Paleofish Mitochondries

Objectif : Lancement de Eager sur référence majoritaire

16/10/2025

CR réunion du 9 oct 2025

- Quel Thymallus est le plus proche phylogénétiquement de Thymallus ligericus dont le whole génome est disponible ? Cf. Joelle. Trancher entre Thymallus thymallus et Thymallus grubli.
- Pour l’Anguilla : NC_006535.3 Anguilla bicolor pacifica avec le whole génome GCA_039654925.2
- Pour Coregonus : le whole génome lavaretus GCA_964263955.1 contient la mito. Pour Coregonus oxyrinchus, Joelle regarde quel est le whole génome le plus proche phylogénétiquement.
- Pour Brachymystax, le second multimapping portait sur les génomes mito de lenox et tumensis. Seul le lenox est ressorti or nous avons uniquement le whole génome de Brachymystax lenox tsinlingensis donc relancer le second multimapping avec les génomes mito de tumensis NC_024674.1 et lenox tsinlingensis NC_018342.1. Comme cela, nous pourrons ensuite prender le whole génome Brachymystax lenok tsinlingensis qui est le seul disponible.
- Vérifier si les mito sont dans les scaffolds du whole génome du Brachymystax lenok tsinlingensis ? Les retirer sinon afin de ne pas les avoir en double et faire un cat du whole génome avec le génome mito. Car dans les autres whole génomes, habituellement, les mito sont identifiées comme chromosome dans le whole génome, pas pour le Brachymystax lenok tsinlingensis.
- Check question sur le NC manquant.
- Lancer Eager dès que Joelle aura fournit les espèces les plus proches pour Thymallus et Coregonus.
- Pour info, le rapport multiQC est disponible via Eager : $ ls multiqc/

  multiqc_data multiqc_report.html

Rappel résultats des multimapping

4 Undeterminated
5 Autres que S. salar et S.trutta
32 S. salar
47 S. trutta

---

88 échantillons

L11_7_SCO_1354_S70:NC_006531.1:182488 → Anguille
NC_006535.3 164611 Anguilla bicolor pacifica

L0_4_EG9_S84:NC_012928.1:19566 → Thymallus thymallus
NC_056303.1 12714 Thymallus ligericus

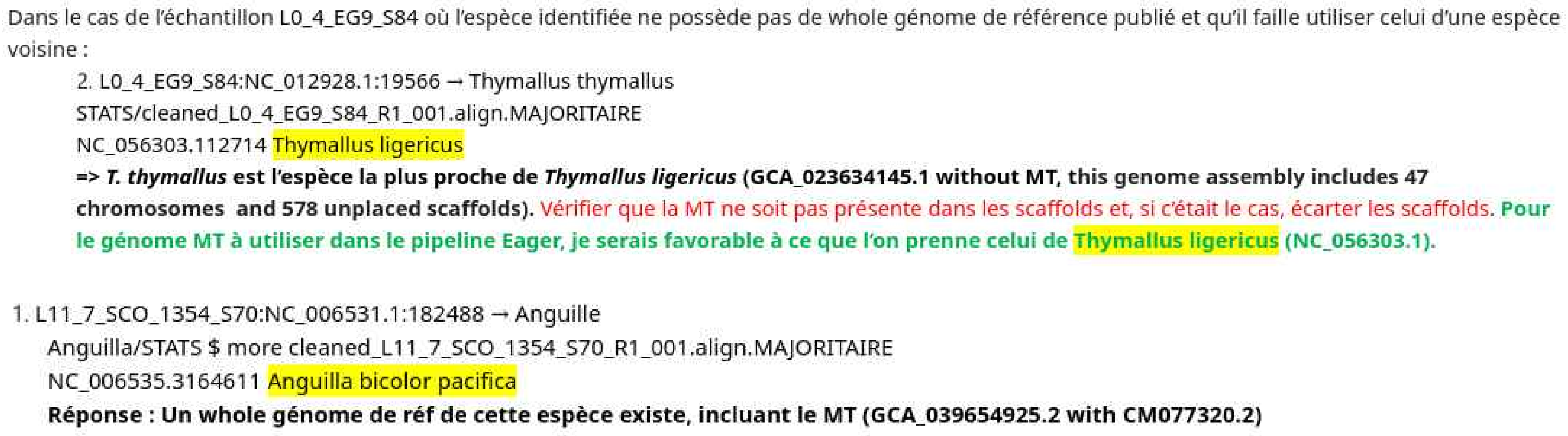
L2_4_CD2_S74:NC_018341.1:5799 → Brachymystax
NC_018341.1 727 Brachymystax lenok

L0_5_EG10_S85:NC_020762.1:38793 → Coregonus
L7_14_22d_S48:NC_020762.1:1255682
L0_5_EG10_S85 : NC_002646.1 6117 Coregonus lavaretus
L7_14_22d_S48 : NC_020765.1 186123 Coregonus oxyrinchus

Point d’avancement

Vérification de la liste des mitogénomes avec comparaison des listes entre le fichiers Excel et les répertoires MITOGENOMES du work → OK

Choix des whole génomes selon l’espèce la plus proche (mails Joelle 9 et 10 oct):



Dans le cas de l’échantillon L0_4_EG9_S84 où l’espèce identifiée ne possède pas de whole génome de référence publié et qu’il faille utiliser celui d'une espèce voisine :

