

LES POPULATIONS GÉNÉTIQUES DE RÉFÉRENCE DE SOURIS DE LABORATOIRE : UNE NOUVELLE GÉNÉRATION DE MODÈLES ANIMAUX POUR LES MALADIES HUMAINES

GENETIC REFERENCE POPULATIONS OF LABORATORY MICE: A NEW GENERATION OF ANIMAL MODELS FOR HUMAN DISEASES

Par Xavier MONTAGUTELLI⁽¹⁾
Communication présentée le 4 Juin 2015
Manuscrit accepté le 19 Novembre 2015

RÉSUMÉ

La souris est l'animal de laboratoire le plus utilisé pour étudier les maladies humaines et développer de nouvelles approches thérapeutiques. Toutefois, les résultats obtenus chez les souris ne sont pas toujours transposables à l'homme. L'une des raisons est la faible diversité génétique des lignées de souris de laboratoire, qui est bien moindre que celle des populations humaines. *A fortiori*, une étude portant sur une seule lignée de souris ne peut pas reproduire la diversité des réponses observées chez l'homme. Les généticiens de la souris ont développé de nouvelles collections de lignées qui comportent une variabilité génétique au moins égale à celle des populations humaines. Elles permettent non seulement d'obtenir des phénotypes nouveaux et très variés mais aussi de développer des approches de biologie des systèmes à une échelle inégalée.

Mots clés : souris, génétique, modèles animaux, maladies humaines, Collaborative Cross.

ABSTRACT

The mouse is the most used laboratory animal to study human diseases and to develop new therapeutic approaches. However, the results obtained in mice are not always translatable to humans. One reason is the reduced genetic diversity of mouse laboratory strains which is much smaller than that of human populations. Moreover, a study carried on a single mouse strain cannot reproduce the diversity of responses observed in humans. Mouse geneticists have developed novel collections of strains which encompass genetic variations at least similar to that of human populations. They provide not only new and highly varied phenotypes but also the possibility to develop systems biology approaches at an unprecedented scale.

Key words: mouse, genetics, animal models, human diseases, Collaborative Cross.

INTRODUCTION

La recherche biomédicale a pour objectifs d'acquiescer des connaissances sur la physiologie et la pathologie de l'homme et des animaux, et de mettre au point des procédés permettant d'améliorer les conditions de vie de ces êtres vivants, de prévenir et de guérir les maladies. L'atteinte de ces objectifs est rendue difficile en raison de la variabilité des propriétés biologiques et des états physiologiques entre les individus d'une même espèce. La recherche chez l'homme se heurte également à l'impossibilité de réaliser certaines investigations pour des rai-

sons pratiques ou éthiques. De ce fait, les chercheurs ont, depuis plusieurs siècles, mis au point des modèles expérimentaux utilisant des animaux faciles à étudier dans l'environnement du laboratoire. Élevés sous forme de populations génétiquement homogènes et dans des conditions très standardisées, ces animaux permettent de réduire la complexité des situations naturelles et l'hétérogénéité des facteurs de l'environnement responsables de variations interindividuelles, donc d'accroître considérablement la puissance des expériences.

(1) Directeur de Recherche, Unité de Génétique fonctionnelle de la Souris, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris cedex 15 – xavier.montagutelli@pasteur.fr

Parmi les animaux de laboratoire, la souris occupe une place prépondérante, même si elle n'est pas toujours le meilleur modèle dans tous les domaines de la biologie. Elle compte en particulier parmi ses avantages une adaptation facile à l'élevage en animalerie de laboratoire, une petite taille qui permet de travailler sur de grands échantillons d'animaux pour un coût raisonnable, un temps de génération très court qui permet d'obtenir quatre générations en une seule année, la possibilité de créer des lignées consanguines (dans lesquelles tous les individus sont génétiquement identiques et homozygotes pour tous les autosomes), de grandes homologues physiologiques avec l'homme, une grande proximité de son génome avec celui de l'homme et la possibilité de créer à façon pratiquement n'importe quelle modification génétique (ceci n'est désormais plus son apanage). Ainsi, la souris a été utilisée depuis plus d'un siècle, et l'est de plus en plus, à la fois en recherche fondamentale mais également pour tester de nouveaux médicaments ou modalités thérapeutiques.

Il faut toutefois pondérer des contributions considérables par des limites notables qui concernent d'ailleurs tous les modèles animaux. En effet, ces modèles ne réussissent pas toujours à reproduire fidèlement les symptômes observés dans la pathologie humaine, en particulier la diversité des situations cliniques entre les patients. Par ailleurs, pour une proportion importante des médicaments en développement, les effets prometteurs observés en phase préclinique sur des modèles animaux ne sont pas confirmés lors d'études cliniques chez l'homme (Kola & Landis, 2004).

Il n'est pas l'objet de cette communication de discuter des raisons qui peuvent être invoquées pour expliquer ces limites et ces échecs. Mentionnons toutefois le choix de l'espèce animale modèle utilisée, qui peut ne pas être pertinent (par exemple à cause de différences physiologiques entre cette espèce modèle et l'homme), ou les modalités d'induction de la pathologie qui peuvent être très éloignées des mécanismes physiopathologiques à l'œuvre chez l'homme. Nous nous intéresserons ici à l'influence de la variabilité génétique ; nous montrerons comment de nouvelles populations de souris permettent de mieux modéliser la diversité génétique des populations humaines. Elles offrent une perspective expérimentale concrète en vue de mieux comprendre les réseaux d'interactions qui structurent le fonctionnement d'un organisme entier et d'envisager le développement d'une médecine personnalisée.

LES MODÈLES ANIMAUX FACE À LA DIVERSITÉ HUMAINE

Les propriétés biologiques d'un individu sont fortement influencées par sa composition génétique. Notre génome modèle nos caractéristiques morphologiques et physiologiques,

notre vulnérabilité aux maladies (quelles qu'en soient leurs causes), notre réponse au vieillissement, aux facteurs de l'environnement, aux médicaments, etc. Ainsi, l'hétérogénéité génétique qui existe au sein d'une population humaine, par exemple de patients atteints de schizophrénie ou d'hypertension artérielle, impose-t-elle de travailler sur de très grandes cohortes afin d'avoir la puissance nécessaire pour détecter des différences liées à un facteur d'intérêt, par exemple l'efficacité d'une drogue ou l'influence d'un facteur de risque supposé.

