

1 Introduction to Integrated Genome Viewer (IGV)

1.1 A quick start guide

1.2 Introduction

Integrative Genome Viewer (IGV) allows you to visualise genomic datasets. This quick start guide will give you a brief overview of IGV, how to load data, navigate the genome and visualise your data. The IGV user guide is really useful and contains information on many more features than we have the chance go through in this quick start guide.

Integrative Genome Viewer (IGV)

Broad Institute and the Regents of the University of California

Download: <https://software.broadinstitute.org/software/igv/download>

User guide: <http://software.broadinstitute.org/software/igv/UserGuide>

1.3 Learning outcomes

By the end of this quick start guide you can expect to be able to:

- Index a reference genome for IGV
- Load a reference genome file into IGV
- Load gene annotations into IGV
- Load alignment files into IGV
- Navigate a genome in IGV

1.4 Authors

This tutorial was written by [Victoria Offord](#).

1.5 Prerequisites

This guide assumes that you have the following software or packages and their dependencies installed on your computer. The software or packages used in this guide may be updated from time to time so, we have also given you the version which was used when writing the guide.

Package name	Link for download/installation instructions	Version
samtools	https://github.com/samtools/samtools	1.6
IGV	(https://software.broadinstitute.org/software/igv/)	2.3.90

1.6 Indexing a reference genome for IGV

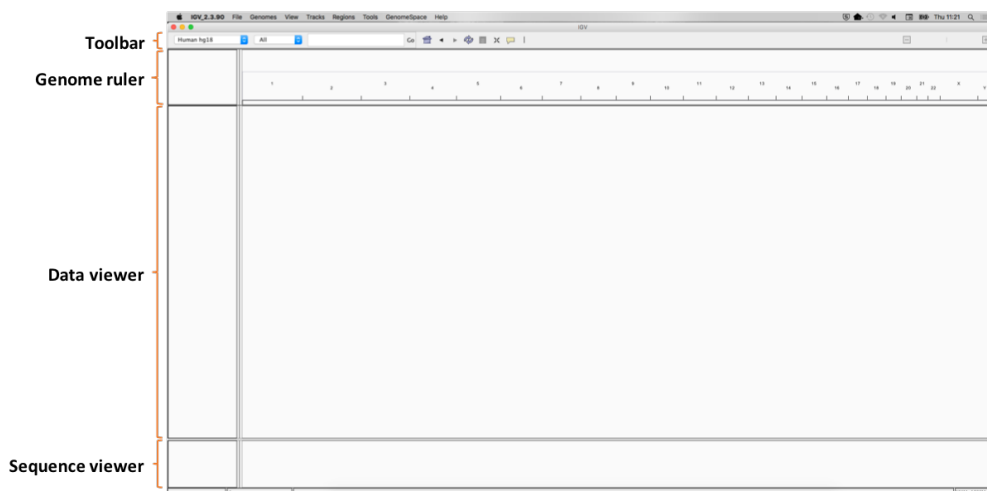
Before you begin, make sure you have an index for the reference genome which IGV will use to traverse the genome. You can do this using [samtools](#).

```
samtools faidx <your_genome_file.fa>
```

The resulting index file will have the extension **.fai** and must be in the same directory as the reference genome.

1.7 IGV main window

When you start IGV, it will open a main window. At the top of this window you have a **toolbar** and **genome ruler** for navigation. The largest area in the main window is the **data viewer** where your alignments, annotations and other data will be displayed. To do this, IGV uses horizontal rows called **tracks**. Finally, at the bottom, there is a sequence viewer which contains the base level information for your reference genome.

