

## QTL affectant la composition fine du lait en protéines dans les trois races bovines Montbéliarde, Normande et Holstein

SANCHEZ M.P. (1), GOVIGNON-GION A. (1), M. FERRAND M. (2), GELE M. (2), POURCHET D. (3), ROSSIGNOL M.N. (4), FRITZ S. (5), MIRANDA G. (1), MARTIN P. (1), BROCHARD M. (2), BOICHARD D. (1)  
(1) INRA, UMR1313 GABI, 78350 Jouy en Josas, France  
(2) IDELE, 75012 Paris, France  
(3) ECEL Doubs - Territoire de Belfort, 25640 Roulans, France  
(4) Labogena, 78350 Jouy en Josas, France  
(5) UNCEIA, 75012 Paris, France

### RESUME

Dans le cadre du programme PhénoFinLait, une analyse du génome complet a été réalisée pour détecter des QTL affectant la composition protéique du lait estimée à partir de spectres moyen infrarouge dans les trois principales races bovines laitières françaises : Montbéliarde (MO), Normande (NO) et Holstein (HO). Les teneurs en protéines  $\alpha$ -lactalbumine,  $\beta$ -lactoglobuline et caséines  $\alpha$ s1,  $\alpha$ s2,  $\beta$  et  $\kappa$  ont été estimées en g/100g de lait (%lait) ou en g/100g de protéines (%prot) sur près de 900 000 échantillons de lait prélevés sur 160 253 vaches. Des génotypages ont été réalisés avec la puce Illumina 50K sur 2 773 MO, 2 673 NO et 2 208 HO. Le phénotype analysé est la moyenne sur les 3 premières lactations des données des contrôles laitiers corrigées pour les effets de milieu. Les analyses de détection de QTL ont été effectuées intra-race en utilisant une approche LDLA combinant analyse de liaison et analyse du déséquilibre de liaison. De très nombreux QTL ont été détectés dans chacune des races. Certains QTL, retrouvés dans les 3 races étudiées, présentent des effets extrêmement significatifs ( $P < 10^{-15}$  après correction de Bonferroni). Ils se situent sur les chromosomes (BTA) 6 (2 régions QTL), 11 et 20. Sur BTA6, la région comprise entre 37 et 46Mb affecte le taux de caséine  $\alpha$ s1 en %prot, celle autour de 87Mb a notamment des effets sur les taux de caséines  $\alpha$ s1,  $\alpha$ s2,  $\beta$  et  $\kappa$  en %lait ou en %prot. La région QTL située autour de 103Mb sur BTA11 a un effet sur les taux de caséines  $\alpha$ s1,  $\beta$  et  $\kappa$  et surtout sur le taux de  $\beta$ -lactoglobuline en %prot. Le taux d' $\alpha$ -lactalbumine en %lait et surtout en %prot est affecté par une région localisée à 58-59Mb sur BTA20. Une région QTL ayant des effets sur de nombreux caractères exprimés surtout en %lait a également été identifiée en races NO et HO sur BTA14 (1,7-1,8Mb). Ces QTL expliquent de 5 (BTA14) à 51% (BTA11) de la variance génétique du caractère. D'autres QTL très significatifs sont également mis en évidence en race NO sur BTA2 et 13, HO sur BTA20 (32Mb) et MO sur BTA25 et 29. Enfin, de nombreuses autres régions atteignant des seuils un peu moins significatifs sont également détectées. Les gènes des caséines (BTA6-87Mb), de la  $\beta$ -lactoglobuline (BTA11), *DGAT1* (BTA14) et *GHR* (BTA20-32Mb) apparaissent comme de bons candidats pour certains de ces QTL. Ces résultats constituent une première étape favorable pour identifier les mutations causales et mettre en place une sélection génomique pour la composition protéique du lait.

