



MakeHub

Creating Individual Assembly Hubs for Display with the UCSC Genome Browser

Plant and Animal Genomes XXVIII, January 14th 2020

Katharina J. Hoff

Contact: katharina.hoff@uni-greifswald.de



Introduction

Pipeline

Connecting Your Hub

References

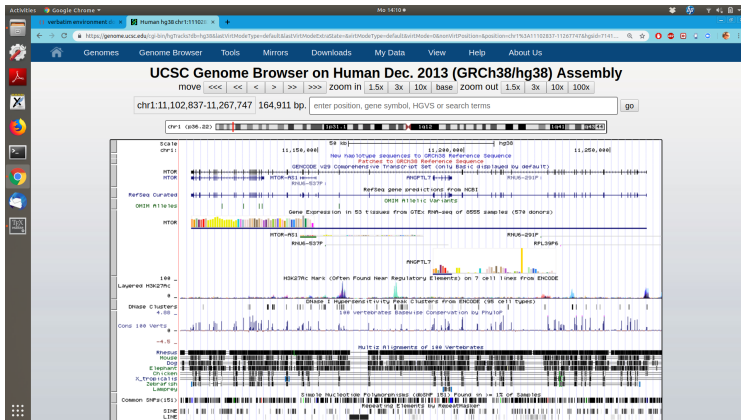
Contents

1 Introduction

2 Pipeline

3 Connecting Your Hub

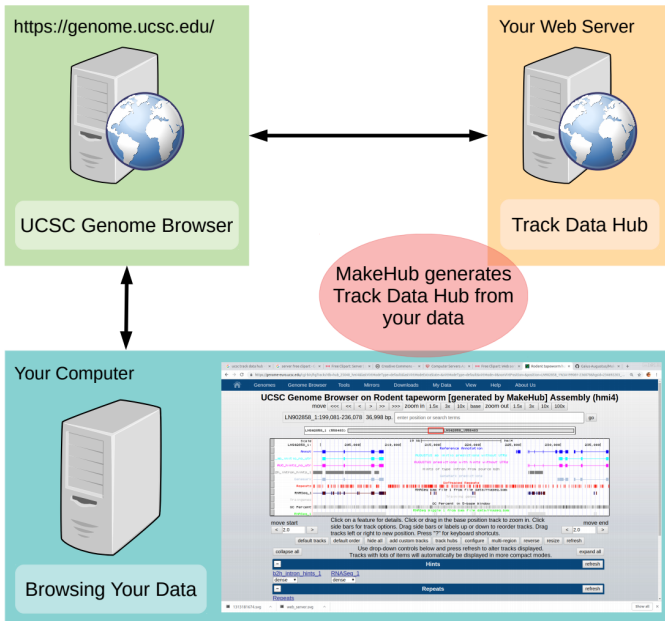
UCSC Genome Browser



- Browser is hosted by UCSC → no installation required
- Assembly Hubs can be locally generated and hosted



UCSC Genome Browser & MakeHub



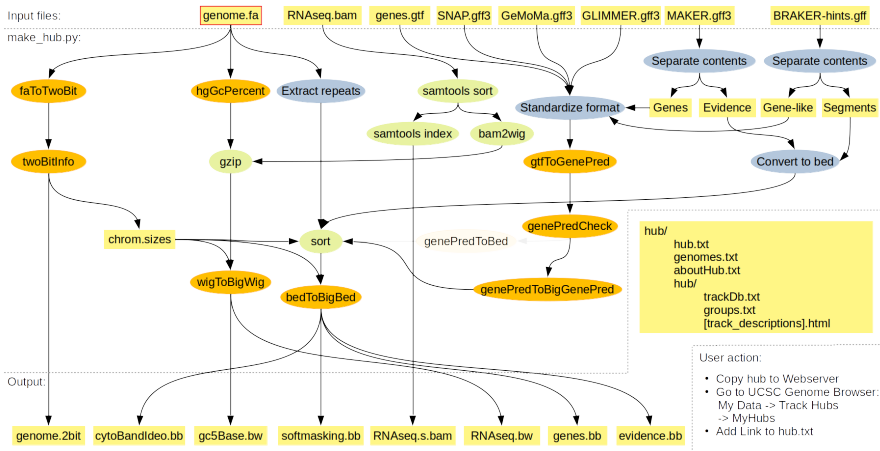
Prerequisites

Unix-System

- **MakeHub** from <https://github.com/Gaius-Augustus/MakeHub>
Obtain with

```
git clone https://github.com/Gaius-Augustus/MakeHub.git
```
- Python 3.5 including biopython & numpy
- SAMtools
- optional: bam2wig from AUGUSTUS
- UCSC tools:
 - bedToBigBed
 - genePredCheck
 - faToTwoBit
 - gtfToGenePred
 - hgGcPercent
 - ixlxx
 - twoBitInfo
 - wigToBigWig
 - genePredToBed

