

Bioinformatics Open Source Conference 2013 @ Berlin, Germany (July 19)

BioRuby updates

Power of modularity in the community-based
open source development model

Toshiaki Katayama <ktym@dbcls.jp>

<http://jp.linkedin.com/in/toshiakikatayama>

*Database Center for Life Science (DBCLS),
Research Organization of Information and Systems (ROIS), Japan*



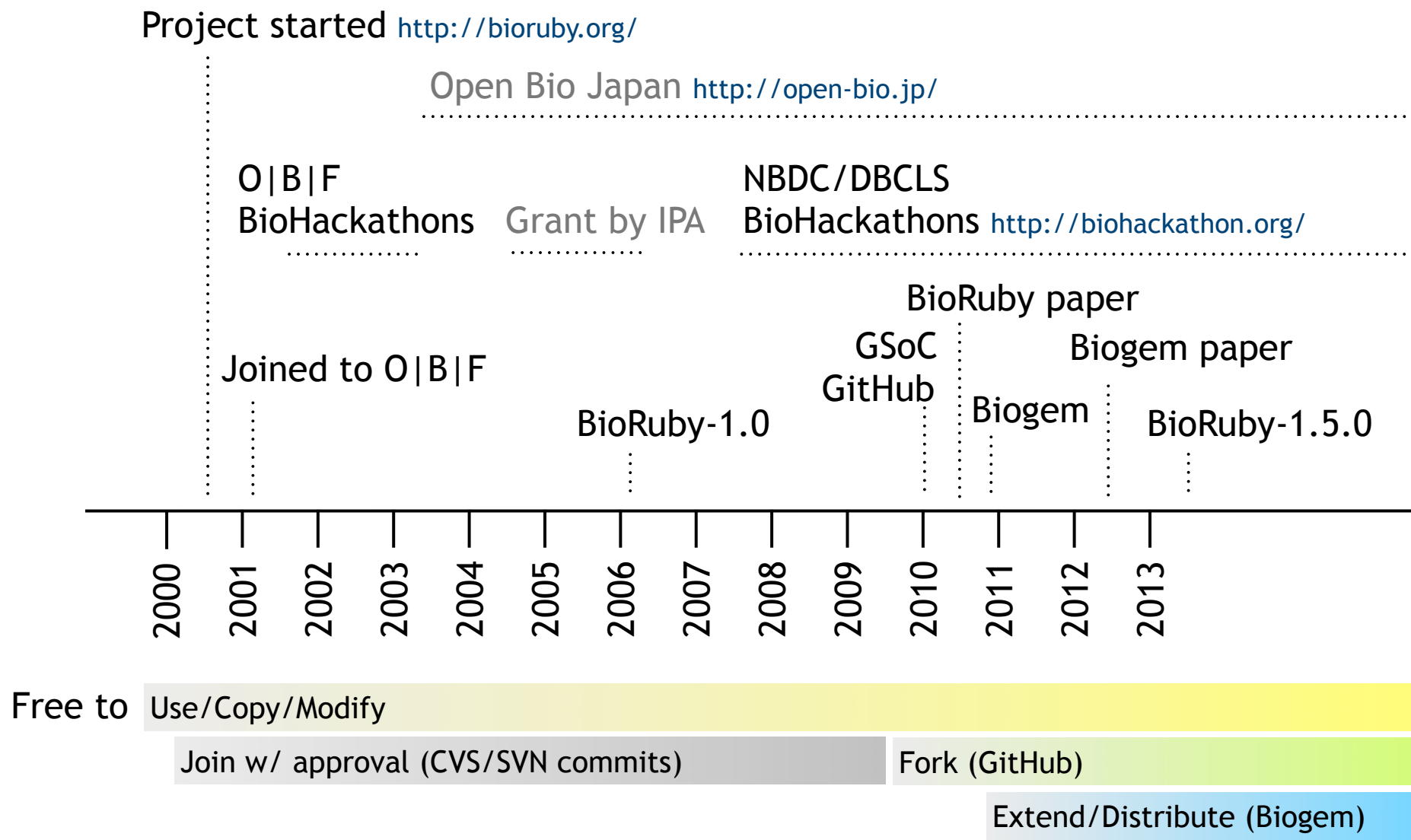
Pjotr Prins *University Medical Center Utrecht, Netherlands*

