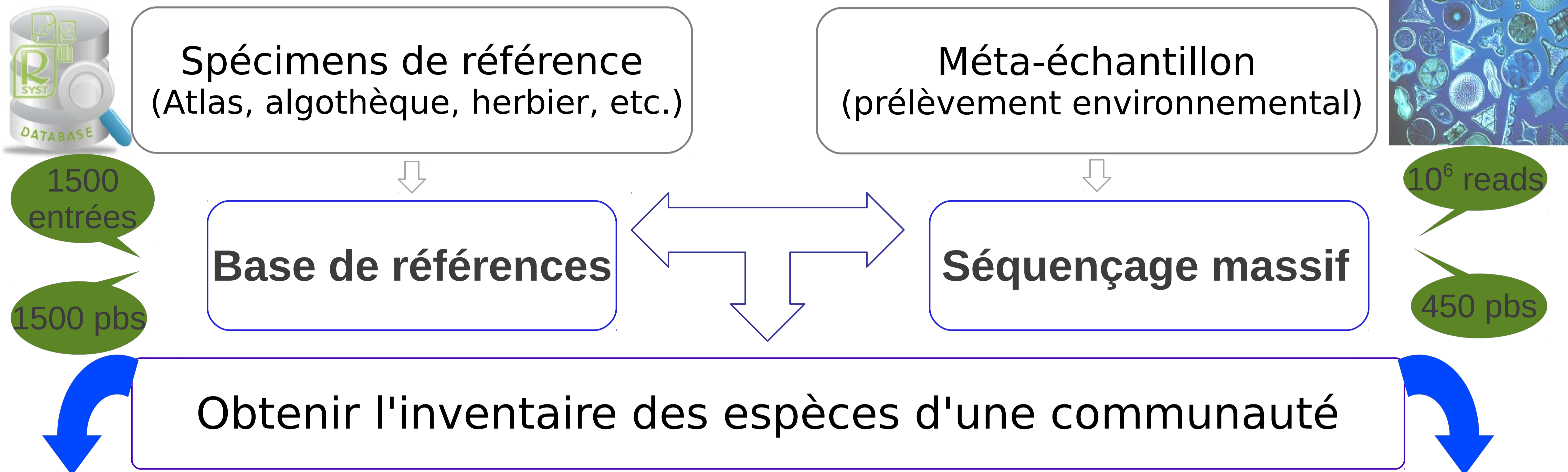


MetaMatch: un algorithme fiable et précis pour l'assignation taxonomique au niveau de l'espèce

Jean-Marc Frigerio (1) Philippe Chaumeil (1), Pierre Gay (2), Lenaïg Kermarrec (3,4), Frédéric Rimet (4), Agnès Bouchez (4), and Alain Franc (1)

¹INRA, UMR BioGeCo, 33610 Cestas, France, ²Université de Bordeaux 1, 33400 Talence, France. ³Asconit Consultants, 66350 Toulouges, France ⁴INRA, UMR CARRETEL, 74200 Thonon-les-Bains,



MetaMatch: Test d'assignation de chaque read sur une Base de Référence

metaMatch parfait: 0/1
Assignation si correspondance parfaite entre le read et une partie de la séquence de référence.

metaMatch imparfait
distance reliée au score de l'algorithme de Smith & Waterman (alignement local)

