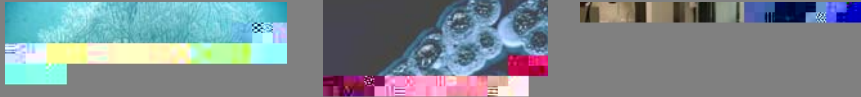
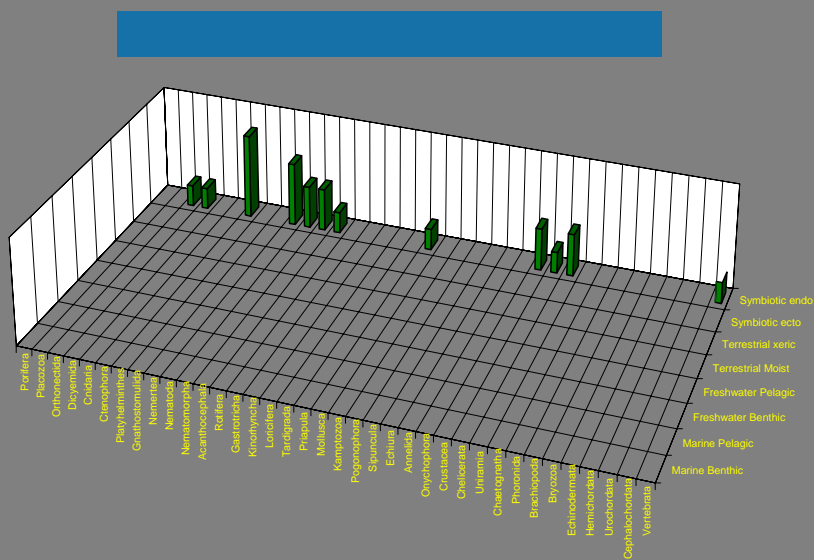


Towards a practical knowledgebase for marine genetic resources



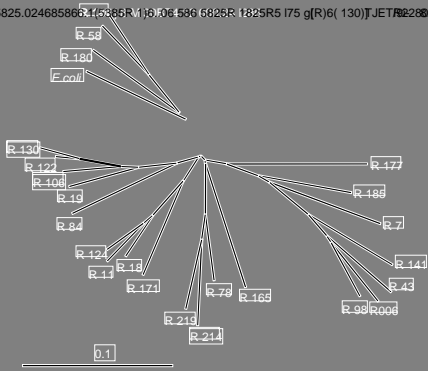
Libby Evans-Illidge

Manager Bioresources Library, Australian Institute of Marine Science,
 PMB 3 Townsville MC, 4810, Queensland, Australia.
e.evansillidge@aims.gov.au



Superimposed with bacterial symbiont diversity

R 5892.2 8.18 1.7 85.9 R 1.76 586 5825.024685863.115355 R 116.06 536 5825 R 1325 R 5 175 g(R)6(130) JET 1220172686 5825.022812c4246 802 10 1218.76 586 58258.24 m43.14 1218.767 85144123823643507507283/



Natural Products “Renaissance” (with a twist)

Paterson and Anderson, *Science* 21 October 2005

A significant cumulative effort in ocean exploration

‘Parents’ of marine science:

- Indigenous observations over millennia
- Aristotle 384-322BC
- Charles Darwin
HMS *Beagle* 1831-36
- Challenger 1872-75



Where is the data? How can we access it?

- Specialist data
- Published Literature
- Portals for metadata
- Networked datasets
- Integrated datasets

The published literature

Literature Databases

- Cambridge Scientific Abstracts
- Aquatic Sciences & Fisheries Abstracts
- Zoological Record
- Biosis
- MarineLit
- Patent databases

Citation & Indexing services

- Web of Science
- Google Scholar

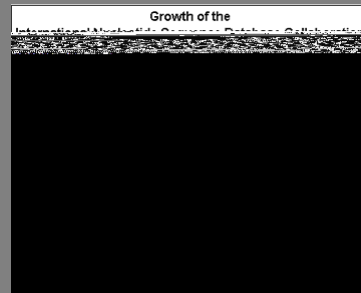
Citation data management

- Ref works
- Endnote



Data Networks/Repositories Nucleotide sequences

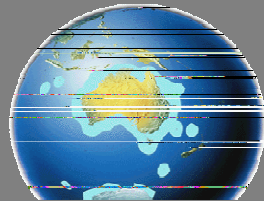
- **INSDC** (international Nucleotide Sequence Database Collaboration)
www.insdc.org
 - **GenBank[®], DDBJ, EMBL**
 - Annotated collection of all publicly available DNA sequences
 - Submission of sequences required by many journals prior to publication
 - Online submission, update and review
 - Country of origin identified
 - No restriction on use or distribution
 - 73078143 loci, 77248690945 bases, from 73078143 reported sequences (June 15 2007)
- **Data searching & analysis tools**
 - eg. BLAST



www.ncbi.nlm.nih.gov/Genbank/

Specialist data – find the people

Location of specimen and data holdings
reflects historical location of the specialists