Raoul Bonnal *Istituto Nazionale Genetica Molecolare, Italy*

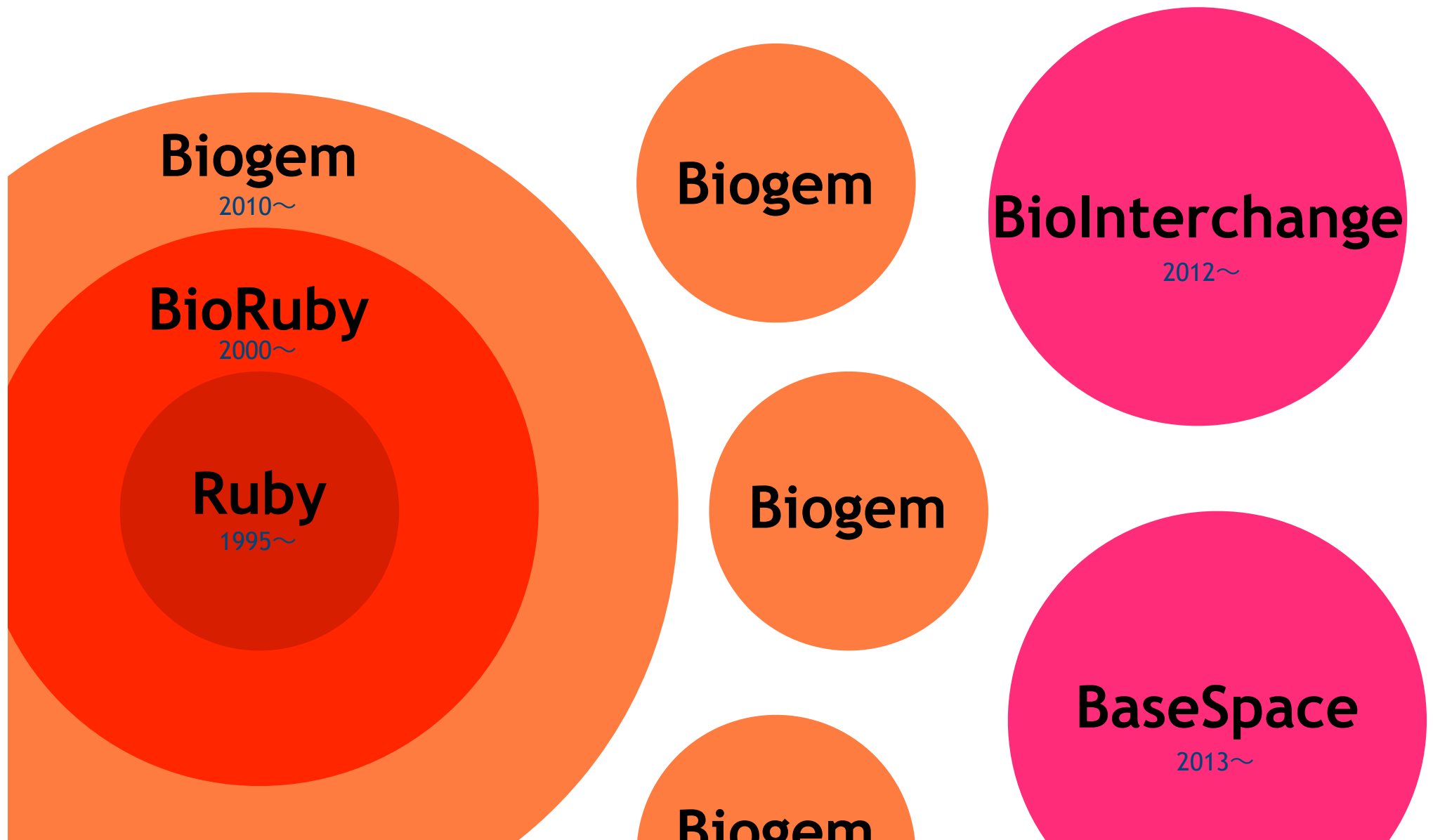
Francesco Strozzi *Parco Tecnologico Padano, Italy*

Naohisa Goto *Osaka University, Japan*

BioRuby stairway to freedom



Biogem ecosystem and beyond



Biogem lowered the entry barrier

- User can freely develop their own libs/apps and distribute them.
- BioRuby core can concentrate on its stability and compatibility.

Biogem developer:

```
% gem install bio-gem
```

```
% biogem yourapp
```

```
% cd bioruby-yourapp
```

```
# develop lib/*.rb and/or bin/* as you like
```

```
% bundle exec release
```

```
# will make yourapp available on GitHub.com and Rubygems.org
```

Biogem user:

```
% gem install bio-yourapp
```

That's it!



http://biogems.info

Biogems.info

| Biogems | Rubygems | BioLinux | People |

The future is bright for Ruby in bioinformatics! Create a Biogem! | BioRuby!

New: CloudBioLinux packages are now listed on biogems.info!

#	biogem	description	by	cite	version	released	stars	issues	source	build	total	90d*)	7d	90d**)
											downloads	commits		
1	bio	Bioinformatics library (...)	BioRuby project		1.4.3.0001	7 weeks				build passing	43678	5436	2	28
2	biodiversity	Parser of scientific names (...)	Dmitry Mozherin		3.1.2	3 weeks	9	5		build passing	22428	2733	0	23
3	bio gem	Biogem is a software generator for Ruby in (...)	Raoul J.P. Bonnal, Pjotr P.											
4	bio samtools	Binder of samtools for Ruby, on the top (...)	Ricardo Ramirez-Gonzalez, Raoul J.P. Bonnal											
5	entrez	Http requests to entrez e-utilities (...)	Jared Ning											
6	bio ucsc api	The Ruby ucsc api accessing the ucsc genome (...)	Hiroyuki Mishima, Jan Aert											
7	intermine	Webservice client library for intermine data-warehouses (...)	Alex Kalderimis											
8	bio gff3	Gff3 parser for big data (...)	Pjotr Prins											
9	sequenceserver	Blast search made easy! (...)	Anurag Priyam, Ben J Wood, Yannick Wurm											
10	bio logger	Log4r wrapper with extra features for roles and (...)	Pjotr Prins											
11	bio maf	Maf parser for BioRuby (...)	Clayton Wheeler		1.0.1	11 months				build passing	7292	1299	0	0
12	bio gadget	Gadgets for bioinformatics (...)	Shintaro Katayama		0.4.8	5 weeks					6313	2710		
13	bio grid	A biogem to submit jobs on a queue (...)	Francesco Strozzi		0.3.3	8 months	5			build unknown	5826	1204	0	0
14	bio blastxmlparser	Very fast blast xml parser and library for (...)	Pjotr Prins		1.1.1	5 months				build passing	5643	674	0	1
15	bio faster	A fast parser for fastq files (...)	Francesco Strozzi		0.4.5	13 months				build passing	5301	726	0	0
16	bio phyta	Pipeline to remove contaminations from est libraries (...)	Philipp Comans		0.9.4	19 months					5164	510	0	0
17	bio alignment	Support for multiple sequence alignments (msa) (...)	Pjotr Prins		0.0.7	12 months				build passing	5001	639	0	1
18	bio assembly	BioRuby assembly plugin (...)	Chase Miller, Francesco Strozzi		0.1.0	2 years	4				4651	429	0	0
19	scaffolder	Genome scaffolding for human beings (...)	Michael Barton		0.4.4	23 months				build passing	4601	432		

