Unique Identifiers in an Online Permit and Monitoring System to support the Nagoya Protocol

Paul Oldham
Hartmut Meyer
Olivier Rukundo

26 May 2016

Introduction

This section discusses the use of unique identifiers for permits and related documents and the application of the principle of triple redundancy.[1]

Unique identifiers associated with permits and associated documents will allow for sets of documents linked to a specific application to be stored within a retrievable register or file history. As such, unique identifiers will enable internal coherence within the permit system.

Unique identifiers linked to a permit and corresponding MAT (an ABS contract) are also the key to monitoring the outcomes of research and the genetic resources. We propose a system consisting of the following unique identifiers:

1. Country Codes, dates and unique numbers (e.g. BS20151234).
2. Bar codes.
3. QR Codes (Quick Response codes).
4. Html embed codes.

In proposing the use of unique identifiers we also suggest the application of the principle of triple redundancy, notably in labelling. The principle of triple redundancy is a well established engineering principle that involves three different systems performing the same function. In the event that one system fails a second system takes over the same function. If the second system fails then the third system takes over. Given that the failure of all three systems is unlikely the function of a particular system within a wider system is maintained. It is important to note that triple redundancy does not guarantee that a system will not fail completely, rather it reduces the likelihood of failure.

The principle of triple redundancy is particularly relevant to an online permit and monitoring system in connection with the use of unique identifiers to maintain the link between a permit, mutually agreed terms and samples of biological material that are subsequently deposited with a collection, transferred to a collection in a third country and potentially utilized by third parties.

The problem that triple redundancy helps to address is in ensuring the maintenance of the connection between a permit and a document containing MAT (an ABS contract) and the materials that are collected and transferred. These materials may be either physical or electronic (e.g. DNA and amino acid sequence data). The scope of genetic resources is likely to be defined in domestic ABS legislation.

[1]The research in this paper was conducted with the support of The Bahamas Environment, Science & Technology Commission (BEST) of the Government of the Bahamas under the UNEP/GEF project "Strengthening Access and Benefit Sharing (ABS) in the Bahamas" as set out in *Oldham, P (2015) Concepts for an Electronic Monitoring Tool. UNEP/GEF project "Strengthening Access and Benefit Sharing (ABS) in the Bahamas"*. The present paper was written with the additional support of The ABS Capacity Development Initiative through the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ). The views expressed are solely those of the authors and should not be interpreted as reflecting the views of the Government of the Bahamas, the ABS Initiative or GIZ.

Country Code Identifiers

All countries possess a unique two letter country code defined in international standard ISO3166-1 alpha 2 (e.g. see this Wikipedia entry and the ISO browsing platform and select Country codes and search).

For example, the standard two letter country code for the Bahamas is BS, Kenya is KE, Uganda is UG and South Africa is ZA. If these country codes are combined with a date (YYYYMMDD) and a unique number (1234) a unique identifier will be generated. In this case we will simply use the year (YYYY) to generate the following identifiers.

- BS20151234
- KE20151234
- UG20151234
- ZA20151234

What is clear from this is that a single and distinctive unique identifier (country - year - number) has been created for each permit and corresponding MAT that is immediately distinguishable using the country code from similar numbers issued by other authorities. The combination of the country code, the date and a numeric identifier (country - YYYMMDD - number) is already used very successfully to keep track of approximately 90 million patent documents in countries around the world and is recommended. The example below is for a patent application from the United States that makes reference to collections from the Bahamas and can be viewed on the main worldwide patent database espacenet operated by the European Patent Office here.

In this example we can see that the unique document identifier US2001049387A1 consists of the following [US] – the country code – [2001] the year of application – [049387] – the unique numeric identifier, and [A1] known as the "kind code" for the type of document (in this case a patent application).

The strength of this system is that the components combine into a unique identifier that is:

a. distinctive, and;
b. easy to retrieve.

The relevance of this type of numbering system becomes apparent when we consider the section of this patent application that makes reference to biological collections in a country. In this case the example is from the Bahamas (See the first paragraph highlighted or read directly here).

In this case, the applicants make reference to a sponge specimen 23-XI-98-3-002 and a HBOI CatNo. 003:00973. Under the proposal advanced in this document the ABS domestic legislation, permit and associated MAT could require recipients of a permit to disclose the permit number (e.g. BS20151234) in the description section of any patent application arising from the research.[2].

Unique Identifiers and Registers

An important feature of this system as deployed within patent databases is that the unique identifier is used to store all documents that relate to a particular application over time. In this case the United States application was also filed in Europe at the European Patent Office and this document can be identified in the patent family of the US document. European Patent Application EP1259502A2 was filed prior to the introduction of the year into the identifier. However, for our purposes the document is important because in Europe it is possible to access all documents linked to that application, including formal communications between the patent office and applicants in the European Patent Register. The figure below displays the list of documents linked to this identifier and can be accessed here.

It is immediately clear that the use of this unique identifier system allows all documents related to that identifier to be linked together into an electronic register (file history) for each application. We propose a

[2]See Oldham and Burton (2010) Defusing Disclosure in Patent Applications, UNEP/CBD/COP/10/INF/44 and Oldham, Hall and Forero (2013) Biological Diversity in the Patent System. PLOS ONE)

