

Utilisation des résultats de cartographie fine de QTL en sélection chez les bovins laitiers

FRITZ S. (1), GUILLAUME F. (2), TARRES J. (3), BAUR A. (1,3), BOUSSAHA M. (3), BOSCHER M.Y. (4), JOURNAUX L. (1), MALAFOSSE A. (1), GAUTIER M. (3), COLLEAU J.J. (3), EGGEN A. (3), BOICHARD D. (3)

(1) Union nationale des coopératives agricoles d'élevage et d'insémination animale - service génétique - 149 rue de Bercy - 75595 Paris Cedex 12 - France

(2) Institut de l'élevage - 149 rue de Bercy - Paris - France

(3) INRA génétique animale et biologie intégrative - équipe génétique et génomique bovin - 78352 Jouy-en-Josas cedex - France

(4) GIE LABOGENA - 78352 Jouy-en-Josas cedex - France

RESUME - Les génotypages de 2 837 taureaux sur 54 001 marqueurs moléculaires ont permis de réaliser des travaux de cartographie fine de QTL (région chromosomique ayant un effet significatif sur un caractère quantitatif) sur vingt caractères (liés à la production laitière, à la fertilité des femelles, au score cellulaire, à la morphologie des animaux, à la locomotion, à la vitesse de traite, aux facilités de naissance et de vêlage et à la mortinatalité) dans les trois principales races bovines laitières françaises. Les résultats de ces travaux ont permis de faire évoluer la sélection assistée par marqueurs en octobre 2008. La précision (CD) des valeurs génétiques estimées avec cette nouvelle méthodologie est uniquement dépendante de la part de variance génétique expliquée par les QTL. Les CD des index des caractères évalués en SAM pour un jeune animal sans performances seront proches de 0,50 dès octobre 2008, estimation confirmée sur un premier jeu de données réelles. Une nouvelle évolution en 2009 devrait encore augmenter les CD des index SAM. Ces gains de précision vont entraîner de profondes évolutions dans l'organisation et le déroulement des programmes de sélection.

Use of QTL fine mapping results for marker-assisted selection in dairy cattle

FRITZ S. (1), GUILLAUME F. (2), TARRES J. (3), BAUR A. (1,3), BOUSSAHA M. (3), BOSCHER M.Y. (4), JOURNAUX L. (1), MALAFOSSE A. (1), GAUTIER M. (3), COLLEAU J.J. (3), EGGEN A. (3), BOICHARD D. (3)

SUMMARY - Genotypes of 2837 bulls for 54001 molecular markers are the basis of a QTL (chromosomal regions with a significant effect on a quantitative trait) fine-mapping project involving 20 traits (dairy traits, female post partum fertility, somatic cell score, type, locomotion, milking ease, calving conditions, stillbirth) in the three main French dairy breeds. The results have been used to update the French Marker-Assisted Selection (MAS) program in October 2008. Reliability of estimated breeding values obtained through this MAS approach is only dependent on the proportion of genetic variance explained by these QTL. For a young animal without performances nor progeny test, the reliability is close to 0.50, as assessed by a first validation data set. Another large evolution of the approach is planned during the year 2009 and should even increase the reliability of the MAS estimated breeding values. With such an accuracy, we can expect dramatic changes in the management of the breeding schemes.

INTRODUCTION

Un programme français de sélection assistée par marqueurs (SAM) conduit par l'INRA, LABOGENA et l'UNCEIA a été mis en place depuis 2001 dans les trois principales races bovines laitières françaises (Montbéliarde, Normande et Holstein) afin d'optimiser les choix des jeunes animaux sans performances dans les étapes précoces de sélection. (Fritz *et al.*, 2003). Jusqu'à présent, ce programme s'appuyait sur des régions QTL (région chromosomique ayant un effet significatif sur un caractère quantitatif d'intérêt) de très grande taille puisqu'elles contiennent plusieurs centaines de gènes. Mais les progrès technologiques récents en matière de génotypages et de cartographie du génome bovin ouvrent de nouvelles perspectives tant au niveau de la recherche que de l'utilisation de ces informations moléculaires en sélection (Fritz *et al.*, 2007).

