



BioRuby and the KEGG API

Toshiaki Katayama

k@bioruby.org

Bioinformatics center, Kyoto U., Japan

Use the source!

<http://bioruby.org/>

BOSC 2003

informatics
BioRuby.org

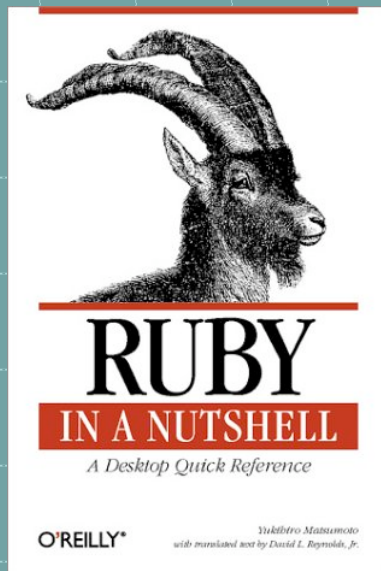
What is BioRuby?

- ◆ Yet another BioPerl written in Ruby
 - ◆ since Nov 2000
- ◆ Developed in Japan
 - ◆ includes support for Japanese resources like KEGG

So, what is Ruby?

- ◆ Scripting language
 - ◆ clean syntax, v. easy to use/read/write
- ◆ Everything is object
 - ◆ Integer, String, Regexp, Exception etc. w/o exception
- ◆ Created by Japanese author 'matz'

So, what is Ruby?



Pre-installed in Mac OS X!



<http://bioruby.org/>

BOSC 2003

informatics
BioRuby.org

What the Ruby code looks like?

```
#!/usr/bin/ruby
```

```
puts "Hello world!"
```

What the BioRuby code looks like?

```
#!/usr/bin/env ruby

require 'bio'

gene = Bio::Seq::NA.new("catgaattattgtagannntgataaagacttgac")
prot = gene.translate

# => "HELL*XW*RLD" (Bio::Seq::AA object)

puts prot.split('X').join(' ').capitalize.gsub(/\s/, 'o') << '!'

# => "Hello World!"
```

Bio::Sequence

- ◆ Bio::Sequence (aliased to Bio::Seq)

- ◆ seq.composition
- ◆ seq.window_search
- ◆ seq.randomize
- ◆ seq.to_fasta

- ◆ Bio::Sequence::NA

- ◆ seq.complement
- ◆ seq.splicing
- ◆ seq.translate
- ◆ seq.to_re

- ◆ Bio::Sequence::AA

- ◆ seq.molecular_weight

```
#!/usr/bin/env ruby
require 'bio'

seq = Bio::Seq::NA.new("atggcttcagt..
seq.window_search(15, 3) do |sub|
  puts sub.translate.molecular_weight
end
```

What BioRuby can do more?

- ◆ Biological sequence manipulations
- ◆ Run applications (Blast etc.) and parse its report
- ◆ Database entry retrieval, parsing
- ◆ PubMed reference DB search, reformatting
- ◆ Accessing OBDA, DAS, KEGG API

Bio::Applications

- ◆ Bio::Blast, Fasta, HMMER, EMBOSS
- ◆ Bio::Genscan
- ◆ Bio::PSORT, TargetP
- ◆ Bio::SOSUI, TMHMM
 - ◆ each contains Report class for parsing the results

