

## (二) 應用範例

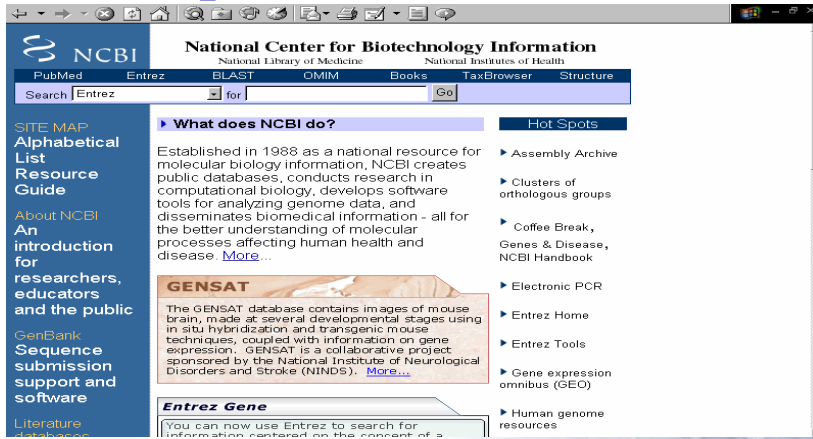
講者：陳亞寧

## 實際應用範例

- **Linking to Full Text**
  - Linking Sequence No. to Bio-sequence Databanks
  - Linking Individual Industrial Codes to the Full Scheme
- **Linking to Descriptive Records**
  - Linking Organism Names to Taxonomic Records
  - Linking Personal Names to Biographical Information



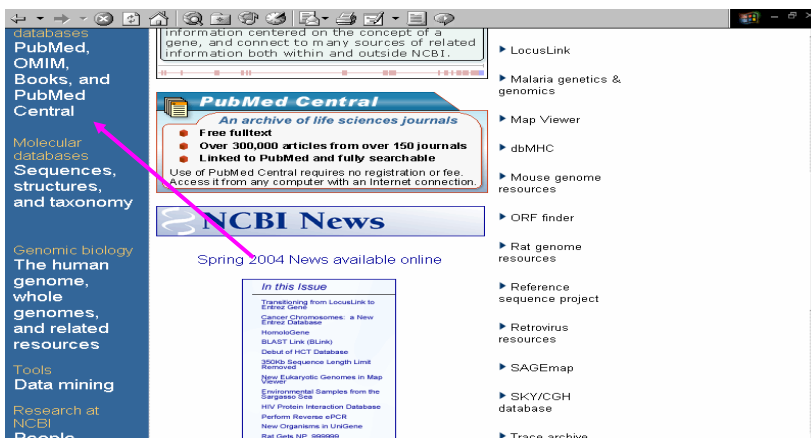
# Linking Sequence No. to Bio-sequence Databanks



