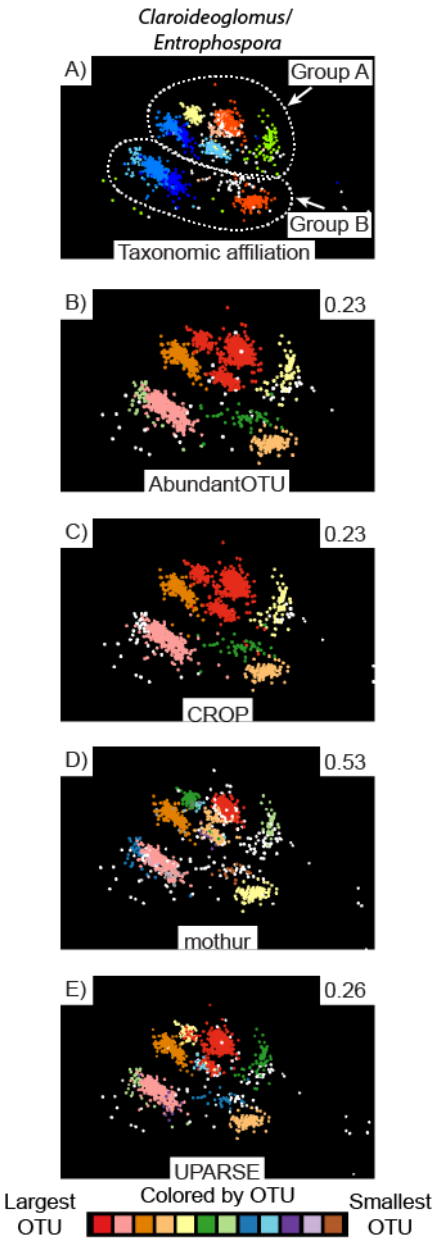

Figure S1: Multidimensional scaling (MDS) visualizations of OTU clusters for isolates in the *Claroideoglomus/Entrophospora* group; sequences (points) in the top row are colored by their species or geographic isolate affiliation (the same colors are used as in Fig. 2) and sequences from each sequence group (see Fig. 2) are circled and labeled. The remaining rows (top to bottom) show the same sequences colored by OTU for each of the four clustering methods that used the common sequence processing pipeline: AbundantOTU (second row), CROP (third row), mothur (fourth row), and UPARSE (fifth row). Only OTUs containing more than 10 sequences are shown, and OTUs are colored according to the number of sequences they contain. The adjusted Rand index value in the upper right-hand corner of each panel for the clustering methods quantifies the fit of the OTU delineations compared to the known species attribution. Larger adjusted Rand index values indicate closer correspondence between each OTU and the

sequences originating from each AM fungal species: mothur performed relatively better for this group, but all four clustering methods gave OTUs that poorly matched species as expected given the gene tree discordance that occurs in these species. Interactive 3-D versions of these MDS visualizations are available online at: https://spidal-gw.dsc.soic.indiana.edu/public/groupdashboard/AM%20fungal%20clustering%20AEM

