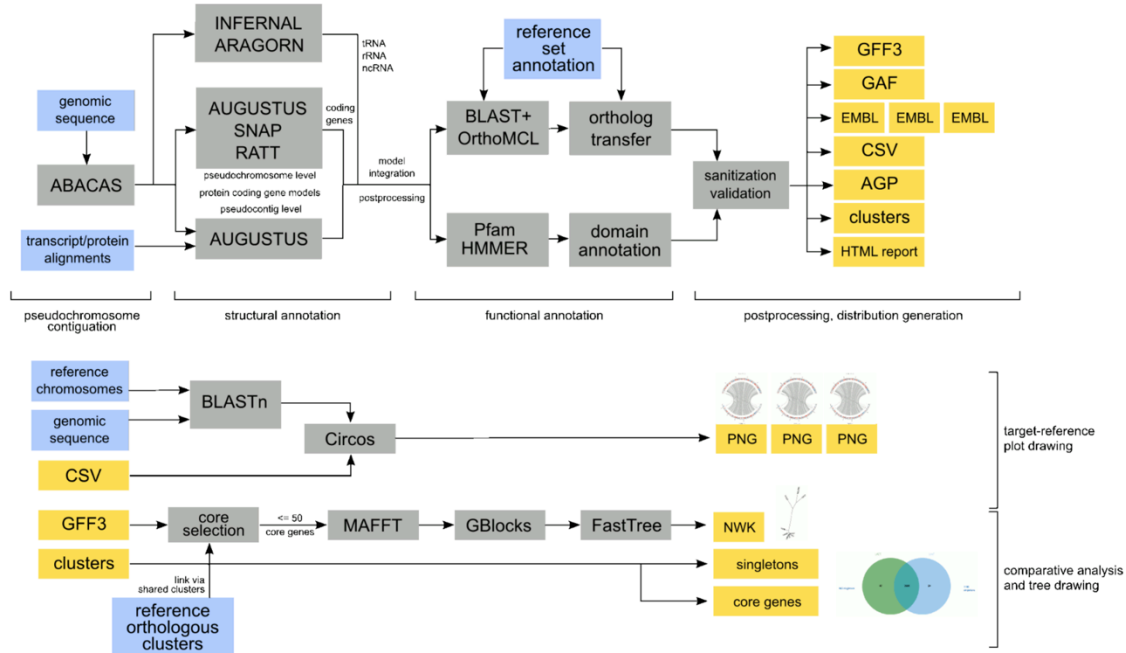


## Genome Annotation with Companion (Part 2)

You should have gotten an email indicating the status of your annotation (ie. job started and job complete). The email contains a link to the annotation output.



- Explore your results with your group and discuss the annotation findings:
  - What does the genome statics tab tell you about your annotation? Are the results surprising? You can explore the reference genome you used in EuPathDB to help you assess the results. (For example, are you getting a reasonable number of genes? What about the GC content? Number of non-coding genes?)

P\_coatneyi\_Hackeri (pcoa)

Completed

This job was submitted 4 days ago and ran for about 4 hours, finally finishing at 2016-02-13 08:49:01 UTC.

Genome statistics	Result files	Orthology	Phylogeny	Synteny	Job parameters	Pipeline logs	Validator report
							Value
Number of annotated regions/sequences							15
Number of genes							5191
Gene density (genes/megabase)							185.11
Number of coding genes							5126
Number of genes with function							4949
Number of non-coding genes							65
Number of genes with multiple CDSs							2704
Overall GC%							39.66
Coding GC%							41.91

- What does the “Result files” tab contain? What is an AGP file? What is a GFF3 file?

**P\_coatneyi\_Hackeri** (pcoa) Completed

This job was submitted **4 days** ago and ran for **about 4 hours**, finally finishing at **2016-02-13 08:49:01 UTC**.

Genome statistics **Result files** Orthology Phylogeny Synteny Job parameters Pipeline logs Validator report

	MD5	Size
<a href="#">Pseudochromosome level genomic sequence (FASTA)</a>	<a href="#">[MD5]</a>	7.81 MB
<a href="#">Pseudochromosome level genes (GFF3)</a>	<a href="#">[MD5]</a>	4.65 MB
<a href="#">Pseudochromosome layout (AGP)</a>	<a href="#">[MD5]</a>	32.6 KB
<a href="#">Scaffold level genomic sequence (FASTA)</a>	<a href="#">[MD5]</a>	7.82 MB
<a href="#">Scaffold level genes (GFF3)</a>	<a href="#">[MD5]</a>	7.95 MB
<a href="#">Scaffold layout (AGP)</a>	<a href="#">[MD5]</a>	165 KB
<a href="#">Pseudochromosome level sequence and annotation (EMBL)</a>	<a href="#">[MD5]</a>	13.5 MB
<a href="#">Functional GO annotation (GAF1)</a>	<a href="#">[MD5]</a>	1.56 MB
<a href="#">Protein sequences (FASTA)</a>	<a href="#">[MD5]</a>	3.56 MB

All files are provided in the following formats: GFF3 (format specification), FASTA (format specification), EMBL (format specification), GAF 1.0 (format specification) and AGP 2.0 (format specification).

- What does the “orthology” tab display? How many predicted proteins from your new genome are in common with ones from the reference genome? How many are unique to yours? What do singletons represent (click on the singleton number to see what these genes are)?

**P\_coatneyi\_Hackeri** (pcoa) Completed

This job was submitted **4 days** ago and ran for **about 4 hours**, finally finishing at **2016-02-13 08:49:01 UTC**.

Genome statistics **Result files** **Orthology** Phylogeny Synteny Job parameters Pipeline logs Validator report

These Venn diagrams show shared and species-specific gene clusters in the target genome pcoa (left, green) and the Plasmodium knowlesi reference (right, blue). Singletons, i.e. genes without orthologs and paralogs, are shown outside the Venn diagram to the left and right.

Click on the numbers to view clusters and singleton gene lists.

Cluster assignments as OrthoMCL 1.4 output file

- What does the phylogeny tab represent? Does it make sense?

[illegible]

- Examine the Synteny tab – are these genomes syntenic?

**P\_coatneyi\_Hackeri** (pcoa) Completed

---

This job was submitted **4 days** ago and ran for **about 4 hours**, finally finishing at **2016-02-13 08:49:01 UTC**.

[Genome statistics](#)   
 [Result files](#)   
 [Orthology](#)   
 [Phylogeny](#)

Synteny

[Job parameters](#)   
 [Pipeline logs](#)   
 [Validator report](#)

---

Each circle below represents a single target-reference pseudochromosome alignment. Click on the thumbnail to zoom in.

[Download all 15 images \(ZIP\)](#)