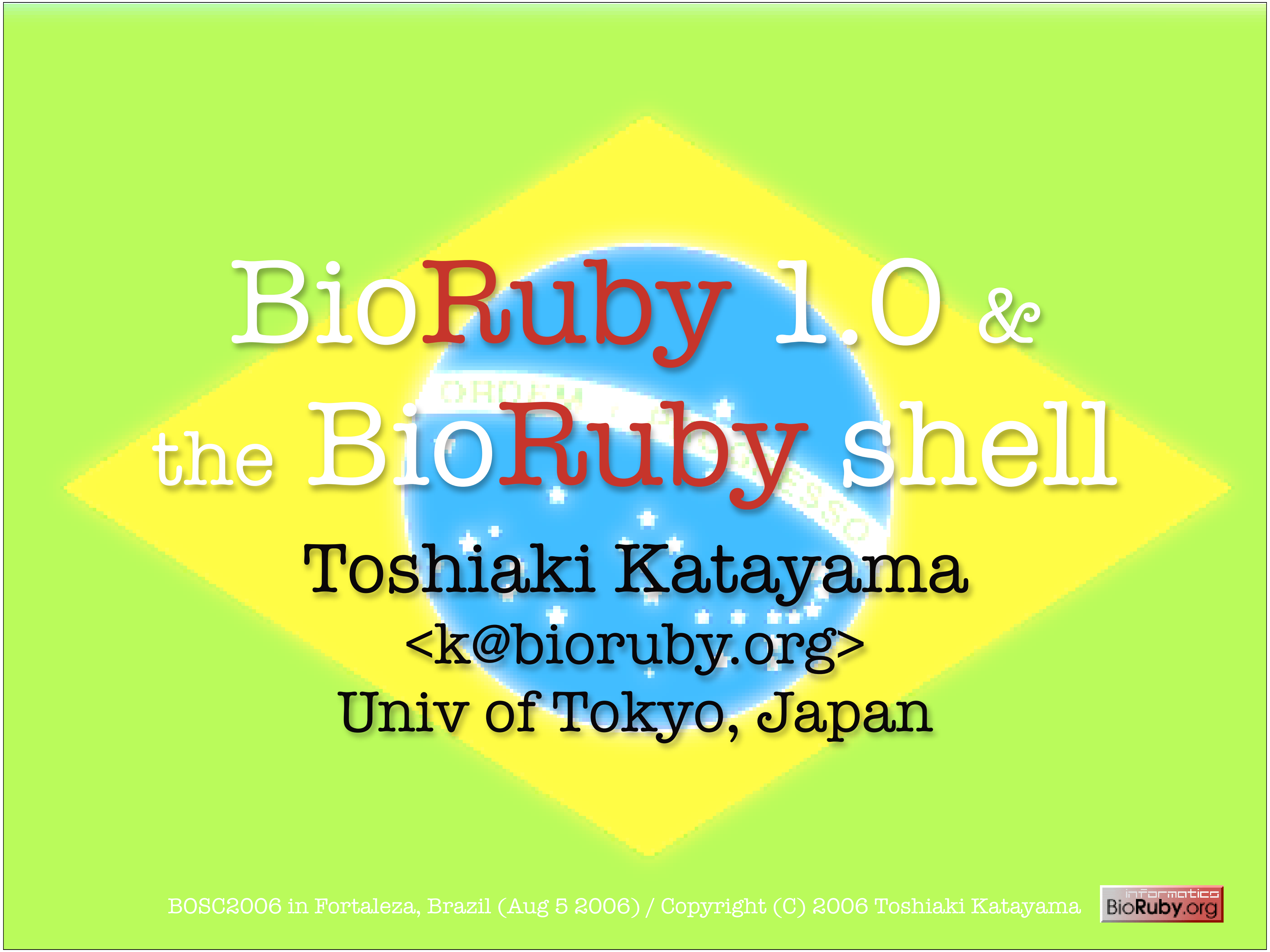
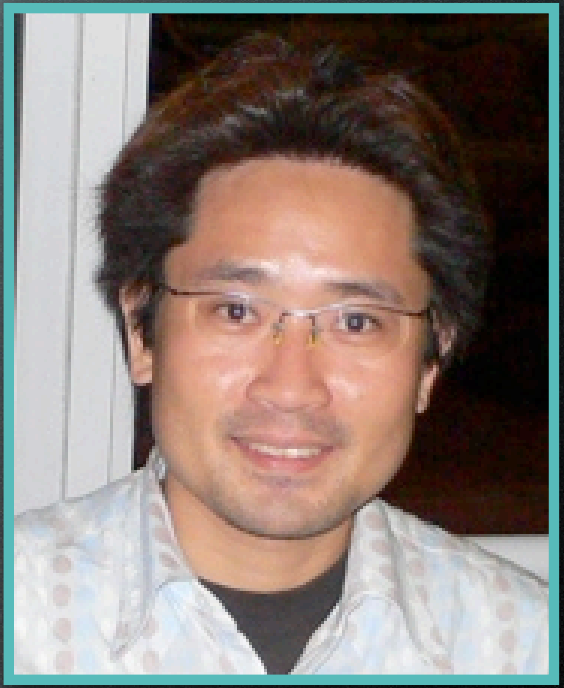


BioRuby 1.0 & the BioRuby shell

Toshiaki Katayama
<k@bioruby.org>
Univ of Tokyo, Japan

Toshiaki Katayama



k@bioruby.org

from

KEGG Kyoto Encyclopedia of Genes and Genomes

Human Genome Center
Institute of Medical Science, University of Tokyo

THE UNIVERSITY OF TOKYO

lab Tokyo branch

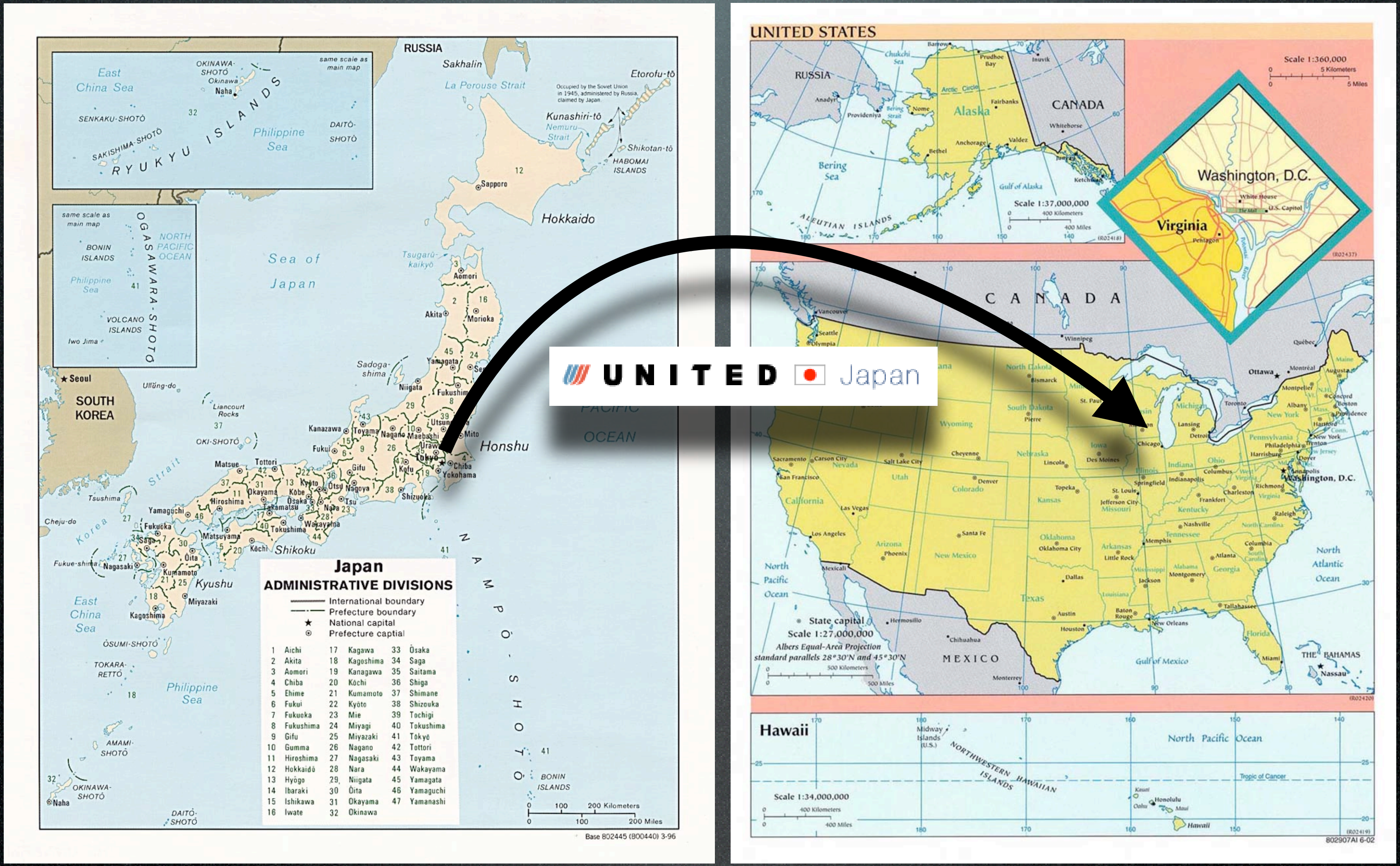
where

I'm developing ...

- KEGG API
  - SOAP/WSDL based web service
    - http://www.genome.jp/kegg/soap/
- KEGG DAS
  - GMOD/GBrowse based genome DB
    - http://das.hgc.jp/

