

Quickload to Trackhub Converter

- [Introduction](#)
- [Converter UI](#)
 - [URL - https://translate.bioviz.org/](https://translate.bioviz.org/)
- [Get Requests](#)
 - [/hub.txt](#)
 - [/genomes.txt](#)
 - [/trackDb.txt](#)
- [Head Request](#)
- [Update UCSC Synonyms.txt](#)

Introduction

This page documents the various endpoints used to convert IGB Quickloads to a UCSC track hub.

Converter UI

URL - <https://translate.bioviz.org/>

Get Requests

There are 3 get requests with different file paths [/hub.txt](#), [/genomes.txt](#), [/trackDb.txt](#).

1. [/hub.txt](#)

- **Url** - https://translate.bioviz.org/api/igbTrackHubs?quickLoadUrl=https://api.bitbucket.org/2.0/repositories/lorainelab_udaya/testquickload/src/main/quickload2&filePath=/hub.txt
- **Logic**
 - This URL returns the following fields as a response: hub, shortLabel, longLabel, genomesFile, and email.
 - **hub** – Generated from quickLoadUrl + [/contents.txt](#). The contents.txt file contains the available genome assemblies and their titles. The first genome name in contents.txt is considered the hub name and is prefixed with "IGB" to make it unique.
 - **shortLabel** – Same as hub. It is displayed as the Hub name when a track hub URL is added in Connected Hubs in UCSC Track Data Hubs.
 - **longLabel** – Displays the first genome name and the count of the other available genomes. It appears as the description in UCSC Track Data Hubs.
 - **genomesFile** – Always set to [genomes.txt](#).
 - **email** – Always set to igbhelper@gmail.com.

2. [/genomes.txt](#)

- **Url** - https://translate.bioviz.org/api/igbTrackHubs?quickLoadUrl=https://api.bitbucket.org/2.0/repositories/lorainelab_udaya/testquickload/src/main/quickload2&filePath=/genomes.txt
- **Logic**
 - This URL returns the following fields: genome, trackDb, organism, defaultPos, and twoBitPath.
 - **genome** – Generated from quickLoadUrl + [/contents.txt](#). The genome name is validated against UCSC genome names by checking [ucsc_synonyms.txt](#) (see the section [Generate UCSC Synonyms.txt](#)). If the genome name matches a UCSC genome name, only genome and trackDb are returned. Otherwise, organism, defaultPos, and twoBitPath are also included.
 - **trackDb** – Constructed as genome_name + [/trackDb.txt](#).
 - **organism** – Required only when the genome name is not a UCSC genome name. The value is fetched from species.txt in the integrated-genome-browser repository. The first two words of the genome name (separated by _) are used to search in species.txt. If a match is found, the organism name from the first column of that row is used.
 - **defaultPos** – Fetched from quickLoadUrl + "/" + genome_name + [/genome.txt](#). The genome.txt file contains chromosome lengths and assembly information. defaultPos is set as <first_chromosome_name>:0-<length_of_first_chromosome>.
 - **twoBitPath** – If a 2bit file exists at quickLoadUrl + "/" + genome_name + "/" + genome_name + ".2bit", that path is used. Otherwise, it is retrieved from [annots.xml](#) within the same directory by selecting the <file> tag whose reference attribute is true.

3. [/trackDb.txt](#)

- **Url** - https://translate.bioviz.org/api/igbTrackHubs?quickLoadUrl=https://api.bitbucket.org/2.0/repositories/lorainelab_udaya/testquickload/src/main/quickload2&filePath=/A_thaliana_Jun_2009/trackDb.txt
- **Logic**
 - The field values are generated from quickLoadUrl + "/" + genome_name + [/annots.xml](#).
 - Consider only <file> tags where reference="none".