Cette difficulté est l'une des justifications des modèles animaux, qui reposent sur le recours à des populations d'animaux génétiquement homogènes. La possibilité de produire et de propager des lignées consanguines chez la souris est l'une des raisons de son succès comme animal de laboratoire. Au fil des décennies, les généticiens ont produit des dizaines (voire des centaines) de lignées consanguines qui constituent un riche patrimoine (Montagutelli & de Vienne, 2008; Guénet *et al.* 2015) dont la stabilité est assurée à la fois par la composition génétique homogène de la population et par la possibilité de congeler des embryons². Certaines d'entre elles possèdent des propriétés biologiques propres qui en font des modèles pour l'étude de pathologies humaines, par exemple la lignée NOD (pour *Non-Obese Diabetes*) qui développe spontanément un diabète de type I. Une cinquantaine de lignées font l'objet d'un programme de phénotypage systématique pour de nombreux paramètres biologiques spontanés ou induits, dont les données sont disponibles dans la base *Mouse Phenome Database* du Jackson Laboratory (Maine, USA; <http://phenome.jax.org/>).

Cette multitude de lignées ne doit pas cacher deux réalités. Premièrement, la caractérisation moléculaire des lignées de souris, qui a commencé il y a 30 ans et a culminé avec le séquençage complet d'une partie d'entre elles (Keane *et al.* 2011), a mis en évidence une faible diversité génétique entre ces lignées³ qui s'explique par le petit nombre d'individus à l'origine des souris de laboratoire et par des liens de parenté étroits entre les fondateurs des différentes lignées (Guénet *et al.* 2015).

Deuxièmement, seul un très petit nombre de ces lignées est activement utilisé en recherche. En effet, afin que différents laboratoires travaillant sur le même sujet puissent comparer leurs résultats et les enrichir mutuellement, il est important que les animaux sur lesquels ils travaillent présentent la même composition génétique. Par ailleurs, tous les domaines de la recherche qui utilisent la souris ont de plus en plus recours à des souris génétiquement modifiées. Pour chaque étude, il faut pouvoir comparer les propriétés de la lignée transgénique ou déficiente pour un gène avec celles d'une lignée qui ne porte pas cette modification. Pour que cette comparaison soit valide et porte sur

(2) Pour assurer la stabilité génétique d'une lignée pendant une longue période, un grand nombre d'embryons est congelé à un instant donné. Toutes les 10 générations d'élevage, les reproducteurs sont remplacés par des individus issus de la décongélation de quelques embryons. On supprime ainsi la faible dérive génétique due aux mutations qui aurait pu apparaître au cours des 10 générations.

(3) La diversité génétique d'une population peut être évaluée par la fréquence des variations nucléotidiques au sein des individus. Elle est d'environ un variant tous les 300 nucléotides entre les lignées de souris de laboratoire (Frazer *et al.* 2007). Il est difficile de donner une estimation similaire pour les populations humaines en raison des nombreux types de variants observés.

la seule modification génétique analysée, il faut que tout le reste du génome, appelé fonds génétique, soit identique (de même, bien sûr, que les paramètres d'environnement physico-chimiques et microbiologiques). Il en va de même lorsque l'on compare les effets de plusieurs modifications génétiques simples ou multiples. Le choix de ce fonds génétique n'est pas sans importance car la littérature scientifique compte d'innombrables exemples de modifications génétiques dont les effets dépendent du fonds génétique dans lequel elles sont étudiées (Montagutelli, 2000).

Ces différents éléments permettent de comprendre pourquoi les chercheurs travaillent sur des fonds génétiques peu variés. De fait, la plupart des modifications génétiques sont soit produites, soit transférées sur le fonds génétique C57BL/6J ou son cousin proche C57BL/6N⁴, qui sont devenus les références. Avoir plusieurs milliers de lignées mutées dans un gène différent sur le même fonds génétique est un atout considérable pour explorer finement le fonctionnement des gènes et leurs interactions. D'autres fonds génétiques sont préférés dans certains domaines (tel que BALB/c pour certains aspects de l'immunologie) mais n'offrent pas la même diversité de modifications génétiques.

Pour autant qu'elle soit justifiée, cette uniformisation a des conséquences qui sont souvent perdues de vue par les chercheurs qui utilisent ces modèles. La comparaison des propriétés biologiques de souris porteuses ou non d'une mutation dans un gène permet d'obtenir des informations précieuses sur la fonction de ce gène. Toutefois, il faut garder à l'esprit que cette observation vaut pour le fonds génétique des souris étudiées et ne peut être généralisée sans risque à l'espèce souris dans son ensemble (Simon *et al.* 2013). De même, les effets d'un traitement mis en évidence chez les souris d'une lignée pourraient-ils être différents (moindres ou plus marqués, associés ou non à des effets indésirables) s'ils étaient étudiés également dans un autre fonds génétique (Leclercq & Kaminski, 2015).

Doit-on alors déconseiller l'utilisation des lignées consanguines en recherche biomédicale, puisque les résultats qu'elles permettent d'obtenir ne procurent de certitude que dans le cas d'une lignée mais peuvent être remis en cause dans un autre fonds génétique ? Comment ces lignées si homogènes génétiquement peuvent-elles répondre à l'objectif des modèles animaux qui est de produire des observations transposables aux populations humaines qui sont, elles, si diverses ?

Certains chercheurs pensent contourner cette difficulté en analysant des populations de souris non consanguines, ces souris dites «Swiss» en raison de l'origine géographique des individus fondateurs. Si ces populations présentent une certaine variation génétique entre les individus, celle-ci est bien moindre que celle des populations humaines et même, moindre que celles qui existent entre différentes lignées consanguines (Chia *et al.* 2005). De plus, les différences génétiques entre individus sont responsables d'une plus grande variabilité des résultats.

La meilleure solution, qui est malheureusement encore peu mise en œuvre, consiste à tester les résultats obtenus avec une lignée consanguine particulière dans plusieurs autres lignées consanguines qui possèdent chacune un fonds génétique propre. Non seulement cette solution met à profit tous les avantages des lignées consanguines (population génétiquement déterminée et stable, résultats reproductibles, expériences puissantes), mais elle permet de définir précisément la variation génétique étudiée par le choix des lignées et de reproduire les expériences dans des conditions strictement identiques (Chia *et al.* 2005). Il est ainsi possible d'évaluer l'influence du fonds génétique sur le phénomène étudié. En explorant des fonds génétiques très différents, on obtient alors une gamme de réponses qui peut refléter la diversité des situations rencontrées chez l'homme (Nadeau, 2001).