example

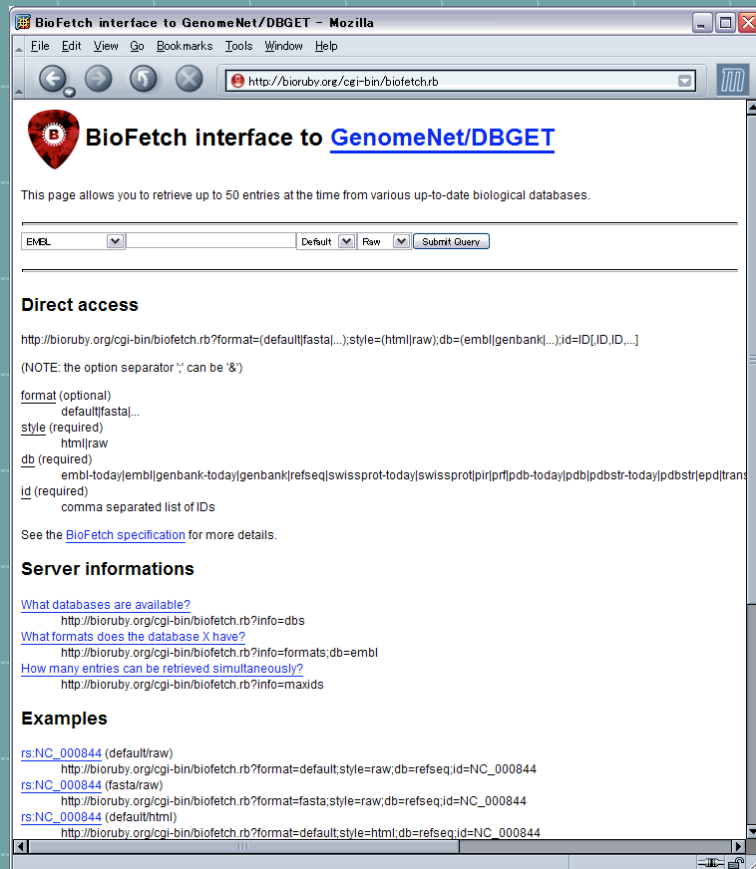
```
#!/usr/bin/env ruby
require 'bio'

File.open("my_blast_output.xml") do |file|
  Bio::Blast.reports(file) do |report|
    report.hits do |hit|
      puts hit.num, hit.query_id, hit.target_id,
            hit.bit_score, hit.evalue, hit.overlap
    end
  end
end
```


Bio::DB

- ◆ Bio::GenBank, EMBL, SwissProt, ...
- ◆ Bio::FastaFormat, GFF, GO
- ◆ Bio::MEDLINE, LITDB
- ◆ Bio::TRANSFAC, PROSITE
- ◆ Bio::FANTOM
- ◆ Bio::KEGG::GENES, ENZYME, Microarray, ...
- ◆ Bio::AAindex

OBDA -- Open Bio* DB Access



◆ Bio::Registry

◆ Bio::Fetch

◆ Bio::FlatFile

◆ Bio::SQL

+

◆ Bio::DAS

OBDA example

◆ ~/.bioinformatics/seqdatabase.ini

```
[embl]
protocol=biofetch
location=http://bioruby.org/cgi-bin/biofetch.rb
dbname=embl
```

```
[swissprot]
protocol=biosql
location=db.bioruby.org
dbname=biosql
driver=mysql
biodbname=sp
```

```
[genbank]
protocol=flat
location=/export/database/
dbname=genbank
```

```
#!/usr/bin/env ruby
require 'bio'
reg = Bio::Registry.new

sp = reg.get_database('swissprot')
puts sp.get_by_id('CYC_BOVIN')

gb = reg.get_database('genbank')
puts gb.get_by_id('AA2CG')
```

What is KEGG?

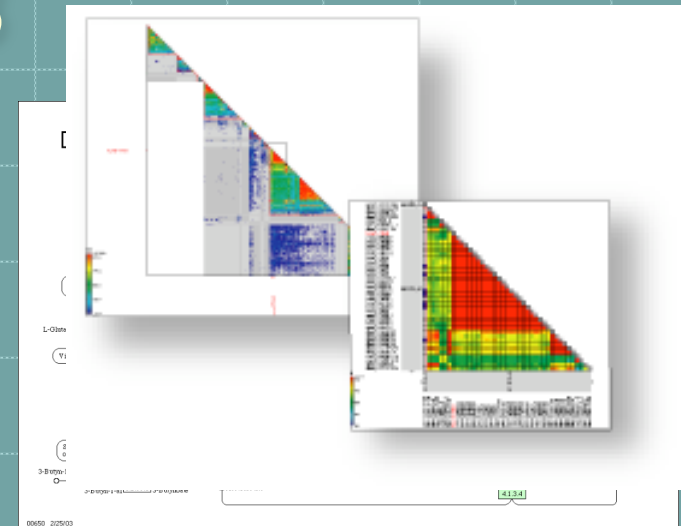


◆ GENOME, GENES

◆ PATHWAY

◆ SSDB

etc.