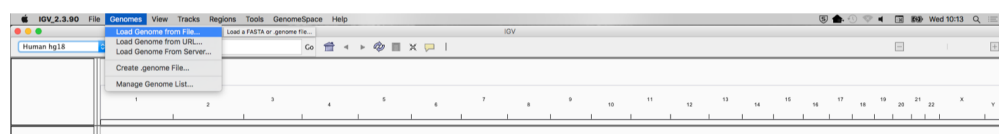


IGV - main window

1.8 Loading a reference genome

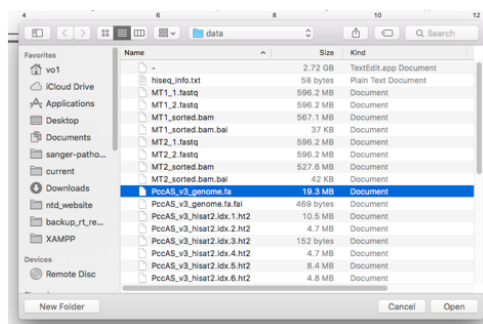
IGV provides several genomes which can be selected with the "*Genome drop-down box*" on the toolbar. However, your reference genome may not always be on this list. When your reference is not available, you will need to load it from a FASTA file.

To load a reference genome from file, go to "*Genomes -> Load Genome from File...*".



IGV - loading genome from file

Select the FASTA file containing the reference genome and click "*Open*".



IGV - loading genome from file

1.8.1 IGV toolbar and genome ruler

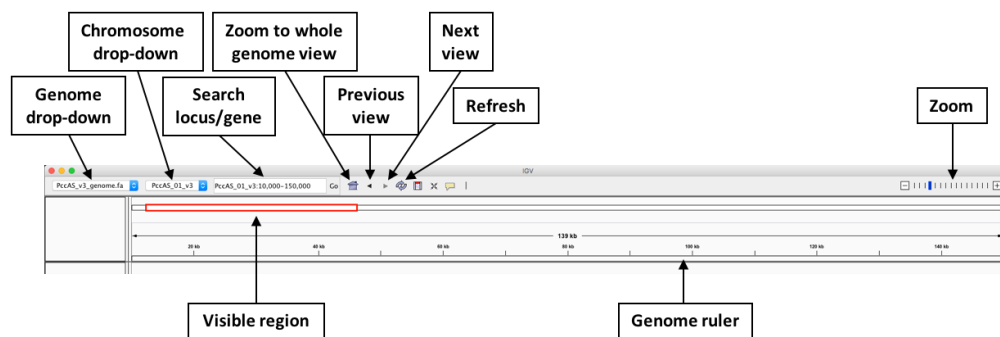
Once the genome has loaded, the chromosomes will be shown on the **genome ruler** with their names/numbers above. When a region is selected, a red box will appear. This represents the visible region of the genome.

Above the genome ruler is the **toolbar** which has a variety of controls for navigating the genome:

- **Genome drop-down** - load a genome
- **Chromosome drop-down** - zoom to a chromosome
- **Search** - zoom to a chromosome, locus or gene

There are several other buttons which can be used to control the visible portion of the genome.

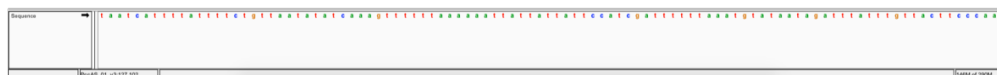
- **Whole genome** - zoom back out to whole genome view
- **Previous/next view** - move backward/forward through views (like the back/forward buttons in a web browser)
- **Refresh** - refresh the display
- **Zoom** - zooms in (+) / out (-) on a chromosome



IGV - toolbar and genome ruler

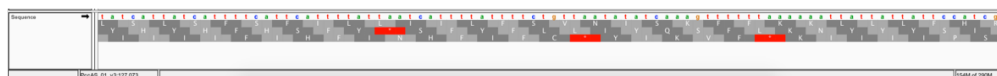
1.8.2 Sequence viewer

The **sequence viewer** shows the genome at the single nucleotide level. You won't be able to see the sequence until you are zoomed in. As you start to zoom in (+), you will see that each nucleotide is represented by a coloured bar (red=T, yellow=g, blue=c and green=a). This makes it easier to spot repetitive regions in the genome. Carry on zooming in (+) and you will see the individual nucleotides.



