

Le Collaborative Cross, un outil révolutionnaire à l'assaut des caractères complexes

Jean-Jacques Panthier, Xavier Montagutelli

► **To cite this version:**

Jean-Jacques Panthier, Xavier Montagutelli. Le Collaborative Cross, un outil révolutionnaire à l'assaut des caractères complexes. médecine/sciences, EDP Sciences, 2012, 28 (1), pp.103-8. 10.1051/medsci/2012281024 . pasteur-01325877

HAL Id: pasteur-01325877

<https://hal-pasteur.archives-ouvertes.fr/pasteur-01325877>

Submitted on 2 Jun 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Le Collaborative Cross, un outil révolutionnaire à l'assaut des caractères complexes.

The Collaborative Cross, a groundbreaking tool to tackle complex traits.

Jean-Jacques PANTHIER¹ et Xavier MONTAGUTELLI

Génétique fonctionnelle de la Souris, URA CNRS 2578.

Institut Pasteur, 25 rue du Docteur Roux, Paris 15, 75015, France

Résumé

Les caractères complexes, comme la vulnérabilité des personnes à une maladie commune, sont contrôlés par des régions génomiques nombreuses dont les effets phénotypiques sont généralement faibles. Partant de ce constat, des généticiens ont développé un nouveau système expérimental pour disséquer la génétique des caractères complexes chez la Souris. Le *Collaborative Cross* est une population génétique de référence comprenant plus de 300 lignées consanguines issues de huit lignées appartenant à trois sous-espèces de *Mus musculus* et ayant capturé 90% du polymorphisme connu chez la souris. Nous décrivons ici la fabrication et les propriétés du *Collaborative Cross* et rapportons les résultats des premières expériences effectuées sur des lignées en cours d'obtention.

Abstract

Complex traits, like the susceptibility to common diseases, are controlled by numerous genomic regions which individual effect is generally weak. These observations led geneticists to develop an experimental system to dissect the genetic of complex traits in the mouse. The Collaborative Cross (CC) is a genetic reference population of over 300 inbred lines derived from eight inbred strains of three *Mus musculus* sub-species, that captures 90% of the genetic variation known in the mouse genome. We present here the generation and the characteristics of the CC and we report the results of the first experiments with partially inbred CC lines.

¹ Tél. 01 45 68 85 55; Fax. 01 45 68 86 34; Courriels: panthier@pasteur.fr & xmonta@pasteur.fr.

Introduction

La majorité des maladies communes, telles que le diabète, l'asthme, les maladies autoimmunes, cardiovasculaires, neurologiques ou psychiatriques, résulte d'interactions fortes entre des facteurs génétiques multiples et l'environnement. Malgré cette complexité, l'élucidation des bases moléculaires de ces maladies semblait à portée de main. Cet espoir reposait en partie sur les succès remportés pour les maladies monogéniques, et sur les progrès technologiques permettant le génotypage de milliers d'individus dans des études d'association pangénomique (GWAS, *genome-wide association studies*). Plusieurs centaines d'associations entre des SNPs (*single nucleotide polymorphisms*) et des phénotypes morbides ont été décrites [1]. Les GWAS ont permis de se forger une idée de l'architecture génétique des caractères complexes de l'Homme. Le risque de développer une maladie commune est presque toujours contrôlé par une myriade de gènes aux effets très réduits², parfois quelques gènes ayant des effets moyens, quasiment jamais par des gènes ayant un effet important [2, 3]. Le contrôle par seulement cinq variants fréquents d'environ 50% du risque de développer une dégénérescence maculaire liée à l'âge semble bien être une exception [4]. De plus, les GWAS réalisées sur quelques milliers de cas et témoins n'ont généralement pas la puissance nécessaire pour identifier les gènes directement responsables d'un risque accru. Pour ces raisons, les bases génétiques des maladies communes demeurent pour l'essentiel inconnues. Le seul recours semble être actuellement d'augmenter la taille des cohortes analysées pour inclure plusieurs dizaines de milliers d'individus [5, 6].

