

Nextflow workflow report

[trusting_tuckerman]

Workflow execution completed unsuccessfully!

The exit status of the task that caused the workflow execution to fail was: **1**.

The full error message was:

```
Error executing process > 'NFCORE_MAG:MAG:BUSCO_QC:BUSCO  
(MEGAHIT-be30by.3.fa)'
```

Caused by:

```
Process `NFCORE_MAG:MAG:BUSCO_QC:BUSCO (MEGAHIT-  
be30by.3.fa)` terminated with an error exit status (1)
```

Command executed:

```
# ensure augustus has write access to config directory  
if [ Y = "Y" ] ; then  
    cp -r /usr/local/config/ augustus_config/  
    export AUGUSTUS_CONFIG_PATH=augustus_config  
fi  
  
# place db in extra folder to ensure BUSCO recognizes it  
as path (instead of downloading it)  
if [ N = "Y" ] ; then  
    mkdir dataset  
    mv dataset/  
fi  
  
# set nullglob: if pattern matches no files, expand to a  
null string rather than to itself  
shopt -s nullglob
```

```

# only used for saving busco downloads
most_spec_db="NA"

if busco --auto-lineage          --mode genome          --in
MEGAHIT-be30by.3.fa            --cpu "8"              --out "BUSCO"
> MEGAHIT-be30by.3.fa_busco.log 2> MEGAHIT-
be30by.3.fa_busco.err; then

    # get name of used specific lineage dataset
    summaries=(BUSCO/short_summary.specific.*.BUSCO.txt)
    if [ ${#summaries[@]} -ne 1 ]; then
        echo "ERROR: none or multiple
'BUSCO/short_summary.specific.*.BUSCO.txt' files found.
Expected one."
        exit 1
    fi
    [[ $summaries =~ BUSCO/short_summary.specific.
(.*)\.BUSCO.txt ]];
    db_name_spec="${BASH_REMATCH[1]}"
    most_spec_db=${db_name_spec}
    echo "Used specific lineage dataset: ${db_name_spec}"

    if [ N = "Y" ]; then
        cp
BUSCO/short_summary.specific.${db_name_spec}.BUSCO.txt
short_summary.specific_lineage.${db_name_spec}.MEGAHIT-
be30by.3.fa.txt

        # if lineage dataset is provided, BUSCO analysis
does not fail in case no genes can be found as when using
the auto selection setting
        # report bin as failed to allow consistent
warnings within the pipeline for both settings
        if egrep -q '$WARNING:          BUSCO did n          o
t find any match.' MEGAHIT-be30by.3.fa_busco.log
; then          echo "WARNING: BUSCO could not find any
genes for the provided lineage dataset! See also M
EGAHIT-be30by.3.fa_busco
.log."          echo -e "MEGAHIT-be30by.3.fa
No genes" > "MEGAHIT-be30by.3.fa_busco.fail
ed_bin.txt"
        fi
        else          # auto lineag
e selection          if { egrep -q '$INFO: \S+ select
ed' MEGAHIT-be30by.3.fa_busco.log          && egrep
-q '$INFO: Lineage \S+ is selected, supported b
y ' MEGAHIT-be30by.3.fa_busco.log ; } ||
{ egrep -q '$INFO: \S+ selected' MEGAHIT-be30by.3.fa_busco.l

