

ACADEMIE UNIVERSITAIRE WALLONIE-EUROPE
UNIVERSITE DE LIEGE
FACULTE DE MEDECINE VETERINAIRE & GIGA-R
DEPARTEMENT DES PRODUCTIONS ANIMALES
UNITE DE GENOMIQUE ANIMALE



CONTRIBUTION A LA GESTION DES DEFAUTS GENETIQUES DANS LA RACE BLANC-BLEU BELGE

CONTRIBUTION TO THE MANAGEMENT OF INHERITED DISORDERS IN BELGIAN BLUE CATTLE BREED



Arnaud SARTELET

THESE PRESENTEE EN VUE DE L'OBTENTION DU GRADE DE
DOCTEUR EN SCIENCES VETERINAIRES

ANNEE ACADEMIQUE 2012-2013

Couverture (au centre): *Vanité du Fond de Bois* et son veau
Filigrane: *Attribut du Fond de Bois* (BBG, indemne pour les tares connues)

« Tout ce qui a trait à la vache m'émeut. »

Vincent Roca

*« Le Blanc Bleu Belge :
pilier de notre agriculture et de notre économie. »*

Guy Lutgen

*« Choisissez un travail que vous aimez
et vous n'aurez pas à travailler un seul jour de votre vie. »*

Confucius

REMERCIEMENTS

J'aimerais tout d'abord avoir une pensée pour le Professeur Roger HANSET qui m'aurait sûrement appris de nombreuses choses et qui a laissé une trace importante pour l'élevage BBB. Au moment d'imprimer ce manuscrit, il est temps pour moi de remercier tous les gens passionnés que j'ai pu rencontrer et qui sont, pour certains, devenus des amis proches. Comme vous pourrez le lire, ils sont nombreux à m'avoir permis de réaliser et finaliser ma thèse avec des résultats aussi passionnants MAIS SURTOUT à m'avoir appris à apprécier cette merveilleuse race, le BLANC-BLEU BELGE. Grâce à toutes ces personnes j'ai passé cinq années extraordinaires au service du BBB et j'espère encore en vivre de nombreuses à leurs côtés.

J'aimerais donc remercier,

*Et exprimer toute ma reconnaissance au **Professeur Michel GEORGES**, mon promoteur, inconnu pour moi avant d'arriver au laboratoire. MERCI Michel d'avoir accepté l'idée folle de Carole de me faire travailler dans votre laboratoire, d'avoir su me remettre dans le bon sens lorsque c'était nécessaire, de m'avoir fait confiance et surtout de m'avoir autorisé à défendre ma thèse.*

*Le **Docteur Carole CHARLIER**, mon co-promoteur, MILLES MERCIS Carole, d'avoir eu cette idée folle de me proposer de travailler sur « ton projet » et de m'avoir permis de changer d'air à un moment où cela devenait nécessaire. Merci de m'avoir transmis cette envie, cette curiosité et ce plaisir de la recherche. Je n'aurais jamais assez de mots pour te dire merci, sans toi, cette thèse n'aurait jamais vu le jour. Et surtout merci pour les moments inoubliables de nos divers périples scientifiques et d'être devenue ma Maman de Science, une maman qui compte beaucoup... Merci à Apolline, Hugues et Blanche pour leur présence et je serais toujours là pour vous.*

*Le **Professeur Daniel DESMECHT**, membre de mon comité de thèse. Daniel vous êtes depuis longtemps à mes côtés, d'abord lors de mon changement chaotique de promoteur et de sujet de DEA, puis lors de ma préparation de ma défense de DEA et enfin vous avez été présent tout au long de cette thèse par vos encouragements, vos conseils et votre soutien. Pour tout cela, Daniel, je vous dis tout simplement MERCI.*

*Le **Docteur Kamal TOUATI**, membre de mon comité de thèse. Kamal, merci de m'avoir mis le pied à l'étrier, de m'avoir lancé dans une carrière scientifique au sein de la Faculté et de m'avoir initié à la chirurgie.*

*Aux **Professeurs Nadine BUYS, Didier BOICHARD, Vincent BOURS, Laurent GILLET, Frédéric ROLLIN** et au **Docteur Dominique CASSART**, mon jury, d'avoir pris le temps de lire et d'apporter les remarques et corrections permettant d'améliorer mon manuscrit.*

***L'Equipe Rilouke** : Corinne dit « Coco », la bimbo du labo, MERCI pour ton initiation..... à la paillasse, Nico dit «le p'tit mousse gay » avec sa bonne humeur légendaire mais toujours prêt à rendre service et d'une efficacité remarquable (plus de 24.000 tests....), Sarah, Charles, Tom, MERCI pour ton soutien et ton aide plus que précieuse, plus qu'un p hautement significatif et nécessaire à ma compréhension de la génétique quantitative, François, arrivé sur la fin mais un vrai GI qui fera de GLASCOW son Vietnam, Wanbo et Zhiyan.*

*Le « **KJB** », Corinne, Florian, Mallo, Nico et Sarah pour ses discussions plus scientifiques les unes que les autres MAIS qui a fini par implorer pour le bien de tout le monde et surtout pour l'avancement de mon travail.... MERCI pour votre soutien précieux.*

A Tom, François, Denis et Fred pour ces réunions secrètes à l'ISEPK ou à l'académie Louis Dreyfus, merci pour votre soutien et vos conseils.

Ludo, mon sparing-partner qui m'a abandonné en cours de route... mais avec qui j'ai pu activement décompresser et échanger mutuellement de temps à autres notre désarroi...

*L'Unité de **Génomique Animale** et particulièrement les **post-docs** : Anne, Anne-So, Benoît, Dim (merci Dim de m'avoir laissé ta paillasse et ton titre de boulet du labo, que j'ai porté avec fierté), Emilie, Haruko, Kayo, Keith, Momo, Mallo, Murali, Rodrigo et Xuwen pour leur accueil, leurs précieux conseils à la paillasse et leur soutien, mes **co-doctorants** : Flo, Ann-Steph, Hui-Jun, Zhyian, Géraldine et Wanbo, nos **secrétaires** de choc : Bernadette puis Sandrina et Nathalie (Nathalie, un TOUT GRAND MERCI pour les relectures et surtout la correction des « quelques » fautes oubliées) de nous faciliter la vie au sein du laboratoire et d'organiser des activités, la **Plateforme de Génotranscriptomique** : Wouter, Cécile, Lati, Nadine, Naïma, Sarah, Benoît, Jérôme et François, MERCI pour votre aide, votre soutien, les bons moments partagés et comme le dit une célèbre citation philosophique : « Quand on vient dans l'UAG on braie deux fois : une fois quand on arrive et une fois quand on repart ».*

*Le **Service d'Anatomo-Pathologie** pour m'avoir permis de réaliser mes autopsies et je voudrais remercier plus particulièrement Mamy, Calixte (Content d'avoir pu t'aider pour SBV) et Sandra, MERCI pour vos autopsies, votre bonne humeur et votre collaboration productive, Mutien-Marie, MERCI pour ton aide et sa collaboration dans les diagnostics PCR, Michel pour ta bonne humeur en salle d'autopsie, Mickaël, le BOULET te remercie du fond du cœur pour ta disponibilité et ton efficacité dans le ramassage de mes chef-d'œuvres, Carine et Nathalie, merci pour l'organisation des transports et le suivi des rapports.*

*Le **Herd-Book Blanc-bleu Belge**, Eveline, Aline, Shanthi, Martine et Pierre pour leur disponibilité, leur aide précieuse dans la fourniture des pédigree, des trends génétiques et de toutes les informations nécessaires à la réalisation de cette thèse ainsi qu'à la gestion des défauts génétiques dans la race BBB.*

*Les centres d'inséminations pour leur confiance, leur transparence, leur disponibilité et leur aide **Belgian Blue Cattle International (BBCI)**, **Belgian Blue Group (BBG)** et particulièrement Eric Lust, **Fabroca** et surtout Benoît Cassart pour son soutien, ses discussions enrichissantes, son partage de la passion du BBB et enfin **Génétique Avenir Belgimex (GAB)** et principalement Luc Mahoux, Jean Ippersiel, Jean-Pierre Monfort et Guy Gaillard pour leur soutien régulier, leurs discussions et leur confiance.*

L'Association Wallonne de l'Elevage et plus particulièrement Xavier Hubin et les agents de terrains pour les informations d'accouplements et pour le relai précieux sur le terrain, surtout en fin de thèse.

A Jean-Pierre et Quentin Monfort ainsi qu'à BBG de m'avoir autorisé à utiliser leurs photos de respectivement *Vanité du Fond de Bois* et son veau et d'*Attribut du Fond de Bois* (indemne des sept tares) afin de pouvoir réaliser cette couverture qui achève magnifiquement ce manuscrit.

Au Ministère de l'Agriculture de la Région Wallonne pour le financement du projet Rilouke.

Je voudrais également remercier, pour avoir pris le temps de me contacter afin de me renseigner les cas observés dans leur clientèle, les Docteurs :

Andriessen, Balthazard, Bataille, Bayrou, Bejot, Bombeek, Bonne, Bonnevie, Bontemps, Botton E., Botton S., Brackman, Brassel, Brasseur, Brassine, Brillon, Bughin, Buron, Buyck, Carlberg, Carnevali, Casalta, Cassart, Cathy, Cauchie, Cherdon, Christiaens, Coghe, Colinet, Collard, Collin, Comté, Conotte, Constant, Cornillie, Crenn, Crévits, Culot, D'Affnay, Dardenne, Daubresse, Daudet, De Knoop, De Marneffe, De Roover, De Stoop, De Swaef, De Thillesse, De Vlamynck, De Vriese, Debarsy, Decoster, Degrande, Degros, Dehon, Delbouvry, Demblon, Demuyndt, Deneyer, Denis, Déom, Deprez, Dernivoix, Deroanne, Desmecht, Desthede, Deswissen, Deveux, Devillers, Devisscher, Devos, Devreux, Devriese, Dewaels, Diez, Dion, Distèque, Disthede, Dive, Dochain, Droulez, Dubuc, Dufrasne, Evrard, Feron, Fisenne, Frankinet, Frennet, Gengoux, Georges, Géron, Ghysens, Gielen, Gielis, Gilis, Glowacki, Graff, Gréban, Guillaume, Haas, Hampert, Hannart, Hellinckx, Hennequin, Heymans, Hoflack, Houbotte, Hubermont, Hubin, Jacquet, Jodogne, Jolly, Jubaru, Kersten, Kesteloot, Laitat, Lambotte, Lampo, Lejoly, Lekenne, Lemaire, Léonard, Leprince, Lequenne, Leroy, Leyder, Leyens, Liégeois, Limbourg, Lisart, Lunsken, Maes, Marcin, Masson, Massonet, Mathieu, Mathot, Mattart, Mazay, Melchior, Meunier, Michel, Misse, Morteveille, Moulin, Neyrinck, Onclin, Otjacques, Pacco, Paindaveine, Pardon, Parmentier, Patz, Pellichero, Pestiaux, Pierret, Pollentier, Ponsar, Putzeys, Quoilin, Remacle, Reynaerts, Roegiers, Roppe, Sacré, Saulmont, Schonbrodt, Schoukens, Scols, Senger, Simon, Snaps, Soenen-Tuytens, Spronck, Thielens, Thomas, Tison, Truffaut, Uystepuust, Van der molen, Van der Noorgate, Van Heule, Van Lerbergh, Van Lierde, Van Loo, Vanacker, Vanderbiest, Vanderheyden, Vanheule, Vercoetere, Vermeulen, Vilenne, Wenquin, Wertz, Weytens, Winand, Wuidar, Wylock.

Pour leur accueil, leur disponibilité, leur coopération et le partage de leur passion de l'élevage du Blanc-bleu Belge, j'aimerais remercier chaleureusement Mesdames et Messieurs :

Abinet F., Adam B. & Michel C., Adam M., Adams A., Aerts A., Aigret M., Algoet B., Andre C., Andre N., Annet G. & N., Annet Th., Anthemus P., Antoine M., Archambeau P., Arnould P., Arnould M. & S., Avet R. & D., Backes R., Baijot P., Balfroid B., Ballaux A., Ballaux N., Balleux V., Barbe V., Bastin R., Baudoin B., Baudoin M., Baudoin Ph., Baudoin R., Baudoin V., Baulard J., Beaudoin E., Beauvois-Ysebaert, Beguin G., Benoit F., Bernard J., Bernard L., Bernier, Bettendorf S., Bihain M., Boclinville A., Bocquel JM., Bodelet D., Bodson B., Bodson M., Boisdenghien V., Boite J., Bombeek R., Borgniet-Piret Ass., Borlon J., Bottemanne Ph., Bouchat A., Bouchez A., Bourdeaud'huy P., Bouvier, Boxus M., Brackman A., Brasseur, Brassine, Bray-Pochet L., Bresmal J., Briffroz F., Brone ML., Budo-Pirson, Brumioul J., Brumioul P., Bughin, Buisseret H., Burgraff JP., Burniat E., Busar, Buyck Ph., C. Tech. Agro., Canivet MF., Cardyn E., Carpentier Th., Cassart B., Cassart Ch., Cassart B., Cassart L., Castelain D., Cathy Jp., Cavillot D., Cavillot M., Chaidron P., Champenois, Charlier F., Chartier V., Chavagne J., Chilliarte, Chisogne M., Christiaens S., Christophe H., Clarence S., Claude J., Clause frères, Cocquyt M., Coheur E., Colaax J., Collard, Collart JP., Collart M., Collignon G., Collignon M., Collignon P., Collin, Collin J., Comes S., Conotte L. & H., Copenaut V., Cordier M., Cornil, Cosse D., Cosse F. & fils, Courtois P., Cravatte, Crépin Ph., Creppe, Crevits P., Crevits R., Crismer R., Crispoux E., Crokaert M., D'haene K., Daems R., Daerden A., Damanet H., Damanet S., Dannels J., Dardenne JP., Dardenne P., David JM., David P., Dawagne P., Dawance V., Dewaele M., Dewez G., De Bonhome B., De Coene A., De Dorlodot G., De Fauw K., De Favereau B., De Smedt D., De Swaef P., De Vriese A., De Waele L., Debock G., Debry F. & T., Debytter, Dechief C., Declercq, Deconinck D., Decruyenaere F., Dedorlodot, Defauw H., Defay C., Defays P., Defeigne, Defenain J., Degand Ph., Degln-Louis Ph., Degros M., Degry C., Dehay H. & G., Dehean-Lanssens, Delbar Ph., Deldeffe LM., Delescaille M. & Bouvier C., Delloy L., Delloy R., Delstanche-Masson, Delvaux B., Delvigne M & L., Delville M., Demarcin A. & B., Demuyndt E., Deneubourg C., Denis, Denis M., Denomerange, Denoo P., Deny, Deny-Verlaine, Depoortere E., Derasse, Deschamps P., Descuter W., Desjardin, Desmet L., Desplentier L., Despy R., Desseilles J., Desselle R. & C., Desthede, Detal E., Detal P., Devers JL., Devillers F., Devlesaver R., Dewaele M., Dewalque B., Dewulf E., Didier JM., Didion J., T. & T., Dion A., Diskheuve, Dochain A., Dochain V., Doclot, Docquier P., Dorignaux frères, Dorlodot, Dorts E., Dourcy JP., Drouguet, Druart O., Dubois M., Dubois P., Dubrulle F., Duchemin Ch., Dufey P., Dufey V., Dumay Ph., Dumont, Dumortier C., Dupont, Dupont Y., Durant O., Durant Y., Dury JP., Duval, L'Ecole Provinciale d'Agronomie et des Sciences de Ciney, Eeckhout R., Eecloo J., Eliard E., Ernst-Dahlen J. E. & R., Etienne Ph. & J., Evrard R., Fagnartj., Ferauge PY., Feryn D., Fiasse S., Fondu, Foucart R., Francois, Frankart J., Fransolet B., Frauenkrone E., Frederic F., Frissen G., Gaillard G., Gallet M., Gallet P., Garroy, Gaspard, Gaspard R., Gaussin JM., Gauthier S., Geens B., Geerts A., Georges F., Georges JP., Georges R., Georges V., Georis, Gerard F., Germain, Germeau A., Ghyse Ch., Gillen, Gilles Ph., Gillis E., Gilson E., Glorieux, Glorieux L., Gob J., Gobert K., Goddeeris, Godefroid J. & M. Ass., Godfriaud B., Godfriaux S., Godfrind C., Godfroid M., Goebels B., Goffart D., Grandhenry A., Grandjean J., Gremeaux M., Grosjean L., Groux JM., Guelle M., Guermont G., Guillaume C., Guillaume Frères, Guilmet C., Guiot E. & B., Guyot O., Haegens G., Hahn, Halleux M., Halloy, Hampert P., Hanegreefs D., Hanlet B., Hanoul E., Hansenne P., Hanssen J., Harzimont F.,

Haufroid B., Heinrichs E., Henneton B., Henrard P., Henry J., Henry N., Herman H., Hocks, Hocks JF., Hodchamps M., Hogge A., Houard O., Houssa JF., Hovine A., Hovine C., Hovine D., Hubert & Wathelet, Hubert J.-L., Hubert S., Huberty F., Hussin P., Hurdebise B., Huylenbroeck, Imbrecht G., Jacob V., Jacoby F., Jacot Ch., Jacquemin L., Jacquet L., Jacques M., Jaspert & Baudoin, Jaumin J., Jaymaert, Jonet B., Jouniaux J., Jurquin A., Kallen, Kersten V., Kessler, Kesteloot, Krier R. & F., Kune M., Labar JM., Labasse P., Labranche P., Laffineur D., Lagneaux B., Lallemand-Dantine J., Laloyaux M., Lambeaux F. & C., Lambert A., Lambert J., Lambert R., Lamontagne L.J., Lange E., Laruelle M., Laurent B., Laurent P., Laval F., Laval R., Lavis T., Leboutte A., Leboutte JM., Leboutte M., Lebrun M., Lebrun Ph., Lecas D., Leclercq B., Lecocq M., Lecomte JM., Lecomte B., Lecomte E. & J., Lecomte M., Lecomte R., Leconte-Leridan D., Lecron GJ., Lecompe JM., Ledent-Jacquet, Ledoux, Lefèbvre B., Lefèvre R. & P., Leglise B., Lejeune P., Lekane T., Leleu A., Leleux C., Lenearts H., Lenoir P., Leonard M., Lepage P., Lepere C., Lequeux C., Lequeux J., Leray P., Leyder D., Lhermitte J. & V., Liagre I., Liagre M., Licop A., Liegeois Y., Lievens J., Lievens M., Limbourg Ph., Lissioir JL., Locquet Ch., Loly, Loneux F., Louis JM., Maes P., Mahoux B., Mahoux L., Mailleux B., Mailleux JM., Mailleux M., Mailleux M. & JM., Mailleux P., Mailleux-Dony & Fils, Majerus, Malaise O., Maldague J., Marchal P., Marchal-Lambert, Marcourt D., Marenne R., Marlier J., Marot M., Marsil D., Martens P., Martin JP., Martin-Etienne JP. & G., Massard B., Masschelein J., Masson J., Masson A., Masson H., Masson JP., Masson-Laloi, Masure I., Materne C., Mathieu E., Mathieu L., Mattart-Rigo, Mazier J., Meersemans E., Meert U., Mehauden Ph., Meuree M., Meyer L., Meyer-Gillet Ass., Michels J., Minet B., Minet J., Minet M., Misse H., Modave C., Modave T., Moïens Fr., Monfort JP., Monin J., Muller-Jates H., Neubergh Ph., Neyrinckx Y., Nicks A., Nicks C., Nisen JP., Noel D. & D., Noel Y., Nuytten V., Nylona S.A., Ophals J., Orban M., Ostyn E., Otjacques L., Otjacques G., Ottelet M., Paquay Ass., Paquay J., Paquay JP., Paquay Y., Paquet A., Paquet F., Parache L., Paris Y., Parmentier M., Pauly F., Pauly J., Pekel P., Pestiaux A., Petit JM., Philippe A., Pierard, Pierard D., Pierard M., Pierrard J., Pierrard L., Pierrard-Toussaint, Pierret M., Piette-Matagne F., Pinsart Ch., Piret G., Piret H., Piron C., Pirson D., Pirson P., Plainchamps, Plennevaux P., Pochet Ass., Pochet B. & S., Polet JP., Poliart, Pollentier W., Poncelet, Poncelet P., Poncin J., Ponsar, Poulipoulis A., Promme G., Prud'homme, Rabeux P., Rabeux J. & JP., Rabeux M-H., Rabeux P., Ramelot B., Rassenfosse, Reginster, Remacle JP., Renard C., Renard L., Renon F., Reuliaux, Richard C., Richard H., Rigaux, Ringlet A., Robijns S., Roche-Masson B., Roetynck A., Romedenne, Romedenne A., Romedenne M., Ronvaux J., Rossenfosse M., Rossignon JP., Rossion & Croibien, Rossion A., Roupin M., Rousseaux-Deschuyteneer C., Rouxhet JF., Rulot T., Sacré, Salpetier F., Scalais JP., Schiffer J., Schneder G., Scholaert J., Schweich O., Semaire J., Servatte, Seutin B., Seutin F., Severin JF., Simal, Simoens P., Simon, Simon F., Simon J., Simon JF., Simon V., Sirjacq L., Sizaïre A., Smeyters D., Snauwaert R., Soete F., Somville D., Son JM., Stavaux JL., Stommeling, Sulbout M., Surlemont T., Taelman J., Taminiaux, Taminiaux L., Tangeten A., Tasiaux L., Tasiaux M., Tassoul E., Teetaert, Thiange E. , Thiange L. & V., Thomas C., Thomson A., Tilkin, Tireur JP., Toussaint M., Troquet F., Uyttendaele J., Vaguet A., Van Belleghem A., Van Der Heeren, Van de Velde P., Van Eyck E. S. & Fr, Van Eyck P., Van Heule JM., Van Hoof G., Van Wijnsberghe, Vanacker C., Vanacker G., Vanaerde L., Vancompernelle M., Vandekerckhove F., Vandecaveye, Vandemergel R., Vandenwaeyenberg F., Vandercruyssen R., Vanderlinden E., Vandermael A., Vandewaetere I, Vandewaetere M, Vandromme M, Vanelderden, Vanguel, Vanherf M. & A., Vanheynde M., Vanmello M., Vanmullen, Vannieuwenhuyse, Vanveuwenhargen Ch., Vanvinckenroye P & B, Vanzuylen, Veracx S., Verberckmoes, Vercruysse, Vereecke R., Verhelst W., Verlaine, Vermeiren, Vermeulen T., Versavel, Verstraelen A., Villeval-Arnould A., Vincent C. & M., Vraux R., Wanson Ch., Warnier J., Warnier-Lambert J., Warzee C., L. & L., Wathelet G., Wathelet R., Wauthélet M., Wenkin , Weron B., Willand A., Willem A., Withagen J. & P., Witry M., Wylock J., Xhonneux R., Zabus A., Zabus JL., Zander D.

LISTE DES ABBREVIATIONS

α -MSHR	<i>Melanocyte-stimulating hormone α receptor</i>
ADN	Acide désoxyribonucléique
AP	Syndrôme d'arthrogrypose / <i>Arthrogryposis syndrom</i>
AR	Arrêté royale
ARNm	Acide ribonucléique messenger
Arsia	Association régionale de santé et d'identification Animale
AWE	Association wallonne de l'élevage
BBB/BBCB	Blanc-bleu Belge / <i>Belgian blue cattle breed</i>
BBCI	<i>Belgian blue cattle international</i>
BBG	<i>Belgian blue group</i>
BLAD	<i>Bovine leukocyte adhesion deficiency</i>
BLUP	<i>Best linear unbiased prediction</i>
BTV-8	<i>Blue tongue virus serotype 8</i>
CIA	Centre d'insémination artificielle
CLCN7	<i>Chloride channel 7 gene</i>
ClC-7	Transporteur Cl ⁻ /H ⁺ ClC-7
CSB	Centre de sélection bovine
CNV	<i>Copy number variant</i>
CTLD	<i>C-type Lectin-like domain</i>
DGZ	<i>Dierengezondheidszorg Vlaanderen</i>
DMC 1 /	Dystonie musculaire congénitale de type 1/
CMD 1	<i>Congenital muscular dystony type 1</i>
DMC 2 /	Dystonie musculaire congénitale de type 2/
CMD 2	<i>Congenital muscular dystony type 2</i>
EBV	<i>Estimated breeding value</i>
FCO	Fièvre catarhale ovine
FMV	Faculté de médecine vétérinaire
GAB	Génétique avenir belgimex
Gb	Gigabase
GBV	Genomic breeding value
GP/PG	Gestation prolongée / <i>Prolonged gestation</i>
GQM	Gain quotidien moyen
GWAS	<i>Genome wide association study</i>
HAM	Hamartome gingival
HBBB	Herd-Book de la race Blanc-bleu Belge
IA/AI	Insémination artificielle / <i>Artificial insemination</i>
INRA	Institut national de la recherche agronomique
IBD	<i>Identical by descent</i>
IBS	<i>Identical by state</i>
Kb	Kilobase
KITLG	<i>c-Kit ligand</i>
KO	<i>Knock-out</i>
MC1R	<i>Melanocortine type 1 receptor</i>

Mb	Mégabase
MGF	<i>Mast cell growth factor</i>
Mh	<i>Muscular hypertrophy</i>
MOET	<i>Multiple ovulation and embryo transfer</i>
MRC2	<i>Mannose receptor C type 2</i>
NAN/DW	Nanisme / Dwarfism
NF-KB	<i>Nuclear factor – kappa B</i>
NGS	<i>Next generation sequencing</i>
NMD	<i>Nonsense-mediated decay</i>
OMIA	<i>Online mendelian inheritance in animals</i>
ORF	<i>Open reading frame</i>
pb	Paire de bases
PIGH	<i>Phosphatidylinositol glycan anchor biosynthesis class H</i>
PCR	<i>Polymerase chain reaction</i>
pNPI	Procollagène I N-protéinase
PSG/PSQ	Parésie spastique des gastrocnémiens / des quadriceps
QTL	<i>Quantitative trait locus</i>
QTN	<i>Quantitative trait nucleotide</i>
RAC	Rigidité articulaire congénitale
SDRA	Syndrôme de détresse respiratoire aiguë
RNA seq	<i>RNA sequencing</i>
RFLP	<i>Restriction fragment length polymorphism</i>
RNF11	<i>Ring finger protein 11</i>
RT-PCR	<i>Reverse transcriptase-polymerase chain reaction</i>
SAM	Sélection assistée par marqueurs
SBV	<i>Schmallenberg virus</i>
SCFR	<i>Stem cell growth factor receptor</i>
SMA	<i>Spinal muscular atrophy</i>
SNP	<i>Single nucleotide polymorphism</i>
SQT/CTS	Syndrôme de la queue tordue / <i>Crooked tail syndrome</i>
STS	<i>Short tag sequence</i>
TDT	<i>Transmission disequilibrium test</i>
UPV	Union Professionnelle des Vétérinaires
VNTR	<i>Variable number tandem repeat</i>
WHD	<i>White heifer disease</i>

TABLE DES MATIERES

TABLE DES MATIERES

RESUME	13
SUMMARY	17
INTRODUCTION	20
1- Introduction générale	21
2- La race blanc-bleu belge	24
A. Origine et évolution	
B. Population	
C. Caractéristiques	
D. Programme de sélection	
E. La race BBB et ses défauts	
F. Identification des gènes majeurs dans la race BBB	
3- Etude du déterminisme génétique	45
A. Introduction	
B. Etude du déterminisme génétique par cartographie ou clonage positionnel	
a) Marqueurs génétiques	
b) Liaison génétique et déséquilibre de liaison	
c) Etudes de liaison	
d) Etudes d'association	
C. Approche du déterminisme génétique par l'approche de type gène candidat	
D. Identification et validation de la mutation causale	
E. Développement d'un test diagnostique	
OBJECTIFS DU TRAVAIL	69
RESULTATS	72
1- Introduction	73
2- Sélection balancée du locus <i>MRC2</i> bovin responsable du Syndrôme de la queue tordue	75
A. Résumé	
B. Articles	
Fasquelle, Sartelet <i>et al.</i> , <i>PLoS Genetics</i> , 2009.	
Sartelet, Klingbeil <i>et al.</i> , <i>Animal Genetics</i> , 2012.	
3- Effet pléiotrope du gène <i>RNF11</i> bovin sur la croissance et la résistance aux maladies	102
A. Résumé	
B. Article	
Sartelet <i>et al.</i> , <i>PLoS Genetics</i> , 2012.	

4- Identification de la mutation causale, dans le gène <i>CLCN7</i>, accélérant l'activation du transporteur ClC-7 et responsable d'hamartomes gingivaux et d'ostéopétrose	122
A. Résumé	
B. Article	
Sartelet, Stauber <i>et al.</i> , <i>soumis pour publication</i> .	
5- Identification de la mutation causale, dans le gène <i>PIGH</i>, responsable d'une déficience en glycosylphosphatidyl inositol et du syndrome d'arthrogrypose létale	159
A. Résumé	
B. Article	
Sartelet, Li <i>et al.</i> , <i>en préparation</i> .	
6- Sélection assistée par marqueurs pour éradiquer la gestation prolongée ..	176
A. Résumé	
B. Article	
Zhang <i>et al.</i> , <i>Bioinformatics</i> , 2012.	
DISCUSSION GENERALE ET PERSPECTIVES	189
ANNEXES	210
Annexe 1 : Standards de la race	
Annexe 2 : Cotations linéaires	
Annexe 3 : Pedigree et cotations linéaires d' <i>Harpon de l'Orgelot</i>	
Annexe 4 : Evaluations génétiques de la race BBB	
Annexe 5 : Bilan des cas collectés par la cellule d'hérédo-surveillance	
Annexe 6 : Durkin <i>et al.</i> , <i>Nature</i> , 2012.	
Annexe 7 : Passeport d'identification SANITEL bovin	
BIBLIOGRAPHIE	246

RESUME

La race Blanc-bleu Belge (BBB) a une place importante dans l'agriculture et la société belge. Elle représente la moitié du cheptel bovin belge et environ 90 % des bovins viandeux. Elle fournit ainsi près de deux tiers de la viande rouge. Elle est surtout reconnue pour son développement musculaire extraordinaire. Au cours des cinquante dernières années, les éleveurs ont favorisé l'accouplement d'individus souvent apparentés présentant les meilleures caractéristiques afin de fixer ces caractères favorables et/ou utilisé massivement l'insémination artificielle (IA) réduisant de manière importante la taille efficace de la population. D'autre part, il est généralement admis que chaque individu est porteur d'une poignée de mutations délétères, à l'état hétérozygote, qui seront alors rapidement disséminées dans la population, entraînant ainsi l'apparition de tares récessives dans les générations suivantes.

Au cours de cette dernière décennie, la génomique, bovine y compris, a connu une évolution spectaculaire des outils technologiques. La cartographie et l'identification rapide des mutations causales pour deux pathologies autosomales récessives ségrégeant en BBB, la dystonie musculaire de type 1 (DMC 1) et la dystonie musculaire de type 2 (DMC 2), ont d'ailleurs prouvé l'efficacité de ces nouveaux outils.

C'est dans ce contexte que la pertinence de mettre sur pied une cellule d'hérédosurveillance de la race BBB est devenue évidente. Cette cellule a pour but de mettre en place et de développer un réseau d'observation en temps réel, de l'émergence des anomalies congénitales. Cela nécessite une collaboration étroite avec les praticiens, les éleveurs, les associations d'élevages, les centres d'insémination et le Herd-Book Blanc-Bleu Belge dans le but i) d'établir un recensement épidémiologique des tares existantes ayant un impact négatif majeur sur l'économie des élevages BBB, ii) d'alimenter une base de données et une biobanque, iii) de cartographier les gènes et mutations responsables pour les défauts dont l'origine génétique est suspectée, et enfin iv) de développer un test diagnostique, de le mettre à la disposition des éleveurs et de permettre la gestion raisonnée de ces tares génétiques.

Depuis 2006, près de 2.600 échantillons ont alimenté la biobanque et la base de données. L'intérêt a été porté sur les cas affectant la rentabilité des élevages de BBB. Nous les avons classés en fonction de leurs conséquences avec i) les défauts congénitaux engendrant une mortinatalité sans répercussion sur la mère, ii) les défauts congénitaux létaux à la naissance affectant la fertilité et/ou la survie de la mère et iii) les défauts altérant les performances zootechniques. Les analyses épidémiologiques des différents phénotypes recensés nous ont amenés à suspecter un mode d'hérédité simple ou complexe pour une dizaine de défauts majeurs en BBB. Pour les caractères simples, les études d'associations sur le génome entier (GWAS) ont abouti à la localisation d'une région homozygote par descendance (IBD) pour cinq d'entre eux, confirmant le mode de transmission autosomale récessive pour le syndrome de la queue tordue (SQT), l'hamartome gingival congénital (HAM), la gestation prolongée (GP) et

l'arthrogrypose (AP). La taille de ce segment IBD partagé varie de 1,2 à 3,6 Mb et peut contenir de ~10 à ~150 gènes. En ce qui concerne les caractères complexes, l'analyse épidémiologique des 206 cas de White Heifer Disease (WHD) a confirmé le rôle majeur du locus Rouan dans le déterminisme de la WHD. Une première GWAS réalisée à partir de 188 cas de WHD pointe deux loci suggestifs qu'il reste à répliquer. Pour le retard de croissance (NAN), caractère considéré comme complexe, 40% des veaux partageaient également un segment IBD homozygote.

Une approche de type gène candidat positionnel, consistant à sélectionner un ou plusieurs gènes contenus dans la région IBD, a été utilisée. Ils sont priorisés en fonction de leur implication dans des phénotypes similaires connus chez l'homme ou dans des modèles animaux tels que la souris. Les régions codantes de ces gènes sont alors séquencées. En l'absence de gènes candidats évidents, un séquençage systématique des régions codantes des gènes de l'intervalle a localisé la mutation causale responsable du SQT et du NAN. Pour le SQT, il s'agit d'une délétion dans le gène *MRC2* qui code pour un récepteur au mannose. Son rôle dans le recyclage du collagène lors de la croissance osseuse et cartilagineuse est compatible avec le retard de croissance observé dans les cas de SQT. Une substitution au niveau d'un site d'épissage du gène *RNF11* est responsable du NAN. *RNF11* fait partie du complexe A20 réprimant l'activation de NF-KB dans la réponse inflammatoire. Cette approche restait lourde et contraignante, surtout lorsque le contenu en gènes des régions IBD était important. Le développement des nouvelles technologies de séquençage à haut-débit a révolutionné cette étape. Trois substitutions privées dans le gène *CLCN7* (Chloride channel-7) ont été identifiées par reséquençage du génome entier, changeant un acide aminé ultra conservé, avec pour conséquence une altération de l'activité de ce transporteur à chlore présent au niveau des ostéoclastes et intervenant dans le remodelage osseux. Des mutations connues au niveau de ce gène provoquent de l'ostéopétrose chez l'homme et chez la souris. Un examen nécropsique approfondi réalisé *a posteriori* sur des veaux homozygotes pour cette mutation a confirmé l'ostéopétrose. Pour l'AP, une substitution au niveau d'un site accepteur d'épissage du gène *PIGH* identifiée par RNA-seq provoque une délétion complète de l'exon 2. Pour la GP, la mutation causale est en cours de validation.

L'identification de la mutation causale permet de développer un test diagnostique direct, grâce à celui-ci les croisements à risque et la naissance de veaux mutants pourraient être évités. Le délai entre la localisation et l'identification de la mutation causale peut cependant être long comme c'est le cas pour la GP et l'AP. A la demande des éleveurs, un test diagnostique indirect, basé sur les marqueurs en déséquilibre de liaison avec la mutation, a été développé pour ces deux pathologies. La fréquence des porteurs au moment du développement de chaque test varie de 5 à 26 %. Dans le cas du NAN, une différence entre la fréquence de la mutation et le nombre de cas recensés par la cellule a été constatée. Une étude prospective, visant à suivre des

accouplements « porteur-porteuse », a mis en évidence une mortalité et une morbidité significativement augmentée entre zéro et un an expliquant cette inconsistance apparente. Ces résultats sont similaires aux observations chez les individus KO (knock-out) pour les autres membres du complexe A20. Enfin, pour le SQT et le NAN, la fréquence élevée de ces mutations (> 20 %) est expliquée par un avantage sélectif qu'elles confèrent aux porteurs. Nous avons démontré que les porteurs de la mutation du gène *MRC2* présentent une musculature supérieure aux non-porteurs, phénotype ayant été préférentiellement sélectionné par les éleveurs. Un argument supplémentaire en faveur de cette hypothèse est l'identification, au cours de l'année 2010, d'une seconde mutation délétère dans le gène *MRC2*, celle-ci ayant probablement été sélectionnée par les éleveurs pour son effet positif sur la musculature. Par contre, si l'avantage sélectif des porteurs de la mutation dans le gène *RNF11* a bien été démontré indirectement, aucun effet sur les cotations linéaires des différents phénotypes mesurés ne l'a été.

Depuis la mise en place des tests, près de 24.000 individus, majoritairement des mâles, ont été testés pour au moins un défaut. La sélection sur les mâles, dans une population comme le BBB où l'insémination représente plus de 50 % des naissances, est plus que suffisante dans la gestion des pertes liées à la naissance d'homozygotes mutants. De plus, malgré l'imperfection des tests indirects dans la détection des individus porteurs, ils permettent de gérer ces défauts de manière quasi aussi efficace. Cette sélection contre les porteurs a rapidement supprimé la naissance de veaux homozygotes pour ces défauts. L'éradication de ces défauts a aussi diminué la mortinatalité de manière significative dans la race BBB par rapport aux races laitières en Wallonie. L'effet sur la diminution du nombre de porteurs dans la population testée est plus progressif, vu qu'aucune sélection sévère n'est exercée sur le réservoir de femelles. Un corolaire négatif de l'élimination systématique des taureaux porteurs d'une ou plusieurs mutations (≈ 50 %) est la probable réduction de la diversité génétique, puisque des lignées entières à risques sont en passe de disparaître. Paradoxalement, cette sélection risque de faire émerger de nouvelles tares, les éleveurs se tournant massivement vers de nouvelles lignées, indemnes pour les mutations testées, mais vraisemblablement porteuses d'autres mutations. En conséquence, afin de maintenir une diversité génétique maximale et conserver le potentiel génétique acquis des taureaux porteurs, des croisements dirigés visant à produire des descendants indemnes doivent être réalisés par des sélectionneurs.

Une cellule d'hérédo-surveillance aspirant à identifier et à gérer les défauts héréditaires dans les races ultra sélectionnées trouve donc une place essentielle dans les programmes de sélection.

SUMMARY

The Belgian blue cattle breed (BBCB) is well known for its exceptional muscular development referred to as « double-muscling ». Belgian Blue cattle account for approximately half of the 2.6 million heads of cattle in Belgium and for approximately two thirds of Belgian beef production. Intense selection for animals with extreme phenotype, including by means of extensive use of artificial insemination, has caused a decrease in the effective population size and concomitant increase in inbreeding coefficient. This has in turn caused recurrent outbursts of inherited defects, imposing considerable economic and welfare burden on breeders.

Recent advances in genomics, including high-throughput SNP genotyping and next generation sequencing (NGS), have greatly accelerated the identification of genes and mutations underlying inherited defects, including in livestock (f.i. Charlier *et al.*, 2008). To maximize the benefit of these new technologies for the Belgian Blue cattle breed, we established a heredo-surveillance platform in close collaboration with breeders, veterinarians, breeding organizations (“Association Wallonne de l’Elevage” and “BBCB Herdbook”) and AI centers. The aim of the platform was (i) to monitor the emergence of genetic defects in the BBCB and quantify their incidence and impact, (ii) to collect epidemiological data and biological samples of cases and controls, (iii) to positionally clone the genes and mutations underlying inherited defects, and (iv) to develop diagnostic tests, and assist breeders in using these to limit the negative impact of the defects.

Since 2006, the platform’s biobank has been populated with more than 2,600 samples, corresponding to 46 anomalies classified as (i) lethal defects affecting the calf, (ii) lethal defects affecting calf and dam, and (iii) non-lethal defects with negative economic impact.

We performed a genome wide association study (GWAS) for 15 defects resulting in the unambiguous identification of risk loci for five, thereby demonstrating their inherited nature. The identified loci explained all the phenotypic variance for the crooked tail syndrome (CTS) on the chromosome 19, the dwarfism (DW) on the chromosome 3, the gingival hamartome (HAM) on the chromosome 25, the prolonged gestation (PG) on the chromosome 18 and the lethal arthrogryposis syndrome (AP) on the chromosome 10 under an autosomal recessive model of inheritance. For dwarfism (DW), the locus identified on the chromosome 3 accounts for 40% of cases of stunted growth in BBCB under the same model. In addition to confirming the major effect of the Roan locus in the determinism of White Heifer Disease (WHD), we have identified two suggestive risk loci with p-value expected to occur fortuitously only once per genome-scan.

Targeted exon sequencing of positional candidates lead to the identification of causative mutations for two conditions. CTS was shown to results from a frameshifting 2-bp deletion in the *MRC2* gene encoding a mannose receptor, while DW was shown to be due to a splice variant in the *RNF11* gene, encoding a subunit of the A20 complex regulating NF-KB signaling. Whole genome NGS was needed to identify the causative mutations for HAM and AP. The substitution of

a highly conserved amino-acid in the chloride channel *CLC-7*, was shown to affect its electrophysiological properties with potential impact on its bone remodeling role in osteoclasts. Mutations in the *CLCN7* gene are known to induce osteopetrosis in humans and mice,ⁿ and this was subsequently shown to also be the case in calves with HAM. A splice site variant 10-bp upstream of the second exon of the *PIGH* gene, encoding PIG-H protein, was shown to cause AP. The causal mutation for PG has not been identified so far.

We were struck by the unusually high population frequency in BBCB of the mutations causing two defects, SQT (25 %) and DW (26 %). We provide strong evidence that this reflects balancing selection due to desirable pleiotropic effects. The *c.2904_2905delAG* mutation in *MRC2* was shown to increase muscle mass in carriers, and this probably accounts for the occurrence of a second, rare *MRC2* loss-of-function mutation discovered subsequently in a distinct BBCB sire line. We provide evidence that the *c124-2A>G* mutation in the *RNF11* gene underwent a selective sweep increasing its frequency, yet were not able thus far to determine the advantage it confers. However, we make the remarkable observation in a prospective study that approximately one third of calves homozygous for the *c124-2A>G RNF11* mutation die before the age of six months, probably as a result of impaired control of the inflammatory response. This finding reconciles the observed high frequency of 26 % carriers, yet limited incidence of dwarfism per se, and emphasizes the higher than expected costs to farmers of the condition.

Diagnostic tests were developed for seven defects and offered to breeders. The tests were based on the detection of the causative mutation for six (CMD 1, CMD 2, both mutations for CTS, DW, HAM and AP), and on the detection of haplotypes in high LD with the causative mutation for PG. By the end of March 2013, more than 24,000 BBCB animals had been tested for at least one mutation. The main use of the diagnostic test has been to select AI bulls free of all known mutations. Consequently, this has led to a drastic and desirable reduction in the incidence of affected calves causing substantial reductions in reported peri-natal mortality. Testing of the dam population has been more limited, and – as a consequence – the population frequencies of the mutations have remained largely unchanged thus far. The cumulative frequency of all deleterious mutations being as high as 50%, selection against carrier bulls causes a substantial genetic bottleneck on the sire side. This may in turn cause the emergence of new defects due to the dissemination of as of yet unidentified deleterious mutations bound to be carried by the selected ones. Moreover, genetic progress for other economically important traits is likely to be hampered as sires with high genetic merit for these traits yet carrying one or more known defect are being culled. Breeders come to realize that it might be better to use the diagnostic test on sire and dam side in order to avoid at risk matings between carriers from the same defect while optimizing genetic progress for desired traits possibly with the aid of genomic selection.

INTRODUCTION

1- INTRODUCTION GENERALE

La population bovine mondiale actuelle, estimée à 1,3 milliards d'individus présents sur tous les continents, est constituée de deux groupes principaux : le genre *Bos indicus* et le genre *Bos taurus*. Le genre *Bos indicus*, réputé pour son adaptation aux climats chauds avec la présence d'une bosse de graisse au niveau du garrot, se retrouve principalement en Asie, en Afrique et en Amérique du Sud. Le genre *Bos taurus*, regroupant les races laitières et viandeuses les plus productives (ex. Holstein, Montbéliarde, Normande, Brune suisse, Blanc-bleu Belge, Angus, Charolais, Limousin, etc.), se retrouve principalement en Europe, en Océanie et en Amérique du Nord. Ces deux genres dérivent d'un ancêtre commun sauvage, l'auroch (*Bos primigenius*) parfois considéré comme une troisième sous-espèce (*Bos taurus primigenius*). Une analyse de microsatellites a estimé l'apparition de la divergence entre *Bos taurus* et *Bos indicus*, il y a 650.000 à 800.000 ans (MacHugh *et al.*, 1997). Les premières traces de domestication de bovins de type *Bos taurus* dateraient, selon des analyses archéologiques et génétiques récentes d'environ 10.500 ans (Bollongino *et al.*, 2012). Elle serait née au sud-est de la Turquie à partir d'un troupeau de 80 femelles d'aurochs (Bollongino *et al.*, 2012). La domestication des bovins mais aussi des autres animaux de production (moutons, chèvres, volailles,...) a été un point déterminant dans l'histoire de l'évolution de l'homme (Ajmone-Marsan *et al.*, 2010). L'homme, à l'origine nomade et chasseur, est devenu grâce à la domestication, sédentaire et éleveur. L'élevage était pour l'homme une source de lait, de viande, de fourrure mais aussi une traction pour le travail de la terre et le transport (Ajmone-Marsan *et al.*, 2010). Depuis ce temps là, l'homme n'a cessé de vouloir améliorer la production de son élevage en sélectionnant les meilleurs individus pour un caractère, dans un environnement donné (climat, relief, alimentation,...) (Andersson, 2001). Cette sélection donnera naissance aux premières races dans des monastères au Moyen-Age. Les premiers herd-books apparaîtront au début du XIX^{ème} siècle. Aujourd'hui, nous dénombrons plus de 800 races bovines différentes, réparties dans quatre finalités : laitière, viandeuse, mixte et de combat (Loftus *et al.*, 1994). Cette sélection, initialement basée sur le phénotype, s'est accélérée au cours du XX^{ème} siècle grâce, d'une part au développement des biotechnologies de la reproduction (Nicholas, 1996) et d'autre part au développement de la génétique quantitative regroupant les méthodes statistiques permettant d'estimer la valeur d'élevage et de choisir les meilleurs individus (ex. Philipsson *et al.*, 1994 ; Golden *et al.*, 2009).

Les biotechnologies de la reproduction au service de la sélection. La première insémination artificielle (IA) a été réalisée chez une chienne à la fin du XVIII^{ème} siècle par Spallanzani mais c'est au début du XX^{ème} que des Russes ont standardisé les premières

procédures d'IA sur de animaux domestiques (Foote, 2002). En 1936, les Danois ont développé la première coopérative d'IA. Elle a rapidement été copiée par l'industrie laitière aux Etats-Unis avec des recherches menées sur des centaines de milliers d'inséminations (Hamilton, 1940). L'IA augmente la capacité de reproduction d'un taureau d'élite et permet d'éviter les inconvénients liés à la saillie naturelle. On peut citer comme exemple la transmission de maladies vénériennes, le risque de blessures tant pour l'éleveur que pour la femelle lors de la monte et la détérioration des systèmes de contentions. A l'heure actuelle, le coût lié à l'achat d'un taureau d'élevage avec les risques que cela implique (ex. taureau non fécond, manque de libido, tares génétiques...) a poussé les éleveurs à recourir préférentiellement à l'IA (Vishwanath, 2003). Depuis, les méthodes de conservation et de manipulation du sperme ont évolué avec notamment la possibilité de produire du sperme sexé, apparu dans les années 80'. Ces dernières permettent à l'éleveur de pouvoir déterminer le sexe des descendants en fonction de ses besoins (femelles de remplacement ou mâles reproducteurs ou pour l'engraissement). L'amélioration des connaissances du cycle ovarien de la vache et la possibilité de recourir à la superovulation ont favorisé le développement de deux méthodes permettant de préserver la génétique des meilleures femelles. Dans un premier temps, la **fécondation *in vivo*** a été développée dans les années 50', avec le premier veau issu d'un transfert d'embryon né en 1951 (Willet *et al.*, 1951). Le développement de la MOET (Multiple Ovulation and Embryo Transfer) dans les années 80' permettait d'obtenir 5 à 10 embryons transférables par collecte (Nicholas et Smith, 1983 ; Nicholas, 1996). La découverte et l'utilisation des prostaglandines amélioreront la synchronisation des receveuses et le résultat du transfert d'embryon. Le premier veau issu du transfert d'un embryon décongelé est né en 1973 et se prénomma *Frosty II* (Wilmot et Rowson, 1973). Le faible nombre d'embryons obtenus et la mauvaise réponse de certaines donneuses de très haute valeur génétique ont poussé les scientifiques à développer la **fécondation *in vitro*** dans les années 80'. Le premier veau issu d'une maturation *in vitro* est né en 1981 (Brackett *et al.*, 1982). Dans un premier temps, les ovocytes ont été prélevés sur des ovaires récoltés à l'abattoir. Puis la technique de l'*Ovum pickup*, permettant de répéter la ponction ovarienne de manière échoguidée par voie transvaginale sur l'individu choisi a été mise au point (Galli *et al.*, 2001). Plus récemment, le clonage et la transgénése ont été développés et les premiers veaux issus des ces technologies sont nés il y a une vingtaine d'années. Ces méthodes ont pour but de cloner certaines races en voie de disparition, d'améliorer génétiquement certaines races ou de créer des individus transgéniques dont les caractéristiques pourraient être utiles à des fins médicales, pharmaceutiques et zootechniques (Basrur et King, 2005). Ces méthodes restent pratiquées à l'heure actuelle dans un nombre limité de pays et ce pour des raisons éthiques (Basrur et King, 2005).

La génétique améliore la sélection des reproducteurs et prédit leur descendance.

Dans les années 30', les chercheurs en zootechnie ont eu la volonté de pouvoir quantifier la valeur génétique des individus. Ces valeurs ou index obtenus permettraient d'améliorer leur schéma de sélection (Hazel, 1943 ; Hazel *et al.*, 1994). Des méthodes statistiques ont été développées afin de calculer ces index à partir des performances du candidat, de son ascendance et de sa descendance, en tenant compte de l'héritabilité du caractère concerné (Goddard et Hayes, 2009). Dans les années 50', Henderson développa la méthode de prédiction BLUP (*Best Linear Unbiased Prediction*) qui estime la valeur génétique d'un individu sur base des performances des apparentés en les corrigeant pour les effets du milieu (Henderson, 1975). Différents modèles ont été appliqués à cette méthode (modèle père, caractère unique, multiples caractères). Toutefois, c'est le modèle animal dans les années 80' qui permet d'avoir la meilleure précision en tenant compte du milieu et de l'ensemble des individus apparentés (Philipsson *et al.*, 1994). Ces méthodes ont eu leurs premières applications dans les races laitières au début des années 70' puis dans les années 80' pour certaines races viandeuses (MacNeil *et al.*, 2009 ; Golden *et al.*, 2009). Ces schémas de sélection, bien qu'efficaces, étaient longs et la collecte des performances coûteuse. Elles sont aujourd'hui associées à la génomique afin d'améliorer et d'accélérer cette sélection (*ex.* Meuwissen *et al.*, 2001).

La sélection, une arme à double tranchant. Les éleveurs ont favorisé l'accouplement d'individus, souvent apparentés, présentant les meilleures caractéristiques afin de fixer les caractères favorables. Comme dans la plupart des races bovines modernes hyper-sélectionnées, cette pratique a réduit considérablement la taille efficace de ces populations. L'introduction progressive de l'IA dans les pratiques d'élevage, avec l'utilisation d'un nombre limité de taureaux d'élite, a encore accentué ce phénomène. D'autre part, il est généralement admis que chaque individu, y compris un taureau d'élite, est porteur d'une poignée de mutations délétères, à l'état hétérozygote, qui seront alors rapidement disséminées dans la population, ceci entraînant l'apparition de tares récessives dans les générations suivantes (*ex.* 1000 Genome Project Consortium, 2010). A titre d'exemple, dans les années 90', la BLAD (Bovine Leukocyte Adhesion Deficiency) en race Holstein pie-noire a coûté près de 5 millions de dollars US par an dans l'industrie laitière nord-américaine (Shuster *et al.*, 1992).

Cette introduction présentera dans un premier temps, les origines et les différentes caractéristiques la race Blanc-bleu Belge (BBB). Ensuite, seront décrits les outils de génomique (outils moléculaires) et de génétique quantitative (méthodes statistiques) utilisés dans les programmes de sélection bovine.

2- LA RACE BLANC-BLEU BELGE

A. Origine et évolution

L'origine de la sélection de la race BBB remonte à la moitié du XIX^{ème} siècle. La création de la race Shorthorn ou Durham (Figure 1), avec sa morphologie en forme de « brique avec une patte à chaque coin », a connu un succès important en Europe et en Amérique du Nord. En 1841, le gouvernement belge favorise son importation et le croisement avec les races locales fortement hétérogènes. Ces croisements étant décevants, la Shorthorn a rapidement perdu de son importance. Elle laissera une trace de son passage encore visible sur le BBB actuel avec notamment la mutation du gène *Rouan*, responsable de la couleur de la robe caractéristique et son rôle majeur dans le déterminisme de la maladie des génisses blanches (*White Heifer Disease* = WHD).

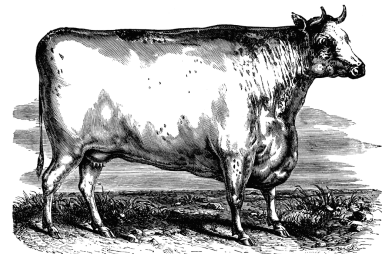


Figure 1 : Vache Durham de la fin du XIX^{ème} siècle.



Figure 2 : Présentation de taureaux Bleus (1900) (Hanset, 1996).

En 1896, une volonté d'uniformisation apparaîtra avec la création du Herd-Book Hesbignon dont l'objectif sera de développer une race locale en se débarrassant des défauts de la Durham. Au début du XX^{ème} siècle, le croisement entre la Durham et la pie-noire de Hollande donnera naissance à la race Bleue du Limon (Figure 2). Ce bétail bleu va progressivement envahir les différentes régions de Moyenne et Haute-Belgique.

Après la première guerre mondiale, le programme de sélection bovine fut remis en question. Le 16 août 1919, un arrêté royal (AR) prévoit la création de syndicats et de fédérations d'élevages dont l'objectif sera de contrôler l'enregistrement des animaux, la gestion des pedigrees et la productivité, avec la mise en place du contrôle laitier. Cet AR reconnaît également officiellement cinq races dont la race Bleue fait partie. La sélection se portera sur un bétail rectangulaire à « deux fins » avec une musculature moyenne et une bonne production laitière (4.000 litres à 3,5 % de matière grasse). Les individus extrêmes : trop ou trop peu musclés (type cul-de-poulain) sont éliminés de ce schéma de sélection.

En 1938, la diversité des couleurs au sein de la race « bleue » pose problème et on préconise de fixer la robe blanche et de la nommer la race « Blanche de la Moyenne et Haute Belgique ». Ceci est principalement suivi en Hainaut avec la recherche d'un type morphologique que l'on appellera le « Grand plat (blanc) d'Ath » (Figure 3). Ce type, à la musculature trop faible, ne suit pas l'évolution de la consommation qui encourage la viande de qualité produite par les

bovins de forme « parallépipède aux angles arrondis ». La race dite « blanche » est alors abandonnée et remplacée par le retour de la race de « Moyenne et Haute Belgique ». Pendant ce temps, la région du Condroz-Hesbaye favorise la sélection de taureaux musclés à la conformation remarquable. Ces taureaux seront présentés aux concours Nationaux de 1939 (*Gamin*) et de 1949 (*Rieur de Rostenne*). En 1945, des « zones raciales » sont définies, limitant l'utilisation de taureaux, lors de montes publiques, à leurs zones. On retrouve la zone de la Moyenne et Haute Belgique comprenant les provinces du Brabant, du Hainaut, de Namur et du Luxembourg. La désignation d'un taureau avec une conformation supérieure comme champion national : *Carlos*, en 1953 va marquer les esprits.



Figure 3 : Bovins de la « race Blanche de la Moyenne et de la Haute Belgique » (www.hbbbbb.org).

L'après-guerre va connaître une transition radicale avec le développement de l'IA et l'ouverture d'un centre d'insémination artificielle (CIA) dans chaque province. Le CIA de Loncin (Province de Liège) verra l'entrée de *Gédéon du Vieux-Château de Maurenne* (Figure 6) (descendant de *Rieur de Rostenne*) au développement musculaire supérieur à celui de *Carlos* qui était probablement hétérozygote porteur de la mutation affectant le gène de la *Myostatine* (*mh/+*), introduisant l'allèle *mh* à large échelle dans la race. Le début des années 50' connaîtra le développement de la technique de la césarienne, ainsi que l'utilisation des antibiotiques en médecine vétérinaire permettant la naissance à terme de veaux hyper-musclés (culards) vivants (Derivaux, 1955 ; Hanzen *et al.*, 1999). De plus, la meilleure valorisation commerciale et bouchère (Hanset, 2004) a rendu la sélection des animaux culards possible et rentable.

En 1969, le Herd-Book de la race de Moyenne et Haute Belgique continuera à imposer la sélection d'une race mixte à dominante viande (Figure 4), notant que la production d'animaux exclusivement à viande est possible mais qu'ils doivent être retirés de la reproduction. Cette voie de sélection ne fait pas l'unanimité chez les éleveurs. Dès 1964, deux taureaux viandeux (probablement



Figure 4 : Vache des années 60' (www.hbbbbb.org).

homozygotes pour la mutation affectant le gène de la *Myostatine* : *mh/mh*), *Ganache de Maufontaine*, arrière-petit-fils de *Gédéon du Vieux-Château de Maurenne* et *Valseur d'Ochain*, petit-fils et arrière-petit-fils de *Gédéon du Vieux-Château de Maurenne* (Figure 6) feront leur entrée au CIA de Loncin. Ces taureaux seront utilisés de manière préférentielle sur des descendantes de *Gédéon du Vieux-Château de Maurenne*. Le type viandeux (*mh/mh*) se fixe de manière croissante dans les élevages belges avec une augmentation du nombre d'élevages adoptant la spéculation de type vache allaitante, passant de 40.000 vaches allaitantes en 1970 à

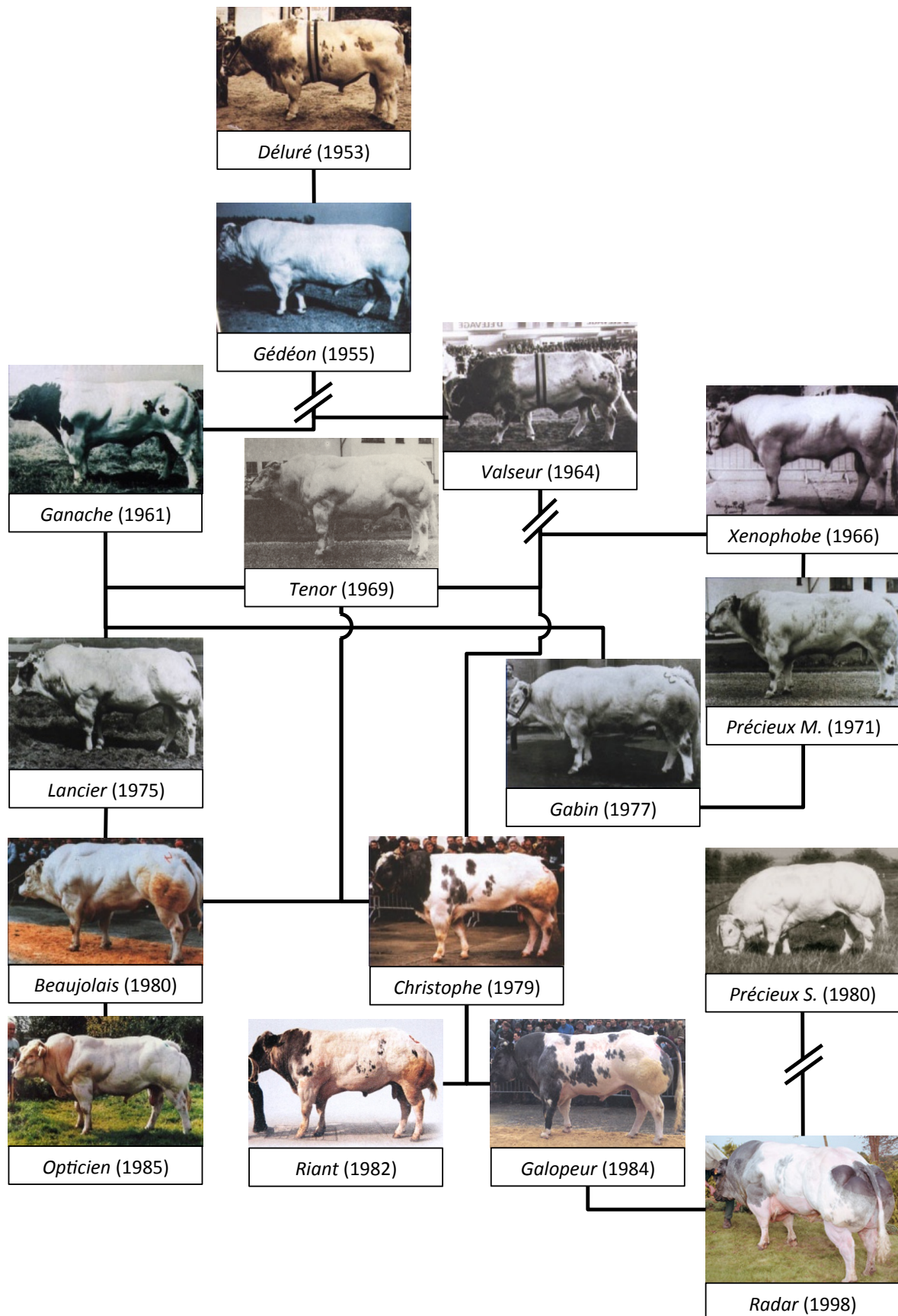
450.000 en 1994 (Hanset, 1996). La race de Moyenne et Haute Belgique devient alors la race Blanc-bleu Belge.



Figure 5: Taureau des années 70' (www.hbbbbb.org).

En 1973, le Herd-Book de la race Blanc-bleu Belge (HBBBB) est créé et devient responsable du livre généalogique, de la sélection et de la promotion de la race. En 1974, l'assemblée générale du HBBBB définit deux rameaux bien distincts, le rameau à viande (Figure 5) et le rameau mixte aux deux fins. De la fixation du gène culard à nos jours, les éleveurs-sélectionneurs n'ont cessé de vouloir améliorer le développement musculaire. Ceci s'est concrétisé par une augmentation rapide de la masse musculaire au cours de ces deux dernières décennies, surtout avec l'utilisation massive d'une poignée de taureaux des années 80' qui ont marqué la race par leur développement musculaire : *Galopeur des Hayons*, *Opticien d'Au Chêne*, *Riant* (tous descendants de *Gédéon du Vieux-Château de Maurenne*) et *Précieux de Somme* (Figure 6).

Aujourd'hui le BBB se sélectionne en race pure avec des rendements de carcasses extraordinaires mais il a également trouvé son utilité en croisement terminal sur des races laitières en Europe et sur des races plus exotiques telles que le zébu afin d'augmenter le rendement des carcasses et la tendreté de la viande.



- 27 -

B. Population

En 2010, la population bovine belge comptait environ 2.600.000 individus. La population BBB en représente 50 %, soit 1.300.000 individus et 95 % du bétail viandeux. Au livre généalogique de la race, 90.000 vaches (17 %) et 2.500 taureaux sont inscrits. On compte environ 25.000 troupeaux de bovins en Belgique dont 17.000 élevant du BBB avec une moyenne de 30 vaches par troupeau (1 à > 100) (URL: <http://economie.fgov.be/fr/modules/pressrelease/statistiques/economie/recensementagricoledemai2010.jsp>). En 2011, on dénombrait environ 200 taureaux d'IA répartis dans les quatre CIA présents en Belgique (Belgian Blue Cattle International = BBCI, Belgian Blue Group = BBG, Fabroca et Génétique Avenir Belgimex). Les troupeaux d'éleveurs-naisseurs se retrouveront principalement dans le sud de la Belgique et les engraisseurs dans le nord.

C. Caractéristiques

a) Conformation et morphologie

La race BBB se distingue par son développement musculaire extrême, tant au niveau des quartiers postérieurs que des quartiers antérieurs. Le poids des taureaux adultes varie en moyenne entre 1.100 et 1.250 kg, avec une taille au garrot allant de 145 à 150 cm. Certains taureaux peuvent atteindre plus de 1.300 kg pour une taille supérieure à 153 cm. Le poids des vaches adultes oscille entre 700 et 750 kg pour une taille au garrot entre 132 à 134 cm. Certaines vaches peuvent atteindre le poids de 1.000 kg avec une taille supérieure à 140 cm. Le poids des veaux à la naissance est en moyenne de 47 kg pour les mâles et de 42 kg pour les femelles. Le poids des veaux a considérablement diminué au cours des ces dernières années avec une sélection pour des plus petits formats mais avec une bonne vitalité, permettant d'éviter les pertes liées à la naissance de gros veaux (cardiomyopathies, rigidité articulaire congénitale, syndrome de détresse respiratoire aiguë...)

Afin de maintenir une qualité et un format viandeux au niveau des animaux, le HB BBB a fixé les standards de la race (Annexe 1) qui reprennent des caractères de conformation musculaire, de morphologie et d'aplombs définissant ce que l'on recherche et ce que l'on exclut.

b) Robe

Malgré son nom, et contrairement à d'autres races, la race BBB a une robe assez polymorphique mais dominée par trois couleurs, le pie-noir, le pie-bleu et le blanc (Figure 7). Ces patterns caractéristiques de la race sont déterminés par trois loci majeurs, le locus *E* (Extension), le locus *R* (Rouan) et le locus *S* (white spotting = pie) (Olson T.A., 1999) (Tableau 1).



Figure 7: Robes et génotypes de trois taureaux d'IA (BBG) : *Adajio de Bray* sous robe noire-pie = $E^d/E^d, r^+/r^+, s/s$, *Lotto VD Vloeikenshoeve* sous robe pie-bleue = $E^d/-, R/r^+, s/s$ et *Séduisant de Fooz* sous robe blanche = $E^d/-, R/R, s/s$.

Le locus Extension est responsable de la pigmentation de la peau dans la plupart des espèces, y compris dans l'espèce bovine. Le gène responsable de la variation de la couleur a été identifié et code pour le récepteur à la mélanocortine de type 1 (melanocortine 1 receptor = MC1R ou melanocyte-stimulating hormone receptor = α -MSHR). Ce récepteur, situé au niveau des mélanocytes, est stimulé par l'hormone mélanotrope et active la synthèse d'eumélanine via l'action de tyrosinases (Lin et Fisher, 2007). En l'absence de stimulation, les mélanocytes produiront de la phéomélanine, un pigment rouge. L'allèle sauvage (E^+) est responsable de la robe brune – brune foncée que l'on retrouve dans les races Jersey, Brunes Suisses. Une mutation non-synonyme (p.Leu99Pro) dans le gène *MC1R* (E^d = dominant) provoque une activation constitutive du récepteur à la mélanocortine et une production continue d'eumélanine, donnant la couleur noire que l'on retrouve en Holstein (Klungland *et al.*, 1995). La robe rouge (Simmental, Hereford, Holstein pie-rouge...) est due à une délétion d'1 bp dans le gène *MC1R* (e = récessif), empêchant la formation d'eumélanine (Klunglang *et al.*, 1995 ; Joerg *et al.*, 1996). En BBB, la plupart des individus sont génétiquement homozygotes noirs (E^d/E^d) (Tableau 1) (Hanset, 1985). Cette couleur a probablement été amenée par les croisements avec des races pie-noires venant de Hollande (Charlier *et al.*, 1996). La couleur rouge, vestige de la race Shorthorn, est exclue des standards de la race BBB. Elle peut se rencontrer occasionnellement chez le produit de certains croisements, signifiant que l'allèle e (Tableau 1) est présent dans la population BBB mais à très faible fréquence.

L'expression de la couleur noire est modifiée par le locus *Rouan* et le locus *S*. Le locus *Rouan* est un héritage de la race Shorthorn importée dans les années 1900 et croisée avec les

rares locales (Charlier *et al.*, 1996). Une mutation non-synonyme dans le gène *MGF* (Mast Growth Factor), aussi appelée *KITLG* (KIT Ligand, le ligand du récepteur c-Kit intervenant dans la migration et la survie des mélanocytes au niveau de la peau) (p.Ala193Asp). Elle est associée à la robe rouanne dans les races BBB et Shorthorn (Seitz *et al.*, 1999). Chez les individus homozygotes sauvages (r^+/r^+) la robe sera noire, chez les individus homozygotes mutés (R/R) elle sera blanche et chez les hétérozygotes pour la mutation (R/r^+) le mélange de poils colorés (noirs en BBB) et de poils blancs, donnera la couleur bleue à la race, d'où, le nom Blanc-bleu Belge. A de nombreuses reprises, les éleveurs ont tenté de fixer cette couleur de robe dans la race comme la Bleue du Limon, mais sans jamais y parvenir. Le caractère co-dominant de l'allèle muté (R) et de l'allèle sauvage (r^+) empêche de fixer la robe bleue dans la race BBB puisque le croisement d'individus bleus donnera 25 % d'individus noirs, 25 % de blancs et 50 % de bleus (Tableau 1).

Dans la race BBB, les individus colorés ne présentent pas une robe uniformément colorée mais présentent un pattern de taches blanches et de taches colorées (bleues ou noires). Ces robes pie-noire ou noir-pie et pie-bleue ou bleu-pie, en fonction de la proportion de zone colorée, (Figure 7) sont causées par le locus *S* (spotting) où l'allèle *s*, associé à la présence de taches blanches, est récessif (Tableau 1). Le locus *S* a été localisé sur le chromosome 6 par étude de liaison d'individus F2 issus du *backcross* entre des individus Hereford et des femelles d'une race composée de Red Angus, de Charolais et de Tarenaise (Grosz et McNeil, 1999). Il s'agit plus que probablement d'un allèle du gène *c-Kit* (gène codant pour le récepteur Kit ou MGFR/SCFR = Mast/Stem cell Growth Factor Receptor) mais le variant causatif n'est pas identifié à ce jour. En BBB, les individus sont probablement tous pies (s/s) (Charlier *et al.*, 1996). La proportion de zone colorée n'est pas fixe et la robe peut varier de quasiment blanche à quasiment colorée. Une étude récente menée dans la race Holstein Pie-noire a identifié trois loci majeurs, KIT sur le chromosome 6, MITF (Microphthalmia Transcription Factor) sur le chromosome 22 et potentiellement PAX5 (Paired box 5) sur le chromosome 8 expliquant 24 % de la variance génétique, confirmant le caractère complexe de la proportion de robe colorée chez les bovins (Hayes *et al.*, 2010). En BBB, la sélection de la robe par les éleveurs a évolué depuis sa création. Les individus blancs qui étaient recherchés dans les années 80' (Hanset, 1985) ont été remplacés dans les CIA et les élevages par des animaux de plus en plus colorés, avec même des individus presque entièrement noirs (Figure 7 : *Adajio de Bray*). Il serait possible d'envisager un jour la sélection d'un individu sans tache (S/S) répondant mieux aux attentes des éleveurs en terme de couleur. Cette évolution est sûrement esthétique mais elle diminuerait aussi la sensibilité à la gale et réduirait le risque pour la maladie des génisses blanches (*cf.* E. La race BBB et ses défauts, p. 39) (Hanset, 1969).

Loci	Gènes	Allèles	Mutation	Génotypes et phénotypes associé	Mode d'hérédité
Extension	MC1R	E^+	WT	E^+/E^+ ou E^+/e = Brun foncé $E^d/-$ = Noire e/e = Rouge	$E^d > E^+ > e$
		E^d	p.Leu99Pro		
		e	Délétion 1 bp		
Rouan	KITLG	R	p.Ala193Asp	$+/+$ = Coloré $R/+$ = Rouan R/R = Blanc	$R = r$
		$+$	WT		
White spotting	KIT ?	S	WT	$S/-$ = Sans tâche s/s = Pie	$S > s$
		s	?		
Blanc-dos	KIT	Cs	Translocation BTA6 > BTA29	$Cs/+$ ou Cs/Cs = Blanc dos $+/+$ = Pie	$Cs > +^*$
		$+$	WT		

Tableau 1 : Loci, gènes et mutations intervenant dans le déterminisme de la robe de la race BBB. Mode d'hérédité : « = » = co-dominance, « > » = dominant > récessif. (WT = type sauvage, = indéterminé, * = chez les individus colorés = semi-dominant)

Récemment, suite à l'entrée au CIA BGG du taureau *Félicien de Courriaulx*, la robe blanc-dos (ligne du dos blanche et côtés colorés) a vu sa fréquence dans les troupeaux augmenter. L'analyse d'individus Blanc-dos ($Cs/-$) a identifié une translocation duplicative, via un intermédiaire circulaire, d'une portion du chromosome 6, contenant le gène codant pour le récepteur Kit, vers le chromosome 29, associée à la robe blanc-dos (Durkin *et al.*, 2012 ; Annexe 6).

c) Reproduction

La principale caractéristique du mode de reproduction en BBB est le vêlage systématique (95 à 99 %) par césarienne élective, sans traction préalable. Cela résulte principalement de la conséquence d'une disproportion fœto-maternelle (Kolkman *et al.*, 2010). Cette disproportion est le fruit de la sélection intensive pour le développement musculaire au détriment de l'ossature qui a conduit à une réduction importante du diamètre de la filière pelvienne (Coopman *et al.*, 2003). La systématisation de la césarienne, dès le début du part, en race BBB,

réduit considérablement la mortalité périnatale. Il a également été démontré qu'il est conseillé d'attendre un effacement complet du col avant de réaliser la césarienne pour avoir une meilleure adaptation à la vie extra-utérine (Uystepuyst *et al.*, 2002).

En terme de performances de reproduction, la race BBB est une race relativement précoce pour une race viandeuse, avec un âge à la puberté d'un an et un âge au premier vêlage de 27-36 mois (Hanzen *et al.*, 1994). Certains troupeaux ont des génisses qui vêlent à 24 mois. La conformation étant le critère déterminant pour la mise à la reproduction, une alimentation soutenue et suivie est nécessaire pour atteindre cet objectif. La durée de gestation est courte, environ 282 jours, contrairement aux autres races viandeuses où l'on atteint en moyenne 288-290 jours. L'intervalle moyen entre deux vêlages est de respectivement 401 jours et 435 jours, pour des troupeaux élevant les veaux au seau et les troupeaux élevant les veaux au pis (Hanzen *et al.*, 1994). Dans la race BBB, l'IA a une place importante dans la reproduction, plus de 50 % des veaux en sont issus. En 2012, un panel d'environ 200 taureaux était proposé aux éleveurs par les quatre centres de commerce de semences en Belgique (BBCI, BBG, Fabroca et Génétique Avenir Belgimex) avec une production annuelle d'environ 1.900.000 doses de sperme.

Au niveau des mâles, peu d'études ont été réalisées afin d'évaluer la fertilité. En 2006, le périmètre scrotal et la semence de taureaux BBB et Holstein-Frisons, évalués avant leur entrée au CIA, ont été comparés. Il ressort de cette étude qu'une grande proportion de taureaux BBB étaient refusés à cause d'un sperme de mauvaise qualité et/ou d'un périmètre scrotal trop petit, suggérant un risque accru d'hypoplasie et/ou de dégénérescence testiculaire (Hoflack *et al.*, 2006). Cette hypothèse a été confirmée avec une augmentation significative de l'index de dégénérescence du sperme et de la quantité de tissus conjonctifs au niveau des testicules (Hoflack *et al.*, 2008). La libido était similaire dans les deux races.

En BBB, l'utilisation de la fécondation *in vitro* se limite à des vaches de haute valeur génétique. Dans les années 90', le nombre de récoltes était d'environ 1.500 par an. Cette méthode de reproduction est en diminution puisqu'en 2011, les vétérinaires de l'AWE ont réalisé 500 récoltes avec un nombre moyen d'embryons transférables de 4,70 par récolte.

De manière générale, le taux de consanguinité élevé et la sélection importante pour la viande, a réduit considérablement les performances de fertilité, tant chez les femelles que chez les mâles. A l'heure actuelle, il est fréquent d'entendre les éleveurs se plaindre « *de bêtes qui ne viennent pas en chaleurs ou des taureaux qui ne saillissent pas, qui saillissent sur la tête ou qui ne sont pas féconds* » (observations personnelles).

d) Production

En Belgique, la consommation moyenne de viande est d'environ 30 kg par an et par habitant, dont 20 kg de viande rouge. La race BBB fournit à elle seule près de 60 % de la viande bovine et près de 75 % de la viande rouge en Belgique. Il est donc important d'améliorer la production et la qualité de viande BBB (Hanset, 2004). De plus, le gène « culard » est responsable à lui seul d'une augmentation de 20 % de la masse musculaire par augmentation du nombre de fibres (hyperplasie) (Grobet *et al.*, 1997).

La sélection des bovins viandeux vise à améliorer les performances de croissance. Les éleveurs désirent donc avoir des bovins avec une croissance rapide et élevée, qui consomment moins d'aliments. Pour évaluer les bovins en engraissement, ils mesurent le poids initial, le poids final, le gain quotidien moyen (GQM) en kg de poids vif par jour d'engraissement, la consommation journalière et le taux de conversion alimentaire en kg d'aliments consommés par kg de croissance. Ces mesures sont difficilement réalisables en ferme. En BBB, le GQM est de 1,6 Kg par jour en station au Centre de Sélection Bovine (CSB) (Gengler *et al.*, 1995) alors qu'en engraissement à la ferme il est d'environ 1,2 kg par jour (Coopman *et al.*, 2007). L'efficacité alimentaire est d'environ 5 kg d'aliment consommé par kg de poids vif (Gengler *et al.*, 1995), soit 9 % de moins par rapport à des races « non culardes » (Hanset *et al.*, 1987). Les taureaux à l'engrais seront abattus vers l'âge de 18 mois avec un poids moyen de 650 kg. Comparativement aux autres races viandeuses, le BBB a un GQM plus faible que le Charolais (1,8 kg/jour). Au niveau de l'efficacité alimentaire, les races plus rustiques peuvent se contenter de grandes quantités de fourrages grossiers tandis que le BBB, limité par son volume d'ingestion, nécessite une ration plus concentrée.

Le point fort de la race BBB, en terme de production, est son rendement à l'abattage (rapport entre le poids de la carcasse dépecée et éviscérée et le poids vif à jeûn), qui peut atteindre 70 % avec 80 % de viande, alors qu'il atteint 60 à 65 % dans les races à viandes conventionnelles (Wheeler *et al.*, 2005). Depuis 2007, l'Union Européenne a établi une grille d'évaluation permettant de déterminer la valeur commerciale d'une carcasse de gros bovin (Règlement CEE N° 1208/81). La classification intègre trois critères : la catégorie (mâles/femelles, castrés/non castrés, vêlées/non vêlées), la conformation et l'engraissement (CW3C : www.cw3c.be). La conformation est évaluée à l'aide de 6 classes (Figure 8):

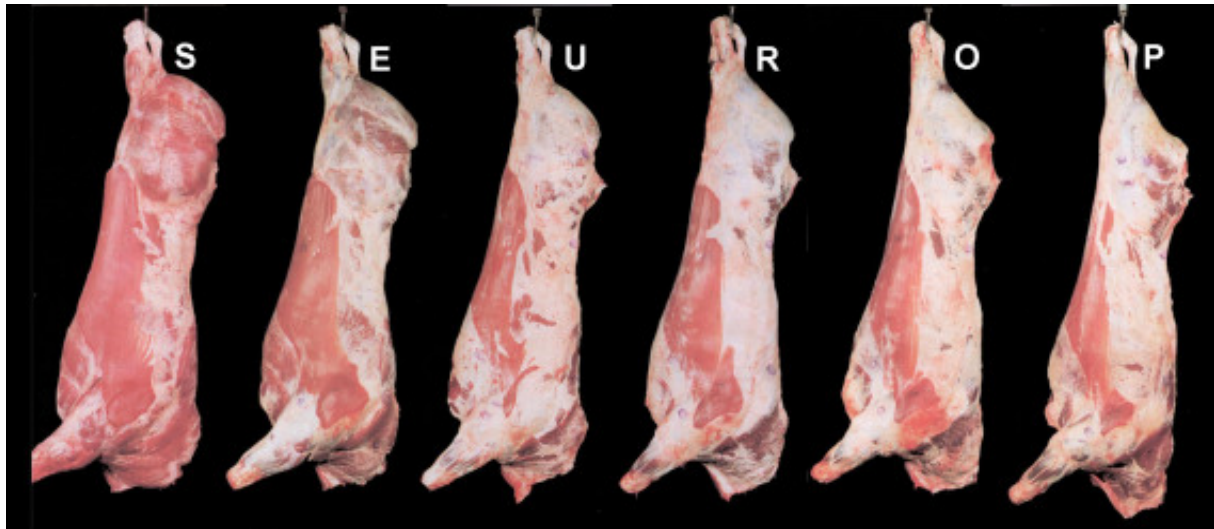


Figure 8: Classification SEUROP pour évaluer la conformation des carcasses

- S: supérieure - profils musculaires extrêmement convexes;
- E: excellente - profils musculaires convexes à super-convexes;
- U: très bonne - profils musculaires convexes dans l'ensemble;
- R: bonne - profils musculaires rectilignes dans l'ensemble;
- O: assez bonne - profils musculaires rectilignes à concaves ;
- P: médiocre - profils musculaires concaves à très concaves.

L'état d'engraissement se détermine en utilisant 5 classes:

- 1: très faible - couverture de graisse très faible à inexistante;
- 2: faible - légère couverture de graisse, muscles apparents;
- 3: moyen - muscles couverts de graisse sauf cuisse et épaule;
- 4: fort - muscles couverts de graisse, cuisse et épaule ± visibles;
- 5: très fort - carcasse recouverte de graisse, dépôt thoracique important.

En Belgique, en 2011, près de 85 % des carcasses sont classées S et E, alors que dans les autres pays européens elles sont majoritairement classées U, R et O (CW3C : www.cw3c.be). De plus, la découpe spécifique au BBB augmente la qualité en bons morceaux de 35 % et réduit les déchets.

En plus de la quantité, la viande de BBB répond aux exigences des consommateurs en terme de santé publique, grâce à son excellente qualité. C'est une viande tendre contenant une forte proportion de fibres musculaires blanches ou de type II (Bellinge *et al.*, 2005) et 20 à 30 % de collagène en moins (Uytterhaegen *et al.* 1994, Georges, 2010). En outre, c'est une viande maigre contenant principalement des acides gras polyinsaturés (De Smet *et al.*, 2000).

D. Programme de sélection

Les critères de sélection de la race BBB sont principalement gérés et définis par le HBBBB. La première mission du HBBBB, en collaboration avec l'Association Wallonne de l'Élevage (AWE), est la tenue du livre généalogique et l'évaluation génétique des individus (taureaux et femelles inscrits). A l'heure actuelle, on dénombre environ 90.000 vaches et 2.500 taureaux inscrits avec en moyenne, par an, 33.500 femelles et 1.900 mâles expertisés au HBBBB. Lors de l'inscription d'un individu au HBBBB, après en avoir vérifié la parenté, un certificat généalogique (pedigree) est délivré au détenteur du bovin. Sur ce certificat, on retrouve les coordonnées du naisseur, du détenteur, le nom de l'individu (nom et suffixe de l'élevage), la date de naissance, la robe, le sexe, la boucle auriculaire mais surtout le pedigree remontant à trois générations (Annexe 3). Ce dernier permet de garantir la traçabilité et les origines de l'individu lors de son utilisation pour la reproduction. Au niveau du HBBBB, il est bien sûr possible de générer le pedigree « complet » de chaque individu inscrit, pour autant que ses ascendants aient été inscrits.

La sélection en BBB, conseillée par le HBBBB, vise à l'heure actuelle à améliorer la rusticité (facilité d'élevage), la fertilité, les aplombs, la viabilité et l'aptitude à boire ; à maintenir une excellente conformation tout en évitant une diminution de la taille, de la longueur et de la largeur au niveau du dos, du format et en réduisant les tempéraments agressifs. Pour atteindre ces objectifs, le HBBBB a développé un système d'index évaluant ces différents critères. L'évaluation génétique a pour but d'estimer, de la manière la plus précise possible, le potentiel génétique de chaque animal afin d'optimiser la sélection. En race BBB, les individus sont évalués sur leur performance propre, avec la cotation linéaire et éventuellement les performances des taureaux élevés en station au CSB. Le potentiel génétique est, quant à lui déterminé grâce à l'évaluation de la descendance (progeny testing). L'IA ayant une grande place dans la reproduction, la race BBB est bien adaptée au progeny testing vu le grand nombre de descendants par taureau (Hanset *et al.*, 1990).

Le système de cotation linéaire en BBB a été développé en 1987, il a été utilisé pour la première fois en 1988 sur des taureaux évalués en station (CSB) et généralisé à toutes les femelles inscrites en 1994 (Hanset *et al.* 1994). Le système évalue 22 caractères, regroupés en trois groupes, évaluant i) le format, ii) la conformation et l'ossature et iii) les aplombs (Annexe 2). Ces caractères sont évalués et notés par des techniciens expérimentés sur une échelle de 1 à 50 avec une moyenne à 25. En plus de ces 22 caractères, cinq notes partielles, évaluant la taille, la musculature, le type viandeux (culard), les aplombs et l'apparence générale, permettent de déterminer la note finale de l'individu. Ces caractères sont notés de 50 à 100. A cette cotation, on ajoute une évaluation des aplombs (déviation et bouleture des antérieurs et des postérieurs,

genoux fléchis et présence de soufflette au niveau des jarrets) avec une note de 0 (absent) à 5 (présent). Après chaque évaluation, un document reprenant cette cotation sous forme d'histogramme est mis à la disposition de l'éleveur. Il lui permet de situer son bovin par rapport à la moyenne nationale (Annexe 3).

L'évaluation sur la descendance se fait en trois étapes : la première à la naissance, la deuxième vers l'âge de 14 mois et la troisième, uniquement chez les femelles adultes, via l'utilisation du système de cotations linéaires. A la naissance, on estime le poids et on évalue la conformation, la durée de la gestation, l'aptitude à boire, la vitalité, les défauts des membres antérieurs et postérieurs, la mortalité et les défauts de gueule. A 14 mois, on estime le poids à partir du périmètre thoracique, on mesure la taille et on évalue la conformation, la combinaison poids-conformation, la mortalité, les défauts des membres antérieurs et postérieurs et les défauts de gueule. Enfin, la cotation linéaire est réalisée sur les femelles inscrites de 15 à 56 mois comme défini précédemment chez les taureaux d'IA.

Plus récemment, l'index poids-carcasse des taurillons descendants de taureaux d'IA, a été développé dans le but d'évaluer et d'estimer les performances de rendement à l'abattage.

Le calcul de la valeur d'élevage (EBV = *Estimated Breeding Value*) permet d'estimer les effets transmissibles des gènes d'un reproducteur à partir de l'évaluation de sa descendance. Cette méthode de calcul doit tenir compte du fait que tous les descendants d'un même taureau ne sont pas élevés dans les mêmes conditions et doit donc tenir compte du milieu. Pour ce faire, on utilise la méthode BLUP qui donne une estimation de la valeur génétique transmissible, indiquant la performance moyenne attendue chez les descendants du taureau. Dans un premier temps, la méthode BLUP a été utilisée dans un modèle père, ne prenant en compte que des informations du pedigree du côté paternel pour évoluer ensuite dans un modèle animal qui tient compte de tous les liens de parentés dans la population BBB (ascendants, descendants et collatéraux). Deux autres modèles, augmentant ainsi la précision, sont utilisés en fonction de la corrélation qu'il peut exister entre deux caractères (Hanset *et al.*, 1994). Dans le cas d'un caractère n'ayant pas de relation avec d'autres ou étant spécifique de la période de mesure, un modèle unicaractère sera utilisé. Pour les caractères corrélés avec d'autres (*ex.* la taille et la longueur), le modèle multicaractère sera beaucoup plus précis pour cette estimation. L'évaluation BLUP est calculée en écart par rapport à la moyenne, dans l'unité du caractère considéré et est ensuite exprimée en déviation standard génétique, grâce à un index dont la moyenne est égale à 100 (10 points de l'index valant une déviation standard). L'index obtenu pour un caractère est associé à une valeur de précision (r^2 ou répétabilité) de cette estimation (www.awenet.be/nouvawe/commun/service/viande/evalgenbbbexplicationgen.php). Cela permet ainsi aux éleveurs de comparer aisément des taureaux entre eux afin de choisir le plus adéquat pour le critère recherché.

Toutes les évaluations génétiques des taureaux d'IA sont publiées semestriellement (Annexe 4). On y retrouve le pedigree de chaque taureau, les évaluations en première et deuxième visites, les cotations linéaires et les index poids-carcasses. De manière à faciliter la lecture de ces index, les notes très favorables seront surlignées en vert et les notes défavorables en rouge, pour un taureau et un caractère donné. Pour être officiels et publiés, les index doivent avoir un seuil minimum de précision de 0,50 pour tous les caractères des premières et des deuxièmes visites. En ce qui concerne les cotations linéaires, tous les index sont publiés pour autant que l'index de la note partielle de musculature ait une précision supérieure ou égale à 0,70. Pour les index poids-carcasse, il faut une précision d'au moins 0,80. En plus de ces règles, le taureau doit avoir au moins 10 descendants répartis dans minimum 2 troupeaux différents. La critique qui pourrait être adressée à ce système, concernerait la subjectivité de ces évaluations qui sont basées sur des critères observables et non mesurables (à l'exception de la taille).

Le calcul de ces index permet également d'évaluer la réponse de la sélection mise en place en BBB sur ces différents critères (Figures 9 & 10). On peut voir, que de manière générale, de 2000 à 2009, les index de la conformation et de la musculature augmentent alors que ceux de la taille et des aplombs diminuent. En réponse à ces diminutions, le programme de sélection a été modifié pour ces deux critères, ce qui a permis d'améliorer cette tendance rapidement.

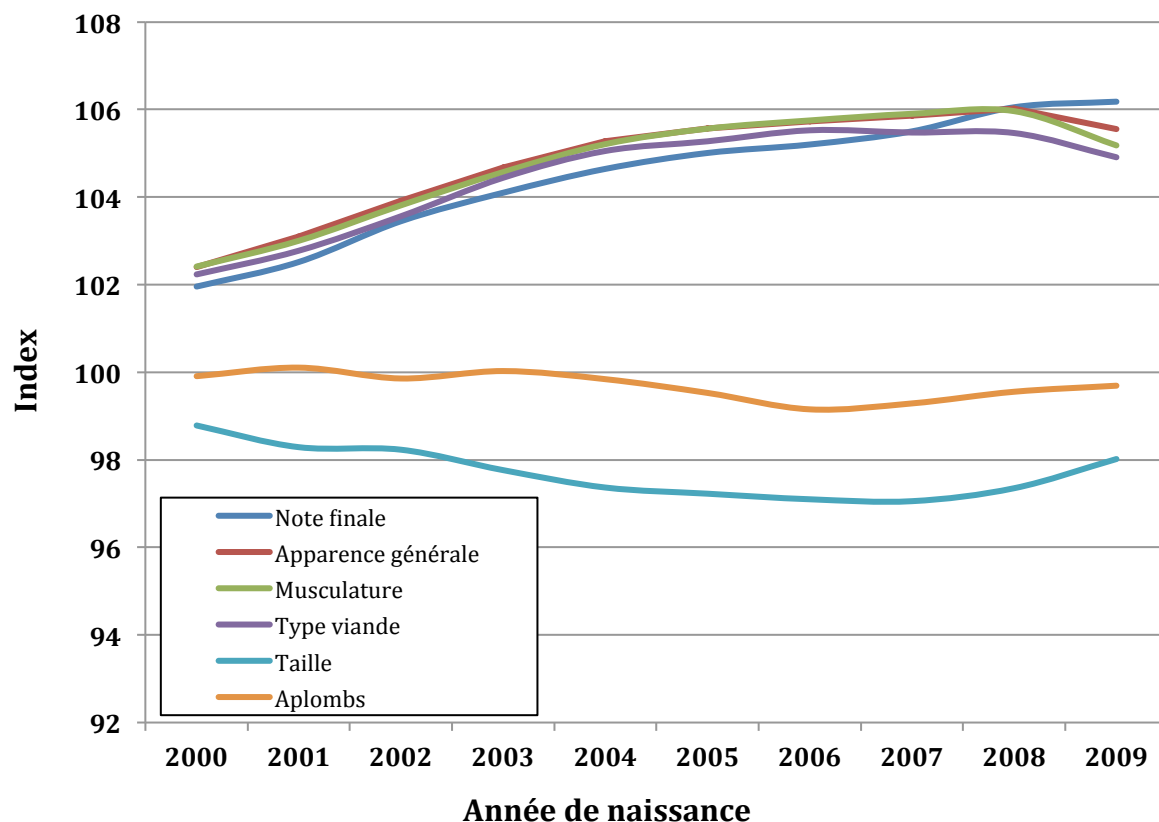


Figure 9 : Evolution des index obtenus en 2^{ème} visite en fonction de l'année de naissance (Source HBBBB).

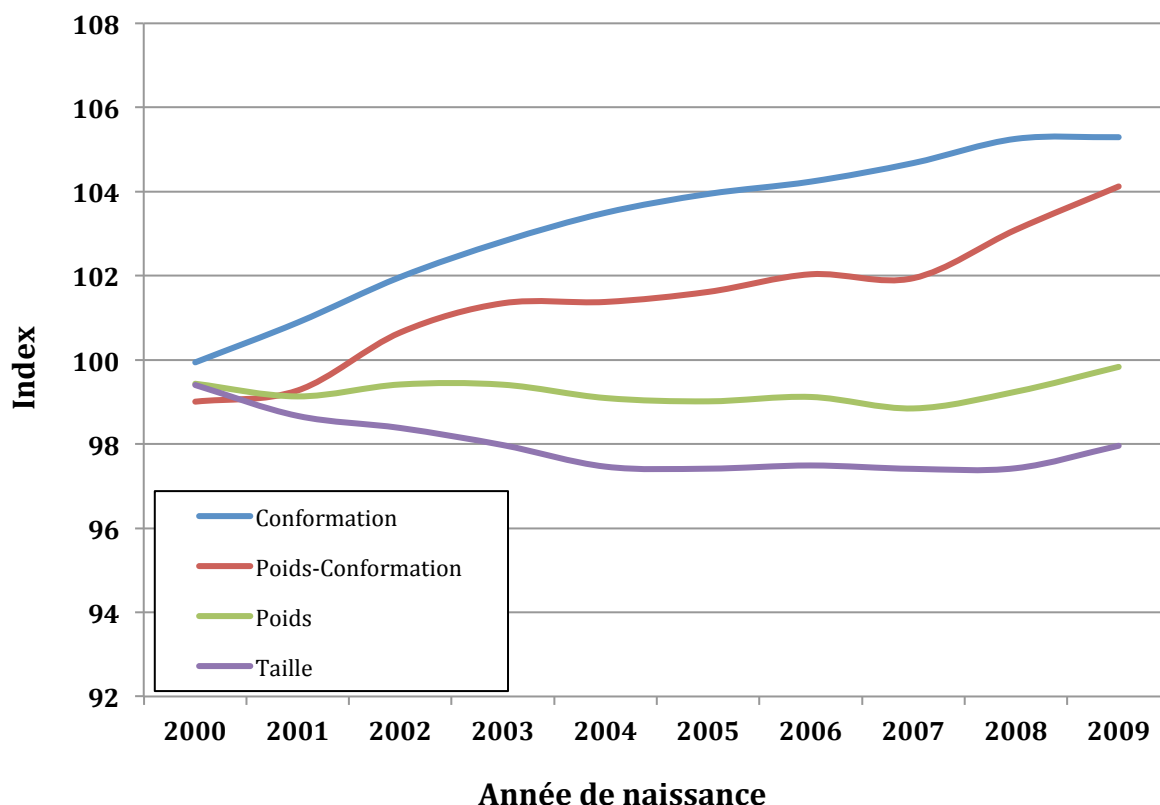


Figure 10 : Evolution des notes partielles des cotations linéaires en fonction de l'année de naissance (Source HBBBB).