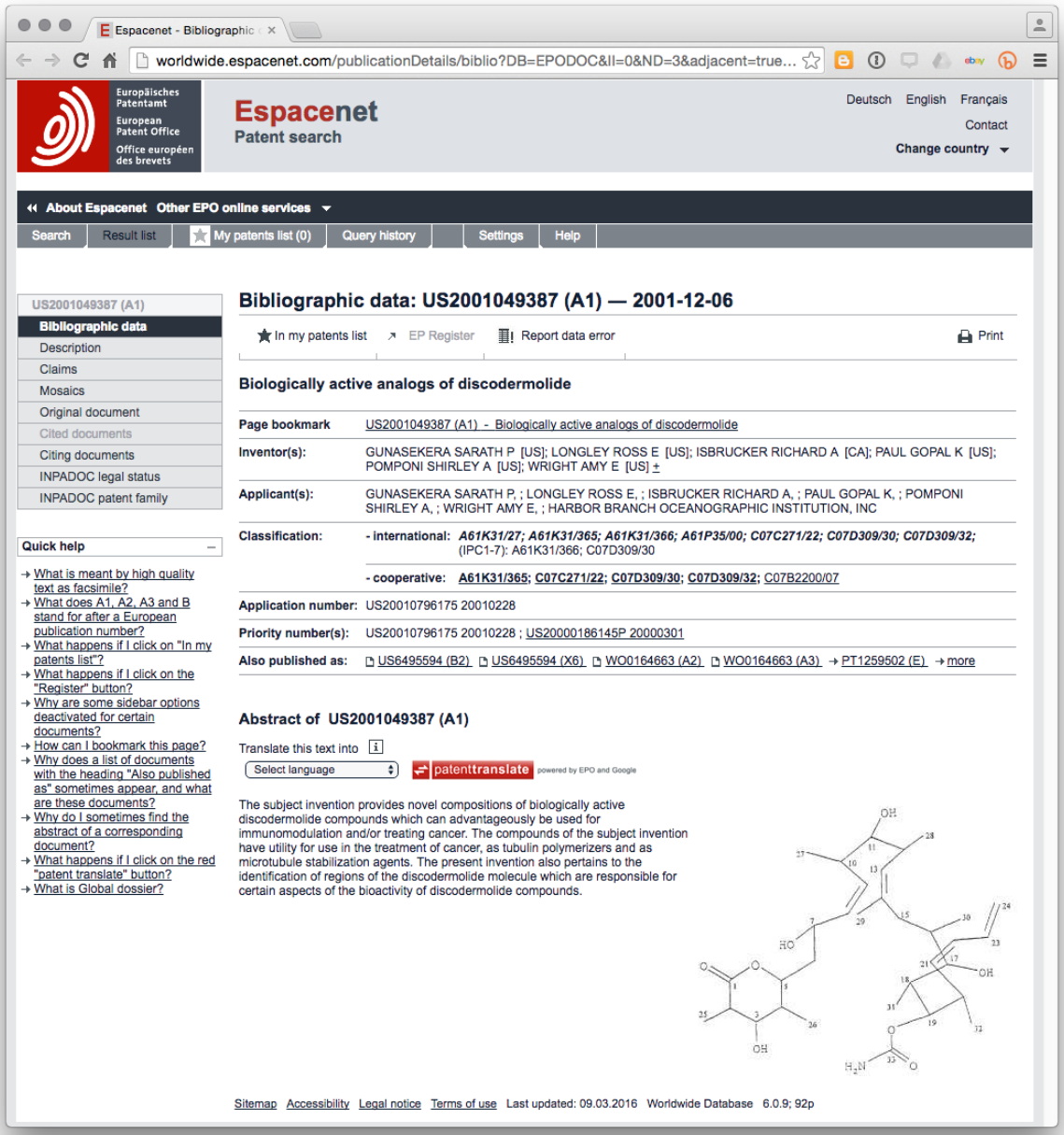
Espacenet - Bibliographic
worldwide.espacenet.com/publicationDetails/biblio?DB=EPODOC&II=0&ND=3&adjacent=true...
Europäisches Patentamt
European Patent Office
Office européen des brevets
Espacenet
Patent search
Deutsch English Français
Contact
Change country
About Espacenet Other EPO online services
Search Result list My patents list (0) Query history Settings Help
US2001049387 (A1)
Bibliographic data
Description
Claims
Mosaics
Original document
Cited documents
Citing documents
INPADOC legal status
INPADOC patent family
Quick help
What is meant by high quality text as facsimile?
What does A1, A2, A3 and B stand for after a European publication number?
What happens if I click on "In my patents list"?
What happens if I click on the "Register" button?
Why are some sidebar options deactivated for certain documents?
How can I bookmark this page?
Why does a list of documents with the heading "Also published as" sometimes appear, and what are these documents?
Why do I sometimes find the abstract of a corresponding document?
What happens if I click on the red "patent translate" button?
What is Global dossier?
Bibliographic data: US2001049387 (A1) — 2001-12-06
In my patents list
EP Register
Report data error
Print
Biologically active analogs of discodermolide
Page bookmark
US2001049387 (A1) - Biologically active analogs of discodermolide
Inventor(s):
GUNASEKERA SARATH P [US]; LONGLEY ROSS E [US]; ISBRUCKER RICHARD A [CA]; PAUL GOPAL K [US]; POMPONI SHIRLEY A [US]; WRIGHT AMY E [US] +
Applicant(s):
GUNASEKERA SARATH P, ; LONGLEY ROSS E, ; ISBRUCKER RICHARD A, ; PAUL GOPAL K, ; POMPONI SHIRLEY A, ; WRIGHT AMY E, ; HARBOR BRANCH OCEANOGRAPHIC INSTITUTION, INC
Classification:
- international: A61K31/27; A61K31/365; A61K31/366; A61P35/00; C07C271/22; C07D309/30; C07D309/32; (IPC1-7): A61K31/366; C07D309/30
- cooperative: A61K31/365; C07C271/22; C07D309/30; C07D309/32; C07B2200/07
Application number:
US20010796175 20010228
Priority number(s):
US20010796175 20010228 ; US20000186145P 20000301
Also published as:
US6495594 (B2) US6495594 (X6) WO0164663 (A2) WO0164663 (A3) PT1259502 (E) more
Abstract of US2001049387 (A1)
Translate this text into
Select language
patenttranslate
powered by EPO and Google
The subject invention provides novel compositions of biologically active discodermolide compounds which can advantageously be used for immunomodulation and/or treating cancer. The compounds of the subject invention have utility for use in the treatment of cancer, as tubulin polymerizers and as microtubule stabilization agents. The present invention also pertains to the identification of regions of the discodermolide molecule which are responsible for certain aspects of the bioactivity of discodermolide compounds.
Sitemap Accessibility Legal notice Terms of use Last updated: 09.03.2016 Worldwide Database 6.0.9; 92p

Figure 1:

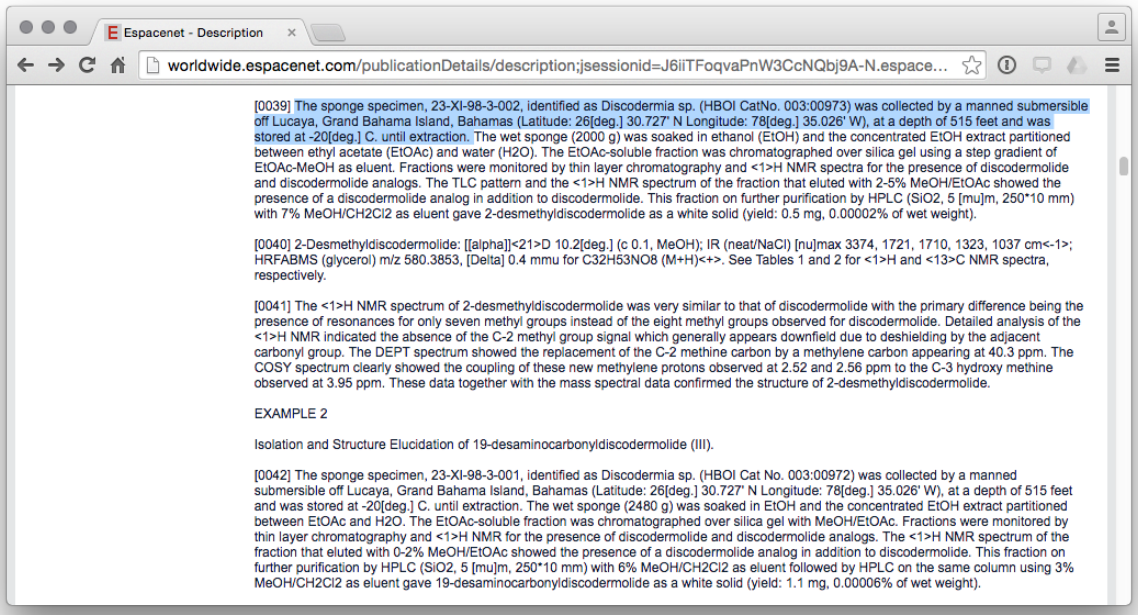

Figure 2:

very similar system where a unique identifier is used to link all documents arising from a permit application both internally within the system and for samples, publications and patent applications arising from the research. Specifically, the use of the unique identifier in labels for samples and specimens would facilitate the monitoring of compliance by users.

This type of unique identifier is simple, easy to use and robust over time. For that reason it is recommended as the first component of a triple system.

Simple Barcodes

A second system is the standard bar code which can be optically scanned to reveal basic information. The bar code system was developed in the 1960s and became ubiquitous for tracking and scanning products from the 1970s onwards. A range of bar code types are available along with free bar code generators. A simple example using a free tool is provided below. This bar code could be attached to documents and samples with basic information that could be encoded into the bar code. The advantage of a simple bar code is that it can be scanned by a machine.

Quick Response Codes (QR Codes)

The third system is QR (Quick Response) codes which provide a much greater level of embedded detail than bar codes and can be used to embed geographic and other information. An example is provided below.

The information in this QR code can be read using free software on a smart phone such as an iPhone, Android phone, or tablet, as can be seen in the image below (using QRReader on the iPhone). QR codes are normally open. However, encryption of data may potentially be desirable so that only authorised users (police, customs, port authorities) can scan the contents.

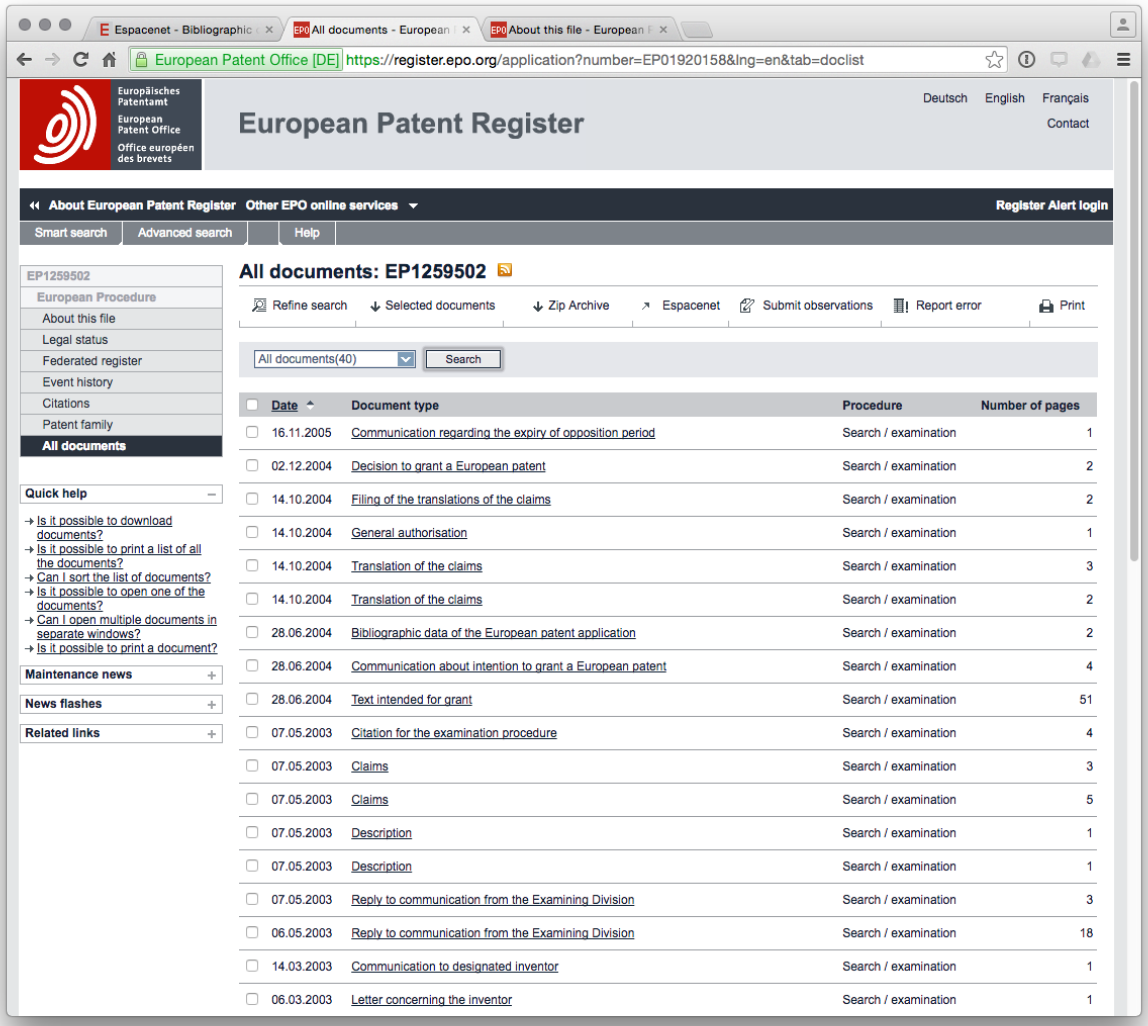

Figure 3:

Figure 4:

Figure 5:

The combination of the three systems would meet the requirement for triple redundancy. The main issue is not likely to be the means to generate the identifiers and codes but with ensuring that both authorities and applicants consistently use the identifiers and codes in documentation (including sample documentation) linked to a permit and associated MAT.

