DSMZ
Deutsche
Sammlung von
Mikroorganismen
und Zellkulturen GmbH

Defining biologically meaningful molecular operational taxonomic units

M. Göker

Why molecular taxonomy?

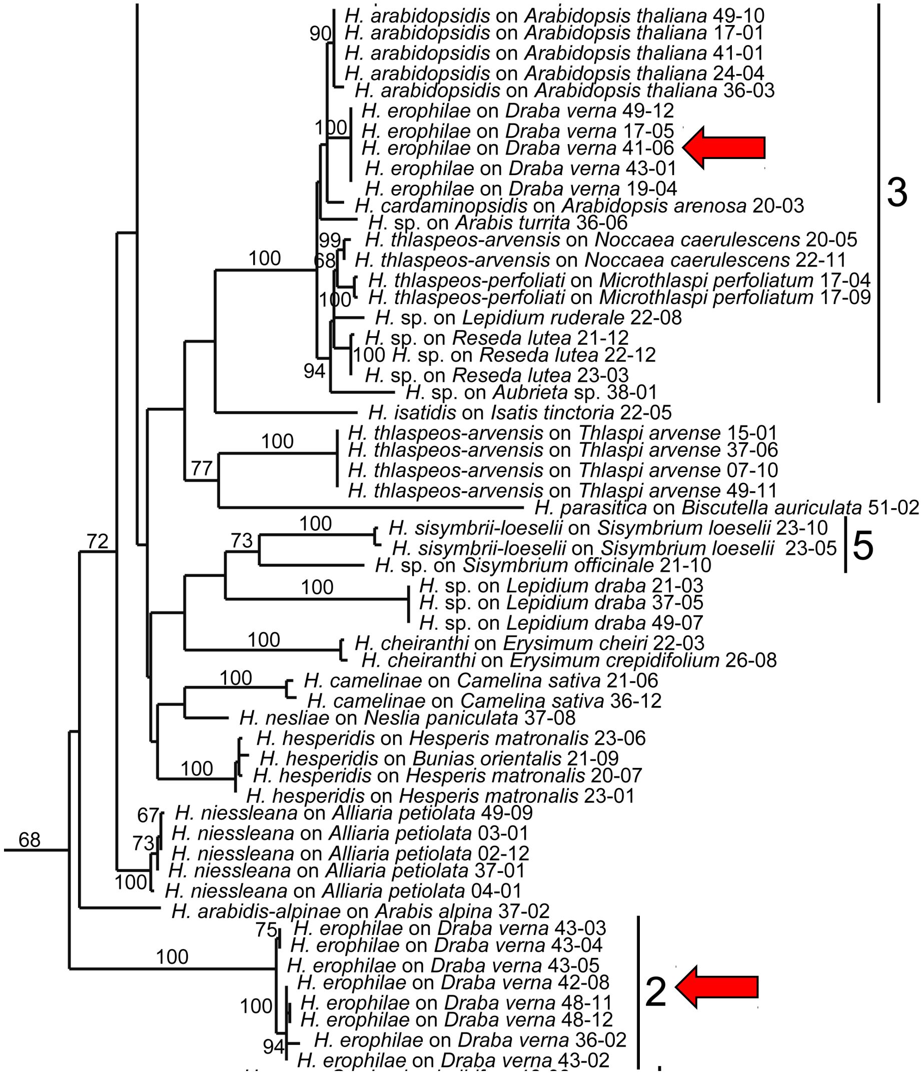
Definition:

Establishing a (informal or even formal) taxonomy of organisms based only on molecular sequences

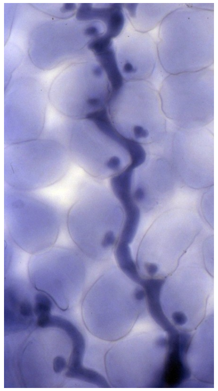
Uses:

- Detection of cryptic and pseudocryptic species
- Detection of misidentifications and mislabelled sequences in public databases
- Identification of juvenile specimens
- Analysis of environmental samples (e.g. metagenomics)

Example: (pseudo-)cryptic species


90
H. arabidopsidis on Arabidopsis thaliana 49-10
H. arabidopsidis on Arabidopsis thaliana 17-01
H. arabidopsidis on Arabidopsis thaliana 41-01
H. arabidopsidis on Arabidopsis thaliana 24-04
H. arabidopsidis on Arabidopsis thaliana 36-03
100
H. erophilae on Draba verna 49-12
H. erophilae on Draba verna 17-05
H. erophilae on Draba verna 41-06
H. erophilae on Draba verna 43-01
H. erophilae on Draba verna 19-04
H. cardaminopsidis on Arabidopsis arenosa 20-03
H. sp. on Arabis turrita 36-06
99
H. thlaspeos-arvensis on Noccaea caerulescens 20-05
68
H. thlaspeos-arvensis on Noccaea caerulescens 22-11
100
H. thlaspeos-perfoliati on Microthlaspi perfoliatum 17-04
100
H. thlaspeos-perfoliati on Microthlaspi perfoliatum 17-09
H. sp. on Lepidium ruderale 22-08
H. sp. on Reseda lutea 21-12
100 H. sp. on Reseda lutea 22-12
94
H. sp. on Reseda lutea 23-03
H. sp. on Aubrieta sp. 38-01
H. isatidis on Isatis tinctoria 22-05
3
100
H. thlaspeos-arvensis on Thlaspi arvense 15-01
H. thlaspeos-arvensis on Thlaspi arvense 37-06
77
H. thlaspeos-arvensis on Thlaspi arvense 07-10
H. thlaspeos-arvensis on Thlaspi arvense 49-11
H. parasitica on Biscutella auriculata 51-02
72
100
H. sisymbrii-loeselii on Sisymbrium loeselii 23-10
73
H. sisymbrii-loeselii on Sisymbrium loeselii 23-05
H. sp. on Sisymbrium officinale 21-10
5
100
H. sp. on Lepidium draba 21-03
H. sp. on Lepidium draba 37-05
H. sp. on Lepidium draba 49-07
100
H. cheiranthi on Erysimum cheiri 22-03
H. cheiranthi on Erysimum crepidifolium 26-08
100
H. camelinae on Camelina sativa 21-06
H. camelinae on Camelina sativa 36-12
H. nesliae on Neslia paniculata 37-08
H. hesperidis on Hesperis matronalis 23-06
H. hesperidis on Bunias orientalis 21-09
100
H. hesperidis on Hesperis matronalis 20-07
H. hesperidis on Hesperis matronalis 23-01
67
H. niessleana on Alliaria petiolata 49-09
68
73
H. niessleana on Alliaria petiolata 03-01
H. niessleana on Alliaria petiolata 02-12
H. niessleana on Alliaria petiolata 37-01
100
H. niessleana on Alliaria petiolata 04-01
H. arabidis-alpinae on Arabis alpina 37-02
75
H. erophilae on Draba verna 43-03
100
H. erophilae on Draba verna 43-04
H. erophilae on Draba verna 43-05
H. erophilae on Draba verna 42-08
100
H. erophilae on Draba verna 48-11
H. erophilae on Draba verna 48-12
2
94
H. erophilae on Draba verna 36-02
H. erophilae on Draba verna 43-02

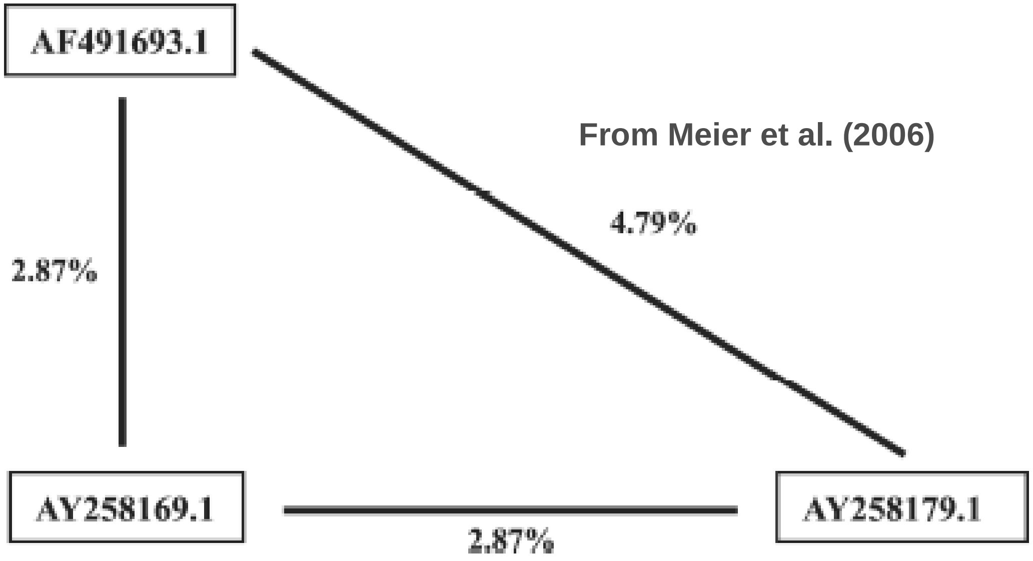
ITS/LSU rDNA data of the genus *Hyaloperonospora* (Peronosporales, Oomycetes) (Göker et al. 2009)



=> Two genetically distinct but microscopically identical species on *Draba verna* host plants

Threshold-based clustering

- Calculate distance *d(i,j)* between each pair of sequences *i* and *j*
- Define a threshold *T*
- Principle: if *d(i,j)* <= *T*, assign *i* and *j* to the same molecular operational taxonomic unit (MOTU)



From Meier et al. (2006)

FIGURE 1. Pairwise distances for three *Anopheles* sequences (AF491693.1, AY258179.1 = *A. maculipennis*; AY258169.1 = *A. messae*). All belong to the same 3% DNA profile, although one pairwise distance exceeds the threshold.

