SUPPORTING INFORMATION

Fungal identification using Molecular Tools: A Primer for the Natural Products Research Community

Huzefa A. Raja,[†] Andrew N. Miller, [‡] Cedric J. Pearce, [§] Nicholas H. Oberlies[†*]

[†]*Department of Chemistry and Biochemistry, University of North Carolina at Greensboro*

[‡] *Illinois Natural History Survey, University of Illinois, Champaign, Illinois 61820, USA*

[§] *Mycosynthetix, Inc., 505 Meadowland Dr., Suite 103, Hillsborough, NC 27278*

CONTENTS

Table S1. Primers and PCR protocols for ribosomal genes and commonly utilized single-copy protein coding genes in fungal identification

Locus	Lineage	Primer	Primer Sequence 5'-3'	Direction	PCR protocol*	Reference
Translation elongation factor 1-alpha (*tef1*) [~1000 bp]	All Fungi	EF1-983F EF1-2218R One of the most broadly used primers. This fragment extends nearly to the 3' end of *tef1*	GCY CCY GGH CAY CGT GAY TTY AT AT GAC ACC RAC RGC RAC RGT YTG	Forward Reverse	1. 94˚C – 2 min 2. 66˚C – 56C touchdown (9 cycles) 3. 94˚C – 30 sec 4. 56˚C – 1 min 5. 72˚C – 1 min 6. Repeat 3–5 for 36 cycles 7. 72˚C – 10 min 8. 4˚C on hold	[1]
Translation elongation factor 1-alpha (*tef1*) [~600 bp]	All Fungi A proposed secondary barcode marker for fungi	EF1-1018F EF1-1620R Note: EF1-1018F and EF1-1620R are internal primers to EF1-983F and EF1-2218R	GAY TTC ATC AAG AAC ATG AT GAC GTT GAA DCC RAC RTT GTC	Forward Reverse	1. 94˚C – 2 min 2. 66˚C – 56C touchdown (9 cycles) 3. 94˚C – 30 sec 4. 56˚C – 1 min 5. 72˚C – 1 min 6. Repeat 3–5 for 36 cycles 7. 72˚C – 10 min 8. 4˚C on hold	[2]
RNA polymerase II subunit 1 (*RPB1*) [~1200 bp]	All Fungi	RPB1af RPB1cr	GAR TGY CCD GGD CAY TTY GG CCN GCD ATN TCR TTR TCC ATR TA	Forward Reverse	1. 96˚C – 5 min 2. 94˚C – 30 sec 3. 52˚C – 30 sec 4. 72˚C – 1 min 5. Repeat 2–4 for 40 cycles 6. 72˚C – 8 min 7. 4˚C on hold	[3,4]
RNA polymerase II subunit 2 (*RPB2*) [~1200 bp]	All Fungi	RPB2-5f RPB2-7cR	GAY GAY MGW GAT CAY TTY GG CCC ATR GCT TGY TTR CCC AT	Forward Reverse	1. 94˚C – 3 min 2. 94˚C – 20 sec 3. 55˚C – 30 sec 4. 72˚C – 1 min 5. Repeat 2–4 for 40 cycles 6. 72˚C – 10 min 7. 4˚C on hold	[5]
beta-tubulin (*tub2/BenA*) [~500 bp]	All Fungi Barcoding marker for *Penicillium*	Bt2a Bt2b	GGT AAC CAA ATC GGT GCT GCT TTC ACC CTC AGT GTA GTG ACC CTT GGC	Forward Reverse	1. 95˚C – 5 min 2. 94˚C – 45 sec 3. 55˚C – 45 sec 4. 72˚C – 1 min 5. Repeat 2–4 for 35 cycles 6. 72˚C – 7 min 7. 4˚C on hold	[6]

