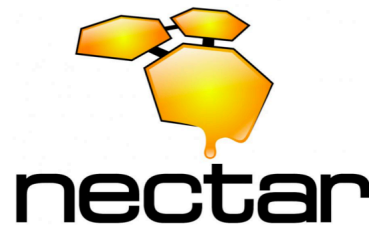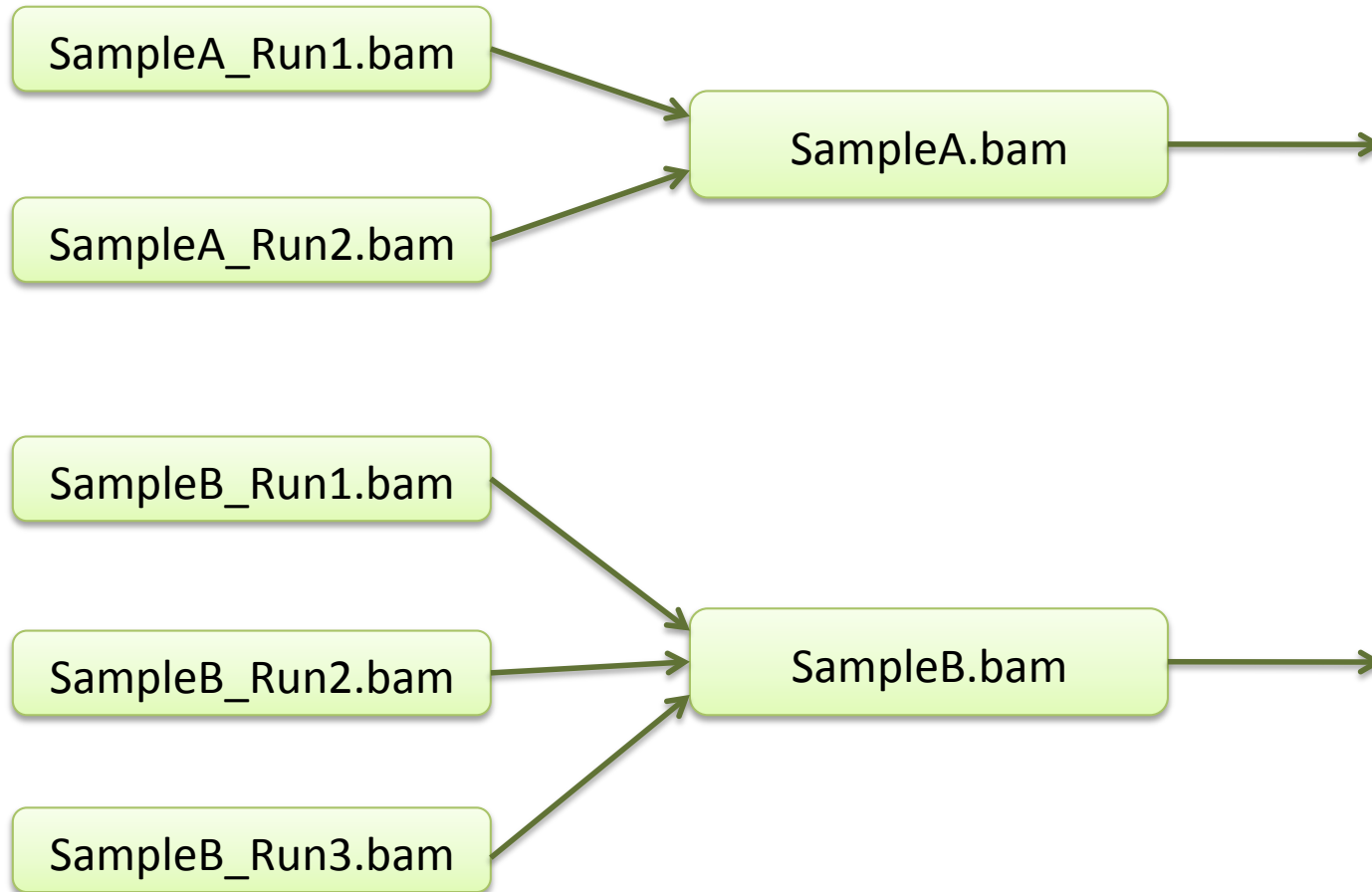# Rubra - flexible distributed pipelines for bioinformatics