=> Can lead to inconsistencies if formulated in that way

Impact of the clustering algorithm

- A distance $d(i,j) <= T$ is called link
- An additional parameter, the “linkage fraction” *F*, determines how many links between an object and a cluster are necessary to include the object in the cluster

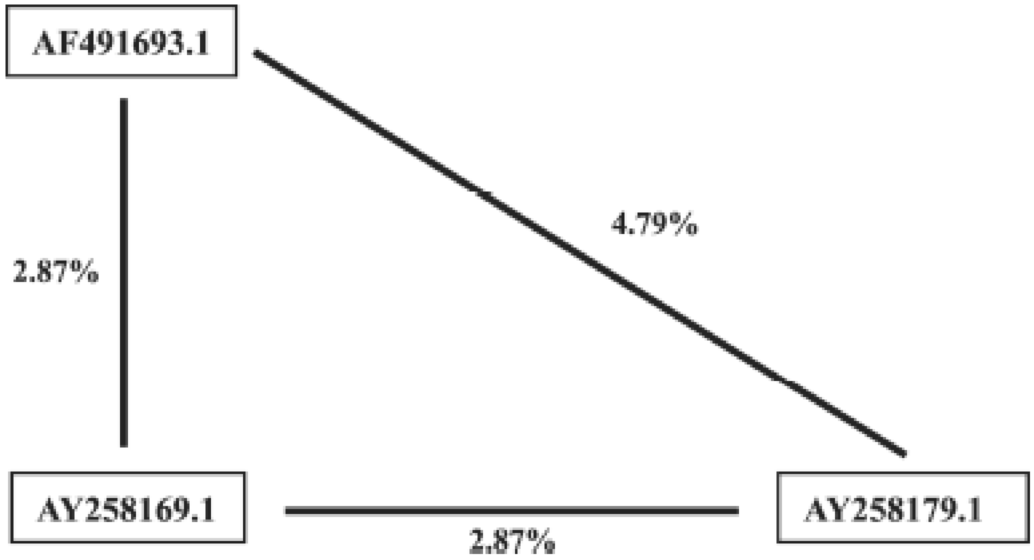


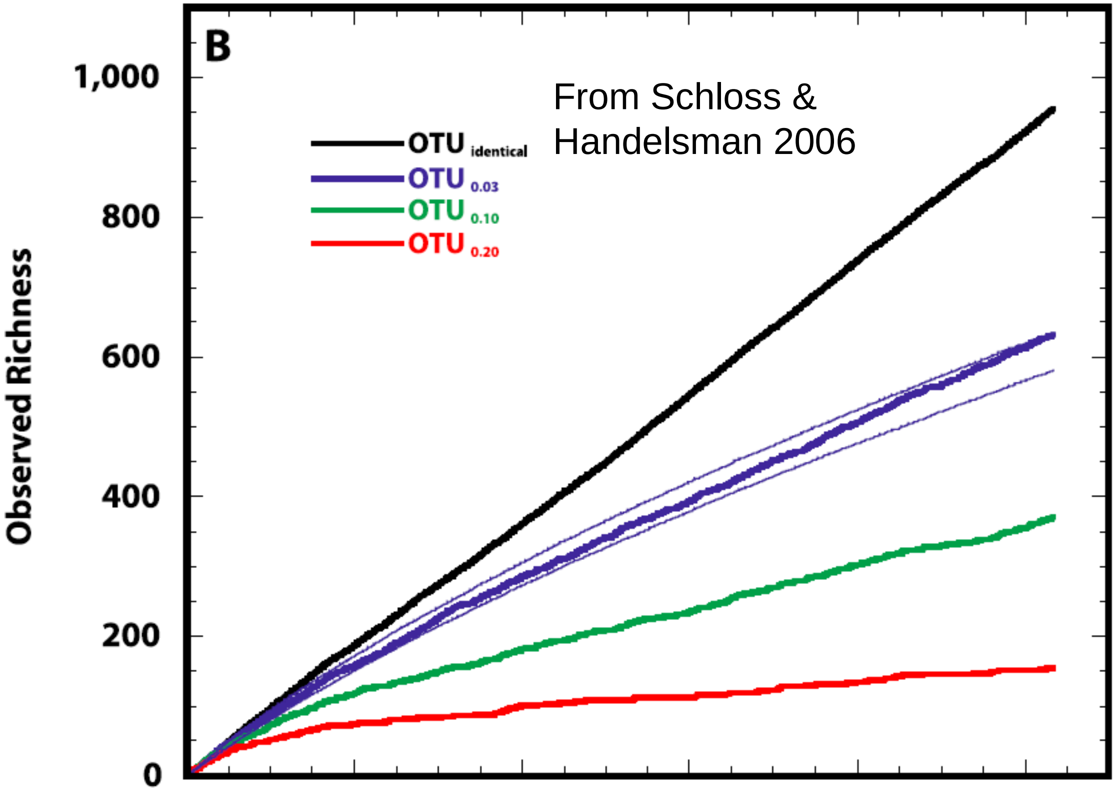
FIGURE 1. Pairwise distances for three *Anopheles* sequences (AF491693.1, AY258179.1 = *A. maculipennis*; AY258169.1 = *A. messae*). All belong to the same 3% DNA profile, although one pairwise distance exceeds the threshold.

=> Here, 1 cluster for *F* <= 0.5, but 2 clusters otherwise!
=> Lower *F* values allow higher within-cluster divergence

How to choose the clustering parameters?

Example:

- Species richness of soil bacteria estimated from 16S rDNA sequences
- Question: Has saturation been obtained?
- Obvious dependency on *T*



From Schloss & Handelsman 2006

=> Choice of parameters has serious consequences for total biodiversity estimates

The debate between traditional and molecular taxonomists

Ongoing intense (and sometimes hostile) debate between molecular taxonomists and traditional morphologists, particularly in the context of DNA barcoding

Criticisms of molecular taxonomy:

- Values of *T* used for clustering differ in the literature, even if applied to the same groups of organisms and molecular markers
- Values of *T* are often based on subjective criteria or on a tradition that emerged in recent years for the sake of comparability between studies
- Genetic divergence may differ between morphologically defined lineages
- A smaller distance (or a higher similarity) does not necessarily indicate a closer phylogenetic relationship

=> How can we maximize the agreement between traditional and molecular taxonomy?

Clustering optimization

- Partition := non-hierarchical, non-overlapping classification

- Many biological data are represented as partitions (e.g. assignment of sequences to species):

Accession number	Organism
EF050035	Pseudoperonospora cubensis
EF174888	Peronospora aestivalis
EF174890	Peronospora sepium
EF174891	Peronospora fulva
EF174894	Peronospora lathyri-verni
EF174944	Peronospora orobi
...	...

- Non-hierarchical clustering also results in a partition, e.g.:

Accession number	Cluster number
EF050035	29
EF174888	26
EF174890	25
EF174891	24
EF174894	27
EF174944	27
...	...

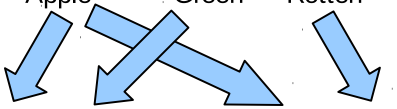
Approach:

- Use set of specimens identified using traditional techniques as reference points
- Determine the clustering parameters that maximize the agreement with a reference partition
- Do not require that full agreement can be obtained

Comparing partitions

3 example partitions of 7 objects

Object	**Fruit type**	**Colour**	**Condition**
A	Apple	Green	Fresh
B	Lemon	Yellow	Fresh
C	Cherry	Red	Rotten
D	Apple	Green	Fresh
E	Cherry	Red	Fresh
F	Lemon	Yellow	Rotten
G	Apple	Green	Rotten

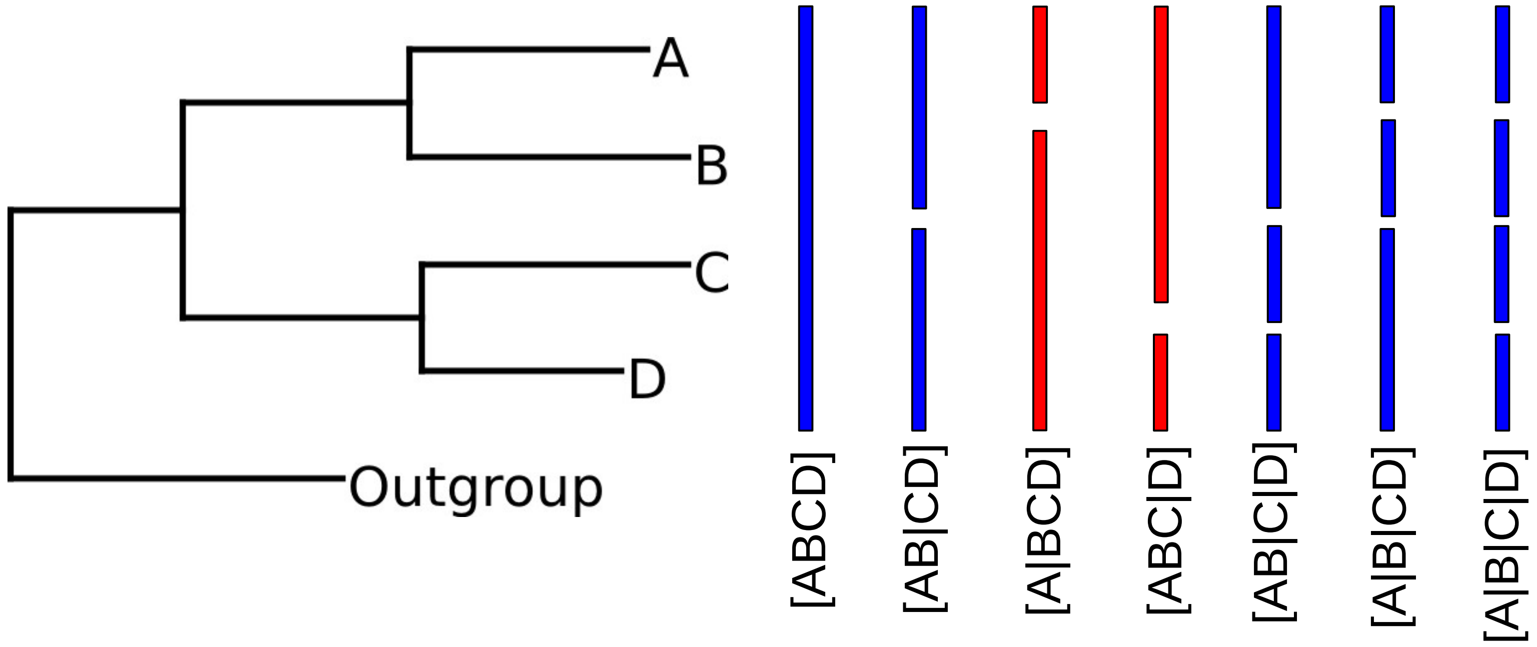


		Same	Different		Same	Different
Observed values	Same	5	0	Same	1	4
	Different	0	16	Different	8	8
		Same	Different		Same	Different
Expected values	Same	1.19	3.81	Same	2.14	2.86
	Different	3.81	12.19	Different	6.86	9.14

