

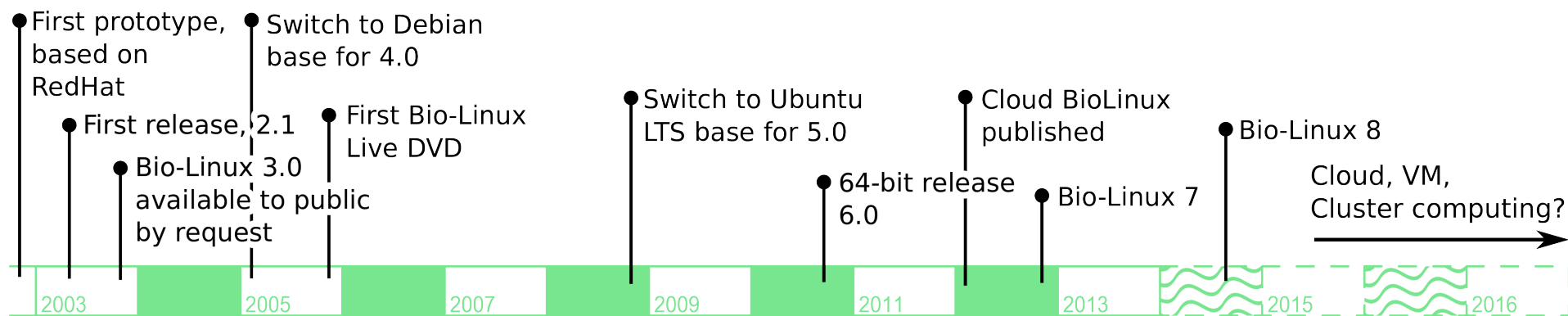
Tim Booth <tbooth@ceh.ac.uk>
NERC Environmental Bioinformatics Centre
CEH Wallingford, UK

Improvements and new features in the 7th
major release of the **Bio-Linux** distro

- History of Bio-Linux?
- Who uses it?
- What's new in Version 7?
- What are we planning for the future?

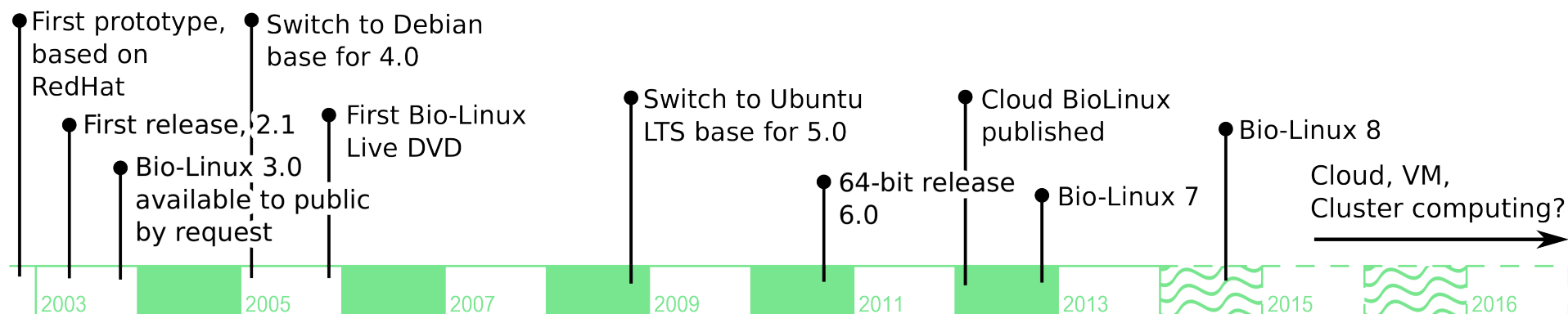
A quick history of Bio-Linux

- Started for a specific group of users
- Single system image, no packages
- EMBOSS + STADEN “killer apps”

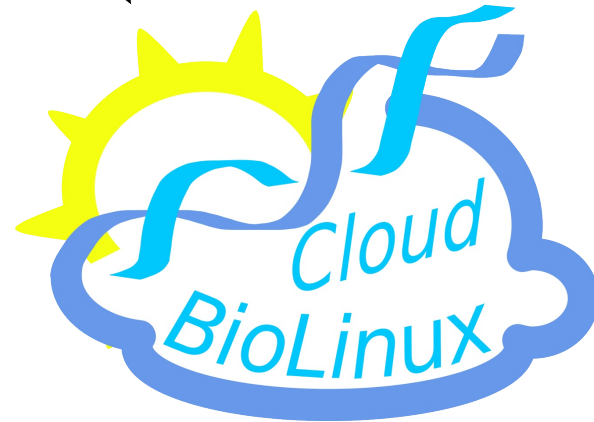


A quick history of Bio-Linux

- Since 2008 a customised Ubuntu image
- Close links with Debian Med



Developer Interactions



Known users of Bio-Linux



What goes on Bio-Linux?

- Must be freely redistributable, ideally FOSS
- Marker gene analysis, metagenomics pipelines
- General conversion and formatting tools
- Things that are already packaged (Debian Med)
- Things that NERC users ask for
- Things that make the system more user-friendly
- Priority is to make things better for existing users than to find new ones

New in version 7

The screenshot displays the Galaxy web interface in a Google Chrome browser window. The address bar shows 'localhost:8080'. The main navigation bar includes 'Galaxy', 'Analyze Data', 'Workflow', 'Shared Data', 'Help', and 'User', along with a 'Using 0 bytes' indicator.