1. CARTOGRAPHIE FINE DE QTL CHEZ LES BOVINS LAITIERS.

Le projet CARTOFINE soutenu par l'ANR et par APIS-GENE dans le cadre du GIS AGENAE prévoit des travaux de cartographie fine de QTL à partir du génotypage de 3 200 taureaux issus du programme SAM (2001-2007) au centre national de génotypages d'Evry sur la puce Illumina

« BovineSNP50 » contenant 54 001 marqueurs SNP (*Single Nucleotide Polymorphism*). Les SNP sont des mutations ponctuelles du génome pouvant être utilisées comme des marqueurs bi-alléliques. Ces marqueurs sont particulièrement intéressants du fait de leur grande abondance et des possibilités d'automatisation permettant d'acquérir plusieurs dizaines de milliers de génotypes à la fois.

1.1. RESULTATS DE GENOTYPAGES UTILISES

Au premier semestre 2008, les génotypages des 2837 premiers taureaux (1575 Holstein, 661 Normands et 601 Montbéliards) sur les puces Illumina étaient disponibles pour les analyses statistiques.

Sur les 54 001 marqueurs SNP génotypés, 51 326 (95 %) ont été positionnés grâce à la dernière carte bovine disponible, les marqueurs non positionnés ne pouvant pas être utilisés dans cette première série d'analyses. Parmi ces marqueurs, seuls les SNP avec une fréquence de l'allèle rare supérieure à 5 % ont été conservés pour une race donnée, soit respectivement 40757 SNP (79,4 %), 38885 SNP (75,8 %) et 39004 SNP (76,0 %) en race Holstein, Normande et Montbéliarde. Cela représente une densité moyenne de marqueurs informatifs de l'ordre de 15 SNP par Mégabase (Mb), autrement dit 15 SNP informatifs couvrent en moyenne une région contenant environ une dizaine de gènes.

1.2. DETECTION DES QTL

Dans un premier temps, des analyses statistiques simples (régressions intra père) permettant de primo-localiser des QTL ont été réalisées sur vingt caractères d'intérêt pour chacune des trois races. Ces caractères sont liés à la production laitière des vaches (5), aux cellules somatiques (1), à la fertilité femelle (2), à la morphologie de la mamelle (4), à la capacité corporelle (2), à la locomotion (1), à la vitesse de traite (1), aux facilités de naissance et de vêlage (2) et à la mortinatalité (2).

Tableau 1 : localisation des QTL (en centi-Morgans) avec les tests les plus significatifs pour chaque groupe de caractères étudiés en cartographie fine en race Holstein (HO), Normande (NO) et montbéliarde (MO)

Groupe de caractères	Race	Chromosomes (cM)
Production laitière	HO	5 (112 cM), 6 (100 cM), 14 (1 cM), 20 (36 cM)
	NO	1 (148 cM), 6 (88 cM), 13 (80 cM), 29 (6 cM)
	MO	1 (70 cM), 6 (105 cM), 14 (1 cM), 19 (44 cM)
Cellules (CEL)	HO	3 (140 cM), 9 (129 cM), 14 (17 cM), 15 (58 cM), 29 (44 cM)
	NO	2 (160 cM), 4 (75 cM), 8 (100 cM), 12 (92 cM)
	MO	3 (60 cM), 6 (108 cM), 12 (70 cM), 16 (86 cM)
Fertilité femelle (FER)	HO	1 (8 cM), 2 (18 cM), 14 (42 cM), Fertilité18 (68 cM)
	NO	2 (18 cM), 4 (69 cM), 7 (92 cM), 20 (70 cM)
	MO	1 (68 cM), 10 (45 cM), 18 (55 cM), 19 (64 cM)
Morphologie mamelle	HO	7 (60 cM), 11 (25 cM), 18 (50 cM), 20 (44 cM)
	NO	10 (105 cM), 14 (71 cM), 20 (18 cM), 25 (18 cM)
	MO	5 (72 cM), 7 (55 cM), 8 (52 cM), 13 (100 cM)
Capacité Corporelle	HO	1 (110 cM), 3 (89 cM), 14 (7 cM), 21 (2 cM)
	NO	4 (130 cM), 6 (70 cM), 20 (8 cM), 28 (30 cM)
	MO	4 (18 cM), 12 (69 cM), 20 (2 cM), 24 (58 cM)
Locomotion	HO	8 (107 cM), 18 (7 cM), 25 (20 cM)
	NO	4 (40 cM), 9 (62 cM), 16 (53 cM), 19 (30 cM)
	MO	7 (92 cM), 18 (77 cM), 19 (1 cM), 28 (51 cM)
Vitesse de Traite	HO	4 (132 cM), 10 (24 cM), 15 (43 cM), 20 (45 cM)
	NO	5 (80 cM), 11 (32 cM), 19 (29 cM), 23 (11 cM)
	MO	4 (88 cM), 5 (60 cM), 14 (5 cM), 28 (22 cM)