<http://www.ncbi.nlm.nih.gov>

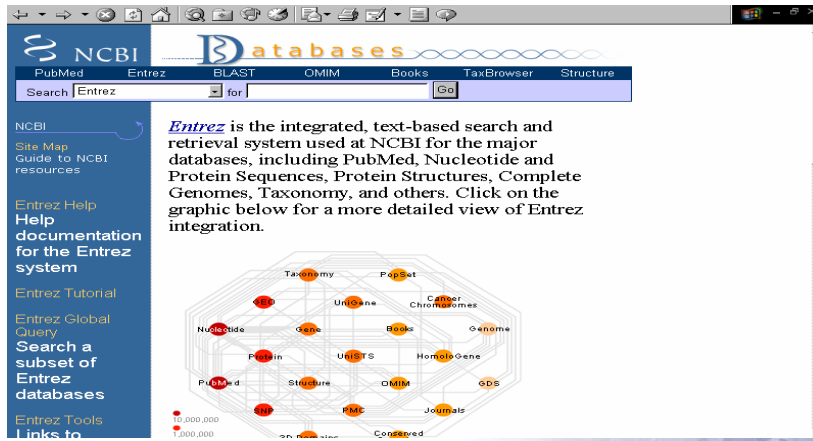


# Linking Sequence No. to Bio-sequence Databanks<sup>2</sup>





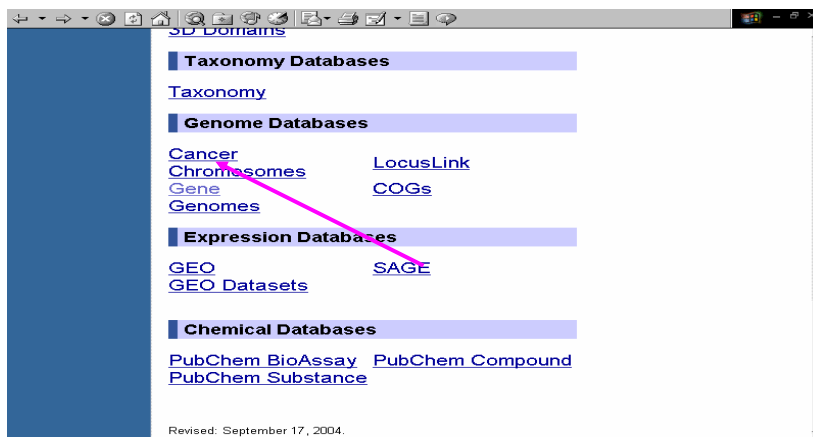
## Linking Sequence No. to Bio-sequence Databanks<sup>3</sup>



The screenshot shows the NCBI Databases homepage. The top navigation bar includes links to PubMed, Entrez, BLAST, OMIM, Books, TaxBrowser, and Structure. A search bar is present with the text "Search Entrez for" and a "Go" button. On the left, there is a sidebar with links to "NCBI Site Map", "Entrez Help", "Entrez Tutorial", "Entrez Global Query", "Search a subset of Entrez databases", and "Entrez Tools". The main content area features a paragraph about Entrez: "Entrez is the integrated, text-based search and retrieval system used at NCBI for the major databases, including PubMed, Nucleotide and Protein Sequences, Protein Structures, Complete Genomes, Taxonomy, and others. Click on the graphic below for a more detailed view of Entrez integration." Below this text is a complex network diagram showing the integration of various databases. The diagram includes nodes for Taxonomy, Protein, Nucleotide, Structure, OMIM, Journals, Genomes, and Cancer Chromosomes, with lines indicating their interconnections. A legend at the bottom left of the diagram indicates that red dots represent 10,000,000 and blue dots represent 1,000,000.



## Linking Sequence No. to Bio-sequence Databanks<sup>4</sup>



The screenshot shows the NCBI Databases homepage with a pink arrow pointing to the "Genomes" link under the "Genome Databases" section. The page is organized into several categories, each with a blue header bar: "Taxonomy Databases" (with a link to "Taxonomy"), "Genome Databases" (with links to "Cancer Chromosomes", "LocusLink", "Gene", "COGs", and "Genomes"), "Expression Databases" (with links to "GEO" and "SAGE"), and "Chemical Databases" (with links to "PubChem BioAssay", "PubChem Compound", and "PubChem Substance"). The "Genomes" link is highlighted with a pink arrow. At the bottom of the page, it says "Revised: September 17, 2004."



## Linking Sequence No. to Bio-sequence Databanks<sup>5</sup>

NCBI Entrez Gene

Search  for    ☒ current records only

Limits Preview/Index History Clipboard Details

- Enter one or more search terms.
- More information about available fields is available [here](#).
- Consider use of the limits and preview/index functions.
- Remember, boolean operators (AND, OR, NOT) must be in uppercase.

**Gene** Background

Gene provides a unified query environment for genes defined by sequence and/or in NCBI's Map Viewer. You can query on names, symbols, accessions, publications, GO terms, chromosome numbers, E.C. numbers, and many other attributes associated with genes and the products they encode.

Because Gene is now an Entrez database, all the familiar and useful functions are now available, including Preview/Index, History, and LinkOut.