Rand Index	(5+16)/(5+0+0+16) = 1.0	(1+8)/(1+4+8+8) = 0.43
Expected Index	(1.19+12.19)/ (5+0+0+16) = 0.64	(2.14+9.14)/(1+4+8+8) = 0.54
Modified Rand Index (MRI)	(1.0-0.64)/(1.0-0.64) = 1.0	(0.43-0.54)/(1.0-0.54) = -0.24

- Rand index (Rand 1971): traverse all pairs of objects and determine proportion of those being in the same cluster in *both* partitions or in a different cluster in *both* partitions
- Modified Rand index (Hubert & Arabie 1985): corrects for chance (by relating to the expected Rand index for two random partitions with the same cluster number and sizes)

Why trees don't help

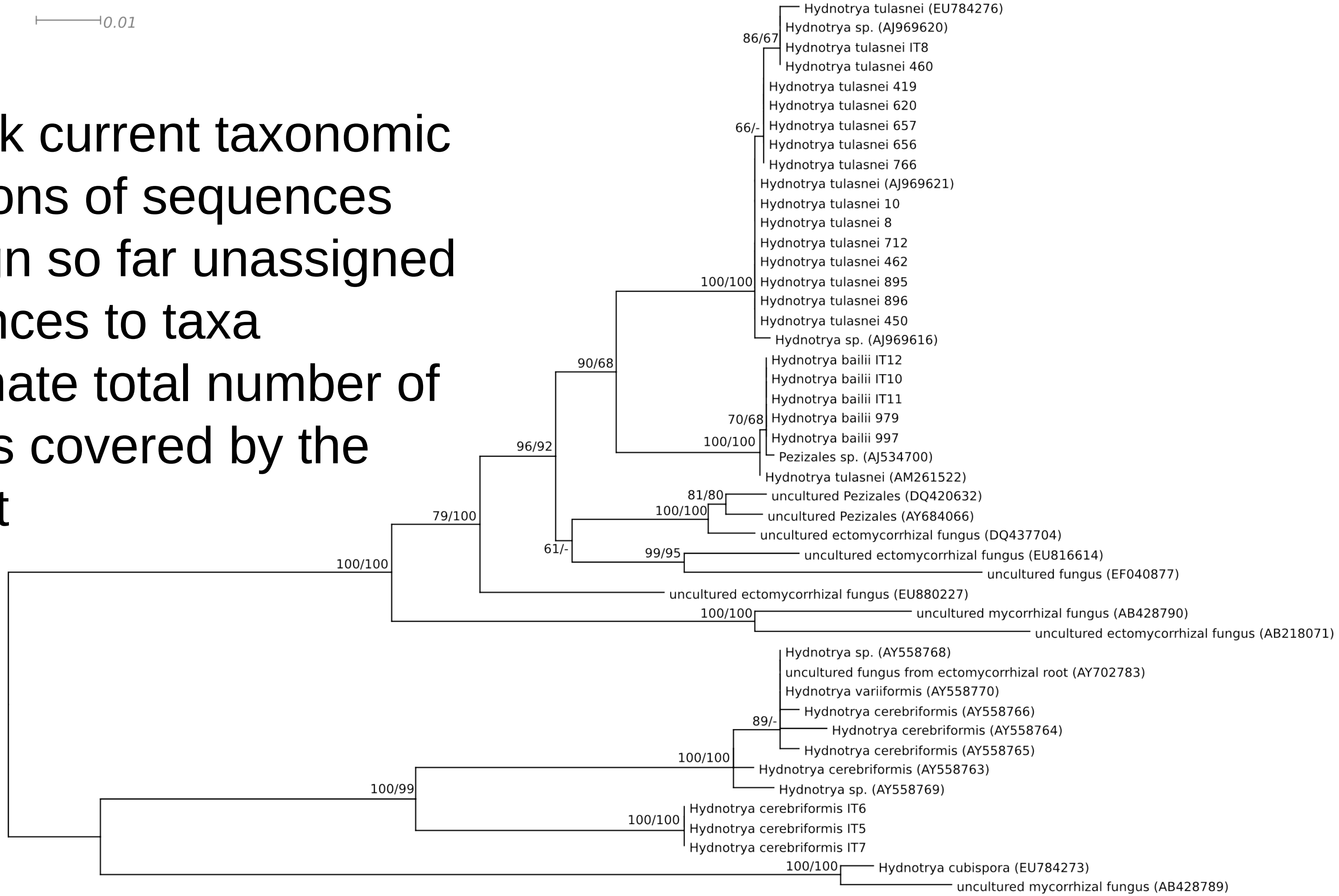


A phylogenetic tree rules out certain classifications (e.g. red ones), but is compatible with many others (blue ones)

Example: *Hydnotrya* ITS rDNA

Tasks

- Check current taxonomic affiliations of sequences
- Assign so far unassigned sequences to taxa
- Estimate total number of species covered by the dataset

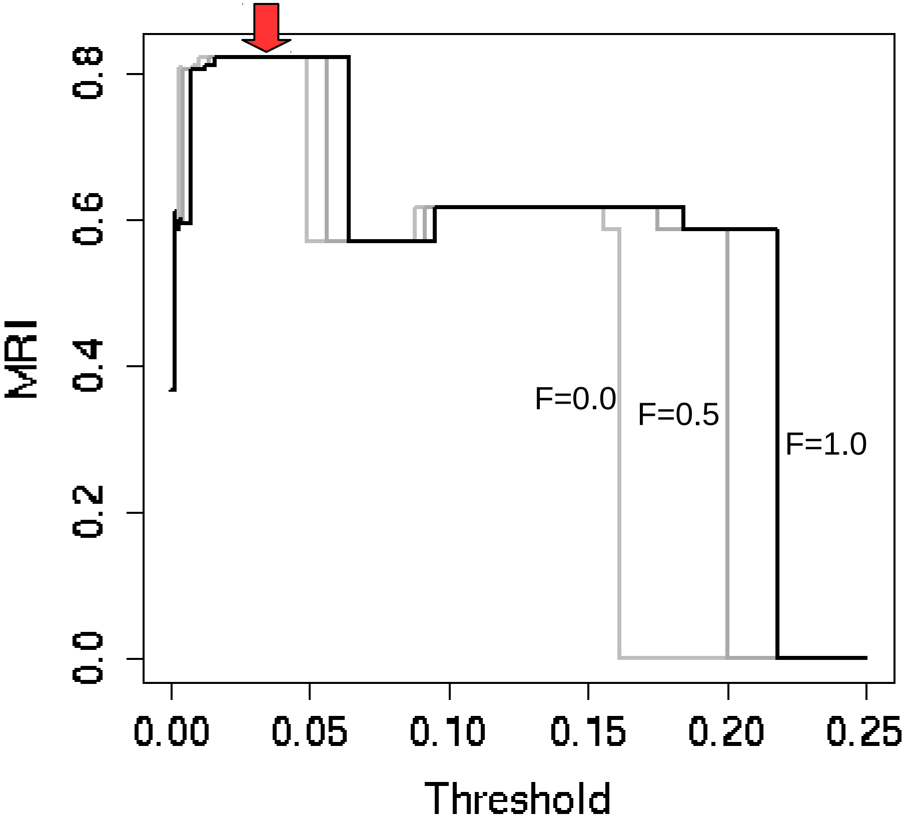

0.01
Hydnotrya tulasnei (EU784276)
Hydnotrya sp. (AJ969620)
86/67
Hydnotrya tulasnei IT8
Hydnotrya tulasnei 460
Hydnotrya tulasnei 419
Hydnotrya tulasnei 620
66/-
Hydnotrya tulasnei 657
Hydnotrya tulasnei 656
Hydnotrya tulasnei 766
Hydnotrya tulasnei (AJ969621)
Hydnotrya tulasnei 10
Hydnotrya tulasnei 8
Hydnotrya tulasnei 712
Hydnotrya tulasnei 462
100/100
Hydnotrya tulasnei 895
Hydnotrya tulasnei 896
Hydnotrya tulasnei 450
Hydnotrya sp. (AJ969616)
90/68
Hydnotrya bailii IT12
Hydnotrya bailii IT10
Hydnotrya bailii IT11
70/68
Hydnotrya bailii 979
96/92
100/100
Hydnotrya bailii 997
Pezizales sp. (AJ534700)
Hydnotrya tulasnei (AM261522)
81/80
uncultured Pezizales (DQ420632)
100/100
79/100
uncultured Pezizales (AY684066)
uncultured ectomycorrhizal fungus (DQ437704)
61/-
99/95
uncultured ectomycorrhizal fungus (EU816614)
100/100
uncultured fungus (EF040877)
uncultured ectomycorrhizal fungus (EU880227)
100/100
uncultured mycorrhizal fungus (AB428790)
uncultured ectomycorrhizal fungus (AB218071)
Hydnotrya sp. (AY558768)
uncultured fungus from ectomycorrhizal root (AY702783)
Hydnotrya variiformis (AY558770)
Hydnotrya cerebriformis (AY558766)
89/-
Hydnotrya cerebriformis (AY558764)
Hydnotrya cerebriformis (AY558765)
100/100
Hydnotrya cerebriformis (AY558763)
100/99
Hydnotrya sp. (AY558769)
Hydnotrya cerebriformis IT6
100/100
Hydnotrya cerebriformis IT5
Hydnotrya cerebriformis IT7
100/100
Hydnotrya cubispora (EU784273)
uncultured mycorrhizal fungus (AB428789)

