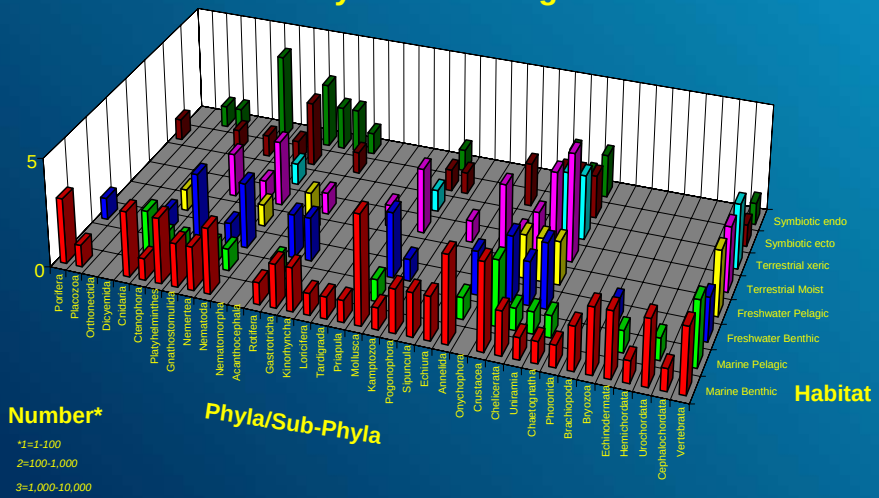
AIMS: Australia's tropical marine research agency.

Towards a practical knowledgebase for marine genetic resources

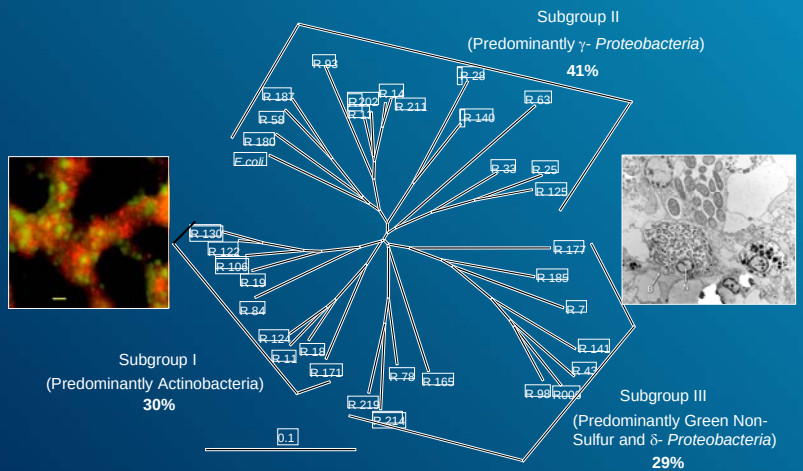
Libby Evans-Illidge

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

The seafloor is the most biodiverse place on earth - Based on diversity of macro-organisms



Superimposed with bacterial symbiont diversity



eg. Microorganisms found in the stable community within the Great Barrier Reef sponge *Rhopaloeides odorabile.* (Nicole Webster)

Marine Natural Products and Related Compounds in Clinical and Advanced Preclinical Trials

Table 1. Status of Marine-Derived Natural Products in Clinical and Preclinical Trials