**Please note:** Entrez Gene is under active development. We welcome your [suggestions](#). We have also added a choice in our Feedback/Corrections [form](#) for suggestions; both paths reach the same staff.

Getting started	Sample queries
Look for genes by name part and multiple species	<a href="#">transporter AND ("Drosophila melanogaster"[orgn] OR "Mus musculus"[orgn]) more...</a>
Look for genes by chromosome and symbol	<a href="#">11[chr] OR 2[chr] AND adh*[sym] more...</a>



## Linking Sequence No. to Bio-sequence Databanks<sup>6</sup>

NCBI Entrez Gene

Search  for    ☒ current records only

Limits Preview/Index History Clipboard Details

Display  Show:  Send to:

☒ 1: **BRCA2** [Links](#)

breast cancer 2, early onset [*Homo sapiens*]

**Other Aliases:** HGNC:1101, FACP, FAD, FAD1, FANCB, FANCD, FANCD1

**Other Designations:** Fanconi anemia, complementation group D1

**Chromosome:** 13; **Location:** 13q12.3

**GeneID:** 675



# Linking Sequence No. to Bio-sequence Databanks<sup>7</sup>

The screenshot shows the Entrez Gene database entry for BRCA2. The left sidebar contains navigation links such as 'Entrez', 'SITE MAP', 'Gene Search', 'FAQ', 'FTP site', 'Related sites', 'Feedback', and 'Subscriptions'. The main content area displays the gene name 'BRCA2' and its description 'breast cancer 2, early onset'. It includes the GeneID (675), Locus tag (HGNC:1101; MIM:600185), and a genomic context diagram showing the gene's location on chromosome 13. The 'Gene type' is listed as 'protein coding', and the 'Gene name' is 'BRCA2'. The 'Gene description' is 'breast cancer 2, early onset'. The 'RefSeq status' is 'Reviewed'. The 'Organism' is 'Homo sapiens'. The 'Lineage' is 'Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo'. The 'Gene aliases' are 'FAD; FADC; FADI; FANCB; FANCD; FANCD1'. The 'Summary' states: 'Mutations in BRCA1 and BRCA2 have been linked to an elevated risk of young onset breast cancer which has been demonstrated to be due to the inheritance of dominant susceptibility genes conferring a high risk of the disease. Unlike BRCA1, BRCA2 has not been linked to ovarian cancer. While BRCA1 mutations are typically microinsertions and point mutations, BRCA2 mutations are typically microdeletions. BRCA2 putatively functions as a tumor suppressor gene, however, its exact function has not been well characterized. The similarity of, and functional'.



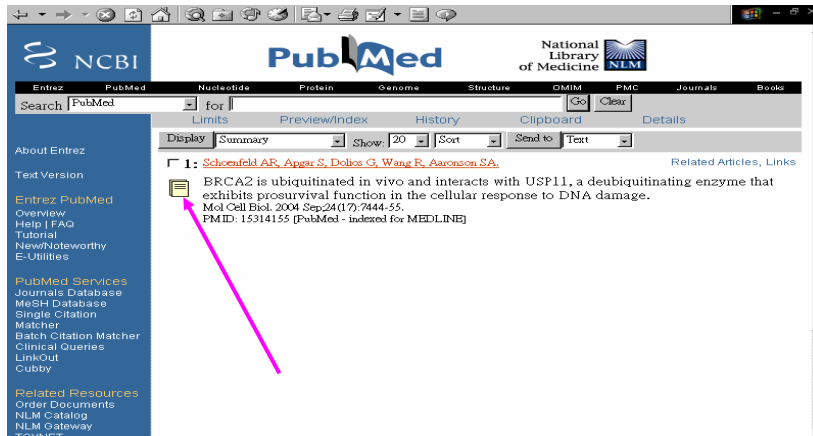
# Linking Sequence No. to Bio-sequence Databanks<sup>8</sup>

