

# NextGENe<sup>®</sup>

*Next Generation Sequencing Software for Biologists*

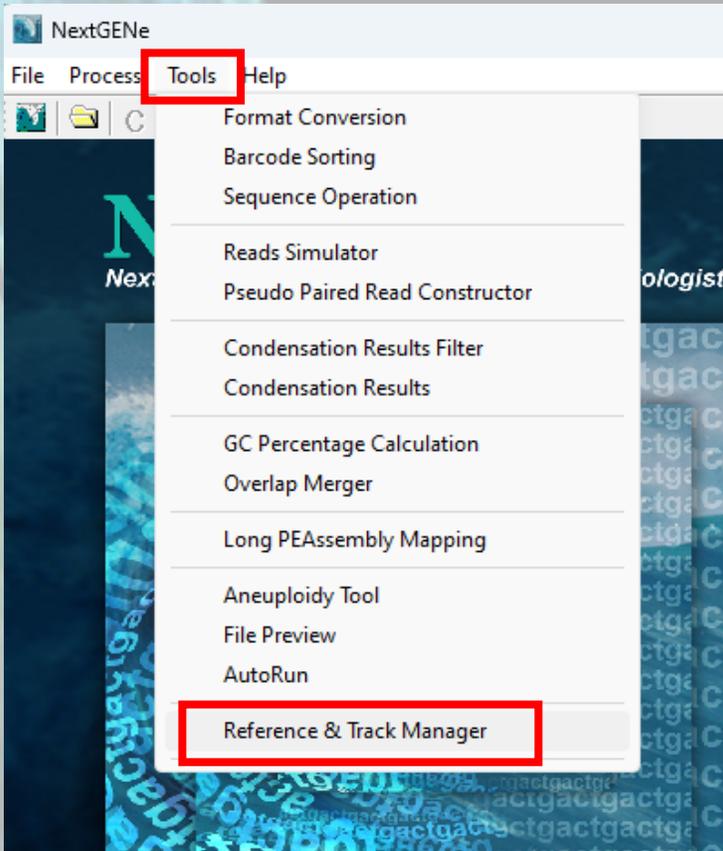
## Importing References



Depending upon your application you may need one of our Indexed references for NextGENe (for genomes larger than 250 Mbp), which may be downloaded and installed as follows.

The installation process of the reference is quite lengthy due to the data size.

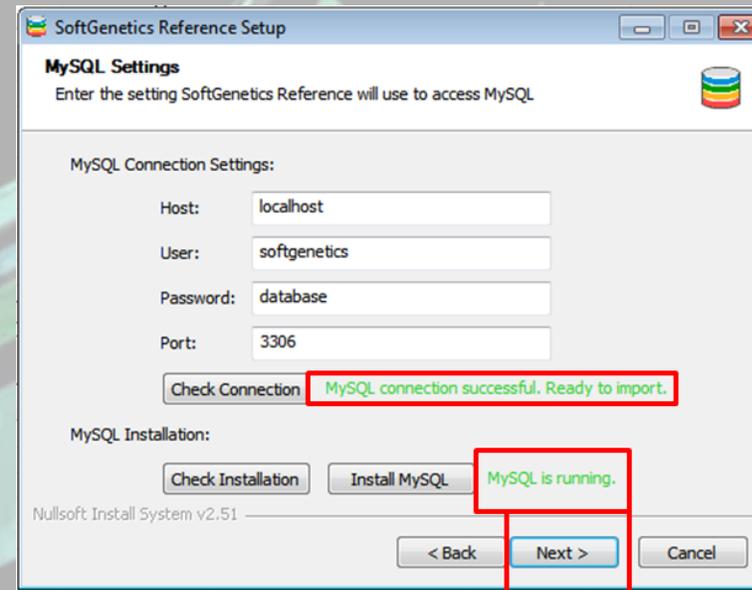
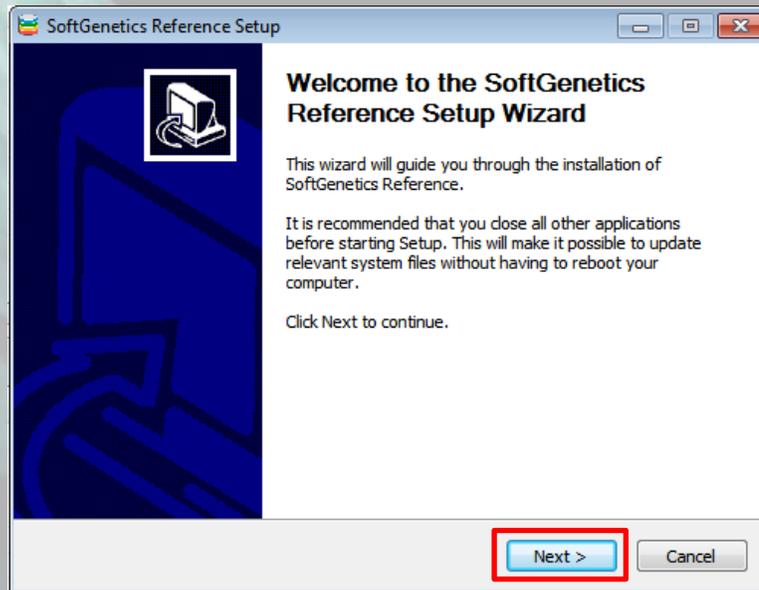
To obtain our pre-indexed reference genomes:



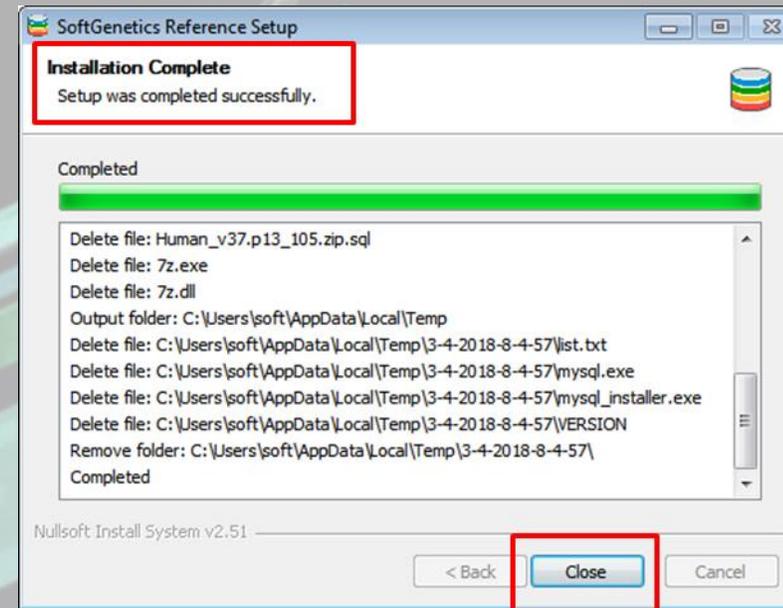
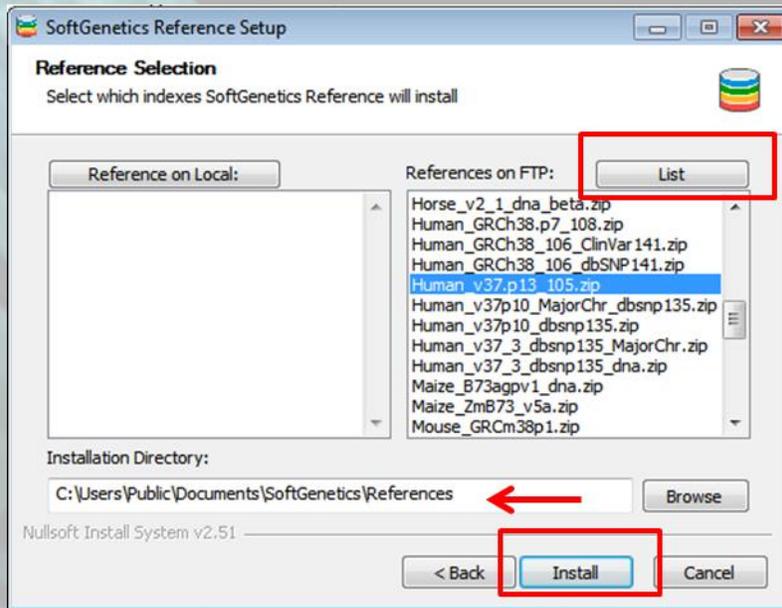
- Right-click on NextGENe and choose "Run as Administrator."
- Go to Tools > Reference and Track Manager
- Click "Import Reference."