À titre d'exemple, nous avons étudié l'effet du fonds génétique sur une mutation que nous avons décrite et qui touche le gène de la ferrocélatase, la dernière enzyme de la voie de la biosynthèse de l'hème (Tutois *et al.* 1991). Un déficit partiel d'activité de cette enzyme est responsable, chez l'homme, de la protoporphyrie érythropoïétique qui se traduit par une photosensibilisation et, dans les formes sévère, une atteinte hépatique pouvant nécessiter une greffe de foie (Gross *et al.* 1998). Dans son fonds génétique d'origine, proche de BALB/c, cette mutation se traduit chez la souris par une anémie, un retard de développement, un fort ictère, une hépato-splénomégalie et une photosensibilisation (Tutois *et al.* 1991). Nous avons transféré cette mutation dans deux autres fonds génétiques en créant deux lignées congéniques. Sur le fonds C57BL/6J, les souris présentent une anémie modérée mais surtout une forte accumulation de pigments dans le foie, sans ictère. Sur le fonds SJL/J, elles présentent une très forte accumulation de protoporphyrine (le substrat de la ferrocélatase) dans les érythrocytes, sans autre symptôme (Abitbol *et al.* 2005). Ces différences considérables de phénotype reflètent la grande variabilité clinique et biologique des patients humains. Aucune des trois lignées de souris ne peut par elle-même prétendre être le modèle de la maladie humaine, mais seulement ressembler à une partie des patients.

Cet exemple illustre bien comment, en travaillant sur plusieurs lignées consanguines de souris, il est possible à la fois de produire des résultats fiables mais également d'explorer l'influence de la diversité génétique sur les propriétés biologiques, qui est un grand défi des études de cohortes de patients.

PRODUIRE UNE NOUVELLE DIVERSITÉ GÉNÉTIQUE

La diversité génétique des lignées consanguines de souris étant limitée et bien moindre que celle des populations humaines, il est nécessaire de développer d'autres lignées possédant de nouveaux fonds génétiques. Deux approches complémentaires ont été utilisées.

(4) C57BL/6J et C57BL/6N sont deux lignées consanguines très proches, séparées depuis les années 1950 et dont les différences génétiques se résument à quelques dizaines de paires de bases, dont certaines sont toutefois responsables de différences phénotypiques.

Parmi les outils développés par les généticiens de la souris figurent des collections de lignées dites « recombinantes consanguines » (RC). Il s'agit bien de lignées consanguines dont elles possèdent toutes les propriétés. Mais elles ont de plus la particularité d'avoir été produites à partir d'individus F2 (hybrides de deuxième génération) obtenus en croisant deux lignées consanguines. Ainsi la première collection de ce type appelée BXD comportait-elle 26 lignées consanguines issues du croisement entre les lignées C57BL/6J et DBA/2J (Bailey, 1971; Taylor, 1989). Elle a été amplifiée et en compte désormais plus d'une centaine (Peirce *et al.* 2004). Chacune des lignées RC a hérité la moitié de son génome de C57BL/6J et l'autre moitié de DBA/2J, mais la composition de ces deux moitiés varie d'une lignée à l'autre (**figure 1**). Chaque lignée porte donc un nouveau fonds génétique qui ne comporte pas de nouveaux allèles mais de nouvelles associations entre les allèles des deux lignées fondatrices.

Cette combinatoire inédite est responsable de nouvelles propriétés biologiques. Lorsque les deux lignées fondatrices présentent un phénotype très contrasté pour un caractère donné, les réponses des lignées RC couvrent généralement une large plage entre les deux extrêmes. Dans le cas où les deux lignées fondatrices ont des phénotypes proches, il est fréquent que ceux des lignées RC présentent une diversité beaucoup plus grande que celle des deux lignées d'origine. Citons à titre d'exemple la réponse à une injection de méthamphétamine (Grisel *et al.* 1997). Alors que la température centrale des lignées C57BL/6J et DBA/2J varie très peu en réponse à une injection intrapéritonéale de 8 mg/kg de méthamphétamine, celle des lignées RC présente des variations allant de -2 à +2°C. Ainsi, en produisant de nouvelles lignées hybrides entre deux lignées consanguines, on obtient de nouveaux phénotypes qui permettent de mieux couvrir la diversité observée dans des populations naturelles.

Les lignées RC ont de surcroît des propriétés génétiques qui permettent d'identifier les gènes qui sont responsables des différences de leurs phénotypes. Grâce à un génotypage de très haute densité (plusieurs centaines de milliers de marqueurs génétiques), leur génome est connu avec une très grande précision. Il est alors possible de rechercher les régions chromosomiques présentant une association entre le génotype des lignées et leur phénotype (Andreux *et al.* 2012).

Un autre atout considérable des lignées RC est la possibilité d'analyser les mêmes lignées dans deux expériences pour des caractères différents, en particulier lorsque ces derniers ne peuvent pas être mesurés chez les mêmes individus. Il devient alors possible de rechercher des corrélations entre ces caractères et d'en tirer des hypothèses quant aux mécanismes sous-jacents et aux possibles relations causales entre ces caractères (Ferris *et al.* 2013).

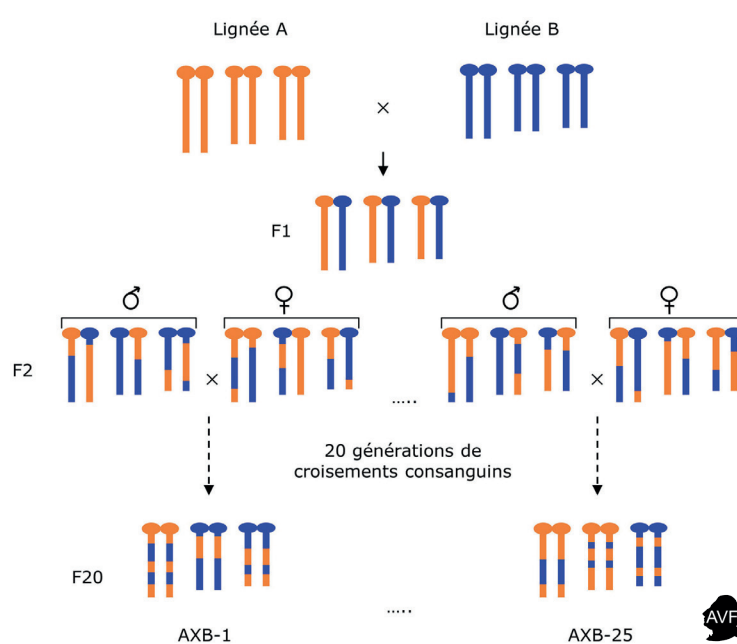


Figure 1 : Obtention et composition génétique des lignées Recombinantes Consanguines (RC). Les lignées RC sont produites en croisant deux lignées consanguines A et B. Des couples d'individus F2 sont formés au hasard et sont le point de départ de nouvelles lignées consanguines obtenues par 20 croisements entre frères et sœurs. Le génome de chaque lignée RC provient pour moitié des lignées A et B, les deux moitiés étant différentes d'une lignée à une autre. Chaque lignée possède donc un assortiment unique des allèles d'origine. Les lignées RC offrent donc une nouvelle diversité génétique.

