**The Cercopithifilaria johnstoni genome (Results section sketched out)**

**General “Genome report” paper size for G3 are rather small**

* **Anywhere from 3-6 figures and 1-3 tables seem to be consistent**

**“Investigations” are full length with no size restrictions**

**From reading a few papers and trying to assemble the results section I think we should aim for an “Investigation” paper taking an evolutionary pitch.**

**This way we could explore the interesting features of the possible CJ model of host switches and pathology. (Don’t think it’s possible in a “Genome report” because they are very concise to just the genome statistics). An investigation paper definitely looks like the format to go but happy for other opinions.**

**I think we agreed that we didn’t want to make this paper just a CJ vs OV comparative paper because it’s kind of a weak and forced comparison and I don’t have a lot of analyses that would strengthen that comparison.**

**My understanding was that we would still keep some of the comparisons I have illustrated, i.e. GC, orthologues, to illustrate that similarity point in discussion.**

**Abstract**

**Introduction**

**Materials and Methods**

\*I was leading in the direction that some papers have done in G3 of combining the results and discussion sections. I obviously haven’t done this before and open to suggestions and/or feedback.\*

\*This would most likely only work if we decided to go the “Genome Report” style of paper. If we agree on writing an “Investigation” it seems more suitable and sensible to split the results and discussion sections”, but again each paper seems to do slightly different styles depending on their data so I’m open to suggestions ☺\*

**Results and Discussion**

**Genome assembly description**

\*Begin by describing the CJ mitochondrial genome. I believe this would be better in the results than in the sup info to discuss the contents of the mtDNA genome\*

**Table 1**: Mitochondrial genome annotation, 13,716 bp

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Gene** | **Type** | **Start** | **End** | **Start codon** | **Length (bp)** | **Length (aa)** |
| trnE | tRNA | 2 | 59 |  | 58 |  |
| trnS1 | tRNA | 67 | 118 |  | 52 |  |
| ND2 | CDS | 119 | 880 | ATG | 762 | 254 |
| trnT | tRNA | 972 | 1027 |  | 56 |  |
| ND4 | CDS | 1028 | 2308 | TTG | 1281 | 427 |
| COX1 | CDS | 2310 | 3902 | ATT | 1593 | 531 |
| trnW | tRNA | 3906 | 3958 |  | 53 |  |
| ND6 | CDS | 3999 | 4427 | ATT | 429 | 143 |
| trnR | tRNA | 4450 | 4503 |  | 54 |  |
| CYTB | CDS | 4569 | 5615 | ATG | 1047 | 349 |
| trnL | tRNA | 5656 | 5710 |  | 55 |  |
| COX3 | CDS | 5711 | 6487 | ATT | 777 | 259 |
| trnA | tRNA | 6908 | 6964 |  | 57 |  |
| trnL | tRNA | 6967 | 7019 |  | 53 |  |
| trnN | tRNA | 7029 | 7091 |  | 63 |  |
| trnM | tRNA | 7090 | 7145 |  | 56 |  |
| trnK | tRNA | 7146 | 7202 |  | 57 |  |
| ND4L | CDS | 7212 | 7442 | ATT | 231 | 77 |
| rrnS | rRNA | 7444 | 8116 |  | 673 |  |
| trnY | tRNA | 8117 | 8171 |  | 55 |  |
| ND1 | CDS | 8178 | 9026 | TTG | 849 | 283 |
| trnF | tRNA | 9032 | 9085 |  | 54 |  |
| ATP6 | CDS | 9106 | 9687 | TTG | 582 | 194 |
| trnI | tRNA | 9693 | 9748 |  | 56 |  |
| trnG | tRNA | 9720 | 9777 |  | 58 |  |
| COX2 | CDS | 9857 | 10471 | ATT | 615 | 205 |
| trnH | tRNA | 10482 | 10536 |  | 55 |  |
| rrnL | rRNA | 10934 | 11442 |  | 509 |  |
| ND3 | CDS | 11443 | 11820 | TTA | 378 | 126 |
| trnC | tRNA | 11828 | 11884 |  | 57 |  |
| trnS2 | tRNA | 11885 | 11938 |  | 54 |  |
| trnP | tRNA | 11940 | 11997 |  | 58 |  |
| trnD | tRNA | 12002 | 12059 |  | 58 |  |
| trnV | tRNA | 12060 | 12115 |  | 56 |  |
| ND5 | CDS | 12124 | 13716 | GTT | 1593 | 531 |

\*Follow by describing the nuclear genome\*

\*Rather than separating BUSCO and CEGMA tables, it appears some papers have combined BUSCO only results, so I have combined the relevant BUSCO genome completeness results within the assembly stats table.\*

\*I have still included the comparative information for all other filarial nematodes. I feel it will still be relevant for discussions and later on with the phylogenetic relationships.\*

\*Contents of genome assembly table will likely changed based on genome improvements\*

