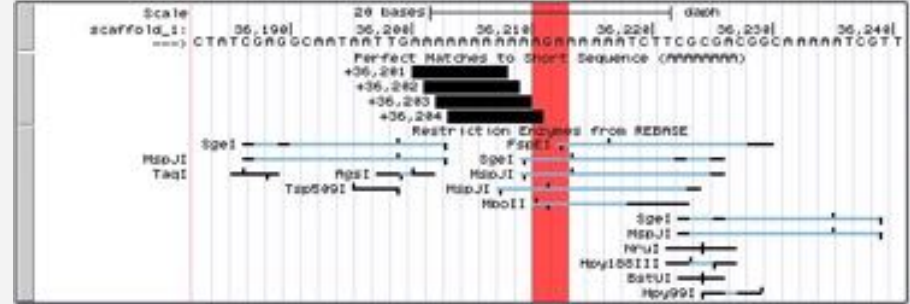


# UCSC Browser Assembly Hubs To Visualize and Analyze New Genomes

```
>scaffold_1
GTTGTAAATACTCTATTCTACAATAAAACCAA
TCATAGGTTGAATTGGCGTTGAAGTAAAAACAAA
...
>scaffold_2
AGTTATGACAAACTATAAAAAGTCGGTAGAGACAAAAG
TCGTTCTGTGGACGAAGCGACCAAAACTGAGCACAAAGAT?
...
>scaffold_3
CATAAATTCATAAATCAATTCATGAAGAATAATT
TAGAAAATTTCCAGGAAGTTTGAAGTTGCTAT
...
```

```
hubDirectory
|__hub.txt
|__genomes.txt
|__hg19
|__trackDb.txt
```





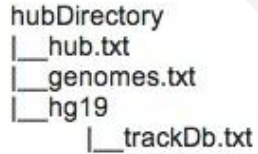
# Examples of Hubs and Resources

## Genbank Assembly Hubs

<http://genome-test.gi.ucsc.edu/gbdb/hubs/genbank/>

**MakeHub** python tool automatically makes a hub

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



```
hubDirectory
|__hub.txt
|__genomes.txt
|__hg19
|__trackDb.txt
```

## UCSC Public Hubs

<http://genome.ucsc.edu/cgi-bin/hgHubConnect>

## G-OnRamp

<http://gonramp.wustl.edu/>

**CyVerse Discovery Environment** byte-range supported data hosting

<https://de.cyverse.org/de/>

# Public Track Hubs

**UCSC Public Hubs** *moderated collection of track hubs*

<http://genome.ucsc.edu/cgi-bin/hgGateway>



A screenshot of the UCSC Genome Browser's "Browse/Select Species" page. The page has a yellow header with the text "Browse/Select Species". Below the header, there is a section titled "POPULAR SPECIES" with six icons: Human (red), Mouse (red), Rat (red), Fruitfly (blue), Worm (blue), and Yeast (purple). Below these icons is a search input field containing the text "duck". Below the search field, there are several search results listed, including "Assembly Hub: CESAR Gene Mappings", "Assembly Hub: Croc and Bird Hub", and "Mouse strain assemblies (16)". A blue arrow points from the left towards the search input field.

Search term

# Public Track Hubs

UCSC Public Hubs *moderated collection of track hubs*

<http://genome.ucsc.edu/cgi-bin/hgHubConnect>

Search term

Public Hubs My Hubs

Enter search terms to find in public track hub description pages:  
duck

Filter hubs by assembly:  Search Public Hubs

Displayed list **restricted by search terms:** duck Show All Hubs

When exploring the detailed search results for a hub, you may right-click on a hub name and select **Connect**. Clicking **Connect** redirects to the gateway page of the selected hub's default assembly.

