

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

# BioRuby

片山 俊明 <k@bioruby.org>

京大化研バイオインフォマティクスセンター  
2003/1/28 infobiologist 第二回研究会 @ 遺伝研



# Open Bio\*

- O|B|F -- Open Bio Foundation
  - バイオインフォマティクスに関するオープンソースコミュニティ
- BioRuby
- BioPerl
- BioPython
- BioJava
- BioDAS
- BioMOBY
- BioPipe
- EMBOSS
- Ensembl
- OmniGene
- GMOD
- GBrowse
- Apollo
- OBDA
- BioCaml
- BioLisp
- BioCyc
- BioConductor
- BioPathways
- BioBlog
- BioLinux
- :

# BioHackathon

- 2002/01 Arizona, 2002/02 Cape Town
- 2003/02 Singapore



- Will be annual event?
  - Hackathon → 1 week intensive hacking
  - BOSC (ISMB) → for open discussion

# Schedule – hack hack hack

BIO HACKATHON TIME MANAGEMENT PLAN			DATE: 24 FEB 02 - 02 MAR 02					
Starting	Ending	SUNDAY	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY
		24-Feb-02	25-Feb-02	26-Feb-02	27-Feb-02	28-Feb-02	1-Mar-02	2-Mar-02
7:00 AM	7:30 AM							
7:30 AM	8:00 AM							
8:00 AM	8:30 AM							
8:30 AM	9:00 AM		Hack Hack Hack	Hack Hack Hack	Hack Hack Hack	Hack Hack Hack	Hack Hack Hack	
9:00 AM	9:30 AM							
9:30 AM	10:00 AM							
10:00 AM	10:30 AM							
10:30 AM	11:00 AM		Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	
11:00 AM	11:30 AM		Hack Hack Hack	Hack Hack Hack	Hack Hack Hack	Hack Hack Hack	Hack Hack Hack	
11:30 AM	12:00 PM							
12:00 PM	12:30 PM							
12:30 PM	1:00 PM	Lunch	Lunch	Lunch	Lunch	Lunch	Lunch	
1:00 PM	1:30 PM							
1:30 PM	2:30 PM		Hack Hack Hack	Off site Touring	Hack Hack Hack	Hack Hack Hack	Hack Hack Hack	
2:00 PM	2:30 PM			Dive / Snorkelling				
2:30 PM	3:00 PM			Kirstenbosch hike				
3:00 PM	3:30 PM			Waterfront				
3:30 PM	4:00 PM		Coffee Break		Coffee Break			
4:00 PM	4:30 PM	Opening Reception &	Hack Hack Hack		Hack Hack Hack	Township Tour &	Coffee Break	
4:30 PM	5:00 PM	Cocktails & Snacks				African Café	Closing remarks	
5:00 PM	5:30 PM			Hack Hack Hack				
5:30 PM	6:00 PM							
6:00 PM	6:30 PM	Table Mountain						
6:30 PM	7:00 PM							
7:00 PM	7:30 PM		Dinner outside	Dinner at Hotel	Dinner at Hotel		Closing Party Outside	
7:30 PM	8:00 PM	at Castle						
8:00 PM	8:30 PM							
8:30 PM	9:00 PM			Hack Hack Hack	Hack Hack Hack	Hack Hack Hack		
9:00 PM	9:30 PM							
9:30 PM	10:00 PM							
10:00 PM	10:30 PM							
10:30 PM	11:00 PM							
11:00 PM	11:30 PM							
11:30 PM	12:00 AM							
12:00 AM	12:30 AM							
12:30 AM	1:00 AM							





	Par 1	Java	Python	Perl	C
Flat File, File	3/4	Matt 0.5	1.0	X	Server/Networking 1.0.
Flat File, RDB	Lindon	X	1.0	X	
BioFetch	Ben Karp	(Dyfed)		Server.	
Registry	Ewan	0.0	0.9	C.	
BioCorba	Ewan 3/4	1.0	0.9	...	
XEMBL	Lindon	1.0	1.0	X	
BioSQL	Mark?	5/4	0.9	1.0	0.9

Test Instance

James + Ewan.

SMT

f



# BioRuby hackathon results

- BioFetch
  - <http://bioruby.org/cgi-bin/biofetch.rb>
- BioSQL
  - EMBL/GenBank/SwissProt in MySQL/PostgreSQL
- BioRegistry
  - `~/.bioinformatics/seqdatabase.ini`
  - `/etc/bioinformatics/seqdatabase.ini`
  - <http://www.open-bio.org/registry/seqdatabase.ini>

# OBDA

- どの Open Bio\* ライブラリで作成したデータベースも共通にアクセスできる
- ユーザは自分の好きな言語で好きなところからデータベースにアクセスすればOK

- Open Bio\* Sequence Database Access
  - BioRegistry (Stanza)
  - BioFlat (Simple index, BDB)
  - BioFetch (CGI/HTTP)
  - BioSQL (MySQL, PostgreSQL, Oracle)
  - XEMBL (SOAP based)
  - BioCORBA (BSANE compliant)

