



Fonds de soutien à l'Obtention Végétale

GPDUR : Génétique et Phénotypage du blé dur pour l'amélioration conjointe du rendement et de la teneur en protéines

Delphine HOURCADE*¹, Boris DEMENOU¹, Damien EYBALIN¹, Matthieu BOGARD¹, Thierry LEFEVRE², Marion CHAPUS³, Carine ANDRE³, Andrea MASSI³, Jacques DAVID⁴, Pierre ROUMET⁴, Céline DUQUE⁵

- 1 - ARVALIS - 6 Chemin de la Cote Vieille - 31450 Bazège
 - 2 - RAGT 2n - 1740 Route du Canal - 82700 Montbartier
 - 3 - SYNGENTA - 12 Chemin de l'Hobit, 31790 Saint-Sauveur / Società Produttori Sementi S.p.A. Via Macero, 1 - 40050 Argelato (Bologna) - Italy
 - 4 - INRAE - Bâtiment Arcad, 10 rue Arthur Young, 34090 Montpellier
 - 5 - LIMAGRAIN Europe - Ferme de l'Etang - BP3 - 77390 Verneuil l'Etang
- *Coordonateur : Delphine HOURCADE, d.hourcade@arvalis.fr

Introduction :

Les exigences qualitatives des utilisateurs ont été à l'origine d'une structuration de la filière au sein de laquelle toutes les parties prenantes ont visé une production de haut niveau qualitatif, recherchant notamment de hauts niveaux de teneur en protéines. Malheureusement, la base génétique du matériel végétal cultivé en France, étroite, semble arriver à ses limites. Le taux de protéines est un critère majeur pour l'accès au marché, le progrès génétique en matière de rendement est de fait très réduit. Evaluer une plus large diversité dans les conditions de culture française, afin d'identifier des sources génétiques pouvant amener à rompre cette corrélation négative entre rendement et protéine semble donc essentiel. L'analyse de la GPD (Grain protein deviation) permet d'évaluer l'écart à la tendance globale de chaque variété à partir de la régression Rendement/Protéine. Il semble utile de mieux caractériser la diversité disponible aux sélectionneurs sur ce critère. Plusieurs sources de diversité étaient disponibles pour constituer un panel de diversité génétique élite, adaptée ou rapidement adaptable aux conditions françaises: le panel de diversité élite mondiale constitué par l'Université de Bologne et la population EPO (Evolutionary Pre-breeding Population) qui a été spécialement conçu dans le but d'intégrer un maximum de diversité élargie et originale par rapport au compartiment élite du blé dur. Il est issu de 14 cycles de brassage avec un taux d'allogamie estimé de 20% d'inter-croisements d'une population composite fondée par l'introgression d'accessions entre les sous espèces de blé dur *T. turgidum* sp. ayant divergé avant l'apparition de la forme moderne de blé dur (*T. turgidum durum*) correspondant aux différentes étapes évolutives. Le panel sera complété par un choix de variétés élites cultivées ou plus anciennes. L'ensemble de ces génotypes est phénotypé dans plusieurs sites, sur 3 ans puis génotypés pour être analysé avec des approches de génétique d'association et de prédiction génomique.

Définition du panel

Différents critères ont permis de sélectionner les génotypes constitutifs du panel d'étude :

