

## Import der Barcodeanalysen für GBOL Rostpilze (Karlsruhe)

### Voraussetzung:

- Die Sammlungsnummern (AccNr.) mit Sammlungsdaten schon vorhanden
- Wirtspflanze als Organismus 1
- Pilz(e) als Organismus 2 (parasitic on Organismus 1), z.Teil als Pilz auf Pilz
- z. Teil mehr als 1 Pilz auf einer Pflanze vorhanden
- Pilzidentifikationen stimmten nicht unbedingt mit den Namen in den Excel Tabellen überein (nachträglich umbestimmt)
- 1 oder 2 Trace forward und/oder reverse in allen möglichen Kombinationen

### Konsequenzen:

- Die Analyse(n) wurden über die AccNr. angehängt („attached“ => Vergleich der AccNr., nicht des“ LastIdentificationName“)
- Nur anhängen der Analyse, kein automatisches Update der Pilz- bzw. Wirtspflanzenbestimmung
- Waren mehr als 2 Pilze vorhanden musste die jeweilige „IdentificationUnitID“ herausgesucht und ergänzt werden, um eine eindeutige Zuordnung zu ermöglichen. Bei Umbestimmungen (Pilzname DC ≠ Pilzname Excel Tabelle) wurde der passendste Pilz ausgesucht (zumindest gleiche Gattung oder eine frühere Bestimmung ist übereinstimmend).
- Im zweiten Schritt wurden zweiten Tracefiles (forward/reverse) als Methoden 4/5 angehängt
- Die Rohdatenfiles (Traces) bzw. der Pfad dorthin wurden als Parameter gespeichert
- Der Pfad zu den Excel Tabellen wurde als Annotation gespeichert.

Übersicht über die Analyse „**Barcode\_Analyse161**“ (mit Methoden und Parametern, zusammengestellt aus verschiedenen Barcodesequenzierungen im Rahmen von GBOL, nicht jede Analyse enthält alle vorhandenen Parameter, diese Analyse gibt es auch in Bonn und München unter dem entsprechenden Namen und der gleichen ID)

### Analyse „Barcode\_Analyse161“

ID: 161  
Methoden: Barcode  
Sequencing  
Taxa: animal, fungus, plant  
No.: Der sequenzierte Genloкус  
Result: der resultierende Barcode

### Methode „Barcode“ (enthält allgemeine Angaben zur Barcodeanalyse)

ID: 12  
MethodenMarker: 1  
Parameter: project [Projektname]

failure	[yes, wenn es einen Fehler bei der Sequenzierung gab]
failure_detail	[Details zur Fehlermeldung, z.B. Kontamination]
region	[der sequenzierte Genlokus z.B. COI, ITS 1, LSU]
sequence_length	[Anzahl der Basenpaare]
trace_count	[Anzahl der sequenzierten Traces]
barcode_compliant	[Verwendbarkeit der Analyse, nur bei 2 oder mehr sequenzierten Traces]
sequence	[der resultierende Barcode]

Methode „Sequencing“ (enthält Details zur Amplifizierung und Sequenzierung)

ID: 16

MethodenMarker: 2-4; muss beim Import angegeben werden, um mehrere Methoden „Sequencing“ auseinanderhalten zu können, muss eindeutig sein

Sequencing 2 – erstes Trace forward

Sequencing 3 – erstes Trace reverse

Sequencing 4 – zweites Trace forward

Sequencing 5 – zweites Trace reverse

Parameter:

- pcr\_primer\_forward\_name
- pcr\_primer\_forward\_sequence
- pcr\_primer\_reverse\_name
- pcr\_primer\_reverse\_sequence
- sequencing\_primer\_name
- sequencing\_primer\_sequence
- sequencing\_timestamp
- sequencing\_labor
- direction [forward oder reverse]
- trace\_filename [nur der Dateiname]
- trace\_file\_url [Pfad zur jeweiligen Tracedatei auf dem Server]
- trace\_file\_org\_length
- trace\_file\_org\_md5
- trace\_file\_encoded
- trace\_file\_encoding [z.B. base64, hex, not\_encoded]
- trace\_file\_enc\_length
- trace\_file\_format [z.B. AB1, SCF]
- trace\_id
- sequence\_id
- well

Zusätzlich:

Import des Pfades zur verwendeten Rohdatendatei (Ausgangs-Excel Tabelle) als Annotation.

