Genome assembly and annotation

Currently our Partek server has human, mouse and zebrafish reference. If sequencing data is from other organisms, or you would like to use a different assembly and annotation from what's available on the server, you would need to find the reference genome and provide that to Partek. There are multiple ways of doing this task, providing Partek with a URL is probably the easiest.

From Ensembl

1. Go to <http://www.ensembl.org/info/data/ftp/index.html> and find your organism

Machine generated alternative text:
Human 
Horn 0 
Mouse 
Mus 
musculus 
Zebra fish 
Damo rerio 
Mus spretus 
Al ca 
Vicugna 
DNA 
(FASTA) 
FAS TAG 
EAS_TArG 
PASTA re 
AS Are 
cDNA 
(FASTA) 
EASIAG 
EASIA6 
FASTA6 
EASIA6 
PASTA 6 
CDS 
(FASTA) 
EASIAtG 
EASTAtG 
FASTA,G 
EASTAtG 
FASTAIG 
ncRNA 
(FASTA) 
EASIBG 
FASTA,G 
FASTAO 
Protein 
sequence 
(FASTA) 
FAS TAG 
PASTA re 
Annotated 
sequence 
(EMBL) 
EMBLte 
Annotated 
sequence 
(GenBank) 
Gentianxte 
Gennanxa 
GenBankrQ 
Gennanxta 
GenBankrQ 
Gene 
sets 
G-LEG 
GFE 36 
Whole Variation 
Variation 
(VCF) 
"CEG 
YCErG 
vcr-rG 
Variation 
NEP) 
VEPLG 
VEPte 
Regulation 
(GFF) 
Regulation G 
(GFF) 
Regulation G 
(GFF) 
Data 
files 
BAWBigWig 
databases 
MYSQLG 
MYSQLtG 
MYSQC6 
MYSQLtG 
MYSQUrG 
(GVF) 
GVF6 
BAM/BiqWiqtG 
data 
BAM/BiqWiq6 
This 's the reference sequence in *.fa format 
This is the annotation file in * 
.gtf or *. 
gff3 format 

1. In Partek Flow, select aligner you would like to use from the task bar, select "New assembly", and fill the pop-up window as below (give it a name so that other people know which version this is, for this example, GRCz10):

Machine generated alternative text:
Home BSR445 v2 STAR 
Select STAR 2.4.1d index 
Configure 
Queue v 
Add STAR 2.4.1d index 
Projects v 
Assembly 
Aligner index 
Alignment options 
Generate unaligned 
reads 
Advanced options 
New assembly 
Whole genome 
Name 
@ Build index 
O Import index 
Create 
Copyright C 2017 Partek Incorporated_ All rights reserved. 
Back 
Option set 
Finish 

1. After clicking "Create", copy and paste the URL for \*.fa file:

Machine generated alternative text:
Queue v 
Home BSR445 v2 STAR Build index Impon reference sequence 
Select file for test from O partek Flow Server O My Computer O URL 
Projects v 
Back 
Input URL ftp-//ftp_ensembl.org/pub/release-gc/fasta/danio rerio/dna,lDanio rerio_GRCz10 x 
Next 

URL can be retrieved from step 1 website. There are many files in that link, what you want to use here should be top-level soft-masked file (see below). Read "README" for more details.

Machine generated alternative text:
07/21/2017 
07/21/2017 
s. 
07/21/2017 
6. 
07/21/2017 
07/21/2017 
B. 
07/21/2017 
9. 
07/21/2017 
07/21/2017 
07/21/2017 
07/21/2017 
02: 
02: 
02: 
02: 
02: 
02: 
02: 
02: 
02: 
02: 
13PM 
13PM 
13PM 
13PM 
13PM 
13PM 
13PM 
13PM 
13PM 
13PM 
23, 
22, 
19, 
23, 
17, 
18, 
9 
439, 
€22, 
989, 
389, 
802, 
39B, 
324, 
s, 
560 
089, 
4, 
880 
025 
420 
375 
262 
IIS 
584 
340 
900 
942 
Dani o 
Dani o 
Dani o 
Dani o 
Dani o 
Dani o 
Danio 
Danic 
Danic 
README 
rerio. 
rerio. 
rerio. 
rerio. 
rerxo. 
rerxo. 
rerxo. 
rer10. 
rer10. 
GRCZIO . 
GRCz10. 
GRCZIO . 
GRCZIO . 
GRCZIO. 
GRCZIO . 
GRCZIO . 
GRCZIO. 
GRCZIO. 
dna 
dna 
dna 
dna 
dna 
dna 
dna 
dna 
dna 
sm. 
sm. 
sm. 
sm. 
sm. 
chromosome . 
chromosome . 
chromosome . 
chromosome . 
chromosome . 
chromosome . 
chromosome . 
7 
fa. 
fa. 
fa. 
. fa. 
fa. 
fa. 
MT. fa .gz 
nonchromosomal . fa. gz 
toplevel . fa . gz 

1. When doing "Quantify to annotation model", retrieve URL for \*.gtf or \*.gff3 similarly from the website. Note that \*.gtf will give one result data node, while \*.gff3 will give 2 data nodes: Gene counts and Transcript counts.

