

- ...we are witnessing exponential growth of scientific data of which only a fraction is being published.
- Scientific community is facing necessity to effectively store, manage and access interdisciplinary machine readable data.
- There are numerous initiatives from local to global scales, but effective interdisciplinary management of scientific (big) data remains a challenge.

UNITE: Curated and Evolving Database for Molecular Identification and for Communicating Fungal Species



UNITE


unite.ut.ee

Sergei Põlme

Run Analysis Search Pages Resources Statistics Notes and News Workbench

unite
community

Communication and identification of DNA based fungal species

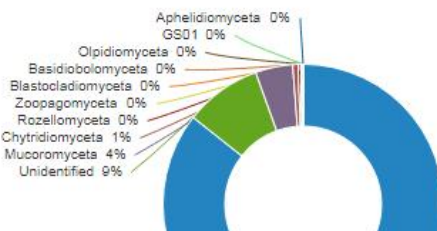


Current version: 8.2; Last updated: 2020-01-15 ([read more](#))
Number of ITS sequences (UNITE+INSD): 2 480 043; Number of UNITE fungal Species Hypotheses with DOIs at 1.5% threshold: 102 100 ([more statistics](#))

Threshold: 1.5% Include: All SH-s Start typing taxon name here ... Go Reset

- ▶ Dikarya (87,449)
- ▶ Unidentified (9,044)
- ▶ Mucoromyceta (4,388)
- ▶ Chytridiomyceta (624)
- ▶ Rozellomyceta (316)
- ▶ Zoopagomyceta (157)
- ▶ Blastocladiomyceta (50)
- ▶ Basidiobolomyceta (33)
- ▶ Olpidiomyceata (23)
- ▶ GS01 (*informal name, see Tedersoo et al., 2017*) (10)
- ▶ Aphelidiomyceta (9)

SH graph: Fungi



Taxon	Percentage
Aphelidiomyceta	0%
GS01	0%
Olpidiomyceata	0%
Basidiobolomyceta	0%
Blastocladiomyceta	0%
Zoopagomyceta	0%
Rozellomyceta	0%
Chytridiomyceta	1%
Mucoromyceta	4%
Unidentified	9%

UNITE provides a unified way for delimiting, identifying, communicating, and working with DNA-based **Species Hypotheses** (SH).

All fungal ITS sequences in the International Nucleotide Sequence Databases (INSD: GenBank, ENA, DDBJ) are clustered to approximately the species level by applying a set of dynamic distance values (<0.5 - 3.0%). All species hypotheses are given a unique, stable name in the form of a DOI, and their taxonomic and ecological annotations are verified through distributed, web-based third-party annotation efforts. SHs are connected to a taxon name and its classification as far as possible (phylum, class, order, etc.) by taking into account identifications for all sequences in the SH.

...the system and the data are updated automatically as the number of public fungal ITS sequences grows.

Agaricus bisporus Species Hypothesis DOI page



Communication and identification of DNA based fungal species

TH021968: **Agaricus bisporus** (J.E. Lange) Pilat | SH1593868.08FU

Distance to the closest SH: 1.5
No. of sequences in SH: 169

Placement in the fungal classification
Fungi; Dikarya; Basidiomycota; Agaricomycotina; Agaricomycetes;
Agaricomycetidae; Agaricales; Agaricaceae; Agaricus
Index Fungorum: urn:lsid:indexfungorum.org:names:292246

Reference sequence: UDB011831
Chosen by: Irja Saar
Date: 2014-11-09 22:04

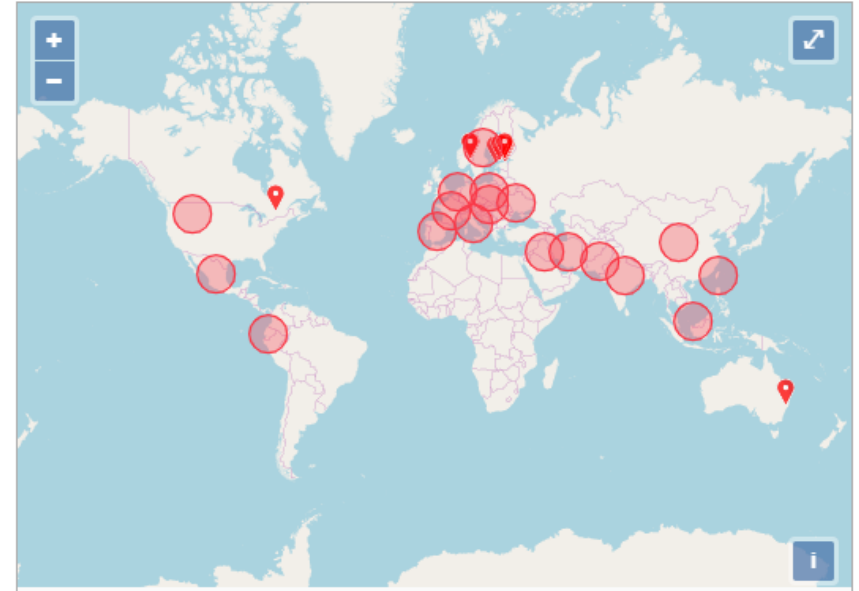
Statistics Taxonomy Ecology

Interacting taxa
Pinus halepensis (1); Epacris longiflora (1); Homo sapiens (1);
EcM lineage
/non-ectomycorrhizal (100);

