

Chinook: A collaborative system for bioinformatics analysis.

VanBUG October 2004

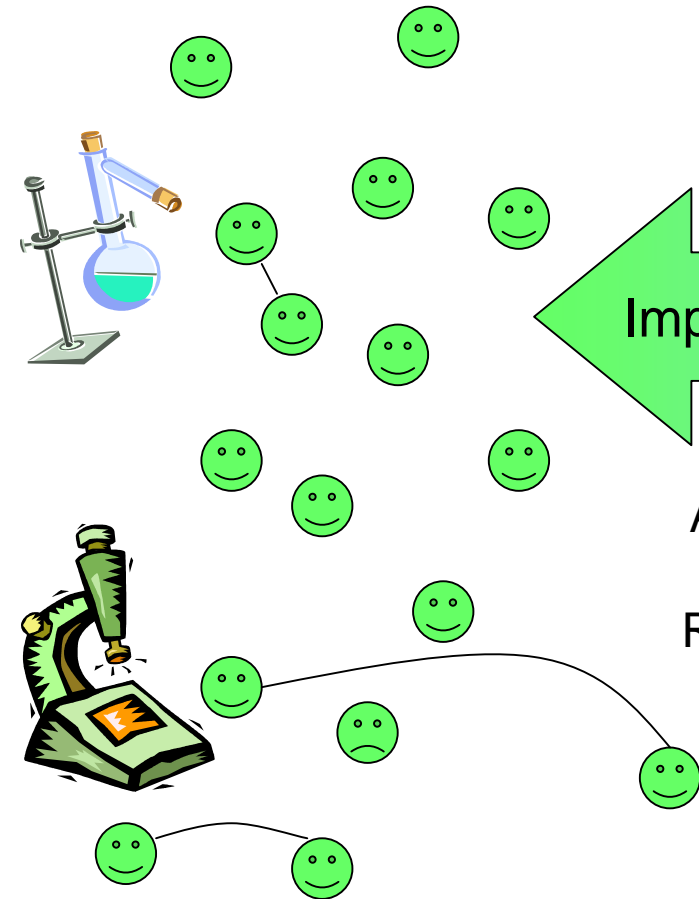
Stephen Montgomery

Things get more complicated...

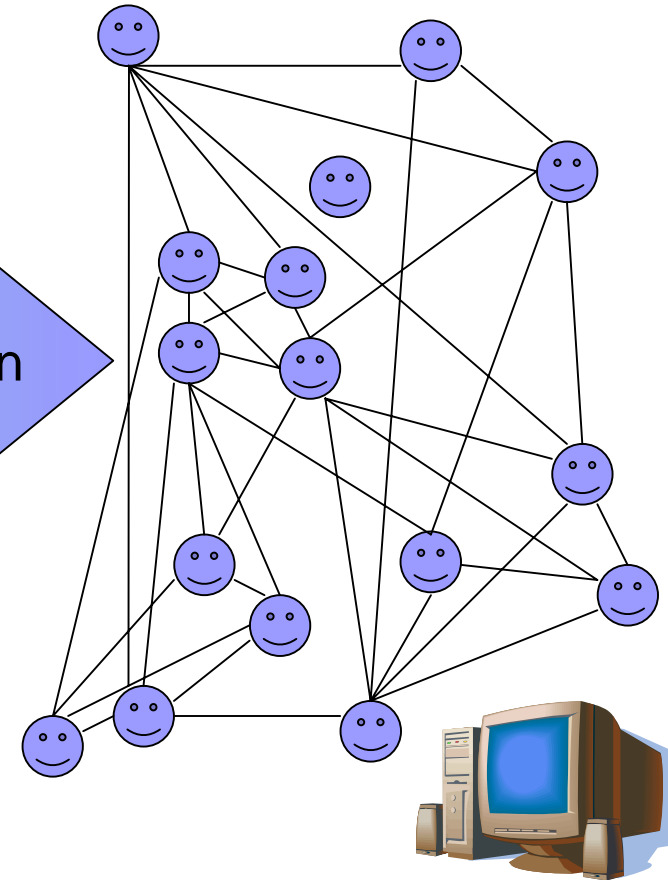
- Each individual has:
 - Access to different resources
 - Computational / Monetary / Personnel
 - Finite time available
 - A different social network
 - Professional obligations
- Each group
 - Organizational boundaries
 - Toolkit (suites and scripts)
 - Method of providing tools (OS, Internet, Interfaces)