to get here

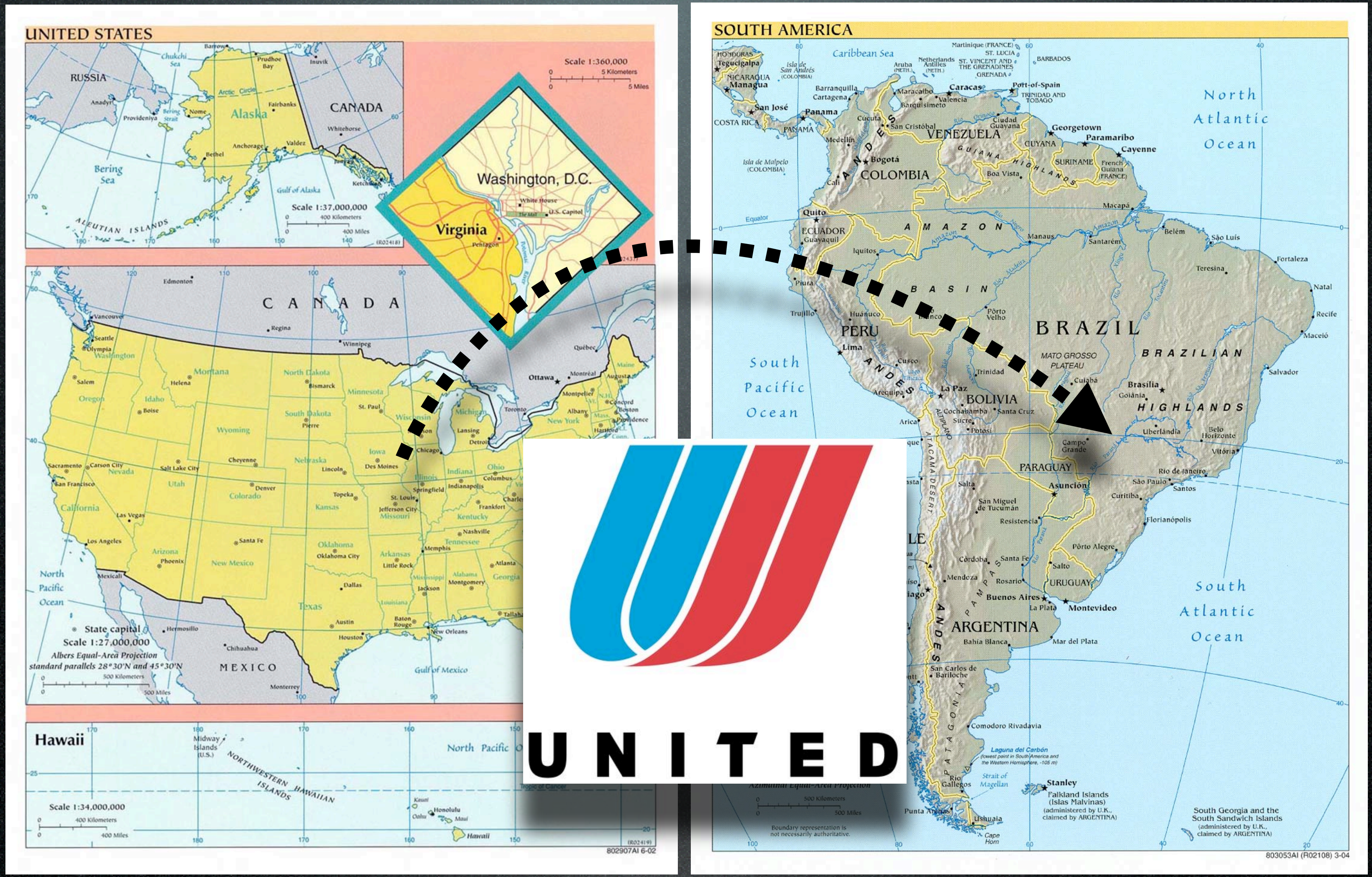
11 hours



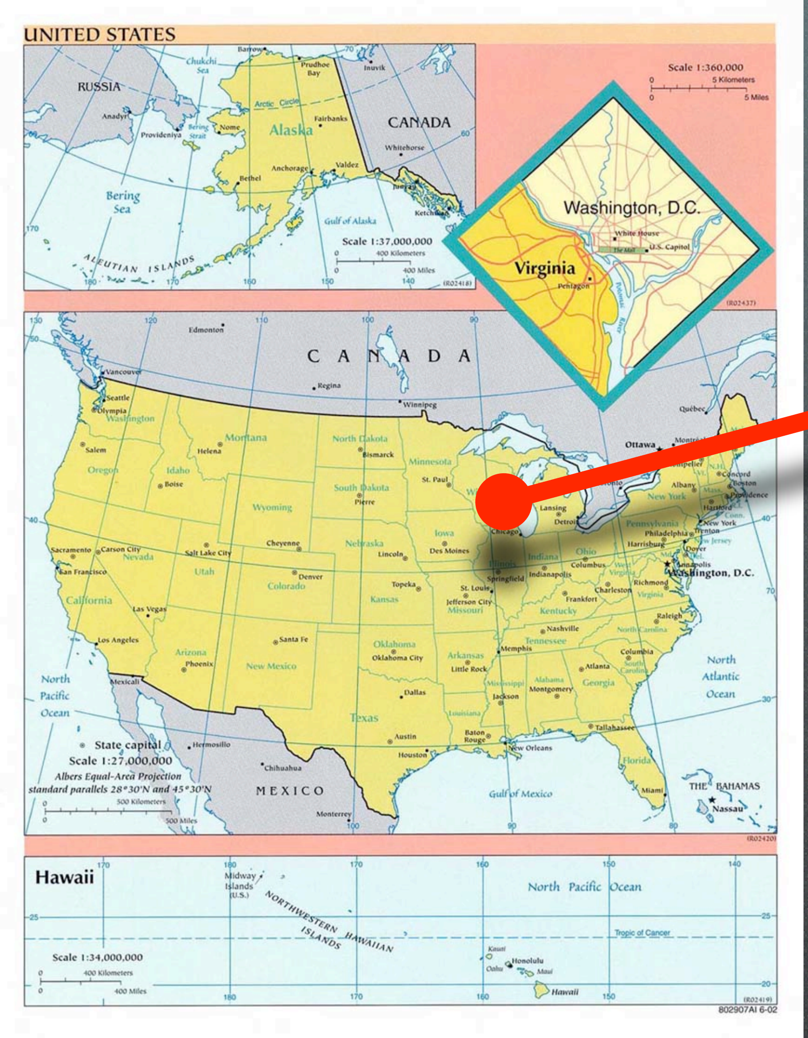
http://www.lib.utexas.edu/maps/

wait for 8 hours

40 overbooked



another 24 hours



UNITED

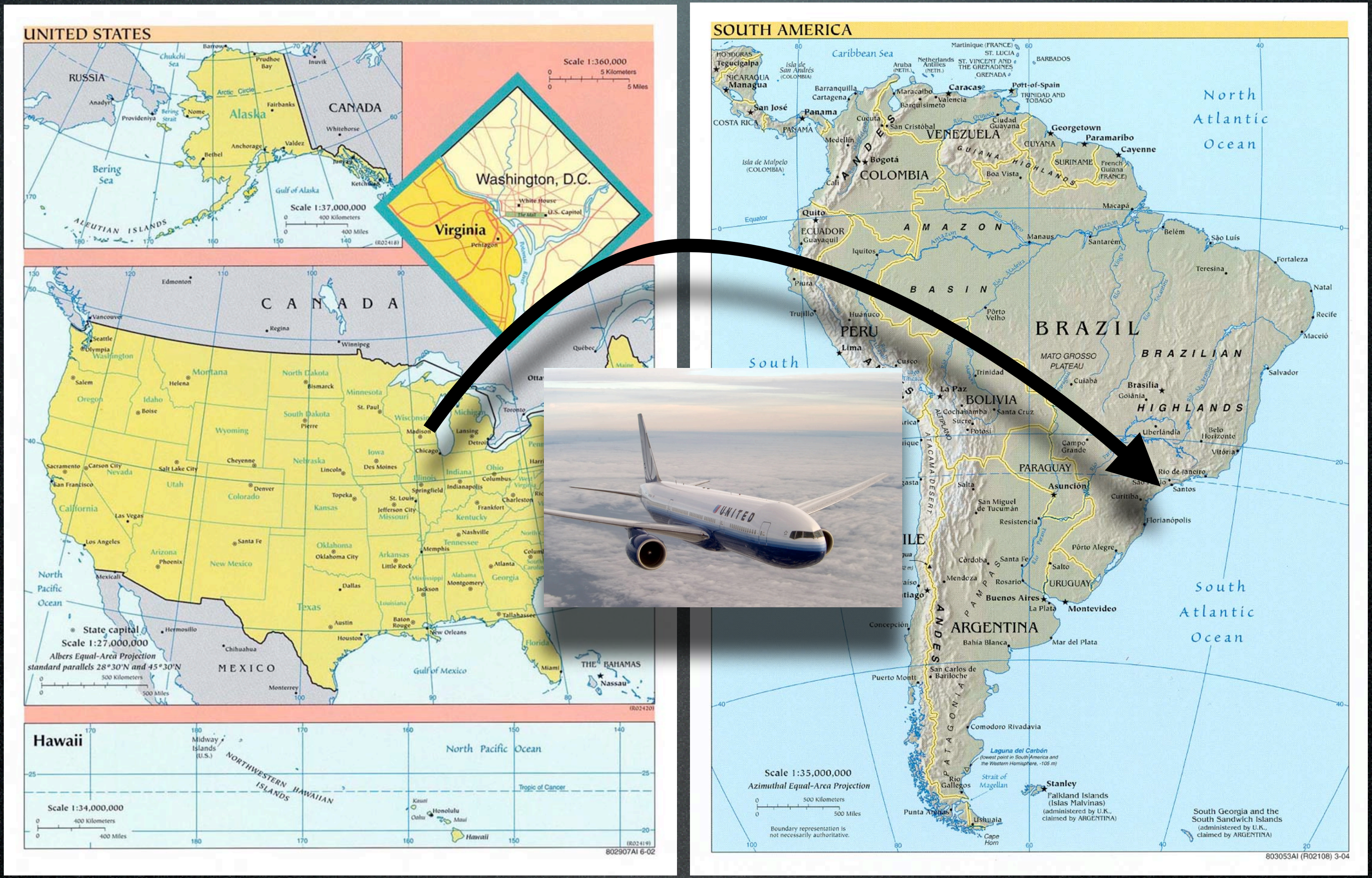
OUR UNITED COMMITMENT (SM)

29JUL06 CHICS
AGENT ID: U128547

PLEASE ACCEPT OUR APOLOGIES FOR BEING UNABLE TO ACCOMMODATE YOU ON FLIGHT 0843 TODAY. OCCASIONALLY- WE HAVE MORE CONFIRMED CUSTOMERS FOR A FLIGHT THAN THERE ARE SEATS AVAILABLE. YOU ARE ENTITLED TO COMPENSATION UNLESS YOU HAVE NOT COMPLIED WITH OUR TICKETING AND CHECK-IN POLICIES. YOU WILL BE PROVIDED A BROCHURE DETAILING OUR DENIED BOARDING COMPENSATION POLICY/F0649

KATAYAMA/TOSHIAKIMR
0843/29JUL/ORDGRU
NK2KFM

10 hours


UNITED STATES
Washington, D.C.
Virginia
SOUTH AMERICA
BRAZIL
Rio de Janeiro
UNITED

lost baggage

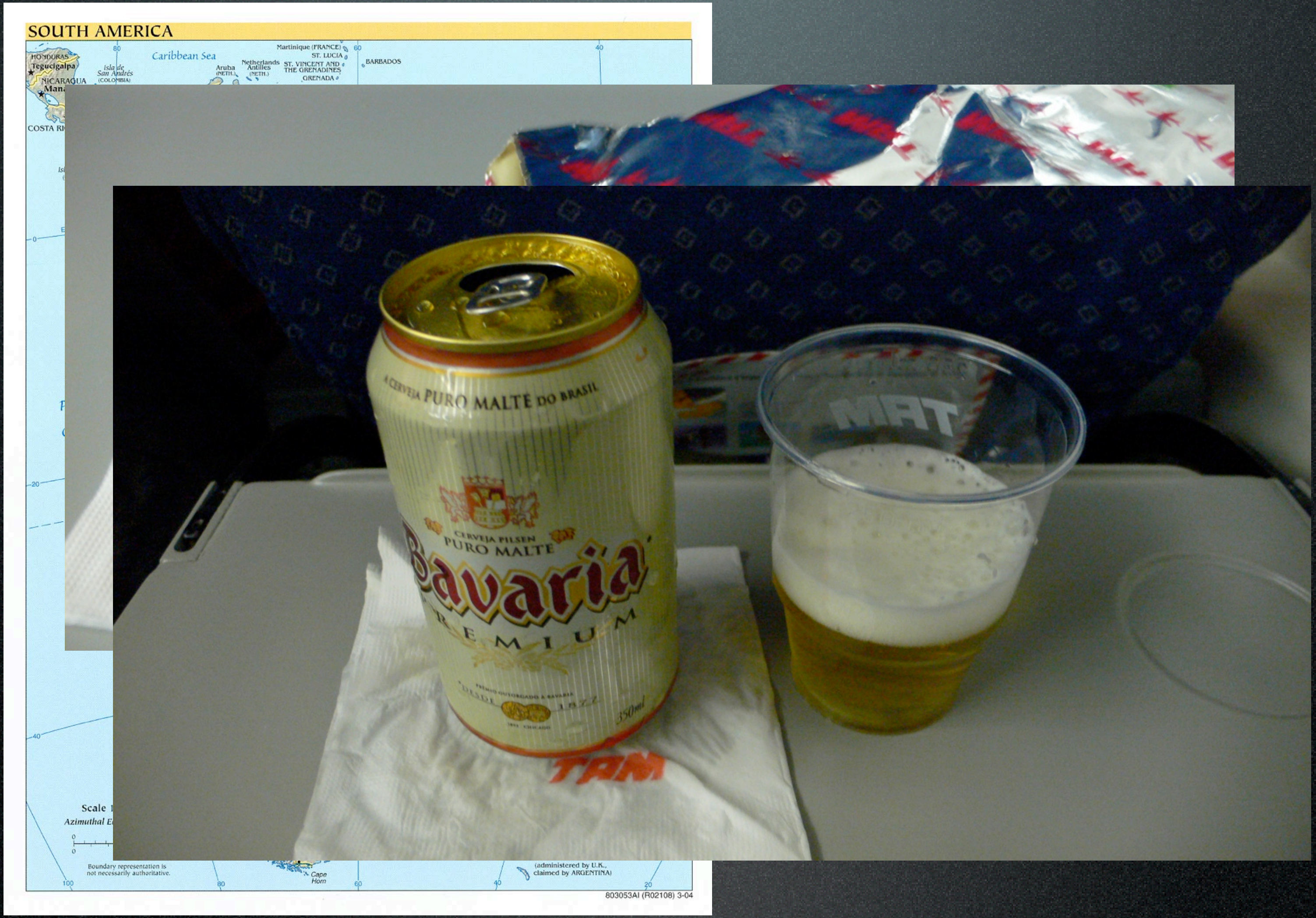
cancelled

no refund...

wait 7 hours


SOUTH AMERICA
Caribbean Sea
North Atlantic Ocean
South Pacific Ocean
South Atlantic Ocean
VENEZUELA
COLOMBIA
ECUADOR
PERU
BRAZIL
BOLIVIA
PARAGUAY
CHILE
ARGENTINA
URUGUAY
GUYANA
SURINAME
French Guiana (FRANCE)
AMAZON BASIN
BRAZILIAN HIGHLANDS
ANDES
MATO GROSSO PLATEAU
Caracas
Bogotá
Quito
Lima
La Paz
Brasília
Asunción
Santiago
Buenos Aires
Montevideo
Georgetown
Paramaribo
Cayenne
São Paulo
Rio de Janeiro
Santos
Curitiba
Porto Alegre
Manaus
Belém
Salvador
Recife
Fortaleza
Equator
Tropic of Capricorn
Falkland Islands (Islas Malvinas) (administered by U.K., claimed by ARGENTINA)
South Georgia and the South Sandwich Islands (administered by U.K., claimed by ARGENTINA)
Scale 1:35,000,000
Azimuthal Equal-Area Projection
500 Kilometers
500 Miles
Boundary representation is not necessarily authoritative.
803053AI (R02108) 3-04

6 hours w/ 2 stops



75 hours to get

to this paradise!

and took
138 hours
(6days)
for my
luggage...

BioRuby

facts

- What: BioPerl for Ruby language
- When: started in Nov 2000
- Who: about 10 developers

  (3 Japanese core developers + others)
- Where: originally in KEGG lab

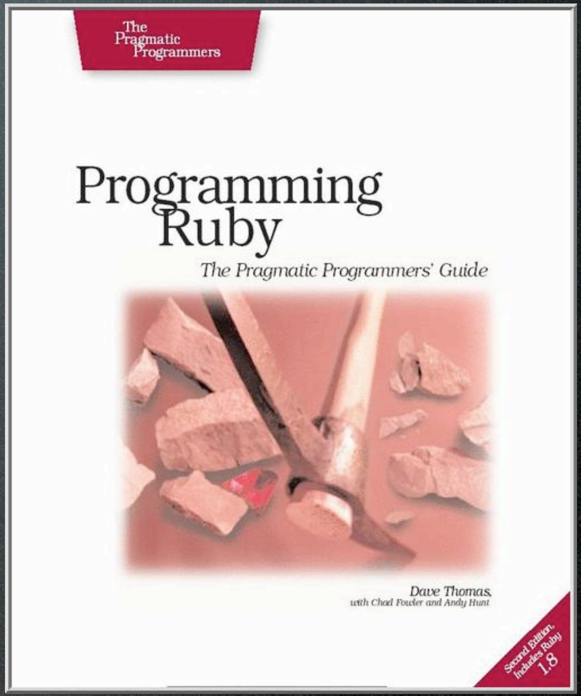
  (now hosted at open-bio.org)

Why Ruby?

(Do you know Ruby?)

Ruby is ...

- object oriented
- scripting language
- born in Japan
- 10 years old



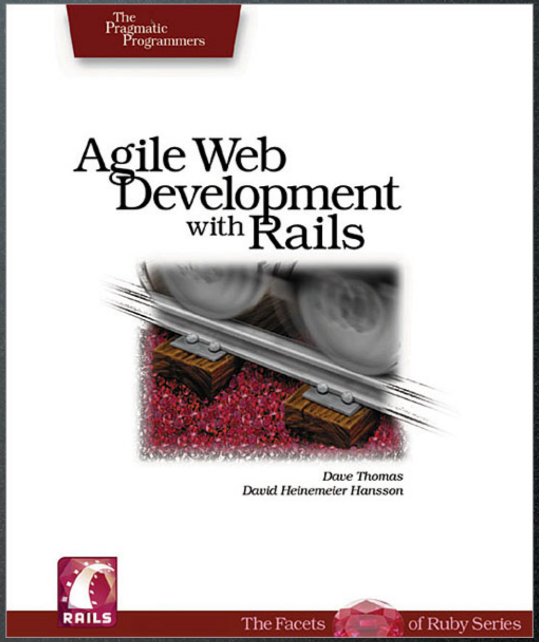
Ruby in five E's

- Everything is an object
- Elegant blocks of inline joy
- Exploring with reflection
- Extending at runtime
- Extensive standard library

cited from http://media.rubyonrails.org/presentations/fisl6.pdf

but now known as ...