MakeHub Workflow



Software: <https://github.com/Gaius-Augustus/MakeHub>

MakeHub: A Command Line Tool

Print Help:

```
make_hub.py -h
```

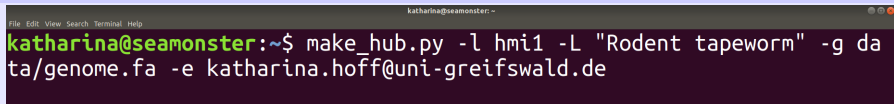
Print Usage Examples:

```
make_hub.py -p
```

Required Input Parameters:

```
make_hub.py -e EMAIL  
              -l SHORT_LABEL  
              -L LONG_LABEL  
              -g GENOME
```

Example: Make Simple Hub



```
katharina@seamster: ~  
katharina@seamster:~$ make_hub.py -l hmi1 -L "Rodent tapeworm" -g data/genome.fa -e katharina.hoff@uni-greifswald.de
```

```
...  
Hub is ready, please copy to a web server, e.g.  
"scp -r ./hmi1 user@server:/target/location"  
Feed the link to hub.txt to UCSC genome browser:  
My Data -> Track Hubs -> MyHubs
```

Hosted at:

```
http://augustus.uni-greifswald.de/bioinf/makehub/examples/hmi1/hub.txt
```

MakeHub: A Command Line Tool

Example: Make Simple Hub

```
katharina@seamster:~$ make_hub.py -l hmi1 -L "Rodent tapeworm" -g da
ta/genome.fa -e katharina.hoff@uni-greifswald.de
```

Resulting Simple Hub

UCSC Genome Browser on Rodent tapeworm [automatically generated assembly hub] Assem

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

LN902858_1:1-15,000 15,000 bp. enter position or search terms go

Scale LN902858_1: 1,000| 2,000| 3,000| 4,000| 5 kb 5,000| 6,000| 7,000| 8,000| 9,000| 10,000| 11,000| 12,000| 13,000| 14,000| hmi1

Repeats Softmasked Repeats

GC Percent in 5-base Window

move start < 2.0 > move end < 2.0 >

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position. Press "?" for keyboard shortcuts.

default tracks default order hide all add custom tracks track hubs configure multi-region reverse resize refresh

collapse all expand all

Use drop-down controls below and press refresh to alter tracks displayed. Tracks with lots of items will automatically be displayed in more compact modes.

Repeats refresh

Mapping and Sequencing refresh

Base Position GC Percent Restr Enzymes Short Match

dense dense hide hide

refresh

MakeHub: A Command Line Tool

Options for RNA-Seq

- b BAM [BAM ...] → wiggle track
- d → bam track
- i HINTS → (RNA-Seq) hints

Parallelization option:

- c CORES

Gene annotation input options:

- X BRAKER_OUT_DIR → all BRAKER gene prediction & hints tracks
- M MAKER_GFF → all MAKER gene prediction & hints tracks
- E GEMOMA_FILTERED_PREDICTIONS
- m GENEMARK
- I GLIMMER_GFF
- S SNAP_GFF
- w AUG_AB_INITIO
- y AUG_AB_INITIO_UTR
- x AUG_HINTS
- z AUG_HINTS_UTR

Options to make hub prettier:

- N LATIN_NAME
- V ASSEMBLY_VERSION

MakeHub: A Command Line Tool

Example: Make Hub with Many Annotation Tracks

```
katharina@seammonster: ~/git/MakeHub
katharina@greif14: ~/git/5c62e0d17094439f14721a
katharina@seammonster: ~/git/MakeHub
katharina@seammonster:~/git/MakeHub$ make_hub.py -l hmi4 -L "Rodent tapeworm" -g data/genome.fa -e katharina.hoff@uni-greifswald.de -a data/annot.gtf -b data/rnaseq.bam -d -X data -M data/maker.gff -E data/gemoma.gff -I data/glimmer.gff -S data/snap.gff -N "Hymenolepsis microstoma" -V GCA_000469805.2
```

```
...
Hub is ready, please copy to a web server, e.g.
"scp -r ./hmi4 user@server:/target/location"
Feed the link to hub.txt to UCSC genome browser:
My Data -> Track Hubs -> MyHubs
```

- 9 gene prediction tracks
- 11 evidence tracks
- Hosted at:
<http://augustus.uni-greifswald.de/bioinf/makehub/examples/hmi4/hub.txt>

MakeHub: A Command Line Tool

Example: Make Hub with Many Annotation Tracks

Galus-Augustus/MakeHub: x | New Tab | augustus.uni-greifswald | Rodent tapeworm hub_2 | +

https://genome.ucsc.edu/cgi-bin/hgTracks?db=hub_2052277_hmi4&lastVirtModeType=default&lastVirtModeExtraState=&virtModeType=default&virtMode=0&nonVirtPosition=&position=LN902858_1163A1-15000&hgid=790046383_FOLDm1tho...

Tracks with lots of items will automatically be displayed in more compact modes.