IGV - sequence viewer

If you right-click on "Sequence" at the left-hand side of the sequence viewer and click "Show translation", you will also see the amino acid sequence for the forward three reading frames.



IGV - sequence translation

You can also see the reverse three reading frames by right-clicking on the track and selecting "Flip strand".

Note: at the bottom right of the main window is the amount of memory available to IGV and how much of this it is currently using - always keep a wary eye on this!!

1.9 Loading gene annotations

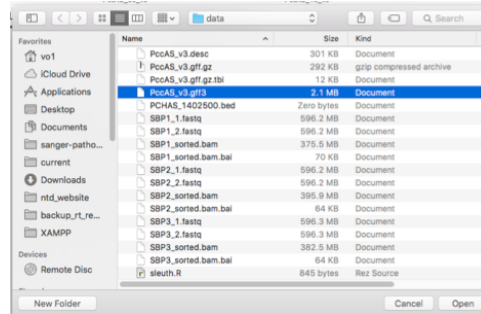
In addition to your genome, you will probably want to load an annotation file that contains information such as gene locations and gene structures (e.g. introns/exons/CDS).

To load a GFF file containing annotations, go to "File -> Load from File...".



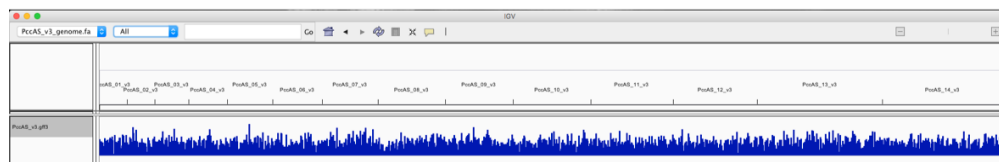
IGV - loading annotation

Select the annotation file and click "Open".



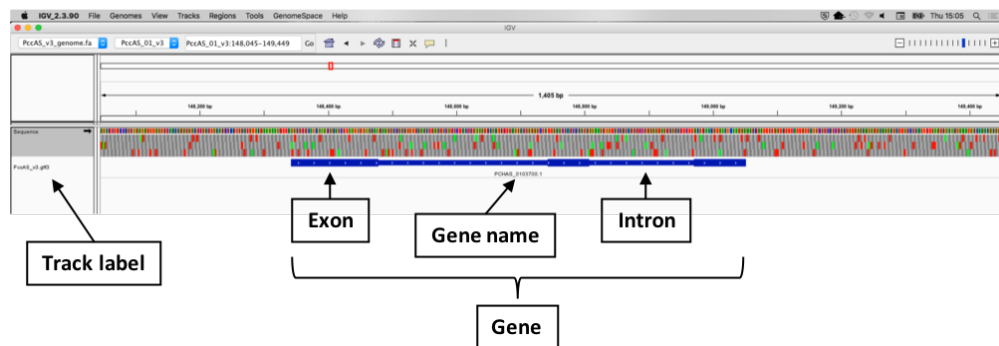
IGV - loading annotation

This will load the **annotation track**. At the genome level, you will see this shown as a density track for the associated annotation. On the left you will see the track label which is the name of the file you just loaded. You can change this label to something more recognisable by right clicking on the label and selecting "Rename Track".



IGV - annotation density

As you zoom in (+), you will start to be able to see the individual genes (shown in blue).

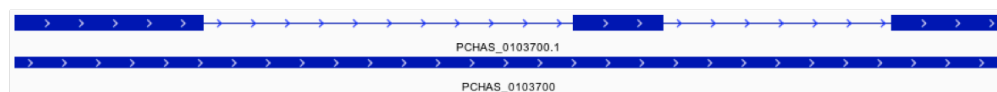


IGV - annotation track

1.9.1 Gene structure

Genes are represented in blue as boxes (exonic regions) and lines (intronic regions). The arrows indicate the strand of the direction in which the gene will be transcribed. The box height indicates whether the region is a coding sequence (taller) or untranslated region (thinner).