### QTL for major milk proteins in three French dairy cattle breeds

SANCHEZ M.P. (1), GOVIGNON-GION A. (1), M. FERRAND M. (2), GELE M. (2), POURCHET D. (3), ROSSIGNOL M.N. (4), FRITZ S. (5), MIRANDA G. (1), MARTIN P. (1), BROCHARD M. (2), BOICHARD D. (1)  
(1) INRA - UMR1313 GABI, 78350 Jouy en Josas

### SUMMARY

A genome wide analysis was performed in the frame of the French PhénoFinLait project to detect QTL affecting milk protein composition estimated from Mid-Infrared spectrometry in the three main French dairy cattle breeds Montbéliarde (MO), Normande (NO) and Holstein (HO). Composition of the main milk proteins ( $\alpha$ -lactalbumin,  $\beta$ -lactoglobulin and  $\alpha$ s1,  $\alpha$ s2,  $\beta$  et  $\kappa$ -caseins) expressed in g/100g of milk (%milk) or in g/100g of proteins (%prot) was estimated on almost 900,000 test-day milk samples from 160,253 cows. Genotyping was performed on 2,773 MO, 2,673 NO and 2,208 HO cows with the Illumina Bovine 50k Beadchip. Individual test-day records of the first three lactations were adjusted for environmental effects and then averaged per cow to define the analyzed phenotypes. QTL detection was carried out within breed by a LDLA (Linkage Disequilibrium and Linkage Analysis) approach. Very numerous QTL were detected in each breed. Some of the QTL regions, found in the 3 breeds, present highly significant effects ( $P < 10^{-15}$  after Bonferroni correction). They were located on chromosomes (BTA) 6 (2 QTL regions), 11 and 20. On BTA6, the region spanning from 37 to 46Mb affected  $\alpha$ s1-casein in %prot while the 87Mb region affected  $\alpha$ s1,  $\alpha$ s2,  $\beta$  and  $\kappa$ -caseins in %prot or %milk. The 103Mb region on BTA11 affected  $\alpha$ s1,  $\beta$  and  $\kappa$ -caseins and particularly  $\beta$ -lactoglobulin in %prot. The proportion of  $\alpha$ -lactalbumin in %milk and especially in %prot was affected by a 58-59Mb region on BTA20. A QTL region affecting numerous traits essentially expressed in %milk was also identified in NO and HO breeds on BTA14 (1.7-1.8Mb). These QTL explained from 5 (BTA14) to 51% (BTA11) of the total genetic variance of the trait. Other QTL very significant were also highlighted in NO (BTA2, BTA13), HO (BTA20-32Mb) and MO (BTA25, BTA29) animals. Finally, numerous other regions with less significant thresholds were also detected. Caseins (BTA6-87Mb),  $\beta$ -lactoglobulin (BTA11), *DGAT1* (BTA14) and *GHR* (BTA20-32Mb) genes are good candidates for some of these QTL. These results are thus a first favorable step in order to identify causal mutations and to include protein composition in future genomic selection programs.

## INTRODUCTION

Le programme PhénoFinLait, initié en 2008, a pour objectif d'étudier la composition fine du lait en acides gras et en protéines, estimée à partir des spectres du moyen infrarouge (MIR) pour les trois espèces bovine, ovine et caprine. Si la composition en acides gras influence surtout les qualités nutritionnelles du lait, *via* notamment le rapport acides gras saturés / insaturés, la composition protéique joue essentiellement un rôle sur les propriétés technofonctionnelles du lait, en particulier sur son aptitude à coaguler, le rendement fromager et la stabilité à la chaleur. Suite au développement des équations de prédiction à partir des spectres MIR et à la mise en place du recueil des données à grande échelle, le déterminisme génétique de ces caractères a pu être étudié. Les résultats de la composition en acides gras pour l'espèce bovine ont fait l'objet d'une synthèse à ces mêmes journées (Govignon-Gion et al., 2012). Le même type d'analyses a été réalisé sur les taux protéiques dans les 3 principales races bovines laitières françaises Montbéliarde (MO), Normande (NO) et Holstein (HO). Nous avons dans un premier temps mis en évidence l'existence d'une composante génétique forte pour ces caractères (Brochard et al., 2013) et nous présentons ici les résultats des analyses de détection de QTL.

## 1. MATERIEL ET METHODES

### 1.1. ANIMAUX ET CARACTERES

Près de 900 000 échantillons de lait ont été prélevés sur 160 253 vaches des trois races MO, NO et HO entre novembre 2009 et août 2012. La composition protéique du lait a été estimée par analyse des spectres MIR *via* des équations de prédiction dont la qualité peut être appréciée par le coefficient de détermination  $R^2$  et l'erreur relative. Les taux des protéines majeures, caséines (CAS)  $\alpha 1$ ,  $\alpha 2$ ,  $\beta$  et  $\kappa$  et des protéines sériques (PS)  $\alpha$ -lactalbumine ( $\alpha$  lact) et  $\beta$ -lactoglobuline ( $\beta$  lact) ont été calculés en g/100g de lait (%lait) et en g/100g de protéines (%prot). Le taux de protéines totales dans le lait (TP) et le taux de dégradation des protéines (protéolyse : PT) ont également été estimés. Les 14 caractères ont été analysés intra-race (Tableau 1).

