



PRECURSOR

CONSORTIUM

Puis PROJET

2024-2028

Coordination

Marie-Laure Martin

UMR IPS2

marie-laure.martin@inrae.fr

Sylvie Coursol

UMR IJPB

sylvie.coursol@inrae.fr

Mots clés

Plantes

Éléments *cis*-régulateurs

Éléments transposables

Base de connaissances

Intelligence artificielle

Unités INRAE impliquées

IPS2

IJPB

URGI

MIA Paris-Saclay

Partenaires

CIRAD (AGAP)

IRD

Université Clermont Auvergne

Mieux comprendre les régions proximales des gènes pour améliorer les modèles de sélection

La transcription des gènes est un processus essentiel dans la réponse adaptative des plantes aux contraintes environnementales. Le consortium scientifique interdisciplinaire PRECURSOR a pour objectif d'explorer et de mieux comprendre ce processus dans les régions proximales des gènes pour, à terme, améliorer le pouvoir prédictif des modèles de sélection.

Contexte et enjeux

La transcription, première étape de l'expression des gènes et de la synthèse des protéines, est finement régulée par de nombreux acteurs moléculaires. Ainsi, les éléments *cis*-régulateurs, courtes séquences d'ADN, régulent l'expression des gènes via des facteurs *trans* qui se fixent sur ces éléments *cis*.

Modifier l'expression de gènes au travers de leurs régulateurs

Les séquences *cis*-régulatrices sont présentes en forte densité dans les régions proximales des gènes, mais leur caractérisation, préalable indispensable à leur exploitation, demeure lacunaire. Des projets récents ont permis d'établir une carte des séquences d'ADN préférentiellement localisées (PLM) dans ces régions (chez *Arabidopsis thaliana* et le maïs), dont près de 80 % ne sont pas encore répertoriées dans les bases de données. Par ailleurs, de nombreuses études montrent que les éléments transposables (ET) peuvent inclure des séquences *cis*-régulatrices. Lorsqu'ils sont insérés à proximité d'un gène, les ET peuvent alors affecter la transcription des gènes adjacents en recrutant des facteurs *trans* supplémentaires.

Ces deux sources de données (PLM et ET) sont prometteuses, car elles permettent une caractérisation à large échelle d'éléments *cis*-régulateurs potentiels. Néanmoins, pour atteindre une véritable compréhension des régions proximales, il est nécessaire de coupler ces informations structurales aux données d'expression. Pour cela, des approches originales utilisant l'intelligence artificielle pourraient être utilisées pour intégrer ces données biologiques dans l'objectif de prédire des gènes clés et leurs réseaux de régulation.

Cependant, ces collectifs scientifiques aux compétences différentes et complémentaires, se rencontrent peu. Le consortium PRECURSOR propose donc de relever ce challenge en rassemblant des compétences interdisciplinaires autour de cette question



© Freepik

Métaprogramme
DIGIT-BIO



Contact: digitbio@inrae.fr

<https://digitbio.hub.inrae.fr/>

Objectifs

PRECURSOR a pour objectif de fédérer des équipes scientifiques à l'interface entre biologie (molécularistes, généticiens, physiologistes) et sciences formelles (statistiques, informatique, bioinformatique), travaillant sur différentes espèces (maïs, blé, sorgho), pour acquérir une vision consolidée des bases génétiques des caractères d'intérêt agronomique, des données structurales aux données d'expression.

L'ambition est de travailler ensemble afin de progresser sur la cartographie et le pouvoir prédictif des éléments *cis*-régulateurs dans les régions proximales des gènes, en considérant la question dans toute sa complexité et prenant en compte les complémentarités / différences entre les espèces envisagées.

PRECURSOR permettra *in fine* de construire un consortium scientifique interdisciplinaire sur l'intégration inédite de données hétérogènes pour mieux comprendre les régions proximales des gènes et, à terme, valoriser de nouveaux allèles d'intérêt agronomique et améliorer le pouvoir prédictif des modèles de sélection.