For a clearer view of the gene structure, right click on the annotation track and click "Expanded".



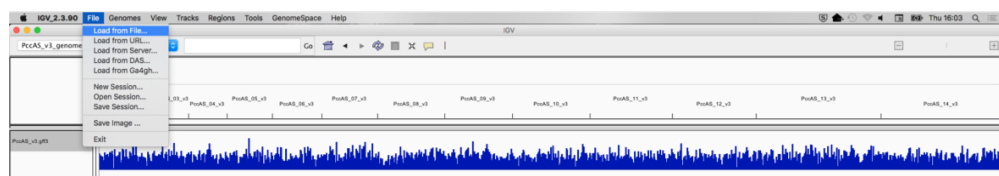
IGV - expanded annotation

Now you will see the annotated isoforms and can more clearly see the arrows that indicate which strand the gene is on. If you zoom in further, you will also see the amino acid sequence superimposed onto the exons.

1.10 Loading alignment files

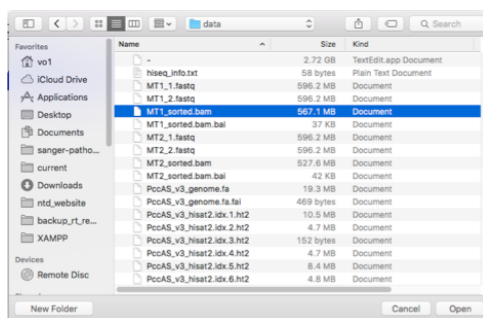
IGV can be used to visualise many different types of data, including read alignments. Each time you load an alignment file it will be added to the **data viewer** as a new major **track**.

To load a read alignment file, go to "File -> Load from File...".



IGV - loading alignment from file

Select a sorted BAM file and click "Open".

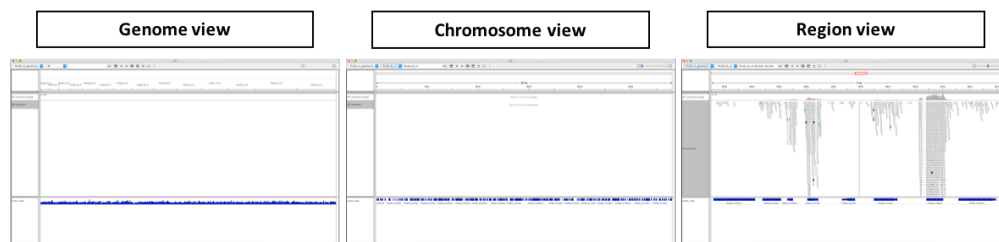


IGV - loading alignment from file

Note: BAM files and their corresponding index files must be in the same directory for IGV to load them properly.

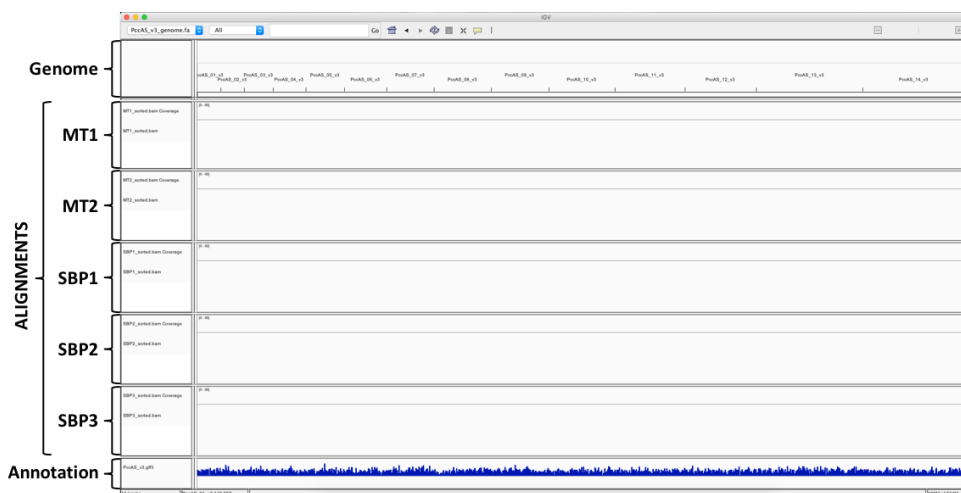
For each read alignment, a major track will appear containing two minor tracks for that sample: coverage statistics and read alignments. For the total number of visible tracks, see the bottom left of main window.