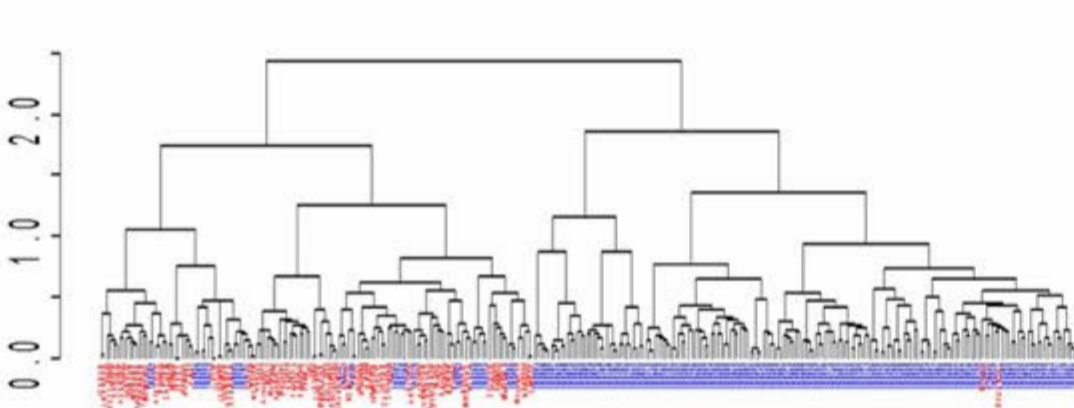
- Analyse de diversité basée sur les données de génotypage, puis sélection des génotypes représentant des groupes génétiques différents (Graphique 1).
- Précocité (variation de 10 jours maximum)
- Hauteur : gamme 60 à 100 cm
- Potentiel rendement
- Connaissance a priori GPD + ou GPD -

Phénotypage du panel

Tableau 1 : Réseau d'essai du panel

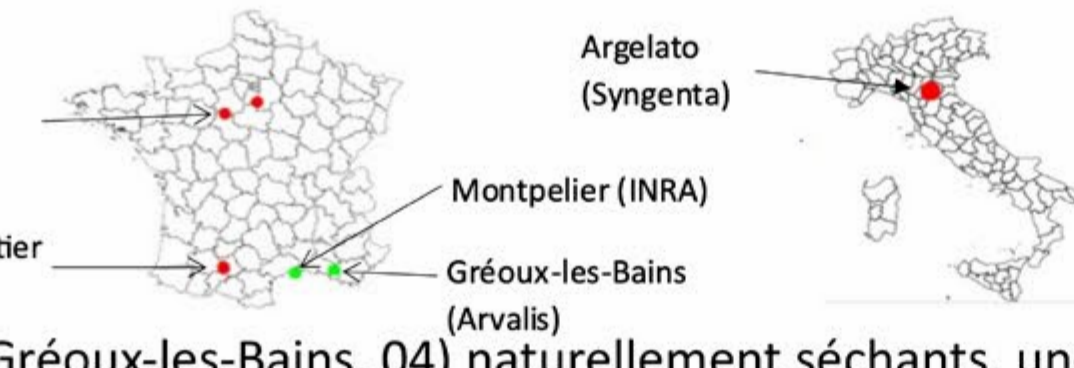
	RAGT	Limagrain puis QUALPREST	INRAE	SYNGENTA	ARVALIS
2016-2017	Multiplication: Rouille jaune et Septoriose	Rouille Jaune		Fusariose	
2017-2018	Rendement /Protéines	Rendement /Protéines		Rendement /Protéines	
2018-2019	Rendement /Protéines	Rendement /Protéines	Rendement /Protéines	Rendement /Protéines	Rendement /Protéines

- Les sites RAGT (Montbartier, 81), QUALPREST (Baignolet et Sancheville, 28) et SYNGENTA (Argelato, Nord Italie) sont conduits en pluvial et fertilisation Azotée optimale
- Les sites INRAE (Mauguio, 34) et ARVALIS (Gréoux-les-Bains, 04) naturellement séchant, une modalité irriguée et une modalité pluviale sont conduites.
- 2 répétitions ont été implantées dans chaque essai ainsi que 4 à 6 témoins de précocité.
- Mesures : Rendement (GY), PMG (TKW), Poids spécifique, Teneur en protéines (GPC), Indice de sédimentation SDS, Gluten humide (WG), Mitadin (Yellow-Berry), Moucheture (Black-point), Hauteur, Epiaison (Z55). Spécifiquement, un suivi NDVI a été conduit sur les essais INRAE et Syngenta (Greenseeker) et ARVALIS (Phénomobile).



