Supplementary INFORMATION

**Sup. 1. Bioinformatic programs and parameters**

*Trimming and k-mer filtering*

The quality of sequencing reads was assessed via FastQC. Following this, the beginning and ends of each read were trimmed to remove adaptor sequences. Next, extremely frequent k-mers (more than 100x), and extremely rare k-mers were removed. Seqtk was used to truncate sequences containing unresolved "N"s and nucleotides of particularly low quality. Seqtk was then used to remove reads that were short, or no longer had a paired read. Reads were then assessed again using FastQC.

*# 1. Trim adaptors from ends of both paired end reads*

**seqtk** trimfq -b 5 -e 5 input\_reads\_pe1.fastq.gz **| gzip >** input\_reads\_pe1.trimmed.fq.gz

**seqtk** trimfq -b 5 -e 5 input\_reads\_pe2.fastq.gz **| gzip >** input\_reads\_pe2.trimmed.fq.gz

*# 2. Interleave Fastqs (required for khmer operation)*

**seqtk** mergepe input\_reads\_pe1.trimmed.fq.gz input\_reads\_pe2.trimmed.fq.gz **>** input\_reads\_pe1-2.trimmed.fq

*# 3. Remove coverage above 100x*

**khmer** normalize-by-median.py -p --ksize 20 -C 100 -M 1e9 -s kmer.counts \

 **-o** input\_reads\_pe1-2.trimmed.max100.fq input\_reads\_p1-2.trimmed.fq

*# 4. Filter low abundance kmers*

**khmer** filter-abund.py -V kmer.counts \

 -o input\_reads\_pe1-2.trimmed.max100.norare.fq \

 input\_reads\_pe1-2.trimmed.max100.fq

*# 5. Remove low quality bases, short sequences, and non-paired reads*

**seqtk** seq -q 10 -N -L 80 input\_reads\_pe1-2.trimmed.max100.norare.fq **|** **\**

 **seqtk** dropse **>** input\_reads\_pe1-2.trimmed.max100.norare.noshort.fq

*# 6. De-interleave filtered reads*

**khmer** split-paired-reads.py input\_reads\_pe1-2.trimmed.max100.norare.noshort.fq -d destination\_directory/

*Bowtie2 read mapping*

Bowtie2 was used to map genome sequencing reads against a previously published *Dientamoeba fragilis* transcriptome and (Barratt *et al.*, 2015), as well as the *Trichomonas vaginalis* shotgun genome (Carlton *et al.*, 2007).

*# 1. Build Bowtie2 index from reference genome/transcriptome*

**bowtie2-build** --threads 6 /path-to/reference.fasta /destination/index\_name

*# 2. Map reads to Bowtie2 index*

**bowtie2** -p 8 -q –very-sensitive-local –score-min G,10,1 -x index\_name. -1 filtered\_reads\_mate1.fq -2 filtered\_reads\_mate2.fq

*metaSPAdes assembly and Bandage visualisation*

MetaSPAdes was used to assemble both the original whole genome sequence reads and the sequencing reads derived from the 28S-18S rDNA long range PCR product

**metaspades.py** -k 21,33,55,77 -1 input\_reads\_pe1.trimmed.max100.norare.noshort.fq -2 input\_reads\_pe2.trimmed.max100.norare.noshort.fq -o output\_directory

The metaSPAdes assembly\_graph.fastg file was then used as an input for the Bandage windows client which visualises the contigs and paths between contiguous nodes.