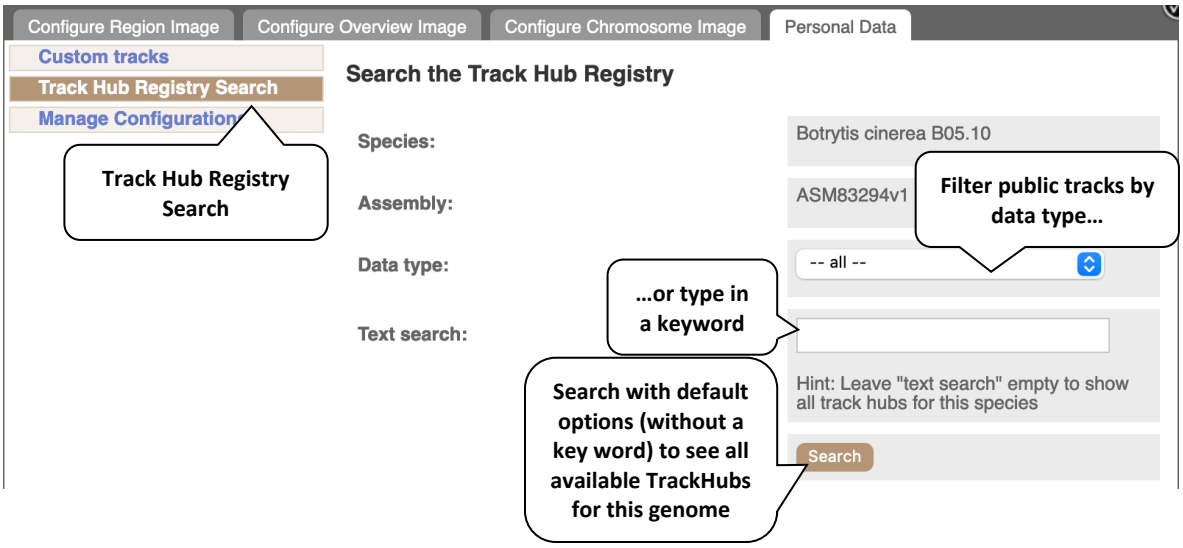
Exercise: Attaching Track Hubs

There are a number of publicly available datasets that are available to add onto views in Ensembl. You can find full lists of these at www.trackhubregistry.org. We're going to search and add these files from within Ensembl.

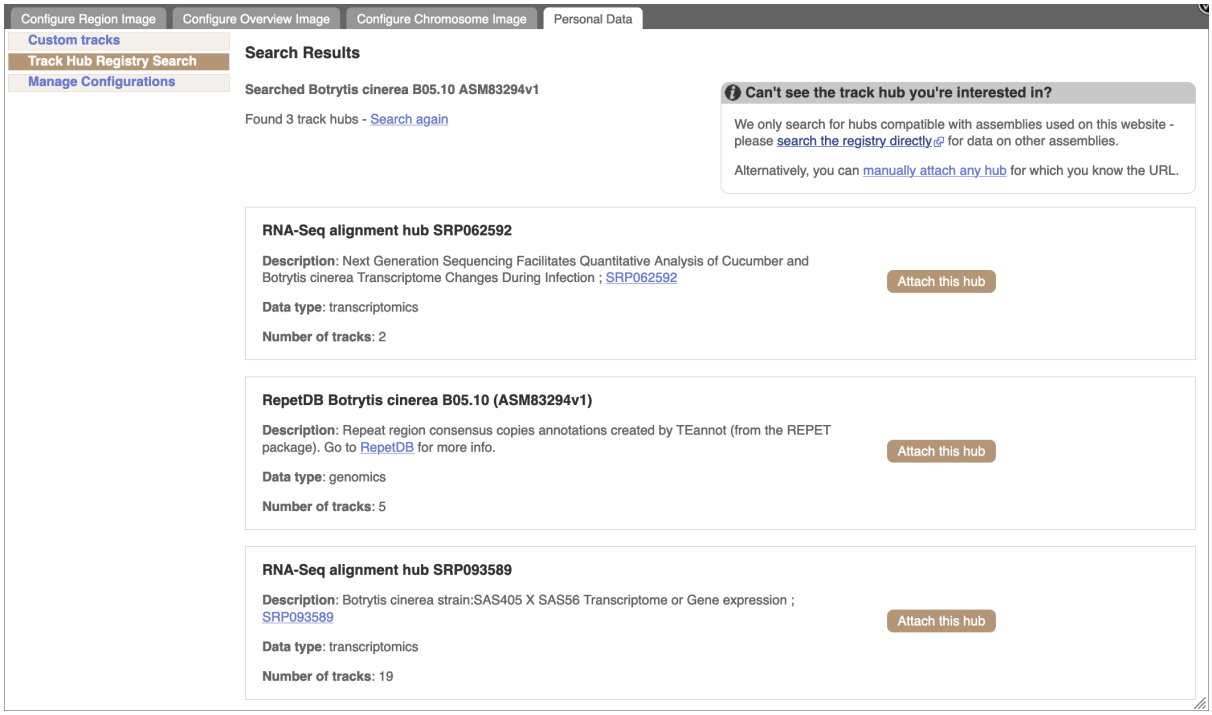
Go to fungi.ensembl.org and search for the region 6:1854110-1894000 in the species Botrytis cinerea B05.10.

