



Betterave2020

## Dossier de presse / press pack

Cappelle-en-Pévèle, le 12 novembre 2015

### Programme AKER

## 2015... l'année de l'information génomique : générer, stocker, valoriser les données

### Synthèse

Après le séquençage du génome des 15 plantes exotiques choisies dans le cadre du programme AKER, la mise au point de marqueurs moléculaires va permettre de détecter des fragments de chromosomes intéressants pour la sélection de nouvelles variétés.

L'étude du déséquilibre de liaison entre marqueurs permettra d'optimiser le génotypage et de développer de nouvelles méthodes comme la sélection génomique.

La mise en place du système d'information AKER offrira la possibilité de collecter, stocker les données brutes puis de filtrer les données essentielles en vue de les valoriser.

### Sommaire

Introduction

1. Séquençage et marquage moléculaire
2. Traitement de l'information génomique
3. Mise en place du système d'information

Conclusion

Annexes

*Illustrations sur demande*

Contact presse : Philippe Pelzer [philippepelzer@gmail.com](mailto:philippepelzer@gmail.com) 06 50 17 05 29  
[www.aker-betterave.fr](http://www.aker-betterave.fr)



Betterave2020

## Introduction

Le programme AKER va fournir à terme du matériel génétique plus performant et, dès maintenant, des outils et des méthodes de sélection. La thématique de cette année 2015 sur l'information génomique se situe dans cette perspective immédiate.

AKER est organisé en 7 WP (work packages) - groupes de travail interdépendants. Cette année, les WP 3, 5 et 6 sont plus spécialement concernés par la thématique « Information génomique » :

- ✓ WP 3 : Séquençage et développement de marqueurs pour le génotypage

Leader : Pierre Devaux (Florimond Desprez)

- ✓ WP 5 : Intégration de nouvelles stratégies de sélection

Co-leaders: Brigitte Mangin (INRA), Ellen Goudemand (Florimond Desprez)

- ✓ WP 6 : Bioinformatique

Co-leaders : Anne-Françoise Adam-Blondon (INRA), Olivier Robert (Florimond Desprez)

*Illustration 1 : Organisation du programme AKER et focus sur les WP concernés*

## 1. Séquençage et marquage moléculaire

### Séquençage

Au départ du programme AKER en 2013, **15 accessions (plantes exotiques) ont été choisies au sein de 3 000 plantes** parmi 10 000 issues des centres de ressources génétiques internationaux, représentant la diversité génétique de la betterave.

**Le génome de ces 15 accessions et du matériel élite a été séquencé**, c'est-à-dire que l'on a détecté le code génétique constitué par les quatre bases (ACGT) de l'ADN des chromosomes. Le génome de la betterave comprend environ 760 millions de paires de bases. Dans la suite du programme AKER, **les croisements entre les 15 accessions exotiques sélectionnées et la lignée élite ont donné les rétrocroisements Back-Cross 1 puis Back-Cross 2**, qui ont été génotypés avec des marqueurs moléculaires judicieusement choisis.

*Illustration 2 : Croisements réalisés dans le cadre du programme AKER*



Betterave2020

**Des millions de lectures** (« reads » en anglais) de 100 bases de longueur dans les deux sens de lecture (« forward » et « reverse ») **sont réalisées sur l'ADN de chacune des 15 accessions**, après coupure en petits fragments. On obtient ainsi pour chaque accession une bibliothèque de courtes séquences de son génome. **On pourrait les comparer à des sacs contenant les pièces d'un puzzle** qu'il faut assembler pour obtenir la séquence de l'ADN de ces accessions. **L'image de fond du puzzle qui permet de les ordonner est une séquence de référence** obtenue préalablement dans un projet qui a pour origine la société SESVanderHave, du Groupe Florimond Desprez, et le programme allemand GABI.

**Les chercheurs partent d'une séquence de référence sur laquelle ils alignent les reads de l'élite.** Puis, quand l'élite est alignée, **les reads des 15 accessions sont alignés sur cette dernière.** On identifie alors **des différences entre la séquence de l'élite alignée sur la référence et la séquence des différentes accessions alignées sur cette dernière** pour définir des marqueurs moléculaires entre l'élite et les accessions.