**Table 2**: Complete genome assembly statistics

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Species** | **Assembly length** | **Sequences (n)** | **Longest sequence** | **N50 length** | **N50**  **(n)** | **BUSCO annotation** |
| Cj NEW | 76,924,992 | 2091 | 588,166 | 99,003 | 232 | C:94.9%[S:94.2%,D:0.7%],F:3.9%,M:1.2%,n:982 |
| C. johnstoni | 79,062,707 |  | 588,165 | 88,758 |  | C:94.1% [D:0.7], F:4.4%, M:1.5%, n = # |
| A. viteae | 77,350,906 |  | 172,453 | 25,808 |  | C:90.5% [D1.7%], F:7%. M:2.5%, n = # |
| B. malayi | 88,235,797 |  | 24,943,668 | 14,214,749 |  | C:96.8% [D:1.1%], F:1.8%, M:1.4%, n = # |
| B. pahangi | 90,545,113 |  | 1,059,003 | 65,666 |  | C89.7% [D:0.7%], F:6.6%, M:3.7%, n = # |
| B. timori | 64,930,714 |  | 45,194 | 4,919 |  | C:53.3% [D:0.4%], F:21.1%, M:25.6%, n = # |
| D. immitis | 88,309,529 |  | 1,085,577 | 71,281 |  | C:91.8% [D:2.1%], F:4.5%, M:3.7%, n = # |
| L. loa | 96,405,338 |  | 1,570,872 | 180,288 |  | C:97.5% [D:1.3%], F:2%, M:0.5%, n = # |
| L. sigmodontis | 64,813,410 |  | 402,953 | 45,863 |  | C:91.6% [D:1.9%], F:5.5%, M:2.9%, n = # |
| O. volvulus | 96,427,137 |  | 28,345,163 | 25,485,961 |  | C:97.6% [D:0.3%], F:1.7%, M:0.7%, n = # |
| O. ochengi | 91,660,559 |  | 230,458 | 16,199 |  | C:85.6% [D:0.5%], F:9.5%, M:4.9%, n = # |
| O. flexuosa | 86,175,476 |  | 55,543 | 2,943 |  | C:47.3% [D:0.2%], F:21.7%, M:31%, n = # |
| W. bancrofti | 76,991,470 |  | 144,416 | 9,917 |  | C:74.9% [D:0.4], F:11.5%, M:13.6%, n = # |

\*Continue describing the CJ genome by discussing repeats\*

**Table 3:** Comparison of the RepeatModeler repeat statistics of *Cercopithifilaria johnstoni* and closely related filarial nematodes. Total percentages (%) of interspersed repeats consisting of LINEs (long interspersed nuclear elements), SINEs (short interspersed nuclear elements), DNA (DNA Transposons), LTR (long terminal repeat) and unclassified repeats. Small RNA, satellites, simple repeats and low complexity repeats are also represented in the table. The percentage of masked repeat bases in each filarial genome are represented as “bases masked”.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Species** | **bases masked** | **SINEs** | **LINEs** | **LTR** | **DNA** | **Unclassified** | **Total interspersed repeats** | **Small RNA** | **Satellites** | **Simple repeats** | **Low complexity** |
| Cj NEW | 5.23 | 0.00 | 0.00 | 0.09 | 0.05 | 1.52 | 1.66 | 0.00 | 0.00 | 2.86 | 0.71 |
| C. johnstoni | 6.23 | 0.00 | 0.15 | 0.00 | 0.07 | 2.20 | 2.41 | 0.00 | 0.01 | 3.11 | 0.70 |
| A. viteae | 3.96 | 0.00 | 0.01 | 0.44 | 0.18 | 0.59 | 1.22 | 0.00 | 0.00 | 2.08 | 0.66 |
| B. malayi | 11.80 | 0.00 | 0.43 | 1.26 | 0.26 | 4.12 | 6.07 | 0.12 | 0.26 | 4.28 | 1.20 |
| B. pahangi | 7.88 | 0.03 | 0.37 | 0.46 | 0.28 | 1.04 | 2.18 | 0.03 | 0.01 | 4.37 | 1.33 |
| B. timori | 5.02 | 0.00 | 0.26 | 0.44 | 0.16 | 0.59 | 1.44 | 0.01 | 0.06 | 2.77 | 0.75 |
| D. immitis | 6.10 | 0.02 | 0.19 | 0.58 | 0.10 | 1.33 | 2.22 | 0.08 | 0.06 | 2.94 | 0.83 |
| L. loa | 13.61 | 0.04 | 0.92 | 2.56 | 0.93 | 3.84 | 8.29 | 0.15 | 0.28 | 4.14 | 0.81 |
| L. sigmodontis | 2.82 | 0.00 | 0.03 | 0.16 | 0.09 | 0.17 | 0.44 | 0.00 | 0.00 | 1.91 | 0.47 |
| O. flexuosa | 8.06 | 0.00 | 0.06 | 1.47 | 0.06 | 3.87 | 5.47 | 0.04 | 0.02 | 2.02 | 0.53 |
| O. ochengi | 10.21 | 0.06 | 0.99 | 1.12 | 0.30 | 4.26 | 6.75 | 0.11 | 0.20 | 2.51 | 0.72 |
| O. volvulus | 10.26 | 0.00 | 0.10 | 1.29 | 0.17 | 5.02 | 6.57 | 0.08 | 0.09 | 2.73 | 0.79 |
| W. bancrofti | 6.96 | 0.06 | 0.08 | 0.86 | 0.19 | 0.48 | 1.68 | 0.00 | 0.08 | 4.12 | 1.09 |

**FLAG!**

\*One point that I made in my thesis but didn’t have any figures or tables for, was the absence of Wolbachia genome.