name	source	status (disease)	comment	
didemnin B	*Tridemnum solidum*	Phase II (cancer)	dropped middle 90s	
dolastatin 10	*Dolabella auricularia*(marine microbe derived;cyanophyte)	Phase I/II (cancer)	no further trials	
girolline	*Pseudaxinyssa cantharella*	Phase I (cancer)	discontinued	
bengamide	*Jaspis* sp.	Phase I (cancer)	licensed Novartis	
cryptophycins	*Nostoc* sp. & *Dysidea arenaria*	Phase I (cancer)	licensed to Lilly	
bryostatin 1	*Bugula neritina*	Phase II (cancer)	licensed to	GPC Bio
dolastatin	*Dollabella*	Phase II (cancer)	melanoma, breast,	
ecteinascidin	*Ecteinascidia turbinata*	Phase II/III (cancer)	licensed to J&J	
aplidine	*Aplidium albicans*	Phase II (cancer)		
E7389 (Hali B)	***Lissodendoryx* sp**	**Phase I (cancer)**	**Eisai**	
discodermolide	*Discodermia dissoluta*	Phase I (cancer)	licensed to Novartis	
kahalalide F	*Eylsia rufescens/Bryopsis* sp.	Phase II (cancer)	licensed to PmMar	
Spisulosine	*Spisula polynyma*	Phase I (cancer)	*Rho*-GTP inhibitor	
HTI-286	*Cymbastella* sp	Phase II (cancer)	licensed to Wyeth	
KRN-7000	*Agelas mauritianus*	Phase I (cancer)		
squalamine	*Squalus acanthias*	Phase II (cancer)	antiangiogenic	
Laulimalide	*Cacospongia mycofijiensis*		preclinical (cancer)	
Curacin A	*Lyngbya majuscula*	preclinical (cancer)		
vitilevuamide	*Didemnum cuculiferum &Polysyncraton lithostrotum*	preclinical (cancer)		
diazonamide	*Diazona angulata*	preclinical (cancer)		
eleutherobin	***Eleutherobia* sp.**	**preclinical (cancer)**		
sarcodictyin	*Sarcodictyon roseum*	preclinical (cancer)		
peloruside A	***Mycale hentscheli***	**preclinical (cancer)**	**Just licensed**	
salicylihalimide	***Haliclona* sp.**		**preclinical (cancer)**	
thiocoraline	*Micromonospora marina*		preclinical (cancer)	
variolins	***Kirkpatrickia variolosa***		**preclinical (cancer)**	
dictyodendrins	*Dictyodendrilla verongiformis*		preclinical (cancer)	licensed to Taiho
manoalide	***Luffariaella variabilis***		**Phase II**	**discontinued**
IPL-576,092	*Petrosia contignata*	Phase II		licensed to Aventis
ziconotide	***Conus magus***		**Phase III pain**	**licensed to WL**
CGX-1160 plus	***Conus geographus, catus, victoriae***	**Phase I (pain)**		

Journal of Natural Products Reviews

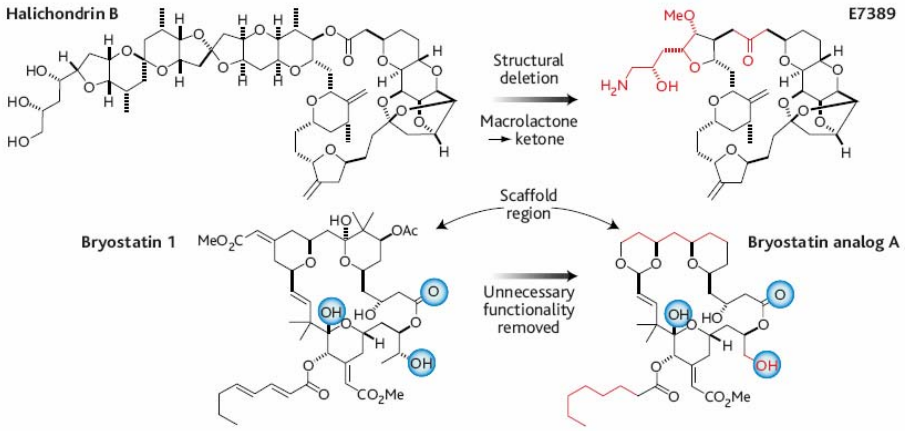
David J. Newman* and Gordon M. Cragg
Natural Products Branch, Developmental Therapeutics Program, NCI-Frederick, P.O. Box B, Frederick, Maryland 21702

Received February 10, 2004; see also Fortman and Sherman CC 2005

@ 38% from Australasia

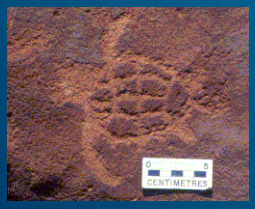
Cytarabine (Ara-A,C) from *Cryptotheca crypta* now routine for leukaemia and lymphoma patients, Glaxo Smith Kline and Pfizer

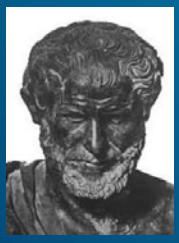
Natural Products "Renaissance" (with a twist)