Si cette première approche permet de créer une infinité de nouvelles combinaisons alléliques, elle ne peut que réassortir les allèles qui existent dans les lignées de souris de laboratoire. Or, ceux-ci sont en nombre limité et certaines régions chromosomiques sont très pauvres en variants alléliques (Yang *et al.* 2007). Les généticiens de la souris ont tiré parti des travaux de quelques évolutionnistes, au premier rang desquels François Bonhomme et son équipe à Montpellier étudiant la phylogénie et la spéciation au sein du genre *Mus*, ils avaient capturé et tenté d'acclimater au laboratoire des individus piégés dans de nombreuses régions de l'Ancien Monde et appartenant à différentes sous-espèces (Bonhomme & Guénet, 1989). Il a été possible, dans quelques cas, de produire de nouvelles lignées consanguines à partir de ces individus sauvages. Alors que les souris de laboratoire ont un génome provenant pour environ 90% de la sous-espèce *Mus musculus domesticus* (Yang *et al.* 2007), de nouvelles lignées dites « de souris sauvages » ont été établies à partir d'individus représentant de façon pure les sous-espèces *M. m. domesticus*, *M. m. musculus*, *M. m. castaneus* et l'espèce *M. spretus* (une espèce génétiquement aussi distante des souris de laboratoire que l'homme l'est du chimpanzé; Dejager *et al.* 2009).

La caractérisation génétique de ces lignées, par de nombreux marqueurs puis par un séquençage complet du génome, a mis en évidence de très nombreux variants génétiques inconnus chez les lignées de laboratoire classiques qui ont été très utiles dès les années 1990 pour la construction des cartes génétiques de la souris (Guénet, 1986; Keane *et al.* 2011). Ces variants

sont non seulement beaucoup plus nombreux qu'entre lignées de laboratoire mais sont également répartis de façon homogène sur l'ensemble des chromosomes.

Ce polymorphisme génétique se traduit par des variations phénotypiques importantes par rapport aux lignées de laboratoire et parfois par l'observation de propriétés jusqu'alors jamais observées. Citons trois exemples tirés de l'étude de la sensibilité aux maladies infectieuses menées dans notre laboratoire à l'Institut Pasteur. Presque toutes les lignées de laboratoire sont très sensibles au virus West Nile (virus de la fièvre du Nil occidental), alors que les lignées de souris sauvages sont naturellement résistantes. Cette différence s'explique par l'existence d'une mutation de perte de fonction dans le gène *Oas1b* portée par presque toutes les lignées de laboratoire, qui l'ont certainement héritée d'un ancêtre commun (Mashimo *et al.* 2002). Par ailleurs, toutes les lignées de souris étudiées jusqu'à récemment se sont montrées très sensibles à des souches virulentes du bacille de la peste. Nous avons décrit pour la première fois une lignée de l'espèce *M. spretus* qui possède un niveau de résistance élevé (Blanchet *et al.* 2011). Enfin, l'étude de plusieurs lignées de souris sauvages nous a permis d'en identifier une qui présente une très grande sensibilité au virus de la fièvre de la Vallée du Rift par comparaison avec les lignées de laboratoire (do Valle *et al.* 2010).

LE COLLABORATIVE CROSS

Chacune des deux approches précédentes a permis d'accroître la diversité génétique des populations de souris dont les chercheurs disposent pour étudier les maladies humaines. En 2001, un groupe de généticiens de la souris a décidé de développer un nouvel outil combinant les avantages de ces deux approches (Threadgill *et al.* 2002; Churchill *et al.* 2004; Threadgill *et al.* 2011; Panthier & Montagutelli, 2012). Ce projet, appelé « Collaborative Cross » (CC) en raison de la collaboration de trois centres de recherche, aux USA (Chesler *et al.* 2008), en Australie (Morahan *et al.* 2008) et en Israël (Iraqi *et al.* 2008), avait pour objectif de créer une collection de plusieurs centaines de lignées RC dérivées non pas de deux mais de huit lignées fondatrices, parmi lesquelles ont été incluses cinq lignées de laboratoire et trois lignées de souris sauvages des sous-espèces *M. m. musculus*, *M. m. domesticus* et *M. m. castaneus*, établies à partir d'individus sauvages capturés respectivement en République Tchèque, dans le Maryland (USA) et en Thaïlande. La **figure 2** illustre le principe de ces croisements. Le polymorphisme génétique en ségrégation était sans précédent, laissant espérer la découverte de phénotypes complè-

tement nouveaux. Le nombre de lignées projeté devait assurer une grande puissance pour localiser des gènes responsables des variations de phénotype (Aylor *et al.* 2011). Une organisation très soignée des croisements a permis d'assurer une représentation équilibrée des allèles des huit lignées fondatrices, tant pour le génome nucléaire que mitochondrial. Près de 15 ans plus tard, une cinquantaine de lignées du CC sont disponibles (en particulier auprès du partenaire américain : <http://csbio.unc.edu/CCstatus/>) et ont déjà été testées par des laboratoires travaillant sur des thématiques très variées telles que le métabolisme, le comportement ou les maladies infectieuses. La caractérisation génétique très fine de ces lignées (disponible sur le site ci-dessus) a montré que les huit lignées fondatrices avaient contribué au génome des lignées RC de façon équilibrée et que la variabilité génétique présente dans le CC, mesurée par la densité des SNP (*single nucleotide polymorphisms*), est au moins égale à celle qui est mesurée dans les populations humaines (Aylor *et al.* 2011; Collaborative Cross Consortium, 2012).