Example: *Hydnotrya*

Procedure

1) Restrict dataset to sequences with taxonomic affiliations

Accession number	Organism	Species name present?
EU784276	Hydnotrya tulasnei	Yes
AJ969620	Hydnotrya sp. G-Ht	No
AJ969621	Hydnotrya tulasnei	Yes
AJ969616	Hydnotrya sp. LB-Ht	No
AJ534700	Pezizales sp. B48	No
AM261522	Hydnotrya tulasnei	Yes
DQ420632	uncultured Pezizales	No
...	...	...

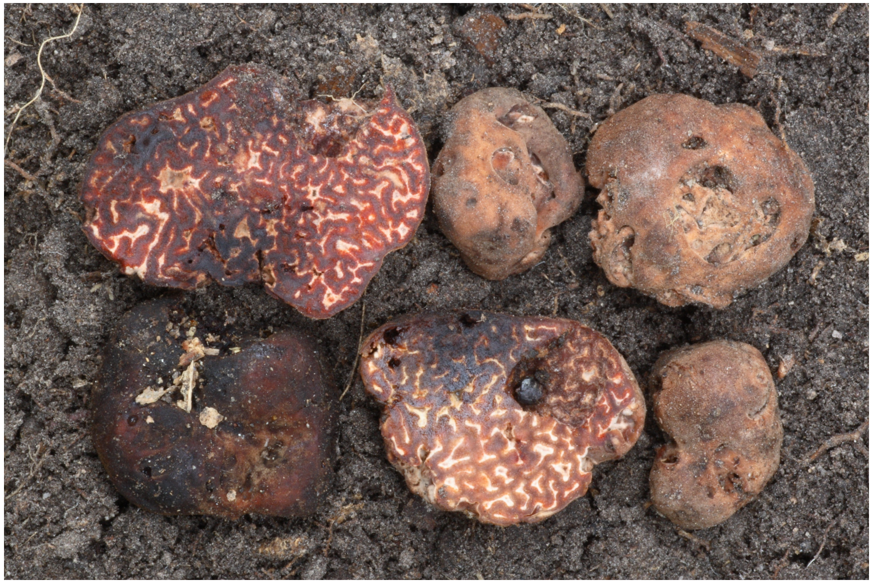


2) Conduct clustering optimization with reduced dataset

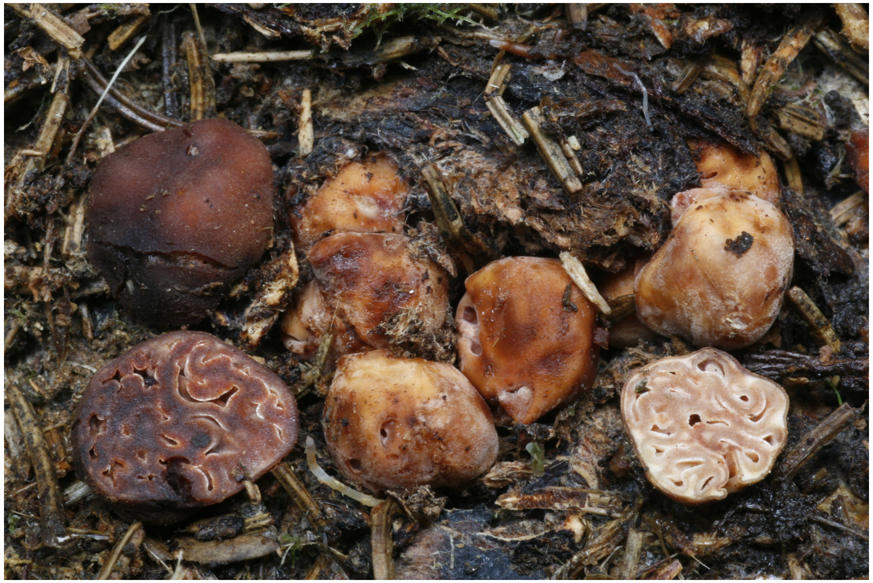
3) Place sequences without taxonomic affiliations back in the dataset

4) Conduct clustering with all sequences and optimized parameters

An overlooked hypogeous fungus



Hydnotrya tulasnei ascocarp
(picture: G.Hensel)



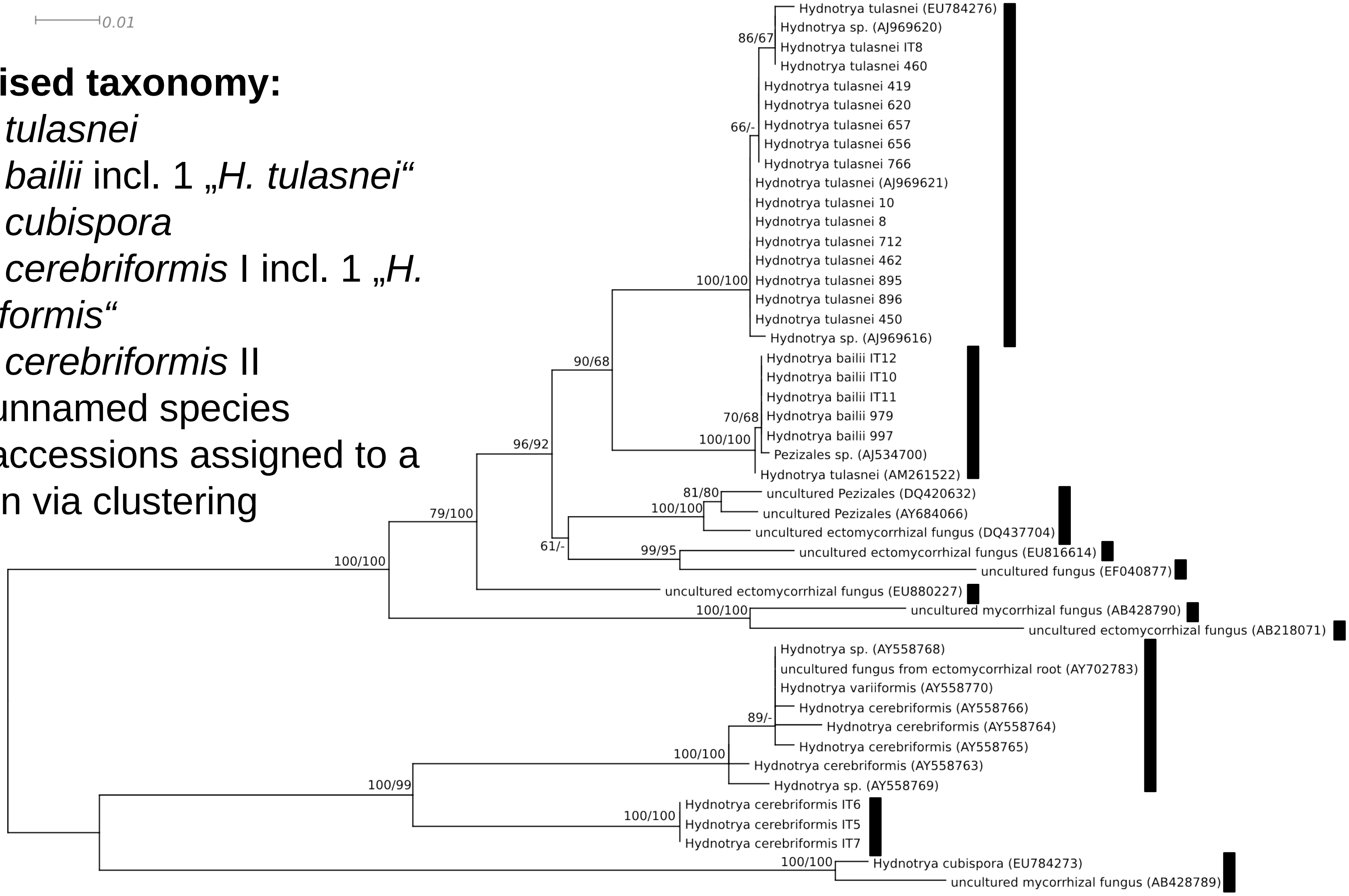
Hydnotrya bailii ascocarp
(picture: G. Hensel)

Stielow et al., under review:
Distinction between *Hydnotrya bailii* Soehner (1959) and *Hydnotrya tulasnei* (Berk.) Berk. & Broome (1846) has been neglected for 50 years!

