

Paleofish Mitochondries

Objectif : Lancement de Eager sur référence majoritaire

CR 17 sept 2025

Présentes: Joelle, Sarah. Points abordés:

- Présentation des traitements réalisés la semaine dernière: Eager pour *S. salar* et *trutta*, Répartition des reads après mapping sur mitogénomes par échantillons.
- Slides d'avancement du travail : https://web-genobioinfo.toulouse.inrae.fr/~sigenae/sarah/PALEOFISH_mito/avancement_3oct25.pdf
- Bilan des mito génomes à conserver pour le second multimapping des échantillons autres que *Salmo* (*salar* ou *trutta*):
https://web-genobioinfo.toulouse.inrae.fr/~sigenae/sarah/PALEOFISH_mito/LINKS_Table_GenomeDeReferenceMito_IntraGenre_Anguilla_Brachymystax_Coregonus_Thymallus.xlsx
- Si le mito génome RefSeq n'est pas disponible, alors intégrer le mito génome Genbank au multifasta.
- Retrait des lignes redondantes et des sous-espèces dans la table des génomes de référence ci-dessus.
- L'étape de multimapping intègre l'ensemble des mito génomes listés dans le fichier xls ci-dessus, même si le whole génome n'est pas disponible pour l'étape Eager. Un whole génome le plus proche sera choisi.
- Regarder si un whole génome assemblé *Hucho hucho* est disponible.
- Utiliser le whole génome *Brachy. lenox* auquel on ajoute le génome mito pour l'étape Eager. Idem pour les autres whole génomes sans mito génome.
- Lancer le second multimapping pour chaque genre, pour les 5 échantillons suivants:
 - L11_7_SCO_1354_S70:NC_006531.1:182488 → Anguille
 - L0_4_EG9_S84:NC_012928.1:19566 → *Thymallus thymallus*
 - L2_4_CD2_S74:NC_018341.1:5799 → *Brachymystax*
 - L0_5_EG10_S85:NC_020762.1:38793 → *Coregonus*
 - L7_14_22d_S48:NC_020762.1:1255682 → *Coregonus*
- Renommage des répertoires de travail
- Les résultats des multimapping seront discutés avant de lancer les étapes Eager.

5 échantillons non *S. trutta* ni *S. salar*

smaman@genobioinfo1 /work/project/crucial/PALEOFISH/01_Multimapping \$ grep 'NC_006531.1' RESULTATS_MAJORITAIRES
L11_7_SCO_1354_S70:NC_006531.1:182488 → Anguille

smaman@genobioinfo1 /work/project/crucial/PALEOFISH/01_Multimapping \$ grep 'NC_012928.1' RESULTATS_MAJORITAIRES
L0_4_EG9_S84:NC_012928.1:19566 → Thymallus thymallus

smaman@genobioinfo1 /work/project/crucial/PALEOFISH/01_Multimapping \$ grep 'NC_018341.1' RESULTATS_MAJORITAIRES
L2_4_CD2_S74:NC_018341.1:5799 → Brachymystax

smaman@genobioinfo1 /work/project/crucial/PALEOFISH/01_Multimapping \$ grep 'NC_020762.1' RESULTATS_MAJORITAIRES
L0_5_EG10_S85:NC_020762.1:38793 → Coregonus
L7_14_22d_S48:NC_020762.1:1255682

=> Joelle : liste des autres génomes mito pour lancer un multimapping sur ce nouveau multifasta.

- 4 Undetermined
- 5 Autres que *S. salar* et *S. trutta*
- 32 *S. salar*
- 47 *S. trutta*

88 échantillons

Résultats du second multimapping pour les 5 échantillons non *S. trutta* ni *S. salar*

01_Multimapping \$ grep 'NC_006531.1' RESULTATS_MAJORITAIRES
L11_7_SCO_1354_S70:NC_006531.1:182488 → **Anguille**
Anguilla/STATS \$ more cleaned_L11_7_SCO_1354_S70_R1_001.align.MAJORITAIRE
NC_006535.3 **164611** *Anguilla bicolor pacifica*

01_Multimapping \$ grep 'NC_012928.1' RESULTATS_MAJORITAIRES
L0_4_EG9_S84:NC_012928.1:19566 → **Thymallus thymallus**
STATS/cleaned_L0_4_EG9_S84_R1_001.align.MAJORITAIRE
NC_056303.1 **12714** *Thymallus ligericus*

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_018341.1 **727** *Brachymystax lenok*

01_Multimapping \$ grep 'NC_020762.1' RESULTATS_MAJORITAIRES
L0_5_EG10_S85:NC_020762.1:38793 → **Coregonus**
L7_14_22d_S48:NC_020762.1:1255682
STATS/cleaned_L0_5_EG10_S85_R1_001.align.MAJORITAIRE
NC_002646.1 **6117** *Coregonus lavaretus*
STATS/cleaned_L7_14_22d_S48_R1_001.align.MAJORITAIRE
NC_020765.1 **186123** *Coregonus oxyrinchus*

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 prendre le whole génome *Brachymystax lenox tsinlingensis* qui est le seul disponible.
- Vérifier si les mito sont dans les scaffolds du whole génome du *Brachymystax lenox 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 lenox 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