E. La race BBB et ses défauts

Le point négatif majeur de la race BBB est, à mon avis, l'image de l'opinion publique liée à sa musculature hors normes et la systématisation de la **césarienne** comme mode de mise-bas. L'hypermusculature, due en partie à la sélection d'une mutation perte-de-fonction dans le gène de la *myostatine*, pourrait être considérée comme un défaut. A une époque où le bien-être animal et l'agriculture biologique prennent une importance non négligeable dans l'opinion publique, la césarienne élective (99 %) des cas rend le BBB fortement critiqué et pointé du doigt. Cependant, des études ont largement démontré que ce mode de vêlage n'est pas plus douloureux qu'un vêlage dystocique avec une traction forcée (Kolkman *et al.*, 2010) et qu'il n'interfère en aucun cas avec la relation mère-veau (Vandenheede *et al.*, 2001). Il est vrai que la relation mère-veau est plus liée à des questions de spéculations (veaux au pis ou veaux nourris artificiellement) et à une sélection négative sur la production laitière pour ces vaches viandeuses. Dans certains pays nordiques, comme la Suède, l'importation de la race BBB est interdite du fait que sa survie est liée à l'homme puisqu'elle nécessite son intervention pour se reproduire.

La sélection du caractère culard a fragilisé les grands systèmes devenant plus sensibles à certaines maladies.

Le **système cardiorespiratoire** est sûrement celui qui a payé le plus lourd tribut lors de la sélection viandeuse, ce qui rend le BBB particulièrement sensible aux maladies respiratoires (bronchopneumonies, laryngo-trachéites, syndrome de détresse respiratoire aiguë) (Gustin *et al.*, 1987 ; Bureau *et al.*, 1999 ; Danlois *et al.*, 2000 ; Bureau *et al.*, 2001 ; Bureau *et al.*, 2001 ; Danlois *et al.*, 2003) et cardiaques (cardiomyopathies) (Amory *et al.*, 1994) avec des répercussions rapides et importantes sur l'état général. Ce développement musculaire nécessite une alimentation et une complémentation en vitamines et en oligo-éléments adéquats. En conséquence, beaucoup de troupeaux BBB en Belgique sont carencés en oligo-éléments (sélénium, iode, cuivre, zinc) et présentent un nombre plus élevé d'individus malades que dans les troupeaux non carencés en oligo-éléments (Guyot *et al.*, 2009).

Par rapport à d'autres races, le BBB présente **une sensibilité accrue à la gale**, avec 74 % de troupeaux affectés en Flandre (Sarre *et al.*, 2012). Ceci serait dû à une réponse immunitaire retardée (Losson *et al.*, 1999 ; Vercruysse *et al.*, 2008) mais aussi à une plus grande finesse et tendreté du cuir moins riche en collagène. La couleur de la robe semble aussi jouer un rôle puisque, dans une même exploitation, les individus BBB sous robe noire semblent moins affectés que ceux de robe blanche (observations personnelles).

Les causes fréquentes de réforme et/ou d'exclusion au niveau du HBBBBB sont les **anomalies de la mâchoire inférieure**. On observe principalement deux types d'anomalies, le brachygnathisme inférieur et la déviation des incisives. Le brachygnathisme ou « gueule de cochon », dû à un développement moindre de la mandibule, est observé vers l'âge de 6 mois. Il affecte le pâturage et est plus fréquent chez les individus culards (*mh/mh*). Ce caractère semble suivre un mode d'hérédité complexe avec l'implication majeure du gène *mh* et une héritabilité de 20 % au sein des individus culards. Le caractère culard semble abaisser le seuil d'expression phénotypique et favoriser l'expression de gènes auxiliaires prédisposant au brachygnathisme (Hanset et Michaux, 1978). La déviation des incisives est, quant à elle, provoquée par la présence d'une hypertrophie de la langue (macroglossie) transitoire ou permanente que l'on rencontre régulièrement chez les veaux BBB. Cette macroglossie empêche également le veau de téter correctement. Comme le brachygnathisme, son hérédité est complexe, avec une héritabilité estimée à 10 % et avec un effet majeur du gène culard (Hanset et Michaux, 1978). Ces deux critères font l'objet d'une attention particulière lors de l'évaluation de la descendance des taureaux d'IA mais aussi des futurs reproducteurs. L'élimination de ces défauts passe donc par une sélection rigoureuse des reproducteurs.

Le **système locomoteur** présente un certain nombre de maladies ayant des répercussions économiques importantes. On peut citer parmi les plus fréquentes, les parésies

spastiques des membres postérieurs, la rigidité articulaire congénitale, les déviations angulaires des principales articulations et les fractures (Hanset, 2007). On définit la parésie spastique des gastrocnémiens (PSG) (Hanset *et al.*, 1993), des quadriceps (PSQ) (Touati *et al.*, 2003 ; De Vlamynck *et al.*, 2012) ou des autres groupes musculaires (Vertenten, 2009). Un mode d'hérédité complexe semble être à l'origine de la PSG. A défaut d'identifier la cause, des traitements symptomatiques, comme la névrectomie du nerf tibial (Vlaminck *et al.*, 2000), ou différentes ténectomies, ont été développés. Malheureusement, pour les parésies des autres groupes musculaires, aucun traitement n'est possible (Vertenten, 2009). La rigidité articulaire congénitale (RAC) est une maladie que l'on retrouve principalement dans les races viandeuses. En BBB, les mâles de poids élevé à la naissance et en présentation postérieure ont beaucoup plus de risques d'être atteints de RAC (Sartelet, 2007). L'origine semble être multifactorielle, avec une hérédité de 20 %, même si certains taureaux montrent une prédisposition à avoir de gros veaux atteints de RAC. Les taurillons à l'engrais à croissance rapide présentent régulièrement des déviations angulaires de type varus (déviations de l'extrémité du membre vers l'intérieur) ou valgus (déviations de l'extrémité du membre vers l'extérieur), principalement au niveau des boulets. L'origine est multifactorielle et dépend fortement de la ration (minéraux, énergie, glucides), du logement et de la génétique (aplombs, développement musculaire, croissance). Enfin, la sélection pour le développement musculaire a augmenté le rapport muscle/os, avec une augmentation de la musculature, notamment au niveau des fessiers, et une diminution de la masse osseuse, rendant le BBB plus sensible aux fractures au niveau du tibia, surtout à la mise à l'herbe (Gangl *et al.*, 2005 ; Sartelet, 2007).

Le principal défaut affectant le **tractus génital femelle** dans la race BBB est la maladie des génisses blanches ou en termes anglo-saxons, la *White Heifer Disease* (WHD). C'est un ensemble d'anomalies congénitales du tractus génital femelle que l'on retrouve quasi exclusivement chez les individus sous robe blanche (Blanc Shorthorn). Cette anomalie est un héritage de l'introduction de la race Shortorn (Rendel, 1952) dans les races locales à la fin du XIX^{ème} siècle (Charlier *et al.*, 1996). Hamoir, en 1897, observait quatre cas d'imperforation vaginale chez des génisses. Les anomalies qu'il constate, il les décrira très justement comme dues à un arrêt de développement des structures Müllériennes.

En 1961, Hanset donne une description détaillée de cette maladie en race BBB. Il a observé 271 tractus génitaux de génisses de robe blanche âgées entre 3 et 5

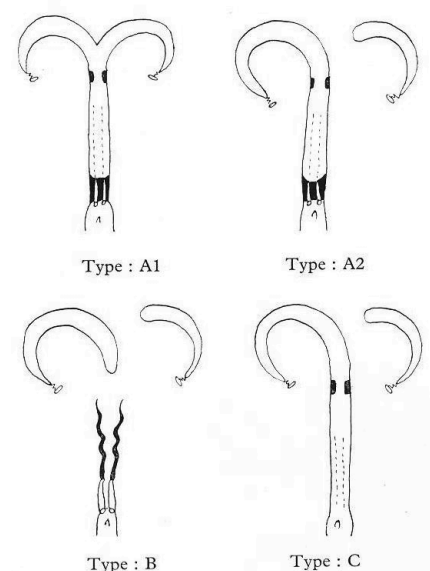


Figure 11: Représentation des trois formes de WHD (Hanset, 1969).

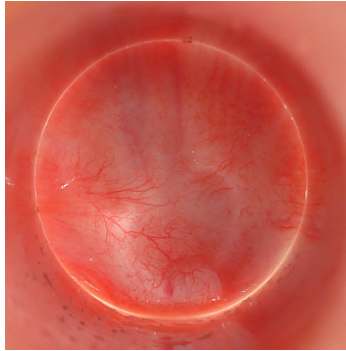


Figure 12: Vaginoscopie d'un vagin aveugle d'une génisse atteinte de WHD de type A.

mois. Sur les 271 observations, il a recensé 39 (15 %) anomalies typiques de la maladie des génisses blanches. Ces anomalies ont été regroupées en trois catégories (Figure 11): i) le type A se caractérise par un fond vulvaire aveugle (Figure 12) contenant deux orifices de chaque côté du plan médian, se prolongeant par un conduit dans une zone intermédiaire fibreuse, à cette zone fait suite un vagin qui se termine en cul-de-sac ; deux sous-types sont décrits, le type A1 qui ne présente pas d'anomalies antérieures (cornes et corps) et le type A2 qui présente des anomalies antérieures ; ii) le type B se caractérise lui aussi par un vestibule aveugle avec cette zone intermédiaire à laquelle fait suite une portion constituée par deux cordons durs, sinueux, fibreux avec une lumière intermittente (« Cordons de Wolff »), la portion antérieure étant représentée par des cornes interrompues ; enfin, iii) le type C se caractérise par l'absence d'anomalies postérieures (vestibule et vagin postérieur) et la présence d'anomalies au niveau de l'utérus et/ou surtout des cornes (Hanset, 1961). Seule cette dernière forme est compatible avec une gestation. Le type A, plus fréquemment observé, représente 60 % des anomalies rencontrées. Quelqu'en soit le type, les anomalies sont limitées aux cornes, au corps, au col et au vagin. Les structures ovariennes sont conservées et fonctionnelles, ce qui se traduit par une expression normale des chaleurs chez ces individus. Les oviductes sont également toujours présents. Un « hymen persistant » n'a jamais été observé, mais de manière anecdotique, un fin voile est parfois décrit. Il faut encore noter que lorsqu'une seule corne est présente, il s'agit le plus souvent de la corne gauche. Ceci est attribué à une migration plus rapide du canal de Müller gauche par rapport à celui de droite lors de l'organogenèse. Le déterminisme complexe de la WHD a été suspecté dans les années 60' avec une implication majeure du locus *Rouan* (Hanset, 1969, Charlier *et al.*, 1996). Dans les années 60', Hanset (1969) fait état d'une incidence de la WHD de respectivement 10 à 15%, 1 à 1,5% et 0,3% chez les génisses sous robe blanche, bleue ou noire. En outre, il prouvera également l'implication d'autres gènes (les gènes auxiliaires) en montrant une variabilité de l'incidence entre taureaux, certains se caractérisant par un pourcentage faible, de l'ordre de 2-3%, de filles atteintes parmi leur descendance blanche, d'autres par un pourcentage élevé, atteignant 35%. Il estimera l'héritabilité de ce caractère au sein de la robe blanche à 20 % (Hanset, 1969).

L'hypoplasie utérine se rencontre fréquemment dans la race BBB. Elle se caractérise par un retard de développement non-irréversible du tractus génital, associé à l'absence d'activité ovarienne chez des génisses d'âge pubère (Boudry, 2002).

F. Identification des gènes majeurs dans la race BBB

La race BBB est sûrement la race à viande la mieux chiffrée en terme de performances, avec le calcul des index, mais aussi caractérisée, avec l'identification de gènes majeurs intervenant dans sa sélection. Les deux phénotypes principaux de la race BBB, la **musculature** et la **robe rouane**, ont fait l'objet de nombreuses recherches afin d'élucider leur mécanisme moléculaire. En 1990, une première étude visant à localiser les deux loci responsables a été réalisée (Georges *et al.*, 1990). Cinq ans plus tard avec le développement des microsatellites et la création d'un croisement expérimental, le locus *mh* et le locus *Rouan* ont été cartographiés. Ce croisement expérimental a été généré par l'accouplement entre des taureaux BBB (*mh/mh*) et des femelles de race Frisonne (+/+). La génération F1 obtenue (*mh/+*) a été croisée avec des individus BBB (*mh/mh*) (Charlier *et al.*, 1995 ; Charlier *et al.*, 1996). Une analyse de liaison réalisée à partir de ce pedigree expérimental a localisé les loci associés à l'hypertrophie musculaire et à la robe rouanne, respectivement sur le chromosome 2 et le chromosome 5 (Charlier *et al.*, 1995 ; Charlier *et al.*, 1996).

L'identification de l'implication du gène *GDF-8* (Growth/Differentiation Factor 8), codant pour la myostatine, dans la régulation négative de la croissance musculaire chez la souris faisait du gène ***MSTN***, compris dans l'intervalle cartographié, le candidat idéal (McPherron *et al.*, 1997). Le séquençage du gène *MSTN* bovin a identifié une délétion de 11 bp responsable de l'hypertrophie musculaire en BBB (Grobet *et al.*, 1997). Le séquençage de ce même gène a permis par la suite d'identifier une série allélique de mutations sélectionnées, responsables de l'hypertrophie musculaire, dans d'autres races (Grobet *et al.*, 1998). Cette mutation explique une augmentation de 20 % de la masse musculaire chez les individus culards (*mh/mh*), cependant on peut voir que depuis l'apparition des premiers *mh/mh* dans les années 60' et sa fixation dans les années 1980, la musculature n'a cessé de croître (Figures 7 & 11), permettant de suspecter l'existence d'autres gènes non identifiés augmentant la musculature (Georges, 2010).

Au niveau du locus ***Rouan***, Charlier et collaborateurs (1996) avaient localisé le locus sur le chromosome 5. Ils avaient également proposé comme candidat le locus *Steel*, codant pour le MGF, connu pour être le ligand du récepteur à la tyrosine kinase c-kit (KITLG) et impliqué dans la mélanogenèse. Le séquençage de ce gène a identifié une mutation non-synonyme (p.Ala193Asp) associée à la robe rouanne responsable de la robe bleue et blanche en BBB (Seitz *et al.*, 1999) (*cf. b) Robe*, p. 29).

Dans les années 60', une centaine de cas de **dermatosparaxie** ont été recensés à la Faculté de Médecine Vétérinaire. Elle se caractérise par une extrême fragilité de la peau et une mortalité dans le premier mois de la vie. L'analyse épidémiologique a permis de suspecter un

mode d'hérédité autosomale récessive sur base d'une analyse épidémiologique. La fréquence de la mutation a été estimée entre 0,6 et 1,2 % (Hanset et Lapierre, 1974). La dermatosparaxie est similaire au syndrome d'Ehlers-Danlos de type VII C chez l'homme, ils sont tous deux liés à une déficience de l'activité de la procollagène I N-protéinase (pNPI) intervenant dans la formation du collagène (Nusgens *et al.*, 1992). Le séquençage du gène codant pour la pNPI, étant le gène candidat de choix, a mis en évidence des mutations non-sens chez des humains atteints du syndrome d'Ehlers-Danlos de type VII C et une délétion de 17 bp chez les veaux souffrant de dermatosparaxie (Colige *et al.*, 1999). Cette maladie a spontanément été éradiquée en BBB, sans la mise en place de test diagnostique au niveau de la population.

Dans les années 90', deux entités relativement similaires au niveau de leurs phénotypes ont émergé. Il s'agit de la **dystonie musculaire congénitale de type 1 et de type 2 (DMC 1 et 2)**. La première (également nommée « veau SMA » pour Spinal Muscular Atrophy) se caractérise chez les veaux dès la naissance, par une raideur généralisée suite à un stimulus externe et une chute caractéristique sur le flanc, sans fléchir les membres d'où l'appellation « veaux planches » (Charlier *et al.*, 2008). En l'absence de stimulus, le veau présente une démarche raide, une fatigabilité accrue après un déplacement et des troubles de la déglutition. Ce défaut est léthal dans les quelques mois après la naissance, principalement suite au développement d'une bronchopneumonie par erreur de lieu. Le second type de dystonie, communément appelée par les éleveurs « veaux électriques », se caractérise par l'apparition des crises tétaniformes (similaires à une intoxication à la strychnine) suite à un stimulus externe. Ce défaut est léthal dans les jours qui suivent la naissance. L'analyse épidémiologique a permis de suspecter un mode d'hérédité autosomal récessif, avec *Christophe de l'Abbaye* pour la DMC 1, et *Rieuse* pour la DMC 2, comme ancêtres fondateurs respectifs suspectés. Ces défauts létaux étaient la source de pertes économiques importantes. La disponibilité de la première séquence de référence bovine et le développement des techniques de génotypage à moyen et à haut débit ont considérablement accéléré les études de cartographie et d'identification de mutations causant des défauts génétiques à ségrégation mendélienne. La première application de ces outils chez les bovins a facilité la cartographie des gènes responsables, sur les chromosomes 25 et 29, respectivement pour la DMC 1 et la DMC 2. Le choix et le séquençage de gènes candidats, localisés dans ces deux intervalles, sur base de leur implication dans des phénotypes similaires chez l'homme (la maladie de Brody et l'hyperkplexie), ont permis d'identifier la mutation causale pour ces deux défauts (Charlier *et al.*, 2008). Il s'agit de deux mutations non-synonymes dans 1) le gène *ATP2A1* (p.Arg559Cys) codant pour une pompe à calcium sarcoplasmique intervenant dans la relaxation musculaire et 2) dans le gène *SLC6A5* (p.Leu270Pro) codant pour un transporteur à glycine intervenant dans le maintien du pool présynaptique de glycine, respectivement pour la DMC 1 et la DMC 2 (Charlier *et al.*, 2008). Le développement d'un test

diagnostique a permis d'estimer la fréquence des porteurs (DMC 1 $\approx$ 15 % et DMC 2 $\approx$ 5 %) et de mettre en place une sélection assistée par marqueurs (SAM) afin d'éviter les accouplements à risque. La mutation responsable de la DMC 2 a également été identifiée dernièrement dans la race British blue, issue de l'export de la race BBB en Angleterre, à partir de deux cas et a été intégrée dans leur programme de sélection (Gill *et al.*, 2012).

Nous avons donc vu au cours de cette partie que la race BBB fait partie intégrante du patrimoine de la Belgique tant au niveau de son histoire, de sa population que de sa production de viande. Le niveau atteint en terme de développement musculaire est le fruit d'une sélection intensive utilisant des outils puissants mis à la disposition des éleveurs. « *La sélection a engendré de bons culards...* » mais elle a emporté avec elle son lot de défauts. Dans la partie suivante, les méthodes permettant de caractériser les défauts d'origine génétique seront détaillées.

3- ETUDE DU DETERMINISME GENETIQUE

A. Introduction

La génétique au sens large est la science qui étudie les gènes, leur transmission et leur variation chez les individus vivants. Elle étudie d'une part la ségrégation (mode de transmission) de phénotypes (caractères) et d'autre part la ségrégation de marqueurs moléculaires (allèles). La génétique se divise en plusieurs branches avec notamment la génomique qui étudie le génome, la génétique quantitative qui étudie la composante génétique de caractères dits quantitatifs, la génétique des populations, la génétique de l'évolution et la génétique du développement.

Tout d'abord, le génome bovin est composé de 60 chromosomes répartis en 29 paires d'autosomes et une paire de chromosomes sexuels. Les premières cartes physiques de marqueurs du génome bovin, établies dans les années 90' avec l'utilisation des microsatellites, constituaient une étape importante dans l'étude de l'hérédité de caractères



Figure 13: L1 Dominette 01449

mendéliens et complexes (Barendse *et al.*, 1994 ; Bishop *et al.*, 1994). Deux découvertes majeures vont permettre d'améliorer la connaissance et l'utilisation de ce génome dans la compréhension de ces caractères. La première est l'établissement de la première séquence complète du génome bovin, obtenue en 2009 à partir de l'acide désoxyribonucléique (ADN) de leucocytes d'une vache de race Hereford nommée L1 Dominette 01449 (Figure 13) (Bovine Genome Sequencing and Analysis Consortium, 2009). Le génome bovin, long de 3 milliards (3 Gb) de paires de bases, contient au moins 22.000 gènes (Bovine Genome Sequencing and Analysis Consortium, 2009). La seconde est le séquençage additionnel de six races de bovins identifiant des milliers de variants supplémentaires de type SNP (Single Nucleotide Polymorphism) (The Bovine HapMap Consortium, 2009). De ces découvertes émaneront le développement de puces de génotypage permettant d'interroger simultanément 50.000, puis aujourd'hui 770.000 SNPs, et d'étudier leur ségrégation (Matukumalli *et al.*, 2009).

Parmi les **caractères** étudiés, on en définit deux types, les simples et les complexes. Les **caractères simples**, ou autrement appelés mendéliens, suite aux découvertes de Gregor Mendel à la fin du XIX^{ème} siècle, sont sous l'influence d'un seul gène. Ces caractères se transmettent d'une génération à l'autre selon des règles bien déterminées énoncées par les lois de Mendel. L'expression de ces caractères (phénotypes) est déterminée par un seul et unique gène qui peut se décliner sous plusieurs formes (allèles) dues à des modifications (mutations). Ce gène sera présent chez un individu diploïde en deux allèles, correspondant à son génotype, provenant l'un du père et l'autre de la mère. Le mode de transmission dépend notamment de la localisation du gène sur les autosomes ou les hétérosomes, chromosomes sexuels, avec des caractères qui ont une hérédité **autosomale** ou **liée à l'X** ou à l'**Y**. On distingue également le mode de transmission **dominant** où l'individu hétérozygote et l'homozygote muté expriment le même phénotype. A l'inverse, dans le cas de caractères à transmission **récessive**, l'individu homozygote muté (deux allèles mutés) exprime le phénotype alors que l'homozygote sauvage (non muté) et l'hétérozygote sont indifférentiables. Entre les deux, un mode de transmission **codominant** ou **additif**, l'hétérozygote présente un phénotype intermédiaire aux deux phénotypes homozygotes (Figure 14). La plupart des cas de maladies mendéliennes ou autosomales récessives sont principalement dus à des variants rares voire privés (limités à une famille au sens large) (Figure 15). Dans certains cas, pour un génotype donné, le phénotype attendu n'est pas observé ou est variable. On détermine la **pénétrance** d'une mutation, que l'on dit incomplète quand pour un génotype donné, homozygote muté, moins de 100 % des individus présentent le phénotype attendu. D'autre part, lorsque le phénotype observé (ex. gravité des lésions, variation dans les paramètres mesurés...) n'est pas identique pour un même génotype, on parle d'**expressivité** variable.

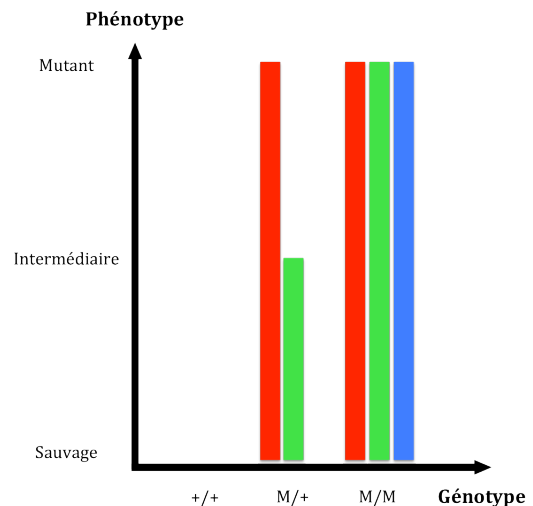
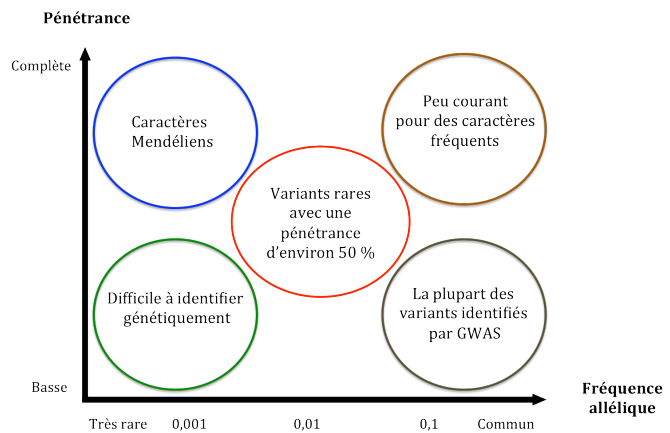


Figure 14: Observation du phénotype en fonction du génotype, bleu = récessif, vert = codominant ou additif et rouge = dominant avec M = allèle muté et + = allèle sauvage.

L'étude des caractères mendéliens a pour but d'identifier le gène et le mécanisme moléculaire responsable de ce phénotype afin d'identifier les individus porteurs de ces mutations et de prévenir les accouplements à risque en cas de maladie récessive ou encore de sélectionner les individus porteurs de mutations bénéfiques. La connaissance du mécanisme

moléculaire des maladies orphelines permet aussi d'envisager la mise en place d'un traitement, y compris la thérapie génique.

Figure 15 : Probabilité d'identifier les variants en fonction de leur fréquence dans la population et de leur effet. A titre d'exemple, les maladies mendéliennes sont causées par des variants très rares mais expliquant une proportion majeure de la variance phénotypique. A *contrario*, les variants identifiés dans les études de maladies plus communes ont une fréquence élevée mais ils expliquent une faible proportion de la variance phénotypique, GWAS = Genome Wide Association Studies (adapté de McCarthy *et al.*, 2008).



Les généticiens ont également démontré que certains caractères, bien que se transmettant d'une génération à l'autre, ne ségrégaient pas suivant les lois de Mendel. Ils se sont rendus compte que ces phénotypes avaient un mode d'hérédité plus complexe. Ces **caractères dits complexes**, que l'on appelle aussi quantitatifs (*quantitative traits*), sont mesurables et sont le résultat de l'effet et de l'interaction de plusieurs gènes modifiés par des facteurs environnementaux. Ils ont également observé que la plupart des caractères importants, étudiés en médecine humaine et en agriculture, étaient des caractères complexes : sensibilité à certaines maladies chez l'homme, production laitière, croissance, fertilité, résistance aux maladies chez les animaux de production (Goddard and Hayes, 2009). Contrairement aux caractères mendéliens, les caractères complexes sont fréquents et associés à des allèles communs dans la population mais ayant des effets très faibles sur le phénotype, ce qui rend leur identification difficile (Figure 15 ; Manolio *et al.*, 2009). L'objectif de l'étude des caractères complexes est d'identifier des loci (QTL = Quantitative Trait Loci) ou mieux des variants nucléotidiques (QTN = Quantitative Trait Nucleotides) influençant des maladies humaines dites complexes (cancer, maladies auto-immunes, cardiovasculaires, psychiatriques, ...) ou la production des animaux de rente. L'objectif est *in fine* de pouvoir prédire le risque d'apparition de ces maladies ou la valeur d'élevage de la descendance à l'aide d'une sélection assistée par marqueurs ou de la sélection génomique. A ce jour, un grand nombre de QTL répartis sur l'entièreté du génome bovin ont été localisés et répertoriés dans une base de données des QTL publiés (www.animalgenome.org/cgi-bin/QTLdb/index). Peu ont été confirmés, répliqués et disséqués au niveau moléculaire, ce qui rend leur utilisation dans un programme de sélection assistée par marqueurs à large échelle risquée et peu rentable (Georges, 2012).

La **sélection assistée par marqueurs** (SAM) a connu ses premières applications dans les années 90' avec l'utilisation de variants du complexe majeur d'histocompatibilité pour identifier les allèles améliorant la résistance aux maladies et la productivité chez des vaches laitières (Weigel *et al.*, 1990). La SAM a ensuite utilisé les résultats obtenus par cartographie, que ce soit des mutations affectant la fonction d'un gène ou que ce soit des QTL. Dans le premier cas, la SAM facilite l'élimination des mutations et loci ayant un effet majeur négatif comme dans le cas de défauts monogéniques ou l'augmentation de la fréquence voire la fixation de l'allèle muté présentant un avantage, comme dans le cas l'hypertrophie musculaire chez les bovins (Dekkers, 2004). Dans le second cas, la SAM utilise des marqueurs en déséquilibre de liaison avec le QTL identifié. Cela permet d'estimer une valeur d'élevage d'un individu à partir de son phénotype, de son pedigree et de son génotype pour le(s) marqueur(s) concerné(s). Ce type de SAM a notamment contribué à l'amélioration des caractères de santé ou de production (production laitière, croissance...) chez les bovins (Dekkers, 2004 ; Goddard et Hayes, 2009) et à réduire l'intervalle entre les générations. Cependant, la prédiction des valeurs d'élevage d'un individu reste limitée puisqu'elle n'utilise qu'un nombre restreint de marqueurs qui n'expliquent qu'une faible proportion de la variance génétique du caractère que l'on veut améliorer (Goddard et Hayes, 2009 ; Hayes *et al.*, 2009).

La **sélection génomique** est une approche qui ne tient pas compte des précédentes détections de QTL et/ou de QTN mais qui utilise l'information générée par une plus grande densité de marqueurs répartis sur l'entièreté du génome avec les puces à SNPs (Meuwissen *et al.*, 2001). L'augmentation de la densité des marqueurs permet d'exploiter l'information de déséquilibre de liaison entre les SNPs et les QTL/QTNs. Le déséquilibre de liaison estime l'effet de ces marqueurs non plus dans une famille mais dans la population d'intérêt (Goddard et Hayes, 2009). La sélection génomique consiste dans un premier temps à diviser en plusieurs segments ou haplotypes l'entièreté du génome et d'en estimer l'effet dans une large population de référence de plusieurs milliers d'individus dont on connaît le génotype et le phénotype. Une équation tenant compte simultanément de tous les marqueurs génotypés et de leur effet sur le phénotype pourra être générée. Ensuite, par exemple, le génotypage des veaux à la naissance suffira à déterminer la présence ou non de ces haplotypes et à estimer leur valeur d'élevage génomique (GEBV = *GE*nomi*c* *B*reedi*ng* *V*alue) à l'aide de l'équation obtenue avec la population de référence (Goddard et Hayes, 2007).

B. Etude du déterminisme génétique par cartographie ou clonage positionnel

Le clonage positionnel a pour objectif d'identifier une région du génome et de localiser le(s) gène(s) candidat(s) responsable(s) de ce phénotype sans hypothèse quant à la fonction de ce(s) gène(s). Le clonage positionnel se décompose en trois étapes : i) la primolocalisation du locus ou des loci influençant le phénotype étudié par analyse de liaison, d'association ou la combinaison des deux, ii) la cartographie fine des loci incriminés et iii) l'analyse de gènes candidats dits positionnels compris dans cet intervalle et la validation de la, voire des mutations identifiées. Dans les années 80-90', l'étape de primolocalisation se décomposait en deux étapes avec la cartographie à l'aide d'analyses de liaison et la réduction de l'intervalle en augmentant la quantité de marqueurs et/ou de méioses informatives dans cet intervalle (Georges, 2012). Les premières méthodes étudiaient la transmission de certains marqueurs au sein d'une famille par **étude dite de liaison**. La pauvreté en marqueurs génétiques a longtemps été un frein au clonage positionnel de gènes et de mutations responsables de variations phénotypiques. La fin du XX^{ème} siècle connaîtra l'avènement des techniques de manipulation de l'ADN. Ces techniques faciliteront l'identification et l'utilisation de variations de l'ADN, à savoir les marqueurs génétiques. L'augmentation de la densité de marqueurs transformera le clonage positionnel en combinant la liaison génétique et le déséquilibre de liaison au sein d'une population dans des **études dites d'association génome entier** (GWAS = *Genome Wide Association Study*) (Goddard et Hayes, 2009).

a) Marqueurs génétiques

L'utilisation des endonucléases de restriction et les **RFLPs** (*Restriction Fragment Length Polymorphisms*), constituera la première évolution de l'identification de variants de la séquence dans les années 80'. Cette méthode est basée sur la présence de motifs nucléotidiques situés dans la séquence d'ADN, appelés sites de restriction et reconnus par des enzymes de restriction qui vont cliver l'ADN en fragments de taille variable en fonction de la présence ou non de mutations au niveau de ces sites. La séparation de ces fragments par électrophorèse permet de déterminer la taille obtenue. La modification de ce site de restriction par substitution ou délétion d'une base empêchera le clivage par l'endonucléase et modifiera la taille du fragment obtenu. La première carte génétique humaine a été établie via l'utilisation de ces RFLPs (Botstein *et al.*, 1980). Au cours de la même décennie, la découverte des **minisatellites** constituera une nouvelle source de polymorphismes. Les minisatellites ou VTNRs (*Variable Tandem Number Repeats*) sont de courtes séquences de 10 à 100 bp répétées. La variation

dépend du nombre de fois où cette séquence est répétée, modifiant ainsi la taille du fragment (Jeffreys *et al.*, 1985). Les minisatellites, situés à 90 % dans les régions sub-télomériques, sont peu informatifs pour la cartographie et sont principalement utilisés en médecine légale (Gill *et al.*, 1985). En 1983, la biologie moléculaire connaîtra une avancée majeure avec le développement de la PCR (*Polymerase Chain Reaction* = Amplification en Chaîne à la Polymérase = ACP) permettant d'amplifier rapidement des fragments d'ADN de plusieurs milliers de bp sur des centaines d'individus et ce de façon simultanée (Saiki *et al.*, 1985). Un nouveau type de polymorphisme, les **microsatellites**, profitera pleinement de cette découverte. Les microsatellites ou STRs (*Short Tandem Repeats*) sont, tout comme les minisatellites, de courtes séquences nucléotidiques de 2 à 10 bp répétées des dizaines de fois. Contrairement aux minisatellites, les microsatellites sont très fréquents et répartis aléatoirement tout le long du génome (Weber et May, 1989). Ces marqueurs ont été largement utilisés dans la cartographie de caractères mendéliens et complexes par étude de liaison tant chez l'homme, que chez les animaux domestiques. Enfin, l'obtention des premières séquences du génome pour de nombreuses espèces, y compris bovine, au début des années 2000, a vu l'émergence d'un nouveau type de marqueur, le **SNP** (*Single Nucleotide Polymorphism*). Les SNPs sont, comme leur nom l'indique, des polymorphismes dus à des mutations ponctuelles (1 bp) et sont principalement bialléliques (seulement deux allèles existent dans la population). Les SNPs sont répartis aléatoirement et de manière uniforme tout au long du génome. Parallèlement à cette découverte, des plateformes de génotypage à haut-débit se sont développées. Celles-ci permettent d'interroger simultanément 50.000 SNPs et plus récemment 770.000 SNPs de dizaines de bovins à moindre coût (Charlier *et al.*, 2008 ; Ding et Jin, 2009 ; Matukumali *et al.*, 2009). La disponibilité d'une grande densité de marqueurs permettra d'étudier leur transmission, non plus uniquement au niveau d'une famille mais aussi au niveau de la population avec l'exploitation du déséquilibre de liaison. Les SNPs ont ainsi contribué à de nombreux succès d'études d'association génome-entier pour des maladies complexes chez l'homme mais aussi chez les animaux domestiques. Chez les bovins, le génotypage à haut débit a considérablement accéléré l'étude des caractères mendéliens, il a aussi amélioré le système de sélection notamment dans les races laitières avec la mise en place de la sélection génomique (Meuwissen *et al.*, 2001 ; Charlier *et al.*, 2008).

Plateformes de génotypage de SNPs à haut débit

Des plateformes de génotypage à haut débit ont ainsi été mises au point afin d'automatiser le génotypage des milliers de SNPs. Pour ce faire, des biopuces à ADN ont été développées afin de génotyper simultanément des dizaines d'individus pour des SNPs préalablement sélectionnés parmi les millions de SNPs connus. A l'heure actuelle, trois puces bovines contenant 3.000, 50.000 et 770.000 SNPs sont disponibles sur le marché (Figure 16) (Charlier *et al.*, 2008 ; Matukumali *et al.*, 2009). Plus récemment, une puce de faible densité de 6.000 marqueurs a été générée et optimisée afin de diminuer le coût du génotypage, tout en permettant une bonne précision d'imputation des SNPs manquants pour obtenir des génotypes à haute densité nécessaires lors de prédictions de valeurs d'élevage en sélection génomique (Boichard *et al.*, 2012). **L'imputation** consiste à prédire les génotypes des marqueurs non génotypés (par erreurs ou par choix d'un nombre limité de marqueurs) en utilisant l'information haplotypique des génotypes connus au niveau d'une population ou d'une famille (Li *et al.*, 2009).

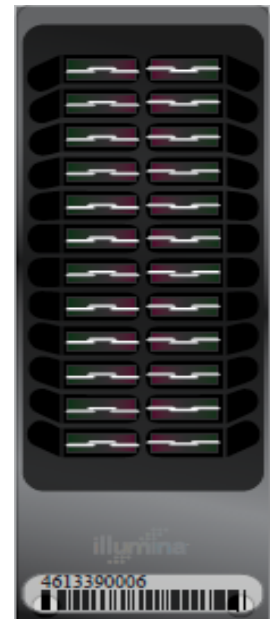


Figure 16: Bovine SNP50 v2 beadchip, puces de génotypage pour 24 échantillons contenant 54.000 SNPs.

La technologie, décrite ci-dessous, qui a été utilisée au cours de ma thèse, est le type « iSelect » développé par Illumina. Les puces de génotypage à ADN sont des supports de verre ou de silicium contenant des pores dans lesquels viennent se loger des billes de silicium portant des séquences oligonucléotidiques. Chaque bille possède à sa surface des milliers d'oligonucléotides identiques spécifiques au SNP à identifier. Chaque oligonucléotide a une séquence totale égale à 73 pb dont l'extrémité 5' du segment est liée de façon covalente à la bille. En partant de cette extrémité, les 23 premières bases sont utilisées pour la localisation des différentes billes. Cette séquence est spécifique et caractéristique pour chaque SNP différent : il s'agit de la séquence adresse. Les 50 bases restantes constituent les séquences complémentaires au locus d'intérêt. Les billes sont placées de manière aléatoire dans les pores et leur localisation est déterminée grâce à la séquence adresse (Gunderson *et al.*, 2005).

Une fois l'ADN extrait à partir de différents tissus (sang, sperme, rate, langue...), une quantité d'ADN (environ 150 ng) est amplifiée. Ici, l'amplification n'est pas réalisée à l'aide de la technique de PCR classique, mais avec une amplification linéaire appelée *Whole-genome amplification*. Cette technique permet d'obtenir les larges quantités d'ADN génomique amplifié (près d'un milligramme) requises pour les études de génotypage à grande échelle, à partir de petites quantités d'ADN génomique de départ (quelques centaines de nano-grammes). L'ADN de départ est donc amplifié de 1.000 à 2.000 fois. Après l'amplification, l'ADN obtenu est fragmenté à l'aide d'enzymes de restriction afin d'obtenir des fragments de 200 à 300 pb dans le but de faciliter l'hybridation aux séquences complémentaires présentes sur les billes. L'ADN fragmenté est ensuite réparti sur la puce et s'hybridera spécifiquement par complémentarité à la séquence de la sonde (Figure 17A). Une fois la capture effectuée, le marquage est réalisé à l'aide d'une polymérase, qui va ajouter le nucléotide complémentaire au nucléotide de l'ADN à génotyper. Cette étape d'extension est possible grâce à l'utilisation de nucléotides marqués à l'aide d'haptènes (Figure 17B). Un haptène est une molécule de faible poids moléculaire, antigénique mais non immunogène. On distingue deux types d'extensions, une spécifique aux allèles des SNP et une non spécifique (Steemers *et al.*, 2006). Pour l'extension spécifique, les allèles pour un SNP déterminé sont présents à l'extrémité 3' de la sonde oligonucléotidique et l'extension s'effectue sur le nucléotide qui suit le SNP (Figure 18A). Cette technologie nécessite l'utilisation de

deux billes par SNP mais dans ce cas une seule couleur de fluorophore est suffisante. Lors d'extension non-spécifique, le SNP n'est pas présent à l'extrémité 3' de la sonde et l'extension s'effectue en ajoutant le nucléotide complémentaire au SNP de l'ADN analysé (Figure 18B). Pour

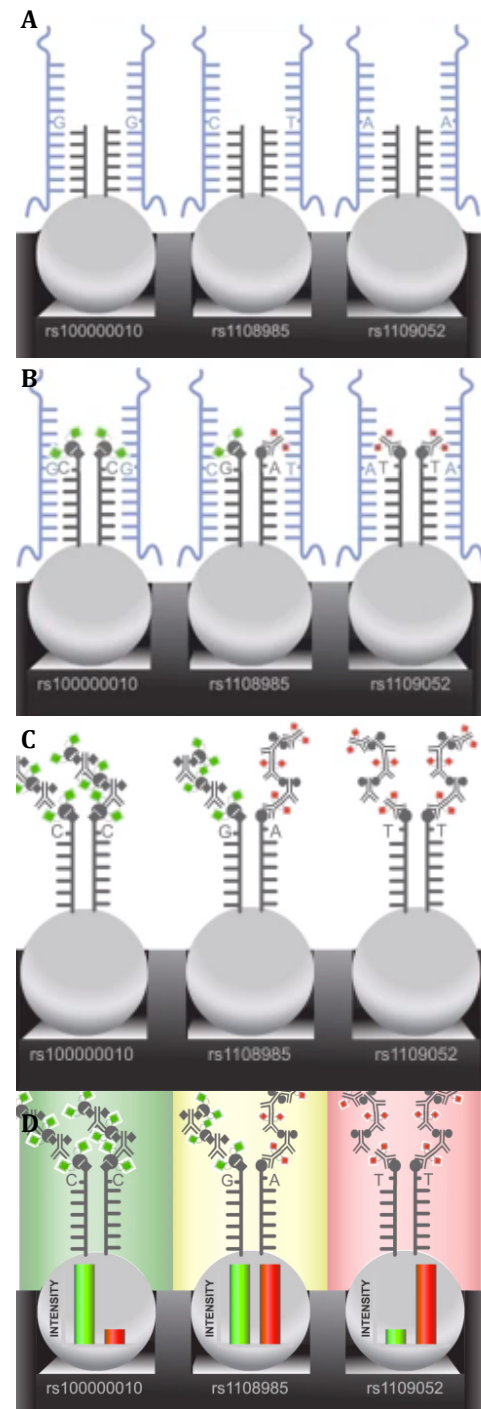


Figure 17: Etapes (A) d'hybridation, (B) d'extension, (C) d'amplification et (D) de lecture réalisées lors du génotypage à l'aide de puces à ADN (Source : www.Illumina.com).

cette technique, une bille par SNP est suffisante ce qui permet de gagner de la place sur la puce, toutefois, l'emploi de deux fluorophores est requis.

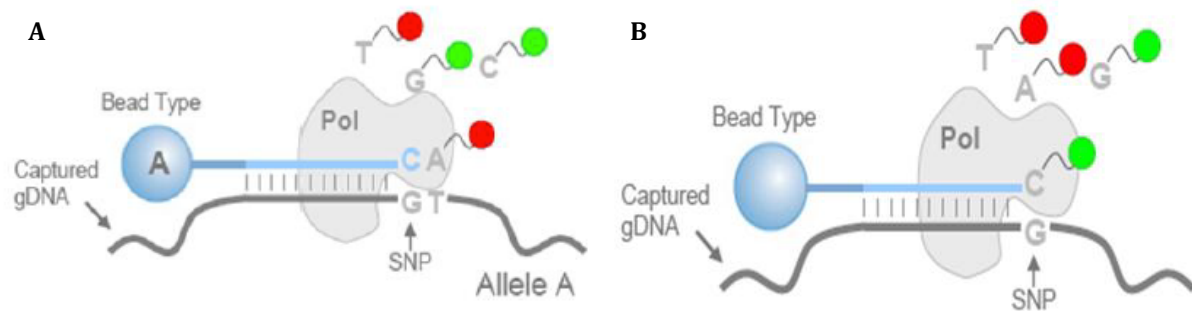


Figure 18 : Extension de type (A) spécifique avec une bille portant une sonde qui contient un des allèles du SNP, dans ce cas le C, et l'extension se fait sur la base suivante = T et de type (B) non spécifique où la bille porte une sonde qui ne possède pas le nucléotide complémentaire du SNP d'intérêt et l'extension se fait avec une nucléotide marqué complémentaire de l'allèle du SNP.

Une fois l'étape d'extension terminée, le signal résultant est amplifié à l'aide d'anticorps porteurs de fluorophore (Figure 17C) et la puce est lue par un scanner à haute résolution (Figure 17D & Figure 19) (Gunderson *et al.*, 2005).

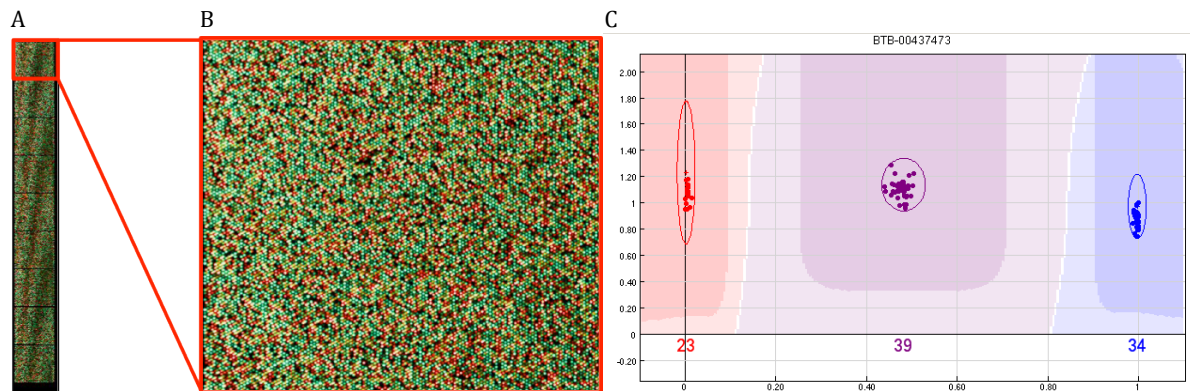


Figure 19 : Image obtenue lors de la lecture d'une puce par le scanner (A) d'une bande (un individu) et (B) le zoom au niveau de cette bande mettant en évidence les 3 couleurs obtenues vert, rouge et jaune. (C) Graphique obtenu après analyse des images au niveau d'un SNP avec en rouge et en bleu les génotypes homozygotes et en violet le groupe des hétérozygotes.

Lors du génotypage, des analyses préliminaires aux analyses statistiques sont nécessaires afin d'éviter les erreurs de génotypage et les associations faussement positives. Dans le cas d'une étude de type cas-contrôles, le **choix de cas les moins hétérogènes possibles** est primordial pour tenter d'identifier le ou les loci responsables. Les manipulations d'un grand nombre d'individus et le génotypage d'une grande quantité de marqueurs simultanément sont des sources d'erreurs. Une fois les individus sélectionnés, la **quantité et la qualité de l'ADN**

extrait nécessaire au génotypage, doivent être déterminées afin d'éliminer les ADN de mauvaise qualité pouvant induire des erreurs de génotypage et par la suite, de fausses associations (McCarthy *et al.*, 2008). Après le génotypage, il est également possible de vérifier i) la concordance au niveau du sexe des individus et ii) la relation de parenté, quand des apparentés sont disponibles et génotypés. La **qualité du génotypage** peut être évaluée par l'estimation du taux de génotypage par individu (*call-rate*) représentant le pourcentage de génotypes assignés sur le nombre total de marqueurs d'un individu. Un *call-rate* minimum de 95 % est nécessaire pour conserver l'individu. Ensuite le taux de génotypage par SNP (*call-freq*) correspond, pour un marqueur donné, au nombre d'individus auxquels un génotype est assigné. Il évalue la qualité de génotypage d'un marqueur. Un *call-freq* de 90 % est nécessaire pour conserver ce marqueur. De plus, les marqueurs ayant une fréquence de l'allèle mineur inférieure à 5 % et ne respectant pas l'équilibre de Hardy-Weinberg dans la population contrôle sont éliminés (Balding, 2006). Une population en **équilibre de Hardy et Weinberg** est une population idéale d'individus diploïdes, de grande taille, panmictiques (accouplements aléatoires entre les individus), en absence de migration, de mutation, de sélection et dont les générations ne se chevauchent pas. Il découle de cet équilibre que la fréquence des allèles, à un locus donné, est définie par la loi de distribution génotypique où $p^2 + q^2 + 2pq = 1$ avec p^2 = fréquence du génotype homozygote pour un allèle, q^2 = la fréquence du génotype homozygote pour l'autre allèle et $2pq$ = la fréquence du génotype hétérozygote. Le déséquilibre de liaison suppose à la fois, la liaison génétique et le déséquilibre gamétique. Une telle association allélique est favorisée par la proximité physique des loci et le caractère récent de la mutation ayant produit l'un des deux allèles.

b) Liaison génétique et déséquilibre de liaison

Les méthodes de cartographie, qu'elles soient de type analyse de liaison ou qu'elles soient de type étude d'association, tiennent compte de deux principes : la liaison génétique et le déséquilibre de liaison.

On parle de **liaison génétique** (*linkage*) entre deux allèles de deux loci différents, définissant un haplotype, lorsque ceux-ci ont tendance à être transmis ensemble d'un individu à sa descendance. Plus ces loci sont proches l'un de l'autre, plus la probabilité d'être transmis ensemble est grande. *A contrario*, plus la distance entre ces deux loci est grande, plus la probabilité de recombinaison (*crossing over*), définie par l'échange de fragments entre deux chromosomes homologues au cours la prophase I de la première division méiotique lors de la formation des gamètes est grande (Figure 20). La proportion d'individus porteurs de chromosomes dits recombinants est dépendante de la distance séparant ces deux loci. C'est de

cette manière que se définit la distance génétique : le centimorgan (cM), où un centimorgan correspond à une fréquence de recombinaison dans la descendance de 1 %. Chez l'homme, 1cM correspond en moyenne à 1 mégabase (Mb). A titre d'exemple, si l'on considère deux gènes X et Y avec respectivement deux allèles Aa et Bb. Si les parents sont respectivement Aa et Bb avec les haplotypes suivants AB et ab et si ces loci sont liés physiquement, c'est à dire très proches l'un de l'autre, la probabilité de recombinaison entre les deux loci est très faible. Les descendants porteront les haplotypes non recombinants AB ou ab. Par contre, si ces deux gènes sont éloignés de 5 cM par exemple, laissant la possibilité de recombinaison, nous retrouverons dans la descendance 5 % d'haplotypes recombinants, à savoir Ab ou aB.

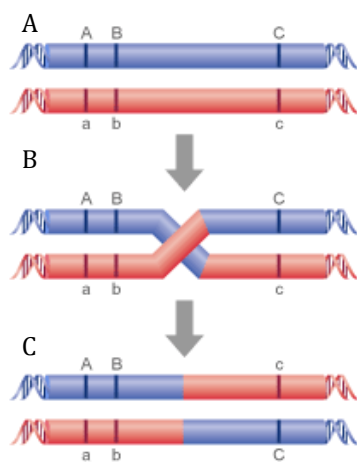


Figure 20 : Liaison, déséquilibre de liaison et recombinaison. A) Soit un individu avec son chromosome d'origine paternelle (bleu) qui comporte 3 loci avec les allèles respectifs A, B et C et son chromosome d'origine maternelle (rouge) avec ces 3 mêmes gènes mais avec les allèles respectifs a, b et c. B) Le phénomène de recombinaison se déroulant lors de la formation des gamètes aura lieu préférentiellement entre les gènes les plus éloignés : B et C. C) Les allèles A et B, et a et b étant proches, sont liés physiquement et seront transmis ensemble de manière préférentielle à la descendance. Les loci Aa et Bb sont donc en déséquilibre de liaison. Les proportions des allèles ABC, Abc, abc et abC retrouvées dépendront de la distance entre Bb et Cc.

La liaison physique entre deux loci peut être déterminée en génétique par la mesure du déséquilibre de liaison (*linkage disequilibrium*) ou son opposé, l'équilibre de liaison. On dit que deux ou plusieurs allèles sont **en déséquilibre de liaison** lorsque deux allèles correspondant à deux loci distincts sont plus fréquemment associés dans une population que ne le voudrait le hasard. Dans le cas de deux loci bi-alléliques avec A/a et B/b comme allèles, il est défini par $D_{AB} = x_{AB} - p_A q_B$, où D_{AB} est le déséquilibre de liaison entre A et B, x_{AB} correspond à la fréquence de l'haplotype AB et p_A ou q_B correspondent à la fréquence de l'allèle A ou de l'allèle B. Lorsque $D_{AB} = 0$, on dit que les allèles A et B sont en équilibre de liaison. Lorsque $D_{AB} \neq 0$, on dit que les allèles A et B sont en déséquilibre de liaison et par conséquent que les haplotypes AB et ab se retrouveront plus fréquemment dans la population. Autrement dit, le déséquilibre de liaison correspond à la différence entre la fréquence haplotypique observée et la fréquence haplotypique attendue, lorsque les deux loci ségrègent de manière indépendante (Ardlie *et al.*, 2002). D est facile à calculer mais il a le désavantage de dépendre de la fréquence des différents allèles. Une alternative proposée par Lewontin (1964) est de normaliser la valeur de D avec le calcul de D' . Si $D_{AB} \geq 0$, $D'_{AB} = D_{AB}/D_{max}$ et D_{max} correspond à la plus petite valeur entre $p_A(1-p_B)$ et $p_B(1-p_A)$ et si $D_{AB} < 0$, $D'_{AB} = D_{AB}/D_{min}$ avec D_{min} la plus grande valeur $-p_A(1-p_A)$ et $-p_B(1-p_B)$. D' est

égal à 1 lorsqu'au moins un des quatre haplotypes possibles est absent. Une autre méthode utilisée, est le calcul du r^2 qui correspond au coefficient de corrélation entre deux allèles de deux loci (Hill et Robertson, 1968) et est défini par $r^2_{AB} = D_{AB}^2 / [p_A(1-p_A) p_B(1-p_B)]$ où $p_A p_a q_B q_b$ correspondent au produit des fréquences des allèles A, a, B et b. $r^2 = 1$ quand au moins deux des haplotypes possibles sont absents et $r^2 = 0$ lorsque les deux loci sont en équilibre de liaison.

Le déséquilibre de liaison est influencé par de nombreux facteurs. Il dépend dans un premier temps de la **liaison physique** entre les allèles de deux gènes situés plus ou moins loin l'un de l'autre sur le même chromosome. Comme dit précédemment, plus ils sont proches plus ils auront tendance à être transmis ensemble et donc l'haplotype se retrouvera préférentiellement dans la population. Le déséquilibre de liaison diminue en fonction de la distance entre les loci, la distance favorisant le phénomène de **recombinaison**. Cependant les zones de recombinaison ne se produisent pas n'importe où mais bien sur des zones dites de « *hot-spots* » où elles ont lieu préférentiellement. Les zones situées entre celles de recombinaison définissent des blocs haplotypiques où tous les loci sont en déséquilibre de liaison. Le déséquilibre de liaison peut également s'observer pour des allèles se trouvant sur des chromosomes différents si ceux-ci sont soumis à la même pression de **sélection**. Le déséquilibre de liaison dépend principalement de l'**histoire de la population** (ex. la migration, l'expansion, la sélection), de sa taille efficace et des événements qui l'ont affectée comme l'apparition de **mutations**, (Slatkin, 2008). La **taille efficace** d'une population est définie comme étant le nombre d'individus dans une population idéale, non sélectionnée avec des accouplements aléatoires qui entrainerait le même taux de consanguinité que dans la population observée. Elle est donc bien plus petite que la population observée car tous les individus ne contribuent pas tous de la même manière à la génération suivante. C'est de cette manière que, dans les races bovines, où la domestication et leur création sont récentes, le nombre de générations est faible, induisant un nombre de recombinaisons peu élevé. La sélection intensive réduit considérablement la taille efficace. On retrouve alors de longs blocs haplotypiques en déséquilibre de liaison de plus d'1 Mb (Farnir *et al.*, 2000 ; de Roos *et al.*, 2008 ; Khaktar *et al.*, 2008). Tandis que chez l'homme où la population efficace a augmenté depuis environ 10.000 ans et où la sélection est moins forte, le déséquilibre de liaison se limite à de courtes distances, environ 60 kilobases (kb) (Reich *et al.*, 2001).

La caractérisation du déséquilibre de liaison dans le génome bovin a des implications majeures dans l'étude du déterminisme génétique de caractères héréditaires. L'existence de blocs haplotypiques autorise la sélection de milliers de marqueurs représentatifs à génotyper, sur les millions connus, afin de diminuer les coûts de génotypage. Ensuite, il sera possible de combler les espaces entre les marqueurs utilisés avec les informations disponibles au niveau de la population par imputation. En cartographie, la notion de déséquilibre de liaison est essentielle

puisqu'elle permet de déterminer des marqueurs génétiques qui sont associés au caractère étudié et en déséquilibre de liaison avec le ou les loci associés à ce caractère.

La connaissance des variations dans la séquence d'ADN (marqueurs), leur localisation sur le génome et la mesure de leur transmission ont favorisé le développement des méthodes de clonage positionnel sur génome entier qui sont de deux types : l'analyse de liaison et l'analyse par étude d'association.

c) Etudes de liaison

La cartographie par analyse de liaison a connu un essor dans les années 80' avec la découverte des premiers marqueurs basés sur l'ADN (Altshuler *et al.*, 2008). L'analyse de liaison étudie la transmission préférentielle d'allèles dans une famille informative au sein de laquelle le caractère d'intérêt ségrège au sein de cette famille (Hischhorn et Daly, 2005). La détermination de l'ensemble des marqueurs (haplotypes) qui se transmettent avec le caractère permet d'identifier dans un premier temps le ou les loci associés. Les études de liaison ont principalement été menées à la fin du XX^{ème} avec l'utilisation de microsatellites au sein de familles, limitant le nombre de recombinaisons et ainsi la résolution de la cartographie (Goddard et Hayes, 2009). On distingue deux types de méthodes statistiques pour mettre en évidence une liaison entre le phénotype et le génotype : les méthodes paramétriques et les méthodes non-paramétriques. Les méthodes dites **paramétriques** consistent à évaluer la transmission d'allèles marqueurs et du phénotype étudié sous un modèle d'hérédité supposé (récessif, dominant...) et à tester, à l'aide d'une méthode de vraisemblance (*likelihood*) si leur transmission est ou non indépendante. La méthode du **LOD score** (*logarithm of the odds* = logarithme des probabilités) est la plus couramment utilisée dans les études de liaison. Cette méthode dépend de la distance entre les deux loci qui est déterminée par le taux de recombinaison (θ) et qui correspond à la proportion d'individus recombinants pour deux loci au sein de la descendance étudiée. Comme décrit précédemment, plus la distance entre les deux loci est importante plus la possibilité de recombinaisons est grande et donc la proportion d'individus recombinants est élevée. Le LOD score correspond au logarithme en base 10 du rapport entre le maximum de vraisemblance sous l'hypothèse d'une liaison génétique entre un locus et le gène responsable ($0 \leq \theta \leq 0,5$) et le maximum de vraisemblance sous l'hypothèse de n'observer aucune liaison ($\theta = 0,5$) ($\text{LOD score} = \log_{10} [L(\theta = 0 \leq \theta \leq 0,5) / L(\theta = 0,5)]$). En général, on estime que la liaison génétique entre le locus étudié et le gène responsable du phénotype est significative, dans le cas de caractère mendélien, lorsque le LOD score est supérieur à 3. La liaison observée a 1000 fois plus de chances de ne pas être observée par hasard. Le LOD score est donc dépendant du

nombre de méioses informatives que l'on observe dans le pedigree étudié. Les méthodes dites **non-paramétriques** quant à elles ne font aucune hypothèse sur le mode d'hérédité mais testent si les individus présentant le phénotype d'intérêt au sein de la famille étudiée partagent le(s) même(s) allèle(s) (*allele sharing*) et si leur transmission dans la descendance ne respecte pas les proportions attendues. On détermine dans ce cas les haplotypes identiques transmis dans la descendance atteinte (*IBD = Identity-by-descent*, en opposition à *IBS = Identity-by-state*, où les allèles sont identiques dans la population mais proviennent d'ancêtres récents différents) (Balding, 2006 ; Ott *et al.*, 2011).

La résolution des études de liaison est dépendante du nombre de recombinaisons au sein de la famille étudiée, de la densité de marqueurs au locus d'intérêt et de l'effet du locus recherché sur la variation du phénotype observé. Plus le nombre de générations est important, plus le nombre de recombinaisons possibles est élevé, meilleure en est la résolution. L'utilisation de cette méthode dans l'étude de caractères mendéliens, pour lesquels la pénétrance est élevée, s'est révélée efficace tant chez l'homme que chez les animaux, avec l'utilisation de quelques dizaines d'individus ayant une résolution de quelques cM soit quelques Mb. Par contre, pour des caractères complexes où les variants n'expliquent qu'une faible proportion de la variation phénotypique, la résolution de la détection des QTL est de l'ordre de quelques dizaines de cM (Goddard et Hayes, 2009). Chez l'homme, l'efficacité est moindre due au faible nombre d'individus utilisés au sein des familles analysées (Hirschhorn et Daly, 2005). Chez les animaux de production, principalement chez les bovins où un taureau d'IA peut avoir des centaines de descendants et où l'on peut générer des pedigrees expérimentaux, l'utilisation de cette méthode de cartographie tant pour l'étude de caractères mendéliens que pour des caractères quantitatifs est efficace (Georges *et al.*, 1993 ; Georges *et al.*, 1995 ; Coppieters *et al.*, 1998). En BBB, cette méthode a permis notamment de cartographier le locus associé à l'hypertrophie musculaire et le locus *Rouan* (Charlier *et al.*, 1995 ; Charlier *et al.*, 1996).

d) Etudes d'association

Les études d'association du génome entier (GWAS) ont connu leur essor au début du XXI^{ème} siècle avec le développement des plateformes de génotypage à haut débit qui a rendu possible le génotypage rapide de centaines d'individus pour des milliers de marqueurs (chez les bovins, 3.000, 6.000, 50.000 puis 770.000 marqueurs de type SNP). Contrairement aux études de liaison, les GWAS consistent à comparer, au niveau d'une population d'individus « non apparentés » (dans les races bovines, les individus sont toujours apparentés à un ancêtre plus ou moins éloigné), la fréquence des allèles pour chaque marqueur entre des cas et des contrôles

pour le phénotype étudié. Dans le cas où un ou plusieurs allèles marqueurs sont surreprésentés dans la population de cas, on dira que ce locus est associé à ce phénotype et qu'il contient la ou les mutations causales. Ce type d'analyses est dépendant du déséquilibre de liaison entre le(s) gène(s) responsable(s) du phénotype étudié et les marqueurs disponibles situés à proximité de ce(s) gène(s) (Hirschhorn et Daly, 2005). Cette méthode ne suppose préalablement aucun mode d'hérédité et la résolution qui dépend de la population étudiée est de l'ordre de centaines de kb à quelques Mb. Chez l'homme, elle est de l'ordre d'une centaine de kb contenant quelques dizaines de gènes (Altshuler *et al.*, 2008).

Analyses statistiques

Il existe quatre types différents de GWAS qui dépendent du caractère étudié et du nombre de marqueurs analysés. On distingue ainsi l'analyse simples points (SNP par SNP) et l'analyse multipoints (haplotype par haplotype). Le principe des études de type GWAS est de comparer la fréquence allélique ou haplotypique entre deux populations, les cas et les contrôles, et de voir si un allèle ou un haplotype est associé significativement (surreprésenté chez les cas ou chez les contrôles dans le cas d'un effet protecteur de cet allèle) au caractère étudié. On définit également les analyses de type cas-contrôles et les analyses pour des caractères continus. Dans le cas d'étude de défauts génétiques, nous détaillerons les études de type cas-contrôles.

L'**analyse simples points**, à savoir marqueur par marqueur, est la méthode la plus élémentaire. Elle consiste à tester l'hypothèse nulle d'absence d'association dans une matrice 2 x 3 qui contient le nombre d'individus des trois génotypes (les deux homozygotes et l'hétérozygote) au sein des deux populations, les atteints et les contrôles. Les tests les plus simples à utiliser dans ce cas sont soit le test de Pearson à deux degrés de liberté, soit le test exact de Fischer (Balding, 2006). L'inconvénient de ce type d'analyse est qu'elle ne tient pas compte de la liaison génétique entre certains loci.

L'**analyse multiples points** permet de tirer toute l'information générée par la grande quantité de SNPs et de tenir compte de la transmission en blocs (haplotypes) de certains allèles. La méthode la plus couramment utilisée est basée sur l'analyse des haplotypes. Cette méthode nécessite préalablement de phaser les génotypes (reconstruire les haplotypes paternels et maternels) en utilisant des informations familiales au sein d'un pedigree (ségrégation mendélienne et liaison génétique) ou des informations au niveau de la population (déséquilibre de liaison) ou les deux (Windig et Meuwissen, 2004 ; Browning et Browning, 2007 ; Druet et Georges, 2010). Chez les animaux de production, la taille des pedigrees a favorisé l'utilisation de

l'information familiale pour le phasage des génotypes, alors que chez l'homme l'information de la population a été préférée. La reconstruction des haplotypes, dans le cas où les individus ne sont pas génotypés avec la même densité de panels (*ex.* 50.000 ou 700.000 SNPs chez les bovins) ou avec des panels différents, nécessite préalablement d'imputer les génotypes manquants.

Correction de la stratification

La principale source de faux-positifs, dans les études d'association et l'identification d'un SNP ou d'un haplotype associé au phénotype étudié, est l'existence de stratification (hétérogénéité due à des sous-populations ayant des fréquences alléliques différentes) dans la population étudiée. La stratification peut être due, soit à la présence d'individus apparentés, soit au mélange de différentes populations au sein des cas ou des contrôles. Cette association reflète la surreprésentation d'un allèle dans une sous-population et non l'association entre celui-ci et le phénotype étudié (Hirschhorn et Daly, 2005). Dans un premier temps, il est nécessaire de détecter la présence ou non de stratification dans la population. Différentes méthodes ont été développées afin de maîtriser cette source d'erreur dans les GWAS.

Le **contrôle génomique** est une méthode qui permet à la fois d'identifier la présence de stratification et de la corriger. Cette méthode est basée sur le calcul d'un facteur d'inflation λ qui est le rapport entre la médiane des valeurs de χ^2 obtenues lors du test d'association et la médiane théorique (Devlin et Roeder, 1999). Une valeur de λ supérieure à 1 indique la présence de stratification, il suffit de corriger les valeurs statistiques obtenues en les divisant par λ . Le contrôle génomique ne s'applique qu'aux études d'association de type simples points (Price *et al.*, 2010). L'inconvénient du contrôle génomique est qu'il corrige de manière uniforme tous les marqueurs à l'aide du même facteur d'inflation λ . Il ne tient pas compte de la différence des fréquences alléliques dans les différentes populations. Ce contrôle génomique sera insuffisant pour les marqueurs ayant une différence de fréquence allélique élevée entre les sous-populations et inutile dans le cas contraire, conduisant à une perte de puissance dans la détection des associations réelles (Price *et al.*, 2006).

L'**association structurée** est une méthode qui identifie d'abord la présence de sous-populations dans la population étudiée et qui vise à regrouper, sur base de marqueurs, les individus dans des clusters d'individus génétiquement proches en utilisant le programme STRUCTURE (Pritchard *et al.*, 2000). L'association est ensuite testée au sein de ces clusters (Pritchard *et al.*, 2000). L'inconvénient de cette méthode est que le programme utilisé pour générer les clusters sur d'importantes quantités d'informations nécessite de grandes capacités

de calculs informatiques. De plus, elle dépend fortement du nombre de clusters et il est difficile de définir le nombre de sous-populations qui composent la population étudiée (Price *et al.*, 2006).

L'**analyse en composantes principales** (ACP) permet de détecter la stratification et également de la corriger. L'ACP consiste à transformer un grand nombre de variables liées entre elles, en un nombre plus petit de nouvelles variables indépendantes les unes des autres, appelées composante principale, expliquant au mieux la variance (la dispersion) des données. La première composante est déterminée afin qu'elle explique une proportion maximale de la variance. La deuxième est perpendiculaire à la première (non liée) et de nouveau la somme des variances doit être la plus grande possible. L'inclusion de ces composantes principales dans l'étude d'association corrige pour les effets de stratification (Price *et al.*, 2010), par exemple avec le programme EIGENSTRAT.

L'utilisation de **modèles mixtes** trouve son origine dans l'évaluation génétique des animaux d'élevages et peut corriger pour la structure au niveau d'une population, d'une famille ou de relations cachées (Price *et al.*, 2010). Cette méthode consiste à modéliser les phénotypes observés comme une somme d'effets fixes (comme les marqueurs, le sexe, l'âge) et d'effets aléatoires (Yu *et al.*, 2006). Le modèle linéaire mixte est défini par : $y = X\beta + S\alpha + Qv + Z\mu + e$ où y correspond à la variable à expliquer à savoir les phénotypes, $X\beta$ correspond aux effets fixes autres que les marqueurs et la structure de la population (sexe, âge), $S\alpha$ aux effets fixes liés aux SNPs, Qv aux effets fixes liés à la structure de la population, $Z\mu$ aux effets polygéniques aléatoires et e est un vecteur rassemblant les erreurs résiduelles (Yu *et al.*, 2006). Cette méthode a l'avantage d'être flexible pour corriger les différents types de stratifications et d'améliorer le contrôle des erreurs de type I et de type II (Yu *et al.*, 2006). L'**erreur de type I** (faux négatif) est l'erreur commise quand l'hypothèse nulle est rejetée, alors qu'elle est vraie. L'**erreur de type II** (faux positif) est quant à elle commise quand l'hypothèse nulle est maintenue, alors qu'elle est fausse.

Les analyses de type étude de liaison ne souffrent pas de ce problème de faux positif lié à la stratification (Ott *et al.*, 2011). Une approche consiste à combiner l'étude de liaison génétique et l'étude du déséquilibre de liaison au sein de familles dans lesquelles le phénotype d'intérêt ségrège. Elles sont regroupées sous le terme d'**études d'associations basées sur la famille** (*family-based association studies*) dont la principale différence est l'utilisation de contrôles ayant un lien de parenté avec les cas (Ott *et al.*, 2011). Le test le plus classique dans ce type de méthode est le test de transmission d'un déséquilibre de liaison (*Transmission disequilibrium*

test = TDT) développé par Spielman et collaborateurs (1993). En l'absence d'association entre un allèle et le phénotype, cet allèle sera transmis de manière uniforme chez les cas et chez les contrôles. L'inconvénient de cette méthode est qu'elle nécessite d'avoir le génotype des parents.

Seuil de significativité

La troisième source d'erreurs, ou de faux positifs (la première étant due à des erreurs de génotypage et la seconde à la stratification) est liée à la détermination du seuil de significativité (Hirschhorn et Daly, 2005). Le seuil permet de confirmer si l'association identifiée entre le phénotype et l'allèle est significative, c'est-à-dire qu'elle est peu probable par hasard sous H_0 (dans l'absence d'association). Dans le cas d'études d'associations, où l'on analyse simultanément des milliers de marqueurs sur des centaines d'individus, la probabilité de trouver une association entre un SNP et un phénotype par hasard sera élevée or il serait erroné de conclure sur cette association. Si l'on considère le seuil de significativité classique de 5 %, cela veut dire que pour une étude réalisée sur 50.000 SNPs, on risquerait de trouver 2.500 associations faussement positives, ce qui est supérieur au nombre de vrais positifs auxquels on pourrait s'attendre. Diverses méthodes statistiques ont été développées afin d'adapter le seuil de significativité et de contrôler les erreurs de type I pour les tests multiples (Balding, 2006).

La méthode la plus simple dans l'analyse de tests multiples est la **correction de Bonferroni**. Elle consiste à diviser le seuil de significativité α de chaque test indépendant par le nombre n de tests indépendants réalisés, la valeur adaptée du seuil α' doit satisfaire à l'équation : $\alpha = 1 - (1 - \alpha')^n$, qui correspond approximativement à $\alpha' = \alpha/n$. A titre de d'exemple, si le seuil utilisé est de 0,05 (5 %) et que l'on utilise 50.000 marqueurs, le seuil α' corrigé est de 1×10^{-6} (Balding, 2006). Cependant, cette méthode est jugée trop conservative puisque l'on considère que les n SNPs sont indépendants, ce qui ne tient pas compte du déséquilibre de liaison entre les marqueurs proches (Gao *et al.*, 2008). Elle va donc éliminer le risque de faux-positifs mais aussi certaines vraies associations, diminuant ainsi la puissance du test (Balding, 2006).

Une alternative à la correction de Bonferroni consiste à faire des **permutations** dont l'objectif est de déterminer la valeur seuil de faux positifs acceptables. Les permutations consistent à échanger les phénotypes entre les différents génotypes afin de tester l'hypothèse nulle, c'est-à-dire l'absence d'association entre les deux. Après des milliers de permutations, on estime le nombre de faux positifs auxquels on peut s'attendre dans ce cas de figure (Balding, 2006). Cette méthode efficace a l'avantage de générer un nouveau seuil de significativité à partir

des données générées. Malheureusement, elle est excessivement chronophage et requiert l'utilisation de systèmes de calculs informatiques très importants (Dudbridge *et al.*, 2006).

C. Etude du déterminisme génétique par l'approche de type gène candidat

La méthode de type gène candidat est une méthode d'analyse de type étude d'association qui consiste dans un premier temps à sélectionner un gène candidat sans localisation préalable en fonction de son implication dans le déterminisme du caractère étudié ou parce que des mutations sont connues pour être responsables de phénotypes similaires dans d'autres organismes. Ensuite, le gène sélectionné sera séquencé (*cf.* D. Identification et validation de la mutation causale, p. 63) afin d'identifier des variants (mutations) modifiant significativement la fonction de la protéine produite. La fréquence de ces variants sera comparée dans la population étudiée (*ex.* les atteints) et dans une population contrôle (Hirschhorn et Daly, 2005). La fréquence élevée de certains variants dans la population étudiée pourrait alors expliquer le phénotype étudié. Cette méthode a montré son succès pour la découverte de mutations responsables de caractères mendéliens (*ex.* identification de la mutation responsable de la BLAD ; Schuster *et al.*, 1992). Concernant l'étude de caractères complexes, son utilisation est limitée. En effet, le nombre important de gènes impliqués dans ces caractères ainsi que le manque de connaissance de certains mécanismes peut rendre ces méthodes très coûteuses et chronophages (Andersson, 2001 ; Georges, 2012). De plus, les méthodes statistiques utilisées manquent de spécificité, ne tiennent pas compte de la structure de la population bovine et les loci n'ont pas été validés et reproduits dans des cohortes indépendantes (Georges, 2012).

D. Identification et validation de la mutation causale

L'étape suivante à la cartographie fine, dans l'identification de la mutation causale, consiste en l'analyse par séquençage du ou des gènes candidats positionnels cartographiés lors d'analyses de liaison ou d'études d'association. Les premières techniques de séquençage ont été développées dans les années 70' avec notamment la méthode Sanger (Sanger *et al.*, 1977). Cette méthode est basée sur la synthèse du brin complémentaire au simple brin, à séquencer avec l'incorporation des quatre désoxyribonucléotides (dATP, dCTP, dGTP, dTTP) et l'incorporation d'une faible proportion d'un des didésoxyribonucléotides (ddATP, ddCTP, ddGTP et ddTTP) stoppant la synthèse du brin néoformé. Ces brins seront ensuite séparés par électrophorèse sur

un gel de polyacrylamide déterminant l'ordre des bases en fonction de la taille du brin synthétisé. Cette méthode suivra l'évolution du développement de technologies comme la PCR, la synthèse d'amorces oligonucléotidiques et permettra d'automatiser et d'accélérer le séquençage (Smith *et al.*, 1986). Le séquençage des régions codantes et des jonctions intron-exons tentera d'identifier des mutations qui, dans la plupart des cas des maladies mendéliennes, sont des mutations de type perte de fonction localisées dans ces régions. Cette méthode restera la plus utilisée jusqu'au milieu des années 2000 et le développement des nouvelles technologies de séquençage.

Le développement des **nouvelles technologies de séquençage** (*Next generation sequencing* = NGS) et la commercialisation de différentes plateformes ont ouvert de nouvelles perspectives à la génomique et à l'identification de variants. Les NGS se décomposent en plusieurs étapes avec i) la préparation de l'ADN génomique en vue du séquençage avec la fragmentation de l'ADN génomique, l'insertion d'adaptateurs aux extrémités et l'amplification de ces fragments (bibliothèques), ii) le séquençage de ces bibliothèques d'ADN préparées, iii) la lecture par bioluminescence des séquences produites et iv) l'alignement des fragments de séquences obtenues sur un génome de référence s'il est disponible ou l'assemblage *de novo* lorsqu'une séquence de référence n'est pas connue ou incomplète (Metzker, 2010). Chaque étape est dépendante du type de plateforme utilisée. Il existe trois plateformes principalement utilisées à ce jour. La technologie développée par **Roche / 454 FLX** qui est basée sur une amplification grâce à une PCR en émulsion où chaque fragment d'ADN est amplifié sur une bille contenue dans une goutte. Après amplification, chaque goutte est placée dans le pore d'une plaque en fibre optique picotitrée et le séquençage est réalisé par une réaction de pyroséquençage. La technologie mise au point par **ABI / SOLiD** (Support Oligonucleotide Ligation Detection) est également basée sur une amplification par PCR en émulsion mais le séquençage se réalise par ligation. Enfin, la technologie utilisée au cours de cette thèse est développée par **Illumina / Genome Analyzer II** puis **Hiseq 2000** avec une amplification en pont sur une plaque solide des fragments immobilisés par une sonde spécifique complémentaire des adaptateurs puis le séquençage se fait par l'ajout de nucléotides fluorescents réversibles et par une lecture de fluorescence. Une fois lu, le marqueur fluorescent est clivé et des agents réducteurs reforment une terminaison hydroxyle à l'extrémité 3' permettant l'incorporation d'une nouvelle base marquée. Ces technologies permettent d'obtenir des milliers, voire des millions de séquences de ces fragments qui seront ensuite assemblés et alignés. Les technologies de NGS ont ainsi réduit le coût du séquençage, passant d'environ 2.400 \$/Mb avec la méthode Sanger à moins d'1\$/Mb avec les NGSs (Metzker, 2010). L'identification de la mutation causale responsable de défauts mendéliens est parfois longue et reste l'étape limitante lorsqu'aucun gène candidat évident ne

ressort et que le nombre de gènes présents dans la région cartographiée reste élevé. Ces nouvelles technologies permettent alors le reséquençage de la région d'intérêt ou du génome entier d'individus atteints afin de les identifier (*ex. Charlier et al., 2012*). Ces NGSs pourront également être appliquées à l'analyse de l'ARN (Acide ribonucléique) via le séquençage massif à haut débit de bibliothèques d'ADN complémentaire (ADNc). Ces méthodes permettront ainsi d'avoir une analyse du transcriptome et de ses modifications, qui peuvent être dues à des mutations affectant des sites d'épissage ou des régions régulatrices (Ozsolak et Milos, 2010). Une autre évolution permettant d'améliorer l'identification de mutations causales, qui sont majoritairement localisées dans les régions codantes, est le NGS associé à la capture de l'exome correspondant à l'ensemble des régions codantes (Bamshad *et al.*, 2011).

Toutes ces méthodes ont pour objectif de générer des données de séquences qui seront alignées sur le génome de référence du bovin afin d'être analysées. Les analyses des séquences obtenues par séquençage classique du gène candidat ou par séquençage du génome entier ont pour but d'identifier les variations ayant des conséquences sur la fonction du ou des gènes concernés.

On dénombre trois types de **mutations ponctuelles**, les substitutions, les insertions et les délétions. Les **substitutions** qui correspondent au changement d'une base azotée en une autre, se déclinent sous deux formes, la **transition** (changement d'une purine en une purine ou d'une pyrimidine en une pyrimidine) et la **transversion** (changement d'une purine en une pyrimidine ou inversement). Lorsqu'une substitution se trouve dans la région codante (exon) d'un gène, elle peut conduire à i) l'absence de changement de l'acide aminé = **mutation synonyme** (*silent mutation*), sans répercussion au niveau protéique, ii) un changement d'un acide aminé en un autre conduisant à une modification de la séquence et altérant les structures et/ou la fonction protéine = **mutation non-synonyme** (*missense mutation*) et iii) un changement d'un acide aminé en un codon-stop induisant la production d'une protéine tronquée voire l'absence de protéine = **mutation non-sens** (*nonsense mutation*). L'**insertion** ou la **délétion** d'une paire de bases, induisent un décalage de la phase de lecture et l'apparition d'un codon-stop prématuré avec la production d'une protéine tronquée voire absente. Il existe également des **polymorphismes structuraux** impliquant plusieurs bases. L'insertion ou la délétion de plusieurs paires de bases lorsque le nombre n'est pas un multiple de 3 provoque un décalage de la phase de lecture et l'apparition d'un codon stop prématuré. L'apparition d'un codon-stop prématuré, quelle que soit son origine, peut avoir deux conséquences en fonction de sa localisation. Lorsque le codon stop apparaît dans le dernier exon, la protéine sera produite et tronquée. Par contre, lorsqu'il intervient dans les autres exons, le mécanisme de dégradation de l'ARNm non-sens (*nonsense-mediated RNA decay = NMD*) est un mécanisme de contrôle, qui en

général détruit l'ARN messager qui comporte un codon stop prématuré et empêche la traduction d'une protéine tronquée (Chang *et al.*, 2007). Ces variations ponctuelles ou structurales peuvent apparaître au niveau des codons start ou des codons stop, ce qui modifiera principalement l'étape de la traduction et la synthèse d'une protéine modifiée. Dans les polymorphismes structuraux, on peut observer également des **inversions**, correspondant à un changement d'orientation d'un segment de chromosome et des **translocations** correspondant à une transposition d'une partie de chromosome vers un chromosome non-homologue. Ces modifications structurelles peuvent modifier l'expression d'un ou de plusieurs gènes, soit en plaçant la séquence dans un environnement différent (régions régulatrices), soit en brisant la séquence aux endroits de cassures des fragments chromosomiques au niveau de la région donneuse et/ou au niveau de la région réceptrice.

Une **variation du nombre de copies** (*Copy Number Variants = CNV*) causée par des **délétions** ou des **duplications** peut faire varier le nombre de copies des gènes pouvant avoir une influence sur le déterminisme de certains phénotypes (Georges, 2007). Ces variants structuraux sont fréquents, chez l'homme, mais bien qu'ils semblent incriminés dans le déterminisme de certaines maladies complexes, le mécanisme et l'implication de tous les CNVs ne semblent pas si clairs (Craddock *et al.*, 2010 ; Stankiewicz et Lupski, 2010).

Ces mutations peuvent également affecter les régions non-codantes, comme les régions régulatrices ou les sites d'épissage. Dans le premier cas, ces mutations vont modifier l'expression du gène qu'elles régulent. Dans le second cas, les mutations affectant les sites d'épissage (sites donneurs ou sites accepteurs) peuvent conduire à la non inclusion de l'exon concerné, à un décalage de la phase de lecture si l'exon n'est pas en phase et à la production d'un ARNm tronqué qui sera ou non dégradé par le NMD. L'apparition de certaines mutations peut également conduire à la création d'un nouveau site d'épissage dans les régions introniques conduisant à l'inclusion d'une partie de l'intron et à la modification de la phase de lecture. Certaines mutations affectant les sites d'épissage peuvent conduire à l'utilisation d'un site cryptique d'épissage présent dans la région codante mais non utilisé normalement (*ex. Sartelet et al.*, 2012). Ces variations peuvent donner lieu à des mutations de type **perte de fonction**, souvent avec un mode de transmission récessif rendant la protéine mutante non fonctionnelle et responsable d'un phénotype mais également à des mutations de type **gain de fonction**, généralement avec un mode de transmission dominant, donnant à la protéine une fonction nouvelle et anormale.

Une fois identifiée, la validation de la causalité de la mutation nécessite i) l'identification de phénotypes similaires dans d'autres espèces liés à des mutations dans le gène orthologue (gène dans deux espèces différentes issu du même gène chez l'ancêtre commun), ii) l'absence de cette mutation dans d'autres races, iii) l'absence d'individus homozygotes mutants dans une

population contrôle saine de la même race, iv) l'absence d'autres mutations pertes de fonctions dans l'intervalle cartographié et enfin v) la confirmation au niveau moléculaire de la modification de fonction de la protéine (expression, structure, fonction) (Georges, 2012).

E. Développement d'un test diagnostique

Le développement d'un test diagnostique permet d'identifier les individus porteurs afin d'éviter les croisements à risque et donc la naissance de veaux mutants.

Au cours de cette thèse, le génotypage des individus a été réalisé grâce la technologie Taqman© basée sur l'activité exonucléasique de la polymérase. Les sondes, une pour l'allèle sauvage et une pour l'allèle muté, sont marquées à leur extrémité 5' par un fluorochrome émetteur (*reporter*) et à leur extrémité 3' par un fluorochrome suppresseur (*quencher*) fluorescent ou non, qui inhibe l'émission du reporter lorsqu'ils sont à proximité. Au cours de la PCR, si la sonde est hybridée sur sa cible, elle est hydrolysée par l'ADN polymérase. Le reporter ainsi séparé du quencher émet un signal proportionnel au nombre de sondes hydrolysées (Figure 21A). L'intensité du signal présent dans l'échantillon est mesuré de manière optique et permet de déterminer le génotype de l'individu (Figure 21B).

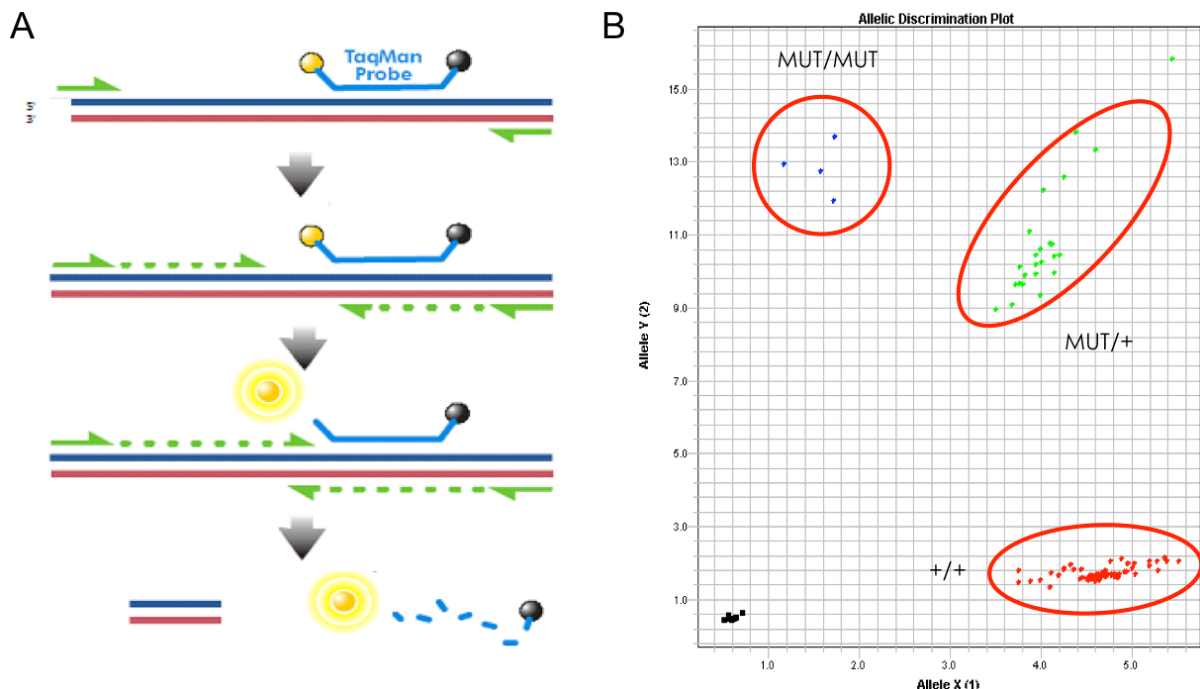


Figure 21 : A. Principe du génotypage avec la technologie Taqman©, avec la sonde (bleue) spécifique du brin sauvage ou du brin muté portant à l'extrémité 5' un fluorophore (jaune) et à son extrémité 3' le suppresseur (noir), une fois liée au brin d'ADN à génotypé. Lors de l'amplification à partir des amorces (vertes), l'activité exonucléasique de la polymérase va cliver la sonde et séparer le fluorophore du suppresseur qui va émettre son signal. B. Résultat de la lecture en fonction du fluorophore et donc de l'allèle détectée avec les 3 populations : homozygotes sauvages (+/+) en rouge, hétérozygotes (MUT/+) en vert et homozygotes mutés (MUT/MUT) en bleu.

Le test direct est basé sur la détection de la mutation elle-même. Sa spécificité et sa sensibilité sont de 100 %. La **spécificité** est la capacité d'identifier un individu non porteur alors qu'il n'est pas porteur de la mutation, l'objectif est d'avoir le moins de faux-négatifs possible. La **sensibilité** est la capacité de détecter un individu porteur à l'aide de ce test alors que celui-ci est bien porteur de la mutation, l'objectif est d'avoir le moins de faux-positifs possible. Dans ce cas, le nombre de faux négatifs et de faux positifs est nul.

Le délai entre la localisation et l'identification de la mutation causale peut être long et atteindre quelques mois, voire des années. Le test indirect peut être envisagé dès lors que le défaut génétique est localisé à un endroit particulier du génome mais que la mutation et le gène responsables ne sont pas encore connus. Le test est alors basé sur l'analyse d'une série de marqueurs génétiques qui signent l'identité de la région localisée. Ce test est aussi connu sous le nom de SAM. L'occurrence de recombinaisons au sein du locus à risque engendre un risque de **faux négatifs**. L'existence de **faux positifs** est possible dans deux cas de figures. A savoir, un individu peut être porteur du même haplotype que celui de la région IBD homozygote mais sans la mutation, lorsque cet haplotype provient d'un ancêtre commun antérieur à l'individu fondateur. La deuxième possibilité se rencontre lorsque les allèles aux marqueurs génotypés sont fréquents dans la population et sont identiques aux allèles à risque dans la population. La sensibilité d'un test indirect est de plus de 95%, ce qui veut dire que dans moins de 5% des cas, un animal réellement porteur sera déclaré non porteur. De même, sa spécificité n'est pas absolue, elle est de l'ordre de 95%, c'est-à-dire que dans 5% des cas, un animal déclaré porteur sera en réalité non porteur. Les risques d'erreurs peuvent être diminués si des informations de pedigree sont disponibles pour l'individu testé, si l'on augmente le nombre de marqueurs ou si ses parents sont aussi testés.

OBJECTIFS DU TRAVAIL

La race BBB a un impact majeur dans le paysage agricole belge et même mondial, tant au niveau économique, que social et culturel. Le haut niveau de sélection imposé à cette race a malheureusement favorisé l'émergence de certains défauts.

L'expérience acquise dans l'identification de mutations causales en utilisant les nouvelles technologies de génotypage pour différentes tares autosomales récessives bovines (DMC 1 et 2 en BBB) a permis d'identifier des facteurs limitants comme la récolte des cas, le phénotypage précis (surtout pour des phénotypes proches) et enfin l'identification de la mutation causale (Charlier *et al.*, 2008).

L'objectif de ce travail était de mettre en place une cellule d'hérédo-surveillance, en temps réel au sein de la race BBB, afin d'identifier et de gérer au plus vite l'émergence des défauts génétiques.

Pour arriver à cet objectif, nous devons :

- 1) développer un réseau facile d'accès entre les principaux acteurs de la scène BBB en Belgique,
- 2) créer une liste complète des défauts et les classer en fonction de leur impact économique (mortalité, réforme, zootechnie) dans les élevages BBB,
- 3) créer une biobanque et une base de données alimentée par les différents cas collectés et phénotypés, les parents et toutes les informations épidémiologiques disponibles,
- 4) identifier, sur bases de l'analyse épidémiologique, les défauts ayant une composante héréditaire,
- 5) identifier par cartographie le locus associé à chaque défaut, à l'aide d'un génotypage du génome entier d'une cohorte de cas préalablement sélectionnés et d'une cohorte de contrôles, en utilisant les méthodes statistiques adéquates,
- 6) identifier la mutation causale, la caractériser et la valider à l'aide de tests fonctionnels,
- 7) développer un test diagnostique indirect ou direct et estimer la fréquence de la mutation dans la race BBB,
- 8) diffuser l'information via les différents canaux de communication à tous les membres du réseau,

9) évaluer l'efficacité de la mise en place de la cellule d'hérédo-surveillance au sein du programme de sélection de la race BBB par rapport au nombre de cas recensés, au taux de mortalité périnatale et à l'évolution de la fréquence de(s) mutation(s),

10) estimer à plus long terme, l'incidence et identifier des loci responsables ou favorisant le développement de défauts dits complexes comme la maladie des génisses blanches.

RESULTATS

1- INTRODUCTION

Une ébauche de cellule d'hérédo-surveillance a vu le jour dans les années 90' à l'initiative des Professeurs Roger Hanset, Frédéric Rollin, Daniel Desmecht, Michel Georges, du Dr Carole Charlier et de quelques praticiens comme les Drs Dive et Frennet. Un premier succès sera l'identification des mutations causales responsables des dystonies musculaires congénitales de type 1 et de type 2 (DMC 1 et 2) (Charlier *et al.*, 2008).