– <http://obda.open-bio.org/>

# BioFlat

- インデックス作成

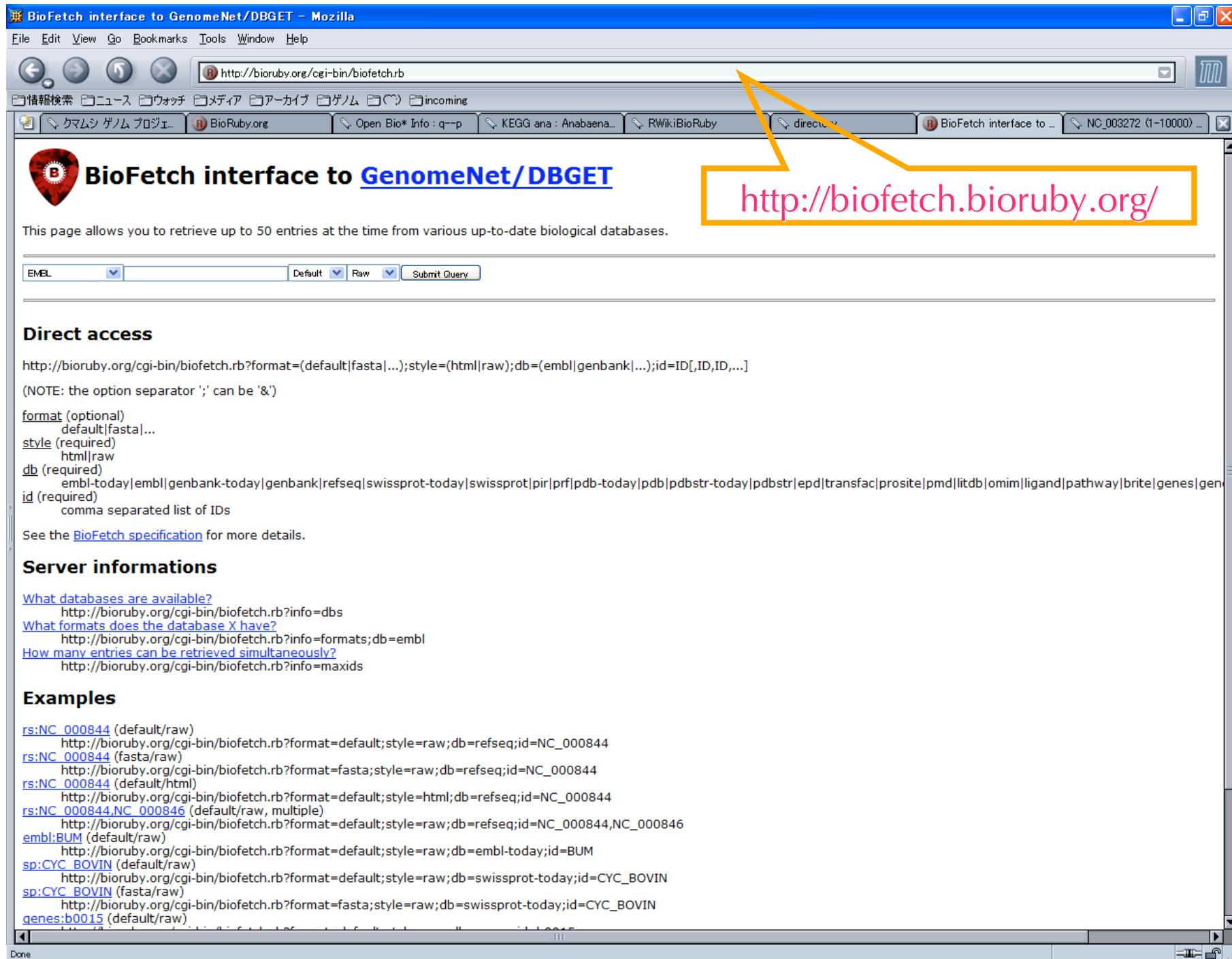
```
% bioflat --makeindex mydatabase gbvr1*.seq
```

- 検索

```
% bioflat mydatabase ENTRY_ID
```

# BioFetch

- WWW (HTTP) の CGI 呼び出しによるエントリ取得
  - サーバ
    - EBI dbfetch
      - <http://www.ebi.ac.uk/cgi-bin/dbfetch>
    - BioRuby biofetch.rb
      - <http://bioruby.org/cgi-bin/biofetch.rb>
  - 引数
    - format=default または format=fasta (必須ではない)
    - style=html または style=raw
    - db=genbank とか db=prosite とか db=pathway ...
    - id=エントリのID を カンマ で区切ったもの
  - 例 <http://bioruby.org/cgi-bin/biofetch.rb?format=default;style=raw;db=embl;id=BUM>

A screenshot of a Mozilla Firefox browser window. The title bar reads "BioFetch interface to GenomeNet/DBGET - Mozilla". The address bar shows the URL "http://bioruby.org/cgi-bin/biofetch.rb". A yellow arrow points from the text "http://biofetch.bioruby.org/" to the address bar. The page content is the BioFetch interface itself, featuring a logo, a "Direct access" section with a query string example, a "Server informations" section with links to database info, and an "Examples" section with several URLs. The bottom right corner of the browser window has a watermark for "informatics.org".

**BioFetch interface to GenomeNet/DBGET**

This page allows you to retrieve up to 50 entries at the time from various up-to-date biological databases.

**Direct access**

`http://bioruby.org/cgi-bin/biofetch.rb?format=(default|fasta|...);style=(html|raw);db=(embl|genbank|...);id=ID[,ID,ID,...]`

(NOTE: the option separator ';' can be '&')

format (optional)  
  default|fasta|...  
style (required)  
  html|raw  
db (required)  
  embl-today|embl|genbank-today|genbank|refseq|swissprot-today|swissprot|pir|prf|pdb-today|pdb|pdbstr-today|pdbstr|epd|transfac|prosite|pmd|litdb|omim|ligand|pathway|brite|genes|gen  
id (required)  
  comma separated list of IDs

See the [BioFetch specification](#) for more details.

**Server informations**

[What databases are available?](#)  
http://bioruby.org/cgi-bin/biofetch.rb?info=dbs

[What formats does the database X have?](#)  
http://bioruby.org/cgi-bin/biofetch.rb?info=formats;db=embl

[How many entries can be retrieved simultaneously?](#)  
http://bioruby.org/cgi-bin/biofetch.rb?info=maxids

**Examples**

[rs:NC\\_000844](#) (default/raw)  
http://bioruby.org/cgi-bin/biofetch.rb?format=default;style=raw;db=refseq;id=NC\_000844

[rs:NC\\_000844](#) (fasta/raw)  
http://bioruby.org/cgi-bin/biofetch.rb?format=fasta;style=raw;db=refseq;id=NC\_000844

[rs:NC\\_000844](#) (default/html)  
http://bioruby.org/cgi-bin/biofetch.rb?format=default;style=html;db=refseq;id=NC\_000844

[rs:NC\\_000844\\_NC\\_000846](#) (default/raw, multiple)  
http://bioruby.org/cgi-bin/biofetch.rb?format=default;style=raw;db=refseq;id=NC\_000844,NC\_000846

[embl:BUM](#) (default/raw)  
http://bioruby.org/cgi-bin/biofetch.rb?format=default;style=raw;db=embl-today;id=BUM

[sp:CYC\\_BOVIN](#) (default/raw)  
http://bioruby.org/cgi-bin/biofetch.rb?format=default;style=raw;db=swissprot-today;id=CYC\_BOVIN

[sp:CYC\\_BOVIN](#) (fasta/raw)  
http://bioruby.org/cgi-bin/biofetch.rb?format=fasta;style=raw;db=swissprot-today;id=CYC\_BOVIN

[genes:b0015](#) (default/raw)

# BioSQL

- GenBank/EMBL/SwissProt エントリ
  - BioPerl の sequence オブジェクトに相当
- MySQL or PostgreSQL or Oracle に格納
- 目的
  - エントリ単位のデータ更新
  - データの正規化、オントロジーとの融合
  - RDBMS のもつ高速インデックスエンジン
- データのロードは BioPerl か BioJava でやってください

# BioRegistry

- 設定ファイル (Stanza フォーマット)

`~/.bioinformatics/seqdatabase.ini`

`/etc/bioinformatics/seqdatabase.ini`

`http://open-bio.org/registry/seqdatabase.ini`

[swissprot]

`protocol=biosql`

`location=db.bioruby.org`

`dbname=biosql`

`driver=mysql`

`biodbname=sp`

- DB name

[genbank] [swissprot] etc.