The left sidebar contains a 'Tools' section with a list of tools, including ClustalW, Metagenomic analyses, FASTA manipulation, NCBI BLAST+, NGS: QC and manipulation, NGS: Picard (beta), and NGS: Mapping. The 'NGS: Mapping' section is expanded, showing tools like Lastz, Map with Bowtie for Illumina, Map with BWA for SOLiD, and Map with BFAST.

The main content area is titled 'Bowtie2 (version 0.1)'. It contains the following configuration options:

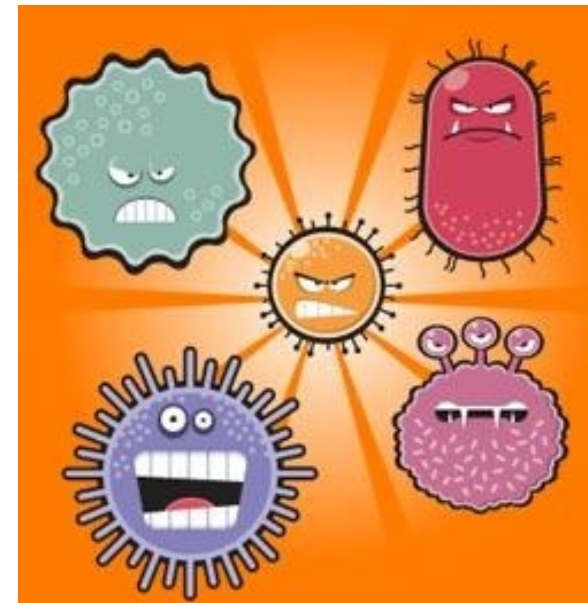
- Is this library mate-paired?:** Single-end
- FASTQ file:** (Dropdown menu)
- Will you select a reference genome from your history or use a built-in index?:** Use a built-in index
- Select a reference genome:** (Dropdown menu)
- Bowtie settings to use:** Use Defaults

Below the configuration options is an 'Execute' button. The right sidebar shows a 'History' section with a list of jobs, including 'Fasting_Example.qual' and 'Fasting_Example.fna'.

mothur

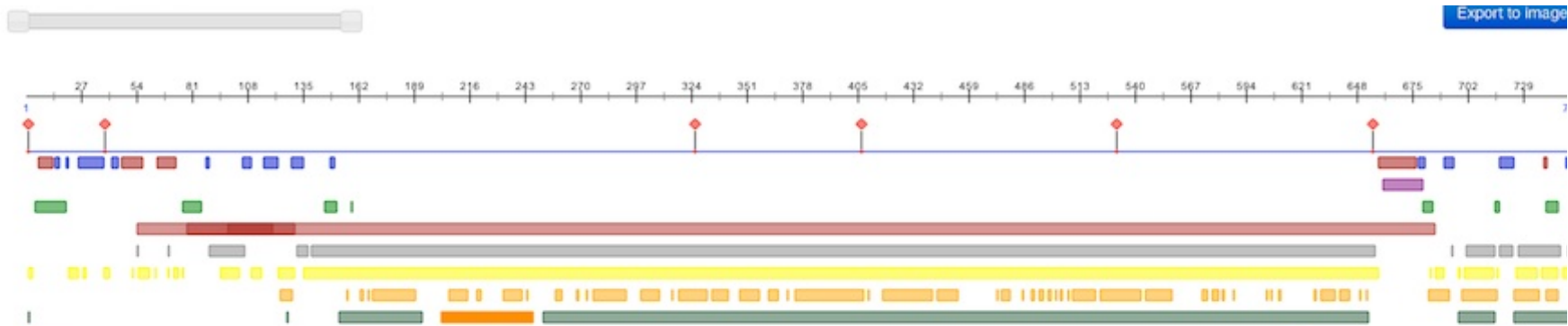


AmpliconNoise



New in version 7

- Binding Sites >
- Cellular Localization >
- Transmembrane Beta-barrels >
- PROTEIN ANALYSIS
 - Post-translational Modifications >
 - Sequence Similarity Search >
 - Protein Structure Prediction >
- Protein Structure Prediction >

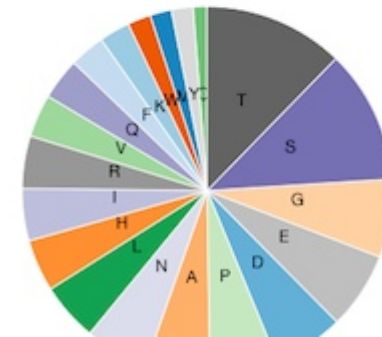


PREDICT PROTEIN

Summary

Recommended Name	P16070
Sequence Length	742
Number of Aligned Proteins	76
Number of Matched PDB Structures	14

Amino Acid composition



- Also...
 - FastQC – nextgen quality assessment
 - Bowtie-Bio tools
 - Several NGS assemblers
 - and of course latest EMBOSS + STADEN (gap5)

Teaching on Bio-Linux



EOS

→ environmentalomics.org

elixir-europe.org ←



- Continue to do what we're good at
- More developer time wanted (always!)
- More tutorial/course material
- More focus on VM/Cloud

Acknowledgements

- Current NEBC team
 - Tim Booth
 - Mesude Bicak
 - Hyun Soon Gweon
 - Dawn Field
 - Katja Lehmann
 - Tracey Timms-Wilson
 - Daniel Pass
 - ***Prof. Dawn Field***
- Past Bio-Linux developers
 - Milo Thurston
 - Daniel Swan
 - Stewart Houten
 - Oliver Buckley
 - Bela Tiwari
- + thanks to various NBAF awardees, esp. Casey Bergman