- This URL returns the following fields: track, bigDataUrl, shortLabel, longLabel, type, and visibility.
 1. **track** – Track should be unique single word in ucscs. It is title with spaces replaced with " _ "
 2. **bigDataUrl** – The name field of the <file> tag if it is already a URL. Otherwise, it is constructed as quickLoadUrl + "/" + genome_name + "/" + name.
 3. **shortLabel** – The last word of the title attribute.
 4. **longLabel** – The description field of the <file> tag.
 5. **type** – Determined based on the file extension of bigDataUrl.
 - a. genomic_extensions = {


```
'bigWig': ['bw', 'bigwig'],
'bigBed 12': ['bb', 'bigbed'],
'bigGenePred': ['bb'],
'bigChain': ['bigchain'],
'bigNarrowPeak': ['bb'],
'bigBarChart': ['bb'],
'bigInteract': ['bb'],
'bigPsl': ['bb'],
'bigMaf': ['bb'],
'hic': ['hic'],
'bam': ['bam', 'cram'],
'HAL': ['hal'],
'vcfTabix': ['vcf', 'vcf.gz']
}
```
 - b. Get the extension of bigDataUrl.
 - c. If the extension is present in genomic_extensions and equals "bb", check whether the word before .bb matches a key in genomic_extensions. If yes, return that key (e.g., bigGenePred).
 - d. Example: GSE63189_Catalog_of_lncRNAs_in_MCF-7cells.bigGenePred.bb bigGenePred
 - e. Otherwise, if the extension matches a value in genomic_extensions, return the corresponding key.
 - f. If the extension is just .bb, return bigBed 12.
 6. **visibility** – Always set to "dense".
- **SuperTrack**
 - If a title has "/" in it, it means it has subfolders. Creating subfolders is not possible in ucsc but we can create top level folder. To do that we need to use superTrack.
 1. **track** – It is first word before "/" in title with " _ " replaced with " _ "
 2. **superTrack on show**
 3. **shortLabel** – It is first word before "/" in title.
 4. **longLabel** – The description field of the <file> tag.
 - a. In side this superTrack we create subtracks.
 - i. **track** – It is similar to normal track name.
 - ii. **bigDataUrl** – It is similar to normal track bigDataUrl.
 - iii. **shortLabel** – It is the title name excluding the first word.
 - iv. **longLabel** – It is similar to normal track longLabel.
 - v. **type** – It is similar to normal track type.
 - vi. **visibility** – Always set to "hide".
 - vii. **parent** – It is the superTrack track name.

Head Request

- **Logic**
 - Returns headers and status_code. The status_code is determined by verifying all GET calls.
- **Purpose**
 - UCSC makes a HEAD request to validate the URL when it is added to Connected Hubs.
 - To test:
 1. Go to [UCSC Track Data Hubs](#) Connected Hubs tab.
 2. When the URL is pasted, UCSC sends a HEAD request to verify the status code.

Update UCSC Synonyms.txt

- **Url** - <https://translate.bioviz.org/api/generateUcscSynonyms>
- **Logic**
 - IGB's source code maintains a list of synonyms for recognizing the same genome version from different genome browsers, called synonyms.txt. We have created a version of this file that matches the IGB specific genome version name to the UCSC genome version name. When a user adds an IGB Quickload as a track hub to the UCSC Genome Browser, translate.bioviz.org checks if the genome versions located in the Quickload match any of the genomes found in UCSC and if so, will load the data in the appropriate genome.
 - When the above URL is clicked on, the translate.bioviz.org server updates the synonyms.txt file following these steps:
 1. Fetch the list of UCSC genomes from the following API: <https://api.genome.ucsc.edu/list/ucscGenomes>
 2. Compare the UCSC genomes to IGB's source code synonyms.txt file (path: core/synonym-lookup/src/main/resources/synonyms.txt).
 3. If a UCSC genome is found in synonyms.txt, add the IGB genome name (first column in synonyms.txt) and the UCSC genome name to the synonyms.txt in the hub_facade repository (path: igb_trackhub/api/synonyms.txt).