Evidence generated by MAKER refresh

augustus hide ▼	augustus_masked hide ▼	blastx hide ▼	est_gff.cufflinks hide ▼	genemark hide ▼	model_gff:augustus hide ▼
protein2genome hide ▼	repeatmasker hide ▼				

Hints refresh

b2h_intron_hints_1 hide ▼	RNASeq_1 dense ▼
--	-------------------------------------

Repeats refresh

[Repeats](#)
dense ▼

Gene Predictions refresh

Annot dense ▼	AUG_ab_initio_no_utr dense ▼	AUG_hints_no_utr dense ▼	GeMoMa dense ▼	Genemark dense ▼	GlimmerHMM dense ▼
maker hide ▼	SNAP hide ▼	Traingenes hide ▼			

Mapping and Sequencing refresh

Base Position dense ▼	GC Percent dense ▼	Restr Enzymes hide ▼	Short Match hide ▼
--	---------------------------------------	---	---------------------------------------

other refresh

[RNASeq_1](#)
dense ▼

refresh

genePredToBigG... Show all x



Connecting Your Hub

- 1 copy hub to a web server
- 2 remember the hyperlink to `hub.txt` file on web server
- 3 go to `https://genome.ucsc.edu/`
- 4 click My Data
- 5 click Track Hubs
- 6 click My Hubs
- 7 insert link to `hub.txt`
- 8 click Add Hub
- 9 ... browse happily ever after

Connecting Your Hub



Introduction

Pipeline

Connecting Your Hub

References

UCSC Genome Browser Gateway - Mozilla Firefox

UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute UCSC Genome Browser Gateway

Genomes Genome Browser Tools Mirrors Downloads My Data Projects Help About Us

Browse/Select Species

POPULAR SPECIES

- Human
- Mouse
- Rat
- Zebrafish
- Fruitfly
- Worm
- Yeast

Enter species or common name

REPRESENTED SPECIES

- fly_utrnaseq (1)
- hml4 (1)
- fly_6.07 (1)
- Starfish (1)
- Human
- Chimp
- Rancho

Find Position

Human Assembly
D...
Track Hubs
Custom Tracks
Track Collection Builder

Position/Search
Enter position, gene symbol or search terms
Current position: chr1:11,102,837-11,267,747

GO

Human Genome Browser - hg38 assembly
view sequences

UCSC Genome Browser assembly ID: hg38
Sequencing/Assembly provider ID: Genome Reference Consortium Human GRCh38.p12 (GCA_000001405.2)
Assembly date: Dec 2012 initial

Connecting Your Hub

Track Data Hubs

Track data hubs are collections of external tracks that can be imported into the UCSC Genome Browser. Hub tracks show up under blue label bar on the main browser page, as well as on the configure page. For more information, including where to [host](#) your track [User's Guide](#). To import a public hub click its "Connect" button below.

NOTE: Because Track Hubs are created and maintained by external sources, UCSC is not responsible for their content.

Public Hubs My Hubs Validate Hub

URL: Add Hub

No Unlisted Track Hubs

Contact [us](#) to add a public hub.

Insert link to your custom
Track Data Hub

Connecting Your Hub



UCSC Genome Browser Gateway - Mozilla Firefox

https://genome-euro.ucsc.edu/cgi-bin/hgGateway/?hgdb=234766299_4768PpUbae190&asB&hcCoVx

UNIVERSITY OF CALIFORNIA SANTA CRUZ Genomics Institute UCSC Genome Browser Gateway

Genomes Genome Browser Tools Mirrors Downloads My Data Projects Help About Us

Browse/Select Species **Find Position**

POPULAR SPECIES

Human Mouse Rat Zebrafish Fruitfly Worm

Yeast

Enter species or common name

REPRESENTED SPECIES

purple sea star Assembly Hub (1) Hub Genomes
hmi4 (1)

Human
Chimp
Bonobo
Gorilla

hmi4 Hub Assembly
[generated by MakeHub]

Position/Search Term
Enter position, gene symbol or search terms
Current position: LN902858_1:1-15,000

GO

Recent tapeworm Genome Browser - hmi4 assembly
view sequences

Start browsing

References

- Raney BJ et al. (2014) Track Data Hubs.
- Kent WJ et al. (2002) UCSC Genome Browser.
- Hoff KJ (2019) MakeHub: Fully automated generation of UCSC Genome Browser Assembly Hubs
- Li H et al. (2009) The sequence alignment/map format and SAMtools
- Stanke M et al. (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding.
- Hoff KJ et al. (2015) BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS.
- Hoff KJ et al. (2019) Whole-Genome Annotation with BRAKER.
- Holt C & Yandell M (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects
- Majoros WH & Salzberg SL (2004) TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders.
- Korf I (2004) Gene finding in novel genomes.
- Keilwagen J et al. (2018) Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi.
- Lomsadze A et al. (2014) Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm.
- Bruna T et al. (2020) GeneMark-EP and -EP+: automatic eukaryotic gene prediction supported by spliced alignments.



Acknowledgements

Mario Stanke
Malte Wellnitz
Matthis Ebel
UCSC Genome Browser Group

The international collaboration between the groups of Mark Borodovsky and Mario Stanke, supported by US National Institute of Health grant GM128145, gave rise to the development of MakeHub.

MakeHub is Available for Download at

<https://github.com/Gaius-Augustus/MakeHub>