KEGG in GMOD

KEGG ana : Anabaena sp. PCC 7120

Showing 20 Kbp from ana, positions 1,555,152 to 1,575,151

Instructions [Hide]: Search using a sequence name, gene name, locus, or other landmark. The wildcard character * is allowed. Examples: ana, palpha, pbeta, pdelta, pepsilon, pgamma, pzeta. [Help]

To center on a location, click the ruler. Use the Scroll/Zoom buttons to change magnification and position. To save this view, bookmark this link.

Landmark or Region Search Scroll/Zoom: <<< Show 20 Kbp + >>>

Overview of ana

1560k 1570k

alr1311r1312 alr1313alr1314alr1315 alr1317r1318 alr1319alr1320alr1322 alr1323 alr1324 alr13

leuB
3-isopropylmalate dehydrogenase [EC:1.1.1.85]

vals
valyl-tRNA synthetase [EC:6.1.1.91]

ana00290
Metabolism: Amino Acid Metabolism; Valine, leucine and isoleucine biosynthesis

ana03090
Using and Degradation: Type II secretion system

ana00770
Information Processing: Translation; Aminoacyl-tRNA biosynthesis

Metabolism: Metabolism of Cofactors and Vitamins;

RNA Pathway

encoding / misc RNAs

Dumps, Searches and other Operations:
Annotate Restriction Sites

Tracks [Hide]
(External tracks italicized)
 CDS scRNA 3-frame translation (reverse)
 Named gene Noncoding / misc RNAs Pathway

http://www.genome.ad.jp/dbget-bin/www_bget?ana:all:319

- ◆ <http://gmod.bioruby.org>
- ◆ Import KEGG/GENES and KEGG/PATHWAY into GMOD browser (converted by BioRuby)
- ◆ Over 100 organisms in unified form
- ◆ linked to GENES and PATHWAY database

KEGG API

- ◆ <http://www.genome.ad.jp/kegg/soap/>
- ◆ SOAP/WSDL based web service
 - ◆ XML, HTTP
- ◆ Proteome and pathway analysis
 - ◆ KEGG/GENES
 - ◆ KEGG/SSDB
 - ◆ KEGG/PATHWAY

example

ISMB poster D-25

```
#!/usr/bin/env ruby
require 'bio'

serv = Bio::KEGG::API.new
puts serv.get_best_neighbors_by_gene('eco:b0002')

# serv.get_best_neighbors_by_gene('eco:b0002', 500)
# serv.get_best_neighbors_by_gene('eco:b0002', 500, ['hin', 'syn'])

list = ['ec:1.1.1.1', 'ec:1.2.1.1']
puts serv.get_genes_by_enzymes(list, ['eco', 'syn'])

list = ['eco:b0600', 'eco:b1190']
puts serv.mark_all_pathways_by_genes('eco', list)
```

Example : Ortholog cluster and common motif

```
#!/usr/bin/ruby
```

```
require 'bio'
```

```
serv = Bio::KEGG::API.new
```

```
homologs = serv.get_all_neighbors_by_gene('hsa:7368', 500)
```

```
pfams = {}
```

```
homologs.each do |g|
```

```
  motifs = serv.get_common_motifs_by_genes(g[0])
```

```
  motifs.each do |m|
```

```
    if m[0] =~ /pf:./
```

```
      if pfams.key?(m)
```

```
        pfams[m] << g[0]
```

```
      else
```

```
        pfams[m] = [g[0]]
```

```
      end
```

```
    end
```