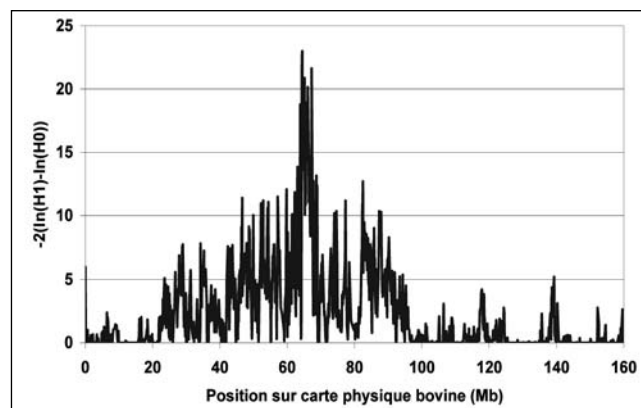
Ces analyses ont permis de détecter respectivement 342 QTL en race Holstein, 282 QTL en race Normande et 301 QTL en race Montbéliarde au seuil de 5 % à l'échelle du chromosome. Parmi tous ces QTL, on retrouve les QTL actuellement sélectionnés dans le cadre du programme SAM. A ce stade, il est toutefois important de noter que pour chaque race, avec un risque de 5 % par analyse et en tenant compte de la multiplicité des tests, on est en droit d'attendre vingt-neuf résultats faux positifs parmi tous ces QTL détectés.

Dans les régions mises ainsi en évidence, des travaux de cartographie fine de QTL (analyses intégrant le déséquilibre de liaison, Meuwissen *et al.*, 2000) ont été réalisés. On parle de marqueurs en déséquilibre de liaison avec le QTL si une même combinaison de marqueurs SNP (haplotype de SNP) est préférentiellement associée au même allèle du QTL dans l'ensemble de la population. Les outils statistiques et informatiques nécessaires à ces travaux ont été développés dans le cadre du projet CARTOFINE (Druet *et al.*, 2008) et testés avec succès sur données réelles à proximité de mutations connues comme la mutation K232A de DGAT (Grisart *et al.*, 2002) ou la mutation F279Y de GHR (Blott *et al.*, 2003). Les principaux QTL cartographiés sont renseignés dans le tableau 1. Dans chaque race, au moins 15 QTL ont été détectés pour chaque caractère, cependant très peu de QTL ainsi détectés sont communs à plusieurs races.

1.3. IDENTIFICATION DE MARQUEURS SNP POUR LA SÉLECTION DES QTL

L'un des objectifs de ces détectations était d'identifier les marqueurs SNP en déséquilibre de liaison avec les principaux QTL. A la position la plus probable du QTL, des combinaisons de quatre à six marqueurs SNP ont ensuite été choisies pour le programme SAM. Par exemple, une combinaison de six marqueurs SNP situés entre 64 et 65 Mb sur le chromosome 1 bovin est en déséquilibre de liaison avec un QTL de production laitière détecté en race Montbéliarde (figure 1). Ces six marqueurs SNP peuvent être utilisés pour sélectionner efficacement ce QTL.

Figure 1 : cartographie fine d'un QTL de production laitière sur le chromosome 1 en race montbéliarde.



2. EVOLUTIONS DU PROGRAMME DE SÉLECTION ASSISTÉE PAR MARQUEURS

Les résultats du projet CARTOFINE permettent d'envisager à court terme l'évolution de la SAM vers une seconde génération de SAM (SAM2) puis rapidement vers une sélection génomique (SG) chez les bovins laitiers.