- The NextGENe Reference Setup Wizard Opens
- Click "Next" on Welcome screen.
- Ensure that MySQL is configured properly. Next to Check Connection, it will state that it is ready for import. If it is ready to import, click "Next" to proceed to next step.
  - If connection failed because MySQL is not installed, click Install MySQL. After this is complete, click Close.
  - Click Check Connection and if prompted with MySQL Connection Successful, then continue with importing of reference genomes. Otherwise, contact [tech\\_support@softgenetics.com](mailto:tech_support@softgenetics.com).



- To view all reference genomes on our FTP server, click "List".
- Highlight the genomes you would like to import.
- Click "Install". When reference and its database are downloaded and imported properly, you will see "Completed".



# Not sure which Human Genome Reference to Import?

**v37.p13 has the most current annotation from the listed v37 references**

**Different annotation. Long name is GRCh37p5 annotation pipeline version 3. Released prior to v37.p10**

**v37p10 annotation released after v37\_3 and prior to v37p13**

**+mRNA to be used for RNA Fusion SV detection**

**Does Not Contain Unlocalized and Unplaced Sequences**

**Reference includes variant annotation from dbSNP b135**

**Reference Used By 1000 Genomes Project**

**SoftGenetics Reference Setup**  
Reference Selection  
Select which indexes SoftGenetics Reference will install

Reference on Local: [Empty]

References on FTP: [List]

- Human\_GRCh38\_106\_dbSNP141.zip
- Human\_RNA.zip
- Human\_v37.p13+mRNA.zip
- Human\_v37.p13\_105.zip
- Human\_v37\_3\_dbSNP135\_MajorChr.zip
- Human\_v37\_3\_dbSNP135\_dna.zip
- Human\_v37p10\_MajorChr\_dbSNP135.zip
- Human\_v37p10\_dbSNP135.zip
- Human\_v38\_p12+mRNA.zip
- Human\_g1k\_v37.zip
- Maize\_B73agp.v1\_dna.zip
- Maize\_ZmB73.v5a.zip

Installation Directory:  
C:\Users\Public\Documents\SoftGenetics\NextGENE\References [Browse]

Nullsoft Install System v2.51

< Back Install Cancel

# Do not see the genome you need listed?

## 1. Get the GCF Accession Number for the genome from NCBI

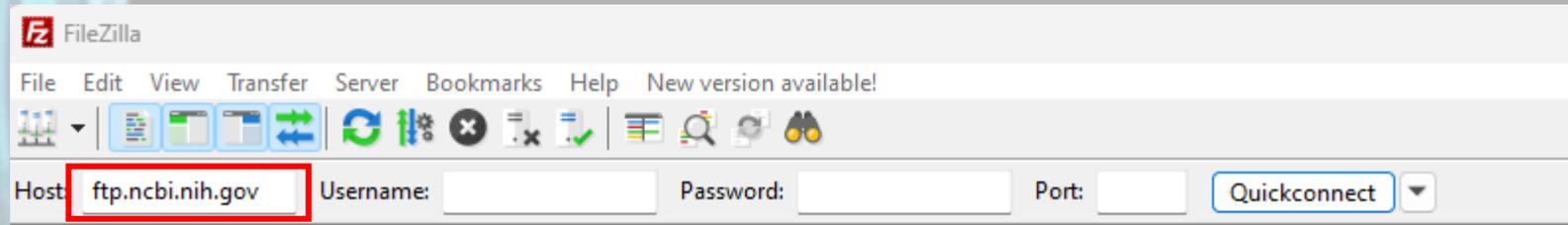
The screenshot shows the NCBI Genome browser interface. At the top, there is the NIH logo and the text "National Library of Medicine National Center for Biotechnology Information". Below this is a search bar labeled "Search NCBI ..." and a "Log in" button. The main navigation menu includes "NCBI Datasets", "Taxonomy", "Genome" (which is selected), "Gene", "Command-line tools", and "Documentation".

The "Genome" section is active, displaying the title "Genome" and a description: "Download a genome data package including genome, transcript and protein sequence, annotation and a data report". Below this is a "Selected taxa" field containing "Homo sapiens" and a "Filters" dropdown menu.