Les données obtenues chez des espèces d'élevage et des animaux de laboratoire confirment ce modèle où l'héritabilité des caractères complexes est due à un nombre important de gènes ayant de faibles effets [7, 8]. En dépit d'une architecture génétique identique, l'étude des caractères complexes chez les rongeurs est rendue plus aisée par la possibilité de contrôler finement les variables environnementales et de réaliser des croisements prédéfinis. Mais l'atout majeur du généticien reste l'existence de lignées consanguines, des populations d'individus génétiquement identiques et homozygotes à tous les locus de leur génome [9]. La stratégie classique pour disséquer la composante génétique d'un caractère complexe chez la Souris est illustrée sur la Figure 1. Deux lignées consanguines, A et B, présentant un phénotype contrasté, sont croisées pour produire des hybrides de première génération (F1) qui portent un jeu de chromosomes A et un jeu de chromosomes B. Les F1 sont alors croisés entre eux pour produire des F2 qui sont phénotypés et génotypés avec des marqueurs répartis sur le génome. On recherche ensuite une association significative entre le génotype à une région génomique et une différence de phénotype. Si une région génomique (ou QTL pour *quantitative trait locus*) contrôle 5% de la

² L'effet d'une région génomique se mesure par la différence de risque, pour un individu, de développer la maladie selon qu'il porte l'un ou l'autre allèle présent dans la population.

variation phénotypique entre les lignées A et B, 1000 animaux F2 permettent de la localiser dans un intervalle d'environ 10 cM [10]. Pour confirmer la réalité de cette association, on produit des animaux dits congéniques (notés A.B-QTL^B) qui sont identiques aux souris A, sauf pour l'intervalle entourant le QTL qui provient lui de la souris B (QTL^B). Ces animaux sont produits en croisant une souris B et une souris A puis, pendant les 9 générations suivantes, une souris portant l'intervalle QTL^B et une souris A [11]. La comparaison du phénotype des souris A et A.B-QTL^B permet de mesurer l'effet du QTL^B dans le fonds génétique A. Les souris A.B-QTL^B peuvent ultérieurement être croisées de nouveau avec des souris A, pour fractionner l'intervalle chromosomique QTL^B en segments de plus petite taille. De proche en proche, l'intervalle QTL^B peut être réduit à environ 1 cM, ce qui correspond en moyenne à 2 Mb chez la souris, un intervalle physique contenant environ 20 gènes. Il reste alors à identifier parmi ces gènes celui qui contribue à la différence de phénotype observé entre les lignées A et B, ce qui peut être réalisé en étudiant la structure et l'expression des gènes de l'intervalle. Plus de 3.500 QTLs contrôlant des caractères variés, tels que le comportement, la réponse immune, le métabolisme, la sensibilité à des pathogènes et la prédisposition à des tumeurs, sont répertoriés dans la base de données *Mouse Genome Informatics* [12]. Toutefois, seule une trentaine de gènes responsables d'un QTL ont pu être identifiés. Ce faible pourcentage s'explique par les efforts et le temps nécessaires pour réaliser ce travail et, là encore, par la faiblesse des effets de ces QTLs. Soulignons également que la stratégie conventionnelle qui utilise les lignées consanguines de laboratoire n'explore qu'une fraction de la diversité génétique présente dans les populations naturelles de souris (Pour une revue sur la cartographie de QTL chez la souris voir [13]).

Pour surmonter ces difficultés, des généticiens de la souris réunis lors de la conférence de l'*International Mammalian Genome Society* en 2001, ont imaginé un nouveau système expérimental idéal dont les principales propriétés doivent être : (1) une très grande diversité génétique, pour augmenter les chances de découvrir des phénotypes contrastés; (2) une excellente résolution de localisation pour faciliter l'identification des gènes responsables des QTLs; (3) une population génétique homogène composée d'individus génétiquement équidistants, pour limiter le risque de découvrir de fausses associations; (4) des fréquences alléliques équilibrées (et une absence d'allèles rares) pour maximiser la puissance statistique; et (5) une parfaite connaissance du génome des individus de la population pour assurer une puissance maximale aux études d'association. Une population ayant toutes ces propriétés permettrait d'atteindre une puissance statistique et une résolution de localisation très élevées, avec un nombre d'individus beaucoup plus faible que celui couramment utilisé en GWAS. Des débats animés ont conduit à la décision de développer ce qui est devenu le *Collaborative Cross* (CC) [14-16]. Pour évaluer le potentiel du CC, des expériences ont été récemment effectuées sur

des lignées en cours de construction. Cet article présente le *CC* et les résultats et enseignements principaux de l'analyse préliminaire.