- the language needed to run Rails



The Pragmatic Programmers

Agile Web Development with Rails

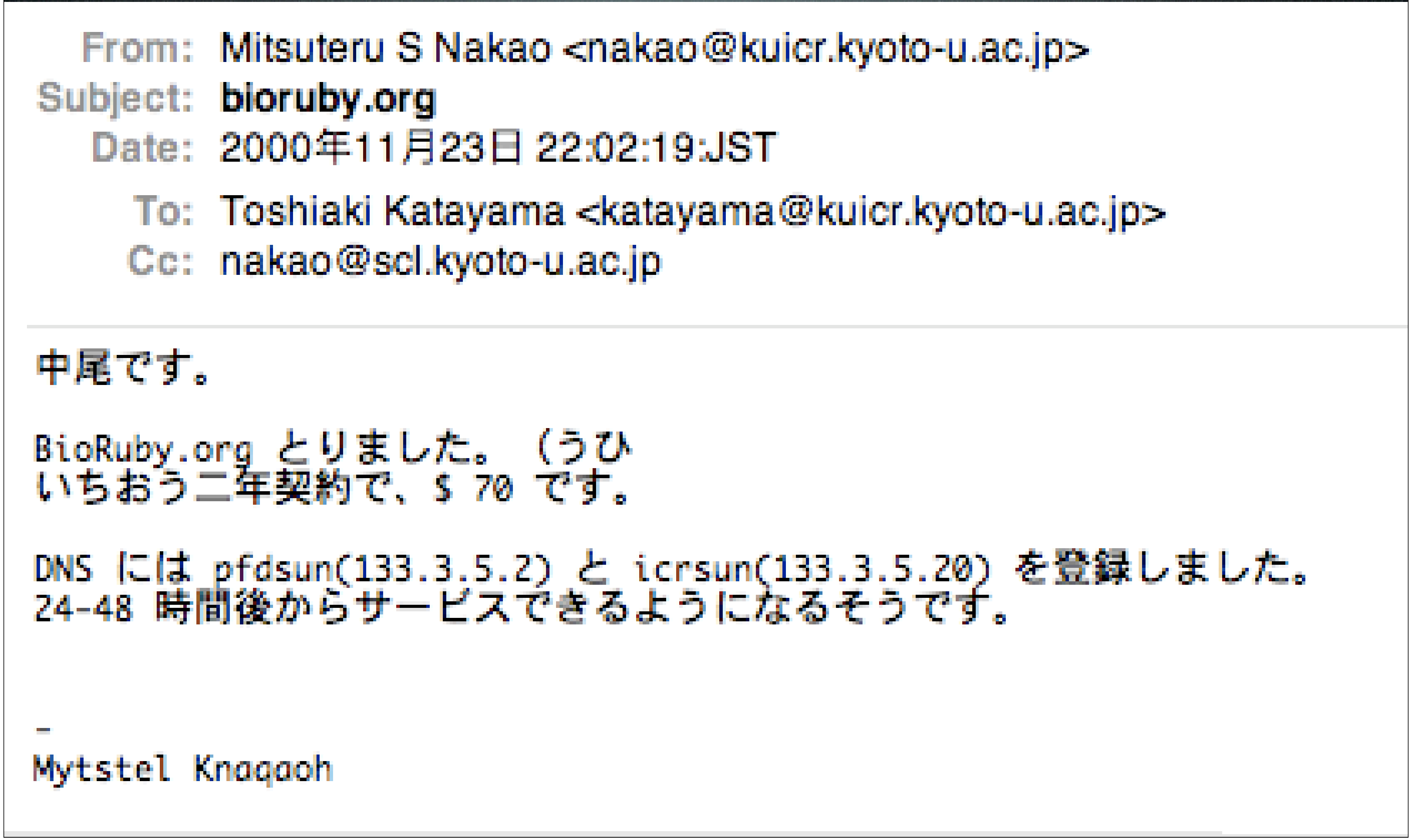
Dave Thomas
David Heinemeier Hansson

The Facets of Ruby Series

Ruby on Rails - http://rubyonrails.org/

history

2000



From: Mitsuteru S Nakao <nakao@kuicr.kyoto-u.ac.jp>
Subject: **bioruby.org**
Date: 2000年11月23日 22:02:19:JST
To: Toshiaki Katayama <katayama@kuicr.kyoto-u.ac.jp>
Cc: nakao@scl.kyoto-u.ac.jp

```
中尾です。

BioRuby.org とりました。（うひ
いちおう二年契約で、＄ 70 です。

DNS には pfdsun(133.3.5.2) と icrsun(133.3.5.20) を登録しました。
24-48 時間後からサービスできるようになるそうです。



-
Mytstel Knaqaoh
```

2000

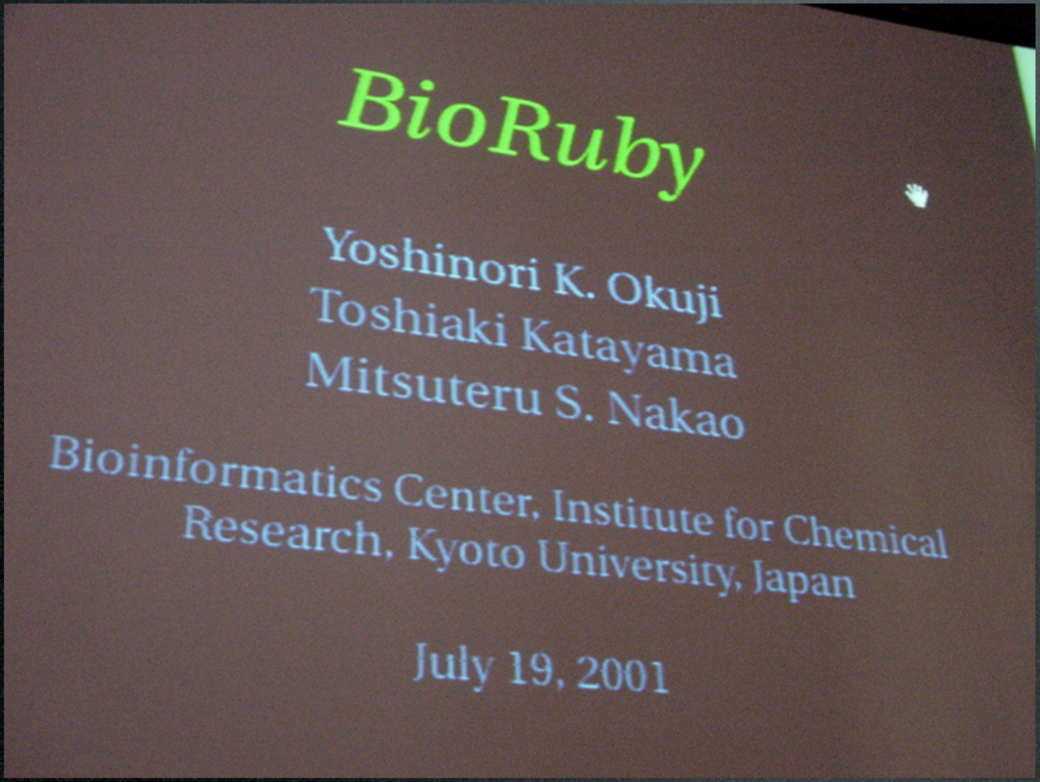
- BioRuby project was started
  - some in-house Perl modules were ported to Ruby
- BioPerl, BioPython, BioJava were already existed

2001

- Jun 21: bioruby-0.2.tar.gz
  - Sequence model
  - GenBank, PROSITE, KEGG parsers
  - DBGET (entry retrieval system)
  - FASTA (summarize results)
- developed a genome browser

2001

- BOSC in Copenhagen



BioRuby

Yoshinori K. Okuji
Toshiaki Katayama
Mitsuteru S. Nakao

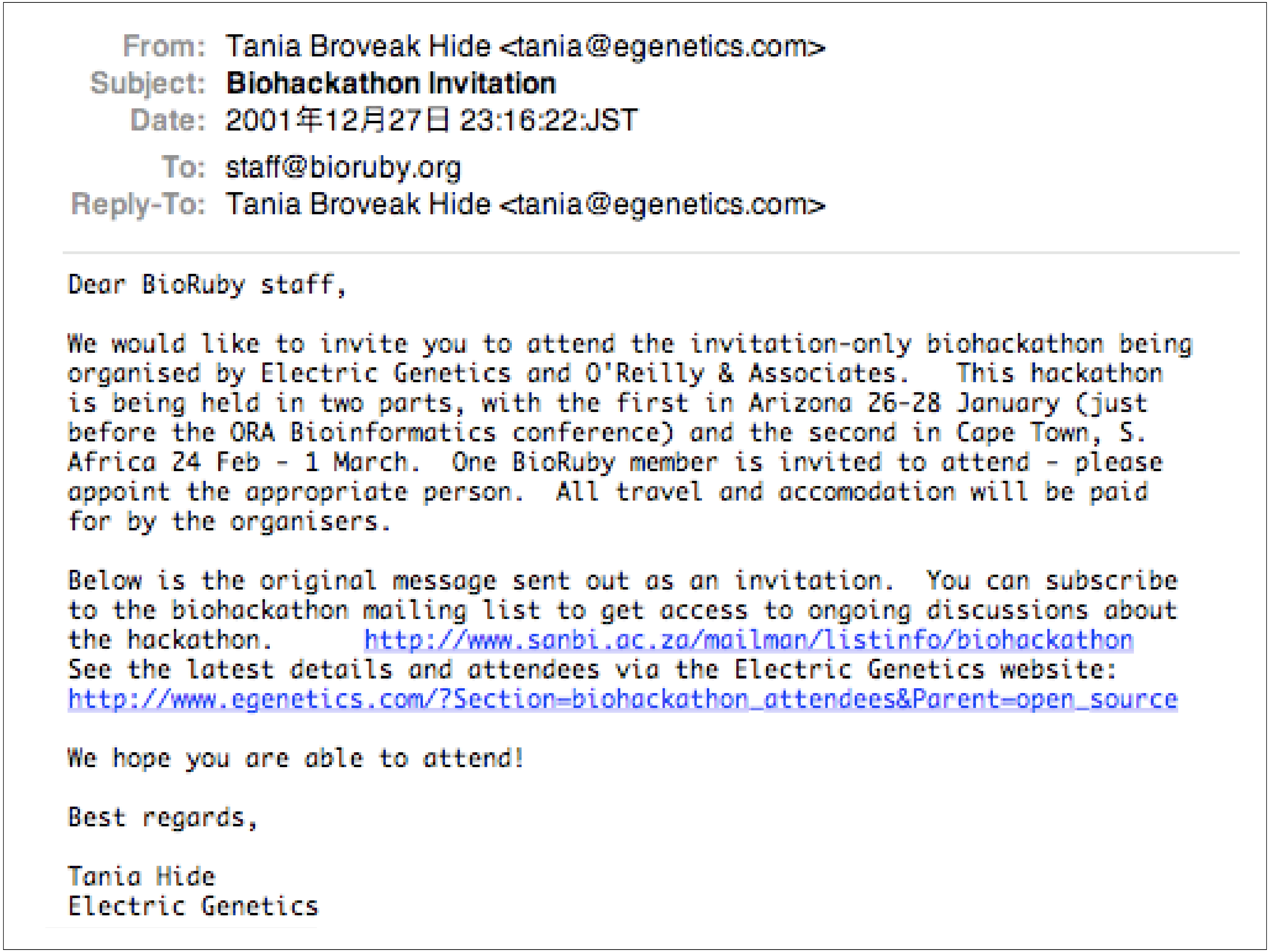
Bioinformatics Center, Institute for Chemical Research, Kyoto University, Japan

July 19, 2001

2001

- visited Ewan@EBI

Bio**Ruby** what?!%$@



From: Tania Broveak Hide <tania@egenetics.com>
Subject: **Biohackathon Invitation**
Date: 2001年12月27日 23:16:22:JST
To: staff@bioruby.org
Reply-To: Tania Broveak Hide <tania@egenetics.com>

Dear BioRuby staff,

We would like to invite you to attend the invitation-only biohackathon being organised by Electric Genetics and O'Reilly & Associates. This hackathon is being held in two parts, with the first in Arizona 26-28 January (just before the ORA Bioinformatics conference) and the second in Cape Town, S. Africa 24 Feb - 1 March. One BioRuby member is invited to attend - please appoint the appropriate person. All travel and accomodation will be paid for by the organisers.

Below is the original message sent out as an invitation. You can subscribe to the biohackathon mailing list to get access to ongoing discussions about the hackathon. http://www.sanbi.ac.za/mailman/listinfo/biohackathon
See the latest details and attendees via the Electric Genetics website:
http://www.egenetics.com/?Section=biohackathon_attendees&Parent=open_source

We hope you are able to attend!

Best regards,

Tania Hide
Electric Genetics

2002

- BioHackathon #1, #2
    - Tucson - O'Reilly
    - South Africa - Electric Genetics

BIOHACKATHON 26-28 JAN 2002, TUCSON, ARIZONA, USA

SAVE

Master Module List

	Spec	Java	Perl	Python	Ruby	StackDB/C	
Registry	✓	Brian 1/2	?? ✓ 7/4	Jeff 2/1			
Index-BDB	✓		Lincoln	Andrew			
Index - Flat	✓	Matt	Michele ✓	Andrew		✓	
BioFetch	✓	Dyfed	Kinakar Heikke	Andrew	server ✓	Antoine	?
CORBA	✓	cli ✓ serv ✓	cli Ewan Server ✓	cli ✓ Serv ✓			
BoSQL	- Chris - ✓	Read ✓ (meta) Write ✓	Read ✓ Bulk Mark Write ✓	Read ✓ Write	X		
Xembl	✓	Cgi Soap	cgi ✗? SOAP ✓	cgi ✓ SOAP X			
UDDI		Matt Brian					

Spec Slave →

Genbank daily

REGISTRY

mysql -u root

192.168.0.95

http://industry.ebi.ac.u

int → filename, ss

key → (fil

Secondary → key

P1234 10

P1234 10

37 minutes → DB

14 + 1:24 minutes → Flat


BIOHAC THON

OBDA

- Open Bio* Database Access
  - BioRegistry
  - BioFlat
  - BioFetch
  - BioSQL

2003

- BioHackathon #3
  - Singapore - Apple

GBrowse (Bio::Graphics), Chado, BioMOBY, ...