Display	Hub Name	Description
<a href="#">Disconnect</a>	Croc and Bird Hub	Croc, Bird, and Archosaur Assembly Hub
Search details ...		
1 Matching Assembly		
Apr. 2013 (BGI_duck_1.0/anaPla1) (anaPla1)		
Assembly Description: ... This is the mallard duck <i>Anas platyrhynchos</i> genome		
<a href="#">Disconnect</a>	CESAR Gene Mappings	Human Exons mapped by CESAR
Search details ...		
1 Matching Assembly		
Apr 2013 (BGI_duck_1.0/anaPla1) (hub)		
<a href="#">Open this assembly</a>		

My Data Help About U

- My Sessions
- Public Sessions
- Track Hubs**
- Custom Tracks
- Track Collection Builder

# Assembly Hubs via Sessions

Public Sessions *community contributed views of the Browser*

<http://genome.ucsc.edu/cgi-bin/hgPublicSessions>

**Public Sessions**

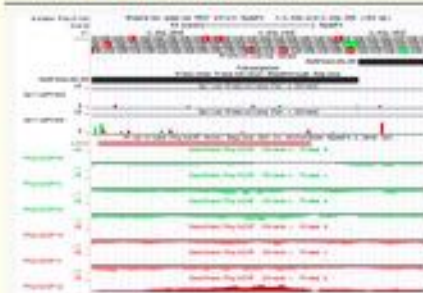
Sessions allow users to save snapshots of the Genome Browser and its current configuration, including displayed tracks, position, and custom track data. The Public Sessions tool allows users to easily share those sessions that they deem interesting with the rest of the world's researchers. You can add your own sessions to this list by checking the appropriate box on the [Session Management](#) page.

See the [Sessions User's Guide](#) for more information.

Sort by: Creation (oldest first)

Show  entries

Search:

Screenshot	Session Properties
	<p><b>Description:</b> This is an assembly hub for the AgamP4 assembly of <i>Anopheles gambiae</i> PEST strain. It includes the assembly; the coding genes and pseudogenes from vectorbase version 4.3; predicted stop codon readthrough regions; PhyloCSF tracks showing evolutionary protein-coding potential; splice-prediction tracks using the maximum-entropy splice-prediction algorithm; and novel coding and pseudogene predictions using PhyloCSF, excluding regions already annotated in vectorbase version 4.3</p> <p><b>Author:</b> #Jungr</p> <p><b>Session Name:</b> AgamP4</p> <p><b>Genome Assembly:</b> hub_102577_AgamP4</p> <p><b>Creation Date:</b> 2016-08-09</p> <p><b>Views:</b> 366</p> <p><b>Description:</b> We sequenced the hermaphroditic freshwater snail, <i>Biomphalaria glabrata</i> (strain BB02), the host for the medically</p>

My Data Help About

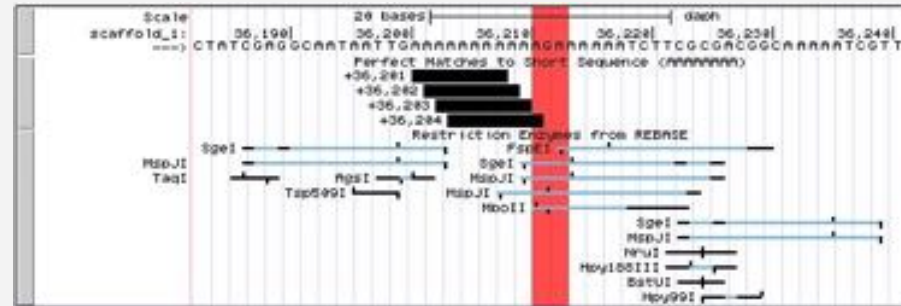
- My Sessions
- Public Sessions**
- Track Hubs
- Custom Tracks