La production et l'entretien des lignées du CC n'ont pas été une tâche facile et demeurent toujours délicats, avec une prolificité très variable d'une lignée à l'autre. Un grand nombre de lignées n'ont pas supporté la mise à l'état consanguin, certainement

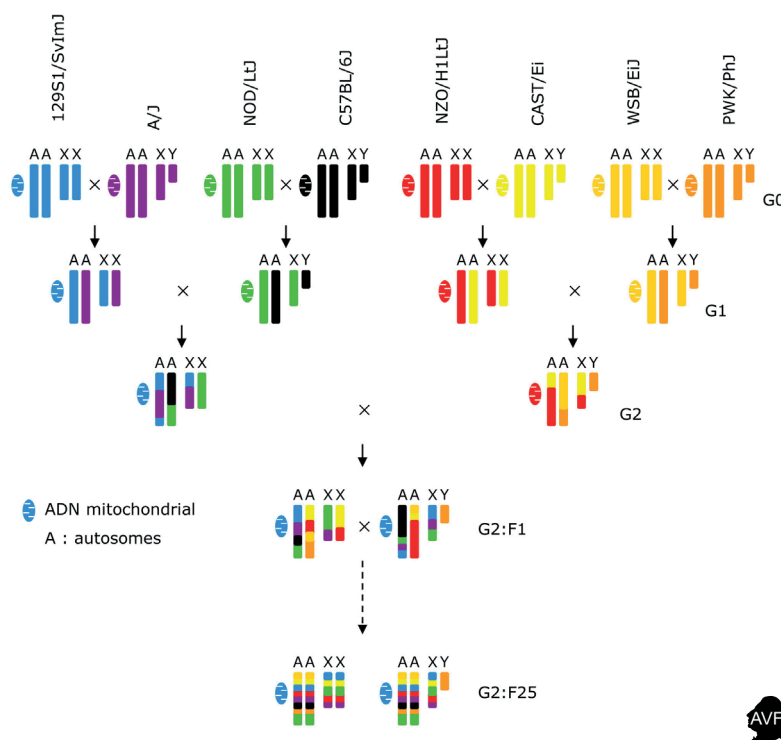


Figure 2 : Obtention et composition génétique des lignées du Collaborative Cross (CC).

Le schéma de croisement conduisant aux lignées du CC est très proche de celui des lignées RC, à cela près qu'il implique huit lignées fondatrices. Celles-ci produisent d'abord des individus appelés G1 (qui ont chacun deux parents), qui sont croisés entre eux pour produire des G2 (à quatre grands-parents). Ceux-ci sont à leur tour accouplés pour produire la génération appelée G2F1, la première qui possède une contribution génétique des huit lignées d'origine et qui sert de point de départ à une nouvelle lignée consanguine par des accouplements entre frères et sœurs. Ce schéma respecte la convention selon laquelle la femelle est indiquée à gauche du symbole de croisement. De ce fait, l'ADN mitochondrial d'une lignée du CC, transmis à chaque génération par les femelles, provient de la lignée à gauche du schéma, alors que le chromosome Y, transmis par les mâles, est hérité de la lignée de droite. Pour cette raison, il est important que la place des lignées dans le schéma change d'une lignée à une autre.

car le brassage génétique entre les sous-espèces parentales a souvent conduit à des combinaisons entre différents allèles qui sont incompatibles avec un développement normal ou la fertilité dans les deux sexes. Il est donc heureux que ce projet ait été lancé à très grande échelle afin que le nombre de lignées survivantes suffise aux besoins des projets.

Le polymorphisme génétique du CC se traduit par une très grande variabilité de phénotype qui dépasse généralement les différences observées au sein des huit lignées fondatrices et couvre une amplitude jamais décrite chez la souris. Ainsi, dans un des premiers articles rapportant des données expérimentales sur le CC, Durrant *et al.* (2011) ont décrit la réponse de 66 lignées à l'administration intraveineuse d'une souche d'*Aspergillus fumigatus*. Alors que les souris de quelques lignées sont mortes en moins de cinq jours, d'autres ont survécu plus longtemps, jusqu'à 30 jours, tandis que 17 lignées se sont montrées complètement résistantes. Un autre exemple spectaculaire qui a eu un fort retentissement a été publié en 2014 sur le virus Ebola. La souris est considérée par la plupart des auteurs comme un modèle médiocre de la maladie humaine causée par ce virus car les lignées de laboratoire couramment utilisées ne développent pas de coagulopathie qui est une des caractéristiques des formes humaines les plus sévères (Nakayama & Saijo, 2013). Dans leur article, Rasmussen *et al.* (2014) ont infecté une cinquantaine de types de F1 entre différentes lignées du CC et ont décrit des variations considérables de sensibilité. Pour une partie des combinaisons génétiques testées, les souris

ont présenté une résistance complète sans développer aucun symptôme (**figure 3**). D'autres ont montré une mortalité partielle. Pour d'autres encore, toutes les souris sont mortes sans symptômes spécifiques. Dans un quatrième groupe, les souris sont mortes et ont développé une hépatite sévère et dans le dernier, elles ont développé de surcroît une coagulopathie. Les auteurs concluent que les conséquences d'une infection par le virus Ebola dépendent très fortement du fonds génétique des animaux. Cet exemple montre que l'on ne peut pas parler « du » modèle souris d'infection par le virus Ebola mais que la souris offre, selon son fonds génétique, la possibilité de travailler de façon reproductible sur les différentes formes cliniques de la maladie. Qui plus est, en recherchant des associations entre la sévérité du phénotype et le génotype des lignées, il sera possible d'identifier des régions chromosomiques, voire des gènes, responsables de ces différences, ce qui est tout à fait hors d'atteinte chez l'homme.

Le polymorphisme génétique du CC est tel que quelques chercheurs ont commencé à l'utiliser pour rechercher des différences de réponse à un médicament, en termes d'efficacité ou de survenue d'effets indésirables. Les premiers résultats (non encore publiés) sont très encourageants et montrent que le CC contribuera certainement au développement de la médecine personnalisée qui prendra en compte le patrimoine génétique du patient pour décider de la meilleure stratégie thérapeutique.