This will take you directly to the Region in Detail page in the location tab. Click on the

button, found just below the Configure this page button on the left. A pop-up will appear, click on Track Hub Registry Search on the left-hand navigation panel.



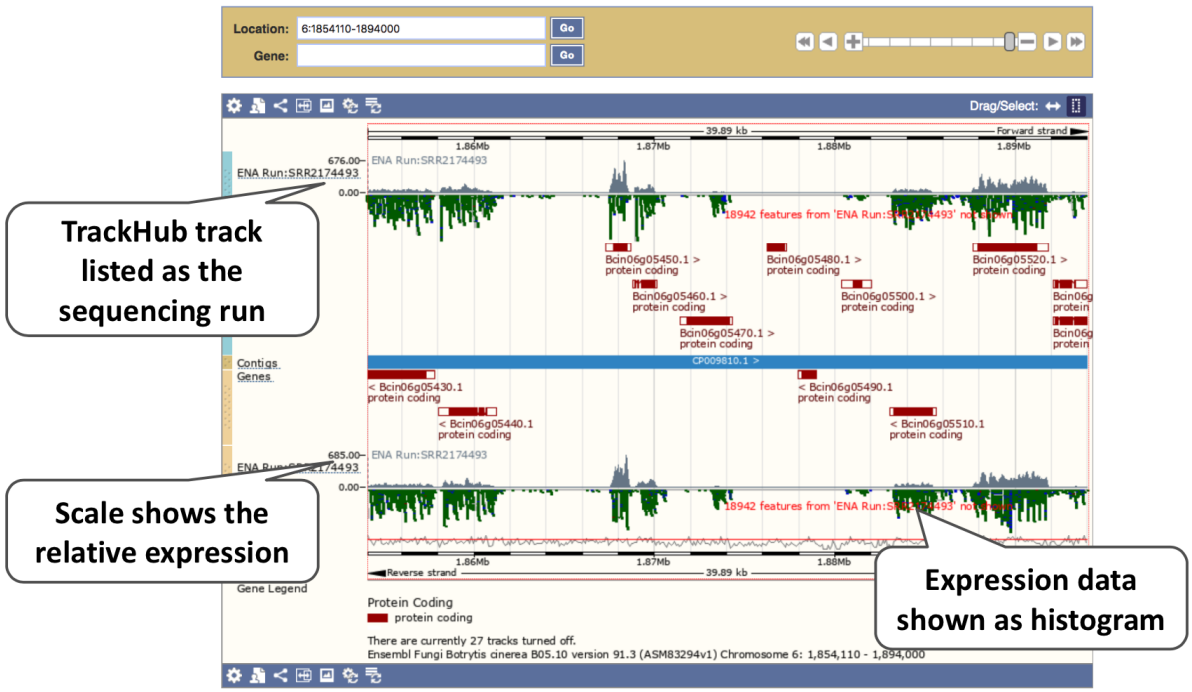
Just click Search with no options selected.



There are three available TrackHubs for this assembly.

