

Compte Rendu Visio Conférence *PepiAnnot*

13 sept 2019 -10h00 / 11h30

<https://rendez-vous.renater.fr>

Membres (28)	Unité
AMSELEM Joëlle	URGI, INRA Versailles
BOISARD Julie	
BOUDET Nathalie	IPS2, MdC UEVE, Gif sur Yvette
BRIONNE Aurélien	INRA, Tours
BRUNAUD Véronique	IPS2, Gif sur Yvette
CANAGUIER Aurélie	INRA, EPGV, Evry
CHARLES Mathieu	
CHOULET Frédéric	INRA GDEC-UCA, Clermont-Ferrand
CORRE Erwan	CNRS Roscoff
DA-ROCHA Martine	INRA, Sophia Agrobiotech, Antibes
DERRIEN Thomas	IGDR - CNRS - UMR6290, Rennes
DEVILLIERS Hugo	INRA, Micalis, Jouy-en-Josas
DIOT Thomas	
FAIVRE RAMPANT Patricia	
JOETS Johann	INRA, Fermes du Moulon, Orsay
KORNOBIS Etienne	
KREPLAK Jonathan	INRA, Dijon
LE DANTEC Loïc	INRA Bordeaux
LEGEAI Fabrice	INRA, BIPAA, Rennes
LEROY Philippe	INRA GDEC-UCA, Clermont-Ferrand
MOLLION Maeva	INRA, Fermes du Moulon, Orsay
NEUVEGLISE Cécile	INRA, Micalis, Jouy-en-Josas
ORJUELA Julie	
RIMBERT Hélène	INRA GDEC-UCA, Clermont-Ferrand
ROGIER Odile	INRA Orléans
SIMON Adeline	INRA, Versailles
TOFFANO-NIOCHE Claire	
VELT Amandine	INRA Colmar

Si des personnes manquent dans la liste ne pas hésiter à contacter Véronique ou Philippe pour une mise à jour. **Si besoin compléter et/ou corriger le tableau ci-dessus. Merci par avance.**

Lors de la visio PepiAnnot du 13 Septembre :

- Il y avait environ 15 personnes connectées
- Thomas Derrien, Fabrice Legeai, Johann Joets, et Maeva Mollion ont été rajoutés au Groupe PepiAnnot (mailing list)
- Visio sous rendez-vous a très bien marché cette fois-ci mais regarder du côté de GoToMeeting utilisé par IFB (info de Joëlle), attention licence payante ...

Ordre du jour :

1. Prochain thèmes abordés 2
2. Thomas Derrien – FEELnc – outil d’annotation des Long non-codingRNA (lncRNA) 2
3. Groupe PepiAnnot en présentiel 3
4. Photo du Jour 3

1. Prochain thèmes abordés

- Nouvel outil sur les détections de variants par Johann Joets
- Un exemple d'outils utilisés pour de la comparaison de transcriptomes (de novo ou pas) par Julien
- LongReads : outils de correction d'erreurs et mapping + splicing alternatif par Philippe & Véro

➔ Prochaine visio sera faite par Johann Joets sur un nouvel outil de détection de variants (GROM), le 27 Novembre de 10h00 à 11h30

2. Thomas Derrien – FEELnc – outil d'annotation des Long non-codingRNA (lncRNA)

Intro générale sur les lncRNA :

- Focus sur les régions codantes/transcrites donc tout part de l'annotation des séquences : régions UTR, ORF, exons, splices sites....
- Caractéristiques des lncRNA font qu'ils ne peuvent pas être annotés de la même façon car peu conservés (plus de conservation via la syntenie que via de la comparaison de séquence) et pas de potentiel codant. Par contre peuvent avoir des introns et une majorité contiennent des queues polyA. Les lncRNA sont peu exprimés et tissu spécifique.
- 2 catégories : lincRNA long-intergenic ncRNA et antisens lncRNA, quelques lncRNA encodent des petits peptides (5-10%)
- Rq : chez l'homme on considère que 60% du génome est transcrit mais seulement 2% en prot (via ENCODE consortium 2015); Chez l'homme on a environ 20 000 loci codants et 16 000 loci lncRNA et par contre beaucoup moins de transcrits alternatifs par locus pour les lncRNA (2 en moyenne)
- Plusieurs actions : activer ou réprimer la transcription d'un mRNA, agir en cis ou trans, éponges à miRNAs, modification de la chromatine...
- Plant Databanks : EnsemblPlant ; Cantata <http://cantata.amu.edu.pl> : *CANTATAdb 2.0 is a database of lncRNAs identified computationally in 36 plant species and 3 algae.*

FEELnc description :

Cet outil vient à la suite de cufflink ou stringTie qui génèrent un transcriptome reconstruit.

Il est organisé en 3 modules :

- 1- Module Filter : pour éliminer les non lncRNA : protein coding, transcrits trop petits (< 200nt), mono-exonic sont moins souvent des lncRNA, on peut garder que les gènes intergéniques (option --linconly) mais ces paramètres peuvent être changés
- 2- Module Coding Potential : utiliser le potentiel codant différent entre gènes codants ou et non codants en calculant les différences entre kmers entre les gènes codants et non codants et la couverture des ORF → utilise du machine learning via des arbres de décision pour définir un score cutoff (par ex < 0.25 pour lncRNA et > 0.59 pour codant pour des prot). Ce module utilise des données mRNA pour la partie apprentissage. Mais il peut simuler sont propre modèle d'apprentissage (mode shuffle)
- 3- Module Classifier : classer les lncRNA en fonction de la localisation par rapport à des gènes codants : intergenic, antisens, overlap etc. et dans intergenic : overlap, convergent et divergents. Attention sans génome de référence pas de classification.

Comparaison avec 5 autres outils :

- FEELnc est le plus sensible mais pas le plus spécifique
- Tout ces outils utilisent du machine learning mais certains ont un jeu d'apprentissage pré-défini pour certaines espèces et ne peuvent donc pas s'adapter à toutes les espèces
- FEELnc, est plus flexible car utilise le jeu d'apprentissage pour s'adapter à chaque espèce. Et quand on n'a pas de lncRNA on peut faire une simulation de lncRNA (via les gènes codants présents) et possibilité d'utiliser des lncRNA d'espèces voisines. Dans la publication, il y a les résultats suivant la proximité des espèces utilisées.
- Utiliser chez le poulet, algues, chien ...environ 80 citations de FEELnc

Dans le groupe PepiAnnot, Erwan a utilisé FEELnc pour des génomes Algues et Jonathan pour le pois. Les 2 sont contents et trouvent performant cet outil. Philippe doit prochaine utiliser FEELnc sur le blé tendre dans le cadre du projet wheatOMICS.

3. Groupe PepiAnnot en présentiel

Pas en 2019 mais plutôt automne 2020

4. Photo du Jour



Thomas DERRIEN