Graphique 1 : Classification hiérarchique de la population EPO (en bleu) et des variétés françaises (en rouge)

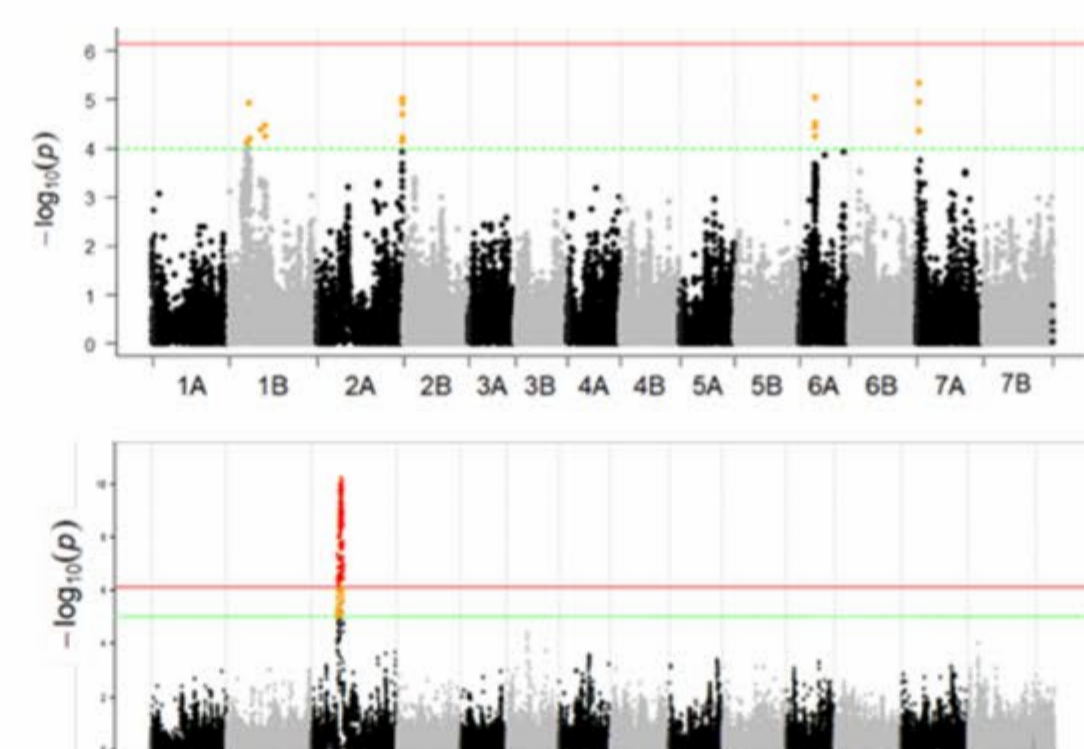
La campagne 2016-2017 a été consacrée à la multiplication des semences (300 variétés) en pépinière et à été l'opportunité d'évaluer la tolérance aux maladies. Les campagnes 2017-2018 et 2018-2019 ont été consacrées à la mise en place de 8 essais Rendement/Protéines sur 235 variétés.



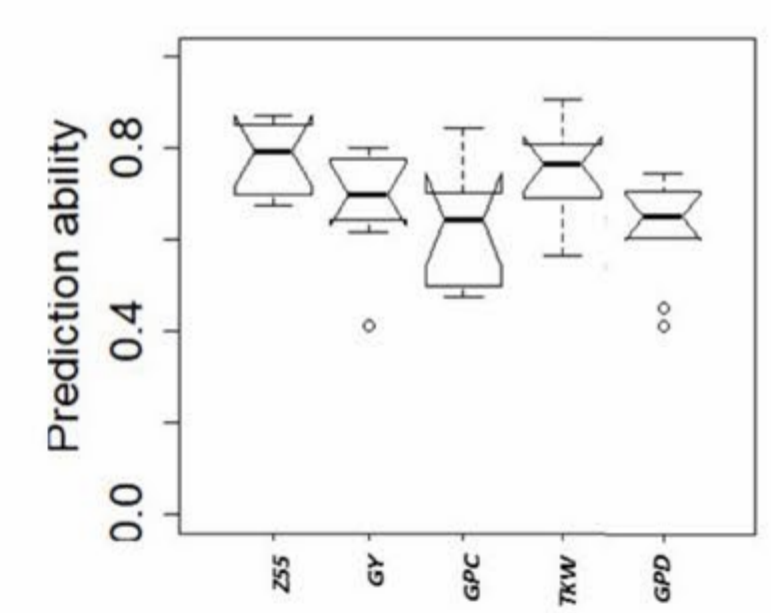
Génétique d'association et prédiction génomique

