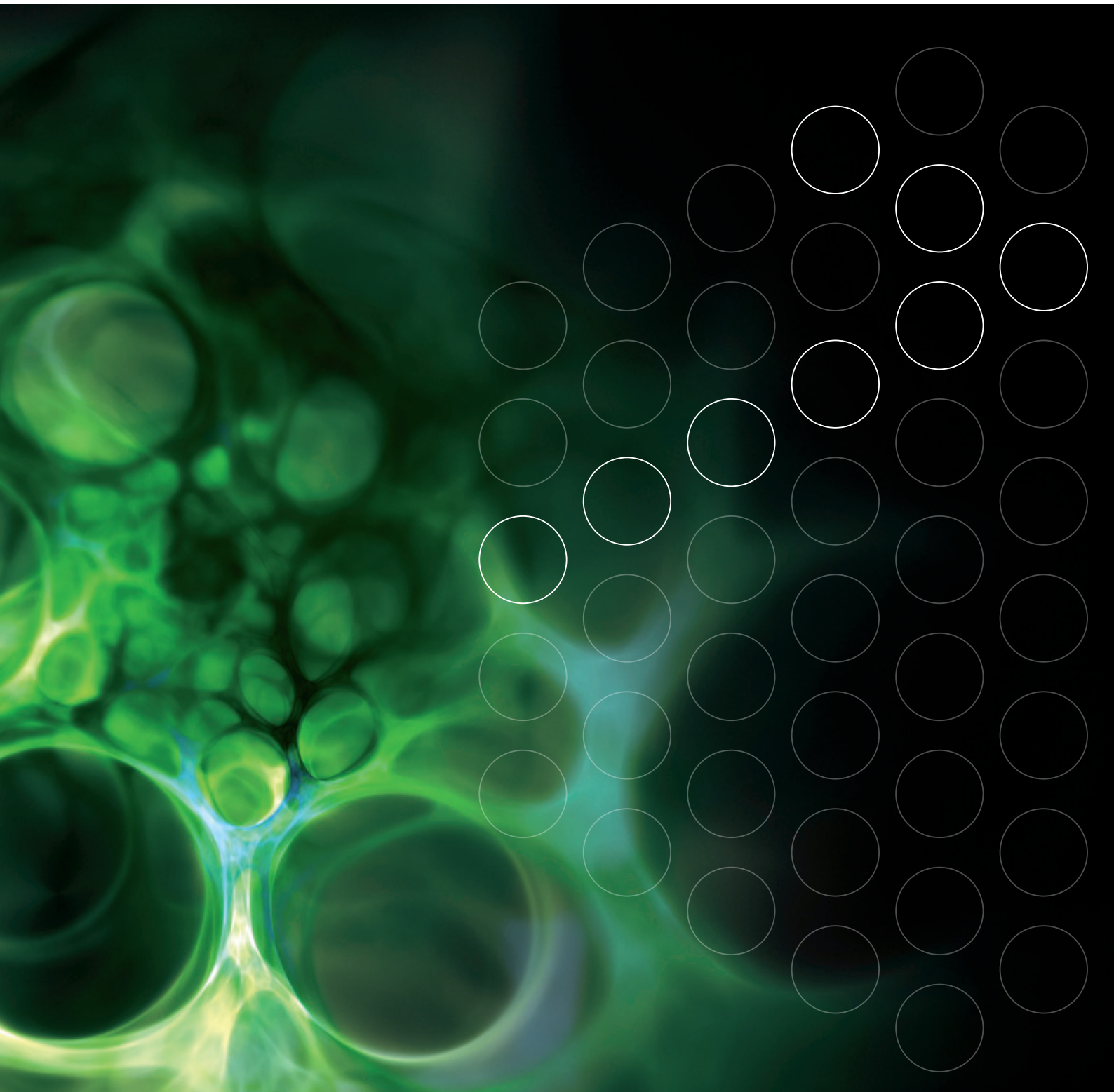


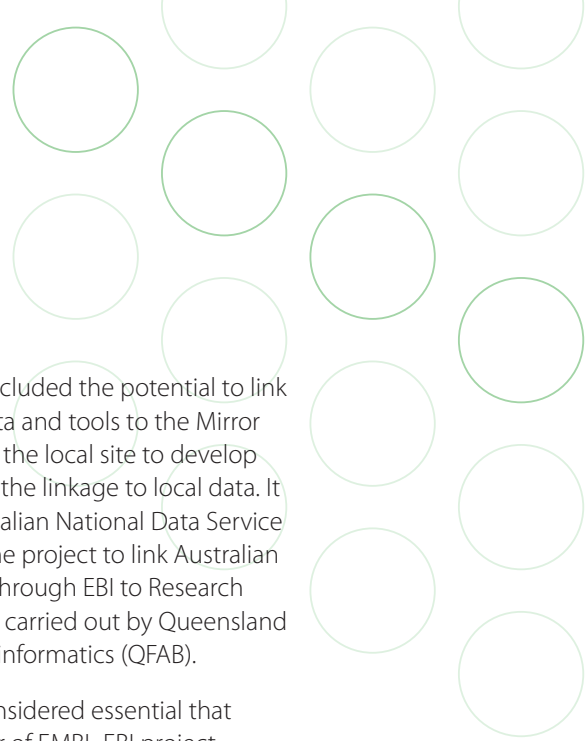
EMBL
Australia



Bioinformatics for the Future – *An update*

August 2011





Executive Summary

EMBL Australia hosted meetings in Brisbane and Canberra to continue the work done in 2009 which explored how Australia's future science can benefit from the developments in bioinformatics.

The current series of workshops also served to present the establishment of the EMBL Australia mirror of EMB-EBI.

Professor Nadia Rosenthal, Head, EMBL Mouse Biology Unit (Monterotondo); Director, Australian Regenerative Medicine Institute; and Head of Science, EMBL Australia, presented an overview of EMBL Australia and the objective to develop a network of laboratories around a central technology core at four centers in Australia.

Dr Ewan Birney, Senior Scientist, EMBL at European Bioinformatics Institute (EBI) presented a talk entitled "Bioinformatics for the Next Decade" which highlighted some of the challenges facing life science research in relation to the quantum of data production, the need to consider what should be stored, how to address the data repositories and generate value from these resources.