Will it be enough to say that there isn’t a intact Wolbachia genome in a paper? Or should I create maybe a table that highlights the low number of reads mapping to Wolbachia?

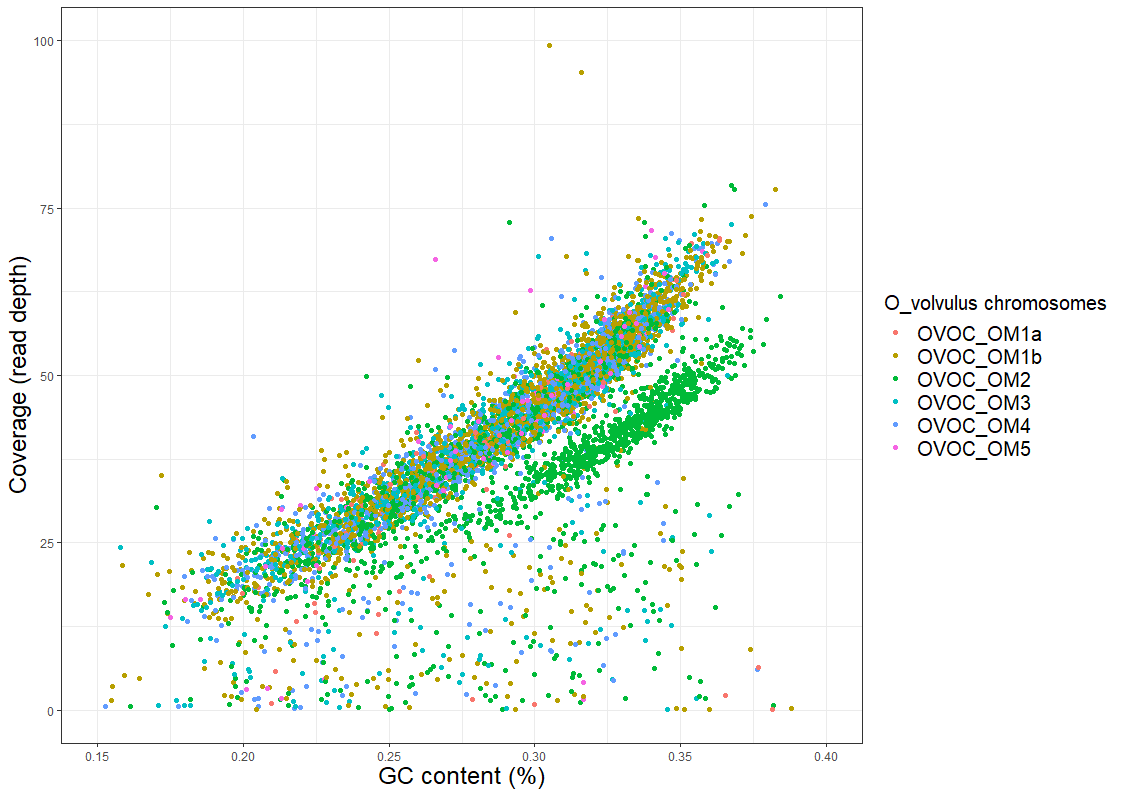
The Wolbachia angle of the paper is the main section that I think is missing information\*

\*By this stage in the paper the complete CJ genome paper has been described, the rest of the paper is now focussing on interesting aspects of CJ biology or relationships, etc\*

\*Discuss the X chromosome analysis, GC vs coverage, CJ possible X-linked chromosome etc\*

\*Interesting discussion of CJ biology\*

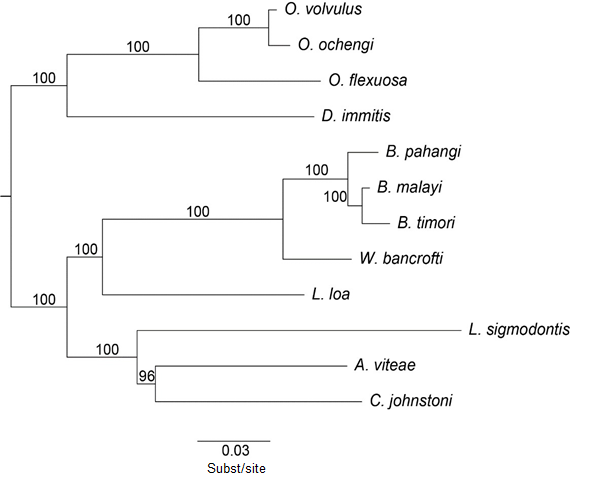
**X chromosome analysis**



**Figure 1**: GC content vs coverage plot of the draft *Cercopithifilaria johnstoni* genome scaffolds with an *Onchocerca volvulus* chromosome projection overlayed onto the *Cercopithifilaria johnstoni* data based on nucleotide similarity. OM2 is the X chromosome for *Onchocerca volvulus*.