### 1.2. CORRECTION POUR LES EFFETS DE MILIEU

Les données retenues pour les analyses sont issues des contrôles élémentaires des trois premières lactations. Seules les vaches ayant au moins trois contrôles par lactation ont été conservées pour les analyses. Un modèle appliqué aux données des contrôles élémentaires a été utilisé pour corriger les données. Pour un caractère donné, il suppose que les performances de chaque contrôle élémentaire sont l'expression d'un même caractère répété. Ce modèle inclut les effets fixés des combinaisons troupeau x jour de contrôle, rang x stade de lactation, mois x année de vêlage, laboratoire x machine d'analyse x mois de contrôle ainsi que les effets aléatoires de la valeur génétique de la vache, de l'environnement permanent et de la valeur résiduelle. Pour chacune des vaches, la moyenne des données pré-corrigées pour tous les effets fixés et pour l'effet de l'environnement permanent a ensuite été calculée.

**Tableau 1** Taux protéiques exprimés en % de lait ou en % de protéines dans les 3 races : moyennes  $\pm$  écart-types et qualité de la prédiction MIR ( $R^2$  = coefficient de détermination et Erreur relative)

| Race                | Unité | TP            | PT             | $\alpha$ lact   | $\beta$ lact    | CAS $\alpha 1$  | CAS $\alpha 2$  | CAS $\beta$     | CAS $\kappa$    |
|---------------------|-------|---------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| MO<br>(n=589 016)   | %lait | 3,4 $\pm$ 0,4 | -              | 0,14 $\pm$ 0,02 | 0,28 $\pm$ 0,05 | 0,94 $\pm$ 0,10 | 0,32 $\pm$ 0,04 | 1,24 $\pm$ 0,11 | 0,33 $\pm$ 0,05 |
|                     | %prot | -             | 10,1 $\pm$ 2,3 | 4,07 $\pm$ 0,28 | 8,25 $\pm$ 1,12 | 27,8 $\pm$ 0,55 | 9,53 $\pm$ 0,30 | 36,6 $\pm$ 0,88 | 9,75 $\pm$ 0,60 |
| NO<br>(n=117 323)   | %lait | 3,6 $\pm$ 0,4 | -              | 0,15 $\pm$ 0,02 | 0,28 $\pm$ 0,05 | 0,99 $\pm$ 0,10 | 0,35 $\pm$ 0,04 | 1,29 $\pm$ 0,11 | 0,35 $\pm$ 0,05 |
|                     | %prot | -             | 10,4 $\pm$ 2,5 | 4,16 $\pm$ 0,36 | 7,94 $\pm$ 1,03 | 27,8 $\pm$ 0,68 | 9,89 $\pm$ 0,33 | 36,2 $\pm$ 1,2  | 9,87 $\pm$ 0,48 |
| HO<br>(n=150 285)   | %lait | 3,3 $\pm$ 0,4 | -              | 0,14 $\pm$ 0,02 | 0,28 $\pm$ 0,05 | 0,92 $\pm$ 0,11 | 0,32 $\pm$ 0,04 | 1,20 $\pm$ 0,13 | 0,31 $\pm$ 0,05 |
|                     | %prot | -             | 11,7 $\pm$ 2,6 | 4,27 $\pm$ 0,42 | 8,46 $\pm$ 1,17 | 27,9 $\pm$ 0,69 | 9,69 $\pm$ 0,39 | 36,2 $\pm$ 1,2  | 9,43 $\pm$ 0,58 |
| $R^2$               |       | 1             | 0,54           | 0,59            | 0,74            | 0,88            | 0,82            | 0,92            | 0,80            |
| Erreur relative (%) |       | 0,73          | 22,2           | 14,36           | 11,7            | 4,7             | 7,5             | 3,7             | 8,4             |

