# Compte Rendu Visio Conférence *PepiAnnot*

## 27 Novembre 2020 -10h00 / 11h30

https://pepi-ibis.inrae.fr/annotation-genomes

| Membres (28) | Unité | Mail |
|---|---|---|
| **AMSELEM Joëlle** | URGI, INRAE, Versailles | joelle.amselem@inrae.fr |
| **BOISARD Julie** | | julie.boisard@edu.mnhn.fr |
| **BOUDET Nathalie** | IPS2, MdC UEVE, Gif sur Yvette | nathalie.boudet@inrae.fr |
| **BRIONNE Aurélien** | INRAE, Tours | aurelien.brionne@inrae.fr |
| **BRUNAUD Véronique** | IPS2, Gif sur Yvette | veronique.brunaud@u-psud.fr |
| **CANAGUIER Aurélie** | INRAE, EPGV, Evry | aurelie.canaguier@inrae.fr |
| **CHARLES Mathieu** | Jouy-en-Josas ? | Mathieu.Charles@inrae.fr |
| **CHOULET Frédéric** | INRAE, GDEC-UCA, Clermont-Ferrand | frederic.choulet@inrae.fr |
| **CORRE Erwan** | CNRS Roscoff | corre@sb-roscoff.fr |
| **DA-ROCHA Martine** | INRAE, Sophia Agrobiotech, Antibes | martine.da-rocha@inrae.fr |
| **DERRIEN Thomas** | IGDR - CNRS - UMR6290, Rennes | thomas.derrien@univ-rennes1.fr |
| **DEVILLIERS Hugo** | INRAE, Micalis, Jouy-en-Josas | hugo.devillers@inrae.fr |
| **DIOT Thomas** | | thomas69.diot@laposte.net |
| **FAIVRE RAMPANT Patricia** | INRAE, | patricia.faivre-rampant@inrae.fr |
| **HILLIOU Frédérique** | INRAE, | frederique.hilliou@inrae.fr |
| **HUNEAU Cécile** | INRAE, GDEC-UCA, Clermont-Ferrand | cecile.huneau@inrae.fr |
| **JOETS Johann** | INRAE, Fermes du Moulon, Orsay | johanne.joets@inrae.fr |
| **KORNOBIS Etienne** | Pasteur, | etienne.kornobis@pasteur.fr |
| **KREPLAK Jonathan** | INRAE, Dijon | jonathan.kreplak@inrae.fr |
| **LASSERRE-ZUBER Pauline** | INRAE, GDEC-UCA, Clermont-Ferrand | pauline.lasserre-zuber@inrae.fr |
| **LE DANTEC Loïc** | INRA Bordeaux | loick.le-dantec@inrae.fr |
| **LEGEAI Fabrice** | INRA, BIPAA, Rennes | Fabrice.Legeai@inrae.fr |
| **LEROY Philippe** | INRA GDEC-UCA, Clermont-Ferrand | philippe.leroy.2@inrae.fr |
| **MARDOC Emile** | INRAE, GDEC-UCA, Clermont-Ferrand | emile.mardoc@inrae.fr |
| **MOLLION Maeva** | INRA, Fermes du Moulon, Orsay | Maeva.Mollion@inrae.fr |
| **MONAT Cécile** | INRAE, GDEC-UCA, Clermont-Ferrand | cecile.monat@inrae.fr |
| **NEUVEGLISE Cécile** | INRAE, Micalis, Jouy-en-Josas | cecile.neuveglise@inrae.fr |
| **ORJUELA Julie** | IRD, Montpellier | julie.orjuela@ird.fr |
| **PALLIER Vincent** | INRAE GDEC-UCA, Clermont-Ferrand | vincent.pailler@inrae.fr |
| **RIMBERT Hélène** | INRAE GDEC-UCA, Clermont-Ferrand | helene.rimbert@inrae.fr |
| **ROGIER Odile** | INRA Orléans | odile.rogier@inrae.fr |
| **SIMON Adeline** | INRAE, Versailles | adeline.simon@inrae.fr |
| **TOFFANO-NIOCHE Claire** | I2BC, CNRS, Gif-sur-Yvette | claire.toffano-nioche@u-psud.fr |
| **VELT Amandine** | INRAE, Colmar | amandine.velt@inrae.fr |

*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 **27 Novembre 2020** :

- Il y avait 13 personnes connectées
- Première visioconférence PepiAnnot avec GoToMeeting (licence PEPI IBIS) .