2.1. PRINCIPE DE LA SAM2

En s'appuyant sur le déséquilibre de liaison entre les haplotypes de SNP retenus et les QTL liés à des caractères d'intérêt, le modèle de l'évaluation génétique assistée par marqueurs peut être considérablement simplifié et la précision des estimations améliorée. En effet, pour chaque région, les animaux ayant hérité du même haplotype de SNP ont une très grande probabilité d'avoir hérité du même allèle au QTL. Les effets de chaque haplotype de SNP peuvent ainsi être estimés globalement au niveau de la population. Dans l'évaluation génétique de la SAM2, les valeurs génétiques sont donc

décomposées en une composante polygénique (due aux gènes non suivis par des marqueurs) et en une somme d'effets des différents haplotypes suivis.

Le modèle génétique de la SAM2 devient :

$$g_i = u_i + \sum_{j=1}^n (H_{ij1} + H_{ij2})$$

où g_i est la valeur génétique de l'animal i

u_i la composante polygénique de l'animal i

H_{ij1} l'effet de l'haplotype paternel du QTL j de l'animal i

H_{ij2} l'effet de l'haplotype maternel du QTL j de l'animal i

Et n le nombre de QTL intégrés au modèle pour le caractère évalué.

Dans ces conditions, il est important de déterminer quelle est la part de la variance génétique qui est expliquée par les QTL du modèle.

2.2. EFFICACITE ATTENDUE DE LA SAM2

Avant de mettre en place la SAM2, l'efficacité du modèle a été testée sur données simulées et vérifiée sur données réelles.

2.2.1. Efficacité estimée sur données simulées

Une étude basée sur 576 jeunes taureaux Holstein a permis de comparer les gains de précision obtenus par différents types d'évaluation génétique. Les généalogies des animaux, les génotypages et les paramètres génétiques utilisés proviennent de données réelles mais les valeurs génétiques vraies des animaux ont été simulées. Les performances des parents sont également simulées comme la somme de l'effet génétique et un résidu dont la variance est fonction de la précision des index réellement disponibles à l'époque de l'entrée en testage sur descendance de ces jeunes taureaux. L'avantage de cette étude réside dans la connaissance des valeurs génétiques vraies des animaux qui restent toujours inconnues pour une étude sur données réelles.

Tableau 2 : CD des valeurs génétiques estimées lors de l'entrée en testage de 576 taureaux pour différentes situations testées.

	h^2	p_{QTL}	n_{QTL}	POL	SAM2
Lait	0,30	40 %	4	0,32	0,53
TB	0,50	60 %	4	0,36	0,67
TP	0,50	50 %	4	0,35	0,56
CEL	0,15	25 %	2	0,27	0,35
FER	0,02	45 %	3	0,22	0,52

Les valeurs génétiques estimées lors de l'entrée en testage dans un modèle polygénique classique (POL) et dans un modèle SAM2 sont comparées avec les niveaux génétiques vrais des 576 taureaux pour déterminer la précision des estimations (CD) de ces deux modèles. L'héritabilité du caractère (h^2), le nombre de QTL simulés (n_{QTL}) et les parts de variances génétiques expliquées par les QTL (p_{QTL}) pour seulement quelques situations testées sont précisés dans le tableau 2.

Dans tous les cas, la SAM2 donne de meilleurs résultats que le modèle polygénique. Ces travaux ont mis en évidence que les CD des valeurs génétiques en SAM2 sont beaucoup plus liés aux parts de variance génétique expliquées par les QTL qu'à l'héritabilité du caractère. Ainsi pour un caractère à

faible héritabilité comme la fertilité mais soumis à quelques QTL à effets forts, l'utilisation d'informations moléculaires semble particulièrement intéressante. Si les parts de variance génétique expliquées par les QTL de la SAM2 sont comparables entre les caractères, on peut donc s'attendre à observer des précisions de valeurs génétiques comparables entre les caractères quelles que soient les héritabilités de ces caractères, ce qui n'est pas le cas dans un modèle polygénique.

2.2.2. Efficacité mesurée sur données réelles

Parmi les taureaux génotypés dans le cadre du projet CARTOFINE, un panel de jeunes taureaux est utilisé pour vérifier l'efficacité du modèle proposé sur données réelles. Il est en effet possible de comparer les valeurs génétiques de ces jeunes taureaux estimées dans les conditions réelles de 2004 avec les valeurs génétiques estimées après testage sur descendance en 2008.

Figure 2 : comparaison des index polygéniques de 2004 des 468 taureaux Holstein pour la quantité de lait avec les index de 2008 après testage sur descendance.