Locus	Lineage	Primer	Primer Sequence 5'-3'	Direction	PCR protocol	Reference
Calmodulin (*CaM*) [~580 bp]	All Fungi Barcoding marker for *Aspergillus*	CMD5 CMD6	CCG AGT ACA AGG AGG CCT TC CCG ATA GAG GTC ATA ACG TGG	Forward Reverse	1. 94˚C – 1 min 2. 94˚C – 1 min 3. 55˚C – 30 sec 4. 72˚C – 1.5 min (90 sec) 5. Repeat 2–4 for 42 cycles or Standard Repeat 2–4 for 35 cycles). 6. 72˚C – 10 min 7. 4˚C on hold	[7]
Mini chromosome maintenance protein (*MCM7*) [~650 bp]	All Fungi	Mcm7-709for Mcm7-1348rev	ACI MGI GTI TCV GAY GTH AAR CC GAY TTD GCI ACI CCI GGR TCW CCC AT	Forward Reverse	1. 94˚C – 5 min 2. 94˚C – 45 sec 3. 56˚C – 50 sec 4. 72˚C – 1 min 5. Repeat 2–4 for 38 cycles 6. 72˚C – 5 min 7. 4˚C on hold	[8,9]
Translation elongation factor 1-alpha (*tef1*) [~1300 bp]	*Trichoderma* Barcoding marker	EF1-728F TEF1LLErev This fragment includes the fourth and fifth introns and a significant portion of the last large exon[10,11]	CAT CGA GAA GTT CGA GAA GG AAC TTG CAG GCA ATG TGG	Forward Reverse	1. 94˚C – 2 min 2. 66˚C – 56˚C touchdown (9 cycles) 3. 94˚C – 30 sec 4. 56˚C – 1 min 5. 72˚C – 1 min 6. Repeat 3–6 for 36 cycles 7. 72˚C – 10 min 8. 4˚C on hold	[12,13]
Translation elongation factor 1-alpha (*tef1*) [~700 bp]	*Fusarium* ID For *Fusarium* identification Intron rich portion of *tef1*	ef1 ef2	ATG GGT AAG GA(A/G) GAC AAG AC GGA (G/A)GT ACC AGT (G/C)AT CAT GTT	Forward Reverse	1. 95˚C – 8 min 2. 95˚C – 30 sec 3. 53˚C – 60 sec 4. 72˚C – 1 min 5. Repeat 2–4 for 35 cycles 6. 72˚C – 5 min 7. 4˚C on hold	[14,15]

Locus	Lineage	Primer	Primer Sequence 5'-3'	Direction	PCR protocol	Reference
*Partial actin [~370 bp]	Sordariomycetes, and Dothideomycetes including Plant pathogenic fungi *Colletotrichum*, *Mycosphaerella* and anamorphs, *Phoma*-like, *Phyllosticta*, *Verticillium*	act-512F act-783R	ATG TGC AAG GCC GGT TTC GC TAC GAG TCC TTC TGG CCC AT	Forward Reverse	1. 95˚C – 8 min 2. 94˚C – 15 sec 3. 55˚C – 20 sec 4. 72˚C – 1 min 5. Repeat 2–4 for 35 cycles 6. 72˚C – 5 min 7. 4˚C on hold	[12]
Internal Transcribed Spacer (ITS) region of the rRNA [~450-800 bp]	Official fungal barcode	ITS1F ITS4	CTT GGT CAT TTA GAG GAA GTA A TCC TCC GCT TAT TGA TAT GC	Forward Reverse	1. 95˚C – 5 min 2. 94˚C – 30 s 3. 52˚C – 30 s 4. 72˚C – 1 min 5. Repeat 2–4 for 35 cycles 6. 72˚C – 8 min 7. 4˚C on hold	[16]
Small Subunit (SSU, 18S) of the rRNA [~1200 bp]	All Fungi	NS1 NS4	GTA GTC ATA TGC TTG TCT C CTT CCG TCA ATT CCT TTA AG	Forward Reverse	Same protocol as ITS region	[16]
Large Subunit (LSU, 28S) of the rRNA [~1200 bp]	All Fungi	LROR LR6	ACC CGC TGA ACT TAA GC CGC CAG TTC TGC TTA CC	Forward Reverse	Same protocol as ITS region	[17,18]

*For protein-coding genes, we have found that using 5µl of DNA works better than lower amounts. Newly extracted DNA usually performs better than DNA stored for several years in -80C for PCR of protein-coding genes.