Hinweise zu den Importschemata („baukastenweise“, jeweils für die jeweiligen Importtabelle anzupassen)

Vorarbeit:

- Ergänzung der Ausgangs-Excel Tabelle um die Spalten „interne Bemerkungen“, „Dateiname“ und „IdentificationUnitID“
- Ausfüllen der Spalte „Dateiname“ mit jeweiligem Dateiname
- Kontrolle der Tracefilenamen, ob Unterstriche statt Leerzeichen verwendet wurden und auf vorhandene Dateiendung (.ab1/.scf)
- ggf. die Absatz
- abspeichern als „Unicode-.txt-Datei“

einige Einstellungen im Import Schema:

Anhängen („attach“) der Analysedaten an die vorhandenen (Sammlungs-)Daten (immer an den zweiten Organismus also den Pilz):

- über **AccessionNumber** (geht solange nur ein Pilz der Pflanze zugeordnet ist), zusätzlich Vergleich ob „TaxonomicGroup“ = „fungus“ ist
- über **IdentificationUnitID** von Organism 2 (muss bei den Fällen herausgesucht werden, bei denen es zwei oder mehr Pilze auf einer Wirtspflanze gibt)

Die übrigen Sammlungsdaten werden nicht verändert (Specimen, Project, Organism1 und Organism2 stehen auf Attach)

#### Methode 2.1.1 *Barcode1*

MethodenID = „12“ (nur Import, wenn ein Barcode in Spalte 19 vorhanden ist)

MethodenMarker = „1“ (nur Import, wenn ein Barcode in Spalte 19 vorhanden ist)

#### Methode 2.1.2 *Sequencing 2*

MethodenID = „16“

MethodenMarker = „2“

direction = forward

alle Angaben aus den Spalten F-T der Ausgangs-Excel Tabelle den jeweiligen Parametern zuordnen

Parameter „sequencing\_primer\_name“= „seq\_for\_primer\_name“

Parameter „sequencing\_primer\_sequenz“= „seq\_for\_primer\_seq“

Parameter „Trace\_filename“= „Rohdatenfile fwd 1“

#### Methode 2.1.3 *Sequencing 3*

MethodenID = „16“

MethodenMarker = „3“

direction = reverse

alle Angaben aus den Spalten F-T der Ausgangs-Excel Tabelle den jeweiligen Parametern zuordnen

Parameter „Trace\_filename“= „Rohdatenfile rev 1“

Parameter „sequencing\_primer\_name“= „seq\_rev\_primer\_name“

Parameter „sequencing\_primer\_sequenz“= „seq\_rev\_primer\_seq“

#### Methode 2.1.4 *Sequencing 4*

MethodenID = „16“

MethodenMarker = "4"

direction = forward

alle Angaben aus den Spalten F-T der Ausgangs-Excel Tabelle den jeweiligen Parametern zuordnen

Parameter „Trace\_filename“= „Rohdatenfile fwd 2“

Parameter „sequencing\_primer\_name“= „seq\_for\_primer\_name“

Parameter „sequencing\_primer\_sequenz“= „seq\_for\_primer\_seq“

#### Methode 2.1.5 Sequencing 5

MethodenID = „16“

MethodenMarker = "5"

direction = reverse

alle Angaben aus den Spalten F-T der Ausgangs-Excel Tabelle den jeweiligen Parametern zuordnen

Parameter „Trace\_filename“= „Rohdatenfile rev 2“

Parameter „sequencing\_primer\_name“= „seq\_rev\_primer\_name“

Parameter „sequencing\_primer\_sequenz“= „seq\_rev\_primer\_seq“

Bei Fehlermeldungen „Fungus not unique“ muss für die jeweilige Sammlungsnummer die „IdentificationUnitID“ herausgesucht werden und in die von DiversityCollection erzeugte „Errortexttabelle“ eingefügt werden. Laden der jeweiligen Basis-Schema, Anpassung des Importschemas (auf Vergleich mit IdentifikationUnitID, s. Abbildungen) und dann können diese Belege importiert werden.