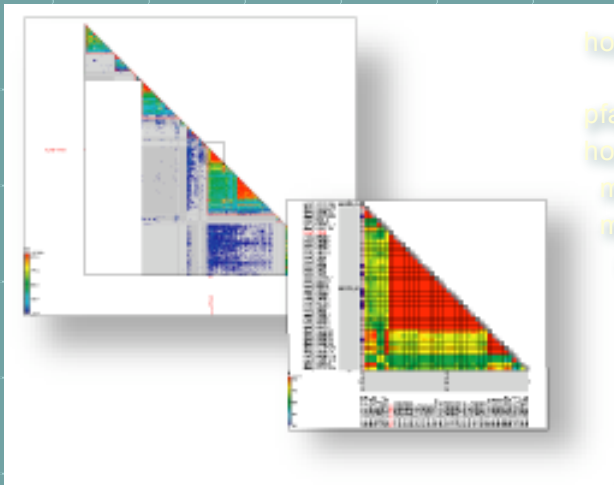
```
  end
```

```
end
```

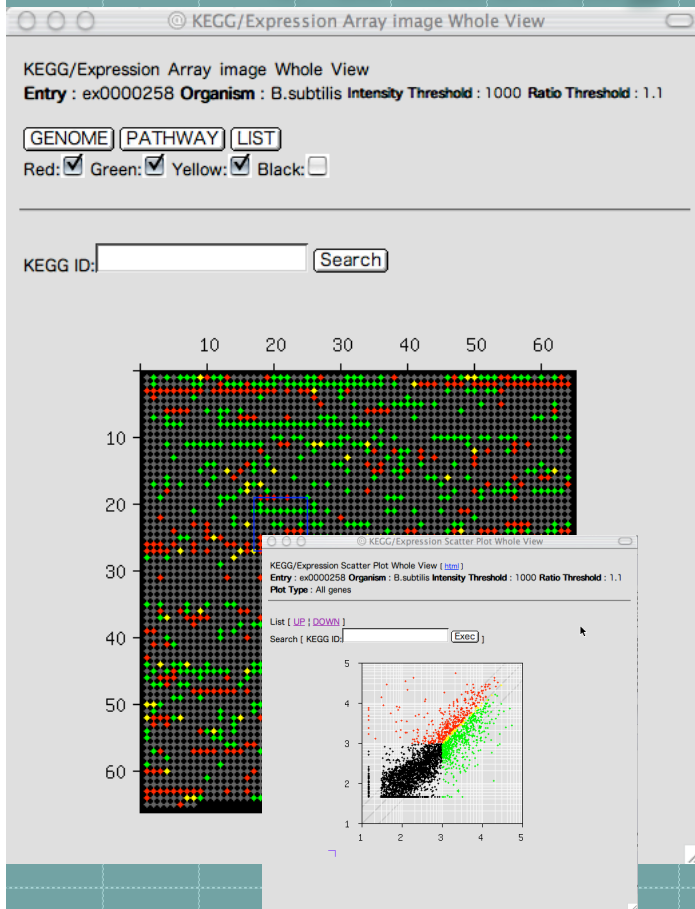
```
pfams.each do |m|
```

```
  p m
```

```
end
```