As part of the program Dr David Green, Implementation Manager of the EMBL Australia mirror of EMBL-EBI presented an overview of the status and functionality of the establishment of the Mirror. The EMBL Australia Mirror of EMBL-EBI is an important new component of Australia's national collaborative research infrastructure. Replication of EBI data, tools and services in Australia will enable research outcomes through high-bandwidth connectivity, programmatic access, proximity to high-performance research computing, and the eventual inclusion of customised and novel data services driven by Australian research priorities.

An open forum was held at each site to discuss and receive the thoughts of users of the EMBL-Australia mirror and help determine the future deployment of features. There was general agreement that local availability and access to the larger datasets will help Australian science. Particular reference was made to mirroring the 1000 Genome dataset. This presents technological challenges and consideration was given to holding a subset of the data and frequency of update. Visualisation using a UCSC type genome browser or linkage to UCSC tracks was considered useful.

Other important ideas included the potential to link Australian developed data and tools to the Mirror with encouragement for the local site to develop its own identity through the linkage to local data. It was noted that the Australian National Data Service funded component of the project to link Australian derived data deposited through EBI to Research Data Australia was being carried out by Queensland Facility for Advanced Bioinformatics (QFAB).

As a final point it was considered essential that the EMBL Australia mirror of EMBL-EBI project communicate with the users through a number of channels, such as through a chat room, to ensure it built on the local needs and develop a strong culture and following within Australia.

Introduction

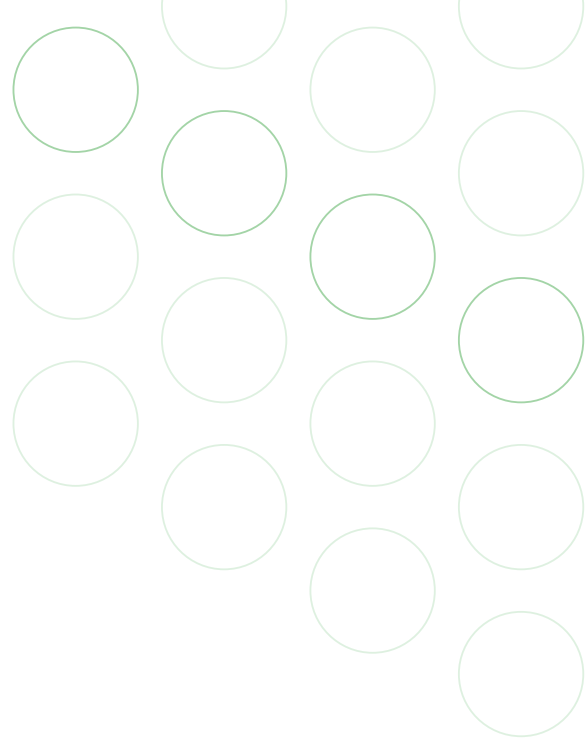
In 2009 I visited Australia and had extensive exposure to the established and sometimes nascent bioinformatics present in Sydney, Melbourne, Canberra, Brisbane, Adelaide and Perth. In 2011 I returned to Australia for a smaller itinerary (Brisbane and Melbourne; Canberra would have been included but for Chilean volcano reasons) but with good input from a variety of different Australian sites; a particular focus was the launch of the EMBL Australia Mirror of the EMBL-EBI mirror in Brisbane.

In this report I will discuss my main impressions and thoughts on this trip, which range from the strategic through to the specific around the EBI mirror. One clear aspect from 2009 is that this area has become more cohesive as a strategic need in Australia in which a collaborative group of scientists across different geographic locations are working towards sustainable solutions.



Dr. Birney is a Senior Scientist at the European Molecular Biology Laboratory working at the European Bioinformatics Institute. He best known as one of the founders of Ensembl. In 2007 his group

merged with Rolf Apweiler's group to form the large PANDA group (Protein and Nucleotide Databases). He has strategic responsibility for all DNA databases. He works closely with a number of independent team leaders at the EBI who have direct responsibility for Ensembl, Ensembl Genomes and the European Nucleotide Archive. He is active with multiple genome projects (in particular Human and Mouse), the Ensembl and Ensembl Genomes projects and the ENCODE project to define functional elements in the Human genome. His main area of research include functional genomics (such as ENCODE), assembly algorithms and statistical methods to analyse genomic information, in particular information associated with individual differences.



Improvements since 2009

A number of key areas stand out as improvements since 2009.

1. The establishment of the EBI mirror in UQ and the depth of service level bioinformatics which occurs at UQ provides a natural centre of gravity for bioinformatics services/data resources in Australia. This centre of gravity for one aspect of bioinformatics should not be confused with a lack of excellence at other locations in Australia (indeed, an outsider's criticism of Australian science is that participants can often be more interested in inter-state/city rivalries than focusing on excellence in science. I found this "backdrop" less intrusive on this visit).
2. At least for the meetings that I attended there was a sense of common purpose across the different institutions (eg, Universities and CSIRO) and states as shown by the strategy meeting conducted by the EMBL Australia Bioinformatics Advisory Committee in UQ and the meeting in Melbourne.
3. There are an increasing number of internationally competitive genomics/big biology data projects occurring in Australia (Cancer genome project, Mouse phenome projects) which demand high end bioinformatics
4. There is a strong provision of storage, compute and bandwidth nationally with particular concentrations in particular cities

