Geneious Plugin – Metadata Importer

Jerry Cooper (Systematics) & Tim Heuer (Informatics), Landcare Research, New Zealand
Version 1.05, December 2016

Purpose

The Metadata Importer (MI) is a plugin component for Geneious versions 9 and later.

The MI is intended to facilitate the importation of meta-data associated with sequenced organisms for downstream analysis and submission of data to GenBank.

The MI allows the user to:

1. Specify a delimited text file of data fields for import
2. Specify the delimiter character in the import file
3. Specify a key pair which uniquely links a column in the import file to a field (standard or user-defined) in a selected set of sequence documents
4. Specify a mapping between import file columns and sequence document fields
5. Copy data from the import file to the sequence documents based on the linked column/field
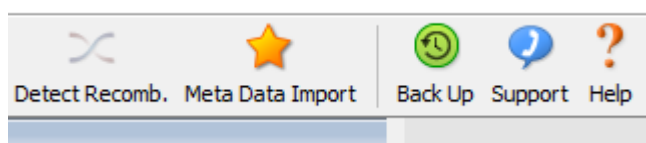
Installation

Save the installation file with the extension gplugin (e.g. GeneiousmetadataImporter-1.05.gplugin).

Install the plugin using the Geneious menu option Tools -> Plugins …

Remember to restart Geneious to activate the plugin.

The MI plugin should appear as an option in the top toolbar.

Use

Prepare a delimited text file containing the metadata to be imported. The file should conform to RFC4180 (https://www.ietf.org/rfc/rfc4180.txt)

Tab stops are the preferred delimiter, rather than commas, because of the non-standard way in which many applications deal with in-field delimiter characters when writing CSV documents formats.

Ensure your metadata file has headings in the first row.

```
qryExportForgeneiousImport.txt - Notepad
File  Edit  Format  View  Help
Name of Fungus   ColID   Association    Date      Site name       country Herbarium      PDDNum  CultureC
Gymnopus            JAC8119        13-Jan-2001    Arthurs Pass    New Zealand     PDD 76639       76639
Laccaria lateritia cf.   JAC8575 Pinus radiata   20-Apr-2003     Bottle Lake, Christchurch      New Zeal
Laccaria paraphysata     JAC8620 Kunzea ericoides      25-May-2003      Maruia, Hut       New Zealand
Rhodocollybia incarnata JAC8621 Kunzea ericoides      25-May-2003      Maruia, Hut       New Zealand
Gymnopus sp. 'Craigieburn (PDD95664)'   JAC8763 Nothofagus fusca       27-Dec-2003      Woolshed Hill Tr
Gymnopus rimutaka        JAC8885        14-Mar-2004    Bealey Chasm Track     New Zealand     PDD 7986
Laccaria glabripes ss    JAC9047 Nothofagus menziesii  15-May-2004    Doughboy Road, track    New Zeal
Laccaria laccata var. pallidifolia      JAC9062 Quercus robur   29-May-2004     Little Hagley Park, Chri
```

The plugin will only function when one or more sequence documents are selected. Then specify the file containing metadata.

You are then asked to specify the delimiter character. The importer does not try to guess the file format.
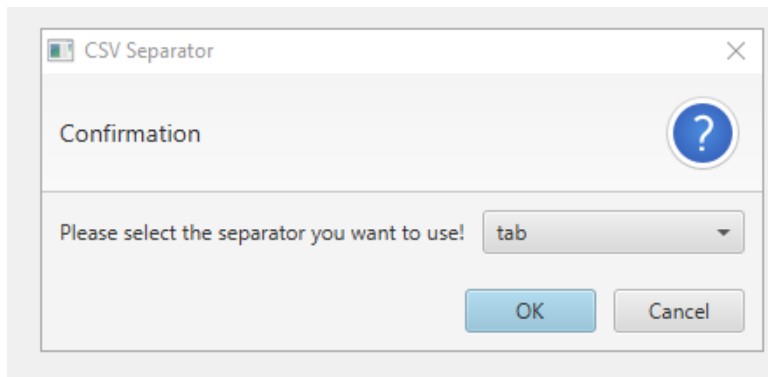
The next stage is to map columns in the import delimited text file to available fields in the set of selected sequence documents, and to specify a key pair which can be used to link the data. If you don't see a sensible looking form, or the plugin aborts, then check that the first line of your import file contains a header.