An improved interaction map of **biologists** and **bioinformaticians**

Biologists



Bioinformaticians



Access to communities
Access to resources
Retain sub-organization

A community-based approach to bioinformatics analysis

- Use the principles of peer-to-peer technology
 - Allow biologists to easily discover and run bioinformatics tools
 - Create a dynamic, reliable network for analysis
 - Reduce overlapping integration efforts
 - Improve communication within/outside organizations
 - Address problems relevant to bioinformatics
 - Attribution
 - Data management
 - Resource distribution



Chinook Client

File Tools Help

Types Filters

Service Types

- Teiresias
- ALIGNMENT
 - LAGAN
 - MLAGAN
 - CLUSTALW
 - Mauve
 - Promoterwise
 - Sim4
 - Orca
- TEST
 - MULTIOUT
 - MOTIF DISCOVERY

Discovered Services

Service	Service Type	Service Lo...	Version	Bandwidth	Jobs In Qu...	Status	Creator	Publisher
SNPSeque...	VARIATION	///localhos...	1	0	0	Connected		UNKNOWN RMI
Teiresias	PATTERN ...	///localhos...	1	0	0	Connected	http://ww...	UNKNOWN RMI
LAGAN	ALIGNMENT	///localhos...	1	0	0	Connected	http://lag...	UNKNOWN RMI
MLAGAN	ALIGNMENT	///localhos...	1	0	0	Connected	http://lag...	UNKNOWN RMI
CLUSTALW	ALIGNMENT	///localhos...	1	0	0	Connected	http://ww...	UNKNOWN RMI
Mauve	ALIGNMENT	///localhos...	02-19-2...	0	0	Connected	http://sm...	UNKNOWN RMI
Promoter...	ALIGNMENT	///localhos...	2.2.3-rc7	0	0	Connected		UNKNOWN RMI
Sim4	ALIGNMENT	///localhos...	2003-09...	0	0	Connected		UNKNOWN RMI
Orca	ALIGNMENT	///localhos...	2003-09...	0	0	Connected		UNKNOWN RMI
MULTIOUT	TEST	///localhos...	1	0	0	Connected		UNKNOWN RMI
Lawrence ...	MOTIF DIS...	///localhos...	1	0	0	Connected		UNKNOWN RMI
EPONINE	TRANSCRI...	///localhos...	1	0	0	Connected		UNKNOWN RMI

Run service Visit homepage Refresh Services

Job Status

Job	Time Sub...	Processing...	Job ID
LAGAN	Tue Sep 1...	12	1741957...

View Results Discard Job

Lightweight Web Browser

Site: http://lagan.stanford.edu/lagan_web/index.shtml

HOME ABOUT PAIRWISE LAGAN MULTI-LAGAN SHUFFLE-LAGAN MANUAL USING AND DOWNLOAD AUTHORS

NEWS

25/05/2004: MLAGAN was used for [multiple alignments of human, mouse, rat and macaque genomes](#) based on the PASH anchoring scheme by the [BRL group at the Baylor College of Medicine](#).

25/02/2004: You can now generate LAGAN alignments through [Sockeye](#), a 3D comparative genome browser that provides visualization for Lagan alignments in the context of user-defined or [Ensembl](#)-imported annotation. [Sockeye](#) is developed at the [Genome Sciences Centre, Vancouver](#).

11/10/2003: The latest version of LAGAN (1.21) is now available under the GNU Public License, free to both academic and commercial organizations. To see the license terms and download the program see the [Citing and Availability](#) page.

9/7/2003: You can now align sequences with [Shuffle-LAGAN](#), our *local* aligner on the web.

4/22/2003: Multi-LAGAN alignments can now be visualized locally with the [Phylo-Vista](#) tool. See details [here](#).

STARTED...

discover and run jobs here or through **bioperl**

Configure service

Data entry

DATA NAME: dna_sequence Edit data

Parameter entry

Name	Value	Type	Description
translate	<input type="checkbox"/>	BOOLEAN	Use transl...
rc	<input type="checkbox"/>	BOOLEAN	Reverse c...
fastreject	<input type="checkbox"/>	BOOLEAN	Abandon ...
Name	Value	Type	Description
chaos		STRING	The conte...
order		STRING	The conte...
recurfl		STRING	Used in re...