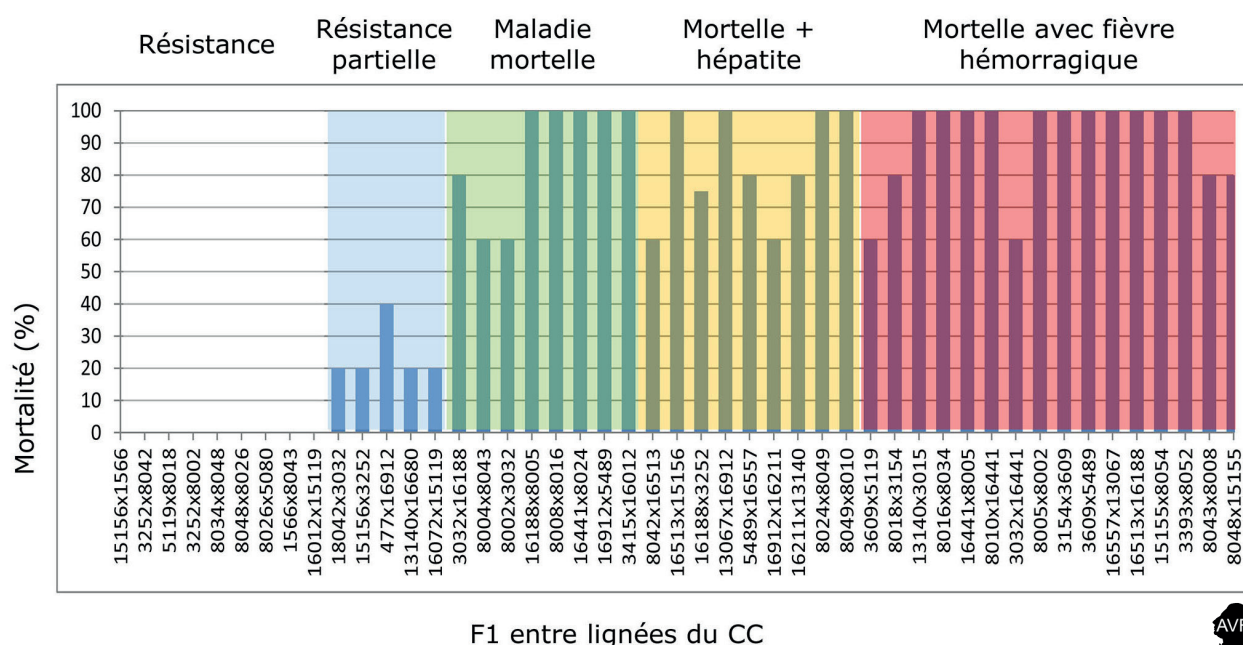


Figure 3 : Diversité de sensibilité au virus Ebola dans les lignées du Collaborative Cross.

Dans leur article publié dans Science en novembre 2014, Rasmussen *et al.* ont rapporté la sensibilité de 47 combinaisons de F1 entre différentes lignées du CC à l'injection d'une souche Ebola adaptée à la souris (Rasmussen *et al.* 2014). Ce graphe représente le taux de mortalité observé pour chacune des combinaisons de F1 testées d'après les données de la publication : (15156x1566 indique que les animaux testés sont des individus F1 issus du croisement entre les lignées 15156 et 1566). Les auteurs ont observé que neuf de ces combinaisons sont complètement résistantes, cinq présentent une faible mortalité, huit développent une infection létale, neuf succombent avec des lésions d'hépatite et 16 meurent avec des signes biologiques de coagulopathie, comme dans les formes humaines les plus sévères.

LES POPULATIONS GÉNÉTIQUES DE RÉFÉRENCE : UN OUTIL PARFAIT POUR LA BIOLOGIE DES SYSTÈMES

Les populations génétiques que nous avons décrites, lignées recombinantes consanguines et Collaborative Cross, sont des exemples parfaits de ce que les généticiens appellent des « populations génétiques de référence » (Schughart *et al.*, 2013). Il s'agit dans les deux cas d'une collection de lignées consanguines, donc dont la composition génétique est précisément connue et très stable, qu'il est possible de perpétuer et sur lesquels de nombreux laboratoires peuvent travailler pour accumuler leurs observations, les confronter, les combiner et rechercher des corrélations entre elles. Un de leurs atouts est que les différentes lignées d'une collection sont dérivées des mêmes lignées fondatrices et sont équidistantes sur le plan génétique, à l'inverse des cohortes humaines dans lesquelles certains individus sont génétiquement beaucoup plus proches que d'autres.

La multiplicité des données qui peuvent être collectées sur ces lignées permet de mettre en évidence des corrélations et

d'émettre des hypothèses sur les mécanismes à l'origine de ces corrélations. Comme le schématise la **figure 4A**, ces données concernent le génotype des lignées (plusieurs centaines de milliers de marqueurs génétiques voire la séquence complète), le niveau d'expression des gènes (qui peut être mesuré dans de multiples tissus, à différents stades du développement et dans différentes conditions expérimentales) et des phénotypes très variés mesurés sur les animaux. Ces données sont collectées par différents laboratoires qui ont accès aux mêmes lignées, et sont centralisées dans une base de données dotée d'outils d'analyse statistique, telle que le site GeneNetwork (<http://www.genenetwork.org>). Selon les types de données qui sont comparées, il est possible de répondre à une infinité de questions de nature très diverse (**figure 4B**), en particulier à des questions très ouvertes telles que : « quels sont tous les gènes dont le niveau d'expression dans le foie est corrélé positivement à celui du gène X ? ». Pour ce faire, on exploite les variations de niveau d'expression du gène X entre les lignées de la collection et on recherche d'autres gènes dont le niveau d'expression varie également entre les lignées, de façon corrélée avec le gène X. En multipliant de telles recherches, il devient possible

