

**Auteurs :**

Clément BIRBES (BIOS4Biol)

Margot ZAHM (BIOS4Biol)

Christophe KLOPP (BIOS4Biol)

Claire KUCHLY (BIOS4Biol)

**Production :**

Assembly ErrorCorrection\_Pipeline

**Résumé :**

Ce pipeline Nextflow permet le polishing de génome à partir d'un simple assemblage et de reads long et/ou court. Il utilise les outils de polishing long reads et short reads les plus efficace et ce de façon optimisé de manière a réduire au maximum les temps d'exécution.

Ce pipeline permet également la correction des erreurs introduites par les assembleurs et les polisher via une étape de « splitting ».

Le pipeline est efficace pour l'ensemble des génomes eukaryotes, quelque soit la taille ou les spécificités de ce dernier.

En plus de sa capacité à améliorer la qualité d'un assemblage, ce pipeline permet d'évaluer les métriques du nouvel assemblage via les outils Busco, Kat et Quast.

**Disponible ici :**

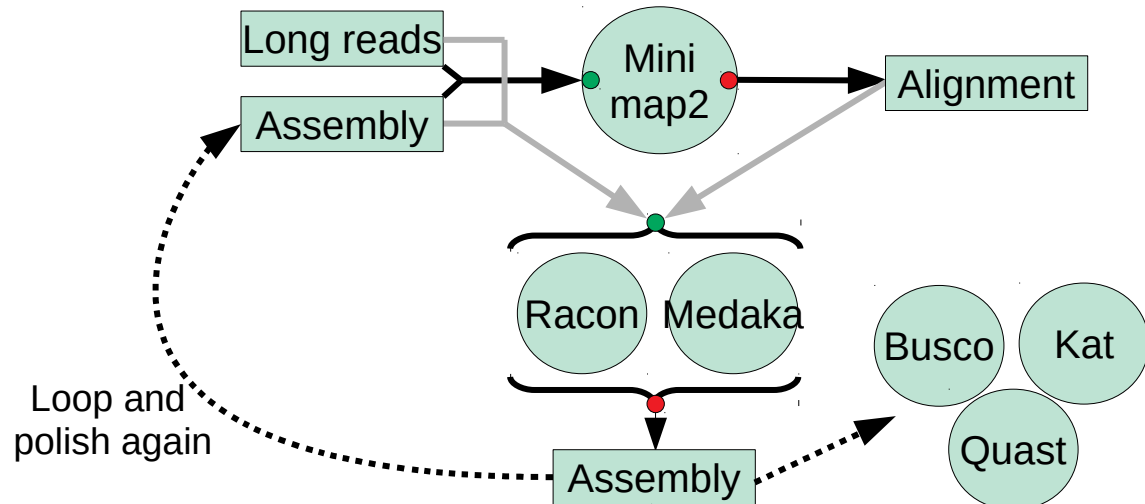
[https://forgemia.inra.fr/seqoccin/axis1-assembly/-/tree/dev/errorCorrection\\_Pipeline](https://forgemia.inra.fr/seqoccin/axis1-assembly/-/tree/dev/errorCorrection_Pipeline)

### Etapes du pipeline :

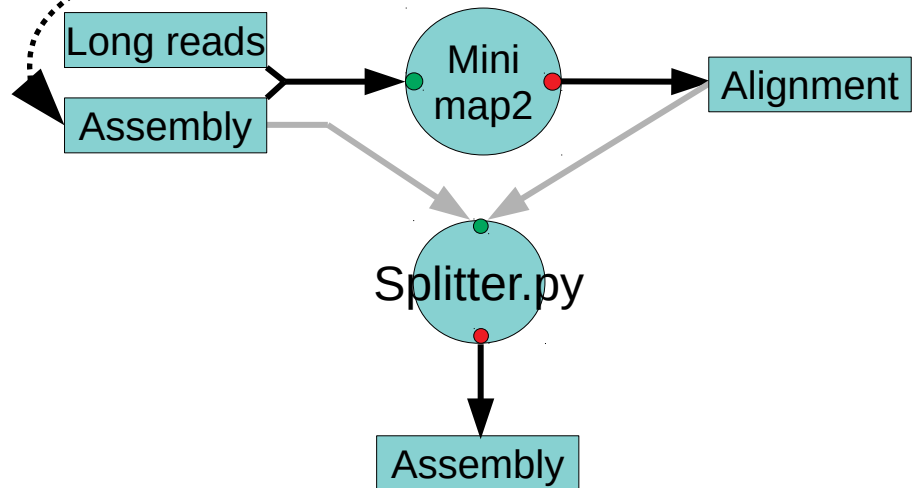
Cette image présente les différentes étapes du pipeline, chacun des 3 parties distincte du pipeline est optionnelle, mais une utilisation complète permet d'obtenir de meilleurs résultats.

Les valeurs par défaut permettent d'obtenir un score busco > à 94 % sur l'ensemble des génomes testés : Bos taurus, Coturnix japonica, Zea mays, Phaseolus vulgaris et Sus crofa

#### Polishing Long reads



#### Split and Correct Long reads



#### Polishing Short reads

