**Taller GBIF.ES: Bases de datos de colecciones y GenBank (Marzo2014)**

Parte práctica

Resumen

Vamos a crear los ficheros necesarios para hacer linkout desde GenBank a nuestra base de datos de la colección en línea.

Ejercicio paso a paso:

1. Crear el "identity file". Para ello usaremos el fichero de ejemplo, "[providerinfo.xml](http://www.gbif.es/ficheros/TallerGenBank2014/providerinfo.xml)", lo editaremos con Notepad++, con nuestra información.
   1. A tener en cuenta:
      1. El número de ProviderId nos debe ser proporcionado GenBank, a efectos del ejercicio introduciremos XXXX, tal como se muestra a continuación:

|  |
| --- |
| Machine generated alternative text: <Providerld>)000k/Providerld> |

1. El nombre de este fichero siempre debe ser "providerinfo.xml"
2. El "free" de "Attribute" quiere decir que el acceso a nuestra base de datos no tiene restricciones, no necesita usuario + contraseña:

|  |
| --- |
| Machine generated alternative text: <Àttribute>Free</Attribute> |

1. Información detallada sobre cómo prepara este fichero, fundamentalmente qué información tiene cabida dentro de "Attribute", se encuentra en:

<http://www.ncbi.nlm.nih.gov/books/NBK3802/#nonbib.File_Preparation_Identity_File>

1. Preparar los ficheros de "resources", estos son los necesarios para hacer linkout desde la BBDD de nucleótidos y desde la BBDD de taxonomía. Los dos ficheros son muy similares, y el proceso es el mismo en ambos casos variando su contenido en detalles mínimos. Por ello el ejercicio lo haremos sólo para hacer linkouts desde la BBDD de nucleótidos (.i.e., como en el fichero de ejemplo "[MAFungiGenbank.ft](http://www.gbif.es/ficheros/TallerGenBank2014/MAFungiGenbank.ft)").
   1. Rellenar una tabla con nuestros linkouts. Para ello utilizamos la de ejemplo, que es la tabla "MA-Fungi en GenBank" de la BBDD de MS-Access "[Ejemplo tabla linkout.mdb](http://www.gbif.es/ficheros/TallerGenBank2014/Ejemplo-tabla-linkout.mdb)". Haremos:
      1. Abrir la tabla mencionada, examinar cómo está rellena, borrar los registros que tiene (pero dejar uno de guía).
      2. Añadir los registros propios, se sugiere hacer búsquedas en GenBank por las siglas de nuestra colección para ello:

Machine generated alternative text: NCBI Resources (3_How To 19
Display Settings: (3 GenBank
oniophora and .8S rRNA gene, 28S rRNA gene, internal
trììšii5fi5er I (ITSI) and internal transcribed spacer 2
(ITS2), specimen_voucher MAFung1477O8 )
Gen5ank AJ4 19196 1
FASTA Graphics
Go to: O
J
Horn e
LI__J Lt, Cc
Pact
T   Ascending
Descending !Adv’
Remove Sort i( Tog
nrt P., FiIt.r
LOCUS
DEFINITION
ACCESSION
VERSION
KEYW)RDS
S OURCE
OR3AN I SM
RE FE RENDE
AJ419196 585 bp DNA linear PLN 17—FEB—2011
Coniophora anda 5.8S rRNA gene 285 rRNA gene, internal
transcribed spacer 1 (ITS1) and internal transcribed spacer 2
MA-Fungi 47708
AJ4 19196
A4 6.1 GI:40642631
28S nibosczal RNA; 285 rRNA gene; 5.8S riboscal RNA; 5.85 rRNA
gene; internal transcribed spacer 1; internal transcribed spacer 2;
ITS1; 1T52
Coniophora anda
Ccnichora crida
Eukaryota; Fungi; Di3carya; Basidiotnycota; Agaricoycotina;
Aganiccmycetes; Aganiccmycetidae; Boletales; Coniophorineae;
Ccnicçhcraceae; Coniophora
1
Martin,M.P. and Raidl,S.
Id
Fcnprip e
‘oniophora anda
2 Hypochnicium guineensis
3 Hypochnicium michelii
4 Hyphodermella rosae
5 Hyphodermella rosae
e
6 Hyphodermella rosae

* + 1. Crear por lo menos 3-4 linkouts (i.e., registros en la tabla)