- Protocol

biosql, index-berkeleydb, index-flat,  
biofetch, bsane-corba, xembl

- Location

[emb1]

`protocol=biofetch`

`location=http://bioruby.org/cgi-bin/biofetch.rb`

`biodbname=emb1`

:



# OBDA を使ったエントリ取得

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

require 'bio'

reg = Bio::Registry.new # 自動的に設定ファイル読み込み

# OBDA のどの取得方法でも (HTTPとかMySQLとか気にしない)
#   db メソッドでデータベースを指定
#   fetch メソッドでエントリを得られる

db = reg.db("swissprot")
entry = db.fetch("TETW_BUTFI")
```

# Creating a bioinformatics nation

A web-services model will allow biological data to be fully exploited.

## Lincoln Stein

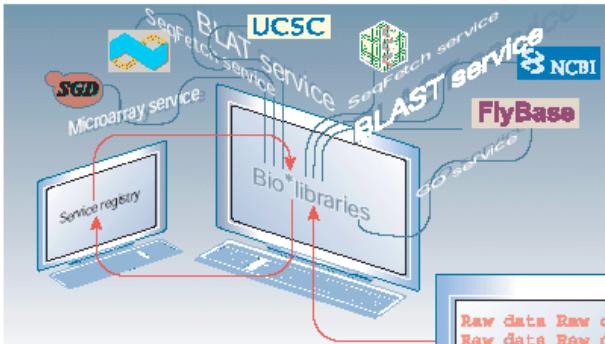
During the Middle Ages and early Renaissance, Italy was fragmented into dozens of rival city-states controlled by such legendary families as the Estes, Viscontis and Medicis. Though picturesque, this political fragmentation was ultimately damaging to science and commerce because of the lack of standardization in everything from weights and measures to the tax code to the currency to the very dialects people spoke. A fragmented and technologically weak society was vulnerable to conquest, and from the seventeenth to the nineteenth centuries Italy was dominated by invading powers.

The old city-states of Italy are an apt metaphor for bioinformatics today. The field is dominated by rival groups, each promoting its web sites, services and data formats. Unarguably, this environment of creative chaos has greatly enriched the field. But it has also created a significant hindrance to researchers wishing to exploit the wealth of genome data to its fullest.

Despite its shaky beginning, the nation of Italy was eventually forged through a combination of violent and diplomatic efforts. It is now a strong and stable component of a larger economic unit, the European Union, with which it shares a common currency, a common set of weights and measures, and a common set of rules for national and international commerce. My hope is that bioinformatics will one day achieve the same degree of strength and stability by adopting a universal code of conduct along the lines I propose here.

## Screen scraping: mediaeval torture

The promise and peril of the bioinformatics landscape is clear to any bench biologist attempting to mine the human genome for information on, say, a favourite genetic region. The online sources of these data each provide remarkable user interfaces and deeply interconnected data sets of great richness. Yet each interface is different, both in the subset of data presented and in organization. The researcher may find herself devoting as much time adjusting to differences in presentation of the data as she does actually thinking about them. The situation is worse when comparing a human gene to its orthologue in another species. This brings the model organism databases into play, each of which has its own type of user interface and format. (See [us.expasy.org/alinks.html](http://expasy.org/alinks.html); [www.stat.wisc.edu/biosci/genome.html](http://www.stat.wisc.edu/biosci/genome.html); and [rbcf.dfc.harvard.edu/cmsmbr/biotools/biotools10.html](http://rbcf.dfc.harvard.edu/cmsmbr/biotools/biotools10.html) for an idea of the scale of the problem.)



**Figure 1** Moving towards a bioinformatics nation. Because each data provider (such as Flybase and UCSC) publishes data in an idiosyncratic form, the Bio\* software package (Bio\* libraries) was created to massage data into a standard internal format. Unfortunately, Bio\* needs to be fixed each time a provider changes its formats. A web-services world would build on the successes of the Bio\* projects by defining standard interfaces to various types of computations and data formats. The Bio\* libraries can be written to recognize these interfaces, allowing them to interoperate easily with all data providers. A service registry would let data providers enter an electronic 'address book', allowing the Bio\* libraries to locate and interact with new data sources automatically.

This inconvenience for the bench biologist is disastrous for the bioinformaticist, who typically needs to aggregate data from many online sources to create a data set for further analysis. When these data reside on different servers using different data formats and access methods, the first step is to write a set of software 'scripts' to fetch them, reformat them and place the extract into a local database. This is not straightforward, because most online biological databases were designed to be accessed by humans, not by machines. Bioinformaticists often find themselves writing scripts to parse the HTML source to extract the data while ignoring graphics links and explanatory text, a process called screen scraping.

Screen scraping is despised for various reasons. First and foremost, it is brittle. Database managers are always tinkering with the user interface, adding a graphic here, moving a button there, to improve the user experience. Each small change in a popular web page breaks dozens of screen-scraping scripts, causing anguished cries and hair-tearing among the bioinformaticists who depended on those scripts for their research and the research of the wet labs they support. Second, it is unreliable. There is no published documentation of what a data source's web pages are supposed to contain, so bioinformaticists must guess from a few

## Bio\* projects reduce the pain

The bioinformatics community has responded to the challenges posed by this 'city-state' situation with the Bio\* projects, a series of freely available open-source projects ([www.open-bio.org](http://www.open-bio.org)), in which nearly a hundred software engineers have developed re-usable code libraries in the Perl, Java, Python and Ruby programming languages (known as BioPerl, BioJava, Biopython and BioRuby, respectively). These libraries automate common bioinformatics tasks, such as manipulating DNA and protein sequences, and provide methods for importing and exporting data between data sources and among file formats. To fetch a piece of data from a database, the bioinformaticist uses the Bio\* libraries to do the fetch, put the information in a standard format, and return the reformatted data to her script. This prevents duplication of effort. No one will ever again

Nature 417:119-120 (2002)





# BioRuby にできること

- Bio::Sequence, Bio::Location, Bio::Feature クラス
  - 塩基配列、アミノ酸配列の取り扱い
  - 組成、スプライシング、翻訳、ウィンドウサーチなど
- Bio::DB 系クラス
  - データベースのパーザ (現在 20 種くらい)
- Bio::Blast, Bio::Fasta クラス
  - Blast/Fasta を実行し結果をペーズするファクトリ
- Bio::PubMed, Bio::Reference クラス
  - 文献データベース検索、BibTeX などのフォーマッタ
- Bio::Registry, Bio::SQL, Bio::Fetch, Bio::FlatFile クラス
  - データベース入出力インターフェイス (OBDA 準拠)
- Bio::Pathway, Bio::Relation クラス
  - グラフ、2 項関係の演算

# コード比較

- FASTA ファイルを BioRuby で読み込む