Older version(s) of this SH is/are available

SH code (Count*/Total count**)
[SH174653.07FU](#) (125/169);
*Number of sequences carried over from previous version
**Total number of sequences composing this SH in current version

Distribution map



*Locations without exact coordinates are displayed as spherical country centroids

Agaricus bisporus SH DOI page

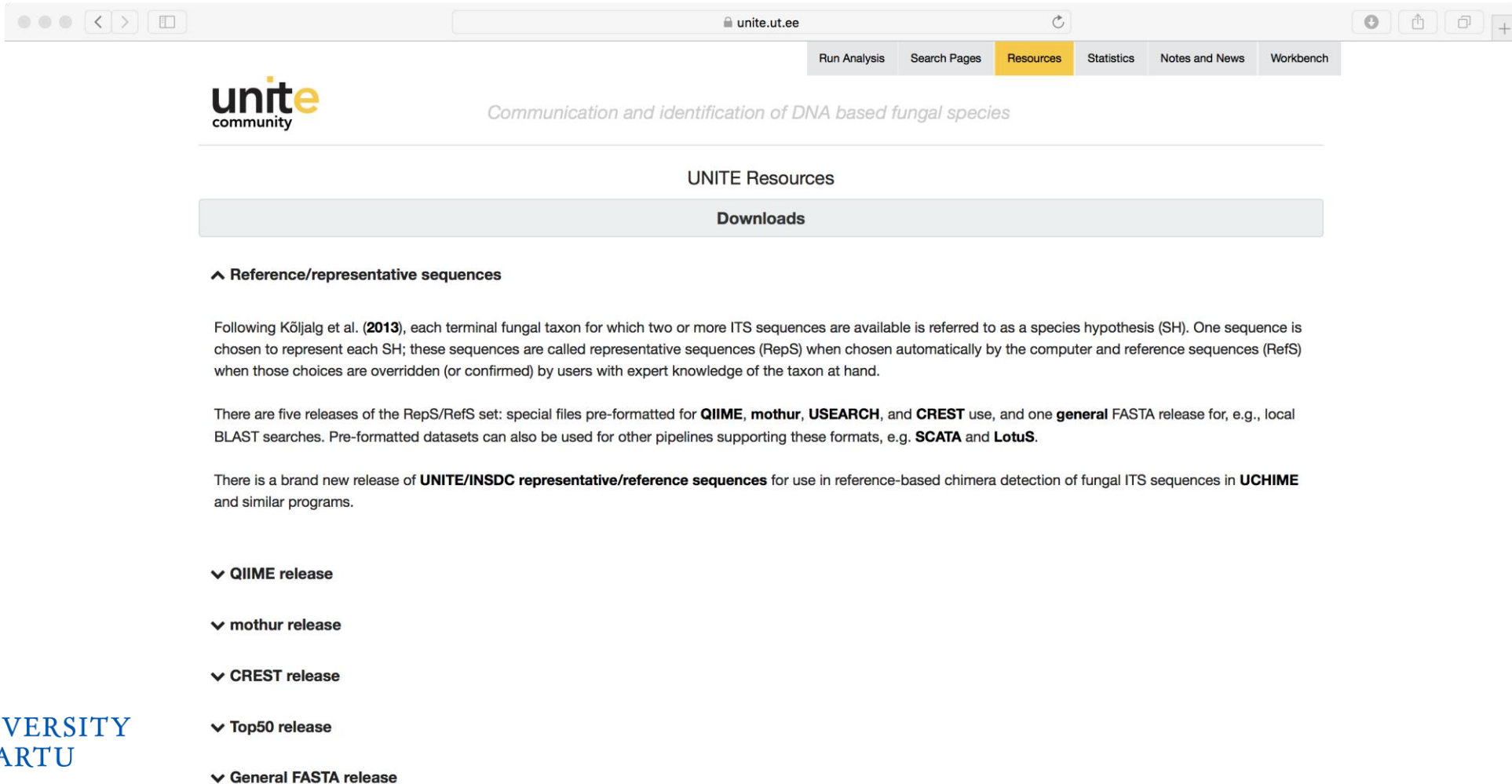


name=SH2396946.08FU#fndtn-panel1

*Locations without exact coordinates are displayed as spherical country centroids

