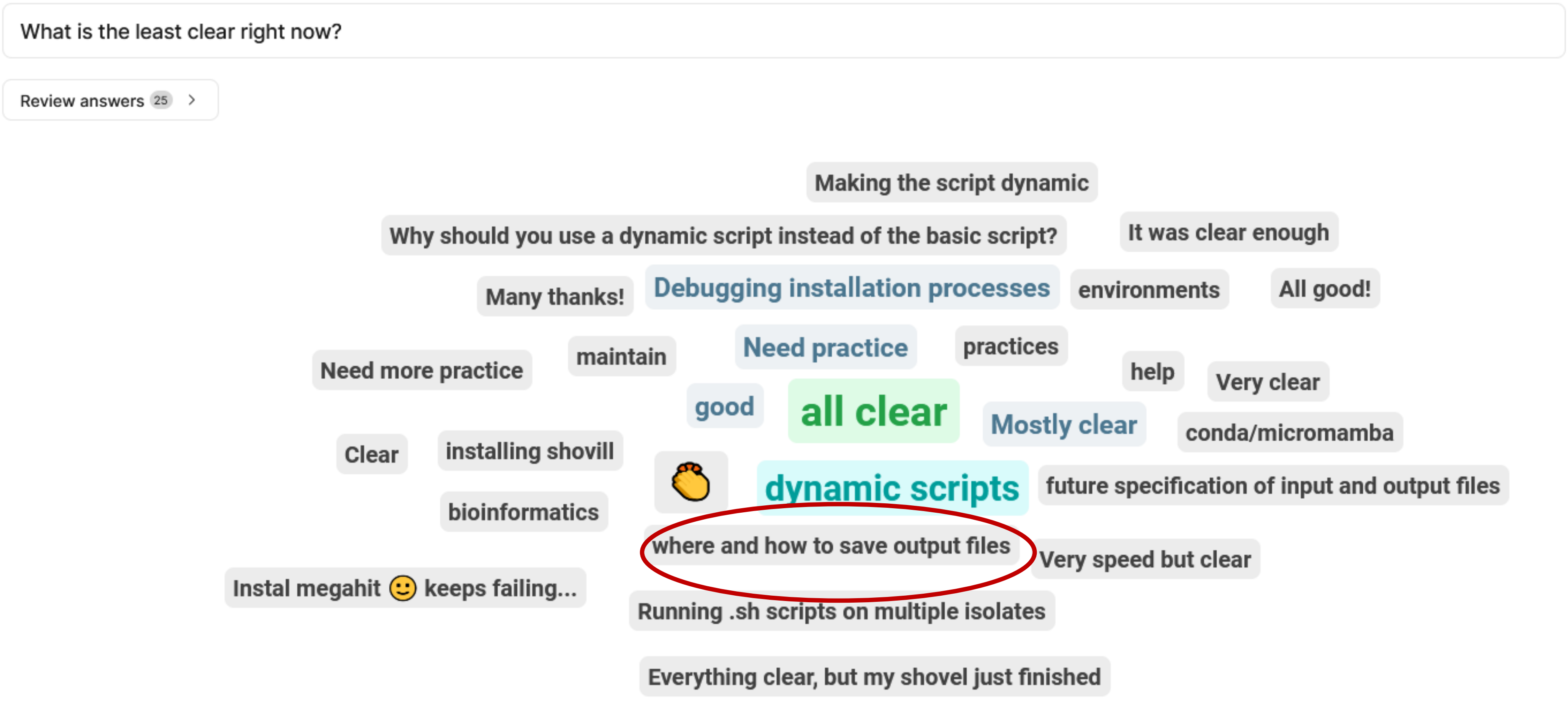
Building Basic Workflows
Day 2

Last time on Basic Workflows…

What did we do?

- Created virtual environments from files with instructions (.yaml)
- Tried out a few bioinformatic commands
- Creating script files to handle commands
- Repurposing the scripts to run with different samples

What is the least clear right now?



What is the least clear right now?

Making the script dynamic

Why should you use a dynamic script instead of the basic script?

