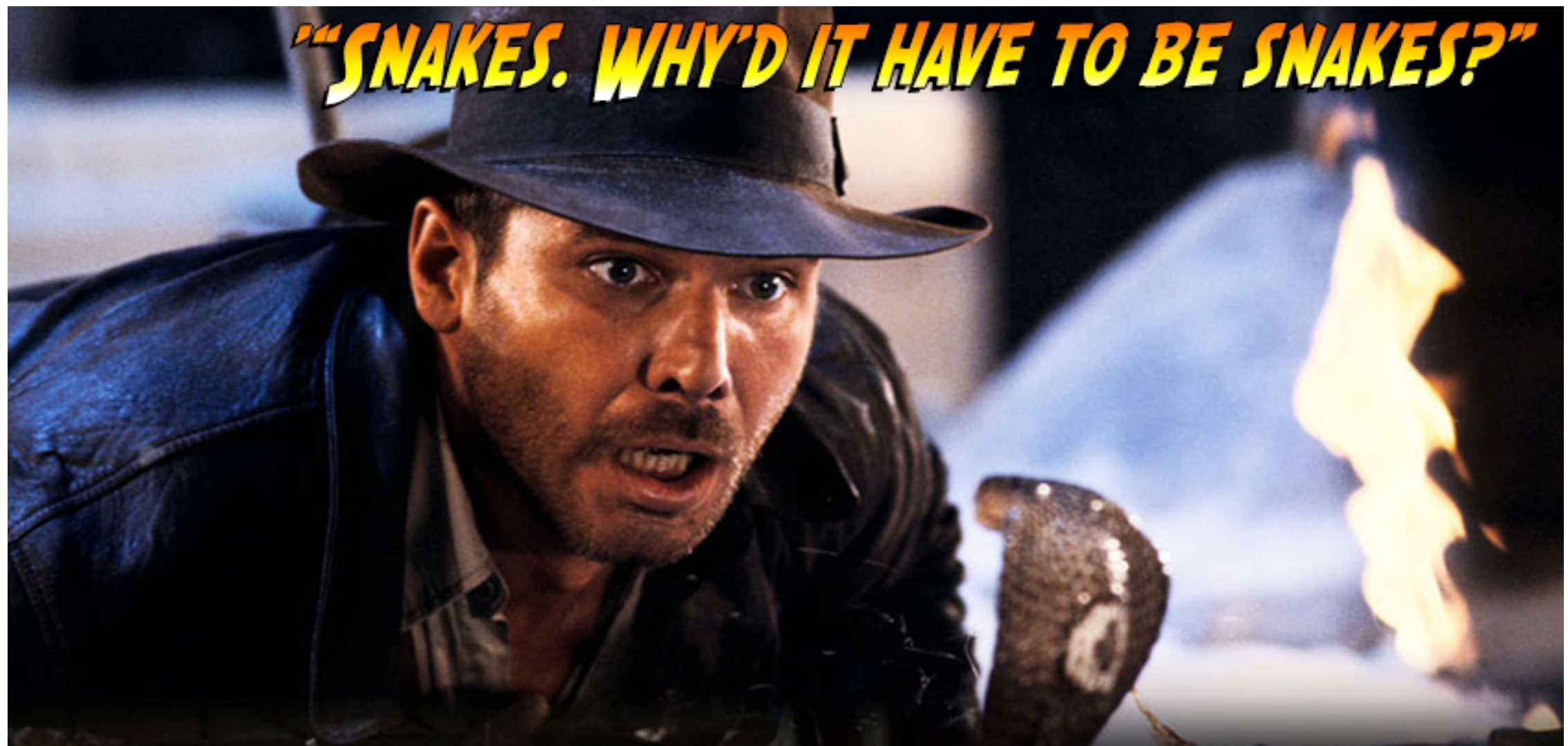


Increasing productivity with Snakemake

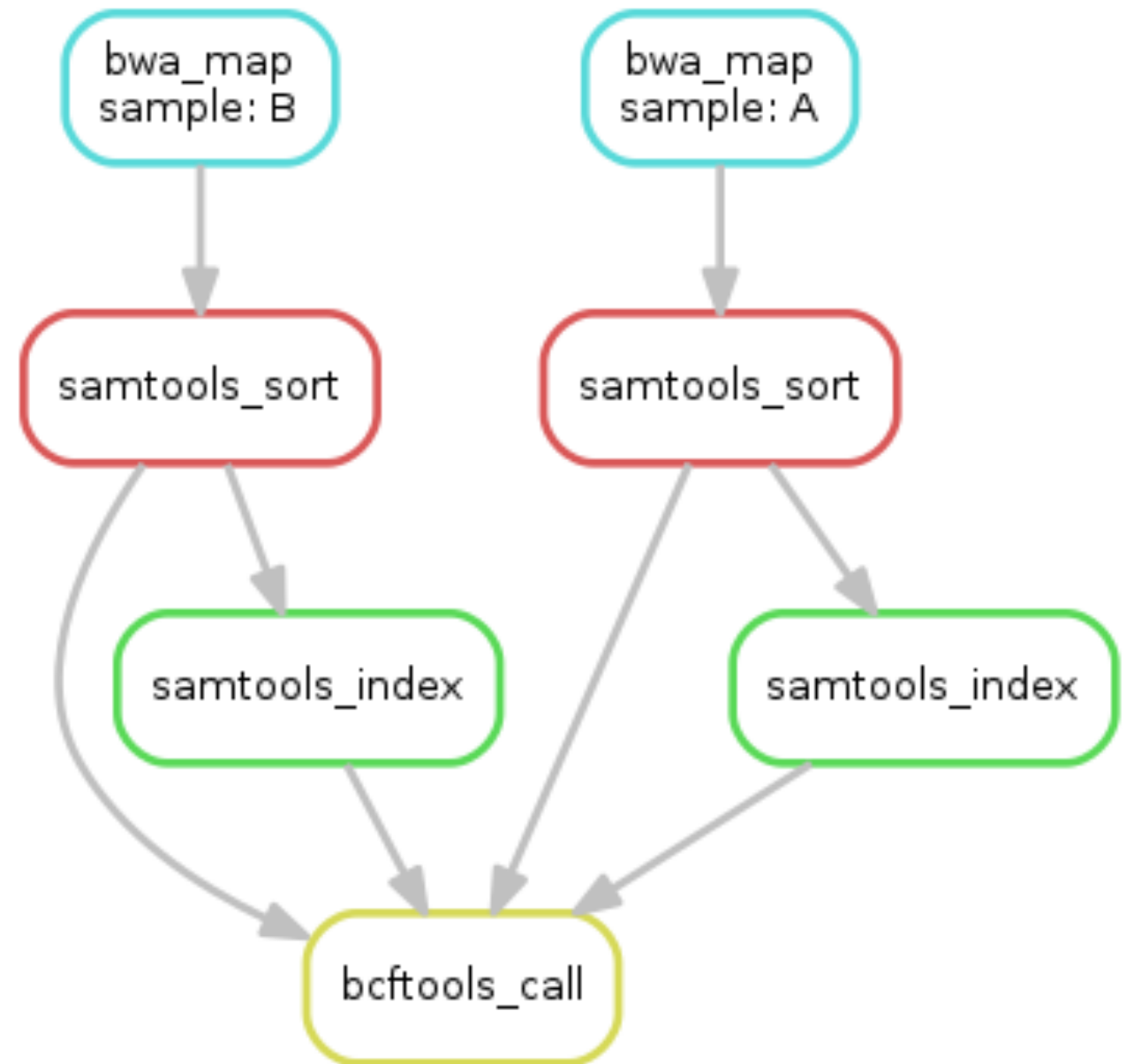


Learning Snakemake will make your life better

- Reproducibility
- Self documenting
- Project management
- Cluster interface
- This afternoon's examples
- Comparative genomics is all about lots

Snakemake?

- “Pythonic” replacement for Make
- You can use python in Snakemake files
- Provides file tracking
- Relatively robust error handling
- Installation (even I managed)
 - Requires Python 3+
 - Comes with Anaconda
 - Pre-installed here at Davis




Becoming a snake charmer


- Do's
 - Write comments (shame on me)
 - Catch STDOUT and STDERR
 - Put your Snakefiles in a Github repository
- Don't
 - Ever-Ever run two snakefiles in the same folder
 - Run snakemake without doing a “dry run” first





Running Snakemake on the command line


snakemake -s Snakefile

snakemake -s Snakefile -n  Dry run

snakemake -s Snakefile -n -p  Print rules

snakemake -s Snakefile -n -p -j 10 jobs  Concurrent

snakemake -s Snakefile -n -p -j 10 -k  keep going

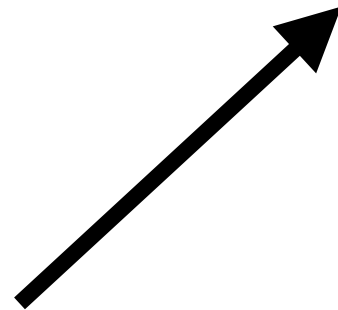
snakemake -s Snakefile -n -p -j 10 -k -w 30  wait



Try running example 1

Write “Snakefiles” backwards

```
rule first:  
    message: “collecting work”  
    input: “hello.world.txt”
```



```
