Sequence analysis software

Jose Blanca
COMAV institute
bioinf.comav.upv.es

NGS answers

For all your NGS questions there's one place to go, the seqanswers forum.

http://seqanswers.com/

seqanswers is a community site with thousands of members.

In seqanswers there are news, questions and answers, discussions, etc.

Most of the content is organized around a forum and a wiki.

NGS software

… is moving fast

ryanhartkopf

NGS workhorses

The most used applications are usually Linux and Command Line based.

Preprocessing
Fastqc
Fastx
Prinseq
sff_extract
(seq_crumbs)

mappers
Bowtie2
BWA
Tophat
GMAP

assemblers
SOAPdenovo
CABOG
Trinity
Mira
Velvet

SAM processing
samtools
picard
GATK

formats
fastq
SAM/BAM (samtools)
vcf/bcf (vcftools)
Bed (bedtools)
GFF

seq_crumbs

Main features:

- Focused on fastx processing
- Unix design:
  - Small utilities
  - Pipe friendly
- Supports gzip, bzip2, and bgzf
- Pair aware.

Trimming:

- trim_by_case
- trim_edges
- trim_quality
- trim_blast_short (for oligonucleotides)

Filtering:

- filter_by_quality
- filter_by_length
- filter_by_name
- filter_by_blast
- filter_by_bowtie2
- filter_by_complexity

Pairing:

- split_matepairs (with a linker e.g. 454)
- pair_matcher
- interleave_pairs
- deinterleave_pairs

Other:

- sff_extract
- convert_format
- guess_seq_format
- seq_head
- sample_seqs
- change_case
- count_seqs

Similar software:

- PRINSEQ
- fastx
- FastqMcf

NGS GUI applications

Galaxy:

- Web frontend on top of the standard applications (bwa, samtools, etc.)
- Somewhat clunky interface, but easy once you know what to do.
- Free software
- The computation of large files in the Galaxy server can be difficult.
- A local server can be set up (informatician required).

Commercial software:

- Very easy to use.
- Windows, Linux and Mac OS versions.
- Costly and proprietary.
- CLC genomic workbench and LaserGene are examples.

NGS pipelines

The pipelines are tailored to an specific set of problems.

Usually Linux and with a Command Line Interface.

Examples:

- Tophat/cufflinks.
  - NGS based expression quantification.
- Gigabayes.
  - SNP calling.
- ngs_backbone.
  - SNP calling and functional annotation.

Jose Blanca
COMAV institute
bioinf.comav.upv.es