Le Collaborative Cross (CC)

Le *CC* combine les propriétés d'individus issus de populations naturelles, des lignées consanguines, et des lignées recombinantes consanguines. Mais commençons par le début. Le génome des lignées classiques de laboratoire est une mosaïque des génomes de trois sous-espèces du genre *Mus*: *Mus musculus domesticus*, *M. m. musculus* et *M. m. castaneus*, où la fraction d'origine *M. m. domesticus* prédomine. Elle représente ainsi 92% du génome de la lignée de référence C57BL/6J [17]. Pour augmenter la diversité génétique de la population, une première décision stratégique a été d'inclure des lignées consanguines provenant de chacune des trois sous-espèces. C'est ainsi que les trois lignées consanguines *M. m. castaneus* CAST/EiJ, *M. m. musculus* PWK/Ei et *M. m. domesticus* WSB/EiJ dérivées d'individus sauvages capturés en Thaïlande, en République Tchèque et dans le Maryland (USA), ont été recrutées pour participer au *CC*. Cinq lignées de laboratoire ont été adjointes : les lignées C57BL/6J et A/J, développées dès 1921 par deux des fondateurs de la Génétique, Clarence C. Little et Leonell C. Strong, la lignée 129S1/SvImJ qui a souvent été utilisée pour produire des souris *knockout*, ainsi que les lignées NOD/LtJ et NZO/H1LtJ qui développent un diabète. Le choix des huit lignées parentales étant fait, il existe plus de 40.000 (factorielle 8) façons de les croiser. Un schéma strict de croisement a été mis en place pour que les allèles des lignées fondatrices soient représentés de façon équilibrée dans la population finale. Pour chaque lignée à produire, la première étape a consisté à croiser les lignées deux à deux (selon les 56 combinaisons possibles) pour obtenir quatre hybrides de première génération (G1, Figure 2). Ces souris G1 furent croisées de 576 façons différentes pour produire des souris G2, qui croisées à leur tour donnèrent naissance aux souris G2:F1. Celles-ci renferment le maximum de la diversité génétique du croisement avec du matériel génétique de chacune des huit lignées parentales à part égale ($1/8 = 12,5\%$) et à l'état hétérozygote. Comment dériver à partir de ces individus G2:F1, tous uniques car différents, une population de référence, stable dans le temps, tout en conservant la fréquence (12,5%) des allèles parentaux ? La seconde décision stratégique a été d'accoupler des frères et sœurs G2:F1 pour produire de nouvelles lignées consanguines par des croisements consanguins ininterrompus sur 20 générations. Plus de la moitié de ces lignées ont cessé de se reproduire au fil des générations, vraisemblablement parce que certaines combinaisons entre des allèles de différentes sous-espèces du genre *Mus* ne sont pas viables pour l'embryon ou le souriceau, ou qu'elles sont associées à une infertilité dans l'un des deux sexes. Des calculs, des simulations et des données expérimentales ont montré qu'il se produit en moyenne 7 méioses informatives au cours des générations successives depuis le croisement entre les lignées parentales jusqu'au

moment où les allèles sont fixés à l'état homozygote. En effet, par les hasards du réassortiment des allèles, certains locus seront fixés à l'état homozygote dès la 2^{ème} ou 3^{ème} génération, alors que d'autres ne le seront que plus tard. Par conséquent, par rapport à une population backcross, on observe 7 fois plus de recombinaison entre deux marqueurs physiquement proches. Cette propriété explique l'excellente résolution de localisation des QTL apportée par le *CC*.

Ce travail considérable a été mené de front aux USA, d'abord au *Oak Ridge National Laboratory* puis à l'Université de Caroline du Nord (Chapel Hill) [18], en Israël, à l'Université de Tel-Aviv [19] et en Australie, à l'Université *Western Australia* [20]. A ce jour, 199 lignées indépendantes sont en cours de construction à Chapel Hill et 169 à Tel-Aviv. Les vingt premières lignées seront disponibles dans quelques semaines pour la recherche académique. Environ 300 lignées *CC* sont attendues pour la fin 2013. Elles seront toutes cryoconservées pour assurer leur pérennité.