- BOSC2003
  - Australia - BioRuby w/ KEGG API
- BioRuby 0.5
  - applications, web services

2004

- BOSC2004
  - Glasgow
    - BioRuby w/ pathway + genome
- BioRuby 0.6
  - fairly stable

2005

- Funded by IPA

IPA®
独立行政法人 情報処理推進機構
INFORMATION-TECHNOLOGY PROMOTION AGENCY, JAPAN

- ChemRuby launched

- pandemic of Rails

Last updated: Oct 31, 2005 6:03 AM PT

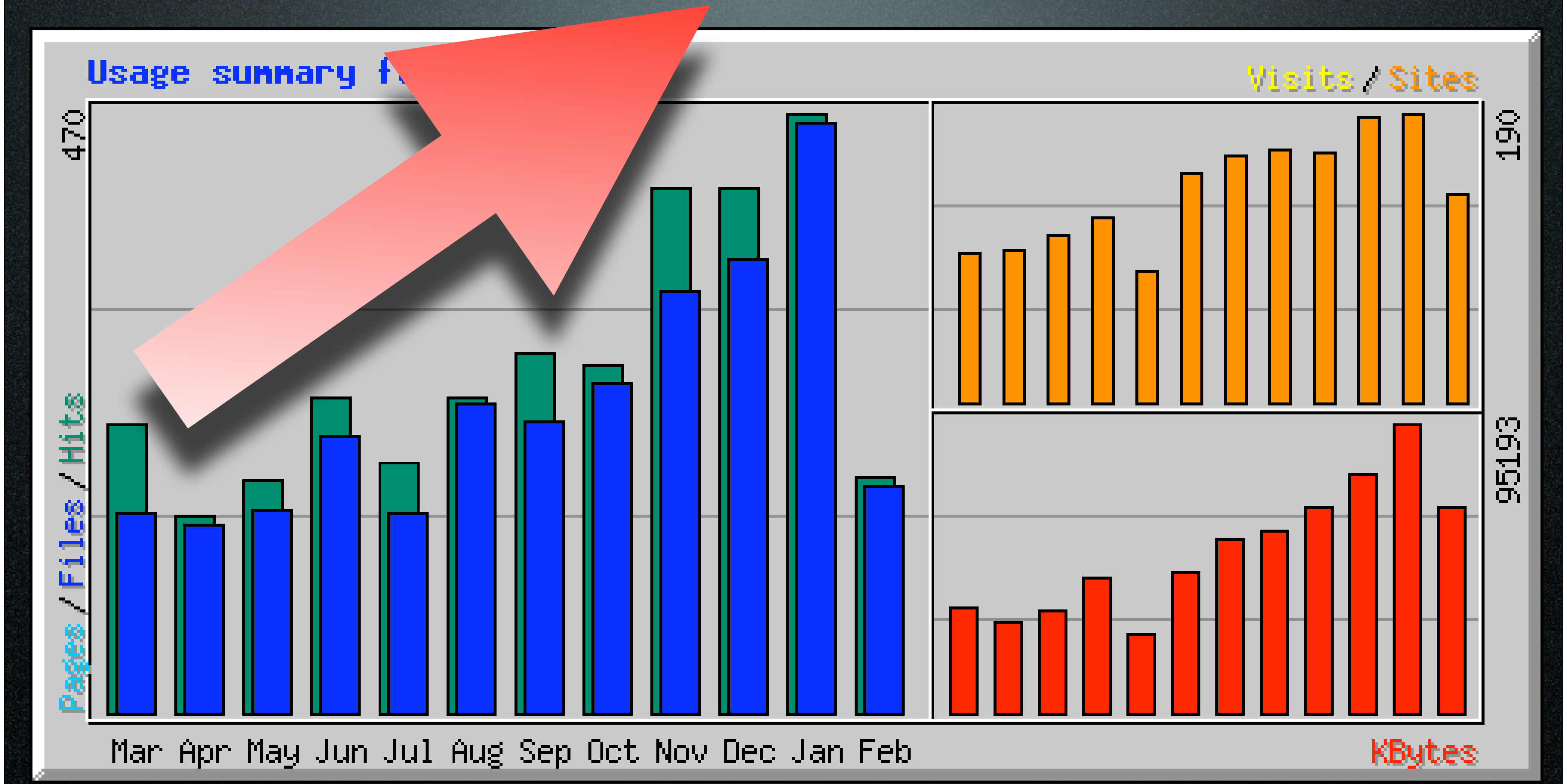
Is the Web ready for prime time?
The future of TV is coming into view as content producers dabble in Internet distribution.
3 hours, 8 minutes ago

Bye complexity, hello Ruby
A solo project called Ruby on Rails has become popular among developers because of its simplicity.
3 hours, 8 minutes ago

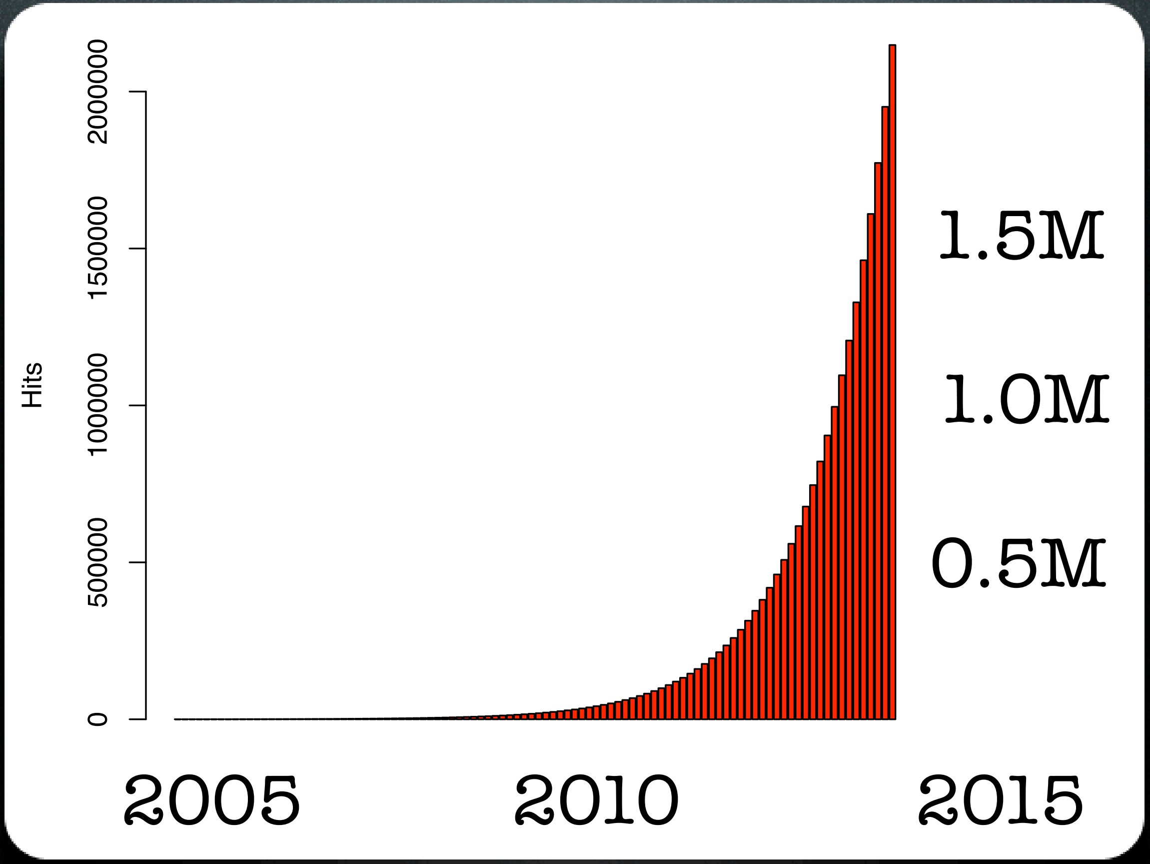
Goo into
Incre Micro hire p open-
3 hour

2006

- BioRuby 1.0.0 is out!


Usage summary f
Visits / Sites
470
Pages / Files / Hits
190
95193
Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb
KBytes

DLs 05-06



1.5M DLs in 2015

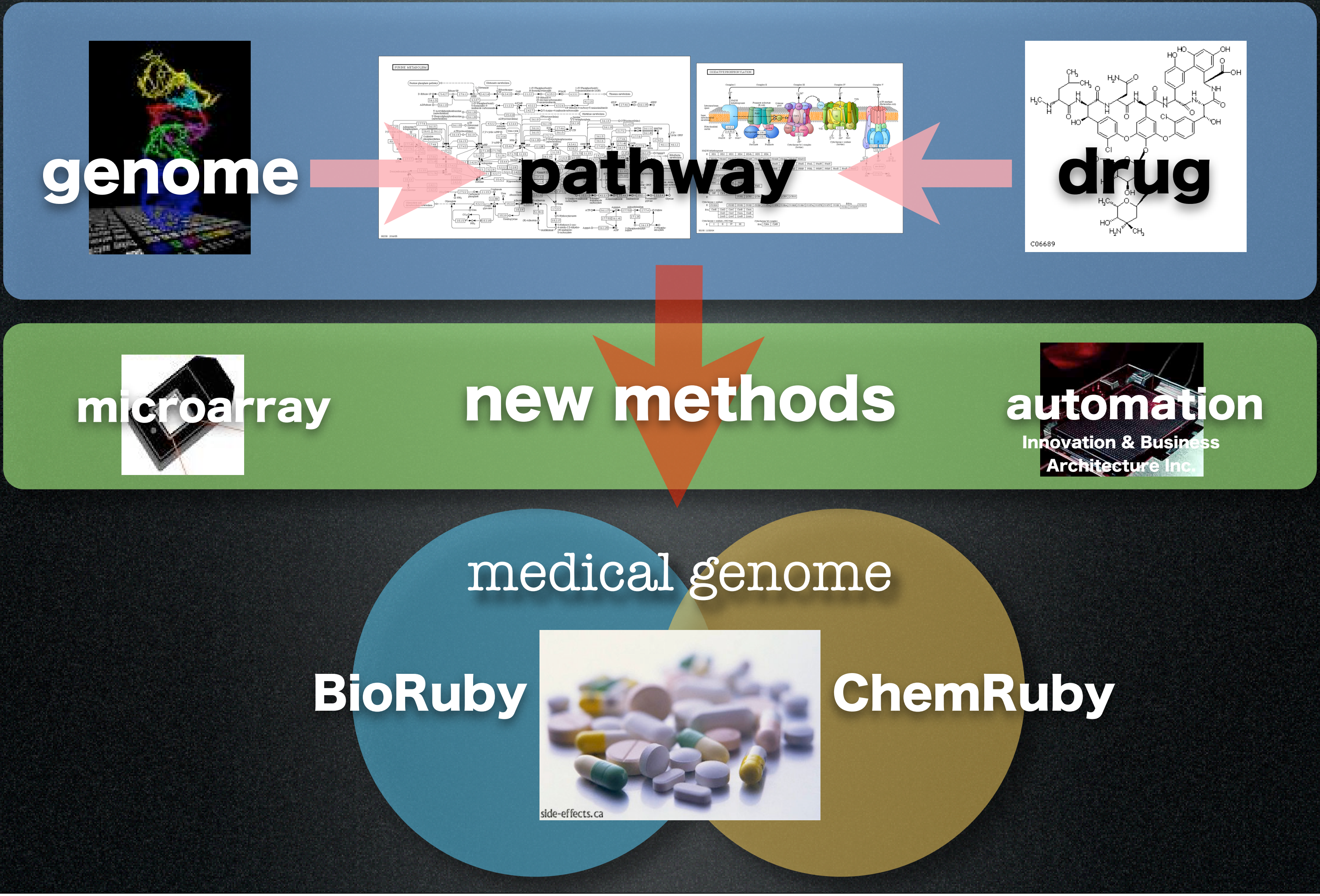
anyway

BioRuby 1.0

- Released: Feb 2006
- New features:
    - interactive shell
    - unit tests
    - documents
    - bug fixes (some bugs are not yet fixed ;-)

purpose

bioinfo + cheminfo



genome → pathway ← drug

microarray

new methods

automation

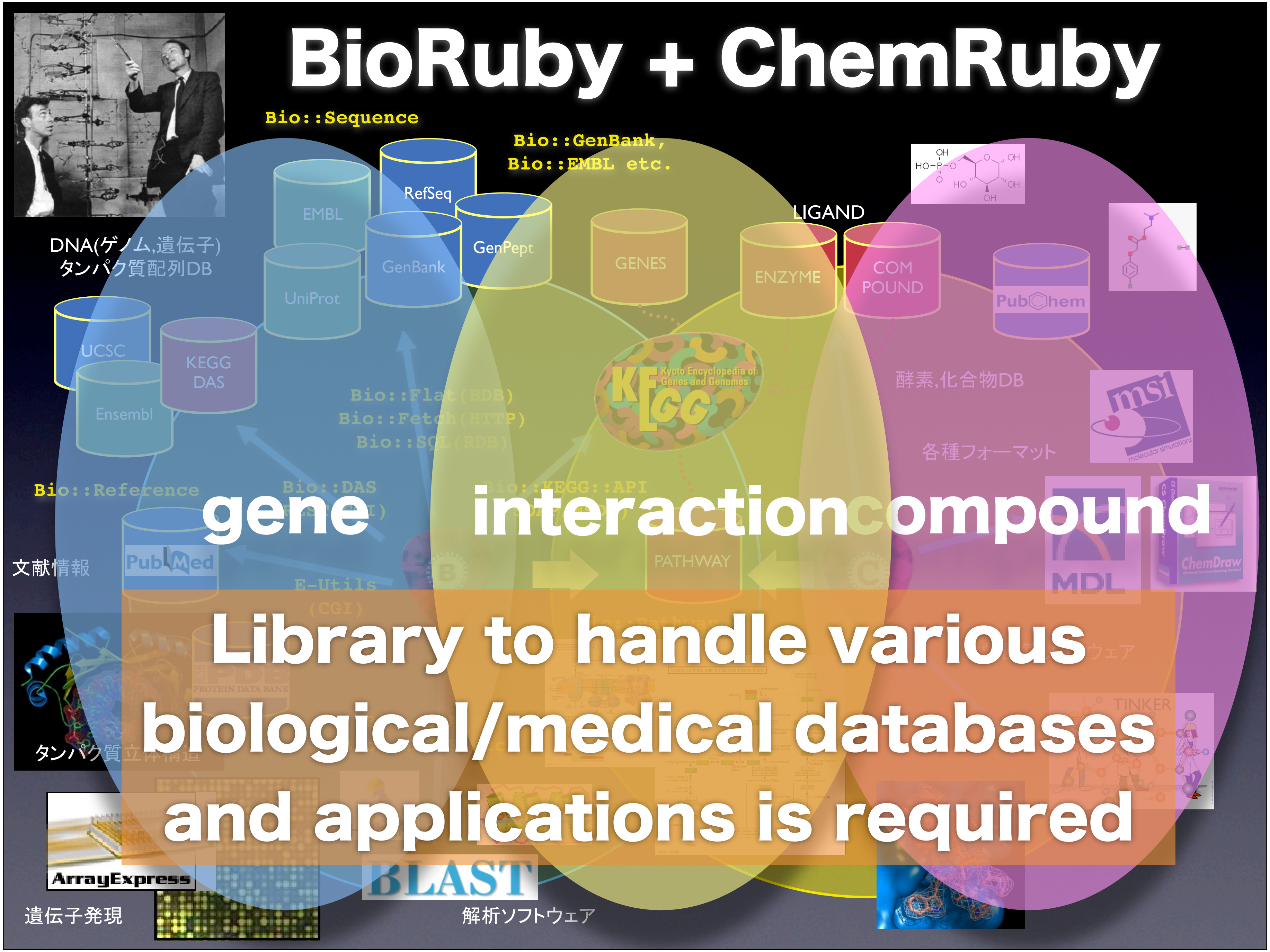
Innovation & Business Architecture Inc.

