**ReadMe\_IGV**

**1.Xiphophorus**

**2.Integrative Genome Viewer (IGV)**

**3.Start the IGV**

**4.Data import**

**5.Data download**

**6.Data processing**

**7.Common problems**

**1. *Xiphophorus***

If you are interested in *Xiphophorus*, please visit the *Xiphophorus* Genetic Stock Center website (www.xiphophorus.txstate.edu/) for more details.

**2. Integrative Genome Viewer(IGV)**

IGV is a genome browser that allows genome hosting, data sharing. It is flexible and enables customization. IGV requires java (1.6+), if you don’t have java, please download Java from www.java.com following Oracle's instructions. IGV is a genome browser made by Jim Robinson at Broad Institute (Nat Biotechnol. 2011 Jan; 29(1): 24–26). We adapted it to this website to host available *Xiphophorus* species genomes. To learn more about IGV, please visit [www.broadinstitute.org/igv/](http://www.broadinstitute.org/igv/).

**3. Start IGV**

The *Xiphophorus* genetic stock center currently hosts genomes for 4 *Xiphophorus* species: *X. hellerii* (sarabia strain), *X. maculatus* (Jp 163 B strain), *X. couchianus* and *X. montezumae*. By clicking “Genome Viewer V3”, a ".jnlp" file will be downloaded to your local computer. This .jnlp file configures the *Xiphophorus* genome and dataset for visualization. By running the .jnlp file, a genome viewer session that has a preloaded genome, gene model track, and genome alignment, as well as tissue specific RNA-Seq data will be initiated. If you have problems downloading or opening IGV, please refer to "common problems" at the end of the page, or contact *Xiphophorus* Genetic Stock Center staff/faculty.

**4. Data** **import**

IGV accepts multiple file formats, including BAM, BED, BEDGRAPH, BIGBED, BIGWIG, BIRSUITE FILES, BROADPEAK, CBS, CN, CYTOBAND, FASTA, GCT, GENEPRED, GFF, GTF, GISTIC, GOBY, GWAS, LGV, LOH, MAF(multiple alignment format), MAF(mutation annotation format),MUT, NARROWPEAK, PSL, RES, SAM, CHROM.SIZES, SEG, SNP, TAB, TDF, TRACK LINE, Type LINE, VCF, WIG and custom file formats. The ".jnlp" file comes with predefined genomes, gene models and data tracks. Users can import other datasets from workstation, user’s data server or website, or from XGSC data server:

**---Import data from workstation:**

To load your own data for visualization, click “File”, “Load from File” and browse to the directory storing data files, and click the file you want to import. Depends on the sized and type of the file to be imported, it can take a while for IGV to build a .index file for fast processing of files.

**---Import data from URL:**

If your data file is store in an URL (e.g., website, ftp, dropbox, AWS storage site), you can choose to import the dataset using a URL or custom data server. To import from URL, click “File”, “Load from URL” and type in the URL of the file to be visualized. To setup your own data server, please see “How to” at the end of this file.

**---Import data from XGSC data server:**

The XGSC shares data with the research community. Currently, XGSC shares a tissue specific RNA-Seq data as well as genomic sequence comparisons between *X. maculatus* and *X. hellerii, X. couchianus, X. montezumae,* Amazon molly, stickleback, spotted gar, cavefish, medaka, Asian arowana, human. To import these data, click “File”, “Load from Server”, then browse the data server in the pop-up window to the file to be visualized. If you experience trouble on getting access to the data server, you can choose to incorporate the data in a running IGV session without downloading them. Click "View"-> "Preference", select "Advanced" tab, and change the Data Registry URL to "https://viewer.xgsc.txstate.edu/dataregistry.txt" to link to the XGSC data server.