Le génotypage de 184 lignées en cours de dérivation (*pré-CC*) pour plus de 620.000 SNPs a permis d'établir, pour chaque lignée, l'origine parentale de chaque région génomique. Il a montré que le génome d'une lignée comprend en moyenne 143 segments chromosomiques issus de l'une des 8 lignées parentales qui concourent, comme prévu, de façon équilibrée (12,5%) à l'ensemble des lignées [21]. Par ailleurs, la séquence du génome de ces 8 lignées parentales vient juste d'être rendue publique [22]. La séquence d'une région d'intérêt dans une lignée donnée pourra donc être déduite de son origine parentale. En terme de diversité génétique, les auteurs estiment que le *CC* aura capturé 89% des variants connus chez la Souris [23], ce qui présage une variabilité des phénotypes jamais atteinte dans un système expérimental chez les mammifères. En terme de puissance, des simulations prédisent que 1.000 individus provenant de 300 lignées *CC* permettront de cartographier un QTL responsable de 5% de la variation phénotypique entre les lignées parentales sur un intervalle de 1,7 cM seulement [24].

Le *CC* est une collection de lignées consanguines, donc d'individus génétiquement identiques et dont les caractéristiques phénotypiques sont stables dans le temps. Les données (phénotypes, expression génique) produites en utilisant ces lignées par des équipes localisées dans des pays différents utilisant des méthodes variées peuvent donc être intégrées et comparées [9]. Sur ce plan, le *CC* constitue une extension du concept de lignées recombinantes consanguines (RI, pour *recombinant inbred*). Les lignées RI sont produites à partir de seulement deux lignées consanguines de laboratoire. Elles présentent donc moins de variabilité génétique que les lignées du *CC*. Elles possèdent également une moindre résolution de localisation du fait que seulement 4 méioses informatives séparent en moyenne le croisement initial entre les deux lignées parentales et la mise à l'état homozygote. Enfin, les jeux existants comportent 15 à 30 lignées, tout au plus, à l'exception de celui qui combine les lignées C57BL/6J et DBA/2J pour lesquelles le nombre de lignées atteint maintenant la centaine. Dès 1978 Donald Bailey a démontré l'intérêt des lignées RI pour analyser un caractère complexe, l'histocompatibilité [25].

Des lignées RI ont été utilisées depuis plus de 20 ans pour étudier de nombreux caractères, tels que le comportement, le vieillissement et la longévité, des maladies autoimmunes, la sensibilité aux infections, ou des comportements addictifs à l'alcool ou aux drogues. Le génotype précis de chaque lignée a été déterminé pour plusieurs milliers de marqueurs. Des lignées RI ont été également utilisées pour analyser les réseaux de régulation d'expression génique par des approches de biologie des systèmes [26]. Des équipes ont établi, pour chacune de ces lignées, le niveau d'expression de tous les gènes dans différents tissus, dans différentes conditions. Toutes ces données sont stockées dans la base de données *GeneNetwork* de l'Université du Tennessee (Memphis, USA) [27] qui comporte des outils d'analyse très sophistiqués. Il est ainsi possible d'identifier *in silico* les régions génomiques qui contrôlent l'expression des gènes (eQTLs, *expression QTLs*) et d'identifier des corrélations d'expression, positives ou négatives, à l'échelle du transcriptome entier, entre gènes pour un même tissu, ou entre différents organes. Le CC offrira les mêmes possibilités d'investigation avec une variabilité, une puissance et une résolution bien supérieures encore.

Que nous dit l'analyse phénotypique et génétique des Pré-CC ?

Dans un premier temps la puissance des lignées du CC a été évaluée en cartographiant un caractère mendélien. Les souris WSB/EiJ sont caractérisées par une tache blanche sur le front, transmise sur un mode récessif, et d'étiologie inconnue. Sur 111 lignées *pré-CC*, six portent cette tache blanche. Une étude d'association a montré que ces six lignées partagent une région de 8,5 Mb sur le chromosome 10 d'origine WSB/Ei. Elle contient 52 gènes dont un seul est associé à des anomalies de la couleur du pelage, *Kitl*, qui code le ligand du récepteur à activité tyrosine kinase KIT. Plusieurs mutations au locus *Kitl* sont connues pour être responsables de taches blanches sur le front. Ainsi l'examen de seulement 111 souris a suffi pour cartographier une mutation dans un petit intervalle physique et identifier un excellent gène candidat [21].