medical genome

BioRuby

ChemRuby

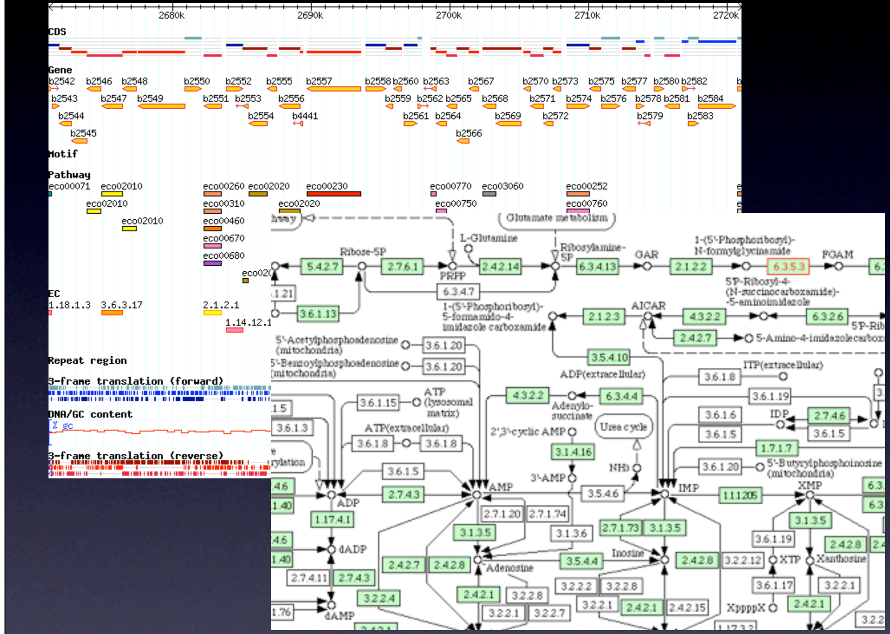
BioRuby + ChemRuby



gene **interaction** **compound**

Library to handle various biological/medical databases and applications is required

biological data formats



Sequences (GenBank, UniProt,......)

```
MMEILRGSPALSAFRINKLLARFQAARLPVHNIYAEYVHFADLNAPLNDDEHAQLERLLK
YGPALASHAPQGKLLLVTPRPGTISPWSSKATDIAHNCGLQQVNRLERGVAYYIEAGTLT
NEQWQQVTAELHDRMMETVFFALDDAEQLFAHHQPTPVTSVDLLGQGRQALIDANLRLGL
ALAEDEIDYLQDAFTKLGRNPNDIELYMFAQANSEHCRHKIFNADWVIDGEQQPKSLFKM
IKNTFETTPDHVLSAYKDNAAVMEGSEVGRYFADHETGRYDFHQEPAHILMKVETHNHPT
AISPWPGAATGSGGEIRDEGATGRGAKPKAGLVGFSVSNLRIPGFEQPWEEDFGKPERIV
TALDIMTEGPLGGAAFNNEFGRPALNGYFRTYEEKVNSHNGEELRGYHK
```

Structures (PDB,......)

```
HEADER    LIGASE                                  15-DEC-04   1VQ3
TITLE     CRYSTAL STRUCTURE OF PHOSPHORIBOSYLFORMYLGLYCINAMIDINE
TITLE    2 SYNTHASE, PURS SUBUNIT (EC 6.3.5.3) (TM1244) FROM
TITLE    3 THERMOTOGA MARITIMA AT 1.90 A RESOLUTION
 :
ATOM      1  N   HIS A  -3      70.636  28.979   8.146  1.00 32.07
ATOM      2  CA  HIS A  -3      70.436  29.954   7.050  1.00 28.18
ATOM      3  C   HIS A  -3      69.098  30.657   7.163  1.00 26.01
ATOM      4  O   HIS A  -3      68.684  31.308   6.214  1.00 29.86
ATOM      5  CB  HIS A  -3      70.492  29.234   5.702  1.00 33.52
```

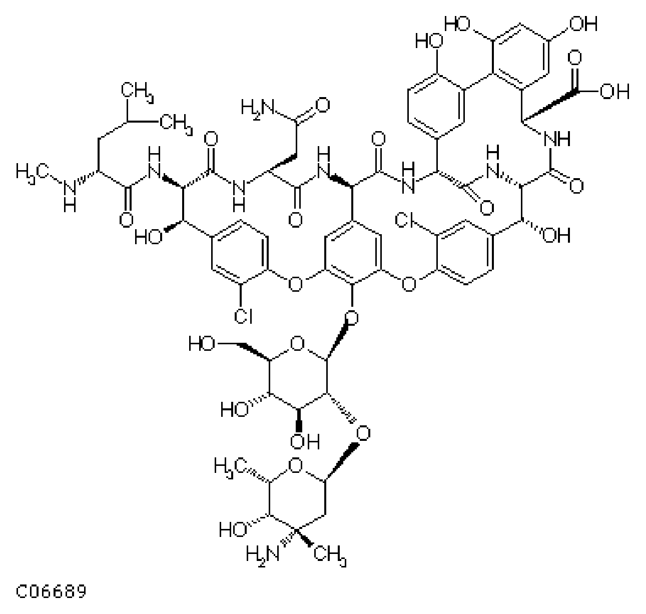
XML (DAS, KGML,......)

```
<?xml version="1.0" standalone="yes"?>
<!DOCTYPE DASGFF SYSTEM "http://www.biodas.org/dtd/dasgff.dtd">
<DASGFF>
<GFF version="1.01" href="http://das.hgc.jp/cgi-bin/das/eco/
features?segment=eco%3A2671072%2C2721071">
<SEGMENT id="eco" start="2671072" stop="2721071" version="1.0">
   <FEATURE id="EC:1.14.12.17/7199" label="1.14.12.17">
      <TYPE id="enzyme:KEGG" category="enzyme">enzyme:KEGG</TYPE>
      <METHOD id="enzyme">enzyme</METHOD>
      <START>2683857</START>
      <END>2685047</END>
```

Expressions (GEO,.......)

```
Affymetrix:CompositeSequence:HG_U95Av2:AFFX-HUMGAPDH/M33197_5_at        AFF
HUMGAPDH/M33197_5_at       IPB000173       1.4.1.16        M33197
ENSG00000111640 7
Affymetrix:CompositeSequence:HG_U95Av2:AFFX-HUMGAPDH/M33197_M_at        AFF
HUMGAPDH/M33197_M_at       IPB000173       1.4.1.16        M33197
ENSG00000111640 7
Affymetrix:CompositeSequence:HG_U95Av2:AFFX-HUMGAPDH/M33197_3_at        AFF
HUMGAPDH/M33197_3_at       IPB000173       1.4.1.16        M33197
ENSG00000111640 7
```

chemical data formats



XML (CML, CDXML,......)

```
<?xml version="1.0" encoding="UTF-8" ?>
<!DOCTYPE CDXML SYSTEM "http://www.cambridgesoft.com/
xml/cdxml.dtd" >
<CDXML
 CreationProgram="ChemDraw 7.0"
 Name="hypericin.mol"
 BoundingBox="67.44 97.35 301.63 315.25"
```

Linear (SMILES, InChI,......)

```
CC1C(C(CC(O1)OC2C(C(C(OC2OC3=C4C=C5C=C3OC6=C(C=C(C=C6)C(C(C
(=O)NC(C(=O)NC5C(=O)NC7C8=CC(=C(C=C8)O)C9=C(C=C(C=C9C(NC(=O)C
(C(C1=CC(=C(O4)C=C1)Cl)O)NC7=O)C(=O)O)O)O)CC(=O)N)NC(=O)C(CC(C)C)
NC)O)Cl)CO)O)O)(C)N)O
```

Network (MDL, Tinker,......)

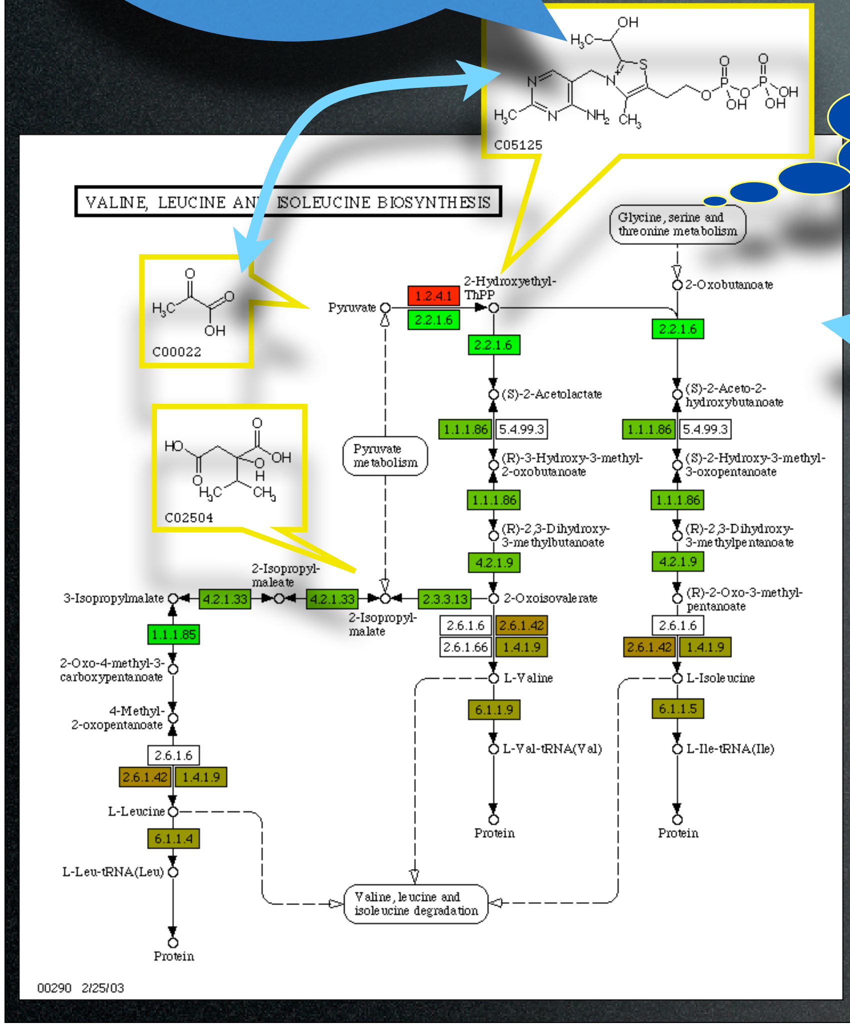
```
  ISISHOST03240423012D 1   1.00000     0.00000     9

  5  4  0     0  0            999 V2000
   -0.0414    0.1586    0.0000 P   0  0  3  0  0  0           0  0  0
   -0.7621   -0.2517    0.0000 O   0  0  0  0  0  0           0  0  0
    0.6759   -0.2517    0.0000 O   0  0  0  0  0  0           0  0  0
    0.1724   -0.6414    0.0000 O   0  0  0  0  0  0           0  0  0
   -0.0414    0.9897    0.0000 O   0  0  0  0  0  0           0  0  0
  1  2  1  0     0  0
  1  3  1  0     0  0
  1  4  1  0     0  0
  1  5  2  0     0  0
M  END
```

Binary (CDX,.......)

```
00000000: 0d6a 4344 3031 3030 0403 0201 0000 0000  .jCD0100........
00000010: 0000 0000 0000 0000 0000 0000 0300 0e00  ................
00000020: 0000 4368 656d 4472 6177 2037 2e30 0800  ..ChemDraw 7.0..
00000030: 0f00 0000 6879 7065 7269 6369 6e2e 6364  ....hypericin.cd
00000040: 7800 0332 0008 00ff ffff ffff ff00 0000  x..2............
00000050: 0000 00ff ff00 0000 00ff ffff ff00 0000  ................
00000060: 00ff ff00 0000 00ff ffff ff00 0000 00ff  ................
```

with KEGG

ChemRuby
find similar
structure



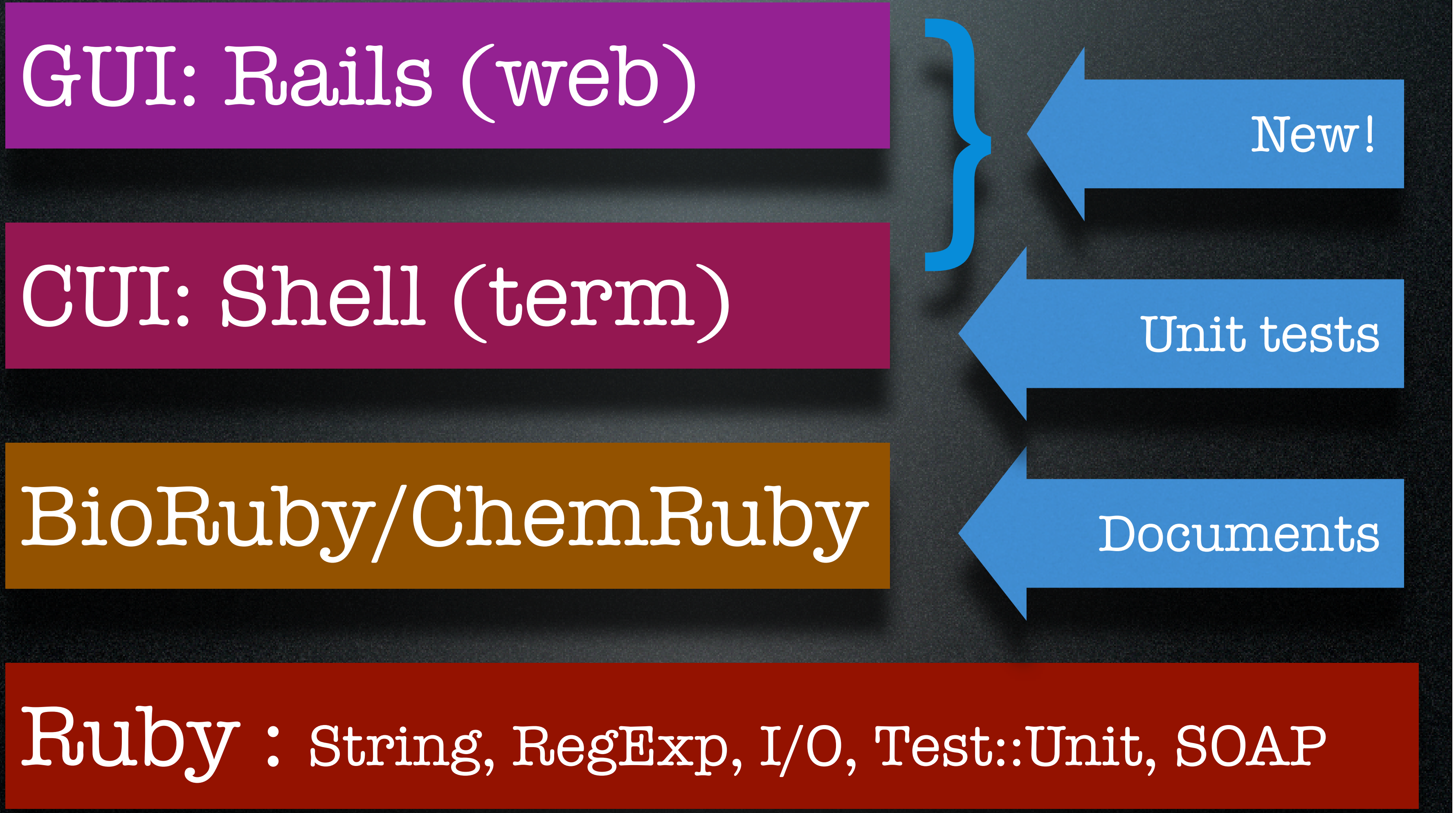
Connect
seamlessly..

