HAL
open science

Les populations expérimentales de cartographie génétique

Xavier Montagutelli, Dominique de Vienne

▶ To cite this version:

Xavier Montagutelli, Dominique de Vienne. Les populations expérimentales de cartographie génétique. Médecine/Sciences, 2008, 24 (1), pp.77-80. 10.1051/medsci/200824177 . pasteur-02552803

HAL Id: pasteur-02552803
https://pasteur.hal.science/pasteur-02552803v1

Submitted on 23 Apr 2020

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.

Les populations expérimentales de cartographie génétique

Xavier Montagutelli[1] et Dominique de Vienne[2]

[1] Unité postulante de Génétique fonctionnelle de la Souris, CNRS URA 2578, Institut Pasteur, 25 rue du Docteur Roux, 75724 Paris cedex 15.

[2] UMR de Génétique végétale, INRA−Univ Paris-Sud−CNRS−AgroParisTech, Ferme du Moulon, 91190 Gif-sur-Yvette.

A l'âge vénérable de près d'un siècle, la cartographie génétique reste une approche centrale en biologie, à laquelle on a recours dans des contextes variés, et avec des objectifs très différents : analyse de l'organisation des génomes (duplications, inversions, etc.), cartographie comparée (étude de la conservation de la synténie entre différentes espèces), cartographie de QTL (*Quantitative Trait Loci*) et de eQTL (QTL d'expression), génétique d'association, isolement de gènes (*positional cloning*), SAM (sélection assistée par marqueurs), etc. Les populations génétiques utilisées, dont le choix répond à des objectifs spécifiques, se sont diversifiées, surtout depuis que l'on dispose de marqueurs moléculaires en nombre quasiment illimité. Il n'est pas toujours facile de s'y retrouver parce que les croisements mis en œuvre sont variés et que, ces populations ayant été développées indépendamment chez les végétaux et les rongeurs de laboratoires, le vocabulaire associé diffère souvent entre les deux domaines. Cette note fait le point sur la terminologie des populations expérimentales les plus courantes, leur mode d'obtention, leurs particularités et leurs intérêts.

Nous considérerons les descendances dont le génome est fixé à l'état homozygote, et qui peuvent donc être perpétuées à l'identique par croisements frères-sœurs ou autofécondation. Comme elles sont issues de croisements entre des populations homozygotes à tous leurs locus (lignées *pures*, ou *fixées*, ou *consanguines*), il n'y a que deux allèles par locus.

Les populations expérimentales peuvent être classées selon :

- la région génomique étudiée : certaines populations sont développées dans le but d'analyser la fonction d'un gène ou d'une région chromosomique, alors que d'autres visent à étudier l'ensemble du génome ;

- le type de croisement ayant permis de les obtenir. On distingue ici les populations expérimentales selon qu'elles sont issues d'une descendance F2 ou d'un rétro-croisement. C'est ce second critère que nous avons préféré, et qui est utilisé dans le tableau et les figures.

D'une façon générale, l'intérêt de travailler sur des lignées génétiquement homogènes et non sur des populations en ségrégation est de pouvoir effectuer des mesures phénotypiques non sur des individus isolés, mais sur un groupe d'individus isogéniques, afin d'accroître la précision de ces mesures, de pondérer l'influence des facteurs d'environnement, d'évaluer la variance intra-groupe et de comparer

l'effet du sexe à génotype constant. Elles permettent également à différents laboratoires de travailler sur le même matériel génétique, pour comparer, corréler, voire compiler leurs résultats.

La plus petite variation génétique qu'il est possible d'étudier concerne un seul gène (ou toute séquence susceptible d'avoir un effet phénotypique). Elle peut avoir pour origine une mutation spontanée, une mutation produite par mutagenèse aléatoire (par un agent physique ou chimique) ou une modification induite par transgenèse ou par recombinaison homologue. Dans ce cas, la lignée obtenue ne diffère de la lignée consanguine de départ qu'au niveau de ce gène, et toute différence phénotypique observée peut donc être lui être attribuée. Ces lignées sont des modèles de choix pour démontrer le rôle d'un gène ou autre séquence dans un caractère, par exemple à l'issue d'une étape de cartographie génétique.

