

Intro to Data QA and Release

Jairo Navarro, UCSC Genomics Institute

September 07, 2018





Routes to the RR

5 different routes to release data

1. **CGI code, related .ra, .js and .css files:** stored in git, moved from dev to beta to the RR via the build process, which can take from one to three weeks.
2. **trackDb:** stored in git, sent to dev/beta/beta-public via makes, pushed to the RR, happens all the time
3. **htdocs static html files:** similar to trackDb, pushed from hgwbeta to the RR all the time
4. **/gbdb:** pushed straight from hgwdev to the RR (because hgwbeta and the RR share their /gbdb)
5. **mysql tables:** pushed from hgwdev to hgwbeta and then from hgwbeta to the RR



Let's start with Assembly/Track QA



Claiming the ticket in Redmine

Ready for QA

 New issue

▾ Filters

▾ Options

 Apply  Clear

<input type="checkbox"/>	#	Tracker	Priority ▾	Subject	Assemblies	Assignee	Updated	Target version	Developer
▣ QA Ready 8									
<input type="checkbox"/>	21079	Track	High	5 NCBI RefSeq Anomalies tracks	hg38		08/31/2018 02:16 PM		Christopher Le
<input type="checkbox"/>	17247	Assembly	Medium	oviAri4 - sheep - Ovis aries	oviAri4		05/03/2018 01:33 PM		Christopher Le
<input type="checkbox"/>	18273	Track	Medium	add more recent RepeatMasker data to hg38	hg38 mm10		06/06/2018 09:33 AM		Christopher Le
<input type="checkbox"/>	18508	Track	Medium	miRNA tissue specific expression	hg38		06/21/2018 04:09 PM		Christopher Le
<input type="checkbox"/>	21852	Track	Medium	GENCODE Mouse VM18 release	mm10		08/30/2018 08:58 AM		Mark Diekhar



Review ticket history

Assembly #20979

Edit Log time Watch Copy



panTro6 - chimp - Pan troglodytes

« Previous | 1 of 17 | Next »

Added by Genome Browser Admin 6 months ago. Updated about 1 hour ago.

Status:	Reviewing	Start date:	02/15/2018
Priority:	Urgent	Due date:	
Assignee:	Jairo Navarro	% Done:	<div style="width: 0%; background-color: #ccc; border: 1px solid #ccc;"></div> 0%
Category:	-	Table List:	/hive/data/genomes/panTro6/pushQ/redmine.panTro6.table.list
Developer:		Release Log Text :	/hive/data/genomes/panTro6/pushQ/redmine.panTro6.releaseLog.txt
Assemblies:	panTro6	Release Log URL:	
Stalled By:	RefSeq assembly release process	Released to RR:	No
File List:	/hive/data/genomes/panTro6/pushQ/redmine.panTro6.file.list		



Copy the “Template: Assembly checklist”

My Drive ▾



Quick Access



panTro6: Assembly Release Chec...

You opened yesterday



README

You opened this month



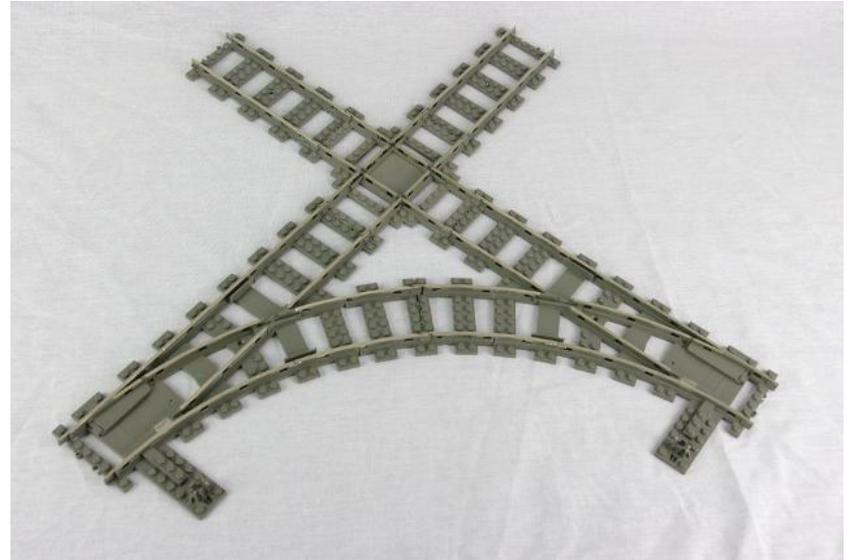
template: Assembly Release Chec...