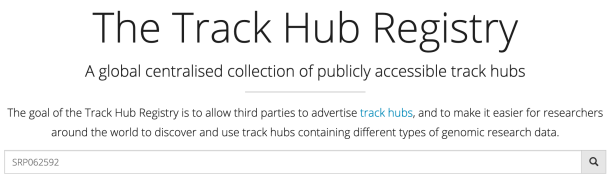
Choose the **RNA-Seq alignment hub SRP062592** by clicking on the 'Attach this hub' button. It is a next generation sequencing quantitative analysis of cucumber and *Botrytis cinerea* transcriptome changes during Infection. Close the pop-up window.

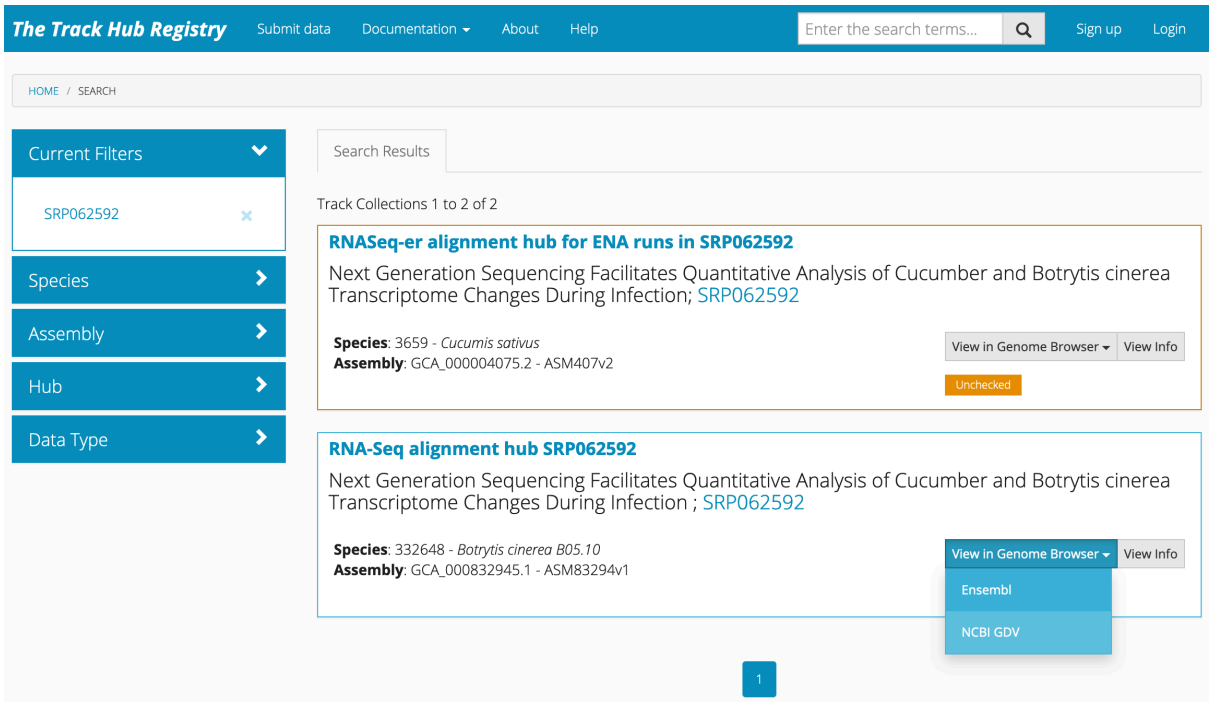
The TrackHub should now load and appear on the most-detailed image at the bottom of the Region in Detail page.



If you zoom in further you can see a more detailed representation of the data.

(a) Go to www.trackhubregistry.org and search for SRP062592. Can you jump to Ensembl directly from the Track Hub Registry page?





If you have your own files, or know a file you want to attach that is not present on the TrackHub registry, you can also attach these. There are two ways to do this, either by URL or by file upload.