From NCBI

Search for the assembly at: <https://www.ncbi.nlm.nih.gov/assembly>, then click from the list on the right "Download the RefSeq assembly" or "Download the GenBank assembly". ("README.txt" explains the differences).

Machine generated alternative text:
NCBI Resources lil Hmy-l.o 
Browse by organism 
Assembly 
Full Report 
GRCz11 
Assembly 
Advanced 
Search 
Help 
Send to: 
Access the data 
Browse in Genome Data Vlewer 
See Genome Information for 
Danio rerio 
Mew the Annotation Report 
Download the RefSeq assembly 
There are 4 assemblies for this 
Download the GenBank assemb 
organism 
BLAST search the assembly 
Download the full sequence report 
Download the statistics report 
Download the regions report 
Description: Genome Reference Consortium Zebrafish Build 11 
Organism name: Danio rerio (zebrafish) 
Infraspecific name: Strain. Tuebingen 
BioSample: 
Submitter. Genorne Reference Consortium 
Date: 2017/05/09 
Assembly type: haploid-with-alt-loci 
Assembly level: Chromosome 
Genome representation: full 
RefSeq category: reference genome 
nnnnnonQK 

\*.fna.gz is equivalent to compressed \*.fa.

Machine generated alternative text:
FTP directory /genomes/a11/GCF/OOO/002/035/GCF 000002035.6 GRCz11 at ftp.ncbi.nlm.nih.gov 
To view this FTP site in File Explorer: press Alt, click View, and then click Open FTP Site in File Explorer. 
Up to higher level directory 
GCE 
GCE 
GCE' 
GCF 
GCF 
GCE 
GCE 
GCF 
GCF 
GCE 
GCF 
GCF 
GCF 
000002035. 
000002035. 
000002035 . 
000002035. 
000002035. 
000002035. 
000002035. 
000002035 . 
000002035 . 
000002035. 
000002035. 
000002035. 
000002035. 
6 
6 
6 
6 
6 
6 
6 
6 
6 
6 
6 
6 
6 
GRCz11 
GRCZII 
GRCZII 
GRCZII 
GRCZII 
GRCZII 
CRCZII 
GRCZII 
GRCZII 
GRCZII 
CRCz11 
GRCz11 
GRCZII 
GRCZII 
assembly 
assembly 
assembly 
asserbly 
cds from 
regions . txt 
report . txt 
stats txt 
structure 
genomic. fna . gz 
07/17/2017 
07/17/2017 
07/17/2017 
07/17/2017 
07/17/2017 
07/17/2017 
07/17/2017 
07/17/2017 
07/17/2017 
08/02/2017 
08/02/2017 
07/17/2017 
07/17/2017 
OE/02/2017 
08/02/2017 
07/17/2017 
07/17/2017 
08/02/2017 
11/29/2017 
0B/02/2017 
252M 
2SPM 
25PM 
25PM 
OB. 
25PM 
OB. 
2SPM 
252M 
2SPM 
25PM 
07. 
52PM 
07 
. S2PM 
OB. 
2SPM 
252M 
07: 
S2PM 
07: 
52PM 
0B. 
25PM 
OB. 
25PM 
07. 
S2PM 
09 : 
19AM 
07: 
S2PM 
79, 
172, 
54,387 
Directory 
25, 
4, 
528, 
706, 
22, 
16, 
34, 
156, 
43, 
115, 
42, 
456, 
479 
667 
373, 
640 
965, 
929 
221, 
011 
660 
, 919 
939, 
526 
317, 
467 
631 
, 937 
145, 
68B 
153, 
681 
25 
410 
14 
297, 
676 
feature table. txt.gz 
encmlc. na. z 
encmic. Eff. z 
encmlc . 
protein . faa.gz 
protein . gpff . gz 
rm. out.gz 
rm. run 
rna. fna . gz 
rna. qbff . qz 
rna from genomic. fna.qz 
README. txt 
annotation hashes. txt 
assembly status . txt 
mdSchecksums . txt 