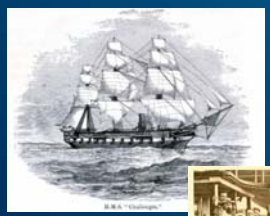


Paterson and Anderson, Science 21 October 2005

A significant cumulative effort in ocean exploration





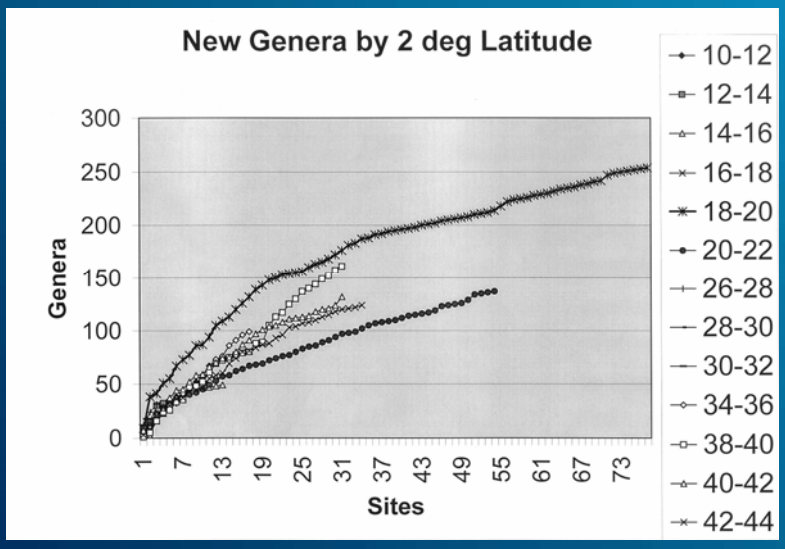


'Parents' of marine science:

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

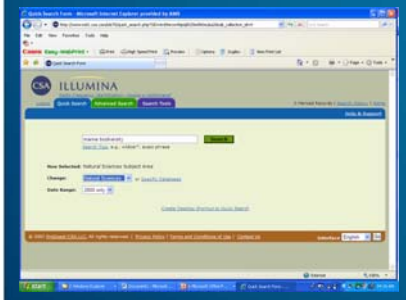
Marine R&D technology in the 21st Century

…but still so much remains unknown



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

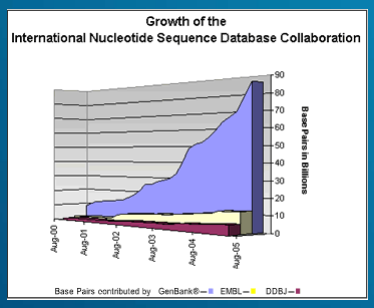
- Web of Science
- Google Scholar

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)



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

- Data searching & analysis tools
  - eg. BLAST

Specialist data – find the people

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

eg. Scleractinian corals

•Major repository at Museum of Tropical Queensland, Townsville Australia

•Doorstep of Great Barrier Reef

•Home of 2 global experts - Dr Carden Wallace (MTQ) and Dr John Veron (AIMS, Townsville)

