

# Analyses quantitatives de formes appliquées à la paléoanthropologie

Jean Dumoncel

En collaboration avec : *A. Beaudet (GAES), J. Braga (AMIS), E. Courcelle (CALMIP), S. Durrleman (ICM), J.-P. Jessel (IRIT), N. Renon (CALMIP), G. Subsol (LIRMM), C. Zanolli (AMIS)*

Journées SUCCES - 23 Novembre 2016



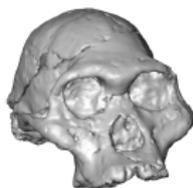
UNIVERSITÉ  
TOULOUSE III  
PAUL SABATIER



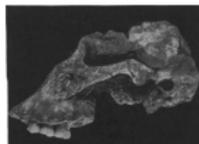
**AMIS** ANTHROPOLOGIE MOLÉCULAIRE  
UMR 5288 du CNRS ET IMAGERIE DE SYNTHÈSE

# LABORATOIRE AMIS

- ▶ Equipe *Phylogénie, sélection et macro-évolution du genre Homo*.
- ▶ Anatomie comparée de vestiges fossiles et de spécimens modernes.

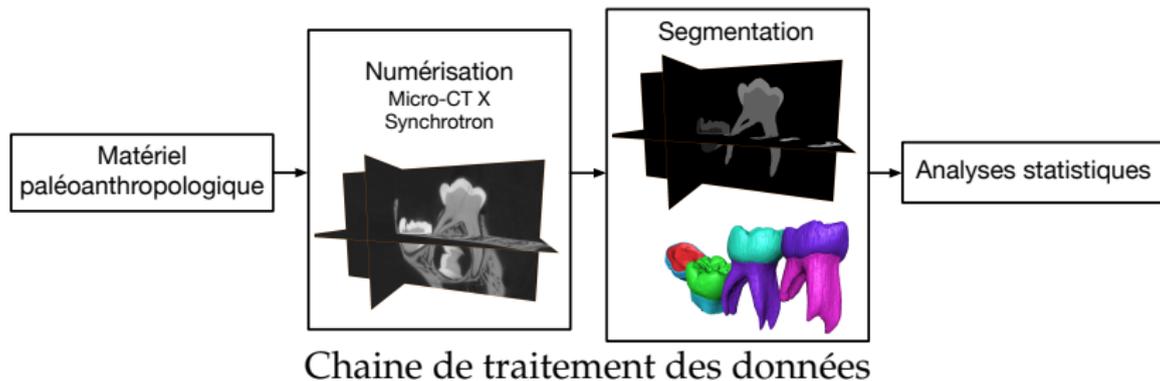


Australopithèques

