BioRuby and distributed development

Ruby project update

Pjotr Prins, *Ben Woodcroft*, Francesco Strozzi,
Hiroyuki Mishima, Toshiaki Katayama, Naohisa Goto,
Raoul Bonnal, *Joachim Baran*

Bioinformatics Open Source FEST (BOSF) 2014

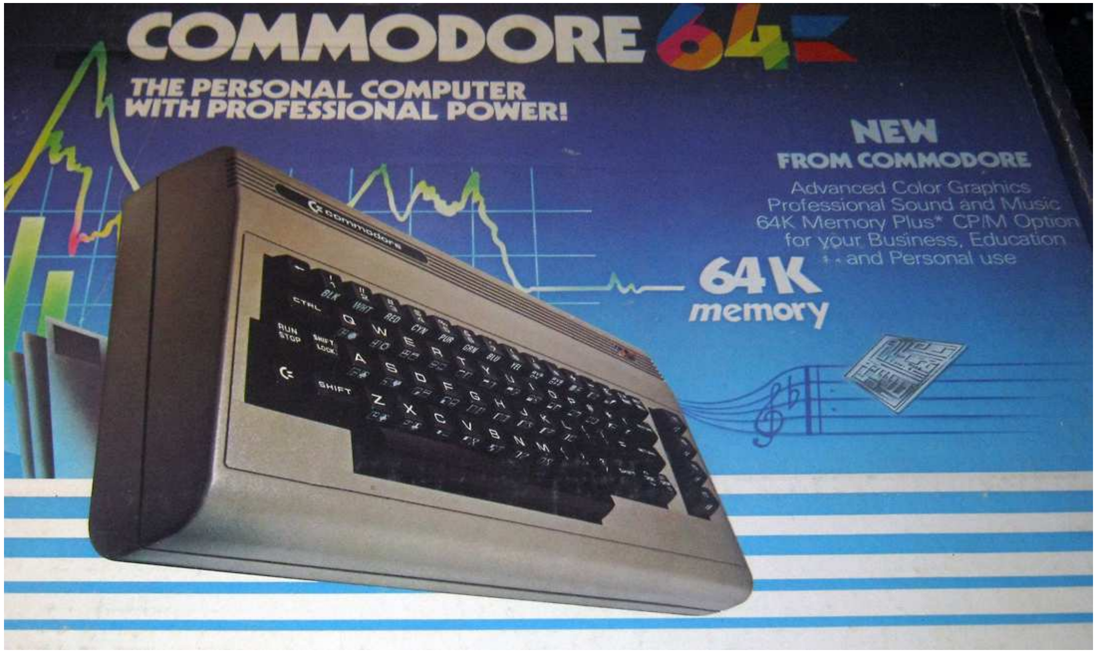
BioRuby

The daydreams of cat herders

Language wars

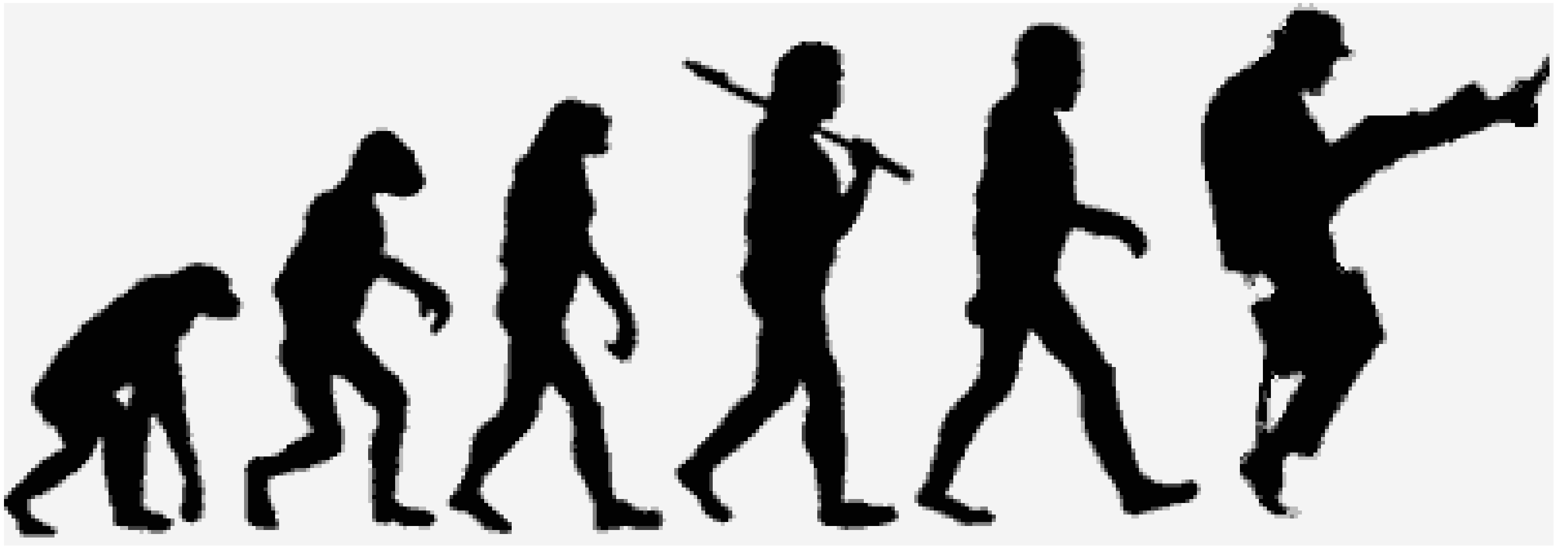
Python has won

I am a Commodore guy



Commodore won!

And now for something. . .



Ruby is different

Why Ruby?

Ruby is my favorite programming language because

- Dynamically typed ’(Ruby LISP Python Perl R JS)
- Fully OOP ’(Ruby Python Perl)
- Pretty good functional programming ’(LISP Ruby Python Perl6 JS ,Coffee)