## 1.3. GENOTYPAGES

Une prise de sang a été réalisée sur environ 8 000 vaches du programme PhénoFinLait et l'ADN a été extrait à Labogena. Sur l'ensemble de ces vaches, 7 500 ont été génotypées pour la puce bovine Illumina 50K et 500 pour la puce bovine Illumina LD. Les génotypages ont été validés, imputés (de la puce LD à la puce 50K notamment) et les phases parentales ont été reconstituées (carte UMD3.1) comme dans Govignon-Gion et al (2012). Un contrôle qualité des données des puces a ensuite été réalisé pour éliminer les marqueurs SNP (« Single Nucleotide Polymorphism ») à faible polymorphisme ou présentant des problèmes techniques de génotypage. Après élimination des vaches ayant moins de 3 contrôles par lactation et des SNP n'ayant pas passé le contrôle qualité, 7 654 vaches (2 773 MO, 2 673 NO et 2 208 HO) et 36 912 SNP en MO, 37 363 SNP en NO et 39 683 SNP en HO ont été utilisés pour les analyses de détection de QTL.

## 1.4. ANALYSES DE DETECTION DE QTL

### 1.4.1. Méthode LDLA

La méthode appliquée pour détecter des associations entre les SNP et les phénotypes (QTL) est la méthode dite LDLA (« Linkage Disequilibrium and Linkage Analysis » ; Meuwissen et Goddard, 2001) qui exploite à la fois l'information apportée par les structures familiales (liaison SNP/QTL propre à la famille) et par la population (déséquilibre de liaison SNP/QTL lié à l'histoire de la population étudiée). Une telle méthode est généralement plus puissante qu'une méthode d'analyse de liaison ou d'analyse d'association pure. L'approche utilisée teste en outre l'effet de chaque SNP en considérant des clusters de 6 SNP (le SNP testé + les 5 SNP voisins), ce qui permet de prendre en compte le déséquilibre de liaison entre les SNP consécutifs et maximise ainsi l'informativité des marqueurs.

Pour chaque SNP, on compare les vraisemblances des données sous deux hypothèses :

- H0 : le SNP n'a pas d'effet sur le caractère
- H1 : le SNP a un effet sur le caractère

Le calcul du rapport des vraisemblances sous les deux hypothèses (LRT = statistique de test) permet de conclure ou non à l'existence d'un QTL à la position du SNP testé. Pour chaque QTL détecté, on peut évaluer l'effet du QTL en calculant la part de variance génétique (VG) expliquée par le QTL :  $\%VG = \sigma_g^2 / (\sigma_u^2 + \sigma_g^2)$ , avec  $\sigma_u^2$ , la variance polygénique estimée sous H0 et  $\sigma_g^2$ , la variance due au QTL estimée sous H1.

### 1.4.2. Seuils de détection des QTL

Pour conclure quant à l'existence ou non d'un QTL à une position donnée, on compare la valeur du LRT calculé à une valeur seuil critique théorique. Si l'on considère que la distribution théorique du LRT est une loi du  $\chi^2$  à 2 degrés de liberté, pour une valeur seuil de 13, le risque de conclure à l'existence d'un QTL alors qu'il n'existe pas (*i.e.* le risque de retenir un faux-positif) est de  $1,5 \cdot 10^{-3}$ . Ce calcul de probabilité ne prend toutefois pas en compte le nombre de tests effectués, très élevé pour ce type d'analyses, qui multiplie d'autant le risque de détecter des faux-positifs.

Nous avons donc appliqué la correction de Bonferroni qui, pour une valeur seuil du LRT, permet de calculer le nombre de faux-positifs attendus en fonction du nombre de tests indépendants réalisés en considérant que nous avons 42 000 tests indépendants (3 000 SNP x 14 caractères). Trois types de QTL ont été définis en fonction du nombre de faux-positifs (FP) attendus et donc du LRT maximum (LRTmax = valeur du LRT maximum obtenue pour une région donnée) :

- QTL *suggestifs* si  $13 \leq \text{LRTmax} < 20$  ( $2 < \text{FP} < 63$ ) ;
- QTL *significatifs* si  $20 \leq \text{LRTmax} < 50$  ( $0 < \text{FP} < 2$ ) ;
- QTL *très significatifs* si  $\text{LRTmax} \geq 50$  ( $\text{FP} \sim 0$ ).

Tous les SNP espacés de plus de 10Mb ont été considérés comme appartenant à des régions QTL différentes.

### 1.4.3. Recherche de gènes candidats

Dans les régions présentant les QTL les plus significatifs, des gènes candidats ont été recherchés à l'aide de l'outil Ensembl : [http://www.ensembl.org/Bos\\_taurus/Info/Index](http://www.ensembl.org/Bos_taurus/Info/Index) (*Bos Taurus* UMD3.1).