UN Atlas of the Oceans

<http://www.oceansatlas.org/>

- UN-Oceans coordination portal
- 14 Global partners and 8000 individual members.
- Peer reviewed.
- Features a Virtual Office, advanced Search, web statistics and monthly newsletter. FREE TO MEMBERS
- New marine genetics page

National data network example

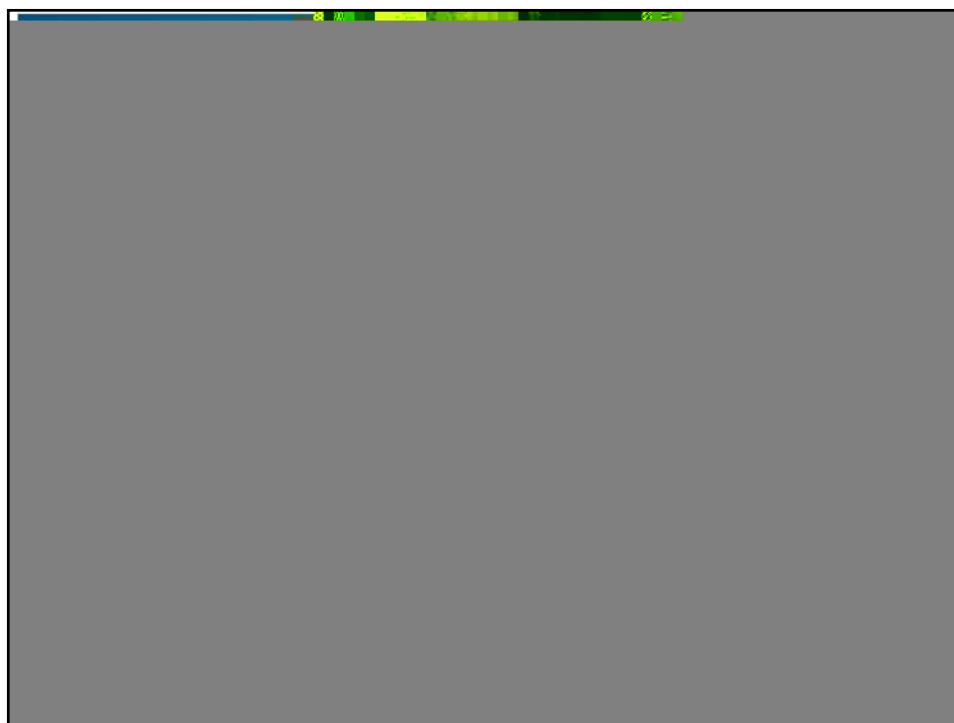


Name <ul style="list-style-type: none">- APNI- Derivation	<i>Castanospermum australe</i> <p>Belongs to the Fabaceae family, commonly referred to as Black Bean or the Moreton Bay Chestnut.</p>	
Description	Beautiful large evergreen tree with glossy dark green pinnate leaves and low spreading branches when grown in the open. During Spring it bears sprays of red and yellow pea-shaped flowers. The large cylindrical pods which are produced in autumn and have 3-5 large bean like seeds inside. The nectar produced by the flowers attracts birds, bats and butterflies.	
Classification <ul style="list-style-type: none">- Barcode of Life- Cladogram- Close relatives	Tree, to 35 m high with trunk up to 1.5 m diam., but usually smaller, mostly glabrous. Leaves 30-60 cm long; leaflets 9-17, obovate-elliptic, 7-20 cm long, 3-5 cm wide, apex acuminate, margins entire, glabrous, upper surface glossy, lower surface paler and dull; petiole 3-6 cm long; lateral petioles 2-5 mm long. Racemes mostly 5-15 cm long; pedicels 20-35 mm long. Calyx c. 10 mm long. Corolla 30-40 mm long, orange to red. Pod mostly 10-20 cm long, 4-6 cm diam., glabrous; seeds 1-5, c. 30 mm diam., brown. Flowers spring.	
Illustrations <ul style="list-style-type: none">- Photographs- Line Drawings		
Identification tools <ul style="list-style-type: none">- Australian Tropical Rain Forest Plants	Ecology <p>Grows in well developed rain forest, but is often found in gallery forest along creeks and rivers.</p> <p>Source: Flora of NSW and Rain Forest Key</p>	Distribution <p>Occurs in Cape York Peninsula (CYP) and North East Queensland (NEQ), and southwards to north eastern New South Wales. Altitudinal range in CYP and NEW from near sea level to 800m. Also occurs in New Caledonia and Vanuatu.</p>
Usage <ul style="list-style-type: none">- Horticulture		
Ethnobotany		
Pharmacology		
Current Research		
Bibliography		