It was clear enough

Many thanks!

Debugging installation processes

environments

All good!

Need more practice

maintain

Need practice

practices

help

Very clear

good

all clear

Mostly clear

conda/micromamba

Clear

installing shovill

👏

dynamic scripts

future specification of input and output files

bioinformatics

where and how to save output files

Very speed but clear

Instal megahit 🙂 keeps failing...

Running .sh scripts on multiple isolates

Everything clear, but my shovel just finished

```
bash FastQC.sh ~/MyExperiment/Dataset/ERR14229040_1.fastq.gz ~/MyExperiment/Dataset/ERR14229040_2.fastq.gz ~/MyExperiment/Results
```

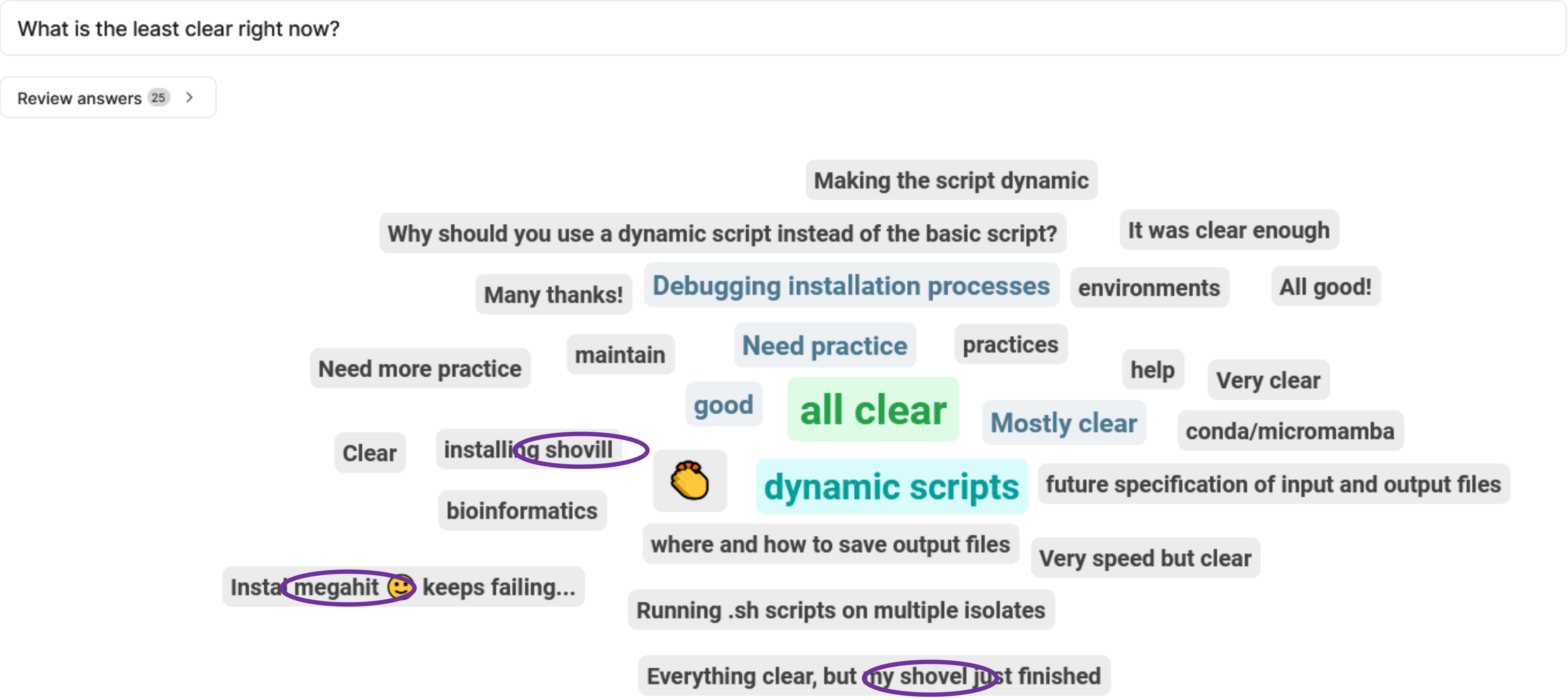
```
#!/bin/bash

# Input
read1=$1
read2=$2

# Output
output_dir=$3

# Variables
outdir=$output_dir/FastQC

mkdir -p $outdir
micromamba run -n QC fastqc $read1 $read2 -o $outdir
```



What is the least clear right now?