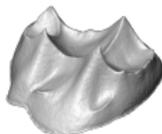


Paranthrope

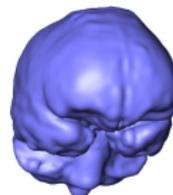
# LABORATOIRE AMIS



Structure dentaire

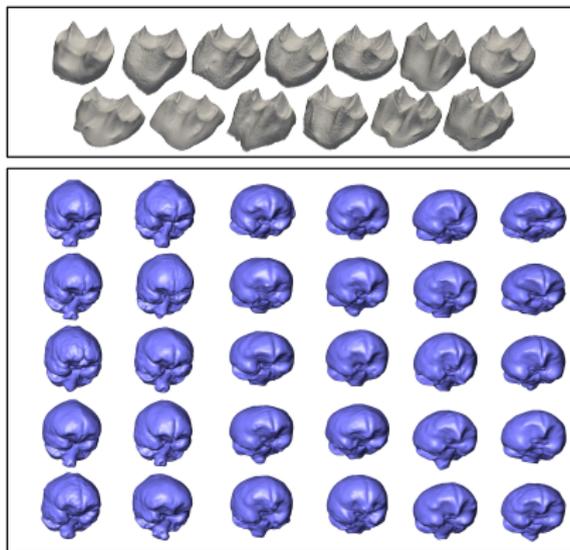


Structure endocrânienne



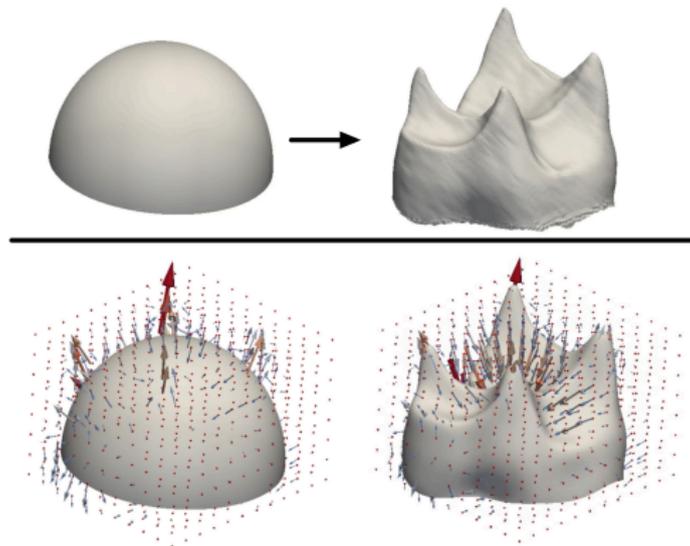
# LABORATOIRE AMIS

- ▶ Etude de la variabilité.
- ▶ Besoin d'une méthode de comparaison de formes.



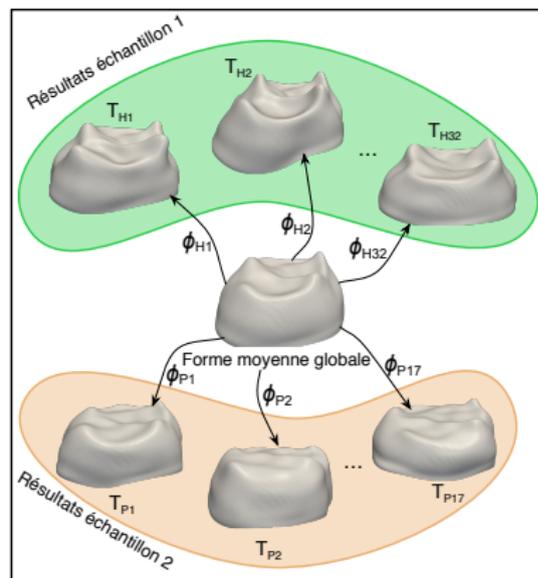
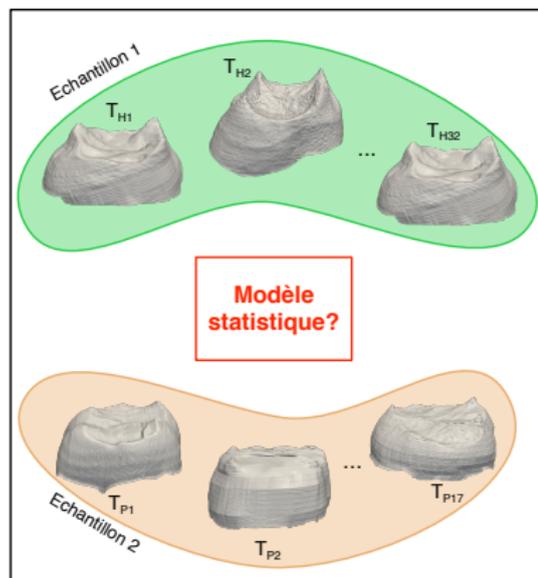
# COMPARAISON DE FORMES

- ▶ Code de calcul : Deformetrica (<http://www.deformetrica.org>).
- ▶ Développé par Stanley Durrleman (Institut du Cerveau et la Moelle épinière).
- ▶ Domaine du recalage de formes.



# COMPARAISON DE FORMES

- Génération des atlas : une forme moyenne et les déformations de la forme moyenne vers chaque spécimen.