*Illustration 3 : Séquençage des accessions*

### Marquage moléculaire

**Un marqueur moléculaire correspond à une courte séquence du code génétique du génome** généralement de l'ordre de 20 bases. Dans le programme AKER, l'objectif a été d'en détecter plusieurs milliers qui soient répartis de manière homogène sur l'ensemble du génome et qui soient polymorphes entre les accessions exotiques et la lignée élite. Dans AKER, **on a identifié, par analyse bioinformatique, plus de 40 millions de variations dans le code génétique du matériel d'étude**, dont des SNP (*Single Nucleotide Polymorphism*) et des IDP (*Insertion Deletion Polymorphism*). Seuls les premiers seront exploités dans un premier temps.

**Tous ces SNP ne peuvent pas devenir des marqueurs moléculaires utilisables** car il existe des contraintes biologiques qui empêchent leur détection dans certains cas. Afin d'éviter ce type de problème, **un important travail de bioinformatique a été effectué pour identifier les SNP qui sont à l'origine de marqueurs moléculaires fiables et robustes.** Cette sélection nécessite l'utilisation de plusieurs étapes et de nombreux logiciels. Afin d'automatiser ces différentes étapes de sélection des meilleurs marqueurs, on a créé un chaîne d'automatisation des différents logiciels dans un environnement approprié (Galaxy). Au final, **200 000 SNP parmi les 40 millions ont franchi toutes les étapes de sélection** et peuvent devenir des marqueurs moléculaires utilisables et potentiellement intéressants.

Lien vers le site : [www.galaxyproject.org](http://www.galaxyproject.org)

*Illustration 4 : Navigateur de génome (JBrowse)*

Un des objectifs du programme AKER est de **générer par croisements un ensemble d'individus qui sont génétiquement quasi identiques à la lignée élite mais qui se**



Betterave2020

**différencient par l'intégration de différents petits fragments du génome des 15 accessions exotiques.** En identifiant des marqueurs moléculaires présents dans les descendants des croisements des parents exotique/élite, on peut maintenant facilement visualiser les différentes parties du génome des parents présentes dans chaque individu. **L'analyse des variations de leur phénotype** (rendement en sucre, résistance aux maladies...) **et de leur génome permettra de découvrir l'effet de ces fragments chromosomiques exotiques sur le phénotype des betteraves**, d'identifier des régions chromosomiques porteuses de gènes d'intérêt et de générer un matériel de base original pour des programmes de sélection.

*Illustration 5 : Détection et sélection des polymorphismes*

Une fois les séquences retenues, une recherche de marqueurs moléculaires du SNP en question est engagée. **Les amorces d'ADN constituant chaque marqueur sont synthétisées chez un prestataire spécialisé.** Ces amorces permettent de mettre en œuvre ces marqueurs moléculaires chez les individus du programme de croisement en utilisant une technique nommée PCR (*Polymerase Chain Reaction* ou Réaction de Polymérisation en Chaîne). Bien entendu, il convient au préalable d'extraire l'ADN de chacun de ces individus. Au total, **15 populations représentées par 6 400 individus chacune vont être suivies** avec une centaine de marqueurs moléculaires, **ce qui représentera plus de 10 millions de points de marquage.**

**En résumé, l'objectif du programme AKER est d'éclater tout le génome de chaque accession dans celui du matériel élite. Le marquage moléculaire permet de repérer des fragments de chromosomes de chaque accession, mais sans savoir *a priori* quels effets ils produiront. Le phénotypage réalisé en fin de programme en 2018 et 2019 permettra alors de le savoir.**

*Illustration 6 : Eclatement des accessions dans le matériel élite*

*Illustration 7 : Profil de marquage SNP*

### **Encadré 1 : Quelques images explicatives**

Les génomes de deux betteraves différentes sont deux copies du même livre dans lesquelles se trouverait un peu partout de petites erreurs typographiques : mots en plus ou en moins, fautes d'orthographe... Ces différences sont repérées par des logiciels de comparaison de texte (logiciels bioinformatiques) et peuvent être utilisés comme des étiquettes spécifiques de chacune des copies du livre (les marqueurs moléculaires). Le marquage moléculaire peut encore se comparer au minage de l'or : on récolte énormément de cailloux pour trouver la pépite.