Continued areas of concern since 2009

1. As noted in my report in 2009, much of bioinformatics “provision” is the mundane business of hiring good people to work on good scientific projects. As in all locations I know, capital funds for hardware can often be easier to release than personnel funds; this imbalance is almost chronic in Australia. I was informed of pleasing changes in the NCRIS concept of “infrastructure” to include this “personnel” component, but this transformation has not occurred yet. The lack of investment in bioinformatics personnel is the single biggest restriction on large scale biological science growth in Australia. Addressing this should remain the top concern of both the funding agencies and the strategic leadership in this area. Occasionally there were comments made about providing more “synergy” across existing projects, but largely I think this is misleading; bioinformatics personnel should of course communicate between themselves productively, but when a new project occurs, even if it is largely similar processing, the actual doing, processing and checking of the results has to occur, and there is a large non-automatable component of this (the analogy is with experimental procedures; for example, cloning DNA is routine, but the fact that someone in Adelaide has mastered cloning of cDNA in plants does not mean that the researcher in Brisbane can clone animal cDNA with nobody to actually do it; of course, the researchers should learn from each other’s experiences, but this process of learning does not remove the need for someone to actually do it).
2. I still believe, as I did in 2009, that a minimal goal is to have 5 to 6 centres of bioinformatics excellence, each with at least 10 bioinformaticians in a coordinated manner, and one site with at least 20 bioinformaticians, if not more. It seems likely, given the Australian funding landscape that the funding would be a mixture of Commonwealth and State funding. It is also clear that having a standard governance structure would be too restrictive given the diversity of local institutions and systems; however, some principles, for example of co-location of these bioinformatics groups should apply.
3. There has been no movement that I can see on a yearly meeting of bioinformatics groups in Australia (probably best merging with the bioinformatics “end” of genomics projects). I think a simple yearly meeting, presumably rotating in locations but with a fixed time each year, would greatly help the nascent community to feel more integrated. I would strongly recommend that this happens
4. There is always an undercurrent of inter-state/ location rivalry. This is less on this trip, but should always be guarded against. One has to balance the fact that there should not be an expectation of identical resources occurring in all locations – after all different locations do have different strengths and depth of scientific support. However, there is excellent science, which occurs in all 7 major scientific locations in Australia, and a system which excludes one of these areas is clearly sub-optimal. To manage this correctly the strategic leadership of bioinformatics in Australia has to be cohesive across these locations whilst not falling into the trap of having simply equal divisions of resources or functionality.

Specific comments on the EBI Mirror

1. The EBI mirror launch is an excellent start for an important component of the bioinformatics service provision in Australia
2. Despite the unity of the web as a delivery mechanism, there is considerable diversity in the delivery components of different EBI projects – mainly due to the need to allow operational freedom across the ~300 service personnel at the EBI broken out into ~20 operations, but also due to the different web of collaborative agreements unique to each project. As such, there is not a single “push button” transfer system of the whole of the EBI, although improvements in the service provision at the EBI is increasing the uniformity of some of our processes. The consequence of this is that each major service has to be considered individually in terms of mirroring requirements.
3. The consequence of (2) is that a careful analysis of costs vs benefits of mirroring services needs to occur on a service by service basis. It is natural that the major benefits will be those projects which have high bandwidth or low latency requirements; this leads to projects such as the 1,000 genomes project, the ftp site and web services providing programmatic access. Given the feedback from the users, the 1,000 genomes project will be a very sensible place to start.
4. The EBI is going through a web redesign process in which the main concepts behind the redesign are finalised in October 2011, and the expected launch of components of the redesign will happen Q1/Q2 in 2012. At this time the UQ mirror must update in anycase, and we would encourage the EMBL Australia and UQ to find an independent but collaborative branding strategy for the website; this would also allow more unique Australian services to be run from the site.
5. By switching away from a “straight mirror” to more a “bioinformatics service portal” this doesn’t lessen the commitment to having a strong collaborative relationship with the EBI – indeed, this arrangement would be the best for the EBI and the best for Australia, EMBL Australia and UQ.

Ewan Birney

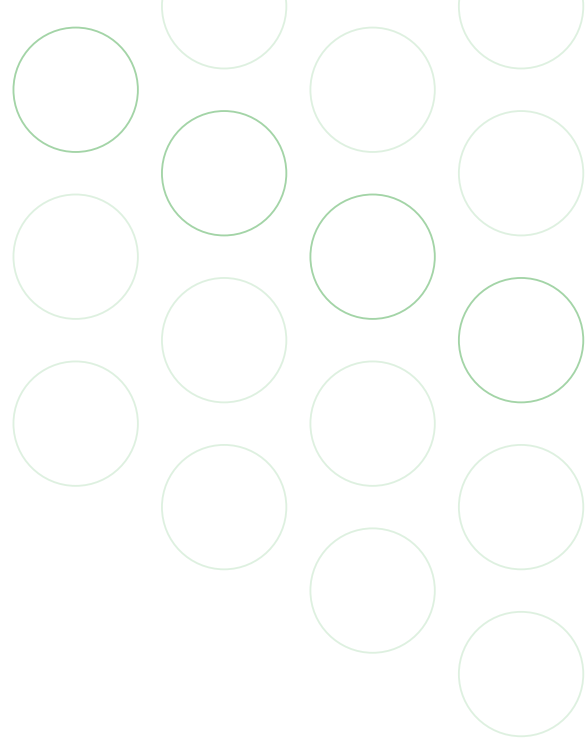
Brisbane

Auditorium, Queensland Bioscience Precinct
The University of Queensland
306 Carmody Rd
St Lucia Queensland 4072

Tuesday, 21 June 2011
1.00pm to 3.00pm

List of Attendees

Anderson, Matthew	<i>Bioinformatics Programmer,IMB</i>
Barker, Jeremy	<i>CEO, QFAB</i>
Bauer, Denis	<i>Post Doc, QBI</i>
Beau, Annabelle	<i>Trainee, QFAB</i>
Bellgard, Matt	<i>Director, CCG, Murdoch Uni</i>
Benzakour, Nouri	<i>Post Doc, SCMB</i>
Berkma, Paul	<i>Student, UQ</i>
Boden, Mikael	<i>Senior Lecturer, SCMB & ITEE</i>
Chan, Kenneth	<i>Research Officer, SAFS</i>
Chaumeil, Pierre	<i>RA, QFAB</i>
Coquery, Jeff	<i>Intern, QFAB</i>
Dietzgen, Ralf	<i>Associate Professor, QAAFI</i>
Fernandez, Selene	<i>PhD Student, IMB</i>
Fink, Lyn	<i>Post Doc, QCMC</i>
Gascoigne, Dennis	<i>PhD Student, IMB</i>
Glazov, Evgeny	<i>Bioinformatics Support, U/LifeTech</i>
Gloaguen, Yoann	<i>Intern, QFAB</i>
Hogan, Ben	<i>PI, IMB</i>
Holmes, Oliver	<i>Dev, IMB</i>
Hudson, Nick	<i>Scientist, CSIRO</i>
Kaas, Quentin	<i>Research Officer, IMB</i>
Kassahn, Karin	<i>Research Officer, IMB</i>
Lai, Kaitao	<i>PhD Student, SAFS</i>
Larbi, Mourad	<i>Intern, QFAB</i>
LeCao, Kim-Anh	<i>Biostatistician, QFAB</i>
Lee, Hong	<i>Research Officer, SAFS</i>
Legaie, Roxane	<i>Bioinformatician, QFAB</i>
Leo, Paul	<i>Scientist, DI</i>
Leong, Lex	<i>PhD Student, CSIRO</i>
Ling, Edmund	<i>Research Officer, SAFS</i>
Little, Bryce	<i>Software Developer, CSIRO</i>
Liu, Chang	<i>Visitor, IMB</i>
Liu, Gangiang	<i>PhD Student, IMB</i>
Lorenc, Michal	<i>PhD Student, SAFS</i>
Ma, Piyush	<i>PhD Student, IMB</i>
Marshall, Mhairi	<i>Bioinformatician, Diamantina Institute</i>
Mattick, John	<i>Professor, IMB</i>
McWilliam, Sean	<i>Bioinformatics Analyst, CSIRO</i>
Miotto, Amanda	<i>Researcher, GU</i>
Morahan, G	<i>WAIMR</i>
Olivier, Poirion	<i>Intern, IMB</i>