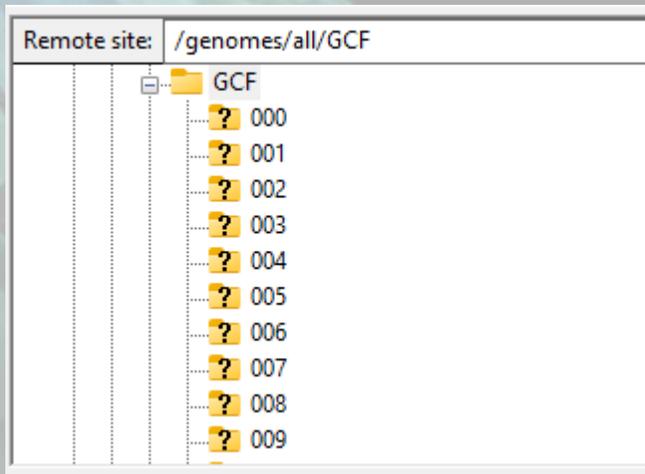
The main content area shows a table of genomes. The table has columns for "Download", "Select columns", "1,090 Genomes", "Rows per page" (set to 20), and "1-20 of 1,090". The table itself has columns for "Assembly", "GenBank", "RefSeq", "Scientific name", "Modifier", "Annotation", and "Action".

Assembly	GenBank	RefSeq	Scientific name	Modifier	Annotation	Action
<input type="checkbox"/> GRCh38.p14	GCA_000001405.29	<b>GCF_000001405.40</b>	Homo sapiens (human)		<a href="#">NCBI RefSeq</a>	⋮
<input type="checkbox"/> T2T-CHM13v2.0	GCA_009914755.4	GCF_009914755.1	Homo sapiens (human)		<a href="#">NCBI RefSeq</a>	⋮
<input type="checkbox"/> Q100 hg002v1.0.1.mat	GCA_018852615.2		Homo sapiens (human)	NA24385 (isolate)		⋮
<input type="checkbox"/> Q100 hg002v1.0.1.pat	GCA_018852605.2		Homo sapiens (human)	NA24385 (isolate)		⋮

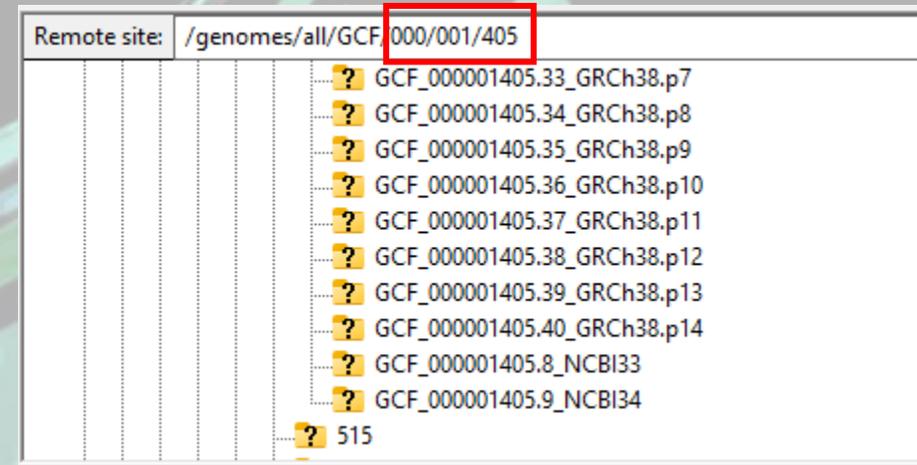
## 2. Use FTP client such as [FileZilla](#) to connect to NCBI's FTP



## 3. Navigate to genomes/all/GCF



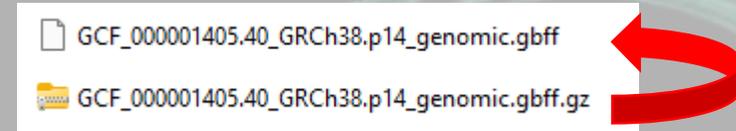
## 4. Navigate through the numbered folders according to the GCF accession number



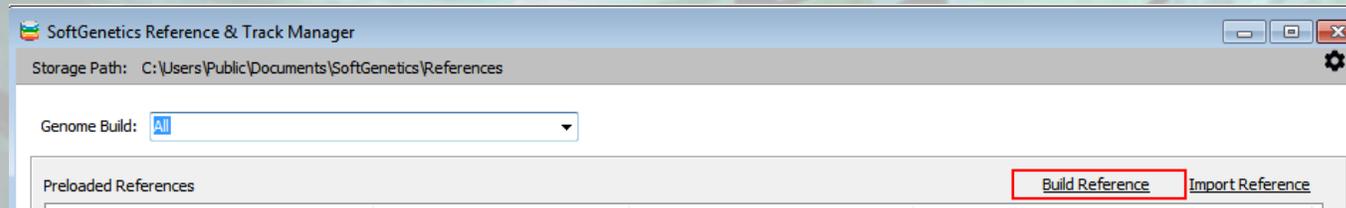
If you do not have an FTP client, open a web browser and go to <https://ftp.ncbi.nih.gov/genomes/all/GCF/> and proceed with step 4.