Clare Sloggett, Gayle Philip, Matthew Wakefield, Bernard Pope
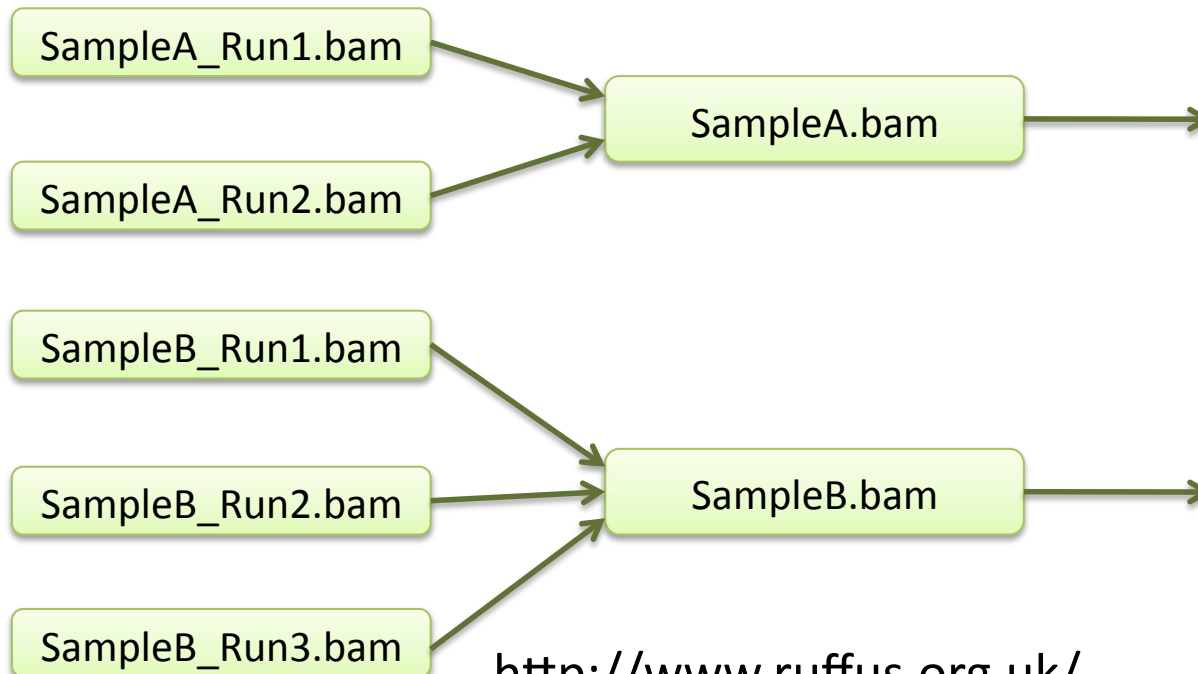
Pipelines

Ruffus

```
@collate(bamfiles, regex(r'(Sample?)_(Run?).bam'), r'\1.bam')
def mergeBamsBySample(inputs, outputs):
    ...
```

SampleA_Run1.bam → SampleA.bam →

SampleA_Run2.bam →

SampleB_Run1.bam → SampleB.bam →

SampleB_Run2.bam →

SampleB_Run3.bam →
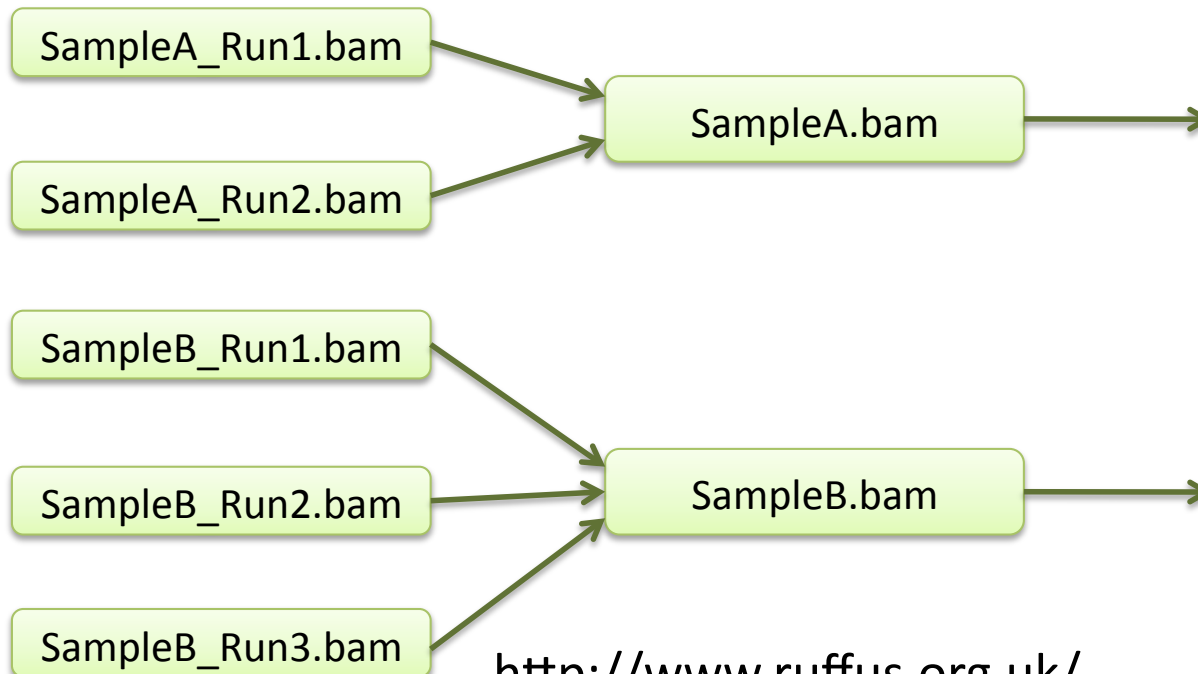
http://www.ruffus.org.uk/

Ruffus: a lightweight Python library for computational pipelines,
Goodstadt L., Bioinformatics 2010 Nov 1;26(21):2778-9

Ruffus

```
@collate(bamfiles, regex(r'(Sample?)_(Run?).bam'), r'\1.bam')
def mergeBamsBySample(inputs, outputs):
    ...
```


- Parallelisation
- Dependency checking
- Logging
- Visualisation

http://www.ruffus.org.uk/

Ruffus: a lightweight Python library for computational pipelines,
Goodstadt L., Bioinformatics 2010 Nov 1;26(21):2778-9

Rubra

```
rubra pipeline.py --config pipeline_config.py
                  --end callSNPs
                  --force dedup
                  --style run
```

Exposes ruffus functionality, plus:
• Config files
• Checkpointing
• HPC job submission

https://github.com/bjpop/rubra

Reusable pipelines: variant calling

- Alignment (BWA)
- Alignment cleaning and variant calling (GATK)
- Annotation (ENSEMBL)
- QC

https://github.com/claresloggett/
variant_calling_pipeline