BioRuby
gene,
expression,
pathway

```
# 枯草菌のパスウェイ106枚中の1
# 栄養源枯渇時の遺伝子発現データを色でマッピング

# process KEGG Expression data here...

serv = Bio::KEGG::API.new
list = serv.list_pathways(org)
list.each do |path|
  pathway = path.entry_id
  fg_list = Array.new
  bg_list = Array.new
  genes = serv.get_genes_by_pathway(pathway)
  genes.each do |gene|
    fg_list << "#000000"
    bg_list << hash[gene] || "#cccccc"
  end
  url = serv.color_pathway_by_objects(pathway, genes, fg_list,
  serv.save_image(url)
end
```

progress

BioRuby dev in 2005



GUI: Rails (web)

CUI: Shell (term)

BioRuby/ChemRuby

Ruby : String, RegExp, I/O, Test::Unit, SOAP

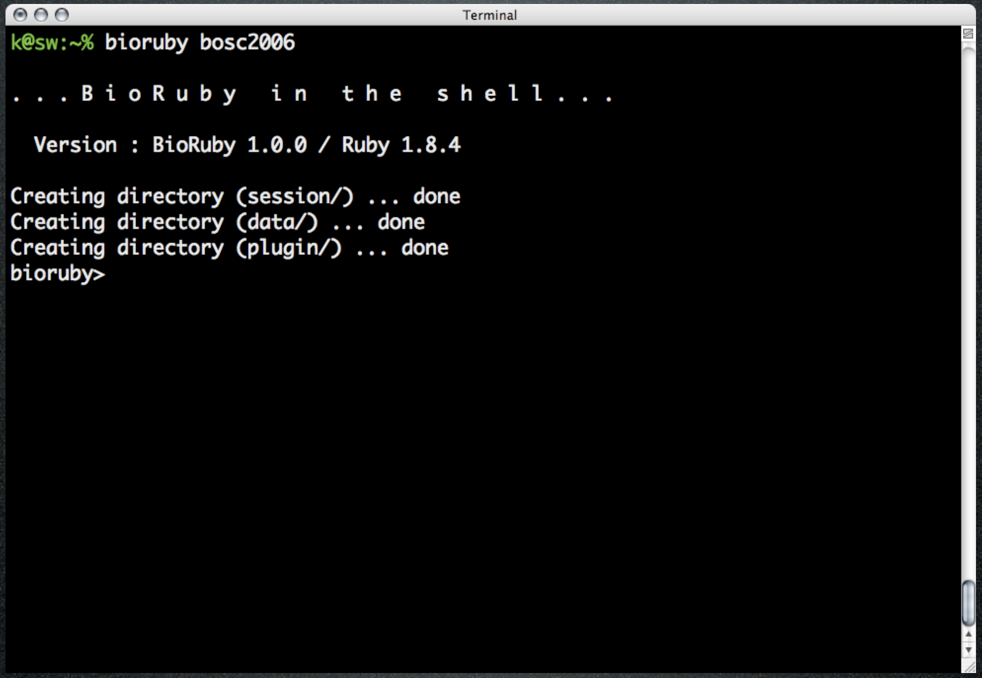
New!

Unit tests

Documents

BioRuby shell

- newly implemented in BioRuby 1.0
- interactive CUI for BioRuby library



```
k@sw:~% bioruby bosc2006

. . . B i o R u b y   i n   t h e   s h e l l . . .

  Version : BioRuby 1.0.0 / Ruby 1.8.4

Creating directory (session/) ... done
Creating directory (data/) ... done
Creating directory (plugin/) ... done
bioruby>
```

features

- persistent object (across sessions)
- history (with timestamps)
- plugins (to extend functionality)
- script generation

  (interactive prototyping to reproduce the procedure)

Library vs Shell (CUI based on 'irb')

- for biologist (!= programmer)
- unified interface

TAMWTDI vs TIOOWTDI

too

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

require 'bio'

# sequence generated from string
dna = Bio::Sequence::NA.new("atgc")

# sequence obtained via KEGG API
kegg = Bio::KEGG::API.new
entry = kegg.bget("hsa:217")
gene = Bio::KEGG::GENES.new(entry)
dna = gene.naseq

# sequence obtained via OBDA BioFetch
obda = Bio::Fetch.new
entry = obda.fetch("gb:AF237819")
gb = Bio::GenBank.new(entry)
dna = gb.naseq
```

```
# sequence obtained via EMBOSS seqret
entry = Bio::EMBOSS.entret("gb:AF237819")
gb = Bio::GenBank.new(entry)
dna = gb.naseq

# sequence obtained from local flat file
Bio::FlatFile.auto("file.txt") do |ff|
  ff.each do |entry|
    dna = entry.naseq
  end
end
```

```
% bioruby
bioruby> dna = seq("atgc")
bioruby> dna = seq("hsa:217")
bioruby> dna = seq("gb:AF237819")
bioruby> flatfile("file.txt") do |entry|
bioruby+   dna = seq(entry)
bioruby+ end
```

to obtain sequence

as a library #1

- Generate sequence from string

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

require 'bio'

# sequence generated from string
dna = Bio::Sequence::NA.new("atgc")
```

as a library #2

- Obtain sequence from KEGG

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

require 'bio'

# sequence obtained via KEGG API
kegg = Bio::KEGG::API.new
entry = kegg.bget("hsa:217")
gene = Bio::KEGG::GENES.new(entry)
dna = gene.naseq
```

as a library #3

- Obtain sequence from BioFetch

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

require 'bio'

# sequence obtained via OBDA BioFetch
obda = Bio::Fetch.new
entry = obda.fetch("gb:AF237819")
gb = Bio::GenBank.new(entry)
dna = gb.naseq
```

as a library #4

- Obtain sequence from EMBOSS

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

require 'bio'

# sequence obtained via EMBOSS seqret
entry = Bio::EMBOSS.entret("gb:AF237819")
gb = Bio::GenBank.new(entry)
dna = gb.naseq
```

as a library #5

- Obtain sequence from EMBOSS

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

require 'bio'

# sequence obtained from local flat file
Bio::FlatFile.auto("file.txt") do |ff|
  ff.each do |entry|
    dna = entry.naseq
  end
end
```

as a shell

- Unified interface

```
% bioruby
bioruby> dna = seq("atgc")
bioruby> dna = seq("hsa:217")
bioruby> dna = seq("gb:AF237819")
bioruby> dna = seq("file.txt")

bioruby> flatfile("file.txt") do |entry|
bioruby+   dna = seq(entry)
bioruby+ end
```

list of shell commands

seq	sequence retrieval	keggapi	access KEGG API server
ent	database entry retrieval	keggdbs	list databases in KEGG
obj	parsed entry retrieval	keggorgs	list organisms in KEGG
seqstat	sequence statistics	keggpathways	list pathways in KEGG
aminoacids	list of amino acids	binfo	database update info
nucleicacids	list of nucleic acids	bfind	search KEGG databases
codontables	translation table	bget	retrieve KEGG entry
flatparse	parse flatfile entry	bconv	convert to KEGG ID
flatfile	open flatfile w/ iteration	obdadbs	list OBDA databases
flatauto	auto detect DB format	biofetch	OBDA BioFeth
flatindex	create index on flatfile	psort	predict localization
flatsearch	search indexed flatfile	blast	execute BLAST search
flatfasta	convert to fasta format	fasta	execute FASTA search

list of util commands

ls	list of objects	script	save procedure in file
rm	delete object	pager	set pager used by disp
savefile	save object in file	config	
cd	change directory	:echo	toggle for irb echo
dir	list of files	:color	toggle for color mode
pwd	print working directory	:message	set opening message
head	peep in file or object	:splash	set opening splash
disp	display file or object	demo	
fold	folding text	:sequence	sequence demo
fill	filling text	:entry	entry retrieval demo
htmlseq	generate colored sequence	:pdb	PDB demo
doublehelix	display animated sequence	:	
midifile	generate DNA music	web	bind to Rails

- obtain gene sequence
- sequence manipulation
- mapping to pathway
- genome music

DEMO - TOC

other enhancements

1.0 features

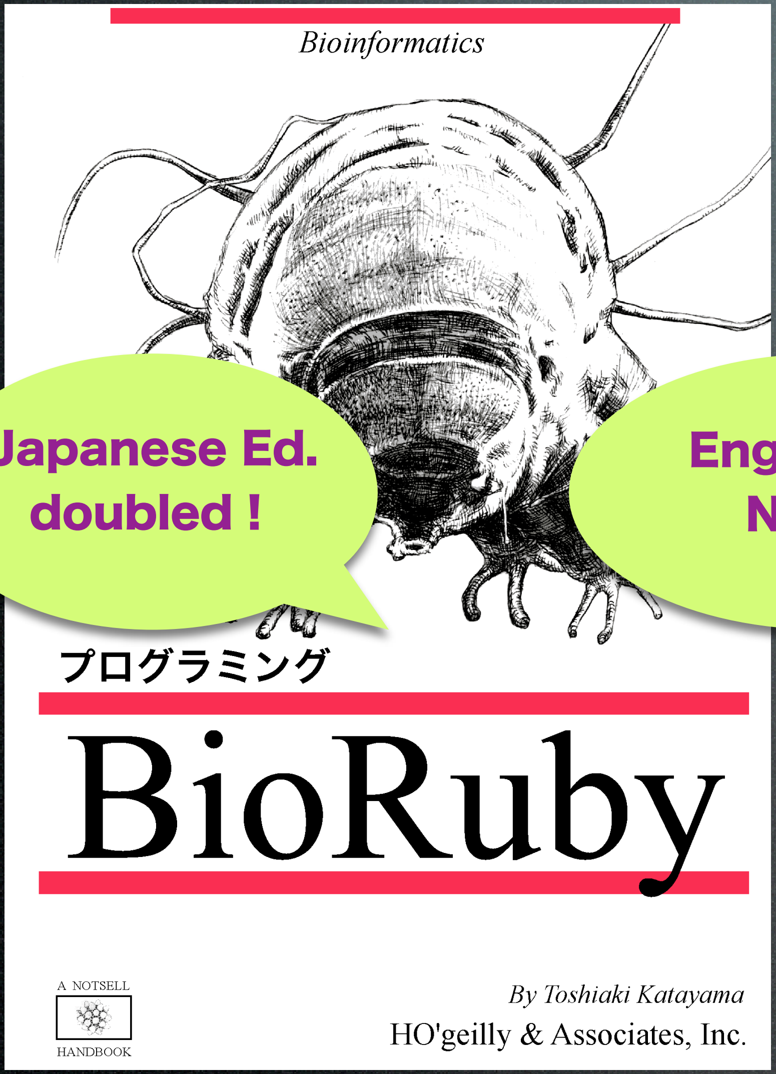
- BioRuby shell
- documentations
- unit tests

documentation

docs

- new English tutorials
- API refs   http://bioruby.org/rdoc/
- doc site   http//bioruby-doc.org/
- guideline doc for developer

Bundled tutorials



Bioinformatics

プログラミング

BioRuby

A NOTSELL HANDBOOK

By Toshiaki Katayama

HO'geilly & Associates, Inc.

Japanese Ed. doubled !

Programming

BioRuby

A NOTSELL HANDBOOK

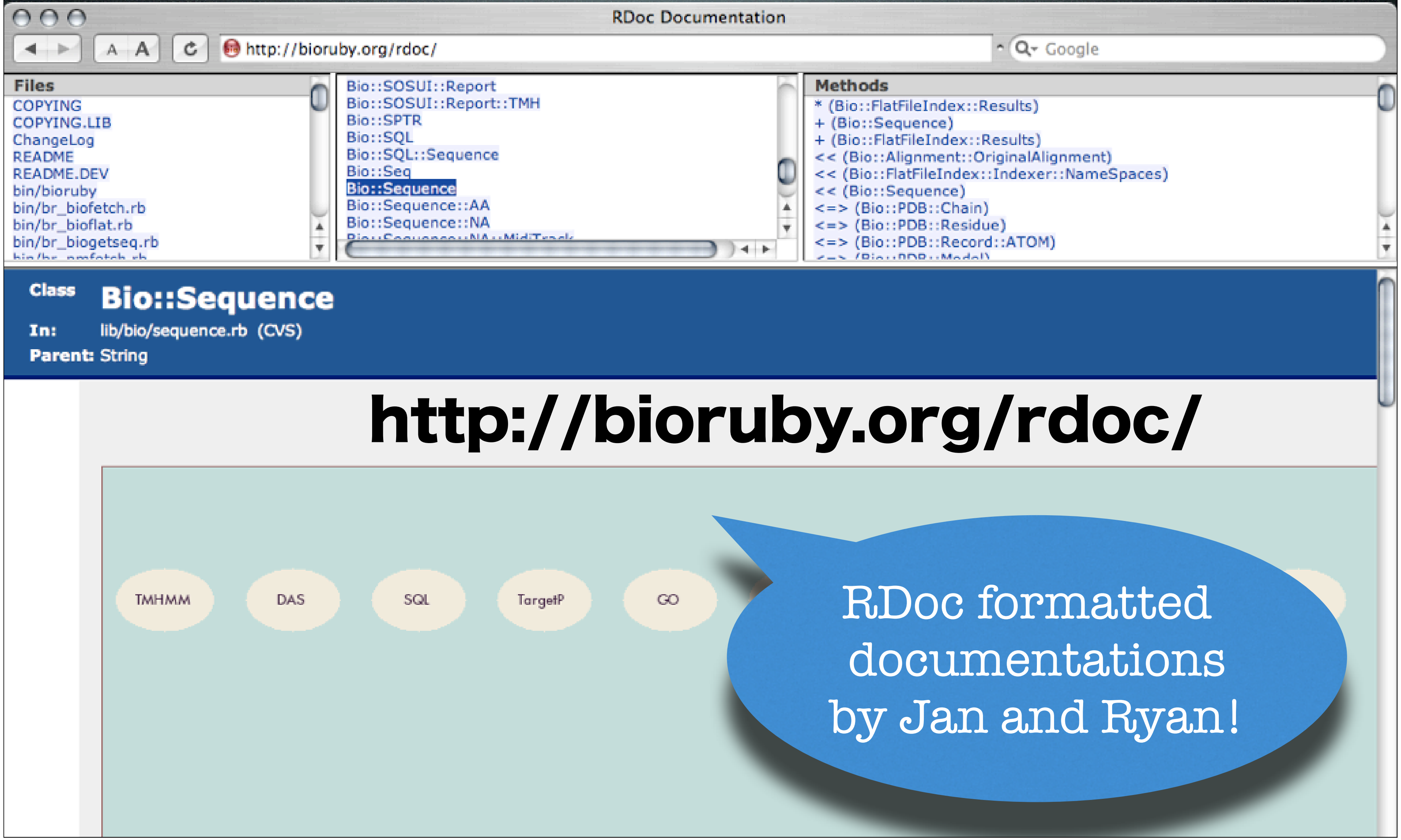
By Toshiaki Katayama

HO'geilly & Associates, Inc.

English Ed. NEW !

Thank you, Pjotr!
(according to Google..)
I miss you.

