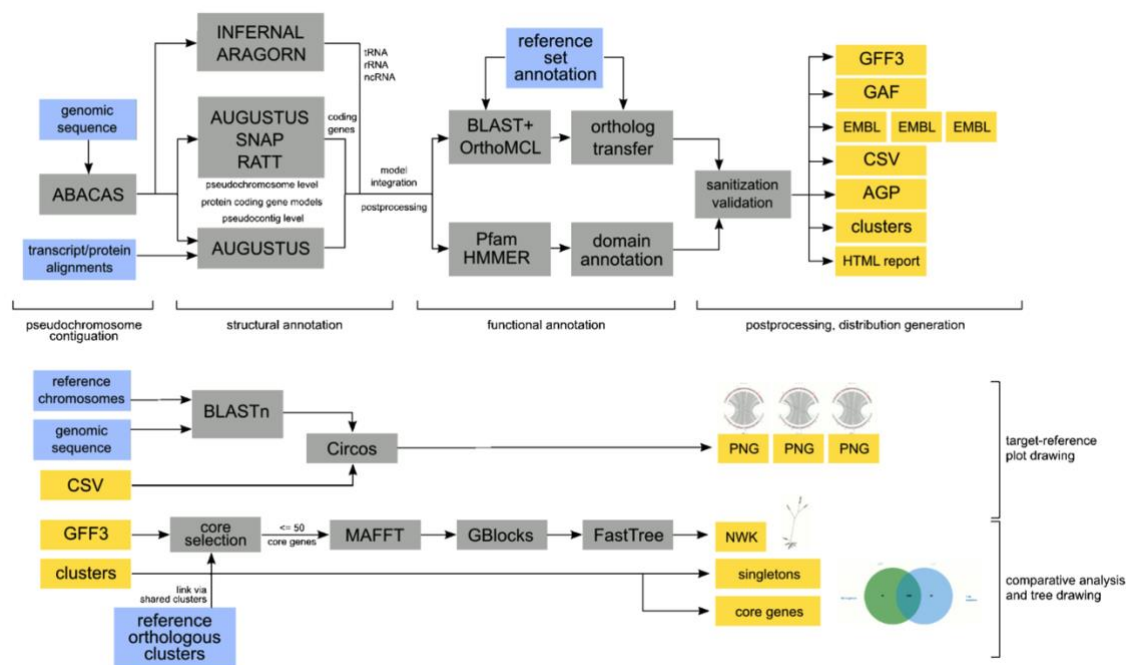


Genome Annotation with Companion (Part 1)

Learning objectives:

- Download genomes and chromosomes from VEuPathDB
- Annotate a genome with Companion
- Interpreting the Companion result
- Download the Companion output and load it into Artemis
- Learn how to use ACT to compare genomes and look at the Companion output

Companion, is an online pipeline that employs different software to annotate and compare an assembled sequence to a reference-annotated genome. The figure below illustrates the Companion pipeline, the software used and the expected output.



For this exercise, we will start with an assembled genome that is unannotated. We will obtain the assembled FASTA files from VEuPathDB sites. Companion can be accessed here: <http://companion.gla.ac.uk/>

Each group will download one of the following genomes (the tinyURL links will initiate the download) and will use Companion to compare with the specified genome as reference.

Group 1 – *Plasmodium coatneyi* Hackeri using *Plasmodium knowlesi* as reference
<https://tinyurl.com/y47vvsoj>

Group 2 - *Plasmodium coatneyi* Hackeri using *Plasmodium falciparum* as reference
<https://tinyurl.com/y47vvsoj>

Group 3 – *Cryptosporidium meleagridis* using *Cryptosporidium parvum* as reference
<https://tinyurl.com/y4fgc3p5>

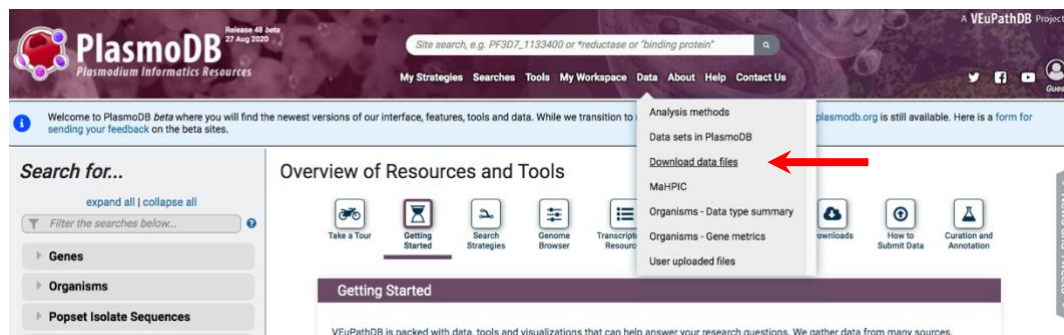
Group 4 *Cryptosporidium baileyi* using *Cryptosporidium parvum* as reference
<https://tinyurl.com/y44ucs5t>