\*Compare the mitochondrial genome phylogeny with the whole nuclear genome phylogeny. Combine these in the same figure for easy and concise comparison in the paper. Main point that the relationships are highly supported and the same. This could then open up discussion on the host switch hypothesis.\*

**Filarial nematode phylogenies**



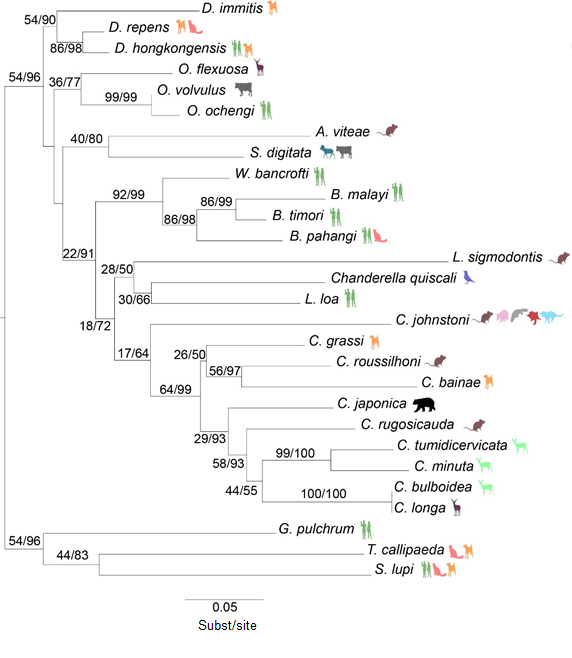
b)

a)

**Figure 2**: Filarial nematode comparison of the a) mitochondrial genome using maximum likelihood/Bayesian inference, and b) whole genome using single-copy orthologue maximum likelihood phylogeny.

\*Originally I was thinking of adding the spp phylogeny in the figure above and illustrate the 3 trees but think because it’s a COI-12S tree should keep it separate? (Thoughts?). This would be the figure that would go into more detail about possible host switching within the Cerco genus and maybe a pathology discussion.\*

**Cercopithifilaria genus phylogeny**

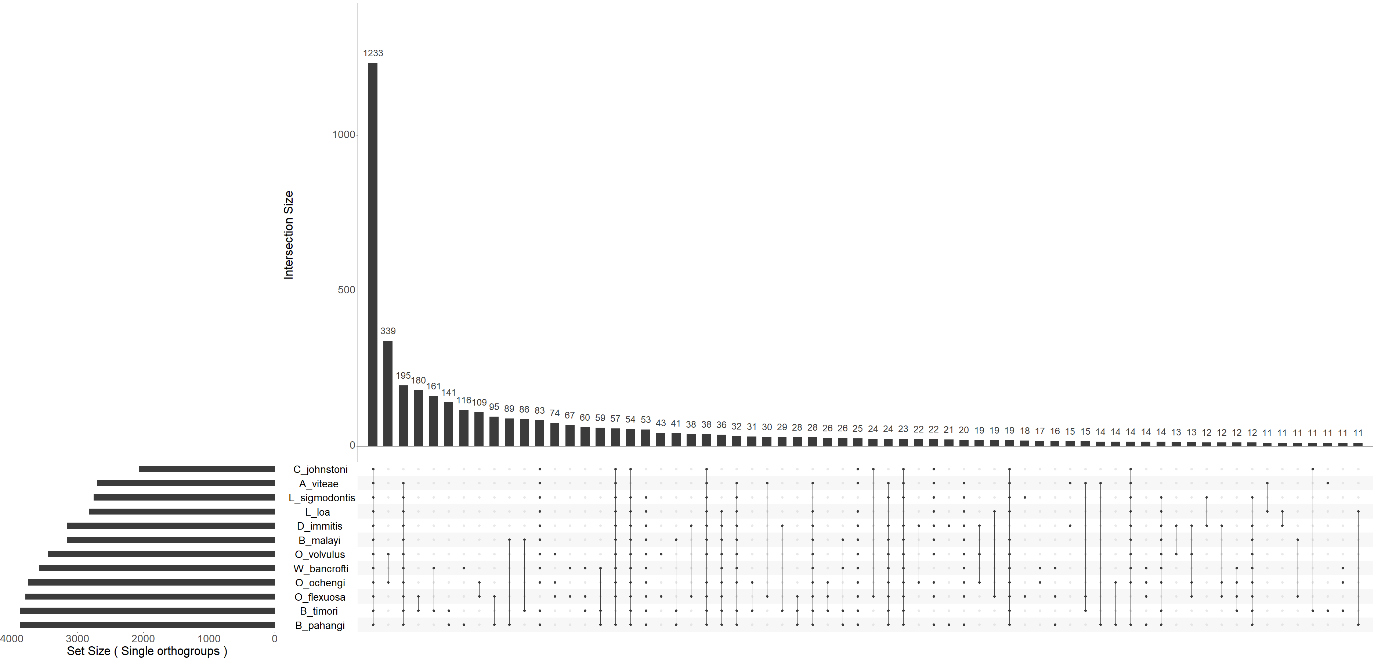


**Figure 3**: Concatenated 12S-COI phylogeny of filarial nematodes to position the *Cercopithifilaria* genus using maximum likelihood/Bayesian inference.