Up to 50% of the MOTUs may be novel species

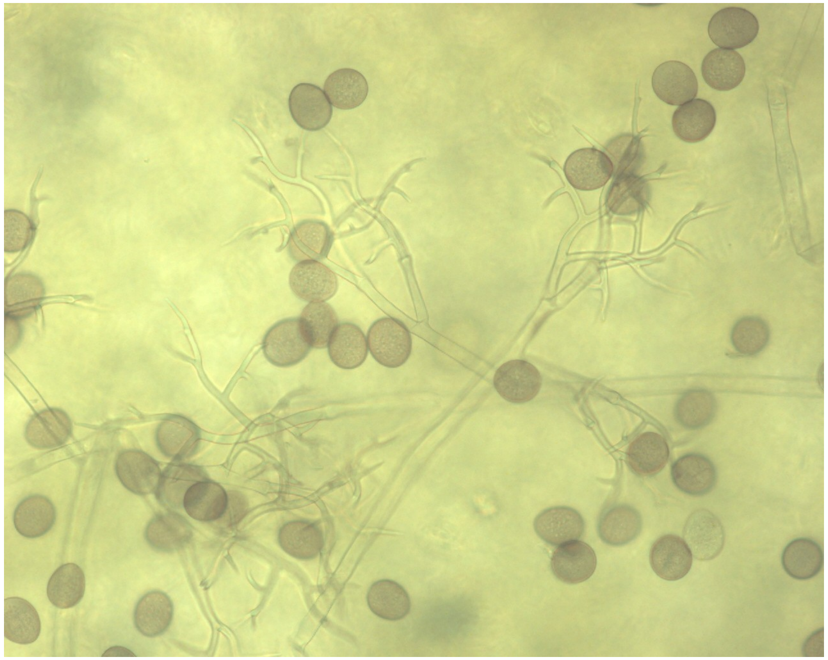
Revised taxonomy:

- *H. tulasnei*
- *H. bailii* incl. 1 „*H. tulasnei*“
- *H. cubispora*
- *H. cerebriformis* I incl. 1 „*H. variiformis*“
- *H. cerebriformis* II
- 6 unnamed species
- 7 accessions assigned to a taxon via clustering



0.01
Hydnotrya tulasnei (EU784276)
Hydnotrya sp. (AJ969620)
86/67
Hydnotrya tulasnei IT8
Hydnotrya tulasnei 460
Hydnotrya tulasnei 419
Hydnotrya tulasnei 620
66/-
Hydnotrya tulasnei 657
Hydnotrya tulasnei 656
Hydnotrya tulasnei 766
Hydnotrya tulasnei (AJ969621)
Hydnotrya tulasnei 10
Hydnotrya tulasnei 8
Hydnotrya tulasnei 712
Hydnotrya tulasnei 462
100/100
Hydnotrya tulasnei 895
Hydnotrya tulasnei 896
Hydnotrya tulasnei 450
Hydnotrya sp. (AJ969616)
90/68
Hydnotrya bailii IT12
Hydnotrya bailii IT10
Hydnotrya bailii IT11
70/68
Hydnotrya bailii 979
96/92
100/100
Hydnotrya bailii 997
Pezizales sp. (AJ534700)
Hydnotrya tulasnei (AM261522)
81/80
uncultured Pezizales (DQ420632)
100/100
uncultured Pezizales (AY684066)
79/100
uncultured ectomycorrhizal fungus (DQ437704)
61/-
99/95
uncultured ectomycorrhizal fungus (EU816614)
100/100
uncultured fungus (EF040877)
uncultured ectomycorrhizal fungus (EU880227)
100/100
uncultured mycorrhizal fungus (AB428790)
uncultured ectomycorrhizal fungus (AB218071)
Hydnotrya sp. (AY558768)
uncultured fungus from ectomycorrhizal root (AY702783)
Hydnotrya variiformis (AY558770)
Hydnotrya cerebriformis (AY558766)
89/-
Hydnotrya cerebriformis (AY558764)
Hydnotrya cerebriformis (AY558765)
100/100
Hydnotrya cerebriformis (AY558763)
100/99
Hydnotrya sp. (AY558769)
Hydnotrya cerebriformis IT6
100/100
Hydnotrya cerebriformis IT5
Hydnotrya cerebriformis IT7
100/100
Hydnotrya cubispora (EU784273)
uncultured mycorrhizal fungus (AB428789)

Example: *Peronospora* ITS rDNA



Peronospora sp. on *Ocimum basilicum*

Tasks

- Revise nomenclature of all Genbank *Peronospora* ITS rDNA sequences
- Check whether a combination of molecular and host plant characters is sufficient to obtain a consistent species concept

Example: *Peronospora*

Procedure

1) Restrict dataset to (a) sequences with taxonomic affiliations and (b) sequences with interpretable host names

Accession number	Organism	Specific host	Species name present?	Host present?
EF614964	Peronospora variabilis	Chenopodium album	Yes	Yes
EF614958	Peronospora sp. SMK20063	Chenopodium ambrosioides	No	Yes
EF614957	Peronospora sp. DAR45530	Chenopodium ambrosioides	No	Yes
EF614955	Peronospora farinosa f. sp. chenopodii	Chenopodium hybridum	Yes	Yes
EF174939	Peronospora sp. GG133		No	No
EF174924	Peronospora sp. HV956		No	No
EF174970	Peronospora trifoliorum		Yes	No
EF174963	Peronospora trifoliorum		Yes	No

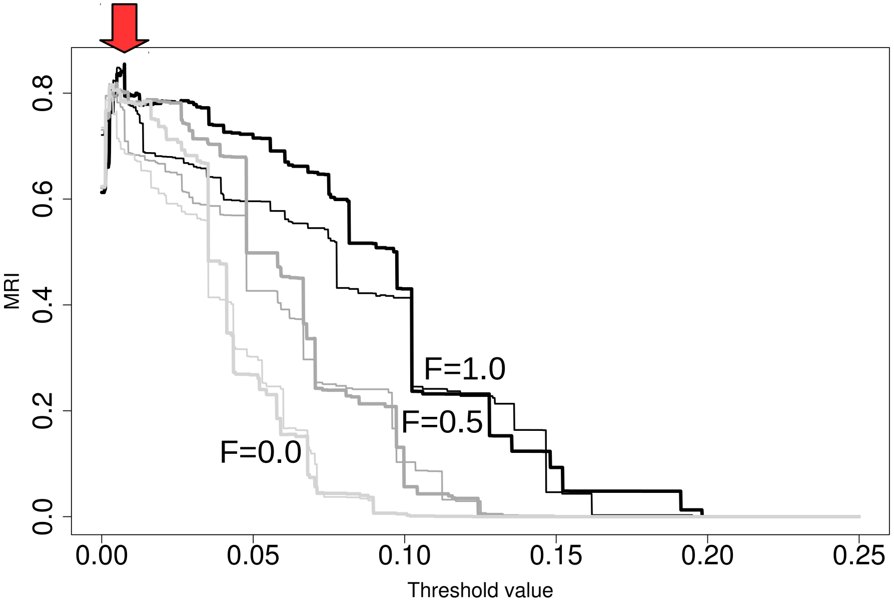
2) Conduct clustering optimization with reduced datasets (a) and (b)

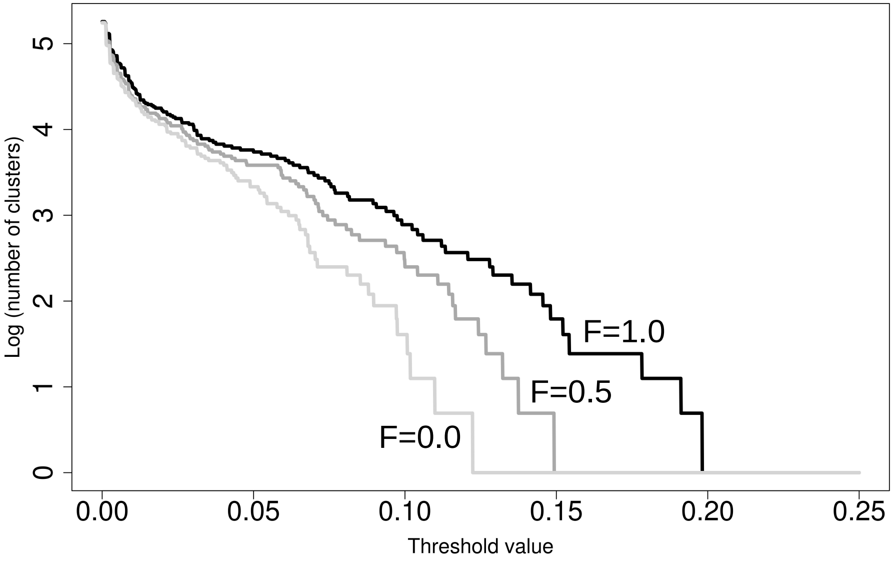
3) Check for coincidence of results (i.e. of optimal clustering parameters)

4) Place sequences without taxonomic affiliations or host information back in the dataset