Making the script dynamic

Why should you use a dynamic script instead of the basic script?

It was clear enough

Many thanks!

Debugging installation processes

environments

All good!

Need more practice

maintain

Need practice

practices

help

Very clear

good

all clear

Mostly clear

conda/micromamba

Clear

installing shovill

👏

dynamic scripts

future specification of input and output files

bioinformatics

where and how to save output files

Very speed but clear

Instal megahit 🙃 keeps failing...

Running .sh scripts on multiple isolates

Everything clear, but my shovel just finished

Shovill bugs up? Use SPAdes

- Test your SPAdes installation by running

```
bin/spades.py --test
```

- A single paired-end library (separate files, gzipped):

```
bin/spades.py -1 left.fastq.gz -2 right.fastq.gz -o output_folder
```

- IonTorrent data:

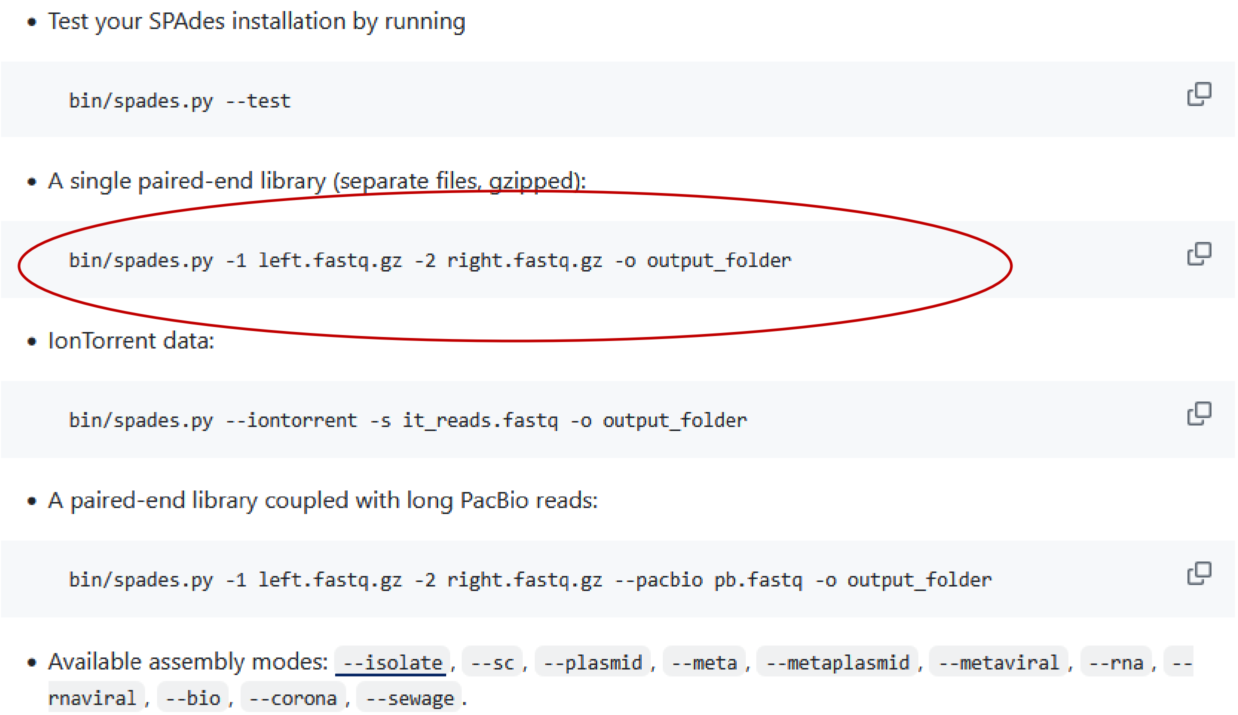
```
bin/spades.py --iontorrent -s it_reads.fastq -o output_folder
```