The screenshot shows a PubMed bibliography for the BRCA2 gene. The left sidebar contains navigation links such as 'Entrez', 'SITE MAP', 'Gene Search', 'FAQ', 'FTP site', 'Related sites', 'Feedback', and 'Subscriptions'. The main content area displays the title 'Bibliography: Gene References into Function (GeneRIF): Submit help'. Below the title, there is a list of 13 numbered entries, each with a 'PubMed' link. A pink arrow points from the 'PubMed' link of entry 5 to the 'PubMed' link of entry 12. The entries are:

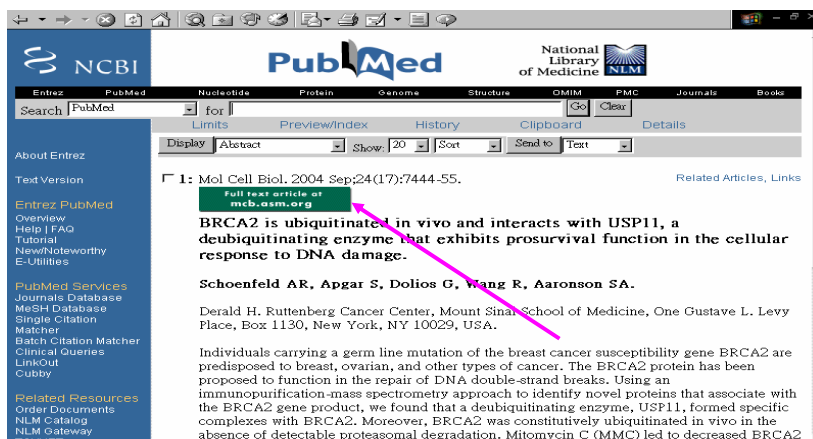
1. Founder mutations are present within the Scottish/Northern Irish population and have implications for the organisation of molecular screening services. [PubMed](#)
2. Results suggest that BRCA2 expression levels are regulated by ubiquitination in response to DNA damage and that USP11 participates in DNA damage repair functions within the BRCA2 pathway independently of BRCA2 deubiquitination. [PubMed](#)
3. Identification and evaluation of 55 genetic variations in the BRCA1 and the BRCA2 genes of patients from 50 Japanese breast cancer families. [PubMed](#)
4. results demonstrate that monoubiquitination of FANCD2, which is regulated by the FA pathway, promotes BRCA2 loading into chromatin complexes. These complexes appear to be required for normal homology-directed DNA repair. [PubMed](#)
5. BRCA2 mutations could not be detected among unrelated non-Ashkenazi-Jewish high risk families in Israel. [PubMed](#)
6. Cancer variation associated with the position of the mutation in the BRCA2 gene in 7 different neoplasms and the differences in cancer risk remain to be explored. [PubMed](#)
7. RAD51D polymorphism is not associated with BRCA1 or 2 genes in breast cancer. [PubMed](#)
8. BRCA mutations were present in 12.7% of the high risk patients, compared with 2.8% of the unselected patients. [PubMed](#)
9. The remarkable clinical overlap between sporadic EMSY amplification and familial BRCA2 deletion implicates a BRCA2 pathway in sporadic breast and ovarian cancer. [PubMed](#)
10. identified 6 children in 5 kindreds exhibiting the co-occurrence of BRCA2 mutations, FA, and early onset acute leukemia. [PubMed](#)
11. , BRCA2 is epistatic to FA genes for ICL repair, but not for damage-induced modification of FANCD2 and may act downstream form FANCD2. [PubMed](#)
12. Allelic loss at the BRCA2 locus may be of use as a negative predictor for metastases-free and overall survival in breast cancer patients. [PubMed](#)
13. Our results suggest that there is a field effect of early genetic events preceding morphologic changes in the mammary glands of BRCA mutation carriers. [PubMed](#)



# Linking Sequence No. to Bio-sequence Databanks<sup>9</sup>



# Linking Sequence No. to Bio-sequence Databanks<sup>10</sup>





# Linking Sequence No. to Bio-sequence Databanks<sup>11</sup>

**MCB** Molecular and Cellular Biology **FREE Back Issues** (now after 6 months)

[HOME](#) [HELP](#) [FEEDBACK](#) [SUBSCRIPTIONS](#) [ARCHIVE](#) [SEARCH](#) [TABLE OF CONTENTS](#)

Institution: Life Science Library / Academia Sinica [Sign In as Member](#)

Molecular and Cellular Biology, September 2004, p. 7444-7455, Vol. 24, No. 17  
0270-7306/04/08.00+0 DOI 10.1128/MCB.24.17.7444-7455.2004  
Copyright © 2004, American Society for Microbiology. All Rights Reserved.