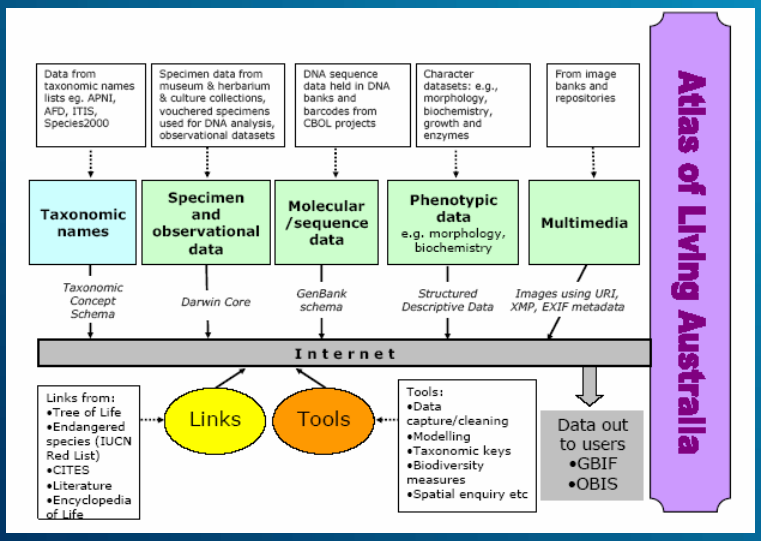
UN Atlas of the Oceans

http://www.oceansatlas.org/

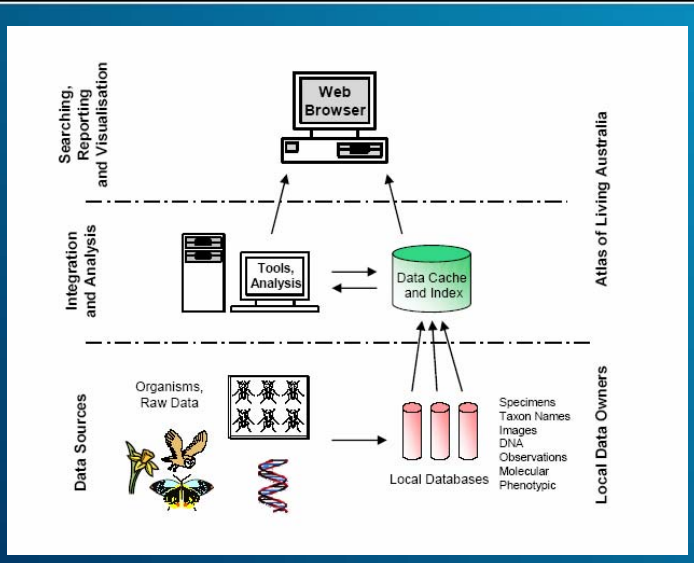
- UN-Oceans coordination portal
- 14 Global partners and 8000 individual members.
- Peer reviewed.
- Features a Virtual Office, advanced Search, web statistics and monthly newsletter. FREE TO MEMBERS
- New marine genetics page

National data network example

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



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



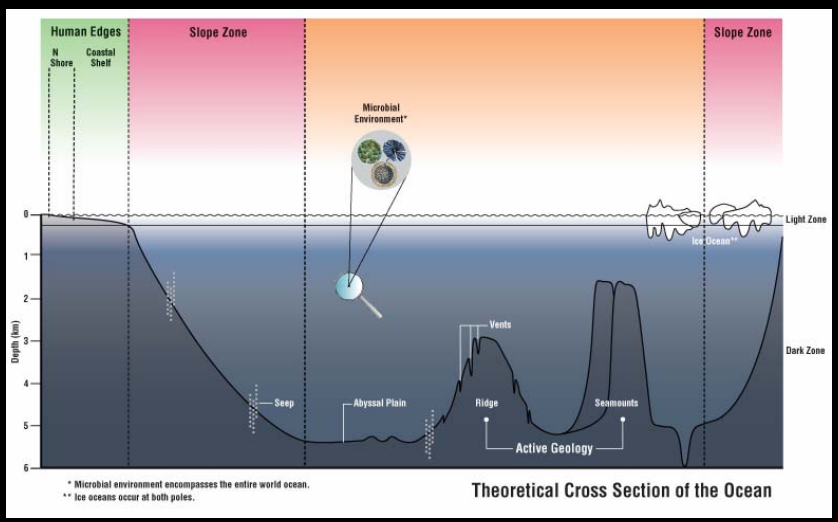
CENSUS OF MARINE LIFE

- 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

Australian Government | Australian Institute of Marine Science

Oceans Present: Realm Projects

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



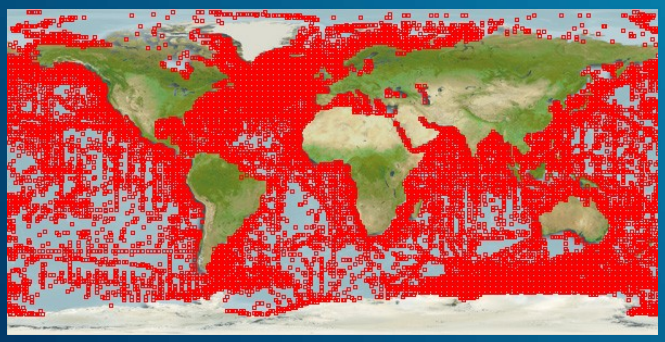
Ocean Realm Field Projects

Human Edges	Nearshore: NaGISA Reefs: CReefs Regional Ecosystem: GoMA Continental Shelf: POST
Hidden Boundaries	Continental Margins: COMARGE Abyssal Plains: CeDAMar
Central Waters	Zooplankton: CMarZ Top Predators: TOPP Ridges: MAR-ECO
Active Geology	Vents & Seeps: ChEss Seamounts: CenSeam
Ice Oceans	Arctic: ArcOD Antarctic: CAML
Microscopic Ocean	Microbes: ICOMM