## 2. RESULTATS

### 2.1. REGIONS QTL DETECTES

Le nombre de résultats *suggestifs* à *très significatifs* ( $\text{LRT} \geq 13$ ) est égal à 172 en race MO, 174 en race NO et 136 en race HO. A ce niveau de seuil, on retrouve des QTL sur quasiment tous les chromosomes dans toutes les races, excepté sur BTA1 et BTA25 en race MO, BTA12, BTA18 et BTA25 en race NO et enfin, BTA7 en race HO (Figure 1). Au-delà d'un seuil de  $\text{LRT} \geq 20$ , le nombre de résultats *significatifs* ou *très significatifs* est à peu près équivalent dans les trois races (67 à 81). En race MO, NO et HO : 24, 21 et 13 régions QTL *significatives* (définies en regroupant les SNP avec un LRT compris entre 20 et 50) sont identifiées sur 16, 17 et 12 chromosomes différents, respectivement. Dix régions QTL sont *très significatives* ( $\text{LRT} \geq 50$  ; Tableau 2). Parmi elles, les 4 régions qui présentent les effets les plus significatifs sont retrouvées dans les 3 races.

- La région **BTA6a** (~37-46Mb) affecte la CAS  $\alpha 1$ . Le SNP le plus significatif de cette région présente des caractéristiques un peu différentes en race HO par rapport aux 2 autres races : il se situe environ 9Mb en aval et son effet est plus important (environ 30% de la VG contre environ 7% en race MO et 10% en race NO).

- La région **BTA6b** (~87Mb) a un effet sur de nombreux caractères surtout en race MO et NO, le caractère le plus affecté étant la CAS  $\kappa$  dans les trois races (16 à 20% de la VG en %lait en MO et en %prot en NO et HO).
- **BTA11** (~103Mb) est la région QTL la plus significative avec un LRT qui atteint 1866 en race MO. Plusieurs caractères exprimés surtout en %prot sont affectés par cette région, les CAS  $\alpha 1$ ,  $\beta$  et  $\kappa$  et surtout la  $\beta$  lact (41 à 51% de la VG selon la race).
- **BTA20b** (~58-59Mb) est la deuxième région QTL la plus significative, elle affecte l' $\alpha$  lact en %lait et surtout en %prot (25 à 32% de la VG selon la race).

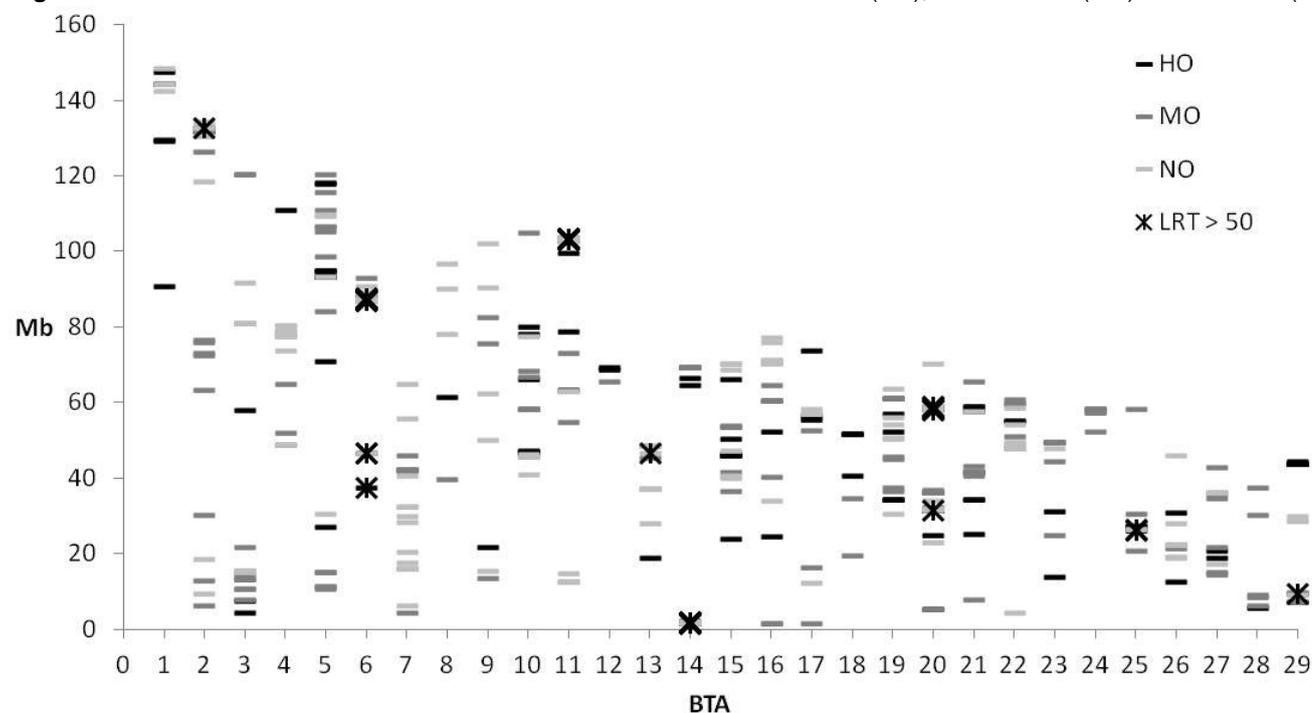
Une autre région se distingue également sur **BTA14** (~1,7-1,8Mb). Elle est *très significative* dans les races NO et HO avec des effets sur de nombreux caractères exprimés essentiellement en %lait (TP,  $\alpha$  lact et toutes les caséines) et *significative* en race MO sur la CAS  $\alpha 2$  en %prot. Les effets de cette région sont toutefois un peu plus modérés que les effets des 4 régions décrites plus haut (3 à 11% de la VG selon le caractère et la race). Enfin, 5 autres régions détectées dans une seule race ont des effets *très significatifs* sur un caractère : **BTA2** (CAS  $\alpha 2$  en %prot) et **BTA13** (PT) en race NO, **BTA20a** (CAS  $\kappa$  en %lait) en race HO et **BTA25** (PT) et **BTA29** (CAS  $\alpha 1$  en %prot) en race MO.