- A paired-end library coupled with long PacBio reads:

```
bin/spades.py -1 left.fastq.gz -2 right.fastq.gz --pacbio pb.fastq -o output_folder
```

- Available assembly modes: `--isolate`, `--sc`, `--plasmid`, `--meta`, `--metaplasmid`, `--metaviral`, `--rna`, `--rnaviral`, `--bio`, `--corona`, `--sewage`.

Shovill bugs up? Use SPAdes



- Test your SPAdes installation by running

```
bin/spades.py --test
```

- A single paired-end library (separate files, gzipped):

```
bin/spades.py -1 left.fastq.gz -2 right.fastq.gz -o output_folder
```

- IonTorrent data:

```
bin/spades.py --iontorrent -s it_reads.fastq -o output_folder
```

- A paired-end library coupled with long PacBio reads:

```
bin/spades.py -1 left.fastq.gz -2 right.fastq.gz --pacbio pb.fastq -o output_folder
```

- Available assembly modes: `--isolate`, `--sc`, `--plasmid`, `--meta`, `--metaplasmid`, `--metaviral`, `--rna`, `--rnaviral`, `--bio`, `--corona`, `--sewage`.

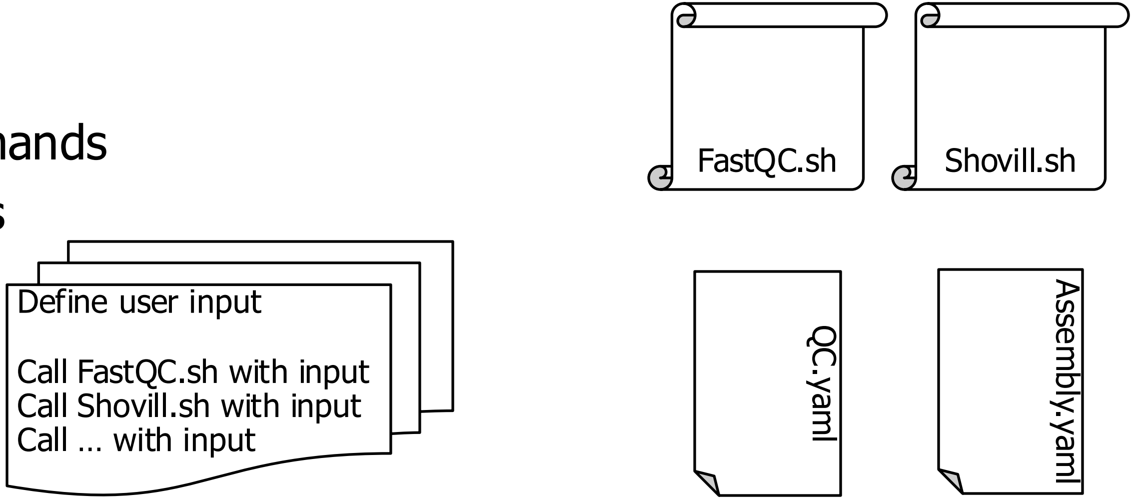
This week on Basic Workflows:

Where are we now?

- 2 scripts files which upon execution will rerun bioinformatic commands
- 2 installation instructions for installing components required by the scripts