Quid des caractères complexes ? Voyons deux exemples. Le premier concerne l'anxiété d'une souris placée dans un environnement ouvert. C'est un caractère complexe qui peut être évalué de façon quantitative après avoir placé délicatement la souris au centre d'une arène (*open field*) en calculant la distance moyenne entre l'animal et le centre de l'arène pendant dix minutes. Plus les souris sont anxieuses, plus elles s'éloignent du centre, donc se rapprochent des parois de l'arène. Ce test est rapide, facile à effectuer, et environ 450 souris mâles et femelles du *pré-CC* ont été phénotypées de cette façon. La distance était en moyenne égale à 5,3 cm mais elle variait considérablement, de 1,2 à 16,5 cm, selon les individus. Un intervalle de 3,6 Mb contenant 45 gènes sur le chromosome 6 est associé de façon significative aux variations de cette distance. Parmi les lignées parentales, les souris PWK/PhJ sont celles qui restent le plus au centre (4,3 cm), tandis que les NOD/LtJ s'en éloignent le plus (8,5 cm), mais les variations observées chez

les souris du *pré-CC* sont bien supérieures à celles qui existent entre les lignées fondatrices [28]. Le second exemple relève de la génétique de la sensibilité aux maladies infectieuses. L'Aspergillose invasive est la seconde cause de mortalité par infection fongique à l'hôpital. Elle affecte principalement des patients immunodéprimés. Soixante six lignées du *pré-CC* ont été infectées par l'agent responsable, *Aspergillus fumigatus*. Les souris ont présenté des sensibilités différentes. Certaines ont survécu à l'infection, d'autres sont mortes dans les 5 jours suivant l'inoculation. La durée de survie est associée de façon significative à 7 régions génomiques [29]. Dans la plupart des cas, ces QTLs sont dus principalement à des différences de résistance ou de vulnérabilité entre les trois sous-espèces de *Mus musculus*. Ceci souligne le bénéfice d'avoir inclus dans le *CC* des lignées présentant un niveau élevé de variation génétique. Les intervalles physiques correspondants, d'une taille comprise entre 6 et 17 Mb contenant de 39 à 139 gènes, peuvent sembler importants, mais cette taille doit être placée en regard du petit nombre d'individus utilisés (n=371) pour cartographier une durée de survie, c'est-à-dire un caractère qui intègre de nombreux mécanismes physiologiques et immunitaires. Des lignées du *pré-CC* ont également été analysées pour plusieurs dizaines d'autres caractères qui concernent le niveau d'expression des ARNm (*eQTLs*), la reproduction, le comportement, la morphologie, la physiologie et la composition de la microflore intestinale [30-32]. Ces études montrent que malgré les conséquences inévitables de la mise à l'état homozygote dans un environnement contrôlé bien différent du milieu naturel, il subsiste entre les lignées *CC* des phénotypes suffisamment contrastés pour permettre une cartographie fine en utilisant un échantillon d'animaux d'une taille raisonnable.

Ne nous leurrions pas : le *CC* ne résoudra pas de façon définitive toutes les difficultés d'identification des QTLs. D'autres ressources complémentaires du *CC* devront probablement être utilisées. Un stock non consanguin maintenu par panmixie entre descendants des huit lignées fondatrices du *CC* est en développement au *Jackson Laboratory* [33]. Il sera très utile pour améliorer la résolution d'une étude d'association réalisée sur le *CC* [8]. En Chine, des lignées consanguines ont été fabriquées qui portent 19 chromosomes identiques d'une lignée de laboratoire et 1 chromosome variant unique provenant de souris sauvages capturées dans la région de Shanghai [34]. Plus près de nous, des lignées recombinantes congéniques ont été fabriquées entre une lignée de laboratoire et des progéniteurs d'une autre espèce de souris, *Mus spretus* [35].