Group 5 *Trypanosoma congolense* using *Trypanosoma brucei* 927 as reference.
<https://tinyurl.com/yxausbhg>

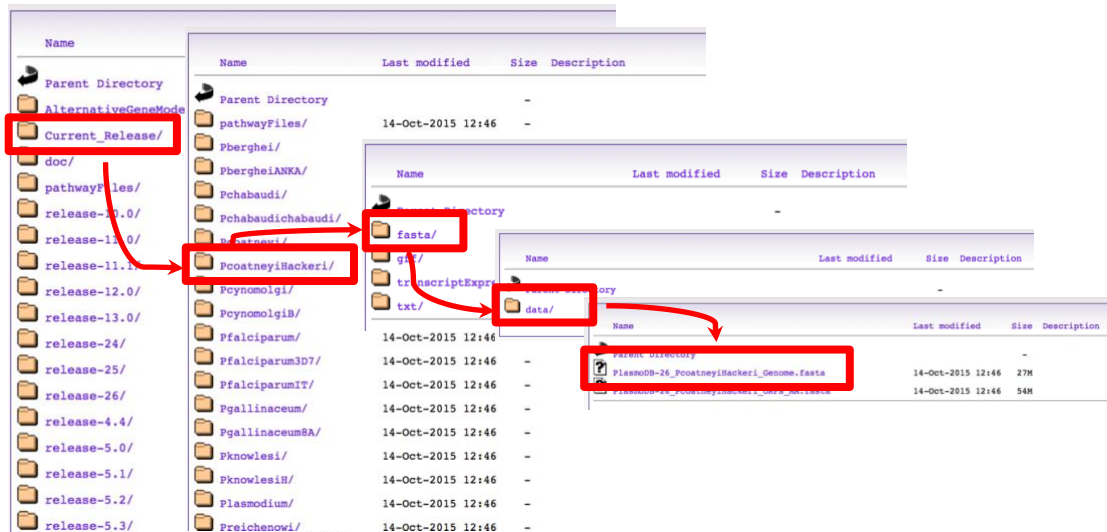
Group 6 *Trypanosoma congolense* 2019 using *Trypanosoma brucei* 927 as reference.
<https://tinyurl.com/y4pqscrm>

A word about downloads:

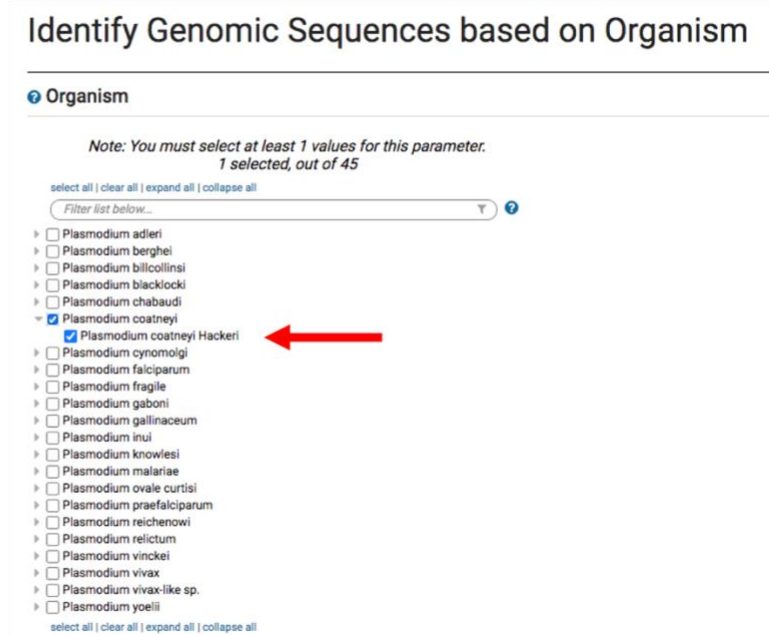
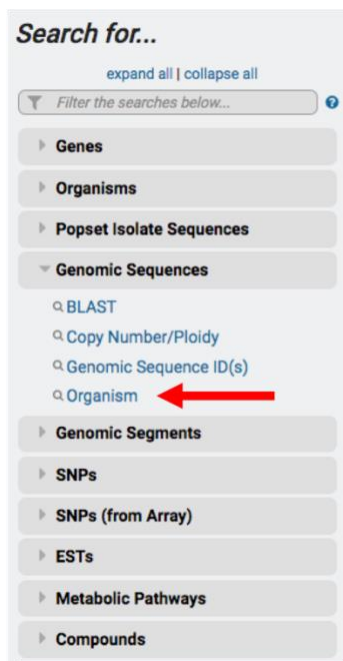
TinyUrls above are direct links to our genome FASTA files in the corresponding VEuPathDB site downloads section. All genomes in VEuPathDB sites are available for download from the “Data File” download section, which you can access from the top menu by clicking on “Data”.