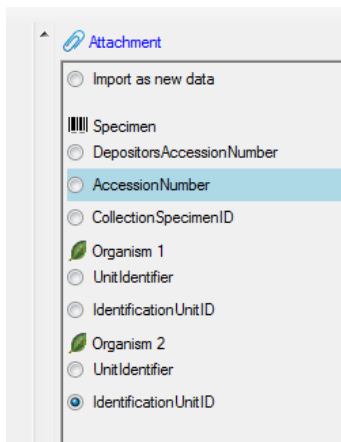
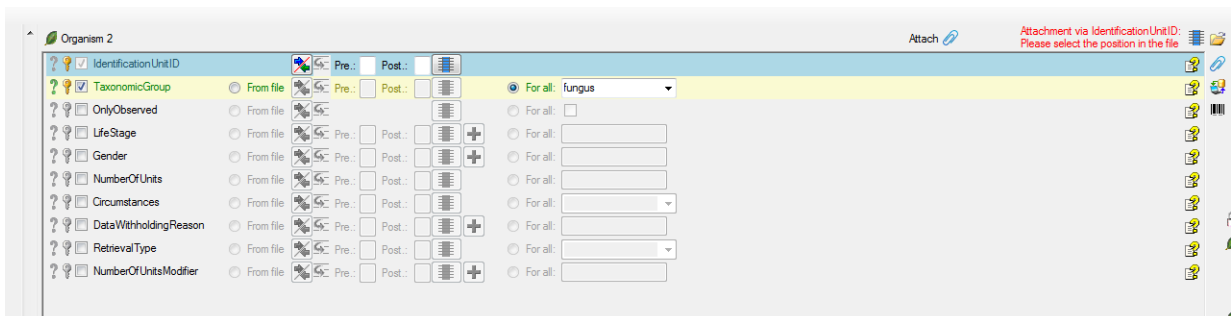


Abb1) Ändern des Attachment von AccessionNumber auf Organism2 / IdentificationUnitID

Abb2) Ansicht bei Organism2 vor der Änderung



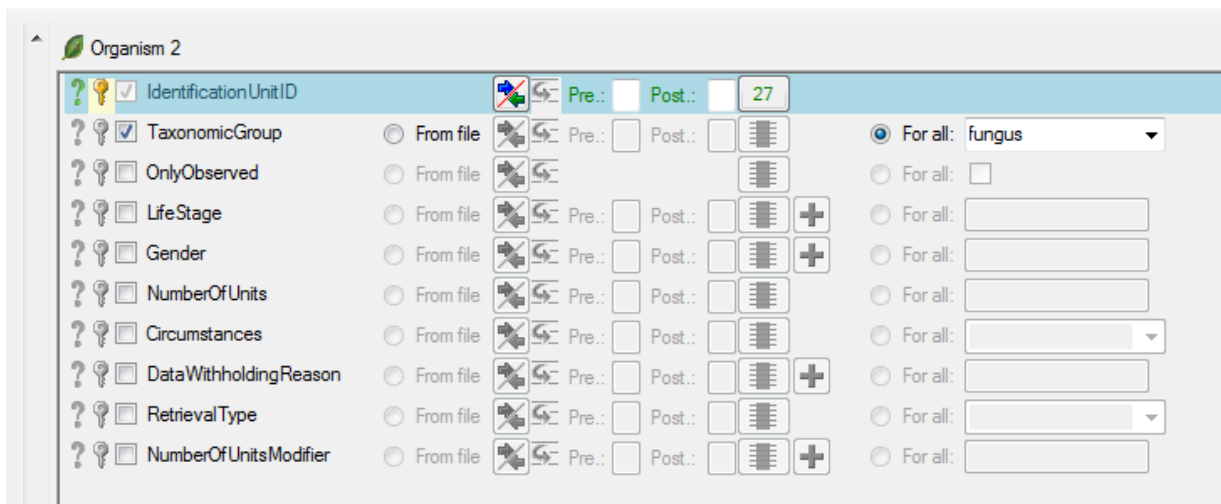


Abb3) Einstellung für Attachment mit IdentificationUnitID (auswählen der Spalte und ändern des Vergleichs)

Import der Methoden 2.1.4 und 2.1.5 in einem zweiten Schritt, da sonst der Import Wizard überlastet ist (mit Schema: 02\_add\_second\_fwd\_or\_rev\_trace.xml)

verwendete Importschemata (nachträglich zusammengefasst):