```
#!/usr/bin/ruby

require 'bio'

flatfile = Bio::FastaFormat.open('filename')

flatfile.each do |entry|
  puts entry.entry_id
  puts entry.seq
  puts entry
end
```

- FASTA ファイルを BioPerl で読み込む

```
#!/usr/bin/perl

use Bio::SeqIO;

my $seqio = new Bio::SeqIO(-format => 'fasta',
                           -file => 'filename');

while ( my $entry = $seqio->next_seq ) {
    print $entry->display_id, "\n";
    print $entry->seq, "\n";
    print ">", $entry->desc, "\n", $entry->seq, "\n";
}
}
```

- FASTA ファイルを BioPython で読み込む

```
#!/usr/bin/python

from Bio import Fasta

iter = Fasta.Iterator(open('filename') , Fasta.RecordParser())

while 1:
    entry = iter.next()
    if not(entry): break
    print entry.title
    print entry.sequence
    print entry
```

# 初めて実行するプログラムはこれでなかった？

```
#!/usr/bin/env ruby
# 由緒正しき hello world プログラム for bioinformatician
require 'bio'

gene = Bio::Seq::NA.new("catgaattattgttagannntgataaagacttgac")
prot = gene.translate    → "HELL*XW*RLD"

puts plot.split('X').join(' ').capitalize.gsub(/\*/,'o') << '!'
    → ["HELL*", "W*RLD"]
        → "HELL* W*RLD"
            → "Hell* w*rld"
                → "Hello world"
    → Hello world!
```

# Blast/Fasta/Hmmer

- BioRuby のスクリプトの中から実行する場合
  - GenomeNet にネットワーク越しに検索をかける
  - 自分のマシンでローカルに実行する
- すでにコマンドラインで実行した結果を使う場合
  - どっちでも一緒に、パースした結果が返る
- EMBOSS は BioRuby の中から実行できるが、個別のプログラムの実行結果への対応はたいへん
  - 今のところ対応済みは Jemboss, Bio-EMBOSS など？

# コード比較2 Blast 編

- BioRuby で Blast を local に実行

```
#!/usr/bin/ruby

require 'bio'

blast = Bio::Blast.local('blastp', 'hoge.pep')
flatfile = Bio::FastaFormat.open('queryfile')

flatfile.each do |seq|
  result = blast.query(seq)
  result.each do |hit|
    puts hit.query_id, hit.target_id, hit.value if hit.value < 0.05
  end
end
```

- BioPerl で Blast を local に実行 (SearchIO はしりませぬ)

```
#!/usr/bin/perl

use Bio::SeqIO;
use Bio::Tools::Run::StandAloneBlast;
use Bio::Tools::BPlite;

my @params = ('program' => 'blastp', 'database' => 'hoge.pep');
my $factory = Bio::Tools::Run::StandAloneBlast->new(@params);

my $input = Bio::SeqIO->new(-format => 'fasta', -file => "queryfile");

while ( my $seq = $input->next_seq ) {
    $result = $factory->blastall($seq);
    while ( my $hit = $result->nextSbjct ) {
        while ( my $hsp = $hit->nextHSP ) {
            print $result->query, $hit->name, $hsp->P, "\n" if $hsp->P < 0.05;
            last;
        }
    }
}
```

- BioPython で Blast を local に実行

```
#!/usr/bin/python

from Bio import Fasta
from Bio.Blast import NCBIStandalone

iterator = Fasta.Iterator(open("queryfile"), Fasta.RecordParser())
while 1:
    query = iterator.next()
    if not(query): break
    open("query.fst", "w").write(str(query))
    out, error = NCBIStandalone.blastall("blastall", "blastp", "hoge.pep", "query.fst")
    parser = NCBIStandalone.BlastParser()
    result = parser.parse(out)
    for alignment in result.alignment:
        for hsp in alignment.hsps:
            if hsp.expect < 0.05:
                print query.title, alignment.title, hsp.expect
```

# PubMed

- PubMed をプログラムからキーワード検索する

```
#!/usr/bin/env ruby
require 'bio'
entries = Bio::PubMed.search(ARGV.join(" "))
puts entries
```

```
% pmsearch.rb genome bioinformatics
→ Medline エントリのリスト
```

- PubMed ID のリストから論文用に BiBTeX を生成する

```
#!/usr/bin/env ruby
require 'bio'
ARGV.each do |pmid|
  entry = Bio::PubMed.query(pmid)
  reference = Bio::MEDLINE.new(entry).reference
  puts reference.bibtex
end
```