- Searchable database, common format
- all known Australian plant & animal species by 2015
- terrestrial and marine

The slide features a header with the word "Atlas" in a purple box. Below it, five boxes list data sources: "Data from taxonomic names lists eg. APNI, AFD, ITIS, Species2000", "Specimen data from museum & herbarium & culture collections, vouchered specimens used for DNA analysis", "DNA sequence data held in DNA banks and barcodes from CBOL projects", "Character datasets: e.g., morphology, biochemistry, growth and", and "From image banks and repositories". The main content area is a large black rectangle.

- It's expensive (\$40M, over 5 years)
- Consistency in taxonomy is an issue

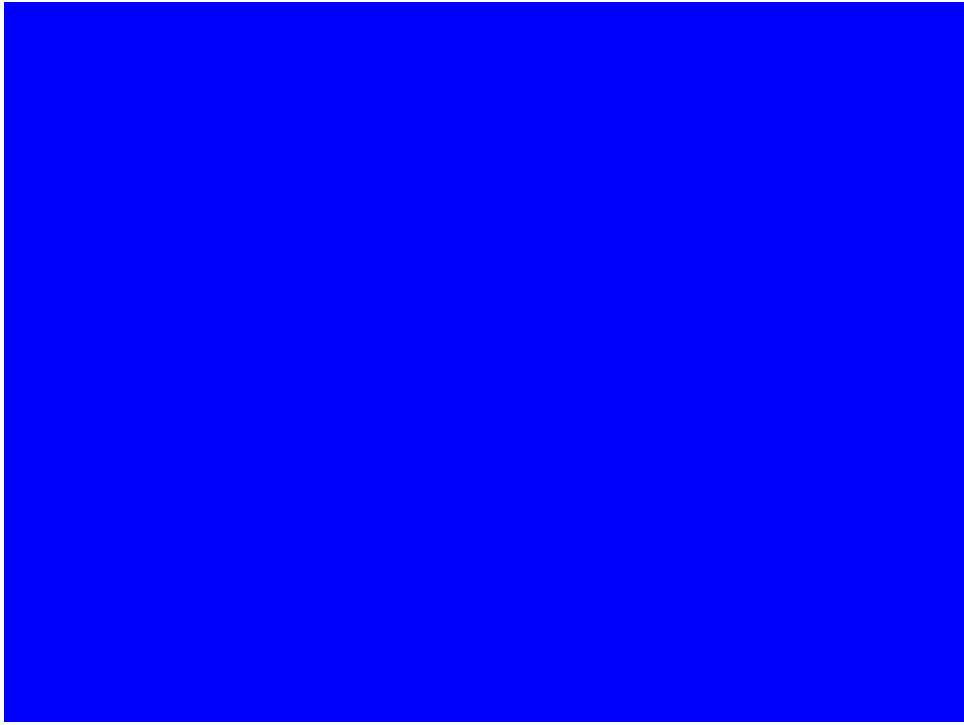


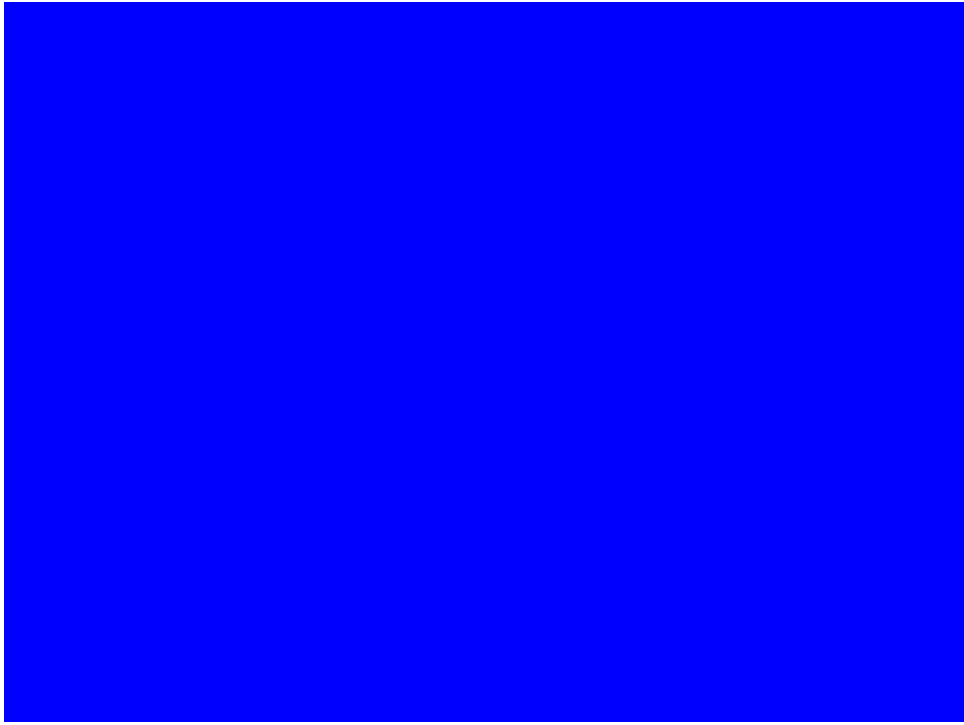
- Growing global network
- Assess and explain the diversity, distribution and abundance of marine life in the world's oceans
- past, present and future
- 2000 to 2010
- 50+ countries, 300+ scientists, 17 major projects
- Link to the Barcode of Life initiative

Oceans Present: Realm Projects

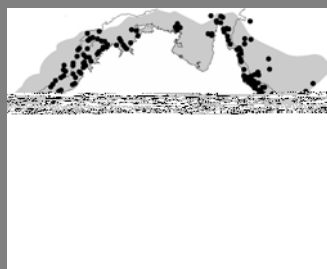
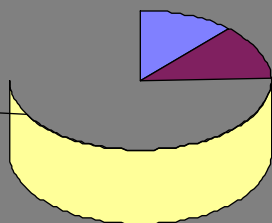
CoML defines its realms & zones in 2003 Baseline Report, *The Unknown Ocean*



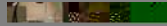




AIMS Bioresources Library – integrated MGR dataset



A Bioresources Library for screening and biodiscovery (past, present, future)