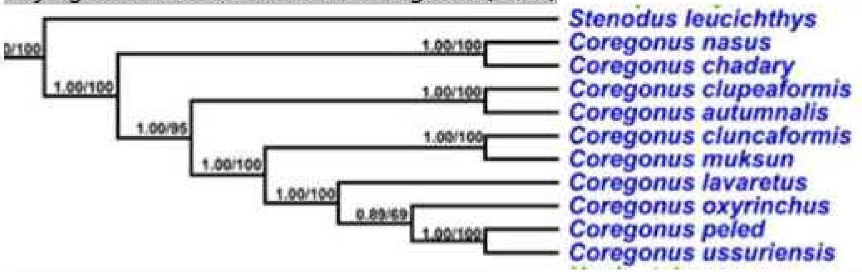
2. L0_4_EG9_S84:NC_012928.1:19566 → Thymallus thymallus
STATS/cleaned_L0_4_EG9_S84_R1_001.align.MAJORITAIRE
NC_056303.112714 Thymallus ligericus
=> *T. thymallus* est l’espèce la plus proche de *Thymallus ligericus* (GCA_023634145.1 without MT, this genome assembly includes 47 chromosomes and 578 unplaced scaffolds). Vérifier que la MT ne soit pas présente dans les scaffolds et, si c’était le cas, écarter les scaffolds. **Pour le génome MT à utiliser dans le pipeline Eager, je serais favorable à ce que l’on prenne celui de Thymallus ligericus (NC_056303.1).**

1. L11_7_SCO_1354_S70:NC_006531.1:182488 → Anguille
Anguilla/STATS $ more cleaned_L11_7_SCO_1354_S70_R1_001.align.MAJORITAIRE
NC_006535.3164611 Anguilla bicolor pacifica
Réponse : Un whole génome de réf de cette espèce existe, incluant le MT (GCA_039654925.2 with CM077320.2)

Point d’avancement

3. L0_5_EG10_S85:NC_020762.1:38793 → Coregonus
   STATS/cleaned_L0_5_EG10_S85_R1_001.align.MAJORITAIRE
   NC_002646.16117 Coregonus lavaretus
   Réponse : Un whole génome de réf de cette espèce existe, incluant le MT (GCA_964263955.1 with OZ182204.1).

4. L7_14_22d_S48:NC_020762.1:1255682 → Coregonus
   STATS/cleaned_L7_14_22d_S48_R1_001.align.MAJORITAIRE
   NC_020765.1186123 Coregonus oxyrinchus
   Phylogénie ci-dessous tirée de Wang et al (2022)



   \+ Tiré de Kroes et al (2023). "According to the IUCN Red List the anadromous houting Coregonus oxyrinchus is categorized as ‘extinct’. Our results suggest that C. oxyrinchus is a junior synonym of C. lavaretus."
   Réponse : Prendre le whole génome de réf de *C. lavaretus*, incluant le MT (GCA_964263955.1 with OZ182204.1).

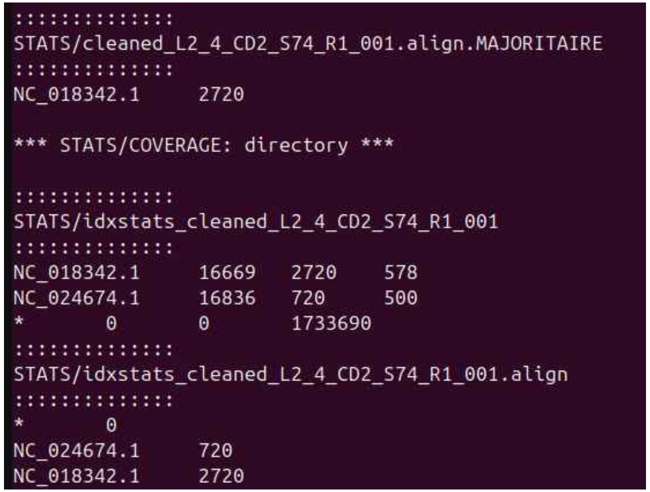
5. 01_Multimapping $ grep 'NC_018341.1' RESULTATS_MAJORITAIRES
   L2_4_CD2_S74:NC_018341.1:5799 → Brachymystax
   STATS/cleaned_L2_4_CD2_S74_R1_001.align.MAJORITAIRE
   NC_018341.1727 Brachymystax lenok
   Réponse : Un whole génome de réf de cette espèce existe sans le MT (GCA_030435695.1 without MT, this genome assembly includes 40 chromosomes and 27 unplaced scaffolds). Vérifier que la MT ne soit pas présente dans les scaffolds et, si c’était le cas, écarter les scaffolds.

Relance du 2nd multimapping pour Brachymystax

Brachymystax lenok		GCA_030435695.1		without	NC_018341.1
Brachymystax tumensis				without	NC_024674.1
Brachymystax lenok tsinlingensis				without	NC_018342.1

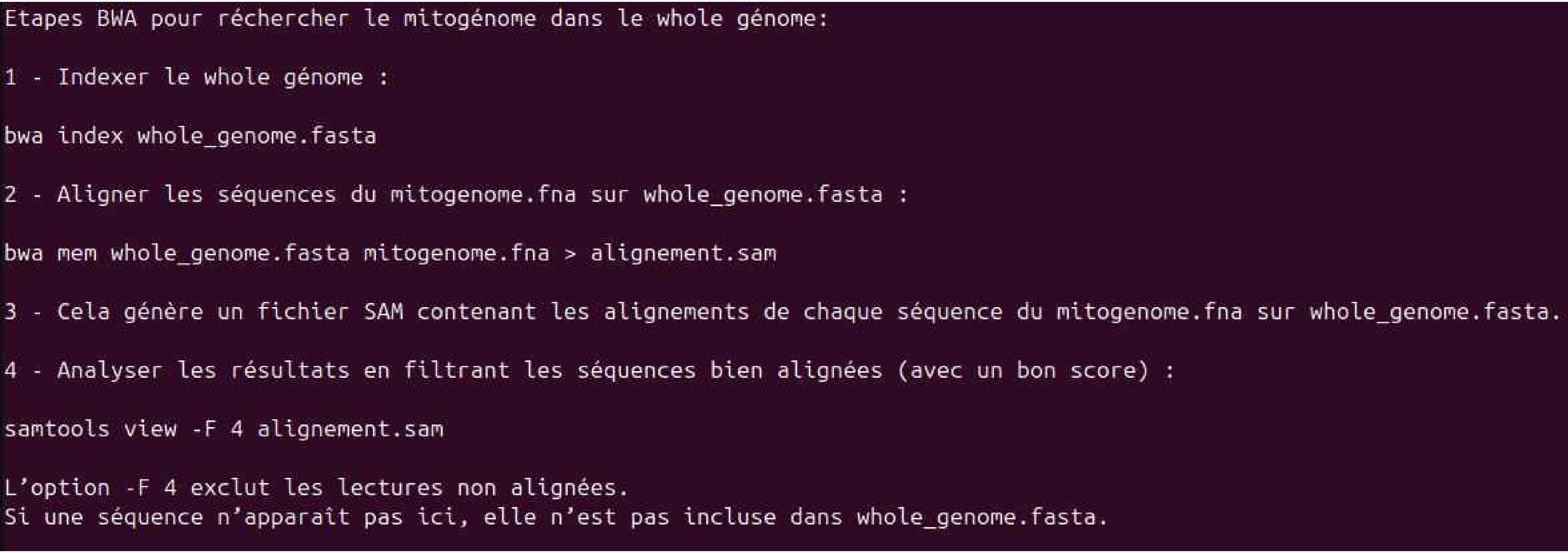
Nous n’avions pas considérer la sous-espèce la première fois or le whole génome est celui de la sous-espèce donc j’ai relancé le 2nd multimapping avec Brachymystax tumensis et Brachymystax lenok tsinlingensis uniquement.