At the genome level, there will be no coverage plot or read alignments visible. At the chromosome level, there are two messages displayed: Zoom in to see coverage/alignments. Finally, once you have zoomed in (+) you will see a density plot in the coverage track and your read alignments.



IGV - alignment views

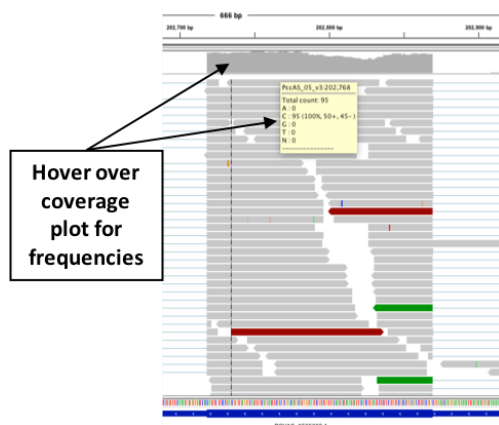
You can open more than one alignment file. Each alignment file will be loaded into a new track with its coverage statistics and read alignments. However, make sure you keep an eye on the memory usage in the bottom right corner or IGV may crash!



IGV - all alignments

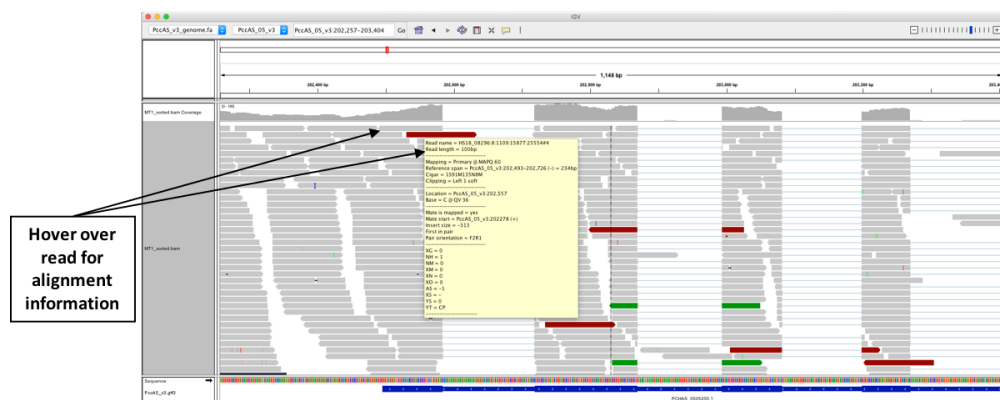
1.10.1 Visualising alignments

Coverage information When zoomed in to view a region, you can get alignment information for each position in the genome by hovering over the coverage track. This will open a yellow box which tells you the total number of reads mapped at that position, a breakdown of the mapped nucleotide frequencies and the number of reads mapping in a forward/reverse orientation. In our example, 95 reads mapped, 50 forward and 45 reverse, all of which called A at position 202,768 on chromosome PccAS_05_v3.



IGV - coverage information

Viewing individual read alignment information Reads are represented by grey or transparent/white bars which are stacked together where they align to the reference genome. Reads are pointed to indicate the orientation in which they mapped i.e. on the forward or reverse strand. Hovering over an individual read will display information about its alignment.