Au 1er octobre 2007, avec le financement du Ministère de l'Agriculture de la Région Wallonne, la cellule d'hérédo-surveillance sous la dénomination « *Rilouke* » (signifiant regarder en wallon) a été mise en place. Une sensibilisation des secteurs intervenant dans la gestion la race BBB (vétérinaires, éleveurs, CIAs, association d'élevages, instances sanitaires : Arsia et DGZ) via les différentes voies de communication (la presse vétérinaire et agricole, les différents sites internet, les mailing listes des différentes associations d'élevages et de vétérinaires et les interventions publiques) a été réalisée, afin de centraliser l'observation des cas d'anomalies congénitales auprès d'une seule personne. Cette cellule aura récolté, avec la collaboration d'environ 200 vétérinaires et 500 éleveurs, au 31 mars 2013 près de 2.600 échantillons biologiques comprenant 1.200 cas et 1.400 parentés avec les informations épidémiologiques s'y rapportant. Ce recensement a identifié, dans un premier temps, les principales anomalies congénitales observées dans la race BBB (Annexe 5). Nous avons ainsi déterminé trois grandes classes d'anomalies congénitales en fonction des pertes engendrées : i) les anomalies létales ou non-viables à la naissance, ii) les anomalies létales ou non viables ayant des répercussions sur l'avenir de la mère, et iii) les anomalies affectant la production de l'individu (*ex.* retard de croissance, fertilité) (Annexe 5).

Pour l'étape de cartographie, nous avons utilisé le programme développé par Druet et Georges (PHASEBOOK, 2010) qui consiste dans un premier temps à phaser les génotypes et ensuite à regrouper en clusters (groupes) d'haplotypes ancestraux en fonction de leurs similarités. La fréquence de ces haplotypes ancestraux est ensuite comparée entre les cas et les contrôles en utilisant un modèle linéaire mixte généralisé développé par Zhang et collaborateurs (2012) qui inclut un effet polygénique pour tenir compte de la stratification de la population. Lorsqu'un locus hautement significatif est mis en évidence, la taille de la région IBD homozygote est visualisée directement au niveau des SNPs dans cette région.

Les résultats qui suivent présentent successivement la caractérisation de cinq défauts héréditaires se transmettant de manière autosomale récessive et leur localisation génomique et pour quatre d'entre eux, l'identification de la/des mutation(s) causale(s) et leur caractérisation.

Une évaluation des effets de la SAM pour les défauts génétiques en BBB a été réalisée sur le nombre de cas observés, la fréquence des porteurs dans différentes populations et enfin la mortalité en Wallonie.

De plus, la cellule d'hérédo-surveillance s'est également intéressée à la couleur de robe blanc-dos ségrégeant dans la race BBB, avec près de 70 individus collectés. L'étude de cette robe a mis en évidence un nouveau mécanisme moléculaire de translocation réciproque, via un intermédiaire circulaire, à l'origine d'une variation du nombre de copies (Durkin *et al.*, 2012 ; Annexe 6).

2- SELECTION BALANCEE DU LOCUS *MRC2* BOVIN RESPONSABLE DU SYNDROME DE LA QUEUE TORDUE

A. Résumé



Entre novembre 2006 et novembre 2007, une centaine de veaux présentant une attache haute et une déviation de la queue (Syndrome de la Queue Tordue = SQT) ont été rapportées à la cellule d'hérédosurveillance. Une analyse phénotypique complète a identifié un retard de croissance, une musculature exacerbée et une tête courte et trapue comme symptômes principaux. De la parésie spastique des membres postérieurs, des membres antérieurs courts et en extension et une scoliose sont présents sur certains veaux. Cette tare n'est pas létale mais dans 25 % des cas, l'euthanasie est préconisée.

L'étude épidémiologique a retracé, du côté paternel et maternel, chaque individu atteint vers un ancêtre commun *Précieux de Somme*, un taureau d'IA né en 1980 et d'ainsi fortement suspecter un mode de transmission autosomale récessive.

Une étude d'association réalisée à partir de 8 cas génotypés avec un panel de 50.000 marqueurs a préalablement identifié une région identique par descendance (IBD) homozygote de 2,42 Mb (Chalier *et al.*, 2008). Cet intervalle a été réduit à 812 kb, en génotypant 105 cas pour cinq SNPs compris dans l'intervalle, réduisant le nombre de gènes candidats à sept. Une délétion de 2 bp (*c.2904-2905delAG*) a été mise en évidence, par le séquençage des régions codantes de ces gènes, dans l'exon 20 du gène du récepteur à mannose de type C2 (*MRC2*) codant pour la protéine Endo180. Cette protéine transmembranaire, principalement des chondrocytes, des ostéocytes et des fibroblastes intervient dans le remodelage de la matrice extracellulaire (Behrendt, 2004). Cette délétion de 2bp induit un décalage de la phase de lecture, l'apparition d'un codon stop prématuré et la production d'une protéine tronquée non-fonctionnelle.

L'apparition d'un codon stop prématuré dans l'exon 21 sur 30 soumet l'ARNm mutant au système de dégradation de l'ARNm non sens (Nonsense-mediated decay = NMD). Le séquençage du cDNA et la RT-PCR quantitative, réalisés à partir d'ARNm extrait d'individus des trois génotypes, ont confirmé l'absence de transcrit de l'allèle muté. La réalisation d'un Western blot, d'un extrait de protéines de tissus d'individus des trois génotypes, a confirmé l'absence de protéine sauvage chez l'homozygote mutant et la réduction chez l'hétérozygote.

Un test diagnostique a été développé et le génotypage de 1.899 individus BBB sains n'a révélé aucun homozygote mutant mais de manière surprenante, il a dénombré 25 % de porteurs y compris le fondateur, *Précieux de Somme*. Cette fréquence élevée de porteurs dans la population laissait suspecter que la mutation conférait un avantage sélectif aux hétérozygotes. Une comparaison des cotations linéaires des différents phénotypes évalués dans le cadre du programme de sélection (Annexe 2) mis en place dans la race BBB, entre les porteurs et les homozygotes sauvages, a été réalisée. Les cotations linéaires évaluant la musculature étaient significativement plus élevées chez les porteurs. Par contre, les cotations relatives aux aplombs et à la taille étaient significativement plus faibles. La sélection intensive menée sur la musculature au détriment de la taille et des aplombs explique en partie la sélection balancée de cette mutation dans le gène *MRC2* et le taux élevé de porteurs dans la population, ce qui confirme que l'augmentation de la fréquence de cette mutation est due à une sélection des individus les mieux conformés.

La mise à disposition du test diagnostique aux éleveurs a rapidement permis d'éviter les accouplements à risque et la naissance de veaux atteints, réduisant considérablement les pertes économiques liées au SQT.



De manière surprenante, 18 cas présentant des symptômes identiques aux SQT précédemment décrits ont été recensés par la cellule d'hérédo-surveillance entre novembre 2009 et mars 2010. Tous sont issus de deux taureaux testés négatifs pour la délétion de 2 bp dans le gène *MRC2*. Le génotypage des deux taureaux a confirmé l'absence de l'allèle muté. Celui des 12 veaux atteints et des douze mères a montré qu'ils étaient tous hétérozygotes pour la délétion de 2bp.

Une substitution d'une thymine en une cytosine (*c.1906T>C*) au niveau de l'exon 13 par reséquençage complet du gène *MRC2* chez deux trios a été identifiée. Elle modifie une cystéine ultra-conservée, intervenant dans la formation d'un pont disulfure dans la troisième lectine de type C (CTLD3) du récepteur Endo180, en une arginine.

Un test diagnostique basé sur cette substitution a montré que les deux pères étaient hétérozygotes pour la substitution, que les 18 veaux étaient des hétérozygotes composés et que la fréquence dans la population BBB était inférieure à 0,3 %.

Au niveau protéique, une oligomérisation des récepteurs Endo180 mutés a été détectée par Western blot. Elle peut être expliquée par la formation de ponts disulfures entre les cystéines laissées libres dans le CTLD3 altérant la fonction de ce récepteur. Une immunofluorescence a montré une modification de la localisation des récepteurs mutés.

Comme démontré pour la délétion de 2bp dans le gène *MRC2*, il est probable que cette substitution ait été également sélectionnée par les éleveurs pour son amélioration de la musculature chez les porteurs.

B. Articles**Balancing selection of a frame-shift mutation at the *MRC2* gene accounts for the outbreak of the crooked tail syndrome in Belgian blue cattle**

Fasquelle C.*, Sartelet A.*, Li W., Dive M., Tamma N., Michaux C., Druet T., Huiljbers I.J., Isacke C.M., Coppieters W., Georges M., Charlier C.

PLoS Genetics, 2009, **5**, e1000666

* Contributed equally

Balancing Selection of a Frame-Shift Mutation in the *MRC2* Gene Accounts for the Outbreak of the Crooked Tail Syndrome in Belgian Blue Cattle

Corinne Fasquelle^{1,2}, Arnaud Sartelet^{1,2}, Wanbo Li^{1,2}, Marc Dive¹, Nico Tamma¹, Charles Michaux², Tom Druet¹, Ivo J. Huijbers³, Clare M. Isacke³, Wouter Coppieters¹, Michel Georges¹, Carole Charlier^{1*}

1 Unit of Animal Genomics, GIGA-R, Department of Animal Sciences, Faculty of Veterinary Medicine, University of Liège, Liège, Belgium, **2** Unit of Bioinformatics, Department of Animal Sciences, Faculty of Veterinary Medicine, University of Liège, Liège, Belgium, **3** Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, United Kingdom

Abstract

We herein describe the positional identification of a 2-bp deletion in the open reading frame of the *MRC2* receptor causing the recessive Crooked Tail Syndrome in cattle. The resulting frame-shift reveals a premature stop codon that causes nonsense-mediated decay of the mutant messenger RNA, and the virtual absence of functional Endo180 protein in affected animals. Cases exhibit skeletal anomalies thought to result from impaired extracellular matrix remodeling during ossification, and as of yet unexplained muscular symptoms. We demonstrate that carrier status is very significantly associated with desired characteristics in the general population, including enhanced muscular development, and that the resulting heterozygote advantage caused a selective sweep which explains the unexpectedly high frequency (25%) of carriers in the Belgian Blue Cattle Breed.

Citation: Fasquelle C, Sartelet A, Li W, Dive M, Tamma N, et al. (2009) Balancing Selection of a Frame-Shift Mutation in the *MRC2* Gene Accounts for the Outbreak of the Crooked Tail Syndrome in Belgian Blue Cattle. PLoS Genet 5(9): e1000666. doi:10.1371/journal.pgen.1000666

Editor: Gregory S. Barsh, Stanford University School of Medicine, United States of America

Received: May 19, 2009; **Accepted:** August 27, 2009; **Published:** September 25, 2009

Copyright: © 2009 Fasquelle et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by grants from the Walloon Ministry of Agriculture (Rilouke), the Belgian Science Policy Organisation (SSTC Genefunc PAI), the University of Liège, and Breakthrough Breast Cancer Research Centre. CC is Chercheur Qualifié of the Fonds National de la Recherche Scientifique. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: carole.charlier@ulg.ac.be

□ Current address: Key Laboratory for Animal Biotechnology of Jiangxi Province and the Ministry of Agriculture of China, NanChang, People's Republic of China.

● These authors contributed equally to this work.

Introduction

The Belgian Blue Cattle breed (BBCB) is notorious for its exceptional muscular development known as “double-muscling”. This extreme phenotype is due in part to an 11-bp loss-of-function deletion in the myostatin gene that has been fixed in the breed (e.g. [1]), as well as to ongoing selection on as of yet unidentified polygenes influencing muscularity. As in other breeds, intense selection has substantially reduced the effective population size. Extensive reliance on artificial insemination (AI), in particular, by allowing popular sires to have thousands of descendants, narrows the genetic basis. The concomitant increase in the rate of inbreeding causes recurrent outbreaks of recessive defects. Inherited defects that have lately afflicted the BBCB include the recently described Congenital Muscular Dystonias (CMD) I and II [2].

As a result of this peculiar demography of domestic animal populations, inherited defects generally involve unique “founder” mutations. Allelic homogeneity greatly facilitates positional identification using identity-by-descent (IBD) mapping, as recently demonstrated using the first generation high density SNP arrays for the bovine [2]. The genes underlying CMD I & II were readily mapped, and the causative mutations in the *ATP2A1* and *SLC6A5* genes identified. The widespread use of the resulting diagnostic

tests allowed immediate and effective control of the corresponding pathologies.

We herein report the positional identification of the mutation causing a novel, recently appeared defect referred to as Crooked Tail Syndrome (CTS). The incidence of CTS has risen very suddenly in the BBCB, and 25% of animals now appear to be CTS carriers. We herein provide strong evidence for exacerbated muscular development of carriers of the CTS mutation, conferring “heterozygote advantage” underlying the selective sweep that raised the causative mutation to alarming proportions.

Results

Crooked Tail Syndrome (CTS) exhibits variable expressivity

We recently established a heredo-surveillance platform operating in close collaboration with field veterinarians to rapidly identify emerging genetic defects. As part of these activities, 105 CTS cases were reported to the platform between November 2006 and November 2007. In addition to the striking deviation of the tail (equally likely to be dextro- or levo-rotatory), detailed clinical examination revealed three symptoms shared by all cases: (i) general growth retardation manifesting itself at approximately one month of age, (ii) abnormal skull shape manifested as a shortened

Author Summary

Livestock are being subject to intense artificial selection aimed at ever-increasing, sometimes extreme, production phenotypes. This is well-illustrated by the exceptional muscular hypertrophy characterizing the “double-muscling” Belgian Blue Cattle Breed (BBCB). We herein identify a loss-of-function mutation of the bovine *MRC2* gene that increases muscle mass in heterozygotes, yet causes skeletal and muscular malformations known as Crooked Tail Syndrome (CTS) in homozygotes. As a result of the “heterozygote advantage”, the *MRC2* c.2904_2905delAG mutation has swept through the BBCB population, resulting in as many as 25% carrier animals and causing a sudden burst of CTS cases. These findings highlight one of the risks associated with pushing domestic animals to their physiological limits by intense artificial selection.

broad head, and (iii) extreme muscular hypertrophy including a conspicuous outgrowth of the gluteus medius anchor. Additional symptoms were observed in a substantial proportion but not all cases: (i) spastic paresis of the hind limbs affecting either the quadriceps only (22%), or quadriceps and gastrocnemius (14%), often associated with straight hocks, (ii) short, straight and extended fore limbs (33%), and (iii) pronounced scoliosis with asymmetric development of the muscles of the back (20%). Figure 1 illustrates the corresponding symptomatology. We performed

complete necropsy of a few selected cases but detected no additional obvious abnormalities. Moreover, radiological examination of crooked tails and scoliotic spines failed to reveal structural defects of the vertebrae (data not shown).

Although the defect is not lethal by itself, the most severe cases (~25%) were euthanized on welfare grounds. The surviving ~75% nevertheless caused important economic losses to their owners as a result of growth retardation and carcass depreciation.

CTS is caused by a fully penetrant, two base pair deletion in the open reading frame (ORF) of the *MRC2* gene

We previously mapped the CTS locus to bovine chromosome 19, in a 2.4 Mb interval shared homozygous-by-descent by the eight analyzed CTS cases [2]. To refine the map location of the CTS locus we genotyped the 105 reported CTS cases for five SNPs covering the 2.4 Mb interval (Figure 2A). The SNPs were selected on the basis of the low population frequency of the disease-associated allele. Genotyping was achieved by first sequencing 35 pools of three animals, followed by individual sequencing of the pools revealing the presence of the major allele and therefore of one or more recombinants. This approach allowed us to confine the critical region to the 812 Kb rs29010018 - AAFC03034831 interval. It comprises seven annotated genes which were ranked on the basis of their perceived relevance with regard to the CTS condition. Coding exons were sequenced in an affected and a matched healthy control individual.



Figure 1. Clinical spectrum exhibited by CTS cases. Crooked tail, growth retardation, stocky head, extreme muscular hypertrophy, spastic paresis of the hind limbs, straight hock, scoliosis.
doi:10.1371/journal.pgen.1000666.g001

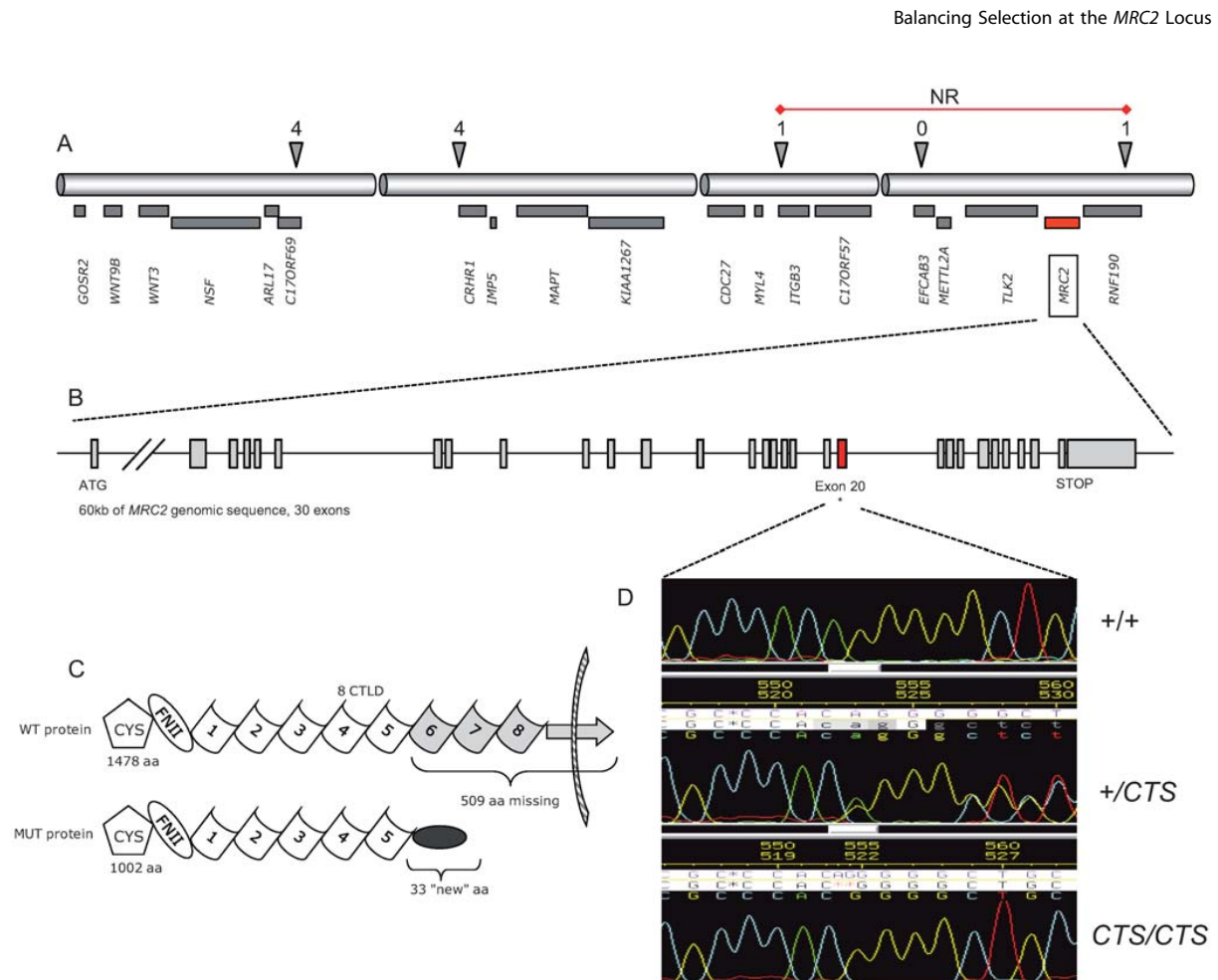


Figure 2. Positional identification of the *c.2904_2905delAG* *MRC2* mutation causing CTS, and its effect of the Endo180 protein. (A) Gene content of the 2.4 Mb interval in which the CTS mutation was located using identity-by-descent mapping. The triangles correspond to five SNPs used to refine the CTS locus position, with corresponding number of recombinant individuals out of 105 CTS cases. The resulting non-recombinant (NR) interval is marked by the red horizontal line. (B) Structure of the *MRC2* gene within that interval. (C) Domain composition of the wild-type (WT) and mutant (MUT) Endo180 protein. (D) Sequences traced obtained from genomic DNA of a homozygous wild-type, carrier and homozygous CTS animal showing the deletion of the ApG dinucleotide in the mutant. doi:10.1371/journal.pgen.1000666.g002

During this process, we identified a 2-bp deletion in the ORF of the mannose receptor C type 2 (*MRC2*) gene. The *MRC2* gene encodes the 180 kDa Endocytic Transmembrane Glycoprotein (Endo180), one of the four members of the mannose receptor family [3,4]. Endo180 is a recycling endocytic receptor that is predominantly expressed in mesenchymal cells such as stromal fibroblasts and in the chondrocytes and osteoblasts/osteocytes in the developing bones, and is proposed to play a role in regulating extracellular matrix degradation and remodelling. It has C-type lectin activity, binds collagen and interacts with urokinase-type plasminogen activator receptor (uPAR) in a trimolecular cell surface complex with pro-urokinase plasminogen activator (pro-PA). The 180 kD Endo180 protein comprises an aminoterminal cysteine-rich domain of unknown function, a fibronectin type II domain which mediates collagen binding, eight C-type lectin-like domains (CTLDs) of which the second mediates Ca^{2+} -dependent lectin activity, a stop-transfer signal anchoring this single-pass transmembrane protein in the membrane, and a carboxyterminal cytoplasmic domain allowing association with adaptor proteins in the clathrin coat. The mutation identified in CTS cases is located

in exon 20 and deletes nucleotides 2904 and 2905 of the *MRC2* cDNA (*c.2904_2905delAG*). It is predicted to append a frame-shifted 30-residue peptide to a truncated Endo180 receptor missing the CTLD6-8 domains, the stop-transfer signal and the cytoplasmic domain (Figure 2B,C,D). As a result, the mutated protein should be unable to localize to the plasma membrane and mediate receptor-mediated endocytosis.

We developed a 5' exonuclease assay for the mutation and genotyped the 105 reported CTS cases. All proved to be homozygous for the *c.2904_2905delAG* mutation. We then genotyped 1,899 healthy Belgian Blue animals. Unexpectedly, 24.7% of animals appeared to be carriers, without a single homozygous mutant ($p < 10^{-12}$ assuming Hardy-Weinberg equilibrium). Taken together, these results allowed us to incriminate the *c.2904_2905delAG* mutation as being causal and fully penetrant.

Mutant *MRC2* mRNAs are targeted by the nonsense-mediated decay (NMD) RNA surveillance pathway

C.2904_2905delAG causes a frame-shift resulting in a premature stop codon in the 21st of the 30-exon *MRC2* gene. Mutant mRNAs

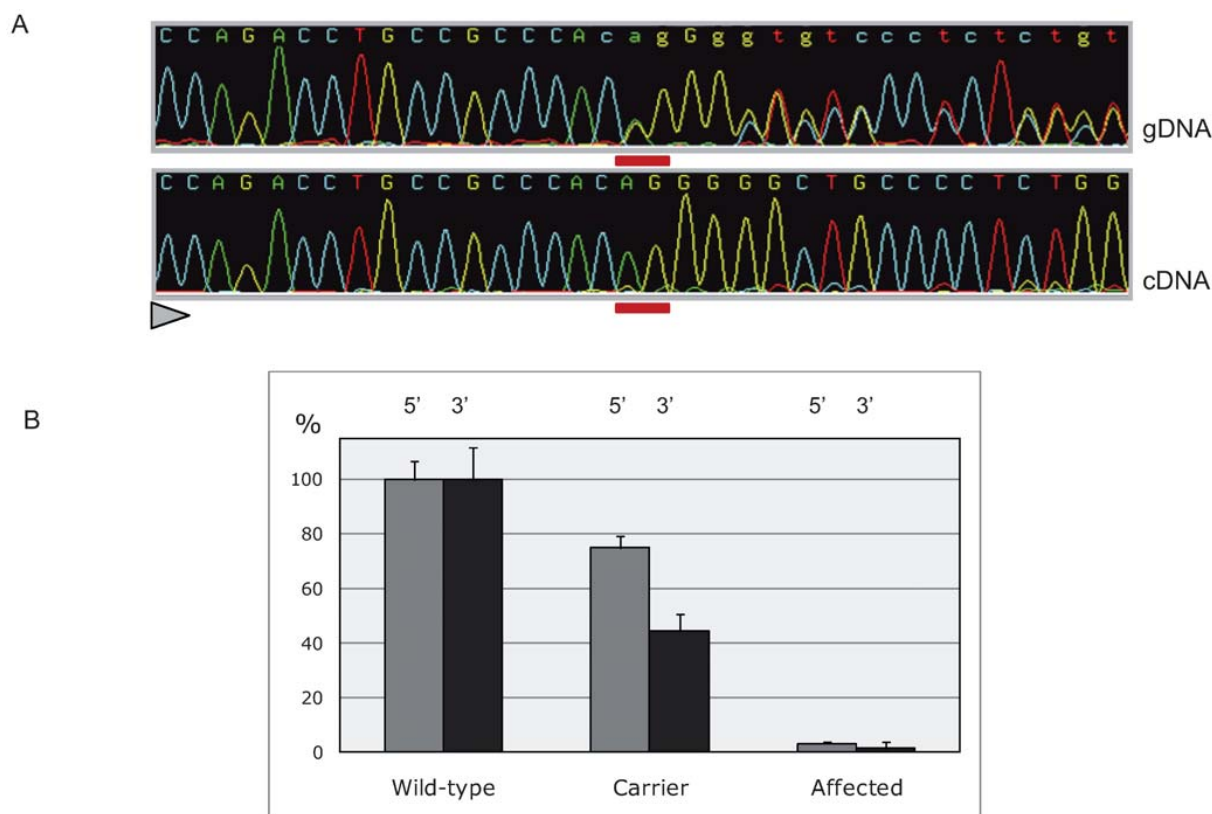


Figure 3. Nonsense-mediated RNA decay of *c.2904_2905delAG* mutant *MRC2* transcripts. (A) Direct sequencing of *MRC2* amplicons spanning the CTS mutation obtained from genomic DNA and pulmonary cDNA of a heterozygous animal, showing the virtually exclusive detection of wild-type allele amongst transcripts (the position of the deleted nucleotides is underlined in red, the sequencing direction is represented by a triangle). (B) Comparing *MRC2* mRNA levels in the lung of +/+, +/CTS and CTS/CTS animals. Data are shown for two amplicons at the 5' and 3' ends of the mRNA, respectively. Error bars correspond to standard errors over three replicates per sample. doi:10.1371/journal.pgen.1000666.g003

are therefore predicted to undergo NMD [5]. To test this, we first compared the levels of wild-type and mutant *MRC2* mRNA in lung and skeletal muscle of a carrier animal by direct sequencing of RT-PCR products encompassing the deletion. As can be seen from Figure 3A, mutant mRNA was barely detectable. We then compared the levels of *MRC2* mRNA in lung tissue of animals of the three genotypes using quantitative RT-PCR performed with primer sets targeting the 5' and 3' end of the mRNA respectively. Highly significant reductions in *MRC2* mRNA levels were observed in carriers relative to homozygous wild-type individuals ($75\% \pm 4\%$ and $45\% \pm 6\%$ of control values for the 5' and 3' systems respectively), while *MRC2* mRNA levels in cases were less than 5% of homozygous wild-types (Figure 3B). Both the allelic imbalance and qRT-PCR experiments thus supported degradation of the mutant transcripts by NMD.

Amounts of full-length Endo180 protein are halved in tissues of carrier animals

From the ten anti-human Endo180 antibodies tested by Western blotting, only one polyclonal rabbit antibody (CAT2) detected the bovine Endo180 protein with sufficient specificity. The CAT2 antibody is directed against the last 19 amino acids of Endo180 of which the last 18 are perfectly conserved between human and cow [6]. CAT2 was thus predicted to allow

recognition of the wild-type but not mutant Endo180. As expected, no wild-type Endo180 was detected in lung tissue of CTS affected animals. In carrier animals, the levels of Endo180 protein were approximately half those observed in homozygous wild-types (Figure 4). Assuming that Endo180 is dosage sensitive,

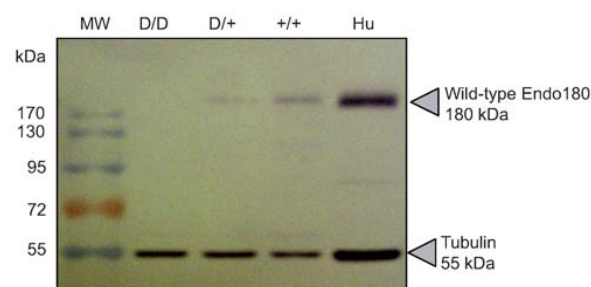


Figure 4. Effect of the CTS mutation on the levels of full-length Endo180. Western blot results from lung of animals of the three genotypes. Hu: human control sample. MW: molecular weight marker. The 55 kDa band corresponds to non-specific binding of the CAT2 antibody to tubulin, used as control for the amount of protein loaded. doi:10.1371/journal.pgen.1000666.g004

Balancing Selection at the *MRC2* Locus

such reduction in the supposedly functional species might affect phenotype.

CTS carrier status increases muscle mass

The unusually high frequency of the CTS mutation in BBCB suggested that it might confer heterozygote advantage in this highly selected population. To test this hypothesis we estimated the effect of carrier genotype on 22 type traits evaluating muscularity, skeletal conformation, size and leg soundness, which are systematically recorded as part of the selection programs implemented in the BBCB. The analysis was conducted on 519 pedigreed bulls, including 148 carrier and 371 homozygous wild-type animals, using a mixed model including fixed effects of *MRC2* genotype, year at scoring, body condition and age at scoring, as well as a random individual animal effect. Variance components and effects were estimated by restricted maximum likelihood (REML) analysis. Highly significant effects were obtained for the four categories of recorded traits (Table 1). CTS carrier animals were smaller, stockier and more heavily muscled. They had a thinner skeleton and more rounded ribs, which are characteristics of beef cattle. *MRC2* genotype accounted for 3.6%, 3.6% and 2.6% for the genetic variance of height, muscularity and general appearance, respectively. These results strongly suggest that CTS carrier frequency is increased by selection programs applied in BBCB.

The *C.2904_2905delAG* mutation is undergoing a selective sweep

To more directly demonstrate the occurrence of a selective sweep, we performed the following analysis. Examination of the

available genealogies of the 105 affected individuals indicated that all of them trace back to Précieux, a popular AI sire, via both sire and dam. This suggested that Précieux, born in 1980, was CTS carrier and that its extensive utilization in the mid eighties spread the CTS mutation in BBCB. Genotyping Précieux and three of his sons for the *C.2904_2905delAG* mutation and the 60K Illumina chip, indeed demonstrated that he carried the CTS mutation embedded in the SNP haplotype shared homozygous-by-descent by the examined cases [2]. Thus, the vast majority of *C.2904_2905delAG* mutations encountered in present-day BBCB animals, trace back to Précieux.

We obtained DNA samples from all BBCB sires (174) born between 2003 and 2005, whose semen had been commercialized by one of the ten major Belgian AI studs. Such AI sires are heavily selected for extreme muscularity. Examination of the pedigrees indicated that 160 of the 174 [2003–2005] AI sires were descendants of Précieux. The number of generations separating these sires from Précieux averaged 5.9 (range: 3 to 8). Genotyping the *C.2904_2905delAG* mutation in this cohort identified 45 CTS carriers, all of them amongst the 160 descendants of Précieux.

Assuming that the CTS mutation indeed underwent a selective sweep, 45 carriers out of the 160 Précieux descendants would be significantly higher than expected by chance alone. To verify this assumption we simulated the segregation of a mutation in the true genealogy of the 160 descendants of Précieux and counted the resulting number of carrier bulls. In these simulations, Précieux was systematically assumed to be carrier, while the frequency of the mutation in animals unrelated to Précieux varied from 0 to 0.05. In the absence of selection (i.e. if a carrier animal is equally likely to transmit either the mutation or the wild-type allele to anyone of its descendants), the probability to obtain 45/160

Table 1. Effect of CTS carrier status on type traits in BBCB.

	Trait or syntetic note	Contrast	Std. error	p value	Carrier characteristics
SIZE	Withers height	2.33	0.320	***	Smaller
	Length	0.36	0.181	N.S.	
	Chest width	−0.68	0.319	*	Larger
	Pelvis width	−0.32	0.188	N.S.	
	Pelvis length	0.25	0.163	N.S.	
MUSC.	Shoulder muscling	−0.58	0.283	*	Increased muscularity
	Top muscling	−1.71	0.409	***	Increased muscularity
	Buttock side	−0.35	0.20	N.S.	
	Buttock rear	−0.50	0.251	*	Increased muscularity
	Synthetic note for muscularity	−0.70	0.246	**	Increased muscularity
SKELETAL CONFORM.	General appearance	−1.73	0.124	***	Better
	Skeleton	−1.18	0.386	**	Thinner
	Rib shape	−1.67	0.456	***	Ronder
	Fore legs stance	0.54	0.157	***	More toed-in
	Rear legs stance	−0.43	0.200	*	More toed-out
OTHER	Skin	−0.34	0.442	N.S.	
	Tail set	−0.24	0.529	N.S.	
	Shoulder bone	0.10	0.108	N.S.	
	Rump	1.25	0.425	**	More horizontal
	Top line	−0.36	0.137	**	More convex
	Hocks stance	0.84	0.367	*	Straighter

* $p < 5\%$, ** $p < 1\%$, *** $p < 0.1\%$.

doi:10.1371/journal.pgen.1000666.t001

carriers was 0.0014, 0.0023 and 0.0130 for mutation frequencies (outside the Précieux lineage) of 0.00, 0.01 and 0.05, respectively (Table S1). Thus we can confidently assert that the *C.2904_2905delAG* mutation indeed underwent a selective sweep in the BBCB.

To have some quantitative assessment of the intensity of the selective sweep, we repeated the “gene dropping” simulations while varying the degree of segregation distortion in favour of the mutant allele. Figure 5 shows the proportion of simulations yielding 45/160 carrier bulls as a function of the transmission probability of the CTS mutation from carrier parents to offspring. It can be seen that the outcome of 45/160 carrier bulls is most likely for a transmission rate between 0.62:0.38 (mutation frequency outside Précieux lineage of 0.05) and 0.67:0.33 (mutation frequency outside Précieux lineage ≤ 0.01). The fact that all 105 CST cases traced back to Précieux both on the dam and sire side, indicates that the mutation frequency outside of the Précieux lineage is closer to 1% than to 5%. Thus, a carrier animal is approximately two times more likely to be selected than a non-carrier sib.

Discussion

We herein describe a frame-shift mutation in the *MRC2* gene causing the CTS syndrome in cattle. Clinical manifestations of CTS are dominated by skeletal and muscular anomalies. Skeletal symptoms including growth retardation, abnormally shaped legs and skulls, are perfectly compatible with the known involvement of *MRC2* in regulating extracellular matrix degradation and remodeling and its strong expression in developing bone [7].

The muscular symptoms, including muscular hypertrophy, tail deviation and spastic paresis are more difficult to rationalize, although a role for the related mannose receptor in myoblast motility and muscle growth has been recently reported [8]. We cannot formally exclude the possibility that the muscular manifestations result from distinct sequence variants in linkage disequilibrium with the CTS mutation, although we favor the more parsimonious hypothesis of a single causative mutation.

It is noteworthy that mice homozygous for a targeted deletion of *MRC2* exons 2 to 6 have been generated in two independent laboratories [9,10]. Both laboratories reported that the mice were viable and fertile, although more recently a minor deficiency in long bone growth, bone mineral density and calvarial bone formation has been demonstrated [7]. Cells derived from these animals show a clear defect in collagen uptake and degradation. One reason for the more pronounced clinical manifestations in cattle than in mice may lie in the distinct nature of the murine and CTS *MRC2* mutations. Cells isolated from the genetically modified mice express a mRNA species in which exon 1 (containing the signal sequence) is spliced in frame onto exon 7 (containing CTLD2), and in embryonic fibroblasts a truncated Endo180 protein missing the cysteine-rich, FNII and CTLD1 domains can be expressed [9]. However, little or no truncated protein is found in postnatal tissues from these knockout mice [10,11]. Alternatively it may be that there are distinct degrees of redundancy between members of the mannose receptor family in different species. Also the more striking phenotype in cattle may be due the different genetic background and particularly the fact that the studied animals were all homozygous for a *MSTN* loss-of-function mutation [1]. This hypothesis could be tested by mating

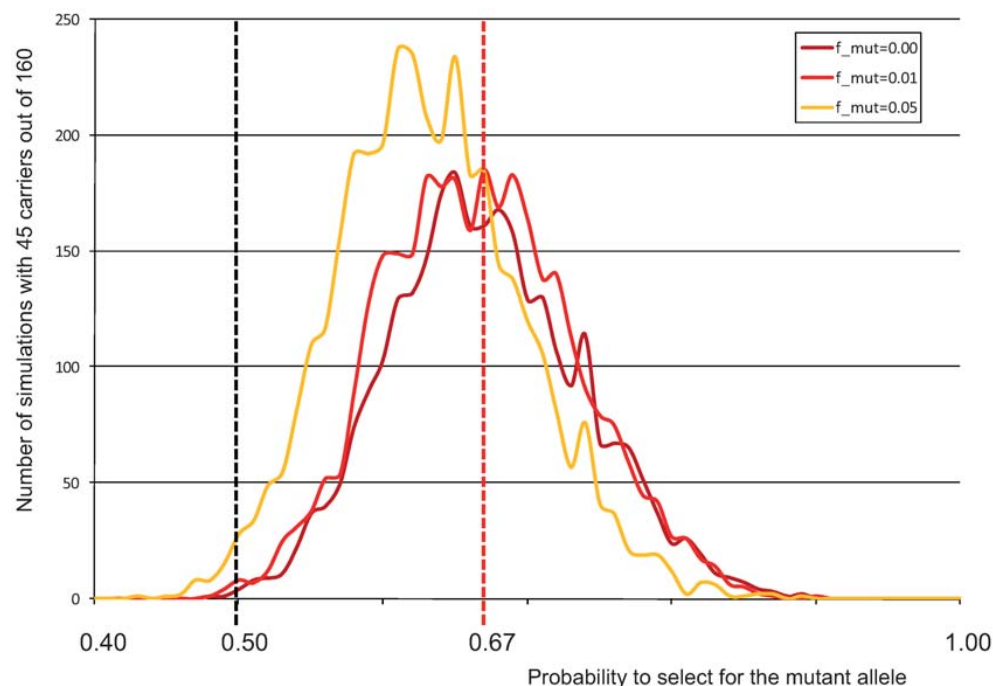


Figure 5. Distribution of the number of simulations (out of 10,000) yielding 45 carriers out of 160 Précieux descendants (Y-axis), as a function of the rate of transmission of the mutation from heterozygous carriers (X-axis). Three curves are given corresponding to frequencies of the mutation outside of the Précieux lineage of 0, 1, and 5%. The dotted red vertical line corresponds to a transmission rate of 67%, maximizing the number of simulations yielding 45 carriers for a mutation frequency (outside of the Précieux lineage) of 1%, considered to be an upper bound in BBCB.
doi:10.1371/journal.pgen.1000666.g005

the available *MRC2* and *MSTN* knock-out mice. Whatever the reason, at this point bovine CTS may be the more informative model to decipher *MRC2* function, and to assist in the identification of as of yet unidentified human pathological conditions resulting from *MRC2* loss-of-function.

We provide very strong evidence of phenotypic manifestations of the CTS mutation in carriers. This is more than likely reflecting dosage sensitivity for Endo180, as NMD causes the mutant protein to be present at near undetectable levels, thus very unlikely to affect cellular function *per se*. Enhanced muscularity of CTS carriers has supposedly contributed greatly to the rapid increase of the CTS mutation in the BBCB. Indeed, we demonstrate that carrier animals have approximately two times more chance to be selected as elite sires than their non-carrier sibs. Note that this is the level of segregation distortion expected for a gene that accounts for ~5% of the genetic variance for a trait with heritability of ~25% and assuming a selection intensity of ~2% (Table S2).

Such selective sweep is reminiscent of the spread of other inherited defects in domestic animals as a result of advantageous traits exhibited by carriers. These include loss-of-function mutations of the porcine ryanodine receptor and equine skeletal muscle sodium channel alpha subunit gene causing, respectively, malignant hyperthermia and hyperkalaemic periodic paralysis in homozygotes, yet increased muscle mass in heterozygotes [12,13], or of a *FGFR3* mutation causing hereditary chondrodysplasia in homozygous sheep and increased size in the carriers [14,15].

A diagnostic test for the CTS mutation has been developed and already applied on more than 4,000 BBCB samples. The resulting information should have an immediate and positive impact on the incidence of CTS, and protect animals and breeders against the pathological condition and ensuing economic losses.

This work is yet another illustration of the value of domestic animal populations in enriching the phenotype-genotype map. It adds to a recent list of positional cloning successes in poultry [16], dog [17,18], horse [19] and bovine [2].

Materials and Methods

Mutation scanning

Coding exons of positional candidate genes were amplified from genomic DNA of a CTS case and a matched control using standard procedures. The primers used for the *MRC2* gene are listed in Table S3. PCR products were directly sequenced using the Big Dye terminator cycle sequencing kit (Applied Biosystem, Foster City, CA). Electrophoresis of purified sequencing reactions was performed on an ABI PRISM 3730 DNA analyzer (PE Applied Biosystems, Foster City, CA). Multiple sequence traces from affected and wild-type animals were aligned and compared using the Phred/Phrap/Consed package (www.genome.washington.edu).

5' exonuclease diagnostic assay of the CTS mutation

A Taqman assay was developed to genotype the CTS mutation, using 5'-GCG CAA CAG CAC CAG AGA-3' and 5'-CTC CCT ACC TTG TTC AGG AAC TG-3' as PCR primers, and 5'-CTG CCG CCC AC[*] GGG-3' (CTS) and 5'-CTG CCG CCC AC[A G]G-3' (wild type) as Taqman probes. Reactions were carried out on a ABI7900HT instrument (Applied Biosystems, Foster City, CA) using standard procedures.

Allelic imbalance test of NMD

Total RNA was extracted from lung, heart and skeletal muscle of a two month old heterozygote *c.2904_2905delAG* animal using Trizol (Invitrogen). The RNA was treated with TurboDNase

(Ambion). cDNA was synthesized using *SuperscriptTMIII* First Strand Synthesis System for RT-PCR (Invitrogen). A portion of *MRC2* cDNA, encompassing the deletion, was amplified using *MRC2* specific primers (Table S4). The PCR products were directly sequenced as described above.

Real-time quantitative RT-PCR test of NMD

Total RNA from lung and skeletal muscle was obtained from animals of the three genotypes (+/+, +/- and CTS/CTS). After DNase-treatment (Turbo DNA-free, Ambion), 500 ng total RNA was reverse transcribed in a final volume of 20 µl using the iScript cDNA Synthesis Kit (Bio-Rad). PCR reactions were performed in a final volume of 15 µl containing 2 µl of 2.5-fold diluted cDNA (corresponding to 20 ng of starting total RNA), 7.5 µl of 2× master mix prepared from the qPCR Core Kit for SYBR green I (Eurogentec), 0.45 µl of 1/2000 SYBR green I working solution prepared from the qPCR Core Kit for SYBR green I (Eurogentec), forward and reverse primers (250 nM each) and nuclease free water. PCRs were performed on an ABI7900HT instrument (Applied Biosystems, Foster City, CA) under the following cycling conditions: 10 min at 95°C followed by 40 cycles at 95°C for 15 sec and 60°C for 1 min. Two primer sets were used to test *MRC2* expression (MRC2_5'QRT_UP/DN and MRC2_3'QRT_UP/DN) and seven genes were included as candidate endogenous controls: (1) Beta Actin (*ACTB*), (2) Glyceraldehyde-3-phosphate Dehydrogenase (*GAPD*), (3) Hypoxanthine Phosphoribosyltransferase 1 (*HPRT1*), (4) Ribosomal Protein Large P0 (*RPLP0*), (5) Ribosomal Protein S18 (*RPS18*), (6) Succinate Dehydrogenase Complex Subunit A Flavoprotein (*SDHA*), and (7) Tyr-3- & Trp-5-Monooxygenase Activation Protein Beta (*1WHAB*). After analysis of the results with geNorm [20], the four following genes were selected as best endogenous controls: *ACTB*, *RPLP0*, *RPS18* and *1WHAB*. The corresponding primer sequences are given in Table S4. All sample/gene combinations were analyzed in triplicate. Relative *MRC2* expression levels, for the 5' & 3' cDNA parts, in the samples of the three genotypes were computed using the qBase software package (<http://medgen.ugent.be/qbase/>) (Hellemans et al., 2007).

Western blotting

A series of available antibodies directed against the human Endo180 were tested by Western blotting for cross reactivity with bovine Endo180 on commercial bovine aortic endothelial cells (BAOEC, Cell Applications). A positive control corresponding to a lysate of MRC5 human fibroblast cell line expressing Endo180 was included in each experiment. The tested antibodies were the following: (i) seven mouse monoclonal antibodies (for details see [21–23]), (ii) a rabbit polyclonal antibody (DEX) directed against the full length human protein [21] and (iii) two rabbit polyclonal antibodies (CAT1 and CAT2) against a peptide from the human C-terminal cytoplasmic domain (CATEKNILVSDMEMNEQ-QE) conjugated to KLH [6]. After initial testing, only the CAT2 antibody was retained for further experiments. Flash-frozen skeletal muscle and lung tissues from animals of the three *MRC2* genotypes (see above) were disrupted and homogenized with a tissue lyser system II (Quiagen). Crude protein extracts were obtained and total protein concentrations determined using a colorimetric test (Pierce BCA Protein Assay kit, Thermo Scientific). Fifteen µg were diluted in 15 µl final volume (1× SDS gel-loading buffer) and loaded on a 5% stacking – 10% resolving Tris-glycine SDS-Polyacrylamide gel. Proteins were separated by electrophoresis at 120 V-250 mA during 3 hours, visualized by Coomassie blue staining, and electro-transferred overnight to Hybond P PVDF membranes (GE Healthcare). Membranes were blocked with 5% skim milk in PBS-Tween 20

Balancing Selection at the *MRC2* Locus

(PBS-T) followed by incubation with primary CAT2 antibodies (1:200) in a total volume of 3 ml for 1 h30 min. After washing, the specific signal was detected by using Alkaline Phosphatase conjugated secondary rabbit antibodies (Sigma) following the instructions of the manufacturer.

Statistical analysis

Phenotypes corresponded to 22 type traits related to muscularity, skeletal conformation, size and leg soundness that are systematically recorded in the BBCB [24]. These were analyzed using a mixed model including genotype at the *MRC2* locus (2), year at scoring (2), body condition (4) as fixed effects, age at scoring as covariate (quadratic regression), the additive genetic animal effect and the residual effect as random effect [25]. The number of animals in the relationship matrix was 6,356. Variance components were estimated using the DFREML method (Derivative-Free Restricted Maximum Likelihood) [26]. The part of the genetic variance due to *MRC2* genotype was estimated as the difference between the variance due to the animal model with and without *MRC2* genotype in the model. The allele substitution effects (contrast) were calculated as the difference between the genotypic means (+/+ and +/M) obtained from the mixed model equations.

Evidencing a selective sweep

We simulated the segregation of a heterozygous mutation from Précieux to its 160 [2003–2005] sire offspring. Variable parameter values were (i) the transmission rate of the mutation from carriers to their offspring (0.5 to 0.75), (ii) the frequency of the mutation outside of the Précieux lineage. 10,000 simulations were conducted for each set of parameter values. Only non-affected genotypes were sampled from matings between heterozygous parents.

Supporting Information

Table S1 Statistics of number of carriers under the neutral model (no selection)(10,000 simulations).
Found at: doi:10.1371/journal.pgen.1000666.s001 (0.07 MB PDF)

Table S2 The table shows, for varying values of δ , the proportion of the phenotypic (P-PV) and genetic variance (P-GV) explained by the QTN in the general population. Assume a normally disturbed trait with 25% heritability, influenced by a QTN with MAF 1 of 0.25 and with two possible genotypes in the population (+/+ and +/M) as is the case for the CTS mutation.

References

- Grobet L, Martin IJ, Poncelet D, Pirottin D, Brouwers B, et al. (1997) A deletion in the bovine myostatin gene causes the double-muscling phenotype in cattle. *Nat Genet* 17: 71–74.
- Charlier C, Coppeters W, Rollin F, Desmecht D, Agerholm JS, et al. (2008) Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nat Genet* 40: 449–454.
- East L, Isacke CM (2002) The mannose receptor family. *Biochim Biophys Acta* 1572: 364–386.
- Behrendt N (2004) The urokinase receptor (uPAR) and the uPAR-associated protein (uPARAP/Endo180): membrane proteins engaged in matrix turnover during tissue remodeling. *Biol Chem* 385: 103–136.
- Chang YF, Imam JS, Wilkinson MF (2007) The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem* 76: 51–74.
- Sturge J, Todd SK, Kogianni G, McCarthy A, Isacke CM (2007) Mannose receptor regulation of macrophage cell migration. *J Leukoc Biol* 82: 585–593.
- Wagenaar-Miller RA, Engelholm LH, Gavard J, Yamada SS, Gutkind JS, et al. (2007) Complementary roles of intracellular and pericellular collagen degradation pathways in vivo. *Mol Cell Biol* 27: 6309–6322.
- Jansen KM, Pavlath GK (2006) Mannose receptor regulates myoblast motility and muscle growth. *J Cell Biol* 174: 403–413.
- East L, McCarthy A, Wienke D, Sturge J, Ashworth A, et al. (2003) A targeted deletion in the endocytic receptor gene Endo180 results in a defect in collagen uptake. *EMBO Rep* 4: 710–716.

Assume that the average phenotype of the +/+ population is $-\delta/2$ and of the +/M population is $+\delta/2$. Assume also that the residual variance is 1. The table shows, for varying values of δ , the proportion of the phenotypic (P-PV) and genetic variance (P-GV) explained by the QTN in the general population. Assume that one selects future AI sires amongst offspring of popular +/M heterozygous sires. The table shows, for five hypothetical phenotypic threshold values for selection $T = 1.00$ – 2.00 , the proportion of sons selected (Prop-Sel), and amongst the selected sons, the ratio of carrier (+/M) versus non-carriers (+/+) (C/NC). Dams were assumed to be +/+ for simplicity. It can be seen that the observed $\sim 2:1$ segregation ratio observed for CTS implies a selection intensity of the order of 0.02 for a QTN that accounts for ~ 0.05 of the genetic variance in the general population. The corresponding cells are highlighted in gray.

Found at: doi:10.1371/journal.pgen.1000666.s002 (0.11 MB PDF)

Table S3 Primer pairs for the *MRC2* gene.

Found at: doi:10.1371/journal.pgen.1000666.s003 (0.08 MB PDF)

Table S4 Allelic imbalance and quantitative RT-PCR primer pairs for the detection of NMD.

Found at: doi:10.1371/journal.pgen.1000666.s004 (0.08 MB PDF)

Acknowledgments

We are grateful to Dimitri Pirottin and Mallory Draye for their assistance with the quantitative RT-PCR and Western blotting, respectively; to David Robertson for his assistance in characterizing the Endo180 antibodies; and to the Walloon Breeding Association (AWE) and the Belgian Blue Beef Herd-Book (HBBBB) for pedigree information and phenotypic data on the 22 type traits.

Author Contributions

Conceived and designed the experiments: MG CC. Performed the experiments: CF AS WL NT IJH CC. Analyzed the data: CF AS WL CM TD IJH CMI MG CC. Contributed reagents/materials/analysis tools: AS MD IJH CMI WC. Wrote the paper: MG CC. Mutation screening, NMD: CF. Case collection, necropsy, phenotypic description, pedigree analysis: AS. Mutation screening: WL. Case collection: MD. Diagnostic test: NT. Animal model: CM. Selective sweep: TD. Western blotting and immunohistochemistry: IJH, CMI. Custom iSelect illumina panel: WC. Designed experiments, analyzed data, wrote the manuscript, and supervised the project: CC, MG.

Balancing Selection at the *MRC2* Locus

18. Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NH, Zody MC, et al. (2007) Efficient mapping of mendelian traits in dogs through genome-wide association. *Nat Genet* 39: 1321–1328.
19. Rosengren Pielberg G, Golovko A, Sundström E, Curik I, Lennartsson J, et al. (2008) A cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse. *Nat Genet* 40: 1004–1009.
20. Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* 8: R19.
21. Isacke CM, van der Geer P, Hunter T, Trowbridge IS (1990) p180, a novel recycling transmembrane glycoprotein with restricted cell type expression. *Mol Cell Biol* 10: 2606–2618.
22. Sheikh H, Yarwood H, Ashworth A, Isacke CM (2000) Endo180, an endocytic recycling glycoprotein related to the macrophage mannose receptor is expressed on fibroblasts, endothelial cells and macrophages and functions as a lectin receptor. *J Cell Sci* 113: 1021–1032.
23. Wienke D, MacFadyen JR, Isacke CM (2003) Identification and characterization of the endocytic transmembrane glycoprotein Endo180 as a novel collagen receptor. *Mol Biol Cell* 14: 3592–3604.
24. Hanset R, Michaux C, Boonen F (1994) Linear classification in the Belgian Blue Cattle Breed: phenotypic and genetic parameters. Ottawa, Canada, International Committee for Animal Recording (ICAR), seminar, Beef performance recording and genetic evaluation. In: Milk and beef recording : State of the art, 1994. Proceedings of the 29th biennial session of the International Committee for Animal Recording (ICAR); Ottawa, Canada; July 31–August 5, 1994. EAAP Publication No. 75. pp 231–237.
25. Lynch M, Walsh B (1997) Genetics and analysis of quantitative traits. Sunderland Massachusetts: Sinauer Associates, Inc.
26. Boldman KG, Kriese LA, Van Vleck LD, Kachman SD (1995) A manual for Use of MTDFREML. A set of Programs to obtain Estimates of Variances and Covariances. United States Department of Agriculture: Agricultural Research Service.



- 1 **Supplemental Table 1:** Statistics of number of carriers under the neutral model (no selection)(10,000
 2 simulations).

f_{mut}	Min	25%	median	mean	75%	max	n>44
0.00	0	2	7	9.54	15	62	14
0.01	0	4	10	11.76	17	58	23
0.05	0	13	19	20.12	26	63	130
0.10	4	22	29	29.41	36	72	688

3

Supplemental table 2: Assume a normally disturbed trait with 25% heritability, influenced by a QTN with MAF of 0.25 and with two possible genotypes in the population (+/+ and +/M) as is the case for the CTS mutation. Assume that the average phenotype of the +/+ population is $-\delta/2$ and of the +/M population is $+\delta/2$. Assume also that the residual variance is 1. The following table shows, for varying values of δ , the proportion of the phenotypic (P-PV) and genetic variance (P-GV) explained by the QTN in the general population. Assume that one selects future AI sires amongst offspring of popular +/M heterozygous sires. The following table shows, for five hypothetical phenotypic threshold values for selection ($T=1,00-2,00$), the proportion of sons selected (Prop-Sel), and amongst the selected sons, the ratio of carrier (+/M) versus non-carriers (+/+) (C/NC). Dams were assumed to be +/+ for simplicity. It can be seen that the observed ~2:1 segregation ratio observed for CTS implies a selection intensity of the order of 0,02 for a QTN that accounts for ~0,05 of the genetic variance in the general population. The corresponding cells are highlighted in gray.

QTN	Population	1/2-sib pedigrees - carrier sire														
		T = 1,00			T = 1,25			T = 1,50			T = 1,75			T = 2,00		
		Prop-Sel	C/NC	Prop-Sel	C/NC	Prop-Sel	C/NC	Prop-Sel	C/NC	Prop-Sel	C/NC	Prop-Sel	C/NC	Prop-Sel	C/NC	
0,00	0,00	0,00	0,00	0,15	1,00	0,10	1,00	0,06	1,00	0,04	1,00	0,02	1,00	0,02	1,00	
0,05	0,00	0,00	0,00	0,15	1,08	0,10	1,10	0,06	1,11	0,04	1,12	0,02	1,13	0,02	1,13	
0,10	0,00	0,01	0,01	0,15	1,17	0,10	1,20	0,06	1,23	0,04	1,26	0,02	1,29	0,02	1,29	
0,15	0,00	0,02	0,02	0,15	1,27	0,10	1,31	0,06	1,36	0,04	1,41	0,02	1,46	0,02	1,46	
0,20	0,01	0,03	0,03	0,15	1,38	0,10	1,44	0,06	1,51	0,04	1,58	0,02	1,65	0,02	1,65	
0,25	0,01	0,05	0,05	0,15	1,49	0,10	1,58	0,06	1,67	0,04	1,77	0,02	1,87	0,02	1,87	
0,30	0,02	0,07	0,07	0,15	1,62	0,10	1,73	0,06	1,85	0,04	1,98	0,02	2,12	0,02	2,12	
0,35	0,02	0,09	0,09	0,15	1,75	0,10	1,89	0,06	2,05	0,04	2,22	0,02	2,41	0,02	2,41	
0,40	0,03	0,12	0,12	0,16	1,90	0,10	2,07	0,06	2,27	0,04	2,49	0,02	2,73	0,02	2,73	
0,45	0,04	0,15	0,15	0,16	2,06	0,10	2,27	0,07	2,51	0,04	2,79	0,02	3,10	0,02	3,10	
0,50	0,04	0,18	0,18	0,16	2,23	0,11	2,49	0,07	2,78	0,04	3,12	0,02	3,51	0,02	3,51	
0,55	0,05	0,21	0,21	0,16	2,42	0,11	2,72	0,07	3,08	0,04	3,50	0,02	3,98	0,02	3,98	
0,60	0,06	0,25	0,25	0,16	2,62	0,11	2,98	0,07	3,41	0,04	3,92	0,02	4,51	0,02	4,51	
0,65	0,07	0,29	0,29	0,16	2,84	0,11	3,27	0,07	3,78	0,04	4,39	0,02	5,12	0,02	5,12	
0,70	0,08	0,34	0,34	0,17	3,08	0,11	3,58	0,07	4,19	0,04	4,92	0,03	5,81	0,03	5,81	
0,75	0,10	0,38	0,38	0,17	3,33	0,11	3,92	0,07	4,64	0,05	5,52	0,03	6,58	0,03	6,58	
0,80	0,11	0,43	0,43	0,17	3,61	0,12	4,30	0,08	5,15	0,05	6,19	0,03	7,47	0,03	7,47	
0,85	0,12	0,48	0,48	0,17	3,92	0,12	4,71	0,08	5,70	0,05	6,93	0,03	8,47	0,03	8,47	
0,90	0,13	0,53	0,53	0,18	4,25	0,12	5,16	0,08	6,32	0,05	7,77	0,03	9,60	0,03	9,60	
0,95	0,14	0,58	0,58	0,18	4,60	0,12	5,66	0,08	7,00	0,05	8,71	0,03	10,89	0,03	10,89	
1,00	0,16	0,63	0,63	0,18	4,99	0,13	6,20	0,09	7,76	0,05	9,77	0,03	12,35	0,03	12,35	

1 **Supplemental table 3:** Primer pairs for the *MRC2* gene

Name	Primer sequence (5'-3')	Gene part	Size (bp)
gUP1	CCGGAGGAAGACGCGAGCCCCT	exon 1 (ATG)	327
gDN1	GGGGGAAAGGAGGAAAAAGTCG		
gUP2	CACAGCCCACTACCAGCGTCAG	exon 2	610
gDN2	CCATGACGATGAAAGAGCTGAC		
gUP3	ACCCTGTGAGAAGCCTTTCCTG	exons 3, 4	907
gDN3	GATGTAGGTCTGCTCGTGGATC		
gUP4	TGTCATGGTGGCAGGTAACGAC	exons 4, 5	576
gDN4	GGGTGGAAATCTGCTGGTCTAG		
gUP5	GGAGGAGGCAAGAGAGCCGAAG	exon 6	366
gDN5	CCTTGTGCTGTGAGGGTGGGTG		
gUP6	AAAGCGTGGTCCCTGTCCAGC	exon 7	813
gDN6	AACGGTAGCACTCCTTGGTGGT		
gUP7	GGCTTGGTGGGAAGAGTGGATCT	exons 8, 9	725
gDN7	GGGGAGGAGGGATTCCGAGAGG		
gUP8	TTGGAGGCATCTGCACAGCTAC	exon 10	407
gDN8	TACCACAGGAGGCTGCGGATTC		
gUP9	CCTGTGCTCAAGCCTGCAGAAA	exon 11	375
gDN9	GCCCTGGAGATAGTTGAAGCTCA		
gUP10	GGTCCCCACTTCCCTGAGCAAG	exon 12	375
gDN10	ATGAAGCCCTAGGTCTCGGTCAT		
gUP11	CAACCCACAGCACATGTCCCT	exon 13	449
gDN11	CTGCTCGGATCATGGCTGGGTC		
gUP12	CCTCGACACCCTGTCCACTGAA	exon 14	398
gDN12	CCTGGCACTAGCAGCAGACACA		
gUP13	CCCGAGTCACAACTGCAAGGATT	exons 15, 16, 17	751
gDN13	GGGTGTGGGATGGACAGGAAGC		
gUP14	CTGCAGCGTGTCTGTCCCTGTT	exon 17, 18	697
gDN14	CGAGTCCCTGCTAGCCATCCAC		
gUP15	GGTCTAACCTGGTGCCCTGTACT	exons 19, 20 (*)	693
gDN15	AGGGGAGAGGGTGGTAGGTTTCAG		
gUP16	GCCAGGTCGGGAGGGTATCAGAG	exons 21, 22, 23	1042
gDN16	CTGCAACCCTGGATGCTCACT		
gUP17	GAGTTGGTCTCTGCCTGCTGTTT	exons 24, 25, 26	995
gDN17	CAGAGTGCAGCACGGGGACTATA		
gUP18	CGTCTCCATGCCATCCTCTATTC	exons 27, 28	743
gDN18	CCCAGGCCTCCCATCCACTGTG		
gUP19	ACTGTATTGTTACTACCACTGTTGTT	exons 29, 30 (STOP)	610
gDN19	AAGGAAACGCCATGCTGCACTC		

2

3

4

- 1 **Supplemental table 4:** Allelic imbalance and quantitative RT-PCR primer pairs for the detection of
- 2 NMD.

Gene	Forward primer	Reverse primer	Size (bp)
MRC2 del	GACAAGAAGTGC GTGTACATGATG	AGAACTGTGCCTCTGACCACTTC	232
MRC2 5' part	CGAGTCTCTCCAGCCTGCAATG	ACTCAGTGCCTCGCGGTCACAC	168
MRC2 3' part	TCCTGCTCCTGGCTCTGCTGAC	CTGCTGCTCGTTCATTTCATG	167
ACTB	TCGCGGACAGGATGCAGAAAGA	GCTGATCCACATCTGCTGGAA	149
GAPDH	TGACCCCTTCATTGACCTTCA	GATGGTGATGGCCTTTCATT	127
HPRT1	TGCTGAGGATTTGGAGAAGG	CAACAGGTCGGCAAAGAACT	154
RPLP0	TGGGCAAGAACACGATGATG	TGAGGTCCTCCTTGGTGAACA	123
RPS18	GCAGAATCCACGCCAATACAA	TCTTCAGGCGCTCCAGGTCTTC	135
SDHA	GCAGAACCTGATGCTTTGTG	CGTAGGAGAGCGTGTGCTT	185
YWHAZ	GCATCCCACAGACTATTTCC	GCAAAGACAATGACAGACCA	120

3

4

Allelic heterogeneity of Crooked Tail Syndrome: result of balancing selection?

Sartelet A.*, Klingbeil P.*, Franklin C.K., Fasquelle C., Géron S.,
Isacke C.M., Georges M., Charlier C.

Animal Genetics, 2012, **43**, 604-607

* Contributed equally



Allelic heterogeneity of Crooked Tail Syndrome: result of balancing selection?

A. Sartelet^{*,1}, P. Klingbeil^{†,1}, C. K. Franklin[†], C. Fasquelle^{*}, S. Géron^{*}, C. M. Isacke[†], M. Georges^{*} and C. Charlier^{*}

^{*}Unit of Animal Genomics, GIGA-R & Department of Animal Sciences, Faculty of Veterinary Medicine, University of Liège, Liège, Belgium. [†]Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, UK

Summary

We report the identification of a second loss-of-function mutation (c.1906T>C) in the bovine *MRC2* gene causing the Crooked Tail Syndrome in Belgian Blue Cattle. We demonstrate that the ensuing substitution of the highly conserved Cysteine 636 with Arginine causes illegitimate receptor oligomerization, which is predicted to impair function of the *MRC2* encoded protein, Endo180. We propose that this second *MRC2* mutation was selected by breeders as a result of its favourable effect on muscularity in heterozygotes.

Keywords Belgian Blue Cattle, Crooked Tail Syndrome, CTS, Endo180, *MRC2*

We recently identified a frame-shift mutation (c.2904-2905delAG) in the *MRC2* gene causing the Crooked Tail Syndrome (CTS) in Belgian Blue Cattle (Fasquelle *et al.* 2009). Unexpectedly, and despite the severity of the condition, we demonstrated that ~25% of Belgian Blue animals are carriers of the c.2904-2905delAG *MRC2* mutation. We demonstrated that this is very likely owing to the fact that the muscularity of c.2904-2905delAG carriers is enhanced, which is a highly desirable trait in Belgian Blue Cattle. Concomitantly, we demonstrated that the c.2904-2905delAG *MRC2* mutation underwent a selective sweep in this breed. Since 2008, 12 000 Belgian Blue animals have been screened for the c.2904-2905delAG mutation and, by selecting against carrier progenitors and/or avoiding at-risk matings, the CTS condition has been largely eliminated from the breed.

Nevertheless, 18 new CTS cases were referred to us between November 2009 and March 2010. All of them had been sired by one of two bulls that had been tested free of the c.2904-2905delAG mutation. The symptomatology was identical to that of previously studied CTS cases, i.e. growth retardation, increased muscular development, tail

deviation, stocky head, short and straight limbs, scoliosis and spastic paresis. The two sires were re-genotyped using DNA extracted from both blood and sperm and were confirmed to be free of c.2904-2905delAG. All 18 calves as well as 12 available dams were shown to be heterozygous for the c.2904-2905delAG mutation (Appendix S1).

We sequenced all 30 *MRC2* exons and intron-exon boundaries in two trios and identified a novel T to C substitution in exon 13 (Fig. 1a,b). The corresponding c.1906T>C mutation is predicted to cause a disruptive substitution (p.Cys636Arg) of one of four extremely conserved Cysteines in the third C-type lectin-like domain (CTLD3), mediating intra-molecular disulphide bonds essential to stabilize the tridimensional structure of the *MRC2*-encoded Endo180 protein (Fig. 1d and Fig. S1) (Zelensky & Gready 2003). We developed a 5' exonuclease assay to genotype the *MRC2* c.1906T>C variant and showed (i) that both sires carried it, (ii) that all their affected offspring were compound c.2904-2905delAG – c.1906T>C heterozygotes, (iii) that the mutation was absent in a breed diversity panel comprising 141 animals representing 11 breeds other than Belgian Blue and (iv) that it had a frequency of ~0.3% (20 carriers in 3188 normal individuals) in Belgian Blue (Appendix S1). RT-PCR amplification of *MRC2* transcripts from skeletal muscle of a compound heterozygous case revealed near-complete absence of c.2904-2905delAG-bearing transcripts, hence confirming their degradation by non-sense mediated RNA decay and demonstrating higher stability of c.1906T>C-bearing transcripts (Fig. 1c).

Address for correspondence

C. Charlier, Unit of Animal Genomics, GIGA-R & Department of Animal Sciences, Faculty of Veterinary Medicine, University of Liège (B34), 1 Avenue de l'Hôpital, 4000-Liège, Belgium.
E-mail: carole.charlier@ulg.ac.be

¹Contributed equally.

Accepted for publication 15 September 2011

2 Sartelet *et al.*

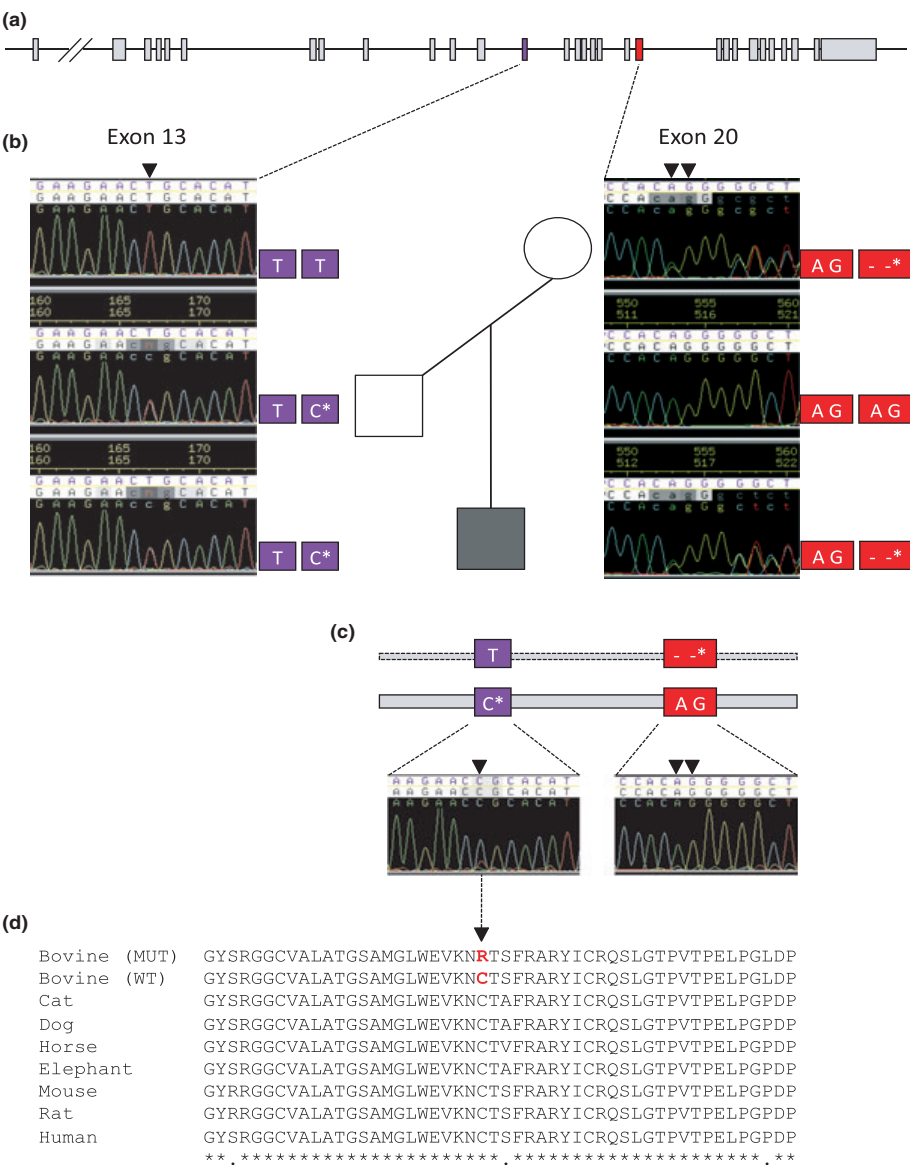


Figure 1 Identification of a missense mutation in exon 13 of the *MRC2* gene. (a) Genomic structure of the *MRC2* gene with the two mutated exons (13 and 20) highlighted in purple and red. (b) Sequence traces obtained from genomic DNA of a trio (mother–father–affected calf) showing the c.1906T>C substitution [left, with respective wild-type (WT) T and mutated C* alleles] and the c.2904-2905delAG frame-shift mutation (right, with respective WT AG and mutated -.* alleles). (c) Corresponding traces from sequencing RT-PCR products of a compound heterozygous affected calf revealing the preferential degradation of the c.2904-2905delAG allele. (d) Alignment of part of the CTLD3 domain around the p.Cys636Arg mutation in different species, illustrating the evolutionary conservation of the mutated Cysteine.

Next we studied the impact of the new mutation on Endo180 protein stability and subcellular localization. We transfected NIH-3T3 mouse fibroblasts with a pcDNA3 vector expressing full-length human Endo180 protein with either the wild-type (WT) Cysteine or the mutant (MUT) Arginine at the corresponding amino acid position [intro-

duced by site-directed mutagenesis (c.1903-1905TGT>CGC)]. Western blot analysis under reducing conditions showed a single expected 180 kDa band expressed at similar levels in WT and MUT cell lysates but not in mock transfected lysates, indicating that the Arginine mutation does not destabilize the Endo180 protein. Interestingly,

Allelic heterogeneity of Crooked Tail Syndrome

3

under non-reducing conditions, in addition to the 180 kDa band, two high molecular bands were clearly visible for the MUT form but absent in the WT (Fig. 2a and Appendix S1). These likely correspond to dimers and possibly oligomers of MUT receptors, formed by inter-molecular disulphide bonds between unpaired CTLD3 Cysteines. Oligomerization was confirmed with a second Endo180 antibody (Fig. 2b).

Immunofluorescent staining of transfected cells showed the expected distribution of WT Endo180 protein to discrete intracellular endosomes present throughout the cytoplasm but concentrated perinuclearly (Sheikh *et al.*

2000; Howard & Isacke 2002). By contrast, the MUT protein showed a more diffuse localization, pointing to a defect in receptor trafficking (Fig. 2c).

Taken together, our data suggest that oligomerization of the Endo180 636R allele compromises its functionality, explaining the MUT phenotype of compound c.2904-2905delAG – c.1906T>C calves.

The frequent bursts of genetic defects in livestock are due primarily to the excessive use of popular sires by artificial insemination. As a consequence, defects are nearly always allelically homogeneous, and it is this feature, typical of genetic isolates, that renders them easy to map in

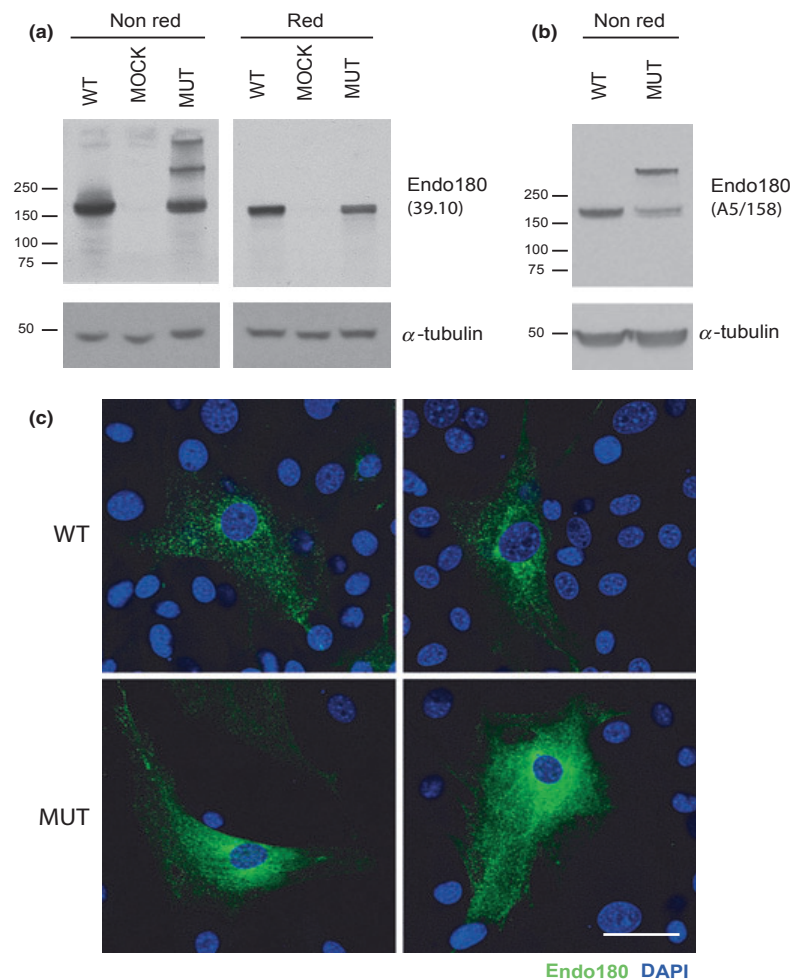


Figure 2 Characterization of the mutant (MUT) Endo180 receptor. (a, b) NIH-3T3 cells were mock transfected or transfected with wild-type (WT) or MUT Endo180 cDNA and whole cell lysates subject to Western blotting under reducing or non-reducing conditions. α -tubulin serves as loading control. (a) Blots were probed with the anti-Endo180 mAb 39.10 showing the presence of additional high molecular weight MUT bands under non-reducing conditions but only a single 180 kDa band under reducing conditions. (b) Equivalent results were obtained with anti-Endo180 mAb A5/158. (c) Immunofluorescent staining of WT and MUT Endo180 protein expressed in transfected NIH-3T3 cells. Scale bar, 50 μ m.

4 Sartelet *et al.*

these populations (Charlier *et al.* 2008). However, if loss-of-function mutations in specific genes confer a selective advantage to carriers, breeders will pick up independent mutations even if initially very rare in the population. It is this phenomenon that underlies the occurrence of an allelic series of loss-of-function mutations in the *MSTN* gene in both cattle and sheep (Georges 2011). We previously showed that functional *MRC2* hemizygosity increases muscle mass. We suspect that the second c.1906T>C *MRC2* mutation was most likely picked up by perceptive breeders as a result of this pleiotropic effect.

Acknowledgements

We are grateful to Anaïs Ullens and to the GIGA Genomic Platform for their technical assistance. CC is Senior Research Associate of the Fonds National de la Recherche Scientifique (Belgium). This work was funded by grants from the Walloon Ministry of Agriculture (Rilouke), the Belgian Science Policy Organisation (SSTC Genefunc PAI), the University of Liege and Breakthrough Breast Cancer Research Centre.

References

- Charlier C., Coppieters W., Rollin F. *et al.* (2008) Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nature Genetics* **40**, 449–54.
- Fasquelle C., Sartelet A., Li W. *et al.* (2009) Balancing selection of a frame-shift mutation in the *MRC2* gene accounts for the outbreak of the Crooked Tail Syndrome in Belgian Blue Cattle. *PLoS Genetics* **5**, e1000666.
- Georges M. (2011) When less means more: impact of the discovery of myostatin in animal breeding. *Immunology Endocrine & Metabolic Agents in Medicinal Chemistry* **10**, 240–8.
- Howard M.J. & Isacke C.M. (2002) The C-type lectin receptor Endo180 displays internalization and recycling properties distinct from other members of the mannose receptor family. *Journal of Biological Chemistry* **277**, 32320–31.
- Sheikh H., Yarwood H., Ashworth A. & Isacke C.M. (2000) Endo180, an endocytic recycling glycoprotein related to the macrophage mannose receptor is expressed on fibroblasts, endothelial cells and macrophages and functions as a lectin receptor. *Journal of Cell Science* **113**, 1021–32.
- Zelensky A.N. & Gready J.E. (2003) Comparative analysis of structural properties of the C-type-lectin-like domain (CTLD). *Proteins* **52**, 466–77.

Supporting information

Additional supporting information may be found in the online version of this article.

Figure S1 Position specific scoring matrix (PSSM) analysis of the mutant C636R CTLD3 domain.

Appendix S1 Supplemental material and methods.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

Supplemental material and methods

5' exonuclease diagnostic assay for the MRC2 c.2904-2905delAG mutation

A Taqman assay was developed to genotype the MRC2 c.2904-2905delAG mutation, using 5'-GCG CAA CAG CAC CAG AGA-3' and 5'-CTC CCT ACC TTG TTC AGG AAC TG-3' as PCR primers, and 5'-CTG CCG CCC AC[A G]G-3' (WT) and 5'-CTG CCG CCC AC[* *] GG-3' (MUT) as Taqman probes (Fasquelle *et al.* 2009). Reactions were carried out on an ABI7900HT instrument (Applied Biosystems, Foster City, CA) using standard procedures.

MRC2 mutation scanning

Coding exons and intron-exon boundaries of the MRC2 gene were amplified from genomic DNA of a trio (CTS calf – mother – father) using standard procedures. Corresponding primer pairs were listed in Fasquelle *et al.* (2009). PCR products were directly sequenced using the Big Dye terminator cycle sequencing kit (Applied Biosystems Foster City, CA). Electrophoresis of purified sequencing reactions was performed on an ABI PRISM 3730 DNA analyzer (PE Applied Biosystems, Foster City, CA). Sequence traces from the trio were aligned and compared using the Phred/Phrap/Consed package (www.genome.washington.edu).

5' exonuclease diagnostic assay of the MRC2 c.1906T>C mutation

We developed a 5' exonuclease assay for the c.1906T>C mutation, using 5'-ACA GGC AGT GCC ATG GG-3' and 3'-GCT CTG GCG GCA GAT GTA G-5' as PCR primers, and 5'-CGA TGT GC[A] GTT CT-3' (WT) and 5'-CGA TGT GC[G] TTC T-3' (MUT) as Taqman probes. Reactions were carried out on an ABI7900HT instrument (Applied Biosystems, Foster City, CA) using standard procedures.

Nonsense mediated RNA decay and allelic imbalance

Total RNA from a compound heterozygous mutant animal was extracted from skeletal muscle using standard procedures (Trizol, Invitrogen). After *DNase*-treatment (Turbo DNA-free, Ambion), cDNA was synthesized using SuperScript III First-Strand Synthesis SuperMix (Invitrogen). cDNA regions encompassing the 2 bp deletion (*c.2904-2905delAG*) and the missense mutation (*c.1906T>C*), were amplified using two *MRC2* specific primer pairs (Supplemental table1). The RT-PCR products were directly sequenced and visualized as described above.

Endo180 mutagenesis and expression

Using a PCR-based approach, the human *MRC2* full length cDNA was mutated at codon position 635 (equivalent to 636 in bovine) *tgt*(Cys) to *cgc*(Arg) by site directed mutagenesis using the primer 5'-AGGTGAAGAACCGCACCTCGTTC-3' and cloned into the pCDNA3 expression vector. Clones were verified by sequencing. NIH-3T3 cells cultured in DMEM plus 10% donor bovine serum were transfected with WT or MUT Endo180 cDNA using Lipofectamine 2000 reagent (Invitrogen). For Western blotting, cells were lysed 2 days post-transfection and total cell lysates subject to SDS-PAGE under non-reducing and reducing conditions. Blots were probed with the human Endo180 specific mAbs 39.10 or A5/158 (Wienke *et al.* 2007; Sheikh *et al.* 2000). mAb 39.10 detects Endo180 under reducing and non-reducing conditions. mAb A5/158 only detects protein under non-reducing conditions. Anti α -tubulin antibody (Sigma) was used as a loading control.

For confocal microscopy, transfected cells cultured on coverslips overnight were fixed in 4% paraformaldehyde, permeabilized with Triton X-100 (Sigma) and stained with anti-Endo180 mAb A5/158 followed by Alexa488-anti-mouse-IgG (Invitrogen). Nuclei were stained with DAPI. Fluorescent images were captured on a Leica Microsystems TCS-SP2 confocal microscope.

Supplemental table 1 :

Primer sequences used to amplify two *MRC2* cDNA regions encompassing the *c.2904-2905delAG* and *c.1906T>C* *MRC2* mutations.

Primer name	Sequence 5'-3'	Product size
MRC2_del2_UP	CGCAACAGCACAGAGAGCAGCAG	145 bp
MRC2_del2_DN	CTAAGGGGTTTGCAATGGTGAC	
MRC2_T>C_UP	CCCTGCAGGACCTCAACTTCAC	319 bp
MRC2_T>C_DN	TCCGAGCTGAACACCTTATAG	

Supplementary figure 1: Position Specific Scoring Matrix (PSSM) analysis of the mutant C636R CTLD3 domain.

A. Schematic representation of the bovine Endo180 receptor displaying the different functional domains, with the mutated C-type lectin-like domain (CTLD) in grey. **B.** Amino acid sequence of the corresponding bovine CTLD3 with the nine most conserved amino acids underlined and the four key cysteines in red; disulfide bonds formed by the Cys (C1-C4 and C2-C3) represented by horizontal lines. **C.** Position-specific scoring matrix (PSSM) output for the CTLD3 domain encompassing C636R, with an alignment stack of 852 sequences for the nine most conserved amino-acids in the CTLD3 (cd00037); a highly negative score (dark red) is obtained when the mutant 636R is substituted for the wild-type C636 (Marchler-Bauer *et al.* 2011).

Supplemental References

Fasquelle C., Sartelet A., Li W., Dive M., Tamma N., Michaux C., Druet T., Huijbers I.J., Isacke C.M., Coppieters W., Georges M. & Charlier C. (2009) Balancing selection of a frame-shift mutation in the

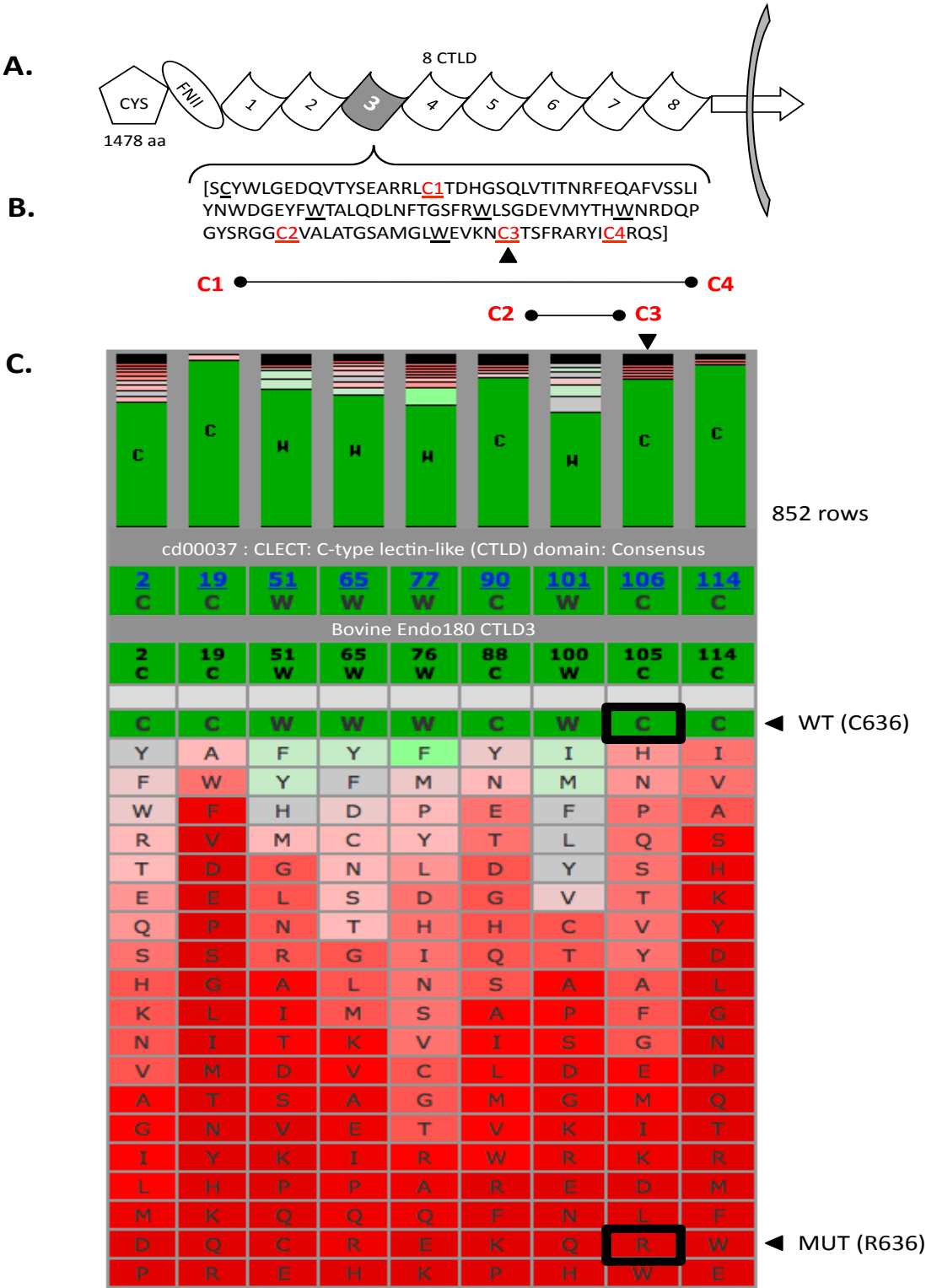
MRC2 gene accounts for the outbreak of the Crooked Tail Syndrome in Belgian Blue Cattle. *PLoS Genetics* 5, e1000666.

Marchler-Bauer A., Lu S., Anderson J.B., Chitsaz F., Derbyshire M.K., DeWeese-Scott C., Fong J.H., Geer L.Y., Geer R.C., Gonzales N.R., Gwadz M., Hurwitz D.I., Jackson J.D., Ke Z., Lanczycki C.J., Lu F., Marchler G.H., Mullokandov M., Omelchenko M.V., Robertson C.L., Song J.S., Thanki N., Yamashita R.A., Zhang D., Zhang N., Zheng C. & Bryant S.H. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Research* 39, D225-9.

Sheikh H., Yarwood H., Ashworth A., & Isacke C.M. (2000) Endo180, an endocytic recycling glycoprotein related to the macrophage mannose receptor is expressed on fibroblasts, endothelial cells and macrophages and functions as a lectin receptor. *Journal of Cell Science* 113, 1021-32.

Wienke D., Davies G.C., Johnson D.A., Sturge J., Lambros M.B., Savage K., Elsheikh S.E., Green A.R., Ellis I.O., Robertson D., Reis-Filho J.S. & Isacke C.M. (2007) The collagen receptor Endo180 (CD280) is expressed on basal-like breast tumor cells and promotes tumor growth in vivo. *Cancer Research* 67, 10230-40.

Sartelet et al., Fig S1



3- EFFET PLEÏOTROPE DU GENE *RNF11* BOVIN SUR LA CROISSANCE ET LA RESISTANCE AUX MALADIES.

A. Résumé



La croissance est un caractère important chez les animaux de production, surtout dans les races viandeuses. C'est un caractère dit complexe qui fait l'objet de nombreuses recherches afin de déterminer des QTL ou QTN l'affectant (*ex. Karim et al., 2011*). Au cours de ces dernières années, un nombre croissant de veaux présentant un retard de croissance important a été recensé par la cellule d'hérédo-surveillance.

Entre 2008 et 2011, 147 individus de race BBB présentant un retard de croissance significatif d'environ 15 % sans modification des proportions, ont été recensés par la cellule. Une étude d'association génome entier réalisée sur 33 cas génotypés avec des puces de moyenne densité (50.000 SNPs) a mis en évidence un locus associé à ce retard de croissance hautement significatif sur le chromosome 3. Une région homozygote IBD de 3,3 Mb partagée par 14 (≈ 40 %) des 33 cas est identifiée à l'examen direct des génotypes des atteints. Le séquençage systématique des régions codantes des gènes compris dans l'intervalle a identifié une substitution au niveau du site accepteur d'épissage du premier intron (*c.124-2A>G*) du gène codant pour la RING finger protein 11 (*RNF11*). L'effet de cette substitution sur le transcrit, évalué par RT-PCR, a démontré la coexistence de deux ARNm, une forme dite majeure attendue avec la non inclusion de l'exon 2 et une forme dite mineure inattendue, présentant une délétion de 7 pb produite par l'activation d'un site cryptique d'épissage dans l'exon 2. Dans les deux formes, un décalage de la phase de lecture et l'apparition d'un codon stop prématuré dans l'exon 3 et l'exon 2, respectivement pour la forme majeure et la forme mineure, sont prédits avec la production d'une protéine tronquée à 75 %. La dégradation de l'ARNm mineur par le NMD a été confirmée par qRT-PCR.

Un test diagnostique a été développé et a estimé la fréquence de la mutation dans la population BBB à 13 %. Cependant, la fréquence élevée de la mutation est incompatible avec le faible nombre de cas de retard de croissance. Une sélection naturelle contre les mutants peut dès lors être suspectée. Une étude rétrospective, évaluant l'effet de l'utilisation de taureaux porteurs sur le taux de non-retour, la mortalité, la morbidité et la réforme précoce des veaux, n'a pas vérifié cette contre-sélection. Une étude prospective à partir d'une centaine d'accouplements porteur-porteuse dans une vingtaine de fermes a été réalisée. Les veaux issus de ces

accouplements ont été suivis en aveugle, sans en connaître les génotypes, de la naissance à environ un an. A la naissance, il n'y a pas de différence avec les proportions mendéliennes attendues. Après un an, dix veaux étaient morts suite à une inflammation sévère et huit ont été réformés précocement. Parmi ces dix-huit veaux, dix-sept étaient homozygotes pour cette mutation. De plus, sur les veaux vivants, nous avons démontré que la mutation, à l'état homozygote, a un effet hautement significatif sur la croissance après l'âge de 6 mois, ceci comparativement aux individus porteurs ou homozygotes sauvages. Ces résultats sont compatibles avec le rôle de RNF11 dans le complexe A20 intervenant dans la régulation négative suite à l'activation de la cascade de NF-KB (Shembade *et al.*, 2009).

Par ailleurs, nous avons démontré, comme pour la mutation responsable du SQT (Fasquelle *et al.*, 2009), que cette mutation au niveau des *RNF11* confère aux porteurs un avantage sélectif. Cependant la comparaison des cotations linéaires des performances propres entre des taureaux d'IA porteurs et des non-porteurs n'a pas identifié de caractère amélioré par cette mutation expliquant la sélection balancée de cette mutation.

Les génotypages des 114 cas restants (sur 147) présentant un retard de croissance confirment que la mutation c124-2A>G à l'état homozygote explique le retard de croissance chez 40 % des individus. Afin d'identifier d'autres loci pouvant expliquer le retard de croissance chez les 60 % restants, une étude d'association a été réalisée sur 86 individus non homozygotes mutés. Aucun autre locus n'a été identifié confirmant ainsi l'architecture complexe du retard de croissance.

B. Article

**A splice site variant in the bovine *RNF11* gene compromises growth
and regulation of the inflammatory response.**

Sartelet A., Druet T., Michaux C., Fasquelle C., Géron S., Tamma N., Zhang Z.,
Coppieters W., Georges M., Charlier C.

PLoS Genetics, 2012, **8**, e1002581

A Splice Site Variant in the Bovine *RNF11* Gene Compromises Growth and Regulation of the Inflammatory Response

Arnaud Sartelet¹, Tom Druet¹, Charles Michaux², Corinne Fasquelle¹, Sarah Geron¹, Nico Tamma¹, Zhiyan Zhang¹, Wouter Coppiepers¹, Michel Georges¹, Carole Charlier^{1*}

1 Unit of Animal Genomics, GIGA-R and Department of Animal Sciences, Faculty of Veterinary Medicine, University of Liège, Liège, Belgium, **2** Unit of Bioinformatics, Department of Animal Sciences, Faculty of Veterinary Medicine, University of Liège, Liège, Belgium

Abstract

We report association mapping of a locus on bovine chromosome 3 that underlies a Mendelian form of stunted growth in Belgian Blue Cattle (BBC). By resequencing positional candidates, we identify the causative c124-2A>G splice variant in intron 1 of the *RNF11* gene, for which all affected animals are homozygous. We make the remarkable observation that 26% of healthy Belgian Blue animals carry the corresponding variant. We demonstrate in a prospective study design that approximately one third of homozygous mutants die prematurely with major inflammatory lesions, hence explaining the rarity of growth-stunted animals despite the high frequency of carriers. We provide preliminary evidence that heterozygous advantage for an as of yet unidentified phenotype may have caused a selective sweep accounting for the high frequency of the *RNF11* c124-2A>G mutation in Belgian Blue Cattle.