Brachymystax lenok tsinlingensis est majoritaire :



```
::::::::::::::
STATS/cleaned_L2_4_CD2_S74_R1_001.align.MAJORITAIRE
::::::::::::::
NC_018342.1     2720

*** STATS/COVERAGE: directory ***

::::::::::::::
STATS/idxstats_cleaned_L2_4_CD2_S74_R1_001
::::::::::::::
NC_018342.1     16669   2720    578
NC_024674.1     16836   720     500
*       0       0       1733690
::::::::::::::
STATS/idxstats_cleaned_L2_4_CD2_S74_R1_001.align
::::::::::::::
*       0
NC_024674.1     720
NC_018342.1     2720
```

Méthodologie pour vérifier si le mitogénome est compris dans le whole génome (scaffolds)



```
Etapes BWA pour réchercher le mitogénome dans le whole génome:

1 - Indexer le whole génome :

bwa index whole_genome.fasta

2 - Aligner les séquences du mitogenome.fna sur whole_genome.fasta :

bwa mem whole_genome.fasta mitogenome.fna > alignement.sam

3 - Cela génère un fichier SAM contenant les alignements de chaque séquence du mitogenome.fna sur whole_genome.fasta.

4 - Analyser les résultats en filtrant les séquences bien alignées (avec un bon score) :

samtools view -F 4 alignement.sam

L’option -F 4 exclut les lectures non alignées.
Si une séquence n’apparaît pas ici, elle n’est pas incluse dans whole_genome.fasta.
```

Résultats : le mitogénome est – il compris dans le whole génome (scaffolds) ?

Espèces	Génomes	Inclus ?
Anguilla bicolor pacifica	GCA_039654925.2_AngPac_1_genomic.fna NC_006535.3.fna	OUI
Brachymystax lenox tsinlingensis	GCA_030435695.1_ASM3043569v1_Brachymystax_lenok_genomic.fna NC_018341.1.fna	OUI
Thymallus_ligericus	GCA_023634145.1_f_Thymallus_Thymallus_.pri.20220222_genomic.fna NC_056303.1.fna	OUI
Coregonus lavaretus	GCA_964263955.1_fCorLav1.hap1.1_genomic.fna NC_002646.1.fna	OUI
Coregonus oxyrinchus	GCA_964263955.1_fCorLav1.hap1.1_genomic.fna NC_020765.1.fna	OUI

- Exemple :

```
smaman@n003 /work/project/crucial/PALEOFISH/02_eager_notSalmo/Anguilla_bicolor_pacifica/GENOMES $ samtools view -F 4 alignement.sam | more
NC_006535.3     0       CM077320.2      3       60      5328S11365M     *       0       0       GTTAACGTAGCTTAAACAAAAGCATGGCACTGAAGATGCCAAGAT
TTAACATCAGTTCTGGCCTGACTTACACATGCAAGTACCCGCGTACCCGTGAGAATGCCCTATATCCCCTCTCGGGGAAAAGGAGCCGGCATCAGGCACACTAAACGTAGCCCAAAACGCCTTGCTCAGCCACACCCACAA
```