hub\_

# Building an Assembly Hub for Platypus

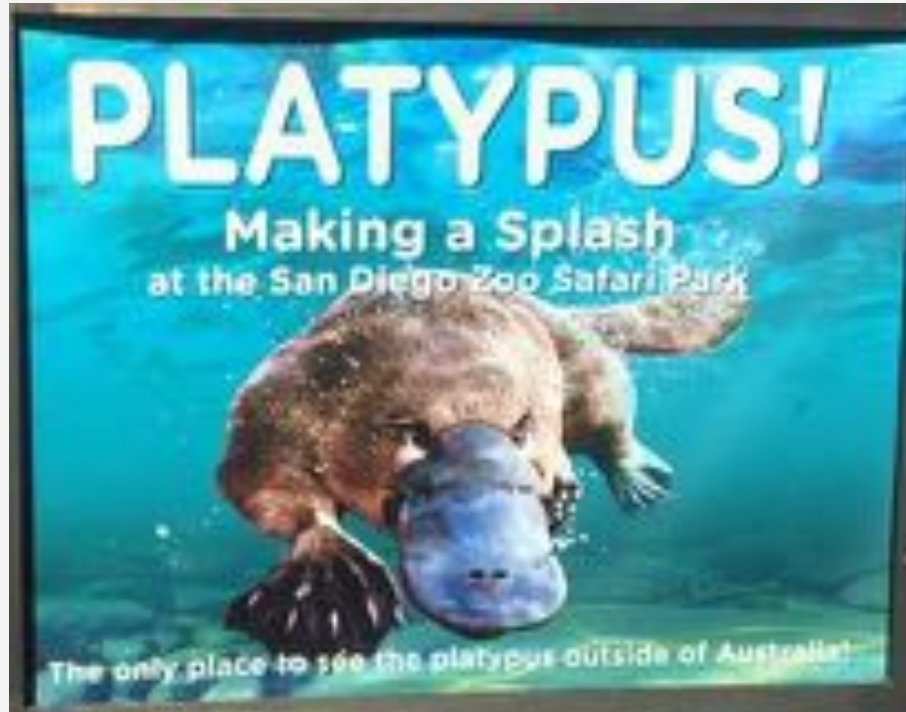
```
>scaffold_1
GTTGTAAATACTCTATTCTACAATAAAACCAA
TCATAGGTTGAATTGGCGTTGAAGTAAAAACAA
...
>scaffold_2
AGTTATGACAAACTATAAAAAGTCGGTAGAGACAAAAG
TCGTTCTGTGGACGAAGCGACCAAACTGAGCACAAGAT
...
>scaffold_3
CATAAATTCATAAATCAATTCATGAAGAATAATT
TAGAAAATTTCCAGGAAGTTTGAAGTTGCTA
...
```

hubDirectory  
|\_\_hub.txt  
|\_\_genomes.txt  
|\_\_hg19  
|\_\_trackDb.txt





# Building an Assembly Hub for Platypus





# Building an Assembly Hub for Platypus



# Build the 2bit for Platypus

Navigate to NCBI and find the latest Platypus Genome:

<https://www.ncbi.nlm.nih.gov/assembly/?term=platypus>



Select mOrnAna1.p.v1 and click the “Download Assembly” to get FASTA

A screenshot of the NCBI assembly page for 'mOrnAna1.p.v1'. The page displays metadata for the assembly, including the organism name 'Ornithorhynchus anatinus (platypus)', isolate 'Pmale09', sex 'male', bioSample 'SAMN08537700', bioProject 'PRJNA489114', submitter 'Vertebrate Genomes Project', and date '2019/05/16'. On the right side, there is a 'Download Assembly' button. A dropdown menu is open, showing options for 'Source database (GenBank or RefSeq)' set to 'RefSeq' and 'File type' set to 'Genomic FASTA (.fna)'. The estimated size is 571.7 MB. A 'Download' button is visible at the bottom of the dropdown menu. A blue box on the left side of the page indicates 'There are 4 assemblies for this organism' and provides a 'See more' link.

# Build the 2bit for Platypus

**Obtain the full FASTA file 500 MB (genomic.fna.gz):**

```
wget
```

```
ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/004/115/215/GCA_004115215.1_mOrnAna1.p.v1/GCA_004115215.1_mOrnAna1.p.v1_genomic.fna.gz
```

**Unzip the fasta file:**

```
gzip -d GCA_004115215.1_mOrnAna1.p.v1_genomic.fna.gz
```

# Build the 2bit for Platypus

Obtain utilities from the UCSC Genome Browser

`faToTwoBit` and other standalone command-line **Utilities** are downloadable tools available on Linux and UNIX platforms.