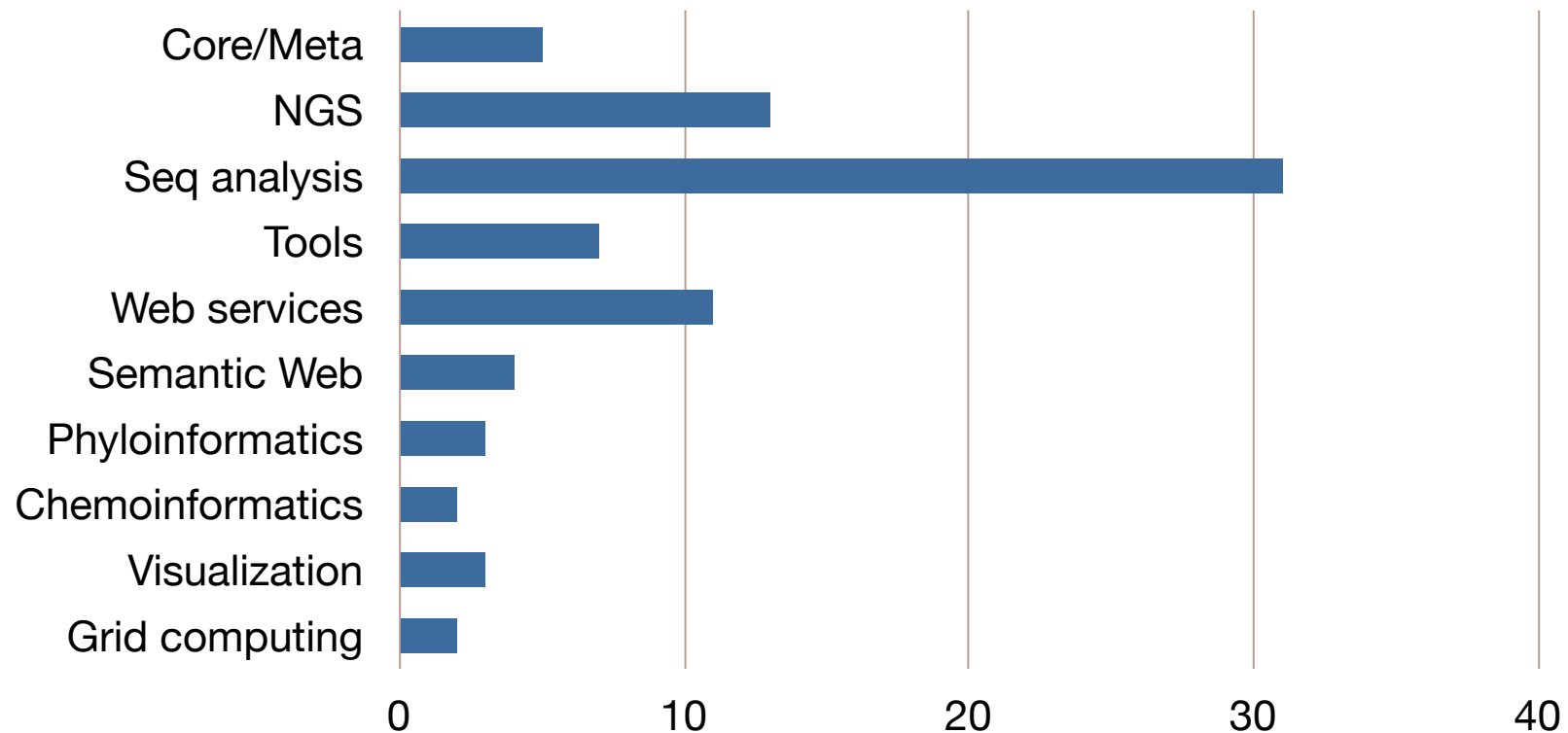
>80 Biogems have been released so far

BioRuby >44K DLs

Biogems vary from ~100 to 100K DLs

Biogem packages break-down

>20 new biogems have been developed since BOSC2012



Biogem packages

Web services

bio-ucsc-api

intermine

entrez, eutils

ruby-ensembl-api

bio-dbsnp

bio-chembl

:

Phyloinformatics

biodiversity

name-spotter

bio-nexml

bio-phyloxml

:

Seq analysis

bio-gff3

bio-hmmer3

bio-maf

bio-phyta

bio-alignment

bio-signalp

bio-tm-hmm

bio-isoelectric-point

bio-genomic-interval

bio-kmer-counter

bio-restriction-enzyme

:

NGS

bio-ngs

bio-samtools

bio-gadget

bio-faster

bio-sambamba

bio-tabix

bio-bgzf

bio-gngm

:

Others

bio-svgenes

biointerchange

:

bio-ucsc-api by Mishima H.

require 'bio-ucsc'

Automatically maps UCSC MySQL schema to Ruby class by ActiveRecord (Rails)

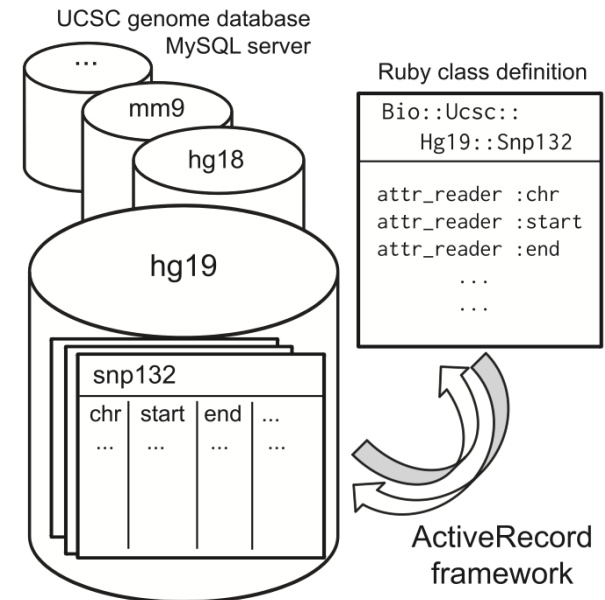
Bio::Ucsc::DB::Table

(e.g., Bio::Ucsc::Hg19::Snp132)



Now integrated in TogoWS
(you don't need to code!)

UCSC Genome Bioinformatics



[http://togows.org/api/ucsc/database/chromosomal-position\[.format\]](http://togows.org/api/ucsc/database/chromosomal-position[.format])

→ <http://togows.org/api/ucsc/hg19/chr1:107,599,267-107,601,915.fasta>

[http://togows.org/api/ucsc/database/table/\[column=query\]\[.format\]/offset,limit](http://togows.org/api/ucsc/database/table/[column=query][.format]/offset,limit)

→ <http://togows.org/api/ucsc/hg19/refGene/name2=UVSSA.json>

→ <http://togows.org/api/ucsc/hg19/snp137/chrom=chr22;refUCSC=A/1,10>

bio-gadget by Katayama S. (not me:)

NGS analysis package

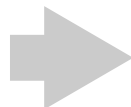
to handle RNA-Seq data with **UMI+barcode+adaptor** reads

```
% gem install bio-gadget  
% bio-gadget <task>
```

Available tasks

- dedup :: Deduplicate fastq (via STDIN)
- demlt :: Demultiplex fastq by barcodes
- fqxz :: automatic (re)compression of *.fq(.gz|.bz2) files
- qvstat :: Statistics of quality values in *.qual file
- rgt2mtx :: Convert cuffdiff read group tracking file into tab-separated matrix
- wig5p :: Convert bam-format alignments into wig-format table
- wigchr :: Extract wiggle track on specified chromosome

mixed-reads



demultiplexed



<http://www.nature.com/nprot/journal/v7/n5/full/nprot.2012.022.html>

Islam S, Kjällquist U, Moliner A, Zajac P, Fan J-B, Lönnerberg P, et al. Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. Nat Protoc. 2012 May;7(5):813-828.

DOI: 10.1038/nprot.2012.022 PMID: 22481528

<http://www.nature.com/nmeth/journal/v9/n1/full/nmeth.1778.html>

Kivioja T, Vähärautio A, Karlsson K, Bonke M, Enge M, Linnarsson S, et al. Counting absolute numbers of molecules using unique molecular identifiers. Nat Methods. 2011 Nov 20;9(1):72-74

DOI: 10.1038/nmeth.1778 PMID: 22101854

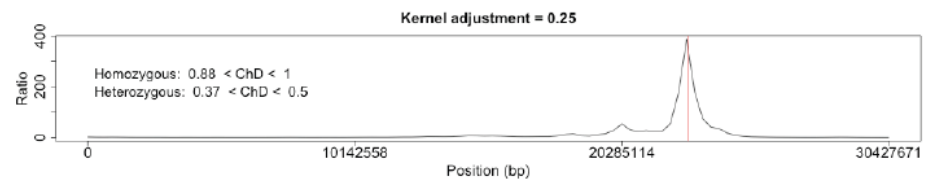
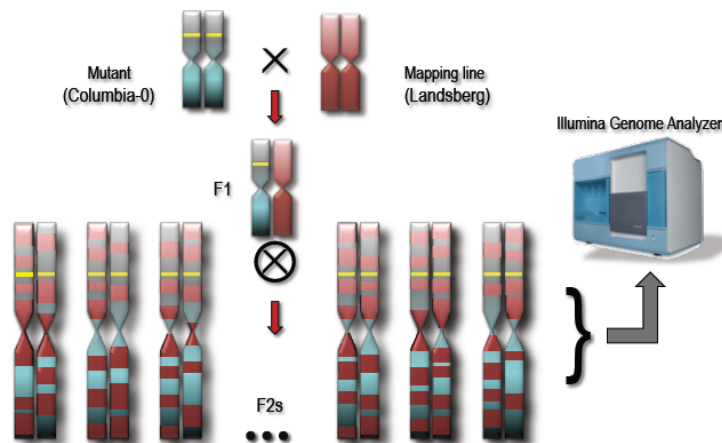
bio-ngm by MacLean D. *et al.*

