CodeFest, BOSC, ISMB

Torsten Seemann
What did I attend?

- **CodeFest**
  - 2 days, \(\sim 10^1\) ppl
  - Loloya Marymount Uni, LA

- **BOSC**
  - 2 days, \(\sim 10^2\) ppl
  - Long Beach Convention Centre, LA

- **ISMB**
  - 3 days, \(\sim 10^3\) ppl
  - Long Beach Convention Centre, LA
CodeFest
What is CodeFest?

- **In past years**
  - ... a great opportunity for open source bioinformatics developers to meet and work collaboratively. We hope to have developers from projects such as BioPerl, BioJava, Biopython, BioRuby, Galaxy and EMBOSS in attendance.

- **This year**
  - Lower attendance
  - Primarily: Galaxy, BioPython, BioLinux
  - Got to know key Galaxy devs (Dan Bannon)
  - Got training from BioLinux head (Tim Booth)
  - UCSD bioinformatics core head (Dawei "Lego" Lin)
University or Mall?
BOSC
BOSC

- Bioinformatics Open Source Conference
  - Satellite meeting of ISMB
  - First held in 2000

- Traditional focus
  - OpenBio projects - Bio::Perl, BioPython, BioJava
  - GMOD projects - Chado, Apollo

- This year
  - Cloud computing: CloudMan, ...
  - Workflow tools: Galaxy, Chipster, Mobyle
  - Data management: HDF5, GenomeSpace
My BOSC highlights

● Jonathan Eisen
  ○ editor in chief of PLoS Biology
  ○ brother of Michael Eisen, founder of PLoS
  ○ prolific blogger/tweeter - @phylogenomics

● "The impact factor is not the goal. The goal is to publish scientific results that are technically sound. And that, after things are published, rather than before, we can evaluate whether they are important or interesting."

● Open-access
  ○ scientific progress will be more efficient if open
  ○ he (now) only publishes in open-access journals
3. Consider being more open pre-publication

- Data
- Papers
- Talks
- Software
- Ideas
- Notebook
My BOSC highlights

- **C. Titus Brown**
  - prolific blogger - *Living in an Ivory Basement*
  - open scientist - grants, reviews, slides - all online
  - khmer software library (python/C)

- **Digital normalization**
  - subsample reads using k-mer frequencies
  - 100/1000-fold reduction in read volume
  - keep primary signal/information, remove noise
  - one-pass algorithm!

- **Applications**
  - can do terabase sized meta-genome assembly
ISMB

- Intelligent Systems for Molecular Biology
  - ISCB: Intl. Society for Computational Biology

- 8 parallel tracks
  - 20 minute talks all day
  - Stream: papers, late-breaking, tech, workshop, ...

- 6 keynotes
  - book-end each day

- 100s of good posters
  - to distract from the lunch on offer
My ISMB highlights

● SEQuel
  ○ *Improving the accuracy of genome assemblies*
  ○ Roy Ronen @ Pevzner Lab

● Summary
  ○ Aligns reads back to correct errors
  ○ Attempts to handle repeats
  ○ Compared Velvet, SOAPdenovo, EULER
  ○ Velvet was least error prone
  ○ SEQuel reduced errors by 65% on average!
  ○ Plan to add to our pipelines
Long Beach Mariner
My ISMB highlights

- **Sensitive metagenomics searches against reduced libraries with FASTA**
  - Prof. William (Bill) Pearson
  - creator of FASTA suite, still actively developed

- **Method**
  - Cluster UniProt
  - Run short read queries via FASTX
    - FASTA can pull in all cluster members for the hit!
    - Much faster than aligning to whole database
  - PAM30 matrix to get significance
    - Low PAM = short segments, high similarity
    - Default BLOSUM62 + BLAST is doomed
The Ion Bus
Would I go again?

- **CodeFest**
  - needed more people I think

- **BOSC**
  - got to meet people i've only interacted on net with
  - many knew my name, built my confidence
  - might try a SIG instead (and move back/forth)

- **ISMB**
  - expensive, but a premier conference
  - usually something on in one track of interest
  - lots of big names in attendance, got to meet some
  - plan ahead if you want to catch up with people!
Thank you!