APIs on the web



RDoc Documentation

http://bioruby.org/rdoc/

Files

COPYING
COPYING.LIB
ChangeLog
README
README.DEV
bin/bioruby
bin/br_biofetch.rb
bin/br_bioflat.rb
bin/br_biogetseq.rb

Bio::SOSUI::Report
Bio::SOSUI::Report::TMH
Bio::SPTR
Bio::SQL
Bio::SQL::Sequence
Bio::Seq
Bio::Sequence
Bio::Sequence::AA
Bio::Sequence::NA

Methods

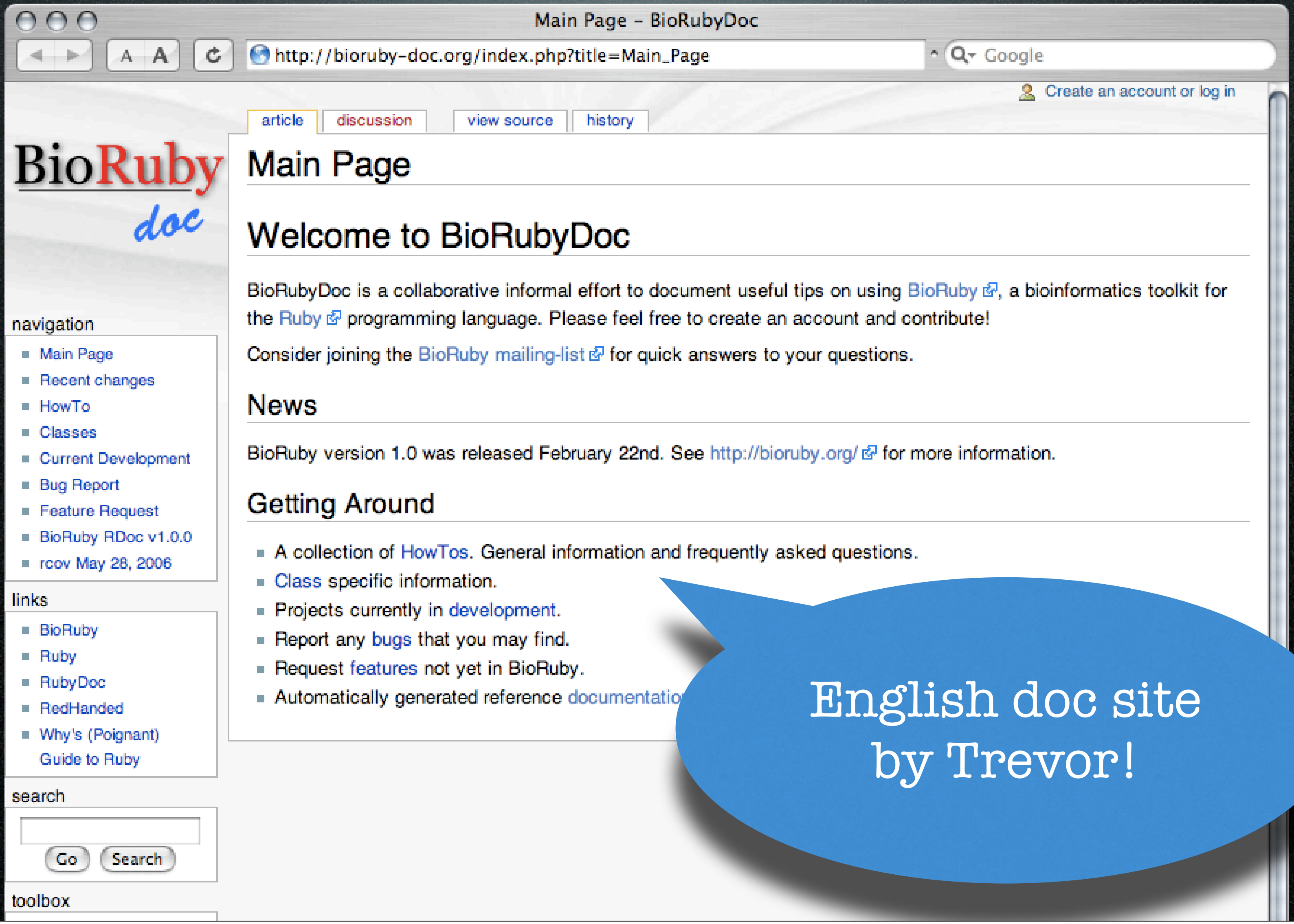
* (Bio::FlatFileIndex::Results)
\+ (Bio::Sequence)
\+ (Bio::FlatFileIndex::Results)
<< (Bio::Alignment::OriginalAlignment)
<< (Bio::FlatFileIndex::Indexer::NameSpaces)
<< (Bio::Sequence)
<=> (Bio::PDB::Chain)
<=> (Bio::PDB::Residue)
<=> (Bio::PDB::Record::ATOM)

Class **Bio::Sequence**
In: lib/bio/sequence.rb (CVS)
Parent: String

http://bioruby.org/rdoc/

RDoc formatted
documentations
by Jan and Ryan!

http://bioruby-doc.org/



Main Page

Welcome to BioRubyDoc

BioRubyDoc is a collaborative informal effort to document useful tips on using BioRuby, a bioinformatics toolkit for the Ruby programming language. Please feel free to create an account and contribute!

Consider joining the BioRuby mailing-list for quick answers to your questions.

News

BioRuby version 1.0 was released February 22nd. See http://bioruby.org/ for more information.

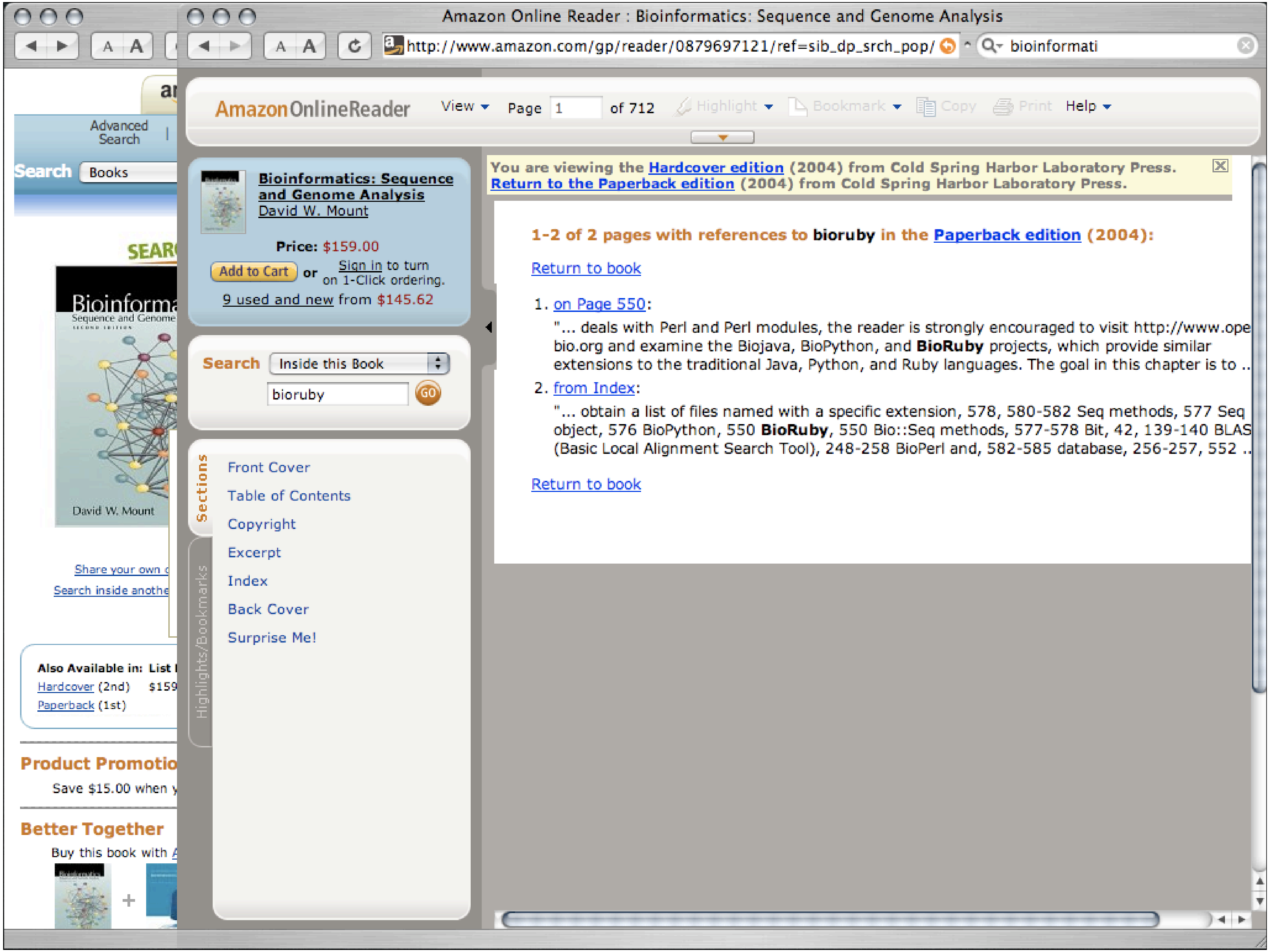
Getting Around

- A collection of HowTos. General information and frequently asked questions.
- Class specific information.
- Projects currently in development.
- Report any bugs that you may find.
- Request features not yet in BioRuby.
- Automatically generated reference documentatio

English doc site
by Trevor!

README.DEV

- guidelines for code contribution
  - license
  - coding style
  - name spaces
  - unit tests
  - autoloads



Bioinformatics: Sequence and Genome Analysis
David W. Mount

Price: $159.00

Add to Cart or Sign in to turn on 1-Click ordering.

9 used and new from $145.62

Search Inside this Book: bioruby GO

Also Available in: List
Hardcover (2nd) $159
Paperback (1st)

Product Promotio
Save $15.00 when y

Better Together
Buy this book with A

You are viewing the Hardcover edition (2004) from Cold Spring Harbor Laboratory Press.
Return to the Paperback edition (2004) from Cold Spring Harbor Laboratory Press.

1-2 of 2 pages with references to bioruby in the Paperback edition (2004):

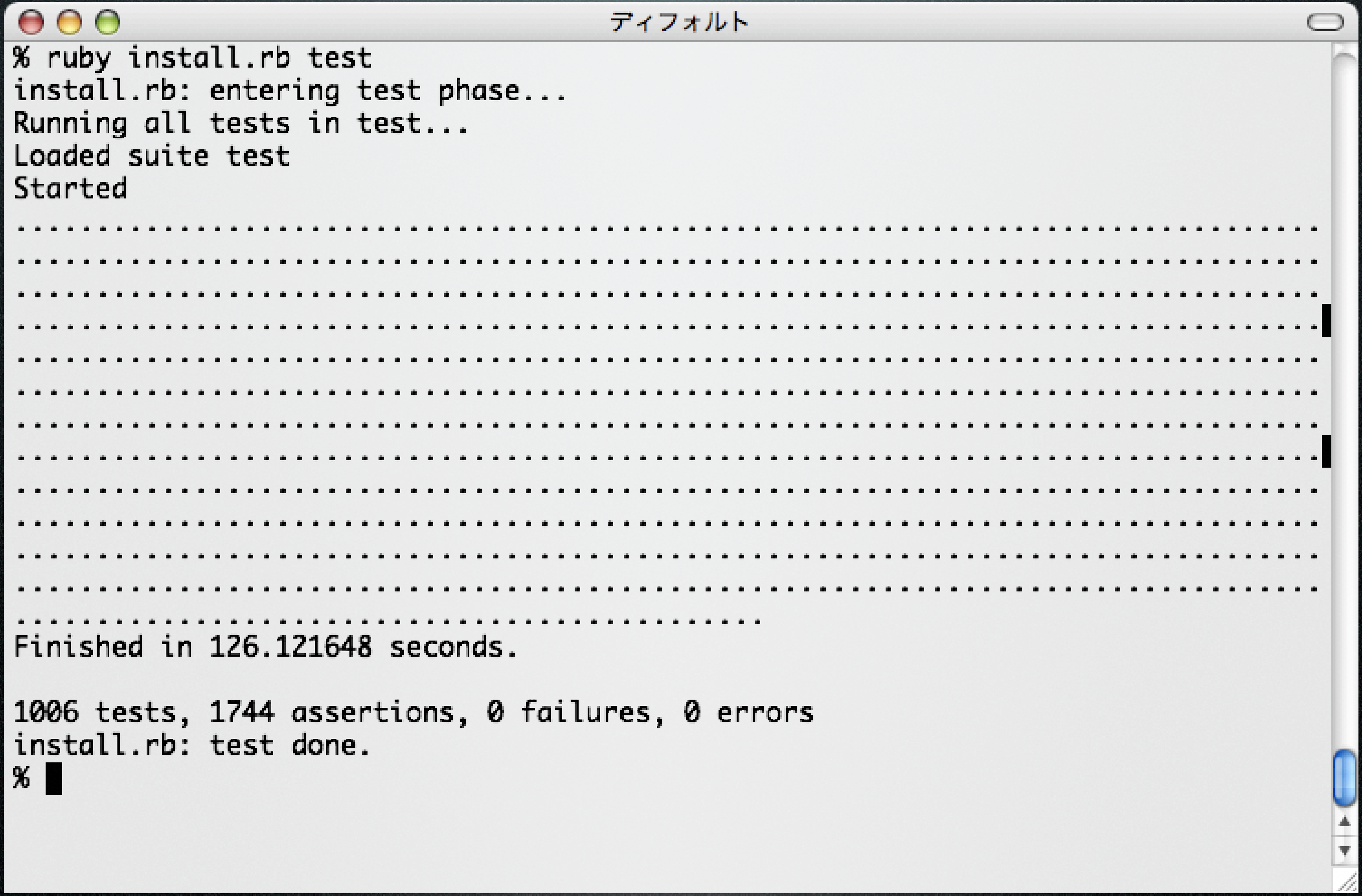
Return to book

1. on Page 550:
   "... deals with Perl and Perl modules, the reader is strongly encouraged to visit http://www.ope bio.org and examine the Biojava, BioPython, and **BioRuby** projects, which provide similar extensions to the traditional Java, Python, and Ruby languages. The goal in this chapter is to ..
2. from Index:
   "... obtain a list of files named with a specific extension, 578, 580-582 Seq methods, 577 Seq object, 576 BioPython, 550 **BioRuby**, 550 Bio::Seq methods, 577-578 Bit, 42, 139-140 BLAS (Basic Local Alignment Search Tool), 248-258 BioPerl and, 582-585 database, 256-257, 552 ..