The image shows a screenshot of the UCSC Genome Browser website. The header includes the University of California Santa Cruz logo and the text "UCSC Genome Browser". Below the header is a navigation bar with links: Home, Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Help, and About Us. The main content area features a large image of a DNA double helix on the left and a dropdown menu on the right. The dropdown menu is open, showing options: Genome Data, Source Code, Genome Browser Store, Utilities (highlighted in blue), FTP, and MySQL Access. An arrow points to the "Utilities" option. At the bottom of the page, there is a link: "download data from the genome Browser database".

# Build the 2bit for Platypus

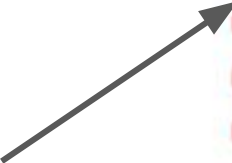
## Obtain utilities from the UCSC Genome Browser

`faToTwoBit` and other standalone command-line **Utilities** are downloadable tools available on Linux and UNIX platforms.

---

### Utilities

The [utilities directory](#) offers downloads of pre-compiled standalone binaries for:

- [LiftOver](#) (which may also be accessed via the [web version](#)). The over.chain lift
  - Other command-line utilities
  - [Blat](#) ▶
- 

# Build the 2bit for Platypus

Obtain utilities from the UCSC Genome Browser

`faToTwoBit` and other standalone command-line **Utilities** are downloadable tools available on Linux and UNIX platforms.

```
faFilter          09-Jan-2018 15:55 4.1M
faFilterN        09-Jan-2018 15:55 4.3M
faFrag           09-Jan-2018 15:55 4.1M
faNoise          09-Jan-2018 15:55 4.1M
faOneRecord      09-Jan-2018 15:55 4.1M
faPolyASize     09-Jan-2018 15:55 4.1M
faRandomize      09-Jan-2018 15:55 4.1M
faBc             09-Jan-2018 15:55 4.1M
faSize          09-Jan-2018 15:55 4.1M
faSomeRecords   09-Jan-2018 15:55 4.1M
faSplit         09-Jan-2018 15:55 4.1M
faToFastq       09-Jan-2018 15:56 4.1M
faToTab         09-Jan-2018 15:56 4.1M
faToTwoBit      09-Jan-2018 15:56 4.2M
faTrans         09-Jan-2018 15:56 4.1M
faToStatsAndSample 09-Jan-2018 15:55 4.2M
faToToFa        09-Jan-2018 15:55 4.2M
faToTwoBits     09-Jan-2018 15:57 16M
faToChromSizes  09-Jan-2018 16:01 2.7K
faToMotif       09-Jan-2018 15:56 4.2M
```



# Build the 2bit for Platypus

Obtain utilities from the UCSC Genome Browser

```
wget http://hgdownload.soe.ucsc.edu/admin/exe/linux.x86_64/faToTwoBit
```

```
wget http://hgdownload.soe.ucsc.edu/admin/exe/linux.x86_64/twoBitInfo
```

```
wget http://hgdownload.soe.ucsc.edu/admin/exe/linux.x86_64/bedToBigBed
```



# Build the 2bit for Platypus

Use faToTwoBit:

```
faToTwoBit GCA_004115215.1_mOrnAna1.p.v1_genomic.fna.gz  
plat.2bit
```



```
>scaffold_1  
GTTGTAAATACTCTATTCTACAATAAAACCAAAGATAACTCTTTATCAG  
TCATAGGTTGAATTGGCGTTGAAGTAAACAAAAATACTGCTCAAAGG  
...  
>scaffold_2  
AGTTATGACAAACTATAAAAAAGTCGGTAGAGACAAAAGCGAAAAGGATCT  
TCGTTTCGTGGACGAAGCGACCAAACTGAGCACAAGATAAATCCCGAATA  
...  
>scaffold_3  
CATAAATTCATAAATCAATTCATGAAGAATAATTTTAGAAAAATGGTTCA  
TAGAAAAATTTCCAGGAAGTTTGAAGTTGCTATAATGATTATTTCTCTTG  
...  
>scaffold_4  
ATACTGCCCCATGCCTCATCTACTCCCTTCTTAACCTTTTCATCAACAACT  
TGCTTAAGTTGGTCATCACTGACTGAGTGAGCCCAAACCTGGAATAGCTGT  
...
```