Larger files, such as BAM files generated by NGS, need to be attached as remote files by URL. There are some BAM files for *Schizosaccharomyces pombe* available at: ftp://ftp.ensemblgenomes.org/pub/misc_data/bam/fungi/Spom/

Let's take a look at that URL.
NOTE: Many internet browsers have recently dropped support for FTP, including the latest Firefox and Google Chrome versions. Firefox v87.0 still contains built-in FTP implementation. If you struggle to open the FTP site, try the HTTP version: http://ftp.ebi.ac.uk/ensemblgenomes/pub/misc_data/bam/fungi/Spom/

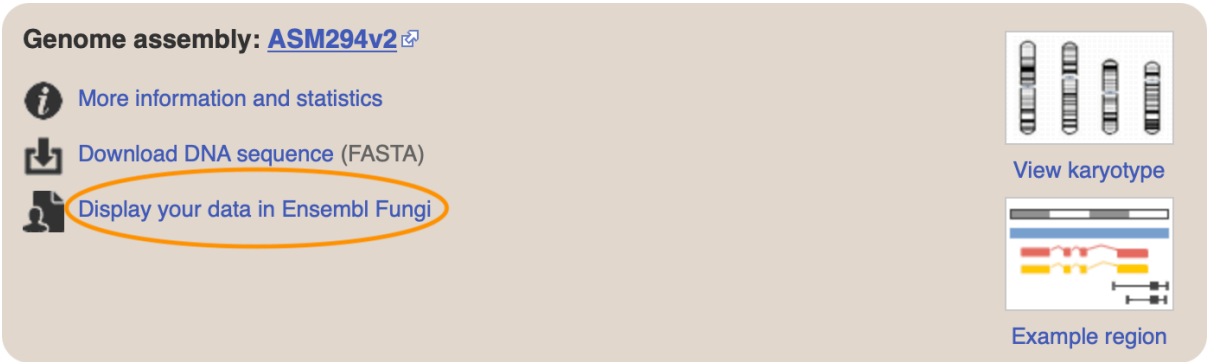
Index of /pub/misc_data/bam/fungi/Spom/

[parent directory]

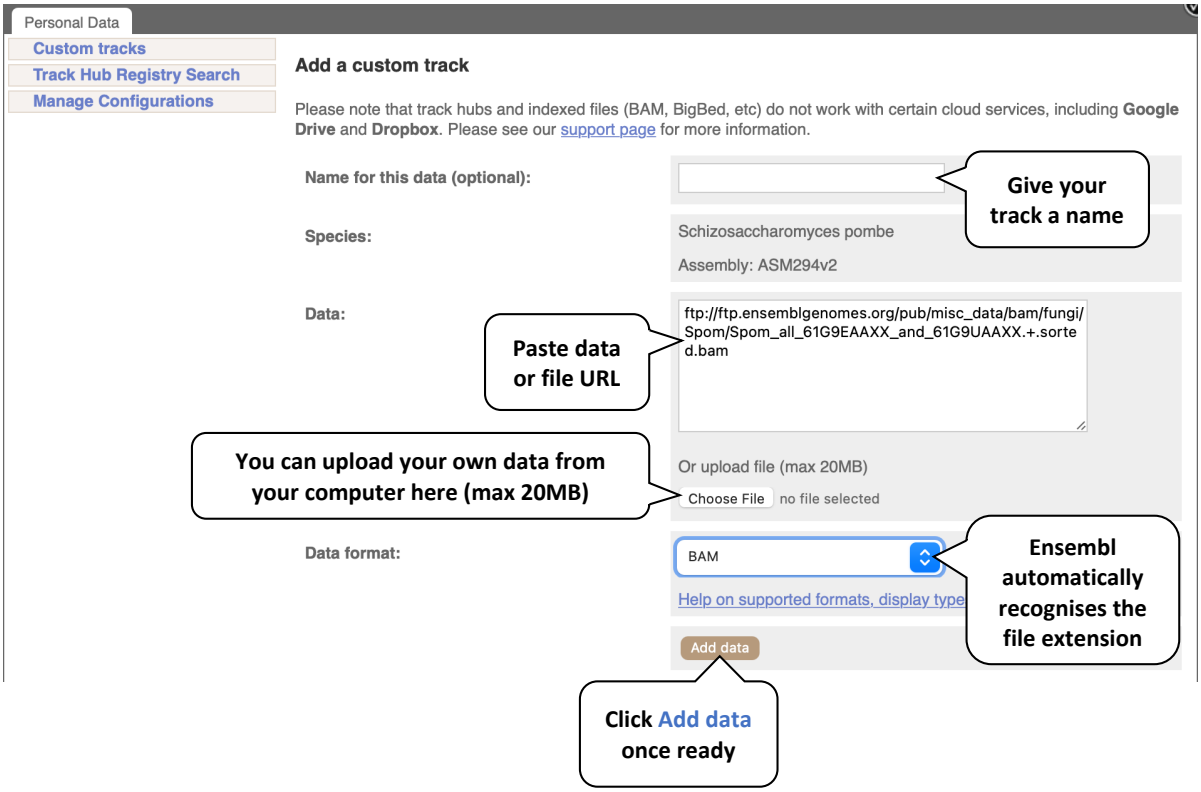
Name	Size	Date Modified
Spom_all_61G9EAAXX_and_61G9UAAXX.+.sorted.bam	3.3 GB	26/11/2014, 00:00:00
Spom_all_61G9EAAXX_and_61G9UAAXX.+.sorted.bam.bai	36.3 kB	26/11/2014, 00:00:00
Spom_all_61G9EAAXX_and_61G9UAAXX.-.sorted.bam	3.8 GB	26/11/2014, 00:00:00
Spom_all_61G9EAAXX_and_61G9UAAXX.-.sorted.bam.bai	36.8 kB	26/11/2014, 00:00:00

