1 | P. tampanensis_2 | P. tampanensis_3 | P. mexicana_Georgia |
|-------------|-------------------------|------------------|------------------|------------------|---------------------|
| PsiK_Contig | K141_2983               | k141_15608       | k141_5119        | k141_12438       | k141_3669           |
| Contig Size | 4,254                   | 9,322            | 8,836            | 1,196            | 4,016               |
| PsiH_Contig | K141_27478              | k141_15608       | k141_5119        | k141_12438       | k141_3669           |
| Contig Size | 40,681                  | 9,322            | 8,836            | 1,196            | 4,016               |
| PsiD_Contig | K141_27478              | k141_11235       | k141_11612       | k141_12438       | k141_17412          |
| Contig Size | 40,681                  | 44,435           | 44,435           | 1,196            | 3528                |
| PsiM_Contig | K141_27478              | k141_1605        | k141_16729       | k141_1150        | k141_18599          |
| Contig Size | 40,681                  | 81,911           | 82,507           | 1105             | 880                |
Figure 6B. Variation map of PsiK, PsiD, PsiM, PsiH, PsiR - Scaffold_7. Grey is no alignment, Blue is reference allele, Aqua is low impact variant and Red represent High Impact SNPs as according to SNPeff.
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 vs SW vs ITW vs 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 7.** Phylogenetic tree of 81 Psilocybe strains displayed using iTOL (https://itol.embl.de). *Panaeolus Copelandia* samples has an *Psilocybe cubensis* ITS sequence is likely a mislabeled sample.
Conclusions
These data support both Reynolds et al. and Awan et al. in that Psilocybin synthesis appears to have evolved both a conserved ~20Kb 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 infecting Massospora 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. 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 4 vendors (Sporeworks.com, Premiumspores.com, Mushroom.com and InnoculateTheWorld.com). Spore preparations utilized a modified DNA isolation procedure described in McKernan et al. Briefly 1.4ml of spores were centrifuged, decanted and resuspended in 200ul of ddH20. 25 µl 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 200µl of supernatant was aspirated and added to 250µl 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 25µl of ddH20

Library construction for whole genome sequencing.
Fragmentation
Genomic DNA (gDNA) was quantified with a Qubit (Thermo Fisher Scientific) and normalized to reflect 4-8ng/µl in 13µl of TE buffer. Libraries were generated using enzymatic fragmentation with the NEB Ultra II kits (NEB part # E7103). Briefly, 3.5µl of 5X NEB fragmentation buffer and 1µl of Ultra II fragmentation enzyme mix are added to 13µl 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 0.75µl of Agilent SureSelect Adaptor Oligo Mix, 0.5µl of ddH20, 15µl of New England Biolabs (NEB) Ultra II Ligation Master Mix, 0.5µl of NEB Ligation enhancer for a total reaction volume of 16.75µl.

Ligation was performed by the addition of 16.75µl of ligation master mix to the 17.5µl 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
to resuspend and add 16μl (approximately 0.45X) 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 200μl 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 10μl of H2O. Transfer 9μl cleaned DNA to a fresh well.

**PCR amplification**
Add 12.5μl 2x NEBNext Q5 Hot Start Master Mix (New England Biolabs #M0492S) to 9μl ligated DNA, then add 3.5μl NEB 8bp index primer/universal primer. The reaction was run the cycling program using 98°C for 30 seconds as an initial denaturation step. 6 cycles of denaturation, 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 15μl of resuspended AMPure XP beads to the PCR reactions (~ 25μl). 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 clear (about 5 minutes), carefully remove and discard the supernatant. Be careful not to disturb the beads that contain DNA targets. Add 200μl 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 15μl of nuclease free H2O, transfer 15μl 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 x 150bp) were generated, averaging over 12 million read pairs per sample and a total sequence of 207Gb.

**DNA Assembly**
Reads were assembled with MegaHit v.1.2.9 (https://academic.oup.com/bioinformatics/article/31/10/1674/177884) (Li et al. 2015; Li et al. 2016). The Nextflow mapping and assembly pipeline is published on GitHub. Quast 5.0 (http://quast.sourceforge.net)was used to calculate the assembly quality statistics (Gurevich et al.
Sequencing data is deposited in NCBI under Project ID PRJNA700437 and PRJNA687911.

**Read Mapping parameters**

Reads were mapped to the P. Envy reference and to their own assemblies to generate BAM files and coverage statistics using bwa-mem (version 0.7.17-r1188).

```
bwa mem -R "@RG\tID:$id_run\tPU:$id_run\tSM:$id_run\tLB:$id_run\tPL:illumina" -t $cpu -M $fasta_ref ${fq[0]} ${fq[1]} | \  
samtools view -hu - | sambamba sort --tmpdir=. /dev/stdin -o $id_sample.bam
```

A workflow for this is deposited in github. [https://github.com/mclaugsf/mgc-public/blob/master/AOAC_TYM_ERV/bwa-pair.nf](https://github.com/mclaugsf/mgc-public/blob/master/AOAC_TYM_ERV/bwa-pair.nf)

**Variant analysis**

Illumina whole-genome shotgun data (McKernan et al. NCBI Project: PRJNA687911) was mapped to the *P. envy* HiFi reference assembly using bwa-mem (version 0.7.17-r1188), samtools (version 1.8), sorted with sambamba (version 0.6.7) and variants were identified using bcftools (v1.10.2) (Danecek et al. 2021).

```
bcftools mpileup -Ou -f $(Hashemi et al.) ${bam} | \  
bcftools call -m -Ov | \  
bgzip -c > ${sample}-gvcf.vcf.gz
```

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-junelst-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/](https://pcingola.github.io/SnpEff/se_buildingdb/))

**Phylogenetic analysis**

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

**Coverage analysis**

Coverage analysis was performed using BAM files mapped to the 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](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](https://www.rdocumentation.org/packages/gplots versions/3.1.1/topics/heatmap.2)).
Attademo L, Bernardini F. 2021. Are dopamine and serotonin involved in COVID-19 pathophysiology? Eur J Psychiatry 35: 62-63.