IGV - read information

Mismatches occur where the nucleotide in the aligned read is not the same as the nucleotide in that position on the reference genome. A mismatch is indicated by a coloured bar at the relevant position on the read. The colour of the bar represents the mismatched base in the read (red=T, yellow=G, blue=C and green=A).



IGV - mismatch

For more information about how reads are coloured and what this means, see the [IGV user guide](#).

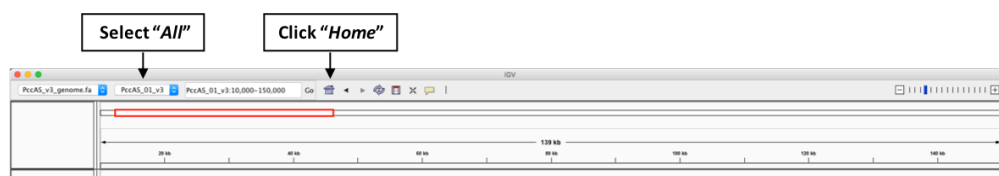
1.11 Navigating the genome

1.11.1 Whole genome view

In IGV you can navigate through different levels of visualisation, from the whole genome, all the way down to a base level resolution.

To return to the whole genome view:

- Select "All" from the chromosome drop-down
- Click the "Home" button

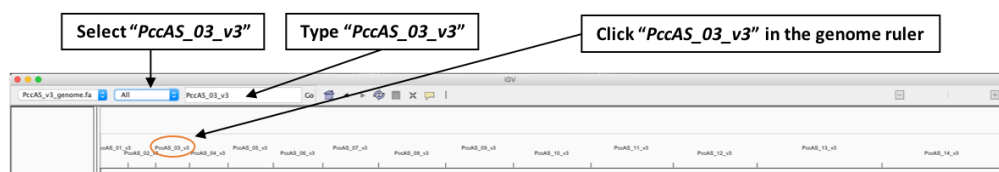


IGV - whole genome view

1.11.2 Chromosome view

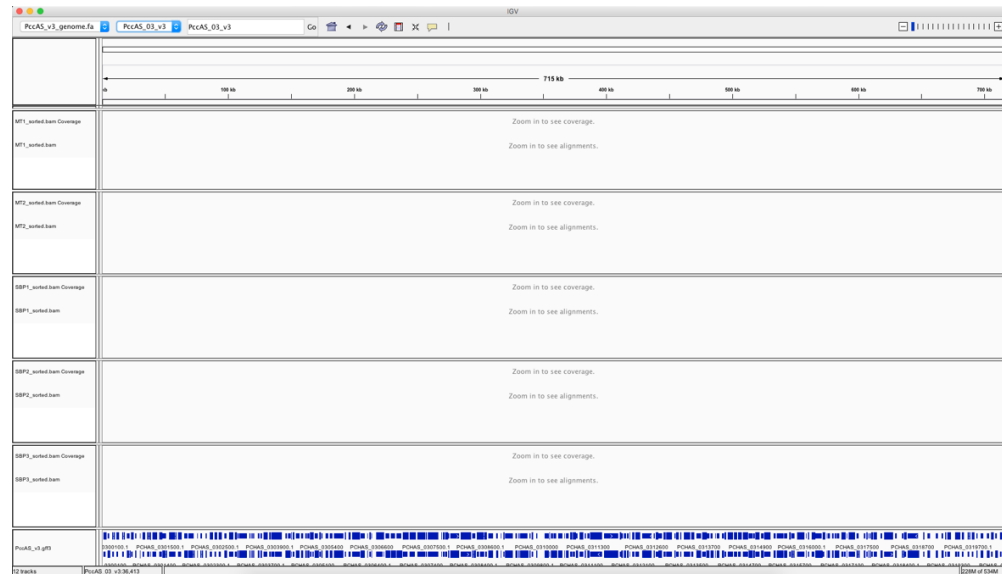
You can also view each of the chromosomes individually. For example, to view PccAS_03_v3 you can:

- Select "PccAS_03_v3" from the chromosome drop-down
- Type "PccAS_03_v3" into the search box
- Click on "PccAS_03_v3" on the genome ruler



IGV - chromosome view

In the chromosome view, the alignment track has changed from a density plot to showing individual gene, the genome ruler is now showing co-ordinates instead of chromosome name/numbers. More importantly, the alignment tracks are saying that to see coverage or read alignments we need to zoom in further.

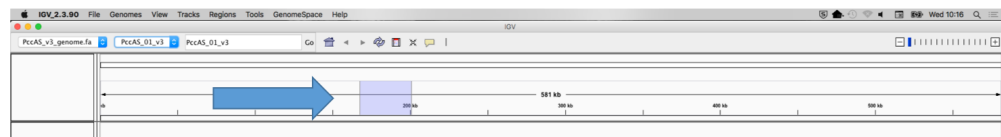


IGV - chromosome view

1.11.3 Region view

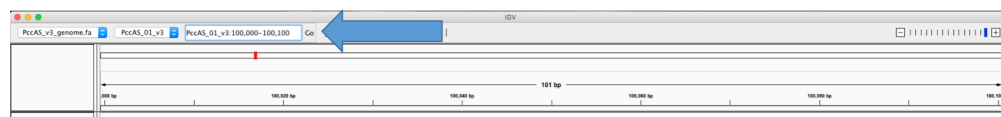
There are several ways to continue to zoom in and view specific regions or base level information.

Select region If you don't know the specific co-ordinates of the region you want to look at, you can click and drag to select a region on the genome toolbar.



IGV - select region

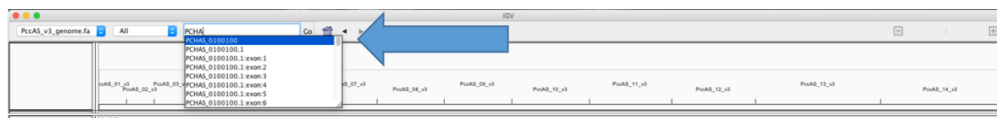
Jump to region If you know the co-ordinates of the region you want to view, you can enter them into the "Search" and click "Go". The format is chromosome:start-stop. For example, to view from 100,000 to 100,100 on PccAS_01_v3, you would enter PccAS_01_v3:100,000-100,100 in the search box.



IGV - search region

Note: the visible region of the chromosome is indicated by the red box on the genome ruler.

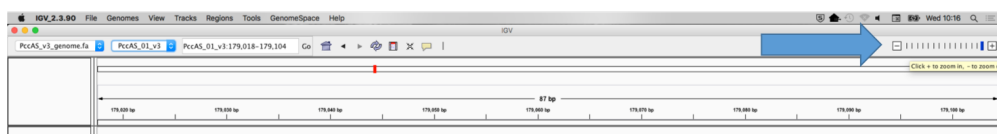
Jump to gene or locus Alternatively, if you know the name of the gene you want to view and you have loaded an annotation file, you can enter the gene name into the "Search" and click "Go". For example, to view PCHAS_0100100, you would enter PccAS_0100100 in the search box.



IGV - search gene

Note: the search box will try to help you by listing options to autocomplete the search box.

Zooming in and out You can zoom in and out from each view by using the "+" and "-" buttons on the zoom control at the right-hand side of the toolbar. This will also work with the "+" and "-" keys on your keyboard.



IGV - zoom in/out

1.11.4 Navigating around the view

There are several ways you can move around the view:

- Left-click and hold on a track in the data viewer. Drag to move left or right.
- Move left right using arrow keys on your keyboard.
- Double click on a gene/feature in the annotation track to zoom in and center on that gene/feature.