Selecting the option “Download data files” takes you to the download directories where you can navigate to the genome and data type you are looking for.



To download specific contigs/scaffolds/chromosomes instead of entire genomes, use a genomic sequence search and place the desired sequences into your basket.



My Search Strategies

Opened (1) All (27) Public (42) Help

Unnamed Search Strategy *

Organism 14 Sequences Step 1

14 Genomic Sequences [Revise this search](#)

Genomic Sequence Results

Rows per page: 20

Download Add to Basket Add Columns

Sequence ID	Organism	Length
CP016250	Plasmodium coatneyi Hackeri	4,930,710
CP016249	Plasmodium coatneyi Hackeri	2,627,280
CP016247	Plasmodium coatneyi Hackeri	2,583,428
CP016246	Plasmodium coatneyi Hackeri	2,003,671

You can access the basket from the top menu.

PlasmoDB Plasmodium Informatics Resources Release 42 Data 27 Aug 2020

Site search, e.g. PF3D7_1133400 or "reductase" or "binding protein"

A VEuPathDB Project

My Strategies Searches Tools My Workspace Data About Help Contact Us

Analyze my data (Galaxy)
 My baskets
 My data sets
 My favorites

My Search Strategies

Opened (1) All (27) Public (42) Help

Unnamed Search Strategy *

My Baskets

Genomic Sequences (2)

Empty Basket Save Basket to Strategy

2 Genomic Sequences

Genomic Sequence Results

Rows per page: 20

Download Add Columns


Sequence ID	Organism	GenbankRecord	Genome Browser	BrowseURL
CP016247	Plasmodium coatneyi Hackeri	N/A	Browser	/plasmo/app/browse?loc=CP016247.1.2583428&data=/plasmo/service/browse/tracks/pcosal-hackeri&track...
CP016250	Plasmodium coatneyi Hackeri	N/A	Browser	/plasmo/app/browse?loc=CP016250.1.4930710&data=/plasmo/service/browse/tracks/pcosal-hackeri&track...

Retrieve Sequence


Nucleotide positions: 1 to 2583428
☐ Reverse & Complement
 Get Sequence

Nucleotide positions: 1 to 4930710
☐ Reverse & Complement
 Get Sequence

Download Genomic Sequence Basket


Choose a Report: ☐ Tab- or comma-delimited (openable in Excel) - choose columns to make a custom table [?](#)
☐ Tab- or comma-delimited (openable in Excel) - choose a pre-configured table [?](#)
☒ FASTA - sequence retrieval, configurable [?](#) 
☐ Standard JSON [?](#)

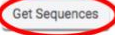
Choose the region of the sequence(s):

☐ Reverse & Complement 

Nucleotide positions: to (0 = end)

Download Type:

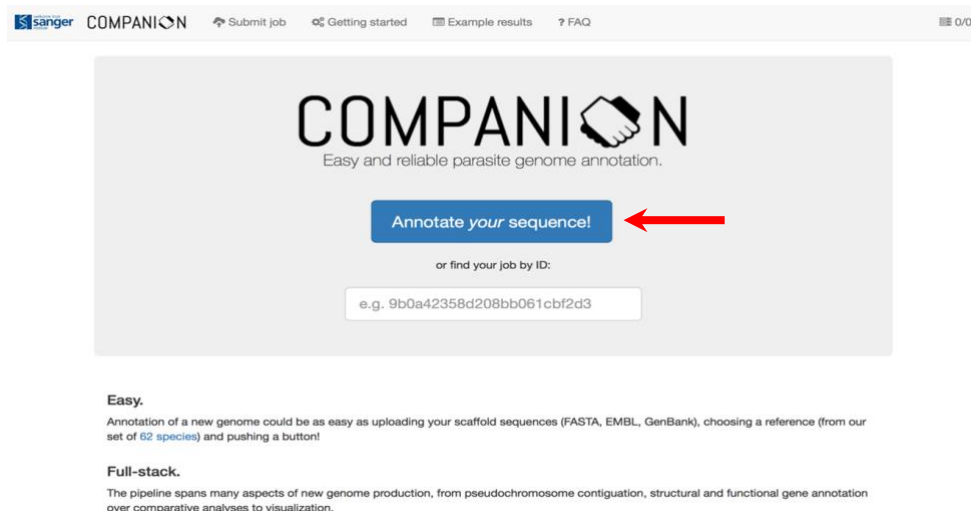
☒ Text File 
☐ Show in Browser

Get Sequences 

-**Back to the Annotation:** Once you have downloaded your sequence file, go to the Companion site:

<http://companion.gla.ac.uk/>

- Click on the “Annotate your sequence” link.