rule second:  
    message: “doing work”  
    output: “hello.world.txt”  
    shell: """  
        echo “test” > {output}  
    """
```



Try running example 2

How to track multiple files?

```
SEQIDS="chr1 chr2 chr3".split()
```

```
rule first:
```

```
    message: "collecting work"
```

```
    input: expand("{seqid}.fa", seqid=SEQIDS)
```

```
rule second:
```

```
    message: "doing work"
```

```
    input: "whole_genome.fasta"
```

```
    output: "{seqid}.fa"
```

```
shell: """
```

```
    samtools faidx {input} {wildcards.seqid} > {output}
```

```
    """
```



Try running example 3

Cluster submission is a snap

cluster.json

submitting the job:

```
{
  "__default__" :
  {
    "account" : "zevk",
    "time" : "00:15:00",
    "c" : 1,
    "n" : 1,
    "partition" : "gc512"
  },
  "snakes" :
  {
    "time" : "00:01:00"
  }
}
```

```
snakemake -p -k -w 25 \
-s Snakefile_4 -j 3 \
--cluster-config cluster.json \
--cluster "sbatch -p {cluster.partition} \
-n {cluster.n} -t {cluster.time} \
-c {cluster.c}"
```



Try running example 3...
on the cluster!

Final example, a full featured Snakefile

- Python modules
- Run block
- External config file

Questions?

Acknowledgements

- Johannes Köster - Author of Snakemake
- Brad Nelson - Snake guru (Eichler lab)
- Matt Settles (for the data)