Citation: Sartelet A, Druet T, Michaux C, Fasquelle C, Geron S, et al. (2012) A Splice Site Variant in the Bovine *RNF11* Gene Compromises Growth and Regulation of the Inflammatory Response. PLoS Genet 8(3): e1002581. doi:10.1371/journal.pgen.1002581

Editor: Jonathan Beever, University of Illinois at Urbana-Champaign, United States of America

Received: December 28, 2011; **Accepted:** January 19, 2012; **Published:** March 15, 2012

Copyright: © 2012 Sartelet et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by grants from the Walloon Ministry of Agriculture (Rilouke), the Belgian Science Policy Organisation (SSTC Genefunc PAI), and the University of Liège. CC is Senior Research Associate and TD is Research Associate of the Fonds National de la Recherche Scientifique (Belgium). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: carole.charlier@ulg.ac.be

Introduction

Growth is one of the economically most important phenotypes in livestock production. While genetic variants with large effects on stature account for part of the between-breed variation [1], within-breed variation is likely to be highly multifactorial and polygenic. Accordingly, quantitative trait loci (QTL) influencing growth are reported on all autosomes in the cattle QTL database (<http://www.animalgenome.org/cgi-bin/QTLdb/BT/index>).

The BBC breed is a beef breed that is famous for its “double-muscling” phenotype caused in part by a disruptive 11-bp deletion in the myostatin (*MSTN*) gene [2]. As in other breeds, growth performances are paramount in BBC as they control duration of the fattening period and final carcass weight, hence directly determining profit.

In recent years, an increasing number of young animals with growth retardation as primary symptoms were reported to our heredesurveillance platform. We established this platform in 2005 to rapidly detect genetic defects emerging in the BBC, identify the culprit genes and mutations, and develop diagnostic tests to limit their negative impact [3]. Animals with growth retardation underwent a standard protocol including a genome-wide association study (GWAS) to identify putative causative loci. We herein report the mapping of a locus accounting for ~40% of growth-retardation cases, and identify the causative loss-of-function mutation in the *RING finger protein 11* (*RNF11*) gene. Moreover, we perform a prospective study that indicates that as much as one

third of homozygous mutants die from infection before six months of age. We finally present evidence that carriers of the mutation might benefit from a selective advantage that may account for its unexpectedly high frequency (~13%) in the BBC population.

Results

A major growth-stunting locus maps to BTA3

Between 2008 and 2011, we collected blood samples and epidemiological data from 147 BBC individuals, aged between 3 months and 3 years old, with pronounced (~15% reduction in stature when compared to contemporaries) yet proportionate growth retardation as primary distinctive feature. We initially genotyped 33 of these with a custom-designed 50 K medium-density bovine SNP array [3]. None of these animals would be homozygous or compound heterozygote for the previously identified c.2904-2905delAG [4] and c.1906T>C [5] *MRC2* mutations causing Crooked Tail Syndrome and known to affect stature. Using the genotypes of the corresponding SNPs (yet obtained with a distinct, high-density bovine SNP array) from 275 healthy sires as control, we performed a GWAS using an approach based on hidden haplotype states with a generalized mixed model accounting for stratification (Zhang *et al.*, submitted for publication). A genome-wide significant signal was obtained on BTA3 driven by haplotype state 17, observed at a frequency of 52% in cases versus 12% in controls (Figure 1A). Fourteen of the 33 cases (42%) were homozygous for the corresponding haplotype, causing

Author Summary

Recessive defects in livestock are common, and this is considered to result from the contraction of the effective population size that accompanies intense selection for desired traits, especially when relying heavily on artificial insemination (as males may concomitantly have a very large number of offspring). The costs of recessive defects are assumed to correspond to the loss of the affected animals. By performing a molecular genetic analysis of stunted growth in Belgian Blue Cattle (BBC), we highlight (i) that the economic impact of recessive defects may outweigh the only loss of affected animals and (ii) that some genetic defects are common for reasons other than inbreeding. We first demonstrate that a splice site variant in the *RING finger protein 11* (*RNF11*) gene accounts for ~40% of cases of stunted growth in BBC. We then show that a large proportion of animals that are homozygous for the corresponding *RNF11* mutation die at a young age due to compromised resistance to pathogens. We finally demonstrate that carriers of the mutation benefit from a selective advantage of unidentified origin that accounts for its high frequency in BBC.

a significant deviation from Hardy-Weinberg expectations in cases (expected: 27%, $p < 0.002$), hence suggesting recessivity.

Retrospective phenotypic analysis of the 14 homozygotes revealed shared features: proportionate growth retardation appearing around 5–6 months of age (not observed at birth), normal muscular development, close forehead, long and thin neck, hairy, long and thin head (Figure 2). Pedigree analysis indicated that the 14 individuals traced back to *Galopeur des Hayons* (a once popular BBC sire) on sire and dam side.

A splice site mutation in the *RNF11* gene is the likely causative mutation

Direct examination of the SNP genotypes of the 14 cases homozygous for hidden state 17 revealed a 3.3 Mb (100,727,788–104,017,608 - Btau 4.0) segment of autozygosity (Figure 1B). It encompassed 19 annotated genes of which none was an obvious candidate (Figure 1C). We thus undertook the systematic re-sequencing of all open reading frames (ORF) and intron-exon boundaries. During this process (and after completion of 14/19 genes), we identified an A to G transition (c124-2A>G) mutating the intron 1 acceptor splice site of the *RNF11* gene (Figure 1D). *RNF11* encodes a highly conserved, ubiquitously expressed protein with 154 amino-acids [6], recently recognized as a subunit of the A20 ubiquitin-editing complex regulating NF- κ B signaling [7]. We developed a 5'-exonuclease assay and genotyped (i) the case-control cohort used for GWAS (33 cases, 275 controls), (ii) a diversity panel encompassing 141 animals from eleven breeds other than BBC, (iii) 549 additional normal adult BBC animals, and (iv) *Galopeur des Hayons*. The c124-2A>G variant appeared in near perfect linkage disequilibrium ($D' = 1$; $r^2 = 0.984$) with haplotype state 17 in the case-control cohort. It was not present in non-BBC animals. It had an allelic frequency of 13% amongst the 824 genotyped healthy adult BBC animals, yet without a single animal being homozygous *GG* ($p < 0.01$ under Hardy-Weinberg equilibrium). *Galopeur* was indeed confirmed to be carrier of the c124-2A>G mutation.

The effect of the c124-2A>G mutation on *RNF11* transcripts was examined by RT-PCR using RNA extracted from skeletal muscle, spleen, mesenteric lymph node, thymus, lung, trachea of one *GG* and one *AA* animal. Using two primers located respectively

in exon 1 and 3 and RNA from wild-type *AA* animals, we obtained a unique 360-bp RT-PCR product in all examined tissues, and showed by sequencing that it encompassed the expected exon 2 sequence (data not shown). The same experiment performed with RNA from a homozygous mutant *GG* animal yielded (i) a major product of ~190 bp, and (ii) a minor product of ~360 bp (Figure 3A). The major product was shown by sequencing to correspond to a transcript skipping exon 2. The minor product missed the first seven base pairs of exon 2, and resulted from the activation of a cryptic splice site in exon 2. RT-PCR conducted with primers located respectively in exon 1 and 2 confirmed the existence of transcripts containing exon 2 in homozygous mutants (Figure 3B). Both forms are expected to cause a frameshift, appending 29 (major product) and 14 (minor product) illegitimate residues to a severely truncated (41/154 amino-acids) *RNF11* protein missing the ubiquitin interaction and RING-finger domains. The transcript corresponding to the minor form is expected to undergo non-sense mediated RNA decay (NMRD) [8], due to the occurrence of a stop codon in exon 2 of three. NMRD is not expected to affect the transcript corresponding to the major form as the corresponding open reading frame terminates in exon 3 of three. We compared the levels of *RNF11* transcript in mesenteric lymph node and spleen of a wild-type *AA* and a mutant *GG* animals, using quantitative RT-PCR with primer sets targeting the second (outside of the 7-bp deletion) and third *RNF11* exons, respectively, as well as three internal control genes. In spleen, we observed a 1.1-fold reduction ($p = 0.4$) in the amount of exon 3 containing transcripts, and a 11-fold reduction ($p < 0.005$) in exon 2 containing transcripts. Assuming NMRD of the minor but not of the major product, this allows us to estimate (i) that ~80% of the *RNF11* pre-mRNAs skip exon 2, while ~20% use the exon 2 cryptic splice site, and (ii) that 55% of exon 2 retaining transcripts are being degraded by NMRD. The same analysis conducted in lymph node reveals a ~2-fold reduction ($p < 0.05$) in exon 3 containing transcripts, and ~37-fold reduction ($p < 0.0005$) in exon 2 containing transcripts, corresponding to (i) ~44% of *RNF11* pre-mRNAs skipping exon 2 and ~56% using the exon 2 cryptic splice site, and (ii) ~95% of exon 2 retaining transcripts being degraded by NMRD (Supporting Information S1).

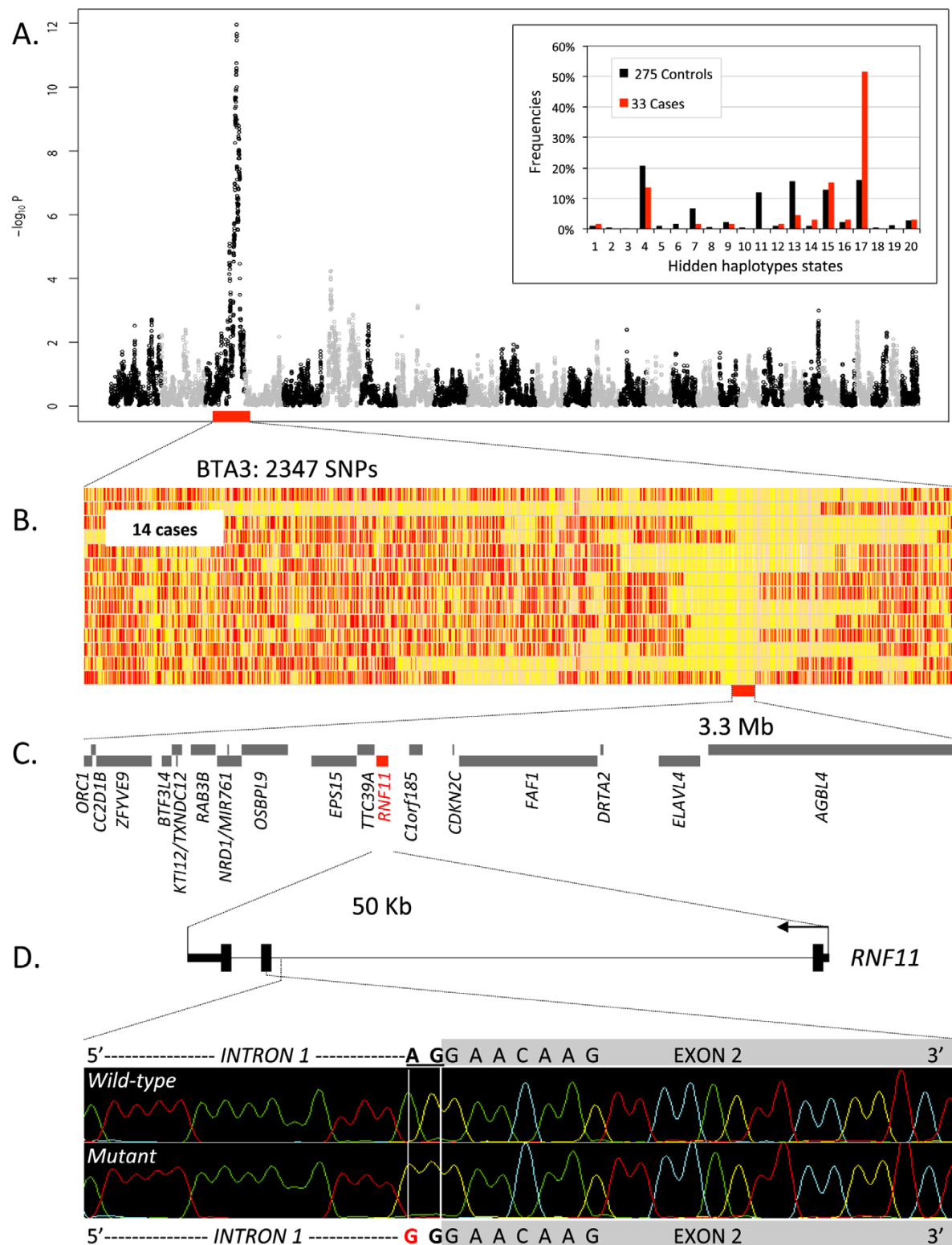
Taken together, our findings strongly support the causality of the c124A>G *RNF11* mutation in determining stunted growth in homozygous *GG* animals.

Increased juvenile mortality accounts for incongruent carrier frequency and disease incidence

The ~26% carrier frequency amongst healthy individuals is incompatible with the number of reported cases of stunted growth. As an example, ~6% of offspring of known carrier bulls should be affected, and such high figures were never recorded. We reasoned that this lower than expected incidence of cases might reflect elimination of mutant animals either before or after birth. Embryonic mortality of homozygous mutant fetuses has been reported for deficiency in uridine monophosphate synthetase (DUMPS) [9], Complex Vertebral Malformation (CVM) [10,11] and Brachyspina (BS) (Charlier *et al.*, submitted for publication).

To test these hypotheses we first examined field data and tested the effect of sire carrier status on (i) "non return (in oestrus) rate" of inseminated cows between 28 and 280 days after insemination, and (ii) rate of mortality, morbidity and culling of offspring between birth and 14 months of age [12]. Non-return rates tended to be slightly decreased when cows were inseminated with semen from carrier sires (i.e. reproductive failure increased), but the effect was not significant ($p = 0.66$). Mortality, morbidity and

RNF11 Splice Site Variant in Cattle



RNF11 Splice Site Variant in Cattle

Figure 1. Genome-wide haplotype-based association mapping of a growth stunting locus on BTA 3. (A) Manhattan plot for the haplotype-based genome-wide association study for stunted growth using a model with 20 ancestral haplotypes. Alternating colors (black and grey dots) mark the limits between autosomes. Inset: frequency of the 20 hidden haplotype states in the 33 cases (red) and the 275 controls (black) at position BTA3:103,391,968 bp. (B) Genotypes of the 14 cases homozygous for hidden haplotype state 17 for 2,347 BTA3 SNPs. Homozygous genotypes are shown in orange or yellow and heterozygous genotypes in red. The limit of the homozygous haplotype shared by the 14 cases is highlighted in red. (C) Gene content of the 3.3 Mb shared interval (19 genes). (D) *RNF11* gene model, and representation of the *RNF11* c124-2A>G splice site variant.
doi:10.1371/journal.pgen.1002581.g001

culling tended to be increased in offspring of carrier sires, but this effect was not significant either ($p = 0.89$) (Supporting Information S1).

As analysis of field data did not provide conclusive results, we performed a prospective study. We identified 105 carrier dams in 22 farms that were pregnant following insemination with semen from known carrier sires. We followed the ensuing 105 calves up to 12 months after birth. The responsible veterinarian (AS) and the

breeders were not aware of the calves' *RNF11* genotype until completion of the study. Genotypic proportions at birth did not deviate significantly from Mendelian expectations (*AA*: 26 (= 24.8%); *AG*: 56 (= 53.3%); *GG*: 23 (= 21.9%); $p = 0.72$). All calves looked normal, and there was no significant effect of *RNF11* genotype on weight or height at birth. However, one year after birth, 10 calves had died and eight had been culled for health-related reasons. Strikingly, all but one of these were homozygous

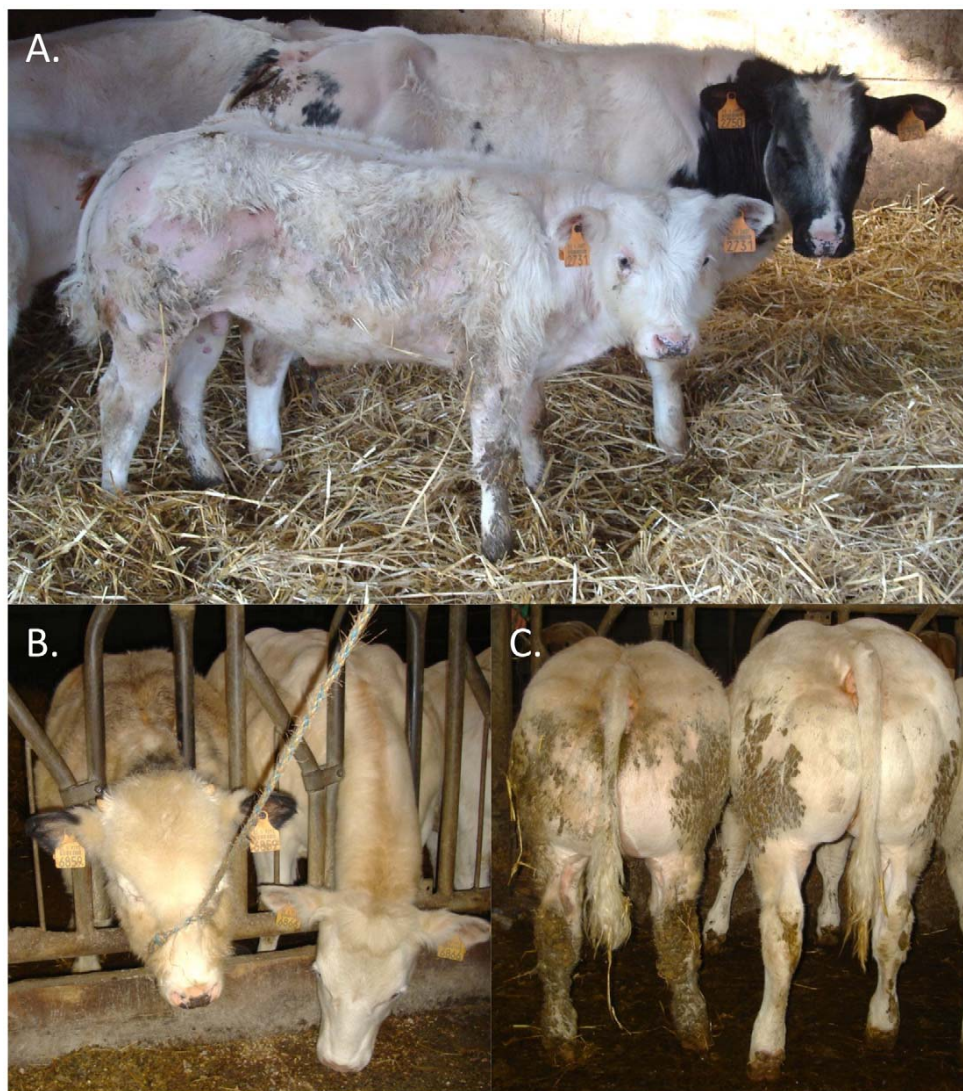


Figure 2. Features of animals homozygous for the *RNF11* c124-2A>G mutation. Affected (front) and control (back) calves of same age, illustrating the proportionate growth retardation, close forehead, and hairy head masking a narrow skull (A). Illustration of the hairy head (B), and normal muscle development (C) of animals homozygous for the *RNF11* c124-2A>G variant.
doi:10.1371/journal.pgen.1002581.g002

RNF11 Splice Site Variant in Cattle

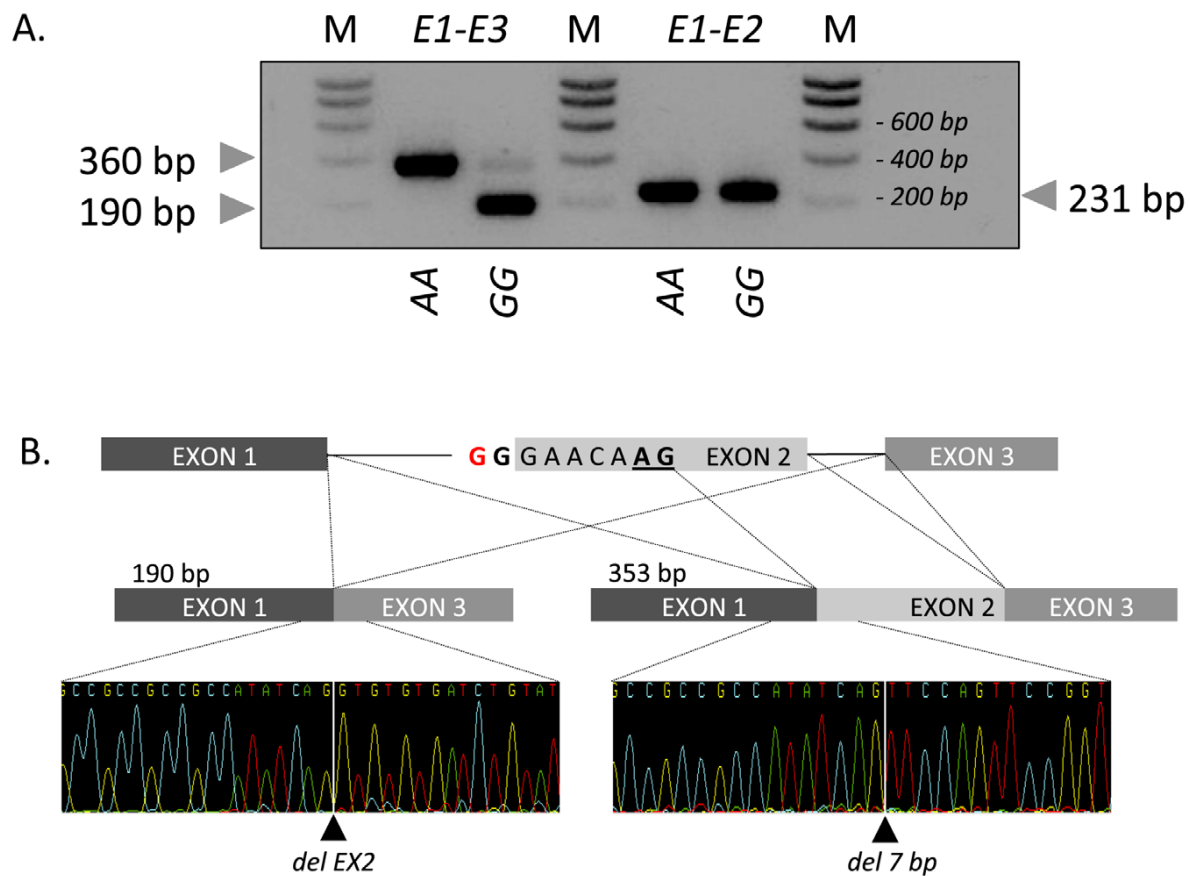


Figure 3. Effect of the c124-2A>G splice site variant on *RNF11* transcripts. (A) Gel electrophoresis of RT-PCR products obtained from mesenteric lymph node from homozygous wild-type (AA) and mutant (GG) animals using primer sets located respectively in exon 1 and 3 (E1–E3) and exon 1 and 2 (E1–E2). M: molecular weight marker. (B) Sequence analysis and structure of the 190-bp and 360-bp RT-PCR products obtained from an affected (GG) animal.

doi:10.1371/journal.pgen.1002581.g003

mutant *GG*, while one was *AG* ($p < 0.0005$) (Figure 4A). While the *AG* animal was euthanized with a limb fracture, the nine deceased *GG* animals died with severe inflammation (primarily pneumonia) (Supporting Information S1). The c124-2A>G genotype had a highly significant ($p \leq 0.001$) effect on post-natal growth. Indeed, all surviving *GG* animals exhibiting stunted development after 6 months (Figure 4B). *A contrario*, the growth pattern of *AG* and *AA* animals was indistinguishable.

Taken together, our data indicate that as much as one third of homozygous *GG* calves die with major inflammation, while all remaining calves exhibit stunted growth and are hence systematically culled prematurely.

Selective advantage of heterozygotes may underlie the high carrier incidence

The 26% carrier frequency amongst healthy BBC animals is puzzling given the observed purifying selection against *GG* animals. This suggests that heterozygotes might benefit from a selective advantage that would maintain the *G* allele at high frequency in the population. Such balanced polymorphism has been demonstrated for *MRC2* loss-of-function mutations causing Crooked Tail Syndrome in homozygotes, yet increased muscle mass in carriers [4,5].

To test this hypothesis, we first used field data and examined the effect of *RNF11* c124-2A>G sire carrier status on own and progeny performances for recorded traits including size, muscularity, type and general appearance [12]. We obtained conflicting results: carrier status appeared to negatively affect the perceived quality of sire, yet improve the quality of its offspring (Supporting Information S1).

As an alternative approach to test for a putative selective advantage benefitting carriers, we evaluated whether the incidence of carriers amongst active AI sires was compatible with Mendelian (0.5:0.5) inheritance of a neutral mutation from the founder bull *Galopeur*. Assuming that the c124-2A>G mutation improves zootechnical performances in heterozygotes, carriers should be over-represented amongst AI sires related to *Galopeur*. Two hundred and six of the 262 BBC AI sires born between 2003 and 2007 were related to *Galopeur* and 58 (=28%) of these proved to carry the *RNF11* c124-2A>G mutation. Using gene dropping in the known genealogies, we computed the probability that 58 or more descendants would be carrier in the absence of selection (no systematic transmission distortion). This probability was 0.0002, 0.0006 and 0.01 assuming a frequency of 0, 0.01 and 0.05 for the c124-2A>G mutation outside the *Galopeur* lineage (Figure 5A). These results suggest that the c124-2A>G mutation indeed

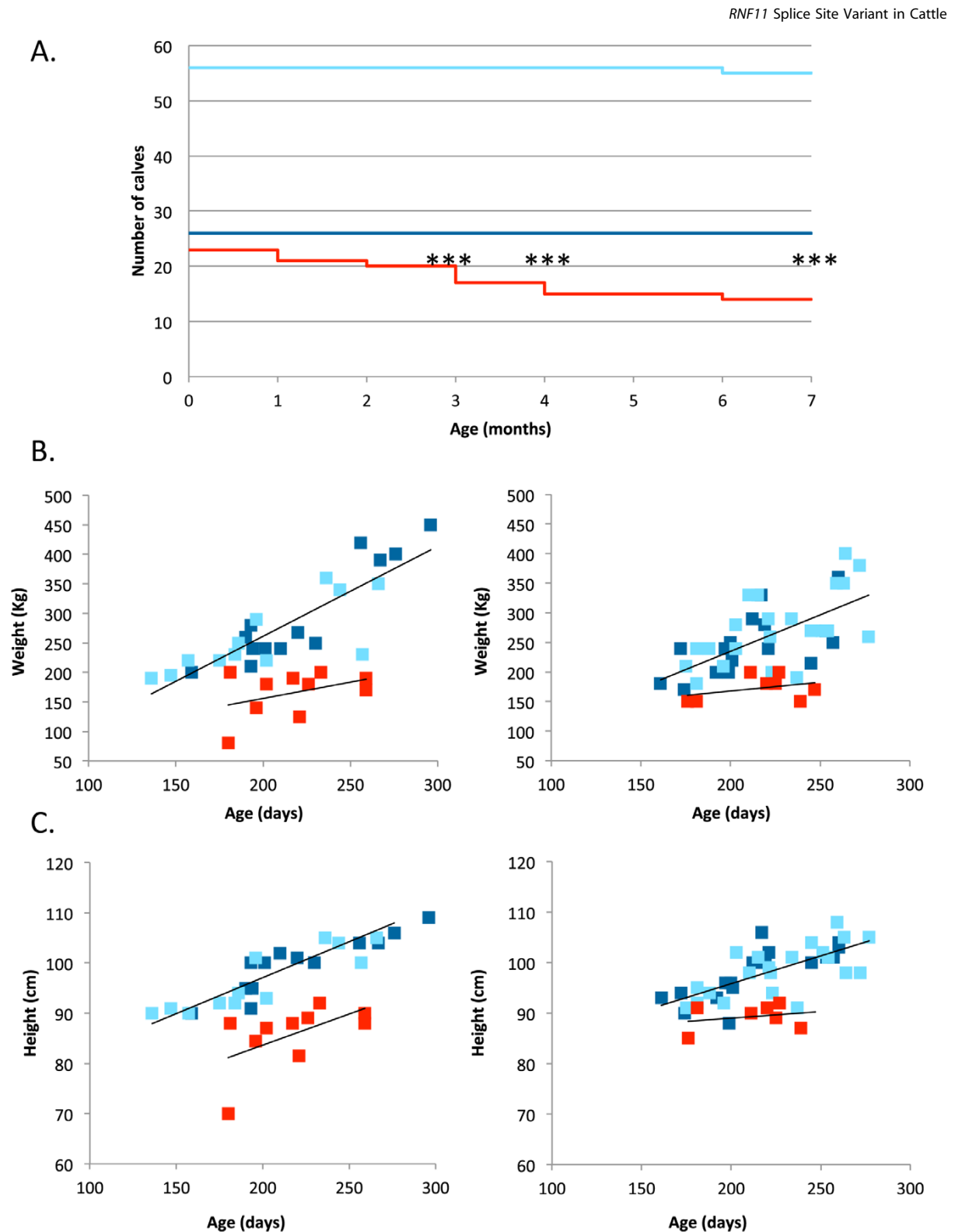


Figure 4. Survival and growth of 105 calves born from matings between carrier sires and dams. (A) Survival (from birth to 7 months of age) of calves sorted by *c124-2A>G* genotype (red: GG, dark blue: AA, light blue: AG) (***: $p < 0.001$). (B) Weight (estimated from heart girth length) and (C) height at withers (from birth to 7 months of age) of calves sorted by *c124-2A>G* genotype (red: GG, dark blue: AA, light blue: AG). Regression lines (black) were fitted separately for affected and non-affected animals.
doi:10.1371/journal.pgen.1002581.g004

underwent a recent selective sweep in the BBC population, although the phenotype that is being selected remains unclear. That 58/206 descendants of *Galopeur* carry the c124-2A>G mutation is best explained by assuming that the mutation has ~10% excess probability (i.e. 60%) to be transmitted by a carrier parent to an AI sire or one of its ancestors (Figure 5B).

Lack of evidence for other major growth-stunting loci

Homozygosity at the *RNF11* c124-2A>G mutation accounted for 14 of the first 33 analyzed cases (i.e. 42%), raising the question of what caused stunted growth in the others. To address this, we genotyped the remaining 114 cases for the c124-2A>G mutation. In agreement with genotypic proportions in the first 33 cases, 47/114 (41%) were homozygous and 23/114 (20%) heterozygous. Therefore, carrier frequency amongst non c124-2A>G homozygous cases was 34% (29/86), which does not differ significantly ($p=0.10$) from the frequency of c124-2A>G carriers in the control cohort (211/829=26%). This suggests that the c124-2A>G mutation is the only common *RNF11* mutation involved in stunted growth in BBC.

To identify putative other loci involved in stunted growth, we genotyped the remaining 67 non c124-2A>G homozygous cases with a medium density 50 K SNP array (Illumina), and rescanned the genome as described before using only non c124-2A>G homozygous cases (86) and the same control cohort (275). As expected, there was no evidence for a residual effect of the *RNF11* locus. Neither was there any genome-wide significant evidence for other loci on any one of the 29 autosomes (Supporting Information S1)

Discussion

We herein demonstrate that a loss-of-function mutation in the *RNF11* gene affects normal growth and disease resistance in calves. This is the first report of a phenotypic effect associated with *RNF11* mutations in any organism, including human and mouse [7].

We postulate that the increased disease susceptibility of homozygous c124-2A>G calves is related to the demonstrated role of RNF11 in feedback down-regulation of NF- κ B by the A20 complex [7]. Indeed, the nine c124-2A>G homozygous calves that underwent necropsy were affected by extensive inflammation of the respiratory tract (eight) or by polyarthritis (one). Of note, A20 knock-out mice die prematurely from multi-organ inflammation [13]. The fact that only ~1/3 of homozygous mutant calves died prematurely is compatible with a defect in the control or resolution of inflammation. External factors, including pathogens, may trigger an intendedly salutary innate and/or adaptive response, that evolves in pathogenic non-resolving inflammation [14].

The effects on growth may be secondary to hidden episodes of uncontrolled inflammation, as proposed for *A20*- and *ITCH*-deficient mice and human [13,15,16]. However, the fact that several of the surviving homozygous c124-2A>G calves appeared perfectly healthy upon clinical examination, suggest that growth retardation might be directly related to alternative functions of RNF11 as modulator of growth factor receptor signaling (particularly TGF- β and EGFR signaling) and transcriptional regulation [6]. It is also noteworthy, that *RNF11* has been found to be highly expressed in bone cells during osteogenesis [17].

Calf mortality is an economically important trait. It is generally considered highly complex and multifactorial, and its heritability is always very low. It is thus difficult to improve using conventional selection strategies. We herein demonstrate that genomic approaches may help dissect such complex phenotypes in sub-

components including some with simple Mendelian determinism amenable to effective “marker assisted selection”. The situation uncovered in this work is reminiscent of bovine leukocyte deficiency (BLAD) in Holstein-Friesian [18], an immune deficiency resulting from CD18 deficiency and causing increased susceptibility to infection in young calves [19].

We provide suggestive evidence that the high incidence of the *RNF11* c124-2A>G mutation in BBC is not only due to drift, but may be due to the superiority of heterozygotes for unidentified selection criteria. Such a situation would be reminiscent of previously described pleiotropic effects on conformation of mutations in the gene encoding the *calcium release channel* (*CRC*) in pigs (causing malignant hyperthermia and porcine stress syndrome in homozygotes) [20] and in the *MRC2* gene in cattle (causing Crooked Tail Syndrome in homozygotes) [4,5]. These examples illustrate some of the issues resulting from the selection of animals with extreme performances.

Materials and Methods

Ethics statement

Blood samples were collected from sires, cows and calves, by trained veterinarians following standard procedures and relevant national guidelines.

Genotyping

Genomic DNA of cases was extracted from 350 μ l of blood using the MagAttract DNA Blood Midi M48 Kit (Qiagen). Genomic DNA of controls was extracted from frozen semen using the MagAttract Mini M48 Kit (Qiagen). The 33 cases of the initial genome scan were genotyped using a custom-made 50 K SNP array [3]. The 67 cases of the second scan (excluding *RNF11* c124-2A>G homozygotes) were genotyped with the BovineSNP50 v2 DNA analysis BeadChip (Illumina). The 275 control sires were genotyped with the BovineHD BeadChip (Illumina). SNP genotyping was conducted using standard procedures at the GIGA genomics core facility.

Genome-wide haplotype-based association studies

Phasing of the SNP genotypes and assignment of the haplotypes to a predetermined number of hidden haplotype states was conducted with PHASEBOOK [21]. Hidden haplotype state-based association analysis was conducted using GLASCOW (Zhang *et al.*, submitted for publication). GLASCOW uses generalized linear models and fits a random hidden haplotype state effect as well as a random polygenic effect to correct for population stratification. Locus-specific p-values were determined from 1,000 permutations assuming a gamma distribution of the used score test (Zhang *et al.*, submitted for publication). We applied a conservative Bonferroni correction assuming 50,000 independent tests to determine the genome-wide significance thresholds.

Mutation scanning

Coding exons of positional candidate genes were amplified from genomic DNA of a homozygous case and a healthy control using standard procedures. The primers used for the *RNF11* gene are listed in the Supporting Information S1. PCR products were directly sequenced using the Big Dye terminator cycle sequencing kit (Applied Biosystem, Foster City, CA). Electrophoresis of purified sequencing reactions was performed on an ABI PRISM 3730 DNA analyzer (PE Applied Biosystems, Foster City, CA). Multiple sequence traces from affected and wild-type animals were aligned and compared using the Phred/Phrap/Consed package (www.genome.washington.edu).

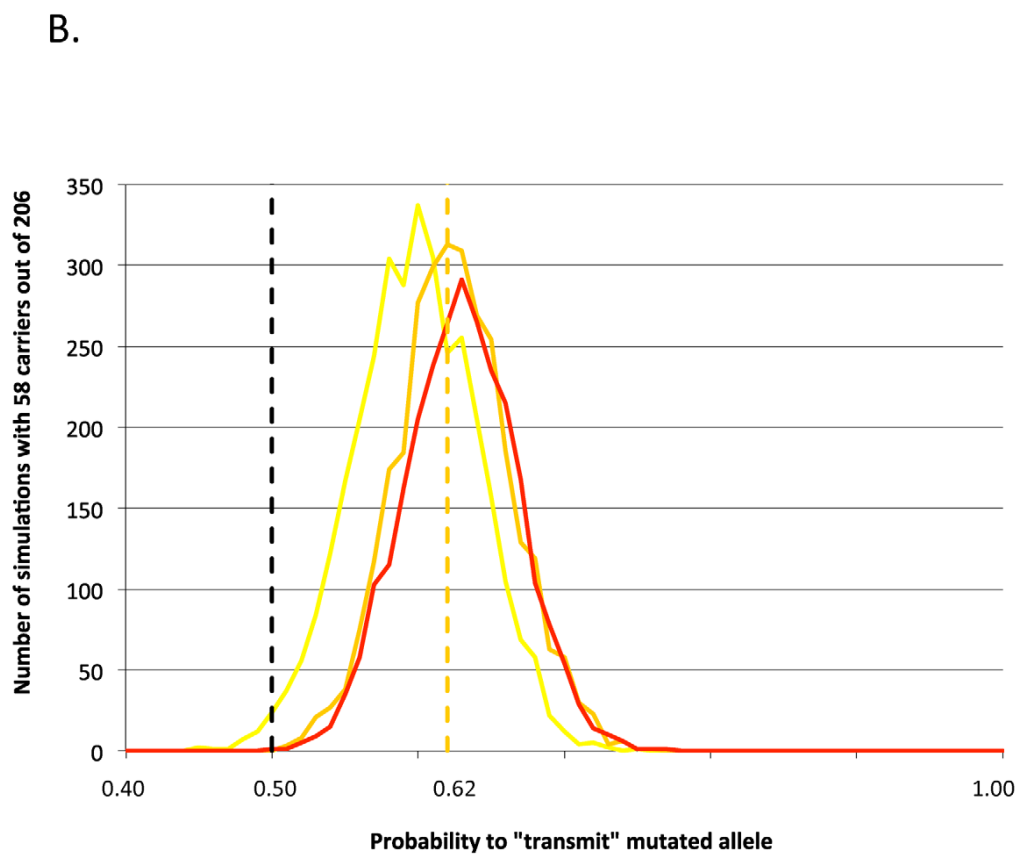
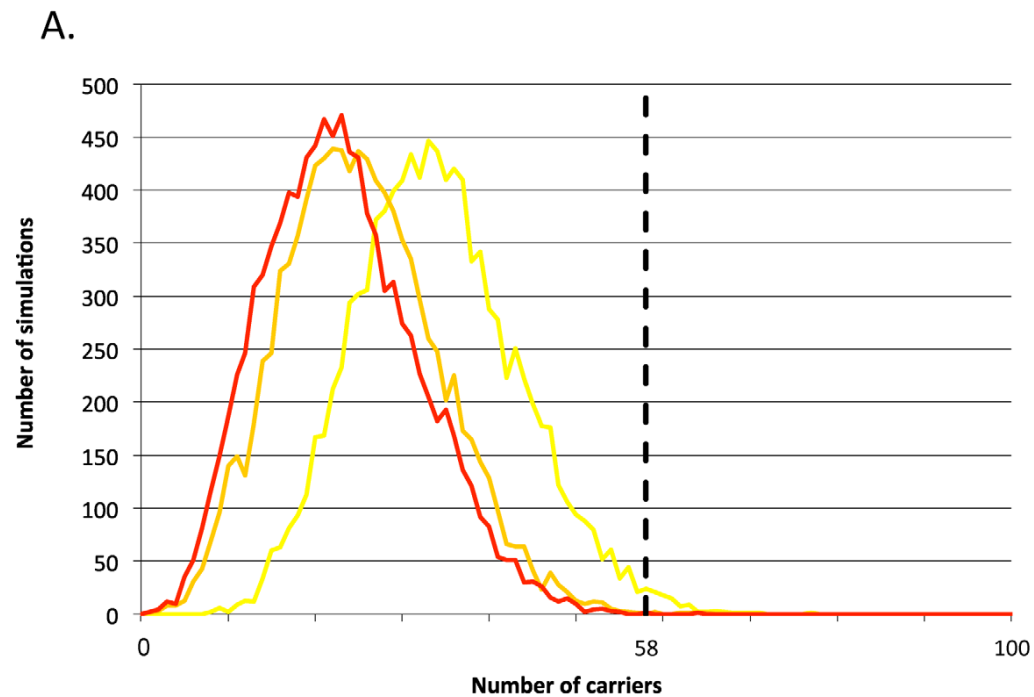
RNF11 Splice Site Variant in Cattle

Figure 5. Signature of selection. (A) Frequency distribution (number of simulations out of 10,000) of the number of sires tracing back to the *Galopeur* founder (total: 206) that are expected to carry the c124-2A>G mutation assuming that it segregates in the corresponding pedigree according to Mendelian expectations, and that the frequency of c124-2A>G outside the *Galopeur* lineage is 0% (red), 1% (orange), or 5% (yellow). The dotted vertical marks the actual number of carrier sires (58) amongst the 206 descendants of *Galopeur*. (B) Distribution of the number of simulations (out of 10,000) yielding 58 carriers out of 206 descendants of *Galopeur* (Y-axis), as a function of the rate of transmission of the mutation from heterozygous carriers (X-axis). Three curves are given corresponding to frequencies of the mutation outside of the *Galopeur*'s lineage of 0% (red), 1% (orange), and 5% (yellow). The dotted orange vertical line corresponds to a transmission rate of 62%, maximizing the number of simulations yielding 58 carriers for a mutation frequency (outside of the *Galopeur*'s lineage) of 1%. doi:10.1371/journal.pgen.1002581.g005

5' exonuclease diagnostic assay of the c124-2A>G *RNF11* mutation

A 5' exonuclease assay was developed to genotype the c124-2A>G *RNF11* mutation, using 5'-AGG AAG AAA CAA AAG GAA AAC ATT ACC TAG A-3' and 5'-TGT TGG ATG ATA GAC CGG AAC TG-3' as PCR primers, and 5'-ACT TGT TCC TAA ATT TT-3' (wild type *A* allele) and 5'-TTG TTC CCA AAT TTT-3' (mutant *G* allele) as probes (Taqman, Applied Biosystems, Foster City, CA). Reactions were carried out on an ABI7900HT instrument (Applied Biosystems, Foster City, CA) using standard procedures.

RT-PCR and cDNA sequencing

Total RNA from *RNF11* c124-2A>G *AA* and *GG* animals was extracted from lung, lymph nodes, spleen, skeletal muscle, thymus and trachea using standard procedures (Trizol, Invitrogen). After *DNase*-treatment (Turbo DNA-free, Ambion), cDNA was synthesized using the SuperScript III First-Strand Synthesis SuperMix (Invitrogen). A cDNA segment was amplified using two *RNF11* specific primers sets: one encompassing exon 2 with primers located in exon 1 and exon 3 (E1–E3) and one encompassing the exon1-exon2 boundary (E1–E2) (Supporting Information S1). PCR products were separated by electrophoresis on a 2% agarose gel containing 0.0001% of SYBR Safe DNA gel stain (Invitrogen) at 100 volts during 40 min and size was evaluated with SmartLadder 200 lanes (Eurogentec). The PCR products were directly sequenced as described above.

Real-time quantitative RT-PCR

Total RNA from *RNF11* c124-2A>G *AA* and *GG* animals was extracted from lymph node, spleen as described above. After *DNase*-treatment (Turbo DNA-free, Ambion), 500 ng of total RNA was reverse transcribed in a final volume of 20 µl using SuperScript III First-Strand Synthesis SuperMix (Invitrogen). PCR reactions were performed in a final volume of 10 µl containing 4 µl of 5-fold diluted cDNA (corresponding to 100 ng of starting total RNA), 1X of Absolute Blue QPCR SYBRE Green ROX Mix 2X (Thermo Fischer Scientific), 0.3 µM forward and reverse primers and nuclease free water. PCR reactions were performed on an ABI7900HT instrument (Applied Biosystems, Foster City, CA) under the following conditions: 10 min at 95°C followed by 40 cycles at 95°C for 15 sec and 60°C for 1 min. Two primers sets were used to test *RNF11* expression and three genes were included as candidate endogenous controls: (1) Beta-Actin (*ACTB*), (2) Ribosomal Protein Large P0 (*RPLP0*), (3) Tyr-3- & Trp-5-Monooxygenase Activation Protein Zeta (*YWHAZ*). The corresponding primer sequences are given in Supporting Information S1. A standard curve with a five point two-fold dilution series (total RNA = 100, 200, 400, 800 and 1600 ng from lymph node and spleen from a *AA* wild-type individual) for each *RNF11* primer set was used to determine the amplification efficiency. All sample/gene combinations were analyzed in triplicate. *ACTB* and *YWHAZ* genes were selected as endogenous controls using geNorm [22]. Normalized relative *RNF11* expression, for exon

2- and exon 3-containing transcripts, in the lymph node and the spleen of a wild-type *AA* and a mutant *GG* animal accounting for primer efficiency were computed using the qbase^{plus} software package (Biogazelle) [23].

Estimating the effect of carrier status for the *RNF11* c124-2A>G mutation on agronomically important traits measured in the field

The effect of the sire's *RNF11* c124-2A>G genotype on non-return rate (NRR) of its mates was estimated using a mixed model including sire's *RNF11* genotype (fixed), year and month at insemination (fixed), mate's herd (random), individual animal effect of the offspring (random) and error. NRR are computed from the AI information collected by inseminators working with the Association Wallonne de l'Elevage (AWE; <http://www.awenet.be/>) at seven time-points after AI. The analysis was performed on 479,674 cows mated to 340 AI sires.

The effect of the sire's *RNF11* c124-2A>G genotype on the rate of mortality, morbidity and culling of its offspring was estimated using a mixed model including sire's *RNF11* genotype (fixed), calf's gender (fixed), year and month of calf's birth (fixed), mate's parity (fixed), calf's *in utero* position (fixed; forward or backward), calf's herd (random), individual animal effect of the calf (random), and error. The corresponding phenotypes are collected by AWE technicians visiting farms, for (i) newborn calves, and (ii) calves having reached the age of 14 months since last visit. The number of records for newborn offspring was 317,350 from 332 AI sires, and for 14 month-old offspring was 126,098 from 288 AI sires.

The effect of the sire's *RNF11* c124-2A>G genotype on its own zootechnical performances was estimated using a mixed model including sire's *RNF11* genotype (fixed), sire's *MRC2* genotype (fixed) [4,5], year and month at scoring (fixed), sire's body condition at scoring (fixed), sire's age at scoring (quadratic regression), individual animal effect for the sires (random) and error [24]. Zootechnical performances of AI sires are recorded between 15 and 56 months of age as 22 linear scores (0–50 score) that are summarized as indexes evaluating size, muscularity, meaty type and general appearance [12]. Three hundred and eleven sires were used in this analysis.

The effect of the sire's *RNF11* c124-2A>G genotype on the zootechnical performances if its offspring was estimated using a mixed model including sire's *RNF11* genotype (fixed), sire's *MRC2* genotype (fixed) [4,5], offspring's gender (fixed), year and month at scoring (fixed), offspring's body condition at scoring (fixed), offspring's age at scoring (quadratic regression), offspring's herd (random), individual animal effect for the offspring (random) and error [24]. The first data set corresponded to the same five global scores (cfr. sire's own performances) measured on 92,475 36-month-old daughters of 306 sires by AWE technicians. The second data set corresponded to weight (Kg), size (cm) and conformation (1–9 score) measured on 95,045 14-month-old offspring of 315 sires.

Covariances between random individual animal effects were assumed to be proportionate to twice the kinship coefficient

RNF11 Splice Site Variant in Cattle

computed from known genealogies. Variance components and fixed effects were computed using MTDFREML [25].

Supporting Information

Supporting Information S1 Supporting figures and tables. (PDF)

Acknowledgments

We are grateful to the GIGA Genomics Platform for their technical assistance, to the breeders for their collaboration, and to the Walloon

Breeding Association (AWE) and the Belgian Blue Beef Herd-Book (HBBBB) for pedigree, performance, and progeny test data.

Author Contributions

Conceived and designed the experiments: MG CC. Performed the experiments: AS CF SG NT. Analyzed the data: AS TD CM ZZ MG CC. Contributed reagents/materials/analysis tools: TD WC. Wrote the paper: AS MG CC.

References

- Karin L, Takeda H, Lin L, Druet T, Arias JA, et al. (2011) Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. *Nat Genet* 43: 405–413.
- Grobet L, Martin LJ, Poncelet D, Pirotin D, Brouwers B, et al. (1997) A deletion in the bovine myostatin gene causes the double-muscling phenotype in cattle. *Nat Genet* 17: 71–74.
- Charlier C, Coppiepiers W, Rollin F, Desmecht D, Agerholm JS, et al. (2008) Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nat Genet* 40: 449–454.
- Fasquelle C, Sartelet A, Li W, Dive M, Tamma N, et al. (2009) Balancing selection of a frame-shift mutation in the MRC2 gene accounts for the outbreak of the Crooked Tail Syndrome in Belgian Blue Cattle. *PLoS Genet* 5: e1000666. doi:10.1371/journal.pgen.1000666
- Sartelet A, Kingbell P, Franklin CK, Fasquelle C, Géron S, et al. (2011) Allelic heterogeneity of Crooked Tail Syndrome: result of balancing selection? *Anim Genet* In press.
- Azmi P, Seth A (2005) RNF11 is a multifunctional modulator of growth factor receptor signalling and transcriptional regulation. *Eur J Cancer* 41: 2549–2560.
- Shembade N, Parvatiyar K, Harhaj NS, Harhaj EW (2009) The ubiquitin-editing enzyme A20 requires RNF11 to downregulate NF-kappaB signalling. *EMBO J* 28: 513–522.
- Chang YF, Imam JS, Wilkinson MF (2007) The nonsense-mediated decay RNA surveillance pathway. *Annu Rev Biochem* 76: 51–74.
- Shanks RD, Robinson JL (1989) Embryonic mortality attributed to inherited deficiency of uridine monophosphate synthase. *J Dairy Sci* 72: 3035–3039.
- Thomsen B, Horn P, Panitz F, Bendixen E, Petersen AH, et al. (2006) A missense mutation in the bovine SLC35A3 gene, encoding a UDP-N-acetylglucosamine transporter, causes complex vertebral malformation. *Genome Res* 16: 97–105.
- Malher X, Beaudeau F, Philipot JM (2006) Effects of sire and dam genotype for complex vertebral malformation (CVM) on risk of return-to-service in Holstein dairy cows and heifers. *Theriogenology* 65: 1215–1225.
- Hanset R, Michaux C, Boonen F (1994) Linear classification in the Belgian Blue Cattle Breed: phenotypic and genetic parameters. Ottawa, Canada, International Committee for Animal Recording (ICAR), seminar, Beef performance recording and genetic evaluation. In: Milk and beef recording: State of the art, 1994. Proc. of the 29th biennial session of the International Committee for Animal Recording (ICAR). Ottawa, Canada, July 31–August 5, 1994. EAAP Publication N° 75: 231–237.
- Lee EG, Boone DL, Chai S, Libby SL, Chien M, et al. (2000) Failure to regulate TNF-induced NF-kappaB and cell death responses in A20-deficient mice. *Science* 289: 2350–2354.
- Nathan C, Ding A (2010) Nonresolving inflammation. *Cell* 140: 871–882.
- Shembade N, Harhaj NS, Parvatiyar K, Copeland NG, Jenkins NA, et al. (2008) The E3 ligase Itch negatively regulates inflammatory signaling pathways by controlling the function of the ubiquitin-editing enzyme A20. *Nat Immunol* 9: 254–262.
- Lohr NJ, Molleston JP, Strauss KA, Torres-Martinez W, Sherman EA, et al. (2010) Human ITCH E3 ubiquitin ligase deficiency causes syndromic multisystem autoimmune disease. *Am J Hum Genet* 86: 447–453.
- Gao Y, Ganss BW, Wang H, Kitching RE, Seth A (2005) The RING finger protein RNF11 is expressed in bone cells during osteogenesis and is regulated by Ets1. *Exp Cell Res* 304: 127–135.
- Shuster DE, Kehrl ME, Jr., Ackermann MR, Gilbert RO (1992) Identification and prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle. *Proc Natl Acad Sci U S A* 89: 9225–9229.
- Nagahata H (2004) Bovine leukocyte adhesion deficiency (BLAD): a review. *J Vet Med Sci* 66: 1475–1482.
- Fujii J, Otsu K, Zorzato F, de Leon S, Khanna VK, et al. (1991) Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science* 253: 448–451.
- Druet T, Georges M (2010) A Hidden Markov Model Combining Linkage and Linkage Disequilibrium Information for Haplotype Reconstruction and Quantitative Trait Locus Fine Mapping. *Genetics* 184: 789–798.
- Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, et al. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3: RESEARCH0034.
- Hellems J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* 8: R19.
- Lynch M, Walsh B (1997) Genetics and analysis of quantitative traits. Sunderland Massachusetts: Sinauer Associates, Inc.
- Boldman KG, Kriese LA, Van Vleck LD, Kachman SD (1995) A manual for Use of MTDFREML. A set of Programs to obtain Estimates of Variances and Covariances. United States Department of Agriculture: Agricultural Research Service.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

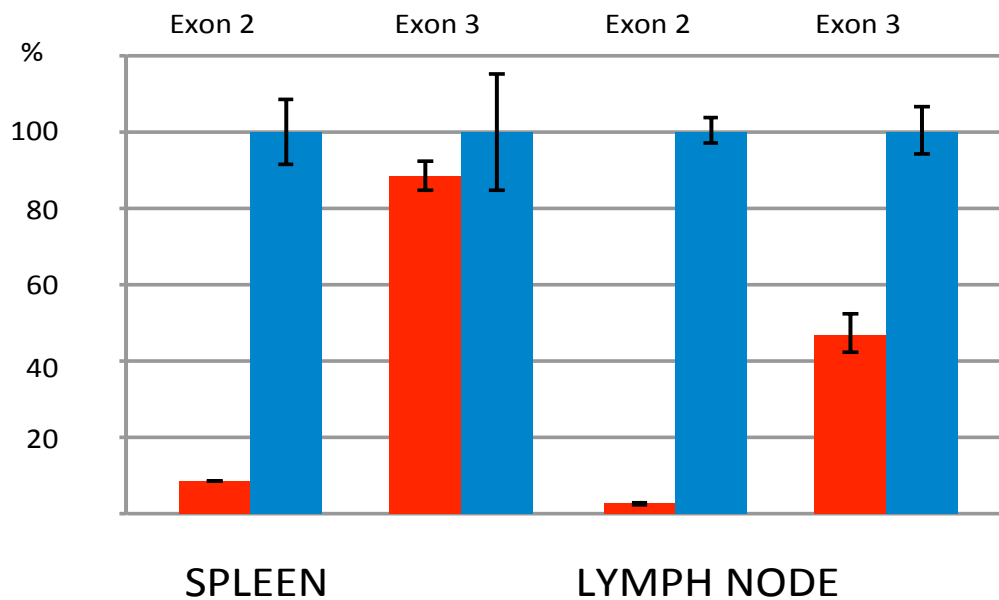
Supporting Information S1

**A splice site variant in the bovine *RNF11* gene
compromises growth and regulation of the inflammatory response.**

Arnaud Sartelet¹, Tom Druet¹, Charles Michaux², Corinne Fasquelle¹,
Sarah Geron¹, Nico Tamma¹, Zhiyan Zhang¹, Wouter Coppieters¹,
Michel Georges¹, Carole Charlier^{1§}.

¹Unit of Animal Genomics, GIGA-R & Department of Animal Sciences, Faculty of Veterinary
Medicine, University of Liège, Belgium.
²Unit of Bioinformatics, Department of Animal Sciences, Faculty of Veterinary Medicine,
University of Liège, Belgium.

1

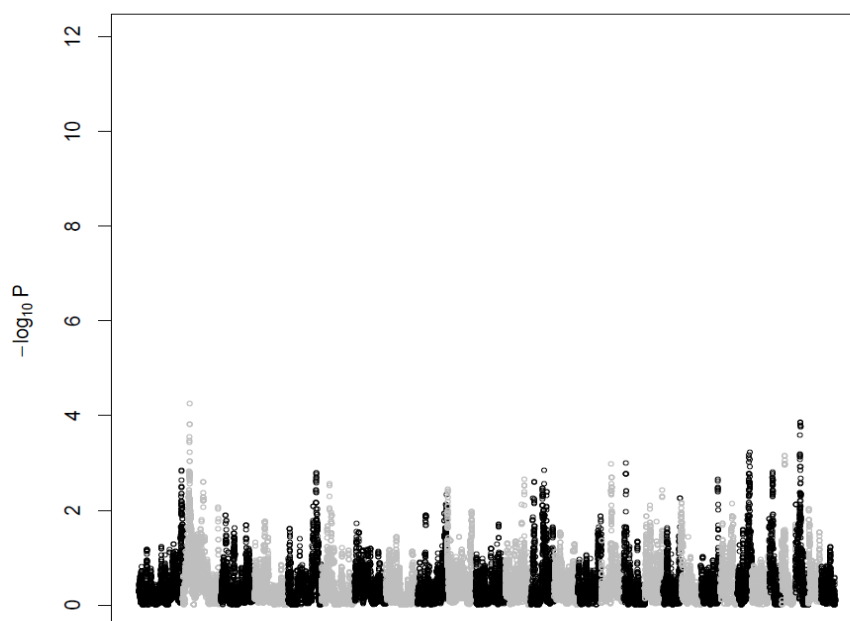


2

3

4 **Supplementary figure 1.** Comparing *RNF11* exon 2 and exon 3 transcript levels in the
 5 spleen (left) and in the mesenteric lymph node (right) of AA wild-type (blue columns)
 6 and GG mutant animals (red columns). Error bars correspond to standard errors over
 7 three replicates per sample.

8



1

2 **Supplementary figure 2. Lack of evidence for other growth stunting loci.**

3 Manhattan plot for the second haplotype-based genome-wide association study for 67
4 stunted-growth cases (excluding c124-A>G homozygous) using a model with 20
5 ancestral haplotypes (20 hidden haplotypes states).

6

- 1 **Supplementary table 1:** Effect of carrier status of AI sires on non return rate of mates
 2 and on the rate of mortality, morbidity and culling of offspring.

3

		Contrast (AA - AG)	p-value
NON RETURN RATE (Days after AI)	28	0.0067	0.7266
	56	-0.0009	0.9602
	84	0.0053	0.7796
	112	0.0060	0.7492
	150	0.0082	0.6602
	200	0.0053	0.7796
	280	0.0065	0.7266
MORTALITY	0 - 3 months	-0.0019	0.9761
	0 - 14 months	-0.0128	0.8966
	1 - 14 months	-0.0101	0.9204
	6 - 14 months	-0.0079	0.9363
MORTALITY + MORBIDITY + CULLING	0 - 14 months	-0.0108	0.9125
	1 - 14 months	-0.0106	0.9046

4

5

- 1 **Supplementary table 2:** Necropsy findings and genotype of the deceased calves in the
- 2 prospective study.

CALF ID	AGE (months)	NECROPSY	GENOTYPE
4322	6	FRACTURE	AG
4334	6	PNEUMONIA	GG
4296	2	PNEUMONIA + MENINGITIS	GG
4477	3	PNEUMONIA + RUMINAL APLASIA	GG
4430	4	PNEUMONIA	GG
4272	< 1	PNEUMONIA + JEJUNAL APLASIA	GG
3806	< 1	LARYNGOTRACHEITIS	GG
4380	2	PNEUMONIA + DIARRHEA	GG
4113	3	POLYARTHRITIS	GG
4548	2	MENINGITIS	GG

3

4

- 1 **Supplementary table 3:** Effect of carrier status of sires on own zootechnical
 2 performances and that of their offspring.

3

		Contrast (AA - AG)	p-value	Carrier characteristics
Performance AI sires (36 months)	Size (cm)	-1.15	0.2766	
	Muscularity	0.28	0.4128	
	Meaty type	0.39	0.1780	
	General appearance	0.45	0.0071	Worse
Performance Daughters (36 months)	Size (cm)	0.26	0.0777	
	Muscularity	-0.36	0.0025	Increased muscularity
	Meaty type	-0.23	0.0103	Increased muscularity
	General appearance	-0.15	0.01006	Better
Progeny test (14 months)	Size (cm)	0.26	0.1372	
	Conformation	-0.13	0.0005	Increased muscularity
	Weight (kg)	3.59	0.0152	Lighter

4

1 **Supplementary table 4:** Primer pairs for the *RNF11* gene.

Name	Primer sequence (5'-3')	Gene part	Size (bp)
gUP1	GATGTAGGAGGATTGGAAAGTG	Exon 1 (ATG)	395 bp
gDN1	CGTGAAGCAGGGAGATGTCATC		
gUP2	CTTTCTTCCTCCCCCAGATCAC	Exon 1 (ATG)	352 bp
gDN2	TTAAAGGTTTCCAAAGTTCAAG		
gUP3	AGAAACAAAAGGAAAACATTAC	Exon 2	506 bp
gDN3	ATGATCAAGTGTGAATAATGTG		
gUP4	GTGATAGAATGACAGGAAGCCG	Exon 3	1292 bp
gDN4	TTGTCCCTTCCAGTGTCCCTTC		
gUP5	TATCGTTGGGGCTGGCTCTATG	Exon 3 (STOP)	1313 bp
gDN5	TGATATTGTGAACTACTGTCTG		

2

3

4 **Supplementary table 5:** Primer pairs for the detection of exon 2 skipping and
5 alternative splicing.

Name	Primer sequence (5'-3')	Size (bp)	
cUP_E1	CCCTGCTTCACGAGTCTCAGTC	231 bp	360 bp
cDN_E2	CATCTCTTCCAGGGTCATAAAC		
cDN_E3	AGGATCTCATCAACCAGTCATC		

6

7

8 **Supplementary table 6:** Quantitative RT-PCR primers .

Gene	Forward primer	Reverse primer	Size (bp)
RNF11_E2	GTTCCGGTCTATCATCCAAC	TCTCTTCCAGGGTCATAAAC	134 bp
RNF11_E3	GTATAGATGACTGGTTGATGAG	TTCTTAATCCCTGGCTCTTTGG	182 bp
ACTB	TCGCGGACAGGATGCAGAAAGA	GCTGATCCACATCTGCTGGAA	149 bp
YWHAZ	GCATCCCACAGACTATTTCC	GCAAAGACAATGACAGACCA	120 bp
RPLP0	TGGGCAAGAACACGATGATG	TGAGGTCCTCCTTGGTGAACA	123 bp

9

10

4- IDENTIFICATION DE LA MUTATION CAUSALE, DANS LE GENE CLCN7, ACCELERANT L'ACTIVATION DU TRANSPORTEUR CLC-7 ET RESPONSABLE D'HAMARTOMES GINGIVAUX ET D'OSTEOPETROSE.

A. Résumé



L'ostéopétrose est une maladie caractérisée par une augmentation de la densité osseuse due à un défaut de la résorption osseuse par les ostéoclastes. La forme sévère de l'ostéopétrose est souvent accompagnée d'anémie et de complications d'origine infectieuse, causées par une hématopoïèse insuffisante dans la cavité médullaire obstruée. Des mutations affectant la formation ou la fonction des ostéoclastes peuvent être à l'origine de l'ostéopétrose. Les pertes de fonctions comprennent des mutations dans les gènes codants pour l'échangeur Cl^-/H^+ : CLC-7 et sa sous-unité β : *Ostm1*, *CLCN7* et *OSTM1*. En plus de sa localisation lysosomale, le tandem $\text{CLC-7}/\text{Ostm1}$ est présent dans la bordure plissée des ostéoclastes. Ce domaine de la membrane plasmique est formé à partir de l'exocytose de vacuoles lysosomales qui est nécessaire à l'acidification de la lacune de Howship au niveau de la surface osseuse. Tant la formation de la bordure plissée que la sécrétion d'acide à ce niveau requièrent la présence d'un tandem $\text{CLC-7}/\text{Ostm1}$ fonctionnel. Le dysfonctionnement de ce tandem au niveau des lysosomes est également responsable de pathologies neuronales, lesquelles ne peuvent être traitées par une transplantation de moelle osseuse, comme c'est le cas pour l'ostéopétrose. Leur fonction au niveau des lysosomes, tout comme leur participation relative à la pathogénie de l'ostéopétrose, soulèvent de nombreuses questions.

Dans cette étude, une nouvelle mutation affectant le gène *CLCN7* est responsable d'un défaut autosomal récessif, avec une forme sévère d'ostéopétrose dans la race Blanc-bleu Belge. Les veaux sont pour la plupart nés morts et présentent des os longs, plus fragiles, sans cavité médullaire, comme observé chez l'homme et les souris déficientes pour le complexe $\text{CLC-7}/\text{Ostm1}$ (Kornak *et al.*, 2001). Ils montraient également des signes de maladie lysosomale. Pourtant, contrairement aux humains et aux souris déficients pour $\text{CLC-7}/\text{Ostm1}$, ils présentaient souvent un hamartome gingival au niveau de la mandibule inférieure et une hydropisie des membranes fœtales chez les mères (Van Loo *et al.*, 2010). Cette mutation n'affecte ni le niveau d'expression du tandem $\text{CLC-7}/\text{Ostm1}$ *in vivo*, ni leur interaction et leur localisation intracellulaire. Des analyses biophysiques ont démontré que la mutation ne réduisait pas

l'échange ionique Cl^-/H^+ , mais par contre, qu'elle accélérât considérablement l'activation et la désactivation, normalement lente de cet échange par le complexe ClC-7/Ostm1.

Il avait été précédemment démontré *in vitro* que certaines mutations causales dans le gène *CLCN7*, chez des patients présentant de l'ostéopétrose et/ou une maladie lysosomale, accélèrent l'activation voltage-dépendante, normalement lente de ClC-7/Ostm1 (Leisle *et al.*, 2011). Mais comme ces mutations diminuent également la quantité protéique de ClC-7 chez ces patients, la causalité de l'accélération de ClC-7 dans l'ostéopétrose restait indéterminée (Kornak *et al.*, 2001). Les résultats obtenus dans cette étude suggèrent non seulement que la perte de fonction, mais aussi qu'une accélération de la cinétique de transport de ClC-7/Ostm1 peut être délétère, révélant par là un nouveau mécanisme par lequel des mutations dans le complexe ClC-7/Ostm1 conduisent à une maladie lysosomale ou osseuse. De manière plus fondamentale, il restera à démontrer comment réconcilier le fait que non seulement des mutations perte de fonction, mais aussi, comme c'est le cas dans notre étude, que des mutations accélérant la cinétique du transporteur mènent au développement de pathologies semblables.

B. Article**A missense mutation accelerating the gating of the lysosomal Cl⁻/H⁺ exchanger ClC-7/Ostm1 causes osteopetrosis with gingival hamartomas in cattle.**

Sartelet A.*, Stauber T.*, Coppieters W., Ludwig C.F., Fasquelle C., Druet T., Zhang Z.,
Ahariz N., Cambisano N., Jentsch T., Charlier C.

Submitted for publication

* Contributed equally

A missense mutation accelerating the gating of the lysosomal Cl⁻/H⁺-exchanger CLC-7/Ostm1 causes osteopetrosis with gingival hamartomas in cattle

Arnaud Sartelet^{1,*}, Tobias Stauber^{2,*}, Wouter Coppieters¹, Carmen F. Ludwig², Corinne Fasquelle¹, Tom Druet¹, Zhiyan Zhang¹, Naima Ahariz¹, Nadine Cambisano¹, Thomas J. Jentsch^{2,§} and Carole Charlier^{1,§}

¹ Unit of Animal Genomics, GIGA-R & Faculty of Veterinary Medicine, University of Liège (B34), 1 Avenue de l'Hôpital, 4000-Liège (Sart Tilman), Belgium

² Leibniz-Institut für Molekulare Pharmakologie (FMP) and Max-Delbrück-Centrum für Molekulare Medizin (MDC), Robert-Rössle-Strasse 10, D-13125 Berlin, Germany

* These authors contributed equally to this work

§ Authors for correspondence (Jentsch@fmp-berlin.de; Carole.Charlier@ulg.ac.be)

Running title: Osteopetrosis with accelerated CLC-7

Key words:

CLCN7, hamartomas, osteopetrosis, lysosomal storage, ion homeostasis, Belgian Blue Cattle

SUMMARY

Chloride/proton exchange by the lysosomal anion transporter ClC-7/Ostm1 is of pivotal importance for the physiology of lysosomes and bone resorption. Mice lacking either ClC-7 or Ostm1 develop a lysosomal storage disease and mutations in either protein have been found to underlie osteopetrosis in mice and humans. Some human disease-causing *CLCN7* mutations accelerate the usually slow voltage-dependent gating of ClC-7/Ostm1. However, it has remained unclear whether the fastened kinetics is indeed causative for the disease. Here we identified and characterized a new deleterious ClC-7 mutation in Belgian Blue Cattle with a severe symptomatology including peri-natal lethality and in most cases gingival hamartomas. By autozygosity mapping and genome-wide sequencing we found a handful of candidate variants, including a cluster of three private SNPs causing the substitution of a conserved tyrosine in the CBS2 domain of ClC-7 by glutamine. The case for ClC-7 was strengthened by subsequent examination of affected calves that revealed severe osteopetrosis and signs of pathological lysosomal storage. The Y750Q mutation preserved the lysosomal localization and assembly of ClC-7/Ostm1, but drastically accelerated its activation by membrane depolarization. These data provide first evidence that accelerated ClC-7/Ostm1 gating per se is deleterious, highlighting a physiological importance of the slow voltage-activation of ClC-7/Ostm1 in lysosomal function and bone resorption.

INTRODUCTION

Osteopetrosis, also known as marble bone disease, is a heterogeneous group of inherited disorders that is characterized by fragile bones generally caused by an impaired bone resorption by osteoclasts (Tolar et al., 2004). Children affected by severe osteopetrosis are often blind due to compression of the optical nerve, and the patients usually die within the first decade of life as a result of secondary defects caused by bone marrow insufficiency. Two of the genes whose mutations cause osteopetrosis are *CLCN7* and *OSTM1*, encoding the Cl^-/H^+ exchanger CIC-7 and its obligate β -subunit Ostm1 (osteopetrosis-associated transmembrane protein 1) (Chalhoub et al., 2003; Kornak et al., 2001; Lange et al., 2006; Leisle et al., 2011). At least in mice, dysfunction of CIC-7/Ostm1 additionally leads to a neuronal ceroid lipofuscinosis (NCL)-like lysosomal pathology (Kasper et al., 2005; Lange et al., 2006), and a severe form of osteopetrosis is accompanied by neurodegeneration and accumulation of ceroid lipofuscin in humans (Steward, 2003).

CIC-7 belongs to the CLC family of chloride channels and transporters which consists of nine mammalian members with diverse physiological roles (Stauber et al., 2012). The CLC family comprises both plasma membrane-localized chloride channels and chloride/proton exchangers that reside predominantly on compartments of the endocytic pathway (Jentsch, 2008; Stauber et al., 2012). CIC-7 and its β -subunit Ostm1 localize to lysosomes of all cells and additionally reside at the ruffled border membrane of bone-resorbing osteoclasts (Kornak et al., 2001; Lange et al., 2006). The latter specialized plasma membrane domain is built up by lysosomal exocytosis and serves to acidify the resorption lacuna, a process required for the degradation of bone material (Teitelbaum, 2000). Both the formation of the ruffled border and the acidification of the resorption lacuna depend on functional CIC-7/Ostm1 (Kornak et al., 2001; Lange et al., 2006). Hence, dysfunction of CIC-7/Ostm1 leads to osteopetrosis in mice and humans (Chalhoub et al., 2003; Kornak et al., 2001; Lange et al., 2006). Contrasting with the impact on the pH of the resorption lacuna, the lysosomal pH is not changed in cells lacking either CIC-7 or Ostm1 (Kasper et al., 2005; Lange et al., 2006; Steinberg et al., 2010; Weinert et al., 2010). In spite of normal lysosomal pH, lysosomal degradation of endocytosed protein is impaired in CIC-7/Ostm1-deficient mice (Wartosch et al., 2009) which develop a neurodegenerative lysosomal storage disease in addition to osteopetrosis (Kasper et al., 2005; Lange et al., 2006; Pressey et al., 2010). Like the other intracellular CLCs, CIC-7/Ostm1

mediates voltage-dependent Cl^-/H^+ exchange, but is unique in its slow activation and deactivation in response to voltage steps (Leisle et al., 2011). Intriguingly, several osteopetrosis-causing human mutations accelerated this gating process (Leisle et al., 2011). Many of these accelerating mutations changed amino-acids close to the interface between cytosolic CBS domains and the transmembrane part of ClC-7 , suggesting that a physical interaction between these parts may be involved in ClC-7 gating (Leisle et al., 2011). However, as we lacked data on the expression levels and subcellular localization of the mutant proteins in affected tissues, it has remained unclear whether accelerated gating per se caused the human pathology.

The Belgian Blue Cattle breed (BBCB; *Bos Taurus*) is a beef breed from Belgium, famous for its hyper-muscled appearance caused by a spontaneous myostatin knock-out allele (Grobet et al., 1997). The BBCB population comprises approximately half a million cows and 2,500 registered bulls. Extensive use of artificial insemination (~half of births), associated with intense selection for traits related to meat production, contracts the effective population size ($N_e \sim 60$), thereby causing frequent outbursts of recessive defects.

In 2005, endorsed by breeders and veterinarians, we established a heredo-surveillance platform to centralize relevant information and biological samples for emerging genetic anomalies, identify responsible genes and mutations, and develop diagnostic tests. Since then, we mapped and identified the causative gene and mutation(s) for twelve and eight diseases respectively, including congenital muscular dystonia I & II, crooked tail syndrome and stunted growth in BBCB (Charlier et al., 2008; Fasquelle et al., 2009; Sartelet et al., 2012a; Sartelet et al., 2012b). These successes were largely due to the development of medium-density single-nucleotide polymorphism (SNP) chips (~50K) for cattle, allowing efficient autozygosity mapping in what equates to a small, genetically isolated population. However, like in other domestic animal species, the downside of the peculiar demography is the limited mapping resolution that can be achieved. Depending on local gene density, segments of autozygosity typically cover tens to a hundred of positional candidate genes. In the absence of obvious functional candidates, pinpointing the causative gene and mutation may remain slow and laborious.

In this study, by combining medium-density SNP arrays and whole-genome sequencing (WGS) we identified a missense mutation in *CLCN7* as responsible for a symptomology in newborn BBCB calves encompassing abnormal skull formation, and often gingival hamartomas

and stillbirth. Further analysis of affected calves revealed a severe osteopetrosis and signs of lysosomal storage. Whereas the mutation neither altered expression levels nor the localization of CLC-7, it accelerated its gating kinetics. These data strongly suggest a functional role of the slow gating kinetics of CLC-7/Ostm1 for lysosomal function and bone resorption.

RESULTS

The gene for gingival hamartomas maps to a 1.3 Mb interval on BTA 25

Between 2008 and 2010, we collected biological material with pedigree records for 63 newborn calves with shared symptomatology: affected calves were mostly stillborn (70% of them), slightly premature (gestation length between 210 and 260 days, 73%), displayed a small body size (45%) and abdominal hydrops (58%), an abnormal skull shape (100%), inferior brachygnathism (100%), protruding tongue (81%) and gingival hamartomas of variable size (up to 15 cm diameter, 80%) located on the lower jaw (Fig. 1 and supplementary material Table S1). All calves born alive were blind and consequently euthanized within days or weeks. At necropsy, we observed liver (82%; Fig. 1) and kidney (59%; not shown) hypertrophies. Mothers of affected calves typically suffered from hydramnios, a condition commonly associated with impaired swallowing in the fetus (Drost, 2007). As a result, ~50% of the dams needed to be culled after parturition, causing considerable extra losses to farmers. Examination of the pedigrees of the 63 cases suggested a recessive mode of inheritance as all of them traced back on both paternal and maternal side to a common ancestor (*Gabin d'Offoux*, a sire born in 1977). Thirty-three cases and 275 healthy controls were initially genotyped using a previously described custom bovine 50K SNP array (Charlier et al., 2008). We performed a genome-wide haplotype-based association study with a generalized linear mixed model accounting for stratification using GLASCOW (Zhang et al., 2012). We identified a single genome-wide significant signal ($p < 10^{-76}$) on the proximal end of chromosome 25 (BTA25) (Fig. 2A). Visual examination of the SNP genotypes defined a non-recombinant autozygous interval of 1.3 Mb (bTau4.0, chr25: 1,084,601–2,288,275) shared by all 33 cases and encompassing more than 100 annotated transcripts (Fig. 2B).

Whole-genome sequencing identifies a likely causative missense mutation in the gene *CLCN7* encoding the chloride/proton exchanger CIC-7

In the absence of any obvious candidate gene, we decided to sequence the whole genome of four affected cases, which were homozygous for the shared identical-by-descent (IBD) segment, and eight unrelated Belgian Blue controls selected as not-carrier for the disease haplotype. Individual paired-end libraries (insert size ~250bp) were generated and ~40 genome-equivalents of paired-end reads (2X110bp) were sequenced on an Illumina GAIIx instrument (supplementary material Fig. S1). Sequence reads were mapped to the bTau4.0 build using the Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009), and resulting alignments directly visualized with the Integrative Genomics Viewer (IGV) (Robinson et al., 2011). An average of ~5.4 Gb were uniquely mapped per animal, resulting in a joint ~8- and ~32-fold coverage of the region of autozygosity in cases and controls, respectively. DNA sequence variants (DSVs) were called with SAMtools (Li et al., 2009). We detected 2001 SNPs and 161 indels for a total of 2162 DSVs; 1733 of these (80%) were filtered-out because they were found in the controls or previously reported in breeds other than BBCB (Coppieters W., personal communication). Out of the remaining 429, 111 mapped to transcribed regions, 8 to ORF (Refseq annotation), of which two caused an amino-acid substitution (data not shown). However, none of these was considered damaging by standard tools for protein-sequence-based prediction of deleteriousness (reviewed in (Cooper and Shendure, 2011)). We further visually scrutinized the entire 1.3Mb region of autozygosity. This revealed three sequence reads encompassing a cluster of three previously undetected nucleotide substitutions [*c2244G>C* + *c2248T>C* + *c2250C>A*] located in exon 23 of the *CLCN7* gene encoding the anion transport protein CIC-7 (supplementary material Fig. S2). Conventional Sanger sequencing of five homozygous cases validated the three DSVs (Fig. 3A). Two of these, [*c2248T>C* + *c2250C>A*], are jointly causing a tyrosine to glutamine substitution (TAC>CAA: Y750Q), while the third [*c2244G>C*] is silent (TCG>TCC: S748). The Y750 residue maps to the second CBS domain (CBS2) of the CIC-7 protein close to the dimer subunit interface (Fig. 3B). It is highly conserved between species (Fig. 3C) but not present in the other eight CLC paralogues (not shown). A position-specific scoring matrix (PSSM) (Marchler-Bauer et al., 2011) output for the CBS2 domain of the mutant bovine CIC-7 is presented in supplementary material Fig. S3.

We developed an assay to directly interrogate the [*c2248T>C* + *c2250C>A*] missense mutations and genotyped the 63 cases, 74 of their parents, 141 animals from 11 breeds other than BBCB, and 6,489 healthy BBCB animals. All cases were homozygous for the Y750Q mutation, while available parents and putative founder (*Gabin*) were all carriers. The mutation was absent in the non-BBCB cohort, and detected at a frequency of 5% (644 carriers) in BBCB controls. None of the genotyped controls was homozygous for the mutant ($p=0.000026$ under Hardy Weinberg equilibrium).

CIC-7 causality is strengthened by severe osteopetrosis and lysosomal pathology of affected calves

As mutations in *CIC-7* can entail osteopetrosis in humans and mice (first described in (Kornak et al., 2001)), we searched for related clinical symptoms in four newly referred cases shown to be homozygous for the Y750Q mutation. This revealed a previously overlooked, severe osteopetrosis of the long bones (Fig. 4). Other features typical for this condition described in human and mouse were also noticed in calves, including: abnormal skull shape, small body size, and visual deficiency or blindness due to retinal degeneration (listed in the supplementary material Table S1). The striking gingival hamartomas displayed by the majority of mutant calves might be related to an impaired tooth eruption, which has been described for *CIC-7* knock-out mice (Kornak et al., 2001) and in anecdotal reports of small odontomas in human patients (Luzzi et al., 2006; Xue et al., 2012). Accordingly, the invariable location of the hamartoma on the lower jaw may be related to the absence of canine and incisor teeth on the upper jaw in cattle.

Besides osteopetrosis, *CIC-7*-deficient mice develop a lysosomal storage disease and display a progressive neurodegeneration accompanied by microglial activation and astrogliosis in the central nervous system (Kasper et al., 2005; Pressey et al., 2010). To assess whether this pathology is also present in homozygous Y750Q mutant calves, we analyzed lysate from available neuronal tissue (cerebellum of a 4-days old calf in comparison to that of a healthy 1-month old calf). We detected an increased amount of the mitochondrial ATP-synthase subunit c (Fig. 5A), which accumulates as lysosomal storage material in *CIC-7*-deficient mice (Kasper et al., 2005). In contrast to *Cln7*^{-/-} mice (Kasper et al., 2005), no increase in cathepsin D levels was detected by immunoblot (Fig. 5B), and probing for microglia (with GSA; not shown) and

astrocytes (immunoblot against GFAP; Fig. 5C) indicated only slight elevations of these cell types. However, considering that the detectable accumulation of storage material precedes neurodegeneration and astrogliosis in *Clcn7*^{-/-} mice (Kasper et al., 2005), this may be explained by the young age of the tested calf. Like in CIC-7-deficient mice (Wartosch et al., 2009), we found an increase in the autophagic marker LC3-II in cerebellum (not shown) and kidney (Fig. 5D).

Taken together, these results strongly supported the causality of the *CLCN7* gene.

The slightly lower amount of CIC-7 protein in Y750Q mutant calves cannot account for the disease phenotype

Total RNA was extracted from kidney and cerebellum from a homozygous mutant calf and an age-matched BBCB control. The RNA was reverse transcribed and three primer pairs (specified in the supplementary material Table S2) were utilized to amplify overlapping amplicons (541, 1088 and 1256bp) covering the entire 2,427bp bovine *CLCN7* ORF. We did not observe obvious qualitative or quantitative differences between case and control amplicons upon agarose gel electrophoresis (data not shown). Sanger sequencing of the three amplicons confirmed the [c2244G>C + c2248T>C + c2250C>A] nucleotide substitutions as only differences (data not shown).

CIC-7 protein destabilization by a disease-causing mutation has previously been reported in a human case (Kornak et al., 2001). To investigate whether the bovine Y750Q mutation might similarly affect CIC-7 stability, we analyzed membrane protein extracts from kidney (Fig. 6A) and cerebellum (not shown) of wild-type and mutant calves by immunoblotting against CIC-7. In both types of tissue, we observed an only modest reduction in CIC-7 concentrations in mutants, which, however, appears very unlikely to cause the observed severe osteopetrotic phenotype of mutant calves because heterozygous *Clcn7*^{+/-} mice display no phenotype (Kornak et al., 2001). We neither detected significant alterations in the levels of the β -subunit of CIC-7, Ostml—neither of the endoplasmic reticulum (ER) precursor form nor of the predominant, processed form (Fig. 6B). By contrast, Ostml levels are drastically reduced in mice lacking CIC-7 ((Lange et al., 2006); Fig. 6B).

The Y750Q mutation does not impair CIC-7–Ostm1 interaction or their correct co-trafficking to lysosomes

Some human osteopetrosis-causing *CLCN7* mutations have been shown to impair trafficking of CIC-7, resulting in retention of the mutant protein in the ER (Leisle et al., 2011; Schulz et al., 2010). To evaluate potential effects of the Y750Q mutation on the subcellular targeting of CIC-7, we inserted the corresponding mutation into available rat and human CIC-7 expression vectors (Y744Q in rCIC-7 and Y746Q in hCIC-7). Upon transfection in HeLa cells, both rat (Fig. 7A) and human (not shown) mutant proteins were correctly targeted to lysosomes like wild-type CIC-7. Ostm1, which is required for protein stability and Cl⁻/H⁺ exchange activity of CIC-7 (Lange et al., 2006; Leisle et al., 2011), needs binding to CIC-7 for ER export and targeting to lysosomes (Lange et al., 2006; Stauber and Jentsch, 2010). To test if the Y750Q mutation affects CIC-7/Ostm1 interaction or Ostm1 trafficking to lysosomes, wild-type and mutant CIC-7 expression vectors were co-transfected with Ostm1 that was labeled by a green-fluorescent protein (GFP) fused to its C-terminus. Immunohistochemistry of transfected cells using antibodies directed against the late endosomal/lysosomal marker protein LAMP-2 showed that Ostm1 properly traffics with mutant hCIC-7^{Y750Q} to lysosomes (Fig. 7B). Furthermore, immunoblots of kidney membrane protein lysate against Ostm1 showed that the bulk of Ostm1 was proteolytically processed to its mature, cleaved ~30 kDa form, with no difference between WT and affected calves (Fig. 6B). As cleavage occurs in, or on the way to lysosomes (Lange et al., 2006), this finding provides support for correct in vivo targeting of Ostm1. Thus, the Y750Q mutation does not inhibit the CIC-7/Ostm1 interaction or its CIC-7-dependent trafficking to lysosomes.

The Y750Q mutation drastically modifies the gating properties of CIC-7/Ostm1

We next assessed the effect of the Y750Q mutation on ion transport by CIC-7/Ostm1. We made use of a human CIC-7 construct that is partially mislocalized to the plasma membrane (hence referred to as CIC-7^{PM}) due to the disruption of two dileucine-based endosomal sorting motifs (Stauber and Jentsch, 2010). The biophysical characteristics of this surface-expressed mutant can be conveniently analyzed in *Xenopus* oocytes by two-electrode voltage-clamp recording (Leisle et al., 2011). CIC-7^{PM}/Ostm1 mediates outwardly rectifying currents that are slowly gated by voltage (Leisle et al., 2011). Introduction of the mutation corresponding to Y750Q

(Y746Q) did not reduce the current amplitude and had no detectable effect on the outwardly rectifying voltage dependence of $\text{ClC-7}^{\text{PM}}/\text{Ostm1}$ (Fig. 8). However, it accelerated the voltage-dependent activation and relaxation kinetics (Fig. 8A) more than three-fold, with an activation rate constant at +80 mV of 104 ± 22 ms for the Y746Q mutant versus 341 ± 64 ms for ‘wild-type’ $\text{hClC-7}^{\text{PM}}/\text{Ostm1}$ (mean values $\pm$ standard deviation from 17 and 13 oocytes, respectively, from three independent batches of oocytes).

DISCUSSION

Identification of the disease-causing gene and mutation

Our previous identifications of genes and mutations underlying various syndromes in BBCB (Charlier et al., 2008; Fasquelle et al., 2009; Sartelet et al., 2012a; Sartelet et al., 2012b) were mainly facilitated by the use of medium-density SNP chips for bovine samples. The disadvantage of this strategy is the size of the relevant segments of autozygosity, which usually contain up to a hundred genes. In human genetics, molecular elucidation of monogenic disorders has benefitted tremendously from next generation sequencing (NGS), either in combination with whole-exome capture or targeted capture of specific genomic regions, or used to directly sequence the entire genome (WGS) (eg. (Lupski et al., 2010; Ng et al., 2009; Volpi et al., 2010)). Whole-exome sequencing was also successfully employed to identify a human *CLCN7* mutation in osteopetrosis patients (Sui et al., 2013). Until recently, targeted capture of specific genomic regions was seriously hampered in cattle as a result of the sub-optimal quality of the reference genome, while exome-capturing reagents were not yet available. We therefore elected to use WGS to tentatively speed up the final positional cloning step. Indeed, even if more costly, WGS was thought to generate extremely valuable “byproducts”, including breed-specific catalogs of common polymorphisms that could be used to filter out non-causative variants (e.g. (Charlier et al., 2012)). Along the same line, WGS can be mined to identify variants that are predicted to disrupt gene/protein function. These include splice site variants, frame-shift mutations, nonsense mutations, and missense mutations predicted to be disruptive. Depending on their frequency, some of them could negatively affect the fitness of the breed. WGS also uncovered thousands of non-synonymous variants of which a number are bound to influence trait of economic

importance. These non-synonymous variants could be directly usable by breeding organizations in genomic selection.

We identified three private base-pair substitutions in the exon 23 of the *CLCN7* gene, for which affected calves were homozygous. These mutations are lying within a 7-bp segment. They most probably originated from a unique mutational event, possibly in the germline of one of the parents of the common ancestor to whom we could trace back all 63 affected calves. Indeed, it was recently shown (Schridder et al., 2011) that up to 3% of *de novo* mutations in the germline are part of a multinucleotide mutational event (MME).

Since we had mapped the gene underlying the severe pathology of the calves, we developed first an indirect, haplotype-based diagnostic test. After the identification of the mutation, we established a direct test to detect carriers. Their successive exploitation to avoid carrier mating immediately dropped down mutant births. The mutation-based diagnostic test has been integrated into the BBCB breeding scheme and is now systematically used to assess carrier-free status before allowing registration of new artificial insemination sires. More than 8,000 animals, mainly males, were genotyped and altogether it contributed to reduce the carrier frequency from 12% in 2008 to ~5% nowadays. Diagnostic tests diffusion, currently for seven recessive diseases, has directly influenced calf mortality rate (before six months), lowering it down from 11% in 2006 to an estimated 7% in 2012. This reduction had a huge economic impact widely recognized by breeders and veterinarians. However, due to the avoidance of carrier mating, we unfortunately lacked tissue from more mutant calves for further analysis. For example, we were devoid of brain sample that could be processed for histological analysis to investigate the lysosomal pathology in more detail.

Osteopetrotic phenotype of the mutant calves

Osteopetrosis has been observed in various cattle breeds. The affected calves are generally born premature (mostly stillborn). Common phenotypes include an abnormal skull shape, inferior brachygnathia (shortened mandible) and a protruding tongue, features also observed with the calves homozygous for the *CLCN7* mutation. Recently, a deletion mutation in the *SLC4A2* gene, encoding the anion exchanger AE2, has been found to underlie osteopetrosis in Red Angus cattle (Meyers et al., 2010). By mediating $\text{Cl}^-/\text{HCO}_3^-$ exchange across the basolateral membrane, AE2 leads to acid uptake into the cytosol (Coury et al., 2013; Wu et al., 2008) which is then extruded

by the H⁺-ATPase of the ruffled border into the resorption lacuna. ClC-7 is thought to enable proton pumping by providing an electric shunt for the proton pump (Kornak et al., 2001). In addition, ClC-7 may be required for the exocytic insertion of proton pumps and other constituents of lysosomal membranes into the ruffled border (Stauber and Jentsch, 2013), a notion supported by the underdevelopment of the ruffled border observed in *Clcn7*^{-/-} mice (Kornak et al., 2001). The hepatomegaly found in most affected calves can be explained by an extramedullary blood production secondary to the osteopetrotic obliteration of the bone marrow cavity (Tolar et al., 2004), and—as mentioned above—the hamartomas may result from the lack of tooth eruption accompanying the osteopetrosis.

The role of slow gating in lysosomal function

Like several human osteopetrosis-associated *CLCN7* mutations (Leisle et al., 2011), the present bovine *CLCN7* mutation drastically accelerated ClC-7/Ostm1 gating and affected a residue that might interfere with contacts of cytosolic CBS domains and the ClC-7 transmembrane block. These observations suggest that the unique slow kinetics of voltage-activation of ClC-7/Ostm1 might be crucial for its cellular function. However, before drawing this conclusion it must be ascertained that these mutations do not change the abundance, localization, or other properties of ClC-7.

So far, the only available expression data of an accelerating mutant (hClC-7^{R762Q}; (Leisle et al., 2011)) show that the mutant protein is unstable in patient-derived fibroblasts (Kornak et al., 2001). Here we show for the first time that a ‘fast’ osteopetrosis-causing ClC-7 mutant displays near-normal expression levels and lysosomal localization in native tissues, strengthening the case for a pathogenic role of accelerated gating. However, it remains unclear how such ‘gain-of-function’ mutations (more current after short times of depolarization) might cause very similar, if not identical phenotypes as a loss-of-function as observed with *Clcn7*^{-/-} mice and several human mutations.

The related endosomal Cl⁻/H⁺ exchangers ClC-3 through ClC-6 display much faster gating kinetics, with a large current component being instantaneously present upon depolarization (Friedrich et al., 1999; Li et al., 2000; Neagoe et al., 2010). It remains to be clarified as to whether the voltage dependence of intracellular CLCs is the same on their native compartments as at the plasma membrane where they could be analyzed. An outwardly

rectifying ClC-7/Ostm1 should be largely inactive at the reported inside-positive voltage of lysosomes (Koivusalo et al., 2011; Leisle et al., 2011). Reductionist model calculations of vesicular acidification, however, predict an inside-negative voltage generated by pH gradient-driven Cl⁻ import (Weinert et al., 2010). The importance of slow ClC-7/Ostm1 gating suggests that the exchanger should be inert to relatively quick voltage transitions, which could be envisaged upon triggered release of luminal calcium or during fusion processes with endosomal compartments. To further approach this interesting issue, new tools to monitor lysosomal voltage and ClC-7/Ostm1 activity will be required. As ion transport by ClC-7/Ostm1 will affect lysosomal ion homeostasis in general and the voltage and osmolarity of this compartment in particular (Stauber and Jentsch, 2013), it remains to be elucidated how an acceleration of the ClC-7/Ostm1 gating impinges upon these parameters.

Conclusion

Using a combination of autozygosity mapping and genome-wide sequencing we have identified a new mutation in the cytosolic carboxy-terminus of the lysosomal Cl⁻/H⁺ exchanger ClC-7 that underlies a recessively inherited severe pathology in neonatal cattle. Like mice and humans with mutations in the genes encoding ClC-7 or its β -subunit Ostm1, affected calves were osteopetrotic and displayed features of lysosomal storage disease. In addition, they often exhibited gingival hamartomas that were likely secondary to the lack of tooth eruption due to the osteopetrosis. Interestingly, in vivo expression levels of the mutant ClC-7 protein were normal and the mutation did not interfere with the subcellular localization of ClC-7 or with the ClC-7/Ostm1 interaction. Since heterologous expression revealed that the mutation did not reduce ClC-7/Ostm1 current magnitudes, there is no evidence that the mutation leads to a loss of function of ClC-7/Ostm1. Instead, the acceleration of the usually slow voltage-dependent activation and deactivation of ClC-7/Ostm1 suggests that a gain-of-function mutation, which results in more current shortly after depolarization, can indeed be deleterious. Several human osteopetrosis-causing *CLCN7* mutations similarly accelerate the gating of ClC-7/Ostm1, but expression levels of the mutant protein could not be studied in patients. In view of the present results we suggest that acceleration of ClC-7 gating suffices to cause osteopetrosis also in humans. Future studies are required to illuminate how the effects of gain-of-function and of loss-of-function mutations merge into the same pathologic phenotype.

MATERIALS AND METHODS

Genome-wide haplotype-based association study

DNA extraction and SNP genotyping using a custom-made bovine 50K SNP array were conducted using standard procedures as described (Charlier et al., 2008). Haplotypes were reconstructed using Beagle (Browning and Browning, 2007). DualPHASE was then used to assign haplotypes to 10 hidden haplotype states (ancestral haplotypes) (Druet and Georges, 2010). Genome-wide association mapping was performed using GLASCOW (Zhang et al., 2012): association between the hidden haplotype states and the phenotype was tested using a generalized linear mixed model (with a logit link function) including a polygenic effect accounting for stratification. The genomic relatedness among individuals was estimated based on hidden haplotype states similarity.