Finally, the use of identifiers is likely to be desirable in cases where ABS domestic frameworks include DNA and amino acid sequence data arising from research under a permit and associated MAT. This requires further exploration but is briefly considered below.

DNA and Amino Acid Sequences

Additional options include a requirement to use the basic unique identifier (BS20151234) in the documentation entered into databases for DNA and amino acid sequence data or html embed codes (for web publications).

In the case of DNA data an example of the existing use of identifiers, and the ability to monitor DNA sequence and amino acid information is provided by the Barcode of Life Database (BOLD).

Information using species names or identifiers can be retrieved using the `taxize` package in R using RStudio which generates the following link for a search for a species name.

```
library(taxize) # load the taxize package
get_boldid(searchterm = "Prunus africana") # search for the dna barcode id for a species
```

```
##
## Retrieving data for taxon 'Prunus africana'
```

```
## [1] "191949"
## attr(,"class")
## [1] "boldid"
## attr(,"match")
## [1] "found"
## attr(,"uri")
## [1] "http://boldsystems.org/index.php/Taxbrowser_Taxonpage?taxid=191949"
```

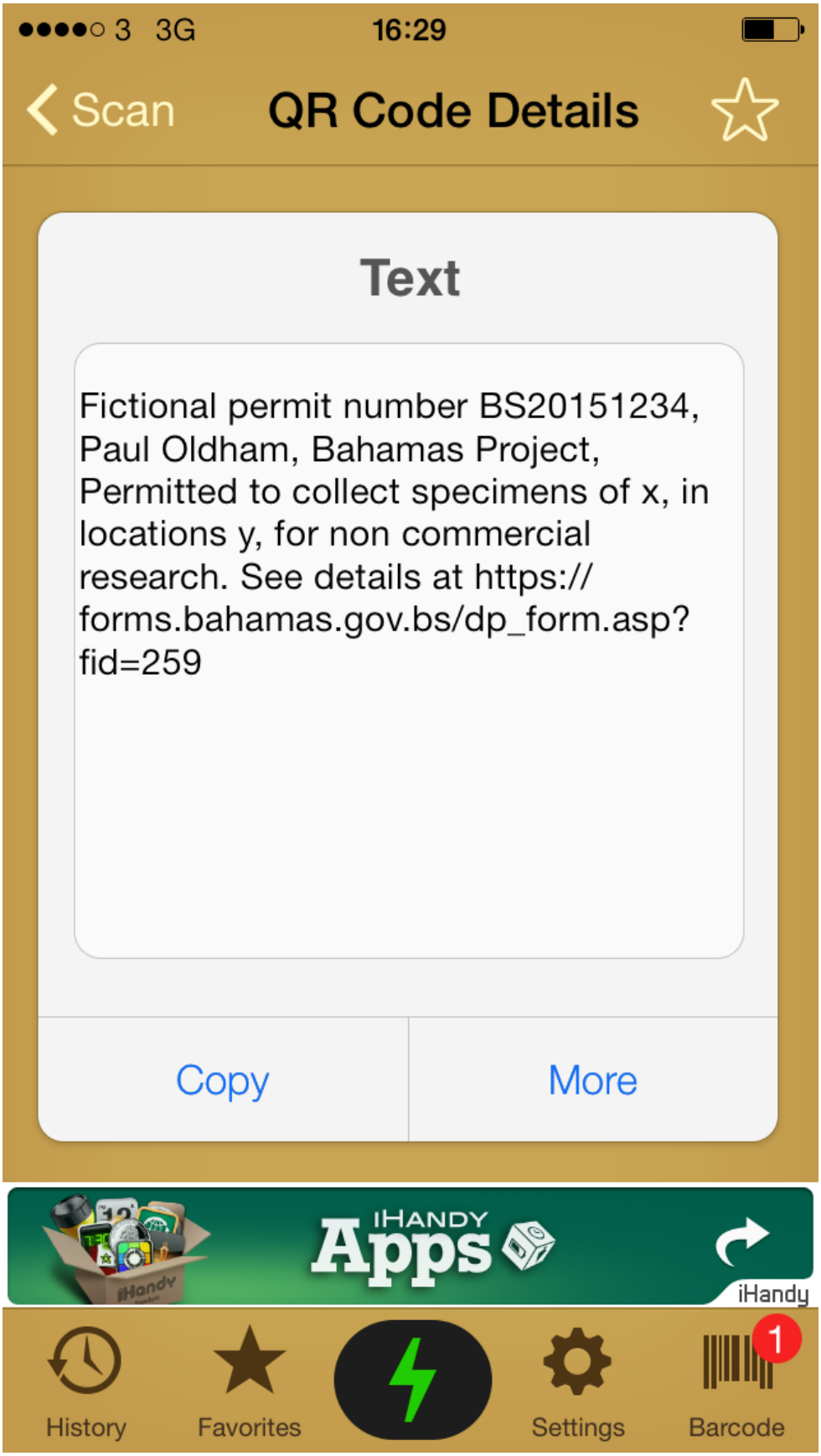
●●●●○ 3 3G
16:29
Scan
QR Code Details
Text
Fictional permit number BS20151234, Paul Oldham, Bahamas Project, Permitted to collect specimens of x, in locations y, for non commercial research. See details at https://forms.bahamas.gov.bs/dp_form.asp?fid=259
Copy
More
iHANDY Apps
iHandy
History
Favorites
Settings
Barcode
1

Figure 6:

More information from the URL can then be accessed in R as a data table or through the BOLD website. On the BOLD website this produces a list of sequence related records. The image below is viewable here.