Accession number	UNITE taxon name	INSD taxon name	Sequence source	Interacting taxa	Sampling area	%	← Alignment →
GU327642(11)	Agaricus	Agaricus (Agaricus bisporus var. bi...					-----TTGA
AY484694	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...					-----GAAGGATCATTATTGA
AY484692	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...					-----AGGATCATTATTGA
AF432885(1)	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...					-----GGAAGGATCATTATTGA
AF465402	Agaricus	Agaricus (Agaricus bisporus var. bi...					-----
AF161014	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...					-----GTGAACCTGCGGAAGGATCATTATTGA
AM930981	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...					-----GGAAGGATCATTATTGA
AJ884646(2)	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...			Ukraine		-----CATTATTGA
AJ133384	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...			Taiwan, Province of		-----
AJ133383(2)	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...			China		-----
AJ133382	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...			China		-----
AJ409229(9)	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...			France		-----AGGATCATTATTGA
AJ131132(2)	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...					-----
HM149322	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...					-----GCAACTTA
HM149318	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...					-----G-----GCAT
HM149317	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...					-----GGGCATC
HM149314	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...					-----GGCTAATTGA
HM149313	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...					-----GGGCACTTT
HM149312	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...					-----TTATTGA
HM561977	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...			Malaysia		-----TTGA
JN222415	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...					TCCTGTAGGTGAACCTGCGGAAGGATCATTATTGA
JN222412	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...					-----TTGA
JN222411	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...					-----TTGA
JN222409	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...					-----TTGA
JN222408	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...					-----TTGA
UDB011831	Agaricus bisporus		TU106622 (Specimen) ↗		Estonia		-----TGGTGACTGCGGAGGATCATTATTGA
JQ982502	Agaricus bisporus	Agaricus bisporus (Agaricus bisporu...			China		-----TTT

UNITE resources, SH datasets are downloadable
UNITE serves as a data provider for a range of metabarcoding software pipelines and regularly exchanges data with all major fungal sequence databases and other community resources.

A screenshot of the UNITE website's "Resources" page. The browser address bar shows "unite.ut.ee". The navigation menu includes "Run Analysis", "Search Pages", "Resources" (highlighted), "Statistics", "Notes and News", and "Workbench". The page title is "UNITE Resources" and the subtitle is "Downloads". The main content area is titled "Reference/representative sequences" and contains three paragraphs of text. The first paragraph explains the concept of species hypothesis (SH) and representative sequences (RepS) versus reference sequences (RefS). The second paragraph lists five releases of the RepS/RefS set: special files for QIIME, mothur, USEARCH, and CREST, and a general FASTA release. The third paragraph mentions a brand new release of UNITE/INSDC representative/reference sequences for use in reference-based chimera detection. Below the text are five expandable sections: "QIIME release", "mothur release", "CREST release", "Top50 release", and "General FASTA release".

unite
community

Communication and identification of DNA based fungal species

UNITE Resources

Downloads

Reference/representative sequences

Following Kõljalg et al. (2013), each terminal fungal taxon for which two or more ITS sequences are available is referred to as a species hypothesis (SH). One sequence is chosen to represent each SH; these sequences are called representative sequences (RepS) when chosen automatically by the computer and reference sequences (RefS) when those choices are overridden (or confirmed) by users with expert knowledge of the taxon at hand.

There are five releases of the RepS/RefS set: special files pre-formatted for **QIIME**, **mothur**, **USEARCH**, and **CREST** use, and one **general** FASTA release for, e.g., local BLAST searches. Pre-formatted datasets can also be used for other pipelines supporting these formats, e.g. **SCATA** and **LotuS**.

There is a brand new release of **UNITE/INSDC representative/reference sequences** for use in reference-based chimera detection of fungal ITS sequences in **UCHIME** and similar programs.

- QIIME release
- mothur release
- CREST release
- Top50 release
- General FASTA release

UNITE Milestones: 2018 – UNITE SH DOIs are implemented in GBIF taxonomic backbone



NEWS | 29 AUGUST 2018

Adding sequence-based identifiers to backbone taxonomy reveals 'dark taxa' fungi

Pilot project with northern European researchers enables inclusion of non-Linnaean 'species hypotheses' aimed at advancing scientific understanding of mycology and functional biodiversity



Hygrocybe conica, observed in Trondheim, Norway by Ole Reitan, via Norwegian Species Observation Service. Photo licensed under CC BY 4.0.





UNITE Milestones

2020 – New UNITE version 8.2 released in January.

New versions cover all Eucaryota (also ITS, not curated yet, absence of representative sequences etc).

Incorporates Sequel (PacBio) full ITS sequences from various eDNA samples, still under development.



UNITE future developments

2020-2022

new UNITE web services and resources available through European Open Science Cloud (H2020 project EOSC-Nordic)

Services: defining alien, invasive, threatened, undescribed, etc. species in DNA based communities

Collaboration with GBIF, iBOL and SILVA communities to develop common space for the identification and communication of DNA based taxon occurrences.



Thank you!

UNITE community is using **flexible PlutoF** platform for their data management and publishing of DNA based species with DOIs.

PlutoF