

---

# **pathos Documentation**

***Release 0.1***

**Michael Panciera**

**Jul 22, 2019**



---

## Contents

---

<b>1</b>	<b>Overview</b>	<b>3</b>
<b>2</b>	<b>Installation</b>	<b>5</b>
<b>3</b>	<b>Configuration</b>	<b>7</b>
<b>4</b>	<b>Single Sample</b>	<b>9</b>
<b>5</b>	<b>Pathos Sheet</b>	<b>11</b>
<b>6</b>	<b>Summary Creation</b>	<b>13</b>
<b>7</b>	<b>Indices and tables</b>	<b>15</b>



Contents:



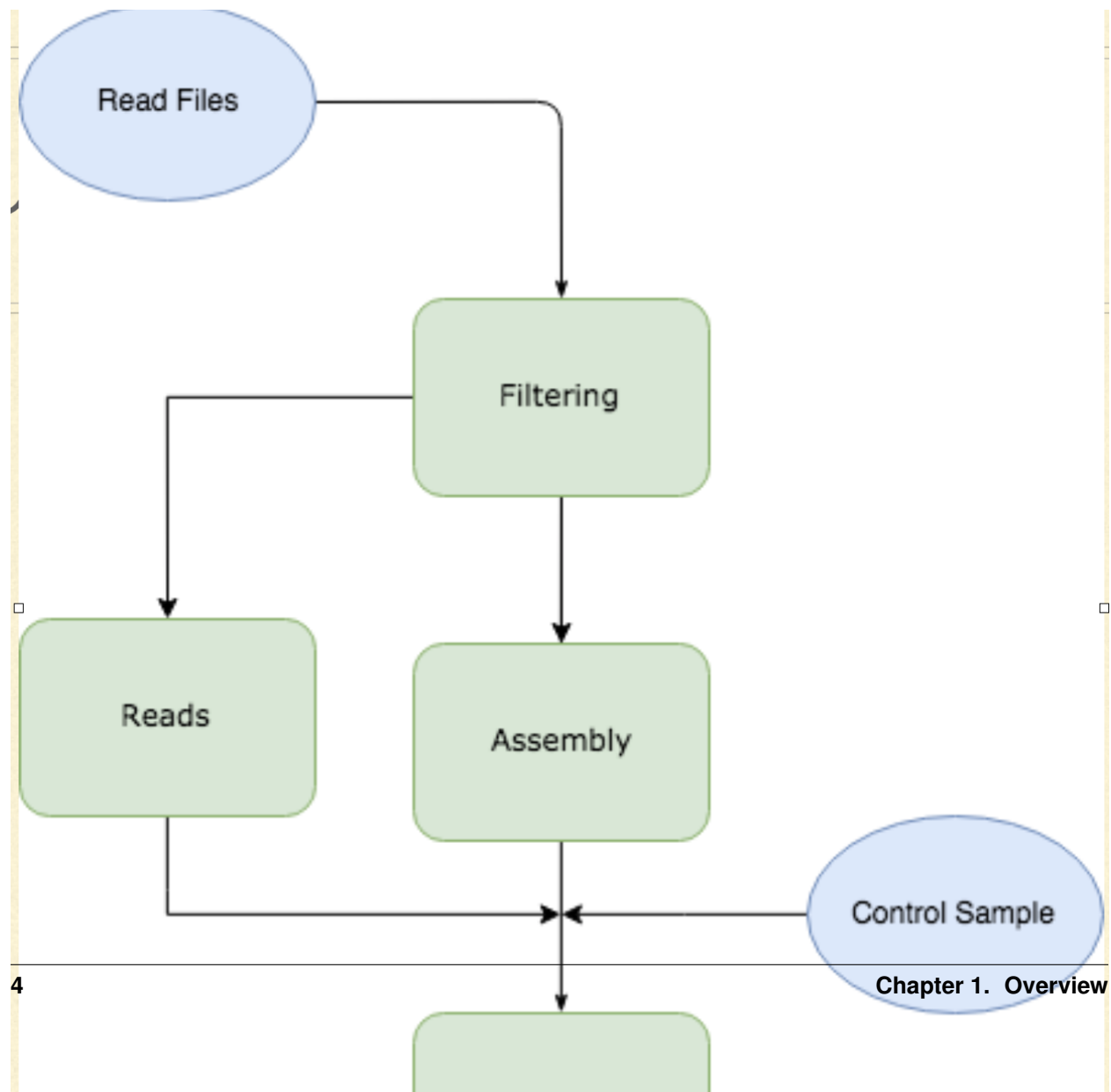


# CHAPTER 1

---

## Overview

---





## CHAPTER 2

---

### Installation

---

#### With Docker:

```
$ docker build -t pathos .  
$ docker run <...>
```

#### Without:

```
curl -sSL https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh -o /  
→tmp/miniconda.sh && \  
conda install -y python=3.6 && \  
git clone https://github.com/averagehat/pathos && \  
cd install && bash assume-conda-install.sh /usr/local/bin/
```



## CHAPTER 3

---

### Configuration

---

The configuration file is specified with the *-c* option when running *pathos\_sheet* or *pathos\_single*. There is an example file at the bottom of the page. Particular options:

*assembler*: should be *ray2* or *abyss*

the assembly *options* are the commandline options used with the chosen assembler.

*param\_file* is the pathdiscov param.txt file used for the ray2/CAP assembly.

The *bowtieDB* and *starDB* filepaths refer to host index databases (they must have already been indexed by *bowtie-build* and *STAR* respectively).

The *minCompressionScore* of *lzwfilter* is the minimum “read complexity” that the filtering algorithm uses to determine to keep the read for further processing. The compression score of a read is calculated by dividing the length of the compressed read by the length of the original read.

The *max\_target\_seqs* flag of *blastn* tells BLAST to stop searching after finding that many sequence matches. A high number results in long BLAST runtimes.

```
pricefilter:
  highQualPercent: 85
  calledPercent: 90
  highQualMin: 0.98
bowtie2:
  bowtieDB: /Users/happyuser/pathos/databses/test/human-bowtie/bowtie-index
star:
  starDB: /Users/happyuser/pathos/databses/test/star-index-dir/
  skip: true
threads: 16
assembly:
  assembler: ray2
  options: " "
  kmer: 25
  minimum_contig_length: 200
cdhitdup:
  minDifference: 5
```

(continues on next page)

(continued from previous page)

```
ncbi:
  ntDB: /Users/happyuser/pathos/databses/test/nt/nt
  nrDB: /Users/happyuser/pathos/databses/test/nr/nr
  ktTaxonomy: /Users/happyuser/pathos/databses/test/krona
lzwfilter:
  minCompressionScore: 0
blastn:
  max_target_seqs: 3
param_file: /Users/happyuser/pathos/param.txt
ete2_db:
```

## CHAPTER 4

---

### Single Sample

---

#### Usage:

```
pipeline.py (--fastq <FASTQS>...) --config <config> [-o <outdir>] [--log <log>] [-  
↪-control <CONTROLS>...]
```

#### Options:

```
-f <FASTQS>, --fastq <FASTQS>  
-r <CONTROLS>, --control <CONTROLS>  
-c <config>, --config <config>  
--o <outdir>, --outdir <outdir>  
--log <log>, -l <log>
```

#### Notes:

For the `--fastq` argument, provide **any** number of paired read files ordered like:

```
--fastq <Samp0_R1> <Samp0_R2> <Samp1_R1> <Samp1_R2> . . .
```

The `--control` argument should be provided **in** the same way.

`--log` **is** the file path to store log output **in**.



## CHAPTER 5

---

### Pathos Sheet

---

```
Usage:
  pathos_sheet.py <samplesheet> --config <config> --sampledir <sampledir> [--log
  ↪<log>] [--outdir <outdir>] [-p] [--qsub]
```

```
Options:
  -c <config>, --config <config>
  --o <outdir>, --outdir <outdir>
  --log <log>
  --sampledir <sampledir>
  --qsub, -q
```

The input samplesheet must be in TSV form with the input sample names in the first column and the control samples on the right column (if there are multiple control samples, each one should be separated by a semi-colon.)

```
191\t132;133
192\t132;133
193;194\t144
```





## CHAPTER 6

---

### Summary Creation

---

Usage:

```
summary.py <indir> <outdir>
```

Run after ``pathos_single`` or ``pathos_sheet`` to get summary information on for that `run`.

`<indir>` is the directory with ``pathos`` output.



## CHAPTER 7

---

### Indices and tables

---

- `genindex`
- `modindex`
- `search`