# COMPARAISON DE FORMES

- ▶ Un recalage : entre  $T=1$  heure et  $T=72$  heures.
- ▶ Un atlas de  $N$  spécimens :  $N \cdot T$  heures environ.
- ▶ Les temps de calculs dépendent de :
  - ▶ Nombre de spécimens.
  - ▶ Paramètres des noyaux.
  - ▶ Nombre de points d'intégrations des équations différentielles.
  - ▶ Nombre de points des surfaces utilisées.
- ▶ Parallélisation dans Deformetrica :
  - ▶ pour l'atlas : utilisation des routines haut-niveau de multi-threading d'ITK.
  - ▶ pour le recalage : routine de bas niveaux exécutables sur carte graphique.

# INFRASTRUCTURE DE CALMIP

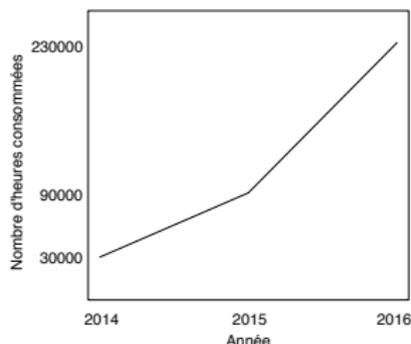
## Mésocentre de CALcul en MIDi-Pyrénées (CALMIP)

- ▶ Supercalculateur EOS (612 processeurs, 12240 coeurs, 39 To de RAM)
- ▶ Système d'ordonnancement des tâches : SLURM.
- ▶ Avantage de l'utilisation de EOS :
  - ▶ Parallélisation.
  - ▶ Calcul déporté.
  - ▶ Grande capacité de mémoire vive.
  - ▶ Outil d'analyse de job.

# PRÉSENTATION DU PROJET

Projet P1440 "Atlas anatomique de primates" :

- ▶ Projet initié en 2014. Utilisation du supercalculateur Hyperion puis de EOS.
- ▶ Utilisateurs : 1 enseignant-chercheur (José Braga), 1 chercheur (Clément Zanolli), 1 chercheur post-doctorant (Amélie Beaudet), 1 ingénieur (Jean Dumoncel).
- ▶ A ce jour : 320 000 heures demandées (environ 36 ans sur un mono-processeur !).



## POURQUOI UTILISER UN SERVEUR DE CALCUL ?

- ▶ Exemple :  $T = 24$  heures et  $N = 50$  : 50 jours de calcul...  
Utilisation de la parallélisation : 1 jour de calcul.
- ▶ Big data à l'échelle du laboratoire AMIS : jusqu'à 150 spécimens.
- ▶ Nécessité de refaire plusieurs fois des calculs similaires :
  - ▶ Problèmes de convergence.
  - ▶ Ajout de spécimens.
- ▶ Besoin de mémoires vives : entre 1 Go et 8 Go pour un calcul,  $N$  fois plus pour  $N$  calculs.
- ▶ Support informatique.

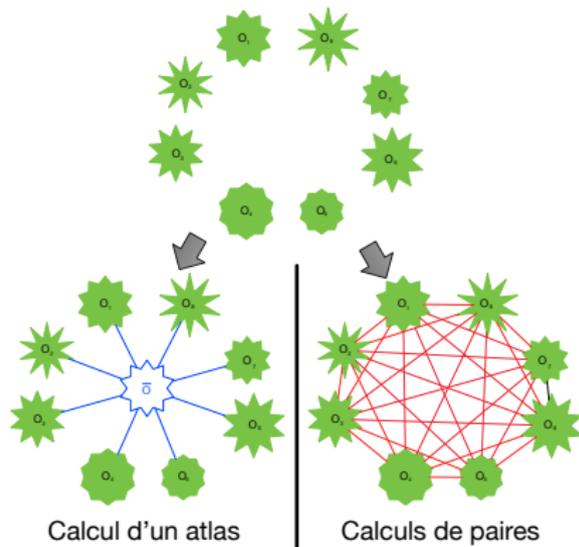
# DIFFICULTÉ RENCONTRÉE

Nombre de coeurs en fonction du nombre de spécimens N :

- ▶ Atlas : N
- ▶ Paires :  $\frac{N(N-1)}{2}$

Problème de gestion de la mémoire :

- ▶ Atlas : partagée
- ▶ Paires : distribuée



# DIFFICULTÉ RENCONTRÉE

## Solutions apportées par CALMIP :

- ▶ Pour le calcul des atlas :
  - ▶ Utilisation du noeud de calcul à grande mémoire : 128 coeurs et 2To de RAM en mémoire partagée.
- ▶ Pour le calcul des paires :
  - ▶ Utilisation d'une partie des 12240 coeurs et la mémoire distribuée.