### 2.2. GENES CANDIDATS

Des gènes ont été recherchés dans les régions *très significatives* et plusieurs gènes candidats ont pu être identifiés dans les 4 régions décrites ci-dessous.

- **BTA6b** : les gènes **CSN1S1**, **CSN2**, **CSN1S2** et **CSN3** localisés entre 87,141 et 87,393Mb sur BTA6 codent respectivement pour les caséines  $\alpha 1$ ,  $\beta$ ,  $\alpha 2$  et  $\kappa$ .
- **BTA11** : le gène codant pour la  $\beta$  lact (**LGB**) est situé entre 103,301 et 103,306Mb, à proximité immédiate des SNP qui présentent les résultats les plus significatifs dans cette région. Le même SNP a un niveau de LRT très élevé dans les 3 races ( $1\,563 < \text{LRT} < 1\,866$ ). Il se situe à 103,289Mb, soit un peu en amont du gène **LGB**.
- **BTA14** : les SNP qui présentent des effets très significatifs dans cette région se situent à proximité ou dans le gène **DGAT1** localisé entre 1,795 et 1,805Mb.
- **BTA20a** : un peu plus de 300kb en aval du QTL détecté dans cette région se trouve le gène **GHR** (31,890-32,200 Mb).

**Figure 1** Positions des SNP avec un  $\text{LRTmax} \geq 13$  dans les trois races Holstein (HO), Montbéliarde (MO) et Normande (NO)



**Tableau 2** Caractéristiques des 10 régions QTL les plus significatives (LRTmax≥50) ; les éléments en gras correspondent aux caractéristiques du SNP le plus significatif de la région