## BRCA2 Is Ubiquitinated In Vivo and Interacts with USP11, a Deubiquitinating Enzyme That Exhibits Prosurvival Function in the Cellular Response to DNA Damage

Alan R. Schoenfeld,<sup>1</sup> Sarah Appar,<sup>1</sup> Georgia Dolios,<sup>2</sup> Rong Wang,<sup>2</sup> and Stuart A. Aaronson<sup>1,\*</sup>

Derald H. Ruttenberg Cancer Center,<sup>1</sup> Department of Human Genetics, Mount Sinai School of Medicine, New York, New York<sup>2</sup>

Received 16 January 2004/ Returned for modification 3 March 2004/ Accepted 3 June 2004

### ABSTRACT

Individuals carrying a germ line mutation of the breast cancer susceptibility gene *BRCA2* are

**This Article**

- Abstract view
- Full Text (PDF)
- Alert me when this article is cited
- Alert me if a correction is posted

**Services**

- Similar articles in this journal
- Similar articles in PubMed
- Alert me to new issues of the journal
- Download to citation manager
- Books from ASM Press

**PubMed**

- PubMed Citation
- Articles by Schoenfeld, A. R.
- Articles by Aaronson, S. A.



# Linking Individual Industrial Codes to the Full Scheme<sup>1</sup>

**U.S. Census Bureau**

## North American Industry Classification System (NAICS)

The North American Industry Classification System (NAICS) has replaced the U.S. Standard Industrial Classification (SIC) system. NAICS will reshape the way we view our changing economy.

NAICS was developed jointly by the U.S., Canada, and Mexico to provide new comparability in statistics about business activity across North America.

NAICS 2002 includes substantial revisions within the Construction and Wholesale Trade sectors, and a number of revisions for the Retail and Information sectors. NAICS 2002 will be implemented in the 2002 Economic Census.

**How NAICS affects data users**

- [New industries](#)
- [New sectors](#)
- [New code system](#)
- [Benefits](#)
- [Breaks in time series](#)

**Implementation**

- [Census Bureau](#)
- [BEA](#)
- [BLS](#)
- [Other agencies](#)

Product

<http://www.census.gov>

**NAICS to SIC**  
SIC to NAICS

334  
NAICS Search

**Federal Register**  
notice describing proposed




## Linking Individual Industrial Codes to the Full Scheme<sup>2</sup>

Results for "334"

Index entry	NAICS Code		2002 U.S. NAICS Title
	2002	1997	
334	<a href="#">334</a>		Computer and Electronic Product Manufacturing
3341	<a href="#">3341</a>		Computer and Peripheral Equipment Manufacturing
33411	<a href="#">33411</a>		Computer and Peripheral Equipment Manufacturing
334111	<a href="#">334111</a>		Electronic Computer Manufacturing
334112	<a href="#">334112</a>		Computer Storage Device Manufacturing
334113	<a href="#">334113</a>		Computer Terminal Manufacturing
334119	<a href="#">334119</a>		Other Computer Peripheral Equipment Manufacturing
3342	<a href="#">3342</a>		Communications Equipment Manufacturing
33421	<a href="#">33421</a>		Telephone Apparatus Manufacturing
334210	<a href="#">334210</a>		Telephone Apparatus Manufacturing
33422	<a href="#">33422</a>		Radio and Television Broadcasting and Wireless Communications Equipment Manufacturing
334220	<a href="#">334220</a>		Radio and Television Broadcasting and Wireless Communications Equipment Manufacturing
33429	<a href="#">33429</a>		Other Communications Equipment Manufacturing