Assignation si distance < seuil

Utilisation des infrastructures

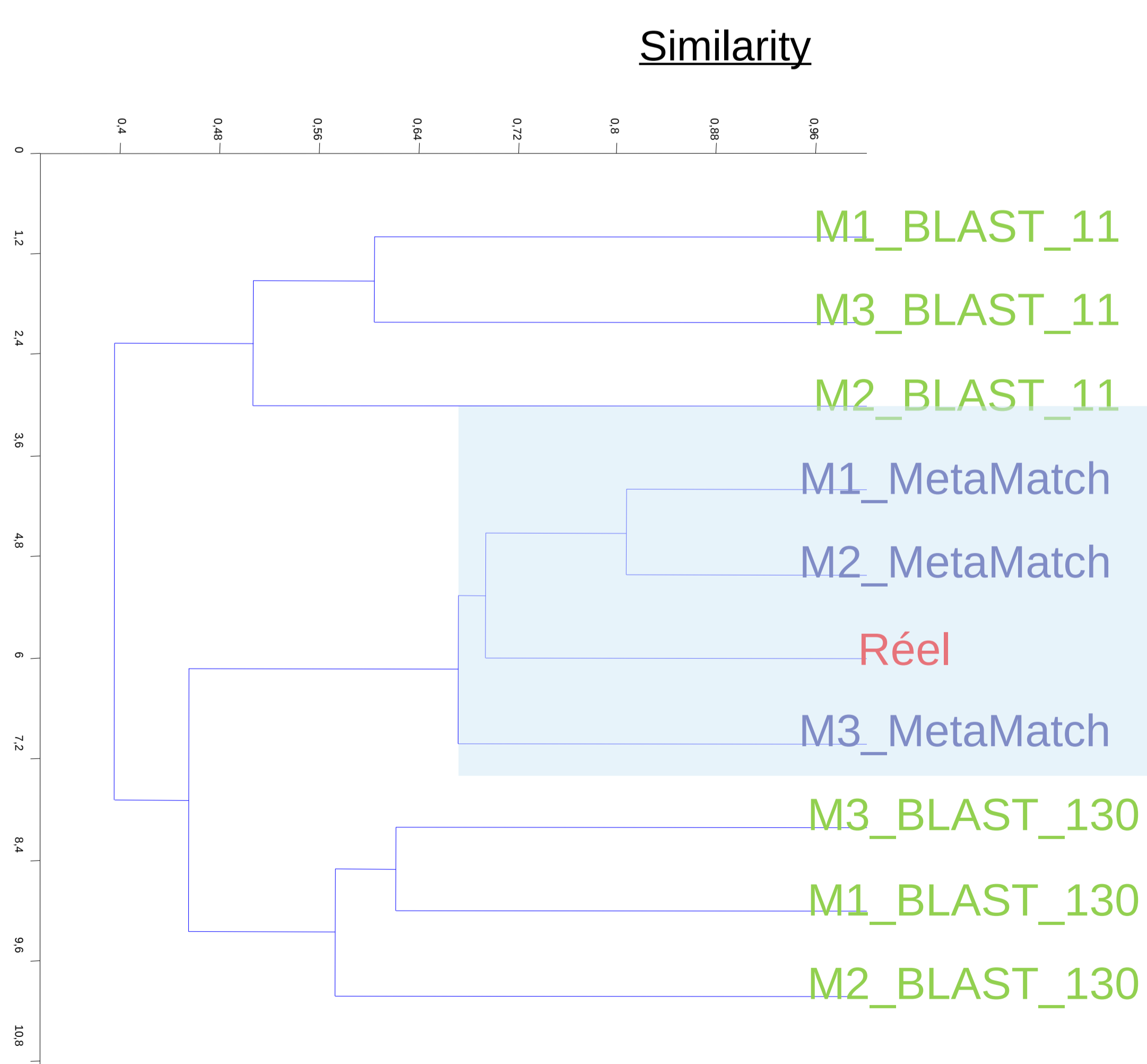
Algorithme très facilement distribuable

Implémenté sur :
cluster Avakas du Mésocentre de Bordeaux (3168 cœurs Intel Xeon X5675) - Ordonnanceur TORQUE+MAUI

Et sur grille EGI, VO France-Grille via l'interface web Dirac et les commandes en ligne Dirac

Perspectives

- Améliorer la vitesse de calcul
- Étudier les cohérences/incohérences taxonomiques morphologie/moléculaire
- Modèle statistique d'assignation
- Étudier la structure de la base (image euclidienne)



Dendrogramme de classification ascendante hiérarchique sur des distances de Jaccard. **Réel**: inventaire assemblé à partir de lignées de diatomées connues. **M(1-3)_BLAST_11**: inventaire obtenu avec Blast (paramètres par défaut, -W 11), **M(1-3)_BLAST_130**, avec Blast paramètre -W 130. **M(1-3)_MetaMatch**: distances produites par metaMatch. M(1-3) sont des mélanges en proportions connues de lignées de diatomées.