```

```

        && egrep -q '$INFO: The results from the P r
odigal gene predictor indicate that your data belongs to the
mollicutes clade. Testing subclades...' MEGAHIT-be30by.3.fa_
        && egrep -q '$INFO: Using local lineag e
s directory ' MEGAHIT-be30by.3.fa_busco.
log ; }; then                # the second statement is neces
sary, because certain mollicute clades use a different gene
tic code, are not part of the BUSCO placement tree, are test
ed separately                # and cause different
log messages                echo "Domain and specific lin
eage could be select
ed by BUSCO."
        cp BUSCO/short_summary.specific.${db_name_sp
ec}.BUSCO.txt short_summary.specific_lineage.${db_name_
spec}.MEGAHIT-be
30b
y.3.fa.txt                d
b_name_gen=""
summaries_gen=(BUSCO/short_summary.generic
.*.BUSCO.txt)            if [ ${#summaries_gen[@]}
-lt 1 ]; then
        echo "No 'BUSCO/short_summary.generic.*.BUSCO.txt
' file found. Assuming selected domain and specific lin
eages a
re the same."
        cp BUSCO/short_summary.specific.${db_name_sp
ec}.BUSCO.txt short_summary.domain.${db_name_spec}.MEGAHIT-be
30by.3.fa.txt                db_name_gen=${
db_name_spec}
        else                [[ $sum
maries_gen =~ BUSCO/short_summary.generic.(.*).
BUSCO.txt ]];                db_name_gen="${BASH
_REMATCH[1]}"                echo "Used generic lin
eage dataset: ${
db_name_gen}"
        cp BUSCO/short_summary.generic.${db_name_g
en}.BUSCO.txt short_summary.domain.${db_name_gen}.MEGAHIT-be
30by.3.fa.txt

        fi
        for f in BUSCO/run_${db_name_gen}/busco_sequences/single
eque
nces/*faa; do
                cat BUSCO/run_${db_name_gen}/busco_sequences/single
sequences/*faa | gzip >MEGAHIT-be30by.3.fa_buscoss.${db_nam
e_gen}.faa.gz
                break
        done
        for f in

```

```

BUSCO/run_${db_name_gen}/busco_sequences/single_copy_busco_se
do
    cat
BUSCO/run_${db_name_gen}/busco_sequences/single_copy_busco_se
| gzip >MEGAHIT-be30by.3.fa_buscoss.${db_name_gen}.fna.gz
    break
done

    elif egrep -q '$INFO: \S+ selected' MEGAHIT-
be30by.3.fa_busco.log && egrep -q '$INFO:      Not enough
markers were placed on the tree \([0-9]*\)'. Root lineage \S+
is kept' MEGAHIT-be30by.3.fa_busco.log ; then
    echo "Domain could be selected by BUSCO, but
no more specific lineage."
    cp
BUSCO/short_summary.specific.${db_name_spec}.BUSCO.txt
short_summary.domain.${db_name_spec}.MEGAHIT-be30by.3.fa.txt

    elif egrep -q '$INFO: \S+ selected' MEGAHIT-
be30by.3.fa_busco.log && egrep -q '$INFO:      Running
virus detection pipeline' MEGAHIT-be30by.3.fa_busco.log ;
then
    # TODO double-check if selected dataset is not
one of bacteria_*, archaea_*, eukaryota_*?
    echo "Domain could not be selected by BUSCO,
but virus dataset was selected."
    cp
BUSCO/short_summary.specific.${db_name_spec}.BUSCO.txt
short_summary.specific_lineage.${db_name_spec}.MEGAHIT-
be30by.3.fa.txt
    else
        echo "ERROR: Some not expected case occurred!
See MEGAHIT-be30by.3.fa_busco.log." >&2
        exit 1
    fi
fi

    for f in
BUSCO/run_${db_name_spec}/busco_sequences/single_copy_busco_s
do
    cat
BUSCO/run_${db_name_spec}/busco_sequences/single_copy_busco_s
| gzip >MEGAHIT-be30by.3.fa_buscoss.${db_name_spec}.faa.gz
    break
done
    for f in
BUSCO/run_${db_name_spec}/busco_sequences/single_copy_busco_s
do
    cat