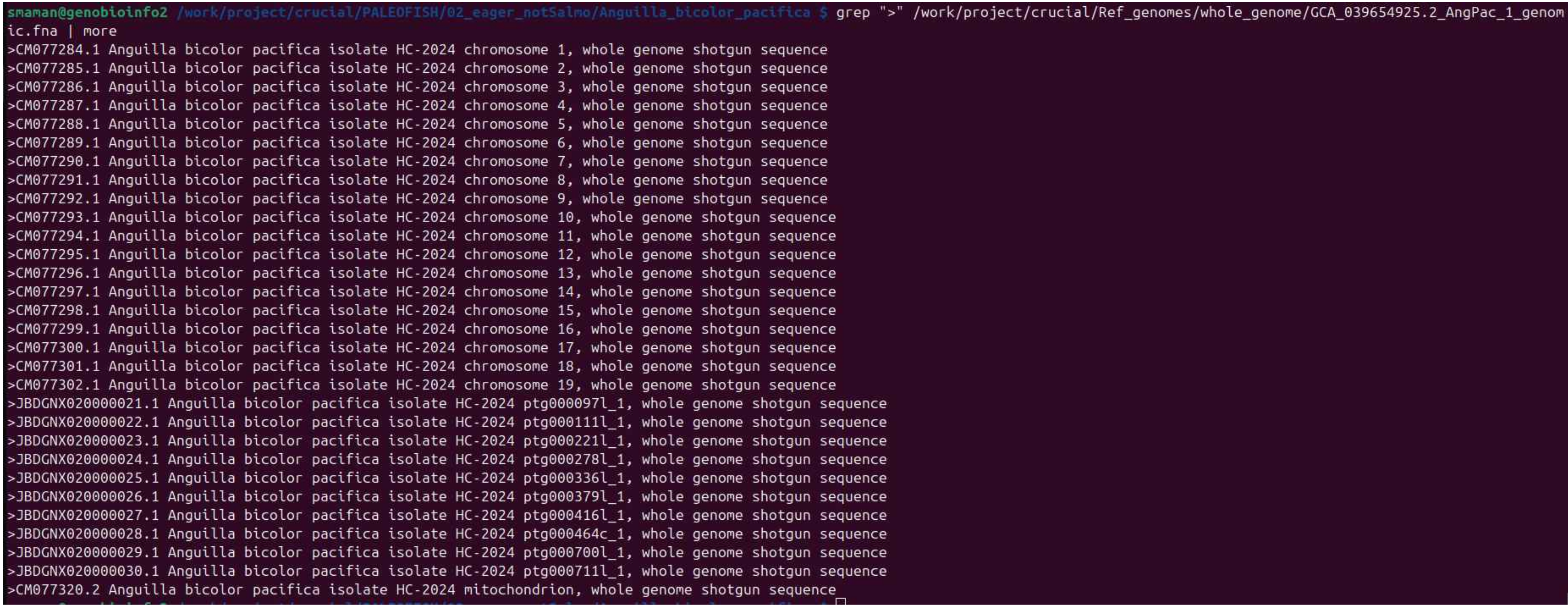
Lancement de Nextflow nf-core/eager double pass

Répertoire de travail : /work/project/crucial/PALEOFISH/02_eager_notSalmo
→ Pas de mito notée NC retrouvée dans le whole génome, donc inclusion dans scaffolds => BAM vide => cat whole+mito et écarter les scaffolds (en cours).

```
smaman@genobioinfo2 /work/project/crucial/PALEOFISH/02_eager_notSalmo/Anguilla_bicolor_pacifica $ more 01_script_eager_part1.sh
#!/bin/bash
#SBATCH -J eagerMito
#SBATCH -p workq
#SBATCH --mem=60G
#SBATCH --time=12:00:00

module purge;
module load bioinfo/NextflowWorkflows/nfcore-Nextflow-v21.10.6;

## Dossiers de travail
dir="/work/project/crucial/PALEOFISH/02_eager_notSalmo/Anguilla_bicolor_pacifica/";
input_file="/work/project/crucial/PALEOFISH/02_eager_notSalmo/Anguilla_bicolor_pacifica/input.tsv";
reference_genome="/work/project/crucial/Ref_genomes/whole_genome/GCA_039654925.2_AngPac_1_genomic.fna";   #GCA_039654925.2  https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_039654925.2
reference_mito="/work/project/crucial/PALEOFISH/01_detect_species_2nd_mapping_notSalmo/MITOGENOMES/NC_006535.3.fna";

#nextflow run nf-core/eager \ Modification du -x true en -x false
nextflow run /work/project/crucial/PALEOTROTA/02_eager/programs/eager/main.nf --input ${input_file} \
                           --fasta ${reference_genome}  \
                           -c '/work/project/crucial/PALEOFISH/02_eager_Salmo/crucial_nextflow_VersionPaleofish.config' \
                           --outdir ${dir} \
                           --udg_type none \
                           --colour_chemistry 2 \
                           --complexity_filter_poly_g true \
                           --complexity_filter_poly_g_min 7 \
                           --mergedonly true \
                           --mapper circularmapper \
                           --circularextension 200 \
                           --circulartarget NC_006535.3 \
                           --circularfilter true \
                           --damage_calculation_tool mapdamage \
                           --dedupper dedup \
                           --run_bam_filtering true \
                           --bam_mapping_quality_threshold 30 \
                           --bam_filter_minreadlength 30 \
                           --run_trim_bam false
smaman@genobioinfo2 /work/project/crucial/PALEOFISH/02_eager_notSalmo/Anguilla_bicolor_pacifica $ grep NC_006535 /work/project/crucial/Ref_genomes/whole_genome/GCA_039654925.2_AngPac_1
_genomic.fna
smaman@genobioinfo2 /work/project/crucial/PALEOFISH/02_eager_notSalmo/Anguilla_bicolor_pacifica $ 
```