The resulting binary file **plat.2bit** will be referenced in **twoBitPath** in the genomes.txt file

# Build a Cytoband Track from the Platypus 2bit

Extract the size of the chromosomes in Platypus for building indexes

```
twoBitInfo plat.2bit plat.chrom.sizes
```



LKOO02000001.1	12661638
LKOO02000002.1	16711
LKOO02000003.1	11978
LKOO02000004.1	5190334
LKOO02000005.1	16581
LKOO02000006.1	25283
LKOO02000007.1	27846
LKOO02000008.1	17515
LKOO02000009.1	19313
LKOO02000010.1	273978

# Build a Cytoband Track from the Platypus 2bit

From the `chrom.sizes` create a BED file that will represent chromosomes

```
cat plat.chrom.sizes | awk '{print $1,0,$2,$1,"gneg"}' >  
yourCytobandIdeo.bed
```



LKUO02000001.1	12661638
LKUO02000002.1	16711
LKUO02000003.1	11978
LKUO02000004.1	5190334
LKUO02000005.1	16581
LKUO02000006.1	25283
LKUO02000007.1	27846
LKUO02000008.1	17515
LKUO02000009.1	19313
LKUO02000010.1	273978



LKUO02000001.1	0	12661638	LKUO02000001.1	gneg
LKUO02000002.1	0	16711	LKUO02000002.1	gneg
LKUO02000003.1	0	11978	LKUO02000003.1	gneg
LKUO02000004.1	0	5190334	LKUO02000004.1	gneg
LKUO02000005.1	0	16581	LKUO02000005.1	gneg
LKUO02000006.1	0	25283	LKUO02000006.1	gneg

# Build a Cytoband Track from the Platypus 2bit

Use `bedToBigBed` to wrap text data into a binary indexed version


```
bedToBigBed -type=bed4 yourCytobandIdea.bed -as=cytoBand.as plat.chrom.sizes  
yourCytobandIdea.bigBed
```



```
LKUO02000001.1 0 12661638 LKUO02000001.1 gneg  
LKUO02000002.1 0 16711 LKUO02000002.1 gneg  
LKUO02000003.1 0 11978 LKUO02000003.1 gneg  
LKUO02000004.1 0 5190334 LKUO02000004.1 gneg  
LKUO02000005.1 0 16581 LKUO02000005.1 gneg  
LKUO02000006.1 0 25283 LKUO02000006.1 gneg
```

```
table cytoBandIdea  
"cytoBandIdea describes...."  
(  
  string chrom;      "Reference.... "  
  uint  chromStart; "Start..."  
  uint  chromEnd;   "End..."  
  string name;      "Name of item"  
  string gieStain;  "Giemsa stain..".  
)
```

Turned  
**yourCytobandIdea.  
bed** text file into a  
binary indexed  
version:  
**yourCytobandIdea.  
bigBed**





# Build the hub.txt

hub ---  
useOneFile on  
shortLabel ---  
longLabel ----  
email ----

genome ----  
description ---  
twoBitPath plat.2bit  
organism ---  
defaultPos ----  
scientificName ---

track cytoBandIdeo  
longLabel Chromosome ideogram with cytogenetic bands  
shortLabel cytoBandIdeo  
bigDataUrl cytoBandIdeo.bigBed  
type bigBed