Example: Gene expression and Pathway analysis by KEGG API

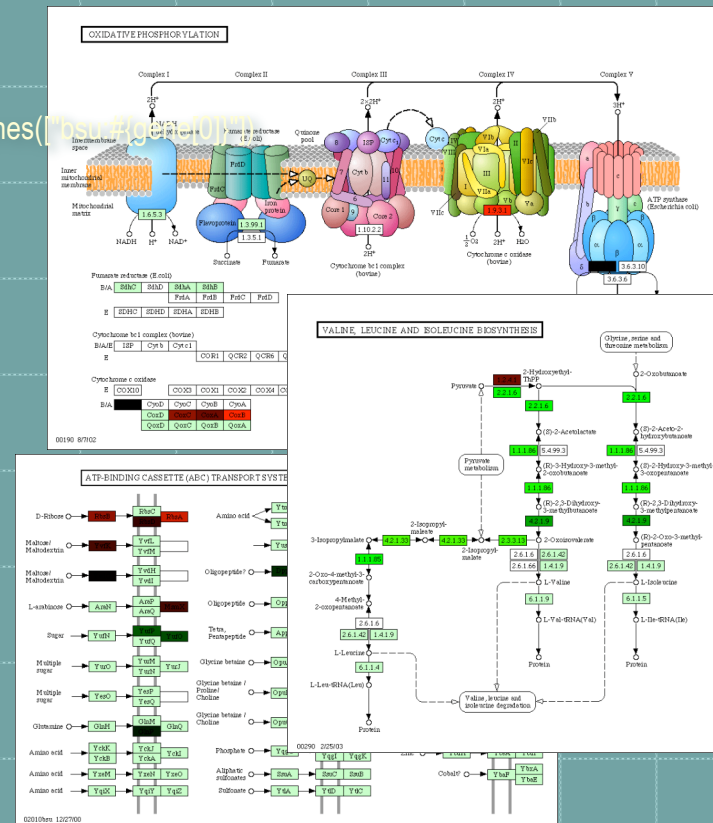


```
serv = Bio::KEGG::API.new
```

```
regulated_genes.each do |gene|
  map_id = serv.get_pathways_by_genes("bsu:#{gene}")
  map_id.each do |entry|
    entry.gsub!('map', 'bsu')
    entry.gsub!('path', 'map')
    if map_gene.key?(entry)
      map_gene[entry] << gene[0]
    else
      map_gene[entry] = [gene[0]]
    end
  end
end
```

```
map_gene.each do |k, v|
  colors = []
  v.each do |g|
    colors << "##{ma.of2rgb[g]}"
  end
end
```

```
p serv.color_pathway_by_genes("bsu:#{k}", v, colors)
end
```



Acknowledgement

- ◆ BioRuby developers
 - ◆ Naohisa Goto, Mitsuteru Nakao, Yoshinori Okuji, Shuichi Kawashima, Masumi Itoh and many more.
- ◆ KEGG and KEGG API curators / developers
 - ◆ Bioinformatics center, Kyoto University, Japan
 - ◆ Yoko Sato, Shuichi Kawashima, Minoru Kanehisa
- ◆ Open Bio* community

http://bioruby.org/

BioRuby.org - Mozilla
http://bioruby.org/

/([biofetch](#) | [gmod](#) | [gb](#) | [g-p](#) | [ura](#) | [cvs](#) | [ftp](#)).bioruby.org/

BIORUBY.ORG

English Thai

About BioRuby

BioRuby プロジェクトは、Bioinformatics に必要な機能や環境を、国産のオブジェクト指向スクリプト言語 Ruby を用いて統合的に実装することを目標としたプロジェクトです。開発や質問が日本語でも行えるほか、KEGG など国内のデータベースにも積極的に対応していくなど、国内の Bioinformatics 研究者も参加しやすいです。BioPerl などと比べると後発ですが、逆にオブジェクト指向の導入や必要な機能を簡単に書き直していきける点や、スクラッチからの開発により独自の発想と必要な機能を自分リットです。できる限り楽しくオープンな開発体制を整える予定なので、積極的に参

News

- 2003/02/17-22 BioHackathon 2003 (k)
- 2003/01/28 infobiologist: 第二回研究会@遠伝研 発表資料 I, II (ng,k)
- 2003/01/28 BioRuby 0.4.0 リリース (ChangeLog)

Roadmap

- 2003/06/27-28 BOSC 2003

Design

BioRuby 設計方針

- ・オブジェクトを適切に切り分け、コードをシンプルに保つ (KISS)
- ・分かりやすい(日本語|英語)ドキュメント (RD) とサンプルコードを付ける
- ・性能よりも分かりやすさを優先 (biologist がすぐに使えるレベル)

Projects

- ・DB パーザ (GenBank, Swiss-Prot, PDB, KEGG etc.) (k,n,s)
- ・FASTA/SSEARCH/BLAST インターフェイス、パーザ (k,n)
- ・DBGET/NCBI/EMBL モジュール mix-in (n,k)
- ・エクスペリメンテーション (s,n,k,o)
- ・パスウェイ解析 (k,s,a)
- ・ピジュアライゼーション: ゲノム、エクスペリメンテーション、クラスタ、パスウェイ (k,o,m,a)
- ・バックエンドの置き換え用高速演算ライブラリ: BioRubyPro (o)
- ・分散データベース: KEGG/GENES - dRuby (k,o,n)
- ・ニュースポータル: Bioinformatics - eRuby (k,m)