Betterave2020

## Encadré 2 : Carte physique et carte génétique

La **carte physique** est une représentation du génome basée sur une **distance en nombre de bases**. La **carte génétique** est une représentation du génome basée sur une **distance qui reflète la probabilité de deux locus** (marqueur ou gène) successifs d'être transmis à sa descendance sous une forme recombinées entre les deux formes des deux parents. **Cette probabilité de recombinaison dépend de la proximité entre les deux marqueurs sur l'ADN** : plus ils sont éloignés, plus cette probabilité est forte. Elle dépend également de la position des marqueurs sur les chromosomes : en général, ces recombinaisons sont plus fréquentes au niveau des télomères que des centromères du chromosome.

*Illustration 8 : Composition d'un chromosome*

**La correspondance entre la distance physique et la distance génétique entre les deux marqueurs est donc variable le long des chromosomes.** Ces cartes génétiques sont des outils indispensables pour mettre en relation des caractères phénotypiques et des marqueurs, mais cela suppose que les marqueurs de cette carte soient bien répartis sur le génome, et ainsi de suivre la transmission dans la descendance de gènes/QTL codant pour des caractères phénotypiques visibles ou mesurables.

*Illustration 9 : Carte physique/carte génétique*

## Encadré 3 : Polymorphisme moléculaire

Le polymorphisme moléculaire exprime une **variation entre individus dans la séquence de leurs génomes**. Ces variations peuvent être à l'origine des différences observées entre les individus d'une population, comme la forme de la racine. Ces polymorphismes correspondent soit à des insertions ou à des délétions IDP (*Insertion Deletion Polymorphism*), soit à des SNP (*Single Nucleotide Polymorphism*).

## Encadré 4 : Mise au point d'une puce à ADN

Florimond Desprez ainsi que sa filiale SESVanderHave ont mis au point une puce à ADN pour la betterave sous l'intitulé Bv15 FDSV. Cette petite plaque de quelques millimètres va permettre de **traiter en un lot plus de 30 000 marqueurs moléculaires**, rendant plus précis le travail d'analyse des individus.

*Illustration 10 : Puce à ADN*



Betterave2020

### Encadré 5 : Le *landscape genetics*

Le programme AKER s'est enrichi d'un nouveau module, le « *landscape genetics* », que l'on peut traduire par « génétique paysagère » et qui consiste à **étudier la manière dont chaque plante s'est adaptée à son environnement** (sol, climat, luminosité, hygrométrie, altitude...). Cette technique doit permettre de cartographier la distribution géographique des ressources génétiques de la betterave. Les travaux sur le *landscape genetics* sont menés par le CNRS et l'Université de Montpellier (Laboratoire biogéographie et écologie des vertébrés), en collaboration avec Florimond Desprez.

## 2. Traitement de l'information génomique

Le programme AKER permet de recueillir de nombreuses données génomiques grâce aux marqueurs moléculaires ; il convient maintenant de les exploiter.

### Déséquilibre de liaison

Le déséquilibre de liaison permet **d'étudier la liaison entre les gènes causaux, c'est-à-dire ceux qui expliquent les différences entre les individus, et les marqueurs génétiques**. Si un gène est en déséquilibre de liaison avec un marqueur, cela signifie que l'information au marqueur permettra de récupérer l'information au gène lié.

**Le déséquilibre de liaison est lié à la proximité physique sur le génome**. Si deux marqueurs sont **proches** physiquement, le déséquilibre entre eux est **important** ; s'ils sont éloignés, le déséquilibre est faible. Ces phénomènes (la liaison et la proximité physique) se conjuguent. Par conséquent, si le déséquilibre de liaison est à **courte portée**, les liaisons gènes/marqueurs détectées seront **précises**, mais il faudra **beaucoup de marqueurs** pour pouvoir les détecter (forte densité de marquage). En revanche, si le déséquilibre est à **longue portée**, il faudra **moins** de marqueurs, mais les liaisons seront **moins précises**.

#### *Illustration 11 : Le concept de déséquilibre de liaison*

Ainsi l'étude du déséquilibre de liaison peut permettre de **déterminer la densité de marquage optimale** permettant d'entreprendre des analyses pour lesquelles tous les gènes causaux doivent être en déséquilibre de liaison avec au moins un marqueur afin de pouvoir capter toute l'information. **On détermine la distance maximale entre deux marqueurs pour obtenir du déséquilibre de liaison entre eux**, d'où on déduit un nombre de marqueurs nécessaires pour analyser le génome correctement. Déterminer la densité de marquage optimale permet de **faire de la planification de génotypage** (si l'analyse est effectuée *a priori*) ou de conclure sur la véracité des résultats obtenus (si l'analyse est réalisée *a posteriori*).