Ocean Biogeographic Information System

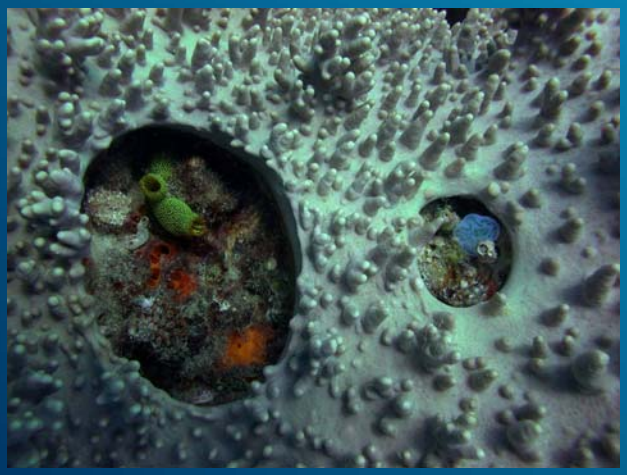
(OBIS) Portal **www.iobis.org**

OBIS strives to document the oceans' diversity,distribution, and abundance of life.

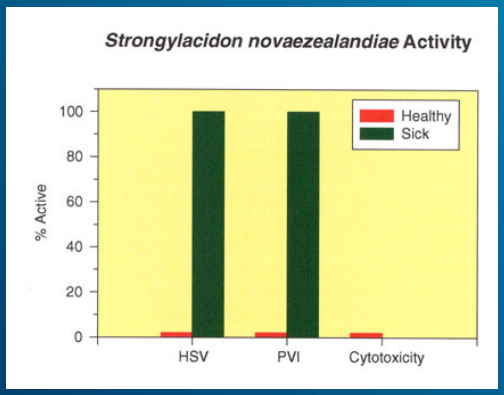


13.2 million records, 81000 species, 210 databases
Searchable on name or geography
Consistency in taxonomic assignment is an issue

Assessment of Marine Genetic Resources
- You need to know more than 'where they occur'



Marine genetic resources are the integrated result of dynamic, complex ecology.

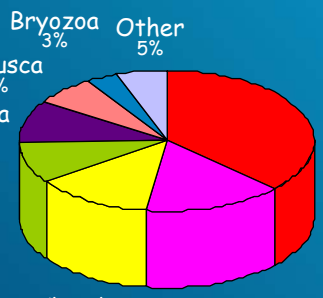


Battershill, unpublished

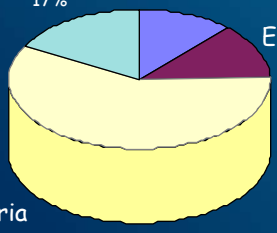
eg.
- Activity of interest (HSV and PVI) in this example is only produced by sick sponges
- Note high selectivity to the target (and absence of cytotoxicity)

AIMS Bioresources Library – integrated MGR dataset

12000 macros



9000 micros



1987 – present

Over 1500 sites around Australia

COLLECTION DATA

- **Precise location**
- **Depth**
- **Size**
- **Best taxonomy (voucher back-up)**
- **Imagery**
- **Coded descriptions**
  - **habitat**
  - **morphology**
  - **substrate**
  - **colour(external/internal)**
  - **epibionts**
  - **symbionts**
- **Descriptive notes and observations**
- **Site information (location, habitat description, community composition)**

Micros:

- **Details of source and selective isolation**

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

Cancer

Viral

Antibiotics

Central Nervous System

Agrichemical

Paints

UV Blocking

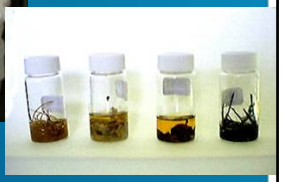
Mineral Processing

Toxin detection

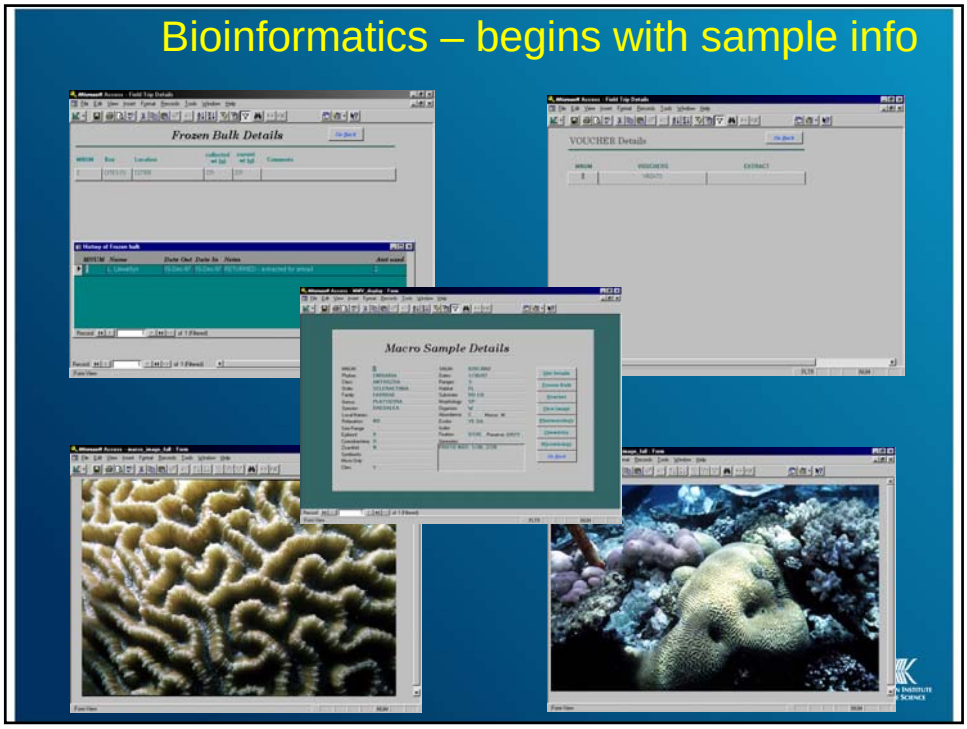
Industrial Enzyme

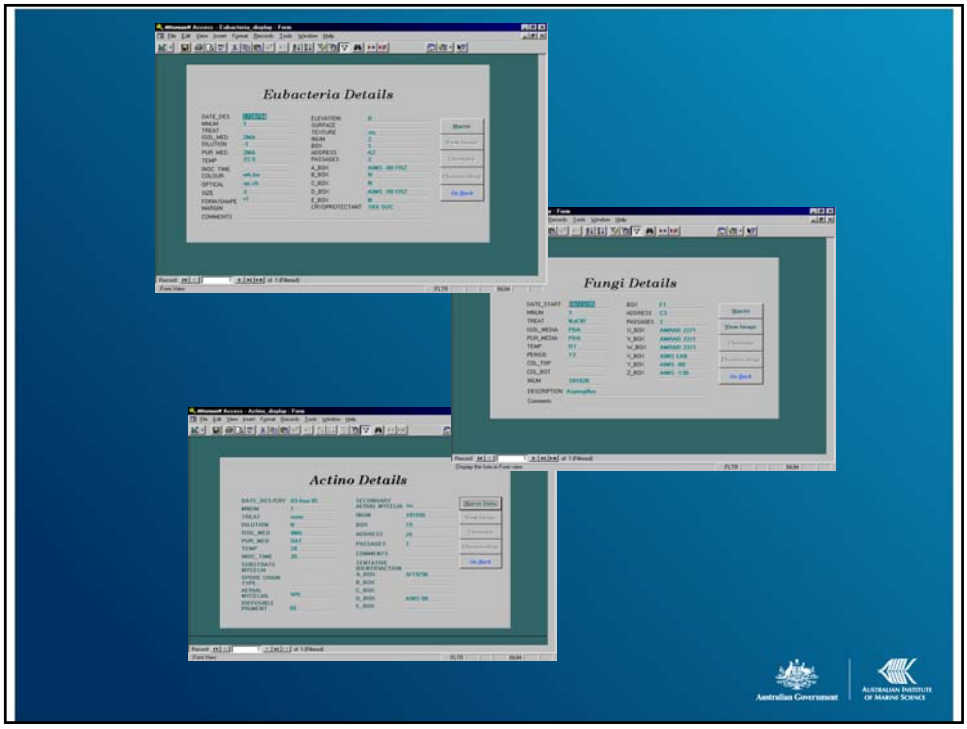
Environmental remediation

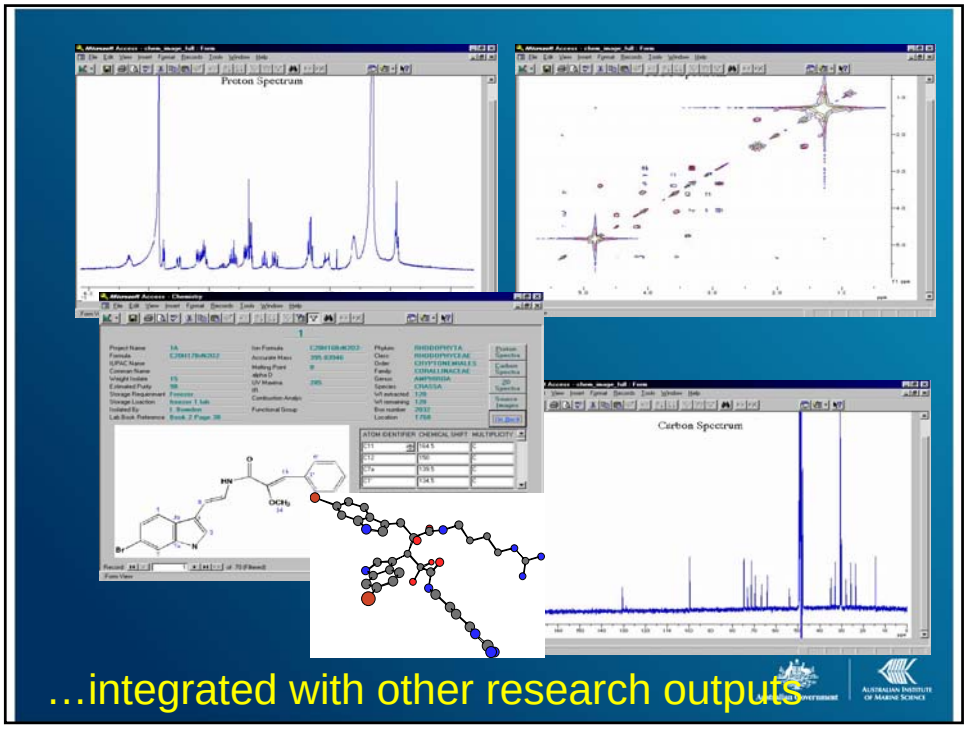
Samples change format (and generate data) many times