## Ordre du jour :

## 1. Prochains thèmes possibles

- **Kmer pour l'annotation** - *CHÂTEAU Annie (à contacter)*
- **R shiny & R markdown** - outils puissants pour l'analyse et la traçabilité en bioinformatique
- **SibeliaZ** - *JOETS Johann*
- Annotation des **TEs** - *CHOULET Frédéric*
- Assemblage **Transcriptome *de novo* short et Long reads** - *BRUNAUD Véronique*
- **DGenies** - DotPlot de chromosomes entiers - *KLOPP Christophe (à contacter)*
- Annotation fonctionnelle - Mercator4/MapMan - *DELANNOY Etienne (à contacter)*
- Intégration des données omiques - *ALAUX Michael (à contacter)*
- Réflexions autour de l'intégration statistique des données omiques - *MARTIN MAGNIETTE Marie-Laure (à contacter)*
- Pan Génomique
- Variants structuraux
- Réseaux de gènes
- Exposé didactique de la théorie des graphes
- Les infrastructures de calcul et de stockage - bonnes pratiques
- Plan de gestion des data avant tout projet !
- États de l'art sur tous les éléments constitutifs connus à ce jour (features) d'un génome
- Apport du « Deep Learning » sur l'analyse des données omiques

## 2. Prochaine réunion PepiAnnot

- **19 Février 2021 - 10h00 - 12h00**
- **Johann Joets** présentera **Liftoff**
- **Hélène Rimbert** présentera **Magatt**
    - Deux pipelines pour le transfert d'annotation d'une séquence de référence sur un nouveau génome assemblé

## 3. Claire Toffano-Nioche - I2BC, CNRS, Gif-sur-Yvette – *The RFAM DataBase*

**Objectifs :**

- Présentation de la base de donnée RFAM - synchronisation RFAM/miRBASE
- Comment créer une nouvelle famille de miRNA dans RFAM avec RFAM cloud

**La base de données RFAM**

- https://rfam.org/
- RFAM : rassemble les données sur tous les RNA, maintenue via EBI (financement Elixir), curation faite par la communauté d'expert des familles
- RFAM travaille sur la caractérisation des différents RNA et leur organisation en familles
- Cette caractérisation dépend de la séquence et de la structure 2D, donc les homologies sont basées sur les deux, la séquence + l'analyse 2D. Cela donne un poids important à la structure 2D qui caractérise les ARN
- Recherche sur RFAM : soit par mot clef, génome, séquence…Il y a une page wiki qui décrit les familles présentes avec une structure 2D, et toutes les caractéristiques (identification phylogénétique).

- Présentation de la publication NAR-database : Actuellement, **RFAM version 14.3** (mise à jour Septembre 2020), 3 444 familles avec 28% de familles en plus et cela grâce à l'inclusion de données venant de 3 bases principalement :
    - **ZWCD (Zasha Weinberg Database)** : comparaison coté séquences + 2D des régions inter géniques et ont trouvé une série des nouveaux RFAM motifs non connus jusqu'ici (un peu comme les PFAM DUF sauf 2D en plus)
    - **EVCB** : virus de l'institut européen. Les virus sont riches en ARNs structurés + exemple du coronavirus
    - **miRBase** : classification basée sur du blast. miRBase compte environ 2000 familles alors que RFAM n'en comptait que 500 … Pour la v14, RFAM a expertisé les alignements de miRBase pour faire son propre classement (ajout de 800 familles RFAM) et renvoie vers miRBase les nouvelles séquences détectées par les analyses RFAM. miRBase décide de les inclure ou non. Hormis 300 alignements miRBase (à analyser + en détail pour devenir des familles RFAM), les 2 bases de données sont maintenant synchronisées par RNACentral (https://rnacentral.org/).