Le panel a été génotypé à l'aide de la puce TaBW420K développée dans le cadre du projet français Breedwheat. Le set de 124923 marqueurs de départ est filtré selon plusieurs critères (MAF (« minor allele frequency ») < 5%, taux d'hétérozygotie > 20%). La matrice finale contient 68093 SNPs. La procédure de GWAS utilisée est similaire à celle proposée par (Segura et al., 2012). Le modèle statistique Q+K a été utilisé prenant en compte la matrice de Kinship et l'appartenance à l'une des 3 origines génétiques. Le seuil de significativité utilisé usuellement est calculé sur la méthode de correction de Bonferroni soit $-\log_{10}(0.05 / 68\ 093) = 6$ mais il est très stringent notamment sur les traits complexes étudiés dans ce projet. Par conséquent un seuil de 4 est communément choisi pour GY, GPD et GPC.

Au total, 12 QTL sont identifiés pour quatre traits parmi les onze étudiés. Pour le trait "Hauteur de la plante", 1 QTL (CFN2441455) détecté sur le chromosome 4B avec un LOD de 12.2 expliquant une variance phénotypique de 19%. Neuf QTL sont identifiés pour le SDS sur les chromosomes 1A, 1B, 2B, 4A, 5B et 6B avec un LOD allant de 7.45 à 14.46 pour une variance phénotypique expliquée de près de 15%. Un QTL est identifié respectivement pour le PMG (CFN1520694, chr. 2A; Graphique 3) et le Mitadin (CFN0559919; chr. 1B) avec un LOD de 10 et 6.1 pour une variance phénotypique expliquée respectivement de 21% et 15%.



Graphique 3 : Manhattan plots issus des analyses de génétique d'association sur les traits GPD et TKW



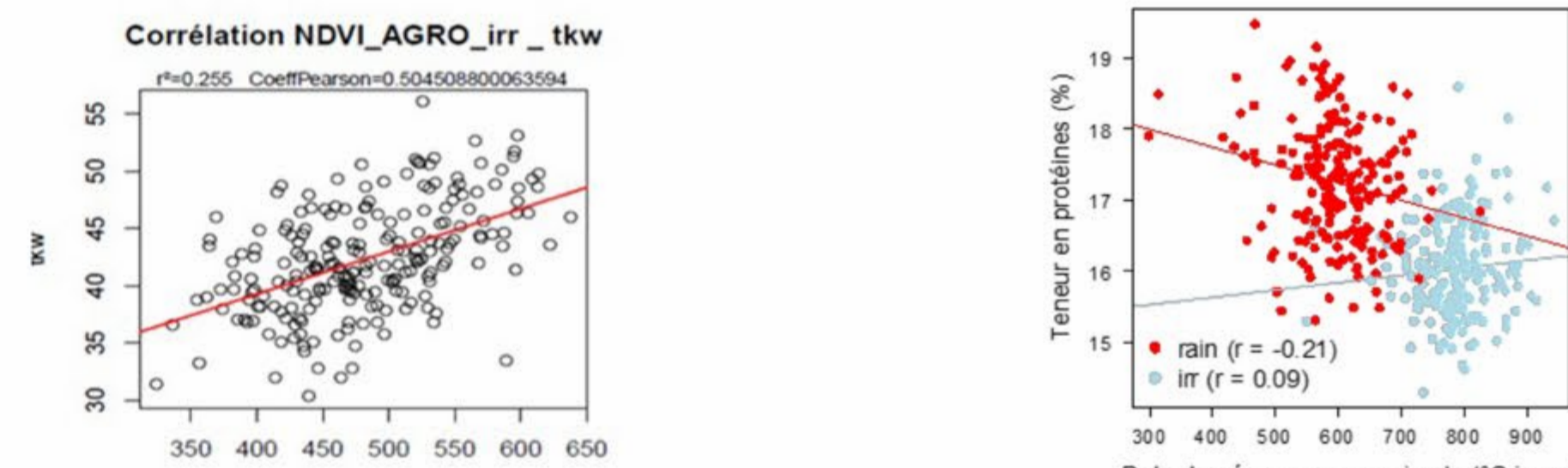
Graphique 4 : Capacités de prédictions génomiques après validation croisée en 10 ensembles du modèle BGLR+QTL