```

```

BUSCO/run_${db_name_spec}/busco_sequences/single_copy_busco_s
| gzip >MEGAHIT-be30by.3.fa_buscoss.${db_name_spec}.fna.gz
    break
done

elif egrep -q '$ERROR:          No genes were recognized by
BUSCO' MEGAHIT-be30by.3.fa_busco.err ; then
    echo "WARNING: BUSCO analysis failed due to no
recognized genes! See also MEGAHIT-be30by.3.fa_busco.err."
    echo -e "MEGAHIT-be30by.3.fa          No genes" >
"MEGAHIT-be30by.3.fa_busco.failed_bin.txt"

elif egrep -q '$INFO: \S+ selected' MEGAHIT-
be30by.3.fa_busco.log && egrep -q '$ERROR:          Placements
failed' MEGAHIT-be30by.3.fa_busco.err ; then
    echo "WARNING: BUSCO analysis failed due to failed
placements! See also MEGAHIT-be30by.3.fa_busco.err. Still
using results for selected generic lineage dataset."
    echo -e "MEGAHIT-be30by.3.fa          Placements failed" >
"MEGAHIT-be30by.3.fa_busco.failed_bin.txt"

    message=$(egrep '$INFO: \S+ selected' MEGAHIT-
be30by.3.fa_busco.log)
    [[ $message =~ INFO:[[:space:]]([[:alnum:]]+)]
    [[[:space:]]selected ]];
    db_name_gen="${BASH_REMATCH[1]}"
    most_spec_db=${db_name_gen}
    echo "Used generic lineage dataset: ${db_name_gen}"
    cp
BUSCO/auto_lineage/run_${db_name_gen}/short_summary.txt
short_summary.domain.${db_name_gen}.MEGAHIT-be30by.3.fa.txt

    for f in
BUSCO/auto_lineage/run_${db_name_gen}/busco_sequences/single_
do
        cat
BUSCO/auto_lineage/run_${db_name_gen}/busco_sequences/single_
| gzip >MEGAHIT-be30by.3.fa_buscoss.${db_name_gen}.faa.gz
        break
    done
    for f in
BUSCO/auto_lineage/run_${db_name_gen}/busco_sequences/single_
do
        cat
BUSCO/auto_lineage/run_${db_name_gen}/busco_sequences/single_
| gzip >MEGAHIT-be30by.3.fa_buscoss.${db_name_gen}.fna.gz
        break
    done

```

```
else
    echo "ERROR: BUSCO analysis failed for some unknown
reason! See also MEGAHIT-be30by.3.fa_busco.err." >&2
    exit 1
fi

# additionally output genes predicted with Prodigal (GFF3)
if [ -f BUSCO/logs/prodigal_out.log ]; then
    mv BUSCO/logs/prodigal_out.log "MEGAHIT-
be30by.3.fa_prodigal.gff"
fi

busco --version | sed "s/BUSCO //" > busco.version.txt
```

Command exit status:

1

Command output:

(empty)

Command error:

ERROR: BUSCO analysis failed for some unknown reason! See also MEGAHIT-be30by.3.fa_busco.err.

Work dir:

/mnt/philippe/teste_pipeline/work/db/699ab81b56c44495e518e592

Tip: when you have fixed the problem you can continue the execution adding the option ``-resume`` to the run command line

Run times

03-Jun-2022 12:35:56 - 04-Jun-2022 08:38:17 (duration: **20h 2m 21s**)

111 succeeded

Nextflow command

```
nextflow run nf-core/mag -profile docker -c process.config --input
'*_R{1,2}.fastq.gz'
```

CPU-Hours

317.4 (4.4% failed)

Launch directory

/mnt/philippe/teste_pipeline

Work directory

/mnt/philippe/teste_pipeline/work

Project directory

/home/philippe/.nextflow/assets/nf-core/mag

Script name

main.nf

Script ID

d8ab3b479c5710afc59c1c28658b47a5

Workflow session

349bbe80-1eb1-4404-872f-750bcc116c5

Workflow

repository

<https://github.com/nf-core/mag>, revision `master` (commit hash `e065754c46eedc85d46cd8a71b32ad73b3b741cb`)

Workflow profile

docker

Nextflow version

version 22.04.0, build 5697 (23-04-2022 18:00 UTC)

Resource Usage

These plots give an overview of the distribution of resource usage for each process.

CPU

Raw Usage

% Allocated

Memory

Physical (RAM)

Virtual (RAM + Disk swap)

% RAM Allocated

Job Duration

Raw Usage

% Allocated

I/O

Read

Write

Tasks

This table shows information about each task in the workflow. Use the search box on the right to filter rows for specific values. Clicking headers will sort the table by that value and scrolling side to side will reveal more columns.

Values shown as: Human readable

(tasks table omitted because the dataset is too big)