You edited at some point

What is the type of release?

- Three types of data releases
 - Assembly release
 - New track for an existing assembly
 - Update to an existing track

- Read the track type description page, i.e., bigBed help page
 - **Is this a new track type?**
 - **Is it a binary indexed track type?**





Review data on hgwdev (command line)

- The bread & butter:
 - qaGbTracks

```
[12:55 PM] jairo@hgwdev trackDb/ $ qaGbTracks
usage: qaGbTracks [-h] [-f TABLEFILE] db [table [table ...]] outFile
qaGbTracks: error: too few arguments
[01:04 PM] jairo@hgwdev trackDb/ $
```

- big* track QA
 - bigBedInfo
 - bigWigInfo
 - bigBedSummary
 - bigWigSummary
 - bigBedToBed
 - bigWigToWig

Review data on hgwdev (website)

- Review gateway page (if a new assembly)
- Review long/short labels
- Make sure data is properly displayed (All metadata is on the details page)
- Review data using our tools (if applicable)
- Check track description pages for HTML styling and typos

UCSC TEST Genome Browser on Chimp Jan. 2018 (Clint_PTRv2/panTro6) Assem

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

chr7:111,311,998-111,585,728 273,731 bp. enter position, gene symbol or search terms go

chr7 7

Scale 100 kb | panTro6
chr7: 111,350,000| 111,400,000| 111,450,000| 111,500,000| 111,550,000|

Some thign
Gap
RefSeq Curated RefSeq gene predictions from NCBI
Other RefSeq Non-Chimp RefSeq Genes
AY064549 Chimp mRNAs from GenBank
AF512947
AY143170 Chimp ESTs That Have Been Spliced
Spliced ESTs
RepeatMasker Repeating Elements by RepeatMasker

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.

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

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



HTML styling review!

- Do not underline - implies hyperlink
- The style in h2 and h3 headers:
 - Capitalize the first word and any proper nouns
 - Everything else should be in lowercase
- Limit lines to a maximum of 100 characters
- Indentation: Indent every child element 2 spaces
- "Genome Browser" should be capitalized whenever it's referring to the application itself.
 - If you're generically referring to genome browsers, no caps are needed.



Push to hgwbeta

Tools to use:

- copyHgcentral
- bigPush.sh
- Cluster-admin :)

[push-request] push gbdb files to hgfnfs1



Inbox x

Push Requests x



Jairo Navarro Gonzalez <jnavarr5@ucsc.edu>

to push-request ▾

Hello Pushers,

Please push the following files:

/gbdb/ce11/crispr/[crispr.bb](#)

/gbdb/ce11/crispr/crisprDetails.tab

from hgwdev --> hgfnfs1/genome-euro/genome-asia

(Note: Items that are symlinked on dev should become real files on hgfnfs1.)

Reason: Staging the CRISPR track for ce11 on hgwbeta, refs #21517

Thanks,
Jairo Navarro



TrackDb release tags

When the data is stored inside a big* file, a “release tag” is needed so the data does not prematurely appear on the RR.

```
track netDm1
release beta,public
shortLabel D. mel. Net
longLabel $o_Organism ($o_date) Alignment Net
group compGeno
priority 134
visibility dense
spectrum on
type netAlign dm1 chainDm1
otherDb dm1
```

```
track xenoClinvar
shortLabel Human Variants
longLabel ClinVar Variants mapped to Mouse
type bigBed 9 +
itemRgb on
bigDataUrl /gdb/mm10/bbi/xenoClinvar.bb
group x
mouseOverField _mouseOver
urls rcvAcc="https://www.ncbi.nlm.nih.gov/clinvar/$$/" geneId="https://www.ncbi.nlm.nih.gov/gene/$$"
release alpha
```



Review on hgwbeta

- Make sure everything is now on the hgwbeta machine
- Review using our tools again
- Check details pages to make sure no supporting tables are missing



Push to the RR

- Ask cluster-admin to move the data to the RR and our official mirror sites
- Validate that all of the data is displayed properly on all three sites

- Finally.... Announce the release on Facebook, Twitter, and genome-announce

[push-request] Push tables for hg19



Inbox x

Push Requests x



Jairo Navarro Gonzalez <jnavarr5@ucsc.edu>

to push-request ▾

Hello Pushers,

Please push the following tables for the hg19 database:

decipherSnvs
decipherSnvsRaw

from mysqlbeta --> mysqlrr/genome-euro/genome-asia

Reason: **Releasing** the DECIPHER SNVs track for hg19, refs #21104

Thanks,
Jairo Navarro



Questions?