Genome-wide resequencing

Four affected individuals, homozygous for the defined IBD haplotype were selected, as well as eight unaffected unrelated individuals from the same breed. The Pair-End Library Prep Kit v2 from Illumina was used to generate a ~200-250 bp paired-end sequencing library from genomic DNA fragments for each animal. Briefly, total genomic DNA was extracted and fragmented by sonication (Bioruptor, Diagenode). Size-selected fragments were end-repaired and ligated to Illumina Paired-End sequencing adapters. Each library was sequenced on one lane of the flow-cell of an Illumina GAIIx with the paired-end module to generate high-quality reads (2X110bp). Reads were mapped and analyzed with publicly available software: *Burrows-Wheeler Alignment Tool* (<http://bio-bwa.sourceforge.net>) and *Samtools* (<http://samtools.sourceforge.net>). The output files were readily uploaded in the *Integrative Genomics Viewer* (IGV; (Robinson et al., 2011)) and visually scrutinized for private variation.

Mutation validation at the DNA and mRNA levels

A primer pair (gUP1-gDN1) was designed to amplify a PCR product encompassing the three private SNP identified by NGS (supplementary material Table S2). It was used to amplify products from genomic DNA of homozygous cases, carriers and unaffected unrelated individuals using standard procedures. Amplicons were directly sequenced using the Big Dye terminator cycle sequencing kit (Applied Biosystems, Foster City, CA). Electrophoresis of purified

sequencing reactions was performed on an ABI PRISM 3730 DNA analyzer (PE Applied Biosystems, Foster City, CA). Sequence traces were aligned and compared to bovine reference using the Phred/Phrap/Consed package (www.genome.washington.edu). Total RNA was extracted from kidney of one homozygous case and one unaffected unrelated individual using Trizol (Invitrogen). The obtained RNA was treated with TurboDNaseI (Ambion) and cDNA was synthesized using SuperscriptTMIII First Strand Synthesis System for RT-PCR (Invitrogen). Full length *CLCN7* cDNA was amplified using three specific primer pairs (cUP1-cDN1, cUP2-cDN2, cUP3-cDN3) (supplementary material Table S2) and the three overlapping amplicons were directly sequenced and analyzed as describe above.

Development of a genotyping test for the missense mutation

A 5' exonuclease assay (Taqman) was developed to genotype the [*c2248T>C* + *c2250C>A*] *CLCN7* mutations, using 5'-CCA TGG ACC TGT CTG AGT TCA T-3' and 5'-ACC CCC CAG CAG TAC CT-3' as PCR primer pair combined with 5'- CCC [TAC] ACG GTG CCC -3' (wild-type) and 5'- CC [CAA] ACG GTG CCC -3' (mutant) respectively labeled with VIC and FAM as Taqman probes. Allelic discrimination reactions were carried out on an ABI7900HT instrument (Applied Biosystems, Fosters City, CA) using standard procedures.

Western blotting

Kidney tissues from calves and mice were homogenized in HEPES-buffered saline (HBS, pH 7.4) with protease inhibitors (complete protease inhibitor cocktail, Roche) and cleared by two successive centrifugation steps at 1,000 x g for 10 minutes. Membrane protein fractions were prepared by centrifugation at 100,000 x g for 30 minutes and subsequent resuspension in HBS supplemented with 2% (w/v) SDS. 80 µg (membrane protein fraction) or 60 µg (whole-organ lysate) per sample were separated by SDS-PAGE and transferred to nitrocellulose. ClC-7 was detected with a polyclonal rabbit antibody directed against the C-terminal peptide RFPPIQSIHVSQDEREC (100% conserved between murine and bovine). The other primary antibodies were guinea-pig anti-Ostm1 (Lange et al., 2006), rabbit antibodies against subunit c of the mitochondrial ATP synthase (gift of E. Kominami, Tokyo), cathepsin D (Calbiochem), LC3B (Cell Signaling) and actin (Sigma), and mouse antibodies against GFAP (clone G-A-5, Sigma) and α-tubulin (clone DM1A, Sigma). After incubation with secondary antibodies conjugated to

horseradish peroxidase (Jackson ImmunoResearch), chemiluminescence signal was detected with a camera system (PeqLab).

Expression constructs

Constructs for expression of rat (Kornak et al., 2001) and human (Leisle et al., 2011) ClC-7, and of fluorescently tagged Ostm1-GFP (Stauber and Jentsch, 2010) in cell culture were as described. Constructs in pTLN for the expression of partially cell surface-localized human ClC-7^{L23A,L24A,L68A,L69A} (hClC-7^{PM}) and Ostm1 in *Xenopus* oocytes, were described previously (Leisle et al., 2011). Point mutations were introduced by PCR and the complete ORF of all constructs were confirmed by sequencing.

Subcellular targeting and interaction with Ostm1

HeLa cells were transfected with plasmid DNA encoding the respective construct(s) using FuGENE6 (Roche). hClC-7 with Ostm1-GFP constructs were co-transfected at a 50 : 1 (ClC-7 : Ostm1-GFP) weight ratio to prevent expression of excess Ostm1-GFP which would not be exported from the ER. Cells were grown for 26-28 hours before fixation with 4% PFA in PBS for 15 minutes. After incubation with 30 mM glycine in PBS for 5 minutes, cells were incubated sequentially with primary and secondary antibodies in PBS/0.1% saponin supplemented with 5% BSA, each for 1 hour at room temperature. Primary antibodies were rabbit anti-ClC-7 (7N4B, (Kornak et al., 2001)), and mouse anti-Lamp-1 or anti-Lamp-2 (clones H4A3 and H4B4, respectively; both from the DSHB). Secondary antibodies conjugated to AlexaFluor 488, 546 or 633 were from Molecular Probes. Images were acquired with an LSM510 confocal microscope with a 63 x, 1.4 NA oil-immersion lens (Zeiss).

Two-electrode voltage clamp measurements

Current measurements of ClC-7^{PM}/Ostm1 were performed as described (Leisle et al., 2011). *Xenopus laevis* oocytes were injected with cRNA encoding for the hClC-7^{PM} construct and for Ostm1 (23 ng each), which was transcribed with the mMessage Machine kit (Ambion) after linearizing of the pTLN plasmid with *Mlu*I. After three days incubation in ND96 saline (96 mM NaCl, 2 mM K-gluconate, 1.8 mM Ca-gluconate, 1 mM Mg-gluconate, 5 mM HEPES, pH 7.5) at 17°C, currents were measured in ND96 using standard two-electrode voltage clamp at room

temperature employing a TurboTEC amplifier (npi electronic) and pClamp10.2 software (Molecular Devices). The holding potential was -30 mV. 2-seconds test pulses between -80 and +80 mV (in 20-mV steps) were followed by a 0.5-second deactivation pulse at -80 mV. Rate constants of the activation kinetics were determined by fitting the currents of the first 250 ms of depolarization to +80 mV to a single-exponential function.

ACKNOWLEDGEMENTS

We are very grateful to Michel Georges for comments and suggestions. We thank the GIGA Genomic Platform, and Nicole Krönke and Janet Liebold for their technical assistance, and E. Kominami for the ATP-synthase subunit c antibody.

COMPETING INTERESTS

The authors declare that they do not have any competing or financial interests.

AUTHOR CONTRIBUTIONS

AS performed case/control collection, pedigree analysis, phenotyping, necropsy and genotyping data analysis; TS performed Western blotting and localization experiments; TS and CFL performed electrophysiological analysis; TS and TJ conceived experiments and analyzed functional data; CF generated paired-end libraries, amplicon sequencing at gDNA and RNA levels and developed the diagnostic test; NC and NA performed SNP genotyping; WC performed bioinformatical analysis of NGS data; TD and ZZ performed genome-wide association mapping; CC conceived and designed the experiments and analyzed the data; TS, TJ and CC wrote the manuscript with contributions of AS.

FUNDING

TD and CC are respectively Research Associate and Senior Research Associate of the Fonds National de la Recherche Scientifique (Belgium). This work was funded by grants from the Walloon Ministry of Agriculture (Rilouke), the Belgian Science Policy Organization (SSTC Genefunc PAI) and the University of Liège, and by grants from the Deutsche Forschungsgemeinschaft (Je164/7 and SFB444) to TJJ.

REFERENCES

- Browning, S.R., and Browning, B.L.** (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet* **81**, 1084-1097.
- Chalhoub, N., Benachenhou, N., Rajapurohitam, V., Pata, M., Ferron, M., Frattini, A., Villa, A., and Vacher, J.** (2003). Grey-lethal mutation induces severe malignant autosomal recessive osteopetrosis in mouse and human. *Nat Med* **9**, 399-406.
- Charlier, C., Agerholm, J.S., Coppieters, W., Karlskov-Mortensen, P., Li, W., de Jong, G., Fasquelle, C., Karim, L., Cirera, S., Cambisano, N., et al.** (2012). A deletion in the bovine FANCI gene compromises fertility by causing fetal death and brachyspina. *PLoS One* **7**, e43085.
- Charlier, C., Coppieters, W., Rollin, F., Desmecht, D., Agerholm, J.S., Cambisano, N., Carta, E., Dardano, S., Dive, M., Fasquelle, C., et al.** (2008). Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nat Genet* **40**, 449-454.
- Cooper, G.M., and Shendure, J.** (2011). Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat Rev Genet* **12**, 628-640.
- Coury, F., Zenger, S., Stewart, A.K., Stephens, S., Neff, L., Tsang, K., Shull, G.E., Alper, S.L., Baron, R., and Aliprantis, A.O.** (2013). SLC4A2-mediated Cl/HCO₃⁻ exchange activity is essential for calpain-dependent regulation of the actin cytoskeleton in osteoclasts. *Proc Natl Acad Sci U S A* **110**, 2163-2168.
- Drost, M.** (2007). Complications during gestation in the cow. *Theriogenology* **68**, 487-491.
- Druet, T., and Georges, M.** (2010). A hidden markov model combining linkage and linkage disequilibrium information for haplotype reconstruction and quantitative trait locus fine mapping. *Genetics* **184**, 789-798.
- Fasquelle, C., Sartelet, A., Li, W., Dive, M., Tamma, N., Michaux, C., Druet, T., Huijbers, I.J., Isacke, C.M., Coppieters, W., et al.** (2009). Balancing selection of a frame-shift mutation in the MRC2 gene accounts for the outbreak of the Crooked Tail Syndrome in Belgian Blue Cattle. *PLoS Genet* **5**, e1000666.
- Friedrich, T., Breiderhoff, T., and Jentsch, T.J.** (1999). Mutational analysis demonstrates that CIC-4 and CIC-5 directly mediate plasma membrane currents. *J Biol Chem* **274**, 896-902.
- Grobet, L., Martin, L.J., Poncelet, D., Pirottin, D., Brouwers, B., Riquet, J., Schoeberlein, A., Dunner, S., M  nissier, F., Massabanda, J., et al.** (1997). A deletion in the bovine myostatin gene causes the double-musled phenotype in cattle. *Nat Genet* **17**, 71-74.
- Jentsch, T.J.** (2008). CLC chloride channels and transporters: From genes to protein structure, pathology and physiology. *Crit Rev Biochem Mol Biol* **43**, 3-36.
- Kasper, D., Planells-Cases, R., Fuhrmann, J.C., Scheel, O., Zeitz, O., Ruether, K., Schmitt, A., Po  t, M., Steinfeld, R., Schweizer, M., et al.** (2005). Loss of the chloride channel CIC-7 leads to lysosomal storage disease and neurodegeneration. *EMBO J* **24**, 1079-1091.
- Koivusalo, M., Steinberg, B.E., Mason, D., and Grinstein, S.** (2011). In situ measurement of the electrical potential across the lysosomal membrane using FRET. *Traffic* **12**, 972-982.
- Kornak, U., Kasper, D., B  sl, M.R., Kaiser, E., Schweizer, M., Schulz, A., Friedrich, W., Delling, G., and Jentsch, T.J.** (2001). Loss of the CIC-7 chloride channel leads to osteopetrosis in mice and man. *Cell* **104**, 205-215.

- Lange, P.F., Wartosch, L., Jentsch, T.J., and Fuhrmann, J.C.** (2006). CIC-7 requires Ostm1 as a β -subunit to support bone resorption and lysosomal function. *Nature* **440**, 220-223.
- Leisle, L., Ludwig, C.F., Wagner, F.A., Jentsch, T.J., and Stauber, T.** (2011). CIC-7 is a slowly voltage-gated $2\text{Cl}^-/1\text{H}^+$ -exchanger and requires Ostm1 for transport activity. *EMBO J* **30**, 2140-2152.
- Li, H., and Durbin, R.** (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.** (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079.
- Li, X., Shimada, K., Showalter, L.A., and Weinman, S.A.** (2000). Biophysical Properties of CIC-3 Differentiate It from Swelling-activated Chloride Channels in Chinese Hamster Ovary-K1 Cells. *J Biol Chem* **275**, 35994-35998.
- Lupski, J.R., Reid, J.G., Gonzaga-Jauregui, C., Rio Deiros, D., Chen, D.C., Nazareth, L., Bainbridge, M., Dinh, H., Jing, C., Wheeler, D.A., et al.** (2010). Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med* **362**, 1181-1191.
- Luzzi, V., Consoli, G., Daryanani, V., Santoro, G., Sfasciotti, G.L., and Polimeni, A.** (2006). Malignant infantile osteopetrosis: dental effects in paediatric patients. Case reports. *Eur J Paediatr Dent* **7**, 39-44.
- Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., et al.** (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* **39**, D225-229.
- Meyers, S.N., McDaniel, T.G., Swist, S.L., Marron, B.M., Steffen, D.J., O'Toole, D., O'Connell, J.R., Beever, J.E., Sonstegard, T.S., and Smith, T.P.** (2010). A deletion mutation in bovine SLC4A2 is associated with osteopetrosis in Red Angus cattle. *BMC Genomics* **11**, 337.
- Neagoe, I., Stauber, T., Fidzinski, P., Bergsdorf, E.Y., and Jentsch, T.J.** (2010). The late endosomal CIC-6 mediates proton/chloride countertransport in heterologous plasma membrane expression. *J Biol Chem* **285**, 21689-21697.
- Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., et al.** (2009). Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**, 272-276.
- Pressey, S.N., O'Donnell, K.J., Stauber, T., Fuhrmann, J.C., Tyynela, J., Jentsch, T.J., and Cooper, J.D.** (2010). Distinct neuropathologic phenotypes after disrupting the chloride transport proteins CIC-6 or CIC-7/Ostm1. *J Neuropathol Exp Neurol* **69**, 1228-1246.
- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P.** (2011). Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26.
- Sartelet, A., Druet, T., Michaux, C., Fasquelle, C., Geron, S., Tamma, N., Zhang, Z., Coppieters, W., Georges, M., and Charlier, C.** (2012a). A splice site variant in the bovine RNF11 gene compromises growth and regulation of the inflammatory response. *PLoS Genet* **8**, e1002581.
- Sartelet, A., Klingbeil, P., Franklin, C.K., Fasquelle, C., Geron, S., Isacke, C.M., Georges, M., and Charlier, C.** (2012b). Allelic heterogeneity of Crooked Tail Syndrome: result of balancing selection? *Anim Genet* **43**, 604-607.

- Schrider, D.R., Hourmozdi, J.N., and Hahn, M.W.** (2011). Pervasive multinucleotide mutational events in eukaryotes. *Curr Biol* **21**, 1051-1054.
- Schulz, P., Werner, J., Stauber, T., Henriksen, K., and Fendler, K.** (2010). The G215R mutation in the Cl⁻/H⁺-antiporter CLC-7 found in ADO II osteopetrosis does not abolish function but causes a severe trafficking defect. *PLoS One* **5**, e12585.
- Stauber, T., and Jentsch, T.J.** (2010). Sorting motifs of the endosomal/lysosomal CLC chloride transporters. *J Biol Chem* **285**, 34537-34548.
- Stauber, T., and Jentsch, T.J.** (2013). Chloride in vesicular trafficking and function. *Annu Rev Physiol* **75**, 453-477.
- Stauber, T., Weinert, S., and Jentsch, T.J.** (2012). Cell biology and physiology of CLC chloride channels and transporters. *Compr Physiol* **2**, 1701-1744.
- Steinberg, B.E., Huynh, K.K., Brodovitch, A., Jabs, S., Stauber, T., Jentsch, T.J., and Grinstein, S.** (2010). A cation counterflux supports lysosomal acidification. *J Cell Biol* **189**, 1171-1186.
- Steward, C.G.** (2003). Neurological aspects of osteopetrosis. *Neuropathol Appl Neurobiol* **29**, 87-97.
- Sui, W., Ou, M., Liang, J., Ding, M., Chen, J., Liu, W., Xiao, R., Meng, X., Wang, L., Pan, X., et al.** (2013). Rapid gene identification in a Chinese osteopetrosis family by whole exome sequencing. *Gene* **516**, 311-315.
- Teitelbaum, S.L.** (2000). Bone resorption by osteoclasts. *Science* **289**, 1504-1508.
- Tolar, J., Teitelbaum, S.L., and Orchard, P.J.** (2004). Osteopetrosis. *N Engl J Med* **351**, 2839-2849.
- Volpi, L., Roversi, G., Colombo, E.A., Leijsten, N., Concolino, D., Calabria, A., Mencarelli, M.A., Fimiani, M., Macciardi, F., Pfundt, R., et al.** (2010). Targeted next-generation sequencing appoints c16orf57 as clericuzio-type poikiloderma with neutropenia gene. *Am J Hum Genet* **86**, 72-76.
- Wartosch, L., Fuhrmann, J.C., Schweizer, M., Stauber, T., and Jentsch, T.J.** (2009). Lysosomal degradation of endocytosed proteins depends on the chloride transport protein CLC-7. *FASEB J* **23**, 4056-4068.
- Weinert, S., Jabs, S., Supanchart, C., Schweizer, M., Gimber, N., Richter, M., Rademann, J., Stauber, T., Kornak, U., and Jentsch, T.J.** (2010). Lysosomal pathology and osteopetrosis upon loss of H⁺-driven lysosomal Cl⁻ accumulation. *Science* **328**, 1401-1403.
- Wu, J., Glimcher, L.H., and Aliprantis, A.O.** (2008). HCO₃⁻/Cl⁻ anion exchanger SLC4A2 is required for proper osteoclast differentiation and function. *Proc Natl Acad Sci U S A* **105**, 16934-16939.
- Xue, Y., Wang, W., Mao, T., and Duan, X.** (2012). Report of two Chinese patients suffering from CLCN7-related osteopetrosis and root dysplasia. *J Craniomaxillofac Surg* **40**, 416-420.
- Zhang, Z., Guillaume, F., Sartelet, A., Charlier, C., Georges, M., Farnir, F., and Druet, T.** (2012). Ancestral haplotype-based association mapping with generalized linear mixed models accounting for stratification. *Bioinformatics* **28**, 2467-2473.

Figure 1 : Clinical features of congenital hamartomas in affected Belgian Blue calves. A. Premature stillborn calf exhibiting gingival hamartoma abnormal skull shape, hydrops and hepatomegaly (inlet : livers from mutant (left) and wild-type calves). B. Sagittal section of a case head revealing a hamartoma within the inferior jaw (arrow). C. Dead case presenting a voluminous hamartoma, note the tooth inclusion (arrow). D. Alive cases with abnormal skull shape accompanied by a protruding tongue.

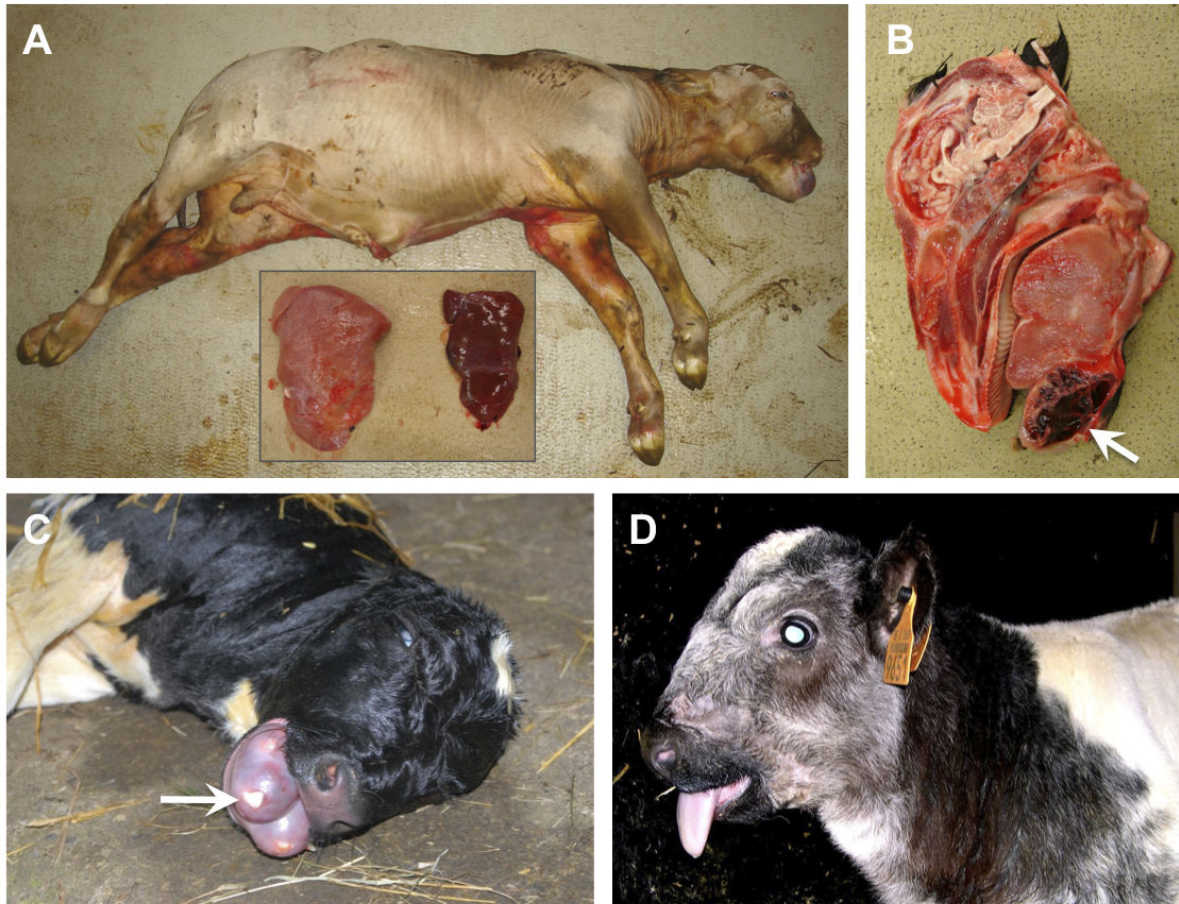


Figure 2: Genome-wide association mapping for the hamartoma disease locus. A. Manhattan plot for case/control GWAS presenting a unique genome-wide significant signal on chromosome 25; the 29 autosomes are alternately labeled in grey or black; inlet: typical hamartoma case. **B.** Cases genotypes for 1,256 chromosome 25 SNP; homozygous genotypes are shown in yellow or white, heterozygous genotypes in red; the centromeric 1.3Mb homozygosity region, identical by state among all cases, is highlighted in red.

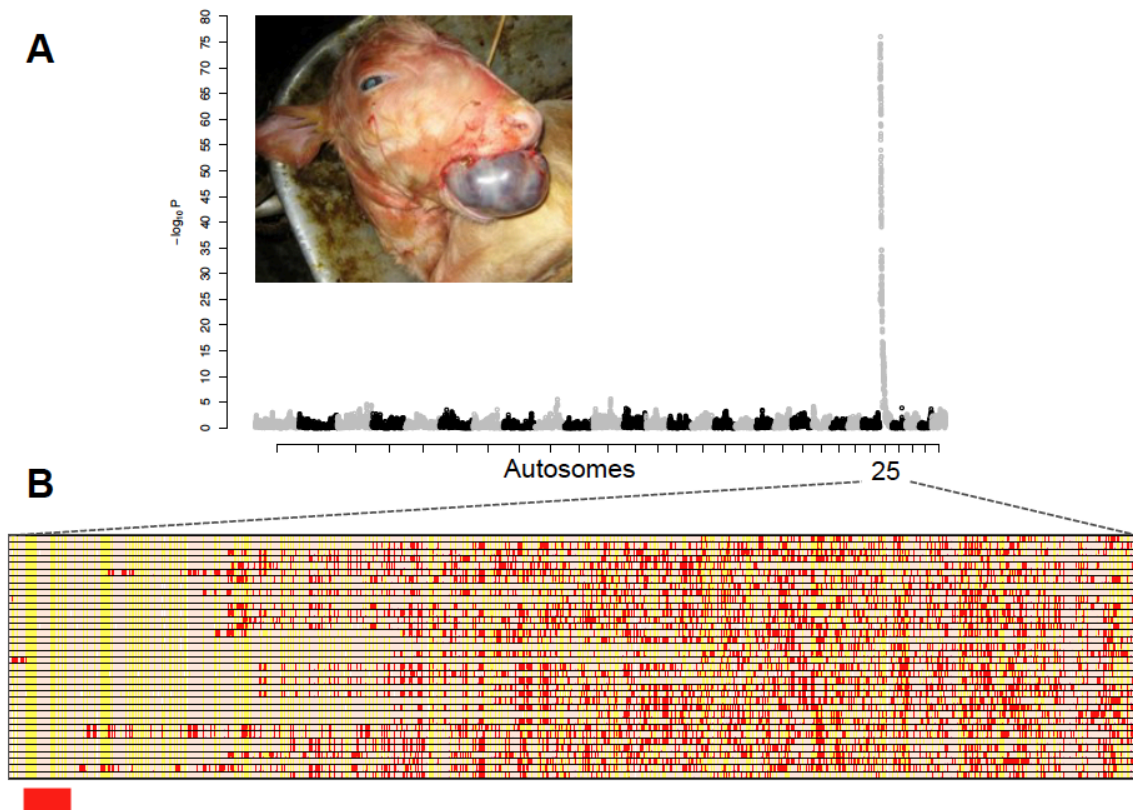


Figure 3: Missense mutations in the second cystathionine β -synthase domain (CBS2) of the CLC-7 protein. **A.** Sequence traces of *CLCN7* exon 23 for wild-type (top) and mutant (bottom) calves; triangles pinpoint three nucleotide substitutions with the corresponding Y750Q amino acid mutation in red. **B.** X-ray structure of CmCLC displaying the localization of the mutated amino acid (red spheres in either subunit of the dimer) according to the published alignment between CmCLC and CLC-7 [46]. The transmembrane core-forming parts are depicted in gray, the cytosolic domains CBS1 in yellow and CBS2 in green, using darker colors for one subunit. **C.** CLC-7 CBS2 domain alignment, from mammals to fish, showing its conservation through evolution with the Y750Q mutation in red.

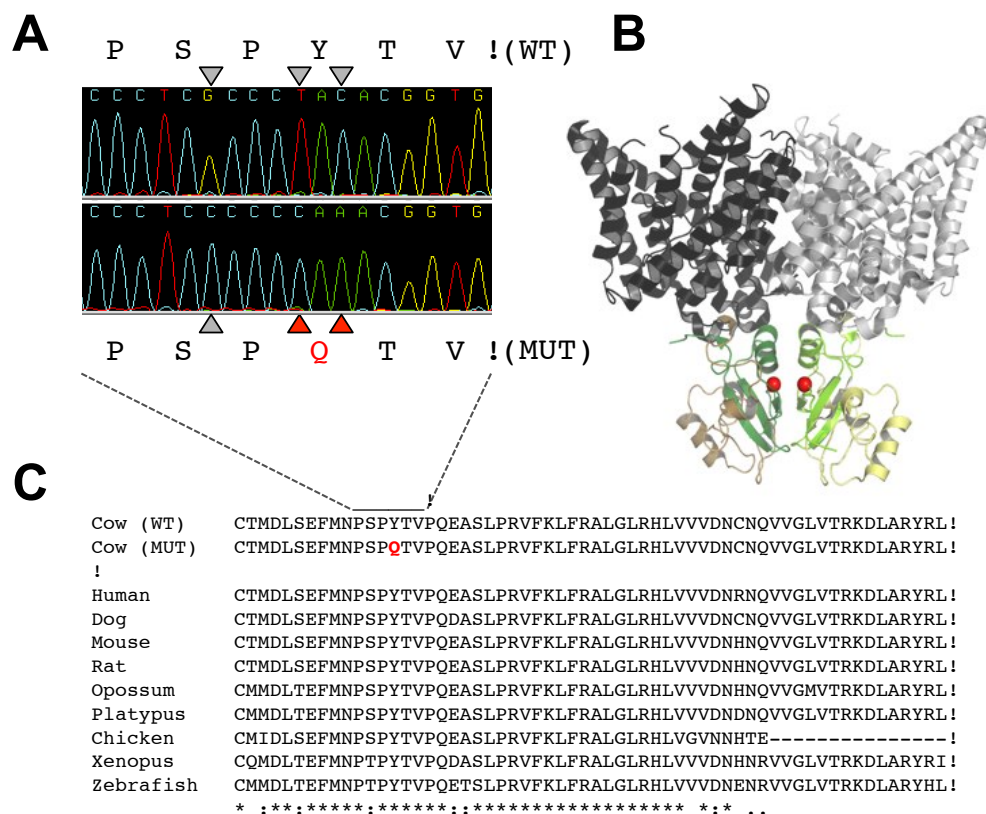


Figure 4: Severe osteopetrotic phenotype exhibited by Y750Q homozygous calves.

A. Dorsoventral radiographs of extended hind legs of one week-old homozygous mutant calf (MUT, right) and a age-matched Belgian Blue control calf (WT, left) are presented. X-rays were performed using a digital radiograph system (Vertix Vet [150 kV/600 mA], Siemens, Germany) with 75 kV and 100 mAs as technical parameters. The bone marrow cavity is clearly visible within metatarsus of the wild-type calf but absente from the mutant corresponding bone (red arrows). The mutant calf also exhibit a tibia fracture (yellow arrow) probably consecutive to the acknowledged increased fragility of osteopetrotic long bones. **B.** Fresh transversal (left) and sagittal (right) sections of long bones (tibia) of aged-matched mutant (MUT) and control (WT) calves calf showing an absence of central bone marrow cavity for the mutant.

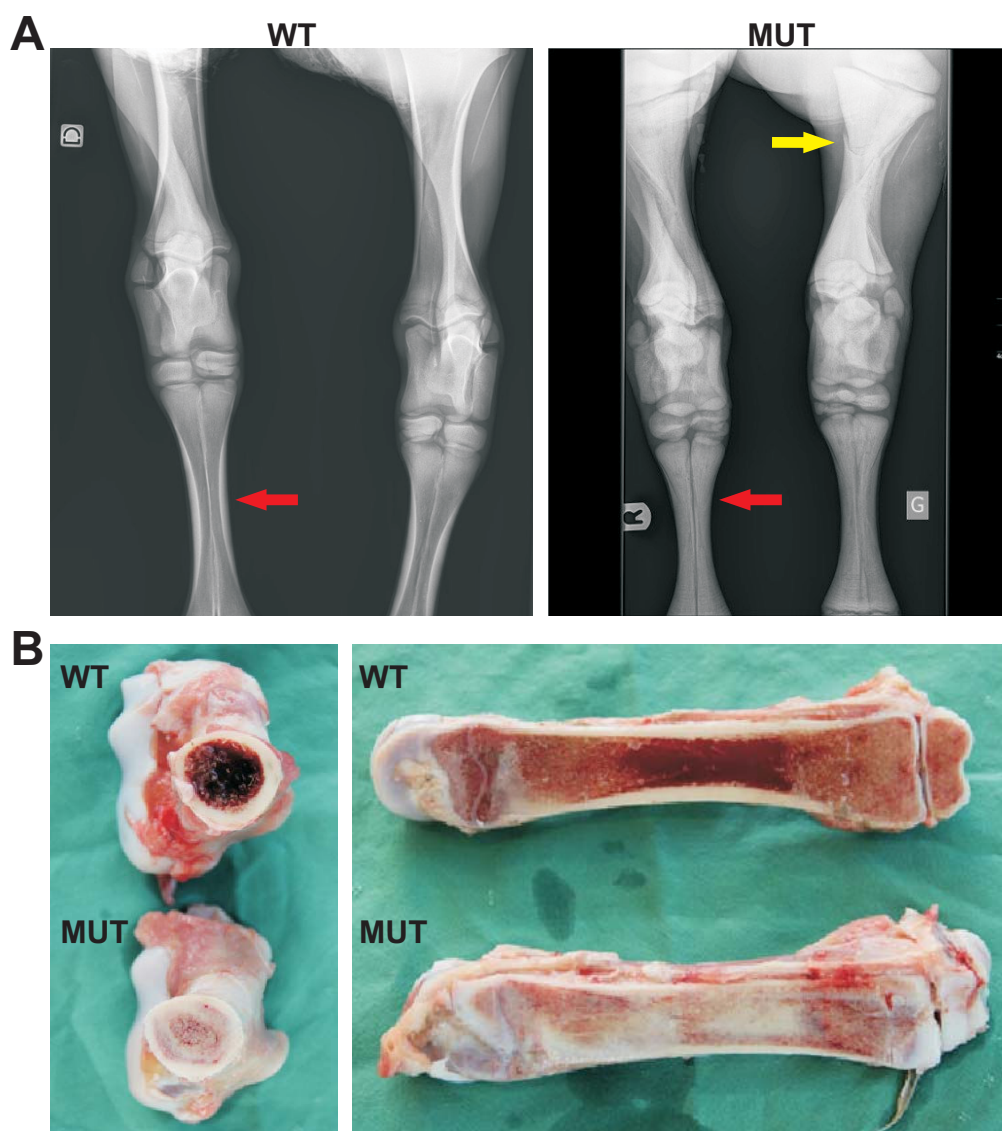


Figure 5: Lysosomal storage in a homozygous CLC-7 (Y750Q) calf. Lysates (60 µg protein per lane) of cerebellum (**A-C**) and kidney (**D**) from a control calf (WT) and from a calf homozygous for the Y750Q mutation (MUT) were analyzed by immunoblot against the indicated proteins. Immunoblotting for α -tubulin or actin served as loading control. **A.** Levels of subunit c of the mitochondrial ATP synthase were increased in the mutated calf. **B.** The animals displayed no differences in the levels of the preprocathepsin D (immature) or the intermediate and mature forms of cathepsin D. **C.** Levels of the astrocyte marker protein GFAP were only moderately increased in the mutated calf. **D.** The autophagic marker LC3-II was increased in the mutated calf whereas overall LC3 levels were unchanged.

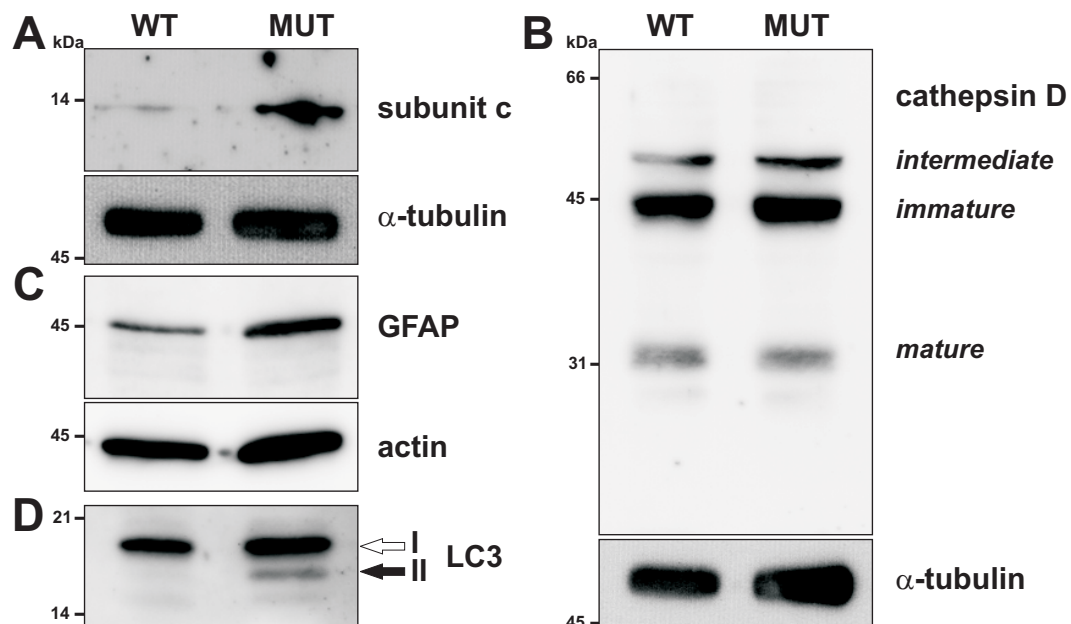


Figure 6: Protein levels of ClC-7 and Ostm1. Membrane protein-enriched lysates (80 µg per lane) of kidney from a control calf (WT) and a calf homozygous for the Y750Q mutation (MUT), as well as from a *Clcn7*^{-/-} mouse (KO) and its wild-type littermate (WT) were analyzed by Western blot for ClC-7 (**A**) and Ostm1 (with an antibody directed against a C-terminal epitope that also recognizes the proteolytically processed transmembrane fragment) (**B**). Immunoblotting for α-tubulin served as loading control. Lack of ClC-7 signal in lysate from the *Clcn7*^{-/-} mouse shows specificity of the antibody. Ostm1, which migrates with an apparently higher molecular weight in the bovine samples, exists predominantly in its proteolytically processed form (~30 kDa). The different sizes may be owed to species-specific differences in glycosylation.

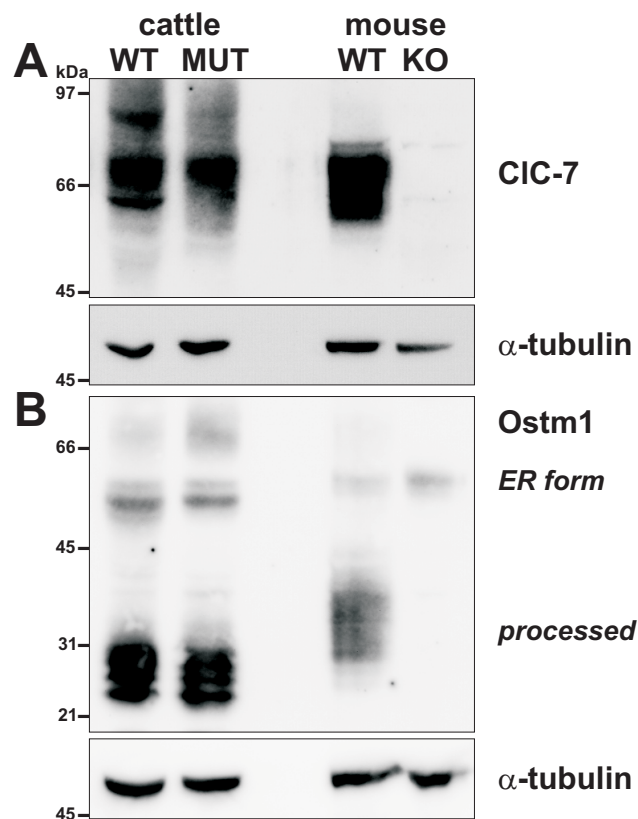


Figure 7: Correct subcellular targeting of ClC-7/Ostm1 upon heterologous expression. **A.** After transient transfection with rat ClC-7, either WT or the Y744Q mutant (MUT), HeLa cells were immunolabelled for ClC-7 (green in overlay) and the late endosomal/lysosomal marker protein LAMP-1 (red). Blue in overlay, DAPI-staining of nuclei. **B.** HeLa cells cotransfected with hClC-7(Y746Q) and mOstm1-GFP (green in overlay) were immunolabelled for ClC-7 (red) and the late endosomal/lysosomal marker protein LAMP-2 (blue).

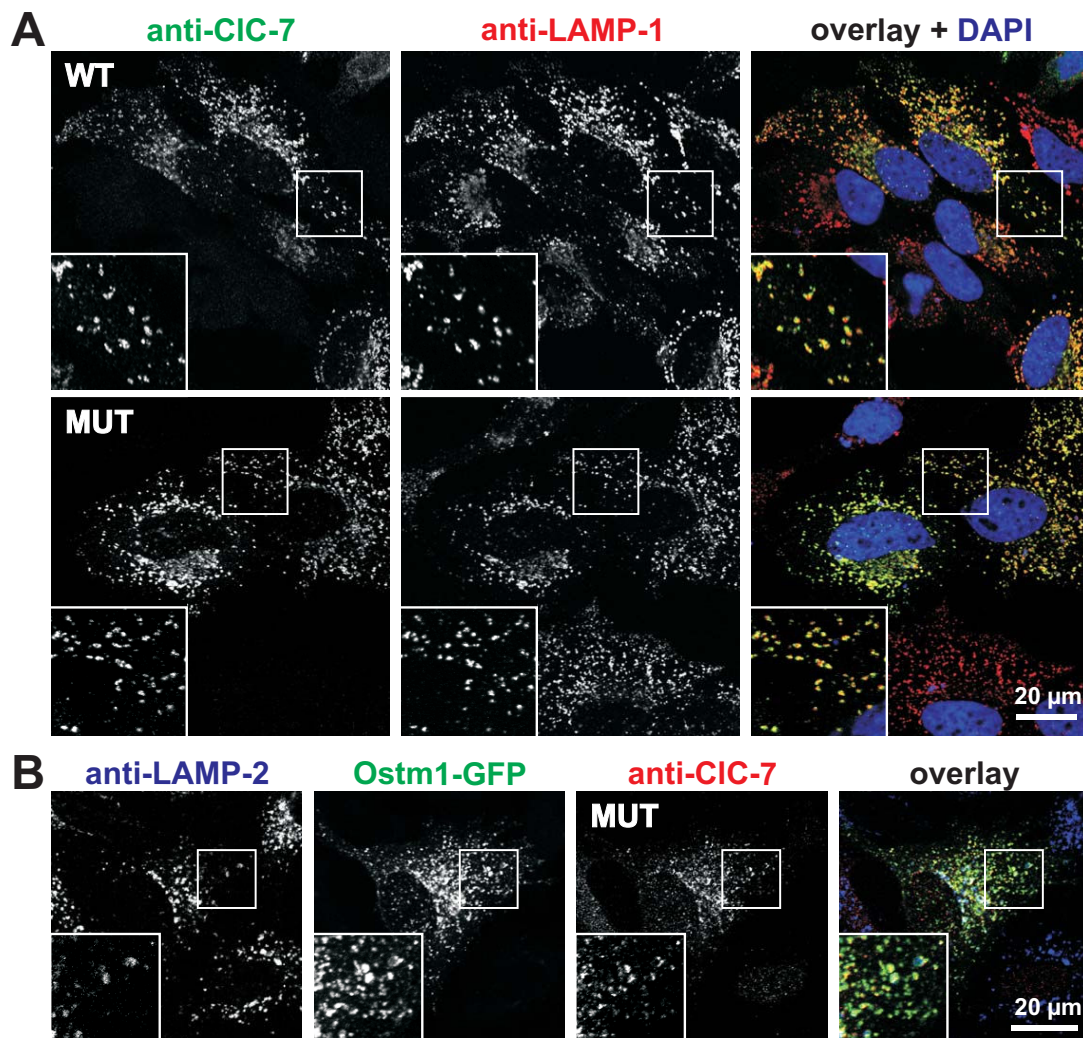


Figure 8: Accelerated gating of ClC-7/Ostm1 by the disease-causing mutation. *Xenopus* oocytes coexpressing Ostm1 and partially plasma membrane-localized hClC-7^{PM}, either without further mutation ('WT') or carrying the Y746Q mutation (MUT), were recorded in two-electrode voltage clamp (clamp protocol in inset). **A.** Current traces (representative for three batches of oocytes) showed that the activation and relaxation kinetics were accelerated by the Y746Q mutation. **B.** Mean currents after 2 s were normalized to the current at +80 mV and plotted as function of voltage. Values are the mean of 13 ('WT') and 18 (MUT) oocytes from three batches of oocytes, small error bars (s.e.m.) are hidden behind the symbol.

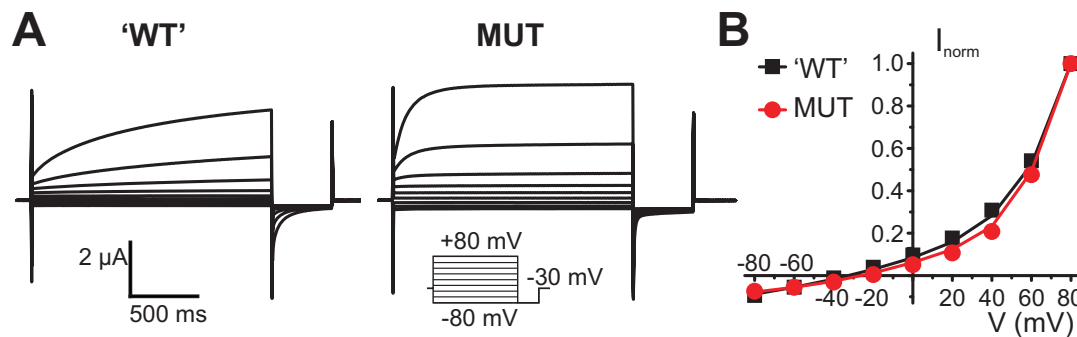


Table S1. Primer pairs for gDNA and cDNA.

Genomic DNA		
Primer name	5'-3' sequence	Size
gUP1	CATCGTCCTACTCAAGCACAAG	666 bp
gDN1	CCCCCTTTCCAAGCCGGTAC	
cDNA		
Primer name	5'-3' sequence	Size
cUP1	CCAACGTTTCCAAGAAGGTGTC	541 bp
cDN1	AGCAGGAGGGAGAAGGACAGTC	
cUP2	CGTGGCCTGCTTCATCGACATC	1088 bp
cDN2	AGGCCAGGAAGAAGTAGACCAG	
cUP3	TGATGGGGAGTACAACCTCGATG	1256 bp
cDN3	GCAGTTGCACAGATTTCCTAAAG	

Table S2. Summary of clinical symptoms. The clinical symptoms for the 63 cases are depicted in this table. A '-' indicates that no data are available.

Case number	Gender	Days before term	Size at birth	Alive/dead at birth	Abnormal skull shape	Inferior brachygnathia	Hamartome (diameter)	Protruding tongue	Blindness	Abdominal hydrops	Liver hypertrophy	Kidney hypertrophy	Hydramnios	Osteopetrosis
1	F	12	small	dead	yes	yes	3 cm	yes	-	yes	yes	yes	yes	-
2	F	0	normal	dead	yes	yes	4 cm	-	-	yes	yes	yes	yes	-
3	M	0	normal	dead	yes	yes	10 cm	-	-	yes	yes	no	yes	-
4	F	15	small	dead	yes	yes	10 cm	-	-	yes	yes	yes	yes	-
5	M	30	small	dead	yes	yes	7 cm	-	-	yes	yes	yes	yes	-
6	F	0	small	alive	yes	yes	5 cm	no	yes	no	no	no	no	-
7	F	0	normal	alive	yes	yes	2 cm	no	yes	no	no	no	no	-
8	F	0	normal	alive	yes	yes	3 cm	yes	yes	no	no	no	no	-
9	F	-	-	dead	yes	yes	yes	-	-	-	-	-	-	-
10	-	0	small	dead	yes	yes	5 cm	-	-	yes	yes	yes	yes	-
11	F	0	small	dead	yes	yes	12 cm	-	-	yes	yes	yes	yes	-
12	F	0	small	alive	yes	yes	no	yes	yes	no	no	no	no	-
13	M	0	normal	alive	yes	yes	no	yes	yes	-	-	-	-	-
14	M	31	small	dead	yes	yes	10 cm	-	-	yes	yes	yes	yes	-
15	F	0	normal	alive	yes	yes	2 cm	yes	yes	no	-	-	no	-
16	F	15	small	dead	yes	yes	10 cm	-	-	yes	yes	yes	yes	-
17	M	0	normal	alive	yes	yes	12 cm	yes	yes	no	no	no	yes	-
18	F	16	small	dead	yes	yes	6 cm	-	-	yes	yes	yes	yes	-
19	M	0	small	alive	yes	yes	no	yes	yes	no	-	-	no	-
20	M	-	small	dead	yes	yes	7 cm	-	-	yes	yes	yes	yes	-
21	M	-	small	dead	yes	yes	2 cm	-	-	yes	yes	yes	yes	-
22	F	-	small	dead	yes	yes	10 cm	-	-	yes	yes	yes	yes	-
23	F	-	-	dead	yes	yes	yes	-	-	yes	-	-	yes	-
24	M	0	normal	dead	yes	yes	3 cm	-	-	yes	yes	no	no	-
25	M	60	small	dead	yes	yes	5 cm	-	-	yes	yes	yes	yes	-
26	M	0	normal	dead	yes	yes	yes	-	-	-	-	-	yes	-
27	M	0	normal	dead	yes	yes	yes	-	-	-	-	-	-	-
28	F	-	-	dead	yes	yes	yes	-	-	-	-	-	-	-
29	M	30	small	dead	yes	yes	yes	-	-	yes	-	-	yes	-
30	F	21	small	dead	yes	yes	4 cm	-	-	yes	yes	yes	yes	-
31	M	0	normal	alive	yes	yes	3 cm	-	yes	no	-	-	no	-
32	M	-	-	dead	yes	yes	yes	-	-	-	-	-	-	-
33	F	7	small	dead	yes	yes	10 cm	-	-	yes	yes	yes	yes	-
34	M	0	normal	dead	yes	yes	yes	-	-	yes	-	-	yes	-
35	M	-	-	dead	yes	yes	yes	-	-	-	-	-	-	-
36	M	0	normal	alive	yes	yes	no	-	yes	no	-	-	no	-
37	M	15	small	dead	yes	yes	1 cm	-	-	yes	-	-	yes	-

38	-	0	normal	alive	yes	yes	no	no	yes	no	-	-	no	-
39	-	0	normal	alive	yes	yes	no	yes	yes	no	-	-	no	-
40	-	-	-	dead	yes	yes	yes	-	-	-	-	-	-	-
41	F	30	small	dead	yes	yes	4 cm	-	-	yes	yes	yes	yes	-
42	-	-	-	dead	yes	yes	yes	-	-	-	-	-	-	-
43	-	0	normal	dead	yes	yes	yes	-	-	no	-	-	no	-
44	-	0	small	dead	yes	yes	10 cm	-	-	yes	yes	no	yes	-
45	M	0	normal	alive	yes	yes	no	-	yes	no	-	-	no	-
46	-	-	-	dead	yes	yes	yes	-	-	-	-	-	-	-
47	-	0	normal	alive	yes	yes	no	yes	yes	no	-	-	no	-
48	-	-	-	dead	yes	yes	yes	-	-	-	-	-	-	-
49	F	0	small	dead	yes	yes	10 cm	-	-	no	-	-	no	-
50	M	0	normal	dead	yes	yes	2 cm	-	-	no	-	-	no	-
51	M	30	small	dead	yes	yes	5 cm	-	-	yes	-	-	yes	-
52	-	-	-	dead	yes	yes	yes	-	-	-	-	-	-	-
53	-	-	-	dead	yes	yes	yes	-	-	-	-	-	-	-
54	M	0	normal	alive	yes	yes	no	yes	yes	no	-	-	no	-
55	M	0	normal	dead	yes	yes	yes	-	-	yes	-	-	yes	-
56	F	0	normal	dead	yes	yes	yes	-	-	yes	-	-	yes	-
57	F	0	normal	dead	yes	yes	yes	-	-	yes	-	-	yes	-
58	M	0	normal	alive	yes	yes	yes	-	yes	no	-	-	no	-
59	M	0	normal	dead	yes	yes	yes	-	-	yes	-	-	yes	-
60	F	0	normal	alive	yes	yes	yes	yes	yes	no	yes	no	no	yes
61	M	0	small	dead	yes	yes	no	-	-	-	yes	-	-	yes
62	M	0	normal	alive	yes	yes	no	yes	yes	no	yes	no	no	yes
63	F	0	normal	dead	yes	yes	no	yes	-	yes	yes	no	yes	yes

Fig. S1. Insert size histograms for the paired-end libraries corresponding to the homozygous mutant calves.

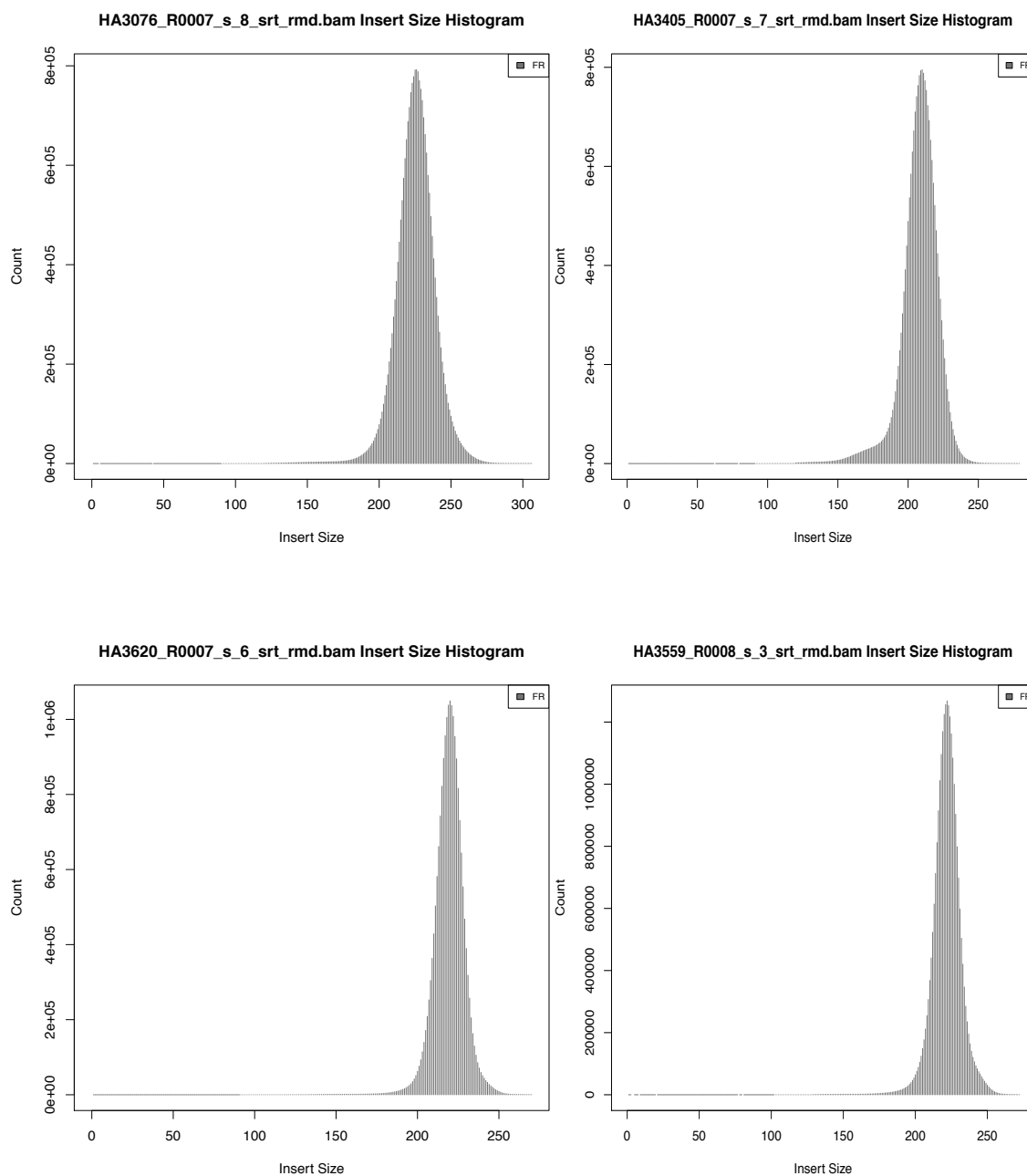


Fig. S2. Identification of three private nucleotide substitutions in *CLCN7* exon 23 from whole genome sequencing of cases. **A.** Screen capture of an IGV output displaying the sequence reads from cases and controls. **B.** Schematic representation of the bovine *CLCN7* genomic organization (adapted from UCSC browser) highlighting exon 23.

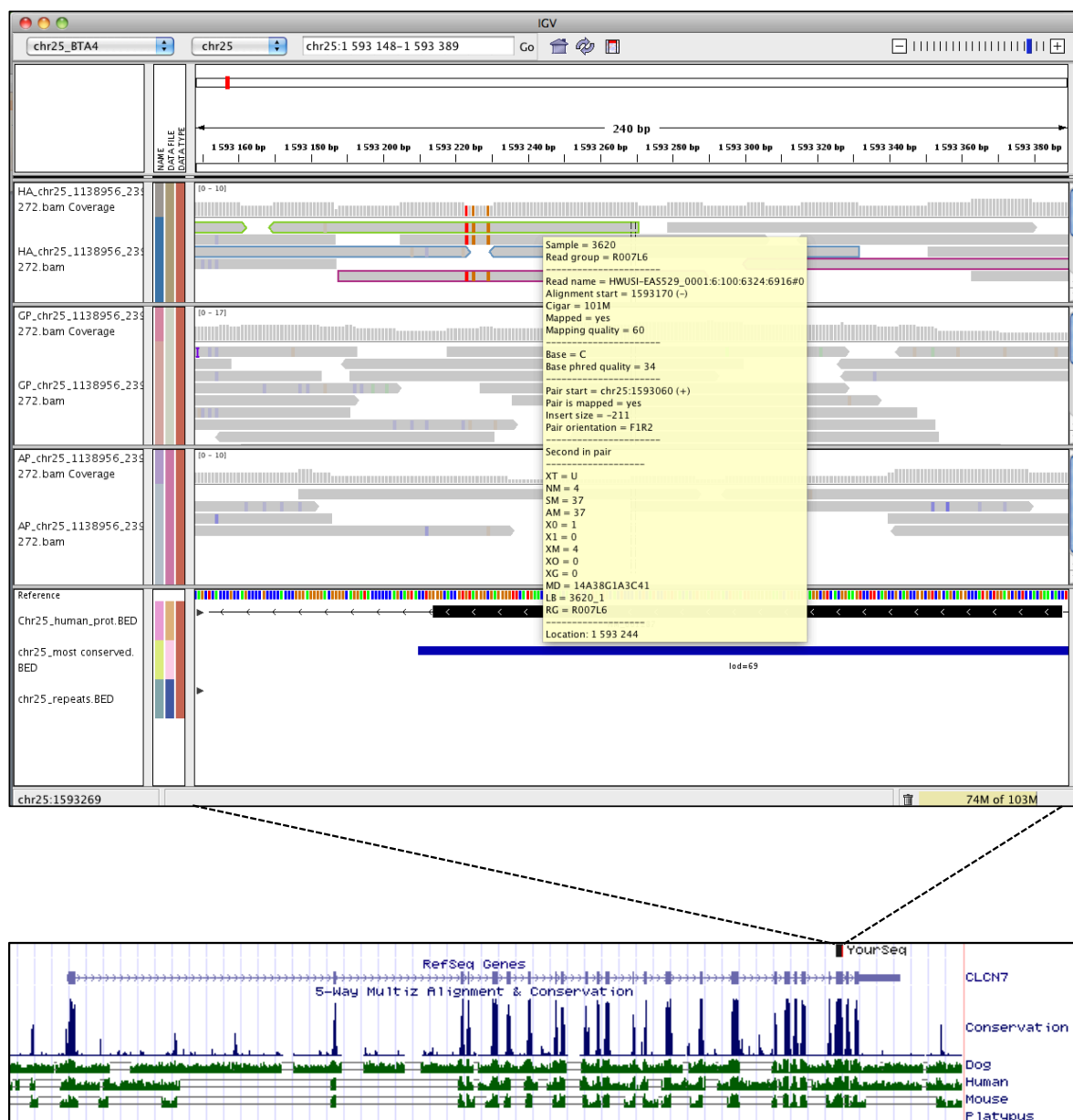
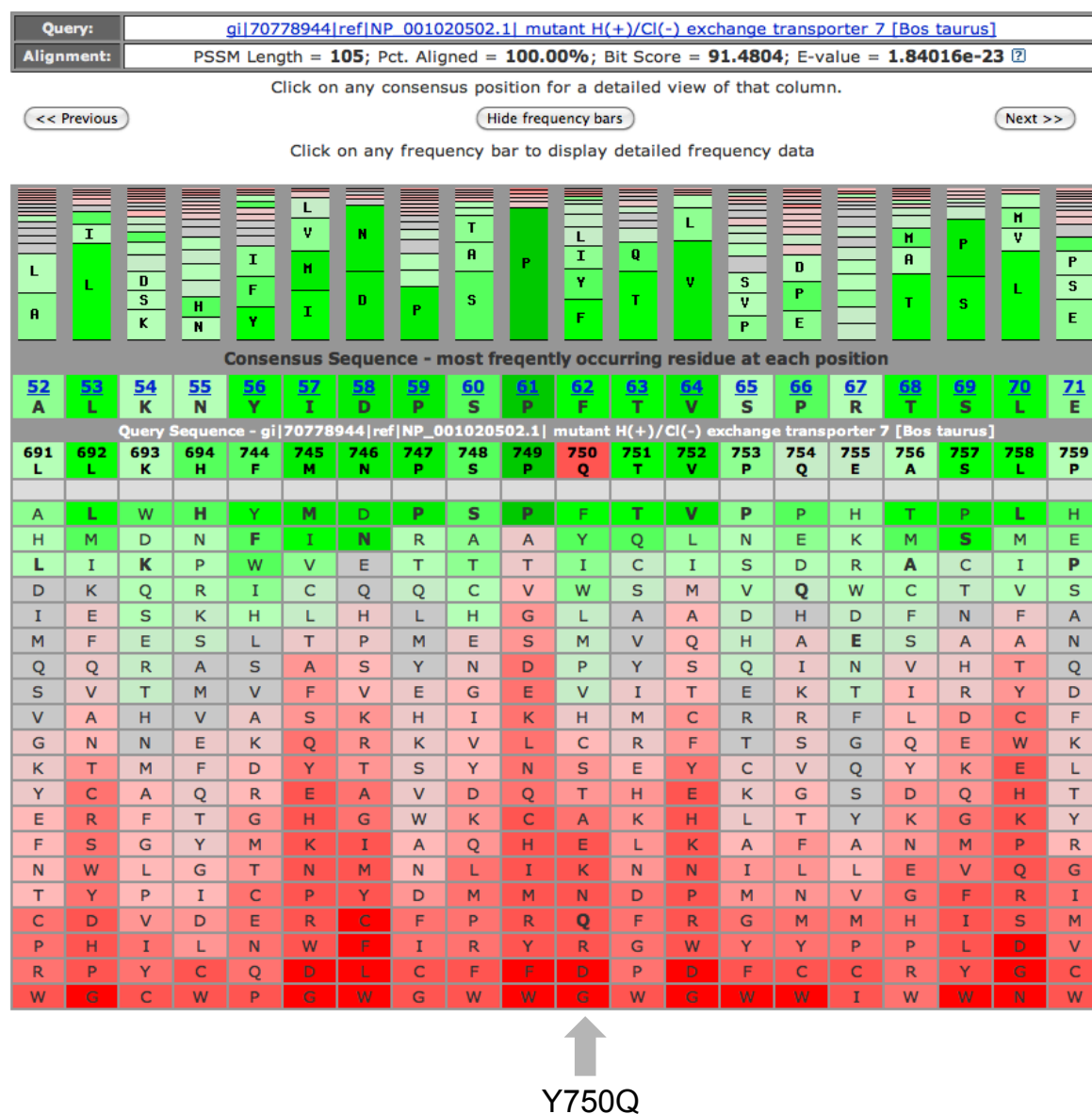


Fig. S3. Position-Specific Scoring Matrix (PSSM) output for the CBS2 domain of the mutant bovine CIC-7. Alignment stack of the cd04591 domain (containing two tandem repeats of the cystathionine beta-synthase (CBS pair) domains in the EriC CLC-type chloride channels in eukaryotes and bacteria) for a 20 amino acid sequence around the Y750Q mutation in CBS2; a highly negative score (dark red) is obtained when the mutant Q750 is substituted for the wild-type Y750 [24].



5- IDENTIFICATION DE LA MUTATION CAUSALE, DANS LE GENE *PIGH*, RESPONSABLE D'UNE DEFICIENCE EN GLYCOSYLPHOSPHATIDYL INOSITOL DU SYNDROME D'ARTHROGRYPOSE LETALE.

A. Résumé



En 2009, la cellule d'hérédo-surveillance a été contactée par des éleveurs et des vétérinaires à propos de l'émergence de veaux nouveau-nés présentant une forme d'arthrogrypose généralisée, létale avec de la macroglossie. L'examen nécropsique de 25 cas a mis en évidence de l'arthrogrypose généralisée associée à de la macroglossie et un défaut d'éruption dentaire et dans une proportion variable de cas, une fente palatine et/ou labiale, un omphalocèle et une opacification de la cornée. Une hydropisie des membranes fœtales a été observée chez certaines mères, elle peut conduire à une réforme précoce, augmentant les pertes liées à ce défaut. Sur base de l'analyse des pedigrees de ces veaux, un mode de transmission autosomale récessive a été suspecté. *Kalimine du Barsy Fontaine*, un taureau d'IA né en 1980, est l'ancêtre commun le plus récent vers lequel on peut remonter par les côtés maternel et paternel.

Une région IBD de 2,2 Mb, associée au syndrome d'arthrogrypose, a été localisée sur le chromosome 10 à l'aide d'une étude d'association basée sur les haplotypes. Aucun gène candidat, sur la vingtaine présents dans l'intervalle, ne s'est révélé potentiellement intéressant. Quatre cas sélectionnés ont fait l'objet d'un séquençage à haut débit afin d'obtenir une couverture complète de la région IBD. De la trentaine de polymorphismes privés identifiés chez les cas, aucune mutation délétère n'a été identifiée. Afin d'élucider le déterminisme génétique de l'arthrogrypose, des embryons ont été générés à partir du croisement entre un taureau porteur de l'haplotype associé à ce défaut et une vache non porteuse. Le séquençage de l'ARN à haut débit d'un embryon porteur a identifié une substitution (*211-10C>G*) au niveau d'un site accepteur d'épissage du premier intron du gène *phosphatidylinositol glycan anchor biosynthesis class H (PIGH)*. Le séquençage du cDNA d'individus mutants a confirmé la non inclusion de l'exon 2 sans décalage de la phase de lecture. La protéine mutante qui en découle est 40 % plus courte et a perdu son domaine central. La protéine PIGH est un membre essentiel du complexe enzymatique et intervient dans l'étape initiale de la biosynthèse du glycosylphosphatidylinositol (GPI) caractéristique d'un sous-ensemble spécifique de protéines attachées à la membrane plasmique. Il existe, chez les mammifères, une liste de plus de 130 protéines avec ancrage GPI,

caractérisées et dépendantes de la biosynthèse de cette ancre. Beaucoup d'entre-elles sont des récepteurs ou des molécules d'adhésion et il semble que la possession d'une ancre GPI oriente leur tri et la polarisation de leur localisation membranaire finale. Des mutations affectant les autres membres du complexe sont connues pour engendrer de la mortalité embryonnaire chez l'homme et la souris. De plus, plusieurs mutations affectant des protéines ancrées au moyen du GPI sont responsables de malformations crânio-faciales, squelettiques et oculaires. De la macroglossie et des hernies ombilicales sont également présentes dans le syndrome d'arthrogrypose létal bovin, ségrégeant en BBB.

En décembre 2009, une sélection assistée par marqueurs, basée dans un premier temps sur l'haplotype, puis depuis octobre 2012 sur la mutation, a été intégrée dans les schémas de sélection, ce qui permet d'éviter les accouplements à risques. La fréquence initiale était de 5 %. La conséquence immédiate est une disparition rapide et quasi complète de veaux mutants, ainsi qu'une diminution lente et constante des porteurs.

B. Article

Genome-wide next-generation DNA and RNA sequencing reveals a mutation that perturbs splicing of the phosphatidylinositol glycan anchorbiosynthesis class H gene (*PIGH*) and causes arthrogryposis in Belgian Blue cattle.

Sartelet A.*, Li W.*, Pailhoux E., Tamma N., Karim L., Fasquelle C.,
Druet T., Coppieters W., Georges M., Charlier C.

In preparation

* Contributed equally

SHORT COMMUNICATION

Genome-wide next-generation DNA and RNA sequencing reveals a mutation that perturbs splicing of the phosphatidylinositol glycan anchor biosynthesis class H gene (*PIGH*) and causes arthrogryposis in Belgian Blue cattle.

Arnaud Sartelet^{1}, Wanbo Li^{1*}, Eric Pailhoux², Nico Tamma¹, Latifa Karim¹, Corinne Fasquelle¹, Tom Druet¹, Wouter Coppieters¹, Michel Georges¹, Carole Charlier^{1§}.*

¹Unit of Animal Genomics, GIGA-R & Department of Animal Sciences, Faculty of Veterinary Medicine, University of Liège, Belgium. ²INRA, UMR 1198 Biologie du Développement et Reproduction, Jouy-en-Josas, France. *Contributed equally

Running head: *PIGH* mutation and lethal GPI deficiency in cattle

Keywords: Arthrogryposis syndrome, *PIGH* gene, splice-site, glycosylphosphatidyl inositol deficiency, Belgian Blue Cattle breed

[§]Correspondence and requests for materials should be addressed to Carole Charlier, Unit of Animal Genomics, GIGA-R & Faculty of Veterinary Medicine, University of Liège (B34), 1 Avenue de l'Hôpital, 4000-Liège, Belgium.

Telephone: (32) 4 366 4126; FAX: (32) 4 366 4198;

E-mail: carole.charlier@ulg.ac.be

We herein report the use of haplotype-based association mapping to identify a locus on bovine chromosome 10 that underlies autosomal recessive arthrogryposis in Belgian Blue Cattle. We identify 31 candidate mutations by resequencing the genome of four cases and 15 controls at ~2.5-fold depth. By analyzing RNA-Seq data from a carrier fetus, we observe skipping of the second exon of the *PIGH* gene, which we confirm by RT-PCR to be fully penetrant in tissues from affected calves. We identify - amongst the 31 candidate variants - a C-to-G transversion in the first intron of the *PIGH* gene (211-10C>G) that is predicted to affect its acceptor splice-site. The resulting PIGH protein is predicted to be non-functional as it misses essential domains, and hence to cause arthrogryposis.

We recently established an “heredo-surveillance platform” to effectively identify and control inherited defects that recurrently emerge as a result of intensive selection in Belgian Blue and other cattle breeds (Charlier *et al.*, 2008 ; Sartelet *et al.*, 2012 a; Sartelet *et al.*, 2012b; Sartelet *et al.*, submitted for publication). Twenty-five Belgian-Blue cases of a new form of arthrogryposis were referred to this platform in 2009 alone. Affected calves were all characterized by generalized arthrogryposis (curved and hooked joints) of the four limbs, severe scoliosis (define) (Figure 1A), and a stocky head with macroglossy and impaired tooth eruption (Figure 1B & 1C). A majority of cases suffered from cleft palate (80 %) and lip (13 %) (Figure 1F), omphalocele (abdominal wall defect with a sac, which is formed from an outpouching of peritoneum, containing intestine; 75 %) (Figure 1D) and corneal clouding (83 %) (Figure 1E). Several dams developed metritis and peritonitis, caused by hydrops (accumulation of excessive fluid in the allantoic or amniotic space) of the fetal membranes due to impaired fetal swallowing.

The 25 cases traced back, on sire and dam side, to the artificial insemination (AI) sire *Kalimine du Barsy Fontaine*, suggesting autosomal recessive inheritance. We therefore genotyped 15 cases with the bovine SNP50 beadchip (Illumina, San Diego) to perform a genome-wide association study (GWAS). We used the

genotypes from 275 Belgian Blue AI sires, obtained with the bovine HD (700K) beadchip (Illumina, San Diego), as controls. The analysis was restricted to 34,368 SNPs shared by both arrays, and conducted with GLASCOW as described (Zhang *et al.*, 2012). This yielded a single genome-wide significant signal ($p = 10^{-40}$) on chromosome 10 (Figure 2A). It resulted from autozygosity of the 15 cases for a 2.2Mb (chr10: 78,424,435-80,602,211; BosTau6/UMD3) identical-by-descent haplotype, hence confirming the suspected mode of inheritance (Figure 2B).

The corresponding interval is collinear with a 2.4Mb segment of human chromosome 14 (chr14: 66,474,214-68,918,385) and encompasses 22 annotated genes. As none of these would be obvious candidates, we generated paired-end libraries for four cases and re-sequenced them to average 2.5-fold depth on a Illumina GAIIX instrument as described (Charlier *et al.*, 2012). Equivalent genomic sequences of 15 healthy Belgian Blues that did not carry the incriminated haplotype were used as controls. Sequence reads were mapped on the BosTau6/UMD3 reference genome using BWA (Li and Durbin, 2009), and variants called with SAMTools (Li *et al.*, 2009). We detected 2,968 variants with quality score > 100 mapping to the 2.2Mb interval. As expected, the five cases appeared homozygous for the 21 Beadchip SNPs defining the associated haplotype. Filtering against known bovine dbSNP SNPs and variants observed in at least one of the 15 controls (-2,442), and demanding homozygosity for the four cases, left us with only 31 candidate variants. Yet, none of these would alter coding sequences (missense or nonsense), or map from an exon-intron junction. These findings suggested that the causative mutation was either regulatory or affecting splicing otherwise. To pursue this hypothesis, we took advantage of available RNA-Seq data from liver and cerebral cortex of a 60-day pc Belgian Blue fetus shown by SNP genotyping to carry the arthrogryposis risk haplotype. Sequence reads were analyzed using TopHat and Cufflinks (Trapnell *et al.*, 2012) and predicted transcripts mapping to the arthrogryposis locus visualized in the IGV browser (Robinson *et al.*, 2011). We readily noticed skipping of the second exon of approximately half (20/45) of the *PIGH* (phosphatidylinositol glycan anchor biosynthesis class H) transcripts (Figure 3B). To verify whether this

observation might be related to the arthrogryposis condition we extracted RNA from skeletal muscle and kidney of an affected calve and an age-matched control, and performed RT-PCR using primers located in exon 1 and 3 of the *PIGH* gene (Figure 4A). While we obtained a unique band of expected 355-bp size for the control, the only band obtained from the case RNA was 145-bp (Figure 4B). Sequencing of the corresponding RT-PCR products revealed the expected *PIGH* exon 1-2-3 sequence for the controls, yet two distinct sequences for the case. The most abundant (~75%) form corresponded to the skipping of exon 2, while the minor (~25%) form was - in addition to missing exon 2 - devoid of the first AAG triplet of exon 3 (Figure 4C). Note the 3' AG end of this lysine codon susceptible to act as a cryptic acceptor splice site.

Interestingly, one of the 31 candidate mutations identified by whole genome sequencing is a C-to-G transversion located in the first intron of *PIGH*, 10-bp upstream of the exon 2 junction (*211-10C>G*) (Figure 3A). The corresponding residue is located in the consensus polypyrimidine track defining canonical "GT-AG" type acceptor splice sites (f.i. Burset *et al.*, 2001) (Supplemental Figure 1). We developed a 3'exonuclease assay to interrogate the *211-10C>G* variant and genotyped 25 cases, 21 parents and 6,987 healthy Belgian Blue animals. All cases were homozygous GG, and all parents CG as expected. 5 % of the healthy Belgian Blues were carriers, while none were homozygous GG ($p = 0.1$).

PIG-H is one of seven highly conserved (yeast to human) subunits of the complex catalyzing the first step in the biosynthesis of glycerophosphatidyl inositol (Watanabe *et al.*, 1998). Mutations in other complex components are known to cause embryonic lethality in human and mice. PIGH is anchored in the membrane of the endoplasmic reticulum (ER) by two trans-membrane helices (residues 38-55 and 60-78), connected by a short ER intra-luminal loop (56-59), and bounded by a short N-terminal (1-37) and a long C-terminal cytoplasmic tail (79-188) (Watanabe *et al.*, 1996). Deleting exon 2 generates a protein missing 70 amino-acids of the of the 188 aa total protein length, therefore unlikely to be properly anchored in the ER membrane and to be functional (Figure 4D).

Taken together, our data strongly suggest that the 211-10C>G variant is causing arthrogryposis in Belgian Blue Cattle, by inducing skipping of exon 2 resulting in the production of a severely shortened, non-functional PIGH.

Materials and Methods

Ethics statement. Biological samples were collected from sires, cows and calves, by trained veterinarians following standard procedures and relevant national guidelines. [Add Embryos collect protocol.]

Genotyping. Genomic DNA of cases was extracted from 350 µl of blood using the MagAttract DNA Blood Midi M48 Kit (Qiagen). Genomic DNA of controls was extracted from frozen semen using the MagAttract Mini M48 Kit (Qiagen). The 15 cases of the initial genome scan were genotyped using a custom-made 50K SNP array (Charlier *et al.*, 2008). The 275 control sires were genotyped with the BovineHD BeadChip (Illumina). SNP genotyping was conducted using standard procedures at the GIGA genomics core facility.

Genome-wide haplotype-based association studies. Phasing of the SNP genotypes and assignment of the haplotypes to a predetermined number of hidden haplotype states was conducted with PHASEBOOK (Druet and Georges, 2010). Hidden haplotype state-based association analysis was conducted using GLASCOW (Zhang *et al.*, 2012). GLASCOW uses generalized linear models and fits a random hidden haplotype state effect as well as a random polygenic effect to correct for population stratification. Locus-specific p-values were determined from 1,000 permutations assuming a gamma distribution of the used score test (Zhang *et al.*, 2012). We applied a conservative Bonferonni correction assuming 50,000 independent tests to determine the genome-wide significance thresholds.

Genome-wide sequencing of cases. Procedure for genome-wide sequencing has been described in details in Sartelet *et al.* (submitted for publication). Shortly, four affected individuals, homozygous for the defined IBD haplotype

were selected, as well as fifteen unrelated controls from the same breed. The Pair-End Library Prep Kit v2 from Illumina was used to generate a ~200-250 bp paired-end library from genomic DNA fragments for each animal. Each library was sequenced on one lane of the flow-cell of an Illumina GAIIx with the paired-end module to generate high-quality reads (2X110bp). Reads were mapped and analyzed with publicly available software: *Burrows-Wheeler Alignment Tool* (<http://bio-bwa.sourceforge.net>) and *Samtools* (<http://samtools.sourceforge.net>). The output files were readily uploaded in the *Integrative Genomics Viewer* (IGV; (Robinson *et al.*, 2011)) and visually scrutinized for private variation.

RNAseq data of heterozygous embryos. As part of a companion project, embryos were produced by directed mating between a sire carrier for the AP haplotype and a wild-type cow. A panel of tissues from embryos at 60 days post fertilization (dpf) was collected. Genomic DNA was extracted and genotyped with the indirect haplotype based test. Succinctly, total RNA was extracted from liver and cerebral cortex, enriched for polyA+ mRNA, cDNA TrueSeq libraries were prepared following manufacturer instructions and sequenced on a Illumina HiSeq instrument (two tissues pooled in one lane). Reads (2X110bp) were mapped and analyzed with the TopHat and Cufflinks open source software tools as described in Trapnell *et al.* (2012). Transcripts isoforms were qualitatively evaluated in IGV browser across the IBD interval.

Mutation validation at the DNA and RNA level. Primers respectively located in exon 1 and exon 3 of the *PIGH* gene encompassing the 211-10C>G mutation was used to amplify products from genomic DNA of homozygous mutant calves and a wild-type calf using. Amplicons were directly sequenced using the Big Dye terminator cycle sequencing kit (Applied Biosystems, Foster City, CA). Electrophoresis of purified sequencing reactions was performed on an ABI PRISM 3730 DNA analyzer (PE Applied Biosystems, Foster City, CA). Sequence traces were aligned and compared to bovine reference using the Phred/Phrap/Consed package (www.genome.washington.edu). Total RNA was extracted from skeletal muscle and kidney of one homozygous case and one unaffected unrelated individual using Trizol (Invitrogen). The obtained RNA was

treated with TurboDNaseI (Ambion) and cDNA was synthesized using SuperscriptTMIII First Strand Synthesis System for RT-PCR (Invitrogen). cDNA was amplified using primers located in exon 1 and exon 3 and the amplicons were directly sequenced and analyzed as describe above.

5' exonuclease diagnostic assay of the *PIGH* 211-10C>G splice acceptor site mutation. A 5' exonuclease assay was developed to genotype the -10C>G *PIGH* mutation, using 5'- ATG GCA GCA GAG AGG ATC ATG -3' and 5'- GGA GTT GAC TTA TTA ACC AGC AGA GA -3' as PCR primers, and 5'-TGT CTG GCT [C]TC TCT TC-3' (wild type C allele) and 5'-TCT GGC T[G]T CTC TTC -3' (mutant G allele) as probes (Taqman, Applied Biosystems, Fosters City, CA). Reactions were carried out on an ABI7900HT instrument (Applied Biosystems, Fosters City, CA) using standard procedures.

References

Burset M, Seledtsov IA, Solovyev VV. (2001) SpliceDB: database of canonical and non-canonical mammalian splice sites. *Nucleic Acids Res.*, 29(1): 255-259.

Charlier C, Coppieters W, Rollin F, Desmecht D, Agerholm JS, Cambisano N, Carta E, Dardano S, Dive M, Fasquelle C, Frennet JC, Hanset R, Hubin X, Jorgensen C, Karim L, Kent M, Harvey K, Pearce BR, Simon P, Tama N, Nie H, Vandeputte S, Lien S, Longeri M, Fredholm M, Harvey RJ, Georges M. (2008) Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nat. Genet.*, 40: 449-454.

Charlier C, Agerholm JS, Coppieters W, Karlskov-Mortensen P, Li W, de Jong G, Fasquelle C, Karim L, Cirera S, Cambisano N, Ahariz N, Mullaart E, Georges M, Fredholm M. (2012) A deletion in the bovine *FANCI* gene compromises fertility by causing fetal death and brachyspina. *PloS ONE*, 7(8):e43085.

Druet T., Georges M. (2010) A Hidden Markov Model Combining Linkage and Linkage Disequilibrium Information for Haplotype Reconstruction and Quantitative Trait Locus Fine Mapping. *Genetics*, 184 (3): 789-798.

Fasquelle C, Sartelet A, Li W, Dive M, Tamma N, Michaux C, Druet T, Huijbers IJ, Isacke CM, Coppieters W, Georges M, Charlier C. (2009) Balancing selection of a frame-shift mutation in the MRC2 gene accounts for the outbreak of the Crooked Tail Syndrome in Belgian Blue Cattle. *PLoS Genet* 5, e1000666.

Li H, Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14): 1754-60.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R and 1000 Genome Project Data Processing Subgroup. (2009) The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics*, 25 (16): 2078-2079.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, and Mesirov JP. (2011) Integrative genomics viewer. *Nat. Biotechnol.* 29 (1): 24-26.

Sartelet A., Druet T, Michaux C, Fasquelle C, Geron S, Tamma N, Zhang Z, Coppieters W, Georges M, Charlier C. (2012) A Splice Site Variant in the Bovine RNF11 Gene Compromises Growth and Regulation of the Inflammatory Response. *PLoS Genet* 8: e1002581.

Sartelet A, Kingbell P, Franklin CK, Fasquelle C, Geron S, Isacke CM, Georges M, Charlier C. (2011) Allelic heterogeneity of Crooked Tail Syndrome: result of balancing selection? *Anim. Genet.*, 43(5): 604-607.

Sartelet A, Stauber T, Coppieters W, Ludwig CF, Fasquelle C, Ahariz N, Cambisano N, Jentsch TJ, Charlier C. (2013) A missense mutation accelerating the gating of the lysosomal Cl⁻/H⁺-exchanger CLC-7/Ostm1 causes osteopetrosis with gingival

hamartomas in cattle. Disease Models and Mechanisms, submitted for publication.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc., 7(3): 562-578.

Watanabe R, Kinoshita T, Masaki R, Yamamoto A, Takeda J, Inoue N. (1996) PIG-A and PIG-H, which participate in glycosylphosphatidylinositol anchor biosynthesis, form a protein complex in the endoplasmic reticulum. J. Biol. Chem., 271(43): 26868-26875.

Watanabe R, Inoue N, Westfall B, Taron CH, Orlean P, Takeda J, Kinoshita T. (1998) The first step of glycosylphosphatidylinositol biosynthesis is mediated by a complex of PIG-A, PIG-H, PIG-C and GPI1. EMBO J., 17 (4): 877-885.

Zhang Z, Guillaume F, Sartelet A, Charlier C, Georges M, Farnir F, Druet T. (2012) Ancestral haplotype-based association mapping with generalized linear mixed models accounting for stratification. Bioinformatics, 28 (19): 2467-2473.

Figure 1: Clinical spectrum of arthrogryposis syndrome. A. Generalized arthrogryposis. B. Brachygnathism and macroglossy. C. Impaired tooth eruption. D. Omphalocele. E. Corneal clouding. F. Hard cleft palate.



Figure 2 : Genetic mapping of the disease-causing gene for arthrogryposis syndrome. **A.** Manhattan plot for the case-control GWAS study; grey and black alternates demarcate the 29 autosomes; **B.** Cases genotypes for 1,916 chromosome 10 SNP; homozygous genotypes are shown in yellow or orange, heterozygous genotypes in red; the 2.2Mb homozygosity region, identical by state among all cases, is highlighted in red.

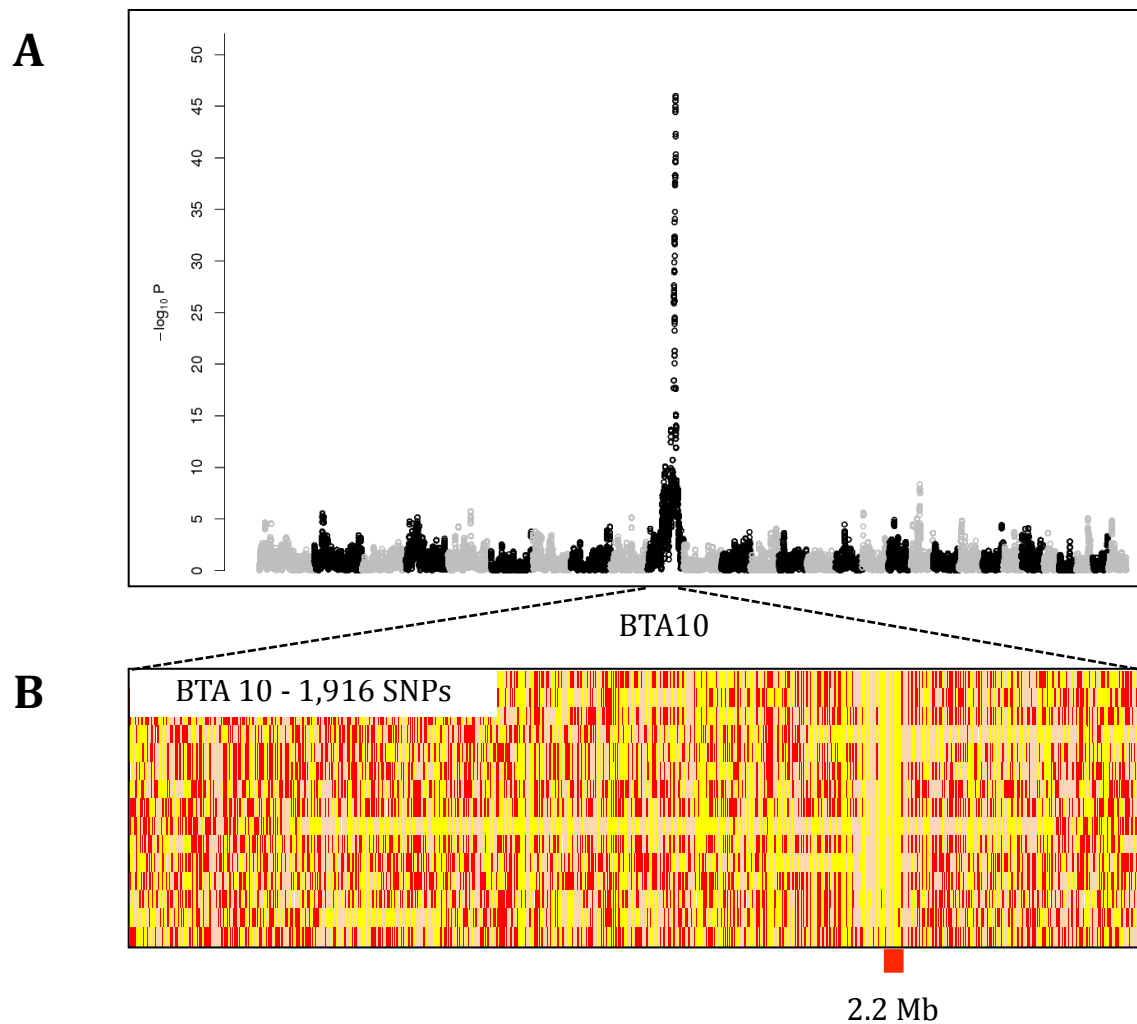


Figure 3 : Private intronic variation and alternative splicing at the *PIGH* locus.

A. Screen capture of an IGV output displaying, on the positive strand (top), a private G to C mutation (blue) from genomic DNA sequence reads of four pooled homozygous cases; on the negative strand (bottom), *PIGH* intron1/exon2 annotation highlighting the mutation position at -10 position of the splice acceptor sequence. **B.** Liver RNASeq data from a heterozygote mutant embryo aligned on the bovine genomic reference sequence at the *PIGH* locus (top), the four exons appear as stacks of grey reads and splicing is denoted by thin blue lines; a complete skipping of *PIGH* exon 2 in ~half of the transcripts is observable (red versus blue arrow height); below, *PIGH* intron/exon annotation with a recap of the two alternative splice events represented by dashed red (skipped) and blue (incorporated) lines.

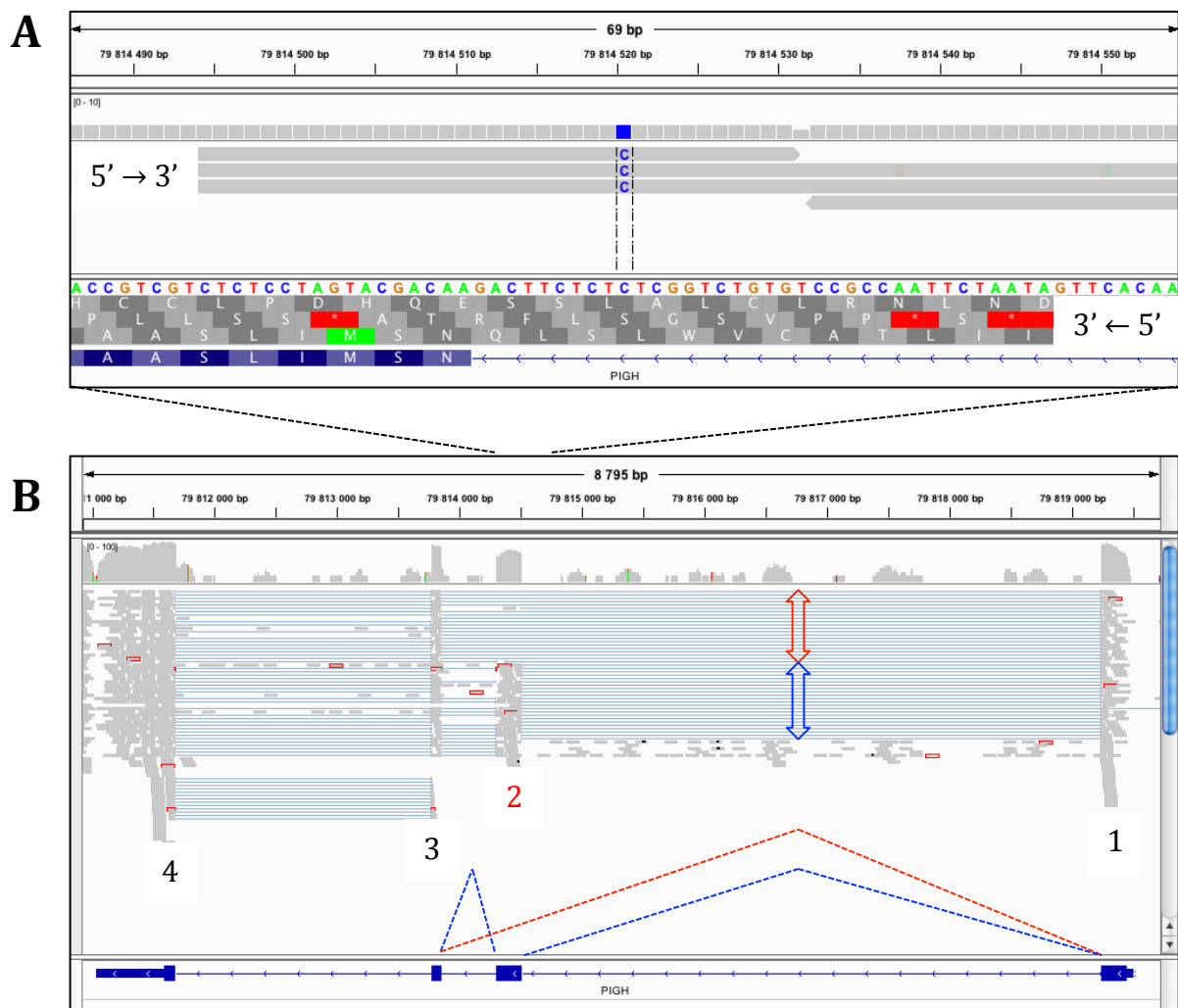
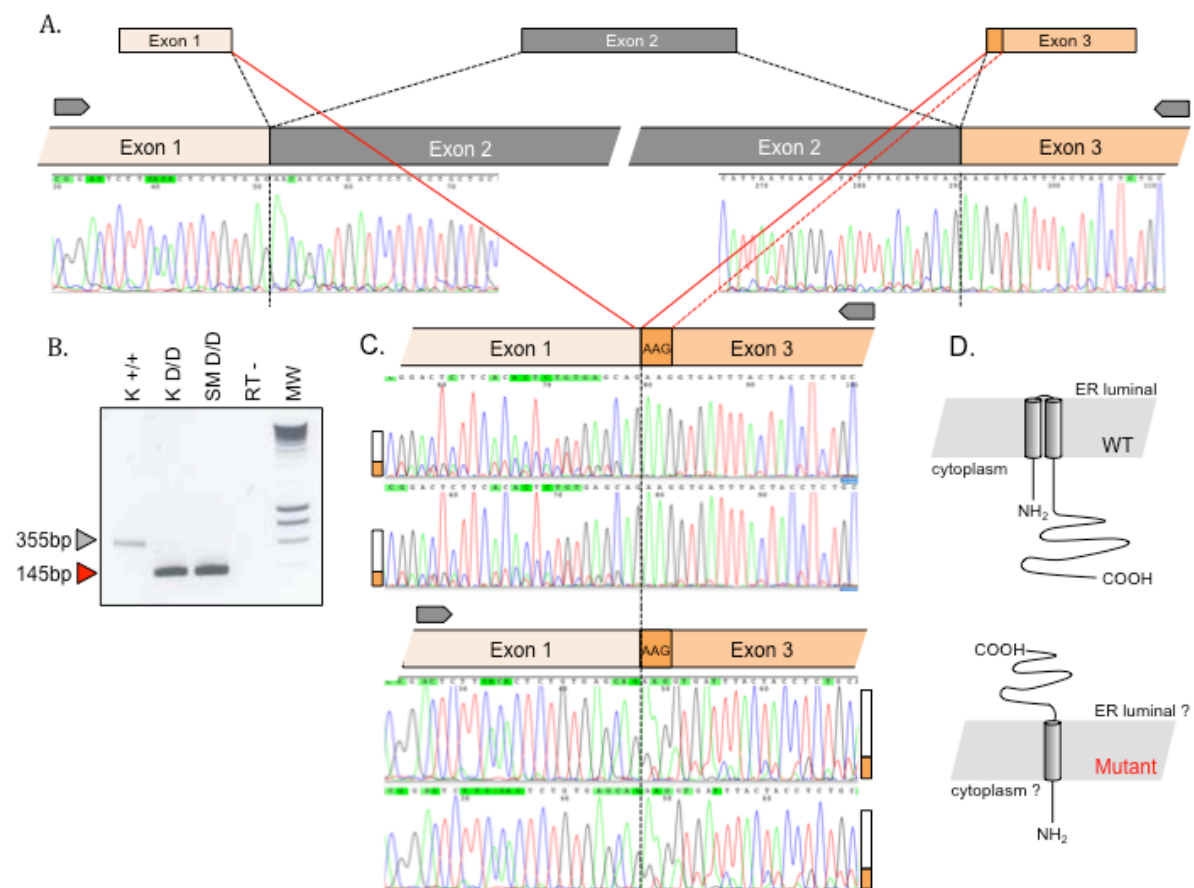


Figure 4 : Mutation effect at the mRNA and protein level. **A.** Partial schematic representation of the *PIGH* organization (genomic and mRNA) accompanied by wild-type cDNA sequence traces across exon1/exon2 and exon2/exon3 junctions obtained respectively with a forward primer and a reverse primer (grey arrows). **B.** Corresponding agarose gel showing cDNA amplification products from kidney (K) for a wild-type (+/+) animal, from kidney (K) and skeletal muscle (SM) for a mutant (D/D) and for a RT-minus control (RT-); MW : molecular weight marker. **C.** Mutant cDNA sequence traces obtained from the ~145bp amplification product with reverse (top) and forward (bottom) primers respectively; the cryptic “AAG” acceptor site in exon 3 is highlighted in dark orange; height of depicted white and orange bars represents the ratio between the two mutant splice forms, with total skipping of exon 2 in white and additional skipping of a bp triplet in orange. **D.** Protein topography for wild-type form (top) in the endoplasmic reticulum (ER) membrane (adapted from Tiede et al., 2000) and putative predicted mutant form (bottom).