La découverte des gènes multiples associés à un caractère complexe, tel qu'une maladie commune, ouvre d'intéressantes perspectives. Il devient possible de construire des lignées congéniques portant un seul allèle « à risque » dans un fonds génétique résistant, et donc de convertir un système multigénique en une série de systèmes monogéniques (mendéliens). La composante phénotypique due à chacun des allèles à risque peut alors être analysée séparément

dans chacune des lignées congéniques [11]. Le succès est complet lorsque le phénotype complexe est réduit en autant de sous-phénotypes ou de mécanismes physiopathologiques qu'il existe de locus le contrôlant. L'exemple le plus achevé de cette stratégie concerne probablement la dissection génétique du lupus érythémateux disséminé [36]. Il est important de coupler cette démarche avec des approches fonctionnelles pour découvrir les gènes qui sous-tendent les allèles à risque. Cela ne nécessite pas de séquençage de gènes, mais une connaissance de la fonction individuelle des gènes contenus dans l'intervalle à tous les niveaux d'analyse, du profil d'expression du gène jusqu'à la physiologie.

Références

1. Hindorff LA, Junkins HA, Hall PN, *et al.* A Catalog of Published Genome-Wide Association Studies. Available at: www.genome.gov/gwastudies.
2. Manolio TA, Collins FS, Cox NJ, *et al.* Finding the missing heritability of complex diseases. *Nature* 2009; 461: 747-53.
3. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; 447: 661-78.
4. Maller J, George S, Purcell S, *et al.* Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nat Genet* 2006; 38: 1055-9.
5. Figueroa JD, Garcia-Closas M, Humphreys M, *et al.* Associations of common variants at 1p11.2 and 14q24.1 (RAD51L1) with breast cancer risk and heterogeneity by tumor subtype: Findings from the Breast Cancer Association Consortium. *Hum Mol Genet* 2011.
6. Prokopenko I, Langenberg C, Florez JC, *et al.* Variants in MTNR1B influence fasting glucose levels. *Nat Genet* 2009; 41: 77-81.
7. Hayes B, Goddard ME. The distribution of the effects of genes affecting quantitative traits in livestock. *Genet Sel Evol* 2001; 33: 209-29.
8. Valdar W, Solberg LC, Gauguier D, *et al.* Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nat Genet* 2006; 38: 879-87.
9. Montagnetelli X, de Vienne D. Les populations expérimentales de cartographie génétique. *Med Sci (Paris)* 2008; 24: 77-80.
10. Darvasi A, Soller M. A simple method to calculate resolving power and confidence interval of QTL map location. *Behav Genet* 1997; 27: 125-32.
11. Montagnetelli X, Abitbol M. Utilisation des lignées congéniques chez la souris. *Med Sci (Paris)* 2004; 20: 887-93.
12. <http://www.informatics.jax.org/>.
13. Flint J, Valdar W, Shifman S, Mott R. Strategies for mapping and cloning quantitative trait genes in rodents. *Nat Rev Genet* 2005; 6: 271-86.
14. Churchill GA, Airey DC, Allayee H, *et al.* The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat Genet* 2004; 36: 1133-7.
15. 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.
16. Threadgill DW, Miller DR, Churchill GA, de Villena FP. The collaborative cross: a recombinant inbred mouse population for the systems genetic era. *ILARJ* 2011; 52: 24-31.
17. Yang H, Bell TA, Churchill GA, Pardo-Manuel de Villena F. On the subspecific origin of the laboratory mouse. *Nat Genet* 2007; 39: 1100-7.
18. Chesler EJ, Miller DR, Branstetter LR, *et al.* The Collaborative Cross at Oak Ridge National Laboratory: developing a powerful resource for systems genetics. *Mamm Genome* 2008; 19: 382-9.
19. 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.
20. Morahan G, Balmer L, Monley D. Establishment of "The Gene Mine": a resource for rapid identification of complex trait genes. *Mamm Genome* 2008; 19: 390-3.
21. Aylor DL, Valdar W, Foulds-Mathes W, *et al.* Genetic analysis of complex traits in the emerging Collaborative Cross. *Genome Res* 2011; 21: 1213-22.
22. Keane TM, Goodstadt L, Danecek P, *et al.* Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 2011; 477: 289-94.
23. Roberts A, Pardo-Manuel de Villena F, Wang W, *et al.* The polymorphism architecture of mouse genetic resources elucidated using genome-wide resequencing data: implications for QTL discovery and systems genetics. *Mamm Genome* 2007; 18: 473-81.