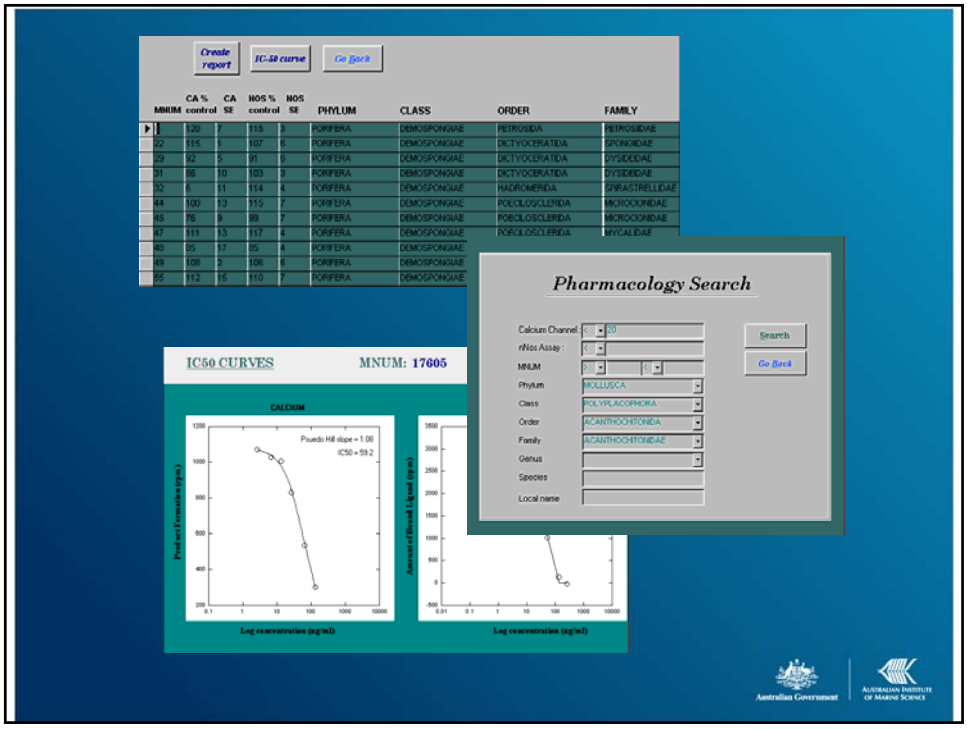


Bioinformatics – begins with sample info





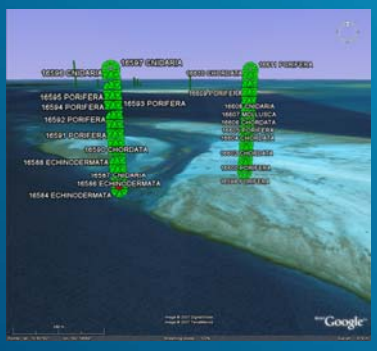

Proton Spectrum
Carbon Spectrum
…integrated with other research outputs


Create report
IC-50 curve
Go Back
MNUM
CA % control
CA SE
NOS % control
NOS SE
PHYLUM
CLASS
ORDER
FAMILY
Pharmacology Search
Calcium Channel
nNos Assay :
MNUM
Phylum
Class
Order
Family
Genus
Species
Local name
Search
Go Back
IC50 CURVES
MNUM: 17605
CALCIUM

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

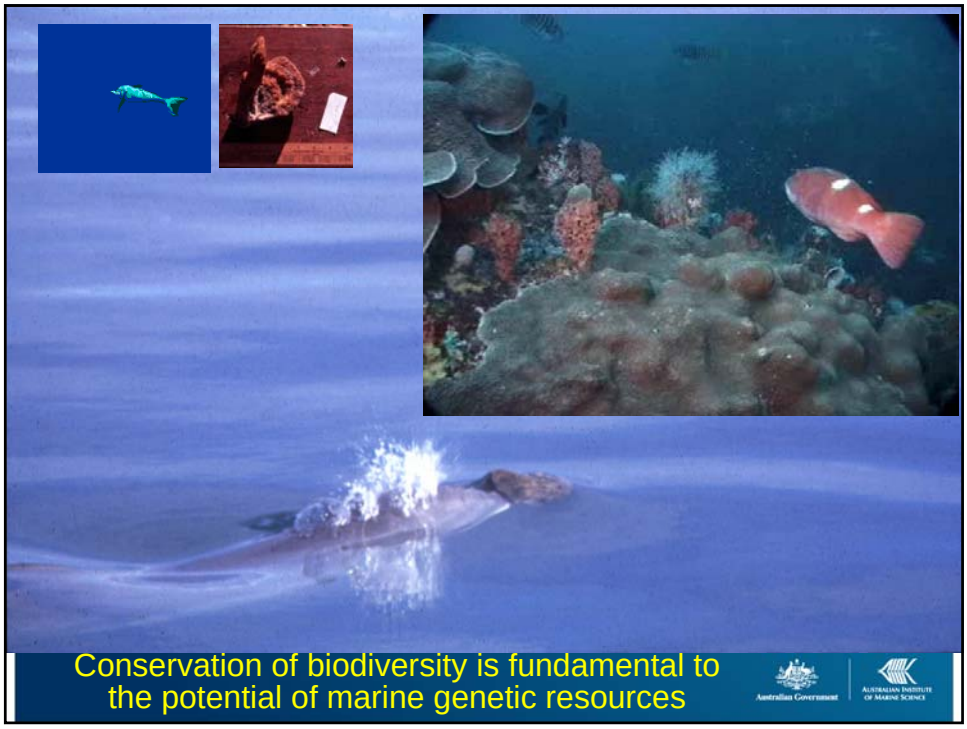
Understanding the chemical ecology
Apply data-mining to enhance biodiscovery

- 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
  - Powerful datamining tool to assess and understand marine genetic resources
  - 'Commercial in Confidence' issues, but even a compilation of non-commercially sensitive information is useful


Conservation of biodiversity is fundamental to
the potential of marine genetic resources