Des croisements frères-sœurs (ou des autofécondations) réalisés à partir d'une descendance F2 conduisent, à terme, à la production d'une série de lignées consanguines dites recombinantes (*RIS* ou *RIL*). Chacune porte, à l'état homozygote, une contribution égale des génomes des deux lignées parentales sous la forme de segments chromosomiques qui diffèrent d'une lignée à l'autre. Plus deux locus sont proches sur le même chromosome, plus ils tendent à être transmis dans un même segment. La capacité de séparer deux locus, qui est à la base de la cartographie génétique, croît avec le nombre de lignées recombinantes. Pour augmenter cette capacité résolutive, on peut réaliser plusieurs générations de croisements aléatoires à partir de la descendance F2, avant de commencer les croisements consanguins. On augmente d'autant le nombre de méioses informatives, donc la survenue d'événements de recombinaison qui réduisent la taille des segments. Cela constitue l'intérêt majeur des lignées (consanguines) hautement recombinantes (*AIL* ou *IRIL*). Ces différents types de populations permettent d'étudier l'ensemble du génome dans des populations possédant une contribution égale des deux lignées parentales.

Les croisements en retour (ou rétro-croisement, ou *backcross*) permettent de produire des lignées auxquelles les deux lignées parentales ont contribué de façon déséquilibrée. On distingue alors une lignée « receveuse », celle dont le génome se retrouve en proportion majoritaire et qui constitue le fonds génétique des lignées, et une lignée « donneuse ». Les lignées congéniques/quasi-isogéniques résultent de croisements en retour répétés avec la lignée receveuse, accompagnés, à chaque génération, de la sélection pour la génération suivante d'individus hétérozygotes à un locus (ou une région chromosomique) d'intérêt. A l'issue de ce processus, qui comprend généralement une dizaine de rétro-croisements, le génome de la lignée produite ne diffère de celui de la lignée receveuse que pour un segment chromosomique contenant le locus d'intérêt mais également quelques dizaines ou centaines de gènes qui lui sont génétiquement liés. L'utilisation de marqueurs génétiques permet de réduire le nombre de générations nécessaires (protocole dit "*speed congenic*") ou de fixer les bornes du segment chromosomique transféré. Ce protocole peut être utilisé pour produire une série de lignées (dites d'introgression) portant des segments chromosomiques partiellement chevauchants qui, de proche en proche, couvrent l'ensemble du génome. Lorsque chaque segment correspond à un chromosome entier et lui seul, on parle de lignée consomique ou lignée de substitution.

Enfin, des croisements frères-sœurs (ou des autofécondations) réalisés à partir d'une descendance issue d'un ou de plusieurs rétro-croisements permettent de produire des lignées recombinantes congéniques/recombinantes déséquilibrées. La contribution de la lignée donneuse dépend du nombre de générations de rétro-croisements qui précèdent la mise à l'état consanguin. Par comparaison avec les lignées recombinantes (consanguines), ces lignées permettent une meilleure dissection génétique des caractères à déterminisme complexe et offrent une plus grande puissance statistique pour mettre en

évidence des interactions épistatiques entre gènes du fait que les différences génétiques entre lignées concernent une plus petite fraction du génome.

L'ensemble de ces matériels génétiques représente une richesse inestimable pour la cartographie des gènes et l'étude des caractères complexes dans les espèces expérimentales. Nous espérons que cette note aidera généticiens « végétalistes » et « animalistes » à mieux se comprendre.

Pour en savoir plus

- *Ouvrages et chapitres d'ouvrages*

J.-J. Panthier, X. Montagutelli et J.-L. Guénet. 2003. Génétique de la souris. Belin, 288 pages.

D. de Vienne (Ed). 2003. Molecular markers in plant genetics and biotechnology. Science Publishers, Inc. Enfield (NH), USA, and Plymouth, UK. 235 pages.

B.A. Taylor. Recombinant inbred strains. *In* M.F. Lyon et A.G. Searle (Ed). 1989. Genetic Variation in the Laboratory Mouse, 2nd ed. Oxford University Press, Oxford, UK, pp 773-796.

- *Articles*

P. Demant et A. A. M. Hart. 1986 Recombinant congenic strains - a new tool for analyzing genetic traits determined by more than one gene. Immunogenetics 24: 416-422.

R.W. Williams, J. Gu, S. Qi et L. Lu. 2001. The genetic structure of recombinant inbred mice: high-resolution consensus maps for complex trait analysis. Genome Biol. 2: RESEARCH0046

E. Wakeland, L. Morel, K. Achey, M. Yui et J. Longmate. 1997. Speed congenics - a classic technique in the fast lane (relatively speaking). Immunology Today 18: 472-477.

R. Mott et J. Flint. 2002. Simultaneous detection and fine mapping of quantitative trait loci in mice using heterogeneous stocks. Genetics 160: 1609-1618.