```
% pm2bibtex.rb 11024183 10592278
→ BiBTeX フォーマットの引用リスト
```

# ペースと正規表現

- `/bb|[^b]{2}/` 実はもっとも重要な言語?
  - to be or not to be
- ここで GenBank のクマムシの遺伝子エントリをペースしてみようと思っていたけれど、
  - GenBank パーザの話は本日何度も出てるので省略
- Location のサンプル
  - bio/location.rb ややこしい locationもあるよ
- <aside>why oh why doesn't Perl have a nice garbage collector. And when Perl 6 comes and Parrot does have one, will Perl 5 be "ported" to Parrot?</aside> said Ewan
  - Ruby を使いましょう:-)

<http://q--p.bioruby.org/>

- 次々とリリース・更新されるバイオインフォソフト
- Open Bio\* の動き
- 最近やたら多いセミナーや研究会の案内
  - とても一人では追いきれない
- Open Bio\* Info → q--p に情報を寄せください

Open Bio\* Info : q--p - Mozilla

File Edit View Go Bookmarks Tools Window Help

http://q--p.bioruby.org/

情報検索 ニュース ウオッチ メディア アーカイブ ゲノム モード incoming

クマシ ゲノム ブ... BioRuby.org Open Bio\* Info ... MEGG ana: Ana... RWikiBioRuby directory BioFetch interfa... NC\_003272 (1-1...)

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

# Open Bio\* Info

<http://q--p.bioruby.org/>

BioLinux.org バイオインフォ用 RPM パッケージ  
Posted via きぬうびい～ on Sun, 26 Jan 2003 00:28:39 +0900 from BioRuby dept.

katayama writes まだ数は少ない, BioLinux.org バイオインフォ用 RPM パッケージ  
(Read More... | 27 hits)

GO annotations for all named genes at SGD  
Posted via きぬうびい～ on Sat, 25 Jan 2003 12:16:30 +0900 from BioRuby dept.

katayama writes 入手だし, GO annotations for all named genes at SGD  
(Read More... | 7 hits)

All of these annotations have been done by SGD curators; GO terms previously assigned by automated methods (evidence code:IEA) have been replaced by referenced, manually curated annotations. ということで、せんぶ目で見てチェックしたみたいですね。

EMBO Practical Course on Analysis and Informatics of Microarray Data  
Posted via きぬうびい～ on Fri, 24 Jan 2003 16:16:10 +0900 (JST) from BioRuby dept.

bono writes not in Japan, EMBO Practical Course on Analysis and Informatics of Microarray Data  
16-22, March 2003 Wellcome Trust Genome Campus, Hinxton, Cambridge, UK  
(Read More... | 9 hits)

[Bioinformatics] G-language  
Posted via きぬうびい～ on Fri, 24 Jan 2003 11:34:40 +0900 from BioRuby dept.

Entrez PubMed

検索

1. Bioinformatics. Private pact ends the DNA data war. (science)