Palfreyman, Robin	<i>Post Doc, AIBN</i>
Pheasant, Michael	<i>Scientist, IMB</i>
Pradeep, R	<i>PhD Student, SAFS</i>
Quenette, Steve	<i>Strategy Manager, Monash Uni</i>
Ragan, Chikoko	<i>PhD Student, QBI</i>
Ragan, Mark	<i>Professor, IMB</i>
Reverter, Toni	<i>Scientist, CSIRO</i>
Robertson, Alan	<i>PhD Student, IMB</i>
Scott, Peter	<i>IT, QFAB</i>
Simons, Cas	<i>Research, QFAB</i>
Skippington, Elizabeth	<i>PhD Student, IMB</i>
Smith, Kelly	<i>Post Doc Fellow, IMB</i>
Smith, Martin	<i>PhD Student, IMB</i>
Song, Sarah	<i>Researcher, IMB</i>
Speed, Terry	<i>Scientist, WEHI</i>
Taft, Ryan	<i>Research Fellow, IMB</i>
Tellam, Ross	<i>Scientist, CSIRO</i>
Tu, Yun	<i>Visitor, IMB</i>
Uhlherr, Alf	<i>Senior Manager, CSIRO</i>
Vanichkina, Darun	<i>Graduate student, IMB</i>
Willadsen, Kai	<i>Research Officer, SCMB, UQ</i>
Williams, Sarah	<i>Bioinformatician, QFAB</i>
Wood, David	<i>Student, QCMC</i>
Wood, S	<i>Sys/Admin, IMB</i>
Yang, Zhe	<i>Post Doc, IMB</i>

Agenda Items

Tuesday, 21 June 2011		
1.00 pm	Introduction to EMBL and EMBL Australia)	Professor Nadia Rosenthal
1.15pm	Overcoming data volume problems in genomic biology	Dr Ewan Birney (Seminar)
2.00pm	EMBL Australia Mirror team: Introduction to the EMBL Australia Mirror of EMBL-EBI	Dr David Green
2.15pm	Open forum on how the Mirror can best serve Australian researchers	Facilitated by Jeremy Barker
3.00pm	Session concludes	

Key Points

Professor Rosenthal presented an overview of EMBL Australia and the connection to EMBL. She outlined the intention of EMBL Australia to support the development of a network of laboratories at four centers around Australia each with a technological core. The EBI Mirror is the technological core in Brisbane based at the University of Queensland.

The aim of the EMBL Australia mirror of the EMBL-EBI is to:

- Boost Australian research through access to data
- Replicate EBI services
- Link unique Australian services to EBI
- Open career opportunities
- Increase funds

For EMBL the most immediate objectives are to;

- Develop the resource
- Appoint a Director of the Australian Bioinformatics Network
- Recruit and embed bioinformatics teams in national initiatives
- Provide links to complementary national facilities
- Promote the mirror and the network
- Identify new funding opportunities

Dr Ewan Birney gave a presentation titled: "Bioinformatics for the Next Decade," which outlined the significant challenges presented by the large volumes of data being produced by genomic biology.

He identified four areas of challenge represented by the letters S,A,I,D:

- **S**ize, storage, security – growing exponentially
- **A**ccess, analysis
- **I**nterpretation, integration
- **D**iscovery, delivery

S governed by available space and the rate of data generation, with the community needing to decide which data is to be deleted;

A and I the ability to access and integrate the complex array of data is difficult and especially when trying to do this from a distance. Complexity provides the opportunity to ask comprehensive questions, however, the process is constrained by the inherent challenges presented by that complexity;

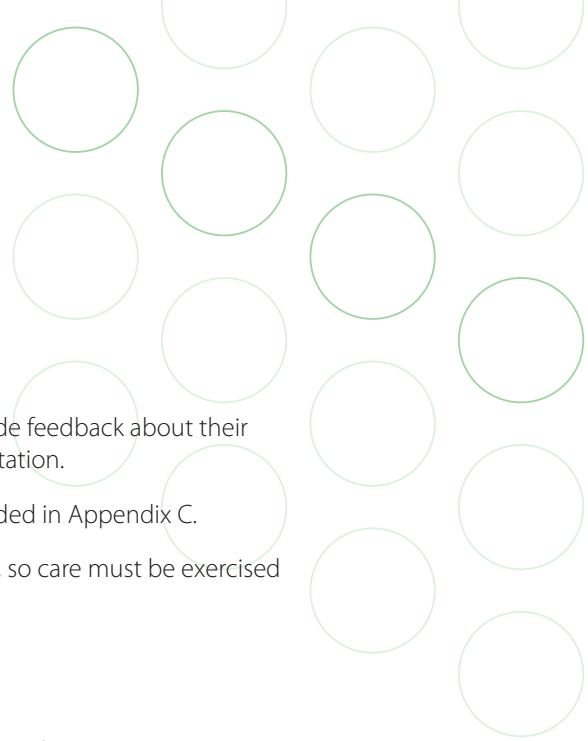
D delivery requires the resources and tools to make the data available with discovery only happening when researchers can use it effectively;

The linkage between the researcher and the data and assistance with this is an essential element of future life science. This can only happen effectively with people trained in both disciplines of information technology and biological science.