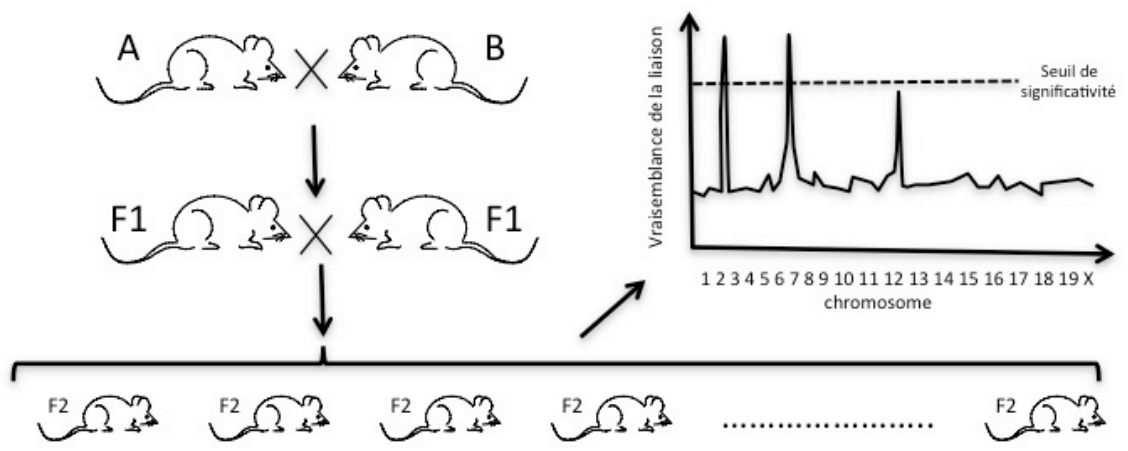
24. Valdar W, Flint J, Mott R. Simulating the collaborative cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. *Genetics* 2006; 172: 1783-97.
25. Bailey DW. Recombinant-inbred strains. An aid to finding identity, linkage, and function of histocompatibility and other genes. *Transplantation* 1971; 11: 325-7.
26. Bystrykh L, Weersing E, Dontje B, *et al.* Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. *Nat Genet* 2005; 37: 225-32.
27. <http://webqtl.org/webqtl/main.py>.
28. Philip VM, Sokoloff G, Ackert-Bicknell CL, *et al.* Genetic analysis in the Collaborative Cross breeding population. *Genome Res* 2011; 21: 1223-38.
29. Durrant C, Tayem H, Yalcin B, *et al.* Collaborative Cross mice and their power to map host susceptibility to *Aspergillus fumigatus* infection. *Genome Res* 2011; 21: 1239-48.
30. Kovacs A, Ben-Jacob N, Tayem H, *et al.* Genotype is a stronger determinant than sex of the mouse gut microbiota. *Microb Ecol* 2011; 61: 423-8.
31. Mathes WF, Aylor DL, Miller DR, *et al.* Architecture of energy balance traits in emerging lines of the Collaborative Cross. *Am J Physiol Endocrinol Metab* 2011; 300: E1124-34.
32. Zombeck JA, Deyoung EK, Brzezinska WJ, Rhodes JS. Selective breeding for increased home cage physical activity in collaborative cross and Hsd:ICR mice. *Behav Genet* 2011; 41: 571-82.
33. <http://jaxmice.jax.org/strain/009376.html>.
34. Xiao J, Liang Y, Li K, *et al.* A novel strategy for genetic dissection of complex traits: the population of specific chromosome substitution strains from laboratory and wild mice. *Mamm Genome* 2010; 21: 370-6.
35. Burgio G, Szatanik M, Guenet JL, *et al.* Interspecific recombinant congenic strains between C57BL/6 and mice of the *Mus spretus* species: a powerful tool to dissect genetic control of complex traits. *Genetics* 2007; 177: 2321-33.
36. Fairhurst AM, Wandstrat AE, Wakeland EK. Systemic lupus erythematosus: multiple immunological phenotypes in a complex genetic disease. *Adv Immunol* 2006; 92: 1-69.