5) Conduct clustering with all sequences and optimized parameters

Example: *Peronospora* ITS rDNA

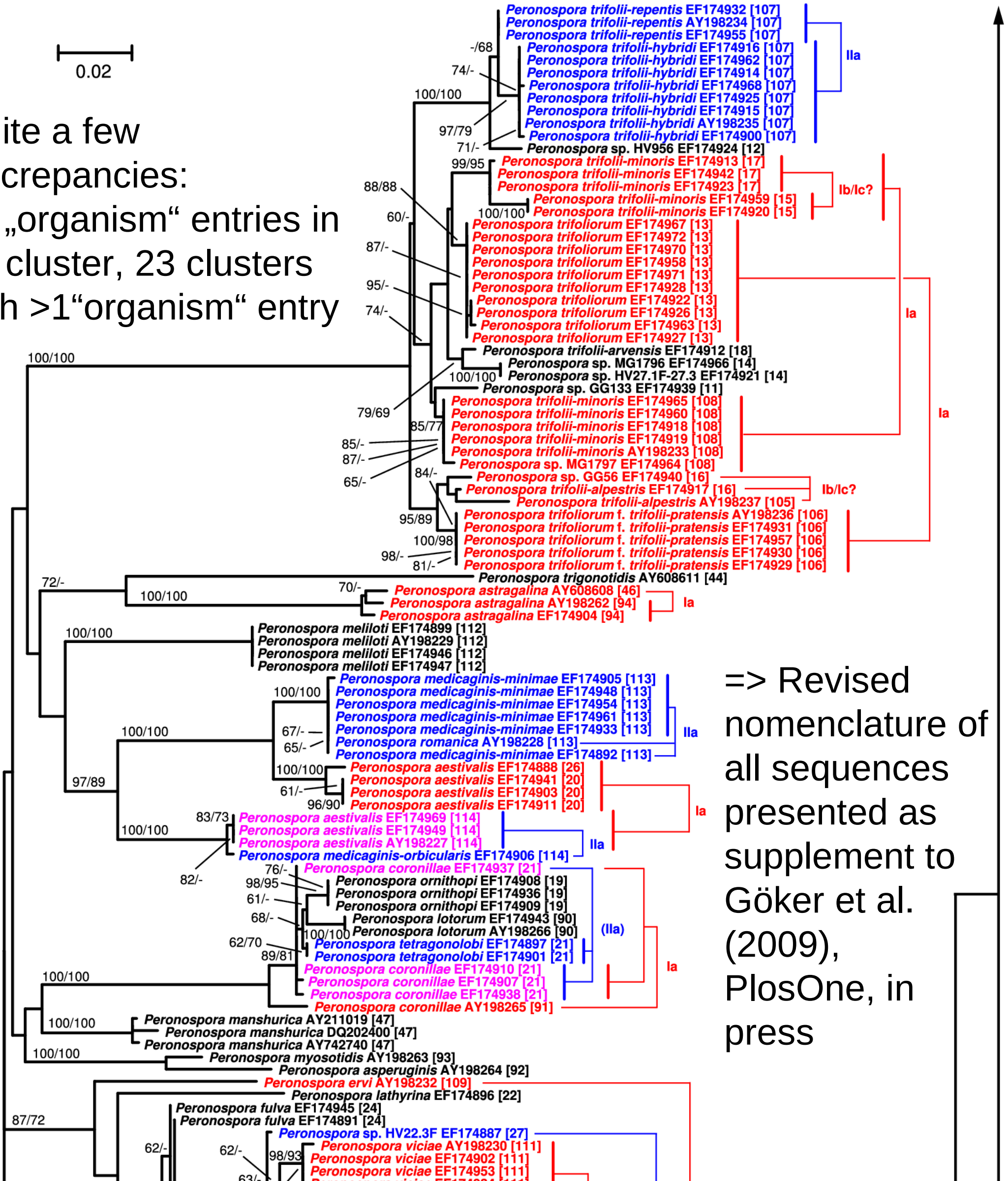




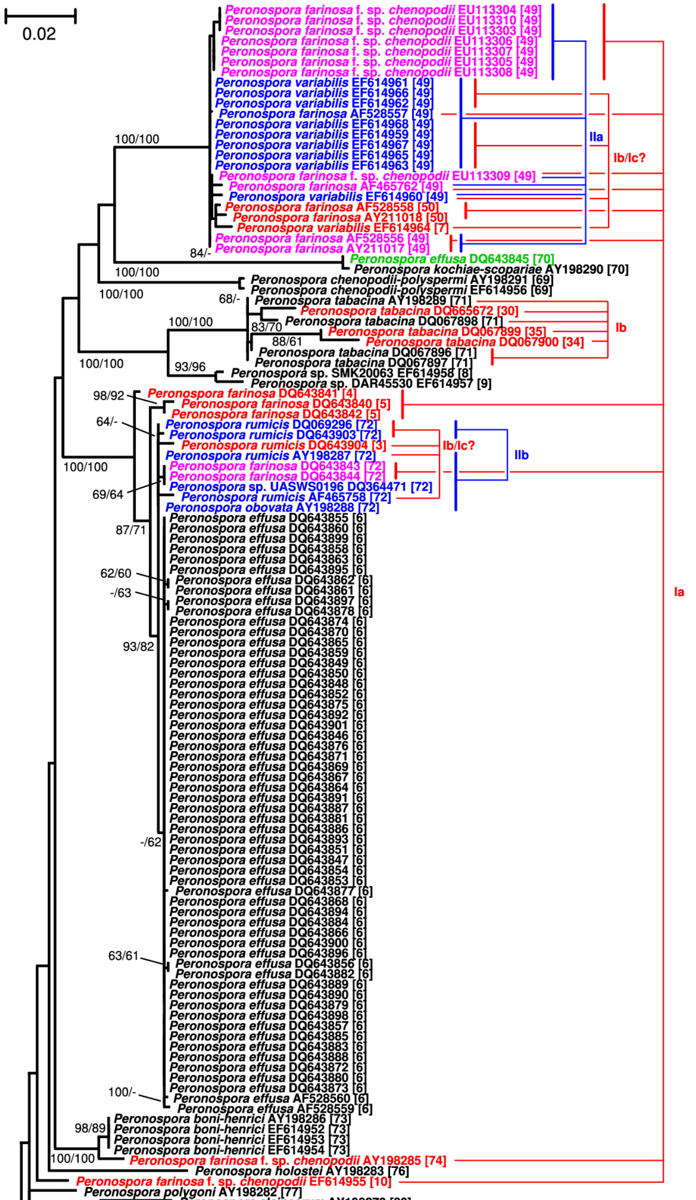
- Taxonomy-based optimization: best result (MRI=0.85485) with T=0.0075 and F=1.0 (left picture: thick lines)
- Host-based optimization: best result (MRI=0.85204) with T=0.0075 and F=1.0 (left picture: thin lines) => *exactly the same optimum*
- Resulting in 117 clusters

Example: *Peronospora* ITS rDNA

Quite a few discrepancies:
20 „organism“ entries in >1 cluster, 23 clusters with >1“organism“ entry


0.02
Peronospora trifolii-repentis EF174932 [107]
Peronospora trifolii-repentis AY198234 [107]
Peronospora trifolii-repentis EF174955 [107]
Peronospora trifolii-hybridi EF174916 [107]
Peronospora trifolii-hybridi EF174962 [107]
Peronospora trifolii-hybridi EF174914 [107]
Peronospora trifolii-hybridi EF174968 [107]
Peronospora trifolii-hybridi EF174925 [107]
Peronospora trifolii-hybridi EF174915 [107]
Peronospora trifolii-hybridi AY198235 [107]
Peronospora trifolii-hybridi EF174900 [107]
Peronospora sp. HV956 EF174924 [12]
Peronospora trifolii-minoris EF174913 [17]
Peronospora trifolii-minoris EF174942 [17]
Peronospora trifolii-minoris EF174923 [17]
Peronospora trifolii-minoris EF174959 [15]
Peronospora trifolii-minoris EF174920 [15]
Peronospora trifoliorum EF174967 [13]
Peronospora trifoliorum EF174972 [13]
Peronospora trifoliorum EF174970 [13]
Peronospora trifoliorum EF174958 [13]
Peronospora trifoliorum EF174971 [13]
Peronospora trifoliorum EF174928 [13]
Peronospora trifoliorum EF174922 [13]
Peronospora trifoliorum EF174926 [13]
Peronospora trifoliorum EF174963 [13]
Peronospora trifoliorum EF174927 [13]
Peronospora trifolii-arvensis EF174912 [18]
Peronospora sp. MG1796 EF174966 [14]
Peronospora sp. HV27.1F-27.3 EF174921 [14]
Peronospora sp. GG133 EF174939 [11]
Peronospora trifolii-minoris EF174965 [108]
Peronospora trifolii-minoris EF174960 [108]
Peronospora trifolii-minoris EF174918 [108]
Peronospora trifolii-minoris EF174919 [108]
Peronospora trifolii-minoris AY198233 [108]
Peronospora sp. MG1797 EF174964 [108]
Peronospora sp. GG56 EF174940 [16]
Peronospora trifolii-alpestris EF174917 [16]
Peronospora trifolii-alpestris AY198237 [105]
Peronospora trifoliorum f. trifolii-pratensis AY198236 [106]
Peronospora trifoliorum f. trifolii-pratensis EF174931 [106]
Peronospora trifoliorum f. trifolii-pratensis EF174957 [106]
Peronospora trifoliorum f. trifolii-pratensis EF174930 [106]
Peronospora trifoliorum f. trifolii-pratensis EF174929 [106]
Peronospora trigonotidis AY608611 [44]
Peronospora astragalina AY608608 [46]
Peronospora astragalina AY198262 [94]
Peronospora astragalina EF174904 [94]
Peronospora meliloti EF174899 [112]
Peronospora meliloti AY198229 [112]
Peronospora meliloti EF174946 [112]
Peronospora meliloti EF174947 [112]
Peronospora medicaginis-minimae EF174905 [113]
Peronospora medicaginis-minimae EF174948 [113]
Peronospora medicaginis-minimae EF174954 [113]
Peronospora medicaginis-minimae EF174961 [113]
Peronospora medicaginis-minimae EF174933 [113]
Peronospora romanica AY198228 [113]
Peronospora medicaginis-minimae EF174892 [113]
Peronospora aestivalis EF174888 [26]
Peronospora aestivalis EF174941 [20]
Peronospora aestivalis EF174903 [20]
Peronospora aestivalis EF174911 [20]
Peronospora aestivalis EF174969 [114]
Peronospora aestivalis EF174949 [114]
Peronospora aestivalis AY198227 [114]
Peronospora medicaginis-orbicularis EF174906 [114]
Peronospora coronillae EF174937 [21]
Peronospora ornithopi EF174908 [19]
Peronospora ornithopi EF174936 [19]
Peronospora ornithopi EF174909 [19]
Peronospora lotorum EF174943 [90]
Peronospora lotorum AY198266 [90]
Peronospora tetragonolobi EF174897 [21]
Peronospora tetragonolobi EF174901 [21]
Peronospora coronillae EF174910 [21]
Peronospora coronillae EF174907 [21]
Peronospora coronillae EF174938 [21]
Peronospora coronillae AY198265 [91]
Peronospora manshurica AY211019 [47]
Peronospora manshurica DQ202400 [47]
Peronospora manshurica AY742740 [47]
Peronospora myosotidis AY198263 [93]
Peronospora asperuginis AY198264 [92]
Peronospora ervi AY198232 [109]
Peronospora lathyrina EF174896 [22]
Peronospora fulva EF174945 [24]
Peronospora fulva EF174891 [24]
Peronospora sp. HV22.3F EF174887 [27]
Peronospora viciae AY198230 [111]
Peronospora viciae EF174902 [111]
Peronospora viciae EF174953 [111]
-/68
74/-
100/100
97/79
71/-
99/95
88/88
60/-
100/100
87/-
95/-
74/-
100/100
79/69
85/77
85/-
87/-
65/-
84/-
95/89
100/98
98/-
81/-
100/100
72/-
100/100
70/-
100/100
100/100
67/-
65/-
100/100
61/-
96/90
97/89
83/73
100/100
82/-
76/-
98/95
61/-
68/-
100/100
62/70
89/81
100/100
100/100
100/100
87/72
62/-
62/-
98/93
63/-
IIa
Ib/Ic?
Ia
Ia
Ib/Ic?
Ia
IIa
Ia
IIa
(IIa)
Ia