Mitogénomes dans whole génome

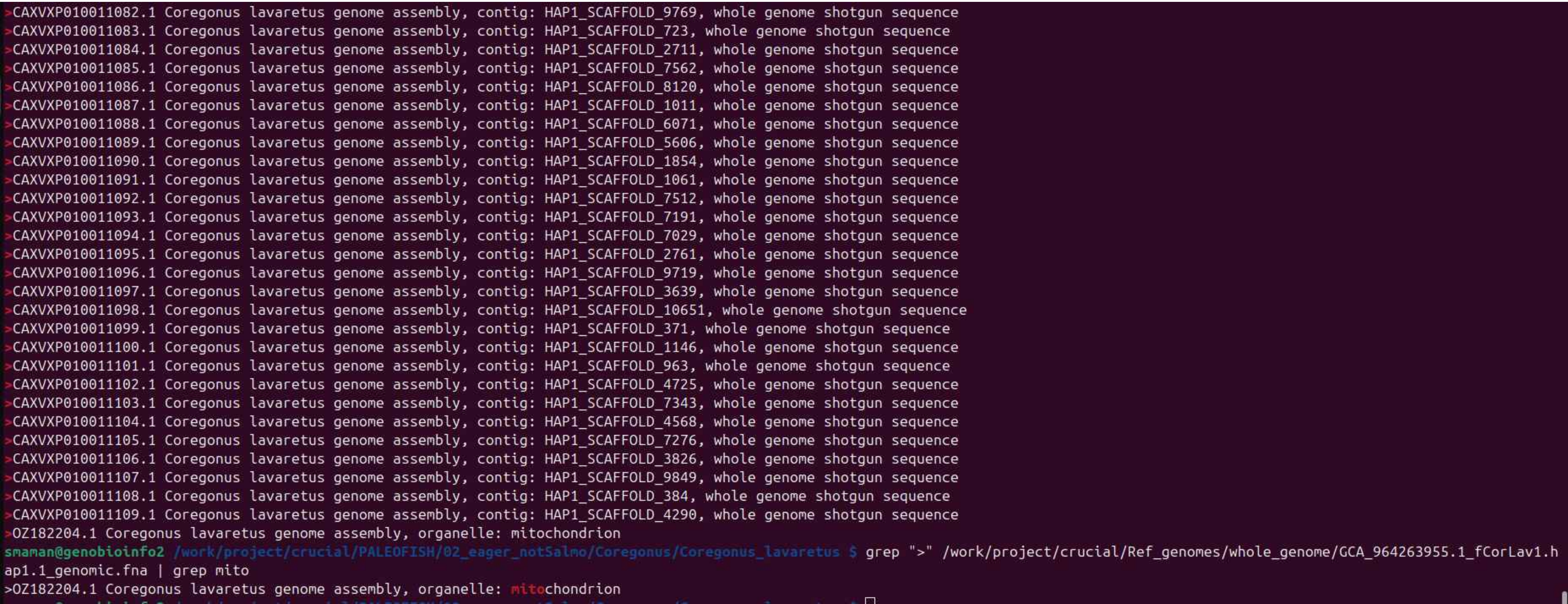
Identification des mitogénomes :
Anguilla : CM077320.2 Anguilla bicolor pacifica isolate HC-2024 mitochondrion, whole genome shotgun sequence
→ Remplacer le mitogénome NC_006535.3 par CM077320.2



```
smaman@genobioinfo2 /work/project/crucial/PALEOFISH/02_eager_notSalmo/Anguilla_bicolor_pacifica $ grep ">" /work/project/crucial/Ref_genomes/whole_genome/GCA_039654925.2_AngPac_1_genom
ic.fna | more
>CM077284.1 Anguilla bicolor pacifica isolate HC-2024 chromosome 1, whole genome shotgun sequence
>CM077285.1 Anguilla bicolor pacifica isolate HC-2024 chromosome 2, whole genome shotgun sequence
>CM077286.1 Anguilla bicolor pacifica isolate HC-2024 chromosome 3, whole genome shotgun sequence
>CM077287.1 Anguilla bicolor pacifica isolate HC-2024 chromosome 4, whole genome shotgun sequence
>CM077288.1 Anguilla bicolor pacifica isolate HC-2024 chromosome 5, whole genome shotgun sequence
>CM077289.1 Anguilla bicolor pacifica isolate HC-2024 chromosome 6, whole genome shotgun sequence
>CM077290.1 Anguilla bicolor pacifica isolate HC-2024 chromosome 7, whole genome shotgun sequence
>CM077291.1 Anguilla bicolor pacifica isolate HC-2024 chromosome 8, whole genome shotgun sequence
>CM077292.1 Anguilla bicolor pacifica isolate HC-2024 chromosome 9, whole genome shotgun sequence
>CM077293.1 Anguilla bicolor pacifica isolate HC-2024 chromosome 10, whole genome shotgun sequence
>CM077294.1 Anguilla bicolor pacifica isolate HC-2024 chromosome 11, whole genome shotgun sequence
>CM077295.1 Anguilla bicolor pacifica isolate HC-2024 chromosome 12, whole genome shotgun sequence
>CM077296.1 Anguilla bicolor pacifica isolate HC-2024 chromosome 13, whole genome shotgun sequence
>CM077297.1 Anguilla bicolor pacifica isolate HC-2024 chromosome 14, whole genome shotgun sequence
>CM077298.1 Anguilla bicolor pacifica isolate HC-2024 chromosome 15, whole genome shotgun sequence
>CM077299.1 Anguilla bicolor pacifica isolate HC-2024 chromosome 16, whole genome shotgun sequence
>CM077300.1 Anguilla bicolor pacifica isolate HC-2024 chromosome 17, whole genome shotgun sequence
>CM077301.1 Anguilla bicolor pacifica isolate HC-2024 chromosome 18, whole genome shotgun sequence
>CM077302.1 Anguilla bicolor pacifica isolate HC-2024 chromosome 19, whole genome shotgun sequence
>JBDGNX020000021.1 Anguilla bicolor pacifica isolate HC-2024 ptg000097l_1, whole genome shotgun sequence
>JBDGNX020000022.1 Anguilla bicolor pacifica isolate HC-2024 ptg000111l_1, whole genome shotgun sequence
>JBDGNX020000023.1 Anguilla bicolor pacifica isolate HC-2024 ptg000221l_1, whole genome shotgun sequence
>JBDGNX020000024.1 Anguilla bicolor pacifica isolate HC-2024 ptg000278l_1, whole genome shotgun sequence
>JBDGNX020000025.1 Anguilla bicolor pacifica isolate HC-2024 ptg000336l_1, whole genome shotgun sequence
>JBDGNX020000026.1 Anguilla bicolor pacifica isolate HC-2024 ptg000379l_1, whole genome shotgun sequence
>JBDGNX020000027.1 Anguilla bicolor pacifica isolate HC-2024 ptg000416l_1, whole genome shotgun sequence
>JBDGNX020000028.1 Anguilla bicolor pacifica isolate HC-2024 ptg000464c_1, whole genome shotgun sequence
>JBDGNX020000029.1 Anguilla bicolor pacifica isolate HC-2024 ptg000700l_1, whole genome shotgun sequence
>JBDGNX020000030.1 Anguilla bicolor pacifica isolate HC-2024 ptg000711l_1, whole genome shotgun sequence
>CM077320.2 Anguilla bicolor pacifica isolate HC-2024 mitochondrion, whole genome shotgun sequence
```