Dr David Green provided an overview of the website and demonstrated the local functionality of the EMBL Australia mirror of EMBL-EBI. Local content is identified through the Australian flag icon and the different header design. Any service which linked through to the UK site was shown with the European Union flag.

Appendix A shows a list of services and datasets that have been made available initially. He noted that the decision to set up this functionality was, in part, determined by the analysis of the Australian use of the EBI services at the home site as shown by the graph in Appendix B.

Dr Green provided a brief overview of some of the local functionality including a side by side comparison of the time to download a dataset from the local and remote FTP sites. This clearly demonstrated the value of having the dataset locally as the processing time was significantly less. It also highlighted the point made by Dr Birney that not having data locally can potentially compromise the ability of the Australian researcher to use the global resource due to latency or network limitations.



Open Forum Outcomes

The open forum sessions were designed as an opportunity for local users to provide feedback about their potential use of the EMBL-EBI Mirror in order to guide the next stage of implementation.

To set the scene the results of a pre-meeting survey were presented and are included in Appendix C.

The number of respondents from all invitees totaled nine (7 Qld, 1 Victoria, 1 ACT), so care must be exercised in drawing conclusions as this was not a representative sample of the audience.

In summary the results show:

- 50% are bioinformaticians (a poll of the audience also suggested this)
- Most respondents use both EBI and NCBI
- Those that use EBI do so primarily for data access and searching, with particular reference to ENSEMBL, BLAST, Interpro, the Short Read Archive, Uniprot, and Arrayexpress.
- Feedback prior to the forum suggested that fast access to data is important especially when it can be integrated with locally available data. Inclusion of UCSC style visualization using 'tracks' was considered desirable. One respondent mentioned the opportunity for the Mirror to assist in building the community.

With this backdrop, discussion opened with questions about the financial arrangements, support and opportunities to link with the Mirror team. As the discussion focused on the requirements of the researchers the following categories were evident:

Data

- Replicate the Short Read Archive data although recognition of the challenge to download and maintain this was identified
- ENCODE
- 1,000 genomes data was a significant request. Potential for a pilot program to establish proof of concept focused on:
 - Variant Call Format files were discussed – these are already being provided
 - Re-calibrated BAM files
- Data which is out of date or is a subset was considered acceptable although there was no resolution of how far out of date and which part of the set was acceptable. Three months was mooted.

Services

- ENSEMBL tracks
- Integration of EBI & UCSC data
- UCSC track information

Other points of discussion

- Broad information / communication of the activity of the Australian Mirror is required
- There was interest in the funding of the project. It was noted that there were numerous Federal sources, Queensland State Government, CSIRO and University of Queensland.
- The governance of the project is through a committee which reports to EMBL-Australia
- Australian research-interest expertise groups could develop around the Mirror
- The ability to integrate local results with the Australian mirror was a point for consideration. It was noted that it was intended that this would be available although the exact method for inclusion needed to be worked out.
- Integration of the Australian component of the EBI data with Research Data Commons using the Atlas of Living Australia to define the relevant species is already being developed by the Queensland Facility for Advanced Bioinformatics (QFAB) with funding by the Australian National Data Service.
- If someone develops a tool that can run on the EBI Mirror then they should get in touch with Professor Nadia Rosenthal or Professor Mark Ragan
- Consideration needs to be given to whether there are any medical/ethical limitations preventing data being made public (exported)?
- Set up a 'Chat room' or equivalent for community feedback

Melbourne

Outback Meeting Room
 Monash University
 Level 1, Building 75
 Wellington Road, Clayton.

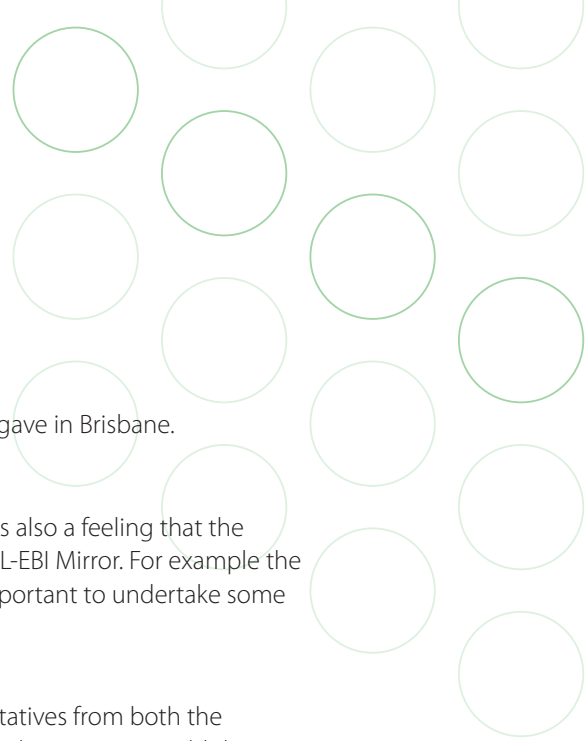
Thursday 23 June
 10:30am – 12:35pm

List of Attendees

Professor Nadia Rosenthal	<i>Scientific Head, EMBL Australia Director, Australian Regenerative Medicine Institute (ARMI) Monash University</i>
Professor Peter Currie	<i>Deputy Director, Australian Regenerative Medicine Institute (ARMI)</i>
Mr Silvio Tiziani	<i>Executive Director, EMBL Australia COO, The Australian Regenerative Medicine Institute (ARMI)</i>
Professor Paul Bonnington	<i>E-Research, Monash University</i>
Mr Steve Quenette	<i>School of Mathematical Sciences, Monash University</i>
Mr David Salgado	<i>Australian Regenerative Medicine Institute (ARMI), Monash University</i>
Dr Ross Wilkinson	<i>Australian National Data Service (ANDS), Monash University</i>
Dr Phil Davies	<i>Department of Business and Innovation</i>
Dr Michael Lynch	<i>Department of Business and Innovation</i>
Dr David Green	<i>Implementation Team for EMBL Australia Mirror of EMBL-EBI Facility, University of Queensland</i>
Dr Andrew Lonie	<i>Head of Life Sciences Computation, The University of Melbourne</i>
Dr Torsten Seemann	<i>Victoria Bioinformatics Consortium, Monash University</i>
Professor Terry Speed	<i>Division Head Bioinformatics, WEHI</i>
Professor Geoff Webb	<i>Infotech, Monash University</i>
Mr Justin Zobel	<i>Compute Science and Software Engineering, The University of Melbourne</i>
Mr Jeff Christiansen	<i>Australian National Data Service (ANDS), Monash University</i>
Professor Paul Hertzog	<i>MIMR Centre for Innate Immunity and Infections, Monash University</i>
Mr Arun Konagurthu	<i>Infotech</i>
Andreas Scherer	<i>Australian Genome Research Facility, WEHI</i>
Assoc Prof Tianhai Tian	<i>ARC Future Fellow, School of Mathematical Sciences, Monash University</i>
Dr Sarah Boyd	<i>Research Fellow, School of Mathematical Sciences, Monash University</i>
Dr Jonathan Keith	<i>Senior Lecturer, School of Mathematical Sciences, Monash University</i>