Here you can see two BAM files (.bam) with corresponding index files (.bam.bai). We're interested in the files Spom_all_61G9EAAXX_and_61G9UAAXX.+.sorted.bam and Spom_all_61G9EAAXX_and_61G9UAAXX.+.sorted.bam.bai. These files are the BAM file and the index file respectively. When attaching a BAM file to Ensembl Genomes, there must be an index file in the same folder.

From the Ensembl Fungi homepage, click on *Schizosaccharomyces pombe*, then on Display your data in Ensembl Fungi.



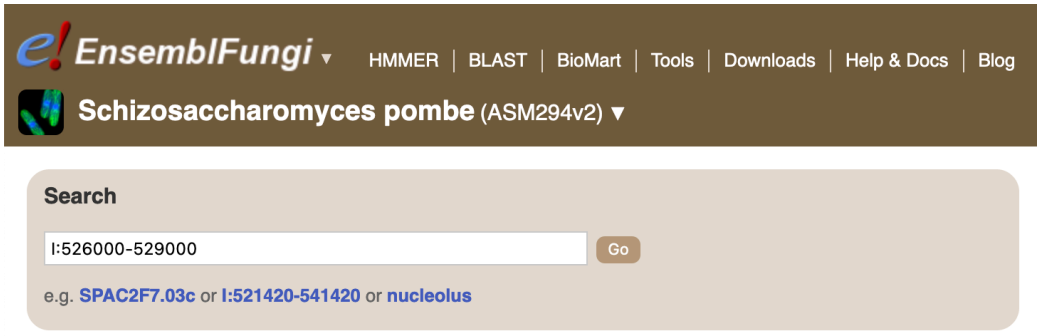
A menu will appear:

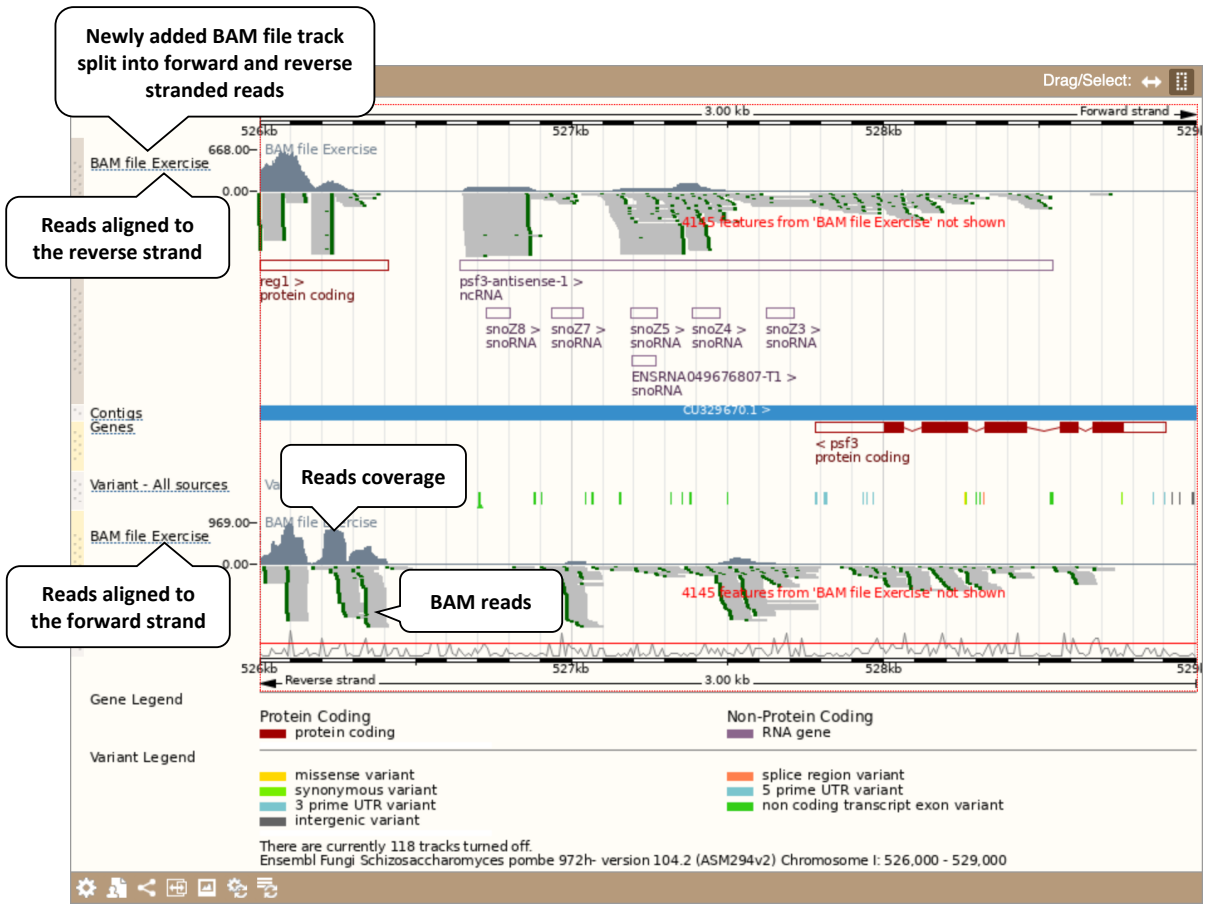


The interface detects file extensions if you upload or attach a file. If you want to upload a file just click on Choose file, choose the file and it should automatically detect the file type you have submitted.

If you have a URL, like the one we located earlier, paste in the URL of the BAM file itself (ftp://ftp.ensemblgenomes.org/pub/misc_data/bam/fungi/Spom/Spom_all_61G9EAAXX_and_61G9UAAXX.+.sorted.bam).

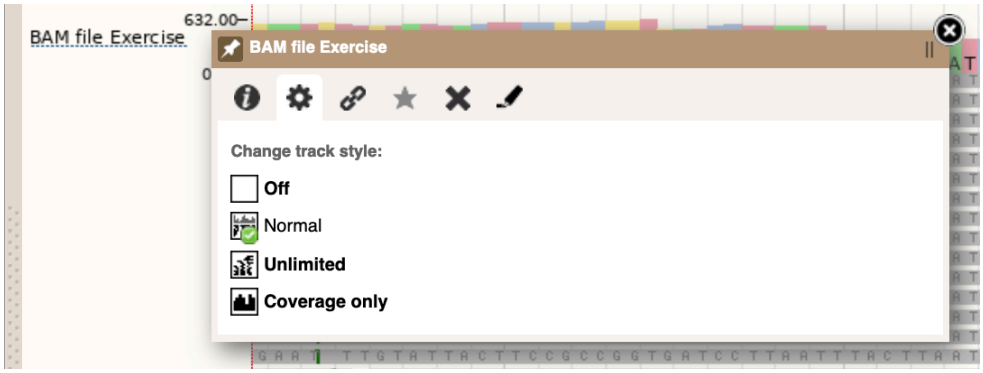
Since this is a file, the interface is able to detect the ".BAM" file extension and automatically labels the format as BAM. Click on Add data and close the menu. It may take a while to load as there is a lot of data (Firefox tends to be fast). Once the data has been uploaded, you'll get a thank you message. Close the window and jump to a Location Tab to see this data. Let's go to I:526000-529000.