=> Revised nomenclature of all sequences presented as supplement to Göker et al. (2009), PlosOne, in press


0.02
Peronospora farinosa f. sp. chenopodii EU113304 [49]
Peronospora farinosa f. sp. chenopodii EU113310 [49]
Peronospora farinosa f. sp. chenopodii EU113303 [49]
Peronospora farinosa f. sp. chenopodii EU113306 [49]
Peronospora farinosa f. sp. chenopodii EU113307 [49]
Peronospora farinosa f. sp. chenopodii EU113305 [49]
Peronospora farinosa f. sp. chenopodii EU113308 [49]
Peronospora variabilis EF614961 [49]
Peronospora variabilis EF614966 [49]
Peronospora variabilis EF614962 [49]
Peronospora farinosa AF528557 [49]
Peronospora variabilis EF614968 [49]
Peronospora variabilis EF614959 [49]
Peronospora variabilis EF614967 [49]
Peronospora variabilis EF614965 [49]
Peronospora variabilis EF614963 [49]
Peronospora farinosa f. sp. chenopodii EU113309 [49]
Peronospora farinosa AF465762 [49]
Peronospora variabilis EF614960 [49]
Peronospora farinosa AF528558 [50]
Peronospora farinosa AY211018 [50]
Peronospora variabilis EF614964 [7]
Peronospora farinosa AF528556 [49]
Peronospora farinosa AY211017 [49]
Peronospora effusa DQ643845 [70]
Peronospora kochiae-scopariae AY198290 [70]
Peronospora chenopodii-polyspermi AY198291 [69]
Peronospora chenopodii-polyspermi EF614956 [69]
Peronospora tabacina AY198289 [71]
Peronospora tabacina DQ665672 [30]
Peronospora tabacina DQ067898 [71]
Peronospora tabacina DQ067899 [35]
Peronospora tabacina DQ067900 [34]
Peronospora tabacina DQ067896 [71]
Peronospora tabacina DQ067897 [71]
Peronospora sp. SMK20063 EF614958 [8]
Peronospora sp. DAR45530 EF614957 [9]
Peronospora farinosa DQ643841 [4]
Peronospora farinosa DQ643840 [5]
Peronospora farinosa DQ643842 [5]
Peronospora rumicis DQ069296 [72]
Peronospora rumicis DQ643903 [72]
Peronospora rumicis DQ643904 [3]
Peronospora rumicis AY198287 [72]
Peronospora farinosa DQ643843 [72]
Peronospora farinosa DQ643844 [72]
Peronospora sp. UASWS0196 DQ364471 [72]
Peronospora rumicis AF465758 [72]
Peronospora obovata AY198288 [72]
Peronospora effusa DQ643855 [6]
Peronospora effusa DQ643860 [6]
Peronospora effusa DQ643899 [6]
Peronospora effusa DQ643858 [6]
Peronospora effusa DQ643863 [6]
Peronospora effusa DQ643895 [6]
Peronospora effusa DQ643862 [6]
Peronospora effusa DQ643861 [6]
Peronospora effusa DQ643897 [6]
Peronospora effusa DQ643878 [6]
Peronospora effusa DQ643874 [6]
Peronospora effusa DQ643870 [6]
Peronospora effusa DQ643865 [6]
Peronospora effusa DQ643859 [6]
Peronospora effusa DQ643849 [6]
Peronospora effusa DQ643850 [6]
Peronospora effusa DQ643848 [6]
Peronospora effusa DQ643852 [6]
Peronospora effusa DQ643875 [6]
Peronospora effusa DQ643892 [6]
Peronospora effusa DQ643901 [6]
Peronospora effusa DQ643846 [6]
Peronospora effusa DQ643876 [6]
Peronospora effusa DQ643871 [6]
Peronospora effusa DQ643869 [6]
Peronospora effusa DQ643867 [6]
Peronospora effusa DQ643864 [6]
Peronospora effusa DQ643891 [6]
Peronospora effusa DQ643887 [6]
Peronospora effusa DQ643881 [6]
Peronospora effusa DQ643886 [6]
Peronospora effusa DQ643893 [6]
Peronospora effusa DQ643851 [6]
Peronospora effusa DQ643847 [6]
Peronospora effusa DQ643854 [6]
Peronospora effusa DQ643853 [6]
Peronospora effusa DQ643877 [6]
Peronospora effusa DQ643868 [6]
Peronospora effusa DQ643894 [6]
Peronospora effusa DQ643884 [6]
Peronospora effusa DQ643866 [6]
Peronospora effusa DQ643900 [6]
Peronospora effusa DQ643896 [6]
Peronospora effusa DQ643856 [6]
Peronospora effusa DQ643882 [6]
Peronospora effusa DQ643889 [6]
Peronospora effusa DQ643890 [6]
Peronospora effusa DQ643879 [6]
Peronospora effusa DQ643898 [6]
Peronospora effusa DQ643857 [6]
Peronospora effusa DQ643885 [6]
Peronospora effusa DQ643883 [6]
Peronospora effusa DQ643888 [6]
Peronospora effusa DQ643872 [6]
Peronospora effusa DQ643880 [6]
Peronospora effusa DQ643873 [6]
Peronospora effusa AF528560 [6]
Peronospora effusa AF528559 [6]
Peronospora boni-henrici AY198286 [73]
Peronospora boni-henrici EF614952 [73]
Peronospora boni-henrici EF614953 [73]
Peronospora boni-henrici EF614954 [73]
Peronospora farinosa f. sp. chenopodii AY198285 [74]
Peronospora holostei AY198283 [76]
Peronospora farinosa f. sp. chenopodii EF614955 [10]
Peronospora polygoni AY198282 [77]
100/100
84/-
100/100
100/100
68/-
100/100
83/70
88/61
100/100
93/96
98/92
64/-
100/100
69/64
87/71
62/60
-/63
93/82
-/62
63/61
100/-
98/89
100/100
IIa
Ib/Ic?
Ib
Ib/Ic?
IIb
Ia