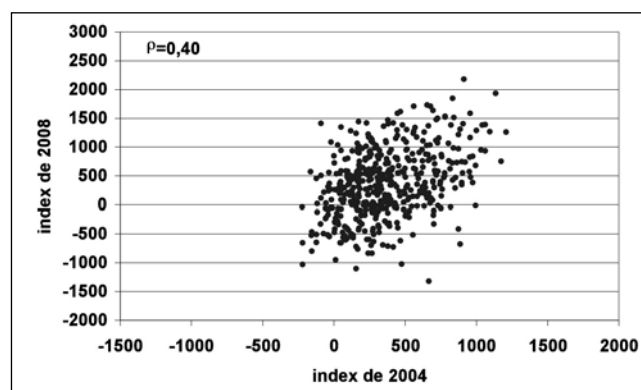
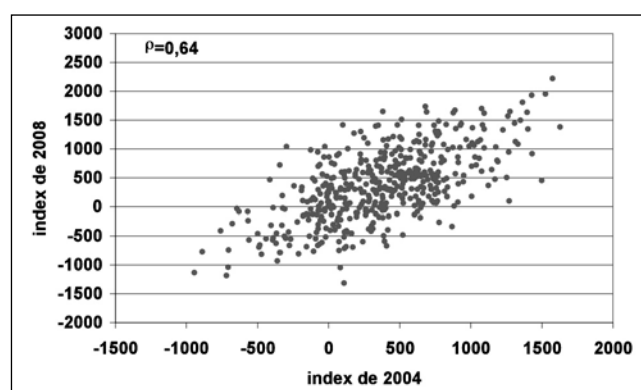


Figure 3 : comparaison des index SAM2 de 2004 des 468 taureaux Holstein pour la quantité de lait avec les index de 2008 après testage sur descendance.



A ce stade du projet, la mise en place de la SAM2 n'est pas terminée. Il n'est possible que de présenter des résultats partiels obtenus pour la quantité de lait et le taux protéique sur un panel de 468 jeunes taureaux Holstein. Les corrélations entre les valeurs génétiques des 468 taureaux estimées en 2004 et celles estimées en 2008 après testage sur descendance sont de 0,40 et 0,52 dans le modèle polygénique pour respectivement la quantité de lait et le taux protéique et de 0,64 et 0,77 dans le modèle SAM2 (figures 2 et 3).

On constate qu'à partir des informations disponibles en 2004, la SAM2 est plus efficace que le modèle polygénique pour repérer les meilleurs mâles avant testage sur descendance. D'ailleurs, les CD des index de ces taureaux calculés dans les conditions de 2004 sont compris entre 0,55 et 0,65 dans le modèle SAM2 alors qu'ils sont inférieurs à 0,40 dans le modèle polygénique. Ces résultats sur données réelles sont tout à fait favorables à la SAM2 et mettent en évidence l'importance de la SAM pour les programmes de sélection chez les bovins laitiers.

2.3. EVOLUTION VERS LA SELECTION GENOMIQUE

Le modèle de la SAM2 est un modèle temporaire puisque l'évolution vers la SG est programmée avant 2010. Globalement, cette évolution consiste à estimer convenablement la composante polygénique à l'aide de marqueurs SNP couvrant l'ensemble du génome. La sélection génomique est en quelque sorte la situation ultime de la SAM2 où l'ensemble des gènes impliqués dans la variabilité génétique d'un caractère sont supposés suivis par des marqueurs ($p_{QTL}=100\%$). On en déduit que les estimations des valeurs génétiques gagnent encore en précision. Avant d'être utilisé en France, l'efficacité du modèle de sélection génomique sera vérifiée sur données réelles.

3. CONSEQUENCES ET PERSPECTIVES

L'INRA, LABOGENA et l'UNCEIA ont signé un nouveau contrat en 2008 pour poursuivre le développement de la SAM chez les bovins laitiers et entreprendre les études nécessaires à l'officialisation des valeurs génétiques utilisant des informations moléculaires. C'est pourquoi depuis août 2008, les animaux du programme sont génotypés à LABOGENA sur la puce Illumina de 54001 SNP. Les entreprises de sélection peuvent dès à présent sélectionner efficacement les jeunes animaux sur près de vingt caractères évalués mensuellement tout en permettant à l'INRA de faire évoluer la SAM2 vers la sélection génomique durant l'année à venir.