For additional information on other fungal primers, see:
http://lutzonilab.org/primer-sequences/
http://www.fungalbarcoding.org/DefaultInfo.aspx?Page=Primers
http://www2.clarku.edu/faculty/dhibbett/Protocols_Folder/Primers/Primers.pdf

Table S2. Do's and Don'ts for molecular identification of Fungi

Do's	Don'ts and its consequences
Do insure culture purity. It is important to insure that the culture, which is going to be used for DNA extraction, is pure and/or contamination free.	Don't use an impure/contaminated culture for DNA work. An impure culture will result in mixed signals during Sanger sequencing and result in poor quality chromatograms and thus preclude proper base calls and will ultimately result in poor quality sequence data, which will make identification via BLAST search and DNA taxonomy inaccurate.
Do ensure DNA extraction, PCR purification kits and molecular biology grade water are free of impurities.	Don't use contaminated solutions/buffers. Any problems with molecular biology kits will most likely cause problems with downstream application such as PCR and Sanger sequencing. Consider making smaller aliquots of water in 1.5mL centrifuge tubes and discarding after each use.
Nucleotide BLAST search: Use published sequences from BLAST search results. Ensure that type-sequences from ex-type cultures are used. Include sequences from type-material.	Don't use sequences that are unpublished without checking for the presence of specimen vouchers and other collection information connected to the unpublished sequences prior to using them.
Use a query coverage of ≥ 80% and ≥ 97–100% sequence similarity as a probable match for identification based on BLAST search; higher is better.	Don't simply use the first result from a BLAST search as your species identification. Don't use ≤ 80% query coverage and ≤ 95% sequence similarity as a species match.
Download the top BLAST matches and align your query sequence and the top BLAST matches to look for sequence homology and again, include authenticate/reference published sequences preferably from type-material.	Don't solely rely on BLAST search results. Use DNA taxonomy and tree-building methods when possible (see section 6).
Do include all citations and accession numbers, including reference strain numbers of fungi, which are identical to your unknown fungus. Provide BLAST search results as a summary of the top BLAST matches.	Don't simply give your fungus a name based on the first line of the BLAST search results. Most likely identifying your fungus based on BLAST search in INSD alone will lead to incorrect identification.
Do sequence singe-copy protein-coding genes when possible.	Don't simply rely on ITS region for identification, especially in lineages such as Hypocreales (e.g. *Fusarium, Trichoderma*) and Eurotiales (e.g. *Aspergillus* and *Penicillium*) as well as genera that are very diverse and poorly represented in GenBank. If you do use ITS, make sure you compare with type sequences/reference strains in curated databases. Don't use 18S rDNA region for species-level identification (see section 2).
Do sequence the entire ITS region (ITS1-5.8S-ITS2).	Don't use only the ITS1 or ITS2 region. Although these have been utilized for identification in metabarcoding studies,[19,20] it might result in inconclusive species-identification.
Do utilize curated sequence databases for BLAST search when possible (See Table1).	Don't simply rely on INSD: GenBank, EMBL, DDJB.
Do submit your newly acquired sequence to GenBank and include the accession number in the manuscript. When submitting a sequence to GenBank, include as much annotation as possible, especially specimen voucher or strain details. After the manuscript is published, inform GenBank of release of sequence and provide details of	Don't forget to contact GenBank after the paper is published. If not, your sequence will remain as unpublished and may not be included in future BLAST search analyses or molecular phylogenies by mycologists and/or natural products chemists.

pagination.	
Do include macro and micro-morphological details of fungal culture when available for the identification of your fungus.	Don't simply use molecular sequence data alone if morphological characters are abundantly available.
Do include citations of papers from which sequences are utilized in your own study.	Not doing so will fail to provide due credit to the fungal taxonomist and systematists who have done all the hard work that enabled you to identify your unknown fungus.
Do realize that fungi are hyperdiverse and not all fungal sequences are deposited in INSD; it is likely that you have an undescribed/unknown or previously not sequenced fungus.	Don't assume that all fungal sequences are available and 100% accurately identified in INSD.
If you do have a new fungus, describe it in a mycology journal. General information on how to describe a new fungus is provided elsewhere.[21] Deposit ITS sequence data from the new fungus in GenBank.	Don't describe a new fungus in a supplementary information section of a chemistry journal. If possible, don't simply describe a new fungus based solely on morphology.
Do collaborate with trained mycologists when possible. Data generated by natural products chemists are useful for fungal taxonomy and vice versa.	Don't work in isolation! Collaborations will foster learning about the fungal kingdom and knowledge about its secondary metabolite chemistry.

