

# Cycle « Initiation à la bioinformatique »

## **Module 3/4 : Prédiction de gènes et annotation de protéines**

Le cycle de formation « Initiation à la bioinformatique » est constitué de 4 modules. Ils peuvent être suivis indépendamment mais ont une cohérence et suivre chaque module peut aider à une meilleure compréhension des modules suivants. Ces modules sont :

- 1) Banques de données et Blast
- 2) Alignements de séquences
- 3) **Prédiction de gènes et annotation de protéines**
- 4) Initiation à la reconstruction phylogénétique en biologie moléculaire

Un autre cycle intitulé « Analyse NGS (Next Generation Sequencing) » est en cours de préparation pour l'année 2018. Les inscriptions pour ce deuxième cycle ouvriront au cours du dernier trimestre 2017.

Cette formation est proposée par la plate-forme de bioinformatique et biostatistique de Lille, *bilille* : <https://wikis.univ-lille1.fr/bilille/>

<b>OBJECTIFS</b>	<ul style="list-style-type: none"> <li>- découvrir les logiciels liés à la prédiction de gènes et à l'annotation de protéines</li> <li>- acquérir la méthodologie pour prédire les gènes présents sur un génome qu'il soit bactérien ou eucaryote</li> <li>- acquérir la méthodologie pour analyser la séquence protéique prédite et en déduire la fonction possible de la protéine, avoir une idée de sa localisation cellulaire</li> <li>- être capable d'analyser les résultats obtenus par les logiciels avec un regard critique</li> </ul>
<b>PUBLIC CONCERNÉ</b>	Biologistes : enseignants-chercheurs, chercheurs, ingénieurs, techniciens, doctorants contractuels, post-doctorants
<b>PRÉ-REQUIS</b>	<p>Savoir utiliser un ordinateur (Windows...) : naviguer sur internet (Internet Explorer ou Firefox), utiliser un traitement de texte (Word ou OpenOffice).</p> <p>Être familier avec les banques de données, Blast, formats de séquences et alignements de séquences (idéalement avoir suivi le module 1/4 « Banques de données et Blast » et le module 2/4 « Alignements de séquences » du cycle d'initiation à la bioinformatique).</p>
<b>CONTENU</b>	<p><b>Annotation structurale de séquences génomiques :</b></p> <ul style="list-style-type: none"> <li>- prédiction des séquences codantes (CDS) présentes sur une séquence nucléique en comparant les résultats de différents logiciels (exemple : ORF Finder, Blast, GeneMark, GeneScan)</li> <li>- analyse de la structure fine des CDS dont la position des introns/exons dans le cas des séquences eucaryotes (exemple : Wise2, Est2Genome)</li> </ul>

# Cycle « Initiation à la bioinformatique »

## **Module 3/4 : Prédiction de gènes et annotation de protéines**

	<p>- traduction des séquences prédites</p> <p><b>Annotations structurale et fonctionnelle de protéines :</b></p> <ul style="list-style-type: none"> <li>- recherche des domaines protéiques présents sur la protéine (InterProScan)</li> <li>- recherche des peptides signaux et étude de la localisation cellulaire (exemple : TopPred, TM-HMM, SignalP, TargetP, psort.org)</li> <li>- comparaison des protéines aux séquences connues dans les banques de données (BlastP)</li> <li>- Classification fonctionnelle des protéines (exemple GO, Kegg...)</li> </ul> <p><b>Visualisation des annotations :</b></p> <ul style="list-style-type: none"> <li>- présentation des formats GFF et BED</li> <li>- représentation graphique des annotations sur le génome/la protéine (exemple : GenomeView)</li> </ul>
<b>MÉTHODE</b>	Quelques brèves notions théoriques essentielles seront abordées au début de chaque partie de cette formation. Néanmoins, l'essentiel de l'atelier consiste en de la pratique sur ordinateur à l'aide de logiciels libres, utilisés via Internet
<b>DATES</b>	Jeudi 11 et Vendredi 12 Janvier 2018
<b>INTERVENANT</b>	<p style="text-align: center;">Sylvain Legrand, MCF Université Lille 1, UMR 8198, Evo-Eco-Paleo</p> <p style="text-align: center;">Cette formation est assurée par la plate-forme de bioinformatique et biostatistique de Lille, <i>bilille</i></p>
<b>EFFECTIF</b>	Un groupe de 12 personnes
<b>LIEU</b>	Cité Scientifique Villeneuve d'Ascq
<b>INSCRIPTIONS</b>	Avant le 11 juillet 2017

# Cycle « Initiation à la bioinformatique »

## **Module 3/4 : Prédiction de gènes et annotation de protéines**

	<p>Tél.: 03 20 33 70 71. Fax : 03 20 33 59 72.          Courriel : <a href="mailto:Anne-Sophie.Grare@univ-lille1.fr">Anne-Sophie.Grare@univ-lille1.fr</a></p>
	<p><b>Dorothee TERRYN</b>          Inserm ADR Lille, 6 rue du Pr. Laguesse, CS 50027, 59045 LILLE Cedex          Tél.: 03 20 29 86 78. Fax : 03 20 49 01 38. Courriel : <a href="mailto:dorothee.terryn@inserm.fr">dorothee.terryn@inserm.fr</a></p>
	<p><b>Pierre SILVEIRA</b>          CNRS - Délégation Nord - Pas de Calais et Picardie, 2 rue des Canoniers, 59046 LILLE Cedex          Tél.: 03 20 12 36 88. Fax : 03 20 12 36 90. Courriel : <a href="mailto:pierre.silveira@dr18.cnrs.fr">pierre.silveira@dr18.cnrs.fr</a></p>
	<p><b>Nadia BENYAHIA</b>          Université Lille 2 - Université du Droit et de la Santé, 42 rue Paul Duez, 59800 LILLE          Tél.: 03 20 96 43 37. Fax : 03 20 88 24 32. Courriel : <a href="mailto:nadia.benyahia@univ-lille2.fr">nadia.benyahia@univ-lille2.fr</a></p>