Mitogénomes dans whole génome

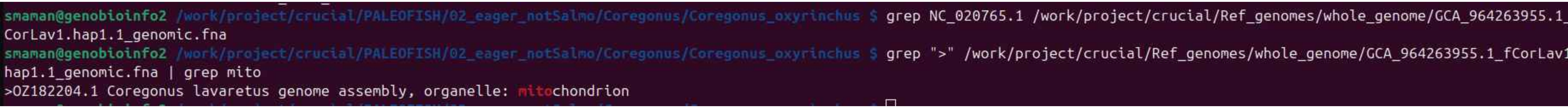
Identification des mitogénomes :
Coregonus l.:
→ Remplacer le mitogénome NC_002646.1 par OZ182204.1



```
>CAXVXP010011082.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_9769, whole genome shotgun sequence
>CAXVXP010011083.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_723, whole genome shotgun sequence
>CAXVXP010011084.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_2711, whole genome shotgun sequence
>CAXVXP010011085.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_7562, whole genome shotgun sequence
>CAXVXP010011086.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_8120, whole genome shotgun sequence
>CAXVXP010011087.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_1011, whole genome shotgun sequence
>CAXVXP010011088.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_6071, whole genome shotgun sequence
>CAXVXP010011089.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_5606, whole genome shotgun sequence
>CAXVXP010011090.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_1854, whole genome shotgun sequence
>CAXVXP010011091.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_1061, whole genome shotgun sequence
>CAXVXP010011092.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_7512, whole genome shotgun sequence
>CAXVXP010011093.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_7191, whole genome shotgun sequence
>CAXVXP010011094.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_7029, whole genome shotgun sequence
>CAXVXP010011095.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_2761, whole genome shotgun sequence
>CAXVXP010011096.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_9719, whole genome shotgun sequence
>CAXVXP010011097.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_3639, whole genome shotgun sequence
>CAXVXP010011098.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_10651, whole genome shotgun sequence
>CAXVXP010011099.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_371, whole genome shotgun sequence
>CAXVXP010011100.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_1146, whole genome shotgun sequence
>CAXVXP010011101.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_963, whole genome shotgun sequence
>CAXVXP010011102.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_4725, whole genome shotgun sequence
>CAXVXP010011103.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_7343, whole genome shotgun sequence
>CAXVXP010011104.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_4568, whole genome shotgun sequence
>CAXVXP010011105.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_7276, whole genome shotgun sequence
>CAXVXP010011106.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_3826, whole genome shotgun sequence
>CAXVXP010011107.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_9849, whole genome shotgun sequence
>CAXVXP010011108.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_384, whole genome shotgun sequence
>CAXVXP010011109.1 Coregonus lavaretus genome assembly, contig: HAP1_SCAFFOLD_4290, whole genome shotgun sequence
>OZ182204.1 Coregonus lavaretus genome assembly, organelle: mitochondrion
smaman@genobioinfo2 /work/project/crucial/PALEOFISH/02_eager_notSalmo/Coregonus/Coregonus_lavaretus $ grep ">" /work/project/crucial/Ref_genomes/whole_genome/GCA_964263955.1_fCorLav1.h
ap1.1_genomic.fna | grep mito
>OZ182204.1 Coregonus lavaretus genome assembly, organelle: mitochondrion
```

Mitogénomes dans whole génome

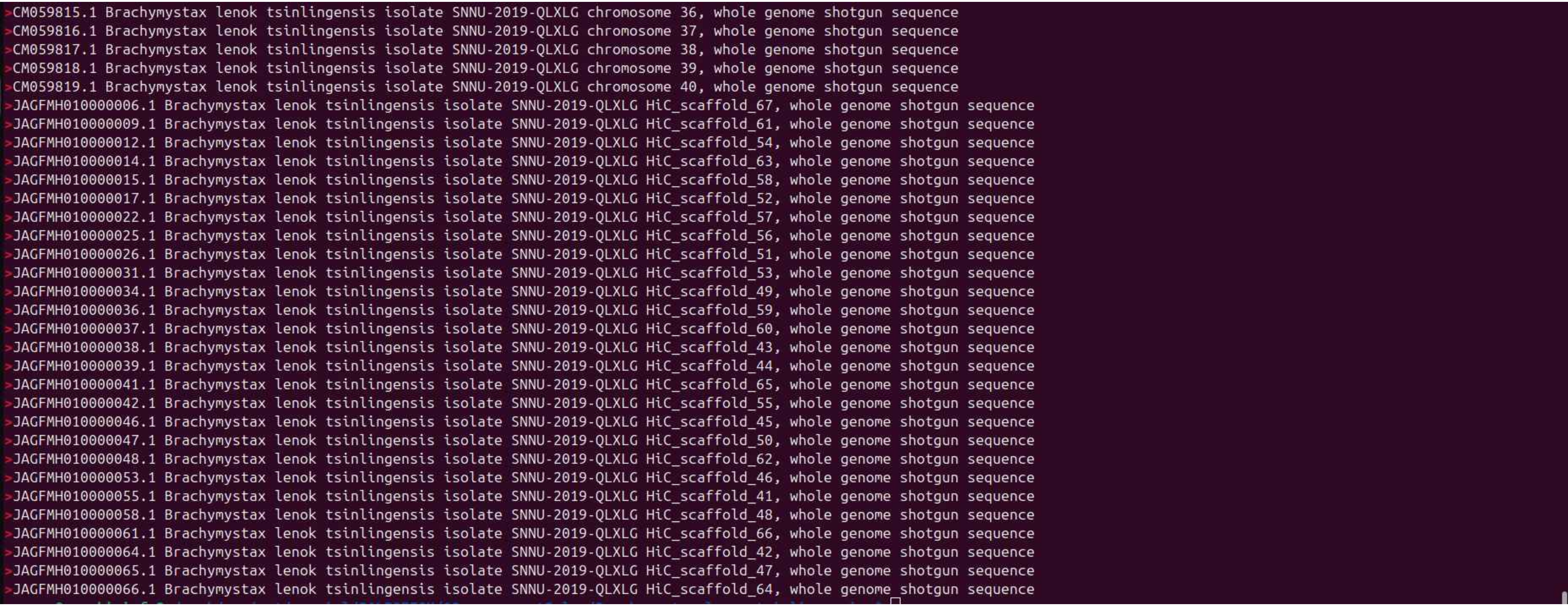
Identification des mitogénomes :
Coregonus l.:
→ Remplacer le mitogénome NC_020765.1 par OZ182204.1



```
smaman@genobioinfo2 /work/project/crucial/PALEOFISH/02_eager_notSalmo/Coregonus/Coregonus_oxyrinchus $ grep NC_020765.1 /work/project/crucial/Ref_genomes/whole_genome/GCA_964263955.1_
CorLav1.hap1.1_genomic.fna
smaman@genobioinfo2 /work/project/crucial/PALEOFISH/02_eager_notSalmo/Coregonus/Coregonus_oxyrinchus $ grep ">" /work/project/crucial/Ref_genomes/whole_genome/GCA_964263955.1_fCorLav1
hap1.1_genomic.fna | grep mito
>OZ182204.1 Coregonus lavaretus genome assembly, organelle: mitochondrion
```

