

# Journée Calcul INRA

## Disseq / MpiDisseq



# Démarche scientifique et calcul

- Idée : construire une image géométrique de la diversité moléculaire d'une communauté (écologie moléculaire) comme forme d'un nuage de points
- Calcul d'une matrice de distances entre séquences 2 à 2 : production d'une matrice pleine
- On a 100 000 séquences, soit une matrice 100 000 x 100 000
- Tous les calculs peuvent être fait en parallèle indépendamment les uns des autres
- Alignement local de Smith-Waterman (algorithme exact, sans heuristiques)

# Disseq et E-Biothon

- Projet initial : parallélisme de jobs
- Machine d'exécution : IBM Blue Gene/P
  - Parallélisme intensif
  - Parallélisme de job très limité, voire impossible
  - Mémoire distribuée
- L'évolution du projet et les échanges avec l'IDRIS ont mis en évidence la nécessité d'évoluer vers un véritable parallélisme pour s'adapter aux configurations des machines d'exploitation (Babel puis Turing)

# Architecture IBM Blue Gene/Q

- Une machine frontale IBM Power7
- 6 racks comportant 1024 nœuds de 16 cœurs chacun, soit 98.304 cœurs
- 64 Go de mémoire par nœud
- Processeurs avec une excellente efficacité énergétique mais plus lents que ceux des architectures classiques (x86)

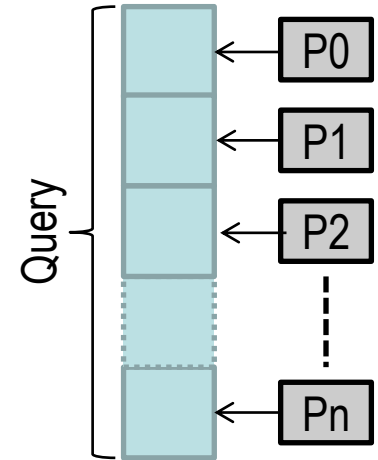


# Pourquoi MPI ?

- Les plus
  - Fonctionne sur n'importe quelle architecture
  - Permet d'utiliser la totalité des cœurs disponibles sur la machine
  - Utilisable avec tous les langages (Fortran, C, C++, Python)
  - Diminue le temps de restitution d'une exécution
  - Un seul job à soumettre, une seule sortie de job à analyser
- Le moins
  - Nécessite une modification du code mais le parallélisme massif de job nécessite aussi du développement

# MpiDisseq

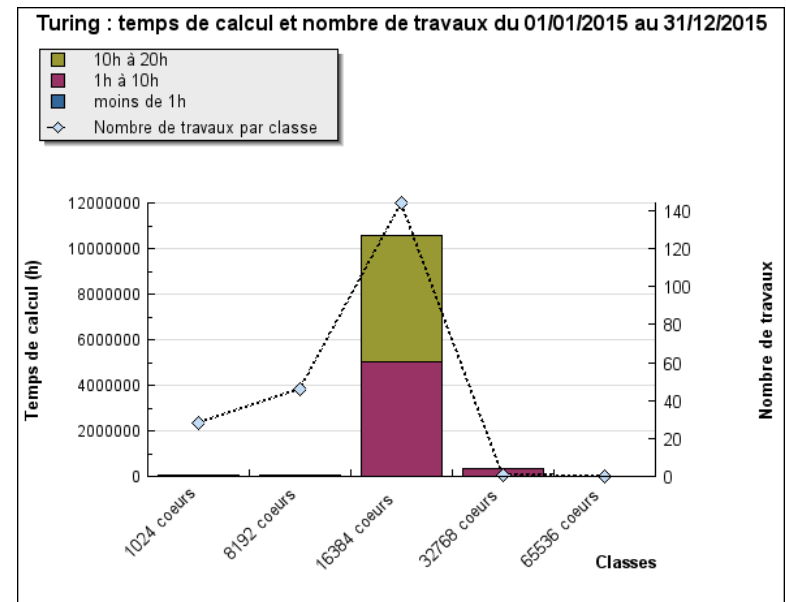
- Comparaison Query contre fichier de Références
  - On divise le fichier Query en autant de parties que de processus
  - Chaque processus récupère  $(nb\_lignes\ totales / nProcs)$  lignes de Query.
  - Après l'initialisation MPI, chaque processus effectue ses calculs sur sa partie de fichier
    - cœur du calcul Disseq inchangé
  - Chaque processus écrit ses résultats dans un fichier
  - Les fichiers sont post-processés ensuite pour n'en former qu'un



# MpiDisseq sur Turing en 2015

- Runs sur 16 384 processus MPI (1/6<sup>ème</sup> de la machine)
  - 11 millions d’heures consommées par environ 200 jobs
  - Temps elapsed total : 671 heures (environ 28 jours)

- En séquentiel sur un processeur :
  - 11 Mh = 1 255 années



# La réussite du passage à l'échelle

- Côté utilisateur
  - Une volonté d'évoluer, de passage à l'échelle
  - Sollicitation de l'IDRIS et de ses ressources humaines
  - Accès préparatoire : [www.edari.fr](http://www.edari.fr)
  - Demande d'heures GENCI : [www.edari.fr](http://www.edari.fr)
- Côté IDRIS
  - Cours MPI, Fortran, C : [cours.idris.fr](http://cours.idris.fr)
  - L'Équipe Support aux Utilisateurs → [assist@idris.fr](mailto:assist@idris.fr)
  - Demande de support avancé :  
[www.idris.fr](http://www.idris.fr) → Espace utilisateurs → Support avancé et IPÉA
  - Des logiciels déjà installés sur les machines (NAMD, GROMACS,...)



# Un premier résultat