## Linking Organism Names to Taxonomic Records

 **ITIS** Integrated Taxonomic Information System

Welcome to ITIS, the Integrated Taxonomic Information System! Here you will find authoritative taxonomic information on plants, animals, fungi, and microbes of North America and the world. We are a [partnership](#) of U.S., [Canadian](#), and [Mexican](#) agencies ([ITIS-North America](#)), other organizations, and taxonomic specialists. ITIS is also a partner of [Species 2000](#) and the [Global Biodiversity Information Facility \(GBIF\)](#).

**Quick search on:**

☒ Any Name or TSN\* ☐ Common Name ☐ Scientific Name ☐ TSN\*

In:  Kingdom

\* Taxonomic Serial Number (TSN)

[Go to Advanced Search and Report](#)

Last Updated: 20-Aug-2002  
[Privacy statement and disclaimers](#)  
<http://www.itis.usda.gov/index.html>

**What's New**  
**About ITIS**  
**Data Access**  
**Submit Data**  
**Tools**  
**TREED**  
**Links**  
**Comments**

<http://www.itis.usda.gov/>





# Linking Organism Names to Taxonomic Records<sup>2</sup>

ITIS Search Results

Home Data Access Submit Data Tools Comment

Results of: Search in every Kingdom for Common Name containing 'beetles'

Kingdom Animalia

- [ambrosia beetles](#) -- Family: Scolytidae -- valid
- [anobiid beetles](#) -- Family: Anobiidae -- valid
- [antlike flower beetles](#) -- Family: Anthicidae -- valid
- [antlike leaf beetles](#) -- Family: Euglenidae -- valid
- [antlike stone beetles](#) -- Family: Scydmaenidae -- valid
- [aromatopid beetles](#) -- Family: Armatopidae -- valid
- [bark beetles](#) -- Family: Scolytidae -- valid
- [bark gnawing beetles](#) -- Family: Scolytidae -- valid
- [beetles](#) -- Order: Coleoptera -- valid
- [bess beetles](#) -- Family: Passalidae -- valid
- [blister beetles](#) -- Family: Meloidae -- valid
- [burrowing water beetles](#) -- Family: Noteridae Regimbart, 1876 -- valid
- [callinid beetles](#) -- Family: Callinidae -- valid
- [carrion beetles](#) -- Family: Silphidae -- valid
- [cebrionid beetles](#) -- Family: Cebrionidae -- valid
- [cedar beetles](#) -- Family: Phloeosinus -- valid
- [cerophytid beetles](#) -- Family: Cerophytidae -- valid
- [cervionid beetles](#) -- Family: Cervionidae -- valid
- [checkered beetles](#) -- Family: Cleridae -- valid
- [cheloniid beetles](#) -- Family: Cheloniidae -- valid
- [click beetles](#) -- Family: Elateridae -- valid
- [comb-clawed beetles](#) -- Family: Alleculidae -- valid
- [crawling water beetles](#) -- Family: Halplidae Latreille, 1802 -- valid
- [cylindrical bark beetles](#) -- Family: Cylindridae -- valid
- [dardine beetles](#) -- Family: Tenebrionidae -- valid
- [dasyderid beetles](#) -- Family: Dasyderidae -- valid
- [desmids beetles](#) -- Family: Desmidiidae -- valid



# Linking Organism Names to Taxonomic Records<sup>3</sup>

Armatopidae

Taxonomic Serial No.: 114281

**Taxonomy and Nomenclature**

Kingdom:	Animalia
Taxonomic Rank:	Family
Synonym(s):	
Common Name(s):	armatopid beetles

**Taxonomic Status:**

Current Standing:	valid
-------------------	-------

**Data Quality Indicators:**

Record Credibility Rating:	unverified
Global Species Completeness:	unknown
Latest Record Review:	unknown