Agenda Items

Thursday, 23 June 2011		
10:30am	Welcome	Professor Nadia Rosenthal
	Overview of EMBL and EMBL Australia (15 mins)	
10:45am	Overcoming data volume problems in genomic biology (45 mins)	Dr Ewan Birney (Seminar)
11:30am	Short demonstration of the EMBL Australia Mirror of EMBL-EBI Facility (25 mins)	Dr David Green
11:55am	Round Table discussion (40 mins) ~ working lunch ~	Facilitated by Professor Paul Bonnington
12:35pm	Session concludes	



Key Points

Professor Rosenthal, Dr Birney and Dr Green repeated the presentations that they gave in Brisbane.

Open Forum Outcomes

There was a general sense that this type of forum is needed and desired. There was also a feeling that the group was too small to pick up the vast array of needs around Victoria for the EMBL-EBI Mirror. For example the Parkville and Bundoora precincts had limited representation. It was considered important to undertake some follow up assessment to capture the needs more completely.

The following represents a summary of the major points of discussion:

- 1000 Genomes dataset would be a useful resource to have on hand. Representatives from both the Parkville and Clayton Precincts agreed. It was indicated that having all 500TB in the country would drive change around how to deal with that data.
- Representative/sample data from all collections at EBI were needed for education purposes with newer data more important than older.
- Derived data should be acquired first (like VCF).
- Uniprot – although it was felt that this would be useful, when asked what components were needed locally because of latency and bandwidth issues, the participants could not advise which parts.
- ENSEMBL – Bacteria needed
- TS Microbial data can be accessed remotely but it was noted that it was already being downloaded locally at Monash.
- UCSC style track information required
- Additional communication channels perhaps initially through a targeted redistribution of the survey to a broader group.

Appendix A: EMBL – EBI Mirror Functionality

As at 21 June 2011, 15 services were accessible on the local copy of the EMBL-EBI mirror website and 148 dataset groups available via the mirror FTP service. Other services and data are available by directly linking to the UK website. A list of the locally available resources is provided below

Services

- ENSEMBL
- InterProScan

Single Sequence Search

- FASTA
- NCBI BLAST
- PSI-BLAST
- PSI-Search
- Ssearch
- WU-Blast

Multiple Sequence Alignment

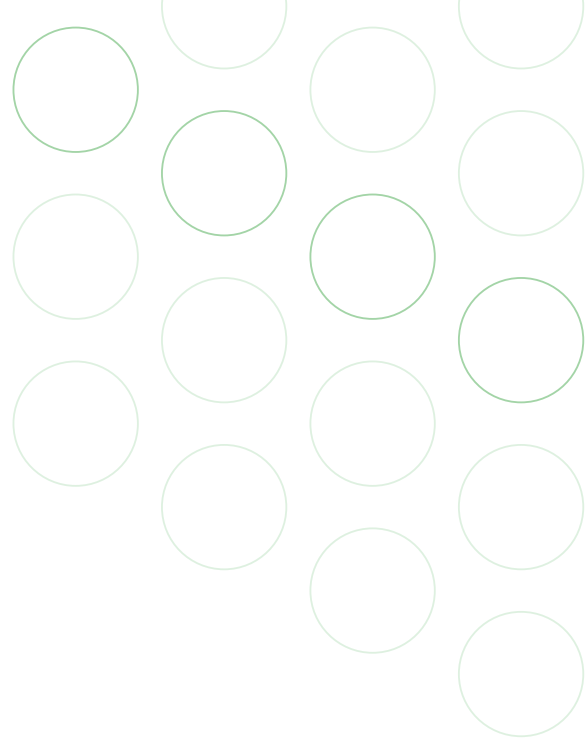
- ClustalW2
- MUSCLE
- T-Coffee
- Kalign
- MAFFT

Databases

Top Level FTP Database Groups

- 16S_RNA
- 3d_ali
- 3Dee
- ace
- alu
- androgenr
- arrayexpress
- ASD
- astd
- backup.info
- berlin
- bii
- bio catal
- biomart
- blocks
- blocks.old
- camoddssp
- camodhssp
- camodpdb
- cd40|base
- chebi
- chembl
- clustr