Mailing Lists

BioRuby 日本語メーリングリスト: ja@bioruby.org
登録方法: ja-request@bioruby.org 宛に Subject: subscribe のメールを送ってください。
解除方法: ja-request@bioruby.org 宛に Subject: unsubscribe のメールを送ってください。

BioRuby 開発者用メーリングリスト (英語): dev@bioruby.org
登録方法: dev-request@bioruby.org 宛に Subject: subscribe のメールを送ってください。
解除方法: dev-request@bioruby.org 宛に Subject: unsubscribe のメールを送ってください。

BioRuby CVS 変更通知リスト (自動): cvs@bioruby.org
登録方法: cvs-request@bioruby.org 宛に Subject: subscribe のメールを送ってください。
解除方法: cvs-request@bioruby.org 宛に Subject: unsubscribe のメールを送ってください。
確認のメールが送られるので、本文を引用して返信してください。

Wiki

BioRuby 開発用 RWiki ページ: <http://ura.bioruby.org/>

Open Bio* Info - q-p - Mozilla
http://q-p-bioruby.org/

[Yendot.org | Slashdot.jp | SourceForge.jp | Yahoo Genome | BizTech | PhRMA | BioInforM | GenomeWeb | BioIT World]

Open Bio* Info

Special FANTOM2 Issue

Entrez PubMed

1. Identification and analysis of chromodomain-containing proteins encoded in the mouse transcriptome. (genome res)

2. Comparative analysis of apoptosis and inflammation genes of mice and humans. (genome res)

3. Antisense Transcripts With FANTOM2 Clone Set and Their Implications for Gene Regulation. (genome res)

4. Staying Straight with A-tracts: A DNA Analog of the HIV-1 Polypurine Tract. (mb)

5. High-frequency gene transfer from the chloroplast genome to the nucleus. (pnas)

6. Conservation of genome content and virulence determinants among clinical and environmental isolates of Pseudomonas aeruginosa. (pnas)

7. In-phase implies large likelihood for independent codon model: distinguishing coding from non-coding sequences. (ib)

8. Genome analysis of a novel Shiga toxin 1 (Stx1)-converting phage which is closely related to Stx2-converting phages but not to other Stx1-converting phages. (baa)

9. Whole-Genome DNA Microarray Analysis of a Hyperthermophile and an Archaeon: Pyrococcus furiosus Grown on Carbohydrates or Peptides. (baa)

10. Suppressor mutations in the study of photosystem I biogenesis: sI0088 is a previously unidentified gene involved in reaction center accumulation in Synechocystis sp. strain PCC 6803. (baa)

11. Human and mouse genomic sequences reveal extensive breakpoint reuse in mammalian evolution. (pnas)

12. Gene function analysis in environmental isolates: The rif regulon of the strict iron oxidizing bacterium Leptospirillum ferrooxidans. (pnas)

Google

先日中先生の回読でこんな本が出てのを知ったのでした。NAR DB issue に掲載していた Andy Baxevanis 監修で BioPerl/GMOD の Lincoln Stein も書いています。目次を見ると1000ページぐらいありそうなき。BLAST はどうかな?

(Read More... | 22 hits)

O'REILLYからBLASTの本が出版される

Posted via きゅうびい on Fri, 20 Jun 2003 16:54:13 +0900 from BioRuby dept.

ngolo writes 表紙は何の魚で、O'REILLYからBLASTの本が出版される

(Read More... | 14 hits)

New dictyBase: An Online Informatics Resource for Dictyostelium

Posted via きゅうびい on Tue, 17 Jun 2003 02:25:02 +0900 from BioRuby dept.

k writes 雑誌 New dictyBase: An Online Informatics Resource for Dictyostelium

GMOD 使ってますね

BioRuby 0.5.0 is out!
25 Jun 2003

<http://bioruby.org/>

BOSC 2003

informatics
BioRuby.org