| BTA | Race | Positions SNP flanquant (Mb) | Caractères <sup>a</sup> affectés par la région QTL |          |           |        |           |          |    |    | LRTmax <sup>b</sup> | %VG <sup>b</sup>   | P <sup>c</sup>    |
|-----|------|------------------------------|--|----------|-----------|--------|-----------|----------|----|----|---------------------|--------------------|-------------------|
|     |      |                              | TP   | PT       | PS        |        | Caséines  |          |    |    |                     |                    |                   |
|     |      |                              |  |          | α lact    | β lact | αs1       | αs2      | β  | κ  |                     |                    |                   |
| 2   | NO   | <b>132,603</b>               |  |          |           |        |           | <b>P</b> |    |    | <b>55</b>           | <b>7,3</b>         | <b>5E-08</b>      |
| 6a  | MO   | <b>46,681</b>                |  |          |           |        | <b>P</b>  |          |    |    | <b>86</b>           | <b>6,6</b>         | <b>9E-15</b>      |
| 6a  | NO   | <b>46,600</b>                |  |          |           |        | <b>P</b>  |          |    |    | <b>86</b>           | <b>10,0</b>        | <b>9E-15</b>      |
| 6a  | HO   | <b>37,653</b>                |  |          |           |        | <b>P</b>  |          |    |    | <b>102</b>          | <b>30,7</b>        | <b>3E-18</b>      |
| 6b  | MO   | <b>87,022</b> - 87,601       | L  | P        | L         | P      | L         | LP       | LP | LP | 93 - <b>310</b>     | 4,8 - <b>16,9</b>  | <b>1E-63</b>      |
| 6b  | NO   | <b>86,905</b> - 87,366       | L  |          | L         |        | L         | LP       | LP | LP | 104 - <b>230</b>    | 6,6 - <b>16,0</b>  | <b>3E-46</b>      |
| 6b  | HO   | <b>87,366</b> - 87,716       |  | P        |           |        |           |          |    | LP | 56 - <b>281</b>     | 9,5 - <b>20,3</b>  | <b>2E-57</b>      |
| 11  | MO   | 103,265 - 103,410            |  |          |           |        | <b>LP</b> | P        |    | P  | 85 - <b>1866</b>    | 6,8 - <b>46,3</b>  | <b>&lt;1E-309</b> |
| 11  | NO   | 103,650 - <b>103,289</b>     |  |          |           |        | <b>LP</b> | P        |    | P  | 62 - <b>1752</b>    | 6,9 - <b>41</b>    | <b>&lt;1E-309</b> |
| 11  | HO   | 103,165 - <b>103,289</b>     |  |          |           |        | <b>LP</b> |          |    | P  | 68 - <b>1563</b>    | 8,4 - <b>51,3</b>  | <b>&lt;1E-309</b> |
| 13  | NO   | <b>46,723</b>                |  | <b>P</b> |           |        |           |          |    |    | <b>63</b>           | <b>7,1</b>         | <b>1.10-9</b>     |
| 14  | NO   | <b>1,801</b>                 | L  |          | L         |        | L         | L        | L  | LP | 53 - <b>117</b>     | 5,2 - <b>11,0</b>  | <b>1E-21</b>      |
| 14  | HO   | <b>1,651</b> - 1,801         | L  | P        | LP        |        | <b>L</b>  | L        | L  | L  | 57 - <b>185</b>     | 5,6 - <b>8,8</b>   | <b>2E-36</b>      |
| 20a | HO   | <b>31,569</b>                |  |          |           |        |           |          |    | L  | <b>51</b>           | <b>7,9</b>         | <b>5.10-7</b>     |
| 20b | MO   | <b>58,241</b>                |  |          | <b>LP</b> |        |           |          |    |    | 58 - <b>404</b>     | 13,2 - <b>31,8</b> | <b>4E-84</b>      |
| 20b | NO   | 58,027 - <b>59,017</b>       |  |          | <b>LP</b> |        |           |          |    |    | 92 - <b>364</b>     | 9,3 - <b>25,0</b>  | <b>2E-75</b>      |
| 20b | HO   | <b>58,293</b>                |  |          | <b>LP</b> |        |           |          |    |    | 77 - <b>392</b>     | 7,6 - <b>25,7</b>  | <b>2E-81</b>      |
| 25  | MO   | <b>26,489</b>                |  | <b>P</b> |           |        |           |          |    |    | <b>63</b>           | <b>7,1</b>         | <b>1.10-9</b>     |
| 29  | MO   | <b>9,341</b>                 |  |          |           |        | <b>P</b>  |          |    |    | <b>66</b>           | <b>4,9</b>         | <b>2.10-10</b>    |

<sup>a</sup> le QTL a un effet sur le caractère exprimé en %prot (P) ou en %lait (L) ; <sup>b</sup> Pour une région, valeurs extrêmes des LRTmax et de l'effet du QTL des caractères affectés (%VG : % variance génétique) <sup>c</sup> Probabilité la plus faible de la région calculée après correction de Bonferroni

### 3. DISCUSSION ET CONCLUSION

De très nombreux QTL ont pu être identifiés grâce au dispositif PhénoFinLait. Certains de ces QTL présentent des seuils de significativité extrêmement élevés, sans doute parce que les caractères d'intérêt sont des taux protéiques et que la protéine est le produit direct des gènes. Les valeurs d'héritabilité estimées dans ces mêmes populations confirment la composante génétique très forte des taux protéiques ( $0,27 < h^2 < 0,86$  ; Brochard et al., 2013).