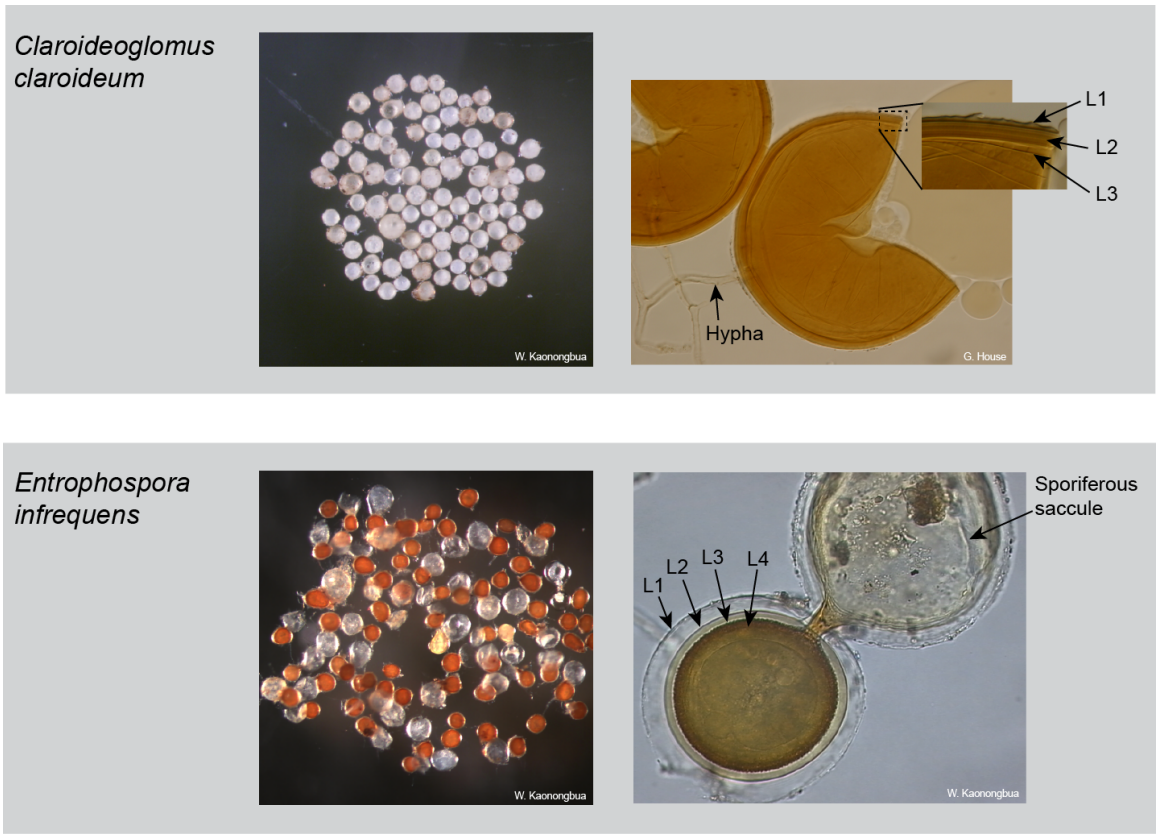

Figure S2: Comparison of spore morphology between *Claroideoglomus claroideum* (top) and *Entrophospora infrequens* (bottom). Groups of spores for each species are shown on the left at the same magnification, and the structure of an individual spore is shown on the right (Note: pictures of spore structure are not at the same magnification). Intact spores show pronounced color differences between the two species, and spores of both species develop differently. In *E. infrequens*, spores develop within the neck of a sporiferous saccule that is identifiable as the hyaline structure attached to the pigmented spore, while spores of *C. claroideum* form from a swelling and thickening of a hyphal tip. The wall structure of mature spores is also different between the two species. Spore walls of *C. claroideum* and *E. infrequens* comprise 3 and 4 main layers, respectively (labeled L1-3 and L1-4 from outside to inside). For *C. claroideum* spores, the L2 layer is laminated unlike in *E. infrequens*. For *E. infrequens* spores, the L3 layer is ornamented with distinctive pentagonal spines and the L4 layer is thick and hyaline. Based on these pronounced differences in morphology and development, these fungi had been placed into different families, however rRNA sequences from each species are largely indistinguishable.

Supplemental Methods: The clustering commands used to create OTUs for each sequence clustering method. For all clustering methods, the clustering output for 97% sequence similarity (or for CROP, settings intended to approximate 97% sequence similarity) was then converted into a common-format OTU table using Python scripts specific to the output format of each clustering method. Only the test sequences (the 454 reads), not the reference sequences, were used in all further analysis of rarefied OTU richness and OTU diversity.

Clustering methods using sequences from the common sequence quality screening pipeline

Clustering methods using the full sequence dataset (51,543 sequences):

AbundantOTU:

```
$ AbundantOTU+ -i FullDatasetFor_AbundantOTU_mothur.fasta -o
FullDatasetFor_AbundantOTU_mothur_AbundantOTU97PercentClustered
```

mothur (command line mode):

```
$ mothur "#align.seqs(candidate=FullDatasetFor_AbundantOTU_mothur.fasta,
template=RefSeqs_AlignedTrimmed_MatchClustering454Dataset.fasta)"

$ mothur "#filter.seqs(fasta=FullDatasetFor_AbundantOTU_mothur.align, vertical=T,
trump=.)"

$ mothur "#unique.seqs(fasta=FullDatasetFor_AbundantOTU_mothur.filter.fasta)"

$ mothur "#dist.seqs(fasta=FullDatasetFor_AbundantOTU_mothur.filter.unique.fasta,
cutoff=0.15, processors=2)"

$ mothur "#cluster(column=FullDatasetFor_AbundantOTU_mothur.filter.unique.dist, name=
FullDatasetFor_AbundantOTU_mothur.filter.names, method=average)"
```