In this example the import text file contains a column with the heading ColID which is mapped to a sequence document field also called ColID, and this is the key field. The key field is selected using the radio buttons. Ideally the import file should contain unique records in the data for this column. The resulting transfer of data is then based on a one-to-one or one-to-many join between the import file and the sequence documents using this key link. If the import file contains multiple records with the same key then the user is prompted to perform the join on either the First or the Last records which are identical in the key column.

In this example the 'Name of Fungus' column in the import file contains data which is copied to the 'Organism' fields in the sequence documents. Similarly the 'Country' column in the import file contains data which is copied to the 'Locality' fields in the sequence documents.

The contents of the target fields are checked for any existing content. If the import would result in overwriting content then the user is prompted "Do you want to overwrite populated field (Y) or skip them (N)". If you answer Yes then populated fields in the selected sequence documents will have their content overwritten. If you select No then data will still be copied for all records and all fields except where it would result in field content being overwritten.

The import process is then initiated. If the text file is large and many sequence documents are selected then importation may take a few seconds. The plugin does not display a process indicator. Wait until the dialogue indicates importation has been completed and then close the plugin.

Key	Imported field	Copy to sequence document field
○	Name of Fungus	Organism | The organism of the document ▾
●	ColID	ColId (CollectionIdentifier) | ▾
○	Association	▾
○	Date	▾
○	Site name	▾
○	country	Locality | Locality ▾
○	Herbarium	▾
○	PDDNum	▾
○	CultureCollection	▾
○	LatLon	▾
○	Collector	▾
○	Determiner	▾

Message ✕

Message ⓘ

Import SUCCESSFUL.

OK

Data should have been copied from the import file to the sequence documents.

Meta Data Import   Back Up   Support   Help

		Name	Organism ▾	ColId (CollectionIdent...	Description
☑		-	Phellodon sinclairii #1	jac14155	2 reads from jac14155
☑		-	Phellodon plicatus #3	jac14092	2 reads from jac14092
☑		-	Phellodon plicatus #2	jac14077	2 reads from jac14077
☑		-	Patellaria cf.	jac14134	2 reads from jac14134
☑		-	Hydnum crocidens var. wellingtonii	jac14173	2 reads from jac14173
☑		-	Hodophilus	jac14165	2 reads from jac14165
☑		-	Gliophorus pallidus	jac14082	2 reads from jac14082
☑		-	Galerina subcerina	jac14123	2 reads from jac14123
☑		-	Deconica merdaria cf.	jac14216	2 reads from jac14216
☑		-	Clavulinopsis archeri	jac14083	2 reads from jac14083
☑		-	Claussenomyces	jac14131	2 reads from jac14131
☑		-	Clathrus archeri	jac14185	2 reads from jac14185
☑		-	Chlorencoelia torta	jac14135	2 reads from jac14135
☑		-	Chlorencoelia torta	jac14068	2 reads from jac14068
☑		-	Cheimonophyllum roseum	jac14183	2 reads from jac14183
☑		-	Austroboletus niveus	jac14213	2 reads from jac14213

Caveats

The importer will only work with the Java JRE shipped with versions 9 and over. You may try replacing the JRE with earlier versions of Geneious.

The import file needs to be well-formed and with header columns in the first row. Any deviation may result in the plugin aborting (by reporting a problem) or throwing an error. We have tried to trap most of the possible errors but some may have escaped. If the plugin produces an error message then please **do not** select the option to Report the problem to the Geneious Development team. The problem will be ours, not theirs. Send Jerry an email and we will see if there is a possible fix. The plugin was developed for our in-house use and we cannot actively support the wider use. However, we would like to make the plugin available to others.

Terms of Use

The authors and Landcare Research do not accept responsibility for any consequences arising from the use of the Geneious Metadata Importer plugin.

The plugin (and associated code on request) is made available under Creative Commons Attribution-NonCommercial-No Derivative Works 4.0 New Zealand License (BY-NC-ND)

Jerry Cooper

CooperJ@LandcareResearch.co.nz