Retrait des scaffolds dans whole génome

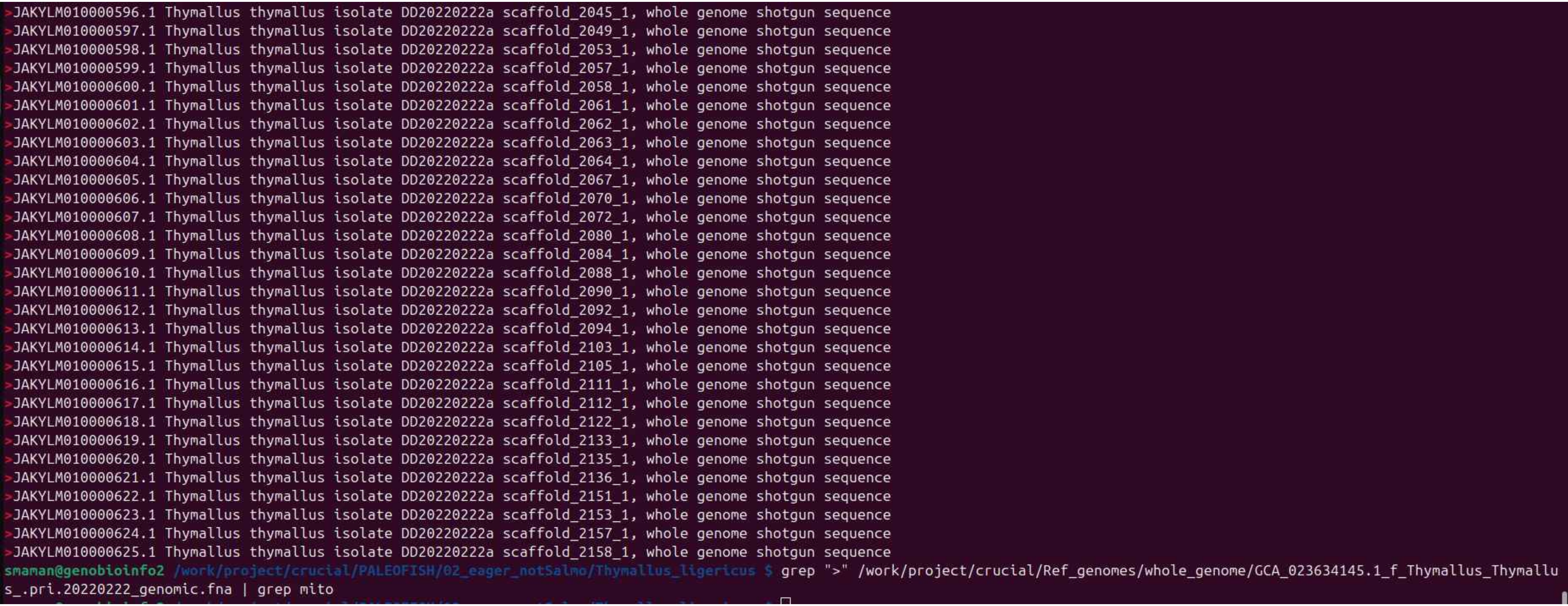
L’objectif est d’identifier puis d’écarter les scaffolds dans les whole génomes.
Brachymystax : pas de mitogénome, scaffolds.



```
>CM059815.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG chromosome 36, whole genome shotgun sequence
>CM059816.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG chromosome 37, whole genome shotgun sequence
>CM059817.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG chromosome 38, whole genome shotgun sequence
>CM059818.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG chromosome 39, whole genome shotgun sequence
>CM059819.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG chromosome 40, whole genome shotgun sequence
>JAGFMH010000006.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_67, whole genome shotgun sequence
>JAGFMH010000009.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_61, whole genome shotgun sequence
>JAGFMH010000012.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_54, whole genome shotgun sequence
>JAGFMH010000014.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_63, whole genome shotgun sequence
>JAGFMH010000015.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_58, whole genome shotgun sequence
>JAGFMH010000017.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_52, whole genome shotgun sequence
>JAGFMH010000022.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_57, whole genome shotgun sequence
>JAGFMH010000025.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_56, whole genome shotgun sequence
>JAGFMH010000026.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_51, whole genome shotgun sequence
>JAGFMH010000031.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_53, whole genome shotgun sequence
>JAGFMH010000034.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_49, whole genome shotgun sequence
>JAGFMH010000036.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_59, whole genome shotgun sequence
>JAGFMH010000037.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_60, whole genome shotgun sequence
>JAGFMH010000038.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_43, whole genome shotgun sequence
>JAGFMH010000039.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_44, whole genome shotgun sequence
>JAGFMH010000041.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_65, whole genome shotgun sequence
>JAGFMH010000042.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_55, whole genome shotgun sequence
>JAGFMH010000046.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_45, whole genome shotgun sequence
>JAGFMH010000047.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_50, whole genome shotgun sequence
>JAGFMH010000048.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_62, whole genome shotgun sequence
>JAGFMH010000053.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_46, whole genome shotgun sequence
>JAGFMH010000055.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_41, whole genome shotgun sequence
>JAGFMH010000058.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_48, whole genome shotgun sequence
>JAGFMH010000061.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_66, whole genome shotgun sequence
>JAGFMH010000064.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_42, whole genome shotgun sequence
>JAGFMH010000065.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_47, whole genome shotgun sequence
>JAGFMH010000066.1 Brachymystax lenok tsinlingensis isolate SNNU-2019-QLXLG HiC_scaffold_64, whole genome shotgun sequence
```