- Great meta programming ’(,LISP Ruby Python Perl6)
- Ruby allows concise but *readable* code
- Ruby attracts the right crowd

Who are Rubyistas?

Ruby has leading initiatives in

- Web frameworks (Rails, Sinatra, Padrino etc.)
- Testing frameworks (Cucumber, Travis CI)
- Web solutions (github, twitter, Hulu, Heroku etc.)
- Software deployment (Homebrew, Chef, Puppet, Vagrant)
- Notably meta: innovation at the process level

Norwegian reminder

SO, what has this to do with BioRuby?

Innovate the process!

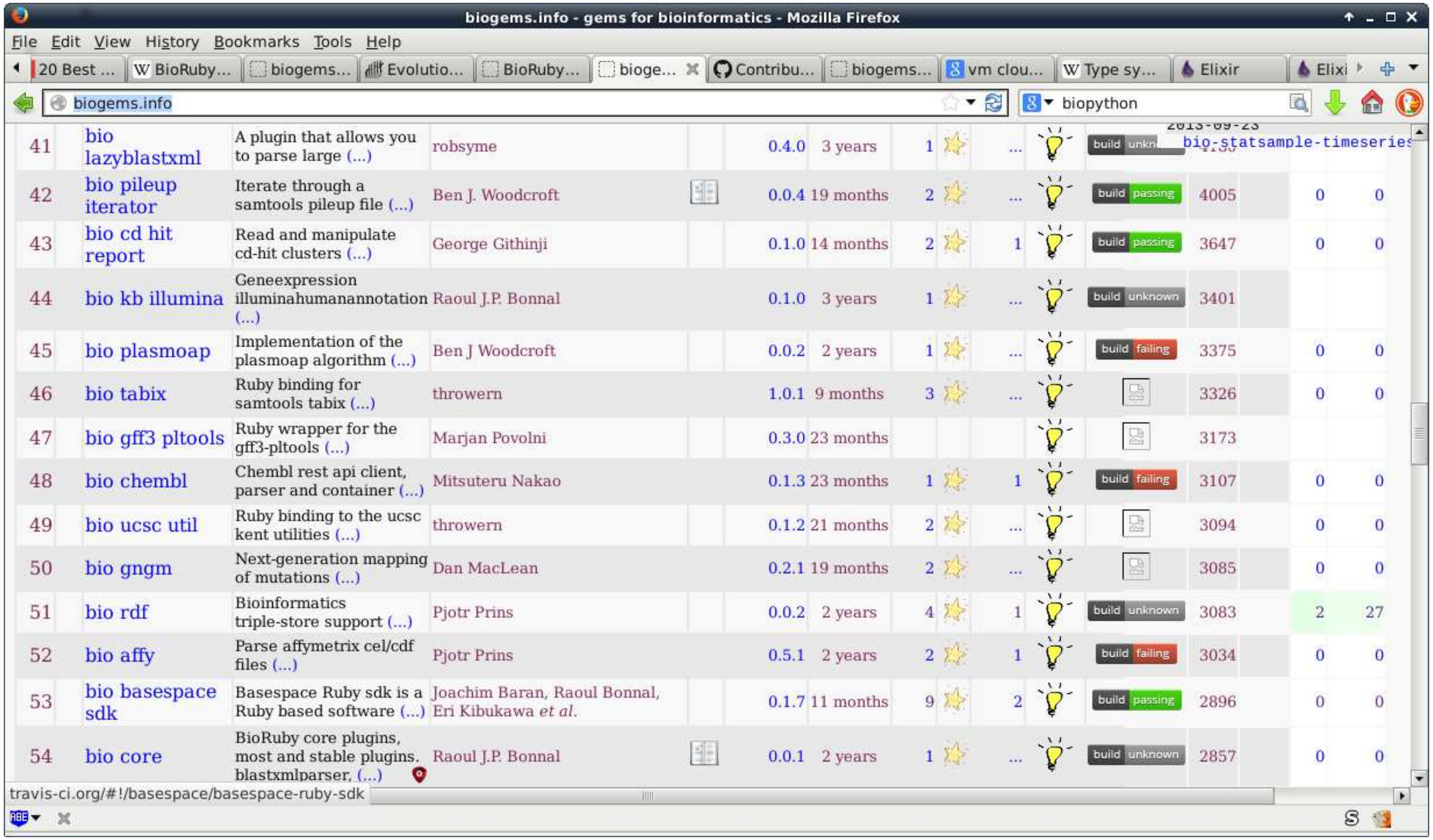
- We leveraged Rubygems to provide BioRuby plugins (meta)
- Easy! Run:

  ```
  biogem -with-bin -with-ffi -cucumber name
  ```