2. G-language Genome Analysis Environment: a workbench for nucleotide sequence data mining. (bi)
3. A task framework for the web interface W2H. (bi)
4. Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. (bi)
5. A generalized global alignment algorithm. (bi)
6. Rank order metrics for quantifying the association of sequence features with gene regulation. (bi)
7. Computational and experimental analysis reveals a novel Src family kinase in the C. elegans genome. (bi)
8. Computational analysis of core promoters in the Drosophila genome. (genome biol)
9. Heterochromatic sequences in a Drosophila whole-genome shotgun assembly. (genome biol)
10. The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. (genome biol)
11. Finishing a whole-genome shotgun: Release 3 of the Drosophila melanogaster euchromatic genome sequence. (genome biol)
12. Heart-specific genes revealed by expressed sequence tag (EST) sampling. (genome biol)
13. Identification and utilization of arbitrary correlations in models of recombination signal sequences. (genome biol)
14. Whole-genome analysis of animal A- and B-type cyclins. (genome biol)
15. The repertoire of protein kinases encoded in the draft version of the human genome: atypical variations and uncommon domain combinations. (genome biol)

bioinformatics BioRuby.org

BioBlog BioC BioC++ BioC# BioCaml BioConductor (R's ML) BioCorba BioCyc BioD BioDAS (ML) BioFortran BioGraph

Done

# q--p への投稿の仕方

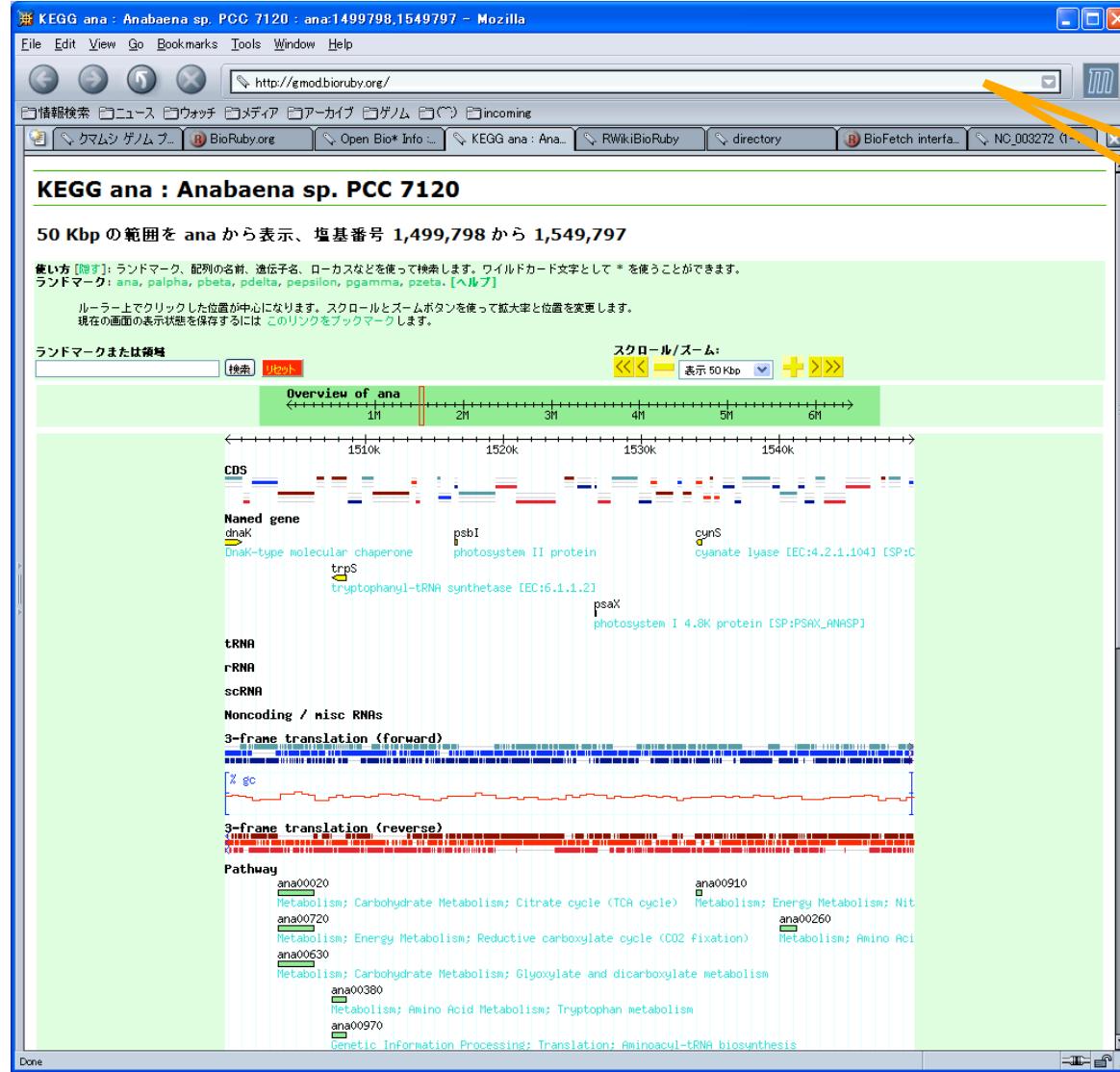
- BioRuby のメーリングリストに入る
  - ja@bioruby.org
- 宛先
  - qp@bioruby.org
- 題名
  - ひとことコメント
- 本文1行目
  - 紹介したい URL, PMID, ISBN
- 本文2行目
  - 簡単な説明
- 本文3行目以降
  - ご自由にコメントを

# q--p アンテナの仕組み

- アンテナ
  - BioRuby の PubMed 検索機能を使用
  - いくつかのジャーナルは頑張って HTML をパース
  - cron で巡回
- たれこみ
  - メールを受け取って MySQL データベースに突っ込む
  - 整形スクリプトが最近のニュースから HTML を生成

# GMOD/GBrowse w/ KEGG アノテーション

- Stein らのゲノムブラウザ、Ensemble, DAS などと統合

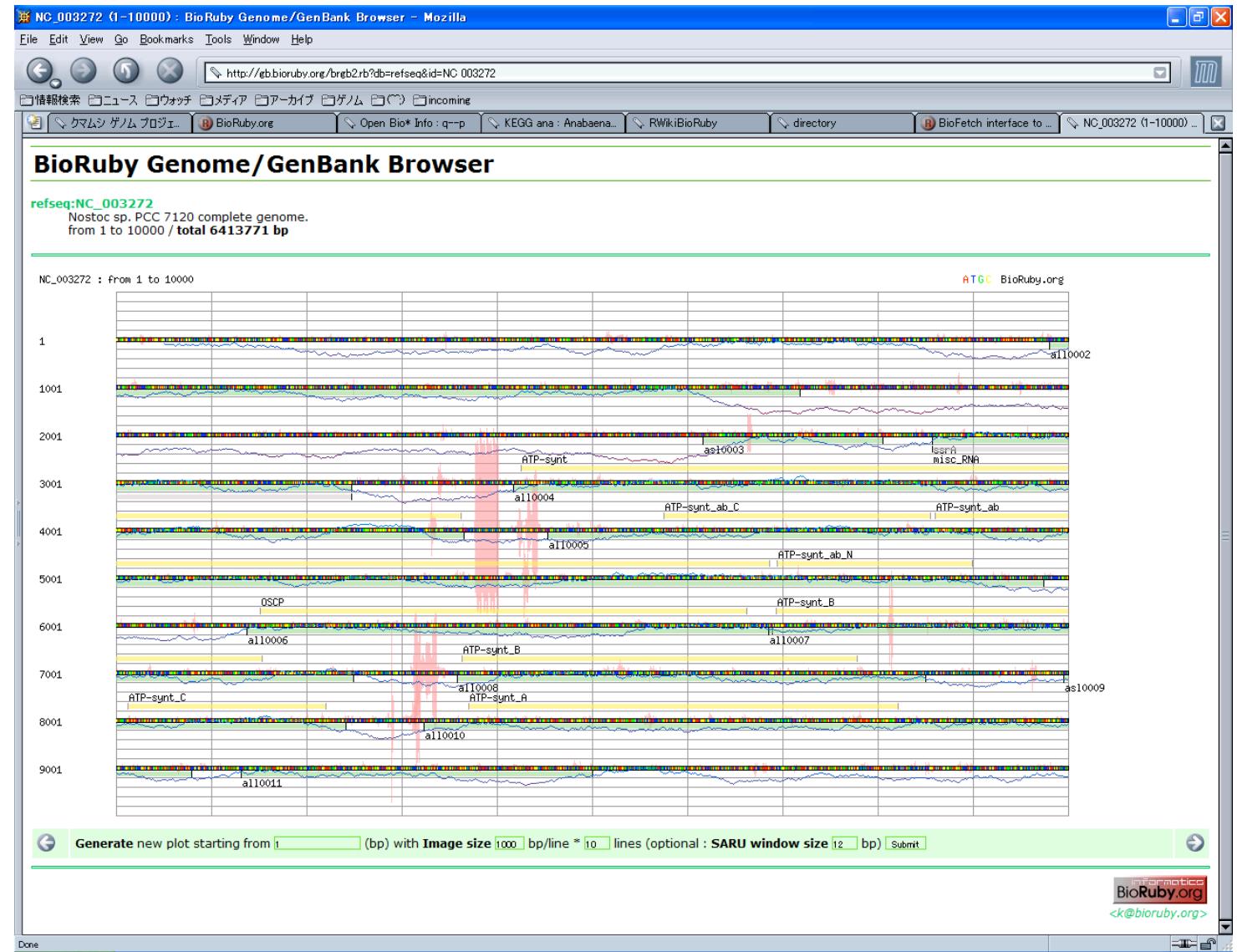


http://gmod.bioruby.org/

# BRGB

- 自家製ゲノムブラウザ。ポスターの作成に。。

- 全 GenBank, RefSeq
- MySQL バックエンド
- GD で透ける PNG
- SVG も生成可
- 全ゲノムを A0 に！



# クマムシゲノムプロジェクト

クマムシゲノムプロジェクト - Mozilla

File Edit View Go Bookmarks Tools Window Help

http://www.kumamushi.net/