Supplemental figure 1 : PIGH splice-acceptor site conservation through evolution.

	INTRON 1															EXON 2										
hg18	t	c	t	g	g	c	t	c	t	c	t	c	t	c	a	g	A	A	C	A	G	C	A	T	G	
panTro2	t	c	t	g	g	c	t	c	t	c	t	c	t	c	a	g	A	A	C	A	G	C	A	T	G	
gorGor1	t	c	t	g	g	c	t	c	t	c	t	c	t	c	a	g	A	A	C	A	G	C	A	T	G	
ponAbe2	t	c	t	g	g	c	t	c	t	c	t	c	t	t	c	a	g	A	A	C	A	G	C	A	T	G
rheMac2	t	c	t	g	g	c	t	c	t	c	t	c	t	t	c	a	g	A	A	C	A	G	C	A	T	G
calJac1	t	c	t	g	g	c	t	c	t	c	t	c	t	t	c	a	g	A	A	C	A	G	C	A	T	G
tarSyr1	t	c	t	g	g	c	t	c	t	-	-	c	t	t	c	a	g	A	A	C	A	G	C	A	T	G
tupBell1	t	c	t	g	g	c	t	c	t	c	t	c	t	t	c	a	g	A	A	C	A	G	C	A	T	G
dipOrd1	t	c	t	g	g	-	-	c	t	c	t	t	t	t	c	a	g	A	A	C	A	G	C	A	T	G
mm9	t	c	t	g	g	c	t	c	t	c	t	c	t	t	c	a	g	A	A	T	A	G	C	A	T	G
rn4	t	c	t	g	g	c	t	c	t	c	t	c	t	t	c	a	g	A	A	T	A	G	C	A	T	G
speTri1	t	c	t	g	g	c	t	c	t	c	t	c	t	t	c	a	g	A	A	C	A	G	C	A	T	G
oryCun1	t	c	t	g	g	c	t	c	t	g	t	c	t	t	c	a	g	A	A	C	A	G	C	A	T	G
vicPac1	t	c	t	g	g	c	t	c	t	c	t	c	t	t	c	a	g	A	A	C	A	G	C	A	T	G
bosTau4 WT	t	c	t	g	g	c	t	c	t	c	t	c	t	t	c	a	g	A	A	C	A	G	C	A	T	G
bosTau4 MUT	t	c	t	g	g	c	t	g	t	c	t	c	t	t	c	a	g	A	A	C	A	G	C	A	T	G
turTru1	t	c	t	g	g	c	t	c	t	-	-	c	t	t	c	a	g	A	A	C	A	G	C	A	T	G
equCab2	t	c	t	g	g	c	t	c	t	c	t	c	t	t	t	a	g	A	A	C	A	G	C	A	T	G
pteVam1	t	c	t	g	g	c	t	c	t	c	t	c	t	t	c	a	g	A	A	C	A	G	C	A	T	G
myoLuc1	t	c	t	g	g	c	t	c	t	c	t	c	t	t	c	a	g	A	A	C	A	G	C	A	T	G
sorAral	t	c	t	g	c	t	-	c	t	c	c	t	t	c	c	a	g	A	A	C	A	G	C	A	T	G
eriEur1	t	t	t	g	g	c	t	c	t	-	-	t	t	t	c	a	g	A	A	C	A	G	C	A	T	G
felCat3	t	c	t	g	g	c	t	c	t	c	t	c	t	t	c	a	g	A	A	C	A	G	C	A	T	A
canFam2	t	c	t	g	g	c	t	c	t	c	t	c	t	t	c	a	g	A	A	C	A	G	C	A	T	G
cavPor3	t	c	t	g	g	c	t	c	t	t	t	c	t	t	c	a	g	A	A	C	A	G	C	A	C	G
ochPri2	t	c	c	g	g	c	t	t	t	g	c	c	t	t	c	a	g	A	A	C	A	G	C	A	T	G
micMur1	t	c	t	g	g	c	t	c	t	c	t	c	t	t	c	a	g	A	A	C	A	G	C	A	T	G
otoGar1	t	c	t	g	g	c	t	c	t	c	t	c	t	t	c	a	g	A	A	C	A	G	C	A	T	G
proCap1	t	c	t	g	g	c	t	c	t	c	t	c	t	t	c	a	g	A	A	C	A	G	C	A	T	G
*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	

consensus splice site

	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	A	G
	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1
% of A	15	6	15	11	19	12	3	10	25	4	100	0
% of U	53	60	49	49	45	45	57	58	29	31	0	0
% of C	21	24	30	33	28	36	36	28	22	65	0	0
% of G	10	10	6	7	9	7	7	5	24	1	0	100

Y ₇₃ Y ₇₅ Y ₇₈ Y ₇₉ Y ₈₀ Y ₇₉ Y ₇₈ Y ₈₁ Y ₈₆ Y ₈₆ NC ₇₁ AG ₇₁ G ₅₂															
	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	1
A	9.0	8.4	7.5	6.8	7.6	8.0	9.7	9.2	7.6	7.8	23.7	4.2	100	0.0	23.9
C	31.0	31.0	30.7	29.3	32.6	33.0	37.3	38.5	41.0	35.2	30.9	70.8	0.0	0.0	13.8
G	12.5	11.5	10.6	10.4	11.0	11.3	11.3	8.5	6.6	6.4	21.2	0.3	0.0	100	52.0
U	42.3	44.0	47.0	49.4	47.1	46.3	40.8	42.9	44.5	50.4	24.0	24.6	0.0	0.0	10.4

6- SELECTION ASSISTEE PAR MARQUEURS POUR ERADIKER LA GESTATION PROLONGEE.

A. Résumé



Les premiers cas de gestation prolongée (GP) en race BBB ont émergé dans les années 90' mais c'est seulement en 2008 que l'autopsie de deux veaux nés après 340 jours de gestation a identifié une hypoplasie de l'hypophyse comme étant à l'origine de ce défaut du part (Cornillie *et al.*, 2008). Entre 2007 et 2008, cinquante veaux nés après une gestation prolongée ont été renseignés, collectés et autopsiés, confirmant par là les observations précédentes. Lorsque la gestation prolongée est confirmée, la césarienne réalisée de manière élective, suite ou non à une induction du part, ou suite à l'« avortement » est une cause majeure de complications post-césarienne. Ces complications, telle qu'une rétention d'arrière-faix, une métrite et une péritonite, fréquemment rencontrées, peuvent compromettre l'avenir de la mère et augmenter les pertes économiques liées à la GP.

Un mode de transmission autosomale récessive a été suspecté sur base de l'analyse épidémiologique, avec des pedigrees remontant par les voies paternelle et maternelle à un ancêtre commun, *Christophe de l'Abbaye*, taureau d'IA né en 1979.

L'étude d'association basée sur les haplotypes réalisée à partir de 32 cas génotypés à l'aide de puces de moyenne densité (50.000 SNPs) a permis de localiser un locus associé à la gestation prolongée sur le chromosome 18 (Zhang *et al.*, 2012) (Figure 22A). A l'examen direct des SNPs de ce chromosome, une région IBD homozygote partagée par tous les cas de 1,6 Mb (Figure 22B) est observée. Elle contient environ 70 gènes sans candidat majeur. Un séquençage génome entier d'homozygotes mutants a été réalisé. Aucune mutation de type "perte de fonction" claire n'a pu être identifiée à ce jour. Des analyses transcriptomiques, à partir d'embryons, issus d'un accouplement porteur-porteuse récoltés à différents stades de gestation sont en cours.

L'information générée par ces différentes analyses a été utilisée pour le développement d'un test indirect performant, basé sur les haplotypes. La fréquence de l'haplotype dans une population BBB d'environ 1.500 individus est de 11,4 %. L'utilisation de ce test indirect a rapidement fait disparaître les cas de GP, justifiant ainsi leur utilisation dans le programme de sélection en attendant l'identification et/ou la validation de la mutation causale.

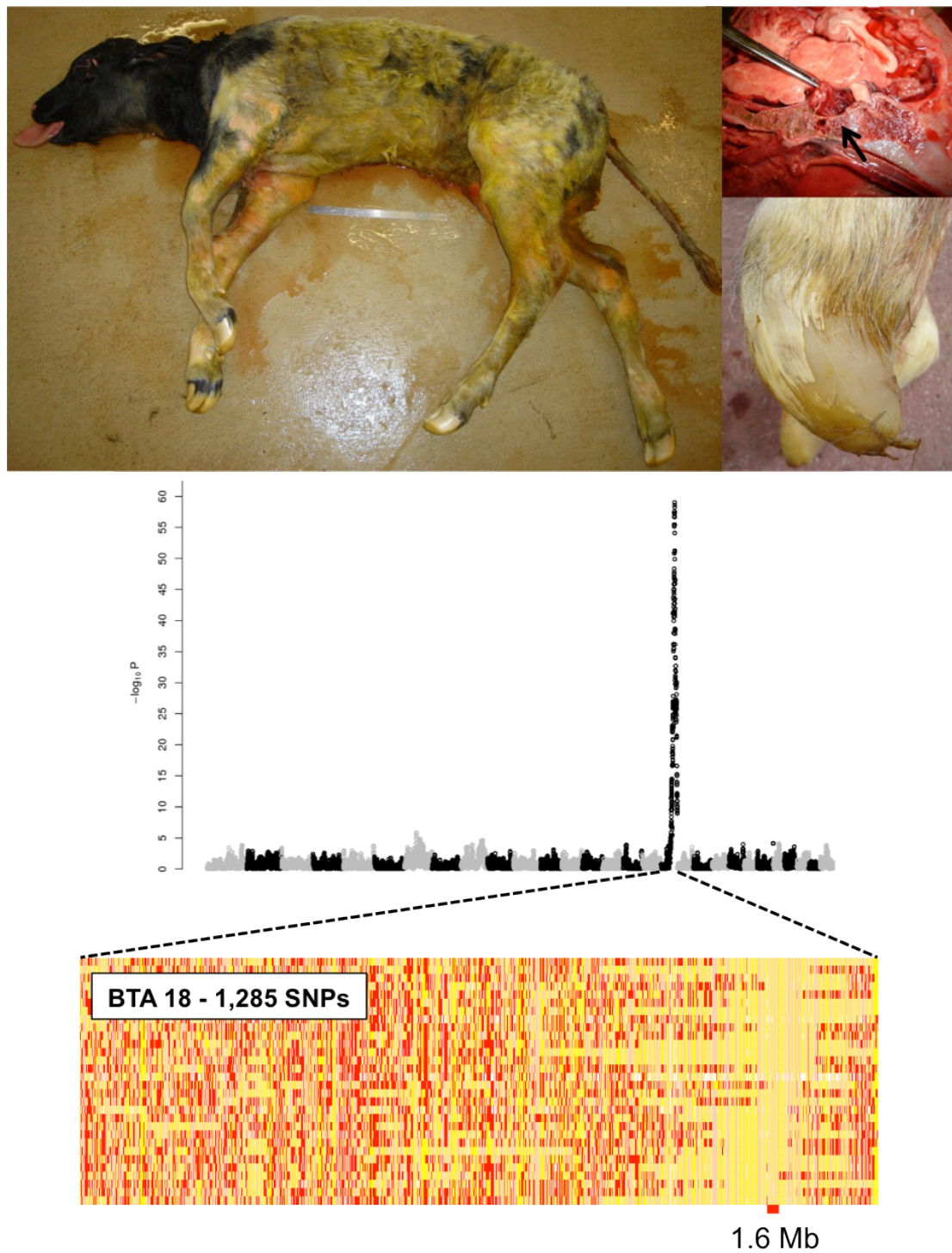


Figure 22 : Cartographie du locus associé à la gestation prolongée. Manhattan plot résultant de l'étude d'association basée sur les haplotypes pour la GP. L'alternance de gris et de noir correspond à la limite des autosomes de 1 à 29. Génotypes de 32 cas des 1,285 SNPs du chromosome 18, Les génotypes hétérozygotes sont en rouge et les génotypes homozygotes sont en jaune et orange. La région IBD de 1,6 Mb est soulignée en rouge.

B. Article**Ancestral haplotype-based association mapping with generalized linear mixed models accounting for stratification.**

Zhang Z., Guillaume F., Sartelet A., Charlier C., Georges M., Farnir F., Druet T.

Bioinformatics, 2012, **28**, 2467-2473

Ancestral haplotype-based association mapping with generalized linear mixed models accounting for stratification

Z. Zhang¹, F. Guillaume¹, A. Sartelet¹, C. Charlier¹, M. Georges¹, F. Farnir^{2,†} and T. Druet^{1,*,†}

¹Unit of Animal Genomics, GIGA-R and Faculty of Veterinary Medicine, University of Liège (B34), 1 avenue de l'Hôpital and ²Unit of Animal Productions, Faculty of Veterinary Medicine, University of Liège (B43), 20 boulevard de Colonster, B-4000 Liège, Belgium

Associate Editor: Jeffrey Barrett

ABSTRACT

Motivation: In many situations, genome-wide association studies are performed in populations presenting stratification. Mixed models including a kinship matrix accounting for genetic relatedness among individuals have been shown to correct for population and/or family structure. Here we extend this methodology to generalized linear mixed models which properly model data under various distributions. In addition we perform association with ancestral haplotypes inferred using a hidden Markov model.

Results: The method was shown to properly account for stratification under various simulated scenarios presenting population and/or family structure. Use of ancestral haplotypes resulted in higher power than SNPs on simulated datasets. Application to real data demonstrates the usefulness of the developed model. Full analysis of a dataset with 4600 individuals and 500 000 SNPs was performed in 2 h 36 min and required 2.28 Gb of RAM.

Availability: The software GLASCOW can be freely downloaded from www.giga.ulg.ac.be/jcms/prod_381171/software.

Contact: francois.guillaume@jouy.inra.fr

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on February 24, 2012; revised on May 24, 2012; accepted on June 12, 2012

1 INTRODUCTION

Genome-wide association studies (GWASs) identify genetic variants (e.g. SNPs, CNV or indels) affecting traits of interest such as those related to human health or of agronomical importance. With the development of high-throughput genotyping and next-generation sequencing, these studies have been particularly successful. Hundreds of loci associated with diseases were detected through GWAS (e.g. Donnelly, 2008). Association studies proved equally valuable in other organisms such as *Arabidopsis thaliana* (Aranzana *et al.*, 2005), mice (Threadgill *et al.*, 2002), dog, crops (Malosetti *et al.*, 2007; Yu *et al.*, 2006) or livestock species.

Although very effective, genetic association studies still face a number of potential pitfalls. One major problem in GWAS comes from the spurious associations that may occur as a

result of relatedness between individuals (e.g. familial relationships or population structure). Another issue is that, especially for complex traits, non-genetic factors (e.g. sex, age, etc.) may have profound impact on the scrutinized phenotype, raising the need for proper modeling of these effects.

An appealing solution to these problems is to use a mixed-model framework. Indeed, this methodology makes it possible to include covariates in the model and to account for the average genomic relatedness among individuals (population or family structure). Such models have been used for many years for QTL mapping especially in animal breeding (George *et al.*, 2000). Recent studies (Kang *et al.*, 2008; Malosetti *et al.*, 2007; Yu *et al.*, 2006; Zhao *et al.*, 2007) have demonstrated that inclusion of such effects in mixed-models properly corrects for stratification and that the use of mixed models to control for stratification resulted in fewer false positives and/or higher power than other techniques such as genomic control (Devlin and Roeder, 1999), structured association (Pritchard *et al.*, 2000) or principal components analysis (Price *et al.*, 2006). In addition, mixed-models were able to capture the multiple levels of population structure and genetic relatedness. All these features make mixed-models a very promising tool to perform association analyses while controlling for relatedness structure.

Linear mixed-models (LMMs) assume that traits are normally distributed. Use of generalized linear mixed models (GLMMs) allows extension of the mixed-model approach to other types of traits, such as binary traits for example. With these models, a linear function of different covariates including polygenic and local genomic effects is used to describe the expected value of the observed phenotype through a so-called link function. Tzeng and Zhang (2007) developed a variance-components score test for association studies which can be used in the GLMM framework.

Analyses can be performed using single SNP or haplotypes of multiple SNPs. Haplotypes are specific combinations of alleles on the same chromosome. They extract more information on the relation between DNA variation and phenotypes than single SNPs and may present higher correlation with underlying mutations [depending on the marker density and the linkage disequilibrium (LD) pattern in the population]. Furthermore, haplotype tests can model allelic heterogeneity or find several (interacting) mutations at different tightly linked sites. However, the power of haplotype-based tests is potentially reduced due to the extra

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the last two authors should be regarded as joint Last Authors.

degrees of freedom needed in these analyses (e.g. Su *et al.*, 2008; Tzeng and Zhang, 2007). Different strategies to minimize this problem have been proposed in the literature, relying mainly on grouping haplotypes based on similarity (e.g. Blott *et al.*, 2003; Durrant *et al.*, 2004; Druet *et al.*, 2008; Seltman *et al.*, 2003). Various clustering algorithms are available. Those relying on sliding window approaches are not optimal as the optimal window size varies from one region to another (Browning, 2008). With the localized haplotype clustering method (Browning and Browning, 2007), clusters of haplotypes are parsimoniously selected. This model allows for greater flexibility because haplotype lengths and the number of haplotypes are variable. Browning (2008) stated that this model is conceptually similar to the clusters of the Scheet and Stephens (2006) model. For each position along the genome, this later model assigns haplotypes to a pre-determined number of ancestral haplotypes present several generations ago from which all haplotypes within a cluster are assumed to have descended. Each haplotype can be associated to a cluster for a different length making the model flexible. In addition, the model can group haplotypes with small difference (missing genotypes, genotyping errors or recent mutations). Su *et al.* (2008) proposed to use these ancestral haplotypes in association testing, whereas we suggested to use them for QTL fine-mapping and genomic selection (Druet and Georges, 2010). In the same study, we showed that these clusters group haplotypes having a recent common ancestor (with short time to coalescence) and high identity-by-descent (IBD) probabilities [as estimated with the method of Meuwissen and Goddard (2001)]. It was as efficient as methods using these IBD probabilities to cluster haplotypes (e.g. Druet *et al.*, 2008). The use of these ancestral haplotypes proved already efficient for QTL fine-mapping (Karim *et al.*, 2011) and genomic selection (de Roos *et al.*, 2011).

In the present study, we develop a haplotyped-based method for association mapping relying on GLMM accounting for stratification and other covariates affecting the modeled trait.

2 METHODS

The proposed method relies on GLMM. To account for stratification and/or polygenic background, the model includes a vector of random polygenic effects (e.g. Yu *et al.*, 2006) in addition to the random haplotype effects:

$$\eta = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \mathbf{H}\mathbf{h} \quad (1)$$

where η is a vector of n linear predictors (with n equal to the number of observed phenotypes), β is a vector of fixed effects including the overall mean, $\mathbf{u}$ is a vector of n' random polygenic effects (with n' equal to the number of individuals for which genomic information is available typically, $n' = n$), $\mathbf{h}$ is a vector of P random ancestral haplotype effects, [with P equal to the chosen number of ancestral haplotypes (Druet and Georges, 2010)]. $\mathbf{X}$, $\mathbf{Z}$ and $\mathbf{H}$ are incidence matrices relating, respectively, fixed effects, polygenic and ancestral haplotype effects to the linear predictor. The variance of the random polygenic effects is denoted $\mathbf{G} = 2\mathbf{K}\sigma_g^2$ (Yu *et al.*, 2006) where $\mathbf{K}$ is a relative kinship matrix obtained from the marker data (see below) and σ_g^2 is the additive genetic variance. The variance of the random ancestral haplotype effects, is denoted $\mathbf{V} = \mathbf{I}\sigma_h^2$ where σ_h^2 is the 'haplotype' variance. The covariances between ancestral haplotype effects were assumed to be zero.

The linear predictors are transformed to the observed scale (e.g., disease status) through $h(\cdot)$, the inverse function of the link function $g(\cdot)$ (e.g. McCullagh and Nelder, 1989). In our analyses, the logit link function was

used to model binomial data such as disease status. The probability for individual i to be affected by the disease $h(\eta_i)$ is therefore obtained through the inverse of the logit function:

$$\mu_i = h(\eta_i) = \frac{e^{\eta_i}}{1 + e^{\eta_i}} \quad (2)$$

Coding healthy individuals with 0 and affected ones with 1, this probability is also the expected value for the trait. The solutions of the GLMM were obtained using an iterative procedure based on the Laplace approximation of the likelihood (Breslow and Clayton, 1993; McCullagh and Nelder, 1989). Indeed, the GLMM equations can be approximated by the following mixed-model equations (MMEs):

$$\begin{bmatrix} \mathbf{X}'\mathbf{W}\mathbf{X} & \mathbf{X}'\mathbf{W}\mathbf{Z} & \mathbf{X}'\mathbf{W}\mathbf{H} \\ \mathbf{Z}'\mathbf{W}\mathbf{X} & \mathbf{Z}'\mathbf{W}\mathbf{Z} + \mathbf{G}^{-1} & \mathbf{Z}'\mathbf{W}\mathbf{H} \\ \mathbf{H}'\mathbf{W}\mathbf{X} & \mathbf{H}'\mathbf{W}\mathbf{Z} & \mathbf{H}'\mathbf{W}\mathbf{H} + \mathbf{V}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{h}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{W}\tilde{\mathbf{y}} \\ \mathbf{Z}'\mathbf{W}\tilde{\mathbf{y}} \\ \mathbf{H}'\mathbf{W}\tilde{\mathbf{y}} \end{bmatrix} \quad (3)$$

where $\mathbf{W}$ is a diagonal matrix with element $w_i = (\phi v(\mu_i)\{g'(\mu_i)\}^2)^{-1}$, with ϕ a dispersion parameters and $v(\cdot)$ is a known variance function; $\tilde{\mathbf{y}}$ is a vector with elements $\tilde{y}_i = \eta_i + g'(\mu_i)(y_i - \mu_i)$. In the case of the logit function (and one record per individual), $\mathbf{W}$ is a diagonal matrix with $\mu_i(1 - \mu_i)$ on the diagonal. Since equations coefficients are functions of the unknowns, an iterative procedure must be used. First, starting values for β , $\mathbf{u}$ and $\mathbf{h}$ are used to compute μ and to build the approximate MME. These are then solved to obtain new estimates of β , $\mathbf{u}$ and $\mathbf{h}$. Parameters of the model (including variance components) are repeatedly estimated by REML until convergence.

The relative kinship matrix $\mathbf{K}$ was estimated based on similarity scores as in Eding *et al.* (2001) and Hayes and Goddard (2008):

$$S_{XY,l} = 0.25[I_{11} + I_{12} + I_{21} + I_{22}] \quad (4)$$

where $S_{XY,l}$ is the similarity score between individuals x and y at locus l and I_{ij} is an indicator variable equal to 1 if allele i on locus l in the individual x and allele j on the same locus in individual y are identical, otherwise it is 0. In our analyses, we replaced SNP alleles by ancestral haplotypes (similar to multi-allelic markers). The relationships are then based on closer founders. S_{XY} , averaged over the whole genome and normalized is then used as an estimator of the kinship relationship f_{XY} as Eding *et al.* (2001):

$$\hat{f}_{XY} = \frac{S_{XY} - s}{1 - s} \quad (5)$$

where s is the minimum value of S_{XY} in the matrix (Hayes and Goddard, 2008). Kang *et al.* (2008) and Zhao *et al.* (2007) concluded that use of similarity score to construct relationship matrices was as efficient as more complex methods and avoided problems of non-positive definite matrices. We tested different methods to construct relationship matrices (based on SNP or haplotypes) but these had little impact on estimation of polygenic effects and even less on residuals.

Associations are tested for every marker position along the genome by a significance test of $\sigma_h^2 = 0$. Explicit evaluation of the likelihood in GLMM is cumbersome, making application of likelihood ratio tests (LRTs) challenging. Therefore we used the score tests as proposed by Verbeke and Molenberghs (2003). Schaid *et al.* (2002) and Tzeng and Zhang (2007) used score tests in haplotype-based association studies with binary traits. The score tests are based on the value of the first derivative of the log-likelihood under the null hypothesis (i.e. the variance of the haplotypes is null). A significant positive first derivative with respect to the variance component indicates that the maximum-likelihood estimator of the haplotypes variance is significantly different from zero.

Tzeng and Zhang (2007) derived a test statistic T based on the score tests for haplotype-based models for GLMM. In the case of the logit function, the test statistic is equal to:

$$T = 0.5(\mathbf{y} - \mu)' \mathbf{H} \mathbf{H}' (\mathbf{y} - \mu) = 0.5(\mathbf{H}'(\mathbf{y} - \mu))' * \mathbf{H}'(\mathbf{y} - \mu) \quad (6)$$

where $\mathbf{y} - \mu$ is a vector of residuals (observations corrected for estimated fixed and random effects) obtained from a GLMM under the null hypothesis (no haplotype effect) where $\sigma_h^2 = 0$. Since T relies on estimation of residuals from a model without haplotype effects, the procedure is similar to the two-step procedure proposed in Aulchenko et al. (2007). Therefore it has the same advantages: the mixed models must be solved only once to obtain the residuals, which considerably speeds up computations, and since residuals are corrected for stratification, they are free from familial correlations and the data become exchangeable (Aulchenko et al., 2007) which means that permutation techniques may be applied.

Tzeng and Zhang (2007) demonstrated that the distribution of the T test statistic under the null hypothesis could be approximated using a gamma distribution. We perform 1,000 permutations of the residuals to estimate the mean and the variance of the gamma distribution (or the shape and scale parameters). Parameters of the gamma distribution are estimated for each tested position (marker) because the distribution is influenced by the structure of the incidence matrix relating haplotypes to residuals, which is potentially position specific. We will refer to this strategy as 'gamma approximation'. In addition, empirical P -values can be computed by repeatedly permuting the phenotypes (residuals) among the individuals (referred to as permutation hereafter).

3 SIMULATIONS

A dataset of 3547 genotyped Holstein, Jersey or crossbred bulls (Karim et al., 2011) was used to simulate case/control studies. Individuals were genotyped for the Illumina Bovine SNP50 SNP chip (Illumina, San Diego, CA). After editing data, 37 647 SNPs were conserved on the 29 autosomes. DualPHASE from the PHASEBOOK package (Druet and Georges, 2010) was used to infer haplotypes from the genotyped individuals and to assign them to K ancestral haplotype clusters (K was set equal to 5, 10 or 20).

To simulate different stratification scenarios including breed (i.e. structured populations) and polygenic (i.e. familial relationships) effects, and mimicking major variant (SNP with large effect) effects, the following model was used:

$$\eta = \mathbf{X}_1\mu + \mathbf{X}_2\beta + \mathbf{Z}_1\mathbf{u} + \mathbf{Z}_2\mathbf{v} \quad (7)$$

where μ is the mean effect, $\mathbf{X}_1$ is a vector of '1', β is the breed effect, $\mathbf{X}_2$ is a vector containing the percentage of Holstein blood (ranging from 0 to 1), $\mathbf{Z}_1$ is a matrix ($n \times 1000$) containing the number of alleles '1' of a set of 1000 SNPs with small phenotypic effect (hereafter called polygenic SNPs), $\mathbf{u}$ is a vector containing the allelic substitution effects of 1000 polygenic SNPs used to simulate a polygenic effect, $\mathbf{Z}_2$ is a vector containing the number of alleles '1' for a SNP with a large phenotypic effect (hereafter called major variant) and $\mathbf{v}$ is the allelic substitution effect for that SNP. The breed effect was equal to 0, 0.2, 0.4 or 0.7 according to the scenario (corresponding to odds ratios (ORs) equal to 1.0, 1.22, 1.5 and 2.0, respectively), the individual SNP effects (1000 SNPs) were drawn from a gamma distribution (shape = 0.4 and scale = 1). The variance of the polygenic effects (the sum of 1000 individual polygenic SNP effects) was then rescaled to 0 or 0.16. Finally, the major variant effects were equal to 0.5 and 0.8 (corresponding to OR equal to 1.65 and 2.22, respectively). The phenotype of each individual was then sampled from a Bernoulli distribution with mean equal to:

$$p_i = h(\eta_i) = \frac{e^{\eta_i}}{1 + e^{\eta_i}} \quad (8)$$

The overall mean effect was set to obtain a prevalence of the disease of 27% in the population. Finally, 500 cases and 500 controls were randomly sampled from the 3547 genotyped individuals. One thousand SNPs were selected as potential major variants and were removed from the dataset prior to phasing. A total of 10 000 and 100 000 simulations were performed per scenario to estimate power and to compute QQ-plots, respectively.

4 RESULTS

4.1 Simulated data

The maximum LD (measured by r^2) between each of the 1000 major variants and the remaining SNPs (hereafter called marker SNPs) or with the 5, 10 or 20 ancestral haplotypes was estimated within a 2 Mb interval (1 Mb on each side of the major variant). The maximum r^2 was on average equal to 0.40 with marker SNPs and to 0.51 and 0.72 with 10 or 20 ancestral haplotypes. However, in some cases, SNPs can still present higher LD. Indeed, using marker SNPs or $K=20$ haplotypes, 7.5 % of the major variants were captured with an r^2 of 1.0 whereas this value dropped to 0% with 5 or 10 haplotypes. From here on, $K=20$ for the remainder of the simulation study.

To test whether our model correctly accounted for stratification, type-I errors were estimated by testing the model under H_0 (the major variant had no effect: v was set equal to 0) and QQ-plots were generated. Simulations were performed including a breed effect of 0.4, but no polygenic effect. Four models were fitted to the generated dataset: using haplotypes or marker SNPs, and including or not a polygenic effect accounting for stratification. In Figure 1, models without polygenic effect clearly present an excess of small P -values. After inclusion of the polygenic effect, the regression slopes of the QQ-plot were below one, indicating that stratification was correctly accounted for but that tests are slightly too conservative.

In Table 1, regression coefficients of QQ-plots obtained with the four fitted models applied to different simulated scenarios are presented. In all cases where stratification was simulated, models without polygenic effect showed excess of small P -values, particularly with haplotypes which tend to capture more

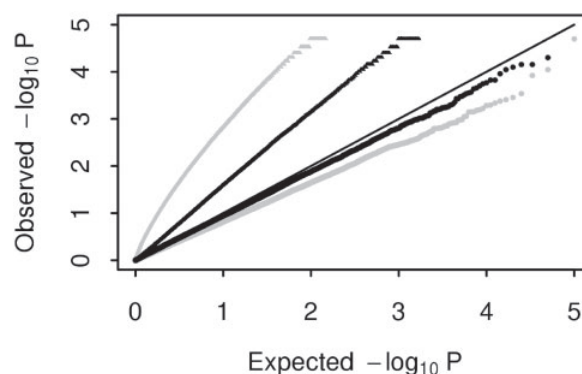


Fig. 1. QQ-plots obtained with marker SNPs (black) or ancestral haplotypes (gray) with (circles) or without (triangles) polygenic terms accounting for stratification included in the model

Table 1. Regression coefficients of QQ-plots obtained with four fitted models on four simulated designs [*P*-values computed with permutation test or with a gamma approximation test (in parenthesis)]

Simulated breed effect	Simulated polygenic variance	Fitted model			
		without polygenic effect		with polygenic effect	
		SNP ^a	Anc. Hap. ^a	SNP	Anc. Hap.
0	0	1.03 (1.01)	0.99 (1.02)	0.97 (0.96)	0.86 (0.87)
0.4	0	1.59 (1.57)	2.54 (3.06)	0.94 (0.94)	0.82 (0.83)
0	0.16	1.39 (1.36)	2.02 (2.22)	0.95 (0.95)	0.86 (0.86)
0.2	0.16	1.65 (1.63)	3.29 (3.89)	0.94 (0.94)	0.82 (0.83)

^aAssociation is performed either with SNP or ancestral haplotypes.

stratification effects. The inclusion of a polygenic effect resulted in regression coefficients below 1.0, even when both breed and polygenic effects were simulated. In these simulations, two methods were used to estimate *P*-values, namely the permutation test and the gamma approximation test, with both yielding approximately the same regression slopes after correction for stratification.

Table 2 compares the power of models with marker SNPs or haplotypes for different OR and frequencies of the major variant (permutations were used to estimate *P*-values). In GWAS, due to multiple testing, low *P*-values must be achieved but the number of simulations allowed us only to estimate the power at $\alpha=0.001$. The marker with the strongest association is not always the closest to the major variant. Therefore, power was tested in a 2 Mb window centered on the major variant, spanning ~30 SNPs. To correct for the resulting multiple testing (and correlation among successive tests along the region), chromosomes were randomly shuffled across individuals 10 000 times. For each permutation, the best *P*-value was stored and the association test was declared significant at $P<0.001$ if one of the *P*-values was lower than 9990 of the best *P*-values obtained by permutation. Major variants, with resulting OR equal to 1.65 and 2.22, accounted only for a small fraction of the variation (r^2 between the SNP and the binary trait below 0.04 and 0.08 according to the SNP effect). Due to incomplete LD with causative SNPs, haplotypes and genotyped SNPs captured an even smaller fraction of that variance. Therefore the power was low (tests are already corrected for ~30 repeated correlated tests by the permutation procedure explained above), particularly when minor allele frequency (MAF) was below 0.2 and for small OR (1.65). For larger SNP effects, power increased, particularly when using ancestral haplotypes in the model which proved better in the present simulations for MAF above 0.10.

The power was also compared for different levels and types of structures (Table 3). In all cases, use of ancestral haplotypes resulted in higher power than for single SNPs and power decreased in datasets presenting structure (particularly for haplotypes and in presence of polygenes).

4.2 Real data

Data from the Belgian Blue cattle breed heredo-surveillance platform were used to test the method on real datasets. Four phenotypes were analyzed: 3 monogenic recessive diseases [gingival

Table 2. Variant detection power ($\alpha=0.001$) with SNP or ancestral haplotypes in a design without structure (no breed and polygenic effects) for different minor allelic frequency (MAF) classes of the causal SNP

MAF Class	OR of variant = 1.65			OR of variant = 2.22		
	r^2 ^a	SNP	Anc. Hap	r^2	SNP	Anc. Hap
0.00–0.10	0.008	0.003	0.002	0.018	0.020	0.016
0.10–0.20	0.017	0.013	0.030	0.040	0.072	0.194
0.20–0.30	0.024	0.033	0.071	0.057	0.129	0.421
0.30–0.40	0.030	0.050	0.113	0.071	0.136	0.520
0.40–0.50	0.032	0.053	0.152	0.076	0.150	0.622

^a r^2 between causative variant and observed phenotype.**Table 3.** Power of association mapping ($\alpha=0.001$) with SNP and haplotypes in different designs with stratification (statistics are provided across all MAFs)

Breed effect	Polygenic variance	OR of variant = 1.65		OR of variant = 2.22	
		SNP	Anc. Hap.	SNP	Anc. Hap.
0	0	0.023	0.078	0.079	0.373
0	0.16	0.025	0.057	0.077	0.322
0.4	0	0.027	0.067	0.077	0.359
0.7	0	0.028	0.071	0.075	0.365
0.2	0.16	0.023	0.053	0.073	0.322

hamartoma (33 cases), arthrogryposis (13 cases) and prolonged gestation (25 cases)] and color-sidedness which is monogenic dominant (8 cases). The causative variants are known for hamartoma (Sartelet *et al.*, in preparation) and color-sidedness (Durkin *et al.*, 2012) whereas for arthrogryposis and prolonged gestation, diagnostic tests have been developed based on markers in LD with the causative variants. In addition to cases, genotypes from 300 controls were available. Individuals were genotyped for a custom made 50 K bovine chip described in Charlier *et al.* (2008). After deleting markers with a call rate below 0.90 or having a MAF below 0.05, 41 878 SNPs mapping to autosomal

chromosomes were used in the study. Haplotypes were reconstructed using DualPHASE (Druet and Georges, 2010) with 10 ancestral haplotypes (we reduced the number of ancestral haplotypes to 10 since the number of individuals is much smaller than in the simulation study). Association studies were also performed with EMMAX (Kang *et al.*, 2010) which performs single point (SNPs) association studies with LMM that account for stratification (through inclusion of a kinship matrix). After Bonferroni correction for $\sim 50\,000$ tests, genome-wide significance was set at 10^{-6} .

For the 3 monogenic recessive diseases P -values (estimated with the gamma approximation) below 10^{-40} were obtained (Manhattan plots are available in Supplementary Figures) in regions in which almost all cases were homozygous for a specific ancestral haplotype whereas almost none of the controls was homozygous for that haplotype, suggesting high LD between this ancestral haplotype and the causative variant. The identified regions were in agreement with the previous findings.

The Manhattan plot for color-sidedness is presented in Figure 2a. The lowest P -value is below 10^{-10} and the corresponding position is located at 0.7 Mb from a CNV causing the phenotype, a copy of a chromosomal segment on BTA6 encompassing the KIT locus which translocated to BTA29 (Durkin *et al.*, 2012). All color-sided individuals carry at least one copy of the same ancestral haplotype (it has dominant behavior). Some controls also carry the haplotype but the phenotype is not observed. Indeed, the phenotype has an incomplete penetrance since it cannot be observed on individuals with completely white coats (homozygous genotypes for a frequent common codominant mutation at the roan locus results in white coats [Charlier *et al.*, 1996]).

Associations performed with EMMAX are presented in Supplementary Figures for monogenic recessive diseases and Figure 2b for color-sidedness. This software relies on LMM and assumes that the traits are normally distributed. As other LMM packages, it can still be applied on binary traits and perform well as shown in Supplementary Figures where several SNP in the region surrounding the causative SNP were highly significant and no other SNP reached such levels of significance.

However, some SNPs reach genome-wide significance in non-causative regions (e.g. association study for arthrogryposis). For color-sidedness, use of EMMAX resulted in a Manhattan plot where the causative region is non-significant and difficult to identify (other regions of the genome show higher significance). These examples illustrate that considering binary trait as normally distributed and using SNP as covariates can result in situations where the region harboring the causative mutation is difficult to identify. In such situations, extension to GLMM and use of ancestral haplotypes resulted in associations where many positions in the region of interest have high level of significance, clearly above the remainder of the genome and with less non-causative regions reaching genome-wide significance.

4.3 Computational performance

To test the computational efficiency of the developed software, we ran an implementation compiled with Intel Fortran using openMP and MKL libraries on a dataset with a total of 4600 individuals genotyped for 500 000 SNPs. The analysis was

performed on a Intel Xeon E5520 processor at 2.27 GHz using four threads. The full analysis lasted 2 h 36 min and required a total of 2.28 Gb of memory.

5 DISCUSSION

Our simulation studies showed that taking into account genomic relationships among individuals through inclusion of a polygenic effect in GLMM accounted for stratification, as previously observed with LMM (e.g. Malosetti *et al.*, 2007; Yu *et al.*, 2006; Zhao *et al.*, 2007). It was also observed that corrections were effective in structured populations (breed effect) and/or when family structure was present (polygenic effect). Most designs in model organisms, plants or animal species present a high level of stratification because either several populations are used in the study or the individuals are closely related, making these robust corrections essential. In addition to correcting for stratification, the GLMM framework offers additional flexibility because it allows for a better modeling of the phenotypes through the inclusion of additional covariates (e.g. sex, age, etc.) and a consequently better association study where all known nuisance factors have been corrected for. Such a possible nuisance factor could be a measure of the population structure, as suggested in some studies (e.g. Price *et al.*, 2006; Pritchard *et al.*, 2000). The effect of adding this correction concurrently with the genomic relatedness structure might be necessary in some populations albeit not always (Kang *et al.*, 2008; Zhao *et al.*, 2007). Another point stressing flexibility of LMM is that association can easily be performed with either SNPs or (ancestral) haplotypes. Extension from LMM to GLMM is important for traits that are not normally distributed. However, in many situations such as large balanced case/control studies where variants are not very rare and have low or moderate effects, LMM perform well with binary traits. GLMM are recommended for strong deviations from normality, when cases or controls are rare within a cell (covariates of the model such as fixed, SNP or haplotype effects). Such situations occur more often in smaller designs with few cases (or controls) and with rare variants (or haplotypes). Such designs are still common in animal or plant species and our applications to real data illustrate that in such cases, use of a GLMM with a logit link function results in cleaner association than LMM.

Slopes of QQ-plots indicated that statistical tests were too conservative resulting in a loss of power. Similar deflation has been described in Amin *et al.* (2007) and is due to the fact that polygenic effects (used for correction) and SNP or haplotype effects are correlated (e.g. SNP or haplotypes are used to estimate K). This correlation is stronger with haplotype which are therefore more affected by the over-correction. One solution would be to correct for deflation with genomic control as in Amin *et al.* (2007) or to perform a test modeling simultaneously polygenic and haplotype effects at most interesting positions. Without such corrections, the proposed model may be more conservative than some other methods relying on LMM.

Although computational efficiency was not the main goal of the present study, the developed method presents interesting computational features. First, GLMM must be solved only once (with the SNP or haplotype variance set to 0). Solving the GLMM equations and inferring the variance components

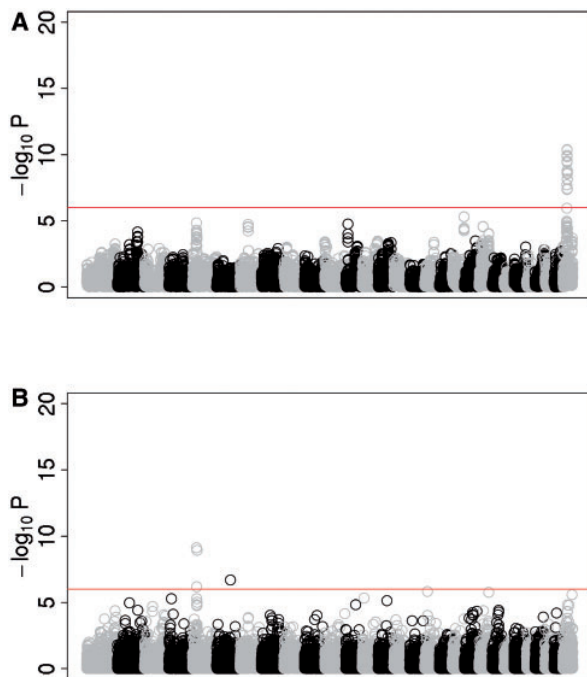
Z.Zhang *et al.*

Fig. 2. Manhattan plot for association study for color-sidedness with (A) GLMM score tests using ancestral haplotypes and (B) EMMAX

are potentially time consuming. With likelihood ratio tests, variances are inferred for a model including polygenic and SNP (or haplotype) random effects for each tested position. With the present method, the test score is obtained from a simple statistic based on records corrected for effects of the model under the null hypothesis (without SNP or haplotypic effect). Performing this test is much faster than inferring variance in the full model. This is very similar to the approach used in GRAMMAR (Aulchenko *et al.*, 2007). In addition, since the data are corrected for family and population structure it becomes exchangeable and permutations can be performed freely, which was not the case for the raw data. Thanks to permutations, empirical *P*-values corrected for multiple testing can easily be obtained. Still, with permutations it would be time consuming to obtain small *P*-values as typically needed in GWAS studies; in that case, approximation of the test score distribution with a gamma distribution (with scale and shift parameters obtained empirically through 1000 permutations) seems to perform well.

As in previous studies (Druet and Georges, 2010; Su *et al.*, 2008), assigning chromosomes to ancestral haplotypes resulted in high LD between haplotype groups and underlying mutations. In the present study, the LD was much higher than when using SNP for a cattle population and with ~50 000 SNPs covering the genome. The picture might change with different marker densities or in other populations. The method for clustering haplotypes has also been successfully applied to the fine-mapping of a QTL affecting bovine stature (Karim *et al.*, 2011) in a crossbred population. In that study, the association between the ancestral haplotypes and the later candidate causative variants was almost perfect (2 misclassified haplotypes out of 1490). More recently,

our method allowed to fine-map a mutation causing dwarfism in cattle which was always associated to the same ancestral haplotype (Sartelet *et al.*, 2012). Other examples in the present study illustrate the high LD between the ancestral haplotypes obtained with DualPHASE (Druet and Georges, 2010) and ungenotyped variants. The association should be better for more recent variants which rapidly increased in frequency due to selection. In that case, the length of the haplotype associated to the variant would be longer than for random variants (as those used in the simulation study), making it easier to identify the haplotype. Ancestral haplotypes can be associated with different types of variants including SNP, multiple alleles, several-linked SNP (a small haplotype), insertions/deletions and duplications. In the real data example on color-sidedness, ancestral haplotypes presented high LD with a trans CNV and other examples of association between ancestral haplotypes and deletions or duplications (either in cis or trans position) can be found in Durkin *et al.* (2012). Note, when large reference populations genotyped at high density (or sequenced) are available, imputation followed by single point association would probably result in higher power than use of ancestral haplotypes (if there is only one causal variant and if SNPs in high LD with this variant are genotyped in the reference panel). However, such reference populations are only available in a few species.

While providing high LD with underlying variants, the use of ancestral haplotypes also controls the number of haplotype groups to be used in the study, which is important to maintain statistical power. This method proves also flexible since there is no need to define arbitrary windows and since haplotype origin can change at any position along the chromosome. For instance, recombinant haplotypes do not create additional haplotypes groups: they are simply potentially assigned to different groups on each side of the cross-over position. Finally, haplotypes with small differences due to genotyping errors or new non-causative mutations can still be grouped together whereas with less flexible methods, new haplotype groups would be defined for each difference, resulting in a loss of power. For the same reason, missing genotypes are easily handled. Our score test framework can easily be applied with other methods for clustering haplotypes, even those modeling a correlation between haplotypic effects. Our method does not rely on a particular biological model but identifies ancestral haplotypes significantly associated with the disease. Therefore it can be applied to monogenic recessive diseases, dominant diseases, phenotypes with complete or incomplete penetrance, oligenic or polygenic diseases or complex traits. It is also robust to misclassified samples which will only reduce slightly the power since there are no strong assumptions such as sharing of an IBD segment in all cases. For instance, the method was used to fine-map a variant causing dwarfism in Belgian Blue cattle (Sartelet *et al.*, 2012) which was the cause only for a subset of cases (14 out of 33). Still, the method detected with high significance (P -value $< 10^{-11}$) the region harboring the causative mutation. Further veterinary examination revealed that dwarfs could be classified into different categories and that the 14 cases corresponded to a specific sub-group. Even without that knowledge, the variant was identified, stressing the robustness of the method. The example of color-sidedness also illustrates that with only a few cases (8), a dominant gene (cases carry only one haplotype) and incomplete penetrance (the phenotype is not

observed on white animals) the method still identifies with high significance the region harboring the causative variant.

ACKNOWLEDGEMENTS

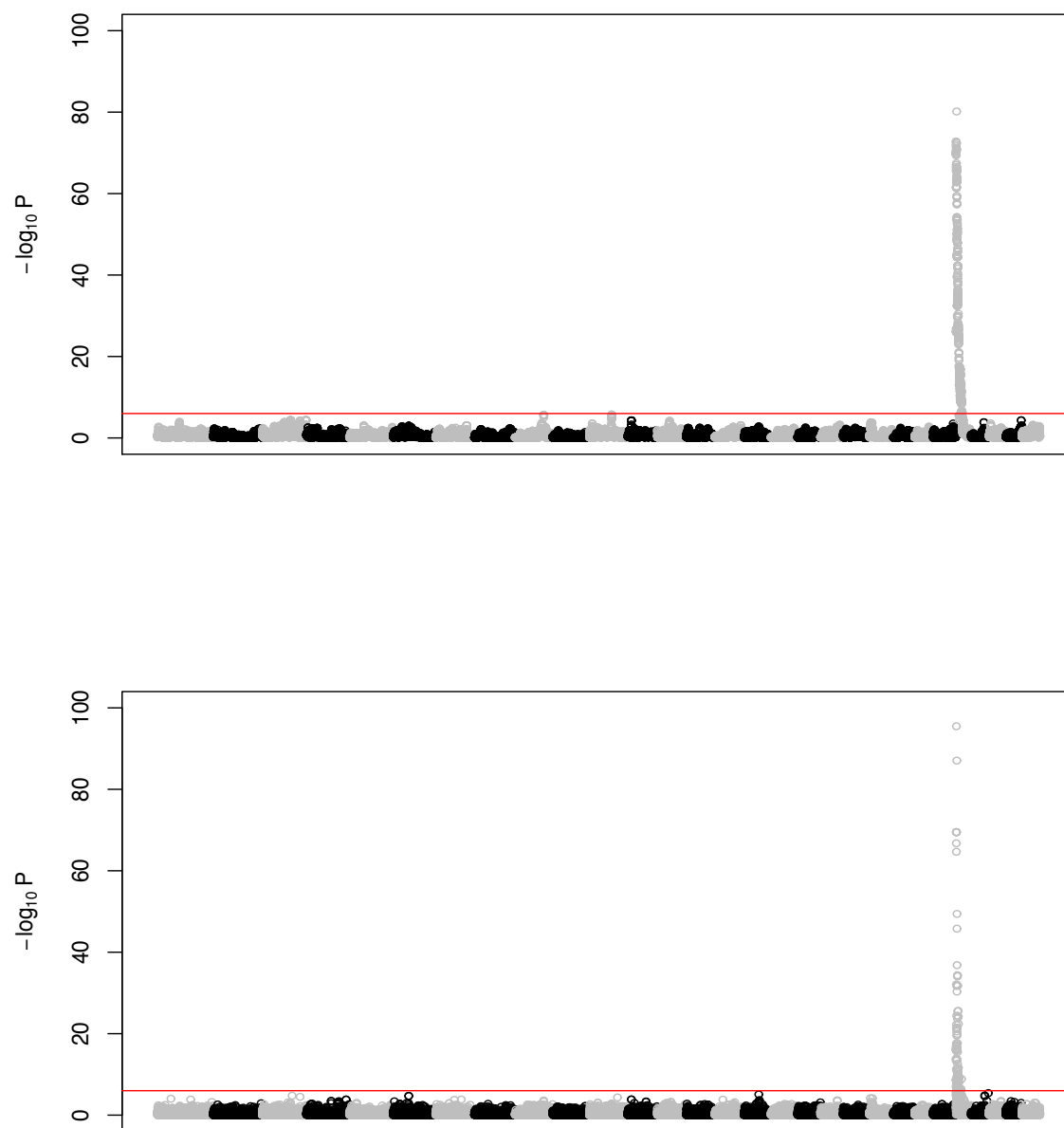
We acknowledge LIC for providing data. Tom Druet and Carole Charlier are, respectively, Research Associate and Senior Research Associate of the Fonds de la Recherche Scientifique—FNRS. The authors thank the GIGA Bioinformatics platform and SEGI (University of Liège) for providing computing resources.

Funding: This work was funded by grants from the Walloon Ministry of Agriculture (Rilouke) and the European Union (PIGENDEF).

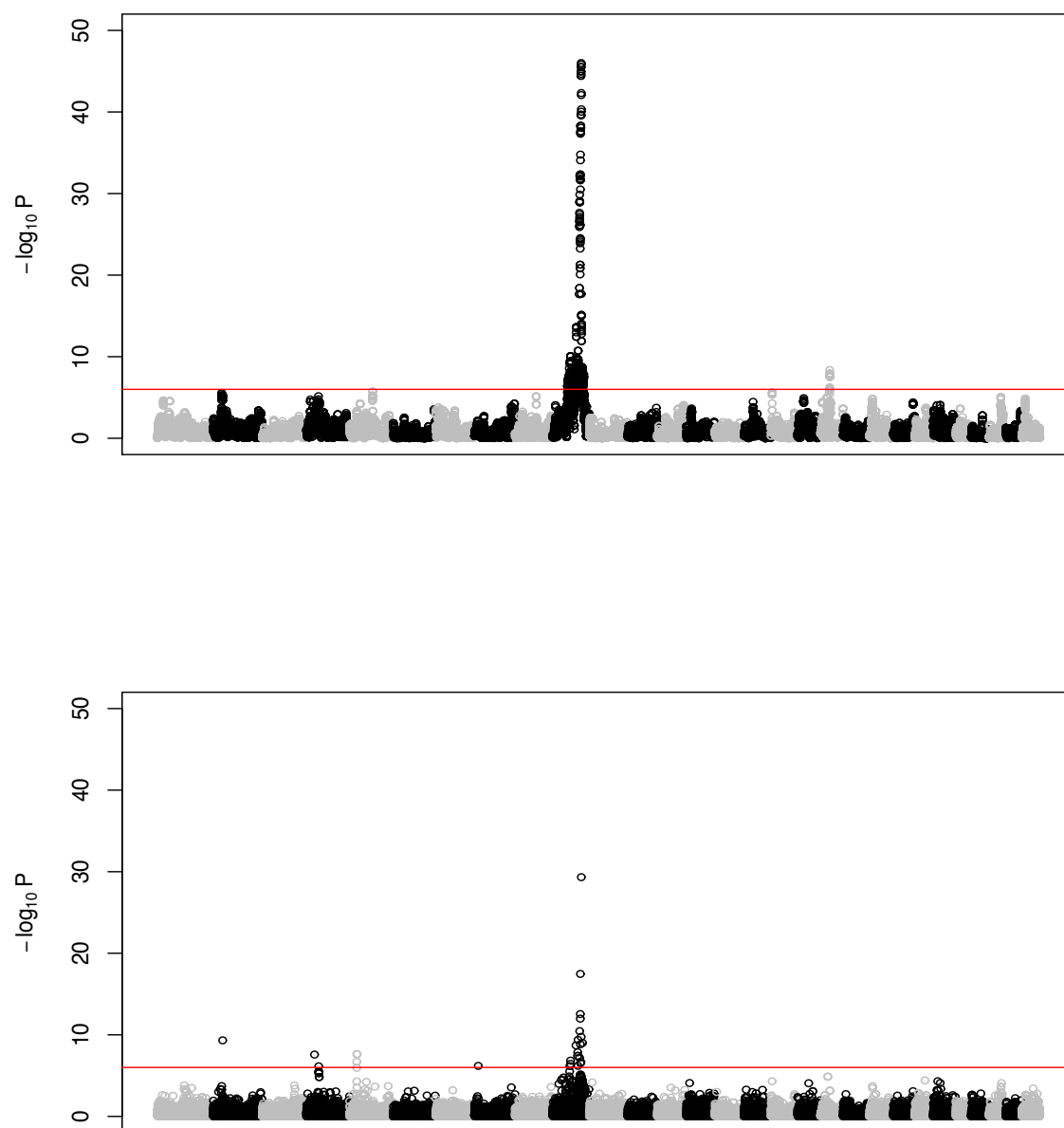
Conflict of Interest: None declared.

REFERENCES

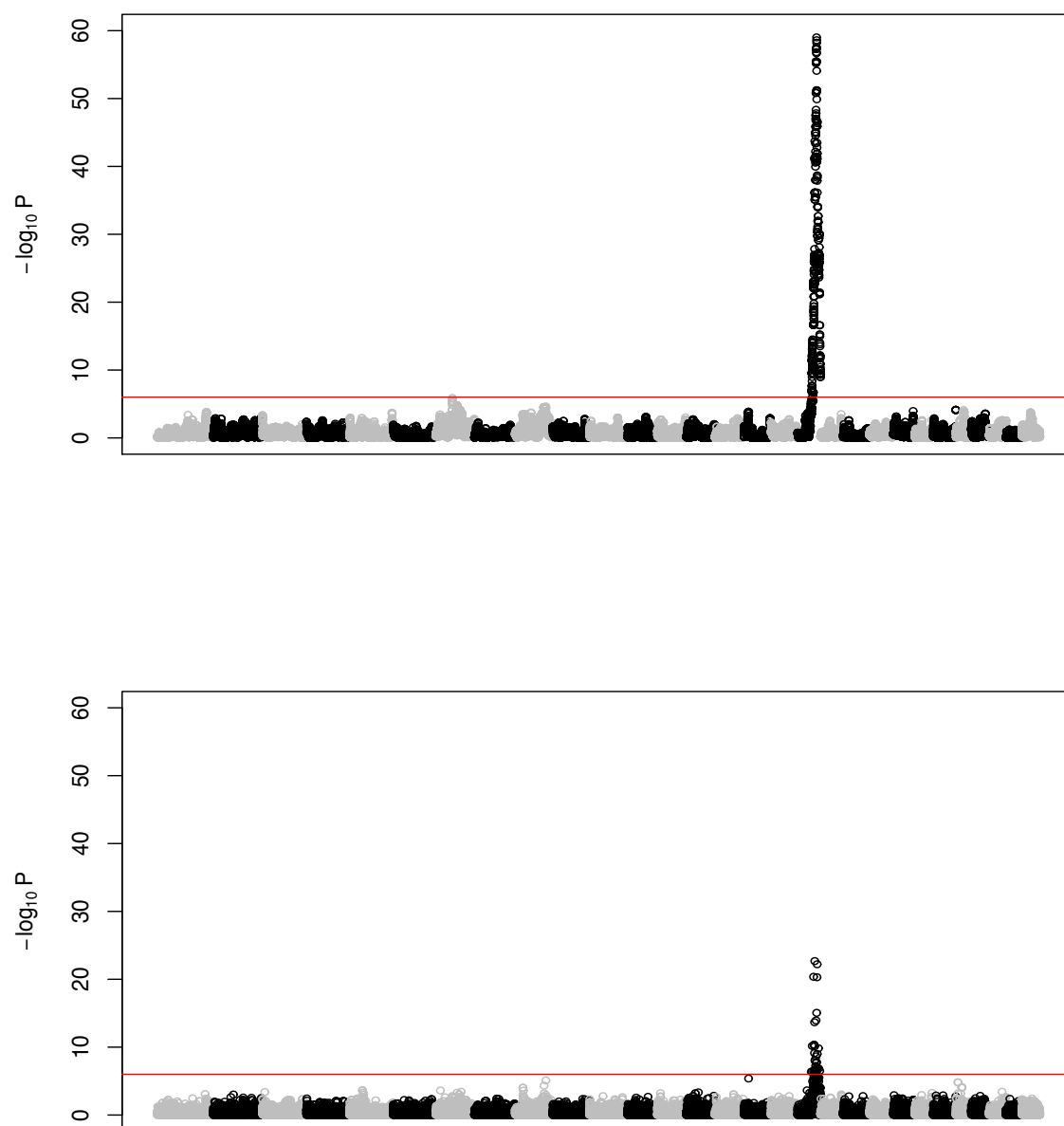
- Amin, N. *et al.* (2007) A genomic background based method for association analysis in related individuals. *PLoS One*, **2**, e1274.
- Aranzana, M.J. *et al.* (2005) Genome-wide association mapping in arabidopsis identifies previously known flowering time and pathogen resistance genes. *PLoS Genet.*, **1**, e60.
- Aulchenko, Y.S. *et al.* (2007) Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. *Genetics*, **177**, 577–585.
- Blott, S. *et al.* (2003) Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics*, **163**, 253–266.
- Breslow, N. and Clayton, D. (1993) Approximate inference in generalized linear mixed models. *J. Am. Stat. Assoc.*, **88**, 9–25.
- Browning, S.R. (2008) Missing data imputation and haplotype phase inference for genome-wide association studies. *Hum. Genet.*, **124**, 439–450.
- Browning, S.R. and Browning, B.L. (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.*, **81**, 1084–1097.
- Charlier, C. *et al.* (1996) Microsatellite mapping of the bovine roan locus: a major determinant of white heifer disease. *Mamm. Genome*, **7**, 138–142.
- Charlier, C. *et al.* (2008) Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nat. Genet.*, **40**, 449–454.
- de Roos, A.P.W. *et al.* (2011) Genomic breeding value estimation using genetic markers, inferred ancestral haplotypes, and the genomic relationship matrix. *J. Dairy Sci.*, **94**, 4708–4714.
- Devlin, B. and Roeder, K. (1999) Genomic control for association studies. *Biometrics*, **55**, 997–1004.
- Donnelly, P. (2008) Progress and challenges in genome-wide association studies in humans. *Nature*, **456**, 728–731.
- Druet, T. and Georges, M. (2010) A hidden markov model combining linkage and linkage disequilibrium information for haplotype reconstruction and quantitative trait locus fine mapping. *Genetics*, **184**, 789–798.
- Druet, T. *et al.* (2008) Fine mapping of quantitative trait loci affecting female fertility in dairy cattle on BTA03 using a dense single-nucleotide polymorphism map. *Genetics*, **178**, 2227–2235.
- Durkin, K. *et al.* (2012) Serial translocation via circular intermediates underlies color-sidedness in cattle. *Nature*, **482**, 81–84.
- Durrant, C. *et al.* (2004) Linkage disequilibrium mapping via clastic analysis of single-nucleotide polymorphism haplotypes. *Am. J. Hum. Genet.*, **75**, 35–43.
- Eding, H. *et al.* (2001) Marker-based estimates of between and within population kinships for the conservation of genetic diversity. *J. Anim. Breed. Genet.*, **118**, 141–159.
- George, A.W. *et al.* (2000) Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. *Genetics*, **156**, 2081–2092.
- Hayes, B.J. and Goddard, M.E. (2008) Technical note: prediction of breeding values using marker-derived relationship matrices. *J. Anim. Sci.*, **86**, 2089–2092.
- Kang, H.M. *et al.* (2008) Efficient control of population structure in model organism association mapping. *Genetics*, **178**, 1709–1723.
- Kang, H.M. *et al.* (2010) Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.*, **42**, 348–354.
- Karim, L. *et al.* (2011) Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. *Nat. Genet.*, **43**, 405–413.
- Malosetti, M. *et al.* (2007) A mixed-model approach to association mapping using pedigree information with an illustration of resistance to phytophthora infestans in potato. *Genetics*, **175**, 879–889.
- McCullagh, P. and Nelder, J. (1989) *Generalized Linear Models (Monographs on Statistics and Applied Probability 37)*, Chapman Hall, London.
- Meuwissen, T.H. and Goddard, M.E. (2001) Prediction of identity by descent probabilities from marker-haplotypes. *Genet. Sel. Evol.*, **33**, 605–634.
- Price, A.L. *et al.* (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.
- Pritchard, J.K. *et al.* (2000) Association mapping in structured populations. *Am. J. Hum. Genet.*, **67**, 170–181.
- Sartelet, A. *et al.* (2012) A splice site variant in the bovine RNF11 gene compromises growth and regulation of the inflammatory response. *PLoS Genet.*, **8**, e1002581.
- Schaid, D.J. *et al.* (2002) Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am. J. Hum. Genet.*, **70**, 425–434.
- Scheet, P. and Stephens, M. (2006) A fast and flexible statistical model for largescale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.*, **78**, 629–644.
- Seltman, H. *et al.* (2003) Evolutionary-based association analysis using haplotype data. *Genet. Epidemiol.*, **25**, 48–58.
- Su, S.-Y. *et al.* (2008) Disease association tests by inferring ancestral haplotypes using a hidden Markov model. *Bioinformatics*, **24**, 972–978.
- Threadgill, D.W. *et al.* (2002) Genetic dissection of complex and quantitative traits: from fantasy to reality via a community effort. *Mamm. Genome*, **13**, 175–178.
- Tzeng, J.-Y. and Zhang, D. (2007) Haplotype-based association analysis via variance components score test. *Am. J. Hum. Genet.*, **81**, 927–938.
- Verbeke, G. and Molenberghs, G. (2003) The use of score tests for inference on variance components. *Biometrics*, **59**, 254–262.
- Yu, J. *et al.* (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.*, **38**, 203–208.
- Zhao, K. *et al.* (2007) An Arabidopsis example of association mapping in structured samples. *PLoS Genet.*, **3**, e4.



Supporting Figure 1: Manhattan plot for association study for gingival hamartoma with GLMM score tests using ancestral haplotypes and EMMAX (from top to bottom).



Supporting Figure 2: Manhattan plot for association study for arthrogryposis with GLMM score tests using ancestral haplotypes and EMMAX (from top to bottom).



Supporting Figure 3: Manhattan plot for association study for prolonged gestation with GLMM score tests using ancestral haplotypes and EMMAX (from top to bottom)

DISCUSSION GENERALE
& PERSPECTIVES

1 – DISCUSSION GENERALE

La cellule d'hérédo-surveillance, un succès belge.

L'objectif premier de ma thèse était le développement et la mise en place d'une cellule d'hérédo-surveillance dans la race BBB. Sept ans plus tard, quand on aime on ne compte pas... le bilan est positif puisque près de 200 vétérinaires praticiens, 500 éleveurs, les quatre centres d'insémination artificielle, les laboratoires régionaux (DGZ et Arsia), l'Association Wallonne de l'Elevage (AWE) et les deux Facultés de Médecine Vétérinaire (FMV) Belges (Liège et Gand) ont contribué au recensement de près de 1200 cas et de 1400 apparentés (Annexe 5). On peut voir que dans d'autres pays, des laboratoires s'intéressent au recensement des défauts ainsi qu'aux caractéristiques raciales dans l'espèce bovine. On peut citer comme exemples, la Suisse avec le laboratoire des Prof. T. Leeb et C. Drögemüller (www.genetics.unibe.ch), la France avec l'Observatoire National des Anomalies Bovines (www.onab.fr) et l'Institut National de la Recherche Agronomique (INRA), le Danemark avec le Prof. J. Agerholm, l'Italie avec le Prof. A. Gentile et l'Australie avec le Prof. F. Nicholas, fondateur de l'OMIA (Online Mendelian Inheritance in Animals) recensant toutes les anomalies congénitales animales. Ils s'impliquent à différents niveaux dans l'identification, le recensement et la gestion de défauts héréditaires dans l'espèce bovine (Ducos *et al.*, 2002 ; Windsor et Agerholm, 2009 ; Agerholm, 2007). Cependant, l'intégration complète de ce type de cellule dans les programmes de sélection, qui peuvent agir indépendamment des associations d'élevage, des gestionnaires de race ou des coopératives de vente de semence, est plutôt rare. Le succès de notre cellule repose sur quelques facteurs qui lui sont propres. Le premier tenant à la superficie de la Belgique qui permet à une seule personne de se déplacer dans tout le pays de façon à récupérer rapidement et à examiner un maximum de cas, ce qui améliore la cohérence du phénotypage de cas rares et isolés. A titre d'exemple, il serait impensable en France, de faire régulièrement 500 kilomètres pour aller chercher un cas, alors que pour la Belgique la distance maximale est d'environ 200 kilomètres, soit deux heures et demie de route. De plus, la race BBB est principalement concentrée sur le territoire Belge. Le deuxième facteur repose sur le fait que 99 % des vêlages dans la race BBB se font par césarienne et donc en présence du vétérinaire, facilitant la détection et le renseignement d'anomalies congénitales visibles à la naissance. Et enfin, les éleveurs et les vétérinaires ont rapidement pris conscience de la nécessité et de l'intérêt de déclarer leurs cas d'anomalies congénitales en toute transparence afin d'essayer d'en déterminer la cause et de pouvoir les gérer au plus vite. Toutes les informations et les échantillons biologiques collectés ont été centralisés au sein de l'Unité de Génomique Animale favorisant par là une analyse immédiate et efficace des différentes cohortes.

Le réseau mis en place dans le cadre de cette cellule a également été utilisé dans le cadre de maladies émergentes comme la maladie de la langue bleue (FCO = Fièvre Catarrhale Ovine) ou plus récemment la maladie de Schmallerberg (SBV). Le recensement rapide de veaux présentant des malformations congénitales a contribué à la compréhension de la pathogénie de ces virus lors de l'infection de bêtes gestantes (Desmecht *et al.*, 2008 ; Garigliany *et al.*, 2012 ; Bayrou *et al.*, soumis pour publication).

Le phénotypage et l'analyse épidémiologique, des étapes clés.

Dans la mesure du possible, tous les cas morts collectés en ferme ont été rapatriés à la FMV de Liège afin de réaliser les examens complémentaires adéquats et nécessaires à la caractérisation la plus complète de chaque individu. Les examens complémentaires réalisés ont été limités à la biochimie, l'imagerie médicale (radiographie, échographie et éventuellement scanner) et surtout à l'anatomo-pathologie et l'histopathologie avec des examens systématiques et approfondis.

La race BBB est une race locale qui utilise massivement l'IA (environ 50 % des naissances). Sa sélection intense pour augmenter la conformation, sa création récente par croisements consanguins à partir d'une poignée de taureaux dans les années 60' (Figure 7) et son livre généalogique bien tenu, permettent une analyse rapide du pedigree des cas d'une cohorte d'animaux présentant un phénotype similaire. Cette analyse de pedigree a pu être réalisée facilement avec la consultation de la base de données des pedigrees disponibles en ligne pour les taureaux d'IA (http://www.hbbbbb.be/index_et_pedigree.php) et sur demande pour les autres individus inscrits. Un mode de transmission récessif est suspecté lorsque l'on peut retracer, dans le pedigree du cas, l'origine du phénotype du côté paternel et du côté maternel vers un ancêtre commun. Avec le faible nombre de taureaux fondateurs fortement utilisés dans le pedigree BBB (Figure 7), il est très fréquent de remonter facilement vers plusieurs ancêtres communs, ce qui peut erronément suggérer une composante héréditaire.

Dès lors, quand une cohorte des cas les plus similaires est constituée et que sur base de l'analyse épidémiologique, un mode de transmission autosomal récessif est suspecté, le génotypage à l'aide de puces à SNPs de moyenne densité devient une évidence.

La cartographie fine, une étape rapide et efficace.

Comme démontré par Charlier et ses collaborateurs (2008), le génotypage du génome entier est devenu une étape rapide et efficace dans la gestion des défauts mendéliens chez les animaux de production. Un nombre très faible d'individus, environ cinq, d'une cohorte suspecte, est suffisant pour cartographier le locus associé avec le défaut étudié. L'augmentation du nombre d'individus génotypés permet surtout de réduire l'intervalle et par conséquent le nombre de gènes. Pour les sept maladies étudiées dans la race BBB, la taille de l'intervalle était comprise entre 1,2 et 3,6 Mb avec un nombre de gènes compris entre 10 et 100 (Tableau 2). Malgré l'augmentation du nombre de cas génotypés, la structure du génome bovin avec ses longs blocs en déséquilibre de liaison (Farnir *et al.*, 2000) n'a pas permis de réduire drastiquement la taille de ces intervalles.

La constitution de cohortes de cas très similaires est donc nécessaire et déterminante pour le succès de la cartographie de ces défauts. Les anomalies i) ayant une expressivité variable comme le SQT, ii) ayant des phénocopies (phénotypes semblables d'étiologies différentes) comme la DMC 1 et la DMC 2, qui ont été *a priori* considérées par certains comme des expressions différentes d'un même défaut, iii) dont l'apparition des premiers cas est concomitante avec l'émergence d'autres maladies non génétiques provoquant des anomalies congénitales, peuvent la compliquer. La dernière situation a été rencontrée à deux reprises dans le cadre de la cellule d'hérédo-surveillance. Une première fois, lors de l'augmentation du nombre de veaux qui présentaient un hamartome gingival congénital et/ou du brachygnathisme. Si le virus de la bluetongue (BTV-8 = *Bluetongue virus* - 8) était peu décrit à l'époque, d'autres membres de la famille de ce virus étaient connus pour engendrer du brachygnathisme congénital (Toussaint *et al.*, 2007 ; Sartelet *et al.*, soumis pour publication). La seconde occurrence s'est présentée plus récemment avec le syndrome d'arthrogrypose létale d'origine génétique et les formes d'arthrogrypose sévère dues à la contamination *in utero* par le virus de Schmallenberg (Bayrou *et al.*, soumis pour publication ; Sartelet *et al.*, en préparation). La présence de phénocopies d'étiologies multiples, y compris multifactorielles, infectieuses, toxiques, mécaniques est une source de perte de puissance dans les résultats de cartographie. Les méthodes statistiques de cartographie ont démontré leur puissance à exclure ces cas d'une cohorte dont le mode de transmission autosomal récessif est confirmé. Elles offrent également la possibilité, dans le cas d'études de caractères ayant une architecture complexe avec un nombre de loci et un effet pour chaque locus indéterminé, d'identifier au sein d'un groupe un sous-groupe d'individus dont le phénotype est expliqué par la présence d'un seul locus ayant un effet majeur. L'étude menée sur des individus présentant un retard de croissance majeur comme seul symptôme, caractère normalement considéré comme complexe, a localisé une région IBD homozygote associée de

manière significative à ce phénotype chez 14 des 33 cas initialement génotypés (Sartelet *et al.*, 2012).

A contrario, la cartographie a également exclu le mode de transmission autosomal récessif simple pour certaines cohortes collectées possédant une épidémiologie suggestive de ce mode d'hérédité pour des anomalies congénitales bien connues, telles que les atrésies digestives, la cardiomyopathie dilatée, les *perosomus elumbis*, etc. (Annexe 5).

L'identification de la mutation causale reste l'étape limitante.

L'identification de la mutation causale reste à ce jour l'étape limitante de cette démarche. Le temps entre la cartographie et l'identification de la mutation causale peut varier de quelques mois à quelques années, comme c'est le cas pour la gestation prolongée où la mutation causale reste non validée à ce jour (Tableau 2). Les principaux facteurs à l'origine de cette difficulté sont la taille de l'intervalle et le nombre de gènes qu'il contient, l'absence de gène candidat fonctionnel vis-à-vis des phénotypes observés, l'annotation de la région cartographiée et le type de mutation. L'absence de gène candidat en fonction des mutations connues dans d'autres espèces et contenues dans les régions orthologues ne permet pas d'ordonner les gènes candidats par ordre d'importance et de choisir les plus intéressants. Lorsqu'en l'absence de gène candidat fort, le nombre de gènes devient trop élevé, le séquençage systématique des régions codantes devient long et contraignant. La méthode classique qui consiste à séquencer les exons et les jonctions introns-exons n'identifie que les mutations situées dans cet espace.

Le développement des méthodes de séquençage à haut débit de l'ADN et de l'ARN, a accéléré et amélioré cette étape par le reséquençage du génome entier, par la capture de la région d'intérêt (*ex.* région IBD homozygote) ou par le reséquençage du transcriptome. De plus, cette méthode permet également d'identifier des mutations localisées à l'extérieur des régions codantes (Charlier *et al.*, 2012 ; Sartelet *et al.*, soumis pour publication ; Sartelet *et al.*, en préparation).

Mode d'hérédité	Tares	Cas ^a	BTA ^b	Gène ^c	Pourcentage de porteurs		Nombre de cas attendus (/2.000 naissances)
					Tx d'IA	Population	
RECESSIF	DMC1 †	81 (12)	25 (2,1 Mb)	<i>ATPA2A1</i>	15 %	16 %	~10
	DMC2 †	21 (7)	29 (3,6 Mb)	<i>SLC6A5</i>	5 %	5 %	~1
	SQT	193 (8)	19 (2,4 Mb)	<i>MRC2-Δ2</i>	25 %	26 %	~37
				<i>MRC2-T>C</i>	< 0,5 %	<1 %	
	NAN	147 (33)	3 (3,3 Mb)	<i>RNF11</i>	26 %	29 %	~40
	HAM †	63 (37)	25 (1,2 Mb)	<i>CLCN7</i>	10 %	10 %	~5
	GP †	63 (32)	18 (1,6 Mb)	-	15 %	12 %	~8
COMPLEXE	AP †	23 (15)	10 (2,3 Mb)	<i>PIGH</i>	5 %	5 %	~1
	WHD	205 (205)	5 + ?	<i>Locus Rouan</i>	-	-	-
	PS	60	19 + ?	<i>MRC2</i> (30 %)	-	-	-
	NAN	147 (33)	3 + ?	<i>RNF11</i> (40 %)	-	-	-

Tableau 2: Identification, localisation et épidémiologie des tares à composante héréditaire en race BBB. ^a Nombre de cas collectés par la cellule d'hérédo-surveillance et, entre parenthèses, le nombre d'échantillons génotypés pour réaliser la cartographie. ^b Pour les caractères complexes, gène majeur impliqué et, ^c entre parenthèses, la proportion des cas expliqués par ce gène. - : gène non connu, † : mortinatalité. BTA : Bos Taurus, correspond au numéro du chormosome, DMC : dystonie musculaire congénitale, SQT : syndrome de la queue tordue, NAN : nanisme, HAM : hamartome gingival, GP : gestation prolongée, AP : arthrogrypose, WHD : maladie des génisses blanches et PS : parésie spastique.

Une sélection assistée par marqueurs compliquée.

Après la validation de la mutation et la mise au point d'un test diagnostique indirect (basé sur l'haplotype) ou direct (basé sur la mutation), une sélection assistée par marqueurs (SAM) visant à identifier les individus porteurs et à éviter les accouplements à risques est possible. Une fois le test mis en place, cette SAM poursuit deux objectifs. Le premier, à court terme, permettra d'éviter 9 mois plus tard (temps moyen de gestation), la naissance de veaux mutés. Le deuxième objectif, plus ambitieux et à plus long terme, pourra permettre l'éradication complète des mutations dans la population. En Belgique, contrairement à d'autres pays comme l'Allemagne, il n'y a eu aucune consigne que ce soit au niveau de la race (Herd-book Blanc-bleu Belge), de la Région ou de l'Etat quant à la gestion de ces défauts (Schutz *et al.*, 2008). Initialement et spontanément, les éleveurs et les CIAs (Tableau 3) ont écarté les mâles porteurs de la reproduction dès la mise à disposition des tests pour DMC 1 et 2, respectivement en décembre

2005 et décembre 2006. Pourtant, suite au développement des tests, respectivement courant 2008 et courant 2009 pour le SQT et le nanisme, cette stratégie d'élimination systématique des taureaux porteurs devenait problématique. En effet, l'élimination pure et simple des reproducteurs porteurs réduisait de plus de la moitié la population des taureaux d'IA (Tableau 4). Si en majorité, les reproducteurs étaient porteurs d'une ou deux tares sur les sept identifiées, moins de la moitié était totalement indemne. La conséquence de cette diminution drastique du nombre de reproducteurs favorisait d'une part l'augmentation de la consanguinité par diminution de la taille efficace de la population et d'autre part l'utilisation accrue de lignées dites « indemnes ». Cette méthode de SAM, qui vise à maîtriser les défauts génétiques et à améliorer la rentabilité des élevages, va dans ce cas favoriser l'apparition de nouveaux défauts et également faire reculer l'amélioration génétique de la race BBB (Figure 7 & 8). L'arthrogrypose létale en est un bon exemple, puisqu'avec seulement 5 % de porteurs, ce défaut a rapidement émergé après l'éradication du SQT et du nanisme. Cette apparition soudaine est liée à l'utilisation de taureaux d'une part négatifs pour DMC1 et 2, SQT et nanisme et d'autre part ayant des origines différentes de celles incriminées pour ces quatre défauts. L'ancêtre commun fondateur incriminé, *Kalimine du Barsy Fontaine*, un taureau né en 1980, est peu présent dans les origines du BBB (Figure 6). De plus, il est assez éloigné dans les pedigrees des individus actuels que pour ignorer la consanguinité sur ce taureau lors des accouplements (Sartelet *et al.*, en préparation).

L'expérience acquise en BBB dans la gestion des tares nous démontre que l'éradication pure et simple des porteurs n'a de sens que lorsqu'une race n'est affectée que par très peu de défauts génétiques et/ou que la fréquence de la mutation est faible dans une population. Lorsque le nombre de défauts et/ou la fréquence deviennent élevés, ce type de SAM devient problématique.

Année de naissance	DMC 1 (559)	DMC 2 (556)	SQT (539)	NAN (528)	HAM (532)	GP (522)	AP (521)
< 2000	17	7	32	17	4	18	1
2000	10	3	41	36	7	11	0
2001	17	4	39	23	8	5	0
2002	14	5	46	28	15	8	5
2003	12	1	32	26	15	12	6
2004	2	5	29	24	11	10	2
2005	2	2	20	26	7	18	5
2006	0	0	13	33	5	5	5
2007	0	0	0	30	14	6	3
2008	0	0	0	6	9	6	3
2009	0	0	0	0	3	3	0
2010	0	0	0	0	0	0	0

Tableau 3: Taux de porteurs (%) pour chaque défaut par année de naissance des taureaux d'IA. Pour chaque défaut, le nombre entre parenthèses correspond au nombre de taureaux génotypés pour ce défaut. La ligne rouge correspond à l'année de naissance des taureaux ayant deux ans, moment de la mise en production de semence, à la mise en place du test de chaque défaut. Exemple: le test diagnostique pour le SQT a été disponible en 2008, et les taureaux sélectionnés pour produire de la semence à ce moment là sont nés, au plus tard aux environs de 2006. DMC = Dystonie Musculaire Congénitale, SQT = Syndrome de la Queue Tordue, NAN = Nanisme, HAM = Hamartome, GP = Gestation prolongée, AP = Arthrogrypose.

De façon inattendue, le SQT et le nanisme atteignaient plus de 25 % de porteurs dans la population. Des fréquences aussi élevées pour des mutations récessives permettent de suspecter une sélection, dite « balancée » de celles-ci dans la population. Pour le SQT, nous avons testé cette hypothèse en réalisant des simulations avec des taux de transmission de la mutation différents (pour simuler ou non un avantage sélectif) dans le pédigrée de 160 individus descendants de *Précieux de Somme* (taureau fondateur) et dont 45 étaient porteurs. Le plus grand nombre de simulations, pour obtenir 45 porteurs sur 160 individus, correspondait à un taux de transmission de 67 % alors qu'en l'absence de sélection, nous nous attendrions à obtenir 50 %. Ces simulations nous prouvent que cette fréquence élevée n'est pas arrivée par hasard et que les porteurs de la mutation du SQT ont été sélectionnés préférentiellement par les éleveurs. Ensuite, sur base d'une comparaison des cotations linéaires et de la performance des descendants de taureaux porteurs et non porteurs, il a été démontré pour le SQT que la mutation causale confère un avantage sélectif aux hétérozygotes en améliorant la conformation et la musculature, critères recherchés par les sélectionneurs, ceci au détriment de la taille et des aplombs (Fasquelle *et al.*, 2009). On pourrait supposer que dans une certaine mesure le BBB a atteint une limite physiologique de

développement musculaire et que la moindre augmentation de conformation peut être à la source d'une augmentation du risque de l'altération de la rentabilité du BBB, en termes de mortinatalité, de fertilité, de croissance, *etc.* Concernant le nanisme, il a également été démontré indirectement, de la même manière que pour le SQT, que la mutation causale a été sélectionnée préférentiellement dans les descendants de *Galopeur des Hayons* par les éleveurs et qu'elle confère un avantage sélectif chez les hétérozygotes. Par contre, aucun phénotype sélectionné n'a pu être mis en évidence sur base de la comparaison des valeurs des cotations linéaires calculées pour les performances propres des individus ou celles de la descendance (*progeny test*). Cet avantage pourrait être lié à une certaine résistance aux maladies, plus particulièrement virales comme cela a récemment été suggéré (Charoenthongtrakul *et al.*, 2013). Dans ces cas de sélection dite « balancée », cela prend plus de temps pour éliminer définitivement ces individus qui auraient un avantage sélectif.

Année de naissance ^a	Nombres de mutations					
	1	2	3	4	≥ 5	0
< 2000 (92)	34	24	4	0	0	38
2000 (27)	37	26	7	0	0	30
2001 (22)	50	14	5	0	0	32
2002 (38)	24	21	8	3	3	42
2003 (66)	48	20	6	0	0	26
2004 (39)	36	15	5	0	0	44
2005 (38)	37	16	5	0	0	42
2006 (38)	29	19	3	0	0	58
2007 (35)	46	3	0	0	0	51
2008 (33)	12	0	3	0	0	85
2009 (40)	0	2	0	0	0	98
2010 (30)	3*	0	0	0	0	97

Tableau 4: Proportion (%) de taureaux portant 0, 1, 2, 3, 4 ou ≥ 5 mutations ou allèles à risque par année de naissance pour 499 taureaux d'IA, nés entre 1975 et 2010, génotypés pour toutes les mutations/allèles. ^a Année de naissance avec le nombre de taureaux génotypés entre parenthèses, * = correspondant à un taureau génotypé préalablement faussement négatif ayant eu des descendants atteints.

Une éradication rapide des mutants et une amélioration de la mortinatalité.

L'objectif principal de la mise en place de cette cellule d'hérédo-surveillance était d'une part, la détection et d'autre part, la gestion de ces défauts génétiques. La sélection basée sur les mâles est suffisante pour l'éradication complète et rapide des veaux mutants quelques mois après la mise en place des tests diagnostiques respectifs (Figure 23). On peut voir que les atteints disparaissent l'année qui suit la mise en place du test, logiquement au minimum 9 mois (équivalants à la durée d'une gestation) après l'utilisation des derniers taureaux porteurs. Une exception toutefois concerne le nanisme, pour lequel la mutation du gène *RNF11* n'explique que 40 % des cas présentant un retard de croissance (Sartelet *et al.*, 2012). Il est donc logique que ce défaut, normalement considéré comme un caractère complexe, soit toujours recensé par la cellule d'hérédo-surveillance.

Le durée comprise entre la cartographie de la région IBD homozygote associée au défaut et la découverte de la mutation peut prendre des mois, voire des années. Des tests indirects peuvent être développés pour permettre de déterminer les individus porteurs du, ou des haplotype(s) associé(s) aux défauts. Ces tests indirects ou haplotypiques souffrent cependant d'une perte de sensibilité et de spécificité (*cf.* E. Développement d'un test diagnostique, p. 66). Leur mise en place dans la race BBB pour l'hamartome, l'arthrogrypose et la gestation prolongée est globalement aussi efficace tant au niveau de l'éradication, qu'au niveau de la diminution de l'allèle à risque dans la population (Figures 23 et 25). Le risque lié aux faux-négatifs a été accepté par les différents acteurs du BBB. Les faux-négatifs, comme cela a été le cas pour deux taureaux d'IA, *Ecrin de Wihogne*, porteur recombinant de l'hamartome ou *Argan de l'Ecluse*, porteur recombinant de la gestation prolongée, ont rapidement été gérés. L'expérience de l'identification de ces faux-négatifs, grâce à la naissance de veaux mutants pour ces deux défauts a amené les éleveurs à admettre l'idée qu'être porteur de tares n'était pas synonyme de mauvais reproducteur. Les « bons produits » obtenus avec ces porteurs recombinants, comme *Argan de l'Ecluse*, confirment qu'il ne faut pas nécessairement « jeter le bébé avec l'eau du bain ». C'est pourquoi, de nombreux éleveurs continuent de les utiliser en portant une attention particulière aux mères sélectionnées pour ces accouplements.

Un corollaire un peu gênant, du succès immédiat de ces tests indirects, est la difficulté d'identifier *a posteriori* des homozygotes mutants. En effet, du matériel biologique spécifique d'individus mutants est nécessaire pour étudier l'impact fonctionnel de la mutation découverte. Il est donc suggéré, dans la mesure du possible de rapidement disposer d'une biobanque contenant un panel assez large de tissus dans différents milieux de conservation en vue de ces analyses. Une telle biobanque est souvent difficile à générer dans le cas de tares ayant une mortinatalité élevée, comme c'est le cas pour l'arthrogrypose ou la gestation prolongée.

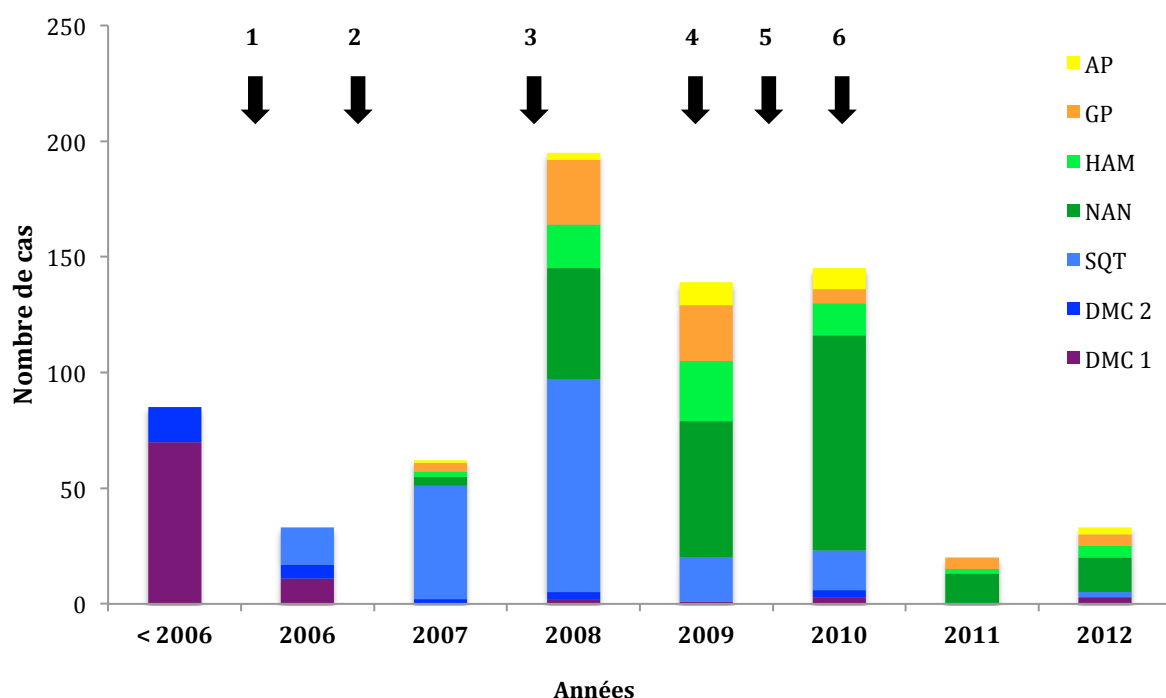


Figure 23: Evolution du nombre de cas rapportés par année à la plateforme d'hérédo-surveillance entre 2006 et 2012 pour la dystonie musculaire congénitale de type 1 (DMC 1) en violet, la dystonie musculaire congénitale de type 2 (DMC 2) en bleu foncé, le syndrome de la queue tordue (SQT) en bleu clair, le retard de croissance et le nanisme (NAN) en vert foncé, l'hamartome (HAM) en vert clair, la gestation prolongée (GP) en orange et l'arthrogrypose (AP) en jaune. Les fleches noires correspondent à la date de mise en place des tests diagnostiques pour 1 = DMC1, 2 = DMC2, 3 = SQT#1, 4 = Nanisme, 5 = hamartome, gestation prolongée et arthrogrypose et 6 = SQT#2.

Etant donné que la plupart de ces tares (DMC 1 et 2, hamartome, gestation prolongée et arthrogrypose) sont létales entre 0 et 6 mois, que 25 % des SQT sont euthanasiés avant 6 mois et que 40 % des nains meurent avant 6 mois, l'éradication de ces défauts devrait améliorer le taux de mortalité dans les élevages BBB. Pour tester cette hypothèse nous avons analysé, à l'aide des données SANITEL (Système d'Identification Animale), le taux de mortalité dans la population BBB. Les données SANITEL sont les seules données disponibles en Belgique pour suivre le mouvement des animaux de production. Toutefois, ce système souffre d'un manque de précision puisque la race n'y est pas renseignée. Sur les passeports des bovins, nous disposons entre autres de l'information relative à l'espèce (bovin, ovin, caprin), du type racial (viandeux, laitier ou mixte), de la date de naissance, de la couleur de la robe, de la date de sortie du troupeau, son motif (vente, abattoir, clos d'équarrissage) et bien sûr l'identification auriculaire (Annexe 7). Les animaux doivent obligatoirement être inscrits dans les sept jours qui suivent la naissance. Sachant que 90 % des individus de type viandeux sont de race BBB, dans les données SANITEL, les bovins de type viandeux sous robe blanche, pie-bleue ou bleu-pie et pie-noire ou noir-pie ont été repris comme appartenant à la race BBB. Le taux de mortalité (nombre de morts/nombre de naissances) chez

les veaux de 0 à 6 mois dans la population « BBB » a été comparé à celui des veaux de la population laitière entre 2002 et 2010, comprenant respectivement 1.461.008 et 752.262 naissances. Il apparaît que depuis la mise en place du premier test diagnostique en 2005, le taux de mortalité en BBB n'a fait que diminuer. Une seule exception, en 2006 et 2007 avec l'apparition en Belgique du virus de la langue bleue responsable d'avortements et de mortinatalité (Toussaint *et al.*, 2007 ; Desmecht *et al.*, 2008). Le taux de mortalité dans la race BBB est passé d'environ 10 % en 2002, à 7 % en 2010 (Figure 24). Dans un premier temps, le nombre de veaux mutants attendus pour chaque tare est estimé en fonction de la fréquence de la mutation, du nombre de veaux nés et de l'année de la mise en place de chaque test (Tableau 2). Pour la fréquence de chaque mutation/allèle à risque, nous avons tenu compte du fait que 50 % des veaux sont issus d'une IA et que les autres 50 % sont le produit d'une saillie naturelle dont la fréquence est estimée à partir des individus génotypés en routine par le laboratoire de l'UAG. Enfin, la proportion de veaux morts due à ces maladies héréditaires, en fonction de leur pourcentage de mortalité, est comparée entre l'année où le test est mis à disposition et l'année suivante (Figure 22). On constate que ces diminutions du taux de mortalité sont du même ordre de grandeur que celui du taux de mortalité expliqué par chaque défaut.