- 100 new projects in 4 years (2/month)
- 6 new contributors in last year

Biogem: an effective tool based approach for scaling up open source software development in bioinformatics, Bonnal *et al.*, Bioinformatics 2012
doi:10.1093/bioinformatics/bts080

Biogems.info



41	bio lazyblastxml	A plugin that allows you to parse large (...)	robsyme	0.4.0	3 years	1	...	build unkn			
42	bio pileup iterator	Iterate through a samtools pileup file (...)	Ben J. Woodcroft	0.0.4	19 months	2	...	build passing	4005	0	0
43	bio cd hit report	Read and manipulate cd-hit clusters (...)	George Githinji	0.1.0	14 months	2	1	build passing	3647	0	0
44	bio kb illumina	Geneexpression illuminahumanannotation (...)	Raoul J.P. Bonnal	0.1.0	3 years	1	...	build unknown	3401		
45	bio plasmoap	Implementation of the plasmoap algorithm (...)	Ben J Woodcroft	0.0.2	2 years	1	...	build failing	3375	0	0
46	bio tabix	Ruby binding for samtools tabix (...)	throwern	1.0.1	9 months	3	...		3326	0	0
47	bio gff3 pltools	Ruby wrapper for the gff3-pltools (...)	Marjan Povolni	0.3.0	23 months				3173		
48	bio chembl	Chembl rest api client, parser and container (...)	Mitsuteru Nakao	0.1.3	23 months	1	1	build failing	3107	0	0
49	bio ucsc util	Ruby binding to the ucsc kent utilities (...)	throwern	0.1.2	21 months	2	...		3094	0	0
50	bio gngm	Next-generation mapping of mutations (...)	Dan MacLean	0.2.1	19 months	2	...		3085	0	0
51	bio rdf	Bioinformatics triple-store support (...)	Pjotr Prins	0.0.2	2 years	4	1	build unknown	3083	2	27
52	bio affy	Parse affymetrix cel/cdf files (...)	Pjotr Prins	0.5.1	2 years	2	1	build failing	3034	0	0
53	bio basespace sdk	Basespace Ruby sdk is a Ruby based software (...)	Joachim Baran, Raoul Bonnal, Eri Kibukawa *et al.*	0.1.7	11 months	9	2	build passing	2896	0	0
54	bio core	BioRuby core plugins, most and stable plugins. blastxmlparser, (...)	Raoul J.P. Bonnal	0.0.1	2 years	1	...	build unknown	2857	0	0

RDF OOP

- BioInterchange: conversion of standardized life sciences data formats into RDF (GFF3, GVF, Newick, Pubmet JSON, PDFx XML, etc.)
  - standardisation & ontologies
  - efficient (batch processing, no in-memory data model)
  - extensible via clear use of OOP design patterns
  - CLI

```
biointerchange -input phylotastic.newick -rdf rdf.phylotastic.newick
```

Joachim Baran & Michel Dumontier `http://www.biointerchange.org/`

RDF JSON

- rdf2json: translating RDF into JSON-LD or minimized JSON alternative
  - goodness of NoSQL (Elasticsearch, MongoDB, RethinkDB, etc.)
  - CLI

```
rdf2json [options] -input filename.nt -output filename.json
```