Betterave2020

**Grâce à la comparaison du déséquilibre de liaison entre des populations élites et sauvages, on peut aussi repérer des traces de la sélection.** En effet, dans certaines zones du génome correspondant à des régions soumises fortement à la sélection, **le déséquilibre de liaison est plus étendu (longue portée) dans les populations élites que dans les populations sauvages.** C'est en quelque sorte la signature du sélectionneur. On a pu notamment mettre en évidence une pression de sélection très forte au niveau des gènes de monogermie, de stérilité mâle ou de résistance à la rhizomanie. Ce déséquilibre de liaison confirme à ce stade des éléments connus sur les populations élites.

*Illustration 12 : Représentation du déséquilibre de liaison dans la population élite*

*Illustration 13 : Représentation du déséquilibre de liaison dans les populations élites et sauvages*

**Le déséquilibre de liaison est donc au service des nouvelles méthodes de sélection basées sur les liaisons marqueurs/gènes causaux, notamment la sélection génomique.**

### **Sélection génomique**

Le programme AKER prend exemple sur une **méthode de sélection qui a fait ses preuves chez les bovins laitiers** et se propose de l'appliquer à la sélection de la betterave sucrière. L'intérêt réside dans le **gain de temps** qu'elle procure à la sélection (accélération des cycles de sélection), ainsi que dans **l'augmentation des précisions de sélection** (sélection directement sur l'ADN des plantes).

L'évolution rapide des techniques de séquençage et de génotypage a soulevé de nouveaux défis dans le développement des méthodes de sélection pour les animaux d'élevage : le concept de sélection génomique a été introduit en 2001. Il propose de **prédire simultanément tous les effets des régions marquées du génome, puis de construire un index génomique en sommant les effets de chaque région et enfin de sélectionner les meilleurs animaux en se basant sur cet index.** Les évaluations génomiques augmentent la précision des évaluations des individus et ont la faculté d'augmenter rapidement le gain génétique des caractères d'intérêt.

Dans le cadre du programme AKER, on met actuellement en place cette **nouvelle méthode de sélection sur les populations élites.** Dans un premier temps, les chercheurs ont souhaité démontrer que cette méthode est aussi applicable à la sélection de la betterave sucrière en faisant une **démonstration de faisabilité** (en anglais : *proof of concept*) par validation croisée.

**Tout débute avec une population globale élite** pour laquelle on dispose d'un génotypage et d'un phénotypage (performances réelles des individus dans les champs). A partir de cette



Betterave2020

population globale, on définit une **population d'entraînement** qui va permettre de calibrer le modèle grâce à la mise en parallèle de son phénotypage et de son génotypage, et une **population de validation** pour laquelle uniquement son génotypage est utilisée. La population d'entraînement est utilisée pour prédire les performances de la population de validation et ces performances prédites sont ensuite comparées aux performances réelles disponibles (phénotypage).

*Illustration 14 : Mise en place d'un schéma de sélection génomique*

**Preuve de concept *in silico* (par l'outil informatique) : la sélection génomique de la betterave fonctionne avec des précisions de prédiction de 75-80 %.** A titre d'exemple, une population d'entraînement de 1 300 lignées de betteraves permet d'obtenir une capacité de prédiction moyenne de 0,775 pour le caractère « rendement en sucre blanc ».

*Illustration 15 : Capacité de prédiction du rendement en sucre blanc par la sélection génomique*

Il convient ensuite d'appliquer ce modèle sur des populations en cours de sélection, dont le phénotype est encore inconnu, provenant de croisements entre deux parents élités. **Deux schémas principaux vont être comparés : sélection phénotypique classique et sélection génomique.** On va observer quel sera le meilleur schéma de sélection *in vivo* en comparant les performances des individus sélectionnés par les différents schémas en fin de programme AKER.

*Illustration 16 : Comparaison de schémas de sélection classique et génomique*

A l'issue du programme, la population AKER issue des croisements entre un parent élite et les accessions sauvages sera prise comme population d'entraînement, et **on pourra appliquer la sélection génomique pour la prédiction et la sélection des individus issus des croisements post-AKER.**

### **Encadré 6 : La sélection génomique chez les bovins de race Holstein**

Jusqu'à 2009, le contrôle des géniteurs bovins de race Holstein se faisait uniquement sur l'évaluation de leurs descendances, c'est-à-dire seulement une fois que leurs premières filles avaient leur première lactation. La diffusion des nouveaux géniteurs ne se faisait donc pas avant l'âge de 5 ans et la précision de prédiction de la valeur des taureaux restait faible (0.75 pour le rendement lait) jusqu'à environ 8 ans (0.95 pour le rendement lait), avant qu'une grande quantité de données ne devienne disponible.