Pour un seuil de LOD de 4, de 3 à 26 QTL sont observés. Quatre QTL sont obtenus pour le GPD avec des LOD de 4,91 à 5,33 sur les chromosomes 1B, 2A, 6A et 7A (Graphique 3). Trois QTL pour le GPC et 5 pour le GY sont également obtenus. Le QTL CFN0126866 (6A) est détecté pour le GPC et le GPD; avec une variance phénotypique expliquée de 7.4% et 10% respectivement, et un effet allélique positif sur les deux traits, impliquant qu'il pourrait augmenter le GPC sans effet négatif sur le rendement. Un QTL associé à un seuil de 4.3 pour Z55 sur le chromosome 2A colocalise avec un QTL majeur de précocité connu Ppd-A1.

Un autre moyen de valoriser ces QTL à effet faible a été évalué. La prédiction génomique du GPD donne une valeur de capacité prédictive moyenne de 0.63 (Graphique 4) et montre que l'intégration des marqueurs génétiques améliore ces capacités. Cette valeur de prédiction du GPD est suffisamment intéressante pour son utilisation par les sélectionneurs dans les programmes de sélection.

Phénotypage haut débit

Concernant le site INRAE, une analyse d'ajustement de courbe logistique a été réalisée par l'INRAE sur les données NDVI pour en extraire différents paramètres. Des corrélations entre les valeurs de ces paramètres et les variables agronomiques ont été calculées. Elles sont nulles avec le GPD quel que soit le paramètre de la courbe étudié mais certaines sont intéressantes avec le PMG par exemple, c'est le cas pour 2 points, l'un correspondant au maximum de surface verte (t0) et l'autre au débit d'inflexion (tinf) pour la modalité irriguée (Graphique 5). En d'autres termes, une variété qui maintient un appareil végétatif vert le plus longtemps possible favorise le remplissage de ses grains. Pour la modalité stressée, les corrélations sont quasi inexistantes pour tous les paramètres et les traits agronomiques.



Graphique 5 : Relation linéaire entre les paramètres de courbes calculées sur les données NDVI et les traits agronomiques mesurés à Mauguio (INRAE) et Gréoux-les-bains (Arvalis)