Joachim Baran & Michel Dumontier `https://github.com/joejimbo/rdf2json`

RDF Functional

- Bio-rdf - introduce macros for SPARQL and RDF using erb
- Bio-vcf - convert VCF to RDF/HTMl/LaTeX
- Bio-table - convert tabular data to RDF/HTML/LaTeX
- More on that in ‘small tools’

`http://biogems.info/`

SciRuby

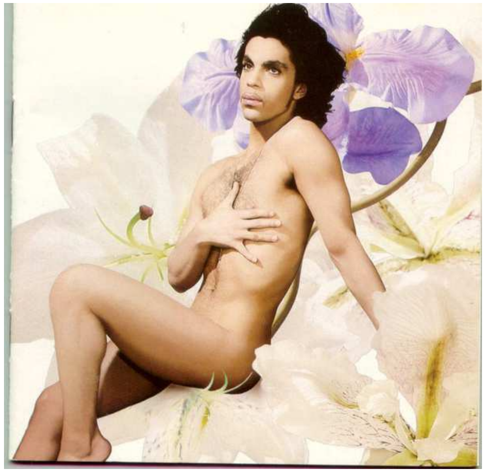
- Small tools manifesto
  `https://github.com/pjotrp/bioinformatics/blob/ma`
- Branching out beyond life sciences
- Cross over to SciRuby
- Attract great developers for GSoC
  - OpenCL
  - Naoki Nishida D3 (demo of interactive plotting
    `https://github.com/domitry/Nyaplotjs`)

‘BioRuby’

- Distributed development (BioRuby is in maintenance mode - note)
- Meta, bin and pipes
- Large data handling
- standards JSON, RDF etc.
- D3 JS library also usable from Python etc. (in fact, Iruby-notebook is Python based)
- Sambamba D language (Artem Tarasov)

BioRuby community

- In all ‘the project formerly known as BioRuby’ is
  - into distributed development (`http://biogems.info/`)
  - has no hangups about computer languages
  - reaching out (SciRuby, RDF) and crossing over…



What is missing in Ruby

- Strong functional support
- Concurrent programming
- Full expression matching
- Macros
- These are also missing in Python, Perl
- Language of the year 2014/2015: Elixir

Where are we heading?

- The world is parallel and multi-core. . .
- Immutable data (Erlang, Scala, D, Clojure, Haskell)
- Higher level abstractions for multi-threading (actors, pipes, continuations/futures)
- More functional programming (expression matching, macros)

Elixir

- Elixir is a dynamic functional language for building concurrent, distributed and fault-tolerant applications (no OOP)
- Elixir has immutable data, pattern matching, higher level abstractions for multi-threading and macros(!) in a readable *Ruby inspired language*
- Elixir is the love child of Ruby and Erlang (and LISP)

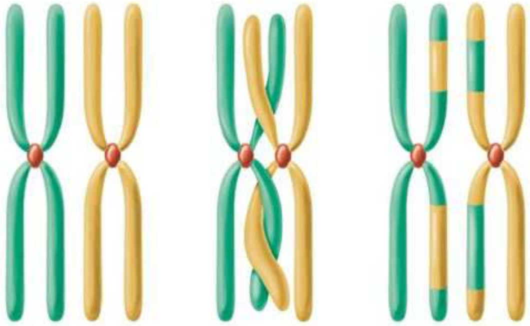
`http://elixir-lang.org/`

Interested? Dave Thomas talk

`http://www.confreaks.com/videos/2591-lonestarruby2013-elixir-power-of-erlang`

Ruby DNA

- Nothing central; distributed development; glue is biogems.info
- Reaching out
- Plan: more tools/libraries/languages
- Plan: GNU Guix and Cloudbiolinux deployment
- Ruby DNA: sticky, reaching out and crossing over!



Acknowledgements

- BioRuby ML, SciRuby and OBF (Biopython, BioPerl) for supporting the project formerly known as BioRuby
- Personal genomics and bioinformatics, University Medical Center Utrecht for supporting FOSS work of Pjotr Prins
- Center for Biomedical Informatics Research, Stanford University, for supporting FOSS work of Joachim Baran
- Bioinformatics Core Facility of Parco Tecnologico Padano for supporting the FOSS work of Francesco Strozzi
- Istituto Nazionale Genetica Molecolare for supporting FOSS work of Raoul Bonnal
- Australian Centre for Ecogenomics for supporting FOSS work of Ben Woodcroft
- and many others. . .