Dorénavant, la sélection génomique permet grâce au génotypage des taureaux dès leur naissance, de diffuser les nouveaux géniteurs dès leur maturité sexuelle, soit 2 ans, avec des précisions de prédiction de l'ordre de 0.70 pour le rendement en lait. La recherche a démontré





Betterave2020

que, dans la plupart des situations, un groupe de 10 géniteurs sélectionnés par sélection génomique a la même moyenne de fiabilité que des taureaux élités.

Les producteurs laitiers ont donc maintenant le choix entre utiliser des jeunes géniteurs qui ont uniquement des informations génétiques ou des taureaux éprouvés pour lesquels les résultats de testage sur descendance sont disponibles. En 2012, 70 % des jeunes taureaux Holstein sur le marché sont issus de la sélection génomique alors que les 30 % restant proviennent de la sélection classique sur descendance. Les entreprises de sélection Holstein ont mis à la disposition des éleveurs des packs regroupant les meilleurs taureaux issus de la génomique et permettant d'augmenter la fiabilité génétique.

### 3. Mise en place du système d'information

Le programme AKER va générer un grand nombre de données. Il convient de les structurer dans un système d'information : ensemble organisé de ressources (logiciels, données et procédures) qui permet de regrouper, classifier, traiter et diffuser de l'information.

#### GnplS (Genoplante Information System)

GnplS (Genoplante Information System) est un **système d'information intégré fonctionnel**, mis au point par l'unité de recherches INRA URGI (Unité de Recherche en Génomique-Info) et permettant de **gérer les données INRA de génétique et de génomique pour les plantes et leurs champignons pathogènes**. En constante évolution depuis 10 ans, il est utilisé dans tous les **Programmes d'Investissements d'Avenir**, Breedwheat (blé), Amazing (maïs), Peamust (pois) et Rapsodyn (colza), où la version maintenue à l'URGI sert d'entrepôt intégré de données et de mise à disposition de données à long terme.

Lien vers le site : <http://urqi.versailles.inra.fr/gnplS>

**GnplS est adapté en permanence aux spécificités de la betterave** par l'URGI, et ces évolutions sont mises en production trois fois au cours de l'année. AKER est le premier programme à développer et maintenir une version du système d'information GnplS spécialisée pour la betterave.

#### AIS (AKER Information System)

Ainsi, **l'architecture de GnplS a été dupliquée pour le programme AKER sous l'intitulé AIS** (AKER Information System) et une infrastructure permettant d'accueillir AIS a été mise en place par le Laboratoire de Génétique et Biométrie (LGB) de Florimond Desprez. La structure de la base de données de GnplS a été alors installée. Elle permet de **centraliser les données produites** (ressources génétiques, génotypage, phénotypage, séquençage, etc.)



Betterave2020

dans le cadre du programme AKER pour les 3 000 accessions étudiées **et ainsi d'assurer leur pérennité dans AIS.**

Toutes les données utiles au programme ou générées par le programme AKER seront donc déposées dans AIS :

- Les **données brutes** au minimum, sachant que les images ne seront pas chargées dans AIS mais qu'il sera possible de les rendre accessible via un URI (*Unique Resource Identifier*). Ce système garantit leur unicité et leur pérennité.
- Les **données élaborées** issues d'un processus d'analyse, de calculs statistiques ou d'un logiciel d'analyse d'image.

En termes de débit, il n'y aura pas de problèmes particuliers pour charger les résultats des expérimentations sur les 3 000 accessions étudiées : la structure d'AIS est déjà adaptée. **Les données publiques issues du projet seront également transférées dans le système d'information de l'URGI, GnpIS,** ce qui contribuera à les rendre visible et les porter à la connaissance de la **communauté internationale des chercheurs en biologie végétale.**

**En résumé, ce système d'information permet d'agrèger les données et de les rendre accessibles.** Il doit permettre de visualiser les données essentielles aux moments clés du programme. Pour cela, il aide à gérer de très grandes quantités (à titre d'exemple, les données qui concernent le séquençage pèsent 4 Téra octets au total : 2 pour les données brutes, 2 pour les données générées) et à pérenniser les données importantes après les analyses. **Le poids des informations à conserver dépendra donc des filtres mis en place pour tamiser ces données.** Pour reprendre des images, on fera la différence entre les brouillons et la rédaction finale, entre les diamants bruts et le diamant taillé.