The screenshot shows the Companion website interface. At the top, there is a navigation bar with the Sanger logo and links for 'COMPANION', 'Submit job', 'Getting started', 'Example results', and 'FAQ'. The main content area features the 'COMPANION' logo with a handshake icon and the tagline 'Easy and reliable parasite genome annotation.' Below this is a large blue button labeled 'Annotate your sequence!' with a red arrow pointing to it. Underneath the button, it says 'or find your job by ID:' followed by a text input field containing the example ID 'e.g. 9b0a42358d208bb061cbf2d3'. At the bottom, there are two sections: 'Easy.' which states that annotation is as easy as uploading sequences and pushing a button, and 'Full-stack.' which describes the comprehensive pipeline from genome production to visualization.

-Follow the instructions as described on the Companion website:

1. Provide basic information about the job you are about to submit. This includes a job name, species prefix (usually the first letter of the genus and the first three letters of the species: *Cryptosporidium parvum* = CPAR).

Submit a new annotation job

Step 1: Basic job properties

First of all, please specify a free-text **name** for your new job. It should reflect the purpose of your job, and should probably include the organism you are annotating.

Example: *My new species annotation*

Job name

Please also give a short **species prefix** that will be used to name entities (such as genes, pseudochromosomes, etc.) generated during the annotation run. It should not contain spaces or special characters.

Example: *LDON*

Species prefix

LFOO

Finally, please provide a **species name** that describes the target species you are annotating.

Example: *Leishmania donovani*

Species name

Leishmania donovani

2. In step 2, choose the assembly file that you downloaded.
3. In step 3, indicate if you will be using RNAseq evidence to guide the annotation – in this exercise we will **not** use any RNAseq data.
4. In step 4, select the reference sequence you would like to use to transfer the annotation and to compare your sequence to. Typically, you would like to use a reference that is closely related, so a phylogenetic tree might be useful to look at. Here are examples of phylogenies for *Plasmodium* and *Cryptosporidium*.

<http://tolweb.org/Cryptosporidium/124803>

<http://tolweb.org/Plasmodium/68071>

Trypanosoma phylogenetic tree

https://projects.exeter.ac.uk/meeg/sites/default/files/pictures/tryp_tree.jpg

Step 2: Target sequence

Please upload a **target sequence file** to be annotated from your local filesystem using the button below. The file (FASTA, EMBL or GenBank format) can be gzip- or bzip2-compressed. In this case it must have a .gz or .bz2 suffix.

Note: The maximal size of your uploaded file is **64 MB**, and the maximum number of individual sequences in it is **3000**.

Choose File no file selected

[Here](#) is an example sequence input file for a *Plasmodium falciparum* IT chromosome 5 sequence that can be used with the *Plasmodium falciparum* 3D7 example reference set (choose below in step 4) for a quick example run. To use it, please download it to your local machine and upload it using the button above.

Step 3: Transcript evidence

The *Companion* pipeline can optionally make use of assembled transcripts in the GTF format as created by Cufflinks.

- ☐ Yes, use transcript evidence.
- ☒ No, do not use transcript evidence.

Step 4: Reference organism

Please pick a (if possible closely related) **reference organism** for this annotation run. This organism will be used to specify the models for gene finding, functional annotation transfer and pseudochromosome contiguation.

Please select a reference species

5. In step 5, there are a few more parameters you may want to examine. For the purpose of our exercise we will keep these at the default values.

Step 5: Pseudochromosome contiguation

The contiguation step will try to orientate the sequences in your input file to align with the chromosomal sequences of the reference organism to build pseudochromosomes, which will then be used as the target sequences for gene annotation. This step is optional; if it is not desired then no modifications will be made to the input sequences.

- ☒ Yes, contiguate pseudochromosomes.
- ☐ No, do not modify my input sequences.

Select minimum required match length for contig placement: 500 bp

200  20000

Select minimum required match similarity for contig placement: 85 %

30  100

6. Enter your email address to get an update when your job starts running and when it is complete. Next, click on the “I’m not a robot” captcha (Completely Automated Public Turing test to tell Computers and Humans Apart). Finally, click on the “Submit Job” link.

Step 6: Advanced settings (click chevron to the right to show/hide)



Your contact information (optional)

You can leave your email address if you want to be notified when your job starts and finishes. This is absolutely optional, if you choose not to share your email address, you can always manually check the status of your job using a private link provided by us after submission.

Email

To protect the service from automated bots, please prove that you are a human by ticking the box below.

☐

I'm not a robot



Submit job