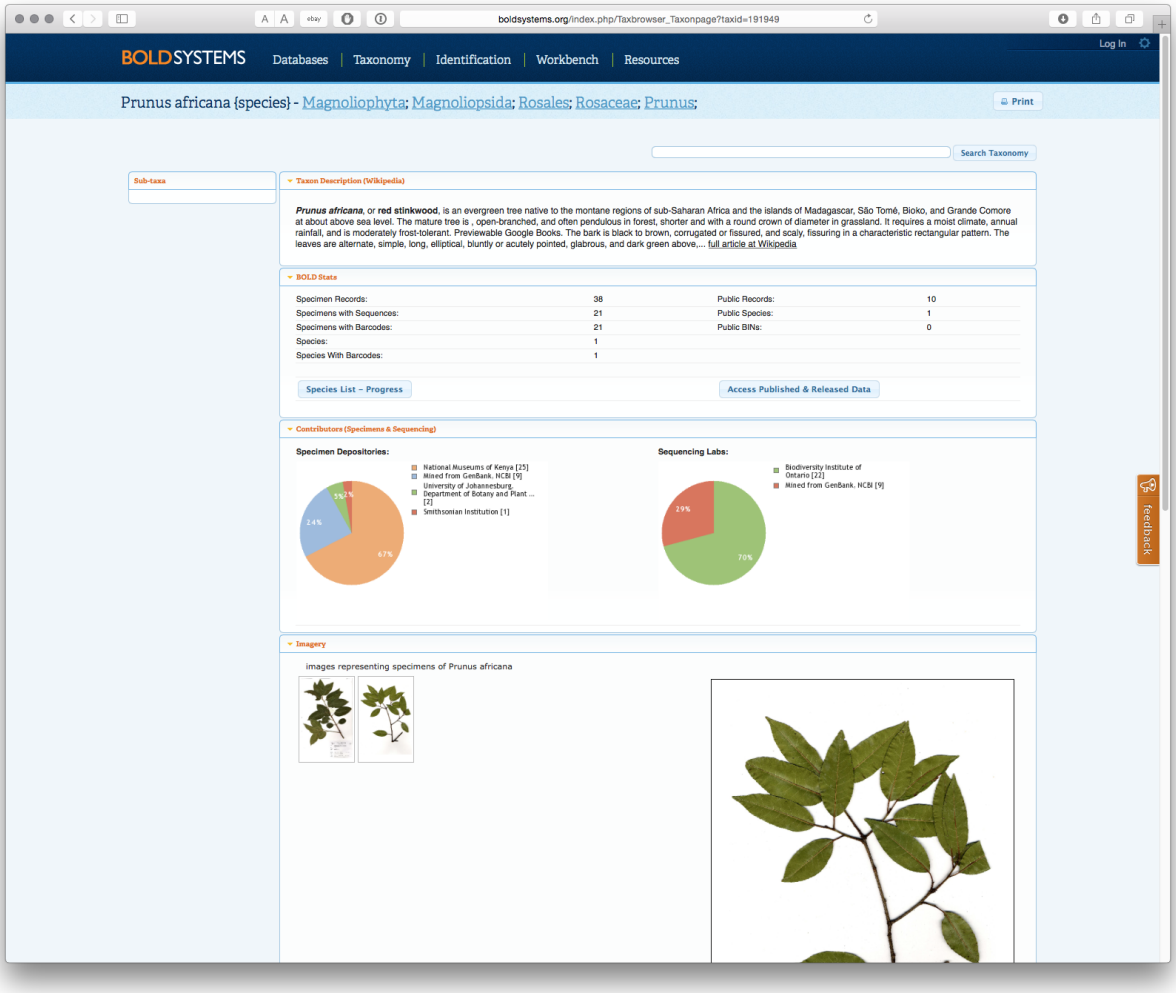

Figure 7:

Selecting `Access Published & Released Data` produces the following list of records.

The final record in this list includes a specimen image and the following information.

A significant amount of information is contained in this record, including the record number, sample ID, Museum ID, where the specimen is located along with where the material was collected, by whom, along with the sequence listing. A link is also provided to the Sequence ID and GenBank Accession number as seen below.

As this demonstrates, it is increasingly possible to rapidly access sequence and associated record information for a particular species or list of species. Given the presence of multiple ID fields it appears reasonable to assume that the simple permit identifier (e.g. BS21051234) in sequence records arising from research could be included in the conditions of the permit and associated MAT. This could readily lead to the creation of an archive of electronic records for biodiversity in a country that contain known sequence data. Uses of such sequences could then become amenable to monitoring using the relevant IDs or sequence searching for identical or similar sequences using BLAST (Basic Local Alignment Search Tool) and associated tools.