- 01\_basis\_barcode\_import\_scheme.xml  
(attach by AccNr.; Methoden Barcode 1, Sequencing 2 und 3)
- 02\_add\_second\_fwd\_or\_rev\_trace.xml  
(attach by AccNr.; ohne Annotation; Methoden Sequencing 4 und 5)

Acc. Nr. **KR-M-0001681** **Melampsora laricis-populina Kleb.** Specimen ID: 12865 Event 9976 Version 4 Withhold. 20 Context: General

2006/3/15 Neustheim, Gewann "Niederfeld"  
MTB: 6517, Quadrant: 1  
Altitude (mNN): 90 m NN  
DE, Baden-Württemberg, Universitätsstadt Mannheim  
Coordinates WGS84 (Longitude / Latitude) Long. (EW): 8.538611111111111 Lat. (NS): 49.4677777777778

**KR-M-0001681**  
Populus spec. [respons.: H. Staub]  
Melampsora laricis-populina Kleb.  
Melampsora laricis-populina Kleb. [respons.: M. Scholler]  
Barcode\_Analyse161 ITS: CTTGGTCATTAGAGGAAGTAAAGTCG  
H. Staub  
P:\GFBio\GBOL\_Rostpilze\1\_Raw\GBOL\_8\_Karlsruhe.xls

**Melampsora laricis-populina Kleb.**  
Tax. group: fungus = Eumycota: mushrooms, sac fungi, ye No. of units: Only obs  
Gender: Substr. rel.: Parasitic on = the taxon is parasitizing its host (regardless of type...) ID 450452  
Family: Order:  
Identifier: Descript.:  
Life stage: TH Circumst.:  
Colon part: Note: PFR  
Retrieval:

Barcode\_Analyse161 ITS: CTTGGTCATTAGAGGAAGTAAAGTCGTAACAAGGTTCCGTAGGTGAACCTGCGGAAGGATCATTATAC  
No.: ITS Result: CTTGGTCATTAGAGGAAGTAAAGTCGTAACAAGGTTCCGTAGGTGAACCTGCGGAAGGATCATTATAC  
Note: Ext. URI:  
Date: Part:

Respons.: B. Bubner

Methods Tools

Barcode	Direction
Sequencing 2	por_primer_forward_name por_primer_forward_sequence por_primer_reverse_name por_primer_reverse_sequence sequencing_labor sequencing_primer_name sequencing_primer_sequence trace_file_url trace_filename well

**KR-M-0001681**  
Melampsora laricis-populina Kleb.  
Populus spec.

Abb4) Vorschau der importierten Daten (bei einem Pilz)

Acc. Nr. **KR-M-0023897** **Tranzschelia discolor (Fuckel) Tranzschel & M.A. Litv.** Specimen ID: 28684 Event 25795 Version 5 Withhold. 20 Context: General

2009/10/19 Grünwinkel, Absiedlung, Abinsel  
MTB: 6916, Quadrant: 3  
Altitude (mNN): 108 m NN  
DE, Baden-Württemberg, Karlsruhe  
**KR-M-0023897**  
Prunus spinosa L. [respons.: M. Scholler]  
Tranzschelia discolor (Fuckel) Tranzschel & M.A. Litv.  
Tranzschelia discolor (Fuckel) Tranzschel & M.A. Litv. [respons.: M. Scholler]  
Barcode\_Analyse161 ITS2 LSU: TAGTAACGGGAGTGAAGAGGAAAAGCCAAATTGTAATCTGGCTCTTTAAGAGTCTGAGTTGTAATTT  
Erysiphe prunastri DC. [respons.: M. Scholler]  
M. Scholler  
P:\GFBio\GBOL\_Rostpilze\1\_Raw\ML\_barcode6.xls

**Tranzschelia discolor (Fuckel) Tranzschel & M.A. Litv.**  
Tax. group: fungus = Eumycota: mushrooms, sac fungi, ye No. of units: Only obs  
Gender: Substr. rel.: Parasitic on = the taxon is parasitizing its host (regardless of type...) ID 454654  
Family: Order:  
Identifier: Descript.:  
Life stage: TH Circumst.:  
Colon part: Note: PFR  
Retrieval:

Barcode\_Analyse161 ITS2 LSU: TAGTAACGGGAGTGAAGAGGAAAAGCCAAATTGTAATCTGGCTCTTTAAGAGTCTGAGTTGTAATTT  
No.: ITS2 LSU Result: TAGTAACGGGAGTGAAGAGGAAAAGCCAAATTGTAATCTGGCTCTTTAAGAGTCTGAGTTGTAATTT  
Note: Germany Ext. URI:  
Date: Part:

Respons.: C. Aime

Methods Tools

Barcode	Direction
Sequencing 2	por_primer_forward_name por_primer_forward_sequence por_primer_reverse_name por_primer_reverse_sequence sequencing_labor sequencing_primer_name sequencing_primer_sequence trace_file_url trace_filename well

**KR-M-0023897**  
Tranzschelia discolor (Fuckel) Tranzschel & M.A. Litv.  
Tranzschelia discolor (Fuckel) Tranzschel & M.A. Litv.  
Prunus spinosa L.  
Erysiphe prunastri DC.

Abb5) Vorschau der importierten Daten (bei zwei Pilzen)

## Übersicht der Spalten in der Ausgangstabelle und zugehörigen Paramtern in DC

Spalte in Excel-Tabelle	Parametername in DC (Analyse, Methode, Specimen)	Beispieldatensatz
Beleg	AccessionNumber (Specimen)	KR-M-0023897
Gattung	LastIdentificationCache (Organism2)	Tranzschelia
Art	LastIdentificationCache (Organism2)	discolor
Wirtsgattung	LastIdentificationCache (Organism1)	Prunus
Wirtsart	LastIdentificationCache (Organism1)	spinosa
well	well (Sequencing 2,3,4,5)	
pcr-for-primer-name	pcr_primer_forward_name (Sequencing 2,3,4,5)	Rust2inv
pcr-for-primer-seq	pcr_primer_forward_sequenz (Sequencing 2,3,4,5)	GATGAAGAACACAGTGAAA
pcr_rev-primer-name	pcr_primer_reverse_name (Sequencing 2,3,4,5)	LR5
pcr-rev-primer-seq	pcr_primer_reverse_sequenz (Sequencing 2,3,4,5)	TCCTGAGGGAAACTTCG
seq-for-primer-name	seq_primer_name (Sequencing 2, 4)	Rust2inv
seq-for-primer-seq	seq_primer_sequenz (Sequencing 2, 4)	GATGAAGAACACAGTGAAA
seq_rev-primer-name	seq_primer_name (Sequencing 3,5)	LR5
seq-rev-primer-seq	seq_primer_sequenz (Sequencing 3,5)	TCCTGAGGGAAACTTCG
Region	region (Barcode 1) No. of Analysis (Analysis 2.2)	ITS2, LSU
Sequencing-Lab	sequencing_labor (Sequencing 2,3,4,5)	C. Aime
Edited by	ResponsibleName (Analyse 2.1)	C. Aime
Annotations	Notes (Analyse 2.1)	Germany
Problems	Failure & failure_detail (Barcode 1)	
Barcode-Sequenz	Result (Analyse 2.1) Sequenz (Barcode 1)	TAGTAACGGCGAGTGAAGA GGGAAAAGCCCATTGTA[...]
Rohdatenfile fwd 1	trace_filename (Sequencing 2, 4) trace_file_format (nur Endung) trace_file_url (+Pfad)	Plate_60_Aime_E03_F.SCF
Rohdatenfile fwd 2	trace_filename (Sequencing 3, 5) trace_file_format (nur Endung) trace_file_url (+Pfad)	Plate_60_Aime_D11_F.SCF
Rohdatenfile rev 1	trace_filename (Sequencing 2, 4) trace_file_format (nur Endung) trace_file_url (+Pfad)	Plate_60_Aime_D12_F.SCF
Rohdatenfile rev 2	trace_filename (Sequencing 3,5) trace_file_format (nur Endung) trace_file_url (+Pfad)	Plate_60_Aime_E04_F.SCF
Withhold	DataWithholdingReason (Specimen)	
Interne Bemerkung		
Dateiname	Annotation (+ Pfad) (Specimen)	ML_barcode_6.xls
IdentificationUnitID	IdentificationUnitID (Organism2)	454654