Another NGS analysis package
to detect causative SNPs affecting WT/mutant phenotypes

GNGM = Generalised NGM (Next-generation EMS mutation mapping)



1. Mapping to reference genome
2. Calculating and grouping allele frequencies
3. Find candidate positions of causative SNPs





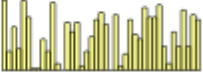




Mutant	Chr	Position	RefN	SNP	RefP	Sub	Depth	Ch _D	Strand	Accession
1	3	20795011	A	T	D	E	16	1.00	-	AT3G56040.1
1	3	20795012	T	C	D	G	19	1.00	-	AT3G56040.1
1	3	22003267	C	T	W	stop	17	0.87	-	AT3G59570.1
1	3	22084518	C	T	R	H	14	1.00	-	AT3G59770.1
1	3	22752769	G	A	G	S	23	0.96	+	AT3G61480.1

<http://bar.utoronto.ca/ngm/description.html>

bio-svgenes by MacLean D.

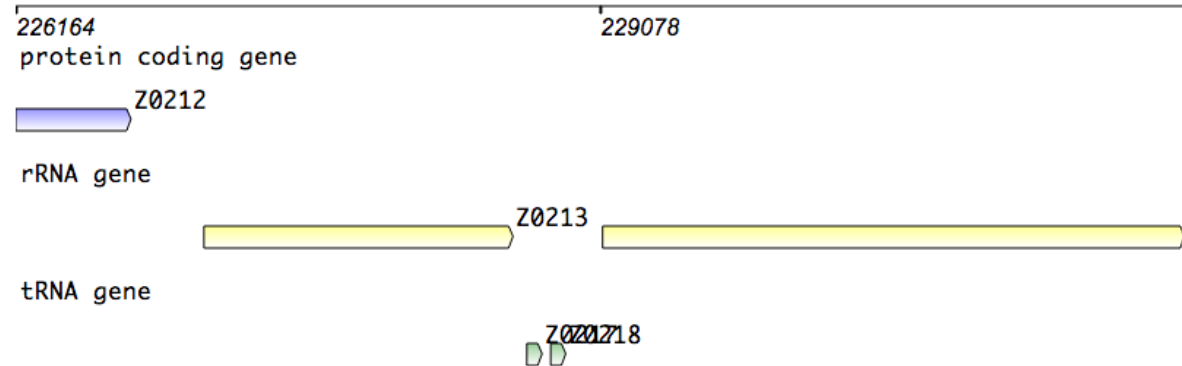
Bio::Graphics for BioRuby

to generate SVG images w/ an intuitive API and w/o dependencies

- `:directed` 
- `:transcript` 
- `:histogram` 
- `:circle` 
- `:down_triangle` 
- `:up_triangle` 
- `:span` 

```
page = Bio::Graphics::Page.new(opts) # sizes etc.  
gene = Bio::Graphics::MiniFeature.new(opts) # positions etc.  
gene_track = page.add_track(opts) # glyphs etc.  
gene_track.add(obj)
```

`page.draw` # => generate a SVG image



bio-diversity by Mozzherin D. *et al.*



Top downloaded
Biogem

Taxonomic scientific name parser to normalize species names from literatures in the best quality

Coeloglossum viride (L.) Hartman x Dactylorhiza majalis (Rchb. f.) P.F. Hunt & Summerhayes ssp. praetermissa (Druce) D.M. Moore & Soó



```
% gem install biodiversity19  
% nnparse find_scientific_names.txt
```

```
{"scientificName":{"parsed":true,"parser_version":"2.1.0","verbatim":"Coeloglossum viride (L.) Hartman x Dactylorhiza majalis (Rchb. f.) P.F. Hunt & Summerhayes ssp. praetermissa (Druce) D.M. Moore & Soó","normalized":"Coeloglossum viride (L.) Hartman x Dactylorhiza majalis (Rchb. f.) P.F. Hunt & Summerhayes ssp. praetermissa (Druce) D.M. Moore & Soó","canonical":"Coeloglossum viride x Dactylorhiza majalis praetermissa","hybrid":true,"details":{"genus":{"string":"Coeloglossum"},"species":{"string":"viride","authorship":"(L.) Hartman","combinationAuthorTeam":{"authorTeam":"Hartman","author":["Hartman"]},"basionymAuthorTeam":{"authorTeam":"L.","author":["L."]}}, {"genus":{"string":"Dactylorhiza"},"species":{"string":"majalis","authorship":"(Rchb. f.) P.F. Hunt & Summerhayes","combinationAuthorTeam":{"authorTeam":"P.F. Hunt & Summerhayes","author":["P.F. Hunt","Summerhayes"]},"basionymAuthorTeam":{"authorTeam":"Rchb. f.","author":["Rchb. f."]}}, {"infraspecies":{"string":"praetermissa","rank":"ssp.","authorship":"(Druce) D.M. Moore & Soó","combinationAuthorTeam":{"authorTeam":"D.M. Moore & Soó","author":["D.M. Moore","Soó"]},"basionymAuthorTeam":{"authorTeam":"Druce","author":["Druce"]}}},"parser_run":1,"positions":{"0":["genus",12],"13":["species",19],"21":["author_word",23],"25":["author_word",32],"35":["genus",47],"48":["species",55],"57":["author_word",62],"63":["author_word",65],"67":["author_word",71],"72":["author_word",76],"79":["author_word",90],"91":["intraspecific_type",95],"96":["intraspecies",108],"110":["author_word",115],"117":["author_word",121],"122":["author_word",127],"130":["author_word",133]}}
```