Retrait des scaffolds dans whole génome

L’objectif est d’identifier puis d’écarter les scaffolds dans les whole génomes.
Thymallus t. : pas de mitogénome, scaffolds.



```
>JAKYLM010000596.1 Thymallus thymallus isolate DD20220222a scaffold_2045_1, whole genome shotgun sequence
>JAKYLM010000597.1 Thymallus thymallus isolate DD20220222a scaffold_2049_1, whole genome shotgun sequence
>JAKYLM010000598.1 Thymallus thymallus isolate DD20220222a scaffold_2053_1, whole genome shotgun sequence
>JAKYLM010000599.1 Thymallus thymallus isolate DD20220222a scaffold_2057_1, whole genome shotgun sequence
>JAKYLM010000600.1 Thymallus thymallus isolate DD20220222a scaffold_2058_1, whole genome shotgun sequence
>JAKYLM010000601.1 Thymallus thymallus isolate DD20220222a scaffold_2061_1, whole genome shotgun sequence
>JAKYLM010000602.1 Thymallus thymallus isolate DD20220222a scaffold_2062_1, whole genome shotgun sequence
>JAKYLM010000603.1 Thymallus thymallus isolate DD20220222a scaffold_2063_1, whole genome shotgun sequence
>JAKYLM010000604.1 Thymallus thymallus isolate DD20220222a scaffold_2064_1, whole genome shotgun sequence
>JAKYLM010000605.1 Thymallus thymallus isolate DD20220222a scaffold_2067_1, whole genome shotgun sequence
>JAKYLM010000606.1 Thymallus thymallus isolate DD20220222a scaffold_2070_1, whole genome shotgun sequence
>JAKYLM010000607.1 Thymallus thymallus isolate DD20220222a scaffold_2072_1, whole genome shotgun sequence
>JAKYLM010000608.1 Thymallus thymallus isolate DD20220222a scaffold_2080_1, whole genome shotgun sequence
>JAKYLM010000609.1 Thymallus thymallus isolate DD20220222a scaffold_2084_1, whole genome shotgun sequence
>JAKYLM010000610.1 Thymallus thymallus isolate DD20220222a scaffold_2088_1, whole genome shotgun sequence
>JAKYLM010000611.1 Thymallus thymallus isolate DD20220222a scaffold_2090_1, whole genome shotgun sequence
>JAKYLM010000612.1 Thymallus thymallus isolate DD20220222a scaffold_2092_1, whole genome shotgun sequence
>JAKYLM010000613.1 Thymallus thymallus isolate DD20220222a scaffold_2094_1, whole genome shotgun sequence
>JAKYLM010000614.1 Thymallus thymallus isolate DD20220222a scaffold_2103_1, whole genome shotgun sequence
>JAKYLM010000615.1 Thymallus thymallus isolate DD20220222a scaffold_2105_1, whole genome shotgun sequence
>JAKYLM010000616.1 Thymallus thymallus isolate DD20220222a scaffold_2111_1, whole genome shotgun sequence
>JAKYLM010000617.1 Thymallus thymallus isolate DD20220222a scaffold_2112_1, whole genome shotgun sequence
>JAKYLM010000618.1 Thymallus thymallus isolate DD20220222a scaffold_2122_1, whole genome shotgun sequence
>JAKYLM010000619.1 Thymallus thymallus isolate DD20220222a scaffold_2133_1, whole genome shotgun sequence
>JAKYLM010000620.1 Thymallus thymallus isolate DD20220222a scaffold_2135_1, whole genome shotgun sequence
>JAKYLM010000621.1 Thymallus thymallus isolate DD20220222a scaffold_2136_1, whole genome shotgun sequence
>JAKYLM010000622.1 Thymallus thymallus isolate DD20220222a scaffold_2151_1, whole genome shotgun sequence
>JAKYLM010000623.1 Thymallus thymallus isolate DD20220222a scaffold_2153_1, whole genome shotgun sequence
>JAKYLM010000624.1 Thymallus thymallus isolate DD20220222a scaffold_2157_1, whole genome shotgun sequence
>JAKYLM010000625.1 Thymallus thymallus isolate DD20220222a scaffold_2158_1, whole genome shotgun sequence
smaman@genobioinfo2 /work/project/crucial/PALEOFISH/02_eager_notSalmo/Thymallus_ligericus $ grep ">" /work/project/crucial/Ref_genomes/whole_genome/GCA_023634145.1_f_Thymallus_Thymallu
s_.pri.20220222_genomic.fna | grep mito
```