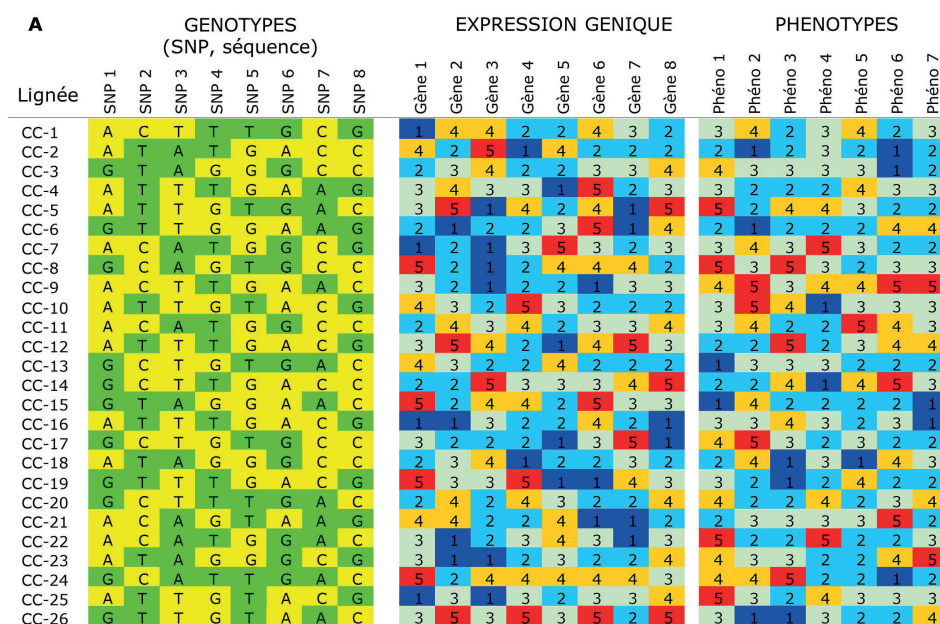


Figure 4 : Utilisation des populations génétiques de référence en biologie des systèmes.

A : Des données de nature variée sont produites dans une collection de lignées constituant une population génétique de référence. La composition génétique de chaque lignée est déterminée par le génotypage de marqueurs tels que des SNP ou par séquençage. Le niveau d'expression de plusieurs dizaines milliers de transcrits est déterminé dans plusieurs organes, à plusieurs stades de développement, dans différentes conditions expérimentales (les niveaux d'expression sont ici notés par intensité croissante par un chiffre de 1 à 5). Enfin, des phénotypes variés peuvent être mesurés chez des individus de ces lignées (avec la même convention, les chiffres de 1 à 5 représentent la valeur du phénotype pour chaque individu). Toutes ces données, obtenues dans les mêmes lignées, sont enregistrées dans une base de données telle que GeneNetwork (<http://www.genenetwork.org>).

B : Les outils statistiques de la base de données permettent de rechercher des corrélations entre ces données. Selon le type des données analysées, il est possible de répondre à des questions variées sans formuler d'hypothèse préalable puisque les corrélations sont recherchées simultanément à partir de toutes les données disponibles. Cette approche « pangénomique » caractérise la biologie des systèmes.

B

Données comparées	Question biologique
Génotype - Expression	Quels gènes contrôlent le niveau d'expression d'un gène ?
Expression - Expression	Quels gènes sont co-exprimés ?
Phénotype - Phénotype	Quels phénotypes sont corrélés ?
Expression - Phénotype	Quels gènes ont un niveau d'expression corrélé à l'intensité d'un phénotype ?
Génotype - Phénotype	Quels gènes contrôlent un phénotype ?



de construire des réseaux de régulation et d'émettre des hypothèses sur les mécanismes sous-jacents (Bystrykh *et al.* 2005). L'analyse simultanée et globale du génome, du transcriptome et du « phénomène » d'un groupe d'individus ou de lignées caractérise la biologie des systèmes qui étudie des réseaux complexes d'interactions pour tenter de décrire un modèle de fonctionnement de la totalité du système. Les populations génétiques de référence que nous avons décrites présentent des qualités remarquables pour servir cette approche.

Enfin, à l'heure où les études d'associations pangénomiques (*genome-wide association studies* ou GWAS), réalisées dans des cohortes de plusieurs milliers ou dizaines de milliers de sujets humains, fournissent des résultats décevants pour identifier les gènes de vulnérabilité aux maladies multifactorielles qui touchent un nombre croissant de personnes (Manolio *et al.* 2009), les populations génétiques de référence apparaissent comme une ressource nouvelle promise à un grand avenir. Analysées dans des conditions d'environnement contrôlées et permettant un large éventail d'investigations, elles fourniront

sans aucun doute des pistes nouvelles qui pourront être testées et validées sur des cohortes humaines de taille raisonnable.

CONCLUSION

Nous avons décrit plusieurs outils développés par les généticiens de la souris qui ont peu à peu permis d'explorer une diversité génétique croissante pour atteindre, voire dépasser celle des populations humaines. Ils montrent à quel point il faut reconsidérer les idées que l'on peut avoir sur les modèles animaux décrits et établis pour des maladies humaines, puisque leurs caractéristiques peuvent varier considérablement en fonction du fonds génétique. Partant de ce constat qui ne doit pas surprendre, il faut maintenant tirer parti de ces nouveaux outils pour mieux modéliser les diverses formes cliniques observées chez les patients et comprendre les mécanismes qui les contrôlent. De même, ces outils permettront de mieux anticiper, dès les études pré-cliniques, la diversité des réponses à une nouvelle molécule et donc, d'améliorer la prédictibilité pour l'homme des données obtenues dans des espèces modèles.