*Illustration 17 : Système d'information AKER (AIS)*

### **Encadré 7 : Une ontologie pour la betterave**

Le programme AKER est l'occasion de mettre au point une ontologie (au sens informatique du terme) pour la betterave, sous la houlette du Laboratoire de Génétique et de Biométrie de Florimond Desprez, ce qui n'existait pas auparavant. Une ontologie est à la fois **un vocabulaire universel et une information structurée** qui permet de mieux partager des données acquises par différents chercheurs et surtout d'y faire des fouilles automatisées. C'est à la fois **un dictionnaire, un lexique, une arborescence, une carte des relations entre les termes qui fonctionne avec une logique en cascade.**

Ce support en anglais va fixer les termes avec lesquels on décrira les observations réalisées sur la betterave, de manière à ce que tous les protagonistes parlent le **même langage.** Par exemple, on mesurera la feuille en centimètres et non en pouces ; on parlera de la racine pour ensuite évoquer sa longueur, sa forme, sa couleur... Actuellement, **une centaine de termes**



Betterave2020

**sont au point et la première version publique est lancée.** Cette ontologie de la betterave est **destinée à être appliquée sur les données de phénotypage générées dans AKER** et elle pourra être améliorée si nécessaire au cours du programme.

Lien vers le site : [www.croponology.org](http://www.croponology.org)

### **Encadré 8 : La bioinformatique**

« La bioinformatique est une science à l'interface des **disciplines numériques** (l'informatique et les mathématiques) et des **sciences de la vie** (biochimie, biologie, microbiologie, écologie, épidémiologie).

Étant donné que les scientifiques de la vie génèrent une quantité croissante de nouvelles données portant sur les génomes, les biomolécules, les organismes, leurs interactions et leur évolution, il y a un **besoin croissant d'approches informatiques** pour la manipulation, le stockage, la visualisation et l'analyse de ces données souvent très complexes. »

### **Encadré 9 : Big Data**

La notion de Big Data recouvre une **combinaison de progrès technologiques** en lien avec la massification des données **et d'innovations d'usage de ces données**. En effet, un ensemble d'innovations technologiques transforme depuis quelques années la façon dont les données sont **générées**, avec notamment des volumes et des débits exponentiels, **transmises, stockées et utilisées**, et cela de façon de plus en plus distribuée (via le cloud). Par ailleurs, ces données sont très **hétérogènes** (vidéo, texte, son, image) et les échanges suivent aussi cette tendance à la **massification**. On observe également des révolutions dans leur structuration, leur visualisation et leurs moteurs de recherche (e.g. Google). Ces nouvelles capacités à « faire parler » de gros volumes de données très hétérogènes constituent la **troisième révolution d'internet** et sont basées sur une « sémantisation » des données.

Cependant, la montée en puissance du Big Data n'est pas uniquement une histoire de technologie. **Les évolutions culturelles vis-à-vis de la génération et du partage d'informations sont des éléments clés de l'augmentation de la richesse et du volume des données.**

### **Conclusion**

Les outils et méthodes de sélection développés pour la connaissance du génome vont permettre de détecter des gènes d'intérêt qui seront ensuite introduits dans les nouvelles variétés issues du programme AKER.



Betterave2020

Ces outils et méthodes conjuguent la biologie avec les autres disciplines telles que l'informatique ou les statistiques dans une vision transversale et collaborative. Elles génèrent des données de plus en plus nombreuses qu'il convient de collecter, trier, stocker, valoriser et diffuser le moment venu.

Les résultats obtenus par le génotypage et les nouvelles méthodes de sélection génomique devront être vérifiés par le phénotypage en fin de programme de manière à valider les hypothèses retenues.