Cancer

Environmental
remediation

Viral

Industrial Enzymes

Antibiotics

Agrichemical

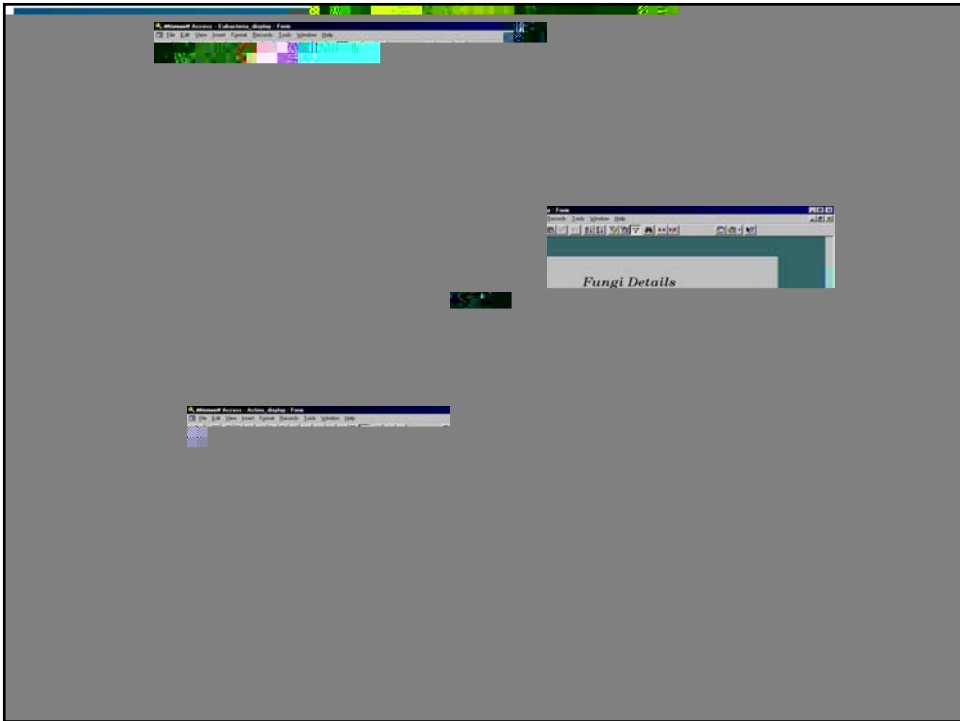
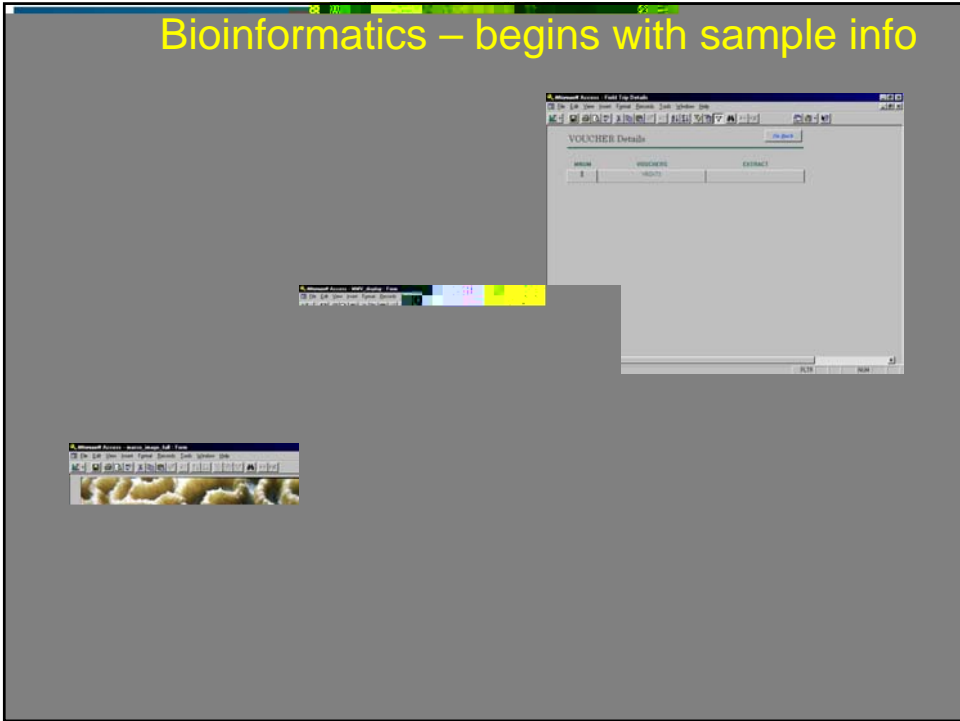
Toxin detection

Paints

Mineral Processing

Central Nervous System

Bioinformatics – begins with sample info



The screenshot shows a multi-windowed software application. At the top, there are two windows: the left one is titled 'Peyton Spectrum' and the right one is titled 'Access'. Below these, there is a window titled 'Chemistry' and another window titled 'Access' at the bottom right. In the center of the main workspace, a ball-and-stick molecular model is displayed, showing a complex organic structure with grey, red, blue, and white atoms connected by bonds.

...integrated with other research outputs

Integrated tool for data mining.

eg. Regional and taxonomic patterns in anti-microbial activity



Analysis at a range of scales



Google Earth interface facilitates good visualisation of multi-variate data

Understanding the chemical ecology Apply data-mining to enhance biodiscovery

- Elaborate leads
 - identifying other material with similar taxonomy/ecology/ screening profile/chemistry
 - Naturally occurring analogues
 - Re-supply without re-collecting
- Predict results of future screening based on past profile
 - Compile list of pre-leads
 - Targeted biodiscovery with ex-situ material

Summary

- Tools exist to access marine biodiversity and genetic resources data that is in the global public domain
- Major networking projects underway to bring together independent geo-referenced datasets (CoML, ALA)
- Consistency in taxonomy is a big issue in networking independent datasets
- Integrated informatics is the ideal
 - Integrate biodiversity data (ecosystems, species (macro and micro), genomes (and meta-genomes)), with natural products research outputs (instrument outputs, structures of compounds, proteins, enzymes etc), and screening results