**Taxonomic Hierarchy**

Kingdom	Animalia -- Animal, animals, animaux
Phylum	Arthropoda -- arthropodes, arthropods, Artrópode
Subphylum	Hexapoda -- hexapods
Class	Insecta -- hexapoda, insectes, insects, insecto
Subclass	Pterygota -- insects ailés, winged insects
Infraclass	Neoptera -- modern, wing-folding insects
Order	Coleoptera -- beetles, besouro, coléoptères
Suborder	Polyphaga
Superfamily	Dynopoidea
Family	Armatopidae -- armatopid beetles
Genus	<b>Direct Children:</b> <a href="#">Eurygaster</a>

## Linking Organism Names to Taxonomic Records<sup>4</sup>

References	
<b>Expert(s):</b>	
Expert:	
Notes:	
Reference for:	
<b>Other Source(s):</b>	
Source:	<a href="#">NODC Taxonomic Code, database (version 8.0)</a>
Acquired:	1996
Notes:	
Reference for:	artematoiid beetles
<b>Publication(s):</b>	
Author(s)/Editor(s):	
Publication Date:	
Article/Chapter Title:	
Journal/Book Name, Vol. No.:	
Page(s):	
Publisher:	
Publication Place:	
ISBN/ISSN:	
Notes:	
Reference for:	

Geographic Information	
<b>Geographic Division:</b>	
<b>Jurisdiction/Origin:</b>	

Comments	
<b>Comment:</b>	

## Linking Organism Names to Taxonomic Records<sup>5</sup>

ISBN/ISSN:  
Notes:  
Reference for:

---

**Geographic Information**

Geographic Division:  
Jurisdiction/Origin:

---

**Comments**

Comment:

---

**Global Biodiversity Information Facility (GBIF) Data Portal** NEW

Retrieve biodiversity records and collection information from GBIF:  
[884 record\(s\)](#), [884 record\(s\) with coordinates](#)  
[Dynamic map](#)

---

**Search on:** ☒ Any Name or TSN ☐ Common Name ☐ Scientific Name ☐ TSN

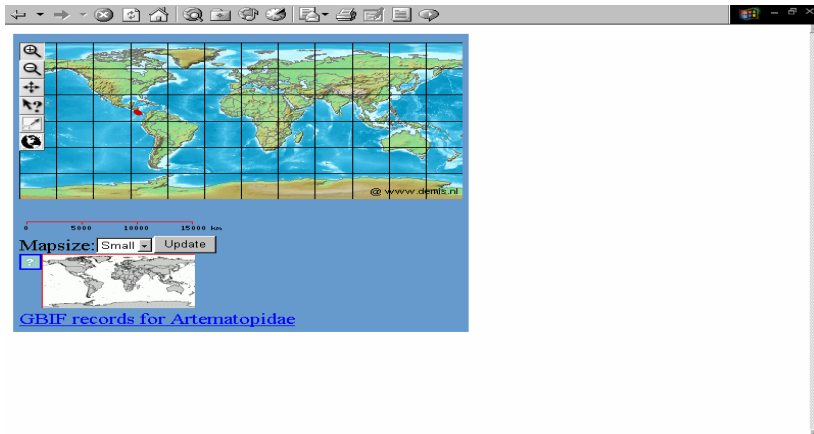
In:  Kingdom:

[Go to Advanced Search and Report](#)

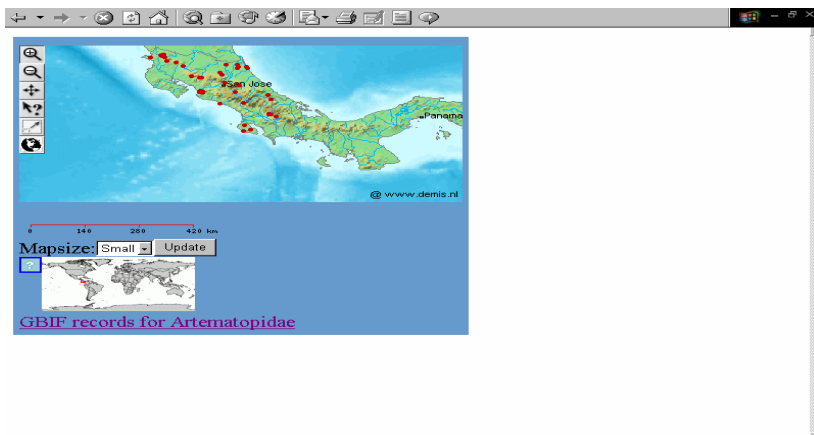
Date Generated:  
Wed Nov 24 2004 01:57:17 MST