Cette analyse n'est qu'une estimation et souffre malheureusement de certaines carences. Tout d'abord, le système SANITEL ne tient pas compte de veaux non-bouclés, ce qui peut représenter une proportion non négligeable de veaux nés morts ou morts dans les sept jours après la naissance. Ces veaux ne sont pas repris dans l'inventaire des élevages et donc dans le calcul du taux de mortalité. La SAM évite les accouplements à risque et élimine la mortinatalité liée aux défauts héréditaires. Les veaux vivants supplémentaires inscrits, obtenus grâce à la SAM sont quant à eux pris en compte et augmentent le nombre de veaux inscrits et diminuent par là le taux de mortalité. L'absence d'indication de la race dans le système est également une source de biais. Parallèlement à la mise en place de la SAM pour la gestion de ces mortalités néonatales d'origine génétique, le HBBBB a mis une priorité sur l'amélioration des caractères fonctionnels chez les veaux tels que la vitalité, la mortalité, l'appétit à boire, *etc.*, dans les critères de sélection (Annexe 4). Les taureaux ayant de mauvaises évaluations pour ces caractères lors des tests sur la descendance ont été écartés du programme de sélection, réduisant ainsi le taux de mortalité.

Même s'il est difficile de conclure que l'éradication de ces sept défauts a significativement amélioré la rentabilité des élevages en diminuant le taux de mortalité, la disparition rapide des cas et l'amélioration globale du taux de mortalité en BBB sont favorables et renforcent l'intérêt de cette cellule. Pour conclure, il n'est pas rare d'entendre dans les fermes : « *si vous n'aviez pas fait quelque chose pour les tares, le BBB serait mort...* », ce qui a pour moi tout autant de valeur que les chiffres précédents.

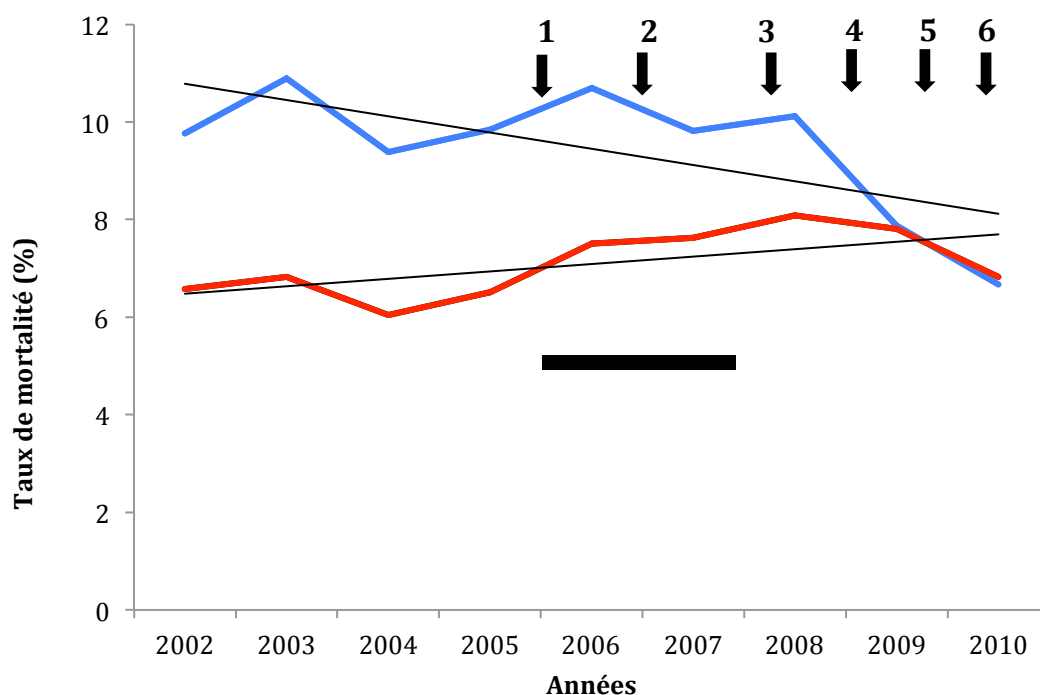


Figure 24: Evolution du taux de mortalité (%) en Wallonie entre les années 2002 et 2010 pour des veaux de 0 à 6 mois dans la population de veaux Blanc-bleu Belge (n = 1.416.008) (courbe bleue) et la population de veaux laitiers (n = 752.262) (courbe rouge) et leur droite de régression correspondante (noire). La barre noire correspond à l'émergence de la langue bleue dans le nord de l'Europe entre 2006 et 2008). Les flèches noires correspondent à la mise en place des tests pour DMC 1 (1), DMC 2 (2), SQT-del2bp (3), nanisme (4), hamartome, gestation prolongée et arthrogrypose (5) et SQT-T>C (6).

Une diminution lente de la fréquence des porteurs.

La gestion des défauts génétiques dans la race BBB se fait principalement par l'éradication des reproducteurs mâles porteurs (92 %) et de quelques vaches de haut potentiel génétique susceptibles de devenir des donneuses d'embryons (8 %). On observe que quelle que soit la fréquence initiale de la mutation, elle diminue lentement comme attendu pour ce type de sélection (Thompson *et al.*, 2006) (Figure 25). La très faible sélection contre les femelles porteuses, et l'utilisation de certains mâles testés pour uniquement l'une ou l'autre tare, voire même non testés, contribuent au maintien de ces mutations dans la population BBB. Pour les défauts ayant une fréquence initiale élevée (25 %), elle est à l'heure actuelle toujours de 15 % comme pour le nanisme et le SQT. Pour les défauts ayant une fréquence initiale de 10 – 15 %, la fréquence actuelle est d'environ 5 % (Figure 25). Une analyse de cette diminution à plus long terme, d'ici une dizaine de générations, permettra d'évaluer la SAM mise en place en BBB.

Cette estimation de l'évolution de la fréquence des porteurs n'est pas optimale puisqu'elle ne tient pas compte de la date de naissance, pour laquelle l'information n'est pas disponible au laboratoire, alors que tous les animaux ne sont pas testés au même âge. De plus, avant la mise en

place des tests groupés directs et indirects, les éleveurs avaient la possibilité de choisir « à la carte », d'ordonner les tests en fonction de la fréquence de la mutation dans la population, de l'historique de la ferme et/ou de la présence de porteurs connus dans le pedigree de l'individu. Cela leur permettait surtout de diminuer le coût du génotypage pour des individus, porteurs d'une ou plusieurs mutations, qui ne seraient pas valorisés dans le commerce en tant que reproducteur. Dans la plupart des cas, le génotypage pour le SQT et le nanisme étaient les plus fréquemment demandés dans un premier temps à cause de leur fréquence élevée et de l'impact de ces défauts dans la population. Les mutations comme DMC 1 et 2 étaient de moins en moins demandées car déjà considérées de manière erronée comme éradiquées. Il serait donc intéressant de tenir compte de l'année de naissance des animaux génotypés mais aussi de prendre en considération les individus génotypés pour toutes les mutations/allèles à risque connus afin de mieux estimer l'évolution de la fréquence de porteurs. De manière plus simple, il serait intéressant d'imposer le génotypage de tous les mâles pour toutes les mutations connues pour accélérer la diminution de toutes les mutations.

En ce qui concerne les taureaux d'IA, les porteurs ont rapidement disparu dès la disponibilité des tests (Tableau 3).

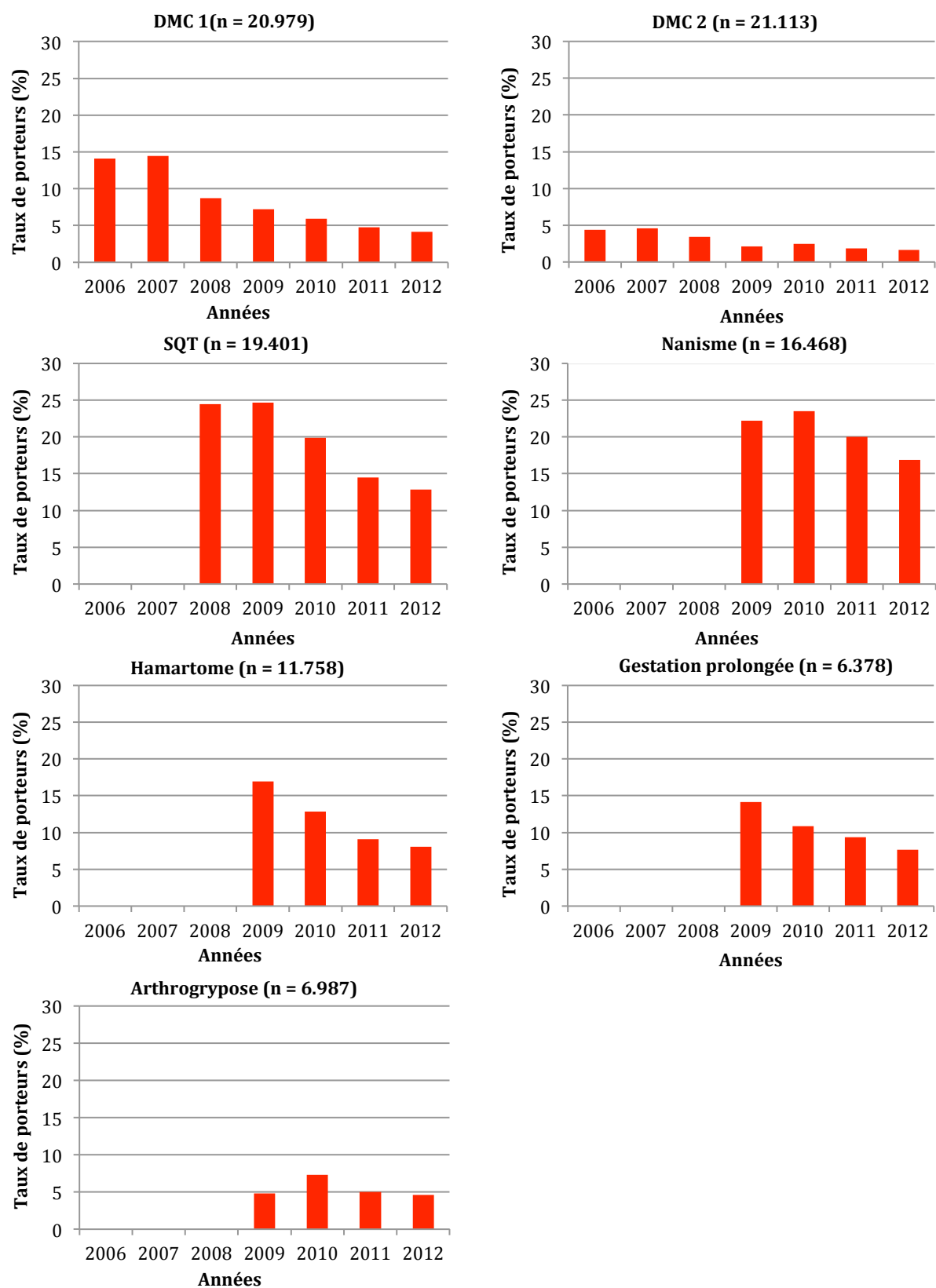


Figure 25 : Evolution du taux de porteurs (%) pour les défauts mendéliens caractérisés dans la race BBB depuis la mise en place des tests indirects (basés sur les haplotypes) ou directs (basés sur la mutation) correspondants dans la population référée au laboratoire de l'UAG pour génotypage. n correspond au nombre d'individus génotypés.

L'étude de caractères complexes reste un défi.

Les caractères complexes affectant la rentabilité des élevages sont nombreux et leur compréhension reste un défi pour les généticiens. Leur dissection et leur caractérisation ont pour objectif, en incorporant ces découvertes dans les programmes de sélection, de pouvoir déterminer les individus améliorateurs ou au contraire à risque. Parmi ceux-ci on retrouve principalement les caractères de résistance aux maladies, de croissance, de fertilité, de production laitière, de qualité de la viande, *etc.*

Le retard de croissance en BBB est un défaut qui a vu une augmentation au cours de ces dernières années. Une cohorte de 147 bovins BBB présentant un retard de croissance prononcé sans aucun autre symptôme majeur et sans étiologie permettant d'expliquer ce retard a été constituée. De manière surprenante, une mutation autosomale récessive a été identifiée. Elle explique près de 40 % des cas de cette cohorte (Sartelet *et al.*, 2012), ce qui peut donner de nouvelles hypothèses pour l'étude de ce genre de caractères. De plus, cette mutation dans le gène *RNF11* interfère également dans la résistance aux maladies, puisqu'elle intervient dans le mécanisme d'auto-contrôle du système inflammatoire. L'effet pléiotrope de cette mutation sur la croissance et la résistance aux maladies ouvre des perspectives dans l'approfondissement de l'étude de ces mutations dites autosomales récessives, qui interviendraient également en partie dans le déterminisme de certains caractères complexes (Charlier *et al.*, 2012 ; Sartelet *et al.*, 2012). Concernant les 60 % restants non expliqués par l'invalidation du gène *RNF11*, l'étude d'association du génome entier n'a identifié aucun autre locus à risque. On peut dès lors suspecter une architecture complexe avec un nombre élevé de loci expliquant chacun une faible proportion de la variance génétique. Les individus souffrant d'un retard de croissance, en excluant les causes infectieuses et métaboliques ainsi que les homozygotes mutants *RNF11*, sont toujours collectés et pourront faire l'objet d'études ultérieures.

La maladie des génisses blanches (WHD = *White Heifer Disease*) est un défaut affectant le tractus génital et donc la fertilité des génisses. Les études épidémiologiques réalisées par le Professeur Hanset dans les années 60' ont permis de d'identifier le caractère héréditaire complexe de ce défaut et l'influence majeure de la couleur blanche de la robe déterminée par une mutation au locus *Rouan* héritée de la race Shorthorn (Hanset, 1969 ; Charlier *et al.*, 1996). Afin de localiser les loci à risque, susceptibles d'intervenir dans l'architecture de la WHD, deux-cent-six génisses affectées par la WHD ont été examinées par fouiller rectal et vaginoscopie puis inventoriées en fonction des anomalies présentes selon la classification déterminée par le Professeur Hanset (Hanset, 1969). Les ovaires et leur activité, la présence ou non de(s) corne(s) utérine(s), la présence d'un mucocolpos (accumulation de liquide dans une corne utérine) ont été identifiés grâce à la palpation transrectale. La présence d'un vestibule aveugle, qui dans le

cas de WHD est d'une dizaine de centimètres, a été confirmée par vaginoscopie. La WHD est à différencier du free-martinisme qui affecte les génisses jumelles d'un mâle et qui se distingue entre autres par l'absence de fonction ovarienne, un tractus rudimentaire et un clitoris hypertrophié. Au niveau épidémiologique, l'incidence de la WHD est devenue relativement faible de nos jours, de l'ordre de moins de 1 % (environ une à deux génisses par an par lot de génisses dans les fermes ayant renseigné une WHD, observations personnelles) contrairement aux années 60' où elle touchait 10 à 15 % des individus de robe blanche, 1 à 1,5 % des individus de robe bleue et 0,5 % des individus de robe noire (Hanset 1969). Cette diminution est principalement due à l'exclusion des génisses atteintes de la reproduction mais aussi au changement de « mode » pour la couleur de la robe. Alors que dans les années 60' la robe blanche était majoritairement recherchée et la noire quasiment interdite, de nos jours les individus pie-noirs voire même entièrement noirs sont préférentiellement sélectionnés... (cf. b) Robe, p. 29). Parmi les trois types de WHD définis par le Professeur Hanset, le type A reste toujours le type le plus fréquemment observé puis le type C et enfin le type B (cf. E. La race BBB et ses défauts, p. 39) (Hanset, 1969).

		COULEUR DE LA ROBE				Hanset, 1969
		Blanche	Pie-bleue	Pie-noire	TOTAL	
TYPES	A	150	11	0	161 (78 %)	23 (59 %)
	B	10	0	0	10 (5 %)	9 (23,1 %)
	C	32	3	0	35 (17 %)	7 (17,9 %)
TOTAL		192 (93 %)	14 (7 %)	0	206	39

Tableau 5 : Répartition des 205 cas de maladie des génisses blanches collectées en fonction de la couleur de la robe et du type (selon la classification du Prof. R. Hanset, 1969) et comparaison avec les proportions obtenues en 1969 dans une étude du Prof. R. Hanset.

L'ADN extrait à partir d'un échantillon sanguin collecté lors de l'examen a été génotypé à l'aide des puces bovines de 50.000 marqueurs développées par l'Unité de Génétique Animale (Charlier *et al.*, 2008). Les études d'associations de type génome entier ont été réalisées sur les haplotypes (Druet et Georges, 2010 ; Zhang *et al.*, 2012) à partir de cette cohorte de WHD et de deux populations de contrôles. La première population contrôle est représentée par 365 taureaux d'IA génotypés à l'aide de la puce *Bovine Illumina HD* contenant 770.000 SNPs. Cette population comprend 188 animaux sous robe blanche (53 %), 124 sous robe bleue (35 %) et 44 sous robe noire (12 %) parmi lesquels 32, 26 et 1 ont respectivement eu au moins une fille atteinte. La deuxième population utilisée se compose de 167 demi-sœurs saines (résultats non

présentés) de robe blanche. L'utilisation de cette cohorte contrôle de robe blanche a pour objectif de soustraire l'effet de la robe blanche pour essayer de faire ressortir les loci influençant le déterminisme de la WHD. Cette cohorte de contrôle souffre néanmoins d'une importante stratification, puisqu'elle représente les filles de 26 taureaux d'IA seulement, alors que la cohorte de WHD comporte plus de 120 pères différents. Ces études n'ont, à ce jour pas permis de localiser un ou des loci significativement associés à ce défaut (Figure 26). Comme attendu, l'étude d'association réalisée avec la population contrôle contenant les 365 taureaux d'IA a identifié un locus sur le chromosome 5 contenant le gène *Rouan*, intervenant dans le déterminisme de la robe en BBB (Charlier *et al.*, 1996). Deux autres loci potentiellement intéressants sur les chromosomes 11 et 16 mériteraient d'être répliqués sur une plus grande cohorte, c'est pourquoi je continue à alimenter cette cohorte (Figure 24).

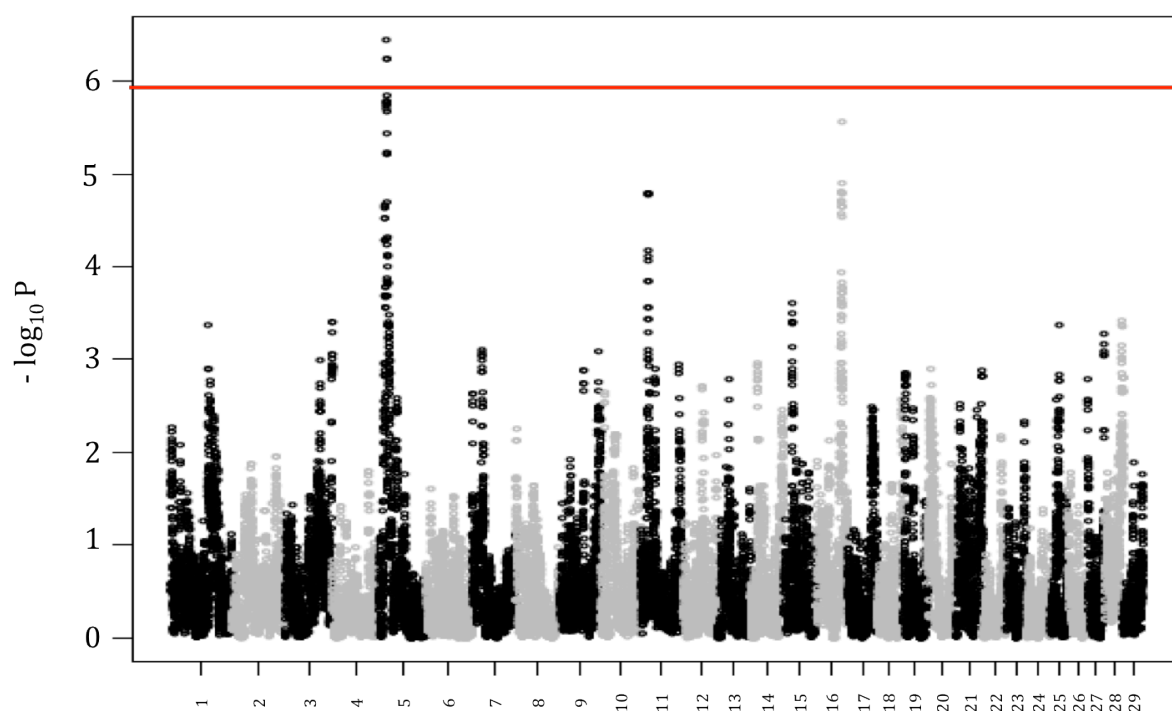


Figure 26 : Manhattan plot résultant de l'étude d'association basée sur les haplotypes avec 188 WHD et une population contrôle de 365 taureaux d'IA pour l'identification du ou des loci associés à la maladie des génisses blanches. L'alternance de gris et de noir correspond à la limite des autosomes de 1 à 29. Le seuil de significativité déterminé par correction Bonferroni avec 34.368 SNPs communs entre la puce de moyenne densité et celle de haute densité à 5 % est de 10^{-6} (ligne rouge).

Comme tous les caractères complexes, le nombre d'individus atteints, nécessaires à la localisation du ou des loci impliqués dans leur déterminisme dépend surtout de l'architecture du caractère étudié. Dans le cas de la WHD, Le nombre de 188 cas analysés s'est avéré insuffisant. L'augmentation du nombre de cas est nécessaire pour augmenter la puissance de détection. Les

analyses réalisées jusqu'ici ne tiennent pas compte des chromosomes sexuels. L'intégration des chromosomes sexuels dans cette étude d'association est donc une des possibilités à envisager.

Dans les caractères complexes, on peut avoir une modification de l'expression d'un génotype en fonction de l'environnement dans lequel l'individu se trouve. On pourrait suspecter que l'environnement utérin influencerait l'expression ou non de la WHD en fonction d'un génotype donné à différents loci incriminés. Récemment, une portion du chromosome Y a été détectée chez des génisses souffrant de troubles de la fertilité, sans anomalies morphologiques et ce même en l'absence de gestation gémellaire mâle-femelle (MacDanel et al., 2012).

La race BBB, une des races les mieux caractérisées au niveau moléculaire.

Avec toutes ces découvertes anciennes et plus récentes, que ce soit pour des caractères zootechniques avec la myostatine et le rouan, ou pour la gestion de défauts héréditaires avec l'identification de sept maladies autosomales récessives, la race BBB est la race viandeuse sûrement la mieux caractérisée au niveau moléculaire. La race BBB est à la production de viande ce que la race Holstein est à la production de lait, tant au niveau de la productivité qu'au niveau de l'implication des éleveurs et des programmes de sélection. Ils ont continuellement travaillé pour améliorer les performances et la rentabilité de ces races de haut niveau de production.

2 - PERSPECTIVES

A mon sens, il est clair que l'existence de cette cellule d'hérédo-surveillance n'est plus à justifier. Il est donc nécessaire, dans un premier temps de poursuivre et de maintenir la vigilance de cette cellule à l'aide des collaborateurs de terrain, vétérinaires et éleveurs, tant au niveau de l'émergence de tares à composante héréditaire, qu'au niveau de l'arrivée de nouvelles maladies infectieuses dans nos latitudes comme il en a été pour la maladie de la langue bleue et le virus de Schmollenberg. Au niveau des défauts génétiques, il est très probable que dans quelques générations, des accouplements consanguins sur des lignées fortement utilisées favorisent l'apparition de nouvelles tares. A titre d'exemple, la lignée d'*Empire d'Ochain* et de son fils *Adajio de Bray* (près de 20.000 doses utilisées en 2011 et environ 40.000 descendants en Wallonie) ou la lignée de *Davina du Haut d'Arquennes* qui compte une bonne dizaine de très bons descendants dans les différents CIAs, ont jusqu'à présent des lignées indemnes des tares connues, mais risquent d'être à l'origine de l'émergence de nouveaux défauts.

De la sélection assistée par marqueurs à la sélection génomique.

Une question récurrente des éleveurs et des CIAs, maintenant que les défauts connus sont gérés, est de savoir s'il n'y a pas moyen de chercher les « bons » gènes. Ceux-ci sont probablement beaucoup plus nombreux et plus difficiles à identifier puisqu'ils n'expliquent majoritairement qu'une faible proportion de la variance phénotypique dans l'architecture de caractères complexes (Georges, 2012). La détection de QTL était une des solutions, néanmoins, à l'heure actuelle, la sélection génomique telle qu'elle est développée dans la race Holstein pie-noire, devient la méthode de choix. La sélection génomique est sûrement l'avenir pour le BBB dans l'ère de la génomique (Druet T., *communication personnelle*). A l'heure où le coût pour génotyper le génome entier d'un individu au moyen d'une puce de moyenne densité est moindre que de le génotyper pour les sept mutations/loci responsables de ces défauts, ne serait-il pas plus intéressant d'utiliser cette méthode et de pouvoir d'une part déterminer les porteurs et d'autre part estimer la valeur génomique de l'individu dès la naissance ? La génomique dans les races viandeuses reste un défi et est plus lente à se mettre en place. Les raisons sont d'une part la taille de la population de référence qui est plus faible dans ces races. D'autre part, les paramètres mesurés en race viandeuse sont moins nombreux, plus difficiles à mesurer et souvent plus subjectifs que dans les races laitières avec les différents paramètres mesurés, liés à la production laitière (Hayes *et al.*, 2013). Même si la sélection génomique sera un plus pour le BBB, de nombreux sélectionneurs vous diront que « *le BBB est une race d'intuition et non de calcul* ». D'ailleurs, peu de sélectionneurs utilisent les cotations linéaires pour choisir leur reproducteur. En conséquence, il est important de sensibiliser les éleveurs de BBB à ce nouveau défi en objectivant la plus-value qu'ils apporteraient à leur élevage et à la renommée du BBB. Il reste à leur démontrer la faisabilité de l'implémentation de la sélection génomique en BBB mais aussi l'importance de leur implication dans la récolte de phénotypes précis pour la constitution d'une population de référence. A l'heure actuelle, les méthodes sont développées, la récolte des phénotypes de la population de référence est en cours et les premiers individus de cette population vont être génotypés.

Les mutations embryonnaires létales et leur impact sur la fertilité.

Rilouke, la cellule d'hérédité-surveillance en BBB visait, dans un premier temps, à recenser des phénotypes, puis à suspecter une étiologie héréditaire pour certains d'entre eux et enfin à localiser la ou les mutations responsables de ces défauts. Au cours de cette dernière décennie, la

baisse de la fertilité en race BBB est devenue un problème majeur. L'identification en HPN de la mutation responsable de cas anecdotiques du syndrome brachyspina mais surtout d'une mortalité embryonnaire des homozygotes mutants se traduisant par un effet négatif sur la fertilité, ouvre de nouvelles perspectives de recherches en BBB concernant les mutations embryonnaires létales (Charlier *et al.*, 2012). Le séquençage haut-débit de bibliothèques préparées à partir de capture d'exomes de taureaux d'élite populaires aurait pour objectif d'identifier des mutations perte-de-fonction dans les régions codantes et d'évaluer ensuite l'éventuelle déplétion en homozygotes mutés dans une population de référence. La fréquence élevée d'une mutation à l'état hétérozygote, en l'absence d'homozygotes mutants traduit une mortalité embryonnaire ou fœtale (Georges, 2012). Dans ce cas, on parle de génétique inverse, vu que l'on part d'une mutation candidate et que l'on réévalue l'effet éventuel du phénotype associé à cette mutation sur la fertilité. C'est pourquoi, similairement aux autres pathologies récessives, un test diagnostique pourrait alors être mis en place afin d'éviter les croisements à risques et améliorer ainsi la fertilité des élevages BBB.

Et pour conclure...

Rilouke, sept ans après sa mise en place, a prouvé que son efficacité n'existe qu'avec une collaboration d'éleveurs passionnés par leur race et de vétérinaires soucieux de comprendre et d'améliorer la rentabilité des élevages. De certaines mutations découvertes, nous avons identifié des effets pléiotropes et de la sélection balancée, rendant la gestion de ces mutations plus difficile. La gestion de ces défauts génétiques, ne se limite pas à l'élimination pure et simple des individus porteurs, mais elle s'intègre entièrement dans un programme de sélection, de conseil et d'accompagnement, visant à continuer la progression de la productivité du BBB, tout en préservant une diversité génétique afin de prévenir l'apparition de nouveaux défauts.

To be continued...

ANNEXES

Annexe 1 : Standards de la race (HBBBB : www.hbbbb.org)

QUALITES	DEFAUTS
TETE	
Légère, bien proportionnée, front plat assez large, mufler large; tête plus large, plus courte et plus massive chez le mâle. Cornes courtes, horizontales, de côté chez les taureaux, recourbées en avant du front chez la vache.	Tête grossière, longue et effilée.
ENCOLURE	
Epaisse et horizontale chez la vache, rebondie et convexe chez le taureau.	
EPAULE	
Bien musclée, en relation harmonieuse avec les régions voisines. Musculature de l'épaule, du bras, de l'avant-bras et de l'angle scapulo-huméral particulièrement développée chez le taureau.	Epaules saillantes et décollées, musculature insuffisante.
GARROT	
Garrot large et musclé, se continuant en ligne droite avec l'encolure, tout au moins chez la vache, et avec le dos.	Garrot étroit, pas assez musclé, mal soudé au dos.
DESSUS : DOS ET REIN	
Horizontal, large et musclé, présence fréquente d'une gouttière médiane au milieu du dos, pouvant se prolonger vers l'arrière (rein double).	Dos ensellé, manque de musculature, mauvaise attache de rein, échine saillante.
FLANC	
Court et plein, le pli de l'aine se prolonge vers l'avant par la corde très épaisse.	Long, creux, paroi abdominale affaissée.
CROUPE	
Large avec les hanches effacées, longue, inclinée, à musculature développée, «double», spécialement chez le mâle, le creux médian étant occupé par le sacrum prolongé par la queue dont l'attache est nette et bien détachée.	Croupe manquant de longueur, de largeur, de musculature, pas assez inclinée, attache de queue noyée.
FESSES ET CUISSSES	
Rebondies et convexes avec sillons intermusculaires apparents; de profil, au départ des pointes ischiales, les fesses prolongent la croupe vers l'arrière en arc de cercle et surplombent le jarret et son tendon. Chez le mâle, de face par derrière, le contour passant par le sommet de la croupe et le bas des fesses tend à prendre la forme d'un cercle.	Musculature insuffisante, fesses trop courtes.
QUEUE	
Développement en rapport avec l'ossature, longueur moyenne, tombe perpendiculairement.	Attache de queue noyée, queue trop courte.

QUALITES	DEFAUTS
MEMBRES & APLOMBS	
Ossature solide et plutôt fine, articulations sèches et souples, onglons sains et résistants. ANTERIEURS : - De profil, l'avant-bras, le genou et le canon forment une colonne droite. - De face, l'avant-bras et le canon forment un angle ouvert en dehors dont le genou est le sommet. POSTERIEURS : - De profil, une perpendiculaire abaissée de la pointe ischiale tombe sur la pointe du jarret. - Vu de derrière, le membre postérieur a une direction parallèle au plan médian du corps. PATURONS : Prolongent la ligne des onglons, de la couronne au boulet; le paturon est naturellement plus droit au postérieur.	Ossature grossière, articulations raides, empâtées, voire tuméfiées. ANTERIEURS : - Les antérieurs portés vers l'avant, le campé du devant, les antérieurs portés vers l'arrière, sous-lui du devant, le genou porté vers l'avant, arqué. - Panard (genoux trop en dedans); cagneux (genoux trop en dehors). POSTERIEURS : - Le campé de derrière, le sous-lui de derrière; l'angle du jarret trop ouvert (jarrets droits), trop fermés (jarrets coudés). - Jarrets fermés ou crochus et animal panard dans le bas des membres; jarrets ouverts et animal cagneux dans le bas des membres. PATURONS : paturons trop peu inclinés (droit jointé); paturons trop inclinés (bas ou long jointé).
TESTICULES	
De volume normal.	Trop petits, tuméfiés, absents (cryptorchidie).
MAMELLES	
Pis carré et symétrique, moyennement développé, trayons bien situés.	Pis pendant, pis en forme de bouteille (pis de chèvre); trayons trop rapprochés.

Annexe 2 : Cotations linéaires (site du HBBBB : <http://www.hbbbb.org>)

	CARACTERE	ECHELLE	INDEX BAS	INDEX ELEVE	h ²
FORMAT	Taille	1 - 50	Plus petite	Plus grande	0,38
	Longueur	1 - 50	Plus courte	Plus longue	0,18
	Largeur de poitrine	1 - 50	Plus étroite	Plus large	0,20
	Largeur de bassin	1 - 50	Plus étroite	Plus large	0,26
CONFORMATION	Peau	1 - 50	Epaisse	Fine	0,17
	Epaule	1 - 50	Moins musclée	Plus musclée	0,30
	Dessus	1 - 50	Moins musclé	Plus musclé	0,30
	Côte	1 - 50	Plate	Arrondie	0,25
	Croupe	1 - 50	Horizontale	Inclinée	0,29
	Longueur de bassin	1 - 50	Plus courte	Plus longue	0,14
	Attache de queue	1 - 50	noyée	Dégagée	0,24
	Fesse de l'arrière	1 - 50	Plates	Très arrondies	0,39
	Fesse de profil	1 - 50	Rectiligne	Très arrondies	0,42
	Ligne de dessus	1 - 50	Ensellée	Convexe	0,14
OSSATURE & APLOMBES	Ossature	1 - 50	Grosse	Fine	0,18
	Epaule : ossature	1 - 50	Saillante	Effacée	0,07
	Antérieurs	1 - 50	Moins panards	Plus panards	0,08
	Paturons antérieurs	1 - 50	Moins long jointés	Plus long jointés	-
	Postérieurs	1 - 50	Ouverts	Crochus	0,08
	Paturons postérieurs	1 - 50	Plus droits jointés	Moins droits jointés	-
	Jarrets	1 - 50	Plus droits	Plus coudés	0,19
	Marche	1 - 50	Très difficile	Très aisée	0,15
NOTES PARTIELLES	Taille	50-100			0,25
	Musculature	50 - 100	Moyenne	Excellente	0,43
	Type viande	50 - 100	Moyenne	Excellente	0,32
	Aplombs	50 - 100	Moyenne	Excellente	0,10
	Apparence générale	50 - 100	Moyenne	Excellente	0,40
NOTE FINALE		50 - 100	Moyenne	Excellente	0,31

Annexe 3: Pedigree et cotations linéaires d'*Harpon de l'Orgelot*, indemne pour les tares connues (BBG)

HERD-BOOK BLANC-BLEU BELGE

A.S.B.L.

Certificat généalogique délivré conformément à la décision 2005/379/CE
de la commission pour les échanges intracommunautaires

A.W.E asbl

Rue des Champs Elysées, 4 - 5590 CINEY (Belgique)
Tél.: 32 Ø 83 23 06 12-16 - Fax: 32 Ø 83 21 28 37

MALE

PEDIGREE OFFICIEL

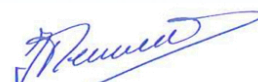
NAISSEUR: 56063026-33 SCOHIEZ André et Bernard	CTI prop. : - PROPRIETAIRE: BBG SCRL chemin du Tersoit, 32 5590 Ciney
No HB : -	DERNIERE MUTATION: 28-08-2012
No ETRANGER : -	
NOM : HARPON DE L'ORGELOT	TATOUAGE : BOUCLE SAUMON: BE 1-55067712 BOUCLE LAITON: BOUCLE TRAV. : 7712
DATE DE NAISSANCE: 14-02-2004	
SEXE : MALE	
ROBE : Blanc	N° ECHANT. GS :
	TYPE: Viandeux
	LIVRE DU H.B.: A 2005

APACHE DE L'ORGELOT 975007720-95 ROBE : Pie-bleu G.S.: 05019363	ISTAMBOUL DE L'ORGELOT 955007140-42 ROBE : Blanc G.S.: 05016928	AGITATEUR DE SOMME 879063870-12
	HARMONIE DE L'ORGELOT 945013241-53 ROBE : Pie-bleu G.S.: 05019454	DEESSE DE L'ORGELOT 905098631-70
EGlantine de l'Orgelot BE 8-54627232 ROBE : Pie-bleu G.S.:	GOLDORAK ET DE FONTENA 959059230-54 ROBE : Blanc G.S.: 13005042	ALBATROS DE BOIS BORSU 896043180-54
	JOSEPHINE DE L'ORGELOT 965011101-05 ROBE : Pie-bleu G.S.: 05020690	ETINCELLE DE L'ORGELOT 915011911-47
		BRUEGEL ET D AU CHENE 936012930-81
		ARDENTE 5797 BOIS BORSU 896561171-65
		MAGICIEN 2277 DU LAURY 929091390-73
		CECILE DE L'ORGELOT 895008141-05

DATE D'EMISSION: 14-09-2012

No D'EMISSION : 7


ing. P. MALLIEU
Secrétaire général


J.-M. WILMET
Président

Les résultats des contrôles de performances et des évaluations génétiques sont disponibles sur le site web du Herd-Book BBB : www.hbbbbb.be

Ce certificat n'est garanti pour exportation
qu'accompagné du certificat de légalisation.

Dit certificaat is enkel dan garantie voor export
als het voorzien is van de H-B BBB stempel.

This certificate is valid for export
when legalize by the H-B BBB stamp.



A.W.E asbl
Contrôle de Performance Viande
Rue des Champs Elysées, 4
B-5590 CINEY
Tel : 083/23.06.11
Fax : 083/23.06.76



HARPON DE L'ORGELOT

Elevage de :

BBG SCRL

chemin du Tersoit, 32

5590 Ciney

Numéro de cti :

Boucle saumon : BE 1-55067712

Robe : Blanc

Tatouage :

Date naissance : 14-02-2004

COTATION LINEAIRE A L'AGE DE 5 an(s) 8 mois 26 jour(s)

		1...5...10...15...20...25...30...35...40...45...50			
Peau	25			NOTES PARTIELLES/100	
Longueur	48			Taille	100.0
Largeur poitrine	29			Musculature	89.0
Largeur bassin	39			Type viande	81.3
Epaule	36			Aplombs	81.0
Dessus	28			Apparence générale	80.0
Côte	20			NOTE FINALE	88.3
Croupe	35			Remarques aplomb :	
Longueur bassin	48			Déviation antérieure	0
Queue	28			Bouleture antérieure	0
Fesse profil	44			Genou fléchi	0
Fesse arrière	41			Déviation postér.	0
Ossature	30			Bouleture postér.	0
Epaule : ossature	25			Soufflette au jarret	0
Ligne dessus	25				
Antérieur face devan	28				
Paturons ant. profil	25				
Postérieur face derr	28				
Paturons pos profil	28				
Jarret	35				
Marche	22				
Taille	50				
Apparence générale	16				
Embonpoint	4				

Taille : 155 cm à 5 an(s) 8 mois 26 jour(s)

ASCENDANCE

NOTES DES PARENTS

ASCENDANCE		NOTES DES PARENTS			
APACHE DE L'ORGELOT 975007720-95	ISTAMBOUL DE L'ORGE	PERE		MERE	
	955007140-42	Taille	93.0	Taille	78.0
EGLANTINE DE L'ORGELOT	HARMONIE DE L'ORGE	Musculature	91.3	Musculature	86.0
	945013241-53	Type viande	83.9	Type viande	81.4
	GOLDORAK ET DE FON	Aplombs	91.0	Aplombs	98.0
	959059230-54	Apparence générale	85.0	Apparence générale	75.0
	JOSEPHINE DE L'ORGE	NOTE FINALE	90.5	NOTE FINALE	85.7
	965011101-05				

Date de classification : 09-11-2009

Le classificateur de l'AWE

Date d'édition : 09-12-2009

GUISSARD Jean-Marc

Nombre d'édition : 1

Annexe 4 : Evaluations génétiques de la race BBB (HBBBB, 2012)

ÉVALUATIONS GÉNÉTIQUES BBB
BWB GENETISCHE EVALUATIES
BB GENETIC EVALUATIONS

Index zootechniques & fonctionnels (1^{ère} et 2^{ème} visites)
Cotations linéaires & Index poids carcasse

Functionele & zootechnische kenmerken (1^{ste} en 2^{de} bezoeken)
Lineaire beoordelingen & Index karkasgewicht

Functional & zootechnical indexes (1st and 2nd visits)
Linear score & Index carcass weight

Juillet - Juli - July
2012

Herd-Book Blanc-Bleu Belge ASBL
Belgisch Witblauw Stamboek VZW
Belgian Blue Herd-Book NPO

www.hbbbbb.be

Index zootechniques et fonctionnels en 1^{ère} visite (naissance):

Nom - Naam - Name	R ² Durée Gestation Drachtduur Gestation length	R ² Poids naissance Geboorte gewicht Birth weight	R ² Conformation Conformatie Conformation	R ² Aptitude à téter Zoogvermogen Suckling ability	R ² Aptitude à boire Drinkvermogen Ability to drink	R ² Vitalité Levenskracht Vitality	R ² Mortalité Sterftecijfer Death rate	R ² Défaut antérieurs Gebrek voorpoten Defect forelegs	R ² Jarrets droits Rechte sprong. Straight hocks	R ² Jarrets coudés Kromme sprong. Bent hocks	R ² Langue sortie Lange tong Protruding tongue	R ² Deviation mäch. Scheve muil Deviation	R ² Prognathisme Snoekmuil Prognathism	R ²
Accord De Wihogne	104	99	71	107	111	110	87	74	76	110	93	99	101	0,85
Adajio De Bray	101	101	121	107	112	110	95	107	98	106	97	99	103	0,93
Agile Pierrecome	104	85	102	103	100	104	109	105	107	102	91	100	109	0,61
Alto 8389 De Reux	96	102	104	110	109	104	87	105	114	85	86	85	97	0,72
Amiral De Ste Helene	90	96	98	98	102	98		103	97	92	105	97	85	0,60
Amor 1843 Vd Lafelderberg	97	95								100	108	104		
Arafat Et Van De Bremberg	99	92	97		88				105		116	108		0,74
Balcon De Berligant	111	105	111					121				101		0,62
Barbier	117	99	91	107	102	106	98	102	82	103	81	81	88	0,56
Baroque De L'ecluse	93	94	109	91	97	99	96	93	113	99	109	107	101	0,96
Bivouac Et 0718 Du Vivier	100	98	110	99	97	95	108	103	115	85	104	105	94	0,93
Blackstar Bringlee Et	97	101	87	106	103	102	102	91	96	107	88	88	95	0,84
Blak 5075 Baty Eprave	108	110	114	111	111	107	97	105	109	100	90	86	98	0,95
Blaky Baty Eprave	98	109	124											
Brilliant Du Pont D'herbais	108	97	101	107	112	103	100	101	99	113	90	95	98	0,82
Cabestan 7791 Grigeoule	104	102	94	106	96	99		99		108	113	122	101	0,61
Canadian Club Kerkenhof.	93	107	123	106	106	100	105	107	104	100	96	88	96	0,79
Cantona Boherard Et	85	97	91	104	112	106	95	104	100	92	100	101	97	0,61
Casanova De Cras Avernas	108	108	120									113		0,55
Climatiseur D'hanotelet	108	104	100								89	87		0,59
Colibri 2701 Du Coin	116	99	97	88	97	98	102	119	108	96	122	113	105	0,86
Colos Van Daisel	96	105	98	102	97	98	105	101	97	113	99	102	96	0,67
Coucou De Cras Avernas	93	91	105							101		90		0,63
Crack Vt Hof Te Landegem	104	113	107											
Crack Du Pont De Messe			97											
Cyclone Bluegrass Et	109													
Dafydd D Ochain	100	95	90	112	114	114	79	91	90	111	102	97	107	0,95
Damier Et De Waleffes	104	94	99	96	102	104	105	97	96	114	102	99	101	0,77
Dartagnan Du Castillon	116	100	90	109	110	109	93	88	102	85	83	82	105	0,96
Davidson De L'ecluse	98	107	97	105	101	103	94	104	102	106	97	100	96	0,76
Debordant Et De Biourge	94	99	108	91	92	94	101	110	105	92	104	113	94	0,96
Decent De Bomal	94	101	100	91	93	91	106	98	103	107	104	111	96	0,56

Index zootechniques et fonctionnels en 2^{ème} visite (14 mois):

2ème VISITE - 2 ^{DE} BEZOEK - 2 ND VISIT																						
Nom - Naam - Name	Conformation Conformatie Conformation	R²	Poids Gewicht Weight	R²	Taille Gestalte Height	R²	Poids - confor. Gewicht - conf. Weight - conf.	R²	Mortalité Sterftecijfer Death rate	R²	Jarrets droits Rechte sprong. Straight hocks	R²	Jarrets coulés Krom. Sprong. Bent hocks	R²	Défaut post. Gebrek achter. Defect rearlegs	R²	Déviat. mach. Scheve muil Deviation jaw	R²	Prognathisme Snoekmuil Prognathism	R²	Brachygnat. Varkensmuil Brachygnatism	R²
Accord De Wihogne	94	0,98	105	0,97	110	0,98	103	0,97	82	0,59	97	0,55	100	0,82	93	0,59	89	0,67	99	0,70	95	0,79
Adajio De Bray	111	0,99	95	0,98	97	0,99	101	0,98			91	0,67	109	0,90	103	0,77	102	0,89	91	0,73	81	0,66
Agile Pierrecome	119	0,93	105	0,90	102	0,94	125	0,92					97	0,55								
Alto 8389 De Reux	94	0,95	96	0,92	99	0,95	89	0,93			119	0,65	93	0,68	104	0,56	98	0,65			93	0,62
Amiral De Ste Helene	106	0,89	87	0,84	96	0,89	83	0,86														
Arafat Et Van De Bremberg	104	0,81	92	0,75	98	0,82	90	0,78											125	0,52		
Balcon De Berligant	115	0,84	90	0,77	97	0,84	96	0,80														
Barbier	84	0,95	98	0,92	103	0,95	84	0,94					98	0,64			101	0,62			108	0,75
Baroque De L'ecluse	107	0,99	90	0,99	99	0,99	88	0,99	97	0,87	110	0,93	103	0,95	102	0,90	96	0,93	92	0,86	109	0,97
Bivouac Et 0718 Du Vivier	112	0,99	105	0,99	91	0,99	119	0,99	105	0,83	111	0,91	97	0,94	102	0,88	101	0,93	92	0,84	106	0,96
Blackstar Bringlee Et	102	0,99	107	0,98	108	0,99	114	0,98	102	0,76	90	0,60	102	0,89	89	0,68	83	0,70	84	0,63	87	0,76
Blak 5075 Baty Eprave	102	0,99	99	0,99	94	0,99	100	0,99	101	0,88	110	0,92	99	0,95	96	0,88	91	0,90	93	0,86	103	0,96
Blaky Baty Eprave	102	0,69	92	0,60	93	0,70	88	0,65														
Brillant Du Pont D'herbais	100	0,97	100	0,95	97	0,97	100	0,96	102	0,58	108	0,72	97	0,80	93	0,58	90	0,66			100	0,81
Cabestan 7791 Grigeoule	92	0,90	96	0,86	97	0,91	87	0,88													110	0,63
Canadian Club Vd Kerkenhofstede	122	0,97	104	0,96	92	0,97	125	0,96			100	0,56	101	0,76	104	0,58	99	0,72	92	0,55	103	0,76
Cantona Boherard Et	84	0,87	85	0,81	97	0,87	62	0,84														
Casanova De Cras Avenas	121	0,63	101	0,56	93	0,64	120	0,60														
Climatiseur D'hanotelet	106	0,72	107	0,63	103	0,73	117	0,68														
Colibri 2701 Du Coin	112	0,97	97	0,95	86	0,97	104	0,96	99	0,59	115	0,75	94	0,77	98	0,61	101	0,79	95	0,59	110	0,87
Colos Van Daisel	113	0,94	102	0,92	97	0,95	114	0,93					109	0,70								
Coucou De Cras Avenas	112	0,85	94	0,79	91	0,86	101	0,82														
Cyclone Bluegrass Et	99	0,64	102	0,53	103	0,65	103	0,58														
Dafydd D'Ochain	88	0,99	104	0,99	109	0,99	96	0,99	97	0,84	81	0,72	103	0,94	87	0,79	99	0,92	122	0,96	84	0,85
Damier Et De Waleffes	111	0,96	98	0,93	95	0,96	106	0,94	105	0,57			110	0,76	100	0,51	97	0,66	107	0,69	99	0,70
Dartagnan Du Castillon	106	1,00	104	0,99	100	1,00	112	0,99	85	0,86	111	0,91	83	0,93	115	0,93	88	0,90	99	0,91	79	0,86
Davidson De L'ecluse	105	0,96	115	0,95	106	0,96	129	0,95	100	0,50			101	0,71			96	0,61			103	0,74
Debordant Et De Biourge	104	1,00	93	0,99	91	1,00	92	0,99	87	0,85	101	0,91	99	0,96	115	0,94	110	0,97	108	0,95	104	0,97
Decent De Bomal	97	0,88	102	0,82	96	0,88	101	0,85					93	0,58			99	0,57			105	0,67
Decisif De Tohogne	112	0,65	100	0,58	94	0,66	109	0,62														
Defiant 6973 Du Molinia	106	0,69	97	0,61	101	0,70	100	0,65														

Cotations linéaires des descendantes (entre 15 et 56 mois) :

COTATIONS LINEAIRES - LINEAIRE BEOORDELING - LINEAR SCORE		Note finale Eindbeoordeling Final score	App. générale Alg. voorkomen General appear.	Type viande Vleestype Meaty type	Musculature NP Bespiering Muscling	Peau Huid Skin	Jarrets Spronggewricht Hocks	Postérieurs Achterpoten Hind legs	Antérieurs Voorpoten Forelegs	Ligne dessus Top lijn Top line	Epaule squelette Shoulder bone Shoulder bone	Ossature Beenwerk Bone structure	Fesses arrière Dijen achteraan. Thighs rear view	Fesses profil Dijen zijaanzicht Thighs side view	Queue Staart implant. Tail set	Longueur bassin Bekkenlengte Pelvis length	Croupe Kruis Rump	Côte Rib Rib	Dessus Rug Top muscling	Epaule : musc. Schouder : resp. Schouder : musc	Largeur bassin Bekkenbreedte Pelvis width	Largeur poitrine Borstbreedte Chest width	Longueur Lengte Length	Taille Gestalte Height	Nom - Naam - Name
	Accord De Wihogne	103	98	96	98	99	99	91	94	100	101	96	100	100	92	102	102	98	97	98	98	97	102	106	
	Adajio De Bray	110	109	101	105	106	101	98	96	96	104	100	111	114	97	102	102	102	101	101	107	98	99	96	
	Agile Pierrecome	105	104	101	105	110	98	94	106	98	93	102	105	109	99	97	97	98	102	105	103	105	96	103	
	Alto 8389 De Reux	88	90	88	91	92	84	99	98	98	101	98	95	95	106	88	90	84	89	90	90	88	89	95	
	Amiral De Ste Helene	103	102	100	101	105	98	96	106	103	100	98	104	102	102	102	96	100	100	101	101	99	95	103	
	Arafat Et Van De Bremberg	98	99	99	98	95	91	102	100	103	97	97	103	105	104	104	100	104	101	105	99	106	91	98	
	Barbier	90	84	86	84	96	100	104	98	101	106	99	86	90	92	112	101	90	83	84	85	83	115	110	
	Baroque De L'ecluse	108	108	105	109	96	106	102	100	102	94	106	107	106	106	90	98	113	114	108	108	107	97	99	
	Bivouac Et 0718 Du Vivier	108	111	116	110	99	92	103	96	96	97	100	111	107	121	89	89	108	104	104	111	107	90	113	
	Blackstar Bringlee Et	100	94	87	93	98	98	100	97	88	103	96	98	98	86	105	104	86	87	90	98	88	107	113	
	Blak 5075 Baty Eprave	107	105	108	104	104	91	103	104	106	103	96	104	102	109	108	108	106	103	104	105	103	107	101	
	Blaky Baty Eprave	99	101	102	101	102	99	100	102	105	101	99	101	100	101	99	99	100	102	102	101	103	99	98	
	Brilliant Du Pont D'herbais	102	101	103	101	105	100	99	101	106	100	97	99	100	105	99	99	103	103	99	96	101	101	101	
	Cabestan 7791 Grigeoule	93	96	99	97	99	97	94	93	109	96	101	97	98	103	88	91	106	103	99	94	101	88	93	
	Canadian Club Kerkenhoefede	108	113	109	112	100	92	91	106	97	104	97	116	113	121	83	102	110	108	114	115	107	92	92	
	Climatiseur D'hanotelet	103	103	106	102	102	92	103	102	99	101	100	99	100	110	100	99	101	106	101	99	103	99	99	
	Colibri 2701 Du Coin	100	107	108	111	122	81	92	105	88	92	106	106	107	105	77	95	118	119	115	103	110	77	82	
	Colos Van Daisel	104	107	105	109	111	94	87	103	95	86	96	111	109	108	84	102	107	110	111	109	102	92	92	
	Dafydd D Ochain	98	93	91	92	109	101	102	99	98	104	110	91	100	100	112	95	90	90	88	87	85	111	109	
	Damier Et De Waleffes	111	110	110	111	104	104	98	105	98	99	94	111	105	105	93	114	110	116	113	114	114	98	101	
	Dartagnan Du Castillon	100	103	99	105	97	87	91	99	93	104	98	105	103	101	86	95	100	108	105	103	101	89	91	
	Davidson De L'ecluse	108	104	108	104	114	99	94	106	96	98	105	103	103	108	108	108	111	106	105	105	101	109	106	
	Debordant Et De Biourge	105	108	107	108	119	97	98	101	108	109	104	106	107	102	98	103	101	110	107	103	102	94	91	
	Decent De Bomal	101	101	100	103	99	110	107	101	108	96	99	100	99	101	101	101	97	104	104	103	102	97	99	
	Decisif De Tohogne	103	105	103	104	104	96	102	97	95	104	99	101	103	101	98	98	97	103	104	101	107	99	99	
	Defiant 6973 Du Molinia	103	102	103	101	96	93	104	101	100	97	94	101	98	104	99	99	102	103	101	104	102	98	98	
	Depute 9265 De Boiron	106	107	109	106	96	97	101	96	101	103	111	103	106	107	97	97	107	108	108	101	101	94	97	
	Destin 9914 Du Coin	94	100	101	100	100	96	107	96	98	101	96	99	99	110	97	97	99	99	99	97	93	98	88	
	Diademe Et. De Fooz	100	105	105	106	103	103	95	100	93	95	96	105	103	100	100	100	98	103	106	106	108	98	90	
	Diamant De Wihogne	100	101	100	103	97	114	88	103	94	100	96	104	104	107	89	89	102	100	106	104	104	91	98	

Index poids carcasse :

Index Poids Carcasse chez les taurillons

Ce tableau contient les index Poids de Carcasse pour les taureaux d'IA BBB. Ils ont été réalisés sur une population de taurillons âgés de 13 à 27 mois (moyenne = 20 mois). Pour être publiés, les index doivent avoir une précision d'au moins 0.80 et avoir au minimum 10 descendants, répartis dans au moins 2 troupeaux.

Les index sont présentés par ordre décroissant. Plus l'index est élevé, plus le poids carcasse est élevé. La moyenne du poids de carcasse (index 100) est de 458 kg et 10 points d'index correspondent à 19 kg.

Karkas gewicht indexen voor de meststieren

Dit tabel stelt de indexen « karkas gewicht » van de KI stieren voor. Ze worden berekend op basis van de stieren tussen 13 en 27 maanden geslacht (gemiddelde = 20 maanden). Om publiceerbaar te zijn, moeten de indexen een betrouwbaarheid van ten minste 0.80 hebben. Bovendien moeten de betreffende stieren minstens 10 nakomelingen geboren in minimum 2 verschillende bedrijven hebben.

De indexen worden in dalende volgorde voorgesteld. Hoe hoger de waarde van de index, hoe zwaarder de karkassen zouden zijn. Een index van 100 vertegenwoordigt het gemiddelde van de populatie (458 kg). 10 punten van indexen stemmen overeen met een verschil van 19 kg.

Carcass Weight indexes for the bull-calves

This table contains the indexes of the carcass weight for the offspring of AI Belgian Blue sires. It is calculated on bull-calves aged from 13 to 27 months (mean age = 20 months). To be published, the indexes must have a precision (r^2) of at least 0.80 and have at least 10 calves, in minimum 2 different herds.

The indexes are presented in decreasing order. The higher the index, the heavier the carcass. The average (index = 100) is 458 kg, and 10 points of the index correspond to 19 kg.

Nom - Naam - Name	Index	R^2
Elan De St Fontaine	116	0,95
Valli Van Terbeck	114	0,96
Dicton Et 8201 Du Coin	113	0,86
Frisko Des Amours Utopiques	111	0,9
Fakir De L'hez Kernai	110	0,94
Joaillier Et Des Pahys	110	0,84
Imprudent 2810 De Somme	109	0,95
Germinal De Fooz	109	0,95
Ebloui De Wihogne	109	0,94
Jalon De Martinpre	109	0,81
Giga Du Bois Remont	109	0,85
Blak 5075 Baty Eprave	108	0,96
Empire D Ochain	108	0,97
Ipon Des Peupliers	108	0,88
Ecrin De St Fontaine	107	0,85
Tison De Wihogne	107	0,88
Genievre De St Fontaine	107	0,95
Fascinant Grigeoule	107	0,93
Illico De Renval	107	0,86
Colos Van Daisel	107	0,86
Bijoutier E.T. De Roupage	106	0,84

Nom - Naam - Name	Index	R^2
Rocky Du Pont De Messe	106	0,85
Nord De Jauvelan	106	0,91
Orme De Somme	106	0,91
Fleurion Et De Maffe	106	0,88
Arlequin Des Peupliers	105	0,85
Danseur D'embise	105	0,92
Bivouac Et 0718 Du Vivier	105	0,88
Dafydd D Ochain	105	0,97
Dartagnan Du Castillon	105	0,98
Jupp De Galhausen	105	0,89
Gratin De Nochet	105	0,83
Fringant Et De Biourge	104	0,95
Diamant De Wihogne	104	0,93
Heros Du Peroy	104	0,93
Important De La Hasse	104	0,85
Blackstar Bringlee Et	104	0,95
Obus De Somme	103	0,92
Raciste 2586 Pachis Mott	102	0,92
Albinos 6546 Grigeoule	102	0,93
Gardien De La Coulbrie	102	0,85
Gamin Des 3 Frontieres	102	0,98

Annexe 5 : Bilan des cas et des parentés collectés et génotypés par la cellule d'Hérédosurveillance du 01.10.2007 au 31.03.2013.

CARACTERES	CAS		PARENTES
	COLLECTES	GENOTYPES 50K	
DMC 1	10	-	3
DMC 2	7	-	4
SQT	195	8	106
Nanisme	233	147	714
Hamartome	68	37	55
Gestation prolongée	72	32	127
Arthrogrypose	26	15	11
White Heifer Disease	206	206	179
Parésies spastiques	63	0	16
Blanc dos	67	15	0
Divers	170	0	96
Taureaux IA	27	27	-
Anourie	10	6	6
Atrésies digestives	9	4	4
Hydrocéphalie	10	4	4
Perosomus elumbis/ Spina bifida	17	7	3
Ataxie progressive	6	5	0
Cardiomyopathie dilatée	4	4	2
Hypotrichose	3	3	2
TOTAL	1167	520	1337

Annexe 6 :**Serial translocation by means of circular intermediates underlies
colour sidedness in cattle.**

Durkin K., Coppieters W., Drögemüller C., Ahariz N., Cambisano N., Druet T.,
Fasquelle C., Haile A., Horin P., Huang L., Kamatani Y., Karim L., Lathrop M., Moser S.,
Oldenbroek K., Rieder S., Sartelet A., Sölkner J., Stålhammar H., Zelenika D., Zhang Z.,
Leeb T., Georges M., Charlier C.

Nature, 2012, 482: 81-84.

LETTER

doi:10.1038/nature10757

Serial translocation by means of circular intermediates underlies colour sidedness in cattle

Keith Durkin¹, Wouter Coppieters¹, Cord Drögemüller², Naima Ahariz¹, Nadine Cambisano¹, Tom Druet¹, Corinne Fasquelle¹, Aynalem Haile³, Petr Horin⁴, Lusheng Huang⁵, Yohichiro Kamatani⁶, Latifa Karim¹, Mark Lathrop⁶, Simon Moser⁷, Kor Oldenbroek⁸, Stefan Rieder⁷, Arnaud Sartelet¹, Johann Sölkner⁹, Hans Stålhammar¹⁰, Diana Zelenika⁶, Zhiyan Zhang¹, Tosso Leeb², Michel Georges¹ & Carole Charlier¹

Colour sidedness is a dominantly inherited phenotype of cattle characterized by the polarization of pigmented sectors on the flanks, snout and ear tips¹. It is also referred to as 'lineback' or 'witrik' (which means white back), as colour-sided animals typically display a white band along their spine. Colour sidedness is documented at least since the Middle Ages and is presently segregating in several cattle breeds around the globe, including in Belgian blue and brown Swiss^{1,2}. Here we report that colour sidedness is determined by a first allele on chromosome 29 (*Cs₂₉*), which results from the translocation of a 492-kilobase chromosome 6 segment encompassing *KIT* to chromosome 29, and a second allele on chromosome 6 (*Cs₆*), derived from the first by repatriation of fused 575-kilobase chromosome 6 and 29 sequences to the *KIT* locus. We provide evidence that both translocation events involved circular intermediates. This is the first example, to our knowledge, of a phenotype determined by homologous yet non-syntenic alleles that result from a novel copy-number-variant-generating mechanism.

To gain insights into the molecular basis of colour sidedness (Fig. 1), we genotyped 21 colour-sided and 30 control Belgian blue animals with a custom-made 50K single nucleotide polymorphism (SNP) array³. As a result of segregation at the *roan* locus, Belgian blue animals are either black spotted ($r^{Bl}r^{Bl}$), blue spotted ($r^{Wh}r^{Wh}$) or white ($r^{Wh}r^{Wh}$)⁴. As white is epistatic to colour sidedness, we selected non-white control animals. We assumed autosomal dominant inheritance (*Cs* allele) and genetic homogeneity in Belgian blue, and thus scanned the genome of colour-sided animals for a shared haplotype (present in at least one copy) using the ASSDOM software (see Methods). This analysis yielded a single genome-wide significant signal ($P < 0.03$), mapping the *Cs* locus to bovine chromosome 29 (BTA29) (Fig. 2a and Supplementary Fig. 1). A sire transmitting the colour-sided phenotype to all its pigmented offspring was homozygous for the corresponding haplotype as expected. The shared haplotype spanned 1.9 Mb, and encompassed *LUZP2*—not known to be involved in pigmentation—as the only gene. Sequencing the *LUZP2* open reading frame (ORF) from colour-sided and control animals did not reveal any protein-sequence-altering variant (data not shown).

In an independent effort, we scanned the genome for copy number variants (CNVs) using a database of 50K SNP genotypes from >4,500 animals of different breeds and the PennCNV software⁵. Intriguingly, this analysis revealed a private duplication encompassing 26 SNPs on chromosome 6 (BTA6), shared by the 21 colour-sided Belgian blue animals (Fig. 2b). We confirmed and refined the boundaries of this ~480 kb duplication by comparative genome hybridization (CGH) of genomic DNA of a *Cs/Cs* and *+/+* Belgian blue animal

on a genome-wide Nimblegen tiling array (Fig. 2c). The corresponding CNV encompassed the *KIT* gene, known to be essential for melanocyte migration and survival⁶, and to be associated with coat colour variation^{7,8}.

To reconcile these apparently discrepant results, we performed fluorescence *in situ* hybridization (FISH) with bacterial artificial chromosome (BAC) clones mapping respectively to the BTA29 association interval (labelled with a red fluorophore), and to the BTA6 CNV interval (labelled with a green fluorophore) on lymphocytes of a *Cs/+* and a *+/+* animal. Overlapping red and green signals on one of the BTA29 homologues of the colour-sided animal demonstrated that the



Figure 1 | Colour-sided Belgian blue and brown Swiss animals. **a**, All represented Belgian blue animals are heterozygous *Cs/+*. In addition to the effect of the *Cs* locus, which causes polarization of pigmented sectors to the flanks, ears and snout, Belgian blue animals exhibit considerable polygenic variation in the degree of white spotting. The colour-sided phenotype of the two nearly completely white animals (extensive degree of white spotting) is apparent from the pigmentation of ear tips and snout. **b**, **c**, In brown Swiss, which are generally devoid of white spotting, homozygous *Cs/Cs* (**b**) and heterozygous *Cs/+* (**c**) animals differ by the extent of pigmentation, explaining why colour sidedness is also referred to as semi-dominant.

¹Unit of Animal Genomics, GIGA-R & Faculty of Veterinary Medicine, University of Liège (B34), 1 Avenue de l'Hôpital, 4000-Liège (Sart Tilman), Belgium. ²Institute of Genetics, University of Bern, Bremgartenstrasse 109a, Postfach 8466, 3001 Bern, Switzerland. ³International Center for Agricultural Research in the Dry Areas (ICARDA), PO Box 5466, Aleppo, Syria. ⁴Institute of Genetics, University of Veterinary and Pharmaceutical Sciences Brno, 612 42 Brno, Czech Republic. ⁵Key Laboratory for Animal Biotechnology of Jiangxi Province and the Ministry of Agriculture of China, NanChang, Jiangxi Province, 330045, China. ⁶Fondation Jean Dausset – Centre d'Etude du Polymorphisme Humain, 27 Rue Juliette Dodu, Paris 75012, France and CEA-IC Centre National de Génotypage, 2 Rue Gaston Crémieux, CP 5721, 91057 Evry Cedex, France. ⁷Swiss College of Agriculture, Bern University of Applied Sciences, 3052 Zollikofen, Switzerland. ⁸Centre for Genetic Resources the Netherlands (CGN), 6701 BH Wageningen, The Netherlands. ⁹Division of Livestock Sciences, University of Natural Resources and Applied Life Sciences, A-1180 Vienna, Austria. ¹⁰Viking Genetics, Örsnro, Box 64, S-532 21 Skara, Sweden.

RESEARCH LETTER

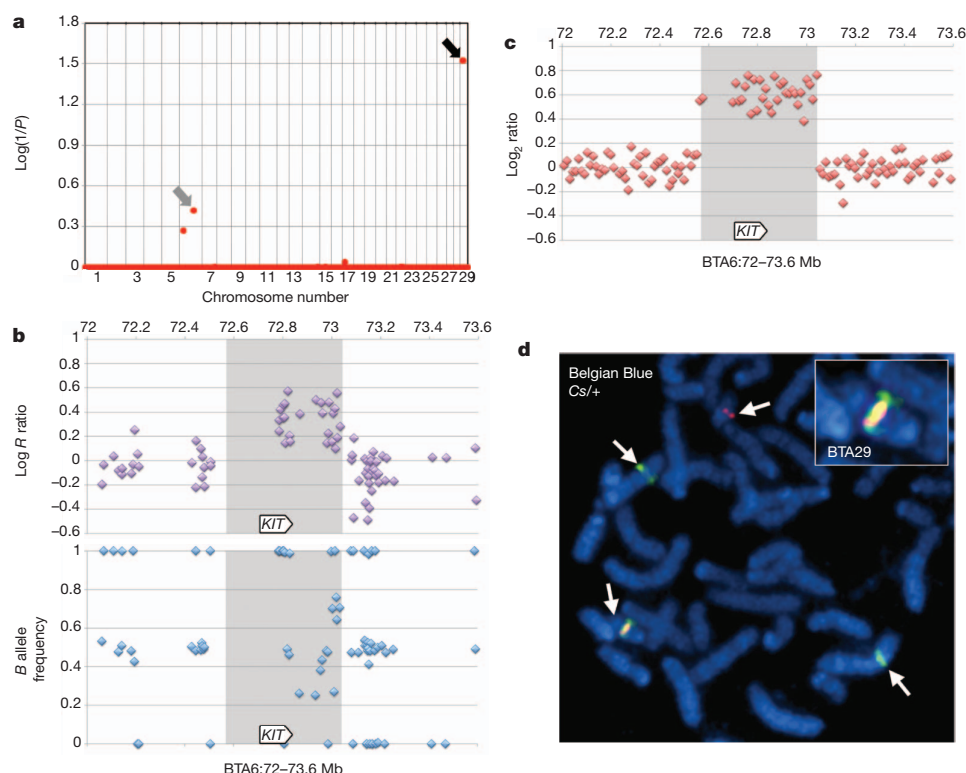


Figure 2 | Identification and mapping of the colour-sided locus in the Belgian blue. **a**, Genome-wide association mapping using ASSDOM (see Methods), revealing a genome-wide significant association ($P < 0.03$) on BTA29 (black arrow) and suggestive association on BTA6 (grey arrow). Vertical lines separate chromosomes. **b**, Detection of a duplication on the BTA6 segment shared by 21 colour-sided Belgian blue animals using 50K SNP array

CNV signal resulted from the translocation of the ~480 kb BTA6 segment to BTA29 (Fig. 2d). Retrospective examination of the SNP genotypes indicated that on average 3.17 of the 26 SNPs mapping to the ~480 kb duplication could not be called in colour-sided animals, yet the remaining genotypes were sufficient to yield a suggestive association signal (genome-wide $P = 0.37$), reflecting the sharing of the extra *KIT* haplotype by colour-sided animals (Fig. 2a).

To define the translocation breakpoints, we generated mate-pair libraries from self-ligated ~5 kb DNA fragments of a homozygous *Cs/Cs* Belgian blue animal, and generated ~10 Gb of sequence on an Illumina GAIIx instrument. We expected two clusters of aberrant mate pairs spanning the left and the right breakpoint resulting from the insertion of an intact BTA6 fragment in a BTA29 segment. We refer to the intact BTA6 fragment as A–B–C–D–E, the BTA29 segment in which it inserts as α – β , and the resulting left and right breakpoints as α –A and E– β , respectively. However, we observed three aberrant mate-pair clusters corresponding respectively to α –D, E–A and C– β fusions (Supplementary Fig. 2). The corresponding topology is most parsimoniously explained by assuming that: (1) the translocated BTA6 fragment circularized (generating fusion E–A); (2) reopened in the C–D interval; and (3) integrated in the α – β BTA29 interval (generating fusions α –D and C– β ; Fig. 3).

Using the genomic coordinates of the clustered mate pairs, we designed primer sets to amplify the three corresponding fusion points. Productive amplification was only achieved using genomic DNA from colour-sided Belgian blue animals, as expected (Supplementary Fig. 3A). We sequenced the corresponding PCR products to define the breakpoints at single-nucleotide resolution. The E–A fusion was characterized by 2 bp micro-homology typical of non-homologous end joining (NHEJ)

data³ and PennCNV³. **c**, Confirmation and boundary definition using CGH on a Nimblegen bovine tiling array. Shaded areas in **b** and **c** mark the boundaries of the CNV as defined by CGH. **d**, Demonstration by FISH of the translocation of a *KIT*-encompassing BTA6 segment onto BTA29 in a heterozygous *Cs*/+ Belgian blue animal. Magnification, $\times 100$.

(Supplementary Fig. 4A), whereas the α –D and C– β fusions exhibited micro-duplications and micro-deletions reminiscent of replication-dependent microhomology-mediated break-induced replication (MMBIR) (Supplementary Fig. 4B)⁹. All breakpoints mapped to interspersed non-homologous repeat elements (data not shown).

The dominance of the *Cs* allele is expected to reflect a gain of function resulting from dysregulated expression of the translocated *KIT* gene. To verify the transcriptional competence of the translocated

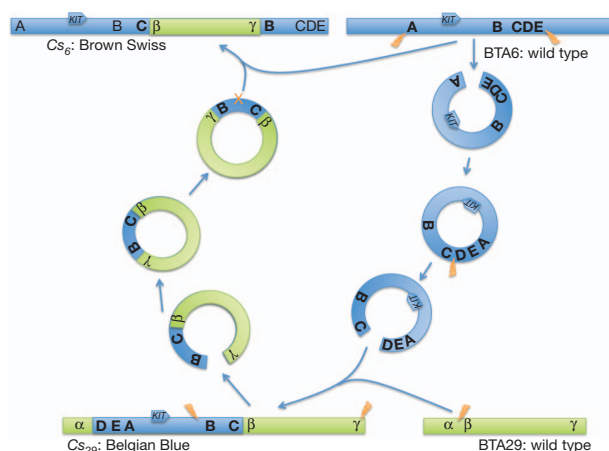


Figure 3 | Model for the generation of the colour-sided *Cs*₂₉ and *Cs*₆ alleles by serial translocation via circular shuttling intermediates.

KIT copy, we performed polymerase chain reaction with reverse transcription (RT-PCR) experiments using amplicons spanning two SNPs: the ss469414206 G-to-A transition in intron 1, and the ss469414207 C-to-T transition in intron 7. We extracted total RNA from pigmented and unpigmented skin sectors of a *Cs/+* colour-sided Belgian blue animal with the GG genotype on BTA6 and the A genotype on BTA29 ((GG)₆/A₂₉) for ss469414206 and the (CC)₆/T₂₉ genotype for ss469414207. The resulting pre-mRNA-dependent RT-PCR products were directly sequenced and the ratio of T/C and A/G species was estimated using Peakpicker¹⁰. T/A transcripts accounted for ~33% of the *KIT* output in both pigmented and unpigmented sectors, demonstrating the transcriptional potential of the translocated gene copy. Long-range RT-PCR analysis of the near complete *KIT* mRNA with primers located respectively in exon 1 and the 3' untranslated region (UTR) followed by amplicon sequencing did not show evidence for alternate transcripts in colour-sided animals (Supplementary Fig. 5).

Linkage analysis performed in a brown Swiss pedigree segregating for colour sidedness with microsatellite markers targeting candidate genes (*KIT*, *KITL*, *MITF*, *EDNRB*, *ADAMTS20*), yielded a log of odds (lod) score of 6.9 maximizing in the immediate vicinity of the *KIT* locus (Supplementary Table 1). We genotyped four colour-sided (three *Cs/Cs* and one *Cs/+*), and five control brown Swiss animals using a 50K SNP array³. Colour-sided *Cs/Cs* brown Swiss animals indeed shared a 2.0 Mb autozygous haplotype encompassing the *KIT* locus. At first glance, these findings suggested a distinct determinism of colour sidedness in Belgian blue and brown Swiss.

We analysed the corresponding SNP genotypes with PennCNV and performed CGH using genomic DNA from brown Swiss *Cs/Cs* versus *+/+* animals. This revealed the duplication of a ~120-kb BTA6 segment nested in the Belgian blue duplication, but excluding the *KIT*

gene. Intriguingly, it also revealed the duplication of a ~418-kb BTA29 segment immediately flanking the Belgian blue insertion site and encompassing the last four of twelve *LUZP2* exons (β - γ ; Fig. 4a, b). Moreover, fusion point C- β (but not α -C and D-A), specifying the Belgian blue *Cs* allele, could be amplified by PCR from genomic DNA of all examined brown Swiss colour-sided but not control animals (Supplementary Fig. 3). These findings established a clear link between the Belgian blue and brown Swiss *Cs* alleles.

We performed FISH analysis on lymphocytes of a *Cs/+* brown Swiss animal using BTA6 and BTA29 BAC clones. Remarkably, we observed overlapping red and green signals on one of the BTA6 homologues, hence revealing the translocation of a BTA29 fragment on BTA6 (Fig. 4c). We generated mate-pair libraries from self-ligated ~5-kb and ~2-kb DNA fragments of a brown Swiss *Cs/Cs* animal and generated ~15 Gb of sequence on an Illumina GAIIx instrument (Supplementary Fig. 2). Analysis of the resulting sequence traces revealed two aberrant mate-pair clusters. The first corresponds to the Belgian blue C- β fusion point, previously detected by PCR. The second corresponds to a novel B- γ fusion point (Supplementary Fig. 2). The most parsimonious explanation accounting for all the data is that the brown Swiss *Cs* allele derives from the Belgian blue *Cs* allele by (1) excision of the B-C- β - γ fragment from the Belgian blue *Cs* allele on BTA29, (2) circularization, and (3) re-integration in the wild-type *KIT* locus by homologous recombination. This would result in a novel *Cs* allele mapping to the *KIT* locus and characterized by tandem duplicates of the B-C segment flanking the translocated β - γ BTA29 fragment (A-B-C- β - γ -B-C-D-E; Fig. 3).

Our model predicts that the C and B BTA6 segments immediately flanking β - γ carry the same haplotype as the *Cs*-allele of Belgian blue. To test this, we developed long-range PCR assays that would specifically

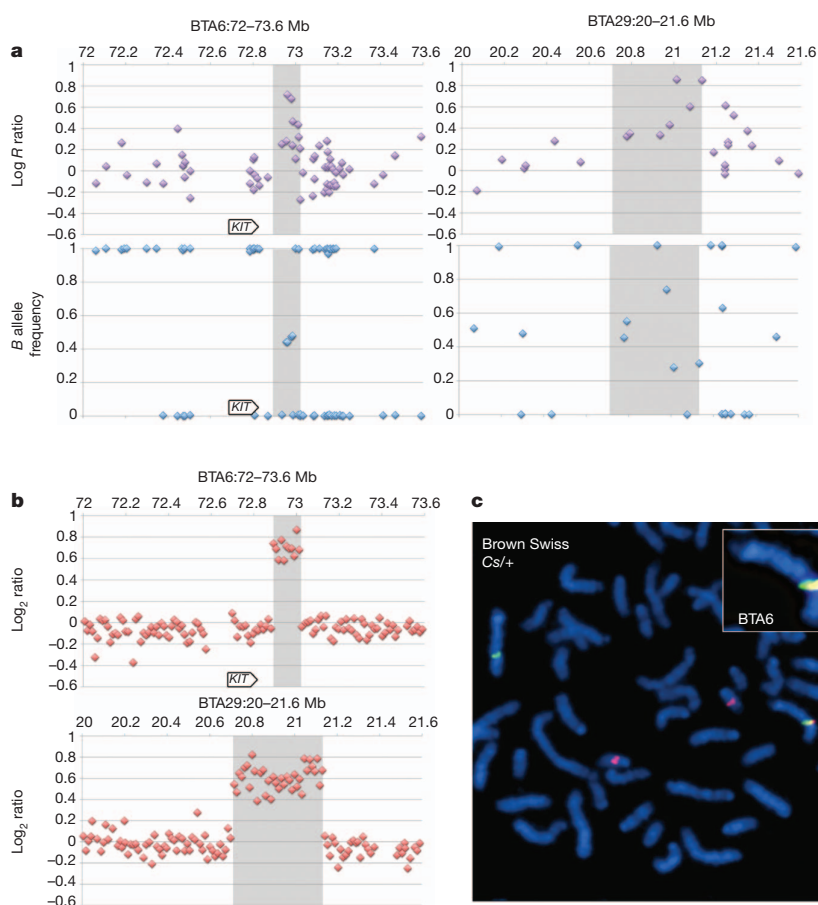


Figure 4 | Identification and mapping of the colour-sided locus in brown Swiss animals.
a, PennCNV evidence for BTA6 and BTA29 duplications in brown Swiss colour-sided animals. The slightly different distribution of BTA6 SNPs when compared to Fig. 2 is due to the use of a different version of the 50K SNP array¹.
b, Confirmation and boundary definition of the *Cs*-associated BTA6 and BTA29 duplication in brown Swiss animals using CGH on Nimblegen bovine tiling arrays. The shaded areas in **a** and **b** mark the boundaries of the CNV, as defined by CGH.
c, Demonstration by FISH of the translocation of a BTA29 segment onto BTA6 in a heterozygous *Cs/+* brown Swiss animal.

RESEARCH LETTER

amplify B, C, B and C segments from genomic DNA of a Cs/Cs brown Swiss animal, and—by sequencing the corresponding amplicons—determined the genotype of B and B and C and C for four SNP positions heterozygous in the high-throughput sequence reads of the brown Swiss Cs/Cs animal. We then determined the genotype of the Belgian blue Cs allele for the corresponding variants and showed that it matched the C and B segments of the brown Swiss Cs allele, in agreement with our model of homologous-recombination-dependent resolution of the circular intermediate (Fig. 3 and Supplementary Fig. 6).

Using the mate-pair genomic coordinates, we designed primer pairs to amplify the γ -B fusion point. As expected, productive amplification was only achieved using genomic DNA from colour-sided brown Swiss animals (Supplementary Fig. 3). We sequenced the corresponding PCR products to define the breakpoints at single-nucleotide resolution. The γ -B fusion presented hallmarks typical of microhomology mediated break induced replication (MMBIR) (Supplementary Fig. 4).

We obtained genomic DNA from colour-sided animals from seven additional cattle breeds and domestic yaks (Supplementary Fig. 7), which we tested by PCR for the presence of the two Belgian-blue-specific fusion points (E-A, α -D), the brown-Swiss-specific fusion point (γ -B), and the Belgian blue/brown Swiss shared fusion point (C- β). Colour-sided Dutch witrik and Ethiopian fogera animals were shown to carry the Belgian blue Cs allele (Cs₂₉), Austrian pustertaler sprinzen, Czech red-spotted cattle and French vosgienne the brown Swiss Cs allele (Cs₆), and Irish moiled, Swedish mountain and domestic yak carried both the Cs₂₉ and Cs₆ alleles (Supplementary Fig. 3). We assume that Cs₂₉ and Cs₆ alleles were introgressed in yak after domestication via well-documented hybridization of *Bos taurus* and *Bos grunniens*. These findings indicate that the Cs₂₉ and Cs₆ alleles account for most if not all colour sidedness in cattle.

Analysis of colour sidedness has revealed a novel CNV-generating translocation mechanism involving circular intermediates. Whether this is a bovine idiosyncrasy or a more common mechanism remains to be determined. That some CNVs reflect translocation events is well established. As an example, 75 probable dispersed duplications in the human genome have been reported¹¹, as well as at least four interchromosomal duplications^{12,13}. We ourselves performed genome-wide association mapping between the SNP and CNV genotype in human and cattle, and observed 21 and 4 putative 'trans-associations' (that is, non-syntenic CNV-defining SNPs and associated SNPs), respectively (Supplementary Material). Some of these dispersed duplications might involve circular intermediates. In support of the occurrence of other instances involving circular intermediates are recent observations of repeated translocation of five clustered ORFs in wine and bioethanol strains of *Saccharomyces cerevisiae*, apparently via resolution of circular intermediates¹⁴. The same mechanism may contribute to somatic mutations in tumours. Indeed, episodes with the *NUP214-ABL1* fusion gene (observed in ~6% of T-cell acute lymphoblastic leukaemia) have been proposed to result from circularization and excision of a chromosome 9 segment bounded by the *NUP214* and *ABL1* genes, and to reintegrate ectopically by the same resolving mechanism proposed for colour sidedness in at least some patients¹⁵. The repatriation of exogenous sequences (including exons) back to the original chromosomal location via circular 'shuttling' intermediates suggests that this mechanism might underlie a specific mode of exon shuffling.

METHODS SUMMARY

A custom 50K SNP array³ was used to genotype 21 colour-sided and 30 control Belgian blue animals. The genome of colour-sided animals was scanned for a shared haplotype using the ASSDOM software. In the brown Swiss, microsatellites adjacent to five candidate genes were genotyped in three half-sibling families and two point linkage calculated using the Merlin software¹⁶. CNVs were identified in the 50K SNP array data using the PennCNV software⁷. Array CGH was carried out on a custom 2.1 M oligonucleotide array (Roche-Nimblegen) with a non-colour-sided Belgian blue used as the reference in each hybridization. Metaphase spreads

were generated from short-term lymphocyte cultures. BACs from the duplicated regions on BTA6 and BTA29 were identified using end sequences from the bovine RPC142 BAC library. BACs were labelled with the appropriate fluorochrome by nick translation (Abbott Molecular), hybridized to the metaphase spreads and examined by fluorescent microscopy. Mate-pair libraries were generated using the Illumina mate-pair library kit v.2 for a Cs/Cs Belgian blue and a Cs/Cs brown Swiss animal. A paired-end library was also generated for a Cs/+ Belgian blue using the Illumina paired-end kit. Sequencing was carried out on an Illumina GAIIx instrument. PCR products spanning regions of interest were purified using QIAquick PCR purification kit (Qiagen), sequenced using Big Dye terminator cycle-sequencing kit v.3.1 (Applied Biosystems) and run on an ABI PRISM 3730 DNA analyser (Applied Biosystems). RNA was extracted from skin using the RNeasy fibrous tissue mini kit (Qiagen) and cDNA was synthesized using SuperScript II first-strand Synthesis SuperMix (Invitrogen). Long-range PCR was carried out using the Expand Long Template PCR System (Roche).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 September; accepted 5 December 2011.

- Olson, T. *The Genetics of Cattle* (CAB International, 1999).
- Porter, V. & Mason, I. L. *Mason's World Dictionary of Livestock Breeds, Types, and Varieties* (CAB International, 2002).
- Charlier, C. *et al.* Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nature Genet.* **40**, 449–454 (2008).
- Charlier, C. *et al.* Microsatellite mapping of the bovine *roan* locus: a major determinant of white heifer disease. *Mamm. Genome* **7**, 138–142 (1996).
- Wang, K. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17**, 1665–1674 (2007).
- Yoshida, H. *et al.* Melanocyte migration and survival controlled by SCF/c-kit expression. *J. Invest. Dermatol. Symp. Proc.* **6**, 1–5 (2001).
- Brooks, S. A. *et al.* A chromosome inversion near the *KIT* gene and the Tobiano spotting pattern in horses. *Cytogenet. Genome Res.* **119**, 225–230 (2007).
- Marklund, S. *et al.* Molecular basis for the dominant white phenotype in the domestic pig. *Genome Res.* **8**, 826–833 (1998).
- Hastings, P. J., Lupski, J. R., Rosenberg, S. M. & Ira, G. Mechanisms of change in gene copy number. *Nature Rev. Genet.* **10**, 551–564 (2009).
- Ge, B. *et al.* Survey of allelic expression using EST mining. *Genome Res.* **15**, 1584–1591 (2005).
- Conrad, D. F. *et al.* Origins and functional impact of copy number variation in the human genome. *Nature* **464**, 704–712 (2010).
- Liu, G. E. *et al.* Analysis of copy number variations among diverse cattle breeds. *Genome Res.* **20**, 693–703 (2010).
- Liu, G. E. *et al.* Analysis of recent segmental duplications in the bovine genome. *BMC Genomics* **10**, 571 (2009).
- Borneman, A. R. *et al.* Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. *PLoS Genet.* **7**, e1001287 (2011).
- Graux, C. *et al.* Heterogeneous patterns of amplification of the *NUP214-ABL1* fusion gene in T-cell acute lymphoblastic. *Leukemia* **23**, 125–133 (2009).
- Abecasis, G. *et al.* Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genet.* **30**, 97–101 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was funded by the Walloon Direction General Operationnelle Agriculture, Ressources Naturelles et Environnement (grants D31-1206 and D31-1214) and the Swiss National Science Foundation (grant 31003A_133034 to T.L.). We are grateful for the support of the GIGA-R Genomics platform, to P. Melvin, to G. Mekuriaw and staff of the Andassa livestock research center, Viking Genetics, Swissgenetics and the Schweizer Braunviehzuchtverband for their assistance in sample collection, to J. Vermeesch and B. Nowakowska for their assistance with array CGH.

Author Contributions K.D. performed the PennCNV, CGH, FISH, next-generation sequencing, breakpoint, DNA sequence variant and expression analyses. W.C. performed the bioinformatic analyses of the high-throughput sequencing reads. A.H., P.H., L.H., K.O., A.S., J.S. and H.S. provided samples from colour-sided animals. C.D., S.M. and S.R. performed linkage analyses in brown Swiss animals. T.L. supervised the work in Switzerland. C.C., M.G., T.D. and Z.Z. performed the association analyses. N.A. and N.C. performed SNP genotyping. C.F. and L.K. performed amplicon sequencing. Y.K., M.L. and D.Z. provided human CNV genotype data. K.D., M.G. and C.C. analysed data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.G. (michel.georges@ulg.ac.be).

METHODS

Association and linkage mapping of the colour-sided locus. SNP genotyping was conducted using custom-made 50K Infinium SNP arrays³ used according to the instructions of the manufacturer. Microsatellite genotyping was conducted as previously described¹⁷. Association analysis was conducted using ASSDOM. ASSDOM searches for chromosome segments devoid of SNPs for which cases have alternate homozygous genotype (say 11 versus 22), excluding the sharing of an identical-by-descent (single-copy) haplotype. Intervals bounded by such excluding SNPs receive a score corresponding to $\sum_{i=1}^k \log(1-p_i)^n$, where p_i is the frequency of the allele missing among n cases, estimated in m controls. The genome-wide statistical significance of the 'non-exclusion' signal is determined by phenotype permutation of the disease status between the n cases and m controls. Two-point linkage analyses were conducted with Merlin¹⁶.

Prediction of CNVs from SNP genotype data. The log R ratio signal intensity and B allele frequency from a custom 50K SNP array³ were obtained using Illumina BeadStudio software. PennCNV, a hidden Markov model based approach that takes into account signal intensity, allelic intensity ratio, distance between markers and allele frequency⁵ was used to call CNVs. Regions of the genome that showed evidence of copy number change were inspected in greater detail. Plots of log R ratio and B allele frequency were examined in BeadStudio (Illumina) and the region checked in the University of California, Santa Cruz (UCSC) genome browser (<http://genome.ucsc.edu/>).

Detection of CNVs by CGH. Array CGH was carried out on a custom 2.1 M oligonucleotide array (Roche-Nimblegen) based on the UMD 3.0 bovine assembly. The array contained 2,152,422 probes (50–75 mers) with a median spacing of 1,160 bp. The reference animal used in hybridizations was a non-colour-sided Belgian blue. Genomic DNA labelling (Cy3 for sample and Cy5 reference), hybridization and washing were performed according to the manufacturer's instructions and have been described elsewhere¹⁸. Slides were scanned using a GenePix 4000B 5 µm microarray scanner (Axon Instruments). Images were processed using NimbleScan software (Roche-Nimblegen). Spatial correction was applied and data normalized¹⁹, segmentation was performed using the DNACopy algorithm²⁰. The log₂ ratios for each oligonucleotide were also examined visually for evidence of a change in copy number in regions of interest.

FISH. Peripheral blood was obtained from colour-sided and wild-type Belgian blue and brown Swiss animals. Pokeweed-stimulated lymphocyte cultures were established and chromosome spreads prepared for colour-sided and wild-type animals in each breed following standard cytogenetic procedures. End sequences from the RPC142 bovine BAC library and the duplicated regions on BTA6 and BTA29 in the colour-sided Belgian blue and brown Swiss were downloaded from the National Center for Biotechnology Information (NCBI; <http://www.ncbi.nlm.nih.gov/>). BLAST was used to identify end sequences located in the duplicated regions. The BACs 160M9 and 156I13 overlap and cover the region chr6:72,566,605–72,817,995 (bosTau4), while the BACs 37P11 and 116G8 also overlap and cover the region chr29:20,772,406–21,035,251 (bosTau4). BAC clones were initially cultured at 37 °C in 1 ml of 2YT media containing 30 µl ml⁻¹ chloramphenicol. The cultures were plated on lysogeny broth (LB) agar plates with 30 µl ml⁻¹ chloramphenicol to obtain single colonies, the identity of the BACs was confirmed by PCR using primers designed within the area encompassed by the respective BAC (Supplementary Table 2). The positive clones were then used to inoculate 100 ml of 2YT media with 30 µl ml⁻¹ chloramphenicol. Following 24 h at 37 °C with constant agitation, DNA was extracted using the Qiagen midiprep kit, following the manufacturer's instructions. The DNA was labelled with the nick translation kit from Abbott Molecular, using the manufacturer's protocol. DNA from BACs 160M9 and 156I13 was mixed and labelled with spectrum green (Abbott Molecular) while 37P11 and 116G8 were mixed and labelled with spectrum orange (Abbott Molecular). Labelled DNA (100 ng) was combined with 1 µg of bovine Cot-1 DNA and 2 µg of bovine genomic DNA, precipitated then resuspended in 3 µl of purified water and 7 µl of hybridization buffer (Abbott Molecular). The separate probes were denatured at 73 °C for 5 min and then combined on a slide containing metaphase spreads. These slides had been denatured in 70% formamide in 2× SSC at 75 °C for 5 min followed by dehydration in ethanol. A coverslip was secured with rubber cement and the slide incubated overnight at 37 °C in a humidified chamber. Slides were then washed in 0.4× SSC/0.3% Tween-20 at 73 °C for 1–3 s followed by washing in 2× SSC/0.1% Tween-20 at room temperature (~18 °C) for 1–3 s and air dried. The slides were counterstained with DAPI II (Abbott Molecular) and visualized by fluorescent microscopy.

Next-generation sequencing of mate-pair and paired-end libraries. Mate-pair libraries with different insert sizes were generated using the Illumina mate-pair library kit, v.2. The manufacturer's instructions were followed except for the step involving fragmentation of the circularized DNA, for which a bioruptor sonicator

UCD-200 (Diagenode) was used. The 300 µl sample was placed in a 1.5 ml Eppendorf tube and sonicated for 8 min with the instrument set to high and a cycle of 30 s on and 30 s off. A 5-kb insert library was generated for a *Cs/Cs* Belgian blue animal and a 2-kb and 5-kb for a *Cs/Cs* brown Swiss animal. A paired-end library with a 400 bp insert size was also generated for a *Cs/+* Belgian blue animal using the Illumina paired-end kit, following the instructions of the manufacturer. The resulting libraries were quantified using Pico-Green (Quant-it, Invitrogen) and the Agilent 2100 Bioanalyzer High Sensitivity DNA kit (Agilent Technologies). Sequencing was carried out on an Illumina GAIIx instrument. Mapping of the 36 bp from each end of the mate-pair libraries and the 110 bp from the ends of the paired-end library was performed using the BWA tool²¹. Breakpoints were identified by visually inspecting the mate pairs using the integrative genomics viewer²² in the ~2-Mb region surrounding the *Cs*-specific duplications in Belgian blue and brown Swiss animals and looking for discordant mate pairs.

PCR amplification of translocation breakpoints. PCR primers were designed to span each of the breakpoints identified by mate-pair sequencing. The primers were tested on genomic DNA from colour-sided and wild-type animals. PCR products were visualized on a 2% agarose gel. Primers with amplification confined to colour-sided animals were purified using the QIAquick PCR purification kit (Qiagen), where multiple bands were observed the relevant band was excised and purified using the QIAquick gel extraction kit (Qiagen). The fragments were then sequenced using Big Dye terminator cycle-sequencing kit v.3.1 (Applied Biosystems) with the purified reaction run on a ABI PRISM 3730 DNA analyser (Applied Biosystems). Primers used are listed in Supplementary Table 2.