## Annexes

### • Publications scientifiques

- ✓ Publication scientifique sur le code informatique du déséquilibre de liaison LDcorSV : Mangin, B., Siberchicot, A., Nicolas, S., Doligez, A., This, P., Cierco-Ayrolles, C. (2012). Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity*, 108 (3), 285-291. DOI : 10.1038/hdy.2011.73
- ✓ Publication scientifique sur le déséquilibre de liaison : Mangin, B., Sandron, F., Henry, K., Devaux, B., Willems, G., Devaux, P., Goudemand, E. (2015). Breeding patterns and cultivated beets origins by genetic diversity and linkage disequilibrium analyses. *Theoretical and Applied Genetics*. On line. DOI: 10.1007/s00122-015-2582-1
- ✓ Publication scientifique sur le *landscape genetics* : Marco Andrello, Karine Henry, Pierre Devaux, Bruno Desprez, Stéphanie Manel. Taxonomic, spatial and adaptive genetic variation of Beta section Beta

### • Posters disponibles

- ✓ WP 3 : Séquençage et développement de marqueurs pour du génotypage haut débit
- ✓ WP 5 : Mise en œuvre de nouvelles stratégies de sélection
- ✓ WP 5 : La sélection génomique
- ✓ WP 6 : Bioinformatique



Betterave2020

- **Présentation des partenaires impliqués**

- ✓ INRA, URGI

L'URGI (Unité de Recherche en Génomique-Info) est une unité de recherche en génomique et bioinformatique de l'Institut National de la Recherche Agronomique (INRA), située à Versailles, dédiée à la génomique des plantes et de leurs pathogènes. Son activité de recherche porte sur la structure et la dynamique du génome. L'URGI comprend à ce jour deux équipes, l'une dédiée au développement d'outils pour l'analyse des génomes de plantes et de champignons, l'autre au développement du système d'information et de l'intégration de données. L'URGI apporte ainsi au programme AKER les outils et l'expertise dont il a besoin en bioinformatique. L'unité héberge une plateforme bioinformatique appartenant à l'Institut Français de Bioinformatique (IFB) <http://www.france-bioinformatique.fr/>, qui est lui-même le nœud français du réseau européen des plateformes de bioinformatique (ELIXIR) <https://www.elixir-europe.org/> Lien vers l'article : <http://www.ncbi.nlm.nih.gov/pubmed/?term=23959375>  
<https://urgi.versailles.inra.fr/>

- ✓ INRA, LIPM

Le Laboratoire des Interactions Plantes-Microorganismes (LIPM), créé en 1981, situé à Toulouse, est une Unité Mixte de Recherche CNRS-INRA, actuellement rattachée à l'Institut des Sciences Biologiques (INSB) et à l'Institut Ecologie et Environnement (INEE) du CNRS ainsi qu'aux départements Santé des Plantes et Environnement (SPE) et Génétique et Amélioration des Plantes (GAP) de l'INRA. Le LIPM explore comme champ de recherche prioritaire les interactions entre les plantes et les micro-organismes symbiotiques ou pathogènes, par des études concertées des partenaires microbiens et végétaux. Ces recherches sont réalisées sur un petit nombre d'espèces modèles (*Arabidopsis thaliana*, *Medicago truncatula*), mais aussi plus récemment sur des modèles d'intérêt agronomique tels que le tournesol ou la tomate.

<http://www6.toulouse.inra.fr/lipm>

- ✓ Florimond Desprez, LGB

Le Laboratoire de Génétique et de Biométrie (LGB) de Florimond Desprez a pour missions d'apporter du matériel génétique nouveau et des outils d'aide à la sélection innovants aux différents laboratoires de la société, de représenter Florimond Desprez auprès des entreprises et des laboratoires de recherche dans le cadre des projets de recherche collaboratifs. Les thématiques développées concernent la veille bibliographique, les analyses statistiques et génétiques, la mise en place et la gestion des bases de données, la bioinformatique et la gestion des séquences d'ADN, etc.

[www.florimond-desprez.com](http://www.florimond-desprez.com)



Betterave2020

✓ Florimond Desprez, LBi

Le Laboratoire de Biotechnologies (LBi) de Florimond Desprez a pour missions d'apporter des outils d'aide à la sélection en cytologie, biologie cellulaire et biologie moléculaire aux différents laboratoires internes et externes de la société, d'améliorer et de rechercher de nouveaux outils pour l'amélioration des plantes. Il participe aux programmes de recherche collaboratifs, et réalise des prestations de services pour d'autres sociétés et instituts de recherche en France et à l'étranger. Le Laboratoire de Biotechnologies expertise et suit également des programmes de recherche internationaux.

[www.florimond-desprez.com](http://www.florimond-desprez.com)