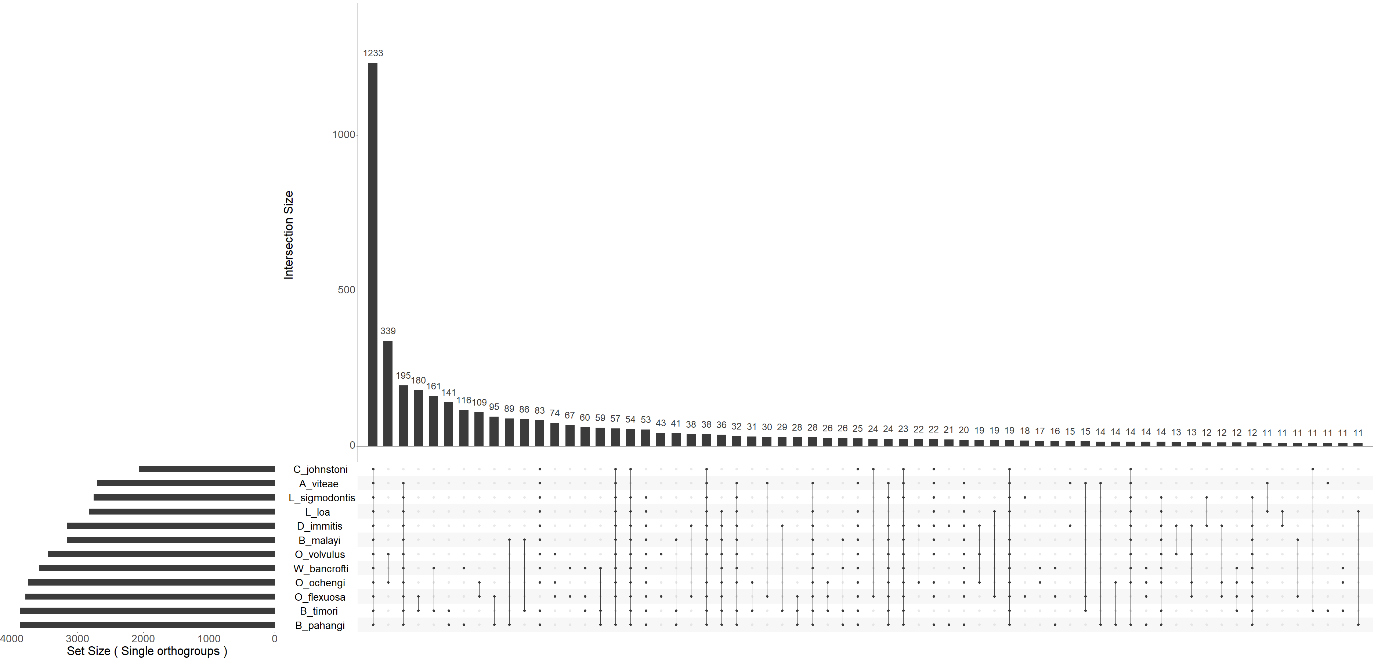
\*Lastly, move onto the immunogenic one-to-one orhtologues CJ and OV share. Combine the following two images to illustrate the single-copy/multi-gene orthogroup relationships rather than a table.\*

\*Wasn’t thinking of including any of the OrthoFinder tables I had in my thesis. What are your thoughts here? I’m unsure how to finish this last section\*

**Orthologues**



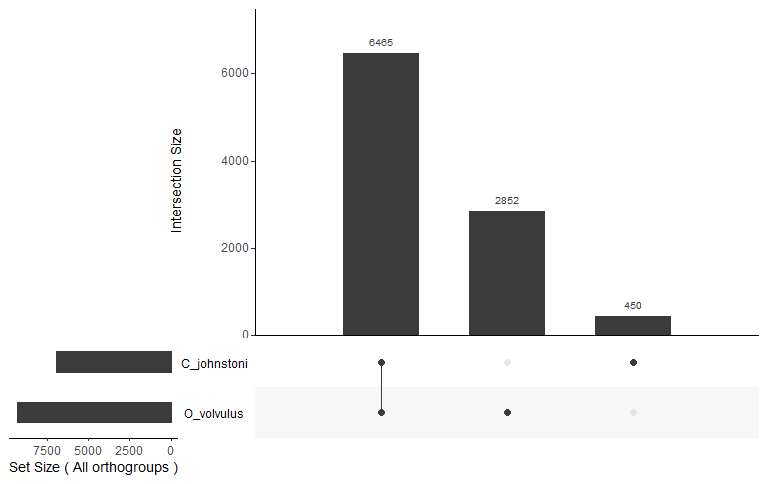
a)



b)

**Figure 4:** a) UpSet plot representing the single-copy orthologue orthogroup relationships across the 12 filarial nematodes. b) UpSet plot representing all the orthogroup relationships with one or more genes in each group across the 12 filarial nematodes.

