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  Instituto Nazionale Genetica Molecolare, Italy
Francesco Strozzi  Parco Tecnologico Padano, Italy
Naohisa Goto  Osaka University, Japan
BioRuby stairway to freedom

Project started [http://bioruby.org/](http://bioruby.org/)

Open Bio Japan [http://open-bio.jp/](http://open-bio.jp/)

O|B|F
BioHackathons

Grant by IPA

NBDC/DBCLS

Joined to O|B|F

BioRuby-1.0

BioRuby paper

GSoC
GitHub

Biogem paper

Biogem

BioRuby-1.5.0

Free to Use/Copy/Modify

Join w/ approval (CVS/SVN commits)

Fork (GitHub)

Extend/Distribute (Biogem)
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!
>80 Biogems have been released so far
BioRuby >44K DLs
Biogems vary from ~100 to 100K DLs
>20 new biogems have been developed since BOSC2012
## Biogem packages

<table>
<thead>
<tr>
<th>Web services</th>
<th>Seq analysis</th>
<th>NGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>bio-ucsc-api</td>
<td>bio-gff3</td>
<td>bio-ngs</td>
</tr>
<tr>
<td>intermine</td>
<td>bio-hmmer3</td>
<td>bio-samtools</td>
</tr>
<tr>
<td>entrez, eutils</td>
<td>bio-maf</td>
<td>bio-gadget</td>
</tr>
<tr>
<td>ruby-ensembl-api</td>
<td>bio-phyta</td>
<td>bio-faster</td>
</tr>
<tr>
<td>bio-dbsnp</td>
<td>bio-alignment</td>
<td>bio-sambamba</td>
</tr>
<tr>
<td>bio-chembl</td>
<td>bio-signalp</td>
<td>bio-tabix</td>
</tr>
<tr>
<td></td>
<td>bio-tm-hmm</td>
<td>bio-bgzf</td>
</tr>
<tr>
<td></td>
<td>bio-isoelectric-point</td>
<td>bio-gngm</td>
</tr>
<tr>
<td></td>
<td>bio-genomic-interval</td>
<td></td>
</tr>
<tr>
<td></td>
<td>bio-kmer-counter</td>
<td></td>
</tr>
<tr>
<td></td>
<td>bio-restriction-enzyme</td>
<td></td>
</tr>
<tr>
<td>Phyloinformatics</td>
<td></td>
<td>Others</td>
</tr>
<tr>
<td>biodiversity</td>
<td></td>
<td>bio-svgenes</td>
</tr>
<tr>
<td>name-spotter</td>
<td></td>
<td>biointerchange</td>
</tr>
<tr>
<td>bio-nexml</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bio-phyloxml</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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!)

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/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

DOI: 10.1038/nprot.2012.022 PMID: 22481528

http://www.nature.com/nmeth/journal/v9/n1/full/nmeth.1778.html
DOI: 10.1038/nmeth.1778 PMID: 22101854
bio-gngm 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

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

```ruby
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
```

---

* :directed
* :transcript
* :histogram
* :circle
* :down_triangle
* :up_triangle
* :span

---

226164
protein coding gene

229078

20212

rRNA gene

20213

tRNA gene

20202718
bio-diversity by Mozzherin D. et al.

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

% gem install biodiversity19
% nnparse find_scientific_names.txt

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

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 OCRed 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!
BioInterchange by Baran J. *et al.*

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

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

http://biointerchange.org/
Creating BaseSpace apps

Project = Samples + AppResults

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

Workflow

Sample (.fq) -> Basecall -> Run (.bcl) -> Sample (.fq) -> FASTQ

AppResult

BAM, VCF

AppSession

App launch

Use Apps in the BaseSpace App
Store to interact with your uploaded or shared data

https://basespace.illumina.com/
https://developer.basespace.illumina.com/
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.git  
(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 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**

**Web app (my_cool_bs_app)**

+ BaseSpace Ruby SDK

Rails/Rack app:
- Controller
- View
  - provides UI
    - file browser
    - parameters
  - show the results
    (text/html/image etc.)

Your NGS tool

Ruby script (which may utilize Biogems) or can be any command line tool written in any other languages (Java, R, Python or C etc.)

BaseSpace

talk to BS
fetch files
execute
results:
- report texts and/or image files
BioBaseSpace by Bonnal R. et al.

Scaffold your BS Web app w/ BaseSpace Ruby SDK inside

Codename: basespace-dojo & basespace-ninja

Done: Authentication & file browsing

Launch your app

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

**BioRuby core**
Naohisa Goto
+ panel members

**Biogem system**
Raoul Bonnal, Pjotr Prins, Francesco Strozzi

**Biogem developers**
Many! [http://biogems.info/](http://biogems.info/)

**BiolInterchange**
Joachim Baran *et al.*

**BaseSpace Ruby SDK**
Toshiaki Katayama, Joachim Baran, Eri Kibukawa, Raoul Bonnal, Francesco Strozzi