## 5. Download the \*\_genomic.gbff.gz file and extract (can use 7-zip)

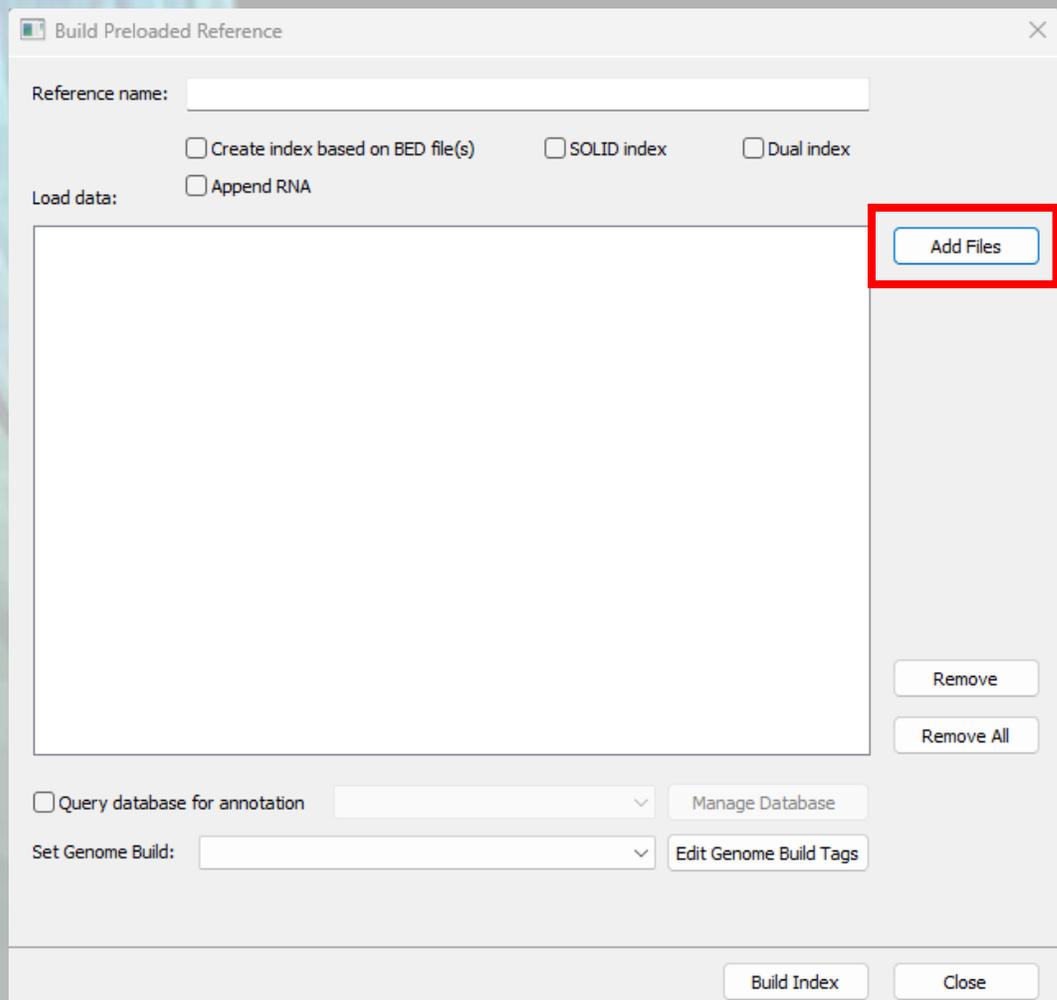
Filename	Filesize	Filetype	Last modified
GCF_000001405.40_GRCh38.p14_feature_count.txt.gz	1,477	Compressed...	10/11/2023
GCF_000001405.40_GRCh38.p14_feature_table.txt.gz	9,075,918	Compressed...	10/11/2023
GCF_000001405.40_GRCh38.p14_genomic.fna.gz	972,898,531	Compressed...	10/11/2023
<b>GCF_000001405.40_GRCh38.p14_genomic.gbff.gz</b>	<b>1,342,909,376</b>	<b>Compressed...</b>	<b>10/11/2023</b>
GCF_000001405.40_GRCh38.p14_genomic.gff.gz	77,767,547	Compressed...	10/11/2023
GCF_000001405.40_GRCh38.p14_genomic.gtf.gz	51,247,158	Compressed...	10/11/2023
GCF_000001405.40_GRCh38.p14_genomic_gaps.txt.gz	11,483	Compressed...	10/11/2023
GCF_000001405.40_GRCh38.p14_protein.faa.gz	28,374,292	Compressed...	10/11/2023
GCF_000001405.40_GRCh38.p14_protein.gbff.gz	166,240,044	Compressed...	10/11/2023