Analysis of *Cs*₂₉-derived *KIT* transcripts. To ensure that no mutations were present in the coding sequence of the *Cs*₂₉-specific *KIT* gene, primers were designed to amplify all the exons and the 3' UTR from genomic DNA (Supplementary Table 2). The resulting PCR products were sequenced as outlined above, which did not reveal any protein altering DNA sequence variant. To examine expression of *KIT* from the *Cs*₂₉ allele, a small biopsy of skin was removed from the back (white skin) and side (pigmented skin) of a colour-sided Belgian blue animal (the relevant ethical procedures were adhered to). The samples were immediately frozen in liquid nitrogen and stored at -80 °C until RNA extraction. The tissue was homogenized using a Tissue Lyser (Qiagen), total RNA was extracted using the RNeasy fibrous tissue mini kit (Qiagen), following the manufacturer's instructions. First-strand cDNA was synthesized using SuperScript II first-strand Synthesis SuperMix (Invitrogen) with a mixture of random hexamers and oligo(dT) primers. Genomic DNA was extracted from whole blood from the same animal using standard phenol-chloroform extraction. The intronic regions of the *KIT* gene were searched for SNPs using the mate-pair sequences produced from the *Cs/Cs* Belgian blue animal. Primers were designed around suitable SNPs and sequenced in the tissue donor. To establish the genotype of the *Cs*₂₉ allele, multiple colour-sided animals were sequenced. For SNP ss469414207 (C/T) 40 colour-sided animals were sequenced. For all *Cs/Cs* animals, the C and T alleles had equal peak height, while in the remaining *Cs/+* animals the T allele produced the smaller peak. Additionally seven non colour-sided Belgian blue animals were also sequenced and none possessed the T allele. For SNP ss469414206 (G/A), two *Cs/Cs* and three *Cs/+* animals were sequenced and showed a pattern of peak heights consistent with the *Cs*₂₉ allele having the A genotype. To determine if the *Cs*₂₉ allele was expressed, the cDNA from the skin biopsies was amplified using primers spanning both SNPs and sequenced as outlined above. Moreover, and to ensure the integrity of *Cs*₂₉-derived *KIT* transcripts, primers located in the first exon and the 3' UTR (Supplementary Table 2) were used with the Expand Long Template PCR System (Roche) to amplify the near full-length *KIT* cDNA. The product was run on 1% gel and examined for evidence of alternative splicing.

Genotyping the duplicated B and C segments of the *Cs*₆ allele. Long-range PCR was carried out using the Expand Long Template PCR System (Roche). The genomic DNA used as template was extracted using QIAamp DNA Mini columns (Qiagen) following the manufacturer's instructions to produce high molecular weight DNA. For each reaction the following mix was prepared: 2 µl Buffer 1, 140 µM dNTP, 120 nM upstream and downstream primer, 0.3 µl enzyme mix and 100 ng genomic DNA, final volume was 20 µl. Extension time and the thermal profile recommended by the manufacturer was followed. Following 30 cycles 10 µl of the product was run on a 0.8% agarose gel to check for amplification, the remaining reaction mix was retained. The PCR primers spanning the relevant SNP were then used in a conventional PCR reaction using between 0.1 and 0.5 µl (amount determined by intensity of band) of the long-range product as template. When long-range PCR product was used as template the number of cycles was reduced to 20. Clean-up and sequencing was carried out as outlined above. Used primers are reported in Supplementary Table 2.

17. Drögemüller, C. *et al.* Genetic mapping of the belt pattern in Brown Swiss cattle to BTA3. *Anim. Genet.* **40**, 225–229 (2009).

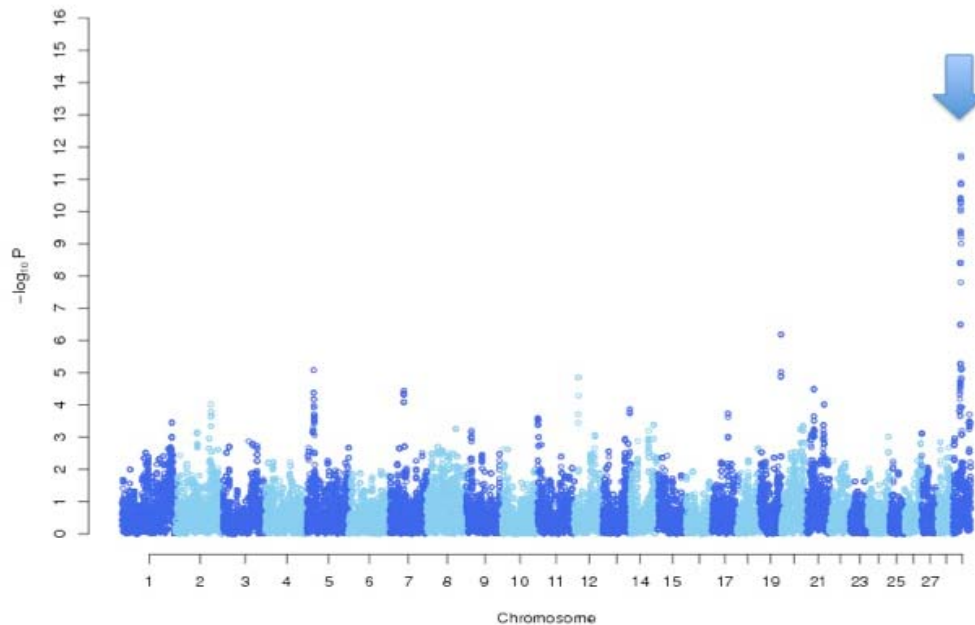
RESEARCH LETTER

18. Selzer, R. *et al.* Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. *Genes Chromosom. Cancer* **44**, 305–319 (2005).
19. Workman, C. *et al.* A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol.* **3**, 0048 (2002).
20. Olshen, A. *et al.* Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–572 (2004).
21. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
22. James, T. *et al.* Integrative genomics viewer. *Nature Biotechnol.* **29**, 24–26 (2011).

SUPPLEMENTARY INFORMATION

doi:10.1038/nature10757

Supplementary Fig. 1:



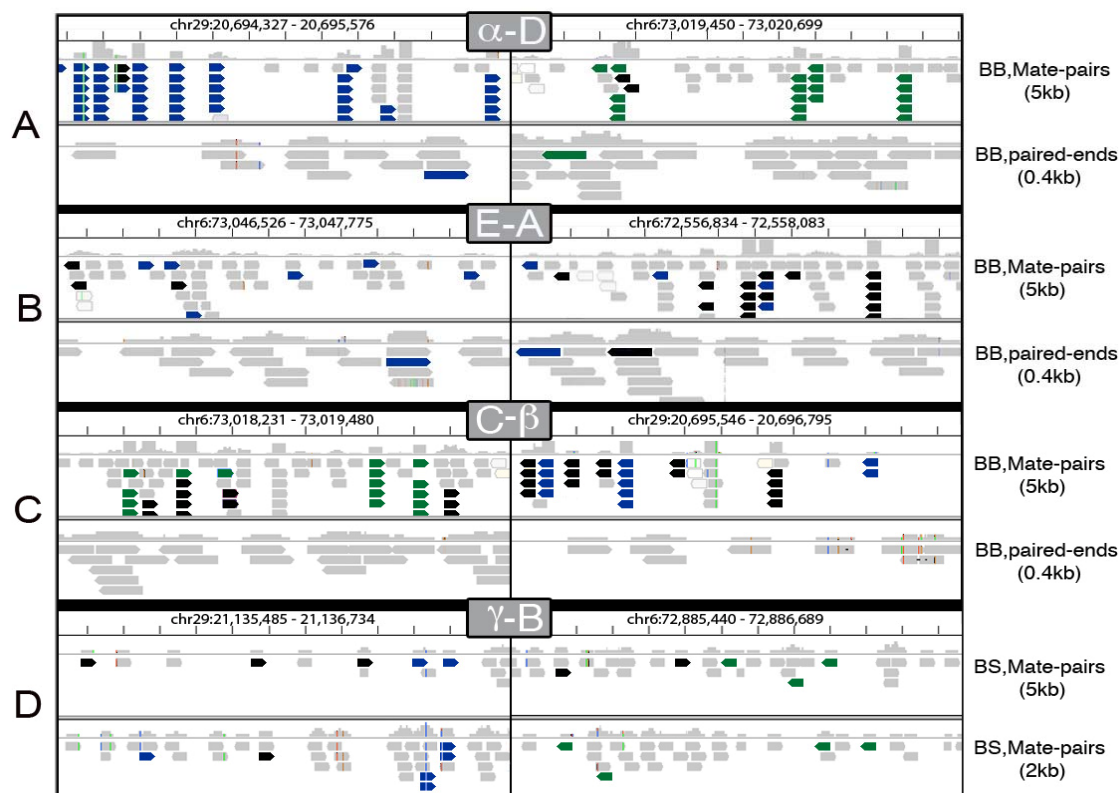
Association mapping of the *Cs* locus conducted with Glasgow (Druet et al., in preparation) in the Belgian Blue Cattle population using 21 color-sided animals and 30 non-white controls. SNP genotypes were first phased using the Hidden Markov Model described by Druet & Georges²³ and implemented with Dualphase. Dualphase assigns haplotypes to a defined number of Hidden States corresponding to clusters of genealogically related chromosomes. Glasgow then searches for association between haplotype states and a binary trait. It uses a Generalized Linear Mixed Model with the logit function as inverse link function to transform an underlying linear variable to a binary trait. The model includes a random haplotype effect (with zero covariance between the effects of distinct haplotypes) as well as a random polygenic effect (with variance-covariance matrix proportionate to genome-wide SNP identity-by-state) to correct for stratification. Following Verbeke & Molenberghs²⁴ and Tzeng & Zhang²⁵, we evaluated the strength of association using a score test. Under the null hypothesis, the score test has an approximate gamma distribution. We empirically determined the mean and variance of the gamma distribution at each marker position from 1,000 permutations of the residuals (observations corrected for modeled effects).

The graph shows the location scores (expressed as $-\log(p)$) that were obtained using Glasgow across the genome. The genome-wide significant signal on chromosome 29 is highlighted by the arrow.

doi:10.1038/nature10757

RESEARCH SUPPLEMENTARY INFORMATION

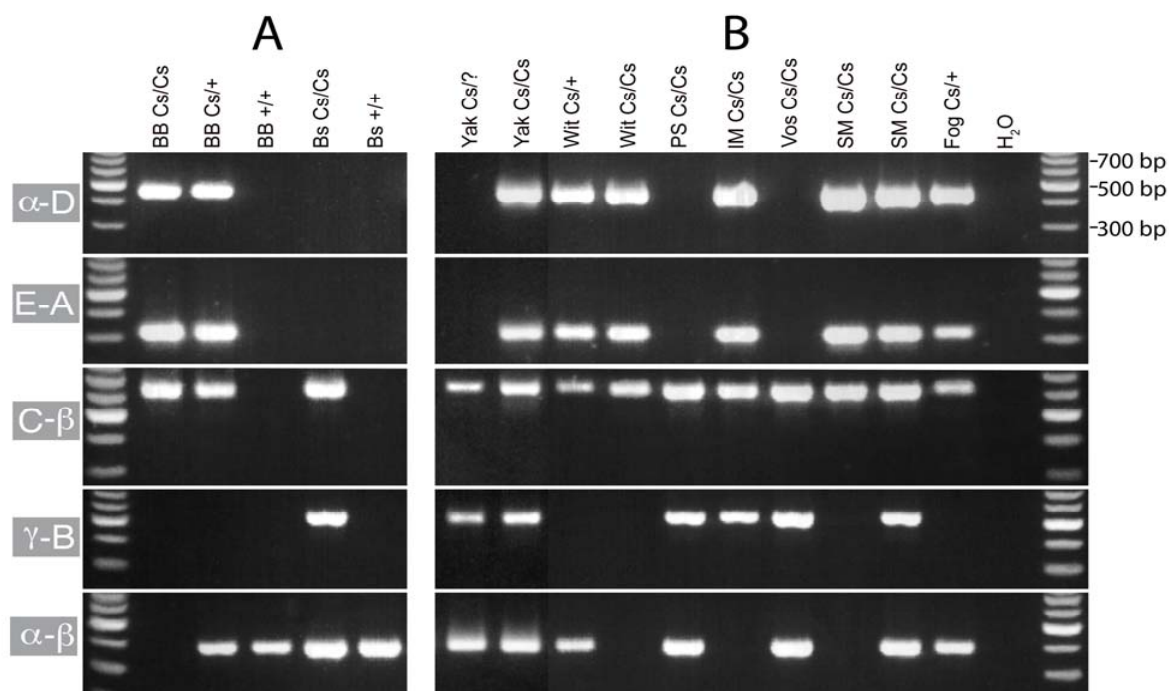
Supplementary Fig. 2:



Integrated Genome Viewer (IGV)²² screen-captures showing aligned mate-pair (~36-bp) or paired-end (~110-bp) reads around the four breakpoints associated with CS alleles (α -D, A-E, C- β and γ -B; cfr. Fig. 3). Reads from aberrantly mapping pairs are color-coded to indicate the chromosome to which the other pair maps (blue: = BTA6; green: BTA29; black: any other chromosome). The vertical line marks the position of the four breakpoints. Positions correspond to *bosTau4* assembly.

doi:10.1038/nature10757

RESEARCH SUPPLEMENTARY INFORMATION

Supplementary Fig. 3:

PCR amplification and agarose gel electrophoresis of amplicons spanning the breakpoints associated with color-sidedness (α -D, A-E, C- β , γ -B and α - β ; cfr. Fig. 3). **(A)** BB: Belgian Blue; Bs: Brown Swiss; **(B)** Yak: domestic yak; Wit: Witrik; PS: Pustertaler Sprinzen; IM: Irish Moiled; Vos: Vosgienne; SM: Swedish Mountain; Fog: Fogera. Data from Czech Red Spotted Cattle not shown.

Wild Type

5'-CTTCATCTTTGGTTTATGACCATG-3'
3'-GAAGTAGAACAACAACTAGGTACGTAGTGGATTGGAGTCGAA-5'

<-1st slip

5'-CTTCATCTTTGGTTTATGACCATGTTTATGACCATGTCATCA-3'
3'-GAAGTAGAACAACAACTAGGTACGTAGTGGATTGGAGTCGAA-5'

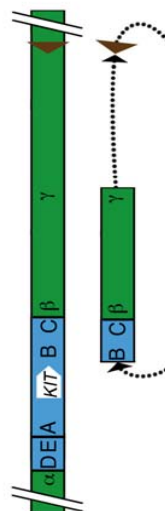
<- 2nd slip

5'-CTTCATCTTTGGTTTATGACCATGGTTTATGACCATGTCATCTCTCA-3'
3'-GAAGTAGAACAACAACTAGGTACGTAGTGGATTGGAGTCGAA-5'

<-...3rd slip

5'-CTTCATCTTTGGTTTATGACCATGGTTTATGACCATGTCATCTCTCACTCAGCTT-3'
3'-GAAGTAGAACAACAACTAGGTACGTAGTGGATTGGAGTCGAA-5'

Combined Sequence



B

α -D

Ref.-BTA29
Prox
Ref.-BTA6

TCCTTG-----GTTTATGACCATGCACTACCTCMCTGAGCTCTATTGTGACATCTC
TCCTTGTTTATGACCATGTTTATGACATGCAATATGGGTCCGACAGGTCAACACACTGAAG
TCCCAAGGACGGGGAGCCTGGTGGGCTCCCTCTGTATGGGGTCCGACAGGTCAACACACTGAAG

Ref-βT29
 Dist
 Ref-βT46
 C-β

TCCTTGGTTTGA-----CATGCATCTCCAACTGCTTATTTGCATCTCTCT
 GAGCTGTGGGGTGCATCATCTGCATGATCTCCAACTGAGTTCATTTGCATCTCTTT
 GAGCTGTGGGGTGCAGTETA TGGGGTGCACAGAGTCAACACAACTGAAGCGATTAGCAGCAG

C

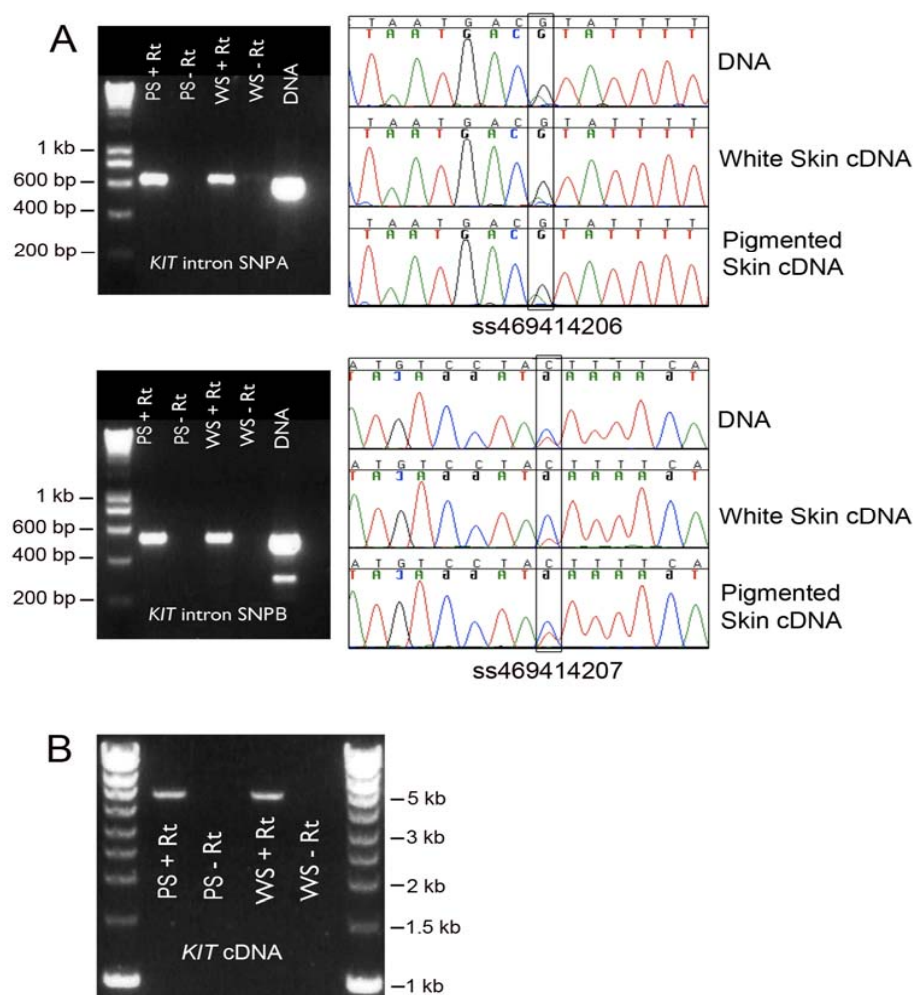
	Ref-RTA29: 21-1*	Breakpoint+ Ref-RTA29: 21-7*
Y-B	AGGTCACCTGAGCTCTTTGAANAATTTGTGACAAACGATCTGCTGCTTCAGAGCTTATG AGGTCACCTGAGCTCTTTGAANAATTTGTGACCTACCTGATTAATTAATCAGCTTGTGAATAC ATGCACCAAAATTTATCTATATGACCTTTTTCAGTTACTGATTAATTAATCAGTTGTAATCT	CTTCTCTCTATTCAAAATCAGCAGCAATCTAATTTTACAGGGAGGGGATCCCCAAACCCAGCA GCTCTATGACAGCAAAATCTGCTCTGCTTTTTCAGGGAGGGGATCCCCAAACCCAGCA GCTCTATGACAGCAAAATCTGCTCTGCTTTTAAAGTTACCTATGCTGATTTGATGATTTGATCC

doi:10.1038/nature10757

RESEARCH SUPPLEMENTARY INFORMATION

(A) Sequence of the E-A fusion showing the GC micro-homology at the two breakpoints, typical of replication-independent non-homologous end joining (NHEJ)²⁶. **(B)** Sequences of the α -D and C- β fusion points showing micro-duplications and micro-deletions reminiscent of replication-dependent micro-homology-mediated break-induced replication (MMBIR)²⁶. Examination of the BTA29 sequences flanking the BTA6 *CS29* insertion, suggests the occurrence of three slippage events that might have resulted from replication fork stalling and template switching (FoSTeS)²⁶. The resulting sequence is characterized by an inverted palindromic 6-bp repeat able to form a hairpin loop. The BTA6 sequences inserted at the base of the loop (marked by the asterisk). **(C)** Sequence of the γ -B fusion showing signatures of replication-dependent micro-homology-mediated break-induced replication (MMBIR) including the micro-insertion (between γ and B) of an inverted 75-bp sequence originally found 644-kb downstream of γ , flanked on either side by tracks of micro-homology with the ends of B and γ respectively.

Supplementary Figure 5:



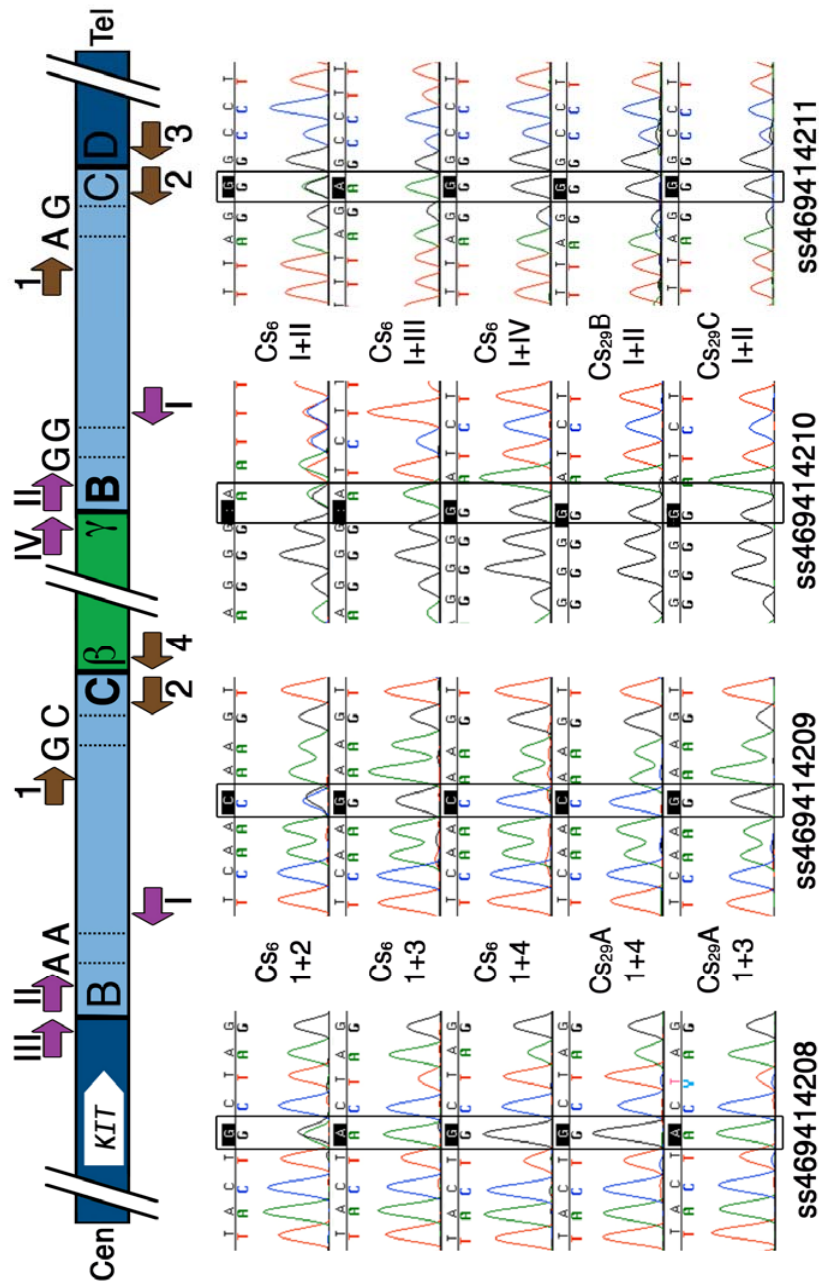
(A) RT-PCR analysis of 673-bp and 533-bp *KIT* amplicons, encompassing the intron 1 "ss469414206" and intron 7 "ss469414207" SNPs, respectively. Direct sequencing of the RT-PCR products obtained from RNA extracted from white (WS) and pigmented (PS) skin of a color-sided animal with (GG)₆/A₂₉ ("rsA") and (CC)₆/T₂₉ ("rsB") genotype reveals BTA₂₉-derived T and A bearing mRNA molecules demonstrating the transcriptional potential of the *KIT* gene on the Cs₂₉ allele. **(B)** Long range RT-PCR analysis of the near complete *KIT* mRNA generates a product with expected size (4,527-bp) using total RNA extracted from white (WS) and pigmented (PS) skin of the same color-sided animal.

doi:10.1038/nature10757

RESEARCH

SUPPLEMENTARY INFORMATION

Supplementary Figure 6:



doi:10.1038/nature10757

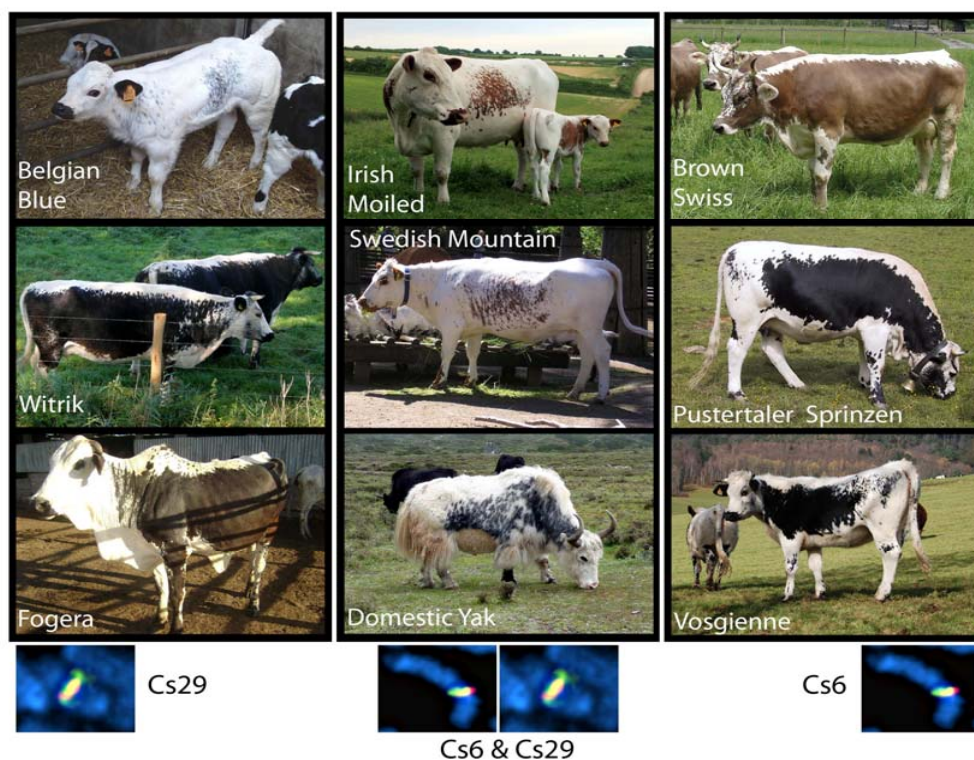
RESEARCH SUPPLEMENTARY INFORMATION

Determining the origin of BTA6 DNA flanking BTA29 DNA in *Cs*₆. Schematic representation of the *Cs*₆ allele: dark blue: non-duplicated BTA6 segments (corresponding to A (left) and D-E (right)); light blue: duplicated portions of BTA6 (B-C and B-C); green: translocated BTA29 segment (β - γ). Primers 1 & 2 and I & II (amplifying segments C & C and B & B, respectively) were used to confirm four SNPs ("ss469414208", "ss469414209", "ss469414210" and "ss469414211") identified in the Brown Swiss *Cs/Cs* next generation sequencing reads using conventional PCR followed by Sanger sequencing. Primers 1 and I were then paired (i) with primers 3 and III to amplify (under long range PCR conditions) the B and C segments flanking the non-duplicated BTA6 sequences, and (ii) with primers 4 and IV to amplify the B and C segments flanking the BTA29-derived sequences. Determining the genotype of the *Cs*₂₉ allele for segment C was achieved by sequencing amplicon 1-4 obtained from genomic DNA of a Belgian Blue color-sided animal. Determining the genotype of the *Cs*₂₉ allele for segment B was achieved by sequencing amplicon I-II obtained from genomic DNA of multiple Belgian Blue color-sided animals. The B and C segments flanking the BTA29 insert (β - γ) in the *Cs*₆ allele were thereby shown to match the B and C segments of the *Cs*₂₉ allele, as predicted from the resolution of a circular intermediate by a non-allelic homologous recombination event in the B-C interval. *Cs*₆ = *Cs*₆/+ animal; *Cs*₂₉A, *Cs*₂₉B, *Cs*₂₉C = *Cs*₂₉/+ Belgian Blue animals.

doi:10.1038/nature10757

RESEARCH SUPPLEMENTARY INFORMATION

Supplementary Figure 7:



Genotypes of color-sided cattle from eight breeds as well as color-sided domestic yaks with regards to the *Cs29* and *Cs6* alleles. Both alleles were introgressed from cattle into domestic yak by hybridization. Czech Red Spotted Cattle not shown

Picture Credits

Belgian Blue: Arnaud Sartelet, Brown Swiss: Tosso Leeb, Irish Moiled: Jane Light, Swedish Mountain: Liftarn, Domestic Yak: Peter Kirwan, Pustertaler Sprinzen: & Fogera: Johann Sölkner, Vosgienne: Carole Charlier, Witrik: Cord Drögemüller.

doi:10.1038/nature10757

RESEARCH SUPPLEMENTARY INFORMATION

Supplementary material: Association mapping of dispersed duplications in the human and bovine genome.**Methods:**

Data. We took advantage of genotype data obtained (i) on 856 individuals of Northern European descent genotyped with the Human 660W-Quad V1 DNA Analysis Beadchip array, (ii) 275 Belgian Blue animals genotyped with the Illumina Bovine 700K HD SNP array, and (iii) 191 Holstein-Friesian animals genotyped with the Illumina Bovine 700K HD SNP array.

CNV detection: CNVs were detected using PennCNV⁵. Overlapping CNVs were merged into CNV-regions (CNVR). Only CNVR encompassing only duplications (copy number 3 or 4) were retained for further analysis.

Phasing: Genotypes were phased using Dualphase²³. As a result, all chromosomes in the data are assigned to a predetermined number (in this case 20) of Hidden Haplotype Clusters at each SNP position.

Association mapping: To perform association analysis between CNV genotype and SNP genotype, we considered all individuals with PennCNV assigned copy number of 3 or 4 as “cases” and all individuals with PennCNV assigned copy number of 2 as “controls”. The association analysis was conducted using Glasgow (described under Suppl. Fig. 1). A random polygenic effect was added to the model for the cow data but not for the human data.

Significance thresholds. We generated QQ-plots using all p-values obtained with marker positions that were not syntenic with the SNPs involved in the corresponding CNVR. In both human and bovine, the QQ-plot revealed a pronounced inflexion at p-values of 10^{-9} , which was therefore selected as significance threshold (**A**).

Results:

We observed 21 putative trans-associations with p-value $< 10^{-9}$ in human and four in the cow. An example of such putative trans-association is shown in panel (**B**). The main features of the corresponding putative dispersed duplications are summarized in table format.

doi:10.1038/nature10757

RESEARCH SUPPLEMENTARY INFORMATION

Summary of putative dispersed duplications identified in human and bovine

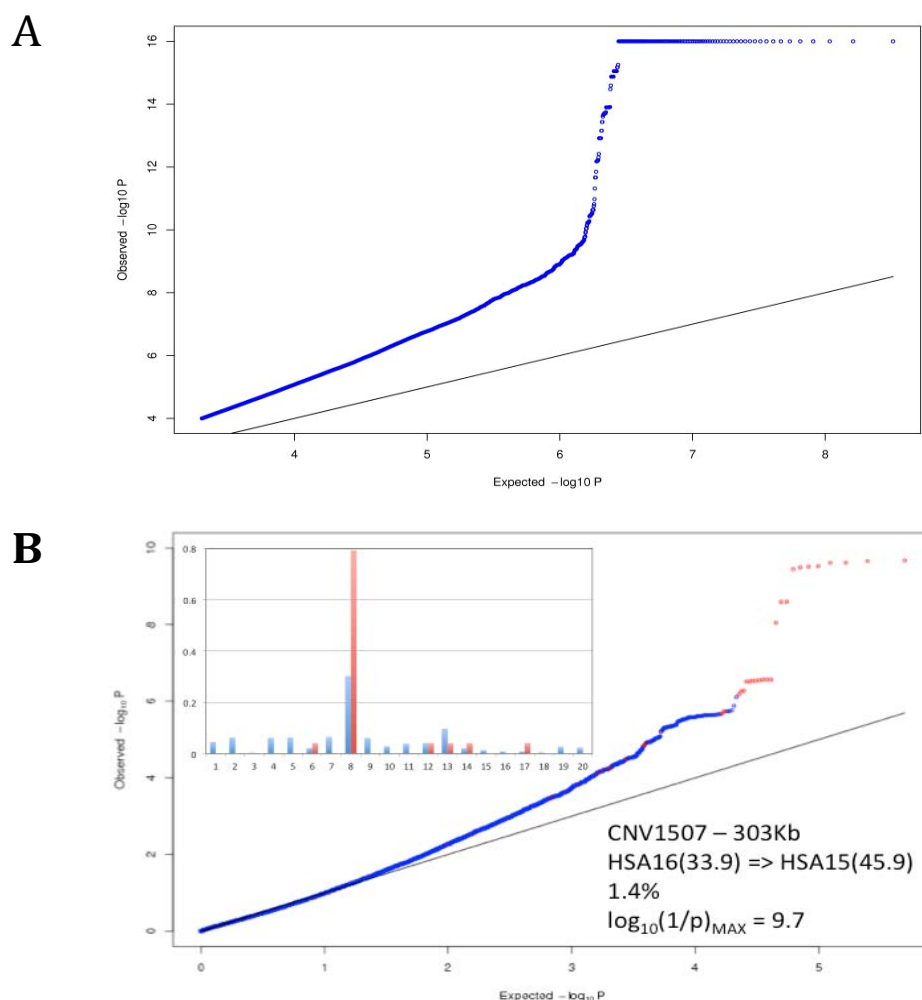
CNVR(egion)						Trans-association		
ID	Chr	Start	End	Length	Frequ.#	Chr	Pos.	Log(1/p)
HUMAN								
1	1	20141	137979	117838	1.6%	17	11395390	10.8
78	1	142536304	142691538	155234	27.8%	4	52811970	16.0
339	3	133969263	133980309	11046	3.9%	12	56365699	10.8
1504	16	32472505	33801989	1329484	21.0%	8	87146973	16.0
383	4	939113	1092019	152906	42.5%	3	192311220	11.8
384	4	1237014	1245277	8263	1.7%	8	38433937	10.5
1190	12	751308	752432	1124	3.3%	3	192311220	10.4
1393	15	20201673	21259305	1057632	53.0%	2	69990290	11.0
1578	17	32058858	32066000	7142	6.6%	1	88973472	16.0
25	1	13374687	13488491	113804	2.0%	13	43798077	9.0
161	2	29335758	29338734	2976	2.3%	21	39227602	9.6
305	3	75605823	75615298	9475	4.7%	13	24007495	9.2
439	4	69371230	69502640	131410	3.5%	6	20487250	9.1
532	5	26219551	26238133	18582	25.7%	7	18935767	9.0
867	8	1947417	1951113	3696	3.5%	18	74779866	9.1
922	8	129763003	129779232	16229	2.2%	11	117322782	9.4
1382	14	104313212	104314910	1698	2.8%	12	56365699	9.6
1451	15	101791262	101792488	1226	2.1%	17	31536843	9.5
1482	16	16460524	16730728	270204	9.0%	9	95753319	9.6
1507	16	33872091	34191016	318925	1.4%	15	45945513	9.7
1654	18	44774027	44788908	14881	1.5%	1	218094151	9.8
BOVINE								
360	18	27914135	28375996	461861	12.6%	12	809826	14.8
434	20	39350230	39388277	38047	15.7%	4	751020	16.0
971	8	51367642	51390418	22776	4.0%	17	12898336	14.4
1543	15	47513169	47522748	9579	3.6%	27	42483035	10.1

#Frequ.: percentage of individuals with Penn-CNV copy number 3 or 4.

Assembly used, Bovine = UMD_3.1, Human = hg19

doi:10.1038/nature10757

RESEARCH SUPPLEMENTARY INFORMATION



Legend: **(A)** QQ-plot showing the distribution of $-\log_{10}p$ values obtained across all CNVR with all non-syntenic SNP positions. A marked inflection was observed at ~ 9 , justifying the choice of $p\text{-values} \leq 10^{-9}$ as significance threshold. Very similar results were obtained with the bovine data (not shown). **(B)** Example of a putative dispersed duplication found by association in human. The main graph shows the QQ-plot of all $-\log_{10}p$ values obtained for the corresponding CNVR. Points in red correspond to all chromosome 15 SNPs, while the points in blue correspond to the SNP positions on all other chromosomes. The bar graph in the inset shows the frequency of the 20 Hidden Haplotype Clusters in “cases” (i.e. individuals with copy number 3 or 4) in red and in “controls” (i.e. individuals with copy number 2) in blue.

doi:10.1038/nature10757

RESEARCH SUPPLEMENTARY INFORMATION

Supplementary table 1: Linkage analysis in Brown Swiss cattle

We selected two microsatellites each for five functional candidate genes from the cattle genome sequence (UMD 3.1 assembly). We genotyped these 10 microsatellites in three two-generation paternal half-sib families segregating for color-sidedness. The three families comprised a total of 37 informative meioses. We calculated two-point linkage using the Merlin software assuming a fully penetrant dominant inheritance. The data clearly indicated that the KIT locus on BTA 6 is linked to color-sided in Brown Swiss cattle.

<i>Gene</i>	<i>BTA</i>	<i>Position (Mb)</i>	<i>Repeat motif</i>	<i>Alpha</i>	<i>Heterogeneity LOD (HLOD)</i>	<i>Two point LOD</i>
<i>KITLG</i>	5	18.567	(AC) ₁₈	0	0	-∞
<i>KITLG</i>	5	19.360	(AC) ₂₀	0.136	0.176	-∞
<i>ADAMTS20</i> ¹⁾	5	37.135	(TA) ₂₇	n.d.	n.d.	n.d.
<i>ADAMTS20</i>	5	37.233	(TA) ₂₁	0	0	-∞
<i>KIT</i>	6	71.789	(TA) ₈ (TG) ₂₀	1	6.894	6.894
<i>KIT</i>	6	71.844	(TG) ₁₁ (TA) ₁₅	0.819	3.306	3.236
<i>EDNRB</i>	12	53.296	(GT) ₁₅	0	0	-∞
<i>EDNRB</i>	12	53.360	(TG) ₁₉	0	0	-∞
<i>MITF</i>	22	31.750	(CA) ₁₈	0	0	-1.548
<i>MITF</i>	22	31.794	(GT) ₁₉	0	0	-0.836

¹⁾ This marker was not informative in the analyzed families.

doi:10.1038/nature10757

RESEARCH SUPPLEMENTARY INFORMATION

Supplementary table 2: Primer sequences

	Product Name	Forward (5'-3')	Reverse (5'-3')	Size bp	Region amplified bosTau4
Primers to ID BACs	KIT_ Ex1B	CTCGAAAGAACAGGGGTCAG	AGGAGCAGCAGAACGAAGAG	274	chr6:72,741,131-72,741,404
	BTA29_20.8	TTTACCCGGAAATCCACAAA	GGTTCAGAGCTAGGGTGTGC	379	chr29:20,899,945-20,900,323
Breakpoint primers	α-D	GGGGAGAATCTGTTTTCCTG	GGAAGGCCTTATTGCACACT	417	BTA29 to 6 breakpoint unique to BTA29
	A-E	GAAGCAACCCAGAGATGAGC	AAGGGAAGCCCATATGATGA	318	Fusion point on BTA6
	C-β	TCAACGAGGGACACATGAA	CAATTGACCCCTCATTTTGG	606	Common breakpoint, found on BTA6 & BTA29
	γ-B	GCTGCAGAAAATGTTATTCCA	TCTTGAAGGGCCATAGCATC	525	BTA29 to 6 breakpoint unique to BTA6
	α-β	GGGGAGAATCTGTTTTCCTG	TAAAGTCGCCAGTGCAAGTG	394	Will not amplify in BTA29 Cs/Cs animals
Primers to check for SNPs in KIT coding sequence	KIT_ Ex1	CTGGGCTCAGCCTTCTACC	TCCTGAAAGACTCGCAGCTC	928	chr6:72,740,730-72,741,657
	KIT_ Ex2	GGAACTTGACCCCGTTGTA	CATACCCGAAGCCACTATGC	743	chr6:72,779,376-72,780,118
	KIT_ Ex3	CCGAAAGGCAACGTCTTAGAT	ATTTTGAGGCTGGGAGAAC	500	chr6:72,782,945-72,783,444
	KIT_ Ex4	CATGGCTGAGGAAAAATGGT	GTGCTATGCAATGGGGAAT	516	chr6:72,784,309-72,784,824
	KIT_ Ex5	GCACTGCAGAGAATTTGGAA	TTGCTTTTGTGCTCTGGTTG	630	chr6:72,788,425-72,789,054
	KIT_ Ex6	TCTTCCGTTTCATTCTGCTG	AGCCCAAACCTTCTCTGT	533	chr6:72,793,394-72,793,926
	KIT_ Ex7	GAGGCTGAACAGAGGACCAG	TCATGTGTCAGCGAATTGT	624	chr6:72,795,451-72,796,074
	KIT_ Ex8	GGAGCTTCAGCATCTTCACC	TCTACCTGCAGGCTGGAAAT	786	chr6:72,811,793-72,812,578
	KIT_ Ex9	CCGATGCCTTCAGTTGATTT	GCCAGTGATGGAATGGACTT	342	chr6:72,813,937-72,814,278
	KIT_ Ex10-11	TGGAGGTGAGAGGTGTTGTG	CTAAAGGCAATGCCATGTGA	459	chr6:72,815,316-72,815,774
	KIT_ Ex12-13	CCACCACCACCATTTATTCC	CCATTTGGGTCAAAATCCTG	458	chr6:72,815,948-72,816,405
	KIT_ Ex14	CTGACCCCTAATCAGGCAGA	GCCTTTCCCATGTTCCCTAT	862	chr6:72,817,391-72,818,232
	KIT_ Ex15	ATAGCCTGCCTCTCACATGC	CAGTGACAACACCAAGG	562	chr6:72,819,627-72,820,188
	KIT_ Ex16	TTCAGCACCTTCCTGTCCCT	TCAAGCGACACTCTGCATT	378	chr6:72,820,095-72,820,472
	KIT_ Ex17	GGCACC GAATGGTTTAAATG	TTCTCCTGCTGTGACCTTCA	567	chr6:72,821,253-72,821,819
	KIT_ Ex18-19	TTGGATCTTTTGCTCTCCA	GCGACCGAAATAACATTTGC	508	chr6:72,824,397-72,824,904
	KIT_ Ex20	GTAAGGGCCAGATGTCCTT	CCAAGAGAATGGAGGTCCTG	494	chr6:72,824,748-72,825,241
	KIT_ Ex21	CATTCCAGCAGAAAAGCACA	TTTCCGCATCAAGGATAAG	791	chr6:72,826,056-72,826,846
	KIT_ 3'UTR-A	GTCCTTCCCAAGGTTTCTCC	TGCTGAAAGCCAGGCTACTT	903	chr6:72,826,694-72,827,596
	KIT_ 3'UTR-B	TGATGCCGTTTGAAAAAGTG	GGAAAGGTGCGAGAGCATAG	712	chr6:72,827,333-72,828,044
	KIT_ 3'UTR-C	TGTTGTCTCGCAGGATTCAG	CATCTGGGAAACCTCACCTT	756	chr6:72,827,909-72,828,664
Expression primers	KIT_ Intron_ Rs1	AAGGAGGCGTTACCAGGTTT	TGCTGGAATGGAAATTAGGC	637	chr6:72,751,987-72,752,623
	KIT_ Intron_ Rs2	TTATTCAGGTCCTGCTTG	AGGGGCATCCTTGAGAGTTT	533	chr6:72,809,532-72,810,064

WWW.NATURE.COM/NATURE114

doi:10.1038/nature10757

RESEARCH SUPPLEMENTARY INFORMATION

	KIT_ Ex1-3'UTR	CTCTTCGTTCTGCTGCTCCT	GGAAAGGTGCGAGAGCATAG	4,527	KIT cDNA
Primers to determine origin of BTA6 DNA flanking BTA29 DNA in Cs6					
SupFig6 1+2	SNP group A	CGTGTGTGAGCATGCTAGGT	GACCAAGACCCCACTTCCT	885	chr6:73,008,888-73,009,772
SupFig6 1+3	SNP group A 6-6*	CGTGTGTGAGCATGCTAGGT	CCGCCTCTCTCATTATCAGC	11,164	chr6:73,008,888-73,020,051
SupFig6 1+4	SNP group A 6-29	CGTGTGTGAGCATGCTAGGT	CAATTGACCCCTCATTTTGG	10,909	BTA6 DNA upstream of common breakpoint
SupFig6 II+I	SNP group B	GCAGTAGATGCCCCATAGA	AGCAGAGGAAGAAGCAGCAG	792	chr6:72,896,736-72,897,527
SupFig6 III+I	SNP group B 6-6#	GCTAGCCAGATCCAGCAGTC	AGCAGAGGAAGAAGCAGCAG	11,792	chr6:72,885,736-72,897,527
SupFig6 IV+I	SNP group B 29-6	GCTGCAGAAAATGTTATTCCA	AGCAGAGGAAGAAGCAGCAG	11,687	BTA6 DNA downstream of BTA6 specific break
6-6* Primers span region C-D			6-6# Primers span region B and non duplicated part of BTA6		
6-29, Primers span breakpoint C-β			29-6, Primers span breakpoint γ-B		
SNP group A: ss469414208, ss469414209			SNP group B: ss469414210, ss469414211		

WWW.NATURE.COM/NATURE115

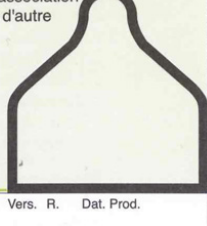
doi:10.1038/nature10757

RESEARCH SUPPLEMENTARY INFORMATION

Supplementary References

23. Druet T & Georges M. A hidden markov model combining linkage and linkage disequilibrium information for haplotype reconstruction and quantitative trait locus fine mapping. *Genetics* **184**,789-798 (2010)
24. Verbeke G & Molenberghs G. The use of score tests for inference on variance components. *Biometrics* **59**, 254-262 (2003)
25. Tzeng J-Y & Zhang D. Haplotype-based association analysis via variance-components score test. *Am. J. Hum. Genet.* **81**, 927-938 (2007)
26. Hastings *et al.* Mechanisms of change in gene copy number. *Nat Rev Genet* **10**, 551-564 (2009).

Annexe 7 : Passeport d'identification d'un bovin par le système SANITEL.

 Validité limitée Beperkte geldigheid Begrenzte Gültigkeit Vers. R. Dat. Prod.	Ne peut circuler en-dehors de Mag zich niet bevinden buiten Verkehr verboten außerhalb et communes avoisinantes en omliggende gemeenten und angrenzenden Gemeinden	BELGIQUE - BELGIË - BELGIEN Statut Statuut Status Passeport Bovin Runderpaspoort Rinderpass Collez ici l'étiquette abattoir jaune si nécessaire Kleef hier indien nodig het gele slachthuisetiket Hier gegebenenfalls die gelbe Schlachthofvignette aufkleben Vers. R. Dat. Prod.	 Vers. R. Dat. Prod.
	Troupeau / Beslag / Herde Date de sortie Vertrekkdatum / Abgangsdatum 20 Signature cédant Handtekening overlater Unterschrift Überlasser	Naissance / Geboorte / Geburt Robe / Haarkleed / Haarkleid Sexe / Geslacht / Geschlecht Type racial / Rastype / Rasstyp Mère / Moeder / Mutter Troupeau Naissance / Geboortebeslag / Geburtsherde: Historique / Historiek / Vorgeschichte: Pays Troupeau Date d'arrivée Troupeau précédent / Vorig beslag / Vorherige Herde:	
Bovin mort: ☐ Date de sortie 20 N° Preneur Preneur  Troupeau Naissance Robe Sexe Type racial Mère Vers. R. Dat. Prod.	Numéro Marque auriculaire N° de version document Date de sortie 20 Nbre de jours depuis le départ: ☐ N° Preneur Preneur Signature N° de communication VRS/WEB Date de notification 20	Volet de sortie Envoi postal dans les délais légaux à l'association agréée à cet effet si vous n'utilisez pas d'autre système de communication agréé par l'autorité compétente.  Vers. R. Dat. Prod.	Troupeau Propriétaire Responsable Collez ici l'étiquette abattoir blanche si nécessaire

BIBLIOGRAPHIE

1 000 GENOME PROJECT CONSORTIUM. A map of human genome variation from population-scale sequencing. *Nature*, 2010, **467**, 1061-1073.

A JMONE-MARSAN P., GARCIA J.F., LENSTRA J.A., THE GLOBALDIV CONSORTIUM. On the origin of the cattle: How aurochs became cattle and colonized the world. *Evol. Anthropol.*, 2010, **19**, 148-157.

ALTSHULER D., DALY M.J., LANDER E.S. Genetic mapping in human disease. *Science*, 2008, **322**, 881-888.

AGERHOLM J.S., BASSE A., CHRISTENSEN K. Investigations on the occurrence of hereditary diseases in the Danish cattle population 1989-1991. *Acta Vet. Scand.*, 1993, **34**, 245-253.

AGERHOLM J.S. Inherited disorders in Danish cattle. *APMIS Suppl.*, 2007, 1-76.

AMORY H., DESMECHT D.J., LINDEN A.S., MCENTEE K., ROLLIN F.A., BEDUIN J.M., GENICOT B.C., D'ORIO V., LEKEUX P.M. Comparison of cardiac function in double-muscled calves and in calves with conventional muscular conformation. *Am. J. Vet. Res.*, 1994, **55**, 561-566.

ANDERSSON L. Genetic dissection of phenotypic diversity in farm animals. *Nat. Rev. Genet.*, 2001, **2**, 130-138.

ARDLIE K.G., KRUGLYAK L., SEIELSTAD M. Patterns of linkage disequilibrium in the human genome. *Nat. Rev. Genet.*, 2002, **3**, 299-309

B ALDING D.J. A tutorial on statistical methods for population association studies. *Nat. Rev. Genet.*, 2006, **7**, 781-791.

BAMSHAD M.J., NG S.B., BIGHAM A.W., TABOR H.K., EMOND M.J., NICKERSON D.A., SHENDURE J. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.*, 2011, **12**, 745-755.

BASRUR P.K., KING W.A. Genetics then and now: breeding the best and biotechnology. *Rev. Sci. Tech.*, 2005, **24**, 31-49.

BARENDSE W., ARMITAGE S.M., KOSSAREK L.M., SHALOM A., KIRKPATRICK B.W., RYAN A.M., CLAYTON D., LI L., NEIBERGS H.L., ZHANG N., GROSSE W.M., WEISS J., CREIGHTON P., MCCARTHY F., RON M., TEALE A.J., FRIES R., MCGRAW R.A., MOORE S.S., GEORGES M., SOLLER M., WOMACK J.E., HETZEL D.J.S. A genetic linkage map of the bovine genome. *Nat. Genet.*, 1994, **6**, 227-235.

BAYROU C., GARIGLIANY M.M., SARLET M., SARTELET A., CASSART D., DESMECHT D. Pathological manifestations and distribution of viral RNA in malformed newborn calves after natural intrauterine infection by emerging Schmallerberg virus. *Soumis pour publication*.

- BEHRENDT N. The urokinase receptor (uPAR) and the uPAR-associated protein (uPARAP/Endo180): membrane proteins engaged in matrix turnover during tissue remodeling. *Biol. Chem.*, 2004, **385**, 103-136
- BELLINGE R.H., LIBERLES D.A., IASCHI S.P., O'BRIEN P.A., TAY G.K. Myostatin and its implications on animal breeding: a review. *Anim. Genet.*, 2005, **36**, 1-6.
- BENNETT R.C., OLDS D., DEATON O.W., THRIFT F.A. Nature of white heifer disease (partial genital aplasia) and its mode of inheritance. *Am. J. Vet. Res.*, 1973, **34**, 13-19.
- BISHOP M.D., KAPPES S.M., KEELE J.W., STONE R.T., SUNDEN S.L., HAWKINS G.A., TOLDO S.S., FRIES R., GROSZ M.D., YOO J., BEATTIE C.W. A genetic linkage map for cattle. *Genetics*, 1994, **136**, 619-639.
- BOICHARD D., CHUNG H., DASSONNEVILLE R., DAVID X., EGGEN A., FRITZ S., GIETZEN K.J., HAYES B.J., LAWLEY C.T., SONSTEGARD T.S., VAN TASSELL C.P., VANRADEN P.M., VIAUD-MARTINEZ K.A., WIGGANS G.R. Design of a bovine low-density SNP array optimized for imputation. *PLoS One*, 2012, **7**, e34130.
- BOLLONGINO R., BURGER J., POWELL A., MASHKOUR M., VIGNE J.D., THOMAS M.G. Modern taurine cattle descended from small number of near-eastern founders. *Mol. Biol. Evol.*, 2012, **29**, 2101-2104.
- BONAITI B., BIBE B., HAVY A., MENISSIER F. Comparaison des races bovines Charolaise, Limousine et Maine-Anjou en race pure et en intercroisement 3. Performances d'abattage des taurillons purs et F1. *Genet. Sel. Evol.*, 1988, **20**, 461-476.
- BOTSTEIN D., WHITE R.L., SKOLNICK M., DAVIS R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.*, 1980, **32**, 314-331.
- BOUDRY B. Hypoplasie utérine chez le génisse Blanc-bleu Belge. (Mémoire de DEA). Université de Liège: Liège, 2002, 31 p.
- BOVINE GENOME SEQUENCING AND ANALYSIS CONSORTIUM. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*, 2009, **324**, 522-528.
- BOVINE HAPMAP CONSORTIUM. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science*, 2009, **324**, 528-532.
- BRACKETT B.G., BOUSQUET D., BOICE M.L., DONAWICK W.J., EVANS J.F., DRESSEL M.A. Normal development following in vitro fertilization in the cow. *Biol. Reprod.*, 1982, **27**, 147-158.
- BROWNING S.R., BROWNING B.L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet*, 2007, **81**, 1084-1097.

- BUREAU F., UYSTEPRUYST C.H., COGHE J., VAN DE WEERDT M., LEKEUX P. Spirometric variables recorded after lobeline administration in healthy Friesian and Belgian white and blue calves: normal values and effects of somatic growth. *Vet. J.*, 1999, **157**, 302-308.
- BUREAU F., DETILLEUX J., DORTS T., UYSTEPRUYST C., COGHE J., LEROY P.L., LEKEUX P. Spirometric performance in Belgian Blue calves: I. Effects on economic losses due to the bovine respiratory disease complex. *J. Anim. Sci.*, 2001, **79**, 1301-1304.
- BUREAU F., MICHAUX C., COGHE J., UYSTEPRUYST C., LEROY P.L., LEKEUX P. Spirometric performance in Belgian Blue calves: II. Analysis of environmental factors and estimation of genetic parameters. *J. Anim. Sci.*, 2001, **79**, 1162-1165.
- C**HANG Y.F., IMAM J.S., WILKINSON M.F. The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.*, 2007, **76**, 51-74.
- CHARLIER C., COPPIETERS W., FARNIR F., GROBET L., LEROY P.L., MICHAUX C., MNI M., SCHWERS A., VANMANSHOVEN P., HANSET R., GEORGES M. The mh gene causing double-muscling in cattle maps to bovine Chromosome 2. *Mamm. Genome*, 1995, **6**, 788-792.
- CHARLIER C., DENYS B., BELANCHE J.I., COPPIETERS W., GROBET L., MNI M., WOMACK J., HANSET R., GEORGES M. Microsatellite mapping of the bovine roan locus: a major determinant of White Heifer disease. *Mamm. Genome*, 1996, **7**, 138-142.
- CHARLIER C., COPPIETERS W., ROLLIN F., DESMECHT D., AGERHOLM J.S., CAMBISANO N., CARTA E., DARDANO S., DIVE M., FASQUELLE C., FRENNET J.C., HANSET R., HUBIN X., JORGENSEN C., KARIM L., KENT M., HARVEY K., PEARCE, B.R. SIMON P., TAMA N., NIE H., VANDEPUTTE S., LIEN S., LONGERI M., FREDHOLM M., HARVEY R.J., GEORGES M. Highly effective SNP-based association mapping and management of recessive defects in livestock. *Nat. Genet.*, 2008, **40**, 449-454.
- CHARLIER C., AGERHOLM J.S., COPPIETERS W., KARLSKOV-MORTENSEN P., DE JONG G., LI W., FASQUELLE C., KARIM L., CAMBISANO N., AHARIZ N., MULLAART E., GEORGES M., FREDHOLM M. A deletion in the bovine FANCI gene compromises fertility by causing fetal death and brachyspina. *PLoS One*, 2012, **7**, e43085.
- CHAROENTHONGTRAKUL S., GAO L., PARVATIYAR K., LEE D., HARHAJ E.W. RING Finger Protein 11 Targets TBK1/IKKi Kinases to Inhibit Antiviral Signaling. *PLoS One*, 2013, **8**, e53717.
- COLIGE A., SIERON A.L., LI S.W., SCHWARZE U., PETTY E., WERTELECKI W., WILCOX W., KRAKOW D., COHN D.H., REARDON W., BYERS P.H., LAPIERE C.M., PROCKOP D.J., NUSGENS B.V. Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis are caused by mutations in the procollagen I N-proteinase gene. *Am. J. Hum. Genet.*, 1999, **65**, 308-317.

- COOPMAN F., KRAFFT A., DEWULF J., VAN ZEVEEREN A., GENGLER N. Estimation of phenotypic and genetic parameters for weight gain and weight at fixed ages in the double-musced Belgian Blue Beef breed using field records. *J. Anim. Breed. Genet.*, 2007, **124**, 20-25.
- COPPIETERS W., RIQUET J., ARRANZ J.J., BERZI P., CAMBISANO N., GRISART B., KARIM L., MARCQ F., MOREAU L., NEZER C., SIMON P., VANMANSHOVEN P., WAGENAAR D., GEORGES M. A QTL with major effect on milk yield and composition maps to bovine chromosome 14. *Mamm. Genome*, 1998, **9**, 540-544.
- CORNILLIE P., VAN DEN BROECK W., SIMOENS P. Prolonged gestation in two Belgian blue cows due to inherited adenohipophyseal hypoplasia in the fetuses. *Vet. Rec.*, 2007, **161**, 388-391.
- CRADDOCK N., HURLES M.E., CARDIN N., PEARSON R.D., PLAGNOL V., ROBSON S., VUKCEVIC D., BARNES C., CONRAD D.F., GIANNOULATOU E., HOLMES C., MARCHINI J.L., STIRRUPS K., TOBIN M.D., WAIN L.V., YAU C., AERTS J., AHMAD T., ANDREWS T.D., ARBURY H., ATTWOOD A., AUTON A., BALL S.G., BALMFORTH A.J., BARRETT J.C., BARROSO I., BARTON A., BENNETT A.J., BHASKAR S., BLASZCZYK K., BOWES J., BRAND O.J., BRAUND P.S., BREDIN F., BREEN G., BROWN M.J., BRUCE I.N., BULL J., BURREN O.S., BURTON J., BYRNES J., CAESAR S., CLEE C.M., COFFEY A.J., CONNELL J.M., COOPER J.D., DOMINICZAK A.F., DOWNES K., DRUMMOND H.E., DUDAKIA D., DUNHAM A., EBBS B., ECCLES D., EDKINS S., EDWARDS C., ELLIOT A., EMERY P., EVANS D.M., EVANS G., EYRE S., FARMER A., FERRIER I.N., FEUK L., FITZGERALD T., FLYNN E., FORBES A., FORTY L., FRANKLYN J.A., FREATHY R.M., GIBBS P., GILBERT P., GOKUMEN O., GORDON-SMITH K., GRAY E., GREEN E., GROVES C.J., GROZEVA D., GWILLIAM R., HALL A., HAMMOND N., HARDY M., HARRISON P., HASSANALI N., HEBASHI H., HINES S., HINKS A., HITMAN G.A., HOCKING L., HOWARD E., HOWARD P., HOWSON J.M., HUGHES D., HUNT S., ISAACS J.D., JAIN M., JEWELL D.P., JOHNSON T., JOLLEY J.D., JONES I.R., JONES L.A., KIROV G., LANGFORD C.F., LANGO-ALLEN H., LATHROP G.M., LEE J., LEE K.L., LEES C., LEWIS K., LINDGREN C.M., MAISURIA-ARMER M., MALLER J., MANSFIELD J., MARTIN P., MASSEY D.C., MCARDLE W.L., MCGUFFIN P., MCLAY K.E., MENTZER A., MIMMACK M.L., MORGAN A.E., MORRIS A.P., MOWAT C., MYERS S., NEWMAN W., NIMMO E.R., O'DONOVAN M.C., ONIPINLA A., ONYIAH I., OVINGTON N.R., OWEN M.J., PALIN K., PARNELL K., PERNET D., PERRY J.R., PHILLIPS A., PINTO D., PRESCOTT N.J., PROKOPENKO I., QUAIL M.A., RAFELT S., RAYNER N.W., REDON R., REID D.M., RENWICK, RING S.M., ROBERTSON N., RUSSELL E., ST CLAIR D., SAMBROOK J.G., SANDERSON J.D., SCHUILENBURG H., SCOTT C.E., SCOTT R., SEAL S., SHAW-HAWKINS S., SHIELDS B.M., SIMMONDS M.J., SMYTH D.J., SOMASKANTHARAJAH E., SPANOVA K., STEER S., STEPHENS J., STEVENS H.E., STONE M.A., SU Z., SYMMONS D.P., THOMPSON J.R., THOMSON W., TRAVERS M.E., TURNBULL C., VALSESIA A., WALKER M., WALKER N.M.,

- WALLACE C., WARREN-PERRY M., WATKINS N.A., WEBSTER J., WEEDON M.N., WILSON A.G., WOODBURN M., WORDSWORTH B.P., YOUNG A.H., ZEGGINI E., CARTER N.P., FRAYLING T.M., LEE C., MCVEAN G., MUNROE P.B., PALOTIE A., SAWCER S.J., SCHERER S.W., STRACHAN D.P., TYLER-SMITH C., BROWN M.A., BURTON P.R., CAULFIELD M.J., COMPSTON A., FARRALL M., GOUGH S.C., HALL A.S., HATTERSLEY A.T., HILL A.V., MATHEW C.G., PEMBREY M., SATSANGI J., STRATTON M.R., WORTHINGTON J., DELOUKAS P., DUNCANSON A., KWIATKOWSKI D.P., MCCARTHY M.I., OUWEHAND W., PARKES M., RAHMAN N., TODD J.A., SAMANI N.J., DONNELLY P. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature*, 2010, **464**, 713-720.
- D**E SMET S., WEBB E.C., CLAEYS E., UYTTERHAEGEN L., DEMEYER D.I. Effect of dietary energy and protein levels on fatty acid composition of intramuscular fat in double-muscled Belgian Blue bulls. *Meat Sci.*, 2001, **56**, 73-79.
- DE ROOS A.P., HAYES B.J., SPELMAN R.J., GODDARD M.E. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics*, 2008, **179**, 1503-1512.
- DE VLAMYNCK C., VLAMINCK L., HAUSPIE S., SAUNDERS J., GASTHUYS F. Ultrasound-guided femoral nerve block as a diagnostic aid in demonstrating quadriceps involvement in bovine spastic paresis. *Vet. J.*, 2012.
- DANLOIS F., ZALTASH S., JOHANSSON J., ROBERTSON B., HAAGSMAN H.P., VAN EIJK M., BEERS M.F., ROLLIN F., RUYSSCHAERT J.M., VANDENBUSSCHE G. Very low surfactant protein C contents in newborn Belgian White and Blue calves with respiratory distress syndrome. *Biochem. J.*, 2000, **351**, 779-787.
- DANLOIS F., ZALTASH S., JOHANSSON J., ROBERTSON B., HAAGSMAN H.P., ROLLIN F., RUYSSCHAERT J.M., VANDENBUSSCHE G. Pulmonary surfactant from healthy Belgian White and Blue and Hin Friesian calves: biochemical and biophysical comparison. *Vet. J.*, 2003, 5-72.
- DING C., JIN S. High-throughput methods for SNP genotyping. *Methods Mol. Biol.*, 2009, **578**, 245-254.
- DEKKERS J.C. Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. *J. Anim. Sci.*, 2004, **82 E-Suppl**, E313-328.
- DESMECHT D., BERGH R.V., SARTELET A., LECLERC M., MIGNOT C., MISSE F., SUDRAUD C., BERTHEMIN S., JOLLY S., MOUSSET B., LINDEN A., COIGNOUL F., CASSART D. Evidence for transplacental transmission of the current wild-type strain of bluetongue virus serotype 8 in cattle. *Vet. Rec.*, 2008, **163**, 50-52.

- DERIVAUX J. L'opération césarienne chez la vache par la flanc gauche. Résultats de 260 interventions. *Rec. Med. Vet.*, 1955, **11**, 937-948.
- DEVLIN B., ROEDER K. Genomic control for association studies. *Biometrics*, 1999, **55**, 997-1004.
- DRUET T., GEORGES M. A hidden markov model combining linkage and linkage disequilibrium information for haplotype reconstruction and quantitative trait locus fine mapping. *Genetics*, 2010, **184**, 789-798.
- DUDBRIDGE F., GUSNANTO A., KOELEMAN B.P. Detecting multiple associations in genome-wide studies. *Hum. Genomics*, 2006, **2**, 310-317.
- DUCOS A., EGGEN A., DARRE R., BOICHARD D. Les anomalies génétiques dans l'espèce bovine. In: *Rend. Rech. Ruminants* (Ed.), Proceedings des 9^{ème} journées 3R, 2002, **9**, 85-91.
- DURKIN K., COPPIETERS W., DROGEMULLER C., AHARIZ N., CAMBISANO N., DRUET T., FASQUELLE C., HAILE A., HORIN P., HUANG L., KAMATANI Y., KARIM L., LATHROP M., MOSER S., OLDENBROEK K., RIEDER S., SARTELET A., SOLKNER J., STALHAMMAR H., ZELENKA D., ZHANG Z., LEEB T., GEORGES M., CHARLIER C. Serial translocation by means of circular intermediates underlies colour sidedness in cattle. *Nature*, 2012, **482**, 81-84.
- F**ARNIR F., COPPIETERS W., ARRANZ J.J., BERZI P., CAMBISANO N., GRISART B., KARIM L., MARCQ F., MOREAU L., MNI M., NEZER C., SIMON P., VANMANSHOVEN P., WAGENAAR D., GEORGES M. Extensive genome-wide linkage disequilibrium in cattle. *Genome Res.*, 2000, **10**, 220-227.
- FASQUELLE C., SARTELET A., LI W., DIVE M., TAMMA N., MICHAUX C., DRUET T., HUIJBERS I.J., ISACKE C.M., COPPIETERS W., GEORGES M., CHARLIER C. Balancing selection of a frame-shift mutation in the MRC2 gene accounts for the outbreak of the Crooked Tail Syndrome in Belgian Blue Cattle. *PLoS Genet.*, 2009, **5**, e1000666.
- FOOTE R.H. The history of artificial insemination: Selected notes and notables. *J. Anim. Sci.*, 2002, **80**, 1-10.
- G**ALLI C., CROTTI G., NOTARI C., TURINI P., DUCHI R., LAZZARI G. Embryo production by ovum pick up from live donors. *Theriogenology*, 2001, **55**, 1341-1357.
- GAO X., STARMER J., MARTIN E.R. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet. Epidemiol.*, 2008, **32**, 361-369.
- GARIGLIANY M.M., HOFFMANN B., DIVE M., SARTELET A., BAYROU C., CASSART D., BEER M., DESMECHT D. Schmallenberg virus in calf born at term with porencephaly, belgium. *Emerg. Infect. Dis.*, 2012, **18**, 1005-1006.

- GENGLER N., SEUTIN C., BOONEN F., VAN VLECK L.D. Estimation of genetic parameters for growth, feed consumption, and conformation traits for double-muscled Belgian blue bulls performance-tested in Belgium. *J. Anim. Sci.*, 1995, **73**, 3269-3273.
- GEORGES M., DIETZ A.B., MISHRA A., NIELSEN D., SARGEANT L.S., SORENSEN A., STEELE M.R., ZHAO X., LEIPOLD H., WOMACK J.E., LATHROP M. Microsatellite mapping of the gene causing weaver disease in cattle will allow the study of an associated quantitative trait locus. *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 1058-1062.
- GEORGES M., DRINKWATER R., KING T., MISHRA A., MOORE S.S., NIELSEN D., SARGEANT L.S., SORENSEN A., STEELE M.R., ZHAO X., WOMACK J.E., HETZEL J. Microsatellite mapping of a gene affecting horn development in *Bos taurus*. *Nat. Genet.*, 1993, **4**, 206-210.
- GEORGES M., NIELSEN D., MACKINNON M., MISHRA A., OKIMOTO R., PASQUINO A.T., SARGEANT L.S., SORENSEN A., STEELE M.R., ZHAO X., WOMACK J.E., HOESCHELE I. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics*, 1995, **139**, 907-920.
- GEORGES M. Mapping, fine mapping, and molecular dissection of quantitative trait Loci in domestic animals. *Annu. Rev. Genomics Hum. Genet.*, 2007, **8**, 131-162.
- GEORGES M. When Less Means More: Impact of Myostatin in Animal Breeding. *Immun. Endoc. & Metab. Agents in Med. Chem.*, 2010, **10**, 240-248.
- GEORGES M. Impact of high-throughput genotyping and sequencing on the identification of genes and variants underlying phenotypic variation in domestic cattle. In: Womack J. (Ed.), *Bovine genomics*. Wiley-Blackwell: Ames, 2012, 234-258.
- GILL J.L., JAMES V.M., CARTA E., HARRIS D., TOPF M., SCHOLLES S.F., HATELEY G., HARVEY R.J. Identification of congenital muscular dystonia 2 associated with an inherited GlyT2 defect in Belgian Blue cattle from the United Kingdom. *Anim. Genet.*, 2012, **43**, 267-270.
- GILL P., JEFFREYS A.J., WERRETT D.J. Forensic application of DNA 'fingerprints'. *Nature*, 1985, **318**, 577-579.
- GODDARD M.E., HAYES B.J. Genomic selection. *J. Anim. Breed. Genet.*, 2007, **124**, 323-330.
- GODDARD M.E., HAYES B.J. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat. Rev. Genet.*, 2009, **10**, 381-391.
- GOLDEN B.L., GARRICK D.J., BENYSHEK L.L. Milestones in beef cattle genetic evaluation. *J. Anim. Sci.*, 2009, **87** (suppl. 14), E3-10.
- GROSZ M.D., MACNEIL M.D. The "spotted" locus maps to bovine chromosome 6 in a Hereford-Cross population. *J. Hered.*, 1999, **90**, 233-236.
- GROBET L., MARTIN L.J., PONCELET D., PIROTTIN D., BROUWERS B., RIQUET J., SCHOEBERLEIN A., DUNNER S., MENISSIER F., MASSABANDA J., FRIES R., HANSET R., GEORGES M. A

- deletion in the bovine myostatin gene causes the double-muscled phenotype in cattle. *Nat. Genet.*, 1997, **17**, 71-74.
- GROBET L., PONCELET D., ROYO L.J., BROUWERS B., PIROTTIN D., MICHAUX C., MENISSIER F., ZANOTTI M., DUNNER S., GEORGES M. Molecular definition of an allelic series of mutations disrupting the myostatin function and causing double-muscling in cattle. *Mamm. Genome*, 1998, **9**, 210-213.
- GUNDERSON K.L., STEEMERS F.J., LEE G., MENDOZA L.G., CHEE M.S. A genome-wide scalable SNP genotyping assay using microarray technology. *Nat. Genet.*, 2005, **37**, 549-554.
- GUSTIN P., BAKIMA M., LEKEUX P., LOMBA F., VAN DE WOESTIJNE K.P. Partitioning of upper airway resistance in Friesian and double-muscled calves. *Respir. Physiol.*, 1987, **69**, 299-306.
- GUYOT H., SAEGERMAN C., LEBRETON P., SANDERSEN C., ROLLIN F. Epidemiology of trace elements deficiencies in Belgian beef and dairy cattle herds. *J. Trace Elem. Med. Biol.*, 2009, **23**, 116-123.
- H**AMILTON J.G. Artificial Insemination of Dairy Cattle. *Can. J. Comp. Med. Vet. Sci.*, 1940, **4**, 323-328.
- HANSET R., ANSAY M. La « White Heifer Disease » : nouvelle description et essai de classification rationnelle des ses différentes formes. *Ann. Med. Vet.*, 1961, 133-146.
- HANSET R. Croisement expérimentaux avec des génisses atteintes de la « White Heifer Disease ». *Ann. Med. Vet.*, 1969, **113**, 3-11.
- HANSET R. La « White Heifer Disease » dans la race bovine de Moyenne et Haute Belgique : un bilan de dix annnées. *Ann. Med. Vet.*, 1969, **113**, 12-21.
- HANSET R., LAPIERE C.M. Inheritance of dermatosparaxis in the calf. A genetic defect of connective tissues. *J. Hered.*, 1974, **65**, 356-358.
- HANSET R., MICHAUX C. Anomalies au niveau du maxillaire inférieur (brachygnathisme et déviation) chez le bovin culard. *Ann. Med. Vet.*, 1978, **122**, 649-655.
- HANSET R. Coat colour inheritance in the Belgian White and Blue cattle breed. *Genet. Sel. Evol.*, 1985, **17**, 443-458.
- HANSET R., MICHAUX C., STASSE A. 1987. Relationships between growth-rate, carcass composition, feed-intake, feed conversion ratio and income in 4 biological types of cattle. *Genet. Sel. Evol.*, 1987, **19**, 225-247.
- HANSET R., MICHAUX C., DETAL G., BOONEN F., LEROY P. Conformation et format dans la sélection du Blanc-Bleu Belge. Introduction d'un système de cotations linéaires. *Ann. Med. Vet.*, 1990, **134**, 197-204.

- HANSET R., MICHAUX C., BOONEN F. Linear classification in the Belgian Blue Cattle Breed: phenotypic and genetic parameters. International Committee for Animal Recording (ICAR), seminar, Beef performance recording and genetic evaluation. In: Milk and beef recording : State of the art, 1994. Proc. of the 29th biennial session of the International Committee for Animal Recording (ICAR). Ottawa, Canada, July 31 - August 5, 1994. EAAP Publication N° 75: 231-237.
- HANSET R. Une histoire longue de quatre millions d'années. In: Buren R., Compère G., Hanset R. François J (Eds), Blanc Bleu Belge, l'aventure d'une race bovine à la conquête du monde. Casterman, Tournai, 1996, 17-28.
- HANSET R. Au cœur de la sélection génétique du Blanc Bleu Belge. In: Buren R., Compère G., Hanset R. François J. (Eds), Blanc Bleu Belge, l'aventure d'une race bovine à la conquête du monde. Casterman, Tournai, 1996, 29-35.
- HANSET R. Emergence and selection of the Belgian Blue breed. 2004, URL : <http://www.hbbbbb.org/pdf/emergenceandselection.pdf>.
- HANSET R. Le gène culard: vers la fin des dystocies? *Point Vétérinaire*, 2007, **38**, 33-36.
- HANZEN C., LAURENT Y., WARD W.R. Comparison of reproductive performance in Belgian dairy and beef cattle. *Theriogenology*, 1994, **41**, 1099-1114.
- HANZEN C., LOURTIE O., ECTORS F. La césarienne dans l'espèce bovine. *Ann. Med. Vet.*, 1999, **143**, 65-88.
- HAYES B.J., BOWMAN P.J., CHAMBERLAIN A.J., GODDARD M.E. Invited review: Genomic selection in dairy cattle: progress and challenges. *J. Dairy Sci.*, 2009, **92**, 433-443.
- HAYES B.J., PRYCE J., CHAMBERLAIN A.J., BOWMAN P.J., GODDARD M.E. Genetic architecture of complex traits and accuracy of genomic prediction: coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. *PLoS Genet.*, 2010, **6**, e1001139.
- HAYES B.J., LEWIN H.A., GODDARD M.E. The future of livestock breeding: genomic selection for efficiency, reduced emissions intensity, and adaptation. *Trends Genet.*, 2012.
- HAZEL L.N. The Genetic Basis for Constructing Selection Indexes. *Genetics*, 1943, **28**, 476-490
- HAZEL L.N., DICKERSON G.E., FREEMAN A.E. The selection index-then, now, and for the future. *J. Dairy Sci.*, 1994, **77**, 3236-3251.
- HENDERSON C.R. Best linear unbiased estimation and prediction under a selection model. *Biometrics*, 1975, **31**, 423-447.
- HERB-BOOK BLANC-BLEU BELGE (HBBBBB). Index zootechniques & fonctionnels (1^{ère} et 2^{ème} visites) – Cotations linéaires & Index poids carcasses. Juillet 2012, 47 p., (<http://www.hbbbbb.org/pdf/evaluationgenetique2012.pdf>).
- HILL W.G., ROBERTSON A. Linkage disequilibrium in finite populations. *Theor. and Appl. Genet.*, 1968, **38**, 226-231.

HIRSCHHORN J.N., DALY M.J. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.*, 2005, **6**, 95-108.

HOFLACK G., VAN SOOM A., MAES D., DE KRUIF A., OPSOMER G., DUCHATEAU L. Breeding soundness and libido examination of Belgian Blue and Holstein Friesian artificial insemination bulls in Belgium and The Netherlands. *Theriogenology*, 2006, **66**, 207-216.

HOFLACK G., VAN DEN BROECK W., MAES D., VAN DAMME K., OPSOMER G., DUCHATEAU L., DE KRUIF A., RODRIGUEZ-MARTINEZ H., VAN SOOM A. Testicular dysfunction is responsible for low sperm quality in Belgian Blue bulls. *Theriogenology*, 2008, **69**, 323-332.

JEFFREYS A.J., WILSON V., THEIN S.L. Hypervariable 'minisatellite' regions in human DNA. *Nature*, 1985, **314**, 67-73.

JOERG H., FRIES H.R., MEIJERINK E., STRANZINGER G.F. Red coat color in Holstein cattle is associated with a deletion in the MSHR gene. *Mamm. Genome*, 1996, **7**, 317-318.

KARIM L., TAKEDA H., LIN L., DRUET T., ARIAS J.A., BAURAIN D., CAMBISANO N., DAVIS S.R., FARNIR F., GRISART B., HARRIS B.L., KEEHAN M.D., LITTLEJOHN M.D., SPELMAN R.J., GEORGES M., COPPIETERS W. Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. *Nat. Genet.*, 2011, **43**, 405-413.

KHATKAR M.S., NICHOLAS F.W., COLLINS A.R., ZENGER K.R., CAVANAGH J.A., BARRIS W., SCHNABEL R.D., TAYLOR J.F., RAADSMA H.W. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. *BMC Genomics*, 2008, **9**, 187.

KLUNGLAND H., VAGE D.I., GOMEZ-RAYA L., ADALSTEINSSON S., LIEN S. The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat color determination. *Mamm. Genome*, 1995, **6**, 636-639.

KOLKMAN I., AERTS S., VERVAECKE H., VICCA J., VANDELOOK J., DE KRUIF A., OPSOMER G., LIPS D. Assessment of differences in some indicators of pain in double muscled Belgian Blue cows following naturally calving vs caesarean section. *Reprod. Domest. Anim.*, 2010, **45**, 160-167.

KORNAK U., KASPER D., BOSL M.R., KAISER E., SCHWEIZER M., SCHULZ A., FRIEDRICH W., DELLING G., JENTSCH T.J. Loss of the ClC-7 chloride channel leads to osteopetrosis in mice and man. *Cell*, 2001, **104**, 205-215.

LEISLE L., LUDWIG C.F., WAGNER F.A., JENTSCH T.J., STAUBER T. ClC-7 is a slowly voltage-gated 2Cl(-)/1H(+)-exchanger and requires Ostm1 for transport activity. *EMBO J.*, 2011,

30, 2140-2152.

LEWONTIN R.C. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics*, 1964, **49**, 49-67

LI Y., WILLER C., SANNA S., ABECASIS G. Genotype imputation. *Annu. Rev. Genomics Hum. Genet.*, 2009, **10**, 387-406.

LIN J.Y., FISHER D.E. Melanocyte biology and skin pigmentation. *Nature*, 2007, **445**, 843-850.

LOFTUS R.T., MACHUGH D.E., BRADLEY D.G., SHARP P.M., CUNNINGHAM P. Evidence for two independent domestications of cattle. *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 2757-2761.

LOSSON B.J., LONNEUX J.F., LEKIMME M. The pathology of *Psoroptes ovis* infestation in cattle with a special emphasis on breed difference. *Vet. Parasitol.*, 1999, **83**, 219-229.

MACHUGH D.E., SHRIVER M.D., LOFTUS R.T., CUNNINGHAM P., BRADLEY D.G. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). *Genetics*, 1997, **146**, 1071-1086.

MACNEIL M.D., REECY J.M., GARRICK D.J. Cattle. In : Cockett N.E., Kole C. (Eds.), *Genome Mapping and Genomics in Domestic Animals*. Springer : Verlag Berlin Heidelberg, 2009, 1-17.

MANOLIO T.A., COLLINS F.S., COX N.J., GOLDSTEIN D.B., HINDORFF L.A., HUNTER D.J., MCCARTHY M.I., RAMOS E.M., CARDON L.R., CHAKRAVARTI A., CHO J.H., GUTTMACHER A.E., KONG A., KRUGLYAK L., MARDIS E., ROTIMI C.N., SLATKIN M., VALLE D., WHITTEMORE A.S., BOEHNKE M., CLARK A.G., EICHLER E.E., GIBSON G., HAINES J.L., MACKAY T.F., MCCARROLL S.A., VISSCHER P.M. Finding the missing heritability of complex diseases. *Nature*, 2009, **461**, 747-753.

MATUKUMALLI L.K., LAWLEY C.T., SCHNABEL R.D., TAYLOR J.F., ALLAN M.F., HEATON M.P., O'CONNELL J., MOORE S.S., SMITH T.P., SONSTEGARD T.S., VAN TASSELL C.P. Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One*, 2009, **4**, e5350.

MCCARTHY M.I., ABECASIS G.R., CARDON L.R., GOLDSTEIN D.B., LITTLE J., IOANNIDIS J.P., HIRSCHHORN J.N. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.*, 2008, **9**, 356-369.

MCDANELD T.G., KUEHN L.A., THOMAS M.G., SNELLING W.M., SONSTEGARD T.S., MATUKUMALLI L.K., SMITH T.P., POLLAK E.J., KEELE J.W. Y are you not pregnant: identification of Y chromosome segments in female cattle with decreased reproductive efficiency. *J. Anim. Sci.*, 2012, **90**, 2142-2151.

MCPHERRON A.C., LAWLER A.M., LEE S.J. Regulation of skeletal muscle mass in mice by a new TGF-beta superfamily member. *Nature*, 1997, **387**, 83-90.

METZKER M.L. Sequencing technologies - the next generation. *Nat. Rev. Genet.*, 2010, **11**, 31-46.

MEUWISSEN T.H., HAYES B.J., GODDARD M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 2001, **157**, 1819-1829.

NICHOLAS F.W. Genetic improvement through reproductive technology. *Anim. Reprod. Sci.* 1996, **42**, 205-214.

NICHOLAS F.W., SMITH C. Increased rates of genetic change in dairy cattle by embryo transfer and splitting. *Anim. Prod.*, 1983, **36**, 341-353.

NUSGENS B.V., VERELLEN-DUMOULIN C., HERMANNNS-LE T., DE PAEPE A., NUYTINCK L., PIERARD G.E., LAPIERE C.M. Evidence for a relationship between Ehlers-Danlos type VII C in humans and bovine dermatosparaxis. *Nat. Genet.*, 1992, **1**, 214-217.

OLSON T.A. Genetics of Colour Variation. In : Fries R., Ruvinsky A. (Eds.), *The Genetics of Cattle*, CAB International : Wallingford, 1999, 33-53.

OTT J., KAMATANI Y., LATHROP M. Family-based designs for genome-wide association studies. *Nat. Rev. Genet.*, 2011, **12**, 465-474.

OZSOLAK F., MILOS P.M. RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.*, 2011, **12**, 87-98.

PHILIPSSON J., BANOS G., ARNASON T. Present and future uses of selection index methodology in dairy cattle. *J. Dairy Sci.*, 1994, **77**, 3252-3261.

PRICE A.L., PATTERSON N.J., PLENGE R.M., WEINBLATT M.E., SHADICK N.A., REICH D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, 2006, **38**, 904-909.

PRICE A.L., ZAITLEN N.A., REICH D., PATTERSON N. New approaches to population stratification in genome-wide association studies. *Nat. Rev. Genet.*, 2010, **11**, 459-463.

PRITCHARD J.K., STEPHENS M., DONNELLY P. Inference of population structure using multilocus genotype data. *Genetics*, 2000, **155**, 945-959.

PRITCHARD J.K., STEPHENS M., ROSENBERG N.A., DONNELLY P. Association mapping in structured populations. *Am. J. Hum. Genet.*, 2000, **67**, 170-181.

RENDEL J.M. White heifer disease in a herd of dairy shorthorns. *J. Genet.*, 1952, **51**, 89-94.

REICH D.E., CARGILL M., BOLK S., IRELAND J., SABETI P.C., RICHTER D.J., LAVERY T., KOUYOUMJIAN R., FARHADIAN S.F., WARD R., LANDER E.S. Linkage disequilibrium in the human genome. *Nature*, 2001, **411**, 199-204

- S ANGER F., NICKLEN S., COULSON A.R. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, 1977, **74**, 5463-5467.
- SARRE C., DE BLEECKER K., DEPREZ P., LEVECKE B., CHARLIER J., VERCRUYSSSE J., CLAEREBOUW E.. Risk factors for *Psoroptes ovis* mange on Belgian Blue farms in Northern Belgium. *Vet. Parasitol.*, 2012, **190**, 216-221.
- SARTELET A. Etudes des principales maladies du système locomoteur chez le veau de race Blanc-bleu Belge (Mémoire de DEA). Université de Liège: Liège, 2007, 33 p.
- SARTELET A., DRUET T., MICHAUX C., FASQUELLE C., GERON S., TAMMA N., ZHANG Z., COPPIETERS W., GEORGES M., CHARLIER C. A splice site variant in the bovine RNF11 gene compromises growth and regulation of the inflammatory response. *PLoS Genet.*, 2012, **8**, e1002581.
- SARTELET A., KLINGBEIL P., FRANKLIN C.K., FASQUELLE C., GERON S., ISACKE C.M., GEORGES M., CHARLIER C. Allelic heterogeneity of Crooked Tail Syndrome: result of balancing selection? *Anim. Genet.*, 2012, **43**, 604-607.
- SARTELET A.*, STAUBER T.*, COPPIETERS W., LUDWIG C.F., FASQUELLE C., DRUET T., ZHANG Z., AHARIZ N., CAMBISANO N., JENTSCH T., CHARLIER C. A missense mutation accelerating the gating of the lysosomal Cl⁻/H⁺ exchanger ClC-7/Ostm1 causes osteopetrosis with gingival hamartomas in cattle. *Soumis pour publication*.
- SARTELET A., LI W., PAILHOUX E., TAMMA N., KARIM L., FASQUELLE C., DRUET T., COPPIETERS W., GEORGES M., CHARLIER C. A splice-acceptor site variant in the bovine PIGH gene causes glycosylphosphatidyl inositol deficiency and lethal arthrogryposis syndrome. *En préparation*.
- SCHUTZ E., SCHARFENSTEIN M., BRENIG B. Implication of complex vertebral malformation and bovine leukocyte adhesion deficiency DNA-based testing on disease frequency in the Holstein population. *J. Dairy Sci.*, 2008, **91**, 4854-4859.
- SEIDEL, G.E. Jr. Overview of sexing sperm. *Theriogenology*, 2007, **68**, 443-446.
- SHEMBADE N., PARVATIYAR K., HARHAJ N.S., HARHAJ E.W. The ubiquitin-editing enzyme A20 requires RNF11 to downregulate NF-kappaB signalling. *EMBO J.*, 2009, **28**, 513-522.
- SHUSTER D.E., KEHRLI M.E., JR. ACKERMANN, M.R. GILBERT R.O. Identification and prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle. *Proc. Natl. Acad. Sci. USA*, 1992, **89**, 9225-9229.
- SEITZ J.J., SCHMUTZ S.M., THUE T.D., BUCHANAN F.C. A missense mutation in the bovine MGF gene is associated with the roan phenotype in Belgian Blue and Shorthorn cattle. *Mamm. Genome*, 1999, **10**, 710-712.

SLATKIN M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genet.*, 2008, **9**, 477-485.

SMITH L.M., SANDERS J.Z., KAISER, R.J., HUGHES P., DODD C., CONNELL C.R., HEINER C., KENT S.B., HOOD L.E. Fluorescence detection in automated DNA sequence analysis. *Nature*, 1986, **321**, 674-679.

SPIELMAN R.S., MCGINNIS R.E., EWENS W.J. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.*, 1993, **52**, 506-516.

STANKIEWICZ P., LUPSKI J.R. Structural variation in the human genome and its role in disease. *Annual review of medicine*, 2010, **61**, 437-455.

STEEMERS F.J., CHANG W., LEE G., BARKER D.L., SHEN R., GUNDERSON K.L. Whole-genome genotyping with the single-base extension assay. *Nat. methods*, 2006, **3**, 31-33.

THOMPSON P.N., HEESTERBEEK J.A., VAN ARENDONK J.A. Changes in disease gene frequency over time with differential genotypic fitness and various control strategies. *J. Anim. Sci.*, 2006, **84**, 2629-2635.

TOUATI K., MULLER P., GANGL M., GRULKE S., PETERS F., SERTEYN D. La parésie spastique du quadriceps fémoral: Une nouvelle entité clinique chez le veau de race Blanc Bleu Belge. *Ann. Med. Vet.*, 2003, **147**, 261 – 265.

TOUSSAINT J.F., SAILLEAU C., MAST J., HOUDART P., CZAPLICKI G., DEMEESTERE L., VANDENBUSSCHE F., VAN DESSEL W., GORIS N., BREARD E., BOUNAADJA L., ETIENNE T., ZIENTARA S., DE CLERCQ K. Bluetongue in Belgium, 2006. *Emerg. Infect. Dis.*, 2007, **13**, 614-616.

UYSTEPRUYST C., COGHE J., DORTS T., HARMEGNIES N., DELSEMME M.H., ART T., LEKEUX P. Optimal timing of elective caesarean section in Belgian White and Blue breed of cattle: the calf's point of view. *Vet. J.*, 2002, **163**, 267-282.

UYTTERHAEGEN L., CLAEYS E., DEMEYER D., LIPPENS M., FIEMS L.O., BOUCQUE C.Y., VAN DE VOORDE G., BASTIAENS A. Effects of double-muscling on carcass quality, beef tenderness and myofibrillar protein degradation in Belgian Blue White bulls. *Meat Science*, 1994, **38**, 255-267.

VAN LOO H., GOVAERE J., CHIERS K., HOOGEWIJS M., OPSOMER G., DE KRUIF A. Hydrops uteri in a BWB heifer combined with a vascular hamartoma of the mandible of the calf. 2010, *Vlaams Diergen. Tijd.*, **79**, 139-142.

- VANDENHEEDE M., NICKS B., DESIRON A., CANART B. Mother-young relationships in Belgian Blue cattle after a Caesarean section: characterisation and effects of parity. 2001, *Appl. Anim. Behav. Sci.*, **72**, 281-292.
- VERCRUYSSSE J., DEPREZ P., EVERAERT D., BASSISSI F., ALVINERIE M. Breed differences in the pharmacokinetics of ivermectin administered subcutaneously to Holstein and Belgian Blue calves. 2008, *Vet. Parasitol.*, **152**, 136-140
- VERSCHOOTEN F., DE MOOR A., DESMET P., WATTE R., GUNST O. Surgical treatment of congenital arthrogryposis of the carpal joint, associated with contraction of the flexor tendons in calves. 1969, *Vet. Rec.* **85**, 140-142.
- VERTENTEN G. Abord de la parésie spastique des veaux. *Le Point Vétérinaire*, 2009, **294**, 39-42.
- VISHWANATH R. Artificial insemination: the state of the art. *Theriogenology*, 2003, **59**, 571-584.
- VLAMINCK L., DE MOOR A., MARTENS A., STEENHAUT M., GASTHUYS F., DESMET P., VAN BRANTEGHEM L. Partial tibial neurectomy in 113 Belgian blue calves with spastic paresis. 2000, *Vet. Rec.*, **147**, 16-19.

WEBER J.L., MAY P.E. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.*, 1989, **44**, 388-396.

- WEIGEL K.A., FREEMAN A.E., KEHRLI M.E., JR., STEAR M.J., KELLEY D.H. Association of class I bovine lymphocyte antigen complex alleles with health and production traits in dairy cattle. *J. Dairy Sci.*, 1990, **73**, 2538-2546.
- WHEELER T.L., CUNDIFF L.V., SHACKELFORD S.D., KOOHARAIE M. Characterization of biological types of cattle (Cycle VII): Carcass, yield, and longissimus palatability traits. *J. Anim. Sci.*, 2005, **83**, 196-207.
- WILLETT E.L., BLACK W.G., CASIDA L.E., STONE W.H. BUCKNER P.J. Successful transplantation of a fertilized bovine ovum. *Science*, 1951, **113**, 247
- WILMUT I., ROWSON L.E. The successful low-temperature preservation of mouse and cow embryos. *J. Reprod. Fertil.*, 1973, **33**, 352-353.
- WINDIG J. J., MEUWISSEN T.H.E. Rapid haplotypes reconstruction in pedigrees with dense marker maps. *J. Anim. Breed. Genet.*, 2004, **121**, 26-39.
- WINDSOR P., AGERHOLM J. Inherited diseases of Australian Holstein-Friesian cattle. *Aust. Vet. J.*, 2009, **87**, 193-199.

YANG J., BENYAMIN B., MCEVOY B.P., GORDON S., HENDERS A.K., NYHOLT D.R., MADDEN P.A., HEATH A.C., MARTIN N.G., MONTGOMERY G.W., GODDARD M.E., VISSCHER P.M.

Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.*, 2010, **42**, 565-569.

YU J., PRESSOIR G., BRIGGS W.H., VROH BI I., YAMASAKI M., DOEBLEY J.F., MCMULLEN M.D., GAUT B.S., NIELSEN D.M., HOLLAND J.B., KRESOVICH S., BUCKLER E.S. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.*, 2006, **38**, 203-208.

ZHANG Z., GUILLAUME F., SARTELET A., CHARLIER C., GEORGES M., FARNIR F., DRUET T. Ancestral haplotype-based association mapping with generalized linear mixed models accounting for stratification. *Bioinformatics*, 2012, **28**, 2467-2473.