3.1. CONSEQUENCES POUR LES PROGRAMMES DE SELECTION

3.1.1. Efforts de génotypages

Avec la SAM, les entreprises de sélection se doivent d'élargir leurs listes initiales de candidats génotypés et de prospecter sur une plus large population de vaches afin d'augmenter la pression de sélection en phase précoce et garantir l'efficacité de leurs schémas. Avec la SAM, la précision des index d'une femelle est équivalente à celle d'un mâle non testé. Ceci peut s'avérer particulièrement intéressant pour la sélection de caractères faiblement héritables et pour le maintien de la diversité génétique.

3.1.2. Réduction de l'intervalle de génération

L'intervalle de génération des bovins laitiers est long en raison de la nécessité de tester les mâles sur descendance avant d'en utiliser certains comme pères de la génération suivante. Avec la SAM, les CD des index d'un taureau non testé devraient être suffisants à terme pour que certains

soient utilisés comme pères à taureaux voire diffusés aux éleveurs. En utilisant plus de taureaux qu'actuellement, la SAM garantit un gain de progrès génétique grâce à une réduction de plus de trois ans de l'intervalle de génération. Des études seront menées au cours de l'année prochaine pour définir la stratégie optimale des schémas d'amélioration génétique pour les entreprises de sélection.

3.2. IMPORTANCE DES PHENOTYPES

3.2.1. Efficacité liée au contrôle de performances

Bien que la SG permette d'estimer efficacement les valeurs génétiques des jeunes animaux sans phénotypes, les phénotypes d'animaux typés sont essentiels pour garantir l'efficacité de la SG. Il est en effet nécessaire de ré estimer régulièrement les effets de toutes les régions du génome en intégrant la dernière génération d'animaux typés. Sans cela, les CD des jeunes animaux baisseraient rapidement au cours des générations (Muir, 2007).

3.2.2. Nouveaux phénotypes étudiés

La mesure de phénotypes et leur enregistrement en routine s'avèrent essentiels pour la sélection génomique. D'ailleurs de nouveaux caractères difficiles à mesurer ou coûteux pourraient en effet être évalués et sélectionnés directement en SG. Des projets de recherche financés par l'ANR et APIS-GENE et menés sur les bovins permettront d'étudier de nouveaux phénotypes liés à des caractères de reproduction (GENIFER), de composition fine du lait (LACTOSCAN) ou de qualité de la viande chez les bovins allaitants (QUALVIGENE).

CONCLUSION

La cartographie génétique à haute densité en marqueurs, appliquée à de grands dispositifs, fournit une information inégalée sur la localisation des gènes responsables du déterminisme des caractères. Elle ouvre la voie à des travaux fondamentaux sur l'identification de ces gènes et la compréhension des mécanismes impliqués mais aussi à la mise en place de nouvelles procédures de sélection qui modifieront considérablement les pratiques actuelles.

Plusieurs pays travaillent actuellement sur le développement et la mise en œuvre de la SG chez les bovins laitiers : Etats-Unis, Canada, Pays-Bas, Nouvelle-Zélande, Danemark, Norvège, etc. La plupart de ces pays ont génotypé plusieurs milliers de bovins sur 20 000 à 54 000 marqueurs SNP. Tous ces travaux mettent en avant le bénéfice apporté par la SG chez les bovins laitiers. Il ne fait désormais plus aucun doute que dans un avenir très proche le modèle de SG sera utilisé dans les évaluations nationales de nombreux pays.

Le projet CARTOFINE est un programme financé conjointement par l'ANR et par APIS-GENE dans le cadre du GIS AGENAE.

- Blott S. *et al.*, 2003. *Genetics*, 163, 253-266
Druet T. *et al.*, 2008. *Genetics*, 178, 2227-2235
Fritz S. *et al.*, 2003. *Renc. Rech. Rum.*, 10, 53-56
Fritz S. *et al.*, 2007. *Renc. Rech. Rum.*, 14, 129-132
Grisart B. *et al.*, 2002. *Genome Research*, 12, 222-231
Meuwissen T.H., Goddard M.E., 2000. *Genetics*, 155, 421-430
Muir W.M., 2007. *J. Anim. Breed. Genet.*, 124, 342-355