Clustering methods using the unique (de-replicated) sequence dataset (10,081 sequences):

CROP (flag '-s' is meant to approximate 97% sequence similarity clustering; the values for flags '-b' and '-z' were chosen based on the number of sequences and their length (350bp) following the suggestions in the CROP v 1.33 Quick Guide):

```
$ CROP -i DereplicatedDatasetFor_CROP_UPARSE.fasta -o
DereplicatedDatasetFor_CROP_UPARSE_CROPClustered -s -b 200 -z 350
```

UPARSE:

```
$ usearch -cluster_otus DereplicatedDatasetFor_CROP_UPARSE.fasta -otus
DereplicatedDatasetFor_CROP_UPARSE_UPARSEOTUCenterSeqs.fasta -sizein -sizeout -relabel
OTU_

$ usearch -usearch_global DereplicatedDatasetFor_CROP_UPARSE.fasta -db
DereplicatedDatasetFor_CROP_UPARSE_UPARSEOTUCenterSeqs.fasta -strand plus -id 0.97 -uc
DereplicatedDatasetFor_CROP_UPARSE_UPARSEClustered.uc
```

Clustering method using independent sequence quality screening pipeline (51,135 sequences):

CD-HIT-OTU:

```
$ cd-hit-otu-454-0.0.2/cd-hit-otu-all.pl -i AllTestAndReferenceSequences.fasta -o CD-
HIT-OTUClustered -p 21 -c 0.97
```

Table S1: Sequences that did not group with other sequences from the expected species in the MDS visualization and instead grouped closely with sequences from a different species. These sequences were removed from the dataset before clustering with AbundantOTU, CROP, mothur, and UPARSE, but were included in the dataset used by the independent sequence quality screening and clustering pipeline of CD-HIT-OTU.

Test (T) or Reference (R) sequence	Sequence barcode	Culture ID	Cultured species name	Number of sequences removed
T	1	INVAM IN212; IU-57	*Racocetra fulgida* (Indiana, USA Site 1)	8
	7	WK-ULBCB Soil sample	*Pacispora scintillans* (Indiana, USA)	46
	9	IU-77	*Funneliformis mosseae* (Illinois, USA)	4
	12	INVAM AU219	*Archaeospora trappei* (Australia)	4
	13	INVAM WV579A	*Claroideoglomus etunicatum* (West Virginia, USA)	3
	18	INVAM IN215; IU-110	*Entrophospora infrequens* (Indiana, USA)	87
	20	Isolate of AZU133	*Entrophospora infrequens* (Arizona, USA)	2
	22	INVAM CA203	*Entrophospora infrequens* (California, USA)	3
	23	INVAM NC258	*Paraglomus occultum* (North Carolina, USA)	2
	24	INVAM MA453B	*Dentiscutata erythropus* (Massachusetts, USA)	1
	25	INVAM IL203A	*Dentiscutata heterogama* (Illinois, USA)	1
	31	IU-112	*Claroideoglomus claroideum* (Indiana, USA Site 2)	1
	33	IU-112	*Claroideoglomus claroideum* (Indiana, USA Site 2)	1
	34	IU-25	*Racocetra fulgida* (Indiana, USA Site 2)	1
	46	INVAM IN215; IU-110	*Entrophospora infrequens* (Indiana, USA)	1
	54	INVAM SA101 or SA112	*Claroideoglomus luteum* (Saskatchewan, Candada)	1
	56	IU-112	*Claroideoglomus*	1

			claroideum (Indiana, USA Site 2)	
	58	IU-112	*Claroideoglomus claroideum* (Indiana, USA Site 2)	3
	60	IU-112	*Claroideoglomus claroideum* (Indiana, USA Site 2)	1
R	N/A	N/A	*Claroideoglomus claroideum*	7
	N/A	N/A	*Claroideoglomus etunicatum*	2
	N/A	N/A	*Claroideoglomus lamellosum*	3
	N/A	N/A	*Entrophospora infrequens*	42
	N/A	N/A	*Funneliformis coronatus*	7
	N/A	N/A	*Funneliformis geosporus*	2
	N/A	N/A	*Funneliformis mosseae*	2
	N/A	N/A	*Rhizoglomus clarum*	2
	N/A	N/A	*Rhizoglomus irregulare*	2
	N/A	N/A	*Septoglomus constrictum*	13