We Need

- A way to execution all commands
- Ability to integrate new tools



FastQC.sh

Shovill.sh

Define user input

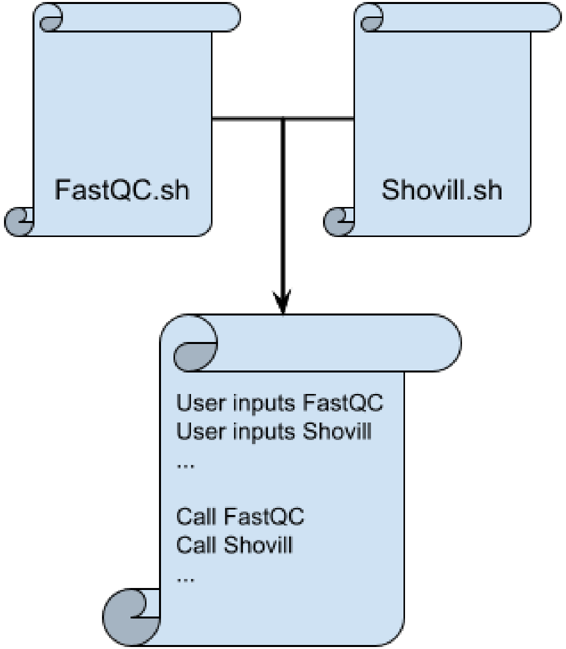
Call FastQC.sh with input
Call Shovill.sh with input
Call … with input

QC.yaml

Assembly.yaml

Centralization vs Modularization

Centralization vs Modularization

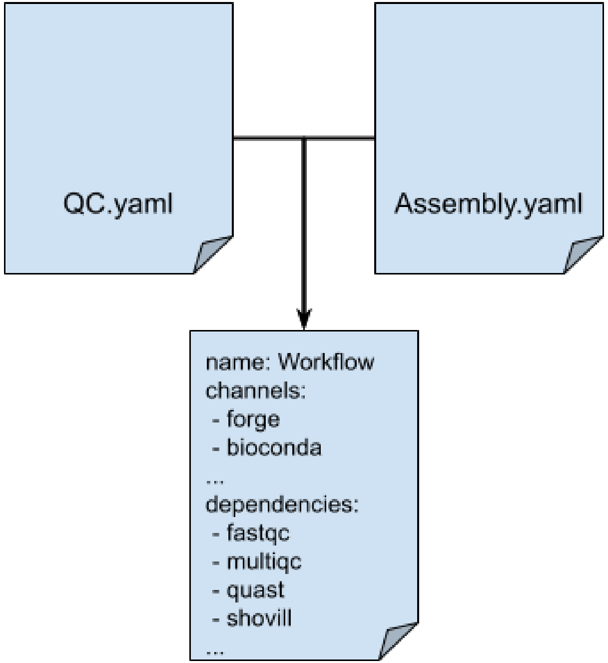


FastQC.sh

Shovill.sh

User inputs FastQC
User inputs Shovill
...

Call FastQC
Call Shovill
...