# DIFFICULTÉ RENCONTRÉE

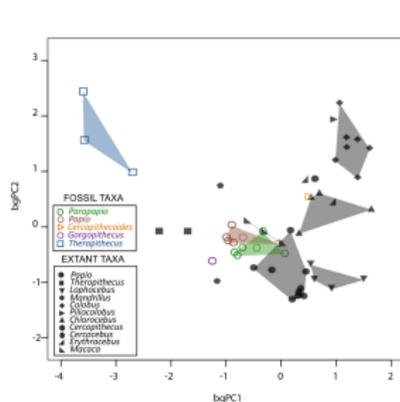
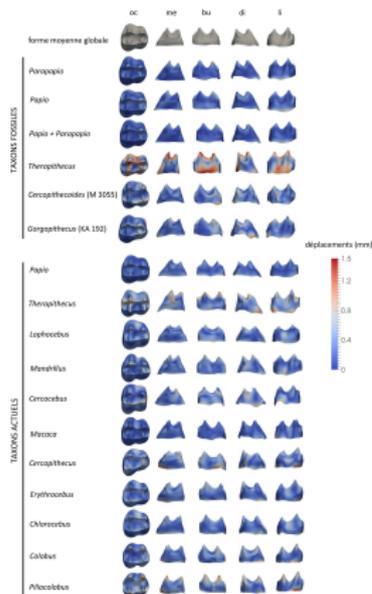
Problèmes engendrés par le calcul par paire :

- ▶ "Embarrassingly parallel problems".
- ▶ Non codé dans Deformetrica.
- ▶ Augmentation en  $N^2$  du nombre de calculs.

## SOLUTION

- ▶ Utilisation de chdb (Gestion automatique des entrées / sorties du programme).
- ▶ Augmentation du nombre d'heures de temps de calcul.

## PUBLICATIONS ASSOCIÉES AU PROJET CALMIP P1440



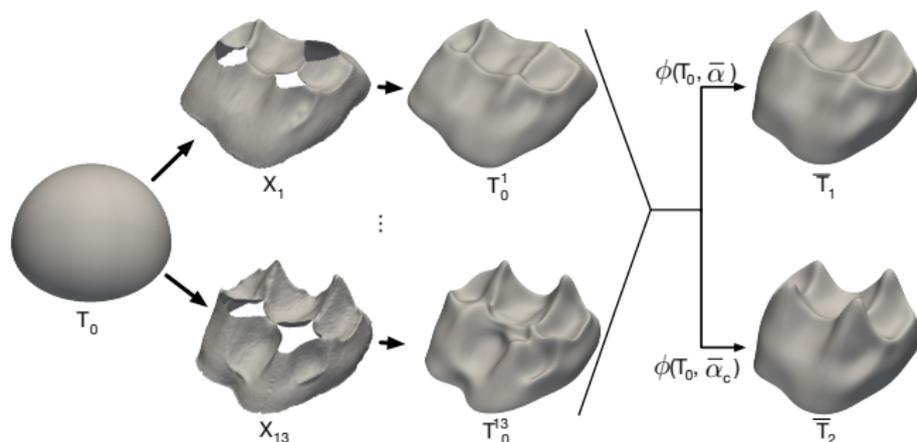
Etude de la variabilité  
avec 79 spécimens

Beaudet et al., 2016, Upper third molar internal structural organization and semicircular canal morphology in Plio-Pleistocene South African cercopithecoids, *Journal of Human Evolution*

