

Recherche de gènes de novo dans le génome d'eucaryotes unicellulaires : les levures

Contact : ingrid . lafontaine @ upmc . fr

Depuis l'obtention de génomes complètement séquencés, dès 1996, des phases codant pour des protéines, ne présentant aucune similarité avec des gènes déjà connus sont annotés dans les génomes. Ces gènes peuvent avoir des homologues qui n'ont pas été détectés soit parce qu'ils appartiennent à des génomes qui n'ont pas encore été séquencés, soit parce que leurs séquences ont trop divergé pour que la similarité soit encore détectable. Il est aussi possible qu'ils correspondent à des erreurs d'annotation, *i.e.* que ces phases ouvertes de lectures ne correspondent pas à des gènes codant pour des protéines.

Enfin, si toutes ces hypothèses sont écartées, certains de ces gènes pourraient correspondre à des gènes nouvellement créés à partir de séquences non-codantes (*de novo*). Depuis une dizaine d'années, plusieurs cas ont ainsi été identifiés chez les eucaryotes, de la levure à l'homme.

Vous travaillerez sur les génomes de levures disponibles à cette adresse :

<http://www.saccharomycessensustricto.org/cgi-bin/s3.cgi?data=Annotations&version=current>

- A. Dresser la liste des ORFans présents dans les génomes des 6 levures *Saccharomyces* : *Saccharomyces cerevisiae*, *S. mikatae* (IFO 1815^T), *S. bayanus* var. *uvarum* (CBS 7001), *S. kudriavzevii* (ZP 591) et *S. paradoxus*.
Pour sélectionner les ORFans candidats, établir des critères de « non-similarité ». Les outils de recherche de similarité comme BLAST, PSI-BLAST, RPS-BLAST, CD-Search, PFAM, etc. pourront être utilisés.
- B. Essayer de caractériser les ORFans, à l'aide de méthodes ne reposant pas sur la similarité des séquences. Par exemple, vous pouvez analyser la composition nucléotidique, des biais d'usage, utiliser des outils de prédiction de structure 2D/3D, etc.

Références bibliographiques (contactez moi si vous en voulez plus):

Vakirlis N, Hebert AS, Opuente DA, Achaz G, Hittinger CT, Fischer G, Coon JJ, Lafontaine I. 2018. A Molecular Portrait of De Novo Genes in Yeasts. *Mol Biol Evol* **35**: 631–645.

McLysaght A, Hurst LD. 2016. Open questions in the study of de novo genes: what, how and why. *Nat Rev Genet* **17**: 567–578.

Bornberg-Bauer E, Schmitz J, Heberlein M. 2015. Emergence of de novo proteins from “dark genomic matter” by “grow slow and moult.” *Biochemical Society Transactions* **43**: 867–873.

Schlötterer C. 2015. Genes from scratch – the evolutionary fate of de novo genes. *Trends in Genetics* **31**: 215–219.

Light S, Basile W, Elofsson A. 2014. Orphans and new gene origination, a structural and evolutionary perspective. *Current Opinion in Structural Biology* **26**: 73–83.