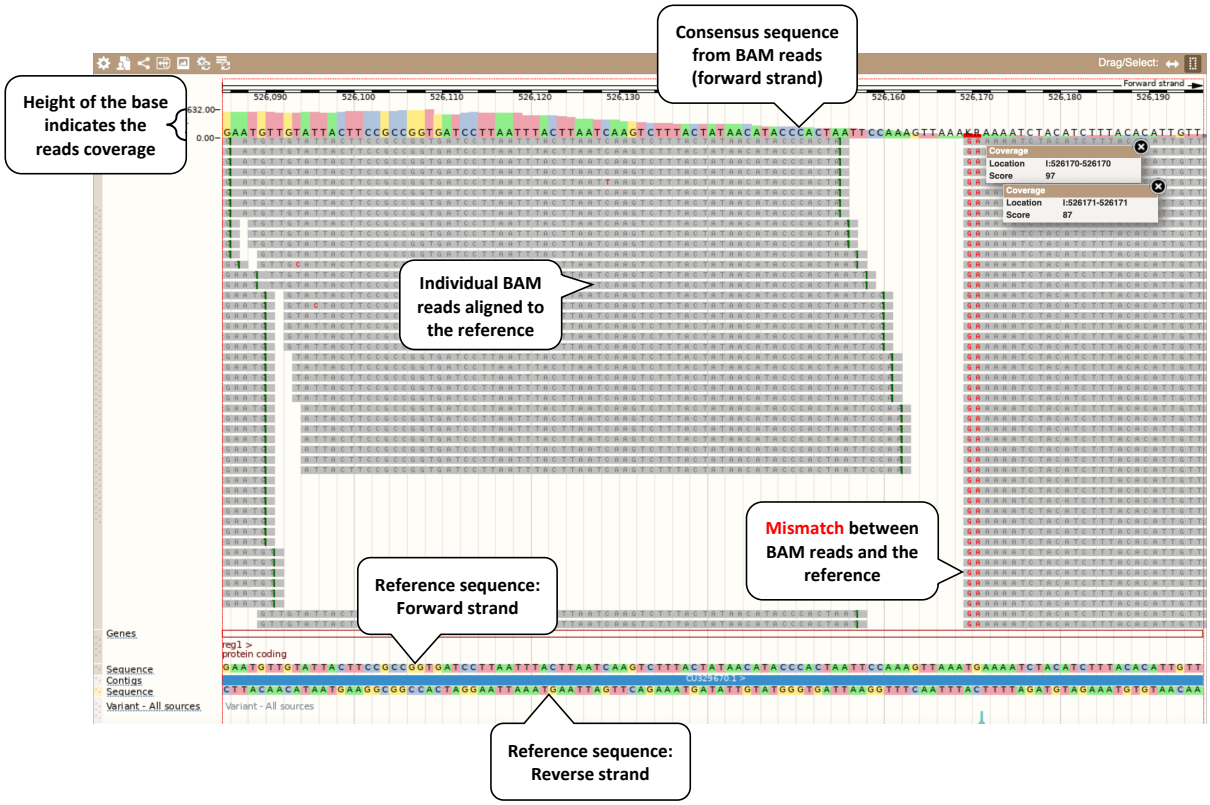




You can zoom in to see the sequence itself. Drag out boxes in the view to zoom in, until you see a sequence of individual reads, or jump to a 110 bp region: I:526085-526195.

(b) Change the track style of the newly added track to Unlimited (showing all reads). Can you spot a site called differently from the reference in our sample? What is its genomic position? What is the read coverage at this position on the forward strand? Would you consider it a real variant or an artefact?




Consensus sequence from BAM reads (forward strand)
Height of the base indicates the reads coverage
Individual BAM reads aligned to the reference
Mismatch between BAM reads and the reference
Reference sequence: Forward strand
Reference sequence: Reverse strand
Drag/Select:
Forward strand
526,090
526,100
526,110
526,120
526,130
526,160
526,170
526,180
526,190
632.00
0.00
Coverage
Location I:526170-526170
Score 97
Coverage
Location I:526171-526171
Score 87
Genes
reg1 >
protein coding
Sequence
Contigs
Sequence
Variant - All sources
Variant - All sources