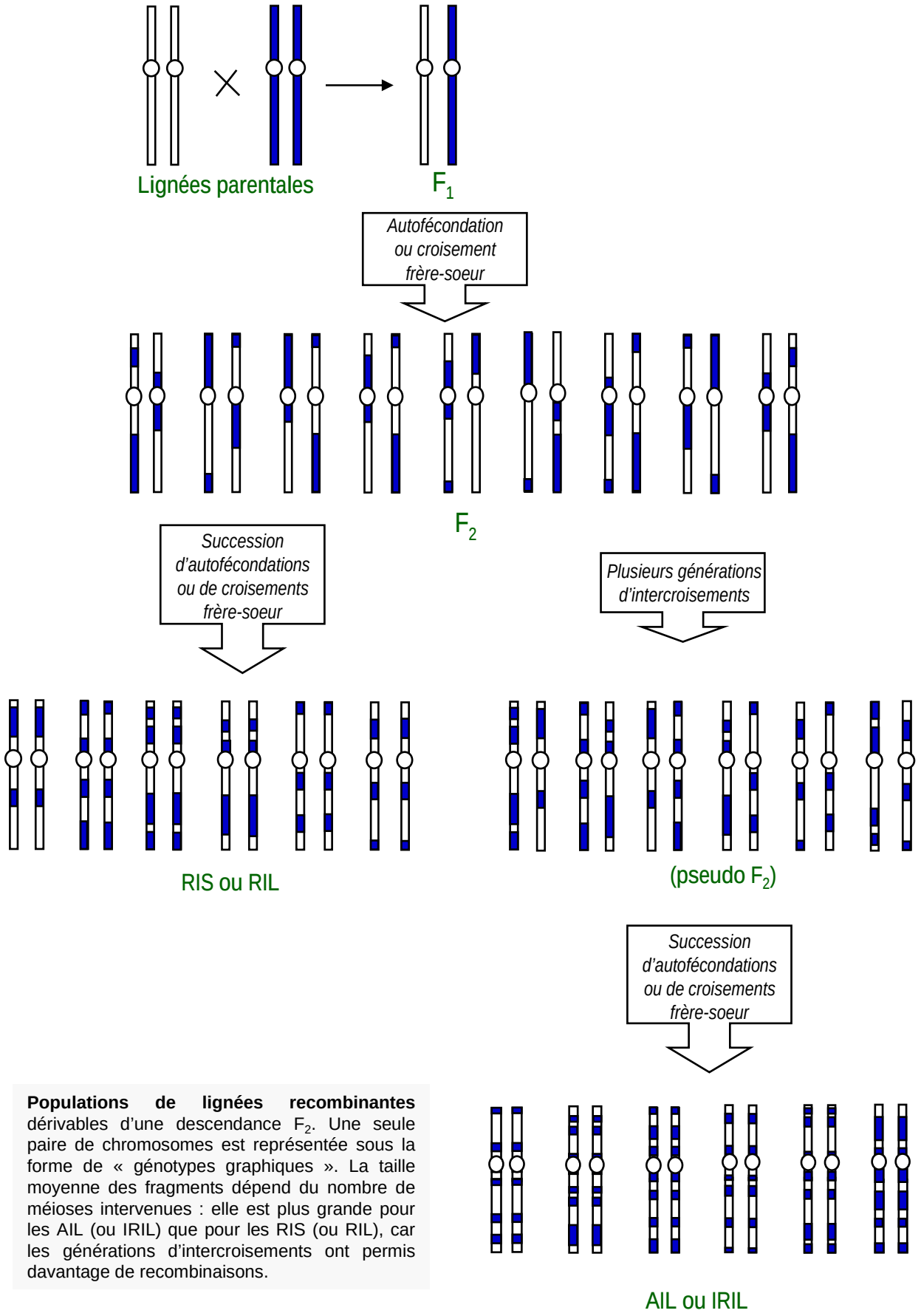
A. Darvasi et M. Soller. 1995. Advanced intercross lines, an experimental population for fine genetic mapping. Genetics 141: 1199-207.

M. R. Tuinstra, G. Ejeta et P. B. Goldsbrough. 1997. Heterogeneous inbred family (HIF) analysis: a method for developing near-isogenic lines that differ at quantitative trait loci. Theor. Appl. Genet. 95: 1005-1011.