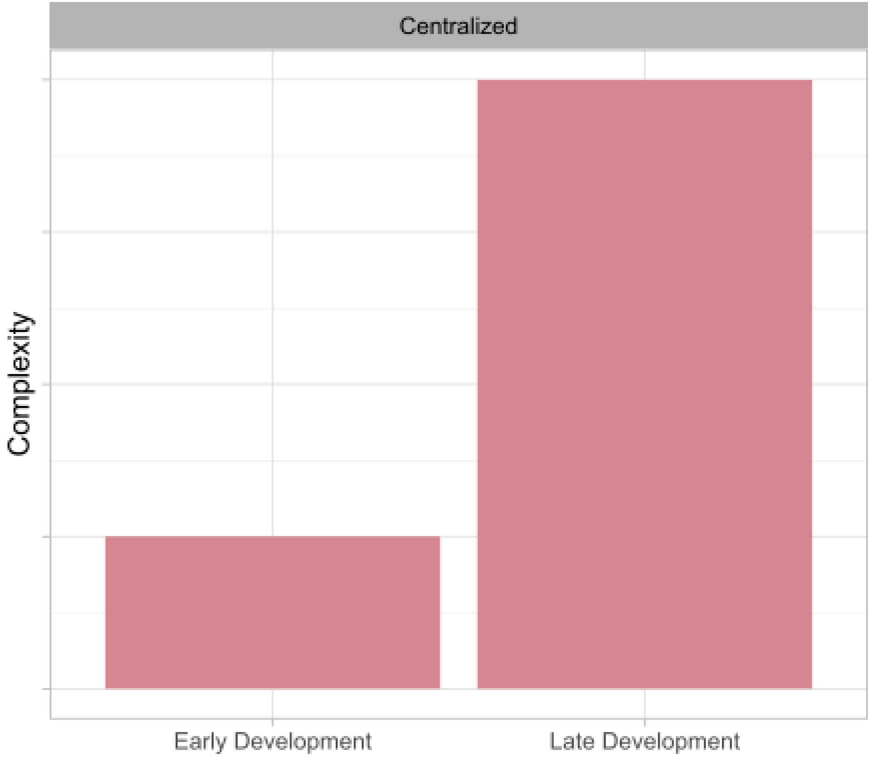


QC.yaml

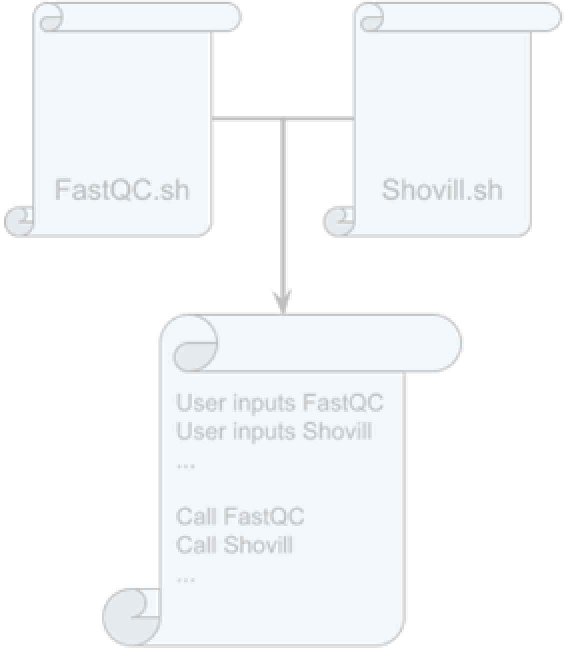
Assembly.yaml

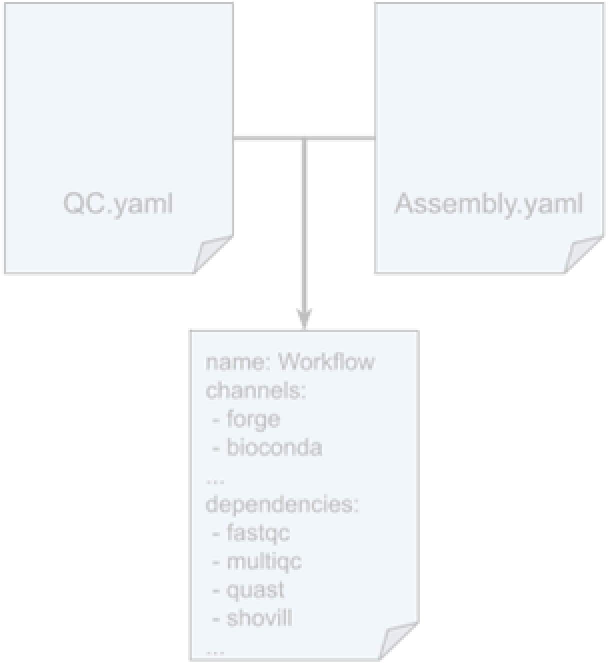
```
name: Workflow
channels:
 - forge
 - bioconda
...
dependencies:
 - fastqc
 - multiqc
 - quast
 - shovill
...
```