\*Optional: include the direction comparison between CJ and OV. I think this would be a good addition to then lead into the last table which will be the CJ-OV 30 orthologues BLAST results. Thoughts?\*



**Figure 5:** UpSet plot representing the complete single-copy and multi-gene orthogroup relationships across *Cercopithifilaria johnstoni* and *Onchocerca volvulus*.

\*Finish with a description of the 30 one-to-one orthologues between CJ and OV that could be useful for future analyses and experimentations\*

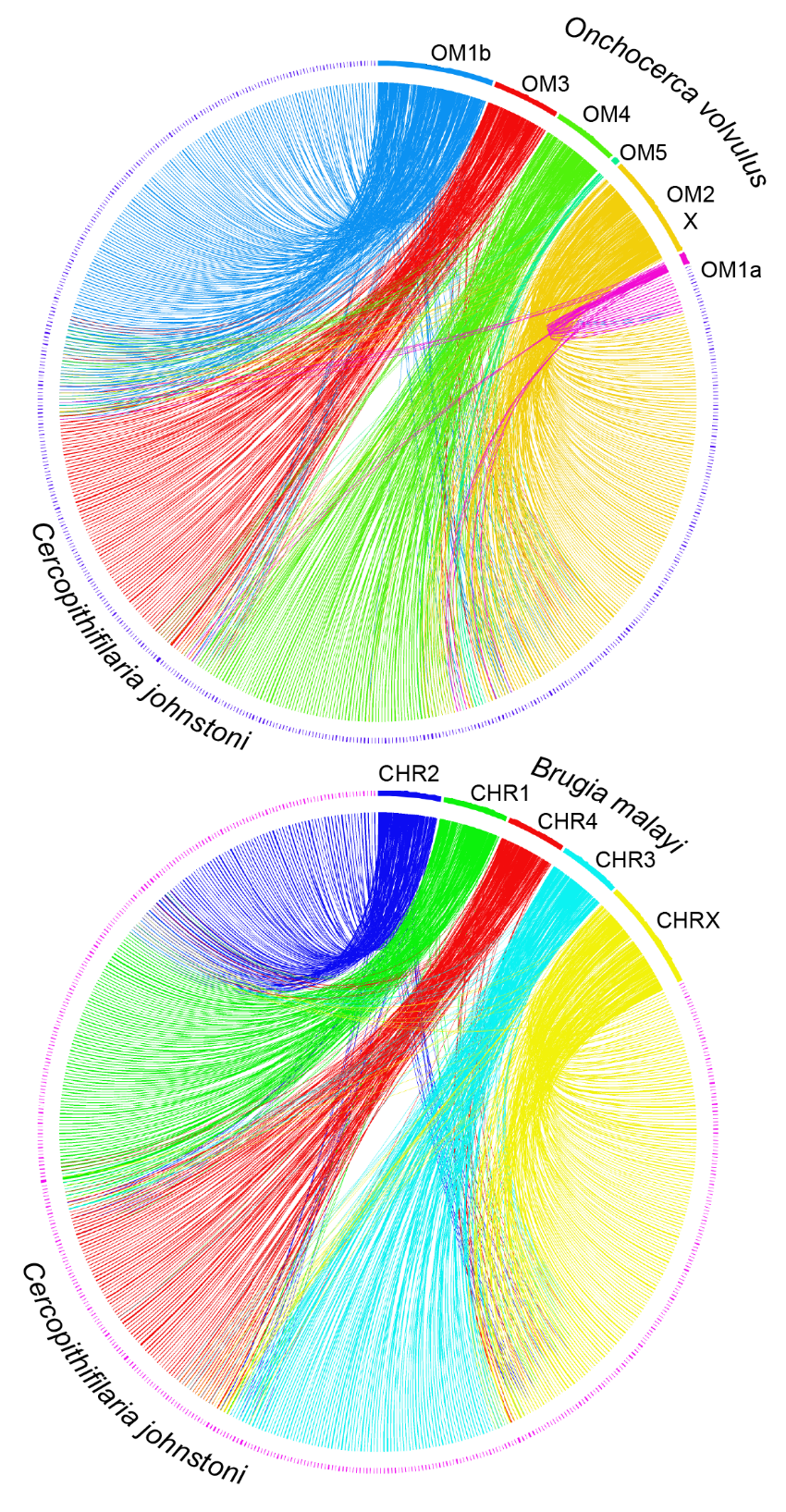
**Table 4:** The 30 *Onchocerca volvulus* immunogenic proteins and corresponding *Cercopithifilaria johnstoni* orthologues are described revealing their top BLAST results to the NCBI database, % identity, protein alignment length, e-value, bit score, location of expression in the cell (derived from *Onchocerca volvulus* experimentation), is the protein secreted and the % amino acid alignment similarity. Lastly, the orthology relationships between *Cercopithifilaria johnstoni* and *Onchocerca volvulus* are described.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Query (*C. johnstoni*)** | **Query (*O. volvulus*)** | **subject (blast NCBI)** | **description (NCBI)** | **OVOC putative function (WBP)** | **% identity** | **alignment length** | **e-value** | **bit score** | **Location in cell, surface, secreted?** | **When expressed** | **Orthology (1-to-1, 1-to-many etc.)** |
| CJg3607.t1 | OVOC860 | P37801.2 | Calponin homolog OV9M | Calponin homolog OV9M | 99.72 | 357 | 0 | 741 | longitudinal muscles below the hypodermis | all life stages | 1-to-1 |
| CJg6299.t1 | OVOC1897 | CRZ21860.1 | BMA-EAT-20 [*B. malayi*] | EAT-20 | 46.687 | 649 | 2.04E-158 | 488 | hypodermis | adult, L3 | 1-to-1 |
| CJg4216.t1 | OVOC1952 | EJW86002.1 | hypothetical protein WUBG\_03087 [*W. bancrofti*] | A0A2K6VRV9 - uncharacterised protein (*C. elegans* orthologue C34E7.4) | 72.973 | 259 | 1.33E-140 | 403 | excretory/ secretory | all life stages | many-to-1 |
| CJg6164.t1 | OVOC2486 | EJD74931.1 | hypothetical protein LOAG\_17826 [*L. loa*] | A0A2K6VUT6 - uncharacterised protein OR ShK domain | 58.497 | 306 | 9.64E-94 | 295 | intracellular | all life stages | 1-to-1 |
| CJg5604.t1 | OVOC3177 | EFO26240.2 | hypothetical protein LOAG\_02238 [*L. loa*] | A0A044SS43 - PHD-type (ZINC finger) domain-containing protein | 79.15 | 1271 | 0 | 1868 | nucleus, secretory vesicles | L3 | 1-to-1 |
| CJg319.t1/g320.t1 | OVOC3203 | XP\_001891888.1 | CAP protein [*B. malayi*] | Adenylyl cyclase-associated protein | 71.494 | 1091 | 0 | 1509 | excretory/secretory | all life stages | 1-to-1 |
| CJg6447.t1 | OVOC5419 | EJW87928.1 | 3-hydroxyacyl-CoA dehydrogenase [*W. bancrofti*] | A0A044TS06 - uncharacterised protein (*C. elegans* orthologue ard-1) | 94.841 | 252 | 1.05E-176 | 494 | hypodermis | all life stages | 1-to-1 |