Figure 1. Stratégie conventionnelle pour étudier la génétique des caractères complexes chez la Souris. Dans une première étape, les segments chromosomiques (QTLs) qui ségrègent avec le phénotype sont identifiés dans un croisement F2 (dans cet exemple des QTLs sont identifiés sur les chromosomes 2, 7 et 12). Dans une seconde étape, chacun des segments identifiés est transféré dans l'autre fonds génétique par 10 croisements en retour, ce qui permet de s'assurer des effets de chaque QTL avant d'analyser les gènes présents dans le segment.

Figure 2. Schéma de croisement ayant conduit aux lignées consanguines du CC. Dans chaque croisement, le femelle est à gauche et le mâle à droite. Les croisements réalisés en première génération déterminent l'origine parentale des chromosomes X et Y et de l'ADN mitochondrial dans la lignée produite (d'après [18]).

Figure 3. Exemple de la composition génétique d'une lignée *Pré-CC*. Elle comprend en moyenne 143 segments chromosomiques provenant chacun d'une des 8 lignées parentales. Certains segments ne sont pas encore fixés à l'état homozygote (d'après [21]).

1. Première étape: Cartographie des régions chromosomiques (QTLs) expliquant une partie de la variabilité du phénotype entre les lignées A et B



2. Deuxième étape: Construction de lignées congéniques

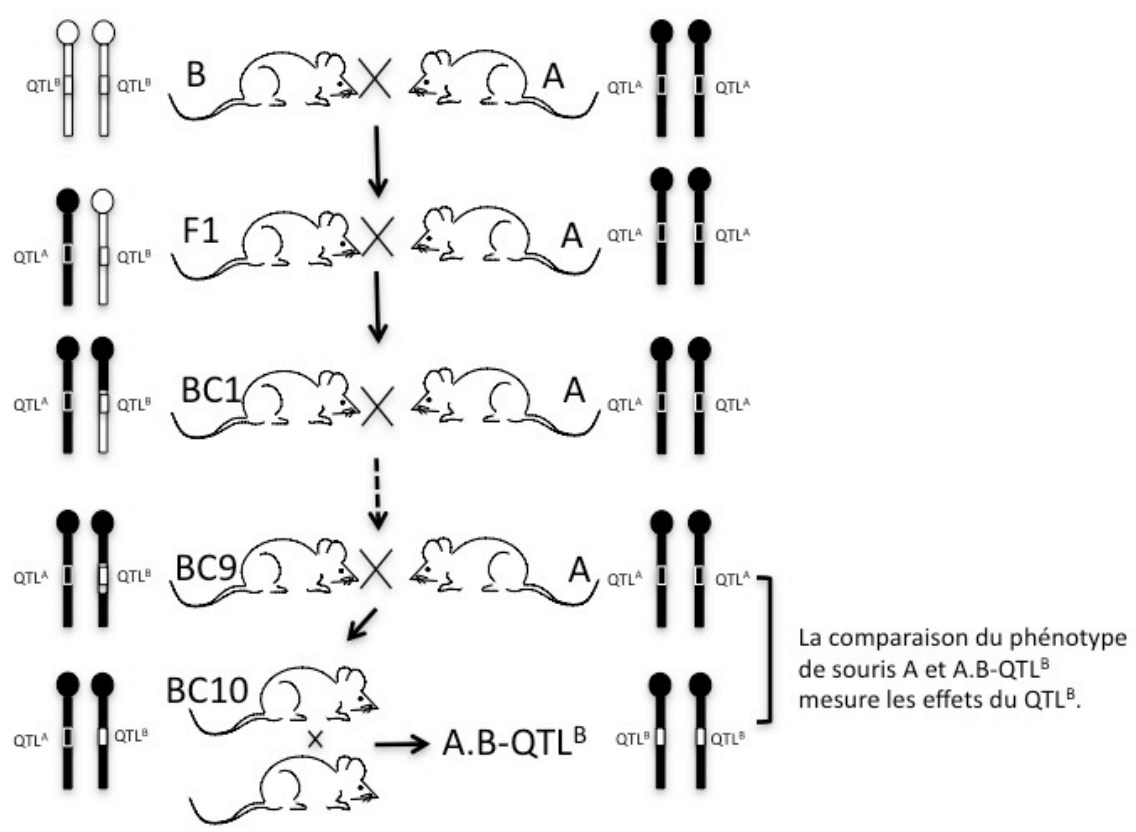


Schéma de croisement pour dériver la population génétique de référence à partir des 8 lignées fondatrices