Les protéines majeures du lait sont au nombre de 6. Nous avons identifié des QTL dans les régions des gènes codant pour 5 de ces protéines (cluster des gènes des 4 caséines sur BTA6 et gène de la β lact sur BTA11). Dans la région du gène de l'α lact (*ALBA*) qui se situe sur BTA5 (~31Mb), aucun QTL n'a été détecté. Il est possible que la qualité de la prédiction du taux d'α lact par spectrométrie MIR ne permette pas de détecter l'effet de ce gène ( $R^2=0,59$  et Erreur relative=14,4%).

Les résultats que nous obtenons avec les taux protéiques estimés par les spectres MIR sont plutôt cohérents avec ceux de Schopen et al. (2011) obtenus sur des taux protéiques estimés par une méthode directe (électrophorèse). Comme eux, nous mettons en évidence des régions QTL sur BTA6, 11 et 14. En revanche, nous n'avons pas retrouvé le QTL avec un effet sur l'α lact sur BTA5, probablement pour les raisons évoquées plus haut. Par ailleurs, nous mettons en évidence de nouvelles régions, dont 2 sur BTA20, qui ont un gros effet sur la composition protéique du lait. Une de ces régions se trouve à proximité du gène *GHR* dans lequel un polymorphisme a été associé au TP (Blott et al., 2003).

Plusieurs gènes candidats sont donc d'ores et déjà identifiés dans les régions BTA6b, BTA11, BTA14 et BTA20a. Dans les autres régions QTL, des gènes n'ont pour le moment pas pu être mis en évidence. Les régions BTA6a et BTA20b méritent tout particulièrement d'être étudiées plus finement en raison de leurs effets très forts sur la composition protéique dans les 3 races. Pour toutes ces régions, la détermination du statut au QTL (homozygote / hétérozygote) des taureaux combinée

à l'analyse de leurs séquences devrait pouvoir nous permettre d'identifier les mutations causales.

Ces résultats, ainsi que ceux des estimations des paramètres génétiques dans ces mêmes populations montrent qu'il est tout à fait possible de modifier la composition protéique du lait de vache par sélection dans les trois principales races laitières françaises en utilisant les données des spectres MIR collectées en routine. En particulier, l'allèle ou l'haplotype favorable situé sur BTA11 pourrait être sélectionné en vue de diminuer le taux de β lact, et par voie de conséquence, d'augmenter la proportion de caséines dans le lait de vache pour améliorer les aptitudes fromagères. La population PhénoFinLait, même si elle est constituée de moins de 3 000 animaux par race, représente dès à présent une population de référence de base pour la sélection génomique, grâce à l'héritabilité élevée des caractères. Avec l'augmentation de la disponibilité des spectres MIR et du génotypage des vaches, la taille de cette population va rapidement s'accroître, ce qui devrait permettre d'estimer dans un avenir proche des valeurs génomiques très précises dans chacune des races MO, NO et HO.

*PhénoFinLait a été financé par l'ANR, Apis-Gène, le ministre de l'agriculture (CASDAR), le CNIEL, FranceAgriMer et FGE.*

**Blott S., Kim J.J., Moisiso S., Schmidt-Küntzel A., Cornet A., Berzi P., Cambisano N., Ford C., Grisart B., Johnson D., Karim L., Simon P., Snell R., Spelman R., Wong J., Viikki J., Georges M., Farnir F., Coppieters W. 2003.** *Genetics*, 163, 253-266  
**Brochard M., Sanchez M.P., Govignon-Gion A., Ferrand M., Gelé M., Pourchet D., Miranda G., Martin P., Boichard D. 2013.** *Renc. Rech. Ruminants*, 20, ces actes.  
**Govignon-Gion A., Fritz S., Larroque H., Chantry-Darmon C., Lahalle F., Brochard M., Boichard D. 2012.** *Renc. Rech. Ruminants*, 19, 65-72.  
**Meuwissen T.H.E., Goddard M.E. 2001.** *Genet. Sel. Evol.*, 33, 605-634.  
**Schopen G.C.B., Visker M.H.P.W., Koks P.D., Mullaart E., Van Arendonk J.A.M., Bovenhuis H. 2011.** *J. Dairy. Sci.*, 94, 3148-315.