| CJg4360.t1 | OVOC5823 | EJW78284.1 | hypothetical protein WUBG\_10805 (CTLH/CRA C-terminal to LisH motif domain) [*W. bancrofti*] | A0A044TYD3 - uncharacterised protein (*C. elegans* orthologue unc-60) | 94.894 | 235 | 2.53E-153 | 434 | intracellular | all life stages | 1-to-1 |
| CJg1898.t1 | OVOC7314 | AAG44695.1 | glutathione S-transferase Ia [*O. volvulus*] | Glutathione S-transferase | 100 | 250 | 0 | 514 | outer surface of hypodermis | all life stages | 1-to-many |
| CJg342.t1 | OVOC7381 | VDK80440.1 | unnamed protein product [*O. ochengi*] | Reticulon-like protein | 82.308 | 130 | 3.25E-66 | 213 | Endoplasmic reticulum | all life stages | 1-to-1 |
| CJg6599.t1 | OVOC7786 | AAD38403.1 | fructose 1,6 bisphosphate aldolase [*O volvulus*] | Fructose-bisphosphate aldolase | 100 | 363 | 0 | 756 | body wall muscle/reproductive tract adult females | L3 | 1-to-1 |
| CJg6746.t1 | OVOC7911 | P11012.2 | Calreticulin (RAL1 antigen) | A0A044UWC0 - uncharacterised protein (*C. elegans* orthologue crt-1 - calreticulin) | 99.408 | 338 | 0 | 681 | Endoplasmic reticulum lumen | larval | 1-to-1 |
| CJg2910.t1 | OVOC8600 | XP\_003143117.1 | SH2 domain-containing protein *[L. loa*] | SH2 domain-containing protein | 70.43 | 186 | 9.32E-86 | 261 | cytoskeleton |  | 1-to-1 |
| CJg10.t1 | OVOC8754 | Q25619.1 | Fatty-acid and retinol-binding protein 1 (Ov20) | Fatty-acid and retinol-binding protein 1 | 100 | 173 | 5.40E-121 | 347 | membrane |  | 1-to-1 |
| CJg4762.t1 | OVOC8985 | CDQ04841.1 | BMA-CYN-16 [*B. malayi*] | PPIase cyclophilin-type domain-containing protein (*C. elegans* orthologues tpi-1) | 85.281 | 462 | 0 | 719 | cytoplasm | all life stages | 1-to-1 |
| CJg2636.t1 | OVOC9225 | CRZ22798.1 | Bm5820, partial [*B. malayi*] | A0A044V9G2 - uncharacterised protein | 90.82 | 512 | 0 | 931 | cuticle, hypodermis, excretory/ secretory | adult | 1-to-1 |
| CJg4836.t1 | OVOC9575 | AAB69625.2 | activation-associated secreted protein-1 [*O. volvulus*] | Activation-associated secreted protein-1 | 96.216 | 185 | 7.11E-131 | 374 | Granules of GE, excretory/secretory | L3 stage | 1-to-1 |
| CJg6815.t1 | OVOC9592 | EJD73678.1 | hypothetical protein LOAG\_18910 (TTR-52) [*L. loa*] | A0A044VD77 - uncharacterised protein (TTR-52) | 71.014 | 138 | 9.20E-57 | 182 | extracellular domain, excretory/ secretory | L3, mf | 1-to-1 |
| CJg1453.t1 | OVOC9752 | AAC77922.1 | peroxidoxin-2 [*O. ochengi*] | A0A044VFS3 - uncharacterised protein (*C. elegans* orthologue prdx-2) | 95.652 | 184 | 2.61E-128 | 367 | excretory/ secretory | all life stages | 1-to-1 |
| CJg4068.t1 | OVOC9984 | CAA31690.1 | immunodominant antigen, partial [*O. volvulus*] | Immunodominant antigen Ov33-3 | 100 | 132 | 3.22E-92 | 275 | extracellular region or secreted | L3 | 1-to-1 |