Partenaires

Départements INRAE	Unités INRAE	Expertises
BAP	<u>IPS2</u>	Bioinformatique des éléments <i>cis</i> -régulateurs, statistiques des données omiques
	<u>IJPB</u>	Biologie des éléments <i>cis</i> -régulateurs, maïs, contraintes environnementales, digestibilité, génomique fonctionnelle
	<u>URGI</u>	Informatique, bases de connaissances, éléments transposables
MathNum	<u>MIA Paris-Saclay</u>	Méthodes d'intelligence artificielle
Partenaires	Expertises	
CIRAD	<u>AGAP (dans la suite : projet exploratoire PRECURSOR 2)</u>	Génétique quantitative, sorgho, génomique fonctionnelle
IRD	<u>DIADÉ</u>	Biologie, céréales tropicales, systèmes racinaires
Université Clermont Auvergne	<u>GDEC</u>	Physiologie moléculaire de la réponse aux stress biotiques et abiotiques, blé, pathogènes fongiques, stress hydrique

Résultats

Structuration d'une communauté interdisciplinaire

Le consortium, constitué en 2024, a réuni 12 membres aux profils complémentaires (biologistes, généticiens, bioinformaticiens, statisticiens et informaticiens), et répartis sur trois sites (Ile-de-France, Clermon-Ferrand et Montpellier). Il a permis de mettre en place un fonctionnement collectif efficace et l'usage d'outils partagés. Cette dynamique a favorisé l'émergence d'un langage commun et une meilleure compréhension mutuelle des contraintes et apports disciplinaires.

Métaprogramme
DIGIT-BIO



Contact: digitbio@inrae.fr

<https://digitbio.hub.inrae.fr/>

Exploration et analyse critique des approches existantes de détection, de caractérisation et de prédiction des CREs

Ces travaux ont permis d'évaluer la pertinence de différentes approches méthodologiques, identifier les besoins en données et définir les conditions nécessaires à une modélisation biologiquement informée. Des premiers tests, notamment autour de la cartographie des CREs chez plusieurs espèces végétales, ont servi de base à ces réflexions.

Cette analyse bibliographique a permis de mettre en évidence les limites des approches actuelles, notamment leur faible interprétabilité biologique, leur transférabilité limitée entre espèces et leur difficulté à prédire des phénotypes complexes. Cette réflexion a contribué à clarifier les verrous scientifiques à lever pour une application efficace en amélioration des plantes. Ce travail a permis la rédaction d'un article d'opinion, actuellement en révision, qui propose un changement de perspective dans l'étude des CREs, en faveur d'approches intégrant explicitement les mécanismes de la régulation transcriptionnelle, leur organisation et leur contexte évolutif, afin de dépasser une logique de prédiction purement statistique et de renforcer l'interprétabilité des modèles.

Perspectives : lancement d'un projet exploratoire Precursor 2 (2026-2028)

PRECURSOR a servi de levier pour le développement de nouvelles perspectives notamment avec le projet exploratoire PRECURSOR-2, financé dans le cadre de l'AMI DIGIT-BIO 2026. Ce projet vise à mobiliser des méthodes d'intelligence artificielle récentes, incluant des modèles d'apprentissage profond et des modèles de langage appliqués aux séquences d'ADN, afin d'identifier et de cartographier de manière systématique les CREs. Ces approches seront appliquées à plusieurs génomes (nucléaires et plastidiaux) et espèces, afin d'évaluer la robustesse et la transférabilité des méthodes. Un second axe majeur de PRECURSOR-2 concerne l'exploration des liens fonctionnels entre CREs et éléments transposables, en combinant informations de séquence, données fonctionnelles et cadres de modélisation adaptés. Enfin, le projet visera également à tester le pouvoir prédictif des CREs pour expliquer et prédire l'expression des gènes, ainsi que des caractères d'intérêt agronomique, notamment dans des contextes de stress. Ces travaux s'inscrivent dans une logique d'intégration des CREs dans des cadres de génétique quantitative et de génétique des systèmes, en complément des approches classiques basées sur les marqueurs anonymes.