Awan. 2018. Convergent evolution of psilocybin biosynthesis by psychedelic mushrooms. BioRXIV doi:DOI: 10.1101/374199.

Bogenschutz MP, Podrebarac SK, Duane JH, Amegadzie SS, Malone TC, Owens LT, Ross S, Mennenga SE. 2018. Clinical Interpretations of Patient Experience in a Trial of Psilocybin-Assisted Psychotherapy for Alcohol Use Disorder. Frontiers in pharmacology 9: 100.

Boyce GR, Gluck-Thaler E, Slot JC, Stajich JE, Davis WJ, James TY, Cooley JR, Panaccione DG, Eilenberg J, De Fine Licht HH et al. 2019. Psychoactive plant- and mushroom-associated alkaloids from two behavior modifying cicada pathogens. Fungal Ecol 41: 147-164.

Carhart-Harris R, Giribaldi B, Watts R, Baker-Jones M, Murphy-Beiner A, Murphy R, Martell J, Blemings A, Erritzoe D, Nutt DJ. 2021. Trial of Psilocybin versus Escitalopram for Depression. The New England journal of medicine 384: 1402-1411.

Creeden JF, Imami AS, Eby HM, Gillman C, Becker KN, Reigle J, Andari E, Pan ZK, O’Donovan SM, McCullumsmith RE et al. 2021. Fluoxetine as an anti-inflammatory therapy in SARS-CoV-2 infection. Biomed Pharmacother 138: 111437.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM et al. 2021. Twelve years of SAMtools and BCFtools. GigaScience 10.

Davis AK, Barrett FS, May DG, Cosimano MP, Sepeda ND, Johnson MW, Finan PH, Griffiths RR. 2021. Effects of Psilocybin-Assisted Therapy on Major Depressive Disorder: A Randomized Clinical Trial. JAMA Psychiatry 78: 481-489.

Dechaumes A, Nekoua MP, Belouzard S, Sane F, Engelmann I, Dubuisson J, Alidjinou EK, Hober D. 2021. Fluoxetine Can Inhibit SARS-CoV-2 In Vitro. Microorganisms 9.

Fricke J, Blei F, Hoffmeister D. 2017. Enzymatic Synthesis of Psilocybin. Angewandte Chemie 56: 12352-12355.

Griffiths RR, Johnson MW, Carducci MA, Umbricht A, Richards WA, Richards BD, Cosimano MP, Klinedinst MA. 2016. Psilocybin produces substantial and sustained decreases in depression and anxiety in patients with life-threatening cancer: A randomized double-blind trial. Journal of psychopharmacology 30: 1181-1197.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29: 1072-1075.

Hashemi H, Haghdooost AA, Haji-Aghajani M, Janbabaee G, Maher A, Noori Hekmat S, Javadi AM, Rahimisadegh R, Emadi S, Rajabalipour MR et al. 2018. A Successful implementation of an idea to a nationally approved plan: Analyzing Iran’s National Health Roadmap using the Kingdon model of policymaking. Medical journal of the Islamic Republic of Iran 32: 46.

Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31: 1674-1676.
Li D, Luo R, Liu CM, Leung CM, Ting HF, Sadakane K, Yamashita H, Lam TW. 2016. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* **102**: 3-11.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094-3100.

Paradis E, Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**: 526-528.

Reynolds HT, Vijayakumar V, Gluck-Thaler E, Korotkin HB, Matheny PB, Slot JC. 2018. Horizontal gene cluster transfer increased hallucinogenic mushroom diversity. *Evol Lett* **2**: 88-101.

Ross S, Bossis A, Guss J, Agin-Liebes G, Malone T, Cohen B, Mennenga SE, Belser A, Kalliontzi K, Babb J et al. 2016. Rapid and sustained symptom reduction following psilocybin treatment for anxiety and depression in patients with life-threatening cancer: a randomized controlled trial. *Journal of psychopharmacology* **30**: 1165-1180.

Schloer S, Brunotte L, Goretzko J, Mecate-Zambrano A, Korthals N, Gerke V, Ludwig S, Rescher U. 2020. Targeting the endolysosomal host-SARS-CoV-2 interface by clinically licensed functional inhibitors of acid sphingomyelinase (FIASMA) including the antidepressant fluoxetine. *Emerg Microbes Infect* **9**: 2245-2255.

Schloer S, Brunotte L, Mecate-Zambrano A, Zheng S, Tang J, Ludwig S, Rescher U. 2021. 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* **178**: 2339-2350.

Tullis. 2021. How ecstasy and psilocybin are shaking up psychiatry. *Nature* **589**, 506-509 (2021).

Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**: 3326-3328.

Zimniak M, Kirshchner L, Hilpert H, Geiger N, Danov O, Oberwinkler H, Steinke M, Sewald K, Seibel J, Bodem J. 2021. The serotonin reuptake inhibitor Fluoxetine inhibits SARS-CoV-2 in human lung tissue. *Scientific reports* **11**: 5890.