Robustness against sampling bias

Figure 1

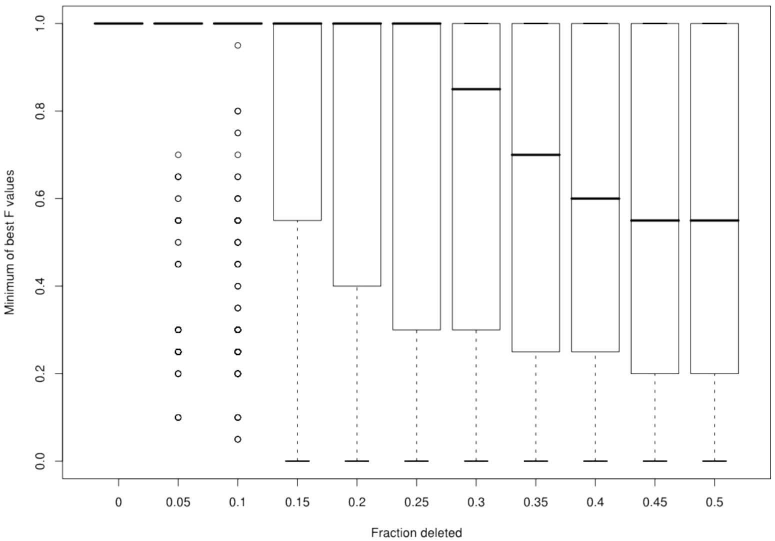


Figure 2

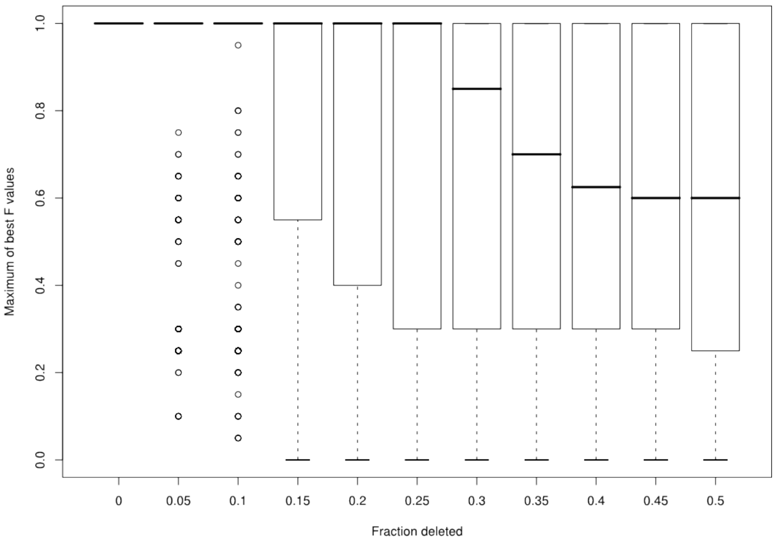


Figure 3

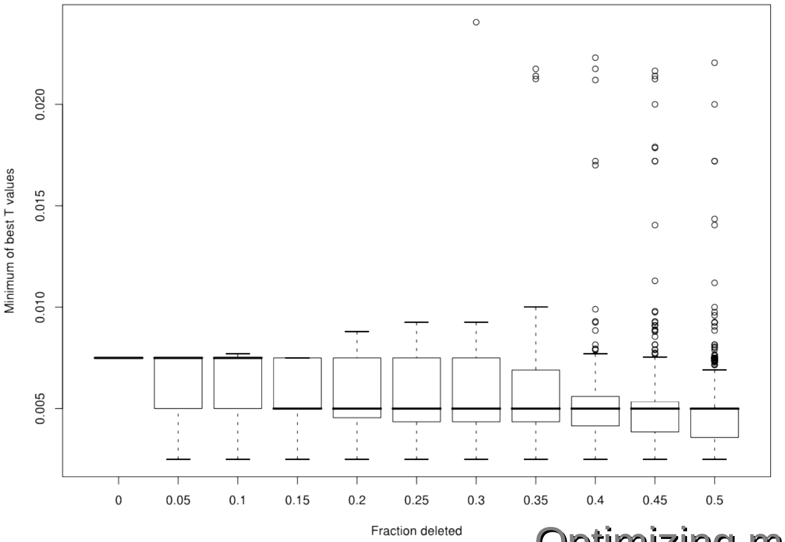
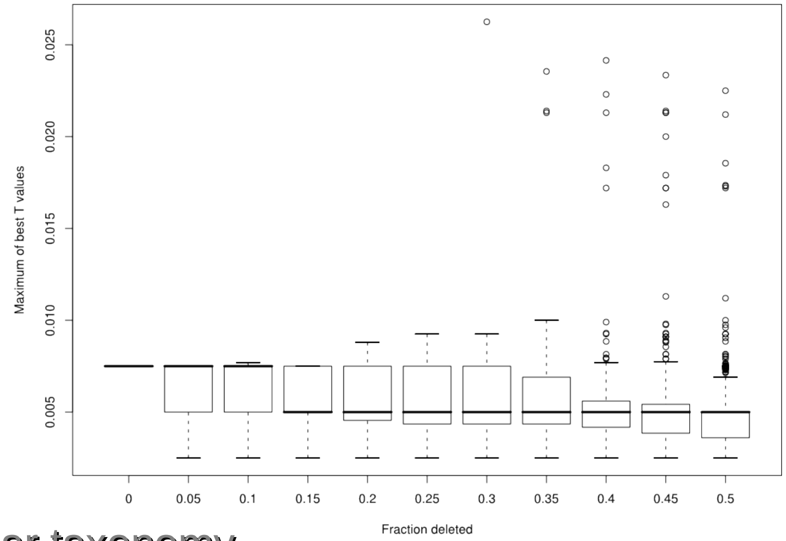


Figure 4



Robustness against errors in the reference partition

Figure 6

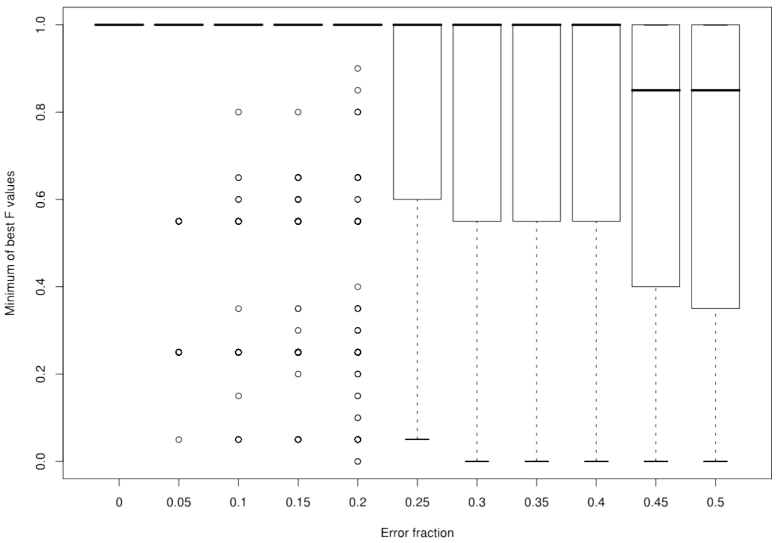


Figure 7

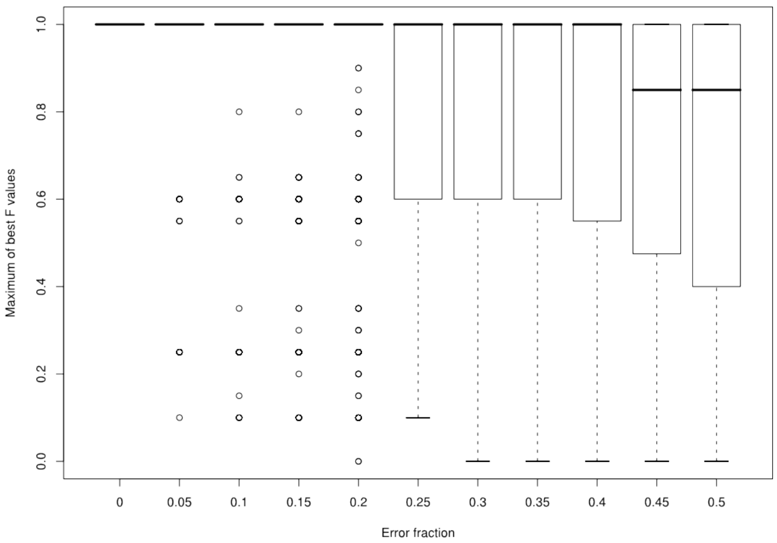


Figure 8

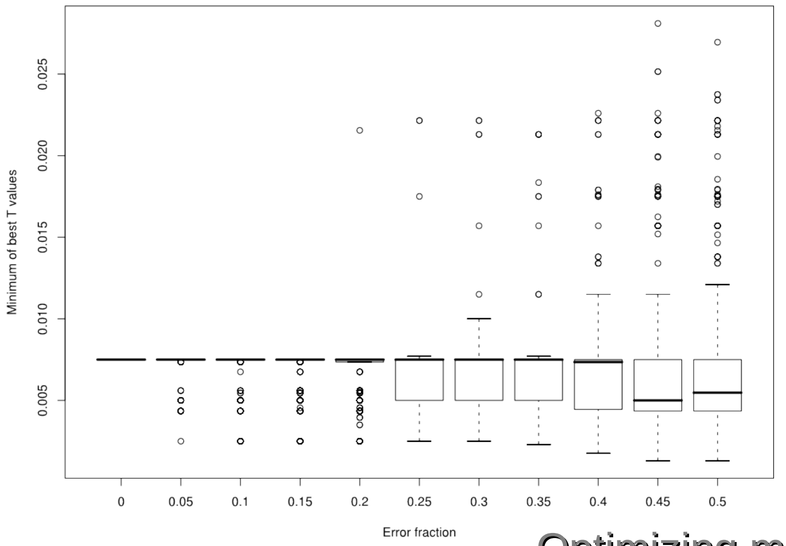
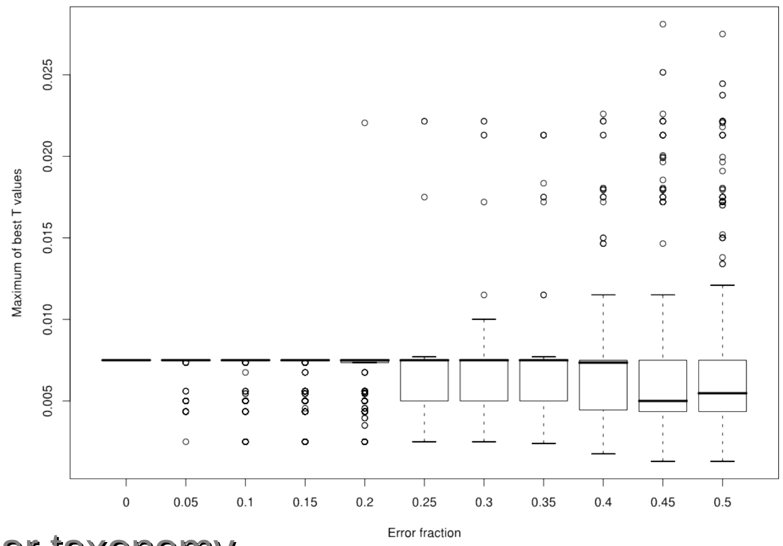


Figure 9



Summary

Clustering optimization based on the agreement between partitions...

- leads to MOTUs with highest agreement to traditional taxonomy, but it is robust against errors in such a reference partition
- connects traditional and modern taxonomic disciplines by specifically addressing the issue of how to optimally account for both traditional species concepts and genetic divergence
- can also be used together with different types of reference partitions (e.g. host species of specialized parasites/mutualists)
- leads to biologically reasonable choices for clustering parameters that are also suitable for sequence identification
- is implemented in the OPTSIL software available at http://www.goeker.org/mg/clustering for all major operating systems