Centralization vs Modularization

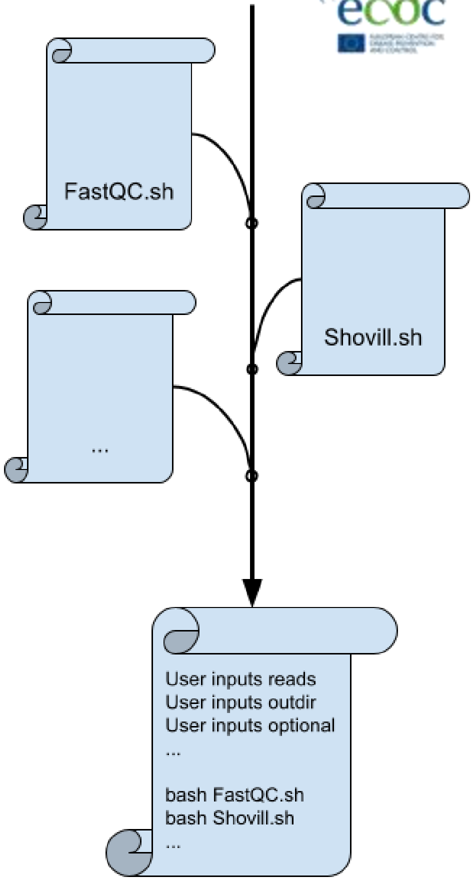


Centralization vs Modularization

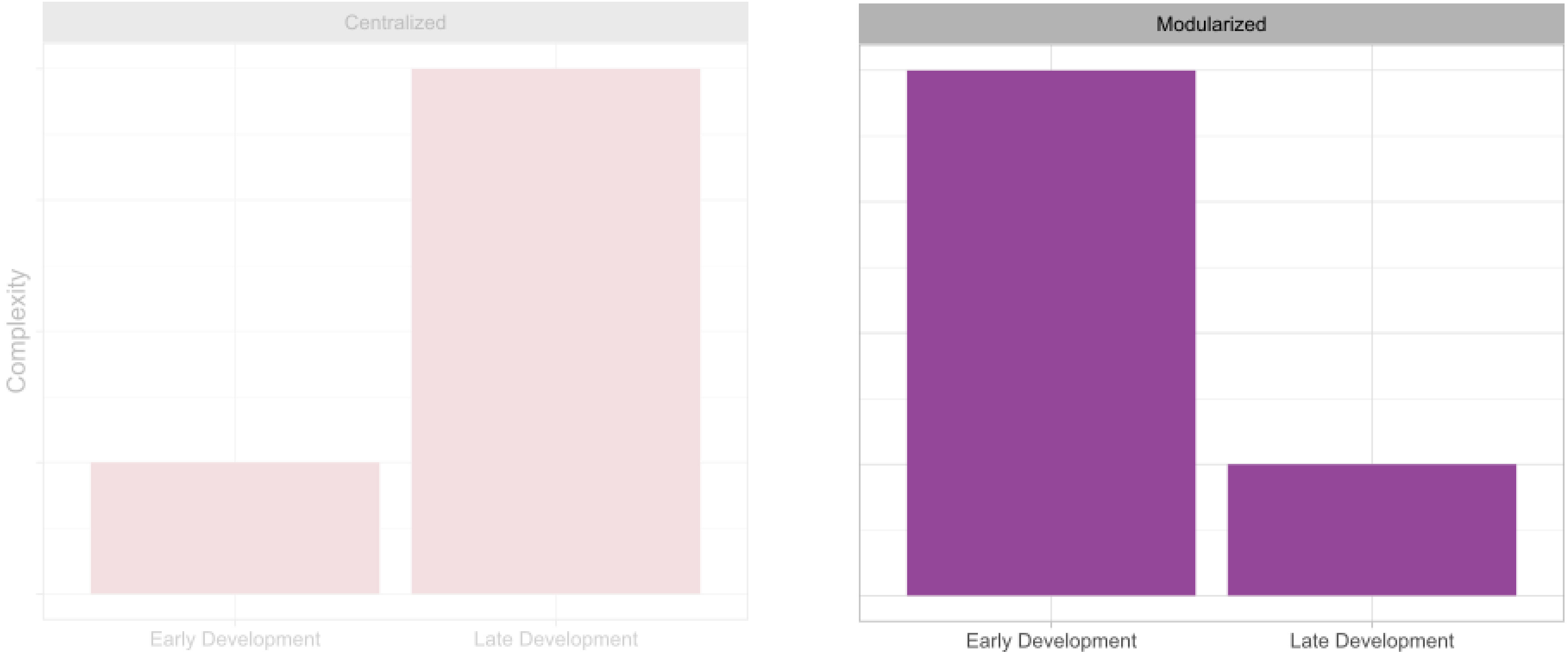




```
name: Workflow
channels:
 - forge
 - bioconda
...
dependencies:
 - fastqc
 - multiqc
 - quast
 - shovill
...
```