Evaluation du critère GPD

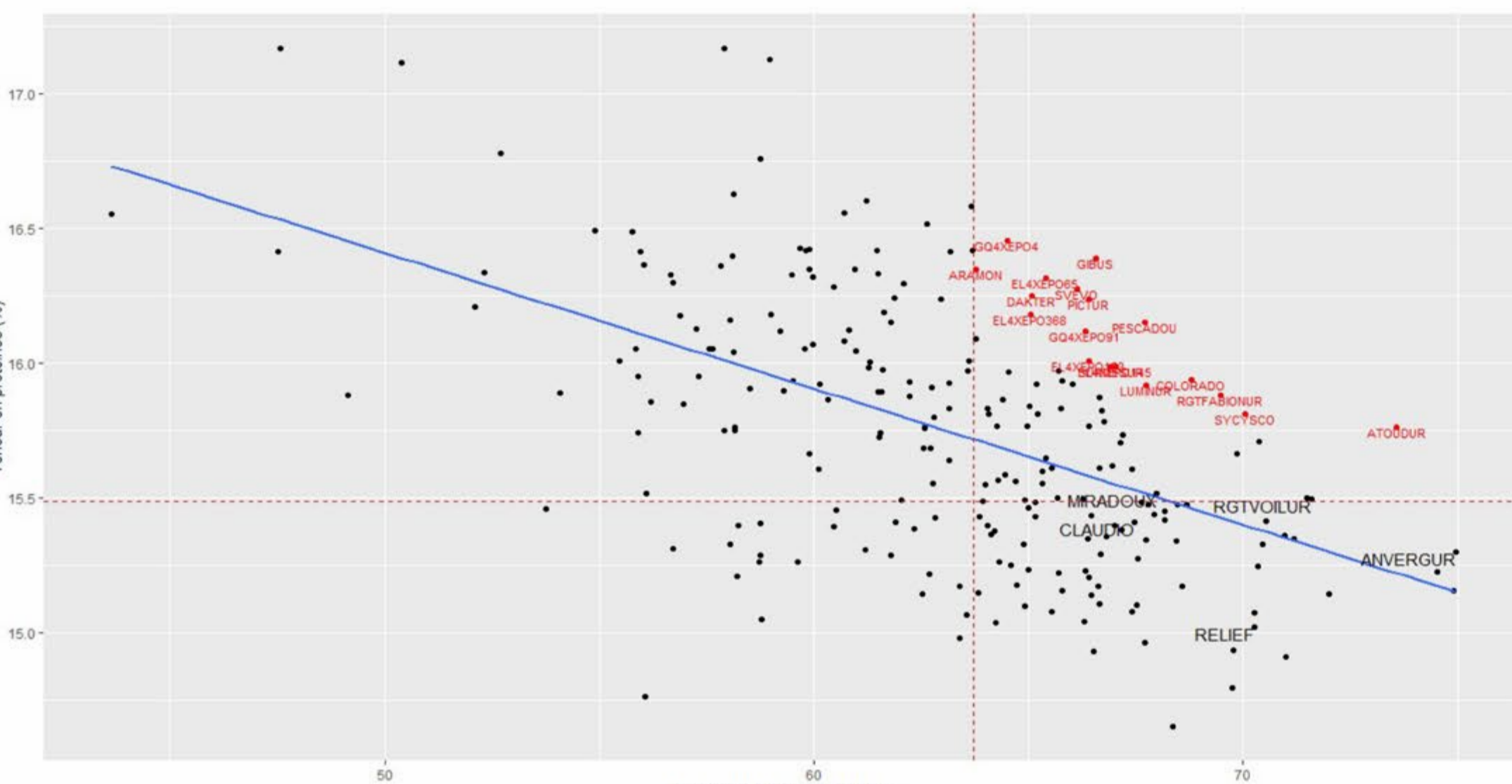
Pour calculer le caractère GPD (grain protein deviation), une méthode itérative est utilisée afin de neutraliser l'effet des valeurs aberrantes sur la pente de la relation rendement-taux en protéines. A la fin, on obtient la régression linéaire qui explique le plus d'observations, on peut calculer les résidus de tous les individus.

Afin d'avoir des données moyennées à l'échelle du réseau d'essai, un modèle mixte a été utilisé avec en effet fixe le génotype et en effet aléatoire chaque essai, selon la formule:

$$Y(ij) = g(i) + \text{essai}(j)$$

Une valeur par variété et pour tout le réseau est ainsi obtenue.

Pour identifier les géneurs qui s'écartent significativement de la régression globale, un seuil est fixé à 15% du meilleur rendement, à la teneur en protéine du meilleur témoin et GPD>0.5.



Graphique 2 : Régression Rendement/Teneur en protéine à l'échelle du réseau d'essai

Sur le graphique 2, les génotypes positionnés au-delà de ces seuils et de la droite de régression sont considérés comme GPD+, car elles présentent un taux de protéines significativement supérieur pour un niveau rendement donné. Ces variétés ont un taux de protéines supérieur de l'ordre de 0.5 à 0.75 points de protéines par rapport à la variété référence du moment, RGT Voilur. Parmi ces génotypes, 6 variétés appartiennent au groupe EPO, 10 sont des variétés élites inscrites en France et 2 appartiennent au panel Unibo. Il n'existe pas vraiment de corrélation entre la nature des populations étudiées et leur écart à la relation « rendement-protéines ». On peut préciser toutefois que les lignées EPO présentent une variabilité plus importante des valeurs de GPD mais aussi les valeurs les plus élevées.