## Linking Organism Names to Taxonomic Records<sup>6</sup>



## Linking Organism Names to Taxonomic Records<sup>7</sup>





## Linking Organism Names to Taxonomic Records<sup>8</sup>

ISBN/ISSN:  
Notes:  
Reference for:

**Geographic Information**

Geographic Division:  
Jurisdiction/Origin:

**Comments**

Comment:

**Global Biodiversity Information Facility (GBIF) Data Portal** **NEW**

Retrieve biological records and collection information from GBIF:  
[884 record\(s\). 884 records\(s\) with coordinates](#)  
[Dynamic map](#)

Search on: ☒ Any Name or TSN ☐ Common Name ☐ Scientific Name ☐ TSN

In:  every  Kingdom  containing

[Go to Advanced Search and Report](#)

Date Generated:  
Wed Nov 24 2004 01:57:17 MST



## Linking Organism Names to Taxonomic Records<sup>9</sup>

HOME | GBIF | BROWSE TAXONOMY | SEARCH | DATA PROVIDERS | COUNTRIES | DATA USE

**GBIF** Prototype data portal  
Global Biodiversity Information Facility

**Family: Artematopidae**

**Specimens/observations**  
Including records from: [Costa Rica](#)

Service	Resource	All	Lat/Long
<a href="#">INBio (200.91.91.109)</a>	Biodiversidad de Costa Rica	884	884
Total		884	884

[Contact info](#) | [Webmaster](#)



## Linking Personal Names to Biographical Information

- <http://authorities.loc.gov>
- <http://catalog.loc.gov/>
- <http://virtua.lib.tku.edu.tw>
- <http://www.lib.ntu.edu.tw/catalog/webpac/webpac.asp>



### (三) 規劃與設計原則

講者：陳亞寧

## 設計前提<sup>1</sup>

- 避免重複工作的投入
  - 採用既有的索引典
  - 以既有的索引典為基礎, 進行小幅度新增, 修改與刪除
  - 發展新的索引典
- 決定索引典的結構與展現格式
  - 扁平式(flat)
  - 階層式(hierarchy)
- 發展方式
  - 委員會: 由一組學科專家組成
  - 經驗式: 從既有文獻中予以分析出所需的詞彙
  - 混合式(hybrid)
- 電腦工具的協助與利用
  - 潛在的詞彙(candidate terms)與停用詞清單(stop list, ex. To, the et al.)
  - 索引典詞彙的實際使用次數
  - 索引典詞彙實際被使用(query)的次數

## 設計前提<sup>2</sup>

- 索引詞的Metadata (Term Records)
  - Descriptor, scope note (SN), 同義詞, non-displayable variations, NT/BT, RT, category/call no., 歷史註(HN)
- 索引詞的品質驗證(Term Verification)
- 索引詞的精確度(Level of Specificity)
  - 數量 & 成本
- 未使用的索引詞(Unassigned Descriptors)
- 公佈與存儲(Announcement and Deposit of Published Thesauri)

## 規劃原則

- 範圍
  - 學科與領域
  - 核心主題—深入, 邊緣主題—粗略
- 資料類型
  - 期刊—精細, 圖書—約略
- 資料量
  - 資料量大小與成長率
  - 量大小 vs. 成本高低
- 使用者
  - 學科專家 vs. 大眾
- 問題類型
  - 概括—粗略, 明確—詳細
- 詞彙組合方式
  - 前組合或後組合
- 詳簡度