Return to book

Unit tests

- added >1000 tests for reliability



```
% ruby install.rb test
install.rb: entering test phase...
Running all tests in test...
Loaded suite test
Started
................................................................................
................................................................................
................................................................................
................................................................................
................................................................................
................................................................................
................................................................................
................................................................................
................................................................................
................................................................................
................................................................................
................................................................................
..............................................
Finished in 126.121648 seconds.

1006 tests, 1744 assertions, 0 failures, 0 errors
install.rb: test done.
% 
```

misc

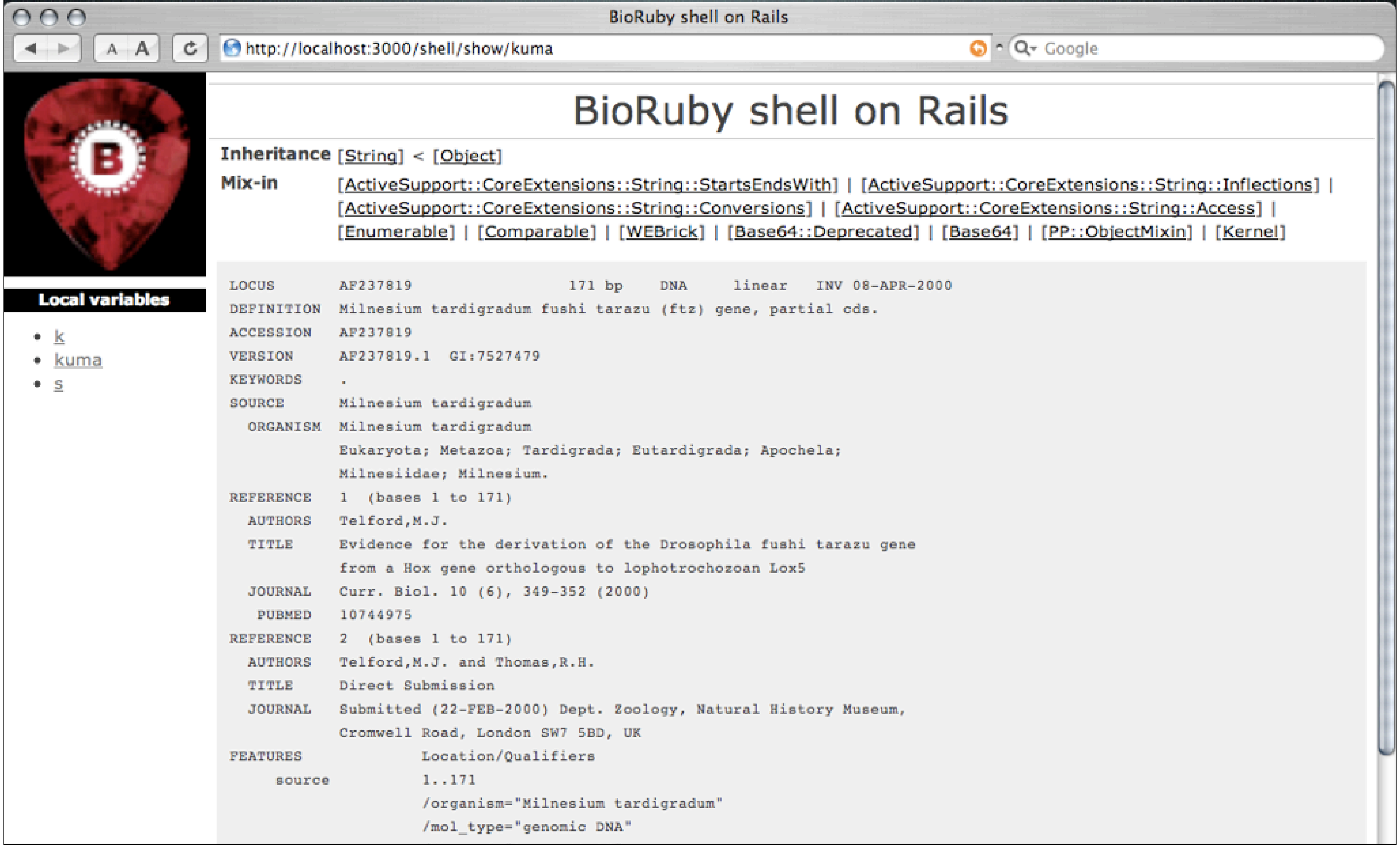
autoload

30x faster to fire up!

```
% repeat 10 { time ruby -r bio -e 0 }
-------------------------------------------------------------
0.6.4:            0.74s user 0.16s system         1.082 total
0.7.0 (9/10):     0.01s user 0.01s system         0.033 total
```

1.082 sec
→ 0.033 sec

BioRuby on Rails



BioRuby shell on Rails

Inheritance [String] < [Object]

Mix-in [ActiveSupport::CoreExtensions::String::StartsEndsWith] | [ActiveSupport::CoreExtensions::String::Inflections] | [ActiveSupport::CoreExtensions::String::Conversions] | [ActiveSupport::CoreExtensions::String::Access] | [Enumerable] | [Comparable] | [WEBrick] | [Base64::Deprecated] | [Base64] | [PP::ObjectMixin] | [Kernel]

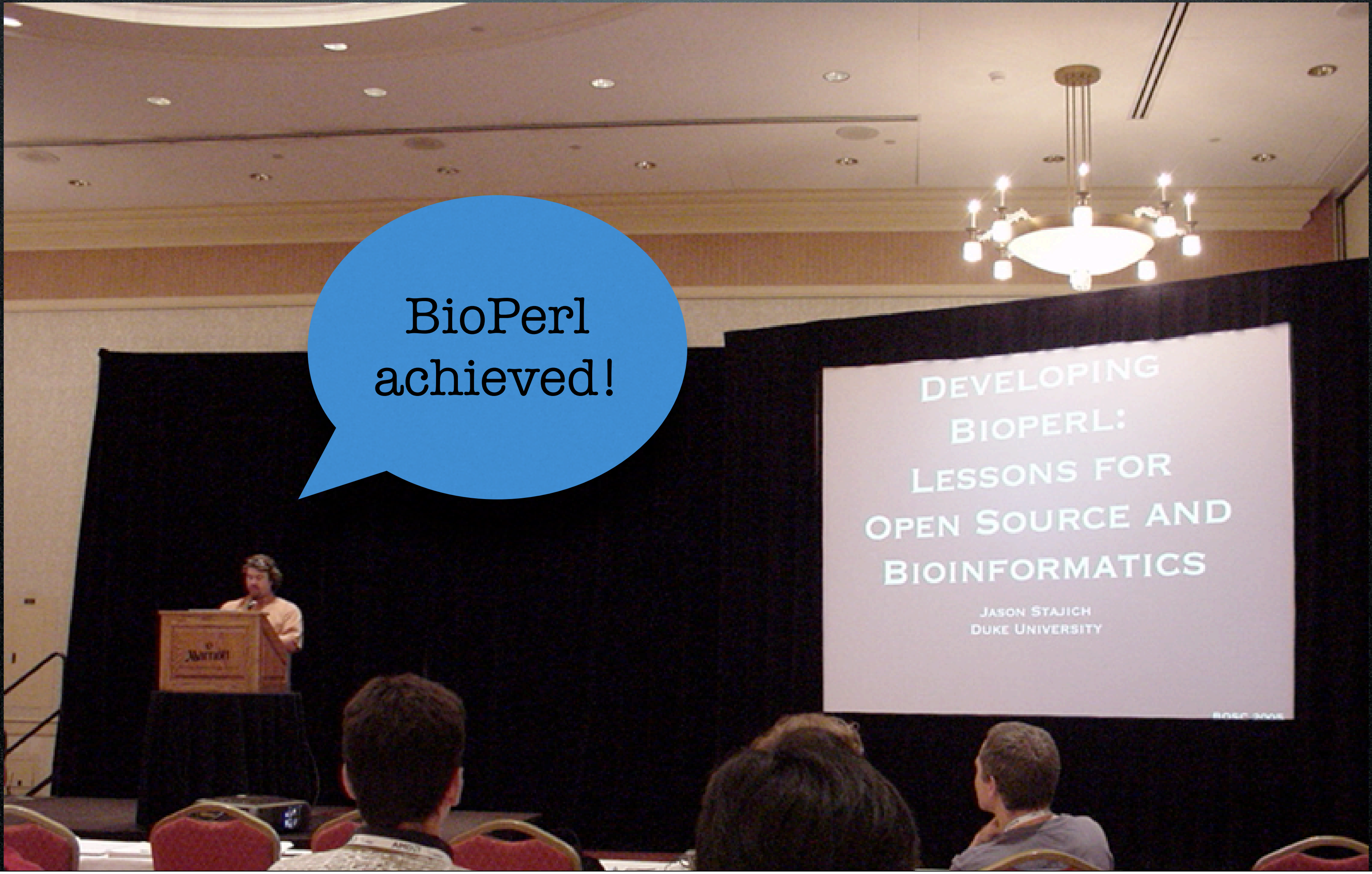
Local variables

- k
- kuma
- s

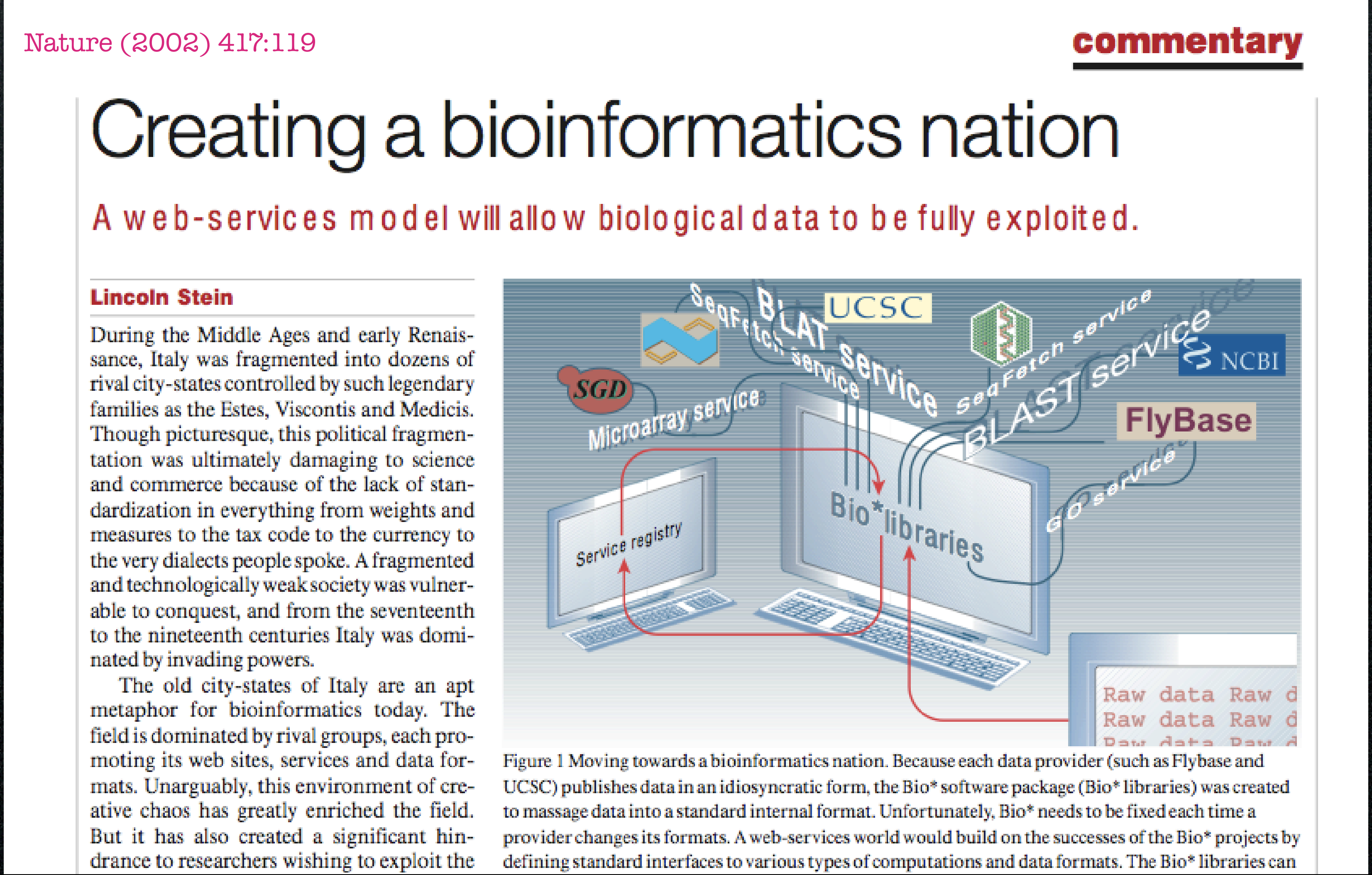
```
LOCUS       AF237819                 171 bp    DNA     linear   INV 08-APR-2000
DEFINITION  Milnesium tardigradum fushi tarazu (ftz) gene, partial cds.
ACCESSION   AF237819
VERSION     AF237819.1  GI:7527479
KEYWORDS    .
SOURCE      Milnesium tardigradum
  ORGANISM  Milnesium tardigradum
            Eukaryota; Metazoa; Tardigrada; Eutardigrada; Apochela;
            Milnesiidae; Milnesium.
REFERENCE   1  (bases 1 to 171)
  AUTHORS   Telford,M.J.
  TITLE     Evidence for the derivation of the Drosophila fushi tarazu gene
            from a Hox gene orthologous to lophotrochozoan Lox5
  JOURNAL   Curr. Biol. 10 (6), 349-352 (2000)
   PUBMED   10744975
REFERENCE   2  (bases 1 to 171)
  AUTHORS   Telford,M.J. and Thomas,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-2000) Dept. Zoology, Natural History Museum,
            Cromwell Road, London SW7 5BD, UK
FEATURES             Location/Qualifiers
     source          1..171
                     /organism="Milnesium tardigradum"
                     /mol_type="genomic DNA"
```

road to open-bio 2.0

BOSC2005@Detroit



what direction to go?



Nature (2002) 417:119

commentary

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

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

web services

Author: Bren Vaughan, ES Group, EMBL-European Bioinformatics Institute

Ruby and WS

SOAP/WSDL in Java

1. download and install Axis library

(...snip...)

2. generate Java classes from WSDL

```
% java -classpath axis.jar:jaxrpc.jar:commons-logging.jar:commons-discovery.jar: saaj.jar: wsdl4j.jar:.
   org.apache.axis.wsdl.WSDL2Java -p keggapi   http://soap.genome.jp/KEGG.wsdl
% javac -classpath axis.jar:jaxrpc.jar:wsdl4j.jar:. keggapi/KEGGLocator.java
% jar cvf keggapi.jar keggapi/*
```

3. code

```
 import keggapi.*;

 class GetGenesByPathway {
      public static void main(String[] args) throws Exception {
           KEGGLocator  locator = new KEGGLocator();
           KEGGPortType serv    = locator.getKEGGPort();

           String   query   = args[0];
           String[] results = serv.get_genes_by_pathway(query);

           for (int i = 0; i < results.length; i++) {
                System.out.println(results[i]);
           }
      }
}
```

4. compile and execute

(...snip...)

SOAP/WSDL in Ruby

1. code

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

require 'soap/wsdlDriver'

wsdl = "http://soap.genome.jp/KEGG.wsdl"
serv = SOAP::WSDLDriverFactory.new(wsdl).create_driver
serv.generate_explicit_type = true

puts serv.get_enzymes_by_pathway(ARGV.shift)
```

2. execute

```
% ruby enzymes.rb map00020
```

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

require 'bio'

serv = Bio::KEGG::API.new

puts serv.get_enzymes_by_pathway(ARGV.shift)
```

WSs in combination

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

require 'bio'

kegg = Bio::KEGG::API.new
seq = kegg.get_aaseqs("hsa:aldh2")

ddbj = Bio::DDBJ::XML::Blast.new
res = ddbj.searchSimple("blastp", "SWISS", seq)
hit = Bio::Blast::Report.new(res).hits.first

ncbi = Bio::NCBI::SOAP::EUtils.new
ent = ncbi.run_eFetch("db" => "protein","id" => hit)

ebi  = Bio::EBI::SOAP::Dbfetch.new
 :
```

KEGG API access log

Statistics on KEGG API v5.0

- 9.5 million hits since Feb 2006

```
3735017 Perl
2818454 Java
2191769 Ruby
 287940 Python
```

Road to BioRuby 2.0

- 1.1: ASAP (need to fix serious bug :)
- 1.2: BioRuby with ActiveRecord
- 1.3: BioRuby <=> ChemRuby

:

- 1.9: BioRuby on Rails
- 2.0: BioRuby on web service

Acknowledgments

Questions?

- If you have, please ask me veeeeery slooooooowly, clearly, word by word.
- Q: ___or type in here___

http://bioruby.org/