## 6. Open the Reference and Track Manager Tool (see page 2) and click “Build Reference”



## 7. Click Add Files



Build Preloaded Reference

Reference name:

Create index based on BED file(s)    SOLID index    Dual index

Load data:  Append RNA

**Add Files**

Remove

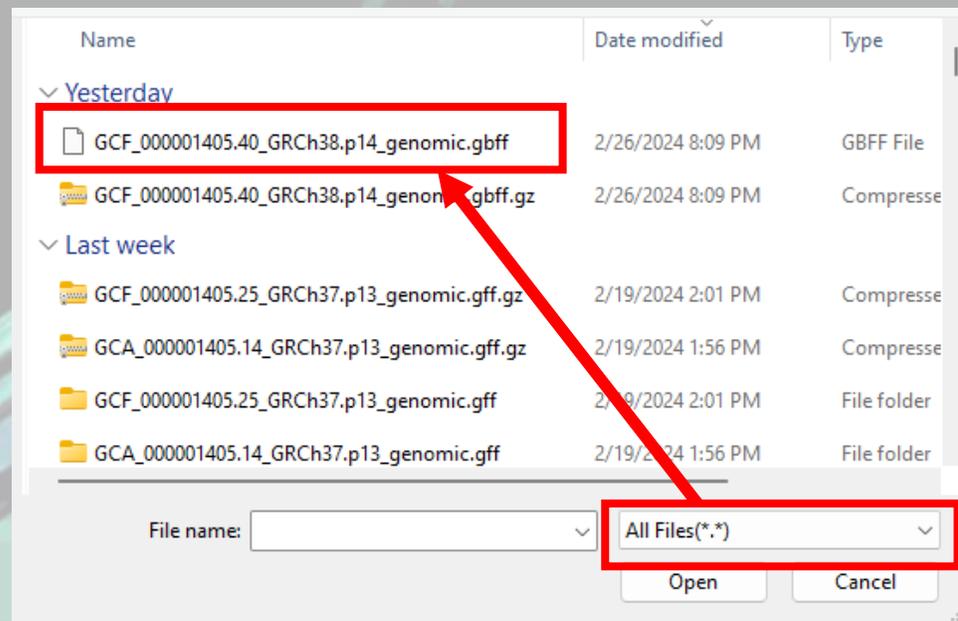
Remove All

Query database for annotation      Manage Database

Set Genome Build:    Edit Genome Build Tags

Build Index   Close

## 8. Use the file type dropdown to select “All Files” and then select the extracted .gbff file



Name	Date modified	Type
Yesterday		
<b>GCF_000001405.40_GRCh38.p14_genomic.gbff</b>	2/26/2024 8:09 PM	GBFF File
GCF_000001405.40_GRCh38.p14_genomic.gbff.gz	2/26/2024 8:09 PM	Compressed File
Last week		
GCF_000001405.25_GRCh37.p13_genomic.gff.gz	2/19/2024 2:01 PM	Compressed File
GCA_000001405.14_GRCh37.p13_genomic.gff.gz	2/19/2024 1:56 PM	Compressed File
GCF_000001405.25_GRCh37.p13_genomic.gff	2/19/2024 2:01 PM	File folder
GCA_000001405.14_GRCh37.p13_genomic.gff	2/19/2024 1:56 PM	File folder

File name:    **All Files (\*.\*)**

Open   Cancel

9. Give your reference a name

Build Preloaded Reference

Reference name:

Create index based on BED file(s)     SOLID index     Dual index

Load data:  Append RNA

Sequence/GBK Files:

C:\Users\caitlin\Downloads\GCF\_000001405.40\_GRCh38.p14\_genomic.gbff

Query database for annotation    Manage Database

Set Genome Build:     Edit Genome Build Tags

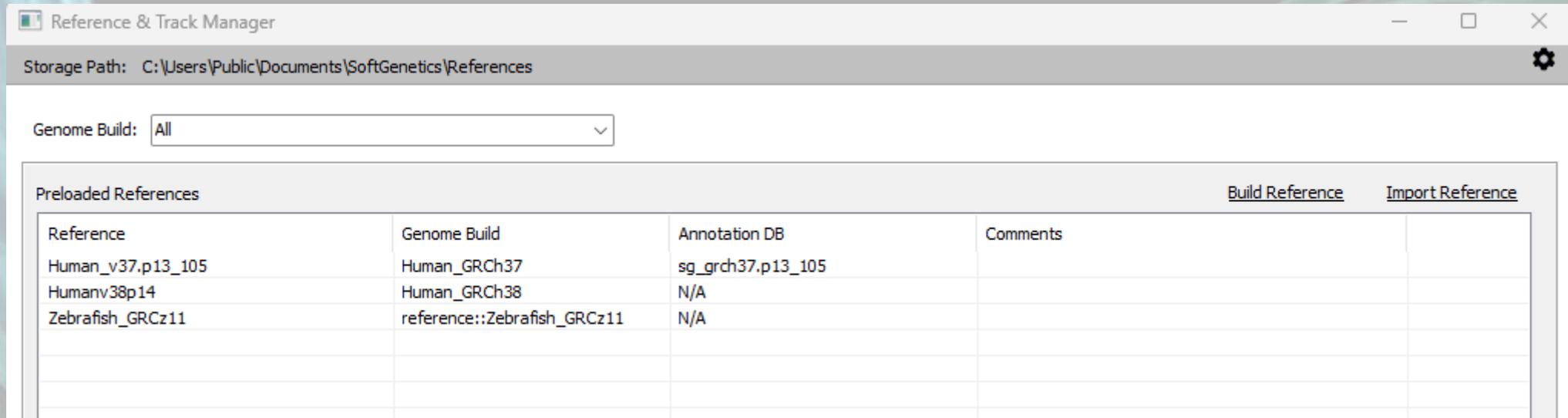
      

10. If applicable, select a genome build

11. Click "Build Index"

In the Reference & Track Manager Tool, set the Genome Build to “All” to display all imported references



The screenshot shows the 'Reference & Track Manager' application window. The title bar includes the application name and standard window controls. Below the title bar, the 'Storage Path' is set to 'C:\Users\Public\Documents\SoftGenetics\References'. A 'Genome Build' dropdown menu is currently set to 'All'. The main area contains a table titled 'Preloaded References' with columns for 'Reference', 'Genome Build', 'Annotation DB', and 'Comments'. There are also two buttons, 'Build Reference' and 'Import Reference', located to the right of the table header.

Reference	Genome Build	Annotation DB	Comments	<a href="#">Build Reference</a>	<a href="#">Import Reference</a>
Human_v37.p13_105	Human_GRCh37	sg_grch37.p13_105			
Humanv38p14	Human_GRCh38	N/A			
Zebrafish_GRCz11	reference::Zebrafish_GRCz11	N/A			



If using the AutoRun Tool, click the Select button to the right of the Reference field to open the list imported references. Select the preferred reference and click OK

The image shows the NextGENe AutoRun Job Editor interface. On the left is a tree view with folders: Job 1, Sample File(s), Preprocessing (None), Reference File(s), Settings File, and Output Path. The main panel contains fields for Job name (Job 1), ID (20240228094623\_526), Template (Choose Template), Input sequence files, Output, Preprocesses (0), Reference (empty), Process & report settings file, Report & Export (0), and Output to Geneticist Assistant. A red box highlights the 'Select' button next to the Reference field, with a red arrow pointing to the 'Load Reference Files' dialog. This dialog shows 'Available preloaded reference:' with three options: Human\_v37.p13\_105 (checked), Humanv38p14, and Zebrafish\_GRCz11. A red box highlights the 'OK' button at the bottom of the dialog.

**Difficulty accessing our FTP or importing your needed reference?**

**Please email us at [tech\\_support@softgenetics.com](mailto:tech_support@softgenetics.com) and let us know what reference you are trying to import and the difficulties being experienced and we will be happy to help.**