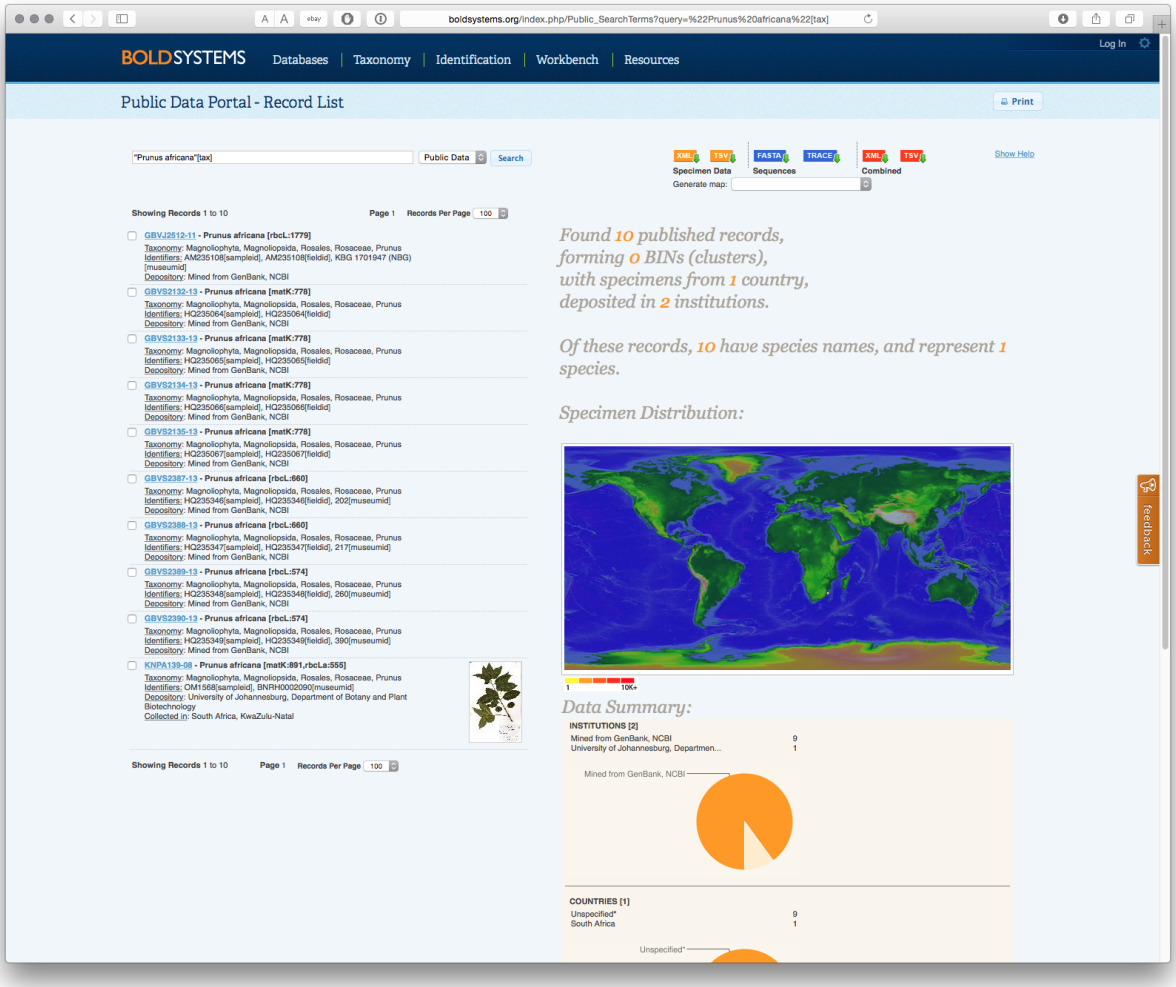
boldsystems.org/index.php/Public_SearchTerms?query=%22Prunus%20africana%22[tax]
BOLDSYSTEMS
Databases | Taxonomy | Identification | Workbench | Resources
Log In
Public Data Portal - Record List
Print
"Prunus africana"[tax]
Public Data
Search
Specimen Data
Sequences
Combined
Generate map:
Show Help
Showing Records 1 to 10
Page 1
Records Per Page 100
GBVJ2512-11 - Prunus africana [rbcL:1779]
Taxonomy: Magnoliophyta, Magnoliopsida, Rosales, Rosaceae, Prunus
Identifiers: AM235108[sampleid], AM235108[fieldid], KBG 1701947 (NBG) [museumid]
Depository: Mined from GenBank, NCBI
GBVS2132-13 - Prunus africana [matK:778]
Taxonomy: Magnoliophyta, Magnoliopsida, Rosales, Rosaceae, Prunus
Identifiers: HQ235064[sampleid], HQ235064[fieldid]
Depository: Mined from GenBank, NCBI
GBVS2133-13 - Prunus africana [matK:778]
Taxonomy: Magnoliophyta, Magnoliopsida, Rosales, Rosaceae, Prunus
Identifiers: HQ235065[sampleid], HQ235065[fieldid]
Depository: Mined from GenBank, NCBI
GBVS2134-13 - Prunus africana [matK:778]
Taxonomy: Magnoliophyta, Magnoliopsida, Rosales, Rosaceae, Prunus
Identifiers: HQ235066[sampleid], HQ235066[fieldid]
Depository: Mined from GenBank, NCBI
GBVS2135-13 - Prunus africana [matK:778]
Taxonomy: Magnoliophyta, Magnoliopsida, Rosales, Rosaceae, Prunus
Identifiers: HQ235067[sampleid], HQ235067[fieldid]
Depository: Mined from GenBank, NCBI
GBVS2387-13 - Prunus africana [rbcL:660]
Taxonomy: Magnoliophyta, Magnoliopsida, Rosales, Rosaceae, Prunus
Identifiers: HQ235346[sampleid], HQ235346[fieldid], 202[museumid]
Depository: Mined from GenBank, NCBI
GBVS2388-13 - Prunus africana [rbcL:660]
Taxonomy: Magnoliophyta, Magnoliopsida, Rosales, Rosaceae, Prunus
Identifiers: HQ235347[sampleid], HQ235347[fieldid], 217[museumid]
Depository: Mined from GenBank, NCBI
GBVS2389-13 - Prunus africana [rbcL:574]
Taxonomy: Magnoliophyta, Magnoliopsida, Rosales, Rosaceae, Prunus
Identifiers: HQ235348[sampleid], HQ235348[fieldid], 260[museumid]
Depository: Mined from GenBank, NCBI
GBVS2390-13 - Prunus africana [rbcL:574]
Taxonomy: Magnoliophyta, Magnoliopsida, Rosales, Rosaceae, Prunus
Identifiers: HQ235349[sampleid], HQ235349[fieldid], 390[museumid]
Depository: Mined from GenBank, NCBI
KNPA139-08 - Prunus africana [matK:891,rbcLa:555]
Taxonomy: Magnoliophyta, Magnoliopsida, Rosales, Rosaceae, Prunus
Identifiers: OM1568[sampleid], BNRH0002090[museumid]
Depository: University of Johannesburg, Department of Botany and Plant Biotechnology
Collected in: South Africa, KwaZulu-Natal
Showing Records 1 to 10
Page 1
Records Per Page 100
Found 10 published records, forming 0 BINs (clusters), with specimens from 1 country, deposited in 2 institutions.
Of these records, 10 have species names, and represent 1 species.
Specimen Distribution:
1
10K+
Data Summary:
INSTITUTIONS [2]
Mined from GenBank, NCBI 9
University of Johannesburg, Departmen... 1
Mined from GenBank, NCBI
COUNTRIES [1]
Unspecified* 9
South Africa 1
Unspecified*
feedback

Figure 8:

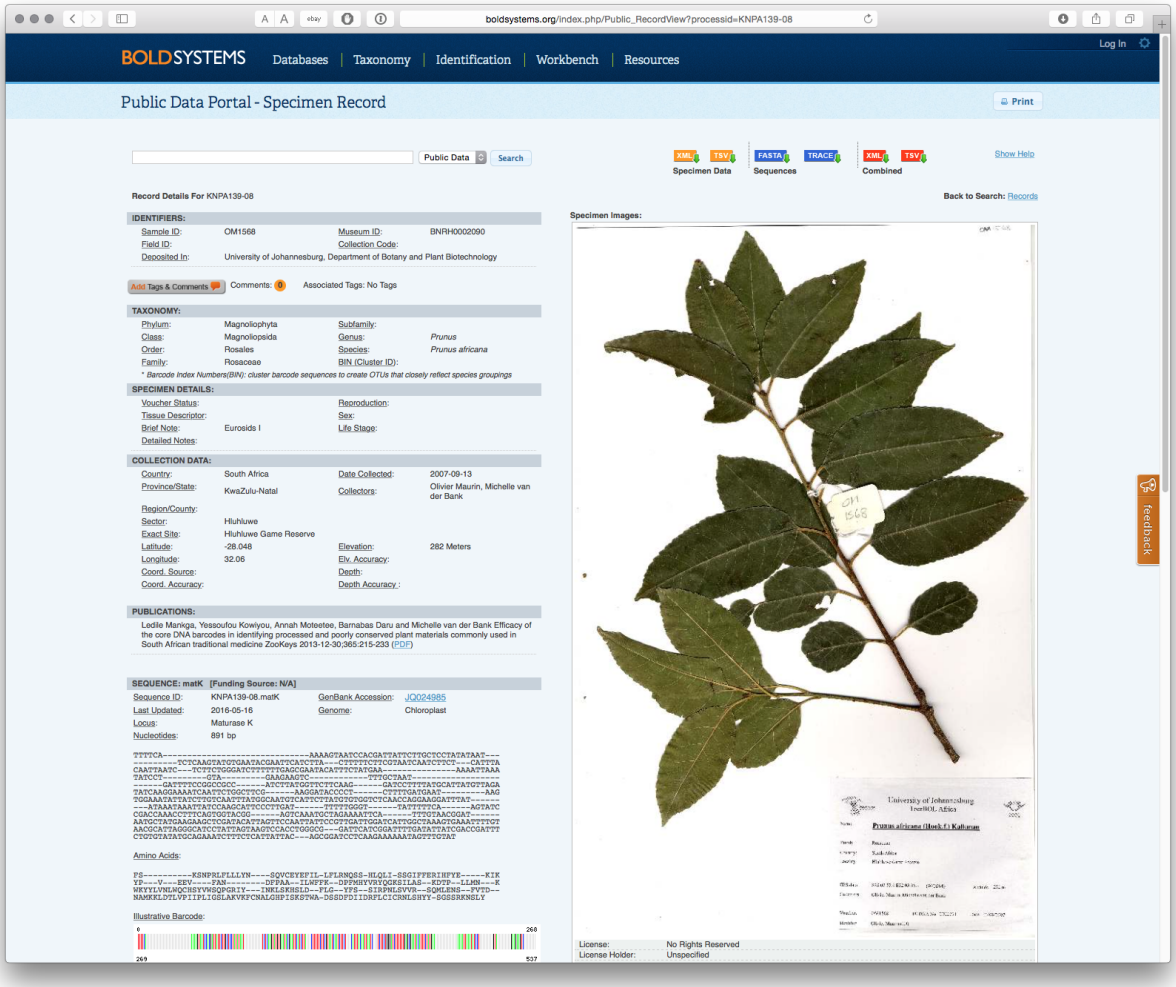
boldsystems.org/index.php/Public_RecordView?processid=KNPA139-08
BOLDSYSTEMS
Databases | Taxonomy | Identification | Workbench | Resources
Log In
Public Data Portal - Specimen Record
Print
Public Data
Search
Specimen Data
Sequences
Combined
Show Help
Record Details For KNPA139-08
Back to Search: Records
IDENTIFIERS:
Sample ID: OM1568
Museum ID: BNRH0002090
Field ID:
Collection Code:
Deposited In: University of Johannesburg, Department of Botany and Plant Biotechnology
Add Tags & Comments
Comments: 0
Associated Tags: No Tags
TAXONOMY:
Phylum: Magnoliophyta
Class: Magnoliopsida
Order: Rosales
Family: Rosaceae
Subfamily:
Genus: Prunus
Species: Prunus africana
BIN (Cluster ID):
* Barcode Index Numbers(BIN): cluster barcode sequences to create OTUs that closely reflect species groupings
SPECIMEN DETAILS:
Voucher Status:
Tissue Descriptor:
Brief Note: Eurosids I
Detailed Notes:
Reproduction:
Sex:
Life Stage:
COLLECTION DATA:
Country: South Africa
Province/State: KwaZulu-Natal
Region/County:
Sector: Hluhluwe
Exact Site: Hluhluwe Game Reserve
Latitude: -28.048
Longitude: 32.06
Coord. Source:
Coord. Accuracy:
Date Collected: 2007-09-13
Collectors: Olivier Maurin, Michelle van der Bank
Elevation: 282 Meters
Elv. Accuracy:
Depth:
Depth Accuracy :
PUBLICATIONS:
Ledile Mankga, Yessoufou Kowiyou, Annah Moteetee, Barnabas Daru and Michelle van der Bank Efficacy of the core DNA barcodes in identifying processed and poorly conserved plant materials commonly used in South African traditional medicine ZooKeys 2013-12-30;365:215-233 (PDF)
SEQUENCE: matK [Funding Source: N/A]
Sequence ID: KNPA139-08.matK
Last Updated: 2016-05-16
Locus: Maturase K
Nucleotides: 891 bp
GenBank Accession: JQ024985
Genome: Chloroplast
Amino Acids:
Illustrative Barcode:
Specimen Images:
University of Johannesburg
Prunus africana (Hook f.) Kalkman
OM 1568
License: No Rights Reserved
License Holder: Unspecified
feedback

Figure 9:

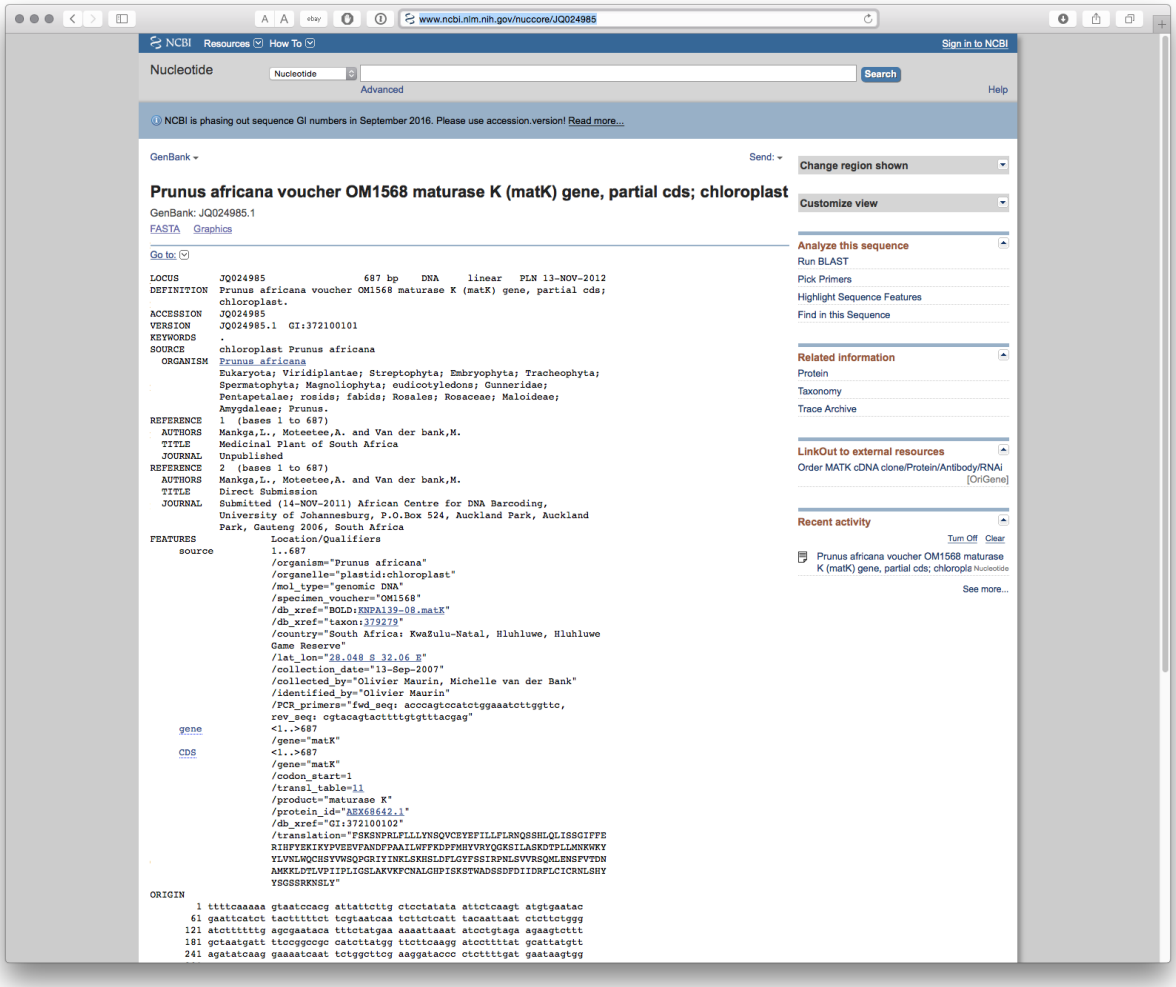
www.ncbi.nlm.nih.gov/nuccore/JQ024985
NCBI Resources How To
Sign in to NCBI
Nucleotide
Search
Advanced
Help
NCBI is phasing out sequence GI numbers in September 2016. Please use accession.version! Read more...
GenBank
Send:
Change region shown
Customize view
Prunus africana voucher OM1568 maturase K (matK) gene, partial cds; chloroplast
GenBank: JQ024985.1
FASTA Graphics
Go to:
LOCUS JQ024985 687 bp DNA linear PLN 13-NOV-2012
DEFINITION Prunus africana voucher OM1568 maturase K (matK) gene, partial cds; chloroplast.
ACCESSION JQ024985
VERSION JQ024985.1 GI:372100101
KEYWORDS .
SOURCE chloroplast Prunus africana
ORGANISM Prunus africana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae; Pentapetalae; rosids; fabids; Rosales; Rosaceae; Maloideae; Amygdaleae; Prunus.
REFERENCE 1 (bases 1 to 687)
AUTHORS Mankga,L., Moteetee,A. and Van der bank,M.
TITLE Medicinal Plant of South Africa
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 687)
AUTHORS Mankga,L., Moteetee,A. and Van der bank,M.
TITLE Direct Submission
JOURNAL Submitted (14-NOV-2011) African Centre for DNA Barcoding, University of Johannesburg, P.O.Box 524, Auckland Park, Auckland Park, Gauteng 2006, South Africa
FEATURES Location/Qualifiers
source 1..687
/organism="Prunus africana"
/organelle="plastid:chloroplast"
/mol_type="genomic DNA"
/specimen_voucher="OM1568"
/db_xref="BOLD:KNPA139-08.matK"
/db_xref="taxon:379279"
/country="South Africa: KwaZulu-Natal, Hluhluwe, Hluhluwe Game Reserve"
/lat_lon="28.048 S 32.06 E"
/collection_date="13-Sep-2007"
/collected_by="Olivier Maurin, Michelle van der Bank"
/identified_by="Olivier Maurin"
/PCR_primers="fwd_seq: acccagtccatctggaaatcttggttc, rev_seq: cgtacagtacttttgtgtttacgag"
gene <1..>687
/gene="matK"
CDS <1..>687
/gene="matK"
/codon_start=1
/transl_table=11
/product="maturase K"
/protein_id="AEX68642.1"
/db_xref="GI:372100102"
/translation="FSKSNPRLPLLLVNSQVCEYEFILLPLRNQSSHLQLISSGIFFE RIHFVEKIKYPVEEVFANDFPAAILWFFKDPFMHVVRVQGKSILASKDTPLLMNKWKV YLVNLWQCHSYVWSQPGRIYINKLSKHSLDFLGYFSSIRPNLSVVRSQMLENSFVTDN AMKKLDTLVPIIPLIGSLAKVKFCNALGHPISKSTWADSSDFDIIDRFLCICRNLSHY YSGSSRKNSLY"
ORIGIN
1 ttttcaaaaa gtaatccacg attattcttg ctcctatata attctcaagt atgtgaatac
61 gaattcatct tacttttttct tcgtaatcaa tcttctcatt tacaattaat ctcttctggg
121 atctttttttg agcgaataca tttctatgaa aaaattaaat atcctgtaga agaagtcttt
181 gctaatgatt ttccggccgc catcttatgg ttcttcaagg atccttttat gcattatgtt
241 agatatcaag gaaaatcaat tctggcttcg aaggataccc ctcttttgat gaataagtgg
Analyze this sequence
Run BLAST
Pick Primers
Highlight Sequence Features
Find in this Sequence
Related information
Protein
Taxonomy
Trace Archive
LinkOut to external resources
Order MATK cDNA clone/Protein/Antibody/RNAi [OriGene]
Recent activity
Turn Off Clear
Prunus africana voucher OM1568 maturase K (matK) gene, partial cds; chloropla Nucleotide
See more...

Figure 10:

Conclusion

In this section we have discussed the use of unique identifiers and the application of the principle of triple redundancy as part of the design of the online permit and monitoring system. Unique identifiers using standardised country codes, dates and sequences of numbers allow for the construction of an internal permit system that establishes and maintains links between a permit application and associated documents (MAT) and communications. This system already works well for millions of patent documents. The principle of triple redundancy was then applied to the generation of labels containing identifiers that could be used to maintain links between the original permit and samples, publications, patent applications, and sequence data originating from the grant of a permit under an ABS framework. The use of free tools (such as `taxize` in R) allows this information to be readily retrieved from a range of different data sources. While requiring further elaboration, the use of unique identifiers combined with the principle of triple redundancy provides a route to cost effective monitoring and is recommended.