BIBLIOGRAPHIE

- Abitbol M, Bernex F, Puy H, Jouault H, Deybach JC, Guénet JL *et al.* A mouse model provides evidence that genetic background modulates anemia and liver injury in erythropoietic protoporphyria. *Am J Physiol Gastrointest Liver Physiol.* 2005; 288:G1208-16.
- Andreux PA, Williams EG, Koutnikova H, Houtkooper RH, Champy MF, Henry H *et al.* Systems genetics of metabolism: the use of the BXD murine reference panel for multiscalar integration of traits. *Cell* 2012; 150:1287-99.
- Aylor DL, Valdar W, Foulds-Mathes W, Buus RJ, Verdugo RA, Baric RS *et al.* Genetic analysis of complex traits in the emerging Collaborative Cross. *Genome Res.* 2011; 21:1213-22.
- Bailey DW. Recombinant-inbred strains. An aid to identify linkage and function of histocompatibility and other genes. *Transplantation* 1971; 11:325-27.
- Blanchet C, Jaubert J, Carniel E, Fayolle C, Milon G, Szatanik M *et al.* *Mus spretus* SEG/Pas mice resist virulent *Yersinia pestis*, under multigenic control. *Genes Immun.* 2011; 12:23-30.
- Bonhomme F, and Guénet JL. The wild house mouse and its relatives. In: Genetic variants and strains of the laboratory mouse. Lyon ML & Searle AG, editors, 2nd ed. Oxford: Oxford University Press; 1989, pp. 649-62.
- Bystrykh L, Weersing E, Dontje B, Sutton S, Pletcher MT, Wiltshire T *et al.* Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. *Nat Genet.* 2005; 37:225-32.
- Chesler EJ, Miller DR, Branstetter LR, Galloway LD, Jackson BL, Philip VM *et al.* The Collaborative Cross at Oak Ridge National Laboratory: developing a powerful resource for systems genetics. *Mamm Genome.* 2008; 19:382-9.
- Chia R, Achilli F, Festing MF and Fisher EM. The origins and uses of mouse outbred stocks. *Nat Genet.* 2005; 37:1181-6.
- Churchill GA, Airey DC, Allayee H, Angel JM, Attie AD, Beatty J *et al.* The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat Genet.* 2004; 36:1133-7.
- Collaborative Cross Consortium. The genome architecture of the Collaborative Cross mouse genetic reference population. *Genetics.* 2012; 190:389-401.
- Dejager L, Libert C, Montagutelli X. Thirty years of *Mus spretus*: a promising future. *Trends Genet.* 2009; 25:234-41.
- do Valle TZ, Billecocq A, Guillemot L, Alberts R, Gommel C, Geffers R *et al.* A new mouse model reveals a critical role for host innate immunity in resistance to Rift Valley fever. *J Immunol.* 2010; 185:6146-56.
- Durrant C, Tayem H, Yalcin B, Cleak J, Goodstadt L, de Villena FP *et al.* Collaborative Cross mice and their power to map host susceptibility to *Aspergillus fumigatus* infection. *Genome Res.* 2011; 21:1239-48.
- Ferris MT, Aylor DL, Bottomly D, Whitmore AC, Aicher LD, Bell TA *et al.* Modeling host genetic regulation of influenza pathogenesis in the collaborative cross. *PLoS Pathog.* 2013; 9:e1003196.
- Frazer KA, Eskin E, Kang HM, Bogue MA, Hinds DA, Beilharz EJ *et al.* A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. *Nature* 2007; 448:1050-5.
- Grisel JE, Belknap JK, O'Toole LA, Helms ML, Wenger CD, Crabbe JC. Quantitative trait loci affecting methamphetamine responses in BXD recombinant inbred mouse strains. *J Neurosci.* 1997; 17:745-54.
- Gross U, Frank M, Doss MO. Hepatic complications of erythropoietic protoporphyria. *Photodermatology Photoimmunology & Photomedicine.* 1998; 14:52-7.
- Guénet JL. The contribution of wild derived mouse inbred strains to gene mapping technology. *Curr Top Microbiol Immunol.* 1986; 127:109-13.
- Guénet JL, Benavides F, Panthier JJ, Montagutelli X. The different categories of genetically standardized populations of laboratory mice. In: *Genetics of the Mouse.* Berlin: Springer; 2015, pp. 319-59.
- Iraqi FA, Churchill G, Mott R. The Collaborative Cross, developing a resource for mammalian systems genetics: a status report of the Wellcome Trust cohort. *Mamm Genome.* 2008; 19:379-81.
- Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 2011; 477:289-94.
- Kola I, and Landis J. Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov.* 2004; 3:711-5.

- Leclercq K, and Kaminski RM. Genetic background of mice strongly influences treatment resistance in the 6 Hz seizure model. *Epilepsia* 2015; 56:310-8.
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ *et al.* Finding the missing heritability of complex diseases. *Nature* 2009; 461:747-53.
- Mashimo T, Lucas M, Simon-Chazottes D, Frenkiel M-P, Montagutelli X, Ceccaldi P-E *et al.* A nonsense mutation in the gene encoding 2'-5'-oligoadenylate synthetase/L1 isoform is associated with West Nile virus susceptibility in laboratory mice. *Proc Natl Acad Sci USA* 2002; 99:11311-6.
- Montagutelli X. Effect of the genetic background on the phenotype of mouse mutations. *J Am Soc Nephrol.* 2000; 11:S101-5.
- Montagutelli X & de Vienne D. Les populations expérimentales de cartographie génétique. *Med Sci (Paris)* 2008; 24:77-80.
- Morahan G, Balmer L and Monley D. Establishment of "The Gene Mine": a resource for rapid identification of complex trait genes. *Mamm Genome.* 2008; 19:390-3.
- Nadeau JH. Modifier genes in mice and humans. *Nat Rev Genet.* 2001; 2:165-74.
- Nakayama E, and Saijo M. Animal models for Ebola and Marburg virus infections. *Front Microbiol.* 2013; 4:267.
- Panthier JJ, and Montagutelli X. Le Collaborative Cross, un outil révolutionnaire à l'assaut des caractères complexes. *Med Sci (Paris)* 2012; 28:103-8.
- Peirce JL, Lu L, Gu J, Silver LM and Williams RW. A new set of BXD recombinant inbred lines from advanced intercross populations in mice. *BMC Genet.* 2004; 5:7.
- Rasmussen AL, Okumura A, Ferris MT, Green R, Feldmann F, Kelly SM *et al.* Host genetic diversity enables Ebola hemorrhagic fever pathogenesis and resistance. *Science* 2014; 346:987-91.
- Schughart K, Libert C and Kas MJ. Controlling complexity: the clinical relevance of mouse complex genetics. *Eur J Hum Genet.* 2013; 21:1191-6.
- Simon MM, Greenaway S, White JK, Fuchs H, Gailus-Durner V, Wells S *et al.* A comparative phenotypic and genomic analysis of C57BL/6J and C57BL/6N mouse strains. *Genome Biol.* 2013; 14:R82.
- Taylor BA. Recombinant inbred strains. In: Genetic Variation in the Laboratory Mouse. Lyon ML & Searle AG, editors, 2nd ed. Oxford:Oxford University Press; 1989, pp. 773-96.
- Threadgill DW, Hunter KW, Williams RW. Genetic dissection of complex and quantitative traits: from fantasy to reality via a community effort. *Mamm Genome.* 2002; 13:175-8.
- Threadgill DW, Miller DR, Churchill GA, de Villena FP. The collaborative cross: a recombinant inbred mouse population for the systems genetic era. *ILAR J.* 2011; 52:24-31.
- Tutois S, Montagutelli X, Da Silva V, Jouault H, Rouyer-Fessard P, Leroy-Viard K *et al.* Erythropoietic protoporphyria in the house mouse: a recessive inherited ferrochelatase deficiency (*fcd*) with anemia and liver disease. *J Clin Invest.* 1991; 88:1730-6.
- Yang H, Bell TA, Churchill GA and Pardo-Manuel de Villena F. On the subspecific origin of the laboratory mouse. *Nat Genet.* 2007; 39:1100-7.