情報検索 ニュース ウオッチ メディア アーカイブ ゲノム incoming

クマムシゲノムプロジェクト BioRuby.org Open Bio\* Info: q--p KEGG ana: Anabaena... RWikiBioRuby Directory BioFetch interface to ... NC\_003272 (1-10000) ...

http://kumamushi.net/

クマムシは、緩歩動物門 *Tardigrada*に属する強靭な生命力を持った奈良な生き物です。特にブルムトゲクマムシ *Echiniscus blumi*は攻殻機動隊やガンダムに出てきそうな装甲がかっこいいです。

イントロダクション：

- クマムシの紹介

文献「五つの王国」に詳しく載っているので、その中からまずクマムシの紹介をまとめてみます。ちなみに、新しいのと比べてみようと思って 1998に出た原著「Five Kingdoms」の 3rd edition を買ってみたところ、クマムシのページが 2 頁から 4 頁に倍増していました～んでもって翻訳間違ってるところ多い～。。『★』は改訂または誤訳により日本語版と原著第3版で内容が異なる所です。英語版でのみ追加されている内容はそのまま書き加えてあります。

- クマムシとは？  
五界説に従って著書「Five Kingdoms」で動物界を 37 の門に分類する Lynn Margulis らはクマムシに独立した緩歩動物門という門を与えています。我々人類から、鳥類、爬虫類、両生類、魚類、木やまでが、たった一つの脊索動物門に押し込められているというのに。昆虫やエビ・カニなどを含む節足動物門とも独立した生物なんです。数え方にもありますが、現在 92 属、750 種あまりが知られています。
- どんな形態をしているのか？  
クマムシはおそらく体腔動物ですが、体腔は生殖腺の周りに限られています。体長は 50μm～1.7mm(★)位。キチン質の(★)厚いクチクラ (cuticle) で覆われた体を持ち、4対の脚に爪または粘着性の円盤状組織(★)などを備えています。Heterotardigrade にはトゲや触角を装備していますが、Eutardigrade にはありません。色は食べ物によって異なり、赤、紫、青、黄、茶、半透明など様々です。
- 体のしくみは？  
呼吸器官、循環系ではなく、体液による拡散で酸素と養分を補給します。酸素、二酸化炭素の交換は透過性のクチクラを通じて体表から直接行います。神経系は、黒または赤色の 1 対の眼と、脳、2 本の縦走神経によって結合された 5 個(★)の腹側神経節を持っています。消化器系は、口から胃、直腸と続き、顆粒状に蓄えられた排出物は脱皮の際に一緒に捨てられます。
- エサは？  
液体や有機堆積物から摂ります。あるいは植物細胞に穴をあけて細胞液をチューチュー吸います。クルマムシ、線虫など、他の動物の体液を吸うものもいます。中には、他のクマムシを共食いするものもいるらしいです。
- 生殖は？  
トゲクマムシのように単為生殖を行うものもありますが、多くの種では雌雄両性で、圧倒的に雌が多いそうです。両性生殖の場合、腸の背側に不对の卵巣または精巢があります。受精は体内で起こる種も体外で起こる種もあります。ある淡水産のクマムシは、雄が新旧のクチクラの間に精子を注入した後、雌が脱皮し 3～35 個の卵を産み落とすことで体外受精します。陸棲のクマムシでは(★)、雄は雌の生殖口や排泄口に射精し、体内受精します。
- 発生は？

informatics  
BioRuby.org

# クマムシとは

- 緩歩動物門（節足動物じゃないよ）を独り占めする生物
- 体長1mm弱、苔や海を中心にうちの庭～南極まで世界中に分布
- 耐熱、耐冷
  - ほぼ絶対零度の  $-272^{\circ}\text{C}$ ～ $151^{\circ}\text{C}$ までOK
- 耐X線
  - 57万レントゲンに耐える（人間の致死線量は500）
- 耐圧
  - 600MPa (6000気圧) で 90% の生存率
  - バクテリアと多細胞生物は 30MPa～300MPaで死んじゃう
- 耐乾燥
  - 乾燥状態になると変形、100年後に水をかけて復活  
→tun 状態、トレハロース：極限状態を生き抜くメカニズム
  - 宇宙旅行の食料、DNA 保存用媒体として？



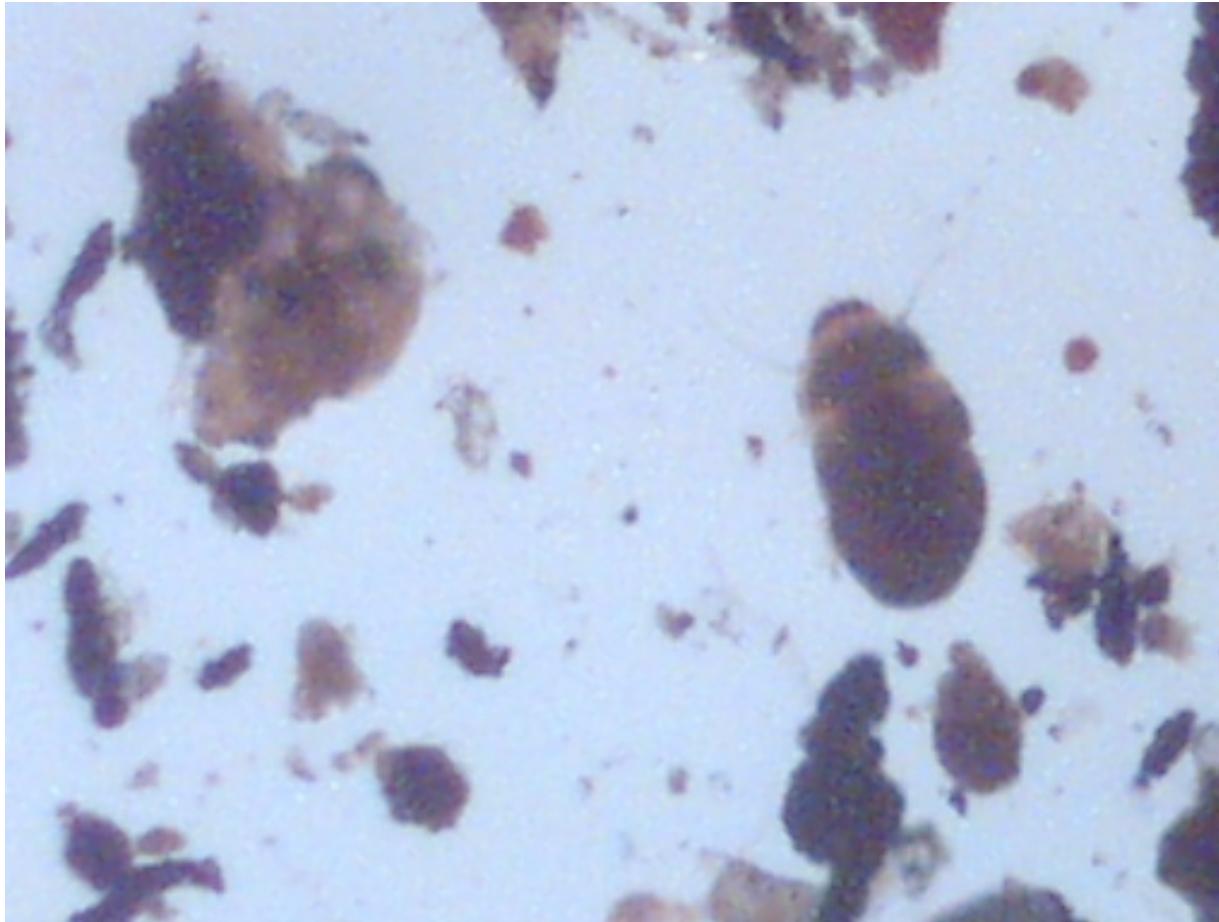
# くまむし観察装置



# オニクマムシ

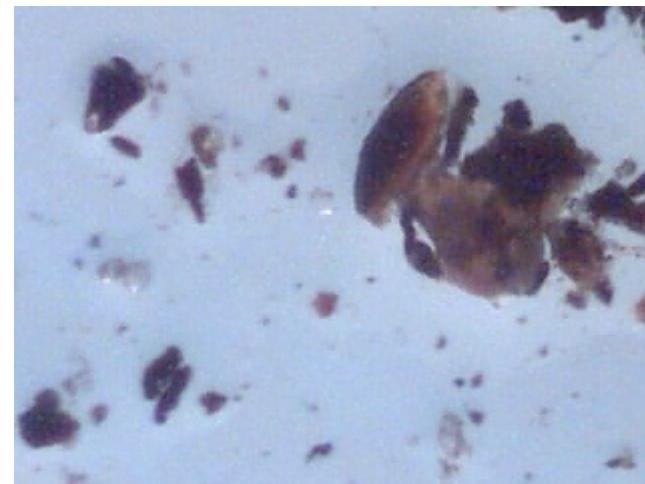
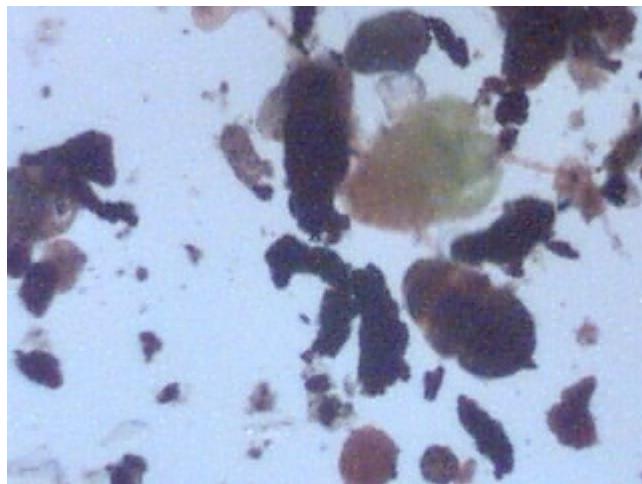


# トゲクマムシ



informatica  
**BioRuby.org**

# 棘くまむしムービー



informatica  
**BioRuby.org**

# 第9回国際クマムシ学会

Untitled Document - Mozilla