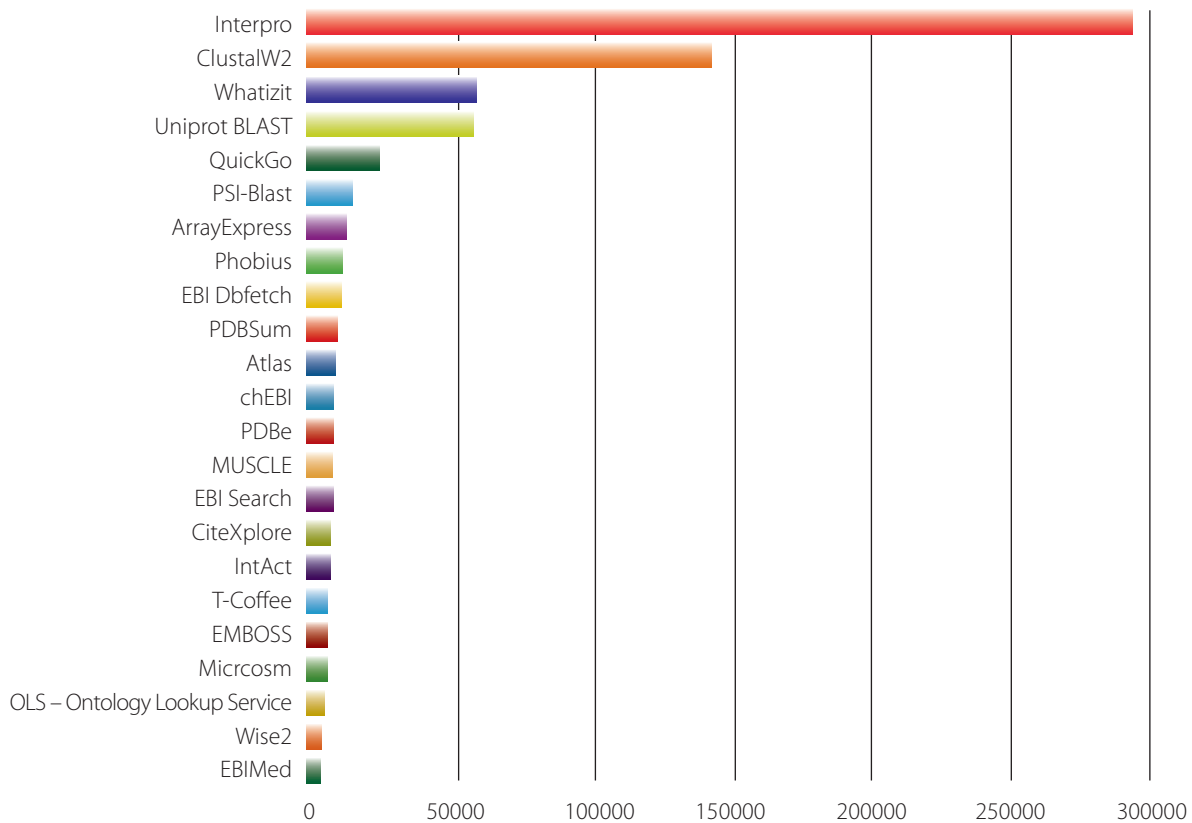
- codonusage
- cpgisle
- cutg
- dali
- databanks
- dbcat
- dbEST
- dbGSS
- dbSTS
- dgva
- domo
- dssp
- ecdc
- edgp
- EGI
- embl
- embnet
- emdb
- emp
- emvec
- ena
- ensembl
- enzyme
- epd
- fastafiles
- FBA2KO
- Fregene_datasets
- geneticcode
- genome_reviews
- genomes
- GO
- gpcrdbsup
- haema
- haemb
- hamsters
- HGNC
- hla
- hovergen
- hssp
- imgt
- info
- intact
- integr8
- intenz
- interpro
- ipd
- IPI
- journals_toc
- kabat
- limb
- lista



- Irgex
- MassSpecDB
- mdm2
- methyl
- microarray
- microcosm
- models
- msd
- ncbi
- nmrshiftdb
- nrdb90
- nrl_3d
- nrsub
- nucleosomal_dna
- p53
- p53APC
- parasites
- patentdata
- pdb_finder
- pdb_select
- pdb_seq
- pir
- pir2sptr
- piraln
- pkcdd
- plmitrna
- pmc
- pride
- primers
- prints
- prodom
- prof_pat
- prosite
- pubchem
- ras
- rcsb
- rdp
- rebase
- reference_proteomes
- reflist
- relibrary
- repbase
- RESID
- RHdb
- rhea
- rldb
- rrna
- sbase
- Science
- seqanalref
- smallrna
- SPproteomes
- srp
- stackdb
- stride
- SubtiList
- supplementary
- swissprot
- taxonomy
- testsets
- tfd
- tmp
- transfac
- transterm
- trembl
- trna
- Unigene
- uniprot
- UTR
- variantdbs
- xray
- yeast

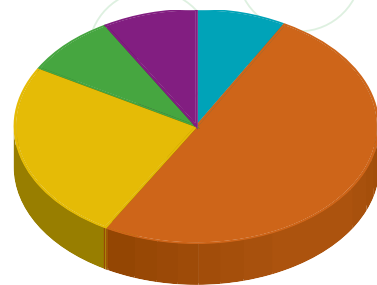
Appendix B: Frequency of access from Australian locations to EBI (UK)

AU Usage of EBI Services



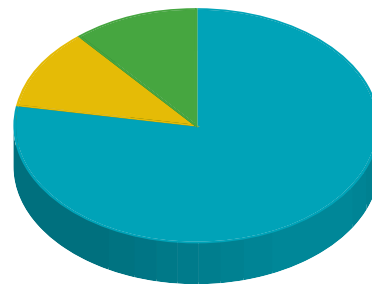
Appendix C: Pre-Workshop Survey: Combined results from all locations

Are you:		
	Response Count	Response %
Management	1	9.09%
Bioinformatician	6	54.55%
Biologist	3	27.27%
Computational Scientist	1	9.09%
Other (Software Engineer)	1	9.09%
<hr/>		
<i>Number of Responses</i>	11	100%
<i>Number of Unique Responses</i>	1	



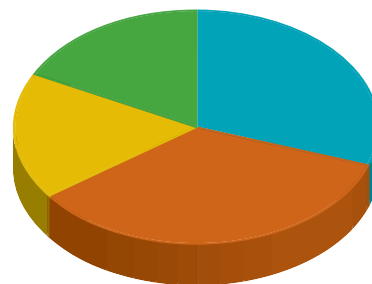
■ Management 1 **9.09%**
 ■ Bioinformatician 6 **54.55%**
 ■ Biologist 3 **27.27%**
 ■ Computational Scientist 1 **9.09%**
 ■ Other (Software Engineer) 1 **9.09%**

Which state or territory are you from?		
	Response Count	Response %
QLD	7	77.78%
NSW	0	0.00%
ACT	1	11.11%
VIC	1	11.11%
TAS	0	0.00%
SA	0	0.00%
WA	0	0.00%
NT	0	0.00%
<hr/>		
<i>Number of Responses</i>	9	1



■ QLD **77.78%**
 ■ NSW **0.00%**
 ■ ACT **11.11%**
 ■ VIC **11.11%**
 ■ TAS **0.00%**
 ■ SA **0.00%**
 ■ WA **0.00%**
 ■ NT **0.00%**

Please indicate which of the following you use:		
	Response Count	Response %
EBI	7	77.78%
NCBI	8	88.89%
ENSEMBL	4	44.44%
UCSC	4	44.44%
<hr/>		
<i>Number of Responses</i>	9	



■ EBI **77.78%**
 ■ NCBI **88.89%**
 ■ ENSEMBL **44.44%**
 ■ UCSC **44.44%**

If you use EBI, what do you primarily use it for?