**Comment créer une nouvelle famille**

- Créer des nouvelles familles demande beaucoup de ressources (RFAM cloud) avec la recherche de structure 2D ➔ pipeline.
    - On part d'une séquence ou de séquences alignées validées (alignement conseillé par RFAM) au format Stockholm (format qui décrit un alignement multiple)
    - Ensuite plusieurs outils tournent : RNAfold, INFERNAL, outils RFAM (rfsearch, rfmake, rfseed…),
    - Puis check la qualité
- Claire nous donne un exemple de l'utilisation du pipeline :
    1. Partir d'un fichier fasta au format Stockholm qui décrit la séquence 2D
    2. Rechercher des séquences similaires dans rfamseq, définir des « seed » d'alignement et calcul significatif via covariance model.
    3. Ce « seed » alignement peut être vérifié à la main et permet d'obtenir la représentation taxonomique, et la structure 2D qui constituent les informations pour contrôler la qualité de la famille.

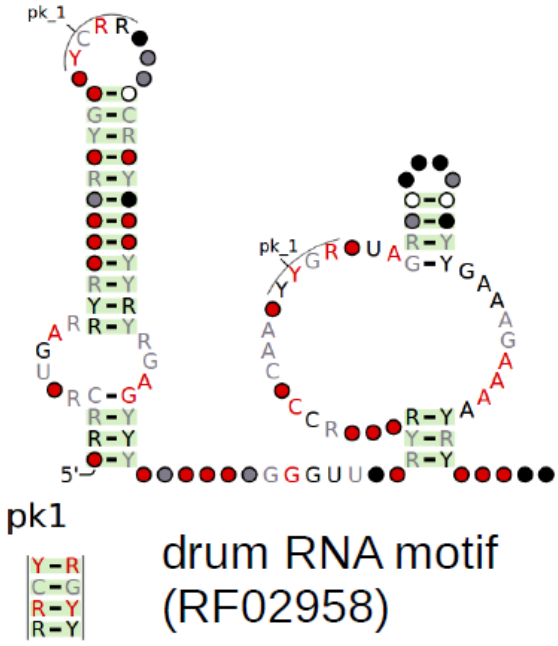
⇨ Application chez les bactéries : à partir de RNAseq de bactéries, sélection de nouveaux candidats ARN conservés à étudier avec le pipeline Rfam. Pour confirmer des petits ARNs bactériens, il est possible de caractériser des Riboswitch (séquence DNA/RNA en 5'UTR des ARNm pour la régulation de la traduction), des small RNAs, de rechercher des structures tige-boucle « terminator » pour fixation de Rho (*Rho-independent transcription terminator at their 3' end*), ...

**Quelques remarques / conclusions**

- Utilisation de RFAM cloud pipeline assez simple, a été fait par des étudiants sur l'application chez les bactéries ➔ mise en évidence de 6 nouvelles familles de smallRNA
- On peut aussi annoter les familles ou faire des Mise à Jour (curation manuelle experte / nouvelles familles)
- Pour la prochaine version de RFAM les perspectives sont de :
    - Compléter EVCB & miRBase
    - Intégrer des données expérimentales de conformation 3D sur les « seed »
    - Connecter la description des familles avec la littérature scientifique (*text mining*)

- Remarque : trouver des familles miRNA c'est aussi trouver les cibles mRNA et les caractères correspondant concernés et régulés

## 4. Photo du Jour




*Claire Toffano-Nioche*