Introduction 1 day: wish list for course contents

Training Day 1: Wish list of course contents

This O'Reilly book may suggest a useful guide framework: http://shop.oreilly.com/product/9781565926646.do

Depending on the exact structure/content of their course, TGAC seem to have planned to take four days to cover a similar scope: http://www.tgac.ac.uk/tgac-summer-school/ ; http://www.tgac.ac.uk/uploads/Training%20Events/tgac%20summer%20school%20programme%20n

(A) Linux

- mkdir, rm, mv, cp, ls, pwd, tail, uniq, wc, grep, man
- Recursion (in context of 'into directories', e.g. cp -R etc.)
- Pipes and redirects
- Shells
- Editing files
- File permissions: chmod

I (LP) prepared a tutorial-style intro to the command-line when helping Remco out, some time ago, at http://lpmacpro/wiki/Tutorial_1_-_20110320_-_Basic_Command-Line - may be of use? It took a couple of hours to go through, I recall.

(B) Data standards / File formats

- FASTA
- FASTQ – Peter's paper [1]
- GFF – specification [2] (mentioning .gtf too confusing?
- .gbk
- VCF
- tsv files (tab separated, including mentioning they're easy to load in Excel?)
- csv files (and why to avoid them in preference of TSV)
- txt (need to clarify difference from Word/RTF)
- word/excel (avoiding)
- Common file conversions/integration (thoughts on teaching a single way to do this, practically, vs. making aware of the multiple ways to do it? For example, EMBOSS seqret, or Biopython's SeqIO.convert) and concatenation of sequence files, using ACT, etc.
- MIAME/MIASE/other standardisation effort awareness?

(C) Databases / Gateways

- Expasy
- NCBI≈Genbank
- EBI
- Short Read Archive (SRA/ENA)
- UniProtKB/SWISSProt
- ArrayExpress and GEO (and BUGS - http://bugs.sgul.ac.uk/node/41 - for the bacteriologists)
- Ensembl/PhytoPath

enter search terms search

POPULAR RECENT COMMENTS ARCHIVES

New in Tablet 1.14.04.10
Tablet has just been updated to version 1.14.04.10. NEW:

New in CurlyWhirly 1.14.04.29
CurlyWhirly has just been updated to version 1.14.04.29.

New in Flapjack 1.14.03.17
Flapjack has just been updated to version 1.14.03.17. NEW: The

New in Flapjack 1.14.04.29
Flapjack has just been updated to version 1.14.04.29. NEW:

New in CurlyWhirly 1.14.04.24
CurlyWhirly has just been updated to version 1.14.04.24.

- PFam
- xBASE
- MISO/Sequence Ontology
- UCSC/VISTA? (too human, I think)
- Local databases/resources: local GBrowse (e.g. http://ppserver/gbrowse2) and Galaxy data libraries
- PDB/RCSB – .pdb/mmCIF; RasMOL, PyMOL, for viewing' SWISSPDBv for threading (too advanced for intro, I think [LP]).

(D) Bioinformatics Tools

- *Sequence comparison and visualization*: BLAST, CLUSTAL, T-COFFEE, SPIDEY, HMMER, FASTA, MAFFT, JALVIEW, Mauve, MUMmer, BLAT, IGV, ACT, Apollo etc. – worth stressing computational (and thence practical!) differences between pairwise and multiple sequence alignment, I think (LP).
- *Gene prediction*: AUGUSTUS, SNAP, PASA, Prodigal, Genewise, Glimmer, GeneMark – use of homology/conservation
- *ORF detection*: NCBI ORF finder (combine with gene prediction as concept?
- *Gene annotation*: GO, InterPro, RAST, BASys
- *Protein sequence annotation*: subcellular localization signal prediction, Pfam, Prosite, ACT, Artemis, Apollo
- EMBOSS
- *Protein/RNA structure prediction*: PSIPred, Vienna, Foldit, Phyre2

(E) General good practice

- Keeping an electronic lab book
- Naming files sensibly
- The importance of metadata
- Version control of code, documents and data (e.g. git – doesn't have to involve remote or local server)
- Scripting sequences of commands/a pipeline – the importance of reuse and reproducibility
- Avoiding unnecessary duplication, and propagation of different downstream versions of modified data (e.g. common raw data files in single location; same arrangement for key stage files – e.g. results of a particular analysis)
- Respecting and recording version numbers
- What questions can bioinformatics (not) answer?
- Approaching computational biology like any other experiment (BLAST examples good here?)
- Stressing the open – and very difficult! – nature of problems, and the widespread collaborative efforts to solve them might be worthwhile, e.g. CASP, CAFA, CAPRI, DREAM, etc.

Contact Us Site Map Cookie Policy Log in Return to top