Global Biodiversity Information Facility / Encyclopedia of Life

Developed for Global Names Index <http://gni.globalnames.org/> supported by GBIF/EOL/NSF

Sister products:

- name-spotter -- Wrapper for name-finding libraries, TaxonFinder (EOL) and NetiNeti (for OCRred text)
- taxamatch_rb -- Tony Rees' algorithm for fuzzy matching of scientific names (compare with corpus)
- dwc-archive -- parser/generator for DarwinCore Archive (CSV + XML) format

Join us!

Without the **BioGem** system, we could not accumulate this variety of apps/libs only from the core **BioRuby** community!

BioRuby -- is a core **library**

BioGem -- can extend BioRuby, use BioRuby or also provide **apps!**

To find biogems -- <http://biogems.info/>

To create a biogem -- <http://biogems.info/howto.html>


Interviews w/ biogem developers -- [coming soon ...](#)

We welcome your contributions especially on

- Statistics
- Semantic Web
- Command line apps
- Web apps and visualization tools
- and something new!

BiInterchange by Baran J. *et al.*

RDF converters for TSV, XML, GFF3, GVF, Newick and other files



BiInterchange
Interchange data using the Resource Description Framework (RDF) and let BiInterchange automatically create RDF triples from your TSV, XML, GFF3, GVF, Newick and other files. BiInterchange helps you transform your data sets into linked data for sharing and data integration via command line, web-service, or API.

<http://biinterchange.org/>

Spin-off project from the BioHackathons in 2012 and 2013

- Developed ontologies for GFF and GVF

Next release:

- Utilizes FALDO location ontology and Identifiers.org URIs

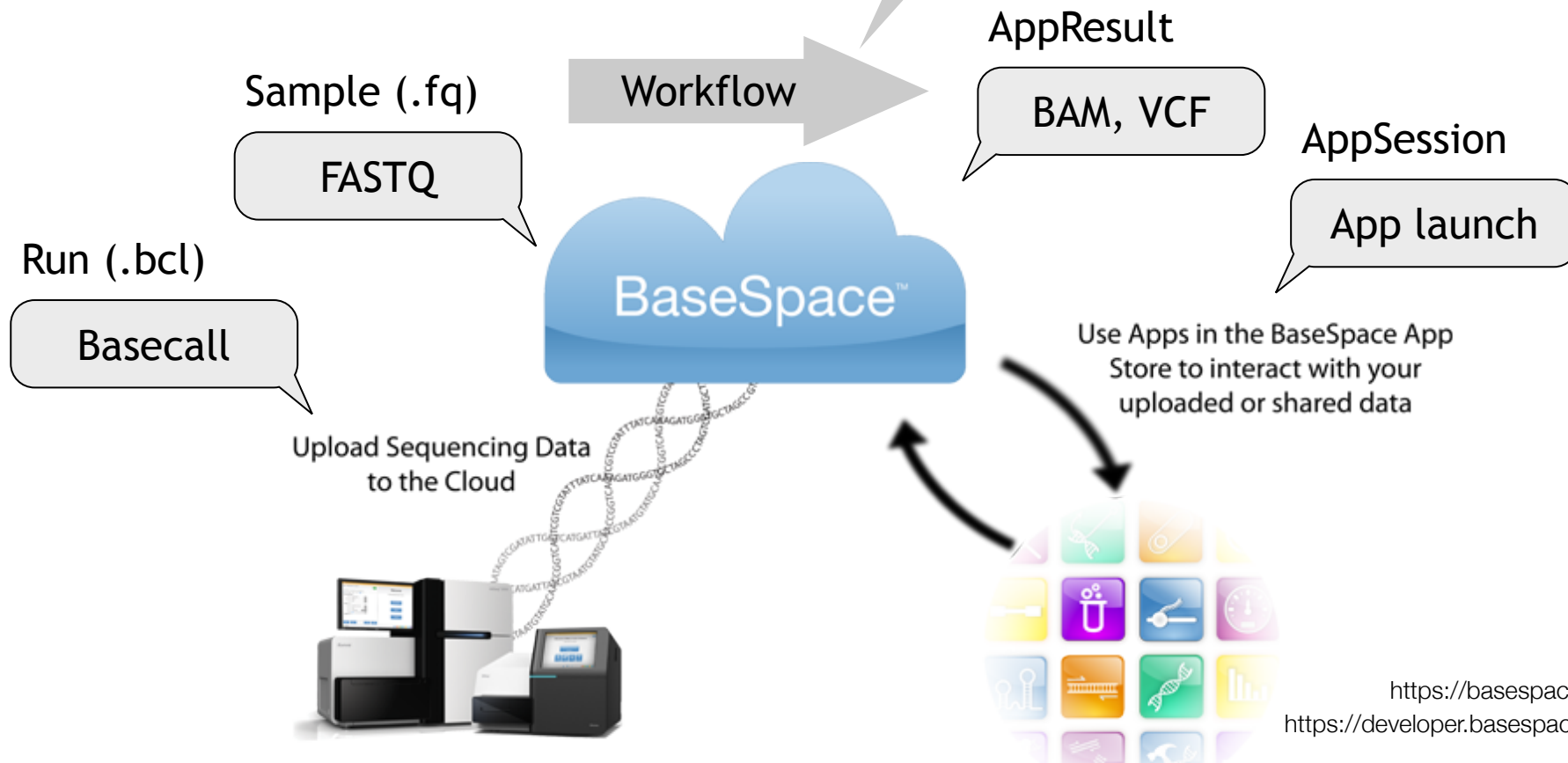
We are also working on converters for

GTF, VCF, PubMed, and INSDC data w/ appropriate ontologies

Creating BaseSpace apps

Project = Samples + AppResults

Mapping: BAM - BWA, Bowtie, SW, iSAAC, ELAND
Variation: VCF - GATK, Somatic Variant Caller, Starling 2
:



BaseSpace Ruby SDK

BaseSpace - Illumina's cloud solution comes w/ Python, Java, R SDKs
Ruby version of SDK is developed by the BioRuby group in 2013
[git clone https://github.com/joejimbo/basespace-ruby-sdk](https://github.com/joejimbo/basespace-ruby-sdk)
(will be available on Illumina's web site shortly)

Developers can create your own app

- You can easily utilize your NGS biogem w/ BaseSpace Ruby SDK
- You will easily obtain much more users

Users can use your app without coding

- Don't need to learn programming. Just a click!

BioBaseSpace for non-Ruby programmers

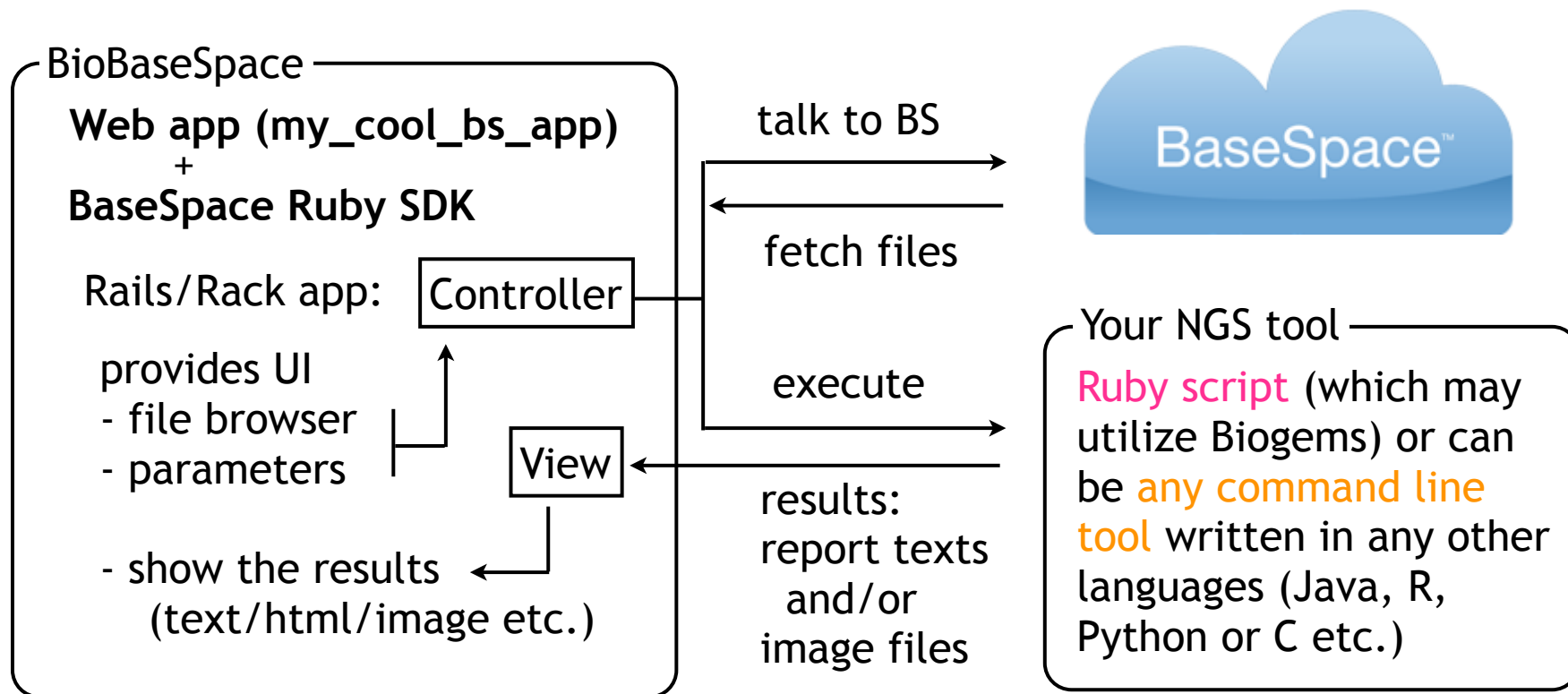
During the Codefest 2013, we found that it can be a burden to create new Web app from scratch on top of your NGS program. So we started new project to provide a Web-app scaffold for BS.



