Galaxy Workshop

An introduction to Galaxy

"Mapping Reads with Galaxy"

IBERS

10 December 2015

https://galaxy.ibers.aber.ac.uk

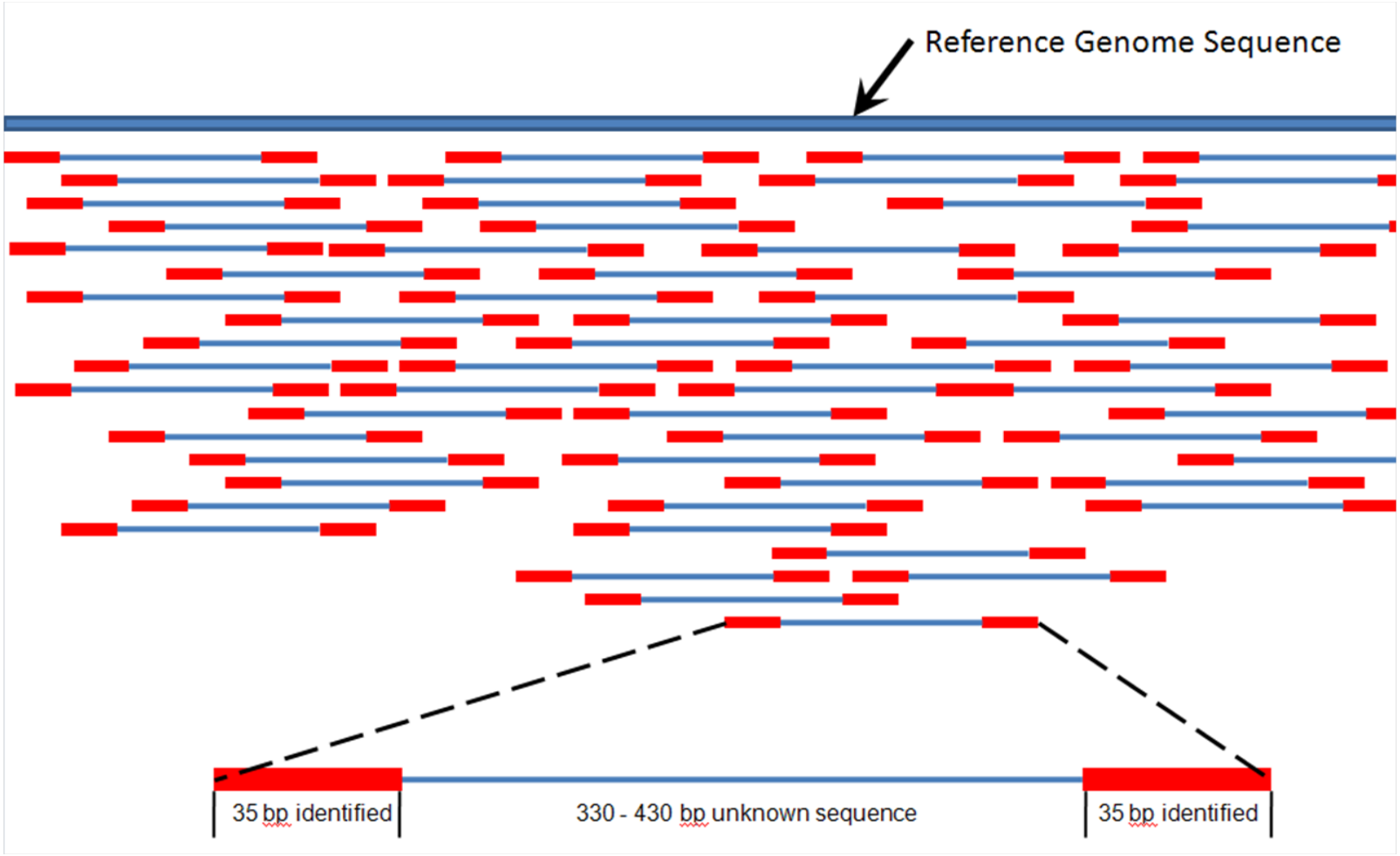
Martin Vickers

Vasilis Lenis

Goals

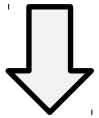
- Introduce to Galaxy
- Introduce to Reads Mapping
- Hands-on experience:
  - Upload data on Galaxy repository
  - Perform bioinformatics analysis with Galaxy
  - Save data
  - Visualize the results

Mapping

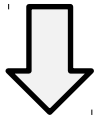

Reference Genome Sequence
35 bp identified
330 - 430 bp unknown sequence
35 bp identified

Workflow

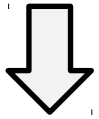
Sequence reads (**fastq** format)



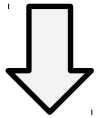
Reads mapping to mtDNA



Sorting **bam** file



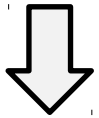
Calculate and plot insert size



Visualize the mapping

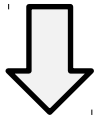
Workflow

Sequence reads (**fastq** format)



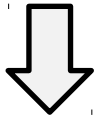
BWA-MEM

Reads mapping to mtDNA



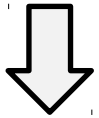
SAMtools

Sorting **bam** file



Picard

Calculate and plot insert size



IGV

Visualize the mapping

File Formats

- **Fastq:** A text-based format for storing both a biological sequence (usually nucleotide sequence) and its corresponding quality scores.

- **BAM:** A compressed binary version of the Sequence Alignment/Map (SAM) format

Tools

- **BWA – MEM:** A new alignment algorithm for aligning sequence reads or long query sequences against a large reference genome.
- **SAMTools:** A collection of utilities for manipulation on SAM and BAM files.
- **Picard:** A collection of tools for manipulating high-throughput sequencing data (HTS) data and formats.
- **IGV:** A high-performance visualization tool for interactive exploration of large, integrated genomic datasets.

Hands On

Mapping a paired–end reads dataset on sheep mitochondrial genome

https://galaxy.ibers.aber.ac.uk/u/mjv08/p/ibers-galaxy-tutorial-1