File Edit View Go Bookmarks Tools Window Help

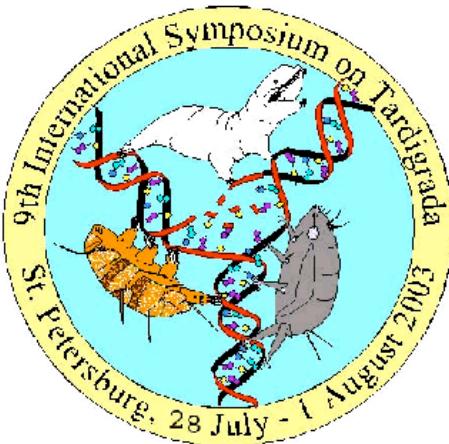
http://chuma.cas.usf.edu/%7Epbnichol/tardsym/index.html

情報検索 ニュース ウオッチ メディア アーカイブ ゲーム incoming

Garey Lab Untitled ... クマムシ ... Tardigrad... The Tardi... Das Bärti... Das Bärti... Das Bärti... Sirata Be... http://...a.txt

## 9th INTERNATIONAL SYMPOSIUM ON TARDIGRADA 28 July - 1 August, 2003 -- St. Pete Beach, Florida USA

[Home](#)  
[Location](#)  
[Room/Rates](#)  
[Directions](#)  
[Pre-Register](#)  
[Workshop](#)  
[Entertainment](#)



Dear Colleagues,

The Ninth International Symposium on Tardigrada will take place July 28 - August 1 2003 at the TradeWinds Sirata Beach Resort. This hotel is located on the beach in St. Pete Beach Florida, approximately 45 km south of the Tampa International Airport . The symposium will begin with a reception Sunday evening July 28th and will end Friday August 1st. The program details and registration information will be announced soon.

Following the Symposium there will be a three day Workshop on Molecular Phylogenetics at the University of South Florida in Tampa from Aug 4-Aug 7. The workshop will be an introduction to laboratory and computer methods of phylogenetic analyses. Space will be limited. More details will be announced soon.

Please fill in the Pre-registration form to help us estimate the number of attendees.

Best Wishes,

Jim Garey, Randy Miller, Frank Romano & Brent Nichols



# これからの BioRuby プロジェクト

- BioSQL, GFF, DAS
- GMOD/GBrowse
- KEGG API
- ドキュメント
- SOAP(DAS, XEMBL, 遺伝研), CORBA アクセスの整備
- EMBOSS, ClustalW, MAFFT など、アライメントクラス
- PATHWAY, SSDB, KO, GO, InterPro
- BioFetch で Entrez E-utils 対応
- PDB – 構造系のデータベースクラス、メンソッドの開発
- マイクロアレイ系

# BioRuby.org

- 総合情報 <http://bioruby.org/>
- 開発情報 <http://ura.bioruby.org/>
- ニュース <http://q--p.bioruby.org/>
- ソース <ftp://bioruby.org/>
- C V S <cvs.bioruby.org>
- M L [ja@bioruby.org](mailto:ja@bioruby.org), [dev@bioruby.org](mailto:dev@bioruby.org)
- 問い合せ [staff@bioruby.org](mailto:staff@bioruby.org)
  - presentation by T. Katayama <[k@bioruby.org](mailto:k@bioruby.org)>
  - クマムシについては <http://kumamushi.net/> へ