BioBaseSpace

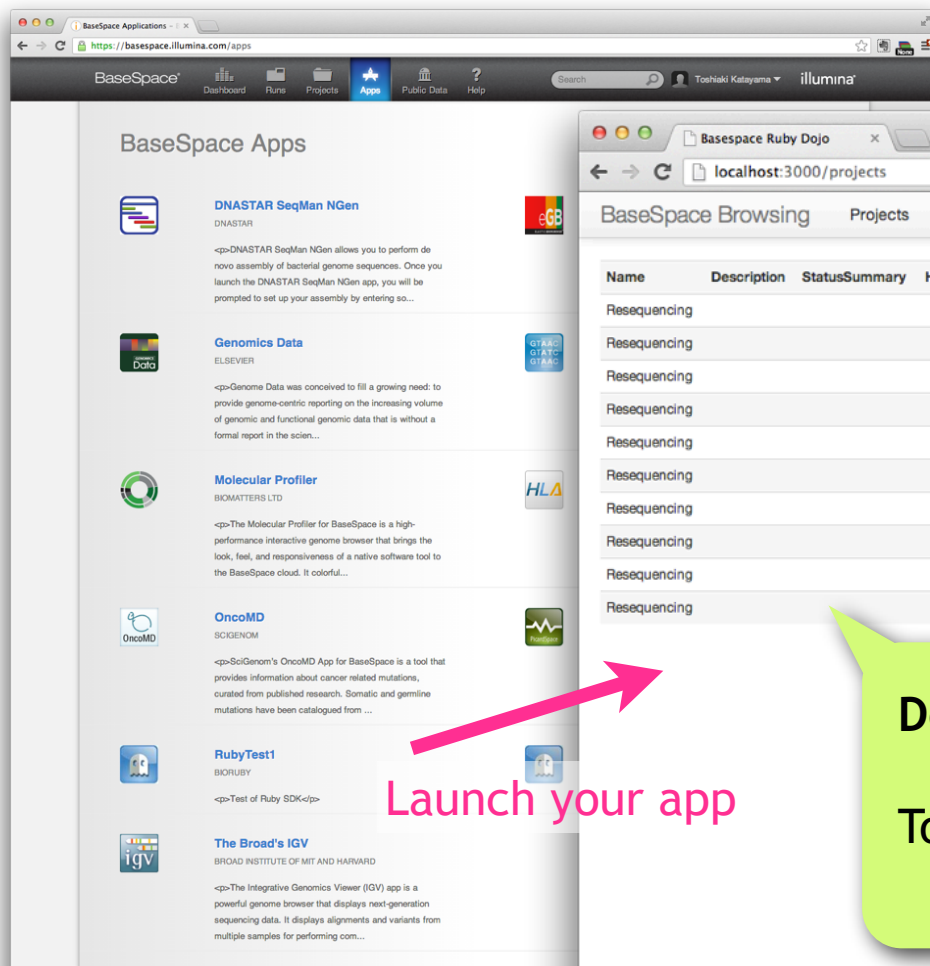
```
% biobasespace create my_cool_bs_app  
% biobasespace deploy my_cool_bs_app --to (AWS|Heroku|others|localhost)
```

just configure the program and parameters to be executed in the app



BioBaseSpace by Bonnal R. *et al.*

Scaffold your BS Web app w/ BaseSpace Ruby SDK inside



Codename: basespace-dojo & basespace-ninja

The screenshot shows the 'BaseSpace Ruby Dojo' web application. The browser address bar is 'localhost:3000/projects'. The application has a navigation bar with 'BaseSpace Browsing', 'Projects', 'Samples', 'Contact Us', and 'About'. Below the navigation bar is a table with the following data:

Name	Description	StatusSummary	HrefFiles	DateCreated	Id	Href	UserOwnedBy	StatusDetail	HrefGenome	AppSession	References
Resequencing				2012-08-19T21:14:13+00:00	1031	v1pre3/appresults/1031	1001: illumina Inc				
Resequencing				2012-08-19T21:14:13+00:00	1032	v1pre3/appresults/1032	1001: illumina Inc				
Resequencing				2012-08-19T21:14:13+00:00	1033	v1pre3/appresults/1033	1001: illumina Inc				
Resequencing				2012-08-19T21:14:13+00:00	1034	v1pre3/appresults/1034	1001: illumina Inc				
Resequencing				2012-08-19T21:14:13+00:00	1035	v1pre3/appresults/1035	1001: illumina Inc				
Resequencing				2012-08-19T21:14:13+00:00	1036	v1pre3/appresults/1036	1001: illumina Inc				
Resequencing				2012-08-19T21:14:13+00:00	1037	v1pre3/appresults/1037	1001: illumina Inc				
Resequencing				2012-08-19T21:14:13+00:00	1038	v1pre3/appresults/1038	1001: illumina Inc				
Resequencing				2012-08-19T21:14:13+00:00	1039	v1pre3/appresults/1039	1001: illumina Inc				
Resequencing				2012-08-19T21:14:13+00:00	1040	v1pre3/appresults/1040	1001: illumina Inc				

Done: Authentication & file browsing

Todo: Configuration of your tool to exec & showing the result

Acknowledgements

BioRuby core

Naohisa Goto

+ panel members

+ many contributors <http://bioruby.open-bio.org/wiki/Contributors>

Biogem system

Raoul Bonnal, Pjotr Prins, Francesco Strozzi

Biogem developers

Many! <http://biogems.info/>

BioInterchange

Joachim Baran *et al.*

BaseSpace Ruby SDK

Toshiaki Katayama, Joachim Baran, Eri Kibukawa, Raoul Bonnal, Francesco Strozzi