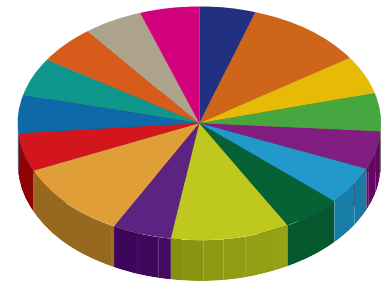
SRA data Download
Swissprot, InterPro
Array Express
data search
Retrieval of sequences and protein-centric analyses
Data set acquisition
Actually, I don't use any of these resources directly, but I manage and work with others who do.

Do you use internet accessible databases? If so, what are your top three?

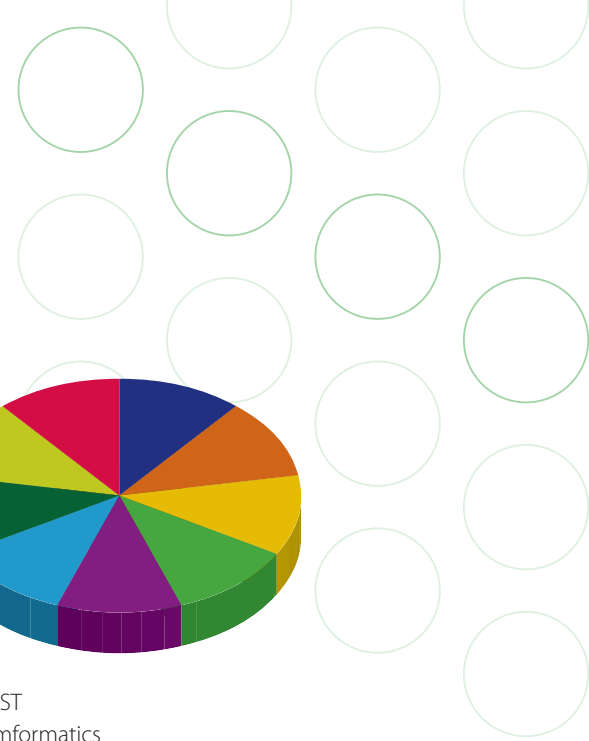
	Response Count	Response %
NT	1	14.29%
NCBI	2	28.57%
Array Express	1	14.29%
UCSC	1	14.29%
EBI	1	14.29%
Uniprot	1	14.29%
<hr/>		
<i>Number of Responses</i>	7	100.00%

SRA	1	16.67%
EBI	2	33.33%
Ensembl	1	16.67%
Geo	2	33.33%
<hr/>		
<i>Number of Responses</i>	6	100.00%

Uniref	1	16.67%
KEGG	1	16.67%
Entrez	1	16.67%
DFCI – Gene Indicies	1	16.67%
AMD	1	16.67%
Ensembl	1	16.67%
<hr/>		
<i>Number of Responses</i>	6	100.00%



- NT
- NCBI
- Array Express
- UCSC
- EBI
- Uniprot
- SRA
- EBI
- Ensembl
- Geo
- Uniref
- KEGG
- Entrez
- DFCI – Gene Indicies
- AMD
- Ensembl



**Do you use internet accessible analysis tools?
If so, what are your top three?**

	Response Count	Response %
BLAST	1	20.00%
Stemformatics	1	20.00%
DAVID	1	20.00%
BioConductor	1	20.00%
Interpro	1	20.00%

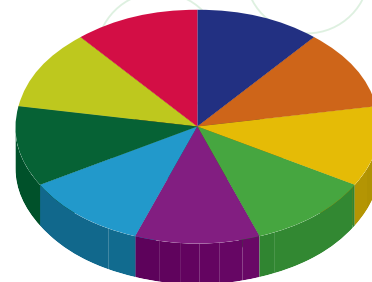
<i>Number of Responses</i>	5	100.00%
----------------------------	---	---------

Galaxy	1	33.33%
BASE	1	33.33%
GSEA	1	33.33%

<i>Number of Responses</i>	3	100.00%
----------------------------	---	---------

Mev	1	100.00%
-----	---	---------

<i>Number of Responses</i>	1	100.00%
----------------------------	---	---------



- BLAST
- Stemformatics
- DAVID
- BioConductor
- Interpro
- Galaxy
- BASE
- GSEA
- Mev

Future growth of the EMBL-Australia EBI mirror will depend on collaborations between Australian research groups and this mirror. How might your research benefit from tightly integrated access to the hardware, software, data and support services of the EBI

Please provide any other comments for discussion at the workshop

- 1 Integration of our in-house datasets with the EBI infrastructure would be made much more possible through the EBI mirror. While I am not currently in a position to determine such a step or collaboration, in coming years as I work to establish my own research group this is certainly something that I will consider.
- 2 Faster access.
- 3 Ability for Stemformatics (hosted by QFAB) to hook into EBI analysis tools / data over fast local network for high performance and service timeliness.
- 4 One assumes that faster return of results and possibly an increased range of tools should be available. Until I have actually used it, it will be hard to tell.
- 5 Faster download etc.
- 6 There is huge need to be able to see next gen sequence data as processed tracks on genome browsers. There is also need to be able to integrate and visualise genomic sequence information from multiple individuals.
- 7 Ease of access to integrated collections of data.
- 8 We are interested in data analysis in a wide range of different formats and different platforms (microarray, sequencing, next gen sequencing datasets) for pathway, GO clustering and network analysis. I am working in Monash Medical Research Centre and involved in tools development for analysis innate immunity and building the IRG network.
- 9 I'm keen to explore the "community building" potential of the mirror.

Participants



Secretariat



Printed on Mega Recycled, an environmentally considered sheet consisting of 50% post consumer recycled waste and 50% FSC certified fibre.