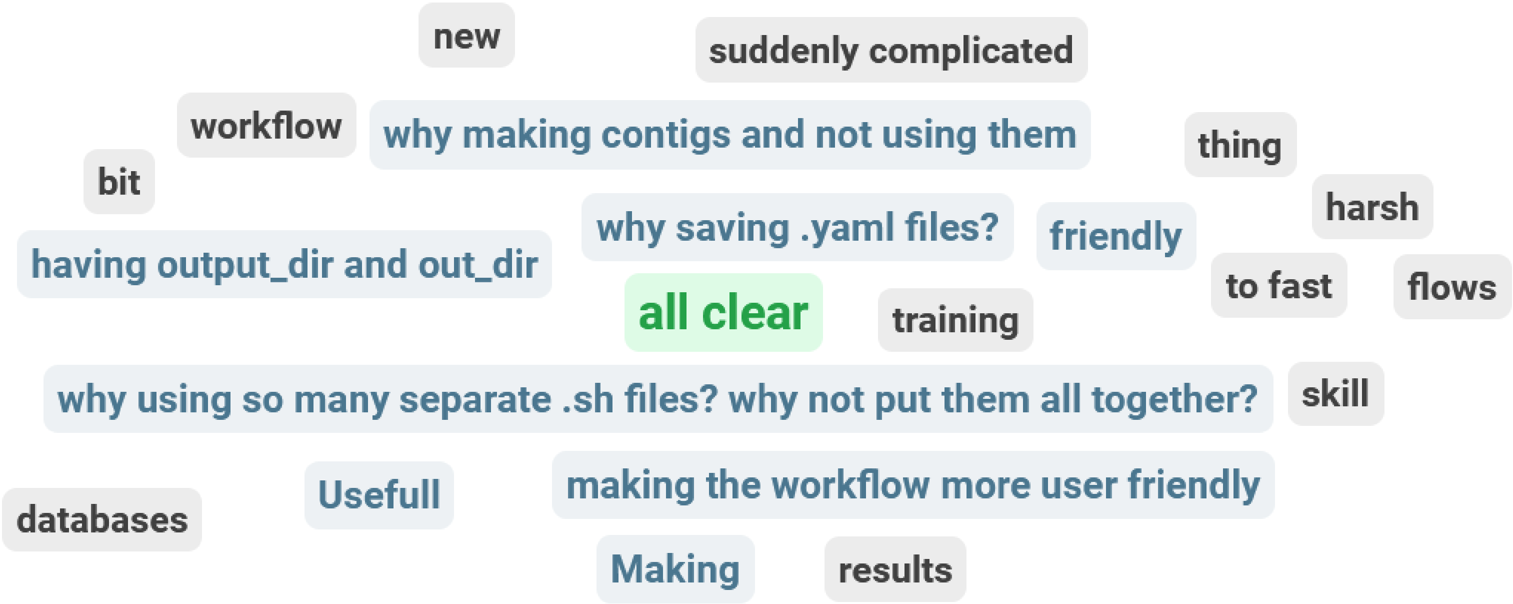
Centralization vs Modularization



Time to get building

What is the least clear right now?

Review answers 27 >



new

suddenly complicated

workflow

why making contigs and not using them

thing

bit

harsh

why saving .yaml files?

friendly

having output_dir and out_dir

to fast

flows

all clear

training

why using so many separate .sh files? why not put them all together?

skill

Usefull

making the workflow more user friendly

databases

Making

results

Don't forget to feedback us!