## PUBLICATIONS ASSOCIÉES AU PROJET CALMIP P1440

## Articles :

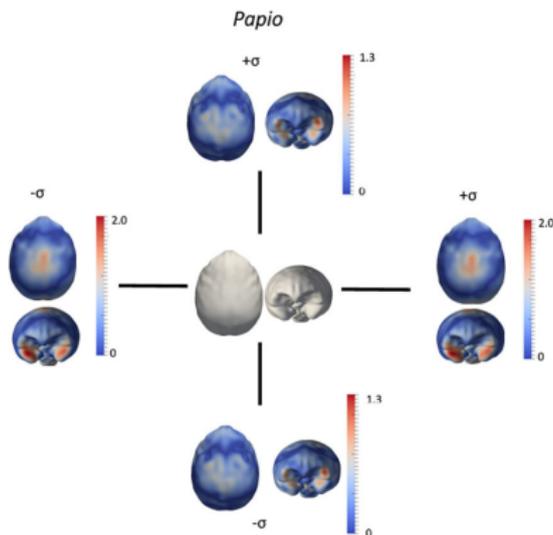
- Dumoncel et al., 2016, "How to Build an Average Model When Samples Are Variably Incomplete? Application to Fossil Data". The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops.



## PUBLICATIONS ASSOCIÉES AU PROJET CALMIP P1440

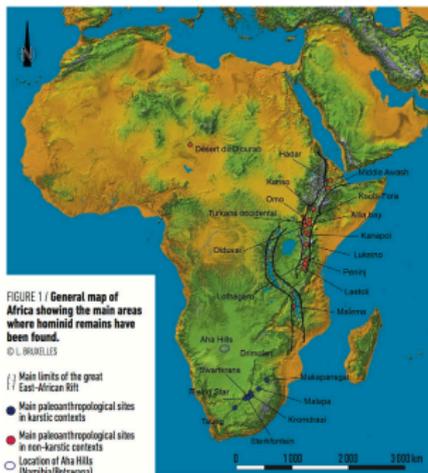
Thèse :

- ▶ Beaudet, 2015, "Caractérisation des structures crânio-dentaires internes des cercopithécoïdes et étude diachronique de leurs variations morphologiques dans la séquence Plio-Pléistocène sud-africaine".



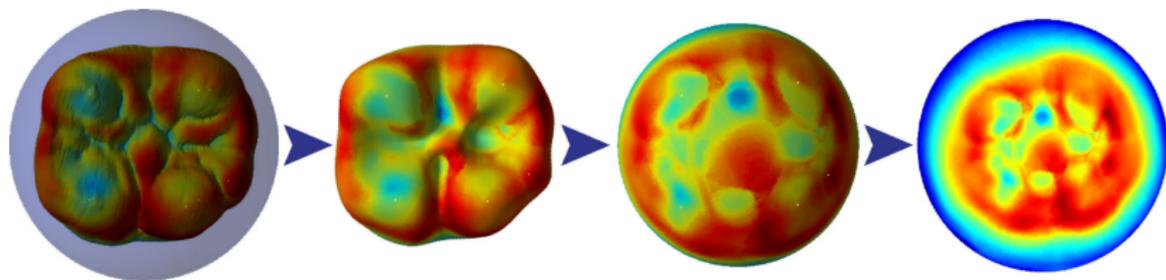
# PERSPECTIVES

Le projet "Kromdraai" (<http://www.kromdraai-origins.org>)



# PERSPECTIVES

## Outil de calcul de l'épaisseur de l'émail



Zannoli et al., 2016, "Innovative approaches to quantify and statistically compare tooth enamel thickness distribution", 6th Annual Meeting of the European Society for the study of Human Evolution (ESHE)

Les perspectives du projet sont :

- ▶ Augmenter le nombre de spécimens.
- ▶ Augmenter la précision des calculs.
- ▶ Tester d'autres possibilités de parallélisation.

# REMERCIEMENTS

Merci aux membres de CALMIP :

- ▶ Boris Dintrans (directeur).
- ▶ Nicolas Renon (responsable calcul scientifique).
- ▶ Emmanuel Courcelle (cellule support utilisateurs).
- ▶ Pierrette Barbaresco (responsable technique).