**5. Data download**

All data incorporated into IGV is publically available through the XGSC data sharing site: https://viewer.xgsc.txstate.edu/, include *X. hellerii,* *X. couchianus, X. montezumae* and *X. maculatus* chromosome assembly (.fasta) sequences; genome annotation files (.gtf/.gff); tissue specific RNA-Seq data (.bam) and genome sequence comparisons (.maf).

**Genome sequence:** https://viewer.xgsc.txstate.edu/data/

**Genome annotation:** https://viewer.xgsc.txstate.edu/data/gtf\_files/

**Tissue specific transcriptomic dataset:** https://viewer.xgsc.txstate.edu/SampleData/

**Genome Sequence Alignments:** https://viewer.xgsc.txstate.edu/GenomeAlignment/

**6. Data processing**

For data to be displayed properly, standardized data processing is required. The bioinformatics tools used to produce the dataset in the current version *Xiphophorus* Genome Viewer is listed below. But other common tools will also work. If user choose to use customized software/scripts, please follow proper file format for IGV.

**---RNA-Seq:**

**Build genome index:** bowtie2-build genome.fa database\_index

**Genome alignment:** tophat2 -o output database\_index pairend1.fastq pairend2.fastq

**Index alignment file:** samtools index output/accepted.bam

**Software:**

**tophat2** (<https://ccb.jhu.edu/software/tophat/index.shtml>; https://viewer.xgsc.txstate.edu/data/tools/tophat-2.0.13.Linux\_x86\_64.tar.gz)

**samtools** (<http://samtools.sourceforge.net/>; https://viewer.xgsc.txstate.edu/data/tools/samtools-1.3.1.tar.bz2)

**---Genome sequence alignment:**

**Genome alignment:** lastz --step=10 --notransition --nogapped --format=maf target.genome.fa query.genome.fa > output.maf

**Join overlapped alignment:** single\_cov2 input.maf > output.maf

**Sort sequence alignment:** maf\_sort input.maf contig.1 > 1.output.maf

 maf\_sort input.maf contig.2 > 2.output.maf

(maf\_sort only sorts alignment on a single subject contig/chromosome, therefore the maf\_sort need to be run on each subject contig/chromosome.)

**Concatenate sorted alignment:** cat \*.output.maf > concatenated.maf

**Label query contig/chromosome:** perl change.query.header\_general.alignment.pl input.maf index output.maf; or perl change.query.header\_detailed.alignment.pl input.maf index output.maf

**Software:**

**lastz** (<https://www.bx.psu.edu/~rsharris/lastz/>; <https://viewer.xgsc.txstate.edu/data/tools/>; https://viewer.xgsc.txstate.edu/data/tools/lastz-1.01.92.tar.gz)

**tba**(<https://www.bx.psu.edu/miller_lab/>; https://viewer.xgsc.txstate.edu/data/tools/multiz-tba.111208.tar.gz)

**change.query.header\_general.alignment.pl** (https://viewer.xgsc.txstate.edu/data/tools/) **change.query.header\_general.alignment.pl**

**change.query.header\_detailed.alignment.pl** (https://viewer.xgsc.txstate.edu/data/tools/ change.query.header\_detailed.alignment.pl)

**7. Common problems**

**Genome session cannot be downloaded:**

For some security setup of Internet browsers, the. jnlp file cannot be downloaded. To activate the download, put our website https://viewer.xgsc.txstate.edu into your safe list of the browser you are using.

**Java is out of date and .jnlp file cannot be launched:**

Update java to 1.6+ to have IGV properly launched. Java 1.7 is preferred. Multiple version of java can be installed in the operating system, make sure the right version is being used to open the .jnlp file. To check java version, you can either open java control panel to find the version number or in terminal (MacOS) or command prompt (windows) and type "java -version".

**Launch IGV is blocked by system.**

In java control panel, click "Security" and put both "http://igv/broadinstitute.org" and "https://viewer.xgsc.txstate.edu" in "exception site list".

If you run into any problems launching IGV, please contact XGSC staff, we will be glad to assist you.