## Sequence duplication

<table>
<thead>
<tr>
<th>Type</th>
<th>Natural</th>
<th>Artificial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>Sample</td>
<td>Contaminant</td>
</tr>
<tr>
<td>Detection</td>
<td>Sequence identity</td>
<td>Alignment identity</td>
</tr>
</tbody>
</table>

Some tools/methods on the course webpage and course repository
Genomic Data Visualization

• Online websites data are also repositories → these run in a web browser: UCSC, Ensembl, GBrowse ...

• Downloadable applications with graphical user interface: IGV, IGB, BamView, Savant, Tablet, GenoViewer, MochiView, SeqMonk, inGAP ...

• Installable web applications: Anno-J, JBrowse
Towards a “better” genome browser

• Writing a better genome browser used to be a “rite of passage”

• There are probably hundreds of applications with various features/applications

• Genomic data visualization is a surprisingly complex matter – users’ needs diverge and can be mutually exclusive
Many are domain specific

Tools developed in a lab tend to suit the tasks performed in that environment:

- Genome variation → IGV, IGB, Tablet
- ChipSeq → MochiView (really nice tool)
- DNA Methylation → ChipMonk and SeqMonk
Visualizing genomes: techniques and challenges

Cydney B Nielsen¹, Michael Cantor², Inna Dubchak²,³, David Gordon⁴ & Ting Wang⁵

As our ability to generate sequencing data continues to increase, data analysis is replacing data generation as the rate-limiting step in genomics studies. Here we provide a guide to genomic data visualization tools that facilitate analysis tasks by enabling researchers to explore, interpret and manipulate their data, and in some cases perform on-the-fly computations. We will discuss graphical methods designed for the analysis of de novo sequencing assemblies and read alignments, genome browsing, and comparative genomics, highlighting the strengths and limitations of these approaches and the challenges ahead.

IGV: Integrative Genomics Viewer

Developed by the Broad Institute – focused on genetic variation studies
IGB (Ig-Bee) Integrated Genome Browser

Integrated Genome Browser
Visualization for genome-scale data

What is IGB?

The Integrated Genome Browser (IGB, pronounced Ig-Bee) is an interactive, zoomable, scrollable software program you can use to visualize and explore genome-scale data sets, such as tiling array data, next-generation sequencing results, genome annotations, microarray designs, and the sequence itself. IGB is implemented using the Java programming language and should run on any computer.

Seems to offers more options than IGV
It is a great tool though with a detailed user guide
Choosing a genome browser

- Data for model organisms may be “pre-filled”

- Custom or less common type of data will need to be loaded manually (we will do this)

- Import your own genome if you are not using a standardized genome build
Import your genome into IGV

If you use a model organism with a well defined genomic build that IGV already knows about then you don’t need to these steps
Visualizing the BAM file

You will need to zoom in to see the data

This is paired view
Right click on the view to set options
Visualize annotations as a new track
One needs to become familiar with at least one genome browser!

This guide describes the Integrative Genomics Viewer (IGV).

- To start IGV, go to the IGV downloads page: [http://www.broadinstitute.org/igv/download](http://www.broadinstitute.org/igv/download).
- For a 10-minute hands-on introduction, see the Quick Start.

Look at a printer-friendly HTML version of the whole User Guide.

- To generate a PDF of the UG, look at the HTML of the whole UG, then Print it. The Print dialog should offer you the ability to print to PDF.
Useful features

• BAM paired end data support

• Supporting opening remote data (data on a webserver)
  
  – place some data on a web location

  – see http://bcc.bx.psu.edu/tmp/

  – IGV ➔ open URL:  http://bcc.bx.psu.edu/tmp/results.bam

  (this is a way to share data with other people)
Homework 18

• Create a custom genome in IGV using the yeast genome.

• Visualize the alignments that you produced for homework 17 (1 million reads from the yeast genome). Show a screenshot of your data that covers a genomic region.

• What is the theoretical base coverage (C)? What is the actually observed coverage?

• Find the highest and lowest observed (nonzero) coverage (hint: `samtools depth`), show a screenshot of these locations.