**The hub stanza sets useOneFile on,**

Rest is information you input

# Build the hub.txt

hub ---  
useOneFile on  
shortLabel ---  
longLabel ----  
email ----

genome ----  
description ---  
**twoBitPath** plat.2bit  
organism ---  
defaultPos ----  
scientificName ---



**The genome stanza** shares where to find the 2bit (and information about the your genome)

The **defaultPos** must have coordinates that exist in your genome.

track cytoBandIdeo  
longLabel Chromosome ideogram with cytogenetic bands  
shortLabel cytoBandIdeo  
**bigDataUrl** cytoBandIdeo.bigBed  
type bigBed



# Build the hub.txt

hub ---

useOneFile on

shortLabel ---

longLabel ----

email ----

genome ----

description ---

twoBitPath plat.2bit

organism ---

defaultPos ----

scientificName ---


track cytoBandIdeo

longLabel Chromosome ideogram with cytogenetic bands

shortLabel cytoBandIdeo

bigDataUrl cytoBandIdeo.bigBed

type bigBed

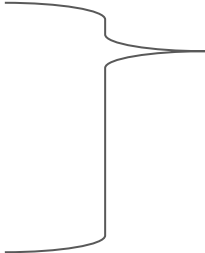


**The cytoBandIdeo track stanza is unique** it expects a certain kind of data to build the navigation bar.

# Adding Blat Servers

You can edit the **genomes stanza** to include blat servers

```
genome ----  
description ---  
twoBitPath plat.2bit  
organism ---  
defaultPos ----  
scientificName ---  
transBlat hgwdev.gi.ucsc.edu 166666  
blat hgwdev.gi.ucsc.edu 166669
```

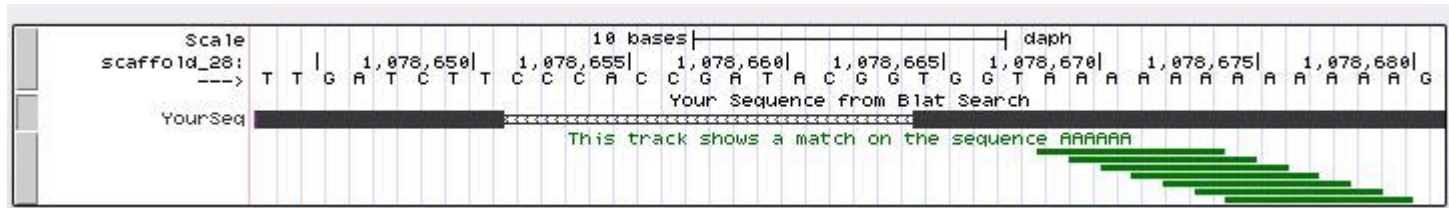
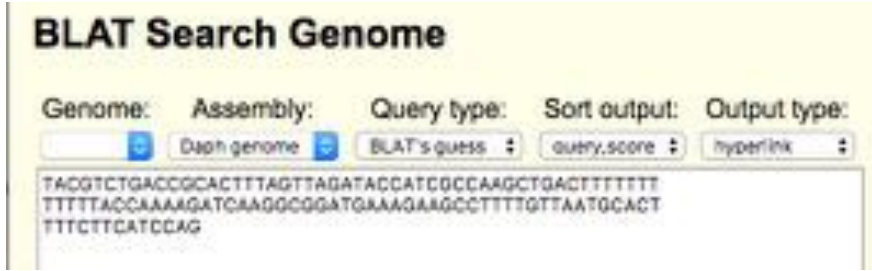


**The genome stanza** can inform the browser where to find blatServers if they are running (or define the blue bar groups)

```
gfServer start localhost 166666 -trans -mask plat.2bit &  
gfServer start localhost 166669 -stepSize=5 plat.2bit &
```

# Adding Blat Servers

- BLAT search sequence or amino-acid matched







Thank You!



UNIVERSITY OF CALIFORNIA  
**SANTA CRUZ** | Genomics  
Institute



SLIDES: [bit.ly/PAG\\_UCSC](https://bit.ly/PAG_UCSC)