| CJg3087.t1 | OVOC9988 | ACB70199.1 | rainforest immunodominant hypodermal antigen [*O. ochengi*] | OV-17 antigen | 99.39 | 164 | 4.19E-114 | 328 | hypodermal layer (excretory/ secretory) | all life stages | 1-to-1 |
| CJg3088.t1 | OVOC9990 | AAB53809.1 | RAL-2 homolog, partial [*A. viteae*] | DUF148 domain-containing protein | 50.806 | 124 | 1.18E-36 | 130 | hypodermal, excretory/ secretory | adult, L3 | 1-to-1 |
| CJg3219.t1 | OVOC10067 | EYC07471.1 | hypothetical protein Y032\_0070g452 [*Ancylostoma ceylanicum*] | A0A044QKJ7 - uncharacterised protein (c. elegans orthologue unc-52) | 49.981 | 2593 | 0 | 2380 | extracellular region or secreted, basement membrane | all life stages | 1-to-1 |
| CJg2309.t1 | OVOC10103 | KHN78795.1 | Neprilysin-2 [*T. canis*] | A0A044QL85 - uncharacterised protein (*C. elegans* orthologue nep-11) | 72.48 | 625 | 0 | 985 | transmembrane | all life stages | 1-to-1 |
| CJg7856.t1 | OVOC10995 | XP\_001898563.1 | hypothetical protein Bm1\_35580 [*B. malayi*] | A0A2K6VGV2 - uncharacterised protein OR ALDH5A1 - succinate-semialdehyde dehydrogenase | 37.615 | 109 | 2.41E-06 | 52 | pharyngeal muscle cell | adult, L3 | 1-to-1 |
| CJg5428.t1 | OVOC11517 | P24706.1 | Superoxide dismutase [Cu-Zn] | Superoxide dismutase [Cu-Zn] | 100 | 151 | 9.05E-105 | 304 | cytoplasm |  | 1-to-1 |
| CJg191.t1 | OVOC11847 | CRZ23009.1 | Bm8924 (Chondroitin proteoglycan 4) [*B. malayi*] | A0A044RF89 - uncharacterised protein OR RNA-binding protein Nova | 45.626 | 423 | 1.80E-92 | 293 | neurons | all life stages | 1-to-1 |
| CJg3662.t1 | OVOC12400 | EJW76014.1 | hypothetical protein WUBG\_13076, partial [*W. bancrofti*] | A0A044RNX4 - uncharacterised protein (*C. elegans* orthologue C08E8.2 - adipocyte plasma membrane associated protein) | 67.273 | 110 | 1.14E-47 | 158 | endoplasmic reticulum | adult, L3 | 1-to-1 |
| CJg2412.t1 | OVOC12449 | CDQ06623.1 | Bm8327 [*B. malayi*] | A0A044RPP5 - uncharacterised protein | 64.865 | 74 | 1.98E-31 | 119 | transmembrane | adult female | 1-to-1 |
| CJg6767.t1 | OVOC12769 | AAD27587.1 | SLAP-1 protein [*O. volvulus*] | Larval 18 kDa protein; Ov64; SLAP-1 protein | 93.443 | 122 | 9.20E-75 | 226 | granules of GE, cuticle | L3 | 1-to-1 |

\*I would like to include one circos plot but are unsure of where this would fit in the story and if it would fit. I’m aware of the abundant microsynteny so understand there is minimal information that can be said about the plot and is probably unnecessary to include.

It is also a direct comparison with O. volvulus. So it would be placed in the context of comparing CJ genome content with OV.

What are your thoughts about including this plot? (Possibly before the GC v Cov plot in the genome assembly description section)\*



**Conclusions**

**Acknowledgements**

**Literature cited**