1. El campo "TextoGenBank" se puede rellenar con una consulta de "actualización de datos" a partir del campo "AccGenBank"
2. Se recomienda rellenar el campo responsable con los datos de quién ha proporcionado la información del "linkout"; la práctica recomendada es que el propio investigador que envía la secuencia al GenBank, proporciona esta información a la colección. Alternativamente y de manera paliativa, el personal de la colección puede tratar de recupera esta información haciendo búsquedas en GenBank. En estos casos se recomienda encabezar la información con la frase "ex officio" ("desde la oficina"), para indicar que se ha hecho en la colección. Por ejemplo:

"ex officio: J. López, 2014"

1. Utilizando las funcionalidades de "Combinar correspondencia" ("Mail Merge") de MS-Word, creamos un documento dinámico para --a partir de la tabla que hemos rellenado en el paso anterior-- generar el texto de la parte de "linkouts”). cf.:

Machine generated alternative text: Lin kouts para Ia BBDD de Nucleótidos
r
j( C:\Datos\Projects\GBLF\GB1F.ES\Formacion\2O14\CoIeccc I dent if j c a do r dei proveedor,
File Edit Search View Encodin
asignado por NCBI
I   1-I——Ji 1
BBDD desde la que se hace “LinkOut”
II
Encabezamiento 2 prid: 4283
Tipo de “LinkOut”
3 dbase: nucleotide
4 ste: culture/stock collections
I L 5 ‘base: http://161.111.170.202/herb/asp/dos.asp?ccnque=YA-Fungi&
J 6 ________________
7 linkid:  Parte constante de ias URL de
8 query: AJ419196 [accn] “LinkOut”
9 base: &base;
10 rule: fnhe=47708 II
11
___________________ Núm. de “LinkOut”; único en ei
12 linkid: 2
13 query: FN552536 [accn] fichero para cada Linkout
“linkOuts” 14 base: &base;
II .
15 rule: fnhe=79156
16
18 query: FN552535[ “Accession number” de Genbank desde
ei que se hace ei Linkout
17 linkid: 3
II
19 base: &base;
20 rule: fnhe—79155
Parte variable de la URL que hace I
referencia al ejemplar dei Linkout
21 ___________________
length :42386 In : 1 Col : 1 Sel :0 Dos\ Windows ANSI INS

1. Copiamos el texto del documento de MS Word generado y lo pegamos en un fichero de texto en blanco que creamos desde el Notepad++
2. Abrimos "[MAFungiGenbank.ft](http://www.gbif.es/ficheros/TallerGenBank2014/MAFungiGenbank.ft)" con el Notepad++, copiamos el encabezamiento (filas 1-6) y lo pegamos en el fichero de texto que hemos creado en el paso "d”; por delante del texto que ya habíamos pegado desde el documento de MS Word.
3. Modificamos el encabezamiento para ajustarlo a nuestros datos; esto supone cambiar el "prid" (a poner aquí nuestro "ProviderID"); y la parte constante de las URL que dan acceso a nuestros ejemplares de la base de datos en línea (línea 5, a cambiar lo que va detrás de "!base: ").
4. En el caso de no tener nuestra colección en línea, podemos utilizar como subrrogado temporal nuestros registros en GBIF; los cambios necesarios para ello se ilustran en el siguiente gráfico:

Machine generated alternative text: [( 9(AColecciones y GenBank\MA-Fungi-Genbank.ft - Notepad.+
File Edit Search View Encoding Language Settings Macro Run Plugins Window ?
! o8   
I M#Fingi-Accessonsft I  MA-FurdGenb*ft  p unnt x’iI  MA-FI.ngl-Genbaridtl
1
2 prid: XXXX
3 dbase: nucleotide
4 stype: culture/stock collections
5 !base: http://jsonviewer. stack.hu/#http://api.gbif.org/vO. 9/occurrence/search?institutionCode=MNCN&co1lectionCode=Q1CN-ADN&
6
7 linkid: 1
query: AJ419196 (accnj
base: &base;
rule: cataloqNuber—23158
11
12 linkid: 

Ya solo queda guardar el fichero y cambiar su extensión de "txt" a "ft", tal como recogen las instrucciones de GenBank al respecto (cf. <http://www.ncbi.nlm.nih.gov/books/NBK3802/#nonbib.File_Preparation_Resource_File__2> )

1. El ejercicio acaba aquí. El procedimiento para preparar linkouts para taxonomía es semejante. En la vida real este proceso se culmina subiendo los ficheros de "provider info" y de “resources” creados por FTP al repositorio indicado por GenBank (con el providerID que nos dieron, y poniéndolos en la carpeta de "holdings".