REFERENCES

(1) Rehner, S. A.; Buckley, E. *Mycologia* **2005,** *97*, 84-98.
(2) Stielow, J.; Lévesque, C.; Seifert, K.; Meyer, W.; Irinyi, L.; Smits, D.; Renfurm, R.; Verkley, G.; Groenewald, M.; Chaduli, D. *Persoonia* **2015,** *35*, 242-263.
(3) Stiller, J. W.; Hall, B. D. *Proc. Natl. Acad. Sci. U.S.A.* **1997,** *94*, 4520-4525.
(4) Matheny, P. B.; Liu, Y. J.; Ammirati, J. F.; Hall, B. D. *Am. J. Bot.* **2002,** *89*, 688-698.
(5) Liu, Y. J.; Whelen, S.; Hall, B. D. *Mol. Biol. Evol.* **1999,** *16*, 1799-1808.
(6) Glass, N. L.; Donaldson, G. C. *Appl. Environ. Microbiol.* **1995,** *61*, 1323-1330.
(7) Hong, S.-B.; Go, S.-J.; Shin, H.-D.; Frisvad, J. C.; Samson, R. A. *Mycologia* **2005,** *97*, 1316-1329.
(8) Schmitt, I.; Crespo, A.; Divakar, P. K.; Fankhauser, J. D.; Herman-Sackett, E.; Kalb, K.; Nelsen, M. P.; Nelson, N. A.; Rivas-Plata, E.; Shimp, A. D.; Widhelm, T.; Lumbsch, H. T. *Persoonia* **2009,** *23*, 35-40.
(9) Raja, H.; Schoch, C. L.; Hustad, V.; Shearer, C.; Miller, A. *MycoKeys* **2011,** *1*, 63-94.
(10) Druzhinina, I.; Kubicek, C. P. *J. Zhejiang Univ. Sci. B* **2005,** *6*, 100.
(11) Druzhinina, I. S.; Kopchinskiy, A. G.; Komoń, M.; Bissett, J.; Szakacs, G.; Kubicek, C. P. *Fungal Genet. Biol.* **2005,** *42*, 813-828.
(12) Carbone, I.; Kohn, L. M. *Mycologia* **1999**, 553-556.
(13) Jaklitsch, W. M.; Komon, M.; Kubicek, C. P.; Druzhinina, I. S. *Mycologia* **2005,** *97*, 1365-1378.
(14) Geiser, D.; Jimenez-Gasco, M.; Kang, S.; Makalowska, I.; Veeraraghavan, N.; Ward, T.; Zhang, N.; Kuldau, G. *Eur. J. Plant Pathol.* **2004,** *110*, 473-479.
(15) Maciá-Vicente, J. G.; Jansson, H.-B.; Abdullah, S. K.; Descals, E.; Salinas, J.; Lopez-Llorca, L. V. *FEMS Microbiol. Ecol.* **2008,** *64*, 90-105.
(16) White, T. J.; Bruns, T.; Lee, S. H.; Taylor, J. W., *PCR protocols: a guide to methods and application*. San Diego, 1990; p 315-322.
(17) Vilgalys, R.; Hester, M. *J. Bacteriol.* **1990,** *172*, 4238-4246.
(18) Rehner, S. A.; Samuels, G. J. *Can. J. Bot.* **1995,** *73*, S816-S823.
(19) Bazzicalupo, A. L.; Bálint, M.; Schmitt, I. *Fungal Ecol.* **2013,** *6*, 102-109.
(20) Blaalid, R.; Kumar, S.; Nilsson, R. H.; Abarenkov, K.; Kirk, P. M.; Kauserud, H. *Mol. Ecol. Resour.* **2013,** *13*, 218-24.
(21) Seifert, K. A.; Rossman, A. Y. *IMA Fungus* **2010,** *1*, 109-116.