OK Cancel



Enter data

Name: A End: 2000
 Start: 1000 Species: Rattus_norvegicus
 Database: rattus_norvegicus_core_22_3b
 Chromosome: 1
 Strand: 1 DatabaseType: ENSEMBL

EMPTY

Add Edit Remove

OK Cancel

CURRENT: 2 MIN: 2 MAX: 2



Enter data

Select retrieval mode
 DNA_LOCATION

Name: A
 Chromosome: 2
 Start: 1000000
 End: 1001000
 Strand: 1
 DatabaseType: ENSEMBL
 Species: Rattus_norvegicus
 Database: rattus_norvegicus_core_22_3b

DATA VALID

OK Cancel

Result information

Job Status

JOB COMPLETED STATUS: success
 Download result files from table

Preview results

Stdout Stderr

Result files

File ID	Description	Downloaded	Local File...
6322.178...	Standard ...	false	
374.1798...	Standard ...	false	

Download

OK Cancel

Preview

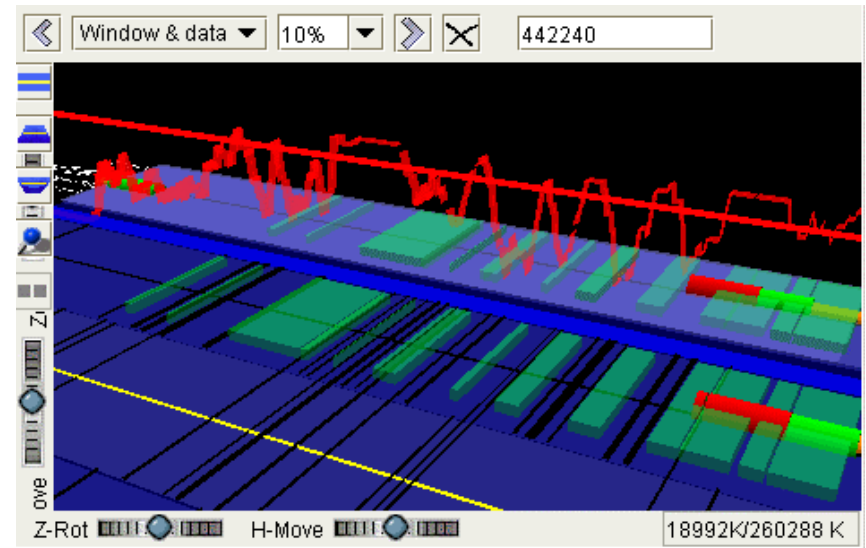
```
Using /tmp/chinook.tmp.-1758111813 /tmp/chinook.tmp.-314156771
(12, 0, 25, 0) x
Using /tmp/chinook.tmp.-1758111813 /tmp/chinook.tmp.-314156771
(13, 1, 30, 0) x
Using /tmp/chinook.tmp.-1758111813 /tmp/chinook.tmp.-314156771
(8, 1, 30, 0) x
Using /tmp/chinook.tmp.-1758111813 /tmp/chinook.tmp.-314156771
(7, 1, 30, 0) x
0 cumulative anchors
read 0 anchs
Total size = 3 * 10^6
```

Algorithms integrated into Chinook

ClustalW	Genscan
Conreal	Sim4
DIALIGN	MSCAN
LAGAN	ANN-Spec
Mauve	Recursive Gibbs Motif Sampler
ORCA	MEME
Shuffle-LAGAN	Motifsampler
T-Coffee	RSAT oligo analysis
Promoterwise	STUBB
Primer3	Teiresias
Eponine	wConsensus
ELPH	

Projects involving Chinook

- **OrthoSEQ** plans to provide analysis through the Chinook/BioPerl Perl API.
- **Sockeye** uses Chinook to deliver state-of-the-art alignment, PCR prediction, and regulatory analysis
- **Pegasys** plans to provide pipeline management to subset of services advertised by Chinook.
- **Bio-Linux** planning to integrate a subset of their algorithms.





Use cases of Chinook

- Grid/Cluster computing.
- Internally connect teams / individuals.
- Collaborate with remote individuals.
- Provide an API layer to your algorithms.
- Insert bioinformatics analysis into applications.
- Show off your tools.

Acknowledgements

GENOME SCIENCES CENTRE

- Steve Jones
- Tony Fu
- Jun Guan
- Keven Lin

- Asim Siddiqui
- Genereg team @ GSC

- Mark Mayo
- Bernard Li

CENTRE FOR MOLECULAR MEDICINE AND THERAPEUTICS

- Wyeth Wasserman
- Jonathan Lim

UBC BIOINFORMATICS CENTRE

- Francis Ouellette
- Graham McVicker
- Sohrab Shah

Funding: MSFHR, Genome Canada

VISIT <http://smweb.bcgsc.bc.ca> OR <http://www.bcgsc.bc.ca/chinook>