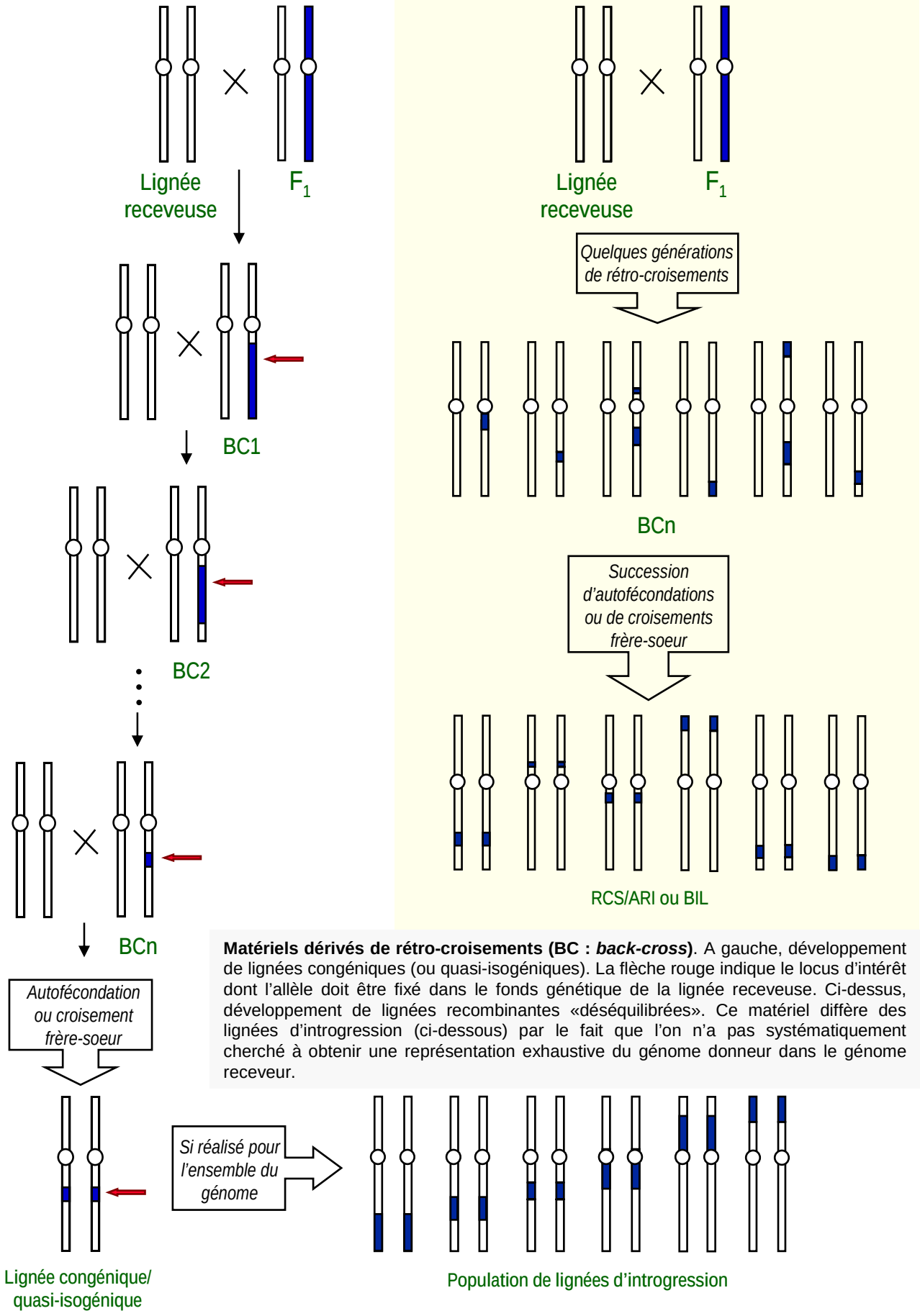
Y. Eshed et D. Zamir. 1995. An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield associated QTLs. Genetics, 141: 1147-1162.

Matériel génétique		Obtention	Propriétés
Nomenclature pour les animaux	**Nomenclature pour les végétaux**		
Lignée consanguine *Inbred strain*	Lignée, lignée pure, ou lignée fixée *Inbred line*	Obtenue par au moins 20 générations de croisements ininterrompus entre individus issus des mêmes parents (frères-sœurs). Chez les espèces végétales autogames (qui s'autofécondent naturellement), ou chez lesquelles on peut forcer l'autofécondation, 6 ou 7 générations sont suffisantes.	Les individus sont en principe tous identiques génétiquement (sauf pour les chromosomes sexuels), et homozygotes à 100% des locus. En pratique, on peut observer une hétérozygotie résiduelle qui décroît au fil des générations.
Lignée co-isogénique *Co-isogenic strain*		Lignée consanguine/pure portant une différence génétique à un seul locus, résultant soit d'une mutation spontanée, soit d'une modification intentionnelle du génome (mutagénèse aléatoire, transgenèse, recombinaison homologue).	La lignée co-isogénique ne diffèrant de la lignée consanguine d'origine que pour le locus considéré, toute modification phénotypique peut être attribuée à la modification génétique portée.
Matériels produits à partir d'une descendance F2			
Lignées recombinantes consanguines *Recombinant inbred strains (RIS)*	Lignées recombinantes *Recombinant inbred lines(RIL)*	Population de lignées consanguines produites par croisements frères-sœurs successifs ou par autofécondation, à partir d'individus F2 dérivés du croisement entre deux lignées parentales.	Le génome de chaque lignée de la collection est un réassortiment aléatoire des génomes des deux lignées parentales, en proportions statistiquement égales, fixé à l'état homozygote.
Lignées consanguines hautement recombinantes* *Advanced intercross lines (AIL)*	Lignées hautement recombinantes* *Intermated Recombinant Inbred Lines (IRIL)*	Les autofécondations, ou les croisements frères-sœurs, sont réalisés non pas à partir de la descendance F2, mais à partir d'une population obtenue en intercroisant aléatoirement, sur plusieurs générations, les individus de la descendance F2. *Nota: l'hétérozygotie résiduelle d'une région d'intérêt dans une population de lignées (hautement) recombinantes permet d'obtenir directement, par autofécondation, un couple de lignées quasi-isogéniques (définition ci-dessous). En anglais : Heterogeneous Inbred Family (HIF) analysis*	Comme précédemment, le génome de chaque lignée de la population est un réassortiment aléatoire des génomes des deux lignées parentales, mais les générations d'intercroisements ont réduit la taille moyenne des segments chromosomiques. La longueur estimée de la carte génétique est donc augmentée en relation avec le nombre de générations d'intercroisements.
Matériels produits à partir d'un rétro-croisement			
Lignée congénique *Congenic strain*	Lignée quasi-isogénique *Near isogenic line*	Obtenue en croisant plusieurs fois en retour (*back-cross*) une lignée donneuse avec une lignée receveuse, et en sélectionnant à chaque génération des individus ayant conservé à un locus ou marqueur d'intérêt l'allèle venant de la lignée donneuse.	La lignée congénique/quasi-isogénique diffère de la lignée receveuse pour un segment chromosomique portant le locus d'intérêt. Ce segment est en moyenne d'autant plus court que le nombre de générations de rétro-croisements est élevé.
	Population de lignées d'introgression *Introgression line population*	Population de lignées congéniques/quasi-isogéniques obtenues en utilisant un jeu de marqueurs qui couvre tout le génome.	Les lignées de la population portent des segments de chromosomes partiellement chevauchants qui couvrent de proche en proche l'ensemble du génome.
Lignée consomique *Consomic strains (Chromosome Substitution Strains)*	Lignées de substitution *Substitution lines*	Même principe que pour les lignées congéniques/quasi-isogéniques, mais en sélectionnant à chaque génération des individus ayant conservé un chromosome entier de la lignée donneuse.	La lignée receveuse et la lignée consomique/de substitution diffèrent pour un chromosome entier. Dans certains cas, ce travail a été fait pour chacun des chromosomes de l'espèce.
Lignées recombinantes congéniques *Recombinant congenic strains (RCS), ou Advanced recombinant inbred (ARI) strains*	Lignées recombinantes déséquilibrées* *Back-cross inbred lines (BIL)*	Population de lignées produites à partir d'individus issus d'au moins trois croisements en retour d'une lignée donneuse avec une lignée receveuse.	Le génome de chaque lignée de la population est un réassortiment aléatoire des génomes des deux lignées parentales, avec une contribution minoritaire de la lignée donneuse, qui dépend du nombre de rétro-croisements.

* Proposition des auteurs, en l'absence de nomenclature établie



Populations de lignées recombinantes dérivables d'une descendance F_2. Une seule paire de chromosomes est représentée sous la forme de « génotypes graphiques ». La taille moyenne des fragments dépend du nombre de méioses intervenues : elle est plus grande pour les AIL (ou IRIL) que pour les RIS (ou RIL), car les générations d'intercroisements ont permis davantage de recombinaisons.



Matériels dérivés de rétro-croisements (BC : *back-cross*). A gauche, développement de lignées congéniques (ou quasi-isogéniques). La flèche rouge indique le locus d'intérêt dont l'allèle doit être fixé dans le fonds génétique de la lignée receveuse. Ci-dessus, développement de lignées recombinantes «déséquilibrées». Ce matériel diffère des lignées d'introgression (ci-dessous) par